VARIABILITY IN CROSS-DIALECTAL PRODUCTION AND PERCEPTION OF CONTRASTING PHONEMES: THE CASE OF THE ALVEOLAR-RETROFLEX CONTRAST IN BEIJING AND TAIWAN MANDARIN

BY

YUNG-HSIANG CHANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

    Associate Professor Chilin Shih, Chair
    Associate Professor Jont B. Allen
    Professor Jennifer S. Cole
    Professor José I. Hualde
    Assistant Professor Silke Hamann, University of Amsterdam

ABSTRACT

The alveolar-retroflex contrast is a critical feature in Mandarin and is often used to differentiate Beijing Mandarin from other dialects of Mandarin like Taiwan Mandarin. While a number of linguistic and sociolinguistic factors have been found to affect the alveolar-retroflex contrast, leading to variation in Taiwan Mandarin, a consistent alveolar-retroflex distinction is described for Beijing Mandarin in the literature on Mandarin phonology. With a series of map tasks, this dissertation examines whether the production of alveolar-retroflex contrast in both dialects is subject to the effects of vowel context and focal prominence. With a discrimination task and a goodness rating task, the categorical and gradient modes of alveolar-retroflex perception in different vowel contexts are investigated for listeners of both dialects. Results of the production study indicate that the acoustic characterization of Beijing vs. Taiwan Mandarin alveolar-retroflex contrast varies by vowel and by how each contrasting phoneme is realized in a particular vowel context. Focal prominence is found to result in longer syllable durations but not increased spectral distinctiveness between the alveolar and retroflex sibilants. The findings are discussed with respect to enhancement theory. The perception study found that Beijing and Taiwan listeners have different perceptual boundaries along the acoustic continuum, with a lower cutoff frication frequency required for the retroflex percepts for Beijing listeners. Listeners' alveolar-retroflex boundaries shift to lower frequencies in the rounded vowel context to normalize for vowel coarticulatory effects. Discrepant within-category sensitivity was found in that while both Beijing and Taiwan listeners perceive all retroflex variants as equally good, Beijing listeners consider the endpoint variant of the alveolar as the best category exemplar. The findings are discussed within the frameworks of quantal theory and exemplar theory as well as with respect to the hyperspace effect in perception. Together, the results show that linguistic (i.e.,

vowel context) and sociolinguistic (i.e., dialect) factors collectively and variably affect the

production and perception of the Mandarin alveolar-retroflex contrast.

Keywords: production variation, perception variation, dialectal variation, phonological contrasts, Mandarin alveolar-retroflex contrast, Beijing Mandarin, Taiwan Mandarin, map tasks, vowel context effect, prosodic strengthening, categorical perception, gradient perception

In memory of my grandfather

ACKNOWLEDGMENTS

Writing this dissertation was a humbling journey. This work could not have been carried to completion were it not for provision from God and support I received from family, teachers and friends over the years.

First and foremost, I would like to show my deepest appreciation for my thesis committee: Chilin Shih, Jennifer Cole, José Hualde, Silke Hamann and Jont Allen. Collectively they have had the greatest impact on the shape and content of this dissertation. My advisor, Chilin, revamped my view and experience of linguistics research. She sets the best role model for me on how linguistics research can and must be rooted in the scientific method. Chilin insists on guidance with just the right amount of input in every step of the entire dissertation so that I can be an independent researcher. The impact of her mentoring will permeate my future research and teaching career. Jennifer is the clearest professor I have ever learned from. Her lectures and talks lit me up every time, demonstrating to me how knowledge can be conveyed in such a structured and thorough manner. I am grateful for her keenness to fine details as well as the big picture of my work. José's teaching and research introduced me to a domain that was new to me yet is so closely related to phonetics and phonology—sound change. I thank him for broadening the scope of my work so as to connect to diachronic phonology, which this cross-dialectal phonetic study certainly ties into. I am thrilled to have Silke on my committee, for her dissertation inspired me to study Mandarin retroflexion. I thank her for all the insight and absolute support for this dissertation and my previous works. Last but not least, Jont came in rescue when I was constructing the auditory stimuli for my perception study. He introduced me to a whole new realm of speech signal processing and manipulation, and I can't thank him enough for generously letting me use *Beren* that he developed with his students. I thank all my committee

Finally, I cannot overemphasize just how important my family has been in helping me get to this point of my graduate student journey. My mom and dad's loving patience and faith in me keeps me going. I thank God for placing Ai-ling Liao, my fiancée, in the last year of my graduate student life. Her unceasing prayers and faithful company during my roller coaster ride of the dissertating write-up, for which I can only commit my life to her in return. At last, grandpa, this dissertation is for you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background and purpose

Variation exists at multiple levels in human speech. At the segmental level, numerous factors, particularly phonetic and social factors, have been identified to affect the realizations of a phonological contrast. Moreover, many of these factors may interact to affect the production or perception of such a contrast. Regardless of considerable variation in speech production, listeners are generally able to either reduce or normalize the variation and arrive at sound categorization successfully. However, perceptual variability can arise as listeners weigh available cues differently or remap the same contrasting sounds from the perspective of a language/dialect that differs from their native phonological system.

Motivated by the variability observed in phonetic realizations and the perception of contrasting phonemes, this dissertation studies a critical place feature in Mandarin, the alveolar-retroflex contrast. The dissertation investigates how this contrast is realized acoustically and perceived under various phonetic conditions by speakers of different dialectal backgrounds (Beijing Mandarin vs. Taiwan Mandarin). The production study uses a series of map tasks to elicit natural yet controlled speech data. To explore possible production variation, two factors are controlled with the stimuli embedded in the map tasks: vowel context and contrastive focus condition. The perception study uses modified speech covering an acoustic continuum between the alveolar and retroflex to investigate the location of category boundaries and best category exemplars in different vowel contexts and across dialect.

The study of variability of contrasting sounds in production and perception, such as the Mandarin alveolar-retroflex contrast, holds implications for both phonological theory and

language acquisition. Investigations of phonetic and perceptual variation of the same phonological contrast between different dialects shed light on diachronic sound change. In this way, this study is significant not only for our understanding of the nature of a phonological categorization, but also for our understanding of variation in speech in general.

1.2 Variation in speech production and perception

Varied realizations of phonological contrasts introduced by phonetic contexts have been widely identified. One of the most well studied phonological contrasts is the voicing contrast[1] in English. Voicing, specified as voice onset time (VOT; the time lag between the release of a consonant closure and the start of vocal fold vibration at the vowel onset), is realized very differently across contexts (Kingston & Diehl 1994). Its realizations can vary by places of articulation (Lisker & Abramson 1964, 1967), the position of a stop within a domain (Umeda 1977; Keating 1984), and as a function of the duration of the preceding vowel (Luce & Charles-Luce 1985). The voicing contrast also adjusts to changes in prosodic prominence like stress and focus, as longer VOT was reported in lexically stressed syllables (Lisker & Abramson 1964, 1967) and under accentuation (Choi 2003; Cole et al. 2007). In addition to phonetic context-induced variation, English voicing contrast is also subject to dialectal variation. For instance, compared to Lisker and Abramson's findings for American English, Docherty (1992) found different VOT durations in British English. Lastly, other languages that distinguish between

---

[1] In English, the voicing feature distinguishes voiceless aspirated stops [pʰ, tʰ, kʰ] from voiced unaspirated stops /b, d, g/, which are generally realized as voiceless unaspirated [p, t, k] in word-initial position. Therefore, in word-initial position, voicing contrast in English is more like a contrast of aspiration. The two categories differ in both fundamental frequency (F0) at the onset of voicing and voice onset time (VOT). However, separate perceptual effect of F0 has been found to be small and dependent on VOT (Lisker & Abramson 1985). Therefore, in characterizing voicing and creating acoustic continua for voicing perception in English, VOT is the most widely used acoustic parameter.

voiced and voiceless categories can have different temporal characteristics from the voicing contrast in English (Cho & McQueen 2005; Shih & Möbius 1998; Shih et al. 1999).

Given the highly variable acoustic cues that mark phonemic distinctions, researchers conducting perception studies attempt to find context-invariant cues to account for successful perception (e.g., Lindblom 1996) or to investigate how phonetic variation is processed. For the latter, it has been found that varied realizations of a phonological contrast introduced by adjacent phonetic context can be compensated for in perception, as listeners were found to change the location of perceptual category boundaries accordingly (Mann & Repp 1980, 1981). Variability as a result of dialectal differences can also be normalized in perception, as listeners adjust their categorization decisions to accommodate different accents (Evans & Iverson 2004). Contrasting with the original strong claim of categorical perception (see §4.3 for more discussion) that there is no fine-grained differentiation within categories, a gradient of perceptual sensitivity to variation in speech signals has been observed with various measures. Listeners have been found to show sensitivity to fine-grained within-category differences and retain in the perceptual system some continuous details in the speech signal (e.g., Goldinger 1998; McMurray et al. 2002), as reflected in measures like goodness rating (Allen & Miller 2001; Miller 1997) and reaction time (Pisoni & Tash 1974). Goodness rating tasks report more fine-grained judgments of the continuous speech signals and can more directly measure listeners' perceptual experience (Massaro & Cohen 1983). Difference in perception reaction time would suggest that not all variants of a sound category are processed equivalently. Together, these measures, as opposed to binary identification responses used to support the strong claim of categorical perception, demonstrate that continuous changes along a given acoustic dimension result in non-linear mapping in the perceptual space.

In light of various effects reported on the production of phonological contrasts, this dissertation continues the pursuit of variation in speech with the Mandarin alveolar-retroflex contrast. The perception of Mandarin alveolar-retroflex contrast will be investigated through the elicitation of judgments on both discrete and continuous scales, permitting this dissertation to contribute to the discussion of categorical and continuous views of speech perception.

A sizable body of work has been dedicated to cross-linguistic speech production and perception with one source of speech variation being differences across speech communities. The production and processing of speech input by non-native listeners has been widely studied with special attention to the perception of a phonological contrast that is absent in one's native language (e.g., Japanese listeners' distinction of English /r-l/ studied by Bradlow et al. 1997 and Cutler et al. 2006, among others). In contrast, little work has been devoted to the investigation of variation in cross-dialectal production and perception. In studying phonological differences between regional dialects of American English, for example, most researchers focused on constructing phonological atlases (Labov et al. 2006) or describing dialectal phonology (e.g., Kurath 1937). Some of the available studies (e.g., Docherty & Foulkes 1999; Thomas 2001) explored the vowel systems of different dialects of English. The acoustic measurements reported in these studies provide a sense of the degree of variation across dialects.

While speech processing by listeners with a different dialect of the same language has received some attention, these perception studies mostly focus on testing how well listeners can identify dialectal pronunciations (e.g., Labov et al. 1972; Flanigan & Norris 2000; Sumner & Samuel 2009) and how stereotyping can shift one's perceptual boundary (e.g., Janson & Schulman 1983; Niedzielski 1999). For the former line of research, the major perceptual variability is that listeners' exposure and experience with a different dialect affects their ability to

accommodate regional variants (Sumner & Samuel 2009). The latter line of research illuminates perceptual variability in that speech perception may not depend solely on acoustic factors, but also on listeners' expectations based on social factors. Relatively few experimental studies investigate variability in sound categorization by listeners from different dialect regions. Two of the relevant works reported that listeners of various dialectal backgrounds differ in their use of acoustic cues in perceiving contrasting sounds (Willis 1972; Peeters 1991). With the case of Mandarin alveolar-retroflex contrast, this dissertation investigates whether a dialect-specific pattern is present in categorization and whether linguistic (i.e., vowel context) and sociolinguistic (i.e., dialect) factors collectively contribute to variation in speech perception.

1.3 Overview of the dissertation

In this dissertation, variability in the production and perception of phonological contrasts is illustrated with the case of the Mandarin alveolar-retroflex contrast, with a focus on cross-dialectal comparison and characterization. The dissertation is divided into six chapters. Following the introduction on variability in speech production and perception (Chapter 1), Chapter 2 reviews the phonetic and phonological properties of Mandarin alveolar and retroflex consonants. In addition, the topics of alveolar-retroflex neutralization and social indexing of the place contrast are discussed. Findings and generalizations on Mandarin dialects and their effects on production and perception are highlighted. Chapter 3 reviews factors that affect the acoustic realizations of the Mandarin alveolar-retroflex contrast and addresses how these factors are controlled in the production experiment. A series of map tasks are introduced to elicit natural yet controlled speech data; the experimental considerations and design of the map task are addressed. Detailed acoustic analysis of the production data is presented alongside discussion on how production patterns vary across different experimental conditions. Chapter 4 reviews the

categorical and continuous aspects of speech perception in preparation for the study of the effects

of dialect difference and vowel context differences on perception. The acoustic cues most

relevant to the discrimination between the alveolar and retroflex consonants are discussed for the

subsequent stimulus construction for the perception study. The chapter then reports findings

from the two perception experiments and ends with discussion on variability in perceptual

patterns. Chapter 5 synthesizes the results of Chapters 3 and 4, while making comparisons to

findings from previous studies. Unresolved issues regarding the patterns observed in the results

of the perception experiment are addressed with a follow-up experiment. Chapter 6 concludes

with the main findings from the production and perception studies and proposes further research.

CHAPTER 2: MANDARIN ALVEOLAR AND RETROFLEX SIBILANTS

Variability observed in speech production and perception contextualizes the current study on variability in the Mandarin alveolar-retroflex contrast. This chapter will first introduce the phonetic and phonological properties of Mandarin alveolar and retroflex consonants. Variation as a result of dialectal differences is addressed throughout the chapter. Finally, the topics of alveolar-retroflex neutralization and social indexing of the place contrast in relation to sound change will be discussed.

2.1 Mandarin syllable structure and sound inventory

The basic structure of a Mandarin syllable can be illustrated as below (Figure 2.1). The onset, glide, nucleus and coda are the building blocks of a Mandarin syllable. Every Mandarin syllable must contain a vowel as the nucleus, whereas all other segments are optional.

Syllable
Onset   Final
Medial (Glide)   Rime
Nucleus   Coda
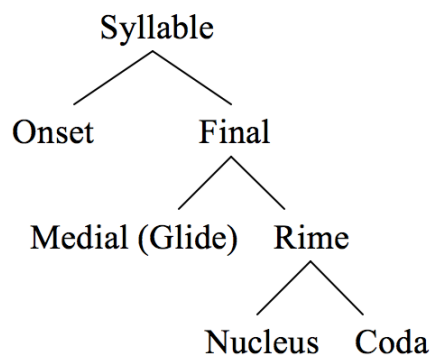
Figure 2.1: Mandarin syllable template

Mandarin Chinese has five monothong vowels /i, y, u, ə, a/ (Chao 1948, Cheng 1973, Lin 1989, Duanmu 2000). The exact number of phonetic vowel categories, however, can be up to 11 (Shih 1995) or 13 (Wu 2011). Mandarin vowels can be classified into three articulatory dimensions: front/back, high/low and rounded/unrounded (see Table 2.1).

	front	central	back
high	i y	ɿ	ʅ u
mid	e ɛ	ə ɚ	o ɔ
low	a		ɑ

Table 2.1: Mandarin vowels

There are three high vowels /i, u, y/. The /i/ vowel surfaces as the central apical vowel [ɿ] after the alveolar fricatives and affricates. The other variant of /i/ is [ʅ], which is also termed the retroflex vowel by Karlgren (1949) and only appears after retroflex consonants. These two allophones [ɿ, ʅ] are traditionally called the apical vowels and are considered the voiced extensions of the preceding consonants by Chao (1948) and Cheng (1973).

The mid vowel /ə/ can have as many as five variants [o, ɔ, ɛ, e, ə]. Some of the variants are considered indistinguishable or dialect-region specific and therefore are not distinguished (Duanmu 2000). But generally, the surface variants arise as a result of /ə/ assimilating in frontness or rounding with the adjacent consonant/glide (Duanmu 2000). Another mid vowel has been suggested to exist in Mandarin—the retroflex (rhotacized) vowel /ɚ/ (Duanmu 2000). It is very limited in distribution and is sometimes alternatively transcribed as /r/.

Different surface values have been proposed for the low vowel /a/ as well—[A[2], ɑ, a, æ, ɐ], according to Xu (1980). However, the most commonly agreed upon surface forms of /a/ are [ɑ, a], as they are acoustically distinct especially in terms of F2 (Shih 1995; Wu 2011). The [ɑ] variant occurs in open syllables and before the velar nasal /ŋ/, whereas [a] occurs only before the alveolar nasal /n/ (Cheng 1973).

There are 19 phonemic consonants in Mandarin (see Table 2.2). Aspiration is a distinctive feature in Mandarin as there are several pairs of aspirated and unaspirated stops and affricates.

[2] [A] is a central low vowel similar to [ɐ]

place / manner	bilabial	labiodental	alveolar	alveopalatal	retroflex	palatal	velar
stop	p pʰ		t tʰ				k kʰ
fricative		f	s	(ɕ)	ʂ		x
affricate			ts tsʰ	(tɕ) (tɕʰ)	tʂ tʂʰ		
nasal		m	l n				ŋ
approximant	(w) (ɥ)		ɹ			(j)	

Table 2.2: Mandarin consonants (allophones are in parentheses)

Phonetically, there is a three-way coronal contrast in Mandarin: the alveolars /s, ts, tsʰ/, the retroflexes /ʂ, tʂ, tʂʰ/ and the palatals [ɕ, tɕ, tɕʰ]. The palatal sibilants are in complementary distribution with the alveolars and retroflexes. The palatals can only be followed by [i] and /y/ and their homorganic glides [j, ɥ], whereas the alveolar and retroflex sibilants can be followed by /a, u, ə/, apical vowels, and the [w] glide.

The alveolar sibilants[3] have been variously termed as dentals (Chao 1968), alveolars (Kratochvil 1968; Ladefoged & Maddieson 1996; Luo & Wang 1981) and denti-alveolars (Lee & Zee 2003) due to impressionistic or articulatory observations of Mandarin speakers from different regions. In this study, the term "alveolar" will be used to refer to consonants whose constriction is made on the lower/upper teeth or in the alveolar region. The transcriptions of Mandarin retroflexes are discordant in the literature. They have been considered to be laminal post-alveolars /ʂ, tʂ, tʂʰ/ (Ladefoged & Maddieson 1996), or as apical post-alveolars /ʃ, tʃ, tʃʰ/ (Lee & Zee 2003) or just retroflexes /ʂ, tʂ, tʂʰ/ (Duanmu 2000). In this study, this series of sounds will be called the retroflex sounds in light of a broader definition of retroflexes in that the term "retroflex" should be used to refer to an articulatory gesture instead of a value for the feature "articulatory place" (Hamann & Fuchs 2010).

---

[3] In this dissertation, the term "sibilants" is used to refer to both fricatives and affricates.

The phonological status of Mandarin palatals has received a lot of attention because they also appear to be in complementary distribution with the velar series /k, kʰ, x/, as velars can only occur before non-high-front vowels and glides in Mandarin. Chao (1934) observed some palatal-velar alternation from some word game data and argued that the palatals should be underlying velars. In Baxter's (1992) analysis, Mandarin palatals have undergone velar palatalization, a common gradual articulatory change (Bloomfield 1933) or perceptually-motivated change (Ohala 1983) across languages in the world. However, Wang (1980) argued that some instances of palatals should be identified with the alveolars, while others with the velars based on historical sources. Wan (2010) conducted some perceptual experiments (similarity rating and sound contraction/extraction tasks) to investigate the psychological status of Mandarin palatals. She found that native Mandarin listeners prefer to replace palatals with alveolars in online processing during phonological experiments. Wan concluded that the palatals are perceptually more similar to the alveolar sibilants. A recent perception study conducted by Lu (2011), however, reported that the perception of the alveolar and palatal is similar to the perception of two contrastive consonants. Her discrimination data indicated perception of two separate phonemes along a synthetic [s-ɕ] continuum, and her similarity rating data showed phonemic-like judgment on the [s-ɕ] pairs, as the ratings were similar to those of the /s-f/ pairs (/s/ and /f/ are separate phonemes in Mandarin). She suggested that the relationship between Mandarin alveolars and palatals may be more phonemic than allophonic[4]. Taken together, the phonological status of Mandarin palatal series is still under debate and in most literature, palatals are still not considered phonemes in Mandarin.

---

[4] It is worth noting a trend of producing a more anterior palatal series or replacing the palatal series with /s, ts, tsʰ/ in the speech of young females in Beijing—a trend commonly known as Nüguoyin ("women's speech in Beijing"). This phenomenon has long been reported in the literature Li (1923). According to Hu (1991), young females considered the [si, tsi, tsʰi] pronunciations more feminine and pleasant as opposed to [ɕi, tɕi, tɕʰi].

Mandarin has three glides [j, w, ɥ]. They are phonologically analyzed as homorganic glides of the /i, u, y/ vowels respectively. The glides occur when they precede another vowel or diphthong, whereas the corresponding vowels occur in syllable nuclear position.

Of the three nasal sounds in Mandarin, only /m, n/ can occur in syllable-initial position, and only /n/ and /ŋ/ can occur in coda position. The approximants /l, ɹ/ both serve as syllable onsets in Mandarin. The retroflex liquid /ɹ/ is also transcribed as /ʐ/ (Dong 1958) as the voiced version of /ʂ/. The transcription discrepancy suggests that some speakers produce the sound as an approximant and others a fricative. Nevertheless, this manner difference between the two variants is not phonemic in Mandarin.

Mandarin is a tonal language and uses pitch to distinguish lexical meanings. The four lexical tones in Mandarin are: high level (Tone 1), high rising (Tone 2), low falling-rising (Tone 3) and high falling (Tone 4) (Chao 1930). Figure 2.2 shows the four tones produced in citation form where duration is normalized. In addition to the four lexical tones, there is a neutral tone in Mandarin that does not have fixed pitch level or contour and only occurs in a weak syllable (Chao 1968). The neutral tone never occurs in word-initial position and must be preceded by at least another non-neutral tone syllable. Mandarin has tone sandhi rules (Shih 1986), where tonal categories change in certain tonal contexts. Contextual tonal variation is subject to both anticipatory and carryover effects (Shih 1988; Xu 1993).
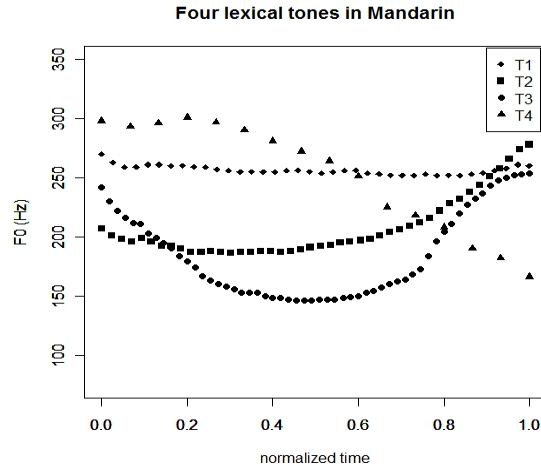
Figure 2.2: Four lexical tones in Mandarin

Mandarin has only around 400 syllables (excluding tone variations) but 3500 frequently used characters (Da 1998). In this regard, Mandarin has very high homophone density, which in principle should give rise to great ambiguity in speech. Nevertheless, given that 73% of Chinese words are multi-syllabic compounds (He & Li 1987)—of which the majority are disyllabic compounds—most ambiguities in speech can be resolved by using context.

2.2 Characteristics of Beijing Mandarin and Taiwan Mandarin

The Chinese dialects are divided into the Mandarin group and non-Mandarin group (which includes Gan, Hakka, Min, Wu, Yue and Xiang) in Ramsey (1987). This dissertation studies two regional pronunciations of Mandarin alveolar and retroflex sibilants, therefore, the word "dialect" will be used to refer to dialects of Mandarin unless otherwise noted. Standard Mandarin is based on the pronunciation of the Beijing dialect, and is the official language spoken in China and Taiwan. See Figure 2.3 for the geographical relationship of China and Taiwan (retrieved from http://mapsof.net/uploads/static-maps/map_of_east_asia.png on August 1, 2012).

Figure 2.3: Geographical relationship between China and Taiwan

In China, Mandarin is more often called *Putonghua*, namely 'common speech'. In Taiwan, Mandarin is referred to as *Guoyu* 'national language'. Although Mandarin is formally taught at school and used in the media, Mandarin varieties spoken in China and Taiwan are not uniform. Beijing Mandarin and Taiwan Mandarin[5] are two major dialects of Mandarin that are different in many ways.

First, a major characteristic of Beijing Mandarin is the extensive use of the diminutive [ɚ] suffix, or r-suffixation. For example, the pronunciation for 'flower' is [xwaɚ] in Beijing Mandarin but [xwa] in Taiwan Mandarin. It should be noted that r-suffixation is not a regular morpho-phonological rule (Li & Thompson 1981), such that not all syllables can take the r-suffix.

---

[5] Beijing Mandarin in this dissertation specifically refers to the dialect spoken by people from the greater Beijing area, instead of *Putonghua*. However, the term used here does not exclusively require speakers to exhibit local features that are unique to the Beijing dialect. Taiwan Mandarin here refers to the standard accent of Mandarin generally recognized by Mandarin native speakers in Taiwan.

In fact, r-suffixation is highly lexicalized. Although some degree of r-suffixation has also been adopted into Taiwan Mandarin, their r-suffixed syllables have different phonetic realizations from their Beijing counterparts. In Wu (2012), it was found that Taiwan Mandarin speaker' r-suffixed syllables were longer in duration than Beijing Mandarin speakers. The durational discrepancy may be due to r-suffixed items (e.g., [xwaɚ]) being produced as one syllable by Beijing speakers but treated as two syllables by Taiwan speakers (i.e., [xua.ɚ]).

The second segmental difference between Beijing Mandarin and Taiwan Mandarin lies in the alveolar-retroflex distinction. Literature on Mandarin phonology generally prescribes a distinctive place contrast made in Beijing Mandarin, whereas a neutralization of the two categories is suggested for Taiwan Mandarin. According to Duanmu (2000, p. 26), "[t]he retroflex series [tʂ, tʂʰ, ʂ, ʐ] is a major characteristic of SC [Standard Chinese] speakers from Beijing. SC speakers from other places often do not have [tʂ, tʂʰ, ʂ, r] in their native dialects, so they often replace [tʂ, tʂʰ, ʂ, r] with the dentals [ts, tsʰ, s, z], or they may use [tʂ, tʂʰ, ʂ] for [ts, tsʰ, s] in hypercorrection". In introducing standard Mandarin spoken in Taiwan, Lin (2007) stated that the first major characteristic of Taiwan Mandarin is that the retroflexes are not distinguished from the alveolars. Some Taiwan Mandarin speakers pronounce the retroflexes as alveolars, and others pronounce both series of consonants as somewhere between the alveolars and the retroflexes (Lin 2007, p. 268).

The third characteristic that has been suggested to distinguish Beijing Mandarin from Taiwan Mandarin is the tonal realizations, even though both dialects use the same four lexical tones. First, Tone 3 is often not fully realized (i.e., no final rise of pitch) in Taiwan Mandarin. In addition, Fon and Chiang (1999) found that tone production in Taiwan Mandarin is characterized by a narrower tonal range, lower tonal heights and flatter tonal contours. Using Chao's (1930)

tone digits where 1 represents the lowest pitch and 5 the highest, they suggest that the tone

values for Taiwan Mandarin should be 44, 323, 312, and 42, as opposed to 55, 35, 214, and 51 as

described by Chao based on Beijing Mandarin speakers' speech (see Table 2.3). The difference

in tonal realizations may be attributed to the influence of Taiwanese (or Southern Min) on

Taiwan Mandarin. About 83 % of the population in Taiwan speaks Taiwanese at different levels

of proficiency (Huang 1993), and the lexical tones in Taiwanese have an overall lower register

than Mandarin (44, 24, 31, 52, 33 according to Chung 1968), which possibly shape the tonal

realizations in Taiwan Mandarin. To investigate this potential dialectal effect, Huang, Wu, and

Fon (2012) had Taiwan Mandarin speakers of high vs. low Taiwanese proficiencies produce the

four Mandarin tones. They found that high-proficiency Taiwanese speakers demonstrated 1) a

lower pitch register for Mandarin level tones, and 2) a narrower pitch range and flatter contour

for Mandarin contour tones compared to low Taiwanese proficiency speakers. Considering the

above discussion, how tones are realized may serve as an indexical property to distinguish the

two dialects.

Mandarin Tone	Beijing Mandarin (Chao 1930)	Taiwan Mandarin (Fon & Chiang 1999)
Tone 1	55	44
Tone 2	35	323
Tone 3	214	312
Tone 4	51	42

Table 2.3: Numeric annotation of Mandarin tones in Beijing vs. Taiwan dialect using Chao's tone digits.

The fourth dialect-distinguishing feature is the occurrence of neutral tone (see C.-Y. Chen

1984; C.-C. Tseng 2004; C.-Y. Tseng 1988 for detailed acoustic descriptions of neutral tone in

Putonghua and Taiwan Mandarin). Duanmu (2000) used tone neutralization as evidence for the

presence of lexical stress in Mandarin, thereby arguing that de-stressing (which is accompanied

by tone neutralization) is less common in Taiwan Mandarin. In the same vein, Lin (2007)

described Taiwan Mandarin speakers as demonstrating less distinction between heavy and light

syllables. According to Lin, words that retain the neutral tone in Taiwan Mandarin are generally

lexicalized (e.g., kinship terms) and the neutral-toned syllable usually has longer duration than

that in Beijing Mandarin. Listed in Table 2.4 are some well-known stress minimal pairs (or full-

toned vs. neutral-toned pairs). In contrast to Beijing speakers, Taiwan speakers generally only

produce the full-toned items listed in Table 2.4 and consider each minimal pair homophones.

pinyin	tone values	gloss
dong xi	55.55 55.0	"east and west" "thing"
ma tou	214.35 214.0	"horse head" "pier"
da.yi	51.51 51.0	"main idea" "careless"

Table 2.4: Stress minimal pairs in Beijing Mandarin (adapted from M. Chen 2000). The notation 55 refers to high level tone (T1); 35 high-rising tone (T2), 214 low-falling-rising tone (T3), 51 high-falling tone (T4), and 0 neutral tone.

2.3 Phonetic properties of Mandarin alveolar and retroflex sibilants

In this section, I will first review the articulatory and acoustic properties of Mandarin

alveolar and retroflex sibilants. Various enhancement strategies for Mandarin retroflexion will

also be discussed.

2.3.1 Articulatory description of Mandarin alveolar and retroflex sibilants

Alveolar fricatives have a constriction in the alveolar region. For languages that do not

distinguish the dental-alveolar place contrast, the alveolar fricatives may have free variants with

a dental constriction. In addition to the constriction location, the tongue portion is also a part of

the articulatory descriptions for alveolar fricatives. To use English /s/ for example, Dart (1991)

found only slightly more laminal articulations than apical articulations on her palatographic and linguographic data from 20 American participants. Therefore, which part of the tongue is used in producing alveolar fricatives may be due to individual anatomical characteristics. According to Ladefoged and Maddieson (1996), two minor articulatory gestures are also often observed across speakers in alveolar articulations: 1) the teeth come close together, and 2) a groove in the center of the tongue is present.

For Mandarin /s, ts, tsʰ/, the available articulatory studies suggest a more dental rather than alveolar articulation[6]. Ladefoged and Wu's (1984)'s X-ray data (see Figure 2.4) from three Beijing speakers indicated that all speakers produced /s/ with the tip of the tongue. The tongue tip was placed at various places of constriction: on the teeth for speaker 1; slightly behind the teeth for speaker 2; further back on the alveolar ridge for speaker 3. The palatographic and linguographic data collected from two speakers from Beijing by Lee and Zee (2003) suggested that the alveolar sibilants can be apico-laminal or laminal denti-alveolar. None of the aforementioned studies provided analysis of acoustic recordings made simultaneously with the articulatory data. However, in light of acoustic studies on English /s/ regarding the tongue portion, Dart (1991) reported that apical /s/ and laminal /s/ do not result in significantly different spectral means of the frication. As for the constriction location, English dental /s/-spectrum would have a flatter spectral shape in the high-frequency range, as opposed to a pattern with distinctive peaks in the alveolar /s/-spectrum (Li 2008).

---

[6] The place of constriction for retroflex fricatives and their corresponding affricates has not been found to be different in Ladefoged and Wu (1984)'s articulatory study of Mandarin, Ladefoged's (1994) study of Toda, and Ohala (1994)'s study of Hindi.
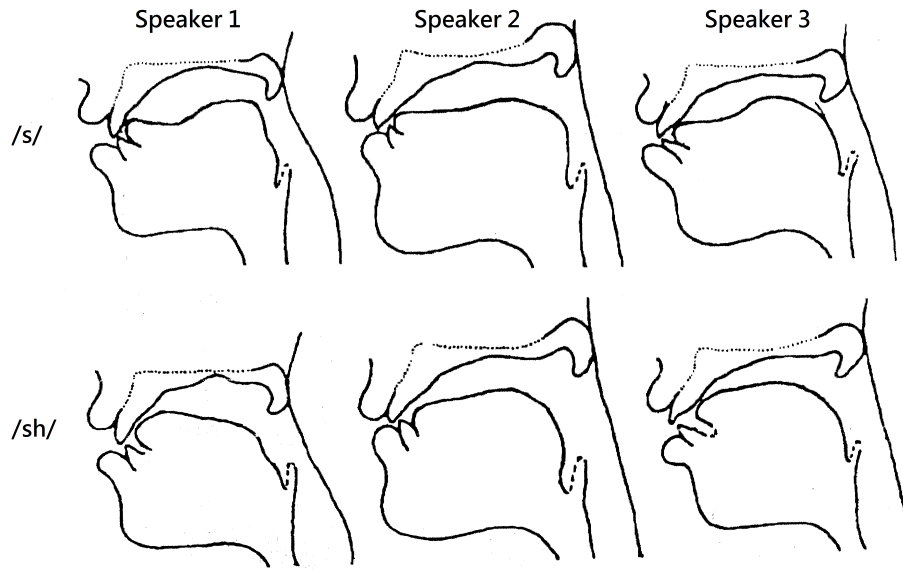
Figure 2.4: Midsagittal X-ray tracings of three male speakers producing /sa/ and /ʂa/ (based on Ladefoged and Wu 1984).

A canonical retroflex sound is defined as being articulated with the tongue tip curling up and back. However, cross-linguistic research has shown great articulatory variation in retroflex consonants in terms of place of constriction and tongue configuration. Two articulatory extremes of retroflex /ʂ/ articulations can be illustrated through the examples of Toda and Tamil. Toda has a subapical articulation in that the tongue tip approaches the palatal region (see Figure 2.5 a). For /ʂ/ in Tamil, it is articulated with the tongue tip against the post-alveolar region (see Figure 2.5 b). In either case, there involves a tongue-tip articulation—termed by Hamann (2003) as apicality. Hamann has proposed another three characteristics of retroflex articulations: posteriority, sublingual cavity, and retraction. Posteriority implies the presence of a sublingual cavity, formed as a result of the backwards displacement of the tongue tip/blade. Retraction refers to the tongue back moving towards the pharynx or velum, during which the tongue middle flattens and lowers.

Figure 2.5: Retroflex fricative /ʂ/ in Toda and Tamil (extracted from Hamann 2003)

For Mandarin retroflexes, Ladefoged and Wu's (1984) X-ray images of /ʂ /(as seen in Figure 2.4) indicated the use of tongue blade instead of tongue tip. The constriction was made around the center of the alveolar ridge for all three subjects, with which Ladefoged and Maddieson (1996) argued Mandarin retroflexes to be post-alveolars. The sublingual cavity was present in both speaker 1's and 3's retroflex productions. Lee and Zee (2003) also reported no retroflex articulation based on their unpublished articulatory data; instead, they found the Mandarin retroflex series to be apical post-alveolars /ʃ, tʃ , tʃʰ/. The available articulatory description of Mandarin retroflexes suggests that the retroflex articulation in Mandarin may be similar to that of Tamil.

2.3.2 Acoustic attributes of Mandarin alveolar and retroflex sibilants

How the spectral energy is distributed in the frication noise distinguishes sibilants of different places of articulation (e.g., Harris 1958; Heinz & Stevens 1961; Hughes & Halle 1956; Shadle 1985). The source of noise in fricatives is produced when the turbulent airflow escapes from a narrow constriction somewhere in the vocal tract. Based on the fricative spectra produced by Stevens' (1989) two-tube model, a shorter front cavity (i.e., a more anterior constriction) would result in a higher frequency of the lowest spectral peak.

For alveolar fricatives, the spectrum generally has a rising slope to the highest peak at around 8 kHz, and most energy of alveolar fricatives is located above 4000 Hz. For retroflexes, the lowering of higher formants is considered the acoustic consequence of retroflex articulation (Ladefoged & Maddieson 1996). Moreover, Stevens and Blumstein (1975, p. 219) noted that "the overall acoustic pattern is characterized by a clustering of F2, F3 and F4 in a relatively narrow frequency region". They further specified that the lowering of F3[7] is characteristic of retroflexion, which has been attested in retroflexes of various languages (e.g., Hamann 2002 on retroflexes in Polish and Russian; Hamilton 1996 on retroflexes in Australian languages; Ladefoged & Maddieson 1996 on English retroflex and bunched /r/).

Mandarin alveolar and retroflex sibilants can be distinguished in terms of different acoustic properties. Shih and Ao (1997) reported that frication duration of alveolars is longer than that of retroflexes based on data from text reading. On the other hand, in Wu and Cao's (1979) and Feng's (1985) experimental studies, frication from /ʂ/ was found to be longer in duration than /s/ in most vowel contexts, whereas the frication from /tʂ, tʂʰ/ was shorter than that of /ts, tsʰ/ in all vowel contexts. The variable durational characteristics of Mandarin sibilants and non-contrastive vowel/consonant length may be the reason why duration has not been reported to be an important perceptual cue in Mandarin alveolar-retroflex distinction. In addition, the mean intensity of the fricative spectra is suggested to be a parameter in characterizing Mandarin sibilants, with alveolars having a higher mean intensity than their retroflex counterparts (Svantesson 1986).

Mandarin sibilants can be differentiated by the frequency characteristics of the noise spectrum. Wu (1963) and Wu and Lin (1989) argued that the primary acoustic cue distinguishing

---

[7] The lowering of F3 can be caused by a posterior constriction, retraction of the tongue and existence of the sublingual cavity. Secondary articulation like lip rounding, velarization and pharyngealization involved in retroflex articulations can also contribute to the F3 lowering.

Mandarin alveolars from retroflexes resides in the position of the lowest spectral prominence, with /s/ higher than /ʂ/. More specifically, Stevens et al. (2004) pointed out that the spectrum of frication noise in Mandarin /s/ is centered around the frequency of F5, compared to the major spectral prominence which is centered around the F2 region for /ʂ/. Besides characterizing local differences in the fricative spectrum, places of articulation can also be differentiated by the measure of centroid frequency, or center of gravity (COG). The COG measure takes overall spectral distribution into consideration and is calculated as the mean of frequencies weighted by intensity. A longer front cavity as a result of a posterior constriction would lower the overall frequency for the major energy concentration in the fricative spectrum, which can be captured by the COG measure. Therefore, COG value is negatively correlated with the length of the front resonating cavity. This measure has been used in Jongman et al. (2000) to characterize English fricatives, in Svantesson (1986) for Mandarin sibilants, and Cho et al. (2002) for Korean fricatives. Figure 2.6 displays the LPC smoothed spectra of the /sa/ and /ʂa/ productions collected from one male Beijing Mandarin speaker. The COG measure is 8521 Hz for /s/ and 2724 Hz for /ʂ/ (see §3.5.2 on the procedure of obtaining the COG value). The spectra indicated that the lowest spectral peak is located at around 4600 Hz for /s/ and 1600 Hz for /ʂ/.
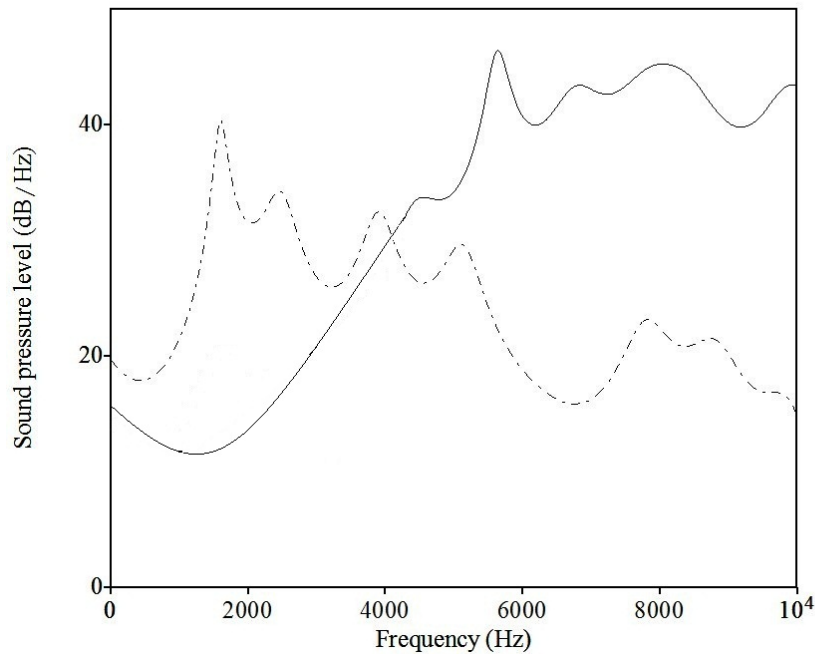
Figure 2.6: Averaged spectra over a 30-ms interval of the fricative portions from /sa/ (solid line) and /ṣa/ (dashed line)

In addition to the spectral characteristics of frication, vocalic formant transitions have been reported to play a role in the identification of sibilants in languages such as English (Whalen 1981) and Shona (Bladon et al. 1987). The vocalic transition between the sibilant and the following vowel has been suggested by Stevens et al. (2004) to discriminate Mandarin alveolar sibilants from their retroflex counterparts. They argued that F2 frequency taken at the onset of the vowel indexes the length of the back cavity in the fricative, for which the alveolar /s/ has a lower starting frequency of F2 and the retroflex /ṣ/ a higher F2 (see Table 2.5 for the onset F2 values taken for the alveolar and retroflex sibilants in /a, i, u/ contexts reported by Stevens et al.). However, the formant transition cue was not found to significantly affect Mandarin fricative perception in Chiu (2010) and Chang (2011). Chiu reported that Mandarin listeners were able to correctly identify /sa/ and /ṣa/ with incongruent formant transitions (e.g., /s/ followed by the

22

vocalic portion crossed-spliced from /ʂa/). In Chang's study, it was found that Mandarin listeners' goodness ratings revealed no sensitivity to mismatched transition in crossed-spliced /sa/ and /ʂa/ tokens. The contribution of formant transitions to fricative perception may be language-specific and depend on spectral similarity among target fricatives (Wagner et al. 2006). In Mandarin, the cue contained in the frication noise is still the most important cue in the alveolar-retroflex distinction.

vowel fricative	/a/	/i/	/u/
/s/	1506	1552	1453
/ʂ/	1740	1810	1521

Table 2.5: Onset F2 frequency (in Hz) for Mandarin alveolar/retroflex sibilants (reproduced from Stevens et al. 2004)

2.3.3 Enhancement of Mandarin retroflexion

In the theory of feature enhancement (Stevens et al. 1986; Keyser & Stevens 2006), gestures producing acoustic effects that are secondary to a feature-defining gesture may be introduced to enhance the perceptual saliency of a sound output. According to Stevens & Keyser (1989), [-anterior] obstruent consonants like retroflexes have weaker acoustic cues compared to their [+anterior] counterparts. To maintain a phonological contrast, these [-anterior] obstruents can be enhanced by additional articulatory gestures like rounding or velarization. In Stevens et al.'s (2004) study of Mandarin /s/ and /ʂ/, they argued that Mandarin speakers would avoid using a lip gesture in retroflex articulations as enhancement. Their argument was a phonological one, in that lip rounding could potentially jeopardize another phonological contrast between /ʂ/ (as in /ʂan/ 'mountain') and /ʂʷ/ (as in /ʂʷan/ 'to tie')[8]. Stevens et al. claimed that raising the tongue blade would be used to create a space under the tongue for enhancement, since the gesture can produce

---

[8] When a consonant is followed by an on-glide /u/, the /u/ vowel is analyzed as a secondary articulation of the initial consonant by Duanmu (2000) such that the CG sequence forms a single segment C^G. Under this analysis, /ʂuan/ 'to tie' is treated as /ʂʷan /, which makes /ʂʷ/ and /ʂ/ (as in /ʂan/ 'mountain') contrastive in terms of rounding. However, /ʂʷ/ by itself is not a phoneme in Mandarin.

an effect that maintains F3 as the front cavity resonance—an acoustic effect lip rounding can similarly achieve. However, Chang's (2010) articulatory data obtained from four Taiwan Mandarin native speakers with Electromagnetic Articulography (EMA) indicated that two participants demonstrated a statistically significant difference between the retroflex and alveolar articulations in terms of lip rounding. That is, lip rounding can be used as an enhancing gesture in Mandarin retroflexion, at least for Taiwan Mandarin. The same lip rounding gesture has also been reported to enhance retroflexion in Swedish (Bruce 2006).

Besides enhancement for the retroflex consonants, the distinct vowel quality of the Mandarin apical vowels following /s/ vs. /ʂ/ was also approached by Stevens et al. (2004) from the perspective of enhancement. They argued that the vowel variants serve to maximize the three-way coronal contrast. In the same vein, Lee and Li (2003) proposed that modifying the vowel that follows a Mandarin coronal would enhance the perceptual saliency of its coronal place of articulation. More specifically, the /i/ vowel is modified to be /ɨ/ after alveolar sibilants and /ʅ/ after retroflex sibilants to enhance the defining articulatory and acoustic attributes of the preceding sibilant.

2.4 Place neutralization

The retroflex sibilants are absent in southern dialects of Chinese (e.g., Z. Chen 2010, based on European missionaries' linguistic documentation in the 19 century; Ramsey 1987), which has been argued to be the reason why southern Mandarin speakers' alveolar-retroflex production is subject to place neutralization (e.g., Chien 1971 and Kubler 1985 for Taiwan Mandarin; L. Zhu 2012 for Shanghai Mandarin). However, according to Y. Zhu (1998) and C.-Y. Chen (1991), alveolar-retroflex neutralization also occurs in many northern dialect areas such as the northeastern provinces in China, Tianjin, Henan and even Beijing.

Alveolar-retroflex place neutralization does not occur on an all-or-none basis, as various degrees of neutralization have been suggested. Li (2009), for example, argued for incomplete alveolar-retroflex neutralization in Taiwan Mandarin. He reported that while production of the alveolar-retroflex contrast is conditioned by multiple contextual factors, there are certain contexts in which neutralization is more likely to occur. According to Li, neutralization tends to occur before high vowels, in lower density neighborhoods and when one's ambient language is Taiwanese rather than Mandarin. In addition, extralinguistic factors like gender (Chuang & Fon 2010; Rau 1996) and speech style (Jeng 2006) were also found to affect the place contrast. Specifically, females better preserve the alveolar-retroflex contrast and show a lower rate of place neutralization than males. This generalization is in accordance with Labov's (1998) observation that "in stable sociolinguistic stratification, men use a higher frequency of non-standard forms than women" (p. 7). In more formal tasks such as wordlist reading, the place contrast is found to be better maintained than in less formal tasks like free conversation.

In the case of alveolar-retroflex neutralization, alveolars are the default forms. That is, retroflexes are replaced with their alveolar counterparts for Mandarin speakers who do not make the place distinction at all. In speech, such neutralization generally does not cause confusion. The functional load (see King 1967 for detailed discussion and computation) of Mandarin fricatives is higher in isolated syllables (0.0207) than in words (0.0051) (Surendran & Niyogi 2003). Therefore, lexical confusion as a result of place neutralization can be better resolved in word context than in isolated syllable context.

There are cases of near neutralization where retroflexion is realized as an intermediate form between alveolar and retroflex—a less retroflexed form (as seen in Chung 2006 and Lin 2007 for Taiwan Mandarin; in C.-Y. Chen 1986 for Singaporean Mandarin). A reduced retroflexed

pronunciation could result in a covert alveolar-retroflex contrast that is perceptually unreliable yet statistically significant in the acoustic measures (see Macken & Barton 1980 for voicing contrast in English; Forrest et al. 1990 for place contrast in English stops). Retroflexion can also occur as a result of overgeneralizing retroflexing to alveolars. Hypercorrect forms are used in response to a situation that calls for formality (Chung 2006). Hypercorrection indicates that speakers are aware of the alveolar-retroflex contrast but cannot make correct distinctions consistently.

Taken together, Mandarin alveolar-retroflex contrast is variably realized within as well as across speakers and dialects. Because the alveolar is the default form, the presence of retroflexion in speech can be used to determine whether the alveolar-retroflex contrast is neutralized. For speakers who make a lesser or inconsistent contrast, retroflexion may be realized as a reduced form or surface as a result of hypercorrection.

2.5 Social indexing of Mandarin alveolar-retroflex contrast

The use of Standard Mandarin pronunciation in popular attitude of the Chinese speaking community is associated with good education, authority and social sophistication (P. Chen 1999). On the other hand, Mandarin with regional dialectal features may be used to signal affinity, sincerity (Su 2005) and social identity (Liao 2008). Among the features that distinguish a standard Mandarin pronunciation and a local dialect-accented Mandarin pronunciation, retroflexion is a canonical index (Chung 2006). While Duanmu (2000) claimed that Standard Chinese "does not carry a superior social prestige" (p. 5) like that of the Received Pronunciation of British English, the lack of an alveolar-retroflex distinction in speech is considered a deviation from the standard form of Mandarin in S. Chen (1990) and an adulterated form of Putonghua in P. Chen (1999). In China, Standard Mandarin that distinguishes the alveolars from the retroflexes

is a High variety[9] used in public domains and in the media throughout the country (P. Chen 1999). In Taiwan, retroflexed speech is also viewed as a prestigious form and receives the label of being "standard" or "correctly pronounced" (Chung 2006).

The alveolar-retroflex contrast has been promoted in various ways. In China, *Putonghua Shuiping Ceshi* (National Putonghua Proficiency Test) is a standardized oral proficiency test at the state level. The test assesses the oral proficiency of Chinese nationals if they wish to apply for jobs in public domains (e.g., schools and government administration). Speakers are placed into three levels based on their oral performance. One of the indicators for Level 2 speakers is to make inconsistent or no alveolar-retroflex distinction (http://www.china-language.gov.cn/87/2007_6_20/1_87_2635_0_1182325529609.html, accessed on June 31, 2012). In Taiwan, there is no official proficiency oral test for nationals. However, retroflexion is prescribed as a feature of "textbook Mandarin", and is the standard for most media broadcasting in Taiwan (Chung 2006). In sum, in both China and Taiwan, the use of retroflexion in speech is indexed for a standard pronunciation and is associated with a higher education level.

Given the social indexical value in retroflexion, speakers have been found to exhibit stylistic retroflexing in speech. For example, Jeng (2006) reported that retroflexion occurs more often in response to a more formal register for Taiwan Mandarin speakers. Chung (2006) pointed out that Taiwan Mandarin speakers are more subject to hypercorrect speech when they try to display respect or formality to the interlocutor. With the examples of stratified variation of retroflexion, particularly in Taiwan Mandarin, it can be seen that "[an] individual creates for himself the patterns of his linguistic behavior so as to resemble those of the group or groups with which from

---

[9] Ferguson (1959) calls a prestigious speech variety used in conjunction with good education or formal occasions the "High variety".

time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished" (Le Page & Tabouret-Keller 1985; p. 181).

2.6 Sound change

In view of the variants of retroflexion and different degrees of the alveolar-retroflex contrast exhibited by a wide spectrum of Mandarin native speakers, it may be speculated as to whether place neutralization is on its way to a complete merger. For native Mandarin speakers who do not have the retroflex phonemes in their sound inventory, the retroflexes have been completely merged into their alveolar counterparts. What is of more interest to diachronic phonological change is speakers whose retroflexion fades in and out of their speech and who use retroflexion in formal speech—which is widely observed in Taiwan Mandarin speakers (Chung 2006).

Common sound change processes such as lenition can be traced back to reduction in casual speech. In order for reductive sound change to take place, casual speech variants need to be conventionalized and phonologically recategorized (Hualde, to appear). That is, if retroflexes are to be completely merged into their alveolar counterparts, the alveolar sibilants must become a conventionalized realization of the retroflex sibilants, and subsequently all variants of retroflexes must be reinterpreted as alveolars. However, as discussed in §2.5, retroflexion is a critical feature of the standard form of Mandarin and is explicitly taught in school in China and Taiwan. In order to type character or transcribe character pronunciations in Pinyin (a romanization system used in China) or Zhuyinfuhao (a set of sound denoting symbols used in Taiwan), Mandarin learners or native speakers must memorize by rote which written Chinese characters have alveolar vs. retroflex initials. Therefore, even though some speakers may be confused about the usage of the retroflex or use retroflexion only in more formal speech registers, it is unlikely that the retroflexes can be completely recategorized as alveolars.

From a sociolinguistic perspective, sound changes can be triggered by hypercorrective behavior (Labov 1973) in that conscious overgeneralized correction of a linguistic form could lead to an overall linguistic change in the language. This trigger would predict Mandarin alveolar fricatives to be merged into retroflexes. However, Chung (2006) argued that hypercorrection only occurs in situations calling for more formal and careful speech in Taiwan Mandarin, as the default form is still the alveolars for speakers who do not make a consistent place distinction. It is not known whether hypercorrect forms will be prevalent enough in Taiwan such that retroflexion would replace the alveolars in production.

As discussed in §2.4, alveolar-retroflex neutralization in Taiwan Mandarin may be best viewed as incomplete neutralization. The place contrast being more subject to neutralization in certain phonetic contexts (Li 2009) can be related to positional neutralization for voicing contrast in Dutch (e.g., Warner et al. 2004). Stylistic retroflexing found in Taiwan Mandarin speakers (Jeng 2006; Chung 2006) is also in line with Port and Crawford's (1989) findings that German speakers were able to make a better word-final voicing distinction in careful speech and control the degree of neutralization depending on the formality of the speech style. In sum, if the acoustic realizations of the alveolar and retroflex sibilants differ somehow in a predictable way, then the neutralization is not phonetically complete. For sound change to take place, where the two categories are merged, any difference between the realizations of the two categories must become unperceivable. Further implications on sound change regarding Mandarin alveolar-retroflex contrast will be drawn from the results of the perception study in Chapter 4.

2.7 Summary

This chapter reviews the phonetic, phonological and sociolinguistic characteristics of Mandarin alveolar-retroflex contrast—a critical feature that distinguishes the two major dialects

of Standard Chinese: Beijing Mandarin and Taiwan Mandarin. The discussion of place neutralization and sound change was focused on Taiwan Mandarin since it is reported in the literature that Taiwan speakers' realizations of the alveolar-retroflex contrast are subject to various factors. The production study in Chapter 3 is hereby designed to systematically investigate the cross-dialectal production of Mandarin alveolar-retroflex contrast, in different vowel contexts and under the influence of contrastive focus.

CHAPTER 3: PRODUCTION STUDY

3.1 Introduction

In Taiwan Mandarin, a number of contextual factors such as vowel contexts (Li 2009) and prosodic prominence (Chuang & Fon 2010) have been identified to greatly influence the magnitude of the alveolar-retroflex contrast. It is unclear whether similar phonetic variability found in Taiwan Mandarin exists in the alveolar-retroflex contrast for Beijing Mandarin as well. Therefore, this production study aims to investigate the cross-dialectal production of Mandarin alveolar-retroflex contrast with regards to phonetic conditioning. In examining the realizations of the Mandarin alveolar-retroflex contrast in various phonetic contexts, I will provide a detailed acoustic characterization of the variation in Mandarin sibilant productions across dialect, prosodic, and vowel contexts.

This chapter will first review factors that have been reported in the literature to affect the acoustic realizations of phonological contrasts, particularly the Mandarin alveolar- retroflex contrast (§3.2). The goal of this study and research questions will be posed in §3.3. The methodology section (§3.4) will address how the factors reviewed in §3.2 are controlled in the production experiment and proceed with the presentation of the experiment design and experimental procedure. Next, the chapter will present analysis and report findings from the experiment (§3.5). Finally, the chapter will discuss the production patterns across various experimental conditions (§3.6) and draw a conclusion (§3.7).

3.2 Factors affecting the acoustic realizations of Mandarin alveolar-retroflex contrast

3.2.1 Prosodic prominence

The acoustic realization of a phonological contrast can be influenced by many factors, among which the effects of prosodic prominence on strengthening phonological distinctiveness have been well studied in the literature. Beckman et al. (1992), Cho (2005), de Jong (1995), and Mo et al. (2009) have reported a significant effect of prosodic prominence (e.g., in an accented syllable or in focus condition) on vowel formant realizations. As a result, a greater phonemic contrast or an expanded acoustic vowel space may be observed (e.g., Hay et al. 2006). While vowels are usually more distinctively realized in the prominent condition, they might not be enhanced in the same way. When accented, Cho (2005) found the English /i/ vowel to be fronter but not necessarily higher in the acoustic dimension. The low back /ɑ/ vowel, in contrast, appears to be both lower and more back in the accented condition. The results suggest that not all distinctive features (i.e., [front/back] and [height]) are enhanced uniformly in prosodically prominent conditions.

Prosodically conditioned acoustic variation has also been found for consonants. Keating (1984) reported increased VOT for English voiced and voiceless stops in domain-initial position. Cole et al. (2007) found that voicing contrast is enhanced for English stops in accented syllables in that accentuation induces a larger contrast between voiced and voiceless categories, although they also noted one of their subject's voicing contrast being diminished under accent. The effects of prosodic prominence on consonantal place contrasts are even more variable. While expecting increased acoustic distinctiveness under accent, Cole et al. (2007) actually found the place

contrast between English /p, t, k/ to be strengthened (i.e., all phoneme categories undergo increases in the relevant acoustic measures to similar extent such that there is no greater separation between categories) instead of being enhanced (i.e., greater separation between categories). The effects of accentuation, as reflected in the acoustic measures of VOT and closure duration, are uniform across stops of different places of articulation in that the two place cues uniformly exhibit greater values under accent. As a result, no greater place distinction is observed under accent through longer VOT and closure duration. On the other hand, the burst spectrum measures show a reduced place contrast in the accented condition, and the F0 measure shows little effect of accentuation on the place distinction. Taken together, Cole et al.'s results suggest that the effects of prosodic strengthening on place contrasts may not be present or uniformly manifested in relevant acoustic measures.

Some evidence of prosodic strengthening effects has been found for the realizations of Mandarin alveolar-retroflex contrast. Chuang and Fon (2010) reported that the Taiwan Mandarin speakers differed in their production of alveolar and retroflex sibilants in accented vs. unaccented conditions in spontaneous speech. Specifically, two different strengthening strategies were observed in their subjects who made the alveolar-retroflex distinction in the accented condition: 1) enlarge the place contrast by producing the alveolar or retroflex sibilants or both toward a more extreme direction (i.e., retroflexes have a lower COG and alveolars have a higher COG), and 2) strengthen the place distinction by increasing the COG of both sibilants[10]. On the other hand, for the subjects who did not make the place distinction in the unaccented condition, their production in the accented condition did not exhibit a place distinction either.

---

[10] Chuang & Fon speculated that increasing COG of both sibilants creates high-frequency noise, which in turn may give rise to the perception of prominence.

3.2.2 Vowel context

Alveolar sibilants are affected by vowel coarticulation differently in the articulatory and acoustic dimensions. In producing English alveolar sibilants, the tongue blade and tongue body tend to move toward the same position regardless of the /a, i, u/ contexts (Kent & Moll 1972). Alveolar sibilants being less subject to vowel-dependent variability has also been found in German (Hoole et al. 1993) and Catalan (Recasens 1991). However, acoustically, significant vowel-related COG change, especially for the rounding dimension, has been found in Niebuhr et al.'s (2011) study of French and English /s/. This articulatory-acoustic relation suggests that although the lingual configuration of alveolar sibilants is more resistant to vowel coarticulation, the effect of vowel coarticulation on lip configuration may still result in differences in the acoustic realizations of the alveolars sibilants.

The vowel context is one of the major factors that result in variability in retroflexion (Hamann 2003). In Hindi, the place of constriction changes when the retroflex stops are followed by different vowels, according to an EPG study conducted by Dixit and Flege (1991). An EPG study of Norwegian retroflex consonants by Simonsen et al. (2008) also reported similar results in that the retroflex consonant in the /a/ vowel context is articulated further back than that in the /i/ context. Acoustically, retroflex consonants are found to exhibit different formant structures depending on the following vowel (Hamann 2003).

Besides the effect of the following vowel, in some literature on retroflexion (Simonsen et al. 2003 for Norwegian retroflex stops in coda position; Tabain 2009 for Central Arrernte retroflex consonants in VC syllables), the cues to retroflexion are argued to be stronger in the preceding vowel than the following vowel. However, since Mandarin alveolar and retroflex sibilants can only serve as syllable onsets, the coarticulatory effect from the preceding segment is reduced

across the syllable boundary such that the effect of the following vowel is more important on the spectral measure of Mandarin sibilants.

In Mandarin, sibilant realizations have been found to be subject to vowel coarticulatory effects. Jeng (2006) reported that alveolar sibilants are generally more prone to coarticulatory influence of the following vowel than their retroflex counterparts. In addition, with all else held constant, the spectral center of gravity measured over the fricative followed by a rounded vowel was lower than when followed by a non-rounded vowel. The vowel context may also be an important factor in the magnitude of the Mandarin alveolar-retroflex contrast. A smaller alveolar-retroflex contrast in the /u/ context was observed in Jeng's (2006) Taiwan Mandarin data. In Li's (2009) study of Taiwan Mandarin sibilant production, he found a smaller COG difference between the two sibilant categories before high vowels, with which he argued the high vowel context to be where place neutralization is more likely to occur. However, it should be noted that Li collapsed sibilant productions in the /i/ and /u/ contexts in the analysis, as he was interested in the factor of vowel height dimension instead of front-back dimension. Hence, it is unclear whether the separation between the two categories in the /i/ and /u/ contexts was both smaller compared to that in the /a/ context.

3.2.3 Word frequency

High-frequency words are produced more quickly than low-frequency words (Oldfield & Wingfield 1965), such that frequent words are shorter in duration than less frequent words (Jurafsky et al. 2001). In addition, high-frequency forms also tend to induce phonetic reduction (e.g., Bybee 2002 for lenition of English consonants; Tseng 2005 for coarticulation in Taiwan Mandarin). Therefore, word frequency is an important conditioning factor in phonological variation and phonological contrasts.

High homophone density in Mandarin raises the question of whether homophony should be considered in this production study. A Chinese *character* is monosyllabic and is generally associated with a morpheme, whereas a Chinese *word* is mostly a disyllabic compound. Many Chinese syllables correspond to a number of different lexical representations, and therefore, homophony may affect lexical retrieval in production. The effects of homophone frequency and word frequency on lexical access performance have been explored in the literature. Jescheniak and Levelt (1994) found that low-frequency homophones in Dutch behaved like high-frequency controls, inheriting the accessing speed of their high-frequency homophone twins. This homophone inheritance effect was challenged by Caramazza et al.'s (2001) study of Spanish and Mandarin productions, as their findings suggest a specific word frequency effect in lexical access. In this study, all the homophones of a target syllable in this study (i.e., alveolar and retroflex syllables) will be pooled in the calculation of its frequency for two reasons: 1) it is a more conservative measure in view of arguments in the literature about the homophone effect, and 2) the study is interested in low-level phonetic measures and therefore should control the frequencies of target phonetic forms (Dell, personal communication).

3.3 Goal and research questions

The goal of this study is to investigate whether vowel context, prosodic prominence and dialect affect the sibilant realizations and the alveolar-retroflex contrast. It is expected that Mandarin sibilant realizations are subject to the coarticulatory effect from the following rounded vowel (i.e., lowering of COG) in view of the literature reviewed in §3.2. As for the vowel effect on the magnitude of the alveolar-retroflex place contrast, while Jeng's (2006) study indicated that the contrast is smaller only in the /u/ context, Li (2009) reported that the alveolar-retroflex contrast is smaller in high vowel contexts. Results from the two studies converged in the

coarticulatory effect of the /u/ vowel on the place contrast, which puts forward a hypothesis that the Mandarin alveolar-retroflex contrast will be smaller in the /u/ context. In terms of the factor of prosodic prominence, it is not known whether prosodic strengthening has differential effects depending on dialect. Therefore, it is hypothesized that prosodically-conditioned strengthening effects found in Taiwan Mandarin (Chuang & Fon 2010) are also present in Beijing Mandarin. Three research questions will be addressed in verifying the aforementioned hypotheses: 1) Do the realizations of the alveolar/retroflex sibilants and the alveolar-retroflex contrast vary across vowel contexts in both Beijing Mandarin and Taiwan Mandarin? 2) Do the realizations of the alveolar/retroflex sibilants and the alveolar-retroflex contrast vary in different prosodic conditions for both Beijing Mandarin and Taiwan Mandarin? 3) Is there any interaction effect between the vowel contexts and prosodic prominence on the realizations of the alveolar/retroflex sibilants and the place contrast for both Beijing Mandarin and Taiwan Mandarin?

3.4 Experimental design

To obtain balanced (i.e., an equal number of alveolar vs. retroflex tokens in various vowel contexts) yet natural speech data for this study, a series of map tasks are used to elicit alveolar and retroflex productions. In a map task (Anderson et al. 1991, based on the MIT Map Task recordings, Human Communication Resource Centre), one participant is typically the giver of instructions while the other is the receiver of such instructions. In §3.4.2.1, I will further address how the factors affecting Mandarin sibilant realizations (as reviewed in §3.2) are controlled in the design of the map tasks.

This study is interested in whether Mandarin alveolar-retroflex contrasts are conditioned by vowel contexts and prosodic prominence for speakers who are able to make such a place distinction. To exclude the population of Taiwan Mandarin speakers who make no alveolar-

retroflex distinction, a screening test is conducted as a part of the production experiment. Speakers are asked to read the Mandarin version of The North Wind and the Sun (see Appendix A), which contains several alveolar and retroflex tokens. The recordings are used as screening materials. The screening process and inclusion criteria will be discussed in detail in §3.4.3.

3.4.1 Participants

The participants included eight Beijing Mandarin speakers (4 males, 4 females) and ten Taiwan Mandarin speakers (6 males, 4 females). All of the Beijing Mandarin speakers were born and raised in Beijing before coming to the United States; the Taiwan Mandarin speakers were mostly from northern Taiwan. While six out of the ten Taiwan Mandarin participants speak some Taiwanese (or Taiwan Southern Min), Mandarin was reported to be the major language used in their families. All participants were undergraduate or graduate students at the University of Illinois, between the ages of 18 and 35. None of them had resided in the United States for over 5 years. All participants had normal self-reported speech and hearing. Further demographic data and language background of the participants are listed in Table 3.1.

Participant	Sex	Age group (yrs)	Home town (born & raised)	Other Chinese dialects	Foreign languages	Length of residence in USA (yrs)
BF1	female	30-35	Beijing		English, Russian	1.5
BF2	female	18-25	Beijing		English	0.3
BF3	female	30-35	Beijing		English	5
BF4	female	18-25	Beijing		English	1
BM1	male	18-25	Beijing		English	0.1
BM2	male	18-25	Beijing		English	0.1
BM3	male	18-25	Beijing		English	0.2
BM4	male	30-35	Beijing		English	0.1
TF1	female	25-30	Taipei	Cantonese	English	4
TF2	female	18-25	Taipei	Taiwanese	English	4
TF3	female	18-25	Taipei		English	0.1
TF4	female	18-25	Taipei	Taiwanese	English	0.1
TM1	male	25-30	Changhua	Taiwanese	English	0.5
TM2	male	18-25	Kaohsiung		English	4.5
TM3	male	30-35	Taipei	Taiwanese	English	3
TM4	male	18-25	Taipei	Taiwanese	English	2
TM5	male	30-35	Changhua	Taiwanese	English	4.5
TM6	male	30-35	Taichung		English	4.5

Table 3.1: Background information of the participants

3.4.2 Production tasks

After the participants completed the language background form (see Appendix B), they proceeded with four production tasks: reading the Mandarin version of *The North Wind and the Sun*, conversation, map tasks and wordlist reading[11]. Given the different orthographic systems used in China and Taiwan, all printed materials came in two versions, simplified and traditional Chinese, and were distributed to the participants based on their reported place of origin. All audio recordings were conducted in a sound-attenuated booth in the Phonetics Lab at the University of Illinois. An AKG C520 head-worn condenser microphone was used to record acoustic signals of the participants' speech onto a Marantz PMD570 recorder.

---

[11] The conversation and wordlist reading data were not analyzed in this study. The alveolar/retroflex productions in these data, however, can be used in future studies that investigate the influence of formality on production.

3.4.2.1 Considerations for design of map tasks

In studying the effects of prosodic strengthening on Mandarin alveolar-retroflex contrasts, one first has to decide how to operationalize degrees of prosodic strength. In Chuang and Fon's (2010) study of conversational speech, they adopted a modified version of the Pan-Mandarin ToBI system to label the prominence levels of alveolar and retroflex syllables. Their prominence labeling is based on three acoustic properties, tonal realization (reduced vs. default tonal shape), duration (shortened vs. default duration) and amplitude (soft vs. default amplitude)[12]. In this study, I chose to experimentally manipulate prominence in terms of contrastive focus so as to ensure that the same stimuli would be produced with and without focal prominence. Map tasks are an appropriate data elicitation method for the current study as Chen (2003) has successfully elicited focused and unfocused productions in Mandarin via map tasks. Regarding the map tasks adapted for this study, focused productions were target syllables uttered in a corrective manner (i.e., participants corrected geographical terms on a map) and unfocused productions were those uttered in a non-corrective manner (i.e., participants gave the route of stops on a map).

Two decisions had to be made in creating the four-syllable stimuli for the map tasks: position of the target syllable in stimulus and the focus domain. In view of Chen's (2003) conclusion of edge-effect lengthening and sensitivity of lengthening to the metrical structure of the focused constituent in Chinese, the position of a target syllable in a word needs to be held constant, as the position would affect phonetic realization. In Chinese, the second syllable of a disyllabic word was found to be more subject to phonetic reduction (e.g., Zhang 1988; Kochanski & Shih 2003), such that making the target syllable the first syllable of a word would avoid syllable reduction or contraction. Therefore, in this study, I consistently designated the target syllable as the first

---

[12] The ToBI labeling criteria Chuang and Fon (2010) used were modified from Peng et al.'s (2007) Pan-Mandarin ToBI system. Chuang and Fon did not specify what their baseline duration and amplitude measures were.

syllable of a four-syllable word. As for the contrastive focus domain, the decision was made in light of the literature on realizations of lexical tones and syllable duration in response to focus. When a Chinese syllable is directly under contrastive focus, the F0 range of the lexical tone is greatly enlarged (Shih 1988; Xu 1999). Furthermore, more robust and uniform lengthening and F0 changes for Mandarin syllables were found within a smaller focus domain (Chen 2003). Taken together, when a focus domain is smaller (i.e., one syllable), a stronger strengthening effect is expected. Therefore, target stimuli and filler items were created to differ only by the first syllable such that the contrastive focus domain is set on the first syllable of a four-syllable word.

Finally, to better control word frequency, the stimuli in this study were devised to be pseudowords. In that regard, it is expected that the participants may produce the same stimuli more easily during a second time (repetition for the purpose of comparing focused vs. unfocused productions). If a stimulus was produced the first time under a non-focus condition and second time under a focus condition, the effect of focus on phonetic realizations might be reduced due to repetition (see Brown 1983 and Hirschberg 1993 for deaccentuation of given information). Therefore, this experiment was designed in a way that focused productions would precede non-focused productions.

3.4.2.2 Map tasks and stimuli

Five maps were designed to naturally elicit alveolar and retroflex productions under contrastive focus and non-focus conditions (see Appendix C for a sample map task). Each map had a different theme and an equal number of target stimuli and corresponding fillers. The instructions were written at the top of each map: "1) Your partner (i.e., the experimenter[13]) is lost

---

[13] The experimenter was the author of this dissertation, who is a native Taiwan Mandarin speaker. The author was aware that within-language convergence (Giles et al. 1991) can occur readily and rapidly, such that the Beijing participants might accommodate their speech to the author's. However, to ensure all focused and unfocused

and has a misoriented map. Based on your map, you need to correct your partner's inaccurate

statements of geographical relations. 2) After the location of all the sites has been confirmed, you

should decide on a route that stops at every site on the map and then describe the route to your

partner." During the map tasks, the participant and the experimenter were seated across from

each other with a plastic board situated between so that the two people could not see each other's

map.

There were 60 target stimuli (2 places of articulation * 2 focus conditions * 3 vowel contexts

/a, i, u/ * 5 samples—of the same place-vowel combination varying by tone and manners of

articulation) for the map tasks. All stimuli were 4-syllable pseudowords. Each target stimulus

was composed of one alveolar/retroflex syllable, a non-alveolar/retroflex syllable and a

disyllabic compound. Neither the second syllable nor the disyllabic compound had an

alveolar/retroflex onset to avoid sound interference with the first syllable. The second syllable of

every 4-syllable stimulus had a plosive onset (differing in voicing and places of articulation) and

one of the diphthongs /ai, ei, ou, au/ as the nucleus. The disyllabic compounds were

meaningful—they were all landscape terms (e.g., river, cliff) or scenic spot names (e.g., castle,

gallery), thereby making the first two syllables a pseudo name of the landscape term or scenic

spot name. The fillers differed from the target stimuli only by the first syllable, and the first

syllable of the fillers was not an alveolar or retroflex syllable. Figure 3.1 gives a schematic

presentation on how the target stimuli were created, followed by a description on how each

constituent of the target stimuli was controlled in terms of syllable and word frequency.

---

productions were successfully elicited given the complexity of the map tasks, the recruitment of a Beijing Mandarin-speaking experimenter was not considered.
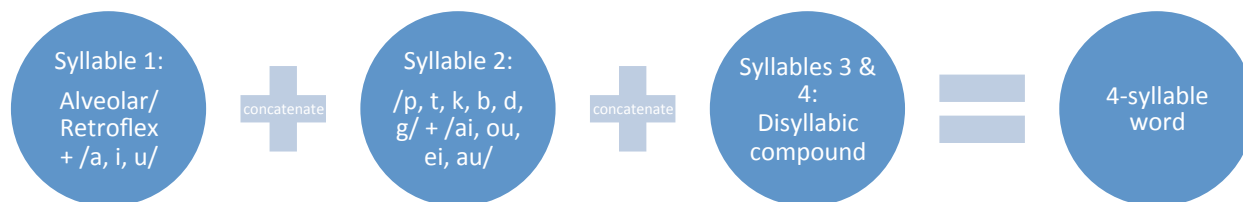
Figure 3.1: The composition of the 4-syallble stimuli

Access to three Mandarin corpora was obtained in order to confirm word frequency:

CALLHOME Mandarin Chinese Lexicon (McEnery & Xiao 2008), Chinese Treebank Version

4.0 (Palmer et al. 2004), and SUBTLEX-CH (Cai & Brysbaert 2010). In the end, only

SUBTLEX-CH was used for confirming word frequency for two reasons: 1) New et al. (2007)'s

and Cai and Brysbaert (2010)'s studies show that word frequencies based on film subtitles (i.e.,

SUBTLEX-CH) correlate more with word processing times than word frequencies based on

written texts from newspapers or internet pages (i.e., Chinese Treebank Version 4.0). 2) The

SUBTLEX-CH corpus has a total number of 5936 different characters and 99121 different words

based on a total of 33.5 million words, which is much bigger than both the CALLHOME and

Treebank corpora.

The lexical frequency of an alveolar or retroflex syllable was derived from the log of the sum

of counts of all the possible orthographic representations (i.e., homophones) for a syllable.

Between the log frequencies of 3.4 and 5, five alveolar-retroflex pairs for each vowel context

were chosen, which resulted in 30 syllables. Following the same log frequency selection

procedure, 30 syllables that had similar log frequencies (3.6-4.8) were selected to be the second

syllables of the stimuli. For the disyllabic compounds, a list of geographical terms and scenic

43

spot names were first generated. These compounds were checked against the SUBLEX-CH corpus and were reduced to 30 compounds that share similar log word frequencies (2.0-3.0). With the 3 lists of constituents, one item was randomly selected from each list and then concatenated into a 4-syllable word. All of the concatenated stimuli were checked against the corpus, and none of the disyllabic names or the 4-syllable words occurred in the corpus, indicating that they were indeed pseudowords. The stimuli then were distributed to one of the 5 map tasks based on the theme of the map they best fit in, making 6 target stimuli and 6 fillers on each map. The final list of target stimuli and fillers can be found in Appendix D. These stimuli were listed in a random order for the wordlist reading task at the end of the production experiment.

3.4.3 Screening results

The screening materials came from speakers reading the Mandarin version of *The North Wind and the Sun* (see Appendix A). The read speech was recorded and all of the alveolar and retroflex tokens were later extracted for an identification task conducted on two phonetically trained native speakers (1 Beijing speaker and 1 Taiwan speaker). Most alveolar and retroflex syllables from the text had a retroflex and alveolar counterpart, hence constituting minimal pairs. The two listeners listened to each retroflex and alveolar production and were presented with one character for each of the two possible values in the minimal pair. The listeners had to choose the character that best represented the sound they heard. In view of previous literature (e.g., Kubler 1985; Chung 2006) that reported alveolars as the default form for Mandarin alveolar-retroflex neutralization, only the accuracy of retroflex productions was computed instead of both alveolar and retroflex productions. It was decided that a speaker's retroflex productions must be correctly identified by the two listeners at least 60% of the time in order to be judged capable of making

the place distinction and for his or her data to be included for further analysis. Figure 3.2 displays the percent accuracy of each speaker's retroflex productions.
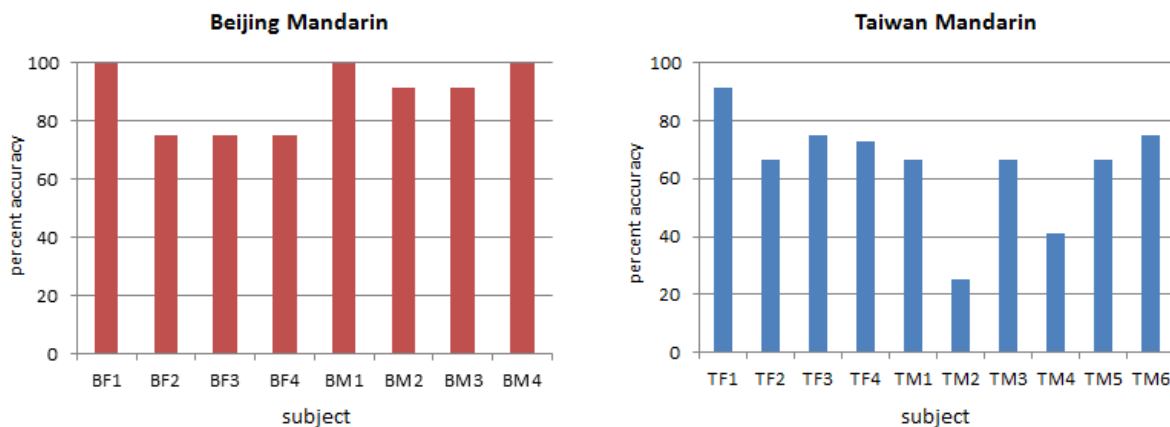
Figure 3.2: Percent accuracy of retroflex productions

For two of the male Taiwan Mandarin speakers (TM2 and TM4), less than 60% of their retroflex productions extracted from reading of *The North Wind and the Sun* were correctly identified[14]. Therefore, their data was not included for further analyses. This study will proceed with data from eight Beijing speakers' (4 male; 4 female) and eight Taiwan (4 male; 4 female) speakers.

3.5 Data analysis

3.5.1 Map task discourse

A sample conversation between the experimenter and the participant is given in this section based on a partial extraction of the experimenter's and the participant's maps (Figures 3.2a and 3.2b respectively). Again, both the target stimuli and the fillers on the maps are four-syllable

---

[14] The agreement rate between the two listeners was 73%, averaged from the identifications of all participants' screening materials.

words; the fillers differ from the target stimuli only by the first syllable. A sample target

stimulus-filler pair can be found in Table 3.2.

stimulus type	pinyin[15]	gloss
target	**zhǐ**dāo gōnglù.	Zhidao Highway
filler	**guā**dāo gōnglù	Guadao Highway

Table 3.2: Sample target stimulus-filler pair

Based on the experimenter's map (Figure 3.3a), *guādāo gōnglù* 'Guadao Highway' is the

place to the lower right of the starting point, instead of *zhǐdāo gōnglù* 'Zhidao Highway',

following the orientation of the subject's map (Figure 3.3b). Another target stimulus-filler pair,

zágài lèyuán 'Zagai Amusement Park' and *qígài lèyuán* 'Qigai Amusement Park', is also

oriented differently between the experimenter's map and the participant's map. In the first part of

a map task, the participant was asked to help orient the experimenter on the map. The

conversation was experimenter-oriented in that it was always the experimenter initiating the

question confirming a site's location. A sample experimenter-participant conversation is given in

Table 3.3 to show how focused productions were elicited through a map task conversation. The

conversation is given in Pinyin in the first line, followed by a word-for-word gloss and English

translation. The first syllables of the target stimuli and the fillers are in bold in the sample

conversation transcription.

---

[15] In Pinyin Romanization, the retroflex *zh*, *ch*, *sh* correspond to [tʂ, tʂʰ, ʂ] in IPA, and the alveolar *z*, *c*, *s* [ts, tsʰ, s].
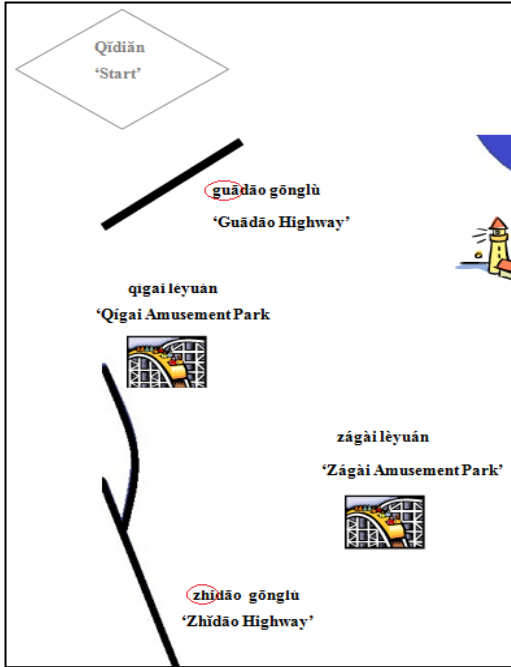
Figure 3.3a: A partial map for the experimenter   Figure 3.3b: A partial map for the participant
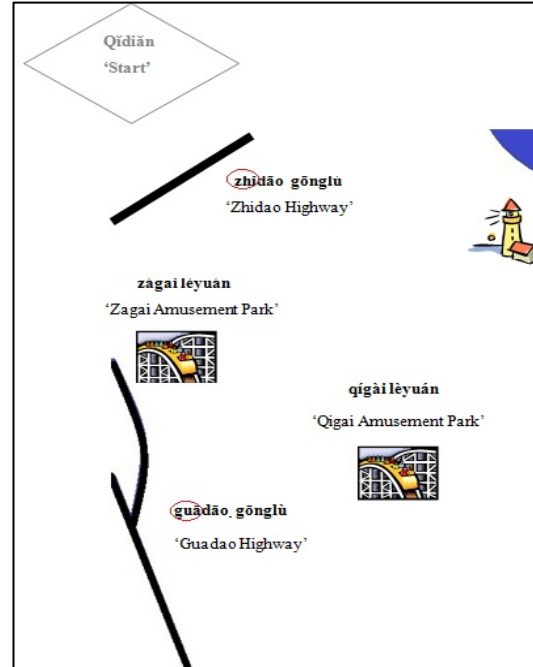
Experimenter:	qǐdiǎn de yòuxiàfāng shì **guā**dāo gōnglù mā? start possessive particle lower-right side be **guā**dāo Highway question particle 'Is to the lower right of the starting point **Guā**dāo Highway? '
Participant:	shì **zhǐ**dāo gōnglù. be **zhǐ**dāo Highway 'It's **Zhǐ**dāo Highway.'
Experimenter:	hǎo nà **zhǐ**dāo gōnglù de zuǒxiàfāng shì **qí**gài lèyuán mā ? ok then **zhǐ**dāo Highway possessive lower-left side be **qí**gài amusement park question particle 'Okay, then to the lower-left side of **Zhǐ**dāo Highway is **Qí**gài amusement park, right?'
Participant:	búshì, shì **zá**gài lèyuán no be **zá**gài amusement park 'No, it's **Zá**gài Amusement Park.'
Experimenter:	ǒu suǒyǐ guādāo gōnglù de yòushàngfāng cáishì **qí**gài Lèyuá duìbúduì oh so **guā**dāo highway possessive upper-right side be **qí**gài amusement park right 'Oh, so to the upper-right side of **Guā**dāo Highway is **Qí**gài Amusement park, right?'
Participant:	Méicuò Correct. 'That's correct.'

Table 3.3: A sample experimenter-participant conversation on orienting the locations on the map

Note that since no carrier phrase or sentence template was imposed on the participants as to

how directions should be corrected or given, it was not surprising to find different types of

utterances in the map tasks. Table 3.4 shows four response patterns that were most often used by

participants in response to the experimenter's question.

Experimenter	qǐdiǎn de yòuxiàfāng shì **guā**dāo gōnglù mā? start possessive particle lower-right side be **guā**dāo Highway question particle 'Is to the lower right of the starting point **Guā**dāo Highway? '
Participant TM1	**Zhǐ**dāo. **zhǐ**dāo '**Zhǐ**dāo'
Participant TF1	Shì **zhǐ**dāo gōnglù. be **zhǐ**dāo Highway 'It's **Zhǐ**dāo Highway.'
Participant BF1	Búshì, **zhǐ**dāo gōnglù. no **zhǐ**dāo Highway 'No, it's **Zhǐ**dāo Highway'
Participant BF2	Búshì, qǐdiǎn de yòuxiàfāng shì **zhǐ**dāo gōnglù no start possessive particle lower-right side be **zhǐ**dāo Highway 'No, to the lower right of the starting point is **Zhǐ**dāo Highway.'

Table 3.4: Various experimenter-participant discourse patterns on orienting the locations on the map

In the second part of a map task, the participant had to draw the route on the map and report

the route to the experimenter. Since the target stimuli were uttered in a non-corrective manner,

the alveolar and retroflex data elicited from this part of the map task were considered produced

under the non-focus condition. In reporting the route, the participants could be very brief or very

detailed in their discourse. A snippet of a route-giving discourse from three participants is given

in Table 3.5.

Experimenter	kěyǐ gàosù wǒ cóng qǐdiǎn dào zhōngdiǎn nǐ yào zěnme zǒu mā？ Can tell me from start to end you want how walk question particle 'Can you tell me the route from the start point to the end point?'
Participant TM1	shǒuxiān qù **zhǐ**dāo gōnglù, **zá**gài lèyuán first go **zhǐ**dāo highway **zá**gài amusement park 'First go to **Zhǐ**dāo Highway, **Zá**gài Amusement Park.
Participant BM1	shǒuxiān qù **zhǐ**dāo gōnglù, ránhòu dào **zá**gài lèyuán first go **zhǐ**dāo highway then arrive **zá**gài amusement park 'First go to **Zhǐ**dāo Highway, then go to **Zá**gài Amusement Park.
Participant BF2	shǒuxiān qù **zhǐ**dāo gōnglù ránhòu wǎngxià zǒu huì jīngguò **zá**gài lèyuán first go **zhǐ**dāo highway then downward walk will pass **zá**gài amusement park 'First go to **Zhǐ**dāo Highway, then go down and you'll pass by **Zá**gài Amusement Park.

Table 3.5: Various route-telling discourse patterns

3.5.2 Segmentation and acoustic measurement

During the audio recordings of the map tasks, only the participant was wearing a microphone. The experimenter's speech could be heard in the recording and was included in the transcription, however, his signals were too weak for acoustical analysis. After removing the experimenter's speech signals, the sound files along with their text transcription were sent to automatic segmentation using the PennPhonetics Lab Forced Aligner (Yuan & Liberman 2008). In order to measure the center of gravity of frication, all the target syllables were further re-segmented manually in Praat (Boersma & Weenink 2008) to separate the frication from the stop closure and aspiration in cases of affricates. The frication interval was measured from the onset of high-frequency noise to the onset of the silent gap before the following vowel, which is characteristic of aspiration. COG[16] was then obtained from the middle 30 ms of the frication interval, to which a high-pass filter set at 1000 Hz was applied to eliminate low frequency noise (Li et al. 2007).

Besides investigating how Mandarin alveolar and retroflex sibilants are realized across vowel contexts and between focus and non-focus conditions, this study is also interested in whether the phonological contrast is conditioned by these two factors. To that end, the phonological contrast was measured in terms of the difference between the COG values of the two sibilant categories (i.e., COG of alveolars minus that of retroflexes). Since the alveolar sibilants exhibit higher COG values than their retroflex counterparts, a greater positive COG difference is expected in case of contrast enhancement.

3.5.3 Realization of contrastive focus

To investigate the phonetic realizations of the Mandarin alveolar-retroflex contrast under different focus conditions, one first has to make sure the map tasks successfully elicited focused

---

[16] The formula used for the COG measurement is documented in the Praat manual.

vs. unfocused productions. It is generally agreed that the most reliable acoustic correlates of focus and stress-marking accentuation are F0, duration and intensity. Given the nature of map task conversations, it is difficult to decide on a baseline of intensity so as to compare focused and unfocused productions. In terms of the raw F0 parameter, since Mandarin is a tone language, tone information also needs to be included in evaluating intonational prominence (Shih 1988). In studying the acoustic manifestation of contrastive focus in Mandarin, Chen (2003) found duration lengthening to be a more robust effect than fundamental frequency change. More specifically, Chen compared data elicited with different degrees of emphasis (i.e., more-focus, on-focus and neutral-focus conditions) in terms of durational adjustments and F0 changes. Her results show that while the amount of durational lengthening increases with the degree of emphasis, changes in F0 pattern exhibited a ceiling effect, in that there is a significant difference between the F0 contours of on-focus and neutral-focus lexical tone productions but no significant difference between the F0 contours of on-focus and more-focus productions. With all of these considerations taken together, I used the duration measure of the target syllables to determine whether contrastive focus was implemented. Focused target syllables are expected to be significantly longer than unfocused ones.

A mixed-design ANOVA was run on the duration data, with FOCUS as a within-subject variable and DIALECT as a between-subject variable. The analysis revealed a significant main effect for FOCUS ($F(1, 478)=183.314$; $p<.001$) and DIALECT ($F(1, 478)=8.056$; $p< .001$), as well as a significant interaction effect ($F(1, 478)=18.571$; $p< .001$). Post-hoc comparison tests show that overall syllable durations of Beijing speakers' productions were significantly shorter than that of Taiwan speakers'. However, in both groups, the focused productions had significantly longer syllable duration than the unfocused ones. For information on individual

durational patterns, see Figure 3.4 (Beijing speakers) and Figure 3.5 (Taiwan speakers) where each speaker's focused and unfocused data were plotted side by side. Since contrastive focus was consistently manifested in syllable duration, I proceed with the spectral analysis regarding the effect of focus on alveolar-retroflex realizations.
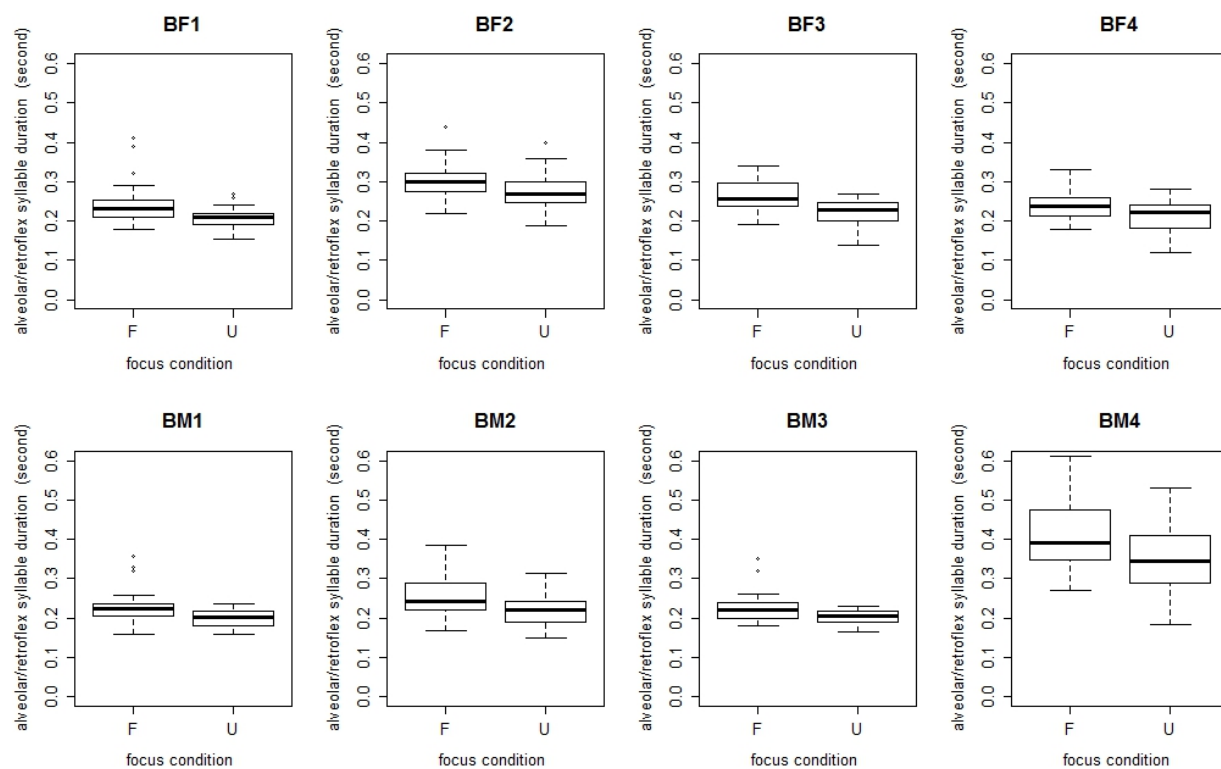
Figure 3.4: Duration of focused (F) vs. unfocused (U) alveolar/retroflex syllables for Beijing Mandarin speakers. (BF1-4 are Beijing female speakers; BM1-4 are Beijing male speakers)
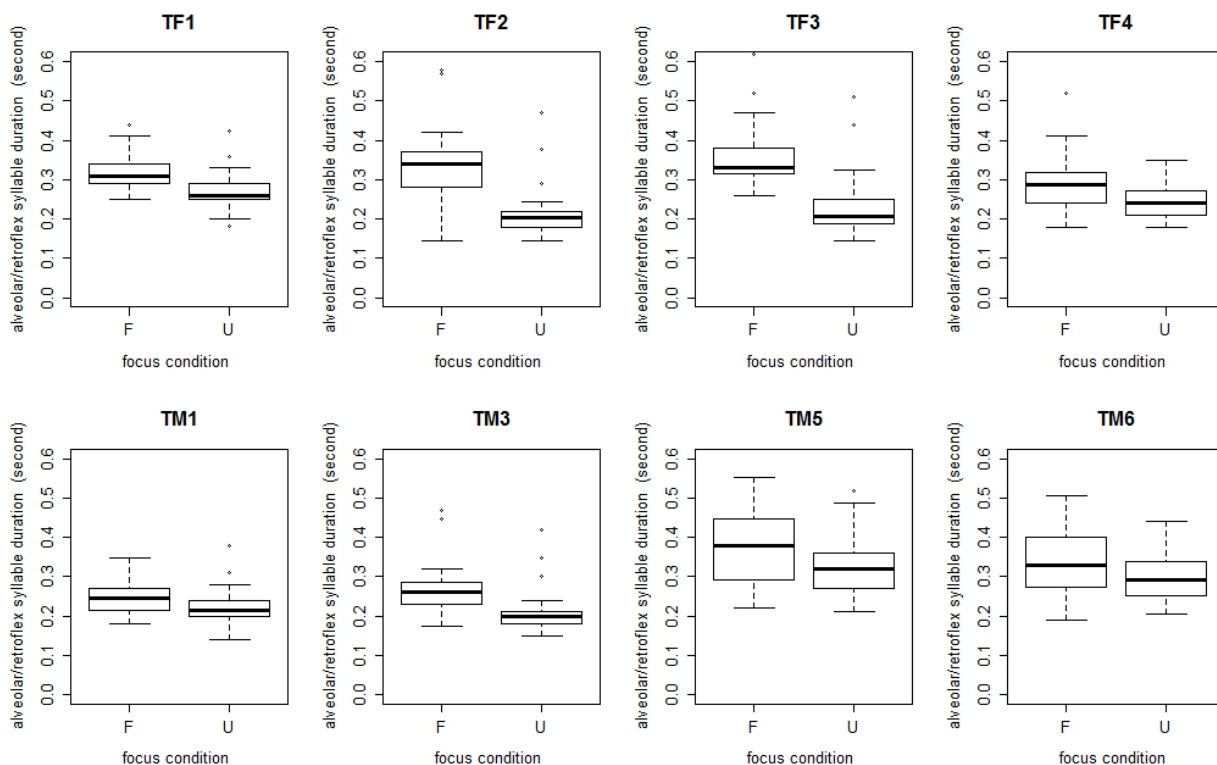
Figure 3.5: Duration of focused (F) vs. unfocused (U) alveolar/retroflex syllables for Taiwan Mandarin speakers. (TF1-4 are Taiwan female speakers; TM1-4 are Taiwan male speakers)

3.5.4 Statistical analysis

Table 3.5 presents the descriptive statistics of Beijing and Taiwan speakers' mean COG of alveolar and retroflex productions, collapsing vowel contexts and focus conditions. It is shown that 1) Beijing speakers' retroflex productions yielded a noticeable lower COG than Taiwan speakers', and 2) Taiwan speakers' data showed a smaller COG difference between the alveolar and retroflex than Beijing speakers'. For both Beijing and Taiwan speakers, retroflex productions had lower mean COG and greater variance than alveolar productions.

dialect place	Alveolar	Retroflex
Beijing	9115.6 (1144.8)	5160.7 (1397.9)
Taiwan	8757.6 (1388.1)	6224.1 (1429.5)

Table 3.6: Mean COG (sd) of alveolar and retroflex productions for Beijing vs. Taiwan Mandarin

A mixed-design ANOVA was run on the COG data, with FOCUS (focused v.s. unfocused condition), VOWEL (/a, i, u/ contexts) and PLACE (alveolar vs. retroflex articulations) as within-subject factors, and DIALECT (Beijing vs. Taiwan Mandarin) as a between-subject factor. The statistical analysis revealed a significant main effect for PLACE ($F(1,78)=718.3$; $p<.0001$), VOWEL ($F(2,78)=37.7$; $p<.0001$) and DIALECT ($F(1,78)=4.5$; $p=0.038$). There was a significant interaction effect between PLACE and FOCUS ($F(1,78)=10.3$; $p < 0.01$), with higher COG for focused alveolar productions than unfocused ones. A significant 3-way interaction between PLACE, VOWEL and DIALECT was found ($F(2,78)=11.4$; $p < 0.001$) in addition to significant 2-way interaction effects for PLACE*VOWEL ($F(2,78)=28.9$; $p < 0.001$), PLACE*DIALECT ($F(1,78)=34.5$; $p < 0.001$) and VOWEL*DIALECT ($F(2,78)=37.3$; $p < 0.001$). A significant 3-way interaction indicates that one factor depends on the levels of the other two factors. Therefore, 2-way interactions were not interpreted; post-hoc comparison tests were conducted to explore only the 3-way interaction. The results showed that 1) in both Beijing Mandarin and Taiwan Mandarin, the alveolars had significantly higher COG than the retroflexes across vowels. 2) In both dialects, alveolar productions in the /u/ context had significantly lower COG than in the other two contexts. 3) In Beijing Mandarin, retroflex productions in the /u/ context had significantly higher COG than in the other two contexts. In contrast, retroflex productions in the /u/ context in Taiwan Mandarin had significantly lower COG than in the other two contexts. 4) Realizations of retroflexion in Taiwan Mandarin varied across vowels, whereas those in Beijing Mandarin were different only between rounded and non-rounded vowel contexts. 5) Alveolar productions in the /u/ context in Taiwan Mandarin had significantly lower COG than that in Beijing Mandarin. 6) Retroflex productions in the /a, i/ contexts in Beijing Mandarin had significantly lower COG than that in Taiwan Mandarin. In sum, alveolar sibilants had higher

COG than retroflex ones. Vowel and dialect variably affected sibilant realizations. As for the focus effect, while alveolar and retroflex realizations differed significantly regardless of focus conditions, focused alveolars did have higher COG than unfocused alveolars.
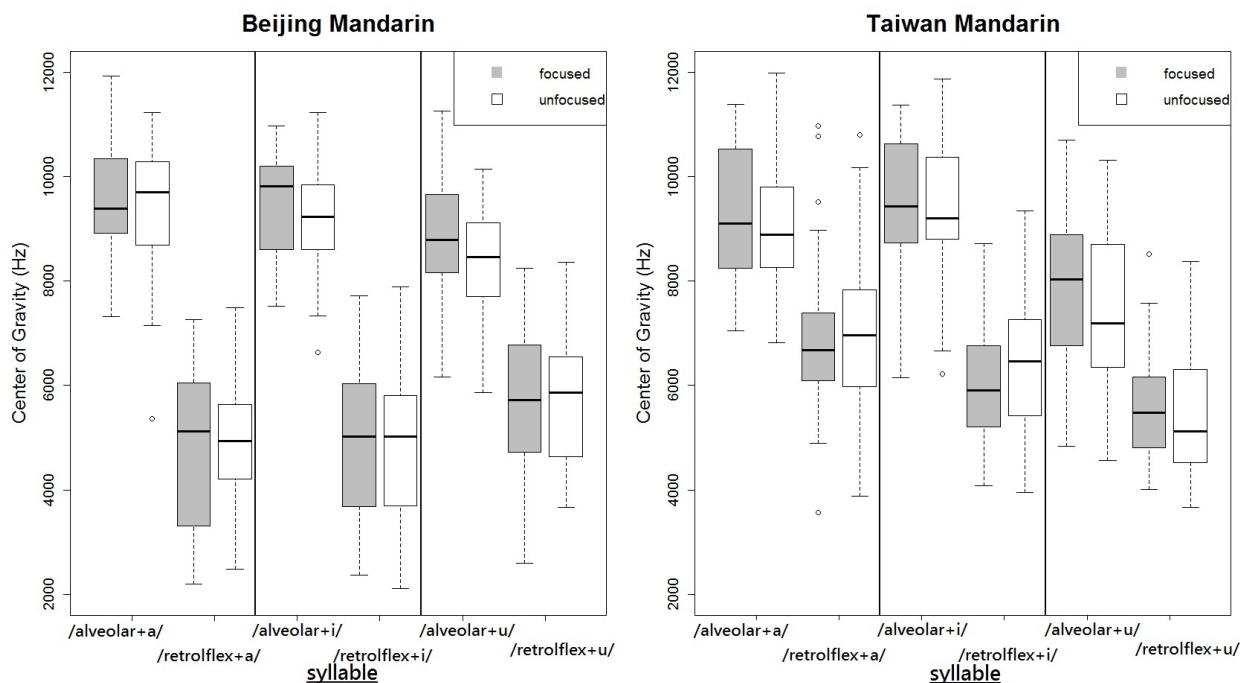
Figure 3.6: Sibilant realizations varied by vowel, place and dialect

After looking at the realizations of the two sibilant categories, it was examined whether vowel contexts, focus, and dialect would affect the realizations of the alveolar-retroflex contrast. A mixed-design ANOVA was run on spectral distance (i.e., each participant's mean COG of alveolars minus that of retroflexes), with FOCUS (focused vs. unfocused condition) and VOWEL (/a, i, u/ contexts) as within-subject factors, and DIALECT (Beijing vs. Taiwan Mandarin) as a between-subject factor. The statistical analysis revealed a significant main effect for VOWEL ($F_{(2,14)}$=10.865; $p<.0001$) and DIALECT ($F_{(1,14)}$=7.9; $p=0.014$). There was a marginally significant interaction effect between VOWEL and DIALECT ($F_{(2,14)}$=3.3; $p=0.051$). Post-hoc comparison tests showed that 1) The place contrast was realized differently

54

between Beijing Mandarin and Taiwan Mandarin only in the /a/ context. 2) For Beijing

Mandarin speakers, the contrast in the /u/ context was smaller than in the other two contexts. For

Taiwan Mandarin speakers, the contrast in the /i/ context was significantly larger than the other

two contexts. In sum, the alveolar-retroflex contrast was realized differently across vowel. The

/u/ context was where the smallest place contrast was observed for both dialects. Beijing

speakers produced a significantly larger alveolar-retroflex contrast than Taiwan speakers only in

the /a/ context.
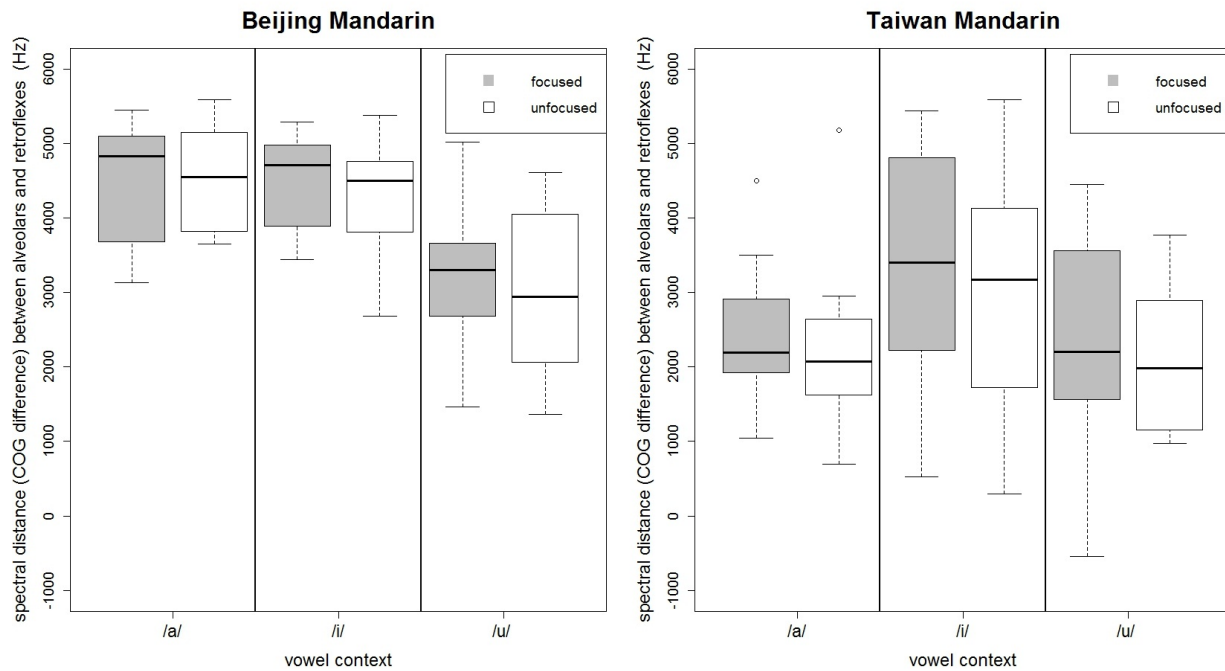
Figure 3.7: The realizations of Mandarin alveolar-retroflex contrast varied by vowel and dialect

3.6 Discussion

The statistical analysis of the sibilant data indicates that place of articulation (alveolar vs.

retroflex), vowel context (the /a, i, u/ vowels that follow the alveolar/retroflex sibilants) and

dialect (Beijing Mandarin vs. Taiwan Mandarin) all affect sibilant realizations. Upon exploration

of the interaction effects, it was found that both alveolar and retroflex sibilants were subject to

the coarticulatory effect from the rounded vowel. The two dialects differed on how the realizations of the alveolar and retroflex sibilants were conditioned by vowel. Particularly, /alveolar+u/ productions in Taiwan Mandarin had significantly lower COG than that in Beijing Mandarin. Beijing speakers realized retroflex sibilants with significantly lower COG in the /a, i/ contexts than Taiwan speakers.

In terms of the focus effect on sibilant realization, only alveolars sibilants were realized differently in terms of the COG measure under focus (see Figure 3.5). The insignificance of the main effect for focus suggests that 1) focal prominence was not consistently or uniformly manifested in the realizations of both sibilant categories, and 2) focal prominence was not uniformly implemented within and across speakers. Because each experimental condition had only five data points for each speaker, statistical analysis on the individual data was not included in §3.3, as robust statistical power may not be achieved. The plots of individual speaker data are displayed in Appendix F. In general, the distributions of the COG values for each sibilant category grouped by focus overlapped to varying degrees for all speakers. Prosodic strengthening may be present in some speakers' data, however, it never occurred across all vowel contexts. For example, speakers BF3 and BM1 had different retroflex realizations in the /u/ context under focus, but not in the other two contexts (see Appendix F). Prosodic strengthening also rarely occurred in both phonemes; in most cases, only alveolars or retroflexes in selective vowel contexts were realized differently in response to focus. In addition, the direction of strengthening is not predictable, even within a speaker. In sporadic cases like speaker BF3, her focused /alveolar+a/ productions had a noticeably higher COG than the unfocused counterparts, but her focused /alveolar+i/ productions had a lower COG than the unfocused counterparts (see Appendix F). Overall, considerable within-speaker variation was observed in the way sibilants

were realized across focus conditions. Great across-speaker variation was also present such that no prosodic strengthening patterns in sibilant realizations could distinguish between Beijing and Taiwan speakers.

For the alveolar-retroflex contrast, the statistical results indicate that the magnitude of the place contrast varied by vowel context and dialect. Both Beijing and Taiwan Mandarin speakers produced a smaller spectral contrast in the /u/ context, as coarticulation from the rounded vowel lowers the COG in alveolar realizations and in turn decreases the spectral distance between the two phonemes. On the other hand, Beijing speakers made a significantly larger contrast in the /a/ context than their Taiwan counterparts. As the effects of anticipatory coarticulation on the spectral characteristics of English coronal fricatives were the smallest when they were followed by the vowel /a/ (Soli 1981), it is reasoned that the small place contrast in the /a/ context in Taiwan Mandarin may not be due to the same coarticulatory effect the /u/ vowel exerts. But rather, some confusion over the place contrast in the /a/ context is suggested for Taiwan Mandarin. Within Taiwan speakers, a larger contrast was made in the /i/ context compared to the other two vowel contexts, indicating that the alveolar-retroflex distinction was less confusable when followed by /i/. It may be that a considerable vowel quality difference between [ɿ] and [ʅ] (a significant difference in F2 and F3 between the two vowels was found in Lee 2011) facilitated the distinction between the two places of articulation.

It should be noted that there are some negative COG values for the Taiwan Mandarin group, as seen in Figure 3.7. Since the COG difference was calculated from the COG values of the alveolar sibilants minus those of the retroflex counterparts, negative values of COG difference would suggest confusion over the place distinction or hypercorrection for some Taiwan Mandarin speakers. Particularly, speakers TM1 and TF2 may stand out as having neutralized the

57

place distinction in most vowel contexts based on the observations of greatly overlapped COG distributions of the two sibilant categories (see Appendix F). Upon closer examination of their sound files, many of TM1's retroflex productions sounded like alveolar ones, and most of TF2's /retroflex+a/ and /retroflex+u/ productions seemed to be replaced by the alveolar counterparts. While both speakers had passed the screening test and were considered able to make the place distinction, their production in the map tasks may not strongly support the presence of such a phonological contrast. It is speculated that TM1's and TF2's alveolar-retroflex contrast was partially lexicalized. They may have successfully encoded certain words with a retroflex, but in cases of uncertainty they defaulted to the alveolars. In future follow-up studies on lexicalized retroflexing in Taiwan Mandarin, the target syllable in the map tasks can be substituted with their homophones to see if the speakers would realize the same syllable differently.

Now, in terms of the focus effect on the alveolar-retroflex contrast, no statistically different realizations in response to focal prominence were found. That is, focus did not lead to an enhancement of the contrast between alveolar and retroflex sibilants in the two dialects of Mandarin Chinese studied. This result was in line with Silbert and De Jong's (2008) findings on the place contrast for English fricatives. They reported that while focus generally increased fricative and syllable durations as well as noise power, it did not result in enhanced spectral features (i.e., the four spectral moment measurements) of the phonological contrast. Therefore, focus does not always induce hyperarticulation and give rise to an enlarged phonological contrast. In fact, if focus affects both categories of a phonological contrast in the same way (i.e., a relevant acoustic measure increases or decreases uniformly), no larger contrast will be observed. This uniform strengthening without place enhancement was reported in Cole et al.'s (2007) study of prosodic effects on English /p, t, k/ contrast. They found that accentuation resulted in a uniform

increase in the acoustic measures of VOT and closure duration, such that no increased acoustic distinctiveness was observed under accent. As for the Mandarin place contrast under study, the absence of contrast enhancement was not a result of uniform strengthening of alveolar and retroflex sibilants, as focus only had a significant effect on alveolar realizations (i.e., a higher COG). It could be that given the variable retroflex realizations under focus, a strengthened articulation for the alveolars was not sufficient to significantly enlarge the separation between the two categories. The question then arises as to why the focus effect was manifested in syllable duration (see §3.5.3) but did not result in contrast enhancement in the spectral domain. Two possible explanations are provided here. First, speech is multidimensional; not all acoustic parameters are subject to change under prosodic strengthening (see §3.2.1 for the review of relevant studies). In the case of fricatives, "longer and louder fricatives likely offer more and better opportunities to hear the relatively static phonetic cues for features like voicing and place of articulation." (Silbert & De Jong 2008; p. 2778). Second, Mandarin has a 3-way coronal contrast within a limited articulatory space. High density of contrasting phonemes may inhibit the effects of focus on spectral place cues so as to maintain acoustic stability of each phoneme as well as a sufficient auditory distinction between the phonemes.

As reviewed in §3.2.1, Chuang and Fon (2010) reported strengthened and enlarged alveolar-retroflex contrast in accentuated condition for some of their Taiwan Mandarin speakers. This study, however, did not find a significant effect of contrastive focus on the same place contrast for either the Beijing or Taiwan participants. Differences between the findings of Chuang and Fon and this study may be due to differences in speech materials and analysis. Chuang and Fon used conversational speech and did not control for vowel contexts and the number of tokens of each phoneme category. They were interested in individual differences and therefore pooled the

alveolar vs. retroflex data across vowel contexts for each speaker in analysis. The current study used map tasks to elicit data, where all alveolar and retroflex stimuli were controlled for lexical frequency and were balanced across vowel contexts. Given the emphasis on cross-dialectal variation, this study grouped speakers' data by their native dialect. That said, this study does not deny potential contrast enhancement through prosodic strengthening in some Mandarin speakers, but rather, the results of this study indicate that prosodic strengthening may not have a clear or stable effect on the Mandarin alveolar-retroflex contrast for most speakers across all vowel contexts.

3.7 Conclusion

This study investigates whether Mandarin sibilants are realized differently across vowel contexts and under different focus conditions in Beijing and Taiwan Mandarin. The results show that sibilant productions were variably affected by the vowel in both dialects. The two dialects can be best distinguished in terms of 1) the alveolar realizations in the /u/ context, as the rounded vowel had a stronger coarticulatory effect on the alveolar production in Taiwan Mandarin, 2) the retroflex realizations in the /a, i/ contexts, as Beijing Mandarin speakers exhibited a stronger degree of retroflexion than Taiwan Mandarin in most vowel contexts except for /u/, and 3) the alveolar-retroflex contrast in the /a/ context, where Beijing speakers produced a significantly larger place contrast than their Taiwan counterparts. Overall, the /u/ vowel context was where the alveolar-retroflex contrast was the least distinct. Therefore, the /u/ vowel context is expected to be the environment where alveolar-retroflex neutralization is more likely to occur.

This study confirms that prosodic prominence, or contrastive focus, does not necessarily lead to an enhancement of a phonological contrast. There was no interaction effect between the vowel context and focal prominence on the sibilant realizations or the alveolar-retroflex contrast.

Therefore, no dialect-specific generalizations can be made in terms of the focus effects on the

Mandarin alveolar-retroflex contrast. To conclude, the characterization of Beijing and Taiwan

Mandarin alveolar-retroflex contrast varies by vowel and by how each contrasting phoneme is

realized in a particular vowel context.

CHAPTER 4: PERCEPTION STUDY

4.1 Introduction

Chapter 3 reported that Beijing Mandarin speakers produced a greater alveolar-retroflex contrast than Taiwan Mandarin speakers in the /a/ context, and that the /u/ context was where both dialects exhibited a lesser degree of the alveolar-retroflex distinction. According to Sumner and Samuel (2009), dialect production is not always representative of dialect perception. This chapter therefore investigates whether the perception of the Mandarin alveolar-retroflex contrast also patterns differently as a result of dialectal differences and whether the vowel context influences the perceptual patterns for both dialects.

The chapter is organized as follows. First, the chapter reviews categorical perception of phonological contrasts (§4.2). Next, exemplar-based models of speech perception that tap within-phoneme gradient perception are reviewed (§4.3). In §4.4, the acoustic cue to the discrimination of place contrast for fricatives will be discussed. The chapter will then review studies on context-dependent perception, as the vowel context is controlled as a between-subject variable for this study (§4.5). Next, based on findings from previous studies, hypotheses and research questions are formulated (§4.6). In §4.7, an experimental overview with a special focus on stimulus construction is given. To test the hypotheses and address the research questions, the chapter will report findings from two experiments (§4.8 and §4.9). Finally, the chapter will discuss perceptual patterns across dialect and vowel context based on the results of the two experiments (§4.10).

4.2 Categorical perception and phonological contrasts

How continuous speech signals are classified into categories is a fundamental issue in the perception of phonological contrasts. When varying acoustic signals of speech are mapped onto

62

discrete phonological categories of a language, the perceptual discontinuity in the identification and discrimination of speech sounds along an acoustic continuum is known as categorical perception. In Liberman et al.'s (1957) study, they had subjects identify synthesized /a/ with the F2 frequency changed in the vowel onset region to simulate transition patterns from a preceding consonant. In their experiment, 13 stimuli were used where the F2 transition started with different frequency values varying by equal interval, yet the listeners were only able to identify three speech sounds /ba, da, ga/. The shifts from the identification response of /ba/ to /da/ to /ga/ were found to be very abrupt, and where the shift occurred along the F2 continuum was considered a phoneme boundary. Liberman et al. also asked their subjects to discriminate among the stimuli, and it was found that the subjects were able to better discriminate between the stimuli that lie on either side of a phoneme boundary than those that fall within the same phoneme category.

In conducting categorical perception research, it is critical to choose the relevant acoustic dimensions along which listeners may perceive different categories. While change in F2 transition of a vowel would allow listeners to identify the place of articulation of the consonants that are not even present in the stimuli, voicing of stops was found to vary as a function of voice onset time (VOT; the time lag between the release of the stop and the vocalic onset) (e.g., Lisker & Abramson 1967; Repp & Lin 1989; Wood 1976). Take English for example, unaspirated voiced stop consonants like /b/ have a shorter VOT, whereas their aspirated voiceless counterparts like /p/ have a longer VOT. Therefore, a discrimination task on voicing along a VOT continuum would show an increase of correct discriminations at the /b-p/ boundary, whereas discrimination performance between sounds within the same category would be relatively poor.

4.3 Gradient perception and exemplars

Exemplar-based models of speech perception are a framework that explains how detailed information from the speech signal is stored as representation of input in the mental lexicon (Goldinger 1996; Johnson 1997; Pierrehumbert 2001). An exemplar view to sound categorization is that every time a listener encounters a new variant (e.g., a coarticulated/reduced form, or variant produced by a speaker of a different gender, dialect) of a phonological category, it is compared to previously encountered instances (i.e., exemplars) of that category. The exemplars are organized by their similarity to previous instances and frequency of occurrences in a given parameter space. Taken together, in the exemplar model, categories are defined by a cluster of exemplars and can capture the plasticity of linguistic knowledge at an individual and a group level.

Research on the internal structure of phoneme categories essentially taps exemplar-based speech categorization, with a focus on whether the subphonemic variants are perceptually equivalent. For example, Pisoni and Tash (1974) reported an increase of reaction times in phoneme identification and same-different discrimination tasks as the stimuli from the /ba-pa/ continuum approach the category boundary[17]. McMurrray et al. (2002) monitored eye movements while participants listened to stimuli from a 9-step /b/-/p/ VOT continuum and looked at pictures that contain /b/ and /p/-initial objects as well as non-/b, p/-initial objects. Their results show that fine-grained acoustic differences in VOT have significant effects on lexical access; the activation of a lexical competitor increases when VOT approaches the category boundary. Miller (1994) and Allen and Miller (2001) investigated the gradient phonetic structure by assessing the within-category goodness. Listeners were asked to rate the stimuli with respect

---

[17] The crossover point in identification, which corresponds to about +30 ms VOT, was where Pisoni and Tash concluded the voicing boundary to be.

to the subjective goodness of English /p/ on a 10-point scale. It was found that higher ratings were given to more prototypical pronunciations. Their data (see Figure 4.1) show an initial increase in goodness rating scores as VOT increases. As the /p/ becomes excessively aspirated (i.e., having exaggerated VOT), a gradual decline in the goodness rating was observed. Therefore, it was concluded that members within a category vary systematically in their goodness, with only a certain range of members judged as the best exemplars.
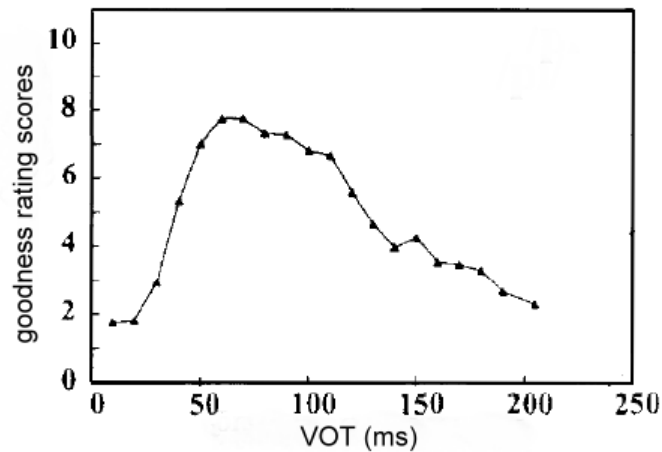
Figure 4.1: Goodness rating of /pi/ as a function of VOT (adapted from Miller 1994)

In sum, a variety of evidence has shown that listeners are sensitive to across-category as well as within-category variation. In gradient perception, some members within a category are perceived as better exemplars than others, suggesting that phonetic categories are much richer in their representational forms than what is revealed by just the category labels.

4.4 Perception of place contrast for coronal fricatives

In distinguishing the place of articulation of coronal fricatives, the location of acoustic energy in the spectrum is the most often noted acoustic cue (e.g., Harris 1958; Heinz & Stevens

1961; Hughes & Halle 1956; Shadle 1985). To take English coronal fricatives for example, most energy in the noise spectrum of /s/ is located above 4000 Hz. In /ʃ/, the spectral peak occurs at about 3.5 Hz. Figure 4.4 displays the AI-gram[18] of /sa/ and /ʃa/ produced by a female native speaker of American English. The x-axis of the AI-grams marks the acoustic event in time (in centisecond) and the y-axis marks frequency (kHz) on the critical band scale. The noise band from around 10 to 25 centiseconds in the AI-gram indicates the fricative. While the alveolar /s/ has high-frequency signals above 4kHz, this frication region (the framed part in the AI-gram of /ʃa/) is also present in /ʃa/. Li and colleagues (Li & Allen 2011; Li, Trevino, Menon & Allen 2012) had their subjects listen to the /ʃa/ tokens with the frication noise above 4kHz removed, yet percepts of /ʃa/ were still reported. On the other hand, when the frication region between 2-4 kHz was removed from /ʃa/ by a time-varying high-pass filter, it was perceived as identical to the /sa/ token produced by the same talker. Modifying the frequency of the lowest bound of the wideband noise while leaving the rest of the frication intact results in the perception of a different fricative. Li and colleagues therefore concluded that the necessary perceptual cue for fricative place of articulation lies in the frequency of the lowest bound of the wideband noise (which will be termed "the cutoff frequency" hereafter). Raising the cutoff frequency of /ʃ/ by removing the frication region between 2-4 kHz will convert /ʃ/ into /s/.

---

[18] AI-gram (Li & Allen 2011; Lobdell 2009; Régnier & Allen 2008) is a time-frequency representation. AI-gram integrates Fletcher's (1950) Articulation Index (AI) model of speech intelligibility and his critical-band auditory model, which can better evaluate the contribution of speech components to particularly consonant perception.
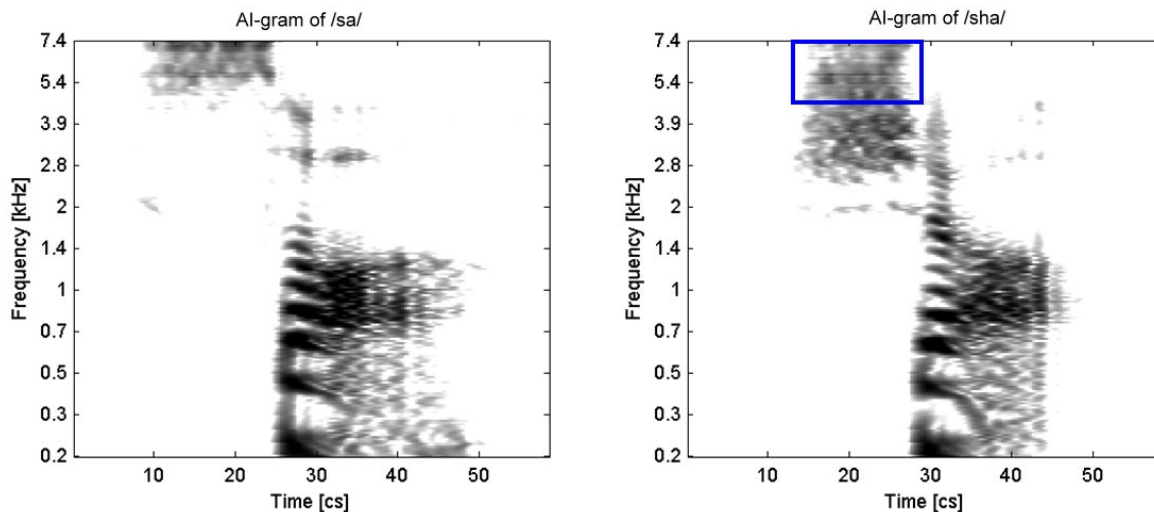
Figure 4.2: AI-grams of /sa/ and /ʃa/ in English (adapted from Li & Allen 2011)

The most important acoustic attribute that distinguishes Mandarin /s/ and /ʂ/ lies in the

frequency domain as well. In Mandarin, the alveolar fricative /s/ has major spectral prominence

in the F4-F5 range, and the spectral noise band frequencies for /ʂ/ are centered between 2-3 kHz

(Svantesson 1986; Stevens et al. 2004). Early works by Wu (1963) and Wu and Lin (1989)

argued that the primary acoustic cue distinguishing Mandarin alveolars from retroflexes is the

position of the lowest spectral prominence. In perception, Jeng (2009) reported that distinction

between Mandarin alveolar and retroflex sibilants is correlated with the distribution of energy

across the noise spectrum, as captured with the COG measure. Parallel to the English /s-ʃ/

spectral difference in Figure 4.2, Figure 4.3 displays the AI-grams of Mandarin /sa/ and /ʂa/ on

the upper panel, as well as the AI-grams of Mandarin /su/ and /ʂu/ on the lower panel. The

horizontal line on each AI-gram indicates the cut-off point of the noise distribution. The

alveolars have higher cut-off frequencies than the retroflexes in both the /a/ and /u/ contexts.

Moreover, the spectral distance (i.e., the distance between the two horizontal lines) between the

alveolar and retroflex in the /a/ context is noticeably greater than that in the /u/ context (2.8 kHz

67

between /sa/ and /ʂa/; 1.9 kHz between /su/ and /ʂu/). The smaller spectral distance between /s/ and /ʂ/ as a result of the coarticulation from the following rounded vowel is consistent with the production data reported in Chapter 3. In the next section, I will further review research on the influence of phonetic contexts on speech perception.
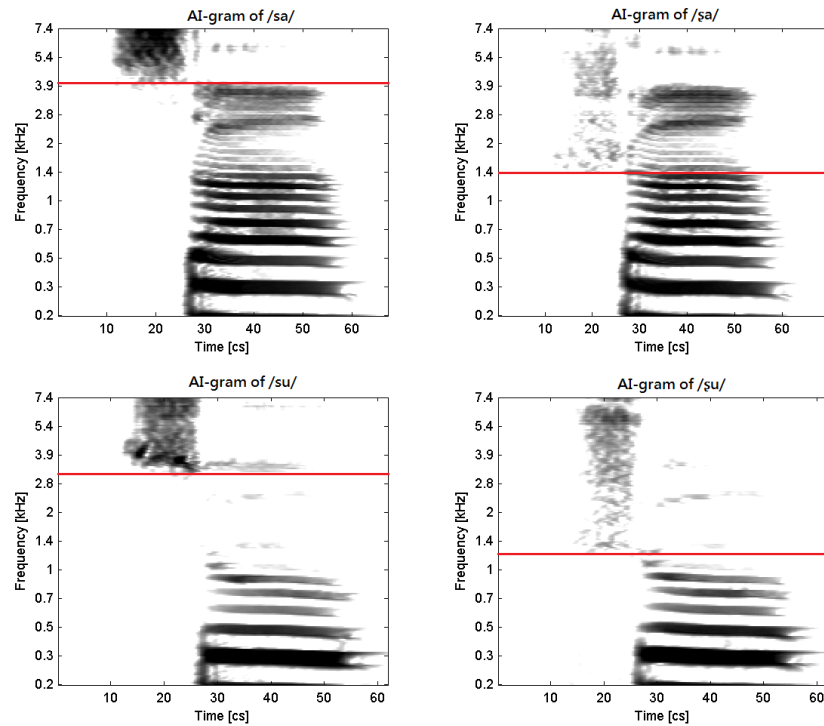
Figure 4.3: AI-grams of /sa/, /ʂa/ (upper panel); /su/and /ʂu/ in Mandarin (lower panel)

4.5 Context-dependent speech perception

Acoustic information characteristic of a given phonetic segment varies extensively as a function of contextual factors such as phonetic environment (e.g., Whalen 1981 and Soli 1981 on the effects of vowel on fricative consonants) and speech rate (e.g., Summerfield 1981 on the effects of speech rate on stop voicing contrast). In perception, research has shown that listeners are remarkably attuned to context-conditioned variation. For example, vowel coarticulation is compensated for in consonant perception (Whalen 1981; Mann & Repp 1980). Reducing the

speech rate of an immediately preceding syllable causes a greater percentage of the target stop consonants to be identified as voiced (Summerfield 1981).

Much of the evidence for contextual effects in perception is drawn from studies examining the location of category boundaries. The general finding is that boundary locations are flexible; context-induced variation results in a systematic change in the boundary location in perception (Repp & Liberman 1987). In the production of fricative consonants, the noise spectrum in the frication has been reported to vary with the following vowel. For example, Carney and Moll (1971) point out that anticipatory rounding would result in a lowering of the noise spectrum for English fricatives. The same acoustic context dependency was also found in the production of Japanese fricatives (Fujisaki & Kunisaki 1978). While production is subject to contextual variation, perceptual compensation for this variation has been reported in that category boundaries shift as a function of vowel context. Whalen (1981) had participants listen to fricative noise tokens synthesized to represent a /ʃ-s/ continuum followed by either /i/ or /u/ and asked subjects in a force-choice experiment to label them as "s" or "sh". His data indicated a lower /ʃ-s/ boundary (in Hz) for /u/ than for /i/. Mann and Repp (1980) replicated Whalen's study by using synthetic fricative noises from a /ʃ-s/ continuum followed by /a/ or /u/. They also found the phoneme boundary to shift toward lower noise frequencies in the /u/ context. In addition, the results of their identification task showed that listeners perceived more instances of /s/ in the context of /u/, especially in case of a fricative noise ambiguous between /s/ and /ʃ/. With that, they suggest that listeners are able to perceptually compensate for these coarticulatory effects.

Besides being reflected in the change of categorical boundary location, contextual effects have also been shown to alter the internal structure of a category in perception. Kawasaki (1986) had her English-speaking subjects rate the nasality of English /ɪ, u, a/ vowels in the /m_m/

69

context. The amplitude of the nasal consonant was attenuated in five steps to create the nasal vs. oral (or less nasal, as the amplitude of /m/ was more attenuated) contexts. She found that subjects gave a higher nasality rating for the nasalized vowels that occurred in the oral context than in the nasal context. That is, the same nasalized vowels were perceived as more nasalized when they were flanked by weaker nasal consonants. Another line of research investigating the context-dependent perceptual structure of phoneme categories lies in voicing contrast, as specified by VOT, varying as a function of speech rate (Miller & Volaitis 1989; Volaitis & Miller 1992; Wayland, Miller, & Volaitis 1994). It was found that a change in target-syllable rate not only alters the voiced-voiceless boundary location, but also changes the within-category structure. When speech rate was slowed, the goodness function for /p/ was found to shift toward longer VOT values, and a wider range of stimuli were considered as good exemplars of /p/.

The above studies indicate that listeners are aware of various types of systematic and context-dependent variation and are able to adjust the perceptual threshold along a given acoustic dimension. In this regard, contextual effects on across-category and within-category variation should both be taken into consideration in perception studies.

4.6 Purposes and research questions

Chapter 3 shows phonetic variation of a phonological contrast as a result of vowel context and dialectal background. It is the goal of this study to investigate whether variable realizations observed in production are reflected in perception such that speakers of a different dialectal background would perceive a phonological contrast differently in a given vowel context.

Different acoustic realizations of the alveolar-retroflex contrast between Beijing and Taiwan Mandarin motivated a hypothesis that the category boundary will lie in different places along the

spectral continuum. In addition, in Taiwan Mandarin, retroflex productions have been found to be acoustically less retroflexed than in Beijing Mandarin (Chang 2011) and the magnitude of the alveolar-retroflex contrast is smaller as well (see Chapter 3). Therefore Taiwan Mandarin listeners are hypothesized to have different best category exemplars from their Beijing counterparts. Finally, given that both Beijing and Taiwan listeners produced a smaller place contrast in the /u/ context, the category boundary location and the internal structure are hypothesized to change across vowel context. Two experiments were designed to verify the aforementioned hypotheses. Experiment 1 was conducted to locate the category boundary between Mandarin alveolars and retroflexes. Experiment 2 looked into the internal structure of Mandarin alveolars and retroflexes.

Four research questions will be addressed:

1) Do Beijing and Taiwan Mandarin speakers share the same alveolar-retroflex boundary in perception?

2) Does the location of the category boundary vary across vowels?

3) Do Beijing and Taiwan Mandarin speakers share the same best exemplars for the alveolar and retroflex categories?

4) Does the location of a category's best exemplars vary across vowel?

4.7 Methodological overview

This study aimed at locating the category boundary as well as investigating the graded internal structure of Mandarin alveolars and retroflexes. To this end, two perception experiments,

a discrimination task and a goodness rating task, were conducted using E-Prime (v2.0; Psychological Software Tools, Pittsburgh, PA).

An ABX paradigm, where the stimuli are arranged in an ABX triad, was used in the discrimination task. More specifically, stimulus A is first paired with stimulus B, which is at a fixed number of steps from stimulus A on a given acoustic continuum. Stimulus X is either identical to stimulus A or stimulus B, thereby yielding two triad combinations (ABB, ABA). If the order of stimulus presentation is counterbalanced, the two triads can be expanded to four (ABB, ABA, BAA, BAB). Within each ABX triad, the three stimuli are played successively and spaced at a fixed time interval. The participants then have to determine whether stimulus X is the same as stimulus A or B. In the goodness rating task, the participants were instructed to rate the subjective goodness of an alveolar or retroflex syllable on a Visual Analog Scale, a continuous scale without discrete references or numerical values.

4.7.1 Participants

The participants for this study were 30 Beijing Mandarin (born and raised in Beijing or its vicinity[19]) and 30 Taiwan Mandarin native speakers (born and raised in Taiwan). They were between 18 and 35 years of age. Figure 4.4 gives further background information of the participants. None of the participants reported any past or present speech or hearing disorders.

---

[19] Seven speakers were from Tianjin or other cities near Beijing. None of these cities is more than 100 miles away from Beijing.

Beijing Mandarin

City of origin

23%
77%

■ Beijing
■ Others

Age

3%
43%   54%

■ 18-25
■ 26-30
■ 31-35

Fluent in another dialect

3% 3%
94%

■ None
■ Shanxi
■ Nanjing

Taiwan Mandarin

City of origin

13%
87%

■ Taipei
■ Others

Age

10%
23%   67%

■ 31-35
■ 26-30
■ 18-25

Fluent in another dialect

50%   50%

■ None
■ Taiwanese
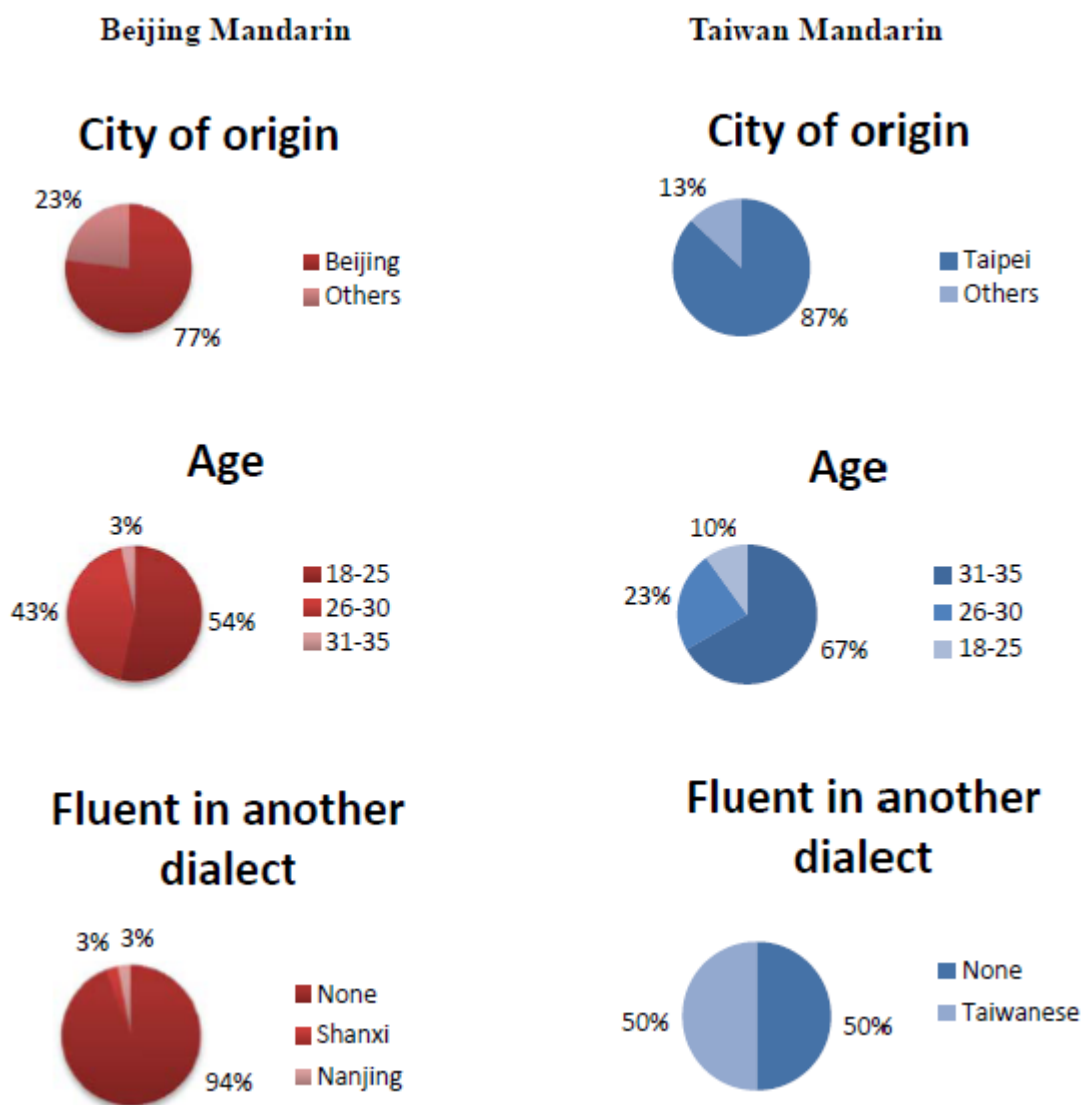
Figure 4.4: Background information of the participants

4.7.2 Experimental considerations

In choosing the auditory stimuli for this study, the /a, i, u/ vowel contexts were considered in the stimulus construction since they were all present in the production study in Chapter 3. However, the phonetic quality of Mandarin /i/ vowel is different when preceded by an alveolar vs. retroflex sibilant (see §2.1 and §2.3.3). In fact, in a continuum of /si-ʂi/, the two distinct allophones may facilitate the alveolar-retroflex distinction in perception (Lee 2011). Therefore,

vowel quality has to be manipulated if the /i/ context is to be included for the perception of the alveolar-retroflex contrast. Since it is not known what the impact of a phonologically non-existent vowel (i.e., a vowel between the apical vowels [ɹ] and [ɭ]) has on discrimination and goodness judgment of CV syllables, the /i/ context was not included in this study.

Another factor that has been considered before constructing the stimuli is whether the stimuli should be produced by a single talker or multiple talkers. Literature on perceptual learning has pointed out that perception is speaker-specific for fricatives (Eisner & McQueen 2005; Kraljic & Samuel 2005). The effects of gender can also cause the boundary location to shift in perception. Mann and Repp (1980), for example, have found that the boundary on their synthetic /ʃ-s/ continuum shifted toward higher frequencies in the context of a female voice. Taken together, stimuli produced by different talkers could result in quite different adjustments as listeners may maintain speaker-specific representations of each category. Experimentally, solving the problem of speaker normalization in speech perception introduces complications in data analysis. Therefore, the stimuli used in this study were produced by a single talker. As the production study in Chapter 3 reported that Beijing Mandarin speakers generally produced a larger alveolar-retroflex contrast than Taiwan Mandarin speakers, a Beijing Mandarin speaker was chosen to record the stimuli in order to maximize the acoustic contrast for subsequent /ʂ-s/ continuum construction.

4.7.3 Stimulus construction

As reviewed in §4.3, the cue in distinguishing between Mandarin alveolar and retroflex sibilants lies in the frequency domain (/s/ has major spectral prominence above 4 kHz, and /ʂ/ centered around the F2 region). In light of the acoustic manipulations introduced in Li and colleagues (Li & Allen 2011; Li, Trevino, Menon & Allen 2012), once the cutoff noise

frequencies (i.e., the frequencies of the lowest bound of the frication noise) of Mandarin /s/ and /ʂ/ are identified (see §4.4), raising the cut-off frequency of /ʂ/ to that of /s/ can morph /ʂ/ into /s/. To create a continuum between /ʂ/ and /s/, one can edit out the frication noise with a high-pass filter starting at the cut-off frequency of /ʂ/ in steps, until reaching the cut-off frequency of /s/. Beren, a software system developed by the Human Speech Recognition (HSR) Group at University of Illinois in Urbana-Champaign, allows such editing of naturally-produced speech signals using the short-time Fourier transform (Allen 1977; Allen & Rabiner 1977) and was used for constructing the stimuli for this study.

In obtaining naturally produced stimuli, a list of monosyllabic Mandarin words (see Appendix F), which all carry tone 1, was spoken in isolation by a male native speaker of Beijing Mandarin. The words were presented in simplified Chinese characters to the talker on a computer screen in a randomized order, and each word was repeated 10 times. The recording took place in the sound-attenuated booth in the Phonetics Lab at University of Illinois. The audio recording and its text transcription were sent to automatic segmentation using the PennPhonetics Lab Forced Aligner (Yuan & Liberman 2008). Using Praat (Boersma & Weenink 2011), the syllable duration of each word was automatically extracted from the segmentation output generated by the Forced Aligner. One pair of /ʂa-sa/ and /ʂu-su/ that shared the same syllable duration (560 ms and 480 ms respectively) were extracted. These tokens were downsampled to 16000 Hz in compliance with Beren's requirement on sampling frequency, and were normalized for RMS amplitude at 65 dB. These tokens were read into Beren, where the AI-gram of each token was generated (as previously seen in Figure 4.3).

To create an 8-step acoustic continuum from /ʂ/ to /s/ in each vowel context, the frication region between the alveolar and retroflex was divided into 7 equal intervals on the log scale,

yielding 8 cutoff frequencies (see Table 4.1 for the cutoff frequencies of each step of the /ṣa-sa/ and /ṣu-su/ continua). The 8 steps of the continuum were constructed by modifying the frication in /ṣa/ at 8 cut-off frequencies[20]. Figure 4.5 shows how this step-by-step high-pass filtering of the noise was carried out for the /ṣa-sa/ continuum. Figure 4.6 shows the same editing procedure for the /ṣu-su/ continuum. It should be noted that since the /ṣ-s/ continuum was constructed for the /a/ and /u/ contexts respectively, to reflect different magnitudes of the alveolar-retroflex contrast across vowel (see Chapter 3), the interval on the /ṣa-sa/ continuum was not equal to the interval on the /ṣu-su/ continuum. For example, step 4 on the /ṣa-sa/ continuum would not correspond to step 4 on the /ṣu-su/ in terms of the noise frequency cut-offs.

continuum steps	/ṣa-sa/ continuum	/ṣu-su/ continuum
Step 1	1.39	1.24
Step 2	1.61	1.4
Step 3	1.89	1.61
Step 4	2.24	1.85
Step 5	2.67	2.12
Step 6	3.15	2.41
Step 7	3.68	2.72
Step 8	4.23	3.12

Table 4.1: Cutoff frequencies (kHz) of each step of the /ṣa-sa/ and /ṣu-su/ continua

---

[20] The cutoff frequency of each step was located using the logarithmic interpolation formula in Deserno (2004).
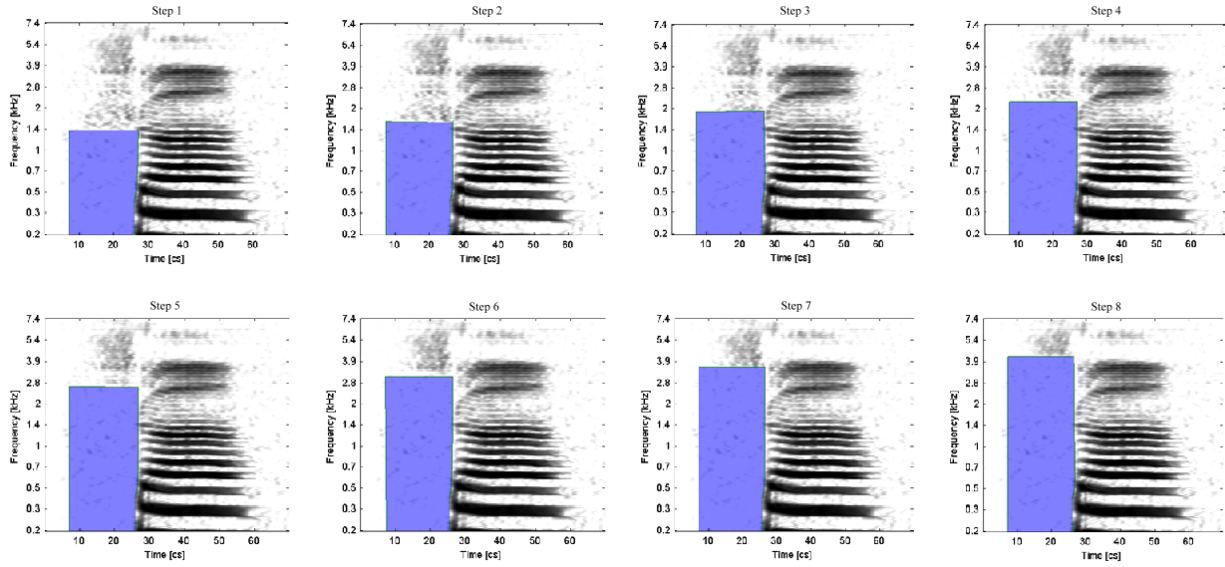
Figure 4.5: AI-grams showing the construction of the 8-step /ṣa-sa/ continuum (The shaded area indicates the region of noise that is being filtered out.)
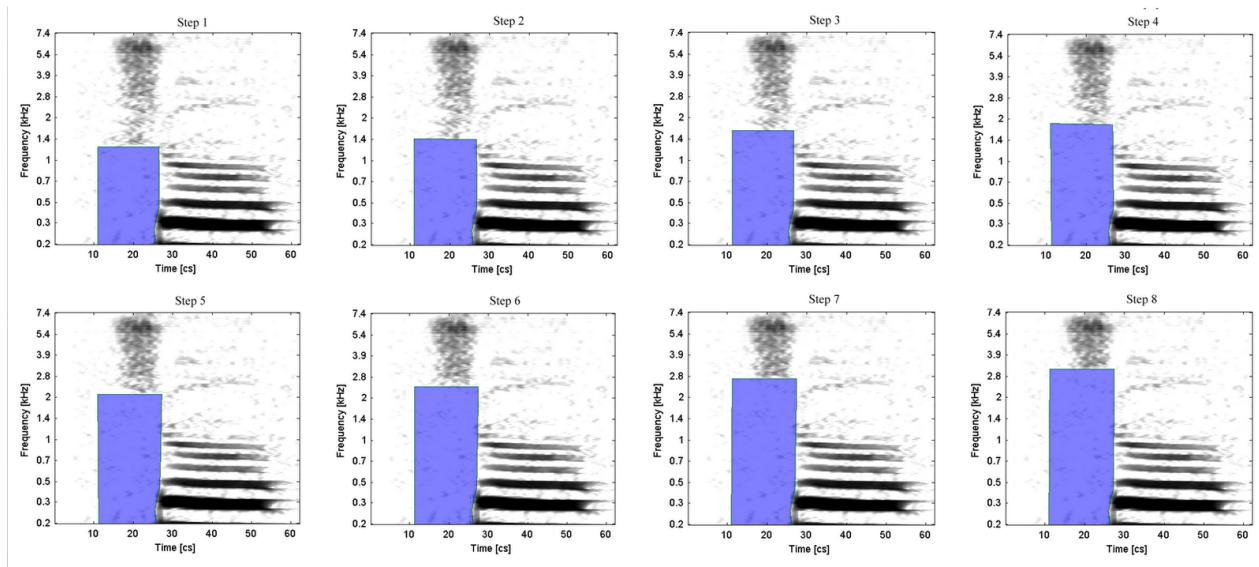
Figure 4.6: AI-grams showing the construction of the 8-step /ṣu-su/ continuum (The shaded area indicates the region of noise that is being filtered out.)

All 8 steps of the continuum stimuli, including step 1 where the weak noise below the cutoff frequency was filtered out, were edited from the naturally-produced retroflex syllables. To verify the goodness of the edited sound files, they were played to two Beijing Mandarin and two

77

Taiwan Mandarin native speakers. The sounds were judged to be natural sounding and the endpoints of the /ʂa-sa/ and /ʂu-su/ continua were unambiguously identified.

In addition to the two fricative continua, two filler continua /i-y/ and /ti-di/ were also created, using the tokens produced by the same male Beijing Mandarin speaker. The /i-y/ continuum was created using the Akustyk package in Praat. The continuum varied by the F3 interpolated between /i/ and /y/ (see Appendix G for the formant trajectories in the synthesis). The /ti-di/ continuum varied by the duration of VOT and was created by splicing off the aspiration from /ti/ at equal time intervals into /di/ (see Appendix H for the VOT continuum).

4.7.4 Procedure

The participants were seated at a computer in a quiet room. There were two experiments, an ABX discrimination task and a goodness rating task. The auditory stimuli were transmitted through earphones connected to the computer, and the participants made their responses by pressing the appropriate buttons on the keyboard (for the ABX task) and using a mouse (for the goodness rating task). Each experiment took about 15 minutes and a 5-minute break was given in between the two experiments. The participants completed one questionnaire about their linguistic background (Appendix D) after completion of the two experiments. The participants received a 7 dollar cash compensation for their participation.

4.8 Experiment 1: ABX discrimination

The goal of this experiment is to locate the category boundary between Mandarin alveolar and retroflex sibilants with an ABX discrimination task. In this experiment, the participants have to match the "X" stimulus with the "A" or "B" stimulus.

4.8.1 Methods

4.8.1.1 Stimuli

This ABX task (also see §4.7 for discussion of the ABX paradigm) consisted of two blocks. In each block, the B stimulus was two steps to the right of stimulus A on the continuum, making up six 2-step AB pairs from each 8-step continuum (see Table 4.2 for illustration; see §4.7.3 for details of the four 8-step acoustic continua). Each AB pair was combined with stimulus X (i.e., stimulus A or B) and presented in the following four combinations: ABB, ABA, BAA, BAB. The inter-stimulus interval within each ABX trial was set at 500 milliseconds. The second block was a repetition of the first block, therefore, each AB pair would be discriminated eight times throughout the task. From the above description, it follows that the total number of trials for this ABX task would be 192 (4 acoustic continua * 6 two-step pairs* 4 presentation combinations *2 blocks). All stimulus trials were presented in different random orders for each participant.

A	B
Step 1	Step 3
Step 2	Step 4
Step 3	Step 5
Step 4	Step 6
Step 5	Step 7
Step 6	Step 8

Table 4.2: Stimulus combination (AB pair) for the ABX task

4.8.1.2 Procedure

At the beginning of the task, the participants were told that they would hear three Mandarin sounds (sounds 1, 2 and 3, which correspond to ABX) in a sequence and they had to judge whether sound 3 was the same as sound 1 or sound 2. A response page (see Figure 4.7) appeared on the monitor after each sound sequence was played. On the response page, the lines connecting sound 3 to sounds 1 and 2 respectively prompt the participants to press the key labeled 1 on the

keyboard if they think sound 3 is identical to sound 1. Or the participants press the key labeled 2 on the keyboard if sound 3 is found to be identical to sound 2. All responses were logged automatically in E-Prime.
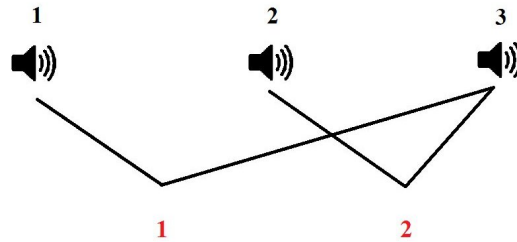

Figure 4.7: A response page prompting for decision in the ABX task

The participants were encouraged to respond as quickly and accurately as possible. There was a 6-trial practice, and the participants had the opportunity to ask clarification questions before proceeding to the experiment.

4.8.2 Analysis

Following categorical perception discussed in §4.2, it is assumed that listeners have the best performance (i.e., the highest accuracy) discriminating two sounds across the phoneme boundary. Therefore, a significant increase in sensitivity (i.e., accuracy) in this task will be considered where the category boundary is located.

Figure 4.8 shows the plot of the mean /ʂa-sa/ discrimination accuracy with a 95% confidence interval for both listener groups. A one-way repeated measure ANOVA, with PAIR (i.e., six 2-step AB pairs) as the within-subject variable and ACCURACY as the dependent variable, was conducted on Beijing and Taiwan listeners' data respectively. For Beijing listeners, the analysis revealed a significant main effect for PAIR ($F(5, 145)=17.734$; $p<.001$), meaning that accuracy was not the same for all stimulus pairs. Post-hoc tests with a Bonferroni correction exploring the

simple main effect of PAIR showed that pairs 1-3, 2-4 and 3-5 were not significantly different from each other, but pair 4-6 was significantly higher in accuracy than pairs 1-3, 2-4, and 3-5 (p<0.001, 0.001, and 0.01 respectively). If listeners could not reliably differentiate step 3 from step 5 with a two-step distance in between, it is assumed that they could not perform better in differentiating between step 4 and step 5 with a one-step distance. On the other hand, listeners performed significantly better differentiating step 4 from step 6, suggesting that the phoneme boundary is between step 4 and step 6. To be more precise, since the phoneme boundary is not between step 4 and step 5, it must be between 5 and 6, which corresponds to 2.91 kHz, halfway between the cutoff frequencies of steps 5 and 6 in the AI-grams (see Figure 4.5 and Table 4.1).

For Taiwan listeners, the discrimination curve rises from a baseline value of approximately 0.5 accuracy to a peak of 0.75 at pair 5-7. The ANOVA analysis revealed a significant main effect for PAIR as well (F(5, 145)=9.25; p<.01). Pairwise comparisons showed that pairs 1-3, 2-4, 3-5 and 4-6 were not significantly different and the accuracy was at chance level, meaning that listeners could not discriminate among steps 1-6. On the other hand, the accuracy rate of pair 5-7 was significantly higher than pair 4-6 (p=0.019), which suggests that the perceptual discontinuity must have occurred between step 6 and step 7. This would correspond to 3.41 kHz, halfway between the cutoff frequencies of steps 6 and 7 in the AI-grams (see Figure 4.5 and Table 4.1).
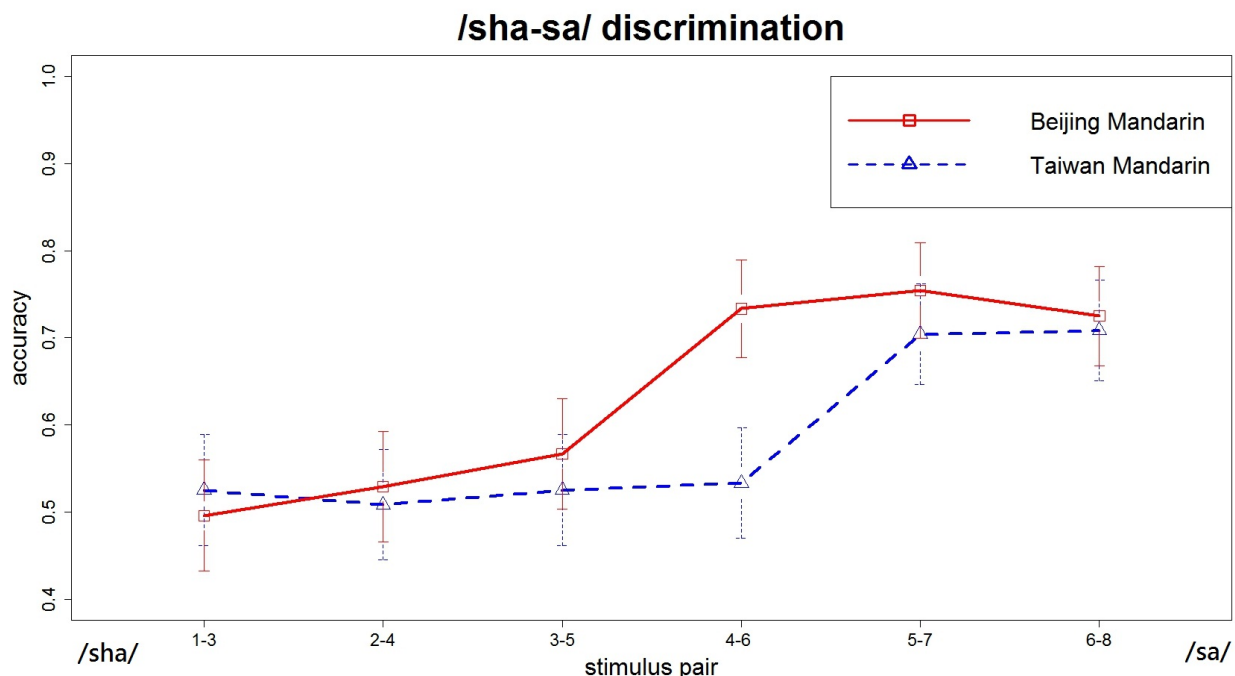
/sha-sa/ discrimination

Figure 4.8: /ṣa-sa/ discrimination for Beijing and Taiwan Mandarin listeners (/sh/ stands for /ṣ/ in the plot)

The plot of the /ṣu-su/ discrimination performance for Beijing and Taiwan listeners is displayed in Figure 4.9. For Beijing listeners, a one-way repeated measure ANOVA analysis revealed a significant main effect for PAIR ($F_{(5, 145)}=14.565$; $p<.001$). Post-hoc tests showed that pairs 1-3, 2-4, 3-5 and 4-6 were not significantly different from each other, but pair 5-7 was significantly higher in accuracy than pairs 1-3, 2-4, 3-5 and 4-6 ($p<0.001$, 0.001, 0.01, and 0.05 respectively). This suggests that the category boundary was located between steps 6 and 7, which corresponds to 2.56 kHz, halfway between the cutoff frequencies of steps 6 and 7 in the AI-grams (see Figure 4.6 and Table 4.1).

For Taiwan listeners, the ANOVA analysis also revealed a significant main effect for PAIR ($F_{(5, 145)}=8.993$; $p<.01$). Post-hoc comparisons showed that only pair 6-8 was significantly higher in accuracy than pairs 1-3, 2-4, 3-5, 4-6 and 5-7 ($p<0.001$, 0.001, 0.001, 0.001 and 0.05 respectively). That only pair 6-8 was discriminated above chance-level accuracy suggests that

the category boundary was located between steps 7 and 8, since listeners could not discriminate among steps 1-7. This would correspond to 2.92 kHz, halfway between the cutoff noise frequencies of steps 7 and 8 in the AI-grams (see Figure 4.6 and Table 4.1).
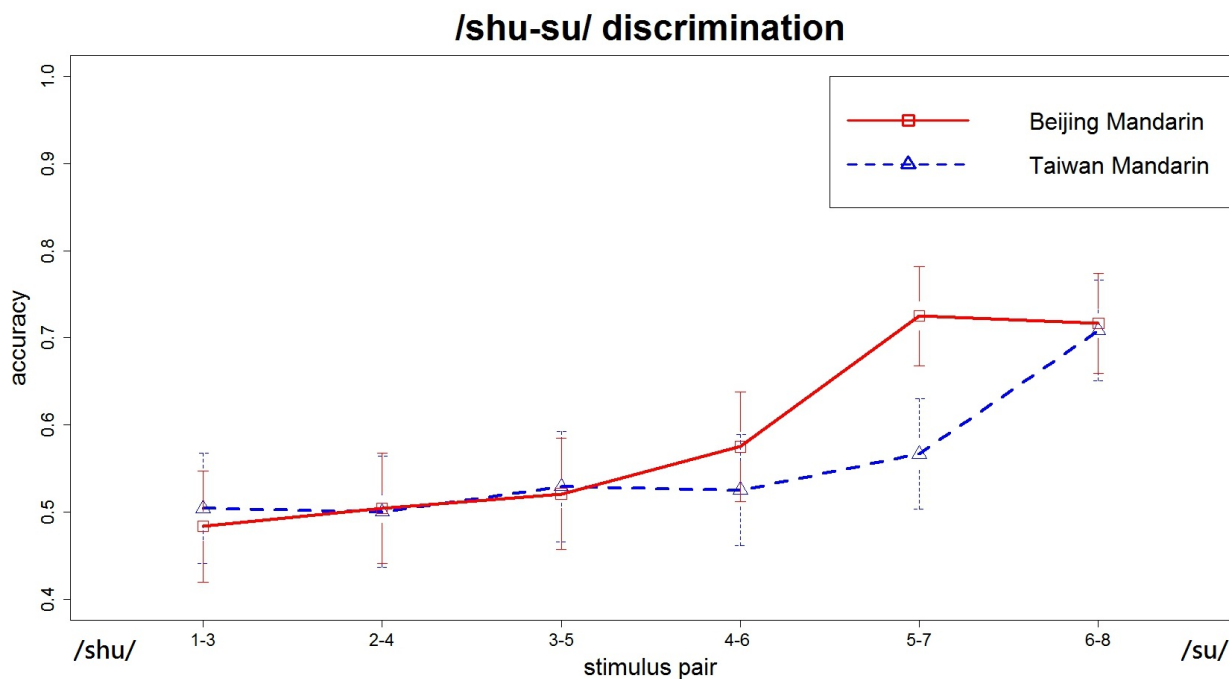
/shu-su/ discrimination

Figure 4.9: /ʂu-su/ discrimination for Beijing and Taiwan Mandarin listeners (/sh/ stands for /ʂ/ in the plot)

4.8.3 Discussion

The current experiment showed clear dialectal differences. Beijing and Taiwan Mandarin listeners had different locations of the category boundary along the /ʂ-s/ continuum, with Taiwan listeners' perceptual boundary located one step closer to the alveolar end of the continuum than Beijing listeners. That is, given the same /ʂ-s/ continuum, Taiwan listeners allocated a wider range of noise frequency band to retroflexion than Beijing listeners.

For both Beijing and Taiwan listeners, the category boundary shifted toward lower noise frequencies in the /u/ context (i.e., from 2.91/3.41 kHz in the /a/ context to 2.56/2.92 kHz in the /u/ context). This boundary shift indicates perceptual compensation for coarticulation from the

83

following rounded vowel (see §4.5 for discussion on compensatory perception). A lower boundary frequency for Beijing listeners than for Taiwan listeners in both vowel contexts suggests that a lower cutoff noise frequency is required for the retroflex percepts for Beijing listeners.

It should be noted that none of the perceptual patterns observed here showed a classical quadratic trend with a peak at the category boundary with tapered ends along the continuum, as reported in categorical perception studies that used a ABX discrimination task (Francis et al. 2003; Liberman et al. 1957; Pisoni & Lazarus 1974). It could be that the stimuli used in those studies were all synthesized stimuli, with exaggerated stimuli that fell out of natural range of speech. In contrast, the present study used stimuli modified from naturally-produced fricatives in the frequency (Hz) domain, and all modifications fell within with the natural range of the Beijing talker. Particularly, the Beijing talker produced an alveolar rather than dental /s/. It is speculated if dental /s/ tokens were included (constructed by raising the cutoff frequency even higher), which in turn expands the frequency range between the alveolar and retroflex on the acoustic continuum, a quadratic perceptual pattern might be observed. I will return to this speculation in §5.2.

Overall, discrimination accuracy remained around the chance level towards the retroflex end (i.e., the left edge of the continuum) but rose significantly towards the alveolar end (i.e., the right edge of the continuum). That is, both Beijing and Taiwan listeners appeared to show little discriminability to the variation within the retroflex category. This asymmetrical within-category discriminability will be further explored in comparison to the goodness rating data in the following section.

4.9 Experiment 2: Goodness rating task

This experiment will investigate the internal structure of the /s/ and /ʂ/ phonemes to see if listeners judge certain exemplars of a phoneme category to be particularly good. The participants will rate the subjective goodness of an alveolar or retroflex syllable on a continuous scale.

4.9.1 Methods

4.9.1.1 Stimuli

The stimuli used in the goodness rating task came from the same acoustic continua described in §4.7.3. In this experiment, the participants heard a stimulus and saw the prompt question on screen asking "How is the pronunciation of this ___?", the blank being the character representation of either syllable from the minimal pair. For example, when step 1 (i.e., /ʂa/) from the /ʂa-sa/ continuum is played, the participants will be asked "How is the pronunciation of this SA?" in one trial, and "How is the pronunciation of this SHA?" in the other. This setup made no assumption that a bad pronunciation of /sa/ entails a good pronunciation of /ʂa/ (cf. Miller 1994 where a stimulus from the /bi/-/pi/-excessively aspirated /pi/ continuum was only judged against /pi/, but not /bi/). Therefore, each step from the four continua was rated twice, making a total number of 64 trials. All stimuli were presented in a random order to each participant.

4.9.1.2. Procedure

Upon hearing a stimulus, a screen with a Visual Analogue Scale prompted the participants to rate the pronunciation (see Figure 4.10). The participants used the mouse to drag the white rhombus along the scale. The participants were asked to move the rhombus towards the left if they believed the sound they heard was a bad pronunciation of the prompted word, and towards the right if they believed that the sound was a good pronunciation of the prompted word. After

they made their decision, two boxes appear below the scale asking them whether they would like to listen and rate again or proceed with the next sound.
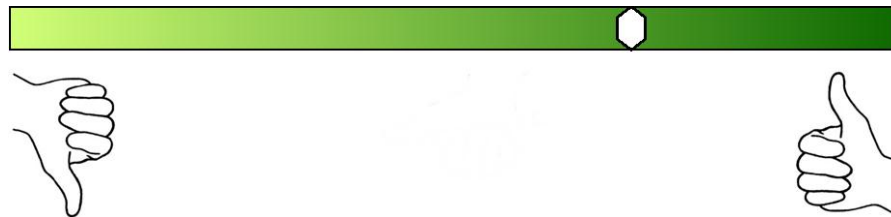

Figure 4.10: The Visual Analog Scale in the goodness rating task

Participants were allowed to listen to each sound as many times as they deemed necessary. There were 16 trials in the practice. The stimuli included both ends of a continuum to implicitly familiarize the participants with the range of acoustic stimuli they would be hearing.

4.9.2 Analysis

As the /ʂ-s/ boundary location was decided based on the ABX data analysis in §4.8.2, in the analysis, the two endpoints on the /ʂ-s/ continuum will be compared to their within-category variants (e.g., the endpoint /ʂ/ being compared to other stimuli on the continuum that are also considered /ʂ/). If an endpoint stimulus has significantly better rating scores than its sub-phonemic variants, then the endpoint stimulus is considered a better exemplar for its category. If the ratings for the sub-phonemic variants and end-point stimulus are not significantly different, then the category is considered to be represented by an exemplar cloud without a better category member.

While participants could listen to the same stimulus and rate it as many times as they like in the goodness rating task, only the last response of each trial was used for data analysis. Since

participants may have used the continuous scale differently—some may fully exploit the scale whereas others may be more conservative raters, all raw rating scores were transformed to z-scores. However, the plots using raw scores vs. standardized z-scores were similar. Therefore, raw scores were used for data plots and statistical analysis in the following.

Figure 4.11 displays the plots of Beijing listeners' goodness judgments of stimuli from the /şa-sa/ continuum. The plot on the left shows listeners' ratings in response to the question "How is the pronunciation of this SHA?", and the plot on the right shows the ratings of the stimuli when being asked the opposite question "How is the pronunciation of this SA?" Each data point in the SA data set is inverted by 100 minus each point (the Visual Analogy Scale corresponds to a 0-100 scale) and correlated with its corresponding data in the SHA data. A strong correlation between the SHA dataset and the inverted SA dataset would suggest that listeners consider a bad instance of / şa/ a good instance of /sa/, and vice versa.

The two sets of data were submitted to a one-way repeated measure ANOVA separately. For the SHA dataset (refer to the left panel in Figure 4.11), there was a significant main effect of STIMULUS ($F(7,203)=46.493$, $p<.001$), which means that goodness ratings were not the same across stimuli steps. Follow-up post-hoc tests revealed no statistical significance among steps 1-5 within the category of retroflex, suggesting that no single member of the retroflex category was better than the other. On the other hand, within the category of alveolar, step 8 was found to be significantly different from step 7 ($p=0.015$) and step 6 ($p<0.001$). That is, step 8 was considered a better exemplar of the alveolar category. Now, the SA dataset (refer to the right panel in Figure 4.11) was submitted to the same statistical analysis and a significant main effect of STIMULUS ($F(7,203)=47.191$, $p<.001$) was also found. Follow-up post-hoc tests revealed no statistical significance among steps 1-5 within the category of retroflex. Within the alveolar category, step

87

8 was found to be significantly different from step 6 (p<0.001) and marginally differently from step 7 (p=0.055). The correlation coefficient between the SHA dataset and inverted SA dataset was 0.797, suggesting the two datasets were strongly correlated.
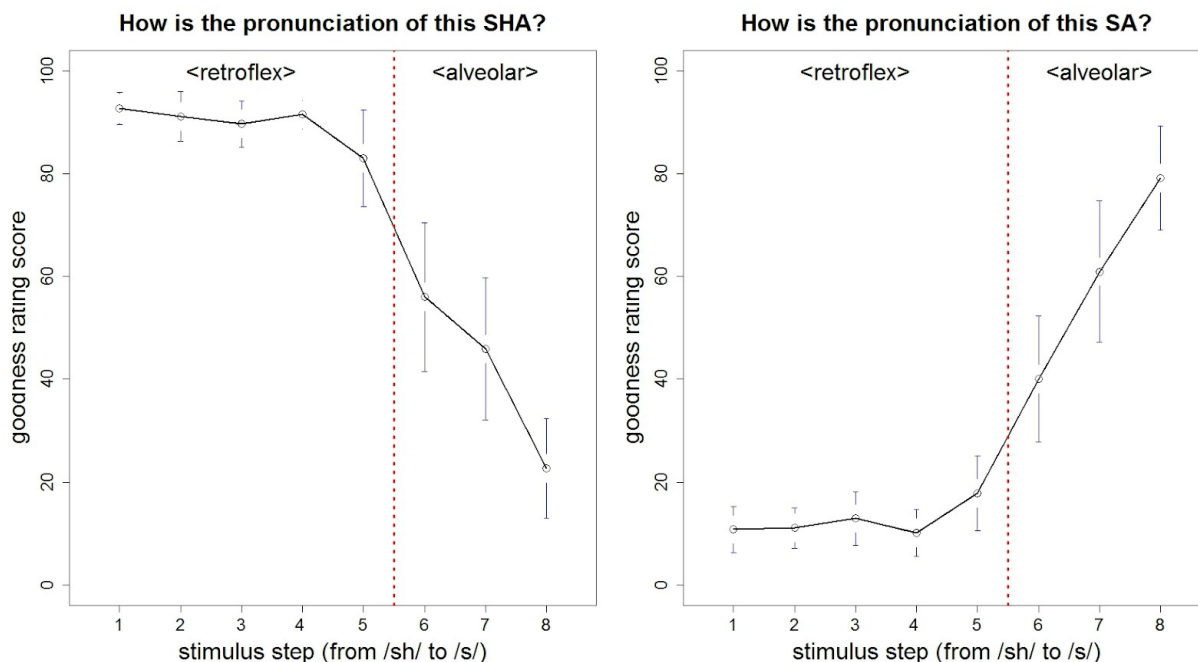

Figure 4.11: Goodness rating of /ʂa/ and /sa/ along a /ʂa-sa/ continuum for Beijing listeners. The dotted line indicates the category boundary based on the ABX data.

For Taiwan listeners' goodness judgments on the /ʂa-sa/ continuum (Figure 4.12), a significant main effect of STIMULUS was found for both sets of data: $F(7,203)=10.11$, $p<.01$ for the SHA data set (refer to the left panel in Figure 4.12); $F(7,203)=11.825$, $p<.01$ for the SA data set (refer to the right panel in Figure 4.12). Pairwise comparisons showed that no within-category variants were significantly different from their corresponding endpoint phonemes. The results suggest that for Taiwan Mandarin listeners, subphonemic variants of both /ʂa/ and /sa/ were perceptually equivalent to their respective endpoint stimulus. The correlation coefficient between

88

the SHA dataset and inverted SA dataset was 0.401, suggesting the two datasets were moderately
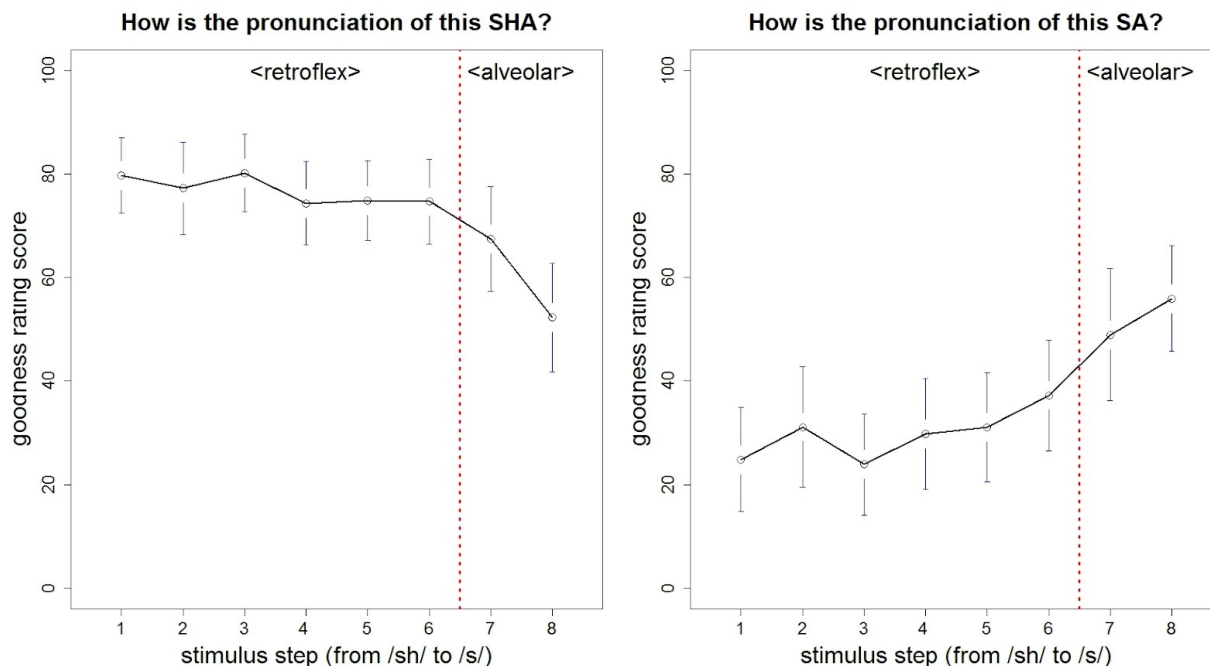
correlated.

Figure 4.12: Goodness rating of /ʂa/ and /sa/ along a /ʂa-sa/ continuum for Taiwan listeners. The dotted line indicates the category boundary based on the ABX data.

Figure 4.13 displays the plots of Beijing listeners' goodness judgments of the stimuli along

the /ʂu-su/ continuum. A one-way repeated measure ANOVA revealed a significant main effect

of STIMULUS for both sets of data: $F(7,203)=56.895$, $p<.001$ for the SHU data set (refer to the

left panel in Figure 4.13); $F(7,203)=40.549$, $p<.001$ for the SU data set (refer to the right panel in

Figure 4.13). For both datasets, pairwise comparisons showed no statistically significant

difference among any steps within the category of retroflex, suggesting that all within-category

members were considered equally good exemplars of /ʂu/. Within the category of alveolar, step 8

was found to be significantly different from step 7 in both sets of data (p <0.01). The correlation

coefficient between the SHU dataset and inverted SU dataset was 0.61, suggesting the two datasets were strongly correlated.
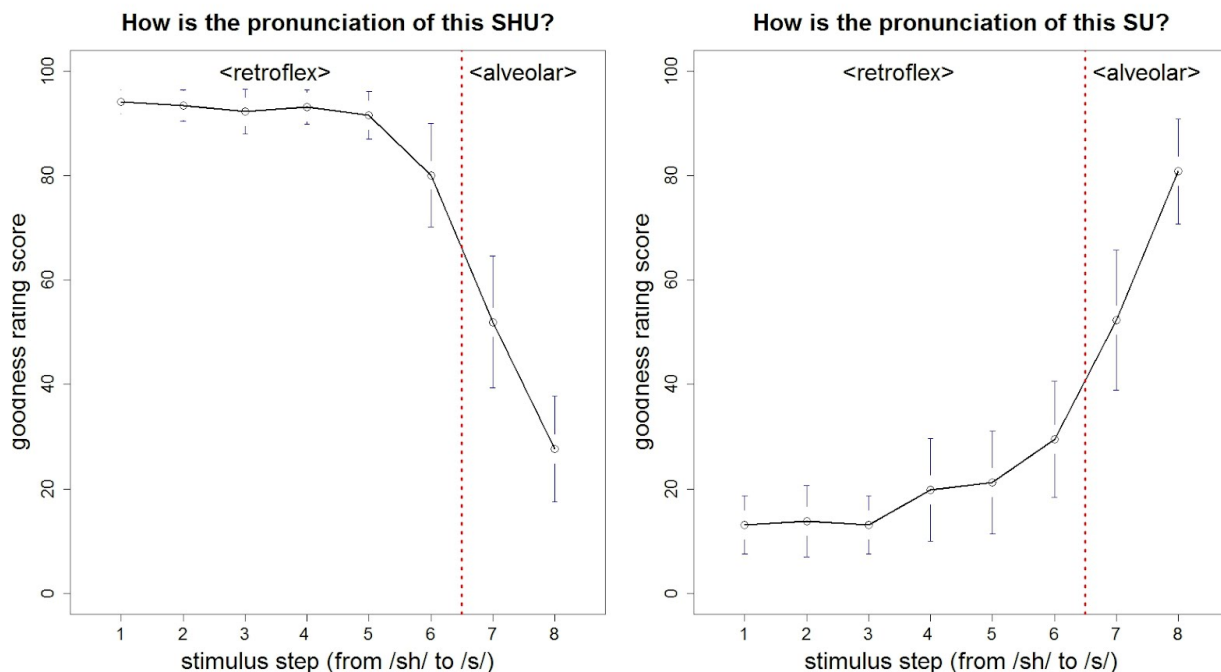

Figure 4.13: Goodness rating of /ʂu/ and /su/ along a /ʂu-su/ continuum for Beijing listeners. The dotted line indicates the category boundary based on the ABX data.

Figure 4.14 shows Taiwan listeners' goodness judgments of the stimuli from the /ʂu-su/ continuum. A one-way repeated measure ANOVA statistical analysis revealed a significant main effect of STIMULUS for both sets of data: $F(7,203)=11.113$, $p<.01$ for the SHU dataset (refer to the left panel in Figure 4.14); $F(7,203)=16.856$, $p<.001$ for the SU dataset (refer to the right panel in Figure 4.14). For the SHU dataset, pairwise comparisons showed no statistically significant difference among any steps within the category of retroflex. For the SU data set, step 1 was only found significantly different from step 7 ($p<0.01$) within the retroflex category. Since only step 8 was considered an alveolar based on the results of the ABX experiment, there was no

other within-category member to compare to. Therefore, step 8 was the only alveolar exemplar in the stimuli being treated. The correlation coefficient between the SHU dataset and the inverted SU dataset was 0.467, suggesting the two datasets were moderately correlated.
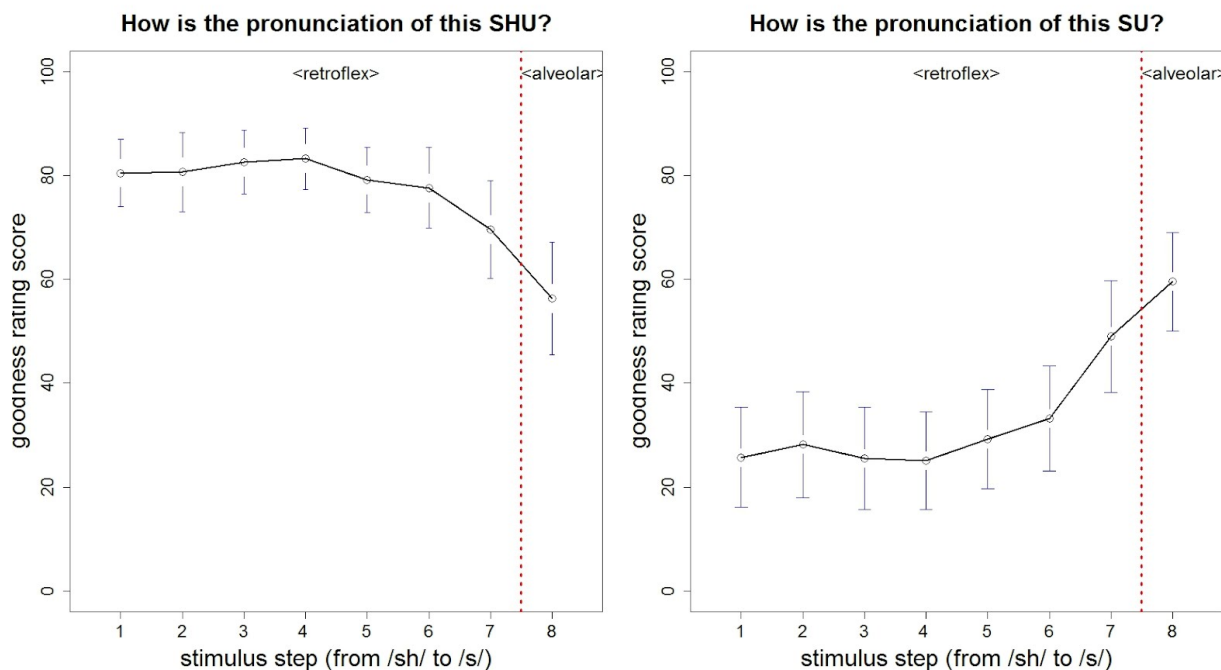

Figure 4.14: Goodness rating of /ʂu/ and /su/ along a /ʂu-su/ continuum for Taiwan listeners. The dotted line indicates the category boundary based on the ABX data.

4.9.3 Discussion

This study investigated the gradient perception along the /ʂ-s/ continuum followed by the /a/ or /u/ vowel. The participants were asked to rate the goodness of a stimulus in response to different prompt questions. The correlation coefficients reported in §4.9.2 suggest that the two sets of goodness data (SHA/SHU vs. SA/SU) were strongly or moderately correlated. For example, when the endpoint /ʂ/ stimulus was rated against the /ʂ/ question, high rating scores were received. In contrast, when the same /ʂ/ stimulus was rated against the /s/ question, low ratings scores were received. Therefore, it can be concluded that a good instance of /ʂ/ would be

perceived as a bad instance of /s/, and vice versa. This result provided support for the literature in which only one prompt question was asked in goodness rating tasks (e.g., Miller & Volaitis 1989; Miller 1997).

Stevens and Blumstein (1975) proposed a quantal relationship between acoustic properties and perceptual characteristics. Within a quantal zone of acoustic event, the perceptual system is relatively insensitive to changes in acoustic attributes. However, a small change outside this zone will give rise to percepts of a different phoneme. In this regard, both Beijing and Taiwan listeners' goodness ratings on within-category variants can be explained by Stevens and Blumstein's quantal theory of speech perception, as greater change in goodness rating was observed near the category boundary. For goodness ratings of variants within the retroflex category, a slowly rising or lowering trend (depending on the prompt question) towards the category boundary was shown (see Figures 4.11-4.14), and the ratings did not significantly vary across the /ʂ/ variants. Within the category of alveolars, Beijing listeners demonstrated clear gradient sensitivity in that the endpoint /s/ stimulus received significantly different ratings from the other /s/ variants. To use the example of Beijing listeners' goodness rating along the /ʂa-sa/ continuum (refer to the left panel in Figure 4.11), although steps 6 and 7 were not categorized as retroflexes, they were not perceived as good exemplars of alveolars either. In contrast, this gradient sensitivity to variants of the alveolar was not observed in Taiwan listeners.

Now, a question arises as to why there was discrepant sensitivity to within-category variation for the alveolar and the retroflex? An articulatory account was provided here to link between speech perception and speech production. Articulatorily, whether constriction is made behind the teeth or against the alveolar ridge, alveolars have a relatively small resonant cavity anterior to the constriction. Alveolars essentially involve only a tongue tip raising gesture. According to Keyser

and Stevens (2006), "[b]efore a front vowel, the F2 starting frequency for an alveolar is only slightly higher than it is before a back vowel, indicating about the same fronted tongue-body position as that for the alveolar preceding a back vowel" (p. 48). The less variable articulatory gestures result in a more specific acoustic landmark for the alveolars—high-frequency frication noise. In contrast, retroflexion has more complex tongue configurations and articulatory properties (e.g., the subliminal space that adds volume and complexity to the resonant cavity), thereby contributing to more variability in its acoustics. In addition, optional enhancing gestures to the retroflexes like tongue blade raising (Stevens et al. 2004) or lip rounding (Chang 2010), can further contribute to the acoustic variability. Perkell and Nelson's (1985) study of variability in production of the vowels /i/ and /a/ may also provide support for the acoustic variability of the alveolar vs. retroflex productions. They reported that a small change in tongue placement in the front cavity, as opposed to the back cavity, would create relatively large percentage changes in the area function at the constriction, which in turn would trigger more perceptually salient formant changes. In sum, the acoustic properties associated with the retroflex consonant are more complex than those of the alveolar consonants (Stevens & Blumstein 1975, p. 230). The acoustic characteristics may in turn impinge on perception such that listeners become more tolerant with a phoneme that has more acoustic variability, but more sensitive to a phoneme that has a more specific acoustic profile.

In studying the perceptual best exemplars, Johnson et al. (1993) found that listeners' choices of perceptual best vowel exemplars to be more extreme (higher high vowels, lower low vowels, farther front front vowels, and farther back back vowels) than their own productions. They called this perceptual vowel space expansion a hyperspace effect. The hyperspace effect may extend to fricative perception given that Taiwan speakers generally produce a lesser degree of retroflexion

(i.e., have a higher spectral mean of the frication noise) than Beijing speakers (see Chapter 3 and Chang's study in 2011), and are expected to give a better rating to retroflex stimuli produced by a Beijing speaker. However, this was not observed in their goodness rating data, as the rating scores on the endpoint /ʂ/ was not significantly different from those on the other /ʂ/ variants. Therefore, Taiwan listeners' retroflex perception data did not provide evidence for the hyperspace effect.

In sum, this study found that Beijing and Taiwan Mandarin listeners perceived all variants of retroflexes as equally good. That is, both groups showed insignificant sensitivity or had great tolerance toward the sub-phonemic variation of the retroflex category. This pattern held true for both the /a/ and /u/ contexts. On the other hand, a gradient structure in the alveolar category was reflected in the goodness judgments of Beijing listeners for both vowel contexts. But this gradient structure was not observed in Taiwan listener's goodness ratings on the alveolar members. To answer the research questions posed in §4.6, both Beijing and Taiwan listeners judged the retroflex, in either vowel context, to be represented by an exemplar cloud without a best category member. For both vowel contexts, the best exemplar for the alveolar category was located in the alveolar end of the continuum for Beijing listeners. In contrast, for Taiwan listeners, the few members (i.e., 2 members along the /ʂa-sa/ continuum and 1 member along the /ʂu-su/ continuum) in the alveolar category were considered equally good exemplars.

4.10 Conclusion

The acoustic continua used in this study represent intermediate stages from the retroflex through the alveolar. The ABX tasks shows that the continuous spectral properties were perceived categorically. Beijing Mandarin and Taiwan Mandarin listeners had different perceptual boundaries along the /ʂ-s/ continua, with a lower cutoff noise frequency required for

retroflex percepts for Beijing Mandarin listeners. In other words, given the same acoustic continuum, fewer steps were perceived as retroflexes by Beijing listeners. The location of the category boundary varied across vowel for both Beijing and Taiwan listeners: the /ʂ-s/ perceptual boundary shifted from 2.91/3.41 kHz in the /a/ context to 2.56/2.92 kHz in the /u/ context. The rounded vowel context shrinks the spectral distance between /ʂ/ and /s/, where most steps along the continuum were perceived as retroflexes for Taiwan listeners. This result suggests that the /u/ context is prone to more perceptual confusion for the alveolar-retroflex distinction for Taiwan listeners.

The goodness rating experiment further investigated the internal structure of each phoneme by examining whether listeners showed gradient sensitivity to fine-grained acoustic differences in the spectral noise distribution. For both Beijing and Taiwan listeners, the perceptual patterns can be characterized by a quantal relationship described by Stevens and Blumstein (1975). More sensitivity to the acoustic variation within the alveolar category, but not within the retroflex category, was observed in Beijing listeners. The discrepant perceptual patterns were further explained in terms of greater articulatory variability in retroflexion such that listeners become more tolerant with its variation in perception.

In conclusion, this study demonstrated the variability in cross-dialectal perception of the Mandarin alveolar-retroflex contrast. The categorical perception aspect of this study allows us to contribute to the body of literature on the influence of dialectal background on the perception of phonological contrasts. The selective gradient perception observed here sheds light on the variable relationship between acoustic variation and perceptual sensitivity as intermediated by articulation. In the next chapter, I will discuss the results from both the production and perception studies.

CHAPTER 5: GENERAL DISCUSSION

This chapter attempts to synthesize the findings from the production study and the perception study (§5.1). Next, I raised some concerns about the rating discrepancy observed in the goodness data. To address these concerns, a follow-up experiment was designed and the results were reported in §5.2.

5.1 Cross-dialectal production and perception of Mandarin alveolar-retroflex contrast

Three similarities were identified between Beijing and Taiwan Mandarin alveolar-retroflex contrast: 1) Alveolar sibilants were realized with higher COG than the retroflex in all vowel contexts. 2) Alveolar productions in the /u/ context had significantly lower COG than in the other two contexts, /i/ and /a/. 3) The /u/ context was where the least alveolar-retroflex distinction was made. The differences that distinguished Beijing Mandarin from Taiwan Mandarin were: 1) Alveolar productions in the /u/ context in Taiwan Mandarin had significantly lower COG than that in Beijing Mandarin. 2) Retroflex productions in the /a, i/ contexts in Beijing Mandarin had significantly lower COG than that in Taiwan Mandarin. 3) Beijing speakers produced a significantly larger alveolar-retroflex contrast than Taiwan speakers in the /a/ context.

One of the most important aspects of the findings lies in identifying the limitation of the focus effects on acoustic cues that mark a phonological contrast. First, focal prominence did not contribute to greater spectral separation between the alveolars and retroflexes. This result is compatible with Cole et al.'s (2007) findings on the place contrast for English stop consonants and Silbert and de Jong's (2008) on the place contrast for English fricatives. Nevertheless, accentuation may cause uniform acoustic strengthening, as reported in Cole et al. (2007). Or

accentuation may enhance dimensions that are not designated as contrastive for a phonological contrast, as the current production study showed that focal prominence was manifested by longer syllable durations. Similarly, in Silbert and de Jong's (2009) study, contrastive focus was found to increase fricative and syllable durations as well as noise power. Although accentuation has been found to increase the distinctiveness of vowels (Beckman et al. 1992; Cho 2005; de Jong 1995) and stop voicing (Cole et al. 2007; Hsu & Jun 1998), the same effects do not appear to extend to place contrasts, suggesting that it should not be taken for granted that prosodic prominence leads to an enhancement of all contrastive features.

The production study also revealed a general pattern of great variability in realizing a place contrast for Mandarin fricatives. The factors of vowel and dialect interacted in affecting the magnitude of Mandarin alveolar-retroflex contrast. Although the two dialects converged in the place contrast in the /i/ and /u/ vowel contexts, the category separation did not result from the same realizations of the alveolars and retroflexes. For example, both alveolar and retroflex productions in the /u/ context in Taiwan Mandarin had lower COG than those in Beijing Mandarin, suggesting that the two classes of sounds were more subject to the coarticulatory effects of rounded vowels in Taiwan Mandarin. In sum, how a place contrast is maintained may vary by language-internal factors (such as vowel context and how each contrasting phoneme is realized), language-external factors (such as dialect) and the interaction of the two.

The perception study investigated the perception of Mandarin alveolar-retroflex contrast across vowel context and dialect with respect to categorical perception, by means of an ABX discrimination task, as well as gradient perception, by means of a goodness rating task. First, Beijing and Taiwan listeners were found to have different locations of the category boundary along the /ʂ-s/ continuum, with Taiwan listeners' perceptual boundary located one step closer to

97

the alveolar end of the continuum than Beijing listeners. For listeners of both dialects, the category boundary shifted to lower frequencies in categorizing /ʂ-s/ in the /u/ context, indicating context-moderated perception. Perceptual compensation for anticipatory vowel coarticulation on the acoustics of fricatives has been demonstrated in Mann and Repp (1980) and Whalen (1981). Second, the location of the best category exemplars did not vary as a result of vowel contexts. Some gradient sensitivity to the alveolar variants was found for only Beijing listeners, whereas no sensitivity to the retroflex variants was found for either Beijing or Taiwan listeners. That is, within the category of the alveolar, only Beijing listeners considered the /s/ endpoint stimuli to be better than the other alveolar members. In contrast, within the category of the retroflex, no member was perceived to be better than the other. An articulatory explanation was provided to account for this disparate within-category sensitivity: alveolars have less variable articulatory gestures, which result in a more specific acoustic landmark; retroflexes involve more complex articulatory properties, thereby exhibiting more acoustic variability. Indeed, retroflex productions were found to exhibit greater variance than alveolar productions for both Beijing and Taiwan speakers (see §3.5.4). Different degrees of acoustic variability in production are in turn suggested to be reflected in sensitivity to within-category variation in perception.

The combined production and perception results lead to a conclusion that although degrees of retroflexion is present in production, it is not reflected in perception such that more retroflexed stimuli receive higher goodness ratings. Then one question is raised as to why some Taiwan speakers, who do make the place distinction, can be perceived as exhibiting little or no retroflexion by Beijing speakers. A possible explanation is that Beijing and Taiwan Mandarin listeners' categorical boundaries were located a step away from each other along the /ʂ-s/ continua (see Figures 4.8 and 4.9). If Taiwan speakers' retroflex productions are reduced overall

and centered around that particular step (i.e., step 6 along the /ʂa-sa/ continuum and step 7 along the /ʂu-su/ continuum), these productions would be considered by Taiwan listeners as equally good as the more retroflexed tokens produced by Beijing speakers. The same productions, however, would actually be perceived by Beijing listeners as possible exemplars of the alveolar. Therefore, the stereotyped dialect perception that Taiwan speakers do not make the alveolar-retroflex distinction may be potentially attributed to different locations of the /ʂ-s/ perceptual boundary for Beijing Mandarin and Taiwan Mandarin.

Finally, for the discussion of a potential alveolar-retroflex merger in Taiwan Mandarin, as seen in the perception study, the acoustic continuum in both the /a, u/ contexts was categorized into two phonemes by Taiwan listeners. Although the less retroflexed stimuli were rated as equally good as the more retroflex stimuli, they were rated differently from the alveolar stimuli. Therefore, while reduced retroflexion may have become a conventionalized variant of retroflex productions in Taiwan Mandarin, it is not phonologically recategorized as an alveolar such that any difference between the realizations of the two categories become unperceivable. In connection with Hualde's (to appear) description of reductive sound changes, a complete merger between the alveolars and retroflexes in the speech community is yet to happen to Taiwan Mandarin.

5.2 Goodness rating discrepancy and follow-up experiment

In addition to discrepant within-category sensitivity reported in the perception study (see §4.9.3), another rating discrepancy was observed for the /ʂ-s/ continuum from the data—listeners appeared more conservative with judgments on the alveolar members. Considering Beijing

listeners' goodness rating scores on both /ṣa-sa/ and /ṣu-su/ continua for example (see Figures 4.11 and 4.13), it can be seen that step 1 /ṣu/ received an average of 90 points when rated against the SHU question and 10 points when rated against the SU question. However, when it comes to rating step 8 /su/, it received an average of 30 points in response to the SHU question and 80 points upon the SU question. Taiwan listeners' data also exhibited a similar pattern. This rating discrepancy may suggest that listeners perceived the retroflex members to be equally good, but were expecting even better alveolar exemplars than the /s/ endpoint stimulus. It is speculated that 1) the endpoint alveolar stimuli were modified from the retroflex stimuli and may not be as good as naturally-produced alveolar stimuli. The modified /s/ stimuli contain partial spectral information (i.e., frication noise and CV transitions) from the retroflex stimuli, which might create a bias in favor of retroflex percepts. Or 2) a hyperspace effect (Johnson et al. 1993) was induced in the ratings of alveolar members such that listeners preferred more extreme /s/ productions, namely dental /s/, even though listeners themselves may produce alveolar /s/. The first account essentially questions the validity of modification of retroflex stimuli to create the alveolar stimuli. The second account can only be tested by including the dental /s/ stimuli. To clarify the nature of this rating discrepancy, a follow-up experiment was conducted.

In the follow-up experiment, two questions are addressed: 1) Are the naturally-produced stimuli /sa/ and /su/ indistinguishable from the modified /sa/ and /su/ (i.e., step 8 of the 8-step /ṣ-s/ continua used in the ABX and goodness rating experiments)? 2) Are dental /s/ stimuli perceived as better exemplars of the alveolar category in Mandarin? To answer the first question, goodness rating of naturally-produced /s/ stimuli will be compared to modified ones. To answer the second question, two more steps will be added to the /ṣa-sa/ and /ṣu-su/ continua to simulate

the dental /s/ by further raising the cutoff frequency of high-pass filters (see §4.7.3 for more detail).

Eight Beijing Mandarin speakers and eight Taiwan Mandarin speakers were recruited for the follow-up experiment. The stimuli were 44 target stimuli (11 steps, including 10 steps of modified stimuli and 1 naturally-produced /s/, * 2 vowels * 2 prompt questions) and 16 filler stimuli (from the /ti-di/ and /i-y/ continua). The experiment was conducted following the same procedure introduced in §4.9.1.2.

The data were submitted to a one-way repeated measure ANOVA. The statistical results can be summarized into two points: 1) The naturally produced /sa/ and /su/ were not scored significantly differently from the modified stimuli (i.e., step 8 from the continua used in the ABX and good rating tasks in Chapter 4), and 2) the goodness ratings on steps 9 and 10 (i.e., the dental tokens) were not significantly different from Step 8 (i.e., the alveolar token).

The goodness ratings of the naturally-produced /s/ vs. the modified /s/ for Beijing listeners and Taiwan listeners were plotted and displayed in Figures 5.1-5.4. The results suggest that the low ratings on the alveolars observed in Chapter 4 (see Figures 4.11-4.14) were not the result of the existence of residue retroflex cues in the alveolar stimuli that may compromise alveolar goodness ratings. In fact, the modified stimuli generally had slightly higher mean (although not at a significant level) and smaller variance than the corresponding naturally-produced stimuli. Therefore, the experimental procedure using Beren to modify naturally-produced retroflexes by filtering out noise band (see §4.7.3 for details) is a valid method that creates natural sounding alveolars.
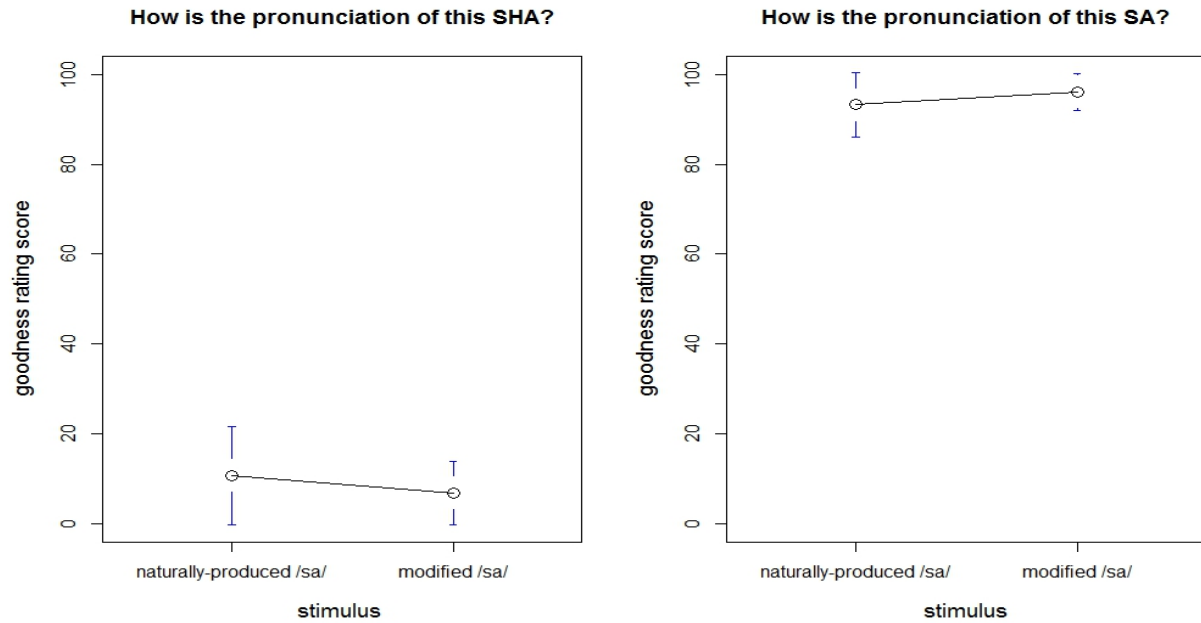
How is the pronunciation of this SHA?

How is the pronunciation of this SA?

Figure 5.1: Goodness rating of naturally-produced /sa/ and modified /sa/ for Beijing listeners.

How is the pronunciation of this SHA?

How is the pronunciation of this SA?
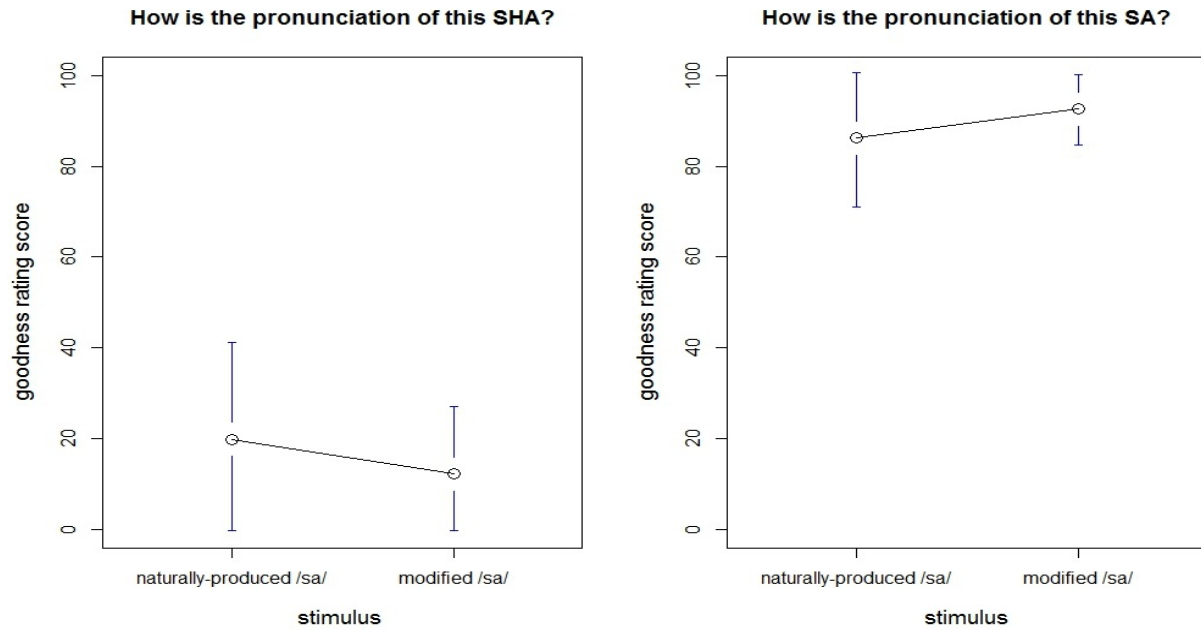
Figure 5.2: Goodness rating of naturally-produced /sa/ and modified /sa/ for Taiwan listeners.

How is the pronunciation of this SHU?

How is the pronunciation of this SU?
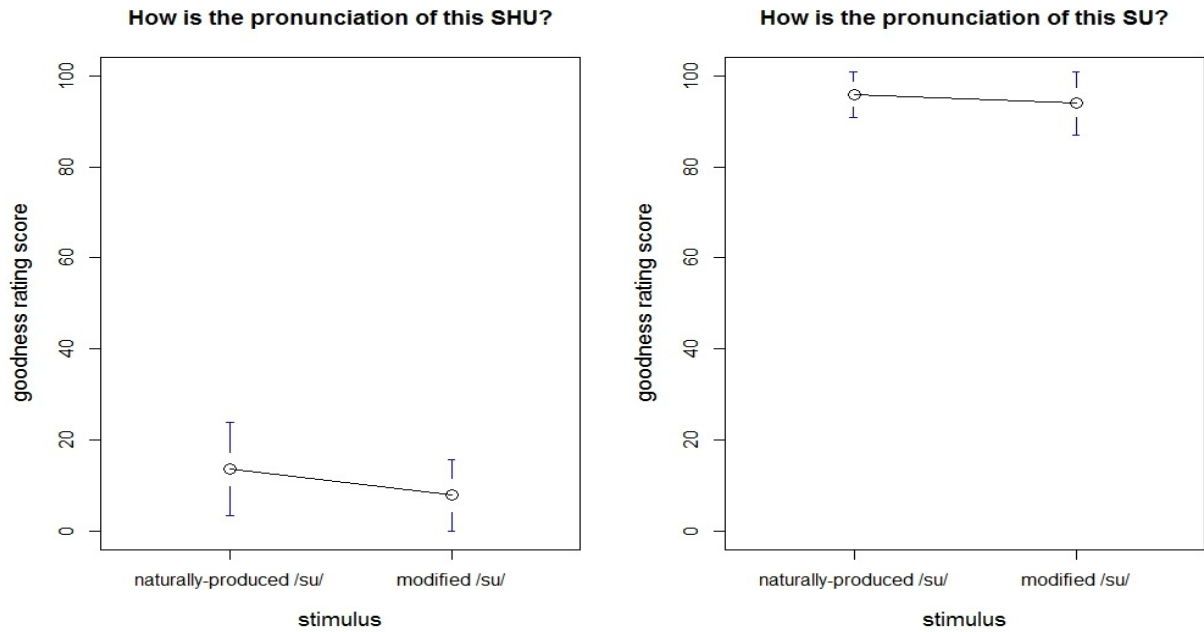
Figure 5.3: Goodness rating of naturally-produced /su/ and modified /su/ for Beijing listeners.

How is the pronunciation of this SHU?

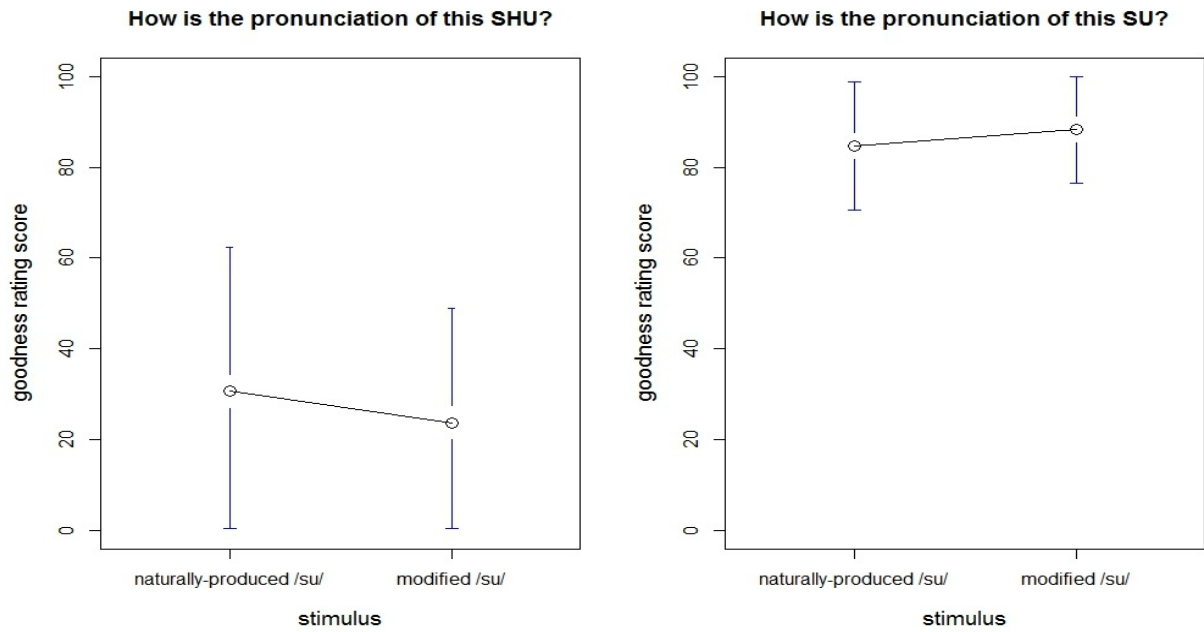**How is the pronunciation of this SU?**

Figure 5.4: Goodness rating of naturally-produced /su/ and modified /su/ for Taiwan listeners.

The goodness ratings of Beijing listeners and Taiwan listeners were plotted and displayed in Figures 5.5-5.8. The ABX task was not included in the follow-up experiment, therefore, the category boundary could not be located for further comparison among within-category variants. However, if the perceptual boundaries obtained in Chapter 4 were imposed on the goodness rating in the current experiment, gradient sensitivity to the alveolar variants could still be observed in Beijing listeners—goodness rating differed between step 6 and steps 8-10 along the /ʂa-sa/ continuum as well as between step 7 and steps 8-10 along the /ʂu-su/ continuum. As the dental tokens (steps 9 and 10) were rated equally good as some other alveolar variants, they were now considered as the best exemplars of /s/ for Beijing listeners. In accordance with the findings from Chapter 4, both Beijing and Taiwan listeners judged the retroflex, in either vowel context, to be represented by an exemplar cloud without a best category member. Overall, the goodness rating task in Chapter 4 and the current experiment share similar trends in that continuous changes along a given acoustic dimension result in both discrete and continuous changes. Another important finding from this follow-up experiment is that Mandarin listeners did not prefer hearing dental /s/ to alveolar /s/, as their goodness rating scores were not significantly different. Hence, the previously speculated hyperspace effect was not found.
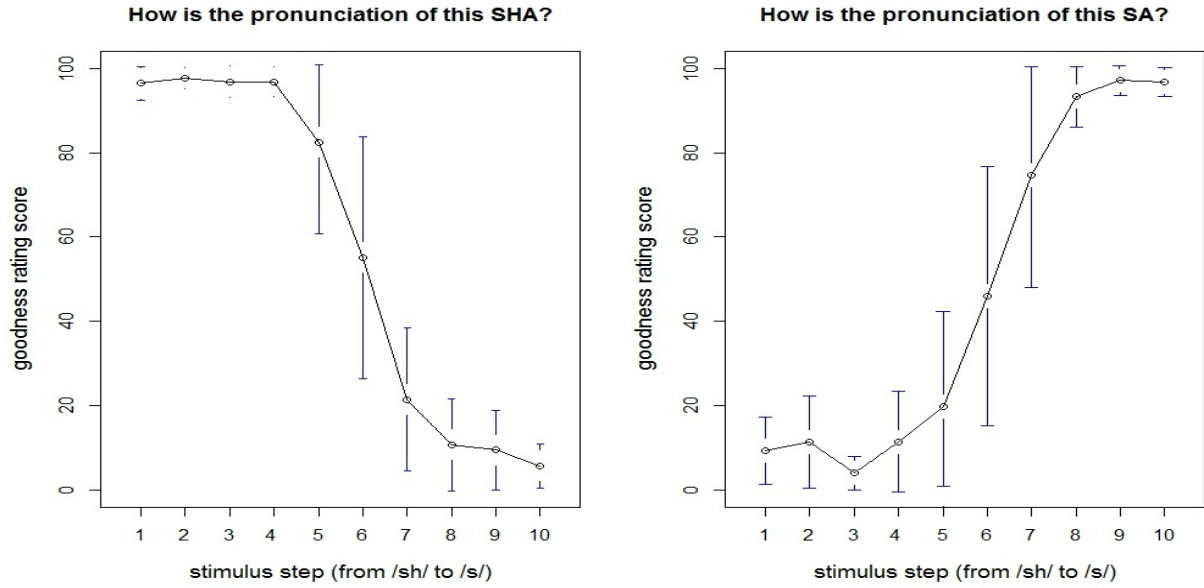
Figure 5.5: Goodness rating of /ʂa/ (left) and /sa/ (right) along a 10-step /ʂa-sa/ continuum for Beijing listeners (step 8 is the alveolar /s/; steps 9 and 10 are the dental /s/)
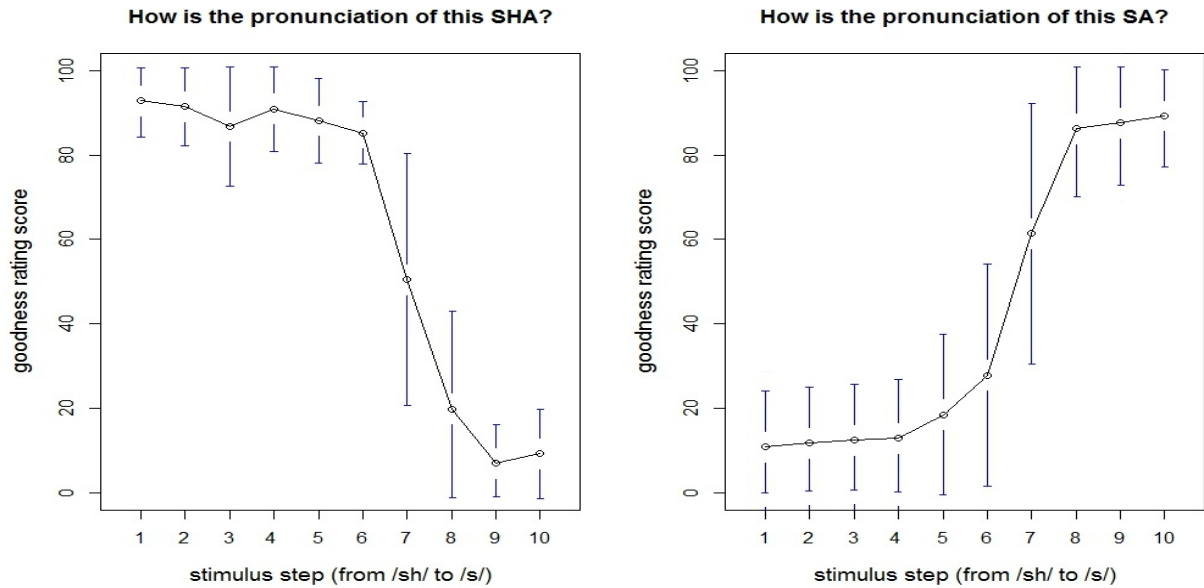

Figure 5.6: Goodness rating of /ʂa/ (left) and /sa/ (right) along a 10-step /ʂa-sa/ continuum for Taiwan listeners (step 8 is the alveolar /s/; steps 9 and 10 are the dental /s/)
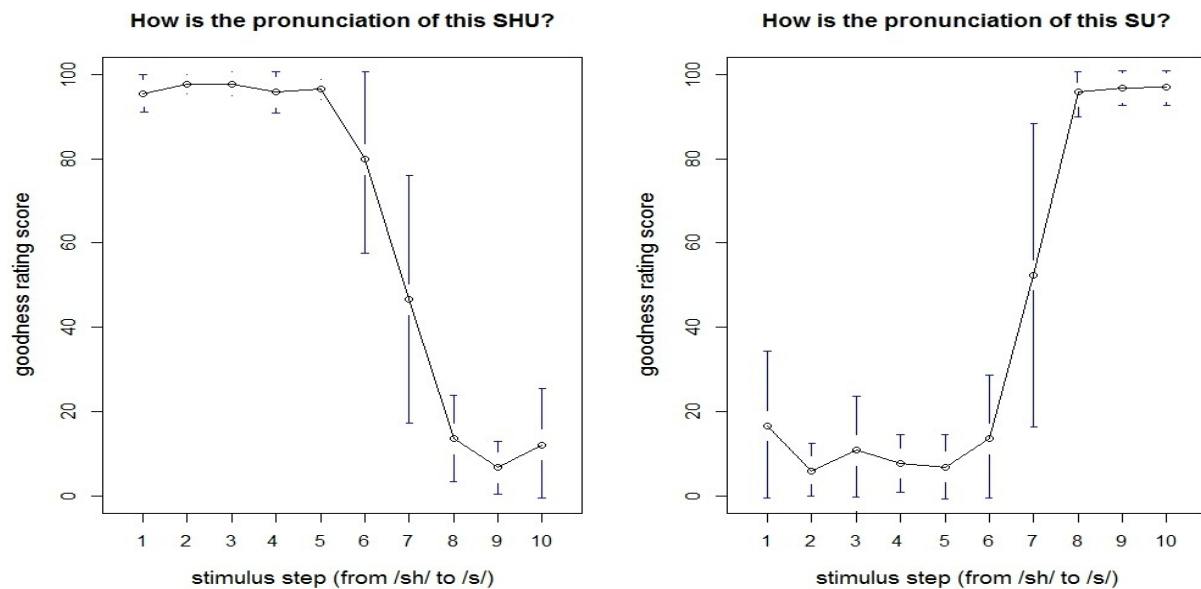
Figure 5.7: Goodness rating of /ʂu/ (left) and /su/ (right) along a 10-step /ʂu-su/ continuum for Beijing listeners (step 8 is the alveolar /s/; steps 9 and 10 are the dental /s/)
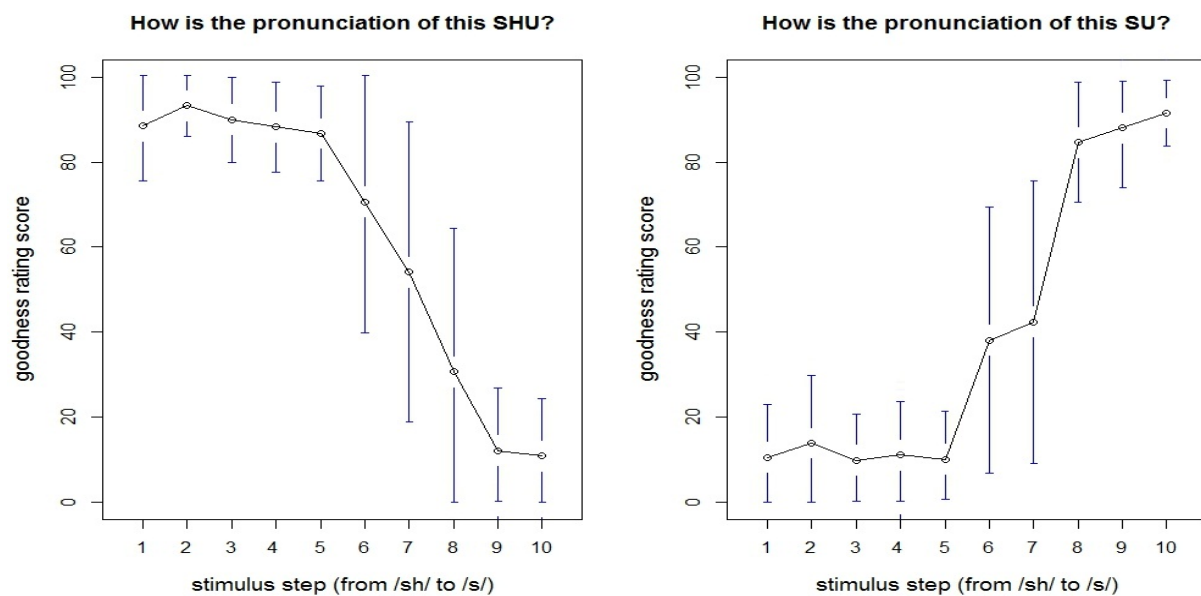

Figure 5.8: Goodness rating of /ʂu/ (left) and /su/ (right) along a 10-step /ʂu-su/ continuum for Taiwan listeners (step 8 is the alveolar /s/; steps 9 and 10 are the dental /s/)

The current experiment attempted to follow up on the discrepant goodness rating patterns observed in Chapter 4. The results confirmed that the modified alveolars stimuli were as good as naturally-produced ones and that dental /s/ was perceived as equally good exemplars as, and not better than, alveolar /s/.

CHAPTER 6: CONCLUSION

This dissertation investigates the cross-dialectal variation in production and perception of the Mandarin alveolar-retroflex contrast. The specific questions addressed in this investigation are: 1) Do the realizations of the two sibilant categories and the alveolar-retroflex contrast vary across vowel contexts in both Beijing Mandarin and Taiwan Mandarin? 2) Do the realizations of the two sibilant categories and the alveolar-retroflex contrast vary in different prosodic conditions for both Beijing Mandarin and Taiwan Mandarin? 3) Is there any interaction effect between the vowel contexts and prosodic prominence on the realizations of the two sibilants and the place contrast for both Beijing Mandarin and Taiwan Mandarin? 4) Do Beijing and Taiwan Mandarin speakers share the same alveolar-retroflex boundary in perception? 5) Does the location of the category boundary vary across vowel? 6) Do Beijing and Taiwan Mandarin speakers share the same best exemplars for the alveolar and retroflex categories? 7) Does the location of a category's best exemplars vary across vowel?

6.1 Summary of findings

The production study found that sibilant productions were variably affected by the vowel context in both dialects, which in turn renders different magnitudes of the alveolar-retroflex contrast. Overall, Beijing speakers produced a stronger degree of retroflexion than Taiwan speakers. However, Beijing speakers were found to only produce a significantly stronger contrast in the /a/ vowel context, and speakers of both dialects produced the least distinct place contrast in the /u/ context. No enlargement of the place contrast was found under contrastive focus: although focused productions were significantly longer than unfocused ones in duration, no increased spectral distinctiveness between the two contrasting phonemes was observed. This result

indicates that prosodic accentuation does not always give rise to an enhancement of a contrastive feature, especially the place feature.

The perception study found that Beijing Mandarin and Taiwan Mandarin listeners had different perceptual boundaries along the /ʂ-s/ continua, with Taiwan listeners' perceptual boundary located at higher frequencies than Beijing listeners. The location of the category boundary varied across vowel for both Beijing and Taiwan listeners. Particularly, the perceptual boundary shifted to lower frequencies in the /u/ context in compensation for the coarticulatory effect. In investigating the internal structure of each phoneme, selective sensitivity to fine-grained acoustic differences in the spectral noise distribution was found. Both Beijing and Taiwan Mandarin listeners perceived all retroflex variants as equally good category exemplars, regardless of vowel context. However, while Beijing listeners perceived the endpoint alveolar stimuli as the best exemplars of the alveolar category in both vowel contexts, Taiwan listeners did not show such sensitivity to the sub-phonemic variation of the alveolar category. Although some dialect-specific within-category sensitivity was observed, the overall findings support the quantal theory of speech perception.

6.2 Future research

As pointed out in Clopper and Pisoni (2005), linguistic variation between speakers due to regional differences is robust, yet only a scarce body of literature focuses on variation from the intersection between phonetic and sociolinguistic factors. With the study of the Mandarin alveolar-retroflex contrast, this dissertation has demonstrated that linguistic (i.e., vowel context) and sociolinguistic (i.e., dialect) factors collectively contribute to the phonological variation. As the detailed acoustic characteristics and perceptual patterns of the alveolar/retroflex sibilants in two major regional varieties of Mandarin are provided in this dissertation, the next step will be to

investigate how the variants of each category are represented in the phonology and if speakers employ this phonological knowledge in dialect identification and discrimination. Given that people in China and Taiwan are increasingly exposed to each other's dialect through the media and travel, whether familiarity with a different dialect causes listeners to develop multiple representations for a phoneme remains to be seen. Whether the dialectal stereotype is integrated into or even has greater influence than (i.e., overrides) the acoustic information listeners receive in dialect perception poses a question of interest as well. Two experimental paradigms are suggested to pursue this line of research. First, a semantic priming paradigm (e.g., Ernestus & Baayen 2007; Sumner & Samuel 2005) can be used to address the issue of representation of dialect variants produced by speakers of different dialects. Second, an alveolar/retroflex sibilant matching task with manipulation of dialect labels, in light of Niedzielski's (1999) study of Canadian Raising, can be conducted. The studies of dialect representation and categorization with careful acoustic manipulation can provide new insight concerning information of linguistic variation that is actually encoded in memory.

REFERENCES

Allen, J. B. (1977). Short-time spectral analysis, synthesis, with modifications, by discrete Fourier transform. *IEEE Transactions on Acoustical Speech and Signal Processing, 25*, 235-238.

Allen, J. B., & Rabiner, L. R. (1977). A unified approach to short-time Fourier analysis, synthesis. *Proceedings of the IEEE, 65*(11), 1558-1564.

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of Acoustical Society of America, 106*, 2031-2039.

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate. *Perception & Psychophysics, 63*, 798-810.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC map task corpus. *Language and Speech, 34*, 351-366.

Baxter, W. H. (1992). *A handbook of Old Chinese phonology*. New York: Mouton de Gruyter.

Beckman, M., Edwards, J., & Fletcher, J. (1991). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: gestures, segment, prosody* (pp. 68-86). Cambridge: Cambridge University Press.

Beckman, M. E., Pierrehumbert, J. B., Local, J., & Temple, R. (2004). Interpreting phonetic interpretation over the lexicon *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 13-38). Cambridge: Cambridge University Press.

Blevins, J. (2004). *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Bloomfield, L. (1993). *Language*. New York: Holt, Rinehart & Winston.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (v. 5.0.40). Amsterdam: Institute of Phonetic Sciences.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (v. 5.3). Amsterdam: Institute of Phonetic Sciences.

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

Brown, G., Ladd, D. R., & Cutler, A. (1983). Prosodic structure and the given/new distinction *Prosody: Models and Measurements* (pp. 67-78). Berlin: Springer Verlag.

Bruce, G. (2006). The phonetic profile of Swedish. *STUF - Language Typology and Universals, 59*(1), 12-35.

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

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*(6), e10729.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The Specific-Word Frequency Effect: Implications for the Representation of Homophones in Speech Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1430-1450.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica, 23*, 193-202.

Chang, Y.-H. (2011). A corpus study of retroflex realizations in Beijing and Taiwan Mandarin. *Proceedings of ICPhS XVII*, 440-443.

Chang, Y.-H. (2010). *Lip Rounding in Taiwan Mandarin Retroflex Sibilants.* Poster presented at the 87th Annual Meeting of the Linguistic Society of America.

Chang, Y.-H. (2011). The role of vowel transitions in the perception of Mandarin sibilants. *Proceedings of the 162nd meeting for the Acoustical Society of America*,130(4), 2570.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chao, Y.-R. (1930). A system of Tone Letters. *Le maitre phonetique, 45*, 24-27.

Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica, 4*, 363-397.

Chao, Y.-R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Chen, C.-Y. (1984). Neutral tone in Mandarin. *Journal of Chinese Linguistics, 12*(2), 299-333.

Chen, C.-Y. (1986). Salient segmental features in Singapore Mandarin. *Journal of Chinese Linguistics, 14*(1), 114-151.

Chen, C.-Y. (1991). The nasal endings and retroflexed initials in Peking Mandarin: instability and the trend of changes. *Journal of Chinese Linguistics, 19*(2), 139-171.

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese dialects*. New York: Cambridge University Press.

Chen, P. (1999). *Modern Chinese: history and sociolinguistics*. Cambridge: Cambridge University Press.

Chen, Y. (2003). *The phonetics and phonology of contrastive focus in Standard Chinese.* Ph.D. Dissertation. Stony Brook University.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese, *Journal of Phonetics*, *34*, 176-201.

Chen, Y. (2008). The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics, 36*(4), 629-648.

Chen, Z. (2010). *19 shiji yilai de fuzhou fangyan [The Fuzhou dialect since the 19th century]*. Fuzhou Shi: Fujian Renmin Publishing Co.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, R. L. (1968). Tone Sandhi in Taiwanese. *Linguistics, 41*, 19-42.

Chien, C. G. (1971). *A Contrastive Study of the Phonological Systems of Mandarin and Taiwanese.* Ph.D. Dissertation. Fu Jen University, Taipei, Taiwan.

Chiu, C. (2010). *Attentional weighting of Polish and Taiwanese Mandarin sibilant perception.* Proceedings of the 2010 Canadian Linguistics Association Annual Conference.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *The Journal of the Acoustical Society of America, 117*(6), 3867-3878.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and Aerodynamic Correlates of Korean Stops and Fricatives. *Journal of Phonetics, 30*(2), 193-228.

Choi, H. (2003). *Prosody induced acoustic variation in English stop consonants.* Proceedings of the 15th international congress of phonetic sciences.

Chuang, Y.-Y., & Fon, J. (2010). The Effect of Prosodic Prominence on the Realizations of Voiceless Dental and Retroflex Sibilants in Taiwan Mandarin Spontaneous Speech. *Proceedings of Speech Prosody 2010.*

Clopper, C. G., & Pisoni, D. B. (2005). Perception of dialect variation. In D. B. Pisoni, E.R. Robert (Eds.), *The Handbook of Speech Perception*. Malden: Blackwell Publishing Ltd.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics, 35*(2), 180-209.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269-384.

Da, J. (1998). Chinese Text Computing, http://lingua.mtsu.edu/chinese-computing/

Dart, S. (1991). Articulatory and Acoustic Properties of Apical and Laminal Articulations. *UCLA Working Papers in Phonetics, 79*.

de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*(1), 53-75.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*(1), 491-504.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. New York: Foris.

Docherty, G. J., & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. J. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 47-71). London: Arnold.

Dong, S. (1958). *Yuyin Changtan [Introduction to Phonetics]*. Beijing: Wenhua Jiaoyu Chubanshe.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages.* Ph.D. Dissertation, MIT.

Duanmu, S. (2000). *The phonology of standard Chinese*. New York: Oxford University Press.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2 ed.). New York: Oxford University Press.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224-238.

Ernestus, M., & Baayen, H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes, 22*, 1-24.

Feng, L., Lin, T., & Wang, L. (1985). Beijinghua yuliu zhong shengyundiao de shichang [Duration of initials, finals, and tones in Beijing dialect] *Beijing Yuyin Shiyanlu [Working Papers in Experimental Phonetics]* (pp. 131-195). Beijing: Beijing University Press.

Flanigan, B. O., & Norris, F. P. (2000). Cross-Dialectal Comprehension as Evidence for Boundary Mapping: Perceptions of the Speech of Southeastern Ohio. *Language Variation and Change, 12*, 175-201.

Fon, J., & Chiang, W. Y. (1999). What does Chao have to say about tones? -- a case study of Taiwan Mandarin. *Journal of Chinese Linguistics, 27*, 15-37.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84*(1), 115-123.

Fox, R. A., & Nissen, S. L. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives. *Journal of Speech, Language, and Hearing Research, 48*(4), 753-765.

Francis, A. L., Valter, C., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029-1044.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 28*, 67-73.

Fujisaki, H., & Kunisaki, O. (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions, 26*, 21-27.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Greenlee, M. (1992). Perception and production of voiceless Spanish fricatives by Chicano children and adults. *Language & Speech, 35*(1), 173-187.

Haley, K. L., Seelinger, E., Mandulak K.C., & Zajac, D.J. (2010). Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics*, *38*(4), 548-554.

Hamann, S. (2002). Postalveolar fricatives in Slavic languages as retroflexes. In S. Baauw, M. Huiskes & M. Schoorlemme (Eds.), *OTS Yearbook 2002* (pp. 105-127). Utrecht: Utrecht Institute of Linguistics.

Hamann, S. (2003). *The phonetics and phonology of retroflexes*. Utrecht: LOT Press.

Hamann, S. (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association, 34*(01), 53-67.

Hamann, S., & Fuchs, S. (2010) Retroflexion of Voiced Stops: Data from Dhao, Thulung, Afar and German. *Language and Speech, 53*(2), 181-216.

Hamilton, P. J. (1996). *Phonetic constraints and markedness in the phonotactics of Australian Aboriginal languages*. Toronto: University of Toronto.

Hardcastle, W. J., & Laver, J. (1997). *The Handbook of phonetic sciences*. Oxford: Blackwell Publishers.

Harris, K. H. (1958). Cues for discrimination of American English fricatives in spoken syllables. *Language and Speech, 1*(1-7).

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language & Speech, 1*(1), 1-7.

Hay, J. F., Seelinger, E., Mandulak, K. C., & Zajac, D. J. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study. *The Journal of the Acoustical Society of America, 119*(5), 3022.

He, K., & Li, D. (1987). *Xiandai Hanyu San Qian Qian Changyong Ci Biao [Three thousand most commonly used words in Modern Chinese]*. Beijing: Beijing Shifan Daxue Chubanshe.

Heniz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonant. *Journal of Acoustical Society of America, 33*, 589-596.

Hirschberg, J. (1993). Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence, 63*.

Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1990). A comparative investigation of Coarticulation in fricatives: electropalatograpahic, electromagnetic and acoustic data. *Language and Speech, 36*, 235-260.

Hsu, C.-S., & Jun, S.-A. (1998). Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?. *UCLA Working Papers in Phonetics, 96*, 69-89.

Hu, M. (1991). Beijinghua Nuguoyin Diaocha [Survey of women's speech in Beijing]. *Yuyianxue lunwenxuan [Digest of Linguistic Research]*, 230-243.

Hualde, J. I. (to appear). Sound change, from https://netfiles.uiuc.edu/jihualde/www/Hualde-Sound_change15.pdf

Huang, S. (1993). *Yuyan, shehui yu zuqun yushi-- Taiwan yuyan shehuixue de yanjiu [Language, society, and ethnic identity-- Studies in language socialogy in Taiwan]*. Taipei: Crane.

Huang, S., Bian, X. W. G., & McLemore, C. (1996). CALLHOME Mandarin Chinese Lexicon, from http://dla.library.upenn.edu/dla/olac/record.html?id=www_ldc_upenn_edu_LDC96L15.

Huang, Y. H., Wu, E. C., & Fon, J. (2012). The effect of Min proficiency on production and perception of tones in Taiwan Mandarin. *Proceedings of Speech Prosody 2012*.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of Acoustical Society of America, 28*, 303-310.

Janson, T., & Schulman, R. (1983). Non-distinctive features and their use. *Journal of Linguistics, 19*, 321-336.

Jeng, J.-Y. (2006). The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies, 40*(1), 27-48.

Jeng, J.-Y. (2009). The Auditory Discrimination of Mandarin Retroflex Contrasts and Spectral Moment Analysis. *Chinese Journal of Psychology, 51*(2), 157-174.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 824-843.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485-499.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*, 127-136.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language, 69*(3), 505-528.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In J. W. Mullenix & K. Johnson (Eds.), *Talker variability in speech processing* (pp. 145-166). San Diego: Academic Press.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*(3), 1252-1263.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W., Bybee, J., & Hopper, P. (2001). Probabilistic relations between words: Evidence from reduction in lexical production *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Karlgren, B. (1949). *The Chinese language: an essay on its nature and history*. New York: Ronald Press.

Kawasaki, H., Ohala, J. J., & Jaeger, J. J. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 81-103). Orlando, FL: Academic Press.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286-319.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech & Hearing Research, 15*, 453-473.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language, 82*, 33-62.

King, R. D. (1967). Functional Load and Sound Change. *Language, 43*(4), 831-852.

Kochanski, G., & Shih, C. (2003). Quantitative measurement of prosodic strength in Mandarin. *Speech Communication, 41*(4), 625-645.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: is there a return to normal. *Cognitive Psychology, 51*, 141-178.

Kratochvil, P. (1968). *The Chinese Language Today*. London: Hutchinson.

Kubler, C. C. (1985). The Influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156-176.

Kuhl, P. K., & Rubel, L. A. W. E. W. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293-332). Washington, DC, US: American Psychological Association.

Kurath, H. (1937). *Handbook of the linguistic geography of New England*. Providence, RI: Brown University Press.

Labov, W. (1973). Hypercorrection by the Lower Middle Class as a Factor in Linguistic Change. *Sociolinguistic patterns* (pp. 122-142). Philadelphia: University of Pennsylvania Press.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics phonology and sound change*. Berlin: Mouton de Gruyter.

Labov, W.(1998). The Intersection of Sex and Social Class in the Course of Linguistic Change In J. Cheshire & P. Trudgill (Eds.), *The Sociolinguistics Reader Volume 2: Gender and Discourse*. London: Arnold.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress*. Philadelphia: U.S. Regional Survey.

Ladefoged, P. (1994). The links between theory and fieldwork in phonetics. *UCLA Working Papers in Phonetics, 88*, 13-28.

Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Oxford: Blackwell Publishers.

Ladefoged, P., & Maddieson, I. (1996). *The sound of the world's languages*. Oxford: Blackwell Publishing.

Ladefoged, P., & Wu, Z. (1984). Places of Articulation: An Investigation of Pekinese Fricatives and Affricates. *Journal of Phonetics, 12*, 267-289.

Lai, Y.-H. (2009). Asymmetry in Mandarin affricate perception by learners of Mandarin Chinese. *Language and Cognitive Processes, 24*(7), 1265-1285.

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. New York: Cambridge University Press.

Lee, C. Y. (2003). An acoustic study of strident fricatives in Mandarin Chinese. *Journal of Acoustical Society of America, 113*(4), 2329.

Lee, C. Y., & Li, Z. (2003). *Enhancement of Phonological Contrast: Acoustics of Apical and Retroflex Vowels in Mandarin Chinese.* Paper presented at the NELS 34.

Lee, S. I. (2011). *An articulatory and acoustic investigation of Mandarin apical vowels.* Poster presented at the 162nd meeting of the Acoustical Society of America.

Lee, W. S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association, 33*(1), 109-112.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkley and Los Angeles: University of California Press.

Li, F. (2008). *The Phonetic Development of Voiceless Sibilant Fricatives in English, Japanese and Mandarin Chinese.* Ph.D. dissertation, Ohio State University.

Li, F., & Allen, J. B. (2011). Manipulation of Consonants in Natural Speech. *IEEE Trans. Audio, Speech and Language processing, 19*, 496-504.

Li, F., Menon, A., & Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *Journal of Acoustical Society of America, 127*(4), 2599-2610.

Li, F., Trevino, A., Menon, A., & Allen, J. B. (2012). A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise. *Journal of Acoustical Society of America, 132*(4), 1-13.

Li, F., Edwards, J., & Beckman, M. (2007). Spectral measures for sibilant fricatives of English, Japanese and Mandarin Chinese. *Proceedings of the XVIth International Congress of Phonetic Sciences*.

Li, J. (1923). *Xinzhu guoyu wenfa [New Mandarin Grammar]*. Shanghai: Shangwu Yinshuguan.

Li, Y. S. (2009). *Incomplete neutralization of alveolar and retroflexed sibilants in Taiwan Mandarin*. Paper presented at BLS 35.

Liao, S. W. (2010). *Identity, Ideology, and Language Variation: A Sociolinguistic Study of Mandarin in Central Taiwan*. Ph.D. Dissertation, University of California, Davis.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Lin, Y. H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology*. Ph.D. Dissertation, University of Texas.

Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge University Press.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of Acoustical Society of America, 99*, 1683-1692.

Lisker, L., & Abramson, A.S (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*, 1-28.

Lisker, L., & Abramson, A. S. (1985). Relative power of cues: F0 shift vs. voice timing In V. Framkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 25-33). New York: Academic Press.

Lu, Y. A. (2011). The psychological reality of phonological representations: the case of Mandarin fricatives. *Proceedings of the 23rd North American Conference on Chinese Linguistics*.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of Acoustical Society of America, 78*, 1949-1957.

Luce, P. A., McLennan, C. T., Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197-214). Oxford, England: Oxford University Press.

Luo, C., & Wang, J. (1981). *Putong Yuyinxue Gangyao [Outline of General Phonetics]*. Beijing: Shangwu Yinshuguan.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice-onset-time in word-initial stop consonants. *Journal of Child Language, 7*, 41-74.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of the [sh]-[s] distinction. *Perception & Psychophysics, 28*, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of Acoustical Society of America, 69*, 548-558.

Mattias, H. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics, 31*(1), 39-62.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33-B42.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271-285.

Miller, J. L. (1997). Internal Structure of Phonetic Categories. *Language and Cognitive Processes, 12*(5-6), 865-870.

Miller, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorisation. *Language and Cognitive Processes, 16*(5), 683-690.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*, 1157-1167.

Mo, Y., Cole, J., & Hasegawa-Johnson, M. (2009). Prosodic effects on vowel production> evidence from formant structure. *Interspeech 2009*.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of Acoustical Society of America, 102*, 1864-1877.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*, 661-677.

Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L (2011). On place assimilation in sibilant sequences—comparing French and English. *Journal of Phonetics, 39*(3), 429-451.

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

Nittrouer, S., & Whalen, D. H. (1989). The perceptual effects of child--adult differences in fricative-vowel coarticulation. *Journal of the Acoustical Society of America, 86*(4), 1266-1276.

Nowak, P. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics, 34*, 139-152.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. M.F. (Eds.), *Papers from the parasession on language and behavior* (Vol. 3, pp. 75-97). New York: Academic Press.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189-216). New York: Springer.

Ohala, M. (1994). Hindi. *Journal of the International Phonetics Association, 24*(1), 35-38.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology, 17*, pp. 273-281.

Palmer, M., Chiou, F. D., Xue, N., & Lee, T. K. (2004). Chinese Treebank Version 4.0, Linguistic Data Consortium, Philadelphia.

Peeters, W. J. M. (1991). *Diphthong Dynamics: A Cross-Linguistic Perceptual Analysis of Temporal Patterns in Dutch, English, and German.* Ph.D. Dissertation, Rijksuniversiteit te Utrecht.

Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. J., Beckman, M. E. (2007). Toward a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Perkell, J. S., & Nelson, W. L. (1985). Variability in production of the vowels /i/ and /a/. *Journal of Acoustic Society of America, 77*(5), 1889-1995.

Pierrehumbert, J, (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America, 55*(2), 328-333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285-290.

Pittman, A. L., & Stelmachowicz, P. G. (2000). Perception of Voiceless Fricatives by Normal-Hearing and Hearing-Impaired Children and Adults. *Journal of Speech, Language & Hearing Research, 43*(6), 1389.

Port, R. F., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*, 257-282.

Psychological Software Tools. E-Prime (v. 2.0). Pittsburgh, PA.

Ramsey, S. R. (1987). *The languages of China*. Princeton: Princeton University Press.

Rau, D. H. V. (1996). Stratified Variation of Retroflexed Initials and the Teaching of Spoken Mandarin. *Proceedings of the First International Conference on Chinese Language Teaching.*

Recasens, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel Coarticulation. *Journal of Phonetics, 19*, 177-192.

Régnier, M. S., & Allen, J. B. (2008). A method to identify noise-robust perceptual features: application for consonants. *Journal of the Acoustical Society of America, 123*(5), 2801-2814.

Repp, B. H., Liberman, A. M., & Harnad, S. (1987). Phonetic category boundaries are flexible *Categorical Perception*. New York: Cambridge University Press.

Repp, B. H., & Lin, H. B. (1989). Effects of preceding context on discrimination of voice onset times. *Perception & Psychophysics, 45*(4), 323-332.

Shadle, C. H. (1985). *The acoustics of fricative consonant* (Report No. 506). Cambridge, MA: Massachusetts Institute of Technology.

Shih, C. (1986). *The Prosodic Domain of Tone Sandhi in Chinese.* Ph.D. dissertation, University of California at San Diego.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers in Cornell Phonetics Laboratory, 3*, 83-109.

Shih, C. (1995). Study of vowel variations for a Mandarin speech synthesizer. *Proceedings of Eurospeech '95*.

Shih, C., & Ao, B. (1997). Duration study for the Bell Laboratories Mandarin Text-to-Speech system. In van J. Santen, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 383-399).

Shih, C., & Möbius, B. (1998). Contextual effects on voicing profiles of German and Mandarin consonants. *Proceedings of the 3rd International Workshop on Speech Synthesis*, 81-86.

Shih, C., Möbius, B., & Narasimhan, B. (1999). Contextual effects on consonant voicing profiles: a cross-linguistic study. *Proceedings of ICPhS XIV*, 989-992.

Shih, C., & Sproat, R. (1996). Issues in text-to-speech conversion for Mandarin. *Computational Linguistics and Chinese Language Processing, 1*(1), 37-86.

Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America, 123*(5), 2769-2779.

Simonsen, H. M. I., & Cowen, S. (2008). Norwegian retroflex stops in a cross linguistic perspective. *Journal of Phonetics, 36*, 385-405.

Soli, S. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of Acoustic Society of America, 70*(4), 976-984.

Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-46.

Stevens, K. (2000). *Acoustic Phonetics*. Cambridge: The MIT Press.

Stevens, K., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language, 65*, 81-106.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics, 3*(4), 215-233.

Stevens, K. N., Keyser, S., & Kawasaki, J. (1986). Toward a phonetic and phonological theory of redundant features In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in Speech Processes* (pp. 426-449). Hillsdale: Lawrence Erlbaum.

Stevens, K. N., Li, Z., Lee, C. Y., Keyser, S. J. (2004). A note on Mandarin Fricatives and Enhancement. In G. Fant, H. Fujisaki, J. Cao & Y. Xu (Eds.), *From Traditional Phonology to Modern Speech Processing*. Beijing: Foreign Language Teaching and Research Press.

Su, H. Y. (2005). *Language Styling and Switching in Speech and Online Contexts: Identity and Language Ideologies in Taiwan.* Ph.D. Dissertation, University of Texas.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1074-1095.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322-338.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487-501.

Sun, Y.-C. (2011). A study on retroflex articulation: Evidence from Sichuan dialect in China. *Asia-Pacific Science and Culture Journal, 1*(2), 14-21.

Surendran, D., & Niyogi, P. (2003). *Measuring the Functional Load of Phonological Contrasts* (Report No. TR-2003-12): Department of Computer Science, University of Chicago.

Svantesson, J. O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics, 14*(1), 53-70.

Tabain, M. (2009). An EPG study of the alveolar vs. retroflex apical contrast in Central Arrernte. *Journal of Phonetics, 37*(4), 486-501.

Thomas, E. R. (2001). An Acoustic Analysis of Vowel Variation in New World English *Publication of the American Dialect Society 85* (Vol. 85). Durham, NC: Duke University Press.

Thomas, E. R. (2002). Sociophonetic approaches of speech perception experiments. *American Speech, 77*, 115-147.

Tseng, C.-C. (2004). *Prosodic Properties of Intonation in Two Major Varieties of Mandarin Chinese: Mainland China vs. Taiwan.* Proceedings of the International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages, Beijing.

Tseng, C.-Y. (1988). Some stress related acoustic features of disyllabic words in Mandarin Chinese. *Bulletin of The Institute of History and Philology, Academia Sinica, Taiwan, R.O.C., 59*(3).

Tseng, S. C. (2005). Contracted syllables in Mandarin: evidence from spontaneous conversations. *Language and Linguistics, 6*(1), 153-180.

Umeda, N. (1977). Consonant duration in American English. *Journal of Acoustical Society of America, 61*, 846-858.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*(2), 723-735.

Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America, 120*(4), 2267-2277.

Wan, I. P. (2010). Phonological experiments in the study of palatals in Mandarin. *Journal of Chinese Linguistics, 38*, 157-174.

Wang, L. (1980). *Hanyu Yinyun [Chinese Phonology]*. Beijing: Zhonghu Shuju.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics, 32*(2), 251-276.

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic category. *Journal of Acoustical Society of America, 95*, 2694-2701.

Whalen, D. H. (1981). Effect of vocalic formant transitions and vowel quality on the English [s]-[sh] boundary. *Journal of Acoustical Society of America, 90*(4), 275-282.

Willis, C. (1972). Perception of Vowel Phonemes in Fort Erie, Ontario, Canada, and Buffalo, New York: An Application of Synthetic Vowel Categorization Tests to Dialectology. *Journal of Speech & Hearing Research, 16*, 246-255.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of Acoustical Society of America, 60*, 1381-1389.

Wu, C.-H. (2011). *The evaluation of second language fluency and foreign accent.* Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Wu, D. (2012). *A Durational Analysis of the Suffix "er" in Mandarin.* M.A. qualifying paper, University of Illinois at Urbana-Champaign.

Wu, Z. (1963). *The Collection of Illustrative Plates of Mandarin*. Beijing: Commercial Press.

Wu, Z., & Cao, J. (1979). Putonghua fuyin shengxue tezheng de jige wenti [Some questions about the acoustic characteristics of Mandarin consonants]. *Di Er Jie Quanguo Shengxue Xueshu Lunwen Zhaiyao [Second National Acoustic Academic Thesis Digest]*.

Wu, Z., & Lin, M. (1989). *Overview of experimental phonetics*. Beijing: Higher Eduction Press.

Xu, S. (1980). *Putonghua Yuyin Zhishi [Phonology of Standard Chinese]*. Beijing: Wenzi Gaige Chubanshe.

Xu, Y. (1993). *Contextual tonal variation in Mandarin Chinese*. Ph.D. Dissertation, University of Connecticut.

Yuan, J. (1989). *Hanyu Fangyan Gaiyao [Outline of Chinese dialects]*. Beijing: Wenzi Gaige Chubanshe.

Yuan, J., & Liberman, M. (2008). *Speaker identification on the SCOTUS corpus.* Journal of Acoustical Society of America, *123*(5), 3878

Zhang, Z. (1988). *Tone and tone sandhi in Chinese.* Ph.D. Dissertation, Ohio State University, Columbus.

Zhou, D., & Wu, Z. (1963). *Articulation album of Putonghua*. Beijing: Shangwu Yingshuguan.

Zhu, L. (2012). *Retroflex and Non-retroflex Merger in Shanghai Accented Mandarin.* M.A. Thesis, University of Washington.

Zhu, Y. K. (1998). Lanqing Guanhua shuolue [An Overview of Non-standard Mandarin]. *Language Studies, 67*(2), 56-60.

APPENDIX A: MANDARIN NORTH WIND AND THE SUN TEXT

有一次，北風跟太陽正在爭論誰的本領大。他們正好看到有個人走過，那個人身上穿了一件厚袍子。他們倆就商量好了，說，誰能先叫這個過路的把他的袍子脫下來，就算是他的本領大。於是，北風就卯足了勁，拼命的吹。可是，他吹得越厲害，那個人就把他的袍子裹得越緊。到最後，北風沒輒了，只好就算了。一會兒，太陽出來一曬，那個人馬上就把袍子脫了下來。所以，北風不得不承認，還是太陽比他的本領大。

(adapted from Lee & Zee 2003)

APPENDIX B: LANGUAGE BACKGROUND FORM

Participant's student ID number:_____   Date:_____

Participant information

Please provide the following information (please circle the options that apply to you)

1. Hearing, Vision, and Speech
1. Do you have normal hearing? Yes   Corrected   No

2. Do you have normal vision? Yes   Corrected   No

3. Do you have normal speech? Yes   Corrected   No

2. General Background
1. Gender: Male   Female

2. Approximate age: 18-25   25-30   30-35   35 and up

3. Which hand do you use for writing? Right   Left

4. Education: high school   college   Master's   Ph.D.

3. Language Background
1. Which city were you born and raised in?

2. How long have you lived in the United States?

4. What is (are) the first language(s) you learned to speak fluently?

3. What language(s) did your parents speak to you while you were growing up?

4. Do you speak any other dialects or foreign languages fluently? Yes   No
(If yes, please list them here.)

APPENDIX C: SAMPLE MAPS

For the experimenter:
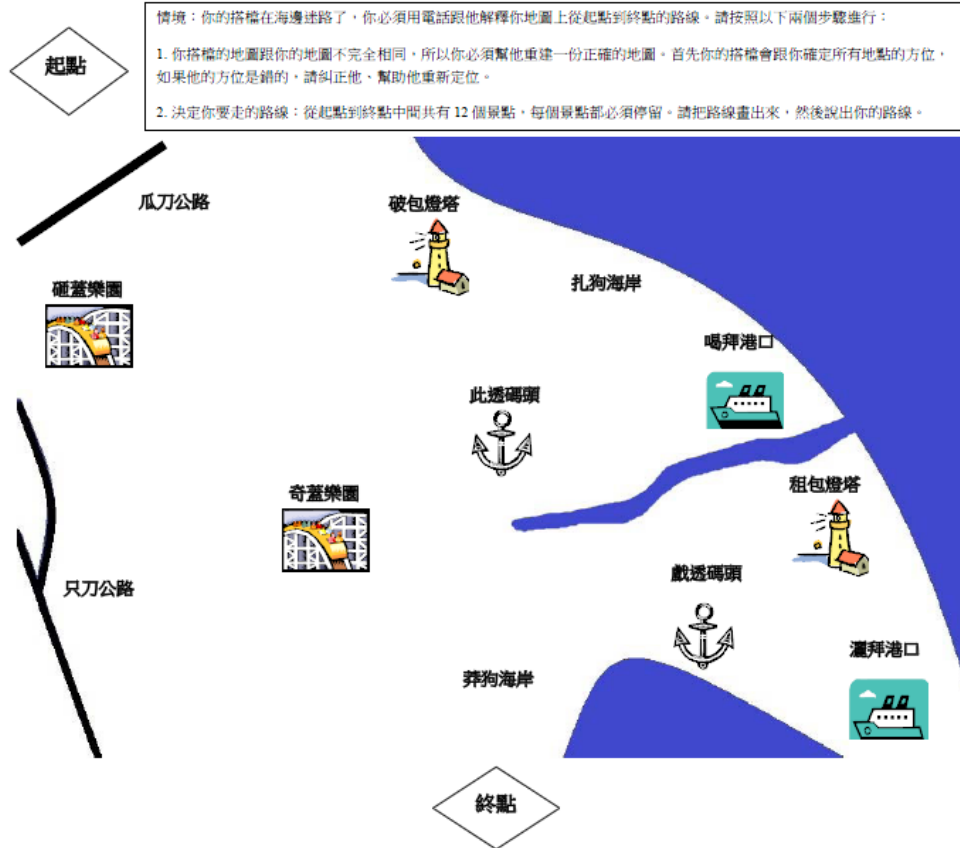
For the participant:
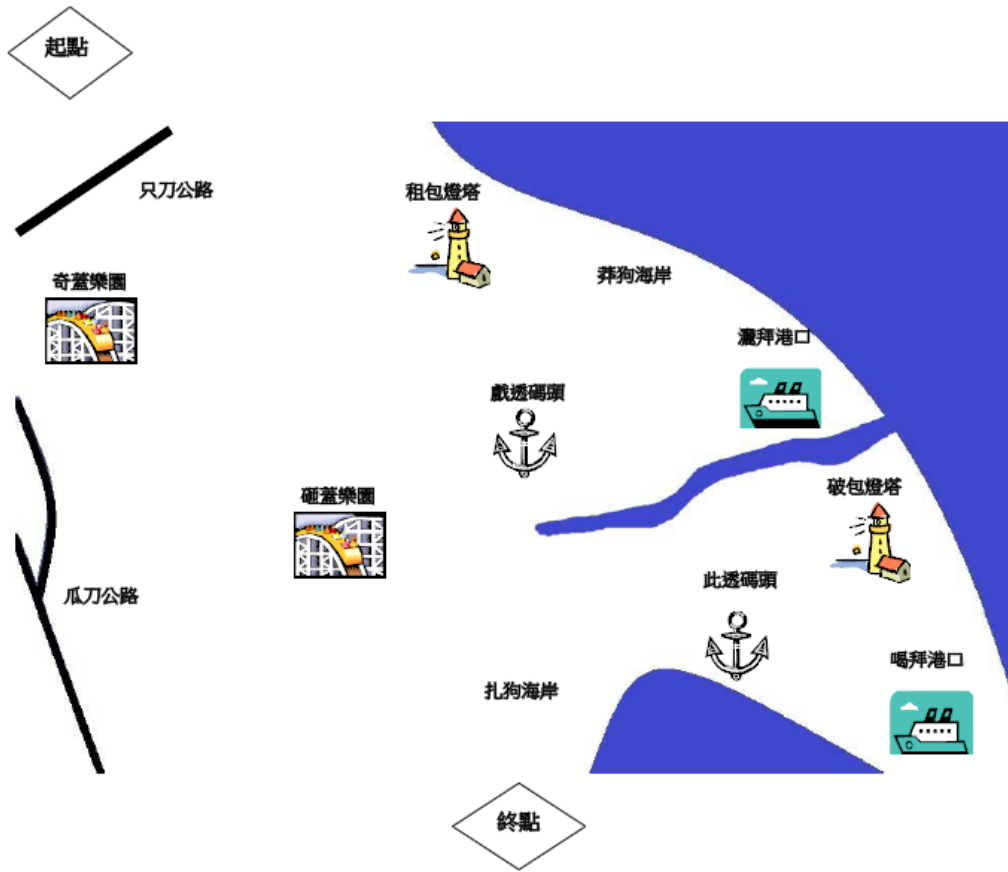
APPENDIX D: WORDLIST

Map task #	Theme	Tokens	Fillers
1	park	薩百懸崖 組逃洞穴 蘇靠瀑布 字代別墅 子告峽谷 次跑河流	婆百懸崖 藍逃洞穴 濫靠瀑布 七代別墅 西告峽谷 羅跑河流
2	beach	扎狗海岸 灑拜港口 租包燈塔 此透碼頭 只刀公路 砸蓋樂園	莽狗海岸 喝拜港口 破包燈塔 戲透碼頭 瓜刀公路 奇蓋樂園
3	campus	眨杯陽台 死派小學 撒套圍牆 楚夠禮堂 插保郵局 足搞大廈	喊杯陽台 孿派小學 伯套圍牆 怡夠禮堂 器保郵局 寡搞大廈
4	resort	炸斗飯店 速排舞廳 直高木屋 廚討噴泉 持考劇院 失拍街道	默斗飯店 寒排舞廳 方高木屋 窩討噴泉 意考劇院 夾拍街道
5	castle	至凱雕像 逐該賓館 叔偷閣樓 祝鬥皇宮 傻呆墓地 擦陪畫廊	賀凱雕像 亂該賓館 摸偷閣樓 過鬥皇宮 家呆墓地 喜陪畫廊

APPENDIX E: PLOTS OF INDIVIDUAL SPEAKER DATA

Beijing speakers:

BM3

BM4

Taiwan speakers:

TF1

TF2

133
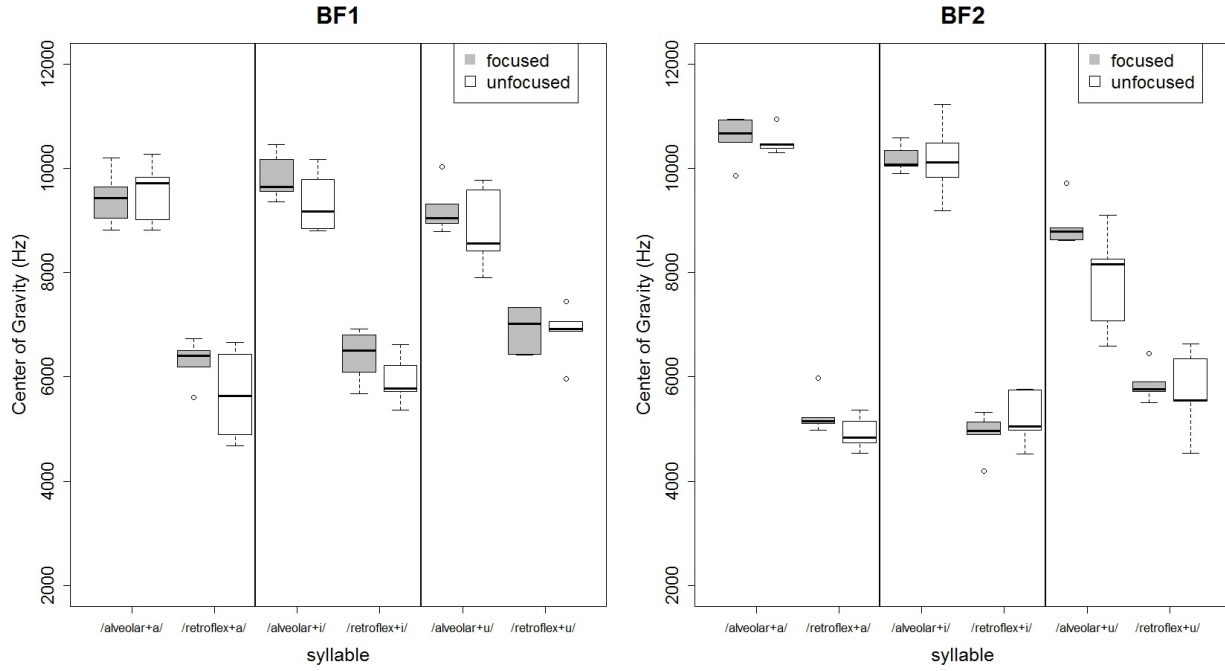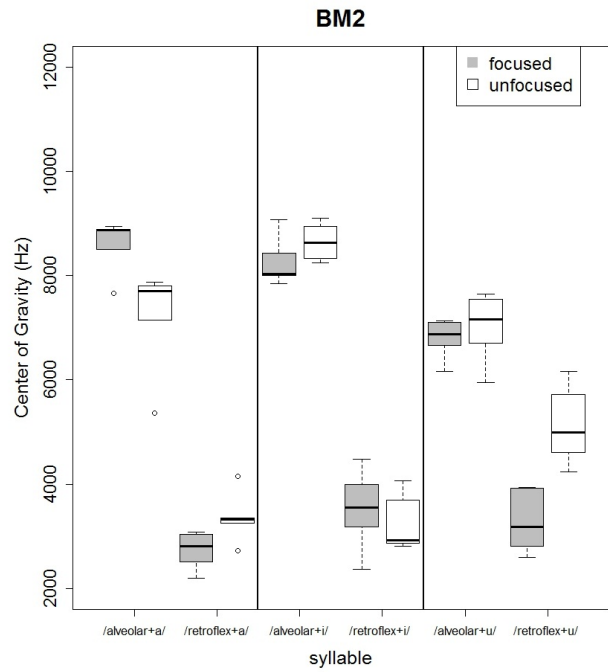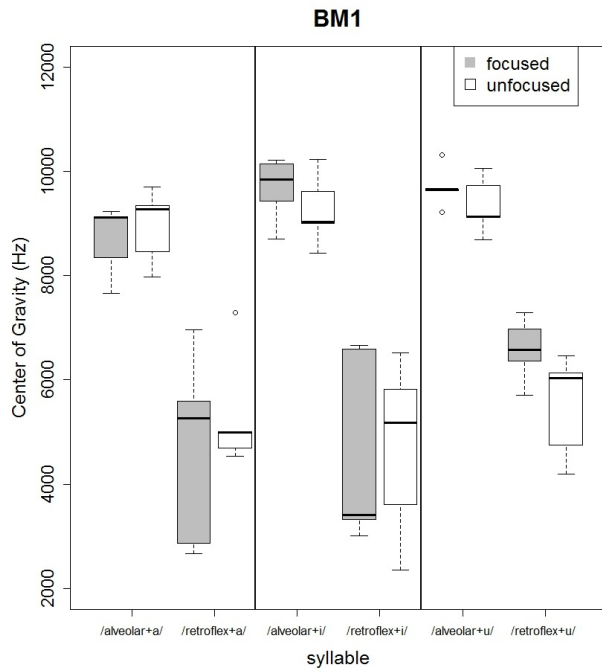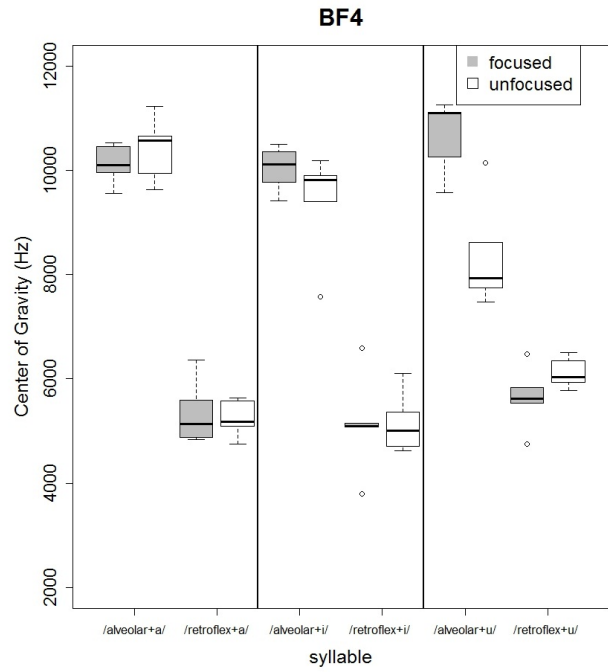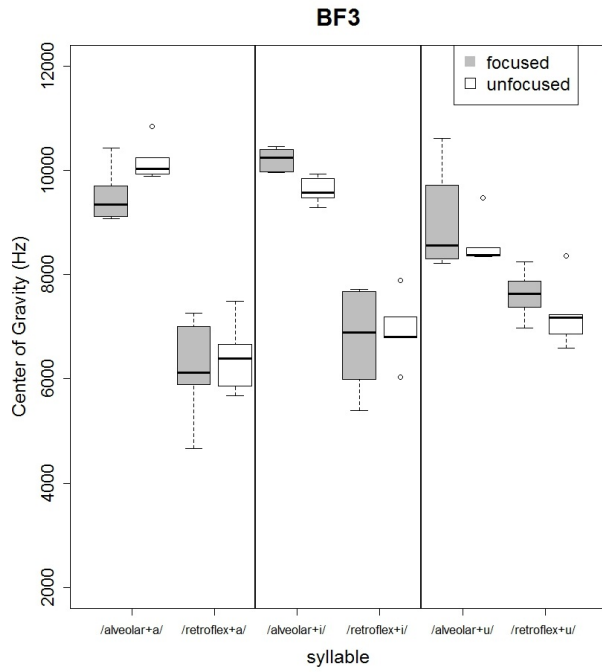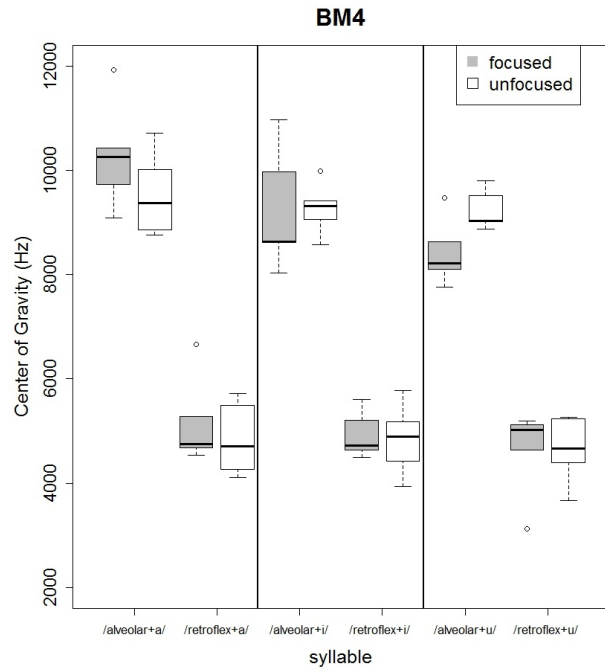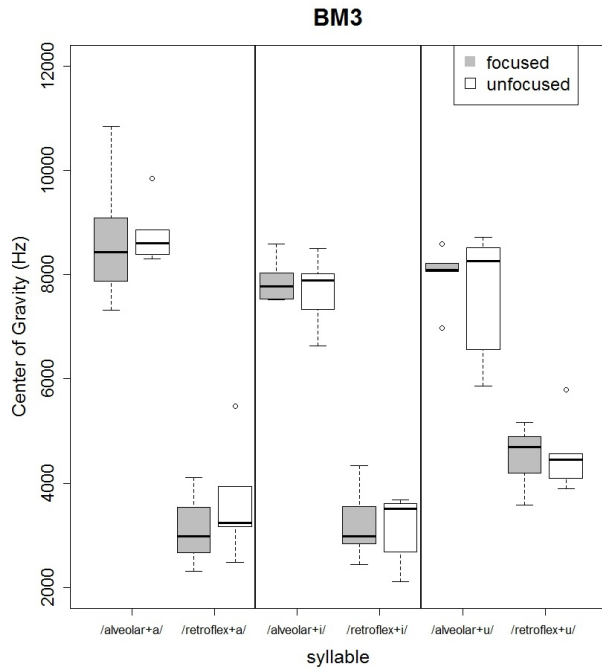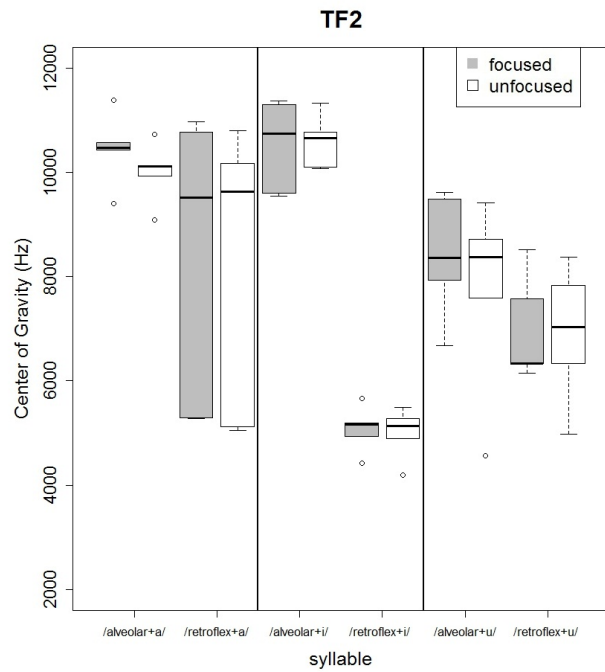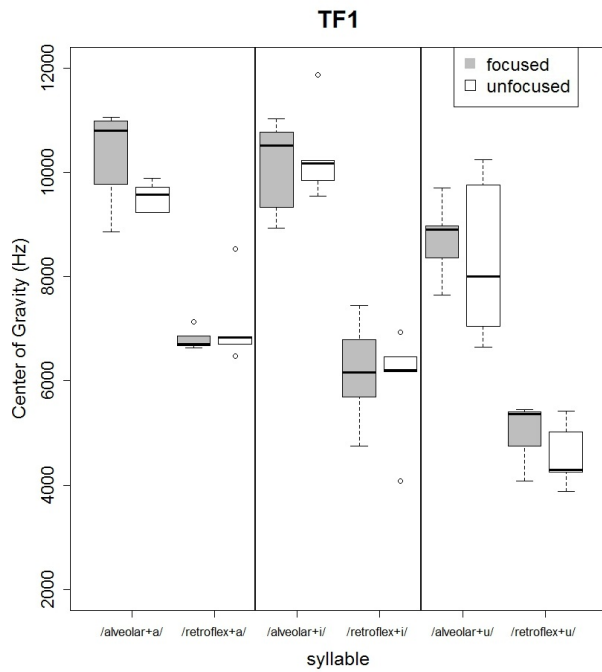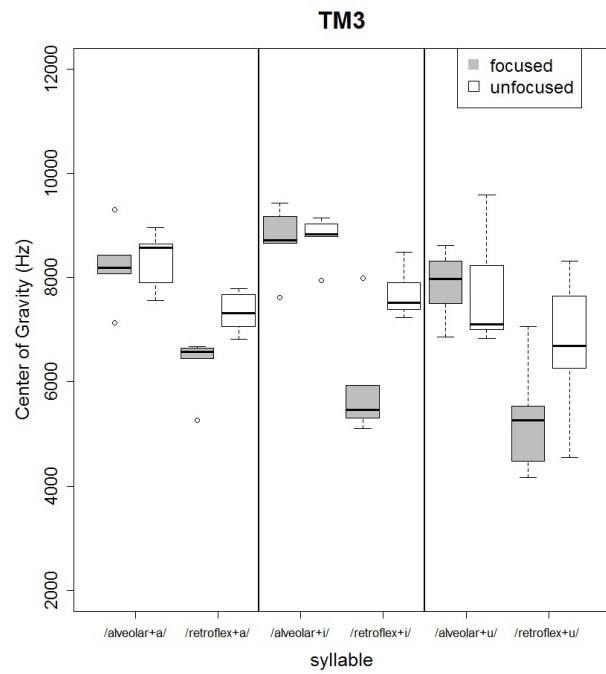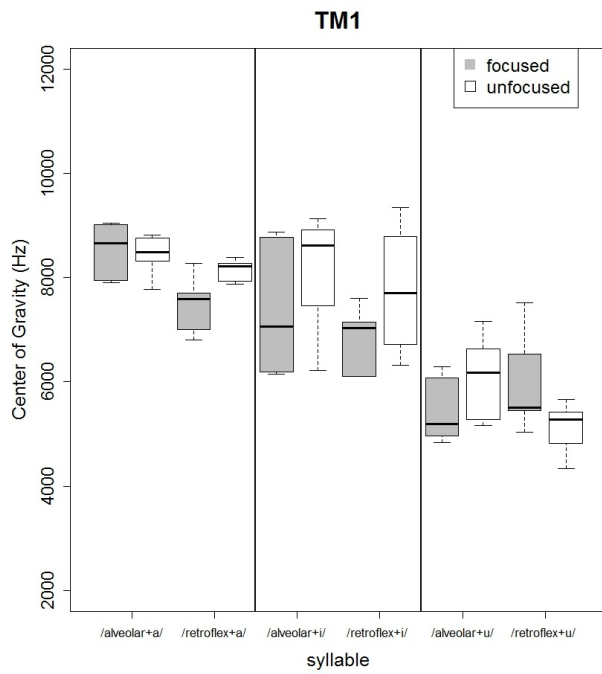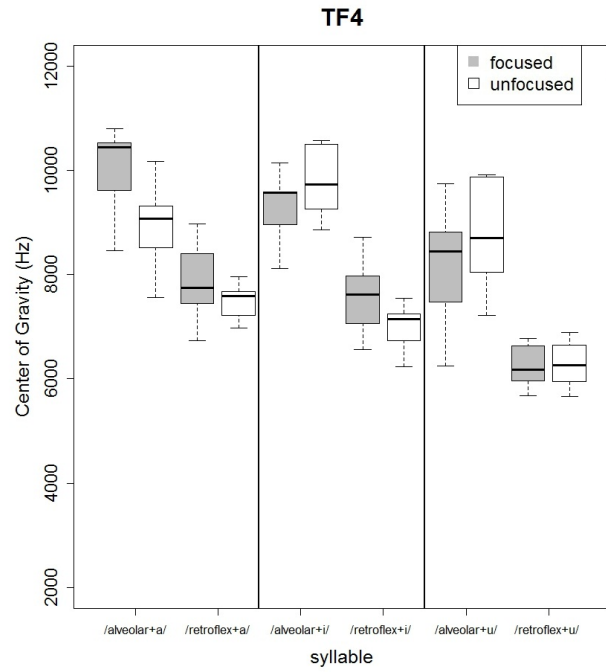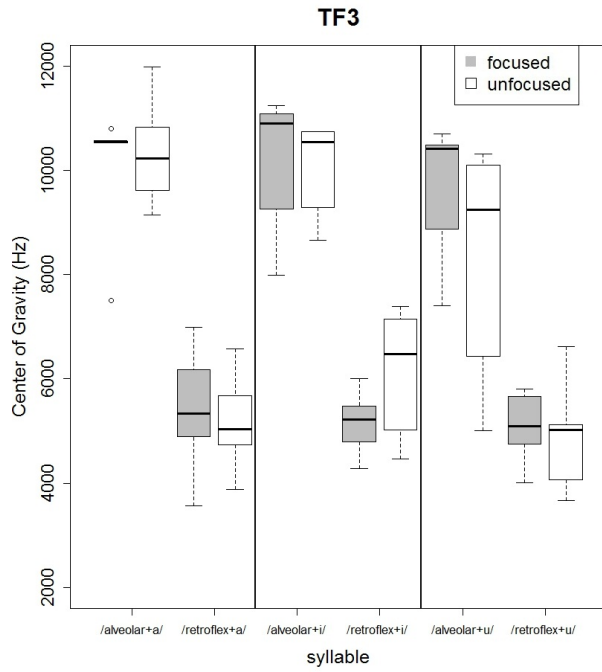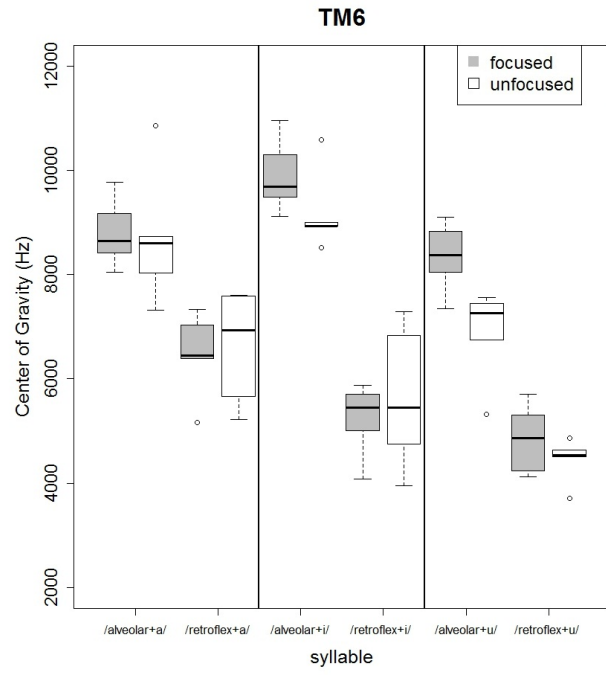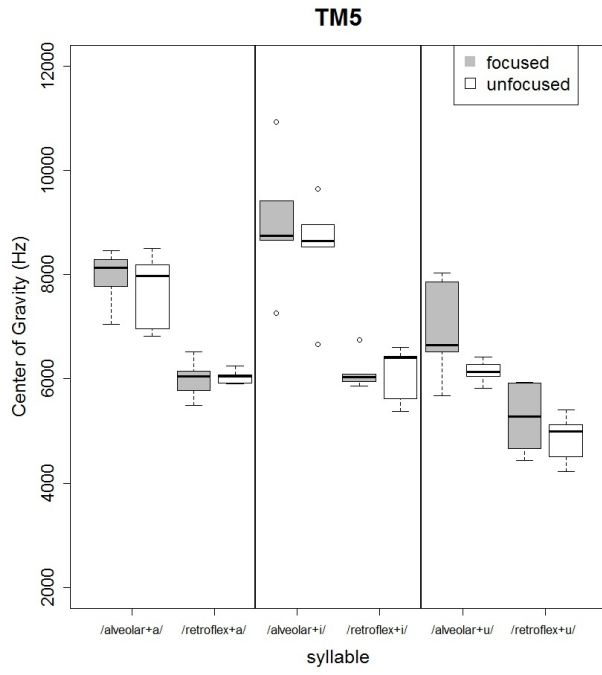
APPENDIX F: WORDLIST

衣 屋 淤 阿 撒 沙 苏 书 司 师 西 需 低 踢 督 秃 搭 他 修 虾 哭 估 趴 八

APPENDIX G: FORMANT TRAJECTORIES FOR THE 8 STEP /i-y/

CONTINUUM
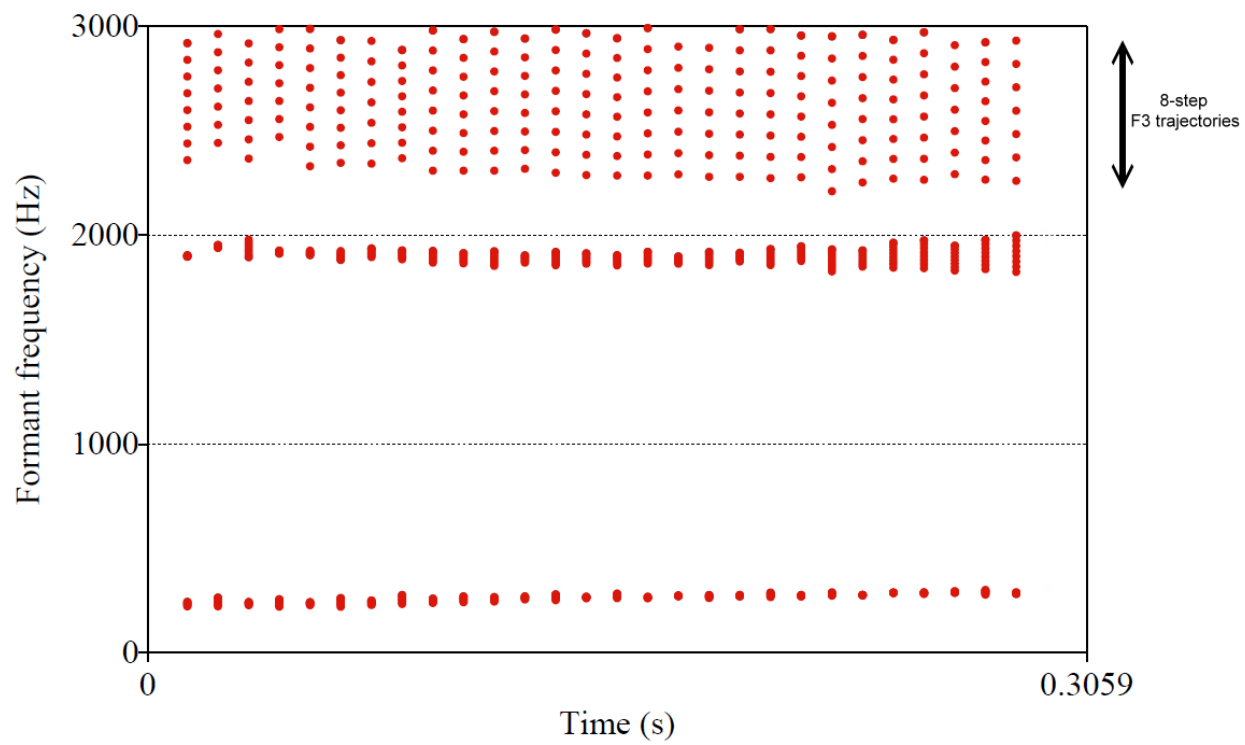
APPENDIX H: THE WAVEFORM OF /ti/

The 8-step /ti-di/ continuum was created by splicing off the aspiration (the framed portion in the waveform) from /ti/ at 7 equal time intervals into /di/.