Sibilant Contrast: Perception, Production, and Sound Change

By

Mingxing Li

M.Phil., The Chinese University of Hong Kong, 2012
M.A., Tianjin Normal University, 2004
B.A., Tianjin Normal University, 2001

Submitted to the graduate degree program in Linguistics and the Graduate Faculty
of the University of Kansas in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Chair: Dr. Jie Zhang

Dr. Allard Jongman

Dr. Joan Sereno

Dr. Annie Tremblay

Dr. Crispin Williams

Date Defended: June 20, 2017

The dissertation committee for Mingxing Li certifies that
this is the approved version of the following dissertation:

Sibilant Contrast: Perception, Production, and Sound Change

Chair: Dr. Jie Zhang

Date Approved: July 19, 2017

Sibilant Contrast: Perception, Production, and Sound Change

Abstract

This study examines sibilant place contrast in the [_i] context in terms of its typology across Chinese dialects and its role in the historical development of Mandarin sibilants. The typology across 170 Chinese dialects reveals that (i) for dialects that have sibilants at three places (dental, palatal, and retroflex), place contrasts in the [_i] context are generally avoided, e.g., */si-ɕi-ʂi/; (ii) for dialects that have sibilants at two places, mostly dental vs. palatal, place contrasts in the [_i] context also tend to be avoided, e.g., */si-ɕi/; (iii) for dialects that do have contrastive dental vs. palatal sibilants in the [_i] context, the place contrast of affricates implies that of fricatives. The first two patterns are mirrored in the sound changes of Mandarin in that contrastive dental and palatal sibilants in the [_i] context that emerged from independent processes have always been enhanced or avoided. In addition, the sound changes also showed an avoidance of contrastive dental vs. palatal in the [_i] context with the shift of palatal sibilants into retroflex sibilants from the 11th to the 14th century.

The connection between the synchronic typology and the diachronic changes raises a number of research questions: (i) Does the vowel context affect the perceptual distinctiveness of sibilant place contrasts, e.g., is [si-ɕi] less distinct than [sa-ɕa]? (ii) Do place contrasts differ in perceptual distinctiveness, e.g., is [si-ɕi] less distinct than [si-ʂi]? (iii) Do different manners of articulation differ in perceptual distinctiveness, e.g., is [tsi-tɕi] less distinct than [si-ɕi]? These issues were investigated through a speeded AX-discrimination experiment, which has been

shown to be able to evaluate the relative perceptual distinctiveness of sound pairs independent of the listener's native phonology. Twenty-nine listeners were put under time pressure to judge if a CV pair is the same or not, where the sibilant onsets of the CV pairs contrast in place (e.g., [si-ɕi]) and the vowels were [i] vs. other vowels (e.g., [si-ɕi] vs. [sa-ɕa] vs. [sɹ̩-ɕi]). Assuming that a longer response time indicates less perceptual distinctiveness, the results showed that (i) the [_i] context reduces the perceptual distinctiveness of the place contrasts of dental vs. palatal sibilants; (ii) the introduction of the apical vowel enhances the perceptual distinctiveness between the contrastive sound pairs; (iii) the dental vs. retroflex contrasts are more distinct than the dental vs. palatal contrasts. These findings match the observations in the cross-linguistic typology and the historical development of Mandarin and support the claim that perceptual distinctiveness regulates the phonological system.

The reduced distinctiveness of dental vs. palatal sibilants in the [_i] context suggests that contrastive dental vs. palatal sibilants are unstable and are likely to be avoided in sound change. A phonetic study was conducted on the sibilants in Xiangtan, a Chinese dialect reported to have the same sound system as 18th century Mandarin (i.e., [sɹ̩ si ɕi ʂɻ̩]) with fully contrastive dental vs. palatal sibilants in the [_i] context. It is predicted that in Xiangtan, the pre-[i] dentals in /si tsi ts^{h}i/ may show signs of being palatalized and thus neutralized with the palatals. Natural productions of /si tsi ts^{h}i/, /ɕi tɕi tɕ^{h}i/, and /sɹ̩ tsɹ̩ ts^{h}ɹ̩/ syllables with matched tones were recorded from 11 native female speakers of Xiangtan. Center of gravity, energy dispersion, intensity, and duration were extracted for three types of sibilants: Canonical dentals as in /sɹ̩/, canonical palatals as in /ɕi/, and pre-[i] dentals as in /si/. A discriminant analysis was performed by first training a classifier on the canonical dentals and canonical palatals and then using the classifier to predict the place (dental vs. palatal) of the pre-[i] dentals. Native Mandarin listeners were also recruited to identify the isolated first half of the pre-[i] dental sibilants as being dental vs. palatal. The results from both studies showed that (i) some Xiangtan speakers have palatalized the dentals in /si tsi ts^hi/, and (ii) certain speakers variably produce dental and palatal sibilants for the same lexical item. Therefore, the results support the contention that dental and palatal contrasts are perceptually less distinct in the [_i] context and the variation in the realization of the pre-[i] dentals indicates that a merger replicating the development of Mandarin is in process.

In general, the perceptual experiment reveals that dental vs. palatal sibilants in the [_i] context form weak contrasts, based on the psychoacoustic similarity of the contrastive elements. The avoidance of weak contrasts is observed in cross-linguistic typology, historical sound change, and speech production. This study thus establishes an empirical connection among the perceptual distinctiveness of sibilant place contrasts, the production of these contrasts, cross-linguistic typology, and historical sound changes.

This dissertation is dedicated to:

Prof. Jialing Wang (1934–2008)

My first mentor in linguistics

Prof. Jie Zhang

My supervisor at KU and my role model

Yanpeng Tang

My wife and my strongest support at all times

Acknowledgements

First and foremost, I have to thank my supervisor Prof. Jie Zhang for his guidance in my Ph.D. study and in doing this dissertation. By 'guidance', I mean listening to me patiently when my idea on the research was by no means clear, asking me questions that challenged my thinking, providing critical comments that forced me to see the bigger picture, and encouraging me to move on when my faith was low. 'Thank' is such a generic word to use and it so falls short of what I mean by it to him. He is the model I should always follow as a scholar and a supervisor.

I am also grateful to my other committee members, Profs. Allard Jongman, Joan Sereno, Annie Tremblay, and Crispin Williams. Their comments and suggestions have always inspired me and helped improve my research to a form I have never imagined. And I feel so fortunate to be able to work with a committee that always support me with knowledge, encouragement, and unending patience. Any errors or failings in this work remain my own and, to my committee members, I owe everything that is good.

One thing I will miss for sure after I leave KU is the weekly meeting of the LING 850 group, where I learnt a lot from my fellow students' research and got their valuable feedback on mine. Thanks to Kelly Berkson, Steve Politzer-Ahles, Hyunjung Lee, Jiang Liu, Goun Lee, Kate Coughlin, Quentin Qin, Yu-Fu Chien, Hanbo Yan, Wen-chi Yeh, Maite Martinez-Garcia, Yingjie Li, Charles Redmon, Katrina Connell, Rustle Zeng, Seulgi Shin, Kunning Yang, among others, for their advice on various parts of the research in this dissertation. It is so nice to work in such a linguistics community where everyone supports each other.

Thanks to Mr. Bin Ma, Dr. Sang-Im Lee-Kim, Prof. Qing Fang, Prof. Huifang Kong, Prof. Yuqing Wei, Prof. Zanhui Huang, among others, for their help in stimulus preparation and data collection. I would never forget that on my fieldwork trip to Xiangtan, Prof. Fang helped me obtain all the data I needed within one day. Thanks to Ms. Corinna Johnson, our office manager at KU Linguistics, for her help in the past five years.

My interest in phonology started sixteen years ago in the classroom of the late Prof. Jialing Wang (1934–2008), who was my first mentor in linguistics. I would never be here without his mentoring. His teaching has forged my interest in phonology and his expectation has been the source of power in my pursuit of linguistics in the past decade.

I owe a debt of gratitude to those who have helped in my life during my Ph.D. study. Thanks to Lian Hee Wee and Winnie Cheung for their timely help when I needed it most, without which I could not have survived in a new place. To Renee and Rich Grubbs for helping me settle down and offering continuous help. To Jia He, Jinfeng Li, Jiang Liu, Yingjie Li, Juqiang Zhang, Shuangteng He, Liyang Zhao, Sha Li, Lori and Graig Miller, Ann Kihm, among others, for their help and support in various aspects.

Thanks to my parents, Guoqin Zhang and Changlin Li, and in-laws, Guiru Tian and Zhiqiang Tang, for their support and patience all these years when I have been so stubborn on my way to become a linguist. And most importantly, to my wife Yanpeng Tang, who has always been with me as my strongest support through good times and bad. My daughter, Mia, was born during the last year of my Ph.D. study. I am so excited to find the happiness and strength she has brought to me.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Contrast distinctiveness and vowel context

A large amount of literature has shown that human languages prefer speech sounds that are more distinct from one another (Martinet 1952; Wang 1968; Liljencrants and Lindblom 1972; Lindblom 1986, 1990a; Flemming 2002, 2004, 2005; Boersma and Hamann 2008, among others) and that certain sound pairs form better contrasts than others. In terms of vowels, for example, a system like /i, a, u/ is believed to be preferable in terms of the perceptual space (e.g., Boersma and Hamann 2008, among others). In the cross-linguistic typology of vowel systems, this is borne out in that nearly all languages have these three vowels (Maddieson 1984:134). Simply put, languages generally prefer vowel contrasts that are more distinct.

In speech, the articulatory gestures for a particular segment usually overlap with the gestures of neighboring segments, and the acoustic properties and perceptual cues of a segment are thus influenced by its contexts (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967, Sereno, Baum, Marean and Lieberman 1987, among others). The perceptual information of a consonant often partially lies on its neighboring vowels, and the quality of the perceptual cues for a segment may differ depending on contexts. In terms of linguistic typology, this is mirrored in the observation that a particular phonological contrast may be licensed in the contexts where the perceptual cue for the contrast can be easily recovered and neutralized in the contexts where the perceptual cue is difficult to recover (Steriade 1997, 1999, 2001, 2008). For example, a distinction between an alveolar /t/ and a retroflex /ʈ/ is better cued by lowered F3 and F4 in a vowel-to-consonant transition and this is consistent with the observation that the /t/ vs. /ʈ/ contrast is typically licensed in the presence of a preceding vowel and neutralized in the absence of it (Steriade 1999).

Of particular interest to the current research is the observation in the phonological literature that, in certain languages, the same consonantal contrast can be licensed in one vowel context but phonetically avoided or shifted in another vowel context. Russian, for example, has the phonological contrast between plain vs. palatalized consonants before most vowels, as in (1a), whereas in the [_i] context, this contrast is phonetically realized as velarized vs. palatalized consonants, as in (1b) (Padgett 2001, 2003). According to Padgett (2001), such a 'contrast shift' was motivated by the relative difficulty to perceptually distinguish a plain vs. palatalized consonant in the [_i] context compared with other vowel contexts. In other words, in the [_i] context, a perceptually less distinct consonant pair (C vs. Cʲ) shifted to a more distinct pair (Cˠ vs. Cʲ). In other words, a contrast can be preserved in a context where it is perceptually less salient by changing one party in the contrast into a form that is more distinct from the other.

(1) The contrast between plain and palatalized consonants in Russian (Adapted from Padgett 2001, 2003)[1]

a. before most vowels

vol	'ox'	**vʲ**ol	'he led'
suda	'court of law'	**sʲ**uda	'here, this way'

b. before the vowel [_i]

vˠitʲ	'to howl'	**vʲ**it	'to twist, weave'
sˠito	'sated'	**sʲ**ito	'sieve'

[1] In dataset (b), the items on the left were conventionally transcribed as vɨtʲ 'to howl' and sɨto 'sated'. Based on his phonetic study, Padgett (2001) argued that the sound [ɨ] is in fact an [i] preceded by a velarized consonant. Following the analysis in Padgett (2001), the items in (b) are transcribed with the velarized consonants to facilitate the comparison with dataset (a).

For sibilant contrasts in particular, a typological study has shown that there is a tendency to avoid place contrasts in the [_i] context (Lee-Kim 2014a). For example, a three-way sibilant place distinction exists among fricatives in Acoma (Miller 1965), Chacobo (Prost 1967), Cashinahua (Kensinger 1963), and Telugu (Lisker 1963). Yet in the [_i] context, the three-way place distinction is reduced to a two-way place distinction, as in (2) (Lee-Kim 2014a).

(2) Reduction of sibilant contrast in the /_i/ context (Lee-Kim 2014a)

Language (Family)	*Sibilants*	*Contrast before /i/*	*Phonological pattern*
Acoma (Kersan)	s-ɕ-ʂ	s-ɕ	ʂ→s or ɕ / _i
Chacobo (Panoan)	s-ʃ-ʂ	s-ʃ	ʂ→ ʃ / _i
Cashinahua (Panoan)	s-ʃ-ʂ	s-ʃ	*ʂi
Telugu (Dravidian)	s-ɕ-ʂ	ɕ-ʂ	s→ɕ/_i/e (optional)

Lee-Kim (2014a) attributed the avoidance of sibilant place contrast in the [_i] context primarily to two factors: the similar spectral properties of the sibilants due to palatalization and the weakening of the formant transition cue in the high front vowel.

Lee-Kim's (2014) generalization was drawn from the typology of languages with three-way place distinctions of sibilants and the languages in her typology typically have the place contrasts of fricatives. In her typology, Mandarin Chinese was referred to as an example of *full contrast plus full enhancement*, i.e., the underlying sibilant contrasts are fully preserved with a full change of the quality of /i/. More specifically, Mandarin Chinese has dental, palatal, and retroflex sibilants, and their surface contrasts in the [_i] context are avoided; instead, the contrasts are realized as /sɹ̩-ɕi-ʂɻ̩/, i.e., the palatal sibilants are followed by [_i] and the dental and retroflex sibilants are followed by homorganic syllabic approximants (aka. apical vowels).

Sibilant contrasts are widely observed across Chinese languages with two typological properties. First, there are more dialects with two-way sibilant place contrasts than dialects with three-way place contrasts (to be detailed in 2.2); second, sibilant affricates are as frequent as sibilant fricatives. For example, Lee and Zee (2015) surveyed the sound systems of 70 Chinese dialects that are geographically widely distributed and representative of 11 Chinese dialectal groups. They observed that the unaspirated and aspirated affricates [ts ts^{h} tɕ tɕ^{h} tʂ tʂ^{h}] are in general as frequent as their fricative counterparts [s ɕ ʂ]. With these two typological properties, the following two questions are raised to complement Lee-Kim's (2014a) generalization: First, are sibilant place contrasts in the [_i] context generally avoided in dialects with two sibilant places and those with three sibilant places alike? Second, do affricates behave the same as their fricative counterparts in their place contrasts in the [_i] context? A typological survey across Chinese dialects is to be conducted to answer these two questions.

1.2 Contrast distinctiveness and sound change

The preference for more distinct contrasts has been observed in diachronic sound change as well, by which a less distinct contrast sometimes develops into a more distinct one. For velar stops, for example, it is difficult to maintain a voicing contrast (/k/ vs. /g/) due to the small supralaryngeal cavity, and consequently, the difficulty to maintain the transglottal airflow needed for vocal fold vibration for /g/. In the avoidance of the weak contrast between /k/ and /g/, some languages changed the pronunciation of /g/ to make it more distinct from /k/, as shown in (3), adapted from the summary in Boersma (1998:384-386).

(3) Avoidance of contrastive /k/ vs. /g/ (Adapted from Boersma 1998:384-386)

Language	*Old contrast*	*New contrast*	*Sound change*
Arabic	k-g	k-dʒ	g→dʒ
Japanese	k-g	k-ŋ	g→ŋ (word-initially)
Low German	k-g	k-ɣ	g→ɣ
Czech, Slovak, Ukrainian	k-g	k-ɦ	g→ɦ

In terms of sibilant contrasts, the historical development of Spanish showed the tendency of contrast enhancement: The sibilant inventory of 16th century Spanish was /s̪ s ʃ/, and it developed into /θ s x/ in Castilian Spanish (Penny 2002; Zampaulo 2013). A similar sound change was observed in the historical development of Polish, by which a [ɕ-ʃʲ] contrast developed into a [ɕ-ʂ̻] contrast (Padgett and Źygis 2007; Źygis and Padgett 2010). Źygis and Padgett (2010) investigated the relative perceptual distinctiveness of [ɕ-ʃʲ] vs. [ɕ-ʂ̻] in an AX discrimination experiment. The stimuli were CV pairs like [ɕa-ʃʲa], [ɕa-ʂ̻a] and VC pairs likes [aɕ-aʃʲ], [aɕ-aʂ̻], and the listeners were 10 native American English listeners and 13 native Polish listeners. Their experimental results showed that the [ɕ-ʃʲ] contrast generally induced a longer response time than the [ɕ-ʂ̻] contrast, indicating that [ɕ-ʃʲ] is perceptually less distinct than [ɕ-ʂ̻]. Therefore, the historical development of Polish sibilants involved an enhancement of the sibilant place contrasts.

Mandarin Chinese has the dental, palatal, and retroflex sibilants, and their developments in various diachronic stages have been reconstructed in the studies of historical Mandarin phonology (Z. Lu 1947, R. Li 1952, Wang 1980, 1985, Ning 1985, among others). It is unclear if contrast enhancement has played a role in the historical sound changes of Mandarin sibilants. A survey on the development of Mandarin sibilants will be conducted and research questions will be raised based on the historical pattern.

1.3 Major themes in the current study: Sibilant place contrasts

For sibilant contrasts, the studies reviewed above have shown that:

(i) Sibilant place contrasts in the [_i] context are generally avoided;

(ii) Less distinct sibilant contrasts can develop into more distinct contrasts in sound changes.

For (i), a typological survey of sibilant place contrasts was conducted across 170 Chinese dialects. The typology showed that:

(a) Sibilant place contrasts in the [_i] context tend to be avoided in dialects with a three-way distinction of sibilant places (e.g., */si-ɕi-ʂi/) as well as in dialects with a two-way distinction of sibilant places (e.g., */si-ɕi/);

(b) In terms of place contrasts in the [_i] context, a place contrast of affricates (e.g., /tsi-tɕi/) implies that of fricatives (e.g., /si-ɕi/).

(c) Across the Chinese dialects, there are also cases where the dental vs. palatal sibilants are reported to fully contrast in the [_i] context, e.g., Xiangtan (Zeng 1993).

For (ii), the historical development of Mandarin sibilants was examined on the basis of previous reconstructions of the consonants and vowels in different historical stages (Wang 1980, 1985, Ning 1985, Li and Zhou 1999). The historical survey showed that:

(d) Sibilant place contrasts in the [_i] context (e.g., /si-ɕi/) that appeared as the result of other independent processes have been continuously avoided via separate sound changes, e.g., vowel change, sibilant change, and contrast neutralization;

(e) In case of sibilant change, dental-palatal contrasts (e.g., /s-ɕ/) developed into dental-retroflex contrasts (e.g., /s-ʂ/), similar to the sound change in Polish (Padgett and Źygis 2007; Źygis and Padgett 2010).

Based on the patterns in the synchronic typology and the diachronic changes, the following research questions were raised about sibilant contrasts:

(f) Do vowel contexts affect the perceptual distinctiveness of sibilant place contrasts, e.g., is [si-ɕi] less distinct than [sa-ɕa]?

(g) Do place contrasts differ in perceptual distinction, e.g., is [si-ɕi] less distinct than [si-ʂi]?

(h) Do different manners of articulation differ in their perceptual distinctiveness, e.g., is [tsi-tɕi] less distinct than [si-ɕi]?

These research questions were to be answered by conducting a speeded AX-discrimination experiment, which has been shown to be able to evaluate the relative perceptual distinctiveness of sound pairs independent of the listener's native phonology. As will be shown in Chapter 3, the results from the perceptual experiment generally showed that the vowel context [_i] reduces the perceptual distinctiveness between the dental and palatal sibilants, confirming the hypothesis that contrasts like [si-ɕi] are perceptually weak. This accounts for the tendency to avoid sibilant place contrast in the [_i] context in linguistic typology and historical sound changes.

Contrastive dental vs. palatal sibilants in the [_i] context have been reported in Chinese dialects like Xiangtan (see the generalization (c) above). A phonetic study was conducted on the contrastive sibilants in Xiangtan on their acoustic and perceptual properties in order to determine whether there is any sign of avoiding contrastive dental vs. palatal sibilants in the [_i] context.

As will be detailed in Chapter 4, acoustic and perceptual data both showed a sign of sibilant neutralization between the dental and palatal sibilants in the [_i] context, replicating the sound changes of Mandarin.

1.4 Significance of the present work

The primary purpose of this dissertation is to provide experimental data on the influence of vowel contexts on the perceptual distinctiveness of sibilant place contrasts and shed light on the relevance of perceptual distinctiveness to typological patterns and sound changes. While previous research has been independently fruitful in experimentation, language typology, and sound changes, there have been relatively few studies that integrate all three, presumably due to the difficulty in obtaining relevant data in all three areas. This dissertation thus provides a case study on sibilant place contrasts connecting all these areas and contributes to the literature by strengthening the empirical connections between experimental data, language typology, and the development of sound systems.

1.5 Structure of this dissertation

This dissertation is divided into six chapters. Chapter 1 introduces the background of the research. Chapter 2 reviews the phonetic and phonological properties of Mandarin sibilants, lays out the typology of sibilant place contrasts and the historical development of Mandarin sibilants, and raises research questions based on the typological survey and the historical development. Chapter 3 reports the design and the results of a Speeded-AX discrimination experiment, which aims to answer the research questions raised in Chapter 2. Chapter 4 reports a phonetic study of sibilants in Xiangtan Chinese, which was reported to fully contrast dental vs. palatal sibilants in the [_i] context, to examine whether there is any sign of sibilant place neutralization in the [_i] context. Chapter 5 is a general discussion of the results of the perceptual study and the phonetic study and their implications to language typology and sound change.

Chapter 2. Sibilant contrasts: Typology and sound change

In this chapter, an introduction is first given to the phonetic and phonological properties of Mandarin sibilants (2.1), and then a typological survey of sibilant contrasts is reported (2.2), followed by a survey of the historical development of Mandarin sibilants (2.3). Research questions are then raised (2.4), based on the patterns observed in 2.2 and 2.3.

2.1 Mandarin sibilants and their place contrasts

Mandarin (Standard Chinese, Putonghua) has dental sibilants [s ts tsʰ], palatal sibilants [ɕ tɕ tɕʰ], and retroflex sibilants [ʂ tʂ tʂʰ]. The dental sibilants are produced with the tongue tip against the back of the upper incisors, the palatal sibilants with the tongue blade against the hard palate, and the retroflex sibilants with the upper surface of the tongue tip approaching the center of the alveolar ridge (Ladefoged and Wu 1984; Cao 1990).[2] In terms of center of gravity (COG, spectral mean), both [s] and [ɕ] have a high frequency peak between 5000Hz and 9000Hz, while the retroflex [ʂ] has a lower COG; in terms of dispersion, [s] has a narrower distribution of energy than [ɕ] and [ʂ] (Svantesson 1986; Wu and Lin 1989:133). The fricative components of the affricates are in general similar to the corresponding fricatives (Svantesson 1986).

In Mandarin, the surface place contrasts of dental, palatal and retroflex sibilants in the [_i] context are avoided with vowel enhancement after dental and palatal sibilants. More specifically, the vowel [i] follows the palatal sibilants but not the dental and retroflex sibilants. As shown in

[2] Mandarin sibilants [ʂ tʂ tʂʰ] are conventionally referred to as retroflexes in the phonological literature. However, they do not involve the curling up of the tongue tip as Indian retroflex sounds (Ladefoged and Wu 1984). The IPA symbols [ʂ tʂ tʂʰ] are used in this dissertation following the convention in the literaure.

(4), the dental and retroflex sibilants are followed by the homorganic syllabic approximants [ɹ̩] and [ɻ̩] (Lee and Zee 2003, Lee-Kim 2014b), which are often referred to as the 'apical vowels'.[3] The articulatory study in Lee-Kim (2014b) showed that, in producing Mandarin [sɹ̩], [ɕi], and [ʂɻ̩], the tongue positions do not change significantly after the tongue reaches the consonantal targets, which confirms that the consonantal and the vocalic gestures are homorganic. In other words, apical vowels are the 'vocalized prolongation' of their preceding consonants (Chao 1934:374). The two apical vowels [ɹ̩] [ɻ̩] and the vowel [i] differ in their first three formants, particularly F2 (Zee and Lee 2001, Lee and Li 2003, Cheung 2004), and there is no obvious consonant-vowel formant transition in the syllables [sɹ̩], [ɕi], and [ʂɻ̩] due to the homorganic relation between the sibilants and the vocalic segments.

(4) The distribution of the vowel [i] and the apical vowels [ɹ̩] and [ɻ̩] in Mandarin

a. ɹ̩ after dentals			b. i after palatals			c. ɻ̩ after retroflexes		
sɹ̩	思	'think'	ɕi	西	'west'	ʂɻ̩	獅	'lion'
tsɹ̩	資	'resource'	tɕi	雞	'chicken'	tʂɻ̩	知	'know'
tsʰɹ̩	差	'uneven'	tɕʰi	七	'seven'	tʂʰɻ̩	吃	'eat'

The avoidance of sibilant place contrast in Mandarin, as in (4), has been argued to be perceptually based, by which less distinct contrasts like /si-ɕi-ʂi/ turn into more distinct contrasts like [sɹ̩-ɕi-ʂɻ̩] (Stevens, Li, Lee, and Keyser 2004; Lee and Li 2003, M. Li 2011, Lee-Kim 2014a, among others). More specifically, it is assumed that the frication noise of the dental /s/, the palatal /ɕ/, and the retroflex /ʂ/ are acoustically close to each other (Stevens et al. 2004) and that,

[3] In the Chinese phonological literature, the two apical vowels are often represented with two non-IPA symbols: ɿ for the post-dental [ɹ̩] and ʅ for the post-retroflex [ɻ̩]. The nature of these two sounds has been described in various ways, e.g., syllabic fricatives (Duanmu 2007:34), fricative vowels (Ladefoged and Maddieson 1996:314), and syllabic approximants (Lee and Zee 2003, Lee-Kim 2014b).

for a contrast triplet like /si-ɕi-ʂi/, the place distinction among the sibilants is likely to be compromised due to the palatalization from the following [i] (Lee and Li 2003, Lee-Kim 2014a). With the contrast at risk, an enhancement gesture is introduced to make /s/ and /ʂ/ more distinct from /ɕ/. In other words, the apical vowels ɹ̩ and ɻ̩ are the continuation of the enhancing gestures on /s/ and /ʂ/ throughout a following /i/ (Stevens et al. 2004; Keyser and Stevens 2006), by which a shift of vowel helps preserve the contrast /si-ɕi-ʂi/ (Lee and Li 2003, Lee-Kim 2014a). As the frication components of affricates are similar to their fricative counterparts (Svantesson 1986), the enhancement analysis for Mandarin fricatives could be easily extended to the affricates.

The way Mandarin avoids sibilant place contrasts in the [_i] context, as in (4), was referred to as *full contrast plus full enhancement* in Lee-Kim's typology (2014a), i.e., the underlying sibilant contrasts are preserved with a complete change of the quality of /i/ after dental and retroflex sibilants, as opposed to the sibilant place neutralization in Acoma, Chacobo, Cashinahua, and Telugu, as in (2). In Lee-Kim's (2014a) typology, the way Mandarin avoids the sibilant place contrasts was assumed to be typologically rare due to the large deviation from the underlying forms, e.g., si → sɹ̩.

Sibilant contrasts are widely observed across Chinese dialects, and there are dialects with two sibilant places and dialects with three sibilant places (to be detailed in 2.2). Also, a Chinese dialect typically has both fricatives and affricates (M. Li 2011; Lee and Zee 2015). Lee-Kim's (2014a) generalization on the avoidance of sibilant place contrast in the [_i] context was based on the typology of languages with sibilants at three places of articulation. In this dissertation, her generalization is to be examined against the typology of Chinese dialects with both three sibilant places and two sibilant places, in particular to examine whether sibilant place contrast in the [_i] context also tends to be avoided in the latter type of dialects. Moreover, given that almost all Chinese dialects have a three-way distinction among fricative, unaspirated affricate, and aspirated affricate, Lee-Kim's typology will also be tested for these three manners of articulation separately to examine whether sibilant place contrast in the [_i] context tends to be avoided across all three manners of articulation.

2.2 Sibilant contrasts across Chinese dialects

To investigate the property of sibilant place contrasts in the [_i] context, a typological survey was conducted across 170 Chinese dialects.[4] *Fangyan* [*Dialects*], a Chinese journal specializing in the description of Chinese dialects, was adopted as the basis of the survey. From all the articles published in *Fangyan* from 1979 to 2012, a total of 155 articles were identified, each describing the syllabic inventory of a Chinese dialect. Geographically, these 170 dialects are widely distributed across Mainland China and belong to different dialectal groups, e.g., Xiang dialects, Gan dialects, Yue dialects, etc.[5] This study focused on the vowel [_i] and the sibilants: The dental [s ts tsʰ], the palatal [ɕ tɕ tɕʰ], and the retroflex [ʂ tʂ tʂʰ].[6] An overview of the 170 dialects is given in Table 1, where the dialects are grouped by the number of sibilant places. More details and a complete list of the 170 dialects can be found in Appendix I and II.

---

4 This is an extension of the survey of 124 Chinese dialects in M. Li (2011).

5 It would be desirable to identify the dialectal group of each dialect in the survey. However, this is difficult because many of the articles did not provide information about the affiliation of the dialect described.

6 A few dialects are reported to have voiced sibilants, e.g., the voiced palatal [dʑ] in *Shibei Chinese* (A. Hiroyuki 2004) and the voiced retroflex [dʐ] in *Xiangtan Chinese* (Y. Zeng 1993). Such onsets generally have the same phonotactic pattern as their voiceless counterparts.

Table 1 Typology of sibilant inventory across 170 Chinese dialects

No. of sibilant place	*No. of dialects*	*Example*
1 Place	35 total	
Dental	28	*Xiamen* (Zhou 1991)
Palatal	7	*Lianzhou* (Cai 1987)
2 Places	84 total	
Dental vs. Palatal	81	*Jiangyong* (Huang 1988)
Dental vs. Retroflex	2	*Jinggangshan Hakka* (Lu 1995)
Palatal vs. Retroflex	1	*Haizhou* (Su 1990)
3 Places	51 total	
Dental vs. Palatal vs. Retroflex	51	*Harbin* (Yin 1995)

Table 1 shows that, for the 35 dialects with sibilants at one place, the majority (28/35) have dentals; for the 84 dialects with two sibilant places, the majority (81/84) have dentals vs. palatals; for the 51 dialects with three sibilant places, all the dialects have dentals vs. palatals vs. retroflexes. As this study focuses on the sibilant place contrasts, the 84 dialects with two sibilant places and the 51 dialects with three sibilant places will be relevant to the typological survey. These two types of dialects are reported below.

2.2.1 Two-way sibilant place contrasts

In the 84 dialects with a two-way distinction of sibilant places, there are 81 dialects that have dental vs. palatal sibilants. In these 81 dialects, there is a clear tendency to avoid the place contrasts in the [_i] context: The contrasts are present in 28% of the dialects (23/81), as exemplified in (5), and absent in 72% (58/81) of the dialects (either with vowel enhancement or with non-combinaiton of certain sibilants with [i]), as exemplified in (6).

(5) Dialects with two sibilant places (I): Fully contrastive dentals vs. palatals – 23 dialects

a. *Jiangyong* (Huang 1988), representing 3 dialects

si	細	'slim'	ɕi	戲	'opera'
tsi	祭	'offer sacrifice'	tɕi	寄	'to mail'
tsʰi	砌	'lay bricks'	tɕʰi	氣	'gas'

All syllables bear a low-falling tone

b. *Changle* (Z. Qian 2003), representing 20 dialects[7]

sɹ̩	獅	'lion'	si	西	'west'	ɕi	稀	'scarce'
tsɹ̩	枝	'branch'	tsi	□	'annoying'	tɕi	飢	'hungry'
tsʰɹ̩	翅	'wing'	tsʰi	妻	'wife'	tɕʰi	欺	'to cheat'

All syllables bear a concave tone

(6) Dialects with two sibilant places (II): Non-contrastive dentals vs. palatals – 58 dialects

a. *Dayü* (L. Liu 1995), representing 56 dialects

sɹ̩	勢	'tendency'	ɕi	西	'west'
tsɹ̩	資	'resources'	tɕi	雞	'chicken'
tsʰɹ̩	滯	'stop'	tɕʰi	欺	'to cheat'

All syllables bear a mid-level tone

b. *Shibei* (Hiroyuki 2004), representing 1 dialect

*si			ɕi	絲	'string'
*tsi			tɕi	疾	'ache'
*tsʰi			tɕʰi	妻	'wife'

Each legitimate syllable bears a high-falling tone.

c. *Jixi* (R. Zhao 1989), representing 1 dialect

sɹ̩	嬉	'amusement'	si	收	'to receive'	*ɕi
tsɹ̩	雞	'fowl'	tsi	周	'circuit'	*tɕi
tsʰɹ̩	癡	'stupid'	tsʰi	抽	'to draw out'	*tɕʰi

Each legitimate syllable bears a high-level tone

Some details need to be noted about the typology of dialects with two sibilant places. As shown in (5), the 23 dialects that preserve the place contrasts of dental vs. palatal sibilants fall into two types: 3 dialects like *Jiangyong* (Huang 1988) in (5a) have contrasts like /si-ɕi/; 20 dialects like *Changde* (Z. Qian 2003) in (5b) have contrasts like /sɹ̩-si-ɕi/.

---

[7] In this data set, the symbol '□' means that the syllable exists in the colloquial version of the language, but there is no corresponding Chinese character.

Of the 58 dialects that disallow the place contrasts of dental vs. palatal sibilants in the [_i] context, 56 dialects like *Dayü* (L. Liu 1995), avoid the contrasts by introducing ‘apical vowels’ after the dental sibilants, e.g., /sɹ̩-ɕi/ in (6a). The other two dialects avoid the contrasts with phonotactic gaps: The rime [_i] does not follow the dentals in *Shibei* (Hiroyuki 2004), as in (6b), and it does not follow the palatals in *Jixi* (R. Zhao 1989), as in (6c).

In the 84 Chinese dialects with a two-way distinction of sibilant places, three dialects differ from the 81 exemplified in (5) and (6) in terms of the places of the sibilants. *Wuhua* (Y. Wei 1997) and *Jinggangshan* (S. Lu 1995) have dental vs. retroflex sibilants in their phonetic inventory, as in (7). Contrastive dental vs. retroflex sibilants are allowed in the [_i] context in *Wuhua* (Y. Wei 1997) but avoided in *Jinggangshan* (S. Lu 1995) with the introduction of apical vowels after the retroflex sibilants.

(7) Dialects with two sibilant places (III): Dentals vs. retroflexes – 2 dialects

a. *Wuhua* (Y. Wei 1997), contrastive dentals vs. retroflexes

sɹ̩	獅	‘lion’	si	西	‘west’	ʂi	詩	‘poem’
tsɹ̩	姿	‘looking’	tsi	擠	‘to squeeze’	tʂi	枝	‘branch’
ts^{h}ɹ̩	粗	‘coarse’	ts^{h}i	妻	‘wife’	tʂ^{h}i	癡	‘stupid’

All syllables bear a high-level tone

b. *Jinggangshan* (S. Lu 1995), non-contrastive dentals vs. retroflexes

sɹ̩	絲	‘thread’	si	西	‘west’	ʃɻ̩	詩	‘poem’
tsɹ̩	資	‘resources’	tsi	擠	‘to squeeze’	tʃɻ̩	知	‘to know’
ts^{h}ɹ̩	雌	‘female’	ts^{h}i	妻	‘wife’	tʃ^{h}ɻ̩	癡	‘stupid’

All syllables bear a high-level tone.

Haizhou (X. Su 1990) has palatal vs. retroflex sibilants in its phonetic inventory, as in (8), and the contrasts between palatal and retroflex sibilants are also enhanced in the [_i] context with the introduction of apical vowels after the retroflex sibilants.

(8) Dialects with two sibilant places (IV): Palatal vs. retroflexes – 1 dialect

Haizhou (X. Su 1990), non-contrastive palatals vs. retroflexes

ɕi	西	‘west’	ʂɻ̩	詩	‘poem’
tɕi	飢	‘hungry’	tʂɻ̩	枝	‘branch’
tɕʰi	妻	‘wife’	tʂʰɻ̩	疵	‘blemish’

All syllables bear a conves tone

In general, the typology of Chinese dialects with two sibilant places shows that, in the majority of the dialects, the place contrasts of dental vs. palatal sibilants in the [_i] context tend to be enhanced or avoided even when there are no retroflex sibilants in the sound system.

2.2.2 Three-way sibilant place contrasts

In the 51 dialects with three sibilant places, the full place contrasts in the [_i] context are present in two dialects, e.g., Qimen (T. Shen 1989) in (9), and enhanced or avoided in the other 49 dialects as in (10) to (12). In other words, the full place contrasts are preserved in 4% (2/51) of the dialects and enhanced or avoided in 96% (48/51) of the dialects.

(9) Dialects with three sibilant places (I): Fully contrastive dentals vs. palatals vs. retroflexes – 2 dialects

e.g., *Qimen* (T. Shen 1989), representing 2 dialects

sɹ̩	死	‘die’	si	膝	‘knee’	ɕi	是	‘is’	ʂi	飾	‘decorate’
tsɹ̩	紫	‘purple’	*tsi			tɕi	己	‘self’	tʂi	纸	‘paper’
tsʰɹ̩	此	‘this’	tsʰi	七	‘seven’	tɕʰi	起	‘get up’	tʂʰi	齿	‘tooth’

Some details need to be noted about the typology of the 49 dialects that avoid the full place contrasts in the [_i] contexts. Generally, these 49 dialects fall into two types. First, 17 dialects preserve the contrasts between two sibilant places in the [_i] context, as in (10) and (11).

(10) Dialects with three sibilant places (II): Dental vs. palatal contrast, with retroflex-[i] avoided – 16 dialects

a. *Xiangtan* (Zeng 1993), representing 13 similar dialects

sɹ̩	四	‘four’	si	細	‘slim’	ɕi	戲	‘opera’	ʂɻ̩	屍	‘corpse’
tsɹ̩	姿	‘pose’	tsi	祭	‘sacrifice’	tɕi	技	‘skill’	tʂɻ̩	枝	‘branch’
ts^{h}ɹ̩	賜	‘to bestow’	ts^{h}i	砌	‘lay bricks’	tɕ^{h}i	氣	‘gas’	tʂ^{h}ɻ̩	痴	‘stupid’

All syllables bear a high-level tone

b. *Changshu* (D. Yuan 2010), representing 3 similar dialects

sɹ̩	始	‘start’	si	死	‘to die’	ɕi	喜	‘happy’	*ʂi
tsɹ̩	紙	‘paper’	tsi	姊	‘sister’	tɕi	舉	‘to lift’	*tʂi
ts^{h}ɹ̩	齒	‘tooth’	ts^{h}i	取	‘to take’	tɕ^{h}i	啟	‘to open’	*tʂ^{h}i

All syllables bear a high-level tone

(11) Dialects with three sibilant places (III): Palatal vs. retroflex contrast, with dental-[i] avoided – 1 dialect

Xuancheng (Shen and Huang 2015)

sɹ̩	詩	‘poem’	ɕi	鮮	‘fresh’	ʃi	西	‘west’
tsɹ̩	枝	‘branch’	tɕi	煎	‘to fry’	tʃi	飢	‘hungry’
ts^{h}ɹ̩	癡	‘stupid’	tɕ^{h}i	千	‘thousand’	tʃ^{h}i	妻	‘wife’

All syllables bear a mid-level tone

In this group, 13 dialects like Nanjing (D. Liu 1994) in (10a) preserve the contrasts of dental vs. palatal sibilants in the [_i] context and introduce apical vowel ɻ̩ after the retroflex sibilants. In the three other dialects like Changshu (D. Yuan 2010) in (10b), the retroflex sibilants are followed by neither [_i] nor apical ɻ̩. In addition, Xuancheng (Shen and Huang 2015) in (11) preserves the place contrasts of palatal vs. retroflex sibilants in the [_i] context rather than that of the dental vs. palatal sibilants.

Second, 32 dialects do not allow any place contrast of sibilants in the [_i] context, as in (12). In this group, there are 28 dialects like Anqing (H. Bao 2012) in (12a), which avoid the two-place sibilant contrasts in the [_i] context with the introduction of the apical ɹ̩ after the dental sibilants and that of the apical ɻ̩ after the retroflex sibilants, i.e., the same way Mandarin avoids the contrasts. On the other hand, in the other four dialects, the rime [_i] does not follow the retroflex sibilants, e.g., Dangtu (Zhang, Yuan, and Shen 2012) in (12b), or the rime [_i] does not follow the palatal sibilants, e.g., Ledu (Cao and Shao 2001) in (12c).

(12) Dialects with three sibilant places (IV): No sibilant place contrasts in the [_i] context – 32 dialects

a. *Anqing* (H. Bao 2012), representing 28 similar dialects

sɹ̩	獅	'lion'	ɕi	西	'west'	ʂɻ̩	詩	'poem'
tsɹ̩	姿	'looking'	tɕi	飢	'hungry'	tʂɻ̩	枝	'branch'
ts^hɹ̩	疵	'blemish'	$tɕ^h$i	妻	'wife'	$tʂ^h$ɻ̩	癡	'stupid'

All syllables bear a mid-falling tone

b. *Dangtu* (Zhang, Yuan, and Shen 2012), representing 2 similar dialects[8]

sɹ̩	獅	'lion'	ɕi	西	'west'	*ʃi
tsɹ̩	枝	'branch'	tɕi	飢	'hungry'	*tʃi
ts^hɹ̩	疵	'blemish'	$tɕ^h$i	妻	'wife'	*$tʃ^h$i

All syllables bear a high-level tone

c. *Ledu* (Cao and Shao 2001), representing 2 similar dialects

sɹ̩	西	'west'	*ɕi	ʂɻ̩	詩	'poem'
tsɹ̩	飢	'hungry'	*tɕi	tʂɻ̩	枝	'branch'
ts^hɹ̩	疵	'blemish'	*$tɕ^h$i	$tʂ^h$ɻ̩	癡	'stupid'

All syllables bear a low-rising tone

In general, the typology above shows that a full contrast among dental, palatal, and retroflex sibilants in the [_i] context is generally enhanced with the introduction of apical vowels after the dental and/or retroflex sibilants. This is consistent with the Lee-Kim's (2014) generalization that three-way sibilant place contrasts in the [_i] context tend to be avoided.

2.2.3 Pair-wise sibilant place contrasts

The typology of two sibilant places (2.2.1) shows that dialects tend to avoid sibilant place contrasts in the [_i] context when there are only a two-way distinction of sibilant places in the inventory. In other words, the non-occurrence of sibilant place contrast in the [_i] context is not

---

[8] According to Zhang et al. (2012), the consonants represented by the symbols [ʃ tʃ $tʃ^h$] in Dangtu are produced with tongue blade approaching the post-alveolar region.

necessarily conditioned by the existence of a three-way distinction of sibilant places in a language. Below, generalizations are made on the contrasts of dental vs. palatal sibilants, dental vs. retroflex sibilants, and palatal vs. retroflex sibilants, respectively.

2.2.3.1 Dental vs. palatal sibilants

Dental vs. palatal sibilants exist in the dialects with two sibilant places as well as those with three sibilant places. Across these two groups, a total of 132 dialects have dental vs. palatal sibilants in their phonetic inventory and an overview of the contrastive pattern is given in Table 2. As Table 2 illustrates, 41 dialects preserve the contrasts of dental vs. palatal sibilants in the [_i] context, including 23 dialects (= 3 + 23) with two sibilant places and 18 dialects (= 2 + 13 + 3) with three sibilant places. On the other hand, 91 dialects avoid the contrasts, including 58 dialects (= 56 + 1 + 1) with two sibilant places and 33 dialects (= 1 + 28 + 2 + 2) with three place sibilants. Overall, in the 132 dialects, contrastive dental vs. palatal sibilants in the [_i] contexts are preserved in 31% of the dialects (= 41/132) and enhanced or avoided in 69% of the dialects (= 91/132). This pattern indicates that, cross-linguistically, the majority of the dialects avoid the place contrast of dental vs. palatal sibilants in the [_i] context.

Table 2 Typology of dental vs. palatal sibilants – 132 dialects

Contrast pattern	*No. of dialects*	*Example dialect*	*Cross reference*
I. Contrast preserved	41 total		
a. **si** **ɕi**	3	*Jiangyong* (Huang 1988)	(5a)
b. sɹ̩ **si** **ɕi**	20	*Changle* (Z. Qian 2003)	(5b)
c. sɹ̩ **si** **ɕi** ʂi	2	*Qimen* (T. Shen 1989)	(9)
d. sɹ̩ **si** **ɕi** ʂɻ̩	13	*Nanjing* (D. Liu 1994)	(10a)
e. sɹ̩ **si** **ɕi** *ʂi	3	*Changshu* (D. Yuan 2010)	(10b)
II. Contrast enhanced/avoided	91 total		
d. **sɹ̩** **ɕi**	56	*Dayü* (L. Liu 1995)	(6a)
e. ***si** **ɕi**	1	*Shibei* (Hiroyuki 2004)	(6b)
f. sɹ̩ **si** ***ɕi**	1	*Jixi* (R. Zhao 1989)	(6c)
g. **sɹ̩** **ɕi** ʃi	1	*Xuancheng* (Shen & Huang 2015)	(11)
h. **sɹ̩** **ɕi** ʂɻ̩	28	*Anqing* (H. Bao 2012)	(12a)
i. **sɹ̩** **ɕi** *ʃi	2	*Dangtu* (Zhang et al. 2012)	(12b)
j. **sɹ̩** ***ɕi** ʂɻ̩	2	*Ledu* (Cao and Shao 2001)	(12c)

2.2.3.2 Dental vs. retroflex sibilants

Dental vs. retroflex sibilants also exist in the dialects with two sibilant places and those with three sibilant places alike. Across these two groups, a total of 52 dialects have dental vs. retroflex sibilants in their phonetic inventory, as summarized in Table 3.

Table 3 Typology of dental vs. retroflex sibilants – 52 dialects

Contrast pattern	*No. of dialects*	*Example dialect*	*Cross reference*
I. Contrast preserved	3 total		
a. sɹ̩ **si** **ʂi**	1	*Wuhua* (Y. Wei 1997)	(7a)
c. sɹ̩ **si** ɕi **ʂi**	2	*Qimen* (T. Shen 1989)	(9)
II. Contrast enhanced/avoided	49 total		
c. sɹ̩ **si** **ʃɻ̩**	1	*Jinggangshan* (S. Lu 1995)	(7b)
d. sɹ̩ **si** ɕi **ʂɻ̩**	13	*Nanjing* (D. Liu 1994)	(10a)
e. sɹ̩ **si** ɕi ***ʂi**	3	*Changshu* (D. Yuan 2010)	(10b)
f. **sɹ̩** ɕi **ʂɻ̩**	28	*Anqing* (H. Bao 2012)	(12a)
g. **sɹ̩** ɕi ***ʃi**	2	*Dangtu* (Zhang et al. 2012)	(12b)
h. **sɹ̩** *ɕi **ʂɻ̩**	2	*Ledu* (Cao and Shao 2001)	(12c)

As Table 3 illustrates, the place contrasts of dental vs. retroflex sibilants in the [_i] contexts are allowed in 3 dialects and enhanced or avoided in 49 dialects. This pattern could be taken as an indication that contrastive dental vs. retroflex sibilants in the [_i] context also tend to be avoided cross-linguistically. However, the avoidance of contrastive dental vs. retroflex sibilants in Table 3 is confounded with the number of sibilant places in a dialect. Except for Wuhua (Y. Wei 1997), all the dialects in Table 3 have sibilants at 3 places. Thus, it is unclear whether the avoidance of contrastive dentals vs. retroflexes is rooted in the avoidance of the full contrasts of dental, palatal, and retroflexes (c.f., 2.2.2).[9] Note that this is different from the avoidance of dental vs. palatal sibilants (c.f., 2.2.1), by which the typology of the 81 dialects with dental vs. palatals only (i.e., to the exclusion of retroflexes) argued independently for the tendency to avoid contrastive dentals vs. palatals in the [_i] context.

2.2.3.3 Palatal vs. retroflex sibilants

Across the dialects with two sibilant places and three sibilant places, a total of 52 dialects have palatal vs. retroflex sibilants in their phonetic inventory. A summary is given in Table 4. As Table 4 illustrates, the place contrasts of palatal vs. retroflex sibilants in the [_i] context are allowed in 3 dialects and enhanced/avoided in the other 49. Similar to the avoidance of contrastive dental vs. retroflex sibilants in Table 3, this observation is also confounded by the number of sibilants in a dialect, and it is unclear if the avoidance of contrastive dentals vs. retroflexes is rooted in the avoidance of the full contrasts of dentals, palatals, and retroflexes (c.f., 2.2.2).[10]

[9] The avoidance of contrastive dental-[i] vs. retroflex-[i] may have another potential cause, i.e., the articulatory incompatibility of a retroflex sibilant and a following [_i] (Pulleyblank 1984:25-26; Hamann 2003:95-96).

[10] Similar to the case of contrastive dental vs. retroflex sibilants, the articulatory incompatibility of a retroflex sibilant and a following [_i] is another potential cause of the typology.

Table 4 Typology of palatal vs. retroflex sibilants – 52 dialects

Contrast pattern	*No. of dialects*	*Example dialect*	*Cross reference*
I. Contrast preserved	3 total		
a. sɹ̩ si **ɕi** **ʂi**	2	*Qimen* (T. Shen 1989)	(9)
b. sɹ̩ **ɕi** **ʃi**	1	*Xuancheng* (Shen & Huang 2015)	(11)
II. Contrast enhanced/avoided	49 total		
c. **ɕi** **ʂɻ̩**	1	*Haizhou* (X. Su 1990)	(8)
d. sɹ̩ si **ɕi** **ʂɻ̩**	13	*Nanjing* (D. Liu 1994)	(10a)
e. sɹ̩ si **ɕi** ***ʂi**	3	*Changshu* (D. Yuan 2010)	(10b)
f. sɹ̩ **ɕi** **ʂɻ̩**	28	*Anqing* (H. Bao 2012)	(12a)
g. sɹ̩ **ɕi** ***ʃi**	2	*Dangtu* (Zhang et al. 2012)	(12b)
h. sɹ̩ ***ɕi** **ʂɻ̩**	2	*Ledu* (Cao and Shao 2001)	(12c)

2.2.4 Generalizations from the typology of sibilant place contrasts

Lee-Kim's (2014a) generalization that sibilant place contrasts in the [_i] context tend to be avoided was drawn from a pool of languages with three-way sibilant place distinction. This generalization is tested above against Chinese dialects with three sibilant places as well as those with two sibilant places. The typological survey in this dissertation shows that:

(a) For dialects with three sibilant places, where the place distinction is primarily among dental, palatal, and retroflex sibilants, there is a clear tendency to avoid the full three-way place contrasts in the [_i] context, primarily with the introduction of apical vowels.

(b) For dialects with two sibilant places, where the place distinction is primarily dental vs. palatal sibilants, there is also a tendency to avoid contrastive dental vs. palatal sibilants in the [_i] context, primarily with the introduction of apical vowels.

While the observation (a) confirms Lee-Kim's generalization, the observation (b) suggests that the avoidance of sibilant place contrasts in the [_i] context is not limited to languages with three sibilant places. In other words, the typology in this dissertation shows that contrastive dental vs. palatal sibilants tend to be avoided in a language even when there is no retroflex sibilants in the inventory. In terms of pair-wise place contrasts, the typology only shows a clear tendency to avoid dental vs. palatal sibilants in the [_i] context (2.2.1). For dental vs. retroflex sibilants and palatal vs. retroflex sibilants, however, the pattern is unclear due to the small number of dialects that only have these two types of place contrasts.

2.2.5 Fricative contrasts vs. affricate contrasts

Lee-Kim's (2014a) typological generalization is tested against the typology of fricatives and affricates alike. For dialects with two sibilant places and three sibilant places alike, it is observed that affricate contrasts generally behave the same as their fricative counterparts, as can be seen from the examples in 2.2.1 to 2.2.3.

However, for dental vs. palatal sibilants, a difference is observed between fricative contrasts and affricate contrasts. As noted in 2.2.3 and 2.2.4, among the different place contrasts, the tendency to avoid place contrasts in the [_i] context is observed only for dental vs. palatal sibilants. Dental vs. palatal sibilants exist in a total of 132 dialects, as shown in Table 2, including 41 dialects with two sibilant places and 91 dialects with three sibilant places. Out of these 132 dialects, contrastive dental vs. palatal sibilants are observed in 41 dialects, e.g., [si-ɕi], [tsi-tɕi], and [ts^{h}i-tɕ^{h}i].

A closer examination of the 41 dialects shows that they fall into 3 groups, as shown in Table 5. First, four dialects like Daoxian (K. He 2003) preserve the place contrast of dental vs. palatal fricatives but not that of the affricates; second, 34 dialects like Jiangyong (X. Huang 1988) preserve the place contrasts of all sibilants; third, three dialects like Datan (Y. Chen 2015) fall between the former two groups in that they preserve the place contrast of either aspirated or unaspirated affricates in addition to the place contrast of fricatives. An overview of these three types of dialects revealed a manner difference: in the [_i] context, the place contrast of affricates implies the place contrast of fricatives. It needs to be noted that this typology is based on a relatively small number of dialects, yet the pattern is clear.

Table 5 Typology of contrastive dental vs. palatal sibilants – 41 dialects

Contrast patterns				*Example dialects*	*No. of dialects*
a.	si-ɕi	* tsi-tɕi	* tsʰi-tɕʰi	*Daoxian* (道縣) (K. He 2003)	4
b.	si-ɕi	* tsi-tɕi	tsʰi-tɕʰi	*Datan* (大坦) (Y. Chen 2015)	2
	si-ɕi	tsi-tɕi	* tsʰi-tɕʰi	*Taishun* (泰順) (Z. Liu 2007)	1
c.	si-ɕi	tsi-tɕi	tsʰi-tɕʰi	*Jiangyong* (江永) (X. Huang 1988)	34

2.2.6 Summary of the typology

To sum up, the typological survey of sibilant place contrasts across Chinese dialects in this dissertation shows that, in the [_i] context:

(a) Three-way sibilant place contrasts tend to be avoided, e.g., */si-ɕi-ʂi/, primarily with the introduction of apical vowels;

(b) Two-way sibilant place contrasts of dental vs. palatal sibilants tend to be avoided, e.g., */si-ɕi/, primarily with the introduction of apical vowels, which holds true independently of whether a language has retroflex sibilants or not;

(c) For contrastive dental vs. palatal sibilants, the place contrast of affricates (i.e., /tsi-tɕi/, /tsʰi-tɕʰi/) implies that of the fricative (i.e., /si-ɕi/).

2.3 Historical development in Mandarin sibilants

The typology in section 2.2 shows that three-way sibilant place contrasts in the [_i] context tend to be avoided, e.g., */si-ɕi-ʂi/. As mentioned in 2.1, Mandarin avoids the three-way sibilant place contrasts in the [_i] context with the introduction of apical vowels, e.g. /sɹ̩-ɕi-ʂɻ̩/. A natural question, then, is how Mandarin contrasts like /sɹ̩-ɕi-ʂɻ̩/ developed into its current form? For the enhancement analysis of Mandarin apical vowels (Stevens et al. 2004, Lee and Li 2003, Lee-Kim 2014a), the question is: Has the formation of apical vowels in Mandarin avoided the sibilant place contrasts in the [_i] context? This section examines the diachronic development of /sɹ̩ ɕi ʂɻ̩/, /tsɹ̩ tɕi tʂɻ̩/, and /tsʰɹ̩ tɕʰi tʂʰɻ̩/ in Mandarin historical phonology.

2.3.1 Method of data collection

Chinese syllables are conventionally grouped by their onsets and rimes in Chinese phonological literature. For example, the syllables [sa] 'to let go' and [san] 'three' have the same onset and the syllables [san] 'three' and [tan] 'greedy' have the same rime. The same is true with ancient Chinese rhyme books like *Qieyun* (published in 601AD and representing the sound systems of Chinese around the 7th century) and *Zhongyuan Yinyun* (published in 1324AD and representing the sound systems of Chinese around 14th century). These documents provide information about the grouping of onsets and rimes rather than the actual pronunciation of the onsets and rimes, because the Chinese writing system does not directly reflect the pronunciation of a written character. Since Karlgren (1917), there have been extensive and fruitful studies that reconstruct the phonetic forms of the onsets and rimes in rhyme books like *Qieyun* and *Zhongyuan Yinyun* (Z. Lu 1947, R. Li 1952, Tung 1954, Wang 1980, 1985, X. Li 1983, Ning 1985, among others).

In terms of time span, the current survey is limited to the historical developments between the 7th century (i.e., the sound system represented by the rhyme book *Qieyun*[11]) and Mandarin Chinese, for which period there has been general agreement on the reconstructed forms of the onsets and rimes. The data were drawn primarily from two resources:[12]

(a) The book *Hanzi Gujin Yinbiao* [*A List of the Historical Pronunciation of Chinese Characters*] (Li and Zhou 1999), which provides the reconstructed forms of 9,000 Chinese characters, with their pronunciations in the 7th century, i.e., the sound system represented by *Qieyun* (601AD), and those in the 14th century, i.e., the sound system represented by *Zhongyuan Yinyun* (1324AD). The reconstructed forms in this book follow primarily the reconstruction in L. Wang (1980) and Ning (1985).

---

[11] Despite the general agreement on the reconstructed consonants and rimes, there has been disagreement on the nature of the sound system represented by *Qieyun*. It is generally believed that the sound system of *Qieyun* represented a lingua franca around the 7th century while some phonologists held that it represented the sound system of a specific dialect. A concise summary of the different views was given in W. Pan (2000:1-13).

[12] The historical survey in this study is an extension of the survey in M. Li (2011), which referred to Li and Zhou (1999) but not L. Wang (1985). The survey in the current study provides a more detailed picture of the development in terms of the phonetic forms across different historical stages.

(b) The book *Hanyu Yuyin Shi* [*The Historical Phonology of Chinese*] by L. Wang (1985). This book provides the reconstructed onsets and rimes in the sound systems of Chinese at various intervals, e.g., 6th to 9th century, 10th to 13th century, 14th to 20th century, etc. Of particular interest to the current study is the reconstruction of the interval between the 10th century and the 13th century, for which the phonetic forms are missing in Li and Zhou (1999).

This study focuses on the diachronic developments of sibilant contrasts relevant to the [_i] context. Thus, the syllables of interest to the current research are those developed into /sɹ̩-ɕi-ʂɻ̩/, /tsɹ̩-tɕi-tʂɻ̩/, and /tsʰɹ̩-tɕʰi-tʂʰɻ̩/ in Mandarin Chinese. A total of 262 Chinese characters were identified from the list in Li and Zhou (1999), as illustrated in Table 6.

Table 6 Examples of Chinese characters collected from Li and Zhou (1999)

Syllable in Mandarin	*Example items*		*No. of characters in Li & Zhou (1999)*
a. sɹ̩	私 'private'		14
b. tsɹ̩	資 'resources'		16
c. tsʰɹ̩	疵 'blemish'		12
d. ɕi	西 'west'	戲 'opera'	32
e. tɕi	飢 'hungry'	雞 'fowl'	54
f. tɕʰi	七 'seven'	欺 'to cheat'	35
g. ʂɻ̩	詩 'poem'		41
h. tʂɻ̩	枝 'branch'		29
i. tʂʰɻ̩	癡 'stupid'		29

2.3.2 Sibilants contrasts in sound change

The diachronic developments of the 262 Chinese characters are laid out in (13), where each Chinese character represents a group of Chinese characters with a similar path of development. The phonetic forms in the 7th century, the 14th century, and the 20th century are taken from Li and Zhou (1999) and the phonetic forms in the 11th century follow the reconstructed onsets and rimes in L. Wang (1985). The phonetic forms in the 18th century in Column III are added following B. Ye (2001:223-225). More specifically, between the 14th century and the 20th century, the velar obstruents in Column III developed into palatal sibilants before [_i], which formed contrasts with the dental sibilants in Column II; then, the dental sibilants in Column II started to be palatalized, i.e., merging with the palatal sibilants in Column III, which was documented in *Yuanyin Zhengkao*, a phonological work complied in 1743 (B. Ye 2001:223-225) and this process was completed by 20th century.

The developments in (13) involved a number of sound changes, as explained below.

(a) The phonology of Chinese around the 7th century had a complex rime system (Wang 1980, 1985) and, according to Li and Zhou (1999) and X. Guo (1986), the 262 Chinese characters belonged to a diverse range of rime categories. (A list of the rime categories of the 262 characters is given in Appendix III). Between the 7th century and the 14th century, there occured a general rime merger (Chen 1976:178; L. Wang 1985:490-525), which neutralized all the diverse rimes into [-i]. In other words, all the 262 Chinese characters developed a rime [-i] between 7th century and 14th century (L. Wang 1985).

(13) Historical development of sibilants in Mandarin (Based on Li and Zhou 1999; L. Wang 1985)

	I.					*II.*			*III.*			*IV.*				
	a.	*b.*	*c.*	*d.*	*e.*	*f.*	*g.*	*h.*	*i.*	*j.*	*k.*	*l.*	*m.*	*n.*	*o.*	*p.*
7th ct.	si 私	tsi 咨	tsʰi 此	dzi 自	zi 嗣	sV	tsV	tsʰV	xV	kV	kʰV	ɕi 屍	tɕi 脂	tɕʰi 齒	dʑi 謚	ʑi 時
11th ct.	① sɹ̩ 私	tsɹ̩ 咨	tsʰɹ̩ 此			② si 西	tsi 齊	tsʰi 妻	xV	kV	kʰV	ɕi 屍	tɕi 脂	tɕʰi 齒		
14th ct.	sɹ̩ 私	tsɹ̩ 咨	tsʰɹ̩ 此			si 西	tsi 齊	tsʰi 妻	xi 喜	ki 其	kʰi 及	③ ʂɻ̩ 屍	tʂɻ̩ 脂	tʂʰɻ̩ 齒		
18th ct.	sɹ̩ 私	tsɹ̩ 咨	tsʰɹ̩ 此			si 西	tsi 齊	tsʰi 妻	④ ɕi 喜	tɕi 其	tɕʰi 及	ʂɻ̩ 屍	tʂɻ̩ 脂	tʂʰɻ̩ 齒		
20th ct.	sɹ̩ 私	tsɹ̩ 咨	tsʰɹ̩ 此			⑤ merged into ɕi	tɕi	tɕʰi	ɕi 喜	tɕi 其	tɕʰi 及	ʂɻ̩ 屍	tʂɻ̩ 脂	tʂʰɻ̩ 齒		

Notes: (i) In Column II and Column III, a symbol like 'sV' represents a group of syllables with a 's' onset and diverse rimes;

(ii) For a particular historical period, e.g., 7th century, the square(s) highlight the sibilant place contrasts in the [_i] context;

(iii) In the same column, light color indicates that there was no sound change in that particular historical stage;

(iv) In the same column, a filled arrow marks a change across two historical stages that led to sibilant place contrasts in the [_i] context; an empty arrow marks a change across two historical stages that avoided the sibilant place contrasts in the [_i] context.

(b) Through the general *rime merger*, all the syllables in (13) neutralized their rimes into [-i] (Chen 1976:178; L. Wang 1985:490-525). It is shown in (13) that the different columns underwent the rime merge at different steps. For examples, the syllables in Column I completed the rime merger around the 7th century while those in Column II completed it around the 11th century, as marked by ②.

(c) In the 7th century, the syllables represented by Column I (*d*, *e*) and Column IV (*o*, *p*.) had voiced sibilants as the onsets, e.g., [dz z] and [dʑ ʑ]. Between the 7th and the 11th century, these onsets underwent a devoicing process and all merged into their voiceless counterparts (Chen 1976).

(d) Dental vs. palatal sibilants formed contrasts in the [_i] context around the 7th century, as highlighted by the boxes. Between the 7th and the 11th century, the apical vowels were formed after the dental sibilants (*apical vowel formation* as marked by ①), which avoided the sibilant place contrasts of dental vs. palatal sibilants.

(e) Around the 11th century, contrastive dentals vs. palatals appeared in the [_i] context again as a result of the rime merger in (a). The contrasts were avoided with ③, by which [ɕi tɕi tɕʰi] developed into [ʂɻ̩ tʂɻ̩ tʂʰɻ̩]. A closer look at ③ revealed that there were in fact two sequential sound changes: *Palatal retroflexion* in the onsets and *apical vowel formation* in the rime.

e.1) In *palatal retroflexion*, the palatal sibilant onsets *merged* into their retroflex counterparts. By 'merge', it means that retroflex sibilants existed in the sound system in the 7th century (Li 1956, Chen 1976) though the retroflexes did not combine with [-i].

e.2) In *apical vowel formation*, the rime [-i] developed into the apical vowel ɻ̩. As noted in L. Wang (1985), all of the apical ɻ̩ developed from a former [-i]. This is similar to the formation of apical vowels after the dental sibilants in ①.

e.3) In other words, the sound changes in ③ applied in two steps, e.g., first *palatal retroflexation* (e.g., ɕi → ʂi) and then *apical vowel formation* (e.g., ʂi → ʂɻ̩). The two-step change is evidenced by the development of Chinese characters like 濕 'wet', e.g., ɕĭĕp (7th c.) > ʂi (14th c.) > ʂɻ̩ (20th c.), for which the phonetic forms are drawn from Li and Zhou (1999).

(f) Between the 14th and the 18th century, the velar obstruents in Column III developed into palatal sibilants, i.e., *velar palatalization*, as marked by ④. With the syllables in Column II, *velar palatalization* turned out to create again the contrasts between dental and palatal sibilants in the [_i] context.

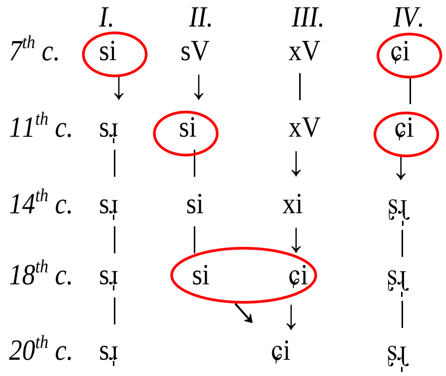
(g) Around the 18th century, the dental sibilants in Column II merged with their palatal counterparts (B. Ye 2001:223-225), i.e., *dental palatalization* ⑤. As a result, the contrasts between Column II and Column III around the 18th century were neutralized.

As the output of all the above sound changes, the phonotactics of sibilant-initial syllables in Mandarin finally came to be its current forms, i.e., /sɹ̩-ɕi-ʂɻ̩/, /tsɹ̩-tɕi-tʂɻ̩/, and /tsʰɹ̩-tɕʰi-tʂʰɻ̩/.

2.3.3 Summary of the historical development

The historical development of Mandarin sibilants in (13) can be summarized as (14) below, using the fricatives to represent all the sibilants.

(14) Sibilant contrasts in Mandarin sound change (Based on Li and Zhou 1999; L. Wang 1985)



The historical development shows that:

(a) In the [_i] context, the place contrasts of dental vs. palatal sibilants have been avoided, e.g., via rime change (7th c. – 11th c.), onset change (11th c.– 14th c.), and contrast neutralization (18th c. – 20th c.), even though they appeared multiple times as the result of other processes;

(b) In the case of onset change (*palatal retroflexion* around the 11th c.), the place contrast of dental vs. palatal sibilants developed into the contrast of dental vs. retroflex sibilants, e.g., si vs. ɕi → si vs. ʂi;

(c) In the case of the sound change around 18th century, the neutralization of the dental vs. palatal sibilants in the [_i] context seemed to be the only way to avoid their contrasts, as there was no room to shift the onsets or rimes as in the 7th century and the 11th century.

2.4 Research questions

Previous studies in language typology and historical sound change have both suggested that sibilant place contrasts tend to be avoided in the [_i] context (Stevens et al. 2004, M. Li 2011, Lee-Kim 2014a, among others) and that less distinct sibilant contrasts sometimes develop into more distinct contrasts in diachronic sound changes (Padgett 2001, 2003, Padgett and Źygis 2007;

Źygis and Padgett 2010, among others).

This study focuses on the sibilant place contrasts. The observations from the typology across Chinese dialects and the development of Mandarin sibilants are summarized below.

(a) The typology shows a clear tendency to avoid dental vs. palatal sibilants in the [_i] context independent of whether there are retroflex sibilants in the sound system (2.2.1.–2.3.2). The typology also shows that the introduction of apical vowels is the most common way to enhance the sibilant place contrasts in the [_i] context (2.2);

(b) The historical sound changes show that the dental vs. palatal contrasts in the [_i] context were once shifted to the dental vs. retroflex contrasts (2.3.2);

(c) For contrastive dental vs. palatal sibilants in the [_i] context, the typology shows that the place contrast of affricates implies the place contrast of fricatives, although the number of dialects in the typology is relatively small.

Assuming a role of contrast distinctiveness in language typology and sound change, the observations listed above raise the following three research questions:

i. For dental vs. palatal sibilants, does vowel context influence the perceptual distinctiveness of sibilant place contrasts? For example, referring to the generalization in (a), is it true that [si-ɕi] is perceptually less distinct than [sa-ɕa] and [sɹ̩-ɕi]?

ii. In the [_i] context, do sibilant place contrasts differ in their perceptual distinctiveness? For example, referring to the generalization in (b), is it true that [si-ɕi] is perceptually less distinct than [si-ʂi]?

iii. For dental vs. palatal sibilants in the [_i] context, do contrasts in different manners differ in their perceptual distinctiveness? For example, based on the generalization in (c), is that true that [tsi-tɕi] and [ts^{h}i-tɕ^{h}i] are perceptually less distinct than [si-ɕi]?

These research questions are to be addressed in the perceptual experiment reported in Chapter 3.

Chapter 3. Perceptual distinctiveness of sibilant contrast

For the three research questions raised in Chapter 2, three hypotheses are laid out in 3.1. The design of a speeded-AX discrimination experiment and its results are reported in 3.2, 3.3, and 3.4. A discussion of the experiment is given in 3.5.

3.1 Hypotheses of sibilant place contrasts

3.1.1 Perceptual distinction in different vowel contexts

The first research question is 'for dental vs. palatal sibilants, do vowel contexts influence the perceptual distinctiveness of sibilant place contrasts?' For this question, the following is hypothesized:

(15) *Hypothesis I* (*Perceptual distinction in different vowel contexts*)

Dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts.

This hypothesis is put forward based on the typology of sibilant place contrasts across Chinese dialects and other languages (Lee-Kim 2014a) and the perceptual analysis of sibilant contrasts in Mandarin (Stevens et al. 2004, Lee and Li 2003). More specifically,

(a) The typology shows that, across Chinese dialects, contrastive dental vs. palatal sibilants in the [_i] context tend to be avoided, primarily with the introduction of apical vowels;

(b) Perceptual accounts assume that the acoustic properties of the dental, palatal, and retroflex sibilants are close to each other (Stevens et al. 2004) and that their place distinction in the [_i] context is likely to be compromised due to palatalization (Lee and

Li 2003, Lee-Kim 2014a). Lee-Kim (2014a) further noted that the [_i] context weakens the formant transition cues to distinguish sibilant places.

With this hypothesis, a pair like [si-ɕi] is predicted to be perceptually less distinct than [sa-ɕa]. Similarly, [tsi-tɕi] and [ts^hi-$tɕ^hi$] are predicted to be less distinct than [tsa-tɕa] and [ts^ha-$tɕ^ha$] respectively.

3.1.2 Perceptual distinction in different place contrasts

The second research question is 'in the [_i] context, do sibilant place contrasts differ in their perceptual distinctiveness?'. For this question, the following hypothesis is made:

(16) *Hypothesis II* (*Perceptual distinction between place contrasts*)

In the [_i] context, the place contrast of dental vs. retroflex sibilants is more distinct than the place contrast of dental vs. palatal sibilants.

This hypothesis was based on the historical development of Mandarin and the experimental study of a similar sound change in Polish (Źygis and Padgett 2010). More specifically,

(a) A survey of historical development of Mandarin sibilants showed that, between the 11^{th} century and the 14^{th} century, the dental vs. palatal contrasts shifted to the dental vs. retroflex contrasts (c.f., 2.3 for details);

(b) In Polish, the [ɕ-$ʃ^j$] contrast developed into the [ɕ-ʂ] contrast (Padgett and Źygis 2007; Źygis and Padgett 2010), and the perceptual experiment by Źygis and Padgett (2010) showed that the [ɕ-ʂ] contrast is perceptually more distinct than the [ɕ-$ʃ^j$] contrast. Simply put, the historical development shifted a less distinct contrast to a more distinct contrast.

Note that, in the survey of historical development, the current work focuses on the sibilant place contrasts in the [_i] context only. Therefore, this study would limit Hypothesis II to the sibilant place contrasts in the [_i] context. With this hypothesis, a pair like [si-ɕi] is predicted to be perceptually less distinct than [si-s̺i]. Similarly, [tsi-tɕi] and [tsʰi-tɕʰi] are predicted to be less distinct than [tsi-ts̺i] and [tsʰi-ts̺ʰi], respectively.

3.1.3 Perceptual distinction in different manners of articulation

The third research question is 'for dental vs. palatal sibilants in the [_i] context, do contrasts in various manners differ in their perceptual distinctiveness?'. For this question, the following hypothesis is made:

(17) *Hypothesis III* (*Perceptual distinction between manners of articulation*)

For contrastive dental vs. palatal sibilants in the [_i] context, the place contrast of affricates is less distinct than the place contrast of fricatives.

This hypothesis was based on the typology of sibilant contrasts across Chinese dialects. More specifically, the typology of the sibilant place contrasts across Chinese dialects showed that, for contrastive dental vs. palatal sibilants in the [_i] context, the place contrast of affricates implies the place contrast of fricatives (c.f. 2.2 for details). Based on this hypothesis, [tsi-tɕi] and [tsʰi-tɕʰi] are predicted to be perceptually less distinct than the pair [si-ɕi].

3.2 Experimental studies of contrast distinctiveness

The evaluation of perceptual distinctiveness between sound pairs can be achieved with various experiments, e.g., similarity rating (Greenberg and Jenkins 1964; Mohr and Wang 1968) and AX discrimination (Pisoni 1973; Johnson and Babel 2010; Babel and Johnson 2010, among others).

The listeners' perceived distinctiveness has been shown to be influenced by both the psychophysical similarity of the sounds in the human auditory system (Pisoni 1973; Werker and Logan 1985; Johnson and Babel 2010) and the contrast and allophony in the listener's native language (e.g., Gandour 1983; Kuhl, Williams, Lacerda, Stevens and Lindblom 1992; Flege, Takagi and Mann 1996; Dupoux, Kakehi, Hirose, Pallier and Mehler 1999; Best, McRoberts and Goodell 2001, Hume and Johnson 2001; Boomershine, Hall, Hume and Johnson 2008).

In AX discrimination tasks, the listeners have been shown to access low-level acoustic information about a speech stimulus (Pisoni 1973; Pisoni and Tash 1974, Werker and Logan 1985, among others). For example, Pisoni and Tash (1974) observed that, among the listeners' 'different' responses, a longer response time was induced by stimulus pairs that were acoustically more similar than by those that were acoustically more different. Yet, studies have also shown the influence of the listeners' language background in AX discrimination. For example, Boomershine, Hall, Hume, and Johnson (2008) tested the discrimination of [ð], [d], and [ɾ] by native listeners of English and Spanish, by which [d-ɾ] are allophonic in English but phonemic in Spanish, and [ð-d] are allophonic in Spanish but phonemic in English. They observed that, in discriminating [d-ɾ], English listeners were slower than Spanish listeners, while in discriminating [ð-d], Spanish listeners were slower than English listeners.

To bypass the influence of the listeners' L1 phonology, Johnson and Babel (2010) and Babel and Johnson (2010) proposed the speeded AX discrimination paradigm, which has the following properties. First, the Inter-Stimulus-Interval is set to be short, with 100 ms as a common duration; second, the listeners are encouraged to respond as quickly as possible, typically under time pressure, e.g., with 500 ms as a goal; and third, they are informed of their response time and accuracy after every trial (see also McGuire 2010). For instance, Johnson and Babel (2010) tested English and Dutch listeners' discrimination of the fricatives [f, θ, s, ʃ, x, h] embedded in the contexts [i_i], [a_a], [u_u] using the speeded paradigm. The phonemic systems of these voiceless fricatives for English an Dutch are different: English has /f, θ, s, ʃ, h/ and Dutch has /f, s, x, h/, with [ʃ] as an allophone of /s/ (Booij 1999, Johnson and Babel 2010). The experimental results showed no evidence of the effect of the listeners' native languages on the response time, which indicates that the speeded nature of the experiment was able to access the psychoacoustic similatity of the stimulus pairs independently of the listners' L1 phonology.

The current study focuses on the perceptual distinctiveness of sibilant contrasts, with the hypotheses listed in 3.1. To test these hypotheses, it is desirable to adopt a method to assess the perceptual distinctiveness of sound pairs that is minimally affected by the L1 background of the listeners. Following Johnson and Babel (2010) and Babel and Johnson (2010), a speeded AX discrimination method was adopted. The experimental studies by Johnson and Babel (2010) and Babel and Johnson (2010) showed that speeded AX discrimination can obtain perceptual distinctiveness independent of the listener's native phonology. Li and Zhang (2017) observed a similar pattern in their speeded AX discrimination experiment that tested the effect of the following vowel on the perceptial distinctiveness of sibilant place contrast with English and Chinese listeners. More specifically, they observed a vowel effect on response time, by which the [_i] context introduced a longer response time than other vowel contexts; on the other hand, there was no evidence showing that the listeners' native languages influenced their responses or modulated the effect of other predicting variables, despite the difference in the sibilant inventories of English and Chinese. This, again, indicates that the speeded AX discrimination paradigm is able to access the psychoacoustic perception of the stimulus pairs independently of the listeners’ L1 phonology.

3.3 A speeded-AX discrimination experiment on contrast distinctiveness

3.3.1 Method

3.3.1.1. Participants

For the convenience of subject recruitment, native English listeners were recruited as listeners in this experiment. As Li and Zhang (2017) observed no difference between Chinese and English listeners in the results of their speeded AX discrimination experiment, the responses from English listeners in this dissertation would be taken to indicate the perceptual distinctiveness independent of the listeners’ native language.

Thirty-one native English listeners were recruited to participate in the experiment, all of whom were undergraduate students at the University of Kansas who received extra course credits for the experiment participation. The participants completed a consent form and questionnaire on their language background. Two participants (F06 and F13) reported hearing deficiencies in their questionaire and their data were excluded from the analysis. In the remaining 29 participants, there were 24 female and 5 male listeners.

3.3.1.2. Stimuli

This experiment aimed to investigate the perceptual distinction of sibilant place contrasts in [_i] and other vowel contexts. Therefore, the stimuli needed to be CV pairs whose onsets are /s ts ts^h/, /ɕ tɕ tɕ^h/, and /ʂ tʂ tʂ^h/ and whose vowels are [_i], [_a].

A. Stimulus syllables

To obtain the stimulus syllables, it is necessary to record the natural production from a language that preserves the contrasts like [si tsi ts^{h}i] vs. [ɕi tɕi tɕ^{h}i] vs. [ʂi tʂi tʂ^{h}i]. As shown in the typology in Section 2.2.2, such languages are extremely rare across Chinese dialects, with three-way sibilant place contrasts reported only in Qimen (T. Shen 1989) and Datan (Y. Chen 2015). Even in these two dialects, T. Shen (1989) reported that /tsi/ is not a legal syllable in Qimen while Y. Chen (2015) reported that /tsi/ and /ʂi/ are illegal in Datan. However, the three-way contrasts of sibilant places, e.g., [si-ɕi-ʂi], are preserved in the speech and singing of Peking opera, a traditional Chinese vocal performance. Therefore, a trained actor of Peking Opera was asked to produce the stimulus syllables needed for the experiment.

A list of the CV syllables that were recorded is given below in (18). Note that, apart from the syllables ending with [_i] or [_a], the syllables /sɹ̩, ʂɻ̩, tsɹ̩, tʂɻ̩, ts^{h}ɹ̩, tʂ^{h}ɻ̩/ were also recorded, which would be used to form stimulus pairs like [sɹ̩-ɕi], i.e., the actual phonetic contrasts as in Mandarin. To investigate the influence of vowel context on sibilant place contrast, a [si-ɕi] pair will be compared with [sa-ɕa] as well as with [sɹ̩-ɕi].

(18) Stimulus syllables for the perceptual experiment[13]

Onsets	I. [_i]	II. [_a]	III. [_ɹ̩] or [_ɻ̩]
a. s	**si** 西 ‘west’	**sa** 撒 ‘to release’	**sɹ̩** 思 ‘to think’
b. ɕ	**ɕi** 兮 particle	**ɕa** 瞎 ‘blind’	
c. ʂ	**ʂi** 世 ‘world’	**ʂa** 殺 ‘to kill’	**ʂɻ̩** 失 ‘to lose’
d. ts	**tsi** 齏 ‘fragment’	**tsa** 咂 ‘to smack lips’	**tsɹ̩** 资 ‘capital’
e. tɕ	**tɕi** 雞 ‘rooster’	**tɕa** 佳 ‘good’	
f. tʂ	**tʂi** 知 ‘to know’	**tʂa** 扎 ‘to tie’	**tʂɻ̩** 枝 ‘branch’
g. tsʰ	**tsʰi** 七 ‘seven’	**tsʰa** 擦 ‘to wipe’	**tsʰɹ̩** 差 ‘uneven’
h. tɕʰ	**tɕʰi** 欺 ‘to cheat’	**tɕʰa** 掐 ‘to pinch’	
i. tʂʰ	**tʂʰi** 吃 ‘to eat’	**tʂʰa** 插 ‘to insert’	**tʂʰɻ̩** 蚩 a surname

Note: All syllables bear a high-level tone except ʂi 世 ‘world’, which has a falling tone. The speaker was asked to produce ʂi with a high-level tone.

B. Stimulus recording

A trained male actor of Peking opera, who is also a native Mandarin speaker, was invited to produce the syllables in (18). The speaker read the Chinese characters in Columns I and III in the normal *speech* of Peking opera (i.e., not the singing style). The characters in Column II were read as they are in Mandarin, not in Peking opera. This is because not all the syllables are legal forms in Peking opera.[14] The target Chinese characters were read in the carrier sentence ‘*wo shuo de shi __ zhe ge zi*’ [‘我說的是__這個字’] ‘what I said was __ this character’. The recording was done at a sampling rate of 44.1 KHz, 16 bits. The speakers produced six tokens of each

[13] In the Chinese Pinyin orthography, the [ɕa tɕa tɕʰa] syllables are represented as *xia*, *jia*, *qia*, where the ‘i’ indicates an onglide. In the phonetic and phonological literature, these syllables are sometimes transcribed as [ɕʲa tɕʲa tɕʰʲa]. However, as Ladefoged and Maddieson (1996:150) pointed out, the alleged onglide [j] involves ‘nothing other than a normal transition between the initial consonant and the following vowel in all these cases’. Therefore, these syllables are referred to as [ɕa tɕa tɕʰa] in this dissertation.

[14] The difference in reading styles, i.e., speaking in Peking opera vs. Mandarin, should not be problematic in the stimulus formation because, for all the stimulus pairs, the two syllables in a pair came from the same reading style. For example, in [si-ɕi] and [sɹ̩-ɕi], all syllables were read in the speaking of Peking opera and, in [sa-ɕa], both syllables were read in Mandarin.

syllable in (18). For each syllable, one token was selected from the six repetitions, whose sibilant intensity, COG, and vowel formants were the closest to the mean values of the six tokens.

C. Stimulus manipulation

In a speeded AX discrimination task, the duration of a stimulus syllable must be controlled to facilitate the comparison of response times across different CV pairs, e.g., [si-ɕi] vs. [sa-ɕa]. Several steps of manipulation were performed on the 24 selected syllables before the formation of CV pairs. (For the naturally produced syllables, the mean durations of the onsets and vowels are given in Appendix IV, along with the durations of CV transitions and the F0, F1, F2, F3 of the vowels.)

First, the duration of the onsets was normalized to a length typical of a sibilant in normal-speed speech. Based on Feng's study of Mandarin consonants (1985), 125 ms was used as the target duration for the fricatives [s, ɕ, ʂ], 50 ms for the unaspirated affricates [ts, tɕ, tʂ], and 100 ms for the aspirated affricates [ts^h, $tɕ^h$, $tʂ^h$]. The duration of a naturally produced onset sibilant was typically longer than its target duration, presumably because the stimulus syllables were produced in a focus position. The sibilant onsets were shortened to the target duration using the Manipulation function in Praat (Boersma 2001). Note that in Feng (1985), the sibilants at different places have slightly different mean duration, e.g., word-initially, [s] = 136ms, [ɕ] = 145 ms, [ʂ] = 134 ms; word-medially, [s] = 110 ms, [ɕ] = 122 ms, [ʂ] = 113 ms. However, a comparison of the dental, palatal, and retroflex sibilants across Chinese dialects shows no consistent pattern of duration difference: The dentals/retroflexes were reported to be longer than the palatals in some dialects and the reverse was reported in others (Ran 2005; C. Liu 2010; Pan

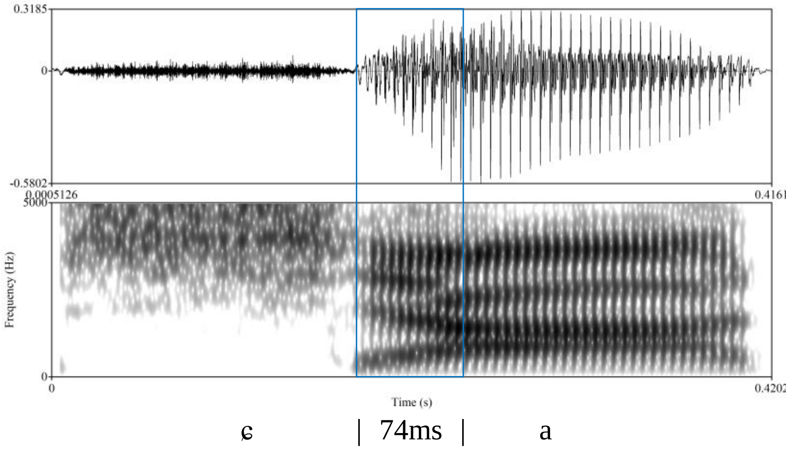
2010, among others). Therefore, in this dissertation, the same duration was used for sibilants in the same manners of articulation.

Second, the vocalic portion of a CV syllable was normalized to 120 ms to match the vowel duration in natural speech (Feng 1985).[15] More specifically, the consonant-vowel transition portion was normalized to 50 ms (Delattre, Liberman, Cooper 1955) and the steady vowel portion (i.e., the steady vowel formants) was normalized to 70 ms. The vowel duration in a naturally produced syllable was typically longer than 120 ms (see Appendix IV), and thus the manipulation was primarily shortening.

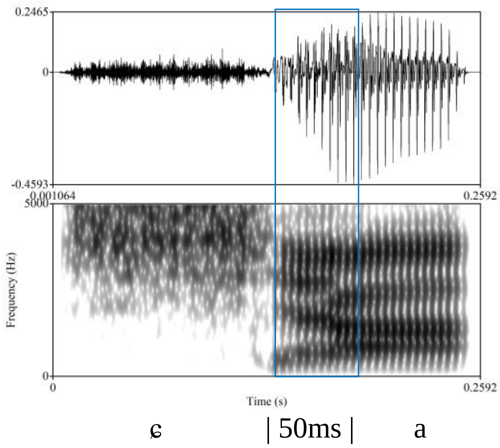
In the manipulation of [ɕa], for example, two intervals were marked on the vowel portion: the CV transition (i.e., from the start of the vocalic part to the start of the steady formants) and the steady vowel (i.e., the rest of the vowel). For the CV transition portions, the duration in the naturally produced tokens was generally between 70 and 90 ms, and this interval was shortened to 50 ms, as illustrated in Figure 1. The shortening was performed in Praat, which adopts the PSOLA (Pitch-Synchronous Overlap-Add) technique (Moulines and Charpentier 1990): A series of frames was created, each centered on a point of maximum excursion, and certain inner frames were eliminated at equal distance, depending on the ratio of 50 ms to the interval duration in the natural token. Then, a waveform was resynthesized by overlapping and adding the remaining frames. As Figure 1 illustrates, the transition portion of the manipulated [ɕa], as in (b), was close

[15] The phonetics literature has observed systematic durational differences among vowels, e.g., low vowels tend to be longer than high vowels (Lehiste 1970; Feng 1985). But, to facilitate the comparison of response time across vowel contexts, the vowels were all controlled to be 120 ms (50 ms CV transition plus 70 ms steady vowel formants).

a. Syllable [ɕa] produced by the speaker



b. Stimulus [ɕa] manipulated from the naturally produced [ɕa]



c. CV transition in Polish [ɕa] syllable (from [ɕali])

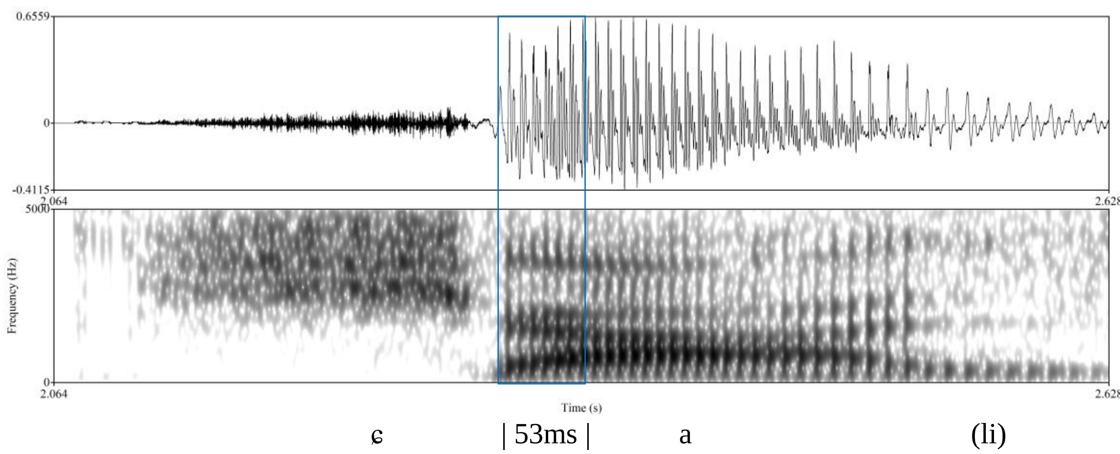


Figure 1 Manipulation of the CV transition, e.g., in [ɕa], focusing on the CV transition

(Polish sound obtained from http://www.phonetics.ucla.edu/course/chapter7/polish/polish.html) to the CV transition in the naturally produced [ɕa] in Polish in terms of duration and F2 onset, as in (c). The duration of the steady vowel in a naturally produced token was generally longer than 70 ms. This interval was shortened to 70 ms, following the procedure as above, to arrive at a vowel of 120 ms (50 ms of CV transition and 70 ms of steady formants). Similar manipulation was applied to all the stimulus syllables in (18). In general, the manipulated syllables all sounded natural in terms of the consonant-vowel transition.

Third, across all the stimulus syllables, a level F0 of 200Hz was superimposed on the vocalic portion, which aimed to control the influence of pitch on the perceived distinctiveness within each CV pair. The pitch manipulation was done in Praat, which uses the PSOLA technique (Moulines and Charpentier 1990) as described above.

Fourth, for all the manipulated syllables, the root-mean-square intensity of the vocalic portion, e.g., [i] as in [si], was normalized to 72dB, which was the mean intensity of the vowel portions of all the naturally produced tokens. The amplitude of the vowel faded out to zero within the last 20 ms. The intensity of the onset sibilants, e.g., [s] as in [si], was manipulated to mean of the onset intensity of the 6 repetitions in the speaker's production, e.g. the 6 [si]'s produced by the speaker. In general, the relative intensity of the consonants and vowels in the stimulus syllables were similar to that in the natural production.

To sum up, with the selection and manipulation listed above, the stimulus syllables used in the perceptual experiment had the following properties:

(a) The stimulus syllables are all manipulated from naturally produced tokens;

(b) For the sibilant onset, the duration was controlled (fricatives = 125 ms, unaspirated affricates = 50 ms, aspirated affricates = 100 ms) and close to the duration in normal-speed speech, the COG was closest to the mean of the 6 tokens in the production, and the intensity was normalized to the mean of the 6 tokens of each syllable.

(c) For the vowel portion, the duration was controlled to be 120 ms, close to the value in normal-speed speech; the transition was controlled to be 50 ms; the formant frequencies were close to the mean values of the 6 tokens in the speaker's production, and the intensity was normalized to 72dB — the mean values of all the naturally produced syllables.

(d) All syllables have a level F0 at 200 Hz.

D. Acoustic properties of the stimulus syllable

Table 7 lists the acoustic properties of the stimulus syllables. For a sibilant onset, the COG and dispersion was measured over the center 80% of the sibilants for the frequency range 0-10kHz, using a praat script written by DiCanio (2013). The intenstity was measured over the entire consonant in Praat. For the aspirated affricates [ts^h] and [$tɕ^h$], the COG was measured on the entire turbulent noise before the aspiration portion. For the vowel portion; the formants were measured at the midpoint of the steady vowel portion. The waveforms and spectrograms of the stimulus syllables after the manipulation are provided in Appendix V.

Table 7 Acoustic properties of the stimulus syllables (after manipulation)

a. Syllables with fricative onsets

Syllable	Onset sibilant			Steady vowel portion		
	COG(Hz)	*Dispersion(Hz)*	*Intensity(dB)*	*F1(Hz)*	*F2(Hz)*	*F3(Hz)*
si	8119	1977	59	321	2570	2904
ɕi	4719	1374	62	318	2590	2977
ʂi	4230	2127	62	315	2547	3025
sɹ̩	8472	1747	61	461	1291	3174
ʂɻ̩	3935	2086	63	342	1786	2460
sa	6221	1383	58	928	1392	2639
ɕa	4588	1149	59	933	1391	2569
ʂa	3385	1363	59	1014	1392	2630

b. Syllables with unaspirated affricate onsets

Syllable	Onset sibilant			Steady vowel portion		
	COG(Hz)	*Dispersion(Hz)*	*Intensity(dB)*	*F1(Hz)*	*F2(Hz)*	*F3(Hz)*
tsi	7850	1939	59	325	2567	2868
tɕi	5535	1509	64	317	2607	2941
tʂi	3972	1782	62	317	2562	2847
tsɹ̩	8588	1917	57	382	1322	3157
tʂɻ̩	4202	2051	63	331	1884	2545
tsa	6828	1747	62	968	1343	2729
tɕa	5504	1146	58	975	1403	2593
tʂa	3743	1563	62	990	1406	2615

c. Syllables with aspirated affricate onsets

Syllable	Onset sibilant			Steady vowel portion		
	COG(Hz)	*Dispersion(Hz)*	*Intensity(dB)*	*F1(Hz)*	*F2(Hz)*	*F3(Hz)*
tsʰi	9318	1471	60	310	2566	3154
tɕʰi	5479	1714	64	315	2557	3048
tʂʰi	3595	1674	65	316	2562	3129
tsʰɹ̩	8406	1925	62	376	1298	3322
tʂʰɻ̩	3306	1614	66	327	1817	2770
tsʰa	6444	1497	57	910	1313	2610
tɕʰa	5228	1098	59	936	1360	2574
tʂʰa	3922	1586	56	943	1348	2692

E. Formation of stimulus pairs

After the manipulation, the CV pairs in (19) were formed. Note that (19) includes the place contrasts of dental vs. palatal sibilants (Group A) and dental vs. retroflex sibilants (Group B). These two place contrasts are relevant to the three hypotheses laid out above. Both place contrasts appeared in three different vowel contexts: the [_i] context, the [_a] context, and the 'homorganic' context by which [_i] follows the palatals, [_ɹ̩] follows the dentals, and [_ɻ̩] follow the retroflexes. The complete combination of all onset pairs and vowel contexts aims to balance the sibilant conditions and the vowel conditions.

(19) Stimulus CV pairs used in the perceptual experiment

Sibilant contrast	I. [_i]	II. [_a]	III. homorganic
A. s-ɕ	si-ɕi	sa-ɕa	sɹ̩-ɕi
ts-tɕ	tsi-tɕi	tsa-tɕa	tsɹ̩-tɕi
tsʰ-tɕʰ	tsʰi-tɕʰi	tsʰa-tɕʰa	tsʰɹ̩-tɕʰi
B. s-ʂ	si-ʂi	sa-ʂa	sɹ̩-ʂɻ̩
ts-tʂ	tsi-tʂi	tsa-tʂa	tsɹ̩-tʂɻ̩
tsʰ-tʂʰ	tsʰi-tʂʰi	tsʰa-tʂʰa	tsʰɹ̩-tʂʰɻ̩

Four stimulus pairs were formed for each CV pair in (19). For the [si-ɕi] contrast, for example, [si-ɕi] and [ɕi-si] were formed as the different pairs and [si-si] and [ɕi-ɕi] were formed as the identical pairs. Thus, the 18 cells in (19) led to a total of 72 stimulus pairs. Within each stimulus pair, the Inter-Stimulus-Interval (ISI) was set as 100 ms to facilitate responses based on the psychoacoustic difference between the two sounds (Pisoni 1973; Werker and Logan 1985; Johnson and Babel 2010). An additional 50 ms was added between the pairs whose onsets were [ts tɕ tʂ] and [tsʰ tɕʰ tʂʰ] to compensate for the duration of oral closure before the release of an affricate.

3.3.1.3. Procedure

The experiment was programmed in Paradigm (Perception Research Systems 2007). The listeners were told that they would listen to sound pairs from an unknown language. On hearing each pair, they were asked to judge if the two sounds are the same by pressing 'same' or 'different' on a button box. The listeners were all right-handed and therefore the same button box setting was used, with 'same' on the left-hand side and 'different' on the right-hand side.

The main experiment was preceded by a practice session. The audio stimuli in the practice session were 16 pairs of nonce syllables like [di-du] [du-di] [di-di] [du-du]. The main experiment was divided into three blocks. Within each block, the 72 CV pairs were played to the listeners in a randomized order.

Figure 2 illustrates the procedure of the speeded AX discrimination task. More specifically, on hearing a pair, the listeners were asked to press the button ('same' vs. 'different') as quickly as possible, with a response time goal of 500 ms, following Johnson and Babel (2010) and Babel and Johnson (2010).

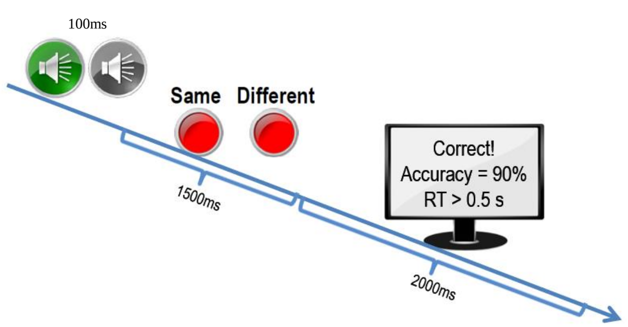


Figure 2 Procedure of the Speeded AX-discrimination

After the listener pressed the button, a feedback screen would appear that informed the listener of the following:

(a) the accuracy of his/her response to the current sound pair (Correct vs. Incorrect),

(b) the overall accuracy of his/her responses up to the current pair (e.g., 90% correct),[16]

(c) whether the response time of the current pair is longer than 500 ms or not (e.g., > 0.5s).

If they did not press the button within 1.5 seconds, the feedback screen would also appear. The feedback screen lasted for 2 seconds, then it disappeared and the next sound pair was played.

The whole experiment, including a language background questionnaire, lasted about 50 minutes. Each listener gave a total of 324 responses (18 comparisons × 4 pairs × 3 blocks) excluding the fillers.

3.3.2 Predictions

In a speeded AX-discrimination task, the response time was assumed to be inversely correlated to the perceptual distinctiveness of the two sounds in a pair. A shorter response time indicates greater perceptual distinctiveness, and vice versa. In addition, a perceptually less distinct pair is expected to introduce more discrimination errors (i.e., a phonetically different pair judged as 'same'). Based on the three hypotheses laid out in 3.1, the predictions are listed below.

3.3.2.1 Perceptual distinction in different vowel contexts

For the effects of vowel context on sibilant distinctiveness, it is hypothesized that dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts. With such a hypothesis, it is predicted that, for the same pair of sibilants, the vowel context [_i] will introduce a longer response time and more discrimination errors than the other vowel contexts, as in (20).

---

[16] Thanks to Bruno Tagliaferri for providing the script to perform this function in Paradigm.

(20) Predicted results of vowel effects for dental vs. palatal sibilants
(where '>' means 'longer RT than' or 'more discrimination errors than')

si-ɕi	>	sa-ɕa, sɹ̩-ɕi
tsi-tɕi	>	tsa-tʂa, tsɹ̩-tɕi
ts^{h}i-tɕ^{h}i	>	ts^{h}a-tɕ^{h}a, ts^{h}ɹ̩-tɕ^{h}i

3.3.2.2 Perceptual distinction in different place contrasts

For the relative perceptual distinctiveness of different place contrasts, it is hypothesized that, in the [_i] context, the contrasts of dental vs. palatal sibilants are less distinct than the contrasts of dental vs. retroflex sibilants. With such a hypothesis, it is predicted that, in the [_i] context, a CV pair with a dental vs. palatal contrast in the onset would introduce a longer response time and more discrimination errors than a CV pair with a dental vs. retroflex onset contrast , as in (21).

(21) Predicted results of place difference for dental-palatal vs. dental retroflex in the [_i] context
(where '>' means 'longer RT than' or 'more discrimination errors than')

si-ɕi	>	si-ʂi
tsi-tɕi	>	tsi-tʂi
ts^{h}i-tɕ^{h}i	>	ts^{h}i-tʂ^{h}i

3.3.2.3 Perceptual distinction in different manners of articulation

For the dental vs. palatal sibilant contrasts in the [_i] context, it is hypothesized that the place contrast in affricates is less distinct than the place contrast in fricatives. With this hypothesis, the results in (22) are predicted. Note that in the stimulus syllables, the affricates and fricatives have an intrinsic difference in the duration of their onsets; therefore, the response times of an affricate-onset pair and a fricative-onset pair cannot be directly compared. The prediction in (22) is made in terms of discrimination accuracy, by which a phonetically different pair is regarded as being perceptually less distinct if it incurs more discrimination errors compared with another pair.

(22) Predicted results of manner difference, where ‘<’ means ‘less accurate in discrimination’

[tsi-tɕi], [tsʰi-tɕʰi] < [si-ɕi]

3.3.3 Results

For each stimulus pair, the response time was calculated from the onset of the sibilant in the second stimulus, e.g., from the start of the frication noise of [ɕi] in the [si-ɕi] pair. The raw response time was transferred into Log Response Time (LogRT) and the analysis only included the listeners’ ‘different’ responses to phonetically different pairs (i.e., the correct responses to the different pairs). To exclude outliers in the responses, the data points outside 2.5 standard deviations from the mean LogRT were trimmed off by *listener* and by *manner of articulation*, more specifically, for the responses of each listener:

(a) the LogRT data were divided into 3 subparts depending on the manner of articulation of the onset sibilants, i.e., fricatives (s-ɕ, s-ʂ), unaspirated affricates (ts-tɕ, ts-tʂ), and aspirated affricates (tsʰ-tɕʰ, tsʰ-tʂʰ). This was necessary because the onset pairs with different manners of articulation differ in durations and the onset duration was included in the response time;

(b) mean and standard deviation of the LogRTs were calculated separately for the three subparts;

(c) within each part, data points outside 2.5 standard deviations were removed.[17]

---

[17] In some previous studies, the trimming was performed on the data points outside 2 standard deviations. A trimming criterion of 2.5 standard deviations was adopted in this study because, for each CV pair in (19), the number of responses from a participant was relatively small. With the time limit of the experiment, each listener was only able to give 6 responses for each of the 18 CV pairs, e.g., 3 responses for [si-ɕi] and 3 responses for [ɕi-si]. Thus, too many data points would be lost if the data were trimmed by 2 standard deviations.

The 5742 phonetically different CV pairs gave a total of 5703 responses from the 29 listeners, with 39 cases of 'no responses', i.e., when the listeners didn't respond within 1500 ms. Out of the remaining 5703 responses, 5080 were correct (= 89.1%) and 623 were incorrect (= 10.9%). The trimming, as described above, was applied to the 5080 correct responses, and a total of 120 data points (= 2.4%) were excluded as outliers (7 below 2.5 standard deviations and 113 above 2.5 standard deviations). The remaining 4960 data points were analyzed.

3.3.3.1 Place distinction in different vowel contexts

A. Response time

It is hypothesized that dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts. The data points relevant to this hypothesis are the response times to the CV pairs whose onsets are dental vs. palatal sibilants. For this dataset, the mean LogRTs of each CV pairs over the 29 listeners are plotted in Figure 3. For the [s-ɕ] contrast, for example, the mean response time was 6.26 (523 ms) in the [_i] context, 6.24 (513 ms) in the [_a] context and 6.20 (492 ms) in the allophonic contexts.

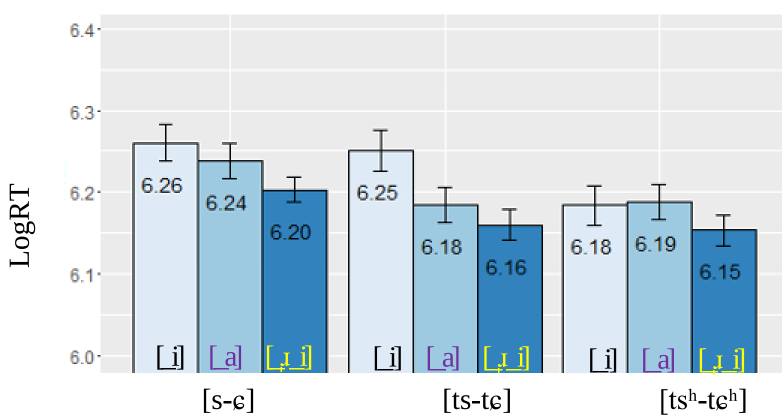


Figure 3 Mean LogRTs: Dentals vs. palatals in different vowel contexts.

(For LogRTs, 6.0 = 403ms, 6.1 = 446ms, 6.2 = 493ms, 6.3 = 545 ms, 6.4 = 602ms)

The LogRT data were analyzed in Linear Mixed Effects Models using the *lmer* function in the R package lme4 (Bates, Maechler, Bolker and Walker 2015a, b) and the p-values were determined by the R package lmerTest (Kuznetsova, Brockhoff, and Christensen 2015). For this dataset, the dependent variable, predicting variables and the random variables are listed in (23), with baselines of the predicting variables indicated.

(23) Variables in the analysis of vowel effects on sibilant distinctiveness: Dentals vs. Palatals

Variable names	*Notes and abbreviation*
Dependent variable	
LogRT	The log-transformed response time
Predicting variables	
Vowel	**[_i]**, **[_a]**, and homorganic vowel (**[_ɹ̩_i]**, ɹ̩ after dentals, i after palatals) (Baseline = [_i])
Onset	Fricative (**s-ɕ**), unaspirated affricate (**ts-tɕ**), aspirated affricate (**tsʰ-tɕʰ**) (Baseline = [**s-ɕ**])
Random variables	
Listener	29 listeners
Listener:CVPair	9 pairs (3 vowels × 3 onset manners) by each listener

The fixed effects in the final model are presented in Table 8. This model was obtained via two steps. First, a null model with Listener and Listener:CVPair as the random factors was compared with two models adding Onset or Vowel as the predicting variable. The addition of both variables significantly improved the model (Onset: $X^2 = 15.6$, $df = 2$, $p < .001$; Vowel: $X^2 = 14.9$, $df = 2$, $p < .001$). Second, a model with Onset and Vowel as the predicting variables and Listener and Listener:CVPair as the random factors was compared with a superset model with the Onset*Vowel interaction added. The addition of the interaction did not lead to significant improvement of the model. Therefore, the final model included Onset and Vowel only.

Table 8 Fixed effects in the mixed-effect linear regression for LogRT: Dentals vs. palatals

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	6.2595	0.0334	36.4	187.18	< .001***
Vowel(_a)	–0.0336	0.0147	225.7	–2.29	.023*
Vowel(_ɹ̩_i)	–0.0580	0.0145	220.2	–3.99	< .001***
Onset(ts-tɕ)	–0.0355	0.0145	216.5	–2.45	.015*
Onset(tsʰ-tɕʰ)	–0.0579	0.0143	209.9	–4.04	< .001***

Model: LogRT ~ Vowel + Onset + (1|Listener) + (1|Listener:CVPair)
Baselines: Vowel = [_i]; Onset = Fricarive (s-ɕ)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Vowel, the difference among the three vowel contexts [_i] [_a], and [_ɹ̩_i] (i.e., the 'homorganic' context in which [ɹ̩] follows the dentals and [i] follows the palatals) was further checked using [_i] and [_a] alternatively as the baseline. The differences among the vowel contexts, as indicated by t values (p values), are summarized in Table 9, which shows that the [_i] context led to a significantly longer response time than [_a] and [_ɹ̩_i], and that there was no significant difference between [_a] and [_ɹ̩_i].

Table 9 Dental vs. palatal sibilants: Differences among vowel contexts in *t* value (*p* value)

	[_a]	[_ɹ̩_i]
[_i]	–2.63 (**.009****)	–3.95 (< **.001*****)
[_a]		–1.33 (.186)

P values appear in parentheses and boldface marks those that reached significance (.05).

Note that, within the subset of dental vs. palatal sibilants, there was no Vowel*Onset interaction. Therefore, there was no evidence that the sibilant pairs with different manners differ in terms of the vowel effect ([_i] > [_a], [_ɹ̩_i]).

For the significant effect of Onset, the differences among the three onset pairs (i.e., [s-ɕ], [ts-tɕ], and [tsʰ-tɕʰ]) were further examined using [s-ɕ] and [ts-tɕ] alternatively as the baseline. The differences among the three onset pairs, as indicated by t values (p values), are summarized in Table 10. It turned out that [s-ɕ] led to a significantly longer response time than [ts-tɕ] and [tsʰ-tɕʰ], and that the latter two pairs had no significant difference.

Table 10 Dental vs. palatal sibilants: Differences among onset pairs in *t* value (*p* value)

	[ts-tɕ]	[tsʰ-tɕʰ]
[s-ɕ]	–2.71 (**.007****)	–3.85 (< **.001*****)
[ts-tɕ]		–1.09 (.276)

P values appear in parentheses and boldface marks those that reached significance (.05).

However, the effect of Onset is not interesting because in the stimulus pairs, the sibilants with different manners of articulation differ in their duration.

B. Accuracy

For the discrimination of dental vs. palatal sibilants in different vowel contexts, Table 11 below summarizes the numbers of discrimination errors, i.e., when a listener incorrectly judged a phonetically different pair as being the same.

Table 11 Numbers of discrimination errors for dental vs. palatal sibilants in different vowel contexts

Onset	_i	_a	_ɹ̩_i
s-ɕ	12	11	6
ts-tɕ	35	19	9
tsʰ-tɕʰ	25	8	9

To examine whether vowel context has an effect on the discrimination accuracy of dental vs. palatal sibilants, the accuracy data of the 29 listeners were analyzed in Mixed-effects Logistic Regression using the *glmer* function in the r package lme4 (Bates, Maechler, Bolker and Walker 2015a, b). To be more specific, the data included the accuracy data of the CV pairs whose onsets were [s-ɕ], [ts-tɕ], [tsʰ-tɕʰ]. For this analysis, the dependent variable, the predicting variable and its baseline, and the random variables are listed in (24).

(24) Variables in the analysis of accuracy: Dental vs. palatal sibilants in different vowel contexts

Variable names	*Notes and abbreviation*
Dependent variable	
Accuracy	0 = Incorrect, 1 = Correct
Predicting variables	
Vowel	[_i], [_a], [_ɻ_i]
	(Baseline = [_i])
Onset	[s-ɕ], [ts-tɕ], [tsʰ-tɕʰ]
	(Baseline = [s-ɕ])
Random variables	
Listener	29 listeners
Listener:CVPair	3 pairs ([si-ɕi] [tsi-tɕi] [tsʰi-tɕʰi]) by each listener

The fixed effects in the final model are presented in Table 12. This model was obtained by comparing a null model with Listener and Listener:CVPair as the random factors with two superset models with Vowel and Onset as the predicting variable respectively. The addition of Vowel and Onset both significantly improved the model (Vowel: X^2 = 27.273, *df* = 2, p < .001; Onset: X^2 = 13.281, *df* = 2, p = .001). Then, a model with Vowel and Onset as the predicting variables and Listener and Listener:CVPair as the random factors was compared with a superset model with the interaction between Vowel and Onset added. The addition of the Vowel*Onset interaction did not significantly improve the model. Therefore, the final model included only Vowel and Onset as the two predicting variables and Listener and Listener:CVPair as the random factors.

Table 12 Fixed effects in the mixed-effect logistic regression: Accuracy for dental vs. palatal sibilants.

	Estimate	*Std.Error*	*z value*	*Pr(>|z|)*
(Intercept)	2.4141	0.2593	9.310	< .001***
Vowel(_a)	0.7521	0.2156	3.489	< .001***
Vowel(_ɻ̩_i)	1.2588	0.2478	5.080	< .001***
Onset(ts-tɕ)	–0.9287	0.2405	–3.861	< .001***
Onset(tsʰ-tɕʰ)	–0.4547	0.2556	–1.779	0.075.

Model: Accuracy ~ Vowel + Onset + (1|Listener) + (1|Listener:CVPair)
Baseline: Vowel = [_i]; Onset = [s-ɕ]
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Vowel, the differences among the three vowel contexts [_i] [_a], and [_ɻ̩_i] were further checked using [_i] and [_a] alternatively as the baseline. The differences among the three vowel contexts, as indicated by z values (p values), are summarized in Table 13. It turned out that the [_i] context led to a lower accuracy (i.e., more discrimination errors) than [_a] and [_ɻ̩_i] and the latter two vowel contexts did not differ significantly in terms of accuracy.

Table 13 Dental vs. palatal sibilants: Differences among vowel contexts in *z* value (*p* value)

	[_a]	[_ɻ̩_i]
[_i]	3.312 (**.001*****)	4.832 (< **.001*****)
[_a]		1.798 (.072)

P values appear in parentheses and boldface marks those that reached significance (.05).

For the significant effect of Onset, the difference among [s-ɕ], [ts-tɕ], and [tsʰ-tɕʰ] was further checked using [s-ɕ] and [ts-tɕ] alternatively as the baseline. The results are reported in Table 14 in terms of z values (p values). It turned out that [s-ɕ] and [tsʰ-tɕʰ] both introduced a significantly higher accuracy (i.e., less discrimination errors) than [ts-tɕ], while [s-ɕ] and [tsʰ-tɕʰ] did not significantly differ from each other in terms of accuracy.

Table 14 Dental vs. palatal sibilants: Differences among onset pairs in *z* value (*p* value)

	[ts-tɕ]	[tsʰ-tɕʰ]
[s-ɕ]	–3.594 (<**.001*****)	–1.706 (.088)
[ts-tɕ]		1.986 (**.047***)

P values appear in brackets and boldface marks those that reached significance level (.05).

In generaly, the results of accuracy were consistent with the results of response time in that, for dental vs. palatal sibilants, the [_i] context introduced a longer response time and a higher likelihood of discrimination errors than the other two vowel contexts.

3.3.3.2 Perceptual distinction between place contrasts

A. Response time

For the relative perceptual distinctiveness of different place contrasts, it is hypothesized that, in the [_i] context, the contrasts of dental vs. palatal sibilants are less distinct than those of dental vs. retroflex sibilants. The relevant subset of the data is the 29 listeners' responses to the CV pairs whose vowel was [_i] and whose onsets are dental vs. palatal sibilants and dental vs. retroflex sibilants, i.e., the responses to the pairs [si-ɕi], [si-ʂi], [tsi-tɕi], [tsi-tʂi], [tsʰi-tɕʰi], [tsʰi-tʂʰi]. For this subset, the mean LogRTs of each CV pairs over the 29 listeners are plotted in Figure 4.

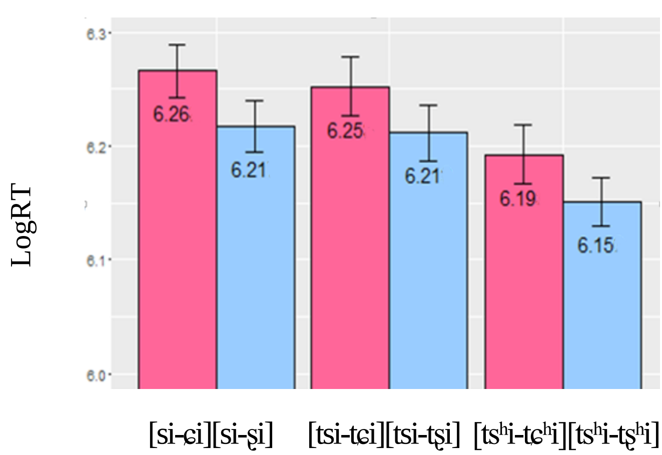


Figure 4 Mean LogRTs: Dental-palatals vs. dental-retroflexes.
(For LogRTs, 6.0 = 403ms, 6.1 = 446ms, 6.2 = 493ms, 6.3 = 545 ms, 6.4 = 602ms)

The logRT data were analyzed in Linear Mixed Effects Models. For this analysis, the dependent variable, predicting variables and random variables are listed in (25), with baselines of the predicting variables indicated.

(25) Variables in the analysis of Dental-palatal vs. dental-retroflex in the [_i] context

Variable names	*Notes and abbreviation*
Dependent variable	
LogRT	The log-transformed response time
Predicting variables	
Place	Dental-Palatal (**DP**), Dental-Retroflex (**DR**) (Baseline = Dental-Palatal (DP))
Onset	Fricative (***Fr***), unaspirated affricate (***Ua***), aspirated affricate (***Aa***) (Baseline = fricative(Fr))
Random variables	
Listener	29 listeners
Listener:CVPair	6 pairs ([si-ɕi][si-ʂi][tsi-tɕi][tsi-tʂi][tsʰi-tɕʰi][tsʰi-tʂʰi]) by each listener

The fixed effects in the final model are presented in Table 15. This model was obtained via two steps. First, a null model with Listener and Listener:CVPair as the random factors was compared with two models adding Place or Onset as the predicting variable. The addition of both variables significantly improved the model (Place: X^2 = 8.52, *df* = 1, p = .004; Onset: X^2 = 20.2, *df* = 2, p < .001). Second, a model with Place and Onset as the predicting variables and Listener and Listener:CVPair as the random factors was compared with a superset model with the Place*Onset interaction. The addition of the interaction did not lead to significant improvement of the model (X^2= 0.26, *df*= 2, p= .880). Therefore, the final model included Place and Onset as the predicting variables and Listener and Listener:CVPair as the random factors.

Table 15 Fixed effects in the mixed-effect linear regression for LogRT: Place contrasts in [_i] context.

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	6.253	0.036	35	171.66	< .001***
Place(DR)	–0.041	0.014	867	–2.90	.004**
Onset(Ua)	–0.005	0.017	867	–0.31	.758
Onset(Aa)	–0.069	0.017	867	–4.10	< .001***

Model: LogRT ~ Place + Onset + (1|Listener) + (1|Listener:CVPair)
Baselines: Place = Dental-Palatal (DP); Onset = Fricative(Fr)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

As shown in the the fixed effect of Place in Table 15, the LogRT of the dental vs. retroflex contrast (DR) is smaller than that of the baseline, i.e., the dental vs. palatal contrast (DP), and the difference reached the significance level. Therefore, for the effect of Place, the contrast of dental vs. palatal sibilants (DP) led to a significantly longer response time than that of dental vs. retroflex sibilants (DR).

For the significant effect of Onset, the differences among the three manners of articulation were further checked using Fricative (Fr) and Unaspirated affricate (Ua) alternatively as the baseline. The results are reported in Table 16 in terms of t values (p values). It turned out that Fricative (Fr) and Unaspirated affricate (Ua) both had a significantly longer response time than Aspirated affricate (Aa), and the difference between Fricative (Fr) and Unaspirated affricate (Ua) was also marginally significant (p = .051).

Table 16 Dental-palatal vs. dental-retroflex: Differences among onset pairs in *t* value (*p* value)

	Unaspirated affricate (Ua)	Aspirated affricate (Aa)
Fricative (Fr)	–0.66 (.051)	–3.99 (<**.001*****)
Unaspirated affricate (Ua)		–3.17 (**.002****)

P values appear in parentheses and boldface marks those that reached significance (.05).

The difference in manner of articulation again is not meaningful as onsets in different manners of articulation have intrinsic difference in duration.

B. Accuracy

For dental vs. palatal sibilants and dental vs. retroflex sibilants in the [_i] context, Table 17 summarizes the number of discrimination errors, i.e., when a listener incorrectly judged a phonetically different pair as being the same.

Table 17 Number of discrimination errors: Dental-palatal vs. Dental-retroflex in the [_i] context

Dental vs. palatal		Dental vs. retroflex	
si-ɕi	12	si-ʂi	7
tsi-tɕi	35	tsi-tʂi	21
tsʰi-tɕʰi	25	tsʰi-tʂʰi	6

To examine whether the difference shown in Table 17 was statistically significant or not, the accuracy data of the 29 listeners were analyzed in Mixed-effects Logistic Regression using the *glmer* function in the r package lme4 (Bates, Maechler, Bolker and Walker 2015a, b). To be more specific, the data set included the accuracy data of the six CV pairs in Table 17, i.e., those whose onsets were the dental-palatal contrast (DP) and the dental-retroflex contrast (DR) and whose vowel was [_i]. For this analysis, the dependent variable, the predicting variables and their baselines, and the random variables are listed in (26).

The fixed effects in the final model are presented in Table 18. This model was obtained by comparing a null model with Listener and Listener:CVPair as the random factors with two superset models with Place and Manner as the predicting variable respectively. The addition of Place and Manner both significantly improved the model (Place: X^2 = 13.929, df= 1, p < .001; Manner: X^2 = 21.807, df= 2, p < .001). Then a model with Place and Manner as the predicting variables and Listener and Listener:CVPair as the random factors was compared with a superset

(26) Variables in the analysis of accuracy: Dental-palatal vs. Dental-retroflex in the [_i] context

Variable names	*Notes and abbreviation*
Dependent variable	
Accuracy	0 = Incorrect, 1 = Correct
Predicting variables	
Place	Dental-Palatal (**DP**), Dental-Retroflex (**DR**) (Baseline = Dental-Palatal (DP))
Manner	Fricative (***Fr***), unaspirated affricate (***Ua***), aspirated affricate (***Aa***) (Baseline = fricative(Fr))
Random variables	
Listener	29 listeners
Listener:CVPair	3 pairs ([si-ɕi] [tsi-tɕi] [tsʰi-tɕʰi]) by each listener

model with the interaction between Place and Manner. The addition of the Place*Manner interaction did not significantly improve the model. Therefore, the final model included only Place and Manner as two predicting variables and Listener and Listener:CVPair as the random factors.

Table 18 Fixed effects in the mixed-effect logistic regression: Place difference in the [_i] context.

	Estimate	*Std.Error*	*z value*	*Pr(>|z|)*
(Intercept)	2.5722	0.2767	9.296	< .001***
Place(DR)	0.7521	0.2236	3.897	< .001***
Manner(Ua)	–1.3092	0.2869	–4.563	< .001***
Manner(Aa)	–0.6004	0.3096	–1.939	.053.

Model: Accuracy ~ Place + Manner + (1|Listener) + (1|Listener:CVPair)
Baseline: Place = Dental-Palatal (DP); Manner = Fricative(Fr)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

As shown in the the fixed effect of Place in Table X7, the accuracy of the dental vs. retroflex contrast (DR) is higher than that of the baseline, i.e., the dental vs. palatal contrast (DP), and the difference reached the significance level. Therefore, for the effect of Place, the contrast of dental vs. palatal sibilants (DP) was more likely to introduce discrimination errors than the contrast of dental vs. retroflex sibilants (DR).

For the significant effect of Manner, the differences among Fricative (Fr), Unaspirated affricate (Ua), and Aspirated affricate (Aa) were further checked using Fricative (Fr) and Unaspirated affricate (Ua) alternatively as the baseline. The results are reported in Table 19 in terms of z values (p values). It turned out that Fricative (Fr) and Aspirated affricate(Aa) both had a significantly higher accuracy (i.e., less discrimination errors) than Unaspirated affricate(Ua), while the difference between Fricative (Fr) and Unaspirated affricate(Ua) is marginally significant.

Table 19 Dental-palatal vs. dental-retroflex: Differences among manners in *z* value (*p* value)

	Unaspirated affricate (Ua)	Aspirated affricate (Aa)
Fricative (Fr)	–4.535 (<**.001*****)	–1.941 (.052)
Unaspirated affricate (Ua)		2.874 (**.004****)

P values appear in parentheses and boldface marks those that reached significance (.05).

In general, the results of accuracy were consistent with the results of response time in that, for sibilant place contrasts in the [_i] context, the dental vs. palatal contrasts had a longer response time and a higher likelihood of discrimination errors than the dental vs. retroflex contrasts.

3.3.3.3 Perceptual distinction in different manners of articulations

For the dental vs. palatal sibilant contrasts in the [_i] context, it was predicted that the CV pairs with affricate onsets should introduce a lower accuracy in discrimination, i.e., more phonetically different pairs being judged as ‘same’ by the listeners. The data relevant to this prediction was thus the listeners' responses to the 3 CV pairs [si-ɕi] [tsi-tɕi] and [tsʰi-tɕʰi]. The total number of discrimination errors over the 29 listeners for these 3 pairs is plotted in Figure 5.

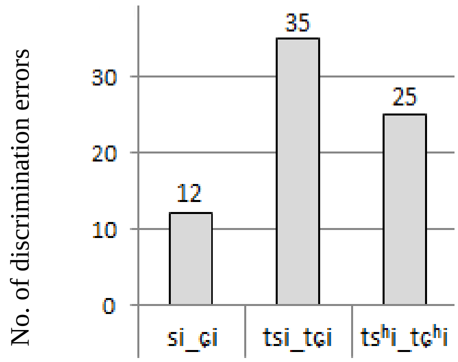


Figure 5 Number of discrimination errors of dental vs. palatal sibilants in the [_i] context

The 29 listeners' data were analyzed in Mixed-effects Logistic Regression using the *glmer* function in the r package lme4 (Bates, Maechler, Bolker and Walker 2015a, b). For this analysis, the dependent variable, the predicting variable and its baseline, and random variables are listed in (27).

(27) Variables in the analysis of manner difference: Dental vs. palatal contrasts in the [_i] context

Variable names	*Notes and abbreviation*
Dependent variable	
Accuracy	0 = Incorrect, 1 = Correct
Predicting variables	
Onset	[s-ɕ], [ts-tɕ], [tsʰ-tɕʰ] (Baseline = [si-ɕi])
Random variables	
Listener	29 listeners
Listener:CVPair	3 pairs ([si-ɕi] [tsi-tɕi] [tsʰi-tɕʰi]) by each listener

The fixed effects in the final model are presented in Table 20. This model was obtained by comparing a null model with Listener and Listener:CVPair as the random factors with a superset model with Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 13.15, *df*= 2, p = .001).

Table 20 Fixed effects in the mixed-effect logistic regression: Manner difference in the [_i] context.

	Estimate	*Std.Error*	*z value*	*Pr(>|z|)*
(Intercept)	2.5692	0.3164	8.120	< .001***
Onset(ts-tɕ)	–1.2481	0.3576	–3.490	< .001***
Onset(ts^h-$tɕ^h$)	–0.8363	0.3721	–2.248	.025*

Model: Accuracy ~ Onset + (1|Listener) + (1|Listener:CVPair)
Baseline: Onset = s-ɕ
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among the three pairs [si-ɕi], [tsi-tɕi], and [ts^hi-$tɕ^h$i] were further checked using [si-ɕi] and [tsi-tɕi] alternatively as the baseline. The results are reported in Table 21 in terms of z values (p values). It turned out that there was no significant difference between [tsi-tɕi] and [ts^hi-$tɕ^h$i] in terms of accuracy.

Table 21 Dental vs. palatal sibilants in the [_i] context: Difference in accuracy in *z* value (*p* value)

	[tsi-tɕi]	[ts^hi-$tɕ^h$i]
[si-ɕi]	–3.490 (< .001***)	–2.248 (**.025***)
[tsi-tɕi]		1.411(.158)

P values appear in parentheses and boldface marks those that reached significance level (.05).

3.3.3.4 Summary of the results

In this section, three hypotheses were put forward on the perceptual distinctiveness of sibilant contrasts based on the cross-linguistic typology and the historical development of Mandarin sbilants. The three hypotheses are generally confirmed by the results from the speeded AX discrimination experiment.

First, for the effects of vowel context on sibilant distinctiveness, it was hypothesized that dental vs. palatal sibilants would be perceptually less distinct in the [_i] context than in other vowel contexts (*Hypothesis I* in (15)). Based on this hypothesis, it was predicted that in a speeded AX discrimination, for dental vs. palatal sibilants, the vowel [_i] should introduce longer response times and more discrimination errors than in the vowel [_a] and the homorganic context, e.g., [si-ɕi] > [sa-ɕa], [sɹ̩-ɕi] as in (20). As can be seen in Table 8, for dental vs. palatal contrasts, [_i] generally led to longer response times than the other two vowel contexts; moreover, the results in accuracy also showed that the [_i] context was more likely to induce discrimination errors than other vowel contexts. This result thus supports the contention that the [_i] context reduces the perceptual distinction of the dental vs. palatal sibilants, as compared to the [_a] context or the homorganic vowel context. This is also consistent with the results of Li and Zhang (2017), who tested the vowel effects on the perceptual distinctiveness of dental vs. palatal sibilants with English and Chinese listeners.

Second, for the perceptual distinction in place contrasts, it was hypothesized that, in the [_i] context, the place contrasts of dental vs. retroflex sibilants would be more distinct than the place contrasts of dental vs. palatal sibilants (*Hypothesis II* in (16)). Based on this hypothesis, it was predicted that, for the stimulus CV pairs in the [_i] context, the dental vs. palatal contrasts would introduce longer response times and more discrimination errors than the dental vs. retroflex sibilants, e.g., si-ɕi > si-ʂi as in (21). This hypothesis was supported by the results from the speeded AX-discrimination experiment: In the [_i] context, the contrasts of dental vs. palatal sibilants led to longer response times than the contrasts of dental vs. retroflex sibilants; in addition, the results of accuracy also showed that, in the [_i] context, the dental vs. palatal contrasts generally induced more discrimination errors than the dental vs. retroflex contrasts. Thus, in the [_i] context, the place contrasts of dental vs. retroflex sibilants are perceptually more distinct than the place contrasts of dental vs. palatal sibilants.

Third, for the perceptual distinction between manners of articulation, it was hypothesized that, for dental vs. palatal sibilants in the [_i] context, the place contrast of affricates would be perceptually less distinct than the place contrast of fricatives (*Hypothesis III* in (17)). Based on this hypothesis, it was predicted that in the speeded AX-discrimination, the stimulus pairs [tsi-tɕi] and [tsʰi-tɕʰi] would be significantly more likely to introduce discrimination errors (i.e., a phonetically different pair judged as the same) than the stimulus pair [si-ɕi]. This hypothesis was also supported by the results from the speeded AX-discrimination. More specifically, the stimulus pairs [tsi-tɕi] and [tsʰi-tɕʰi] introduced a significantly lower accuracy (i.e., more discrimination errors) than [si-ɕi]. Thus, [tsi-tɕi] and [tsʰi-tɕʰi] are perceptually less distinct than [si-ɕi].

3.4 Vowel effect and the phonetic basis of the results

3.4.1 Dental vs. retroflex sibilants: The effect of vowel contexts

For dental vs. palatal sibilants in particular, the speeded AX discrimination showed that the responses times turned out to be longer in the [_i] context compared with other vowel contexts. What remains unclear, however, is whether the same vowel effect holds for other sibilant place contrasts. For example, are dental vs. retroflex sibilants perceptually less distinct in the [_i] context than in other vowel contexts? As (19) illustrates, the stimuli used for the perceptual experiment in fact included contrastive dental vs. retroflex sibilants embedded in the [_i] context, the [_a] context, and the homorganic context (where dental and palatal sibilants are followed by homorganic syllabic approximants, also known as apical vowels). To examine if the same vowel effect for dental vs. palatal sibilants holds for the dental vs. retroflex sibilants, the response time data of the dental vs. retroflex sibilants were examined. Below, the mean Log Response Time for each stimulus pair is plotted in Figure 6.

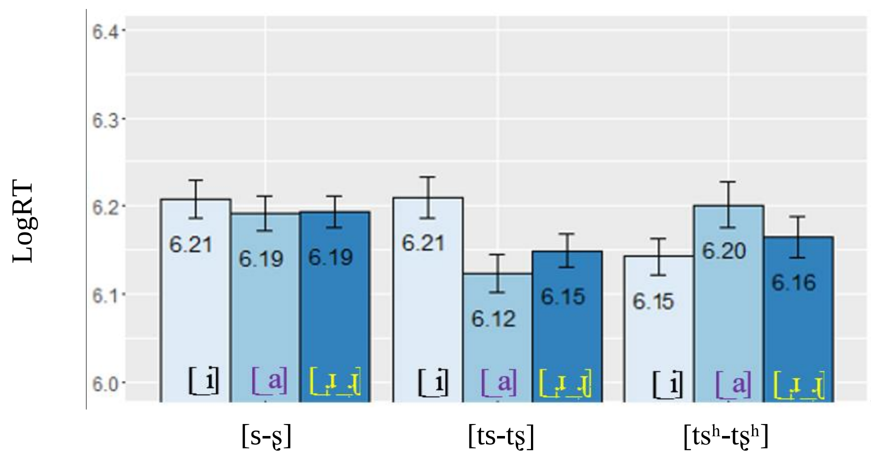


Figure 6 Mean LogRTs of CV Pairs: Dentals vs. retroflexes in different vowel contexts.
(For LogRTs, 6.0 = 403ms, 6.1 = 446ms, 6.2 = 493ms, 6.3 = 545 ms, 6.4 = 602ms)

The LogRT data were analyzed in Linear Mixed Effects Models using the *lmer* function in the R package lme4 (Bates, Maechler, Bolker and Walker 2015a, b) and the p-values were determined by the R package lmerTest (Kuznetsova, Brockhoff, and Christensen 2015). For this analysis, the dependent variable, the predicting variables and their baselines, and the random variables are listed in (28).

(28) Variables in the analysis of vowel effects on sibilant distinctiveness: Dentals vs. Retroflexes

Variable names	*Notes and abbreviation*
Dependent variable	
LogRT	The log-transformed response time
Predicting variables	
Vowel	**[_i]**, **[_a]**, and homorganic **[_ɹ̩_ɻ̩]** (ɹ̩ after dentals, ɻ̩ after retroflexes) (Baseline = [_i])
Onset	Fricative (**s-ʂ**), unaspirated affricate (**ts-tʂ**), aspirated affricate (**tsʰ-tʂʰ**) (Baseline = Fricative (**s-ʂ**))
Random variables	
Listener	29 listeners
Listener:CVPair	9 pairs (3 vowels × 3 onset manners) by each listener

The fixed effects in the final model are presented in Table 22. This model was obtained via two steps. First, a null model with Listener and Listener:CVPair as the random factors was compared separately with two models adding Vowel or Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 9.47, *df*= 2, p = .009) but not the addition of Vowel. Second, a model with Vowel and Onset as the predicting variables and Listener and Listener:CVPair as the random factors was compared with a superset model with the Vowel*Onset interaction. The addition of the interaction led to a significant improvement of the model (X^2 = 12.2, *df* = 4, p = .016). Therefore, the final model included Vowel, Onset, and their interaction as the predicting variables and Listener and Listener:CVPair as the random factors.

Table 22 Fixed effects in the mixed-effect linear regression for LogRT: Dentals vs. retroflexes.

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	6.210	0.035	42	175.18	< .001***
Vowel(_a)	–0.021	0.023	1400	–0.94	.348
Vowel(_*h*)	–0.015	0.023	1400	–0.68	.499
Onset(ts-tʂ)	–0.006	0.024	1400	–0.24	.808
Onset(tsʰ-tʂʰ)	–0.060	0.023	1400	–2.64	.008**
Vowel(_a):Onset(ts-tʂ)	–0.055	0.033	1400	–1.68	.094
Vowel(_ɹ̩_ɻ̩):Onset(ts-tʂ)	–0.042	0.033	1400	–1.28	.200
Vowel(_a):Onset(tsʰ-tʂʰ)	0.067	0.033	1400	2.03	.043*
Vowel(_ɹ̩_ɻ̩):Onset(tsʰ-tʂʰ)	0.027	0.032	1400	0.82	.411

Model: LogRT ~ Vowel * Onset + (1|Listener) + (1|Listener:CVPair)
Baselines: Vowel = [_i]; Onset = Fricative (**s-ʂ**);
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant Vowel*Onset interaction, sperate analyses were given to the effect of Vowel on the subset of the data where the onsets were [s-ʂ], [ts-tʂ], and [tsʰ-tʂʰ]. For the onset pair [s-ʂ], for example, the data set included the response times for the three pairs [si-ʂi], [sa-ʂa], and [sɹ̩-ʂɻ̩]. This dataset was submitted to Linear Mixed Effects Models using the *lmer* function in the R package lme4 (Bates, Maechler, Bolker and Walker 2015a, b). More speficially, a null model with Listener and Listener:CVPair as the random factors was compared with a superset model with Vowel ([_i], [_a], and [_ɹ̩_ɻ̩], with [_i] as the baseline) as the predicting variable. It turned out that, for the onset pairs [s-ʂ] and [tsʰ-tʂʰ], the addition of Vowel did not significantly improve the model; for the onset pair [ts-tʂ], on the other hand, the addition of Vowel significantly improved the model (X^2 = 11.6, *df*= 2, p = .003).

For the dataset with [ts-tʂ] as the onset, the fixed effects in the final model are presented in Table 23.

Table 23 Fixed effects in the mixed-effect logistic regression: Vowel effect for the [ts-tʂ] contrasts.

	Estimate	*Std.Error*	*df*	*t value*	*Pr(>|t|)*
(Intercept)	6.2058	0.0384	35	161.77	< .001***
Vowel([_a])	–0.0817	0.0233	406	–3.50	< .001***
Vowel([_ɹ̩_ɻ̩])	–0.0544	0.0229	406	–2.37	.018*

Model: LogRT~ Vowel + (1|Listener) + (1|Listener:CVPair)
Baseline: Vowel = [_i]
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Vowel of the [ts-tʂ] contrast, the differences among the three vowel contexts (i.e., [_i], [_a], and [_ɹ̩_ɻ̩]) were further checked using [_i] and [_a] alternatively as the baseline. The results are reported in Table 24 in terms of z values (p values).

Table 24 Vowel effect for the [ts-tʂ] contrast: Difference among the vowels in *z* value (*p* value)

	[_a]	[_ɹ̩_ɻ̩]
[_i]	–3.50 (< .001***)	–2.37 (**.018***)
[_a]		1.22(.225)

P values appear in parentheses and boldface marks those that reached significance (.05).

The results in Table 23 and Table 24 showed that, for the [ts-tʂ] contrast, the [_i] context led to a significantly longer response time than the [_a] context and the 'homorganic' context [_ɹ̩_ɻ̩], while the latter two vowel contexts did not significantly differ from each other.

In general, the results from the perceptual experiment showed that, for the dental vs. retroflex sibilants, the vowel effect (e.g., [_i] context introducing longer response times than other vowels) holds only for the unaspirated affricates, i.e., [tsi-tʂi] > [tsa-tʂa], [tsɹ̩-tʂɻ̩], but not for the contrasts [s-ʂ] and [ts^h-$tʂ^h$]. Considering the results reported in 3.3.3.1, the vowel effect was observed for the dental vs. palatal sibilants across three manners of articulation. Simply put, the results from the perceptual experiment showed that the vowel effect holds for sibilant place contrasts between some places of articulation, but not others.

3.4.2 Dental vs. palatal sibilants: The phonetic basis of the vowel effect

The results from the speeded-AX discrimination showed that, for dental vs. palatal sibilants, the [_i] context generally introduced a longer response time across all three manners of articulation. Assuming that a longer response time indicates reduced distinctiveness, the observed results meant that the [_i] context reduced the perceptual distinctiveness of the dental vs. palatal sibilants compared with other vowel contexts. The perceptual basis for the vowel effect (i.e., reduced distinctiveness in the [_i] context) could be attributed to three possible factors: the acoustic properties of the onset sibilants, the steady vowels, and the consonant-vowel transitions. Below, these three possibilities are examined based on the acoustic properties of the manipulated stimulus syllables used in the perceptual experiment.[18]

First, in terms of the acoustic properties of the sibilants, it could be that the reduced perceptual distinctiveness of dental vs. palatal sibilants was caused by the smaller acoustic difference in the [_i] context than in the other vowel context, as assumed in Lee and Li (2003) and Lee-Kim (2014a). In Figure 7 below, the acoustic differences of dental vs. palatal sibilants in [_i] and [_a] contexts are plotted, based on the acoustic measurements in Table 7.[19] Note that, in Figure 7, the homorganic context (e.g., [sɹ̩-ɕi]) was not included because the two syllables differ in both onsets and vowels and thus the comparison of sibilants would not capture the difference in the relevant pairs.

---

[18] The discussion of the phonetic basis of the vowel effect is limited to the three measurements as detailed below. It should be admitted that there remains the possibility that the relative changes in one or more acoustic properties of the consonants and vowels, i.e., some integrated property (Blumsterin and Stevens 1979), may have contributed to the observed vowel effect. This possibility is to be investigated in further studies.

[19] For the acoustic properties in Table 7, details on the measurements were introduced in section 3.3.1.2.

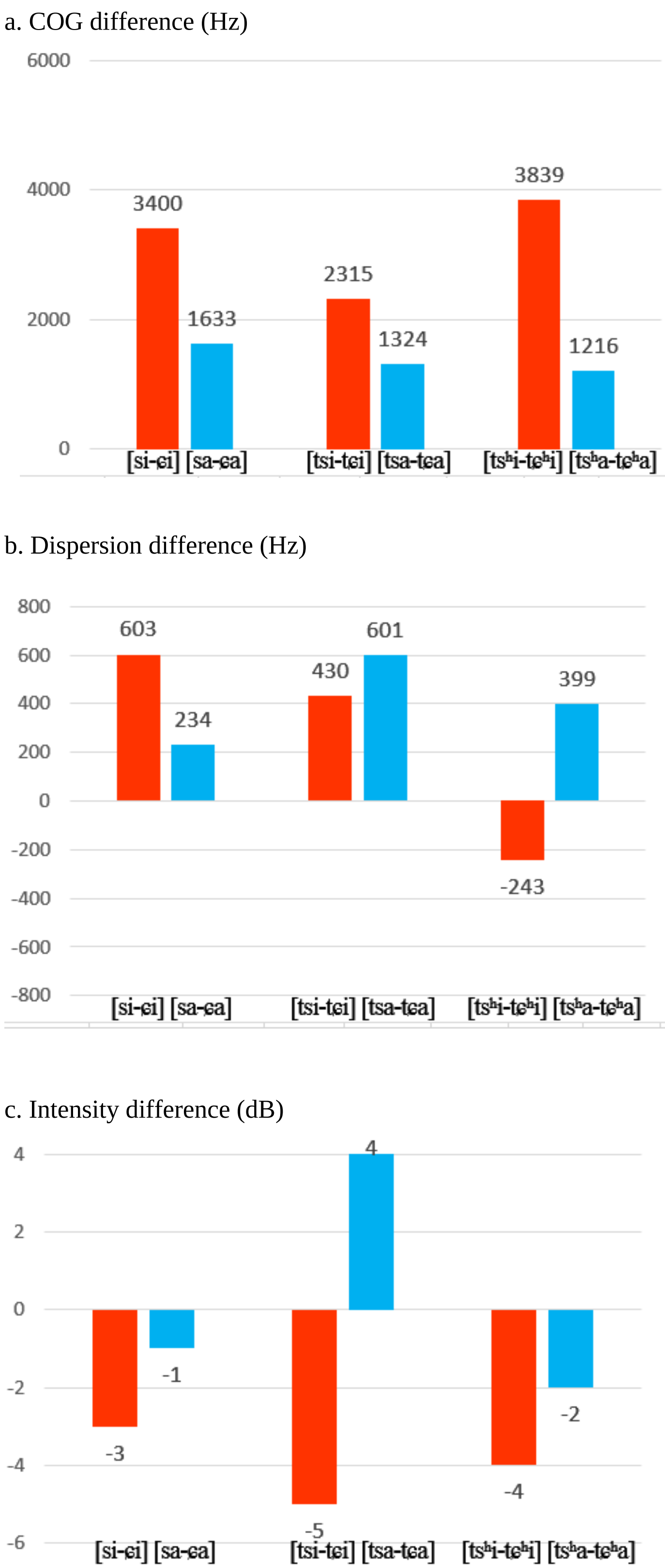


Figure 7 Acoustic difference between dental vs. palatal sibilants in the [_i] and [_a] contexts.

As in Figure 7, the COG difference between the dental and palatal sibilants was in fact larger in the [_i] context than in the [_a] context, e.g., the COG difference of the sibilant onsets was 3400Hz in the [si-ɕi] pair and 1633Hz in the [sa-ɕa] pair. The same is true for the intensity difference. Moreover, there is no consistent pattern between the [_i] and [_a] contexts in terms of the dispersion difference of dental vs. palatal sibilants. These observations contradict the contention that the observed vowel effects (e.g., place distinction is less distinct in the [_i] context than in the [_a] context) are rooted in the smaller difference of the onset sibilants in the [_i] context.

Second, it could be that the observed vowel effects are rooted in the smaller difference between the two [_i]s (e.g., in [si-ɕi]) than the two [_a]s (e.g., in [sa-ɕa]). To test this possibility, the vowel formant differences based on the measurements of F1, F2, and F3 in Table 7 are given in Table 25. In Table 25, the formant values are measured at the middle of the steady vowel portion, and a ΔF value indicates the difference between the two formant values measured at the middle points of the steady vowel portions. For example, for the [si-ɕi] pair, 73 (Hz) is the difference between the F3 of [_i] in [si] and the F3 of [_i] in [ɕi] (by subtracting the palatal value from the dental value).

Previous studies have shown that, for isolated vowel formants, the Just Noticeable Difference (JND) was generally 3%–5% of the reference formant frequency (Flanagan 1955, Kakusho and Karo 1968; Nord and Sventelius 1979), though a JND as low as 1.5% has also been reported (Kewley-Port and Watson 1994). For vowels in consonantal contexts, Mermelstein

Table 25 Acoustic difference in the vowel formants of the stimulus pairs: Dental vs. palatal (Hz).

Vowel		[s]	[ɕ]	*ΔF*	*(%)*	[ts]	[tɕ]	*ΔF*	*(%)*	[ts^h]	[tɕ^h]	*ΔF*	*(%)*
[_i]	*F3*	2904	2977	–73	(2.5%)	2868	2941	–73	(2.5%)	3154	3048	106	(3.5%)
	F2	2570	2590	–20	(0.8%)	2567	2607	–40	(1.6%)	2566	2557	9	(0.4%)
	F1	321	318	3	(0.9%)	325	317	8	(2.5%)	310	315	–5	(1.6%)
[_a]	*F3*	2639	2569	70	(2.7%)	2729	2593	136	(5.2%)	2610	2574	36	(1.4%)
	F2	1392	1391	1	(0.1%)	1343	1403	–60	(4.5%)	1313	1360	–47	(3.6%)
	F1	928	933	–5	(0.5%)	968	975	–7	(0.7%)	910	936	–26	(2.9%)

Note: A *ΔF* indicates the difference between the two vowel formant values on its left. The percentage in parentheses indicates the value of ΔF divided by the *lower* formant value (underlined) on the left.

(1978) reported mean difference limens of 60 Hz for F1 and 176 Hz for F2. To examine if the formant difference is perceivable, a percentage value was calculated for each ΔF value in Table 25 by dividing the ΔF value by the lower of the two formant values on its left. For example, for the pair [si-ɕi], ΔF3 = |–73| and the percentage value is |–73| ÷ 2904 = 2.5%. As shown in Table 25, most of the percentage values were below 4% and smaller than the limens reported in Mermelstein (1978). Moreover, all the vowels appear in CV syllables, where the JND of the vowel formants should be even larger than the JND of isolated vowel formants. It is thus unlikely that the formant differnces have led to salient perceptual differences. Put simply, in each stimulus pair like [si-ɕi], the two vowels are close to each other, and therefore, the observed vowel effects are unlikely to be rooted in the steady vowel formants.

Third, it could be that the transitional properties of dental and palatal sibilants were more similar in the [_i] context than in the [_a] context (Lee-Kim 2014a). Formant transitions have been shown to be important in the place identification of consonants (Delattre et al. 1955; Whalen 1981, 1991; Nowak 2006; Babel and McGuire 2013) and Lee-Kim (2014a) argued that vowel effects on consonant distinctiveness can be reduced to the relative magnitude of formant transitions specific to each vowel. For the perception of palatal fricative, for example, a low/back vowel may provide a greater palatal transition and thus a more robust perceptual cue while a high/front vowel may provide a smaller palatal transition and thus a less robust perceptual cue (Lee-Kim 2014a).

To investigate this possibility, comparisons were made first of the transitonal properties of the dental vs. palatal sibilants in each stimulus pair, and then of the transitional properties between the dental-palatal differences in [_i] vs. [_a] context. Note that, in the stimulus manipulation, the duration of the formant transition in each stimulus syllable was controlled to be 50 ms. Therefore, in this study, the comparison of CV transitions within a pair can be reduced to the comparison between the F2 onset values of the two syllables or that between the F2 offset values of the two syllables.[20] In Table 26 below, the $F2_{onset}$ and $F2_{offset}$ indicate the formant values at the beginning and end of the consonant-vowel transition, i.e., the vocalic portion before the steady formant structures of a following vowel. $\Delta F2_{onset}$ is the F2 difference between the dental and palatal sibilants at the beginning of the CV transitions, where a larger value indicates a larger onset F2 difference. The same holds for $\Delta F2_{offset}$.

[20] To evaluate the transitional property within a CV sequence, an alternative measurement would be the difference between the F2onset and F2offset, i.e., $F2_{onset} – F2_{offset}$. In the current study, the focus is on the difference between the two syllables in a CV pair. Therefore, the $F2_{onset}$ difference was adopted as a straightforward way to evaluate the transitional difference of the two syllables within a CV pair.

Table 26 Acoustic difference of formant transitions in the stimulus pairs: Dental vs. palatal (Hz).

Vowel		[s]	[ɕ]	$\Delta F2_{ons,\,off}$	[ts]	[tɕ]	$\Delta F2_{ons,\,off}$	[tsʰ]	[tɕʰ]	$\Delta F2_{ons,\,off}$
[_i]	$F2_{onset}$	1977	2452	–475	1922	2461	–539	2405	2412	–7
	$F2_{offset}$	2543	2540	3	2481	2506	–25	2553	2527	26
[_a]	$F2_{onset}$	1289	1894	–605	1310	1949	–639	1287	1820	–533
	$F2_{offset}$	1415	1484	–69	1355	1468	–113	1325	1440	–115

Note: $F2_{onset}$ indicates the value of formant at the beginning of the vocalic transition;
$\Delta F2_{onset}$ indicates the formant onset difference between the dental and palatal sibilants in a vowel context.
$F2_{offset}$ indicates the value of formant at the end of the vocalic transition;
$\Delta F2_{offset}$ indicates the formant offset difference between the dental and palatal sibilants in a vowel context.

The values of $\Delta F2_{offset}$ are generally small, considering the JND of F2 value (176Hz for vowels in consonantal contexts) as reported in Mermelstein (1978). Therefore, the transitional difference between the two syllables in a stimulus pair is mostly determined by $\Delta F2_{onset}$. As shown in Table 26, $\Delta F2_{onset}$ for the same sibilant pair is generally smaller in the [_i] context than in the [_a] context. In other words, the transitional difference of the dental vs. palatal sibilants is smaller in the [_i] context, which is consistent with the observation that the [_i] context induced less perceptual distinctiveness between the dental vs. palatal sibilants.[21]

To summarize, based on the measurements of onset COG, dispersion, and intensity, the observed vowel effect is unlikely to be rooted in the acoustic differences in the sibilants, nor is it likely to come from the acoustic properties of the steady vowel formants. On the other hand, the vowel effect is consistent with the observation that the formant transition properties of the dentals and palatals are acoustically more similar in the [_i] context than in other vowel contexts, as schematized in Figure 8.

[21] No statistical analysis was conducted because the measurements and calculations in Table 20 were based on the same 12 stimulus syllables used in the perceptual experiment, i.e., [si-ɕi] [sa-ɕa] [tsi-tɕi] [tsa-tɕa] [tsʰi-tɕʰi] [tsʰa-tɕʰa].

a.[si-ɕi] b. [sa-ɕa]

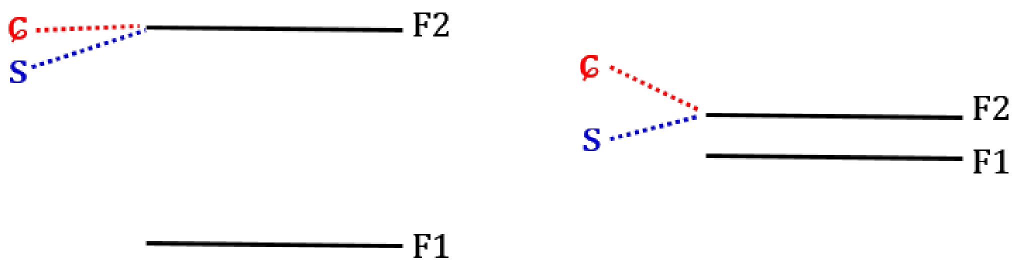


Figure 8 Schematic illustration of the vowel effect on the consonant place distinction: [_i] vs. [_a].

For the stimulus syllables used in the experiment, the COG and intensity differences between dental and palatal sibilants are larger in the [_i] context than in the [_a] context, while the transitional difference is smaller in the [_i] context than in the [_a] context. Thus, the results suggest that, in the listeners' discrimination of the CV pairs, the transitional difference has generally overidden the COG/intensity difference in the sibilants onsets. This conclusion has two caveats: First, the sibilant measurements were restricted to COG, dispersion, and intensity, and it is possible that the onset pairs are more similar (or more distinct) in other acoustic aspects; second, these conclusions were drawn from the specific stimuli used in the experiment, and it is possible that the realization of the dental-i sequence in a different language could be different (e.g., with more palatalization on the dental sibilants).

3.4.3 Different sibilant place contrasts: The acoustic basis

The historical development of Mandarin sibilants (2.3.2) showed that around the 11th century, contrastive dentals vs. palatals in the [_i] context developed into dental vs. retroflex sibilants. The change occurred when the palatal sibilants merged into the retroflex sibilants (which did not combine with the rime [_i] in previous historical stages). This is similar to a sound change in

Polish, by which the [ɕ-ʃʲ] contrast shifted to the [ɕ-ʂ] contrast (Padgett and Źygis 2007; Źygis and Padgett 2010) and [ɕ-ʂ] is shown to be perceptually more distinct than [ɕ-ʃʲ] (Źygis and Padgett 2010). Following Źygis and Padgett (2010), it was hypothesized in this study that in the [_i] context, dental vs. retroflex sibilants would be perceptually more distinct than the dental vs. palatal sibilants. This hypothesis is supported because as reported in 3.3.3.2, in the [_i] context, the dental vs. retroflex sibilants generally introduce shorter response times (i.e., more perceptually distinct) than their dental vs. palatal counterparts.

To examine the basis of the place difference, the acoustic measurements of the stimulus pairs [si-ɕi] [tsi-tɕi] [tsʰi-tɕʰi] from Table 7 are compared with those of [si-ʂi] [tsi-tʂi] [tsʰi-tʂʰi], respectively. The measurements of sibilant acoustic difference are plotted in Figure 9; the formant measurements of the steady vowels (at the middle points) are given in Table 27; the measurements of transitional properties in terms of $F2_{onset}$ and $F2_{offset}$ are given in Table 28.

Figure 9 shows that in the [_i] context, the dental vs. palatal sibilants have a smaller COG difference and a larger dispersion difference than the dental vs. retroflex sibilants. There is no consistent pattern in terms of intensity. Therefore, out of these three acoustic measurements, only the COG difference is consistent with the directionality of the place difference, i.e., dental vs. palatal sibilants being less distinct than the dental vs. retroflex sibilants.

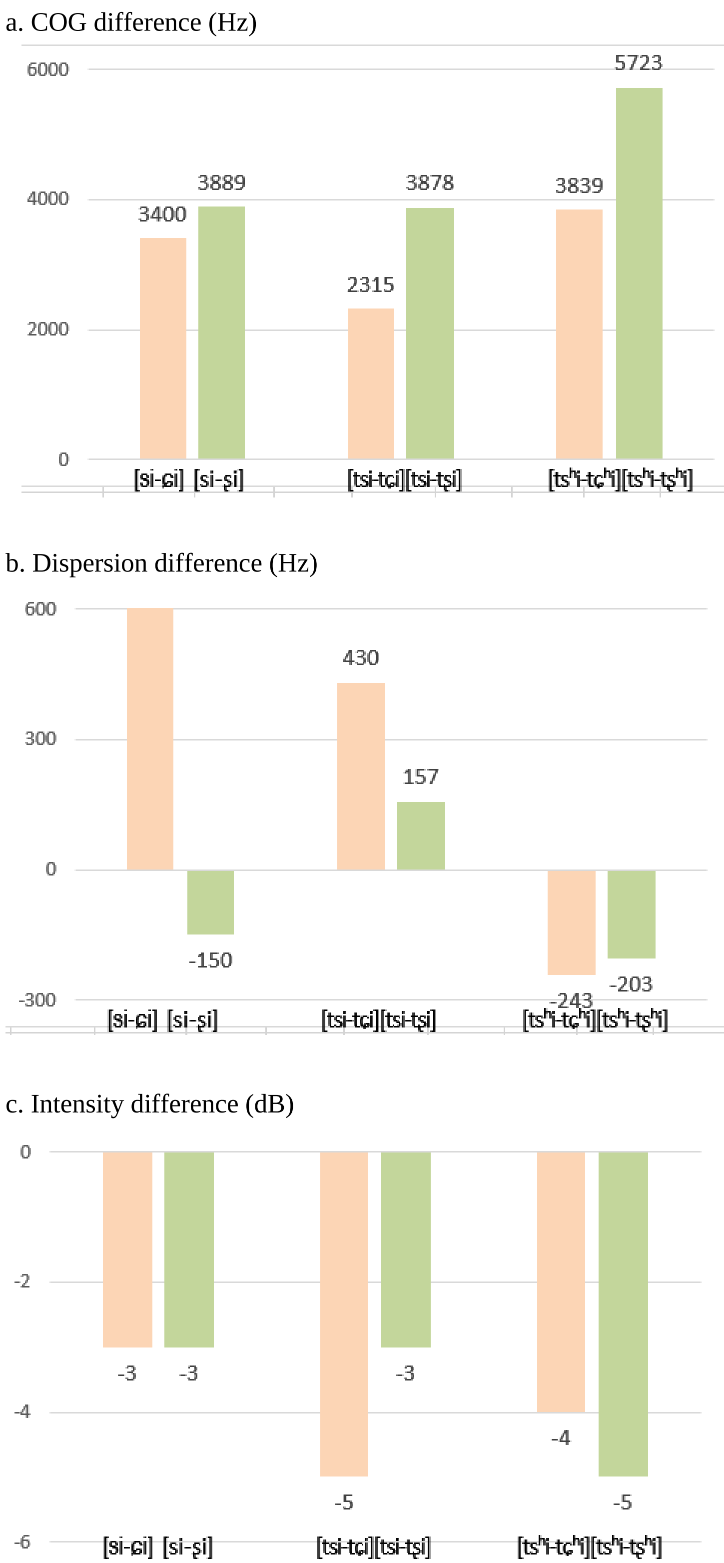


Figure 9 Acoustic difference of the sibilants in [si-ɕi] [si-ʂi] [tsi-tɕi] [tsi-tʂi] [ts^{h}i-tɕ^{h}i] [ts^{h}i-tʂ^{h}i].

Table 27 Acoustic difference in the vowel formants of the stimulus pairs (Hz).

	s[i]	ɕ[i]	*ΔF*	*(%)*	ts[i]	tɕ[i]	*ΔF*	*(%)*	tsʰ[i]	tɕʰ[i]	*ΔF*	*(%)*
F3	2904	2977	–73	(2.5%)	2868	2941	–73	(2.5%)	3154	3048	106	(3.5%)
F2	2570	2590	–20	(0.8%)	2567	2607	–40	(1.6%)	2566	2557	9	(0.4%)
F1	321	318	3	(0.9%)	325	317	8	(2.5%)	310	315	–5	(1.6%)
	s[i]	ʂ[i]	*ΔF*	*(%)*	ts[i]	tʂ[i]	*ΔF*	*(%)*	tsʰ[i]	tʂʰ[i]	*ΔF*	*(%)*
F3	2904	3025	–121	(4.2%)	2868	2847	21	(0.7%)	3154	3129	25	(0.8%)
F2	2570	2547	23	(0.9%)	2567	2562	5	(0.2%)	2566	2562	4	(0.2%)
F1	321	315	6	(1.9%)	325	317	8	(2.5%)	310	316	–6	(1.9%)

Note: A *ΔF* indicates the difference between the two vowel formant values on its left. The percentage in parentheses indicates the value of ΔF divided by the lower formant value (underlined) on the left.

Table 27 shows the formant values of the vowel [i]s in each stimulus pair, measured at the mid-point of the steady formants, and the formant value difference within each pair (ΔF). For example, for the pair [si-ɕi], the F3 of the vowel in [si] is 2904 Hz and that in [ɕi] is 1977 Hz; therefore the F3 difference (ΔF3) between the two vowel [i]s in the [si-ɕi] pair is 73 Hz. To evaluate whether the formant difference (ΔF) will lead to a perceptual difference, a percentage value (in parentheses) is calculated by dividing the ΔF value by the lower value (underlined) in the pair. For the ΔF3 value in the [si-ɕi] pair, for example, the ΔF3 value (73 Hz) is divided by 2904 Hz, i.e., 73/2904 = 2.5%. As shown in Table 27, across all of the CV pairs, the ΔF values are generally lower than 4% and therefore unlikely to have led to perceptual differences.

Table 28 shows the measurements of the transitional difference within each CV pair in terms of $F2_{onset}$ and $F2_{offset}$. For example, for the pair [si-ɕi], the $F2_{onset}$ of [si] is 1977 Hz and that of [ɕi] is 2452 Hz, with a difference of 475 Hz ($ΔF2_{onset}$). Comparing the [si-ɕi] pair with the [si-ʂi] pair, it can be seen that the $F2_{onset}$ difference in the [si-ɕi] pair (475 Hz) is larger than that in the [si-ʂi] pair (27 Hz), which is not consistent with the observation that the [si-ɕi] is perceptually less distinct than [si-ʂi]. The same is true for the comparison between the [tsi-tɕi] pair and the [tsi-tʂi] pair. Therefore, the observation that, in the [_i] context, the dental vs. retroflex contrast is perceptually more distinct than the dental vs. palatal contrast is unlikely to be rooted in the transitional properties of the CV pairs.

Table 28 Acoustic difference between formant transitions in the stimulus pairs (Hz).

	s[i]	ɕ[i]	$\Delta F2_{ons,\ off}$	ts[i]	tɕ[i]	$\Delta F2_{ons,\ off}$	tsʰ[i]	tɕʰ[i]	$\Delta F2_{ons,\ off}$
$F2_{onset}$	1977	2452	–475	1922	2461	–539	2405	2412	–7
$F2_{offset}$	2543	2540	3	2481	2506	–25	2553	2527	26
	s[i]	ʂ[i]	$\Delta F2_{ons,\ off}$	ts[i]	tʂ[i]	$\Delta F2_{ons,\ off}$	tsʰ[i]	tʂʰ[i]	$\Delta F2_{ons,\ off}$
$F2_{onset}$	1977	2004	–27	1922	1877	45	2405	2215	190
$F2_{offset}$	2543	2575	–32	2481	2498	–17	2553	2592	–39

Note: $F2_{onset}$ indicated the value of formant at the beginning of the vocalic transition;
$\Delta F2_{onset}$ indicates the formant onset difference between the dental and palatal sibilants in a vowel context.
$F2_{offset}$ indicates the value of formant at the end of the vocalic transition;
$\Delta F2_{offset}$ indicates the formant offset difference between the dental and palatal sibilants in a vowel context.

To summarize, from the acoustic measurements above, the larger perceptual distinctiveness in the dental vs. retroflex contrasts than the dental vs. palatal contrasts is unlikely to be rooted in the acoustic properties of the vowels or the formant transition. Instead, the difference is more likely rooted in the larger COG difference in [si-ʂi] [tsi-tʂi] [tsʰi-tʂʰi] than in [si-ɕi] [tsi-tɕi] [tsʰi-tɕʰi], respectively.[22]

[22] Note that the role of onset COG difference in the place effect here is different from its role in the vowel effect in 3.4.2. More specifically, for the perceptual distinctiveness of dental vs. palatal sibilants, it has been noted that the directionality of the vowel effect (i.e., [_i] with reduced distinctiveness) is opposite to the fact that the COG difference within a stimulus pair is in fact larger in the [_i] context than in the [_a] context. Therefore, it was suggested that for dental vs. palatal sibilants, the vowel effect is unlikely to be rooted in the COG difference of the contrastive onsets.

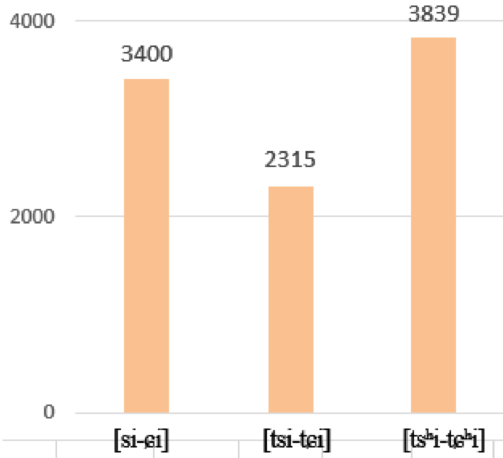
3.4.4 Differences in sibilant manners: The acoustic basis

For the contrastive dental vs. palatal sibilants in the [_i] context, the typological survey in this dissertation shows that a place contrast in affricates (e.g., [tsi-tɕi] [ts^{h}i-tɕ^{h}i]) implies the place contrast in fricatives (e.g., [si-ɕi]). Based on this typology, it was hypothesized that [tsi-tɕi] and [ts^{h}i-tɕ^{h}i] are perceptually less distinct than [si-ɕi]. This hypothesis is supported by the results from the experiments in that the [tsi-tɕi] and [ts^{h}i-tɕ^{h}i] pairs are significantly more likely to introduce discrimination errors (i.e., mroe phonetically different pairs judged as the same) than the [si-ɕi] pair.

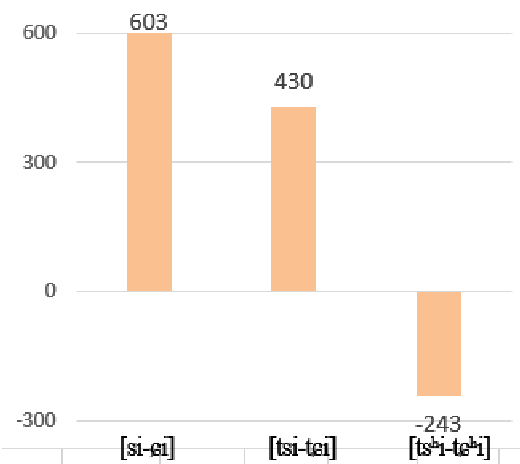
The acoustic differences in the onsets of the three sibilant pairs [si-ɕi] [tsi-tɕi] [ts^{h}i-tɕ^{h}i] are plotted in Figure 10, based on the measurements in Table 7. As discussed in 3.4.2, the steady vowels in each stimulus pair are close to each other and thus unlikely to lead to perceptual differences. The acoustic differences in the transitional properties of [si-ɕi] [tsi-tɕi] [ts^{h}i-tɕ^{h}i] from Table 26 are recalled in Table 29 below.

Compared with [si-ɕi], the [tsi-tɕi] pair has smaller COG and dispersion differences in the onsets, but a larger intensity difference in the onsets and a larger transitional difference. The fact that [tsi-tɕi] incurred more discrimination errors than [si-ɕi], then, is consistent with the smaller COG and dispersion differences of the onset sibilants in [tsi-tɕi] than in [si-ɕi].

a. ΔCOG (Hz)



b. ΔDispersion (Hz)



c. ΔIntensity (dB)

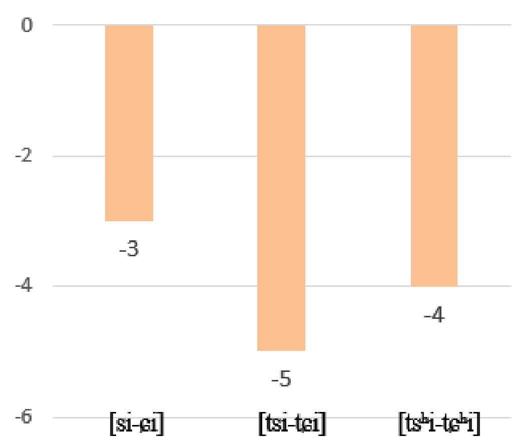


Figure 10 Acoustic difference of the sibilants in [si-ɕi] [tsi-tɕi] [ts^{h}i-tɕ^{h}i].

Table 29 Acoustic differences in formant transitions in the stimulus pairs (Hz) – Repeated from Table 26.

Vowel		[s]	[ɕ]	$\Delta F2_{ons, off}$	[ts]	[tɕ]	$\Delta F2_{ons, off}$	[tsʰ]	[tɕʰ]	$\Delta F2_{ons, off}$
[_i]	$F2_{onset}$	1977	2452	–475	1922	2461	–539	2405	2412	–7
	$F2_{offset}$	2543	2540	3	2481	2506	–25	2553	2527	26

Compared with [si-ɕi], the [tsʰi-tɕʰi] pair has a slightly larger COG and intensity difference in the onsets, but a smaller dispersion difference in the onsets and a considerably smaller transitional difference. The fact that [tsʰi-tɕʰi] incurred more discrimination errors than [si-ɕi], then, is consistent with the smaller dispersion and transitional differences in [tsʰi-tɕʰi] than in [si-ɕi]. The reason for the small $F2_{onset}$ difference [tsʰi-tɕʰi] in could be that the aspiration in the onsets of the [tsʰi-tɕʰi] pair reduces the formant transition, and the breathiness induced by the aspiration may further weaken the formant transition cue.

3.5 Discussion

The cross-linguistic typology across Chinese dialects shows that, for dental vs. palatal sibilants, there is a tendency to avoid their place contrasts in the [_i] context. The typological survey further shows that in most dialects, contrastive dental vs. palatal sibilants in the [_i] context are avoided with the introduction of apical vowels after the dental sibilants.

Following the perceptual accounts of similar typological and diachronic patterns (Stevens et al. 2004, Lee-Kim 2014a, Źygis and Padgett 2010), it is hypothesized that, in the [_i] context, sibilant place contrasts are perceptually less salient than in other vowel contexts. In a speeded AX discrimination, the hypothesis was tested by examining the response time to CV pairs whose onsets are dental vs. palatal sibilants. The predicted vowel effect, i.e., [_i] introducing longer response times and more discrimination errors, was observed, confirming the hypothesis.

Therefore, the results from the perceptual experiment show that phonetic pairs like [si-ɕi] form perceptually weak contrasts, compared with pairs like [sa-ɕa], which is consistent with the contention that the sibilant distinctiveness is reduced in the [_i] context (Lee and Li 2003; Stevens et al. 2004, Lee-Kim 2014a, among others). On other other hand, the results also show that phonetic pairs like [si-ɕi] are less distinct than pairs like [sɹ̩-ɕi], which supports the claim that the introduction of apical vowels enhances the perceptual distinction of the contrasts between /si/ and /ɕi/ (Stevens et al. 2004, Keyser and Stevens 2006, Lee-Kim 2014a, among others).

The vowel effects for dental vs. palatal sibilants turned out to be consistent with both the typological pattern and the diachronic sound change. For the typological pattern, it has been shown that contrastive dental vs. palatal sibilants tend to be avoided in the [_i] context. Significantly, this tendency holds true for dialects that have dental vs. palatal sibilants to the exclusion of retroflex sibilants (2.2). In other words, at least for dental vs. palatal sibilants, the avoidance of their place contrasts in the [_i] context should be rooted in their perceptual similarity of the two contrastive elements independently of whether retroflex sibilants exist in the sound system. A similar pattern was observed in the diachronic development of Mandarin, by which the dental vs. palatal contrasts in the [_i] context have been enhanced or avoided throughout different historical stages (2.3). As in (14), in the 7th century, there was no retroflex-[i] syllables in the sound system, and the sound change still avoided the place contrasts between dental and palatal sibilants in the [_i] context.

3.6 Summary of the perceptual study

The perceptual experiment in this chapter tested three hypotheses on the perceptual distinctiveness of sibilant contrasts, which were proposed based on the typology across Chinese dialects and the historical sound changes of Mandarin sibilants, as repeated below:

(a) *Hypothesis I* (Vowel effect): Dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts;

(b) *Hypothesis II* (Place effect): In the [_i] context, the place contrast of dental vs. retroflex sibilants is more distinct than the place contrast of dental vs. palatal sibilants;

(c) *Hypothesis III* (Manner effect): For dental vs. palatal sibilants in the [_i] context, the place contrast of affricates is less distinct than the place contrast of fricatives.

To test these three hypotheses, a speeded-AX discrimination experiment was conducted, which has been shown to be able to access psychoacoustic perception independent of the listeners' native language. For the first two hypotheses, longer response times and more discrimination errors were assumed to be an indication of less perceptual distinctiveness; for the third hypothesis, more discrimination errors were taken as the measurement to indicate less perceptual distinction. The results from the experiment generally confirmed the three hypotheses:

(a) For dental vs. palatal sibilants, the [_i] context generally introduced longer response times and more discrimination errors than other vowel contexts;

(b) For the [_i] context, the place contrast of dental vs. palatal sibilants introduced longer response times and more discrimination errors than the place contrast of dental vs. retroflex sibilants;

(c) For contrastive dental vs. palatal sibilants in the [_i] context, the place contrast in affricates introduced more discrimination errors than the place contrast in fricatives.

Additionally, a further examination of the results from the speeded AX discrimination also showed that the vowel effect (i.e., [_i] with reduced perceptual distinctiveness) was not observed for all sibilant place contrasts. For example, for the contrasts [s-ʂ] and [ts^{h}-tʂ^{h}], the [_i] context did not introduced longer response times than other vowel contexts.

For the phonetic basis of the observed results, an examination of the acoustic properties of the stimulus syllables indicated that:

(d) For dental vs. palatal sibilants, the reduced distinctiveness in the [_i] context is most likely rooted in the smaller transitional difference of the CV pairs in the [_i] context than in other vowel contexts;

(e) For the [_i] context, the larger perceptual distinctiveness of the dental-retroflex contrast than the dental-palatal contrast was most likely rooted in the larger COG difference of the onsets in [si-ʂi] [tsi-tʂi] [ts^{h}i-tʂ^{h}i] than in [si-ɕi] [tsi-tɕi] [ts^{h}i-tɕ^{h}i];

(f) For dental vs. palatal sibilants in the [_i] context, the relative fewer discrimination errors in the [si-ɕi] pair than in the [tsi-tɕi] [ts^{h}i-tɕ^{h}i] pair is related to the larger COG difference in [si-ɕi] than [tsi-tɕi] and [ts^{h}i-tɕ^{h}i].

Chapter 4. Phonetic properties of sibilant contrasts in the [_i] context

The perceptual experiment in Chapter 3 shows that, for dental vs. palatal sibilants, the place contrasts in the [_i] context are perceptually less distinct than in other vowel contexts. This is consistent with the observation that the same contrasts tend to be avoided in cross-linguistic typology and the historical sound changes of Mandarin. An implication of this observation is that contrastive dental vs. plalatal sibilants in the [_i] context are likely to be avoided in sound changes. Such contrasts are reported to be present in Xiangtan Chinese (Zeng 1993), a dialect with the same sibilant system as 18th century Mandarin. The production of dental vs. palatal sibilants in Xiangtan is examined in this chapter, to investigate if there is a sign of avoiding this contrast in the [_i] context.

4.1 Introduction

4.1.1 Dental vs. palatal sibilants in the [_i] context as weak contrasts

For the place contrasts of dental vs. palatal sibilants in the [_i] context, the studies in this dissertation have shown the following properties:

(a) The typological survey (2.2.1 – 2.2.3) shows that, across Chinese dialects, the place contrasts of dental vs. palatal sibilants in the [_i] context tend to be avoided. For example, of the 81 dialects with dental vs. palatal sibilants only, contrastive [si tsi tsʰi] vs. [ɕi tɕi tɕʰi] are present in 23 dialects and enhanced or avoided in 58 dialects;

(b) The historical development of Mandarin also shows the tendency of avoiding contrastive dental vs. palatal sibilants in the [_i] context (2.3.2): the contrast emerged three times as the result of other sound changes but was subsequently enhanced or avoided with a vowel change or a sibilant change;

(c) The perceptual experiment (Chapter 3) shows that, for dental vs. palatal sibilants, the [_i] context reduces the perceptual distinctiveness of the consonantal place contrasts compared with other vowel contexts.

These observations suggest that (i) dental vs. palatal sibilants form perceptually weak contrasts in the [_i] context, i.e., [si tsi ts^{h}i] vs. [ɕi tɕi tɕ^{h}i]; and (ii) such contrasts are ‘unstable’ in a sound system and are likely to be enhanced or avoided in sound change.

4.1.2 The neutralization of dental and palatal sibilants in the [_i] context

In the typology survey across Chinese dialects, it has been shown that about one-third of Chinese dialects were reported to have the contrasts between dental and palatal sibilants in the [_i] context, e.g., Xiangtan (Zeng 1993) as in (29). Based on the properties of the contrasts laid out in 4.1.1, it is predicted that the sibilant dental vs. palatal contrasts in the [_i] context are likely to be avoided in sound change.

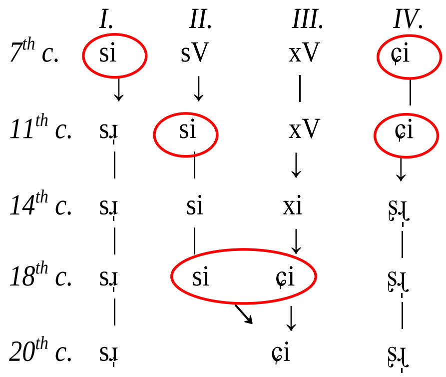
(29) Sibilant-initial syllables in Xiangtan (Zeng 1993)

sɹ̩	四	‘four’	**si**	細	‘slim’	**ɕi**	戲	‘opera’	ʂɻ̩	屍	‘corpse’
tsɹ̩	姿	‘pose’	**tsi**	祭	‘sacrifice’	**tɕi**	技	‘skill’	tʂɻ̩	枝	‘branch’
ts^{h}ɹ̩	賜	‘to bestow’	**ts^{h}i**	砌	‘lay bricks’	**tɕ^{h}i**	氣	‘gas’	tʂ^{h}ɻ̩	痴	‘stupid’

All syllables bear a high-level tone

The avoidance of contrastive dental vs. palatal sibilants in the [_i] context in the historical development of Mandarin, as summarized in (14), is repeated below in (30).

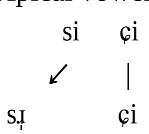
(30) The historical sound changes in Mandarin repeated from (14)
(Based on the reconstructions from Li and Zhou 1999; L. Wang 1985)



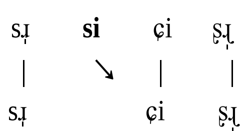
As shown in the survey across Chinese dialects, the most frequent way to avoid dental vs. palatal sibilants is to develop apical vowels after the dental sibilants, as in the historical development of Mandarin sibilants in (31a). Between the 18th century and the 20th century, however, the contrastive dental vs. palatal sibilants in the [_i] context in Mandarin were avoided with the neutralization of the dental and palatal sibilants, as in (31b), rather than the formation of the apical vowel as in (31c).

(31) The avoidance of dental vs. palatal sibilants in the [_i] context

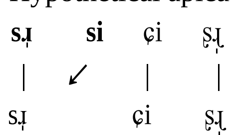
a. Apical vowel formation (7th c. – 11th c.)



b. Neutralization of dental vs. palatal sibilants in 18th century Mandarin



c. Hypothetical apical vowel formation in 18th-century Mandarin



A look at (31) shows that, in both (31b) and (31c), the sound changes would involve the neutralization of syllables, and hence the neutralization of morphemes, because most morphemes in Chinese are monosyllabic (Packard 2015). Instead of being neutralized with [sɻ̩] (31c), syllables like [si] were neutralized with [ɕi] (31b). In terms of perceptual distinctiveness, the sound change in (31b) turned out to involve the neutralization of two elements (e.g., [si] and [ɕi]) that are perceptually more confusable compared with the two parties in (31c) (e.g., [si] and [sɻ̩]), where [i] and [ɻ̩] differ by F2 values (Zee and Lee 2001; Lee and Li 2003; Cheung 2004).[23]

As shown in (29), Xiangtan Chinese (Zeng 1993) has the same syllabic inventory as 18th-century Mandarin. It is thus predicted that in Xiangtan Chinese, the place contrasts of dental vs. palatal sibilants in the [_i] context are likely to show a sign of being neutralized, with the palatalization of the dental sibilants, e.g., [si] → [ɕi]. To test this prediction, two phonetic studies were conducted: An acoustic analysis of dental vs. palatal sibilants in Xiangtan Chinese and a perceptual experiment testing the perceptual properties of the dental sibilants in the [_i] context.

---

[23] It is obvious that there could be multiple factors that have contributed to the choice of (31b) over (31c). For example, previous studies have found that phonological contrasts with higher functional loads are less likely to be neutralized in sound changes (Wedel, Jackson, and Kaplan 2013, among others). The analysis in this dissertation is limited to the role of perceptual distinctiveness in the sound change and the investigation of other factors is reserved for future studies.

4.2 Acoustic measurements: Sibilant contrasts in Xiangtan Chinese

4.2.1 Recording

4.2.1.1 Speakers

Eleven female native speakers of Xiangtan Chinese participated in this study. Their age ranged from 38 to 65 years old, with a mean age of 46.5 years. At the time of the recording, all speakers were living in the district of Bantang of Xiangtan City in Hunan Province, China. Most of the speakers can understand Mandarin but do not speak in Mandarin in their daily life, and a few younger speakers can speak both Xiangtanese and Mandarin.

4.2.1.2 Material

The word list in (32) was used in the recording, where the syllables were chosen from Zeng's (1993) phonetic description of Xiangtan Chinese and checked by a female native speaker of Xiangtan Chinese who has linguistic training. (She did not participate in the recording.) Three groups of syllables were included: (i) syllables like [sɹ̩] in (32a), where dental sibilants precede homorganic apical vowels (i.e., syllabic approximants); (ii) syllables like [si] in (32b), where dental sibilants precede the vowel [_i]; (iii) syllables like [ɕi] in (32c), where palatal sibilants precede the vowel [_i]. The syllables in (32a) served as the baseline for canonical dental sibilants, while the syllables in (32c) served as the baseline for canonical palatal sibilants.

(32) Wordlist used for the recording of Xiangtan Chinese

a. Canonical dentals	b. Pre-[i] dentals	c. Canonical palatals
sɹ̩ 四 'four'	**s**i 細 'slim'	**ɕ**i 戲 'opera'
tsɹ̩ 姿 'looks'	**ts**i 祭 'to sacrifice'	**tɕ**i 技 'skill'
tsʰɹ̩ 賜 'to bestow'	**tsʰ**i 砌 'to lay bricks'	**tɕʰ**i 氣 'gas'

All syllables have a high-level tone.

The dental sibilants in (32b), i.e., those before the vowel [_i], are the focus of this production study. They were reported to fully contrast with the canonical palatals in (32c) (Zeng 1993). Yet, based on the discussion in 4.1, it is hypothesized that these dental sibilants in the [_i] context are likely to show a sign of being palatalized, and thus neutralized with the canonical palatal sibilants in (32c).

4.2.1.3 Recording procedure

The recording was done in a quiet room in the place where the speakers were living. The recording devices included a Shure SM10A head-mounted microphone and a Marantz Portable Solid State Recorder (PMD 671). The microphone was adjusted to ensure that the speakers were comfortable before the recording.

The syllables in (32) were embedded in the carrier sentence '*wo du* __ *qi ci*' ('I read __ seven times'). The speakers were asked to read the sentences in Xiangtan Chinese at a normal speed and each sentence was read three times in a row and, after reading all the sentences, the speakers were asked to read all the sentences in the same way again. Thus the 11 speakers each produced 6 tokens for each syllable in (32) and a total of 594 syllables were recorded. A native female Xiangtan speaker helped to explain the instructions to the native speakers.

The recordings were segmented in Praat (Boersma 2001) using a Praat script (Lennes 2002), and each token of the syllables in (32) was saved as a separate .wav sound file.

4.2.2 Measurements

4.2.2.1 Acoustic properties measured

Four acoustic properties were measured for the onset sibilants: COG (center of gravity or spectral mean), dispersion of energy (spectral variance), intensity, and duration. These four measurements were adopted because each has been shown in previous studies to be able to differentiate dental vs. palatal sibilants, as detailed below.

(a) COG has been shown to be an important cue for the place distinction of sibilants (Heinz and Stevens 1961, Jongman, Wayland, Wong 2000, among others). Dental vs. palatal sibilants in Mandarin are also reported to differ in COG. For example, the COG of [s] is higher than that of [ɕ] (Svantesson 1986), and the lower boundary of energy concentration in [s] (3100Hz) is higher than that in [ɕ] (1800Hz) (Wu and Lin 1989);

(b) In terms of energy dispersion, the dental [s] was reported to have a smaller variance (i.e. narrower dispersion) than the palatal [ɕ] (Svantesson 1986);

(c) Intensity has been reported to differentiate Mandarin sibilants at different places (Li and Gu 2015), with greater intensity values for more posterior sibilants;

(d) For dental vs. palatal sibilants, different patterns of duration differences have been reported in certain Chinese dialects, i.e., the dentals have been reported to be longer than the palatals in some dialects but the pattern is the reverse in others (Ran 2005; C. Liu 2010; Pan 2010). For a particular dialect like Xiangtan, therefore, it remains a possibility that duration could serve as a cue to the place distinction between dental and palatal sibilants.

4.2.2.2 Palatalization vs. coarticulation

In normal speech, the articulatory gestures for a consonant usually overlap with a neighboring vowel. The acoustic properties of the fricatives, for example, are usually influenced by the following vowel, particularly at the portion close to the onset of the vowel (Liberman et al. 1967, Sereno, Baum, Marean and Lieberman 1987, among others). An acoustic measurement over the whole sibilant, therefore, would involve the acoustic properties of the sibilant itself as well as the coarticulatory influence from the following vowel. In the current study, the focus is on whether the dental sibilants in the syllables [si] [tsi] [ts^{h}i] have undergone phonological palatalization. Therefore, it is crucial to tease apart palatalization as the result of a *sound change* from the coarticulation between the dental sibilants and a following [_i]. In other words, it is important to minimize the influence of consonant-vowel coarticulation in the acoustic measurements.

In terms of acoustic properties, Soli (1981) observed that the fricative [s] has different peak frequencies in the vowel contexts [_i], [_a] and [_u] in the portion close to the vowel; but the peak frequencies are indistinguishable in the portion of the [s] 60 ms before the onset of a following vowel. In terms of articulatory gestures, Iskarous, Shadle, and Proctor (2011) observed that, for /s/ in a high-vowel context vs. /s/ in a low-vowel context, the jaw rose a comparable amount across the first half of the fricative and that the jaw position across the two vowel contexts became significantly different at around 75% of the fricative duration. In other words, the articulatory gesture across the first half of [s] was not greatly affected by a following vowel. Based on these studies, the first half of the sibilants was measured in this study to minimize the influence of consonant-vowel coarticulation.

4.2.2.3 Measurement tool

The sibilants in (32) were measured using a praat script written by DiCanio (2013). The script was revised to take measurements over the first half of the sibilants. Three intervals at equal distance across the first half of a sibilant were measured, each with a window size of 20 ms. Time-averaged measurements were generated over the three intervals by the script for each token, including COG (center of gravity), dispersion (variance), intensity, and duration.

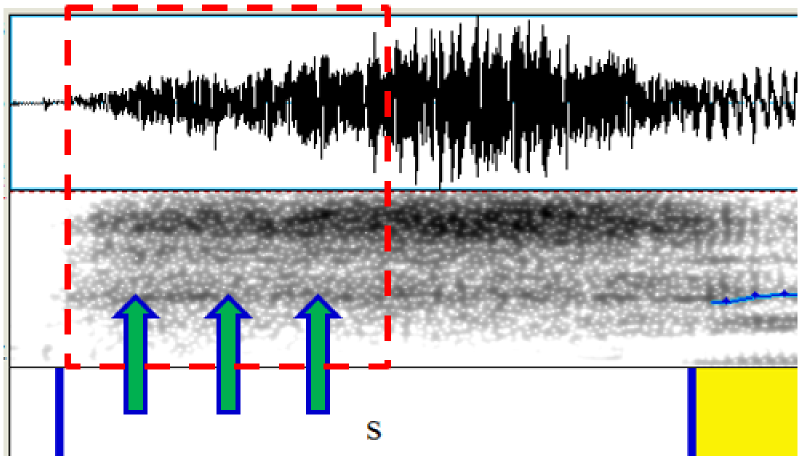


Figure 11 Time-averaged measurements over the first half of the sibilant.

4.2.3 Predictions

Based on the historical development of Mandarin, it is predicted that, in Xiangtan Chinese, there will likely be a sign of neutralization between the dental and palatal sibilants in the [_i] context and that the dental sibilants in [si tsi tsʰi] will likely show a sign of being palatalized (4.1). Thus, it is predicted that, in terms of acoustic properties, at least some of the dental sibilants in [si tsi tsʰi] should be indistinguishable from the canonical palatal sibilants in [ɕi tɕi tɕʰi]. Below, the acoustic properties of the sibilants are first reported (4.2.4), and then linear discriminant analyses were performed to test the predicted palatalization of the sibilants in [si tsi tsʰi] (4.2.5).

4.2.4 Acoustic properties

The focus of this study is the contrast between dental and palatal sibilants in the [_i] context. Thus, comparisons are made between the canonical dental sibilants, the dental sibilants before [_i], and the canonical palatal sibilants. Below, separate reports are given to the results of the measurements on fricatives, unaspirated affricates, and aspirated affricates.

4.2.4.1. Fricatives

The acoustic properties of the three fricatives are plotted in Figure 12, where ‘s’ refers to the canonical /s/ in [sɨ], ‘s before [_i]’ refers to the /s/ in [si], and ‘ɕ’ refers to the canonical /ɕ/ in [ɕi].

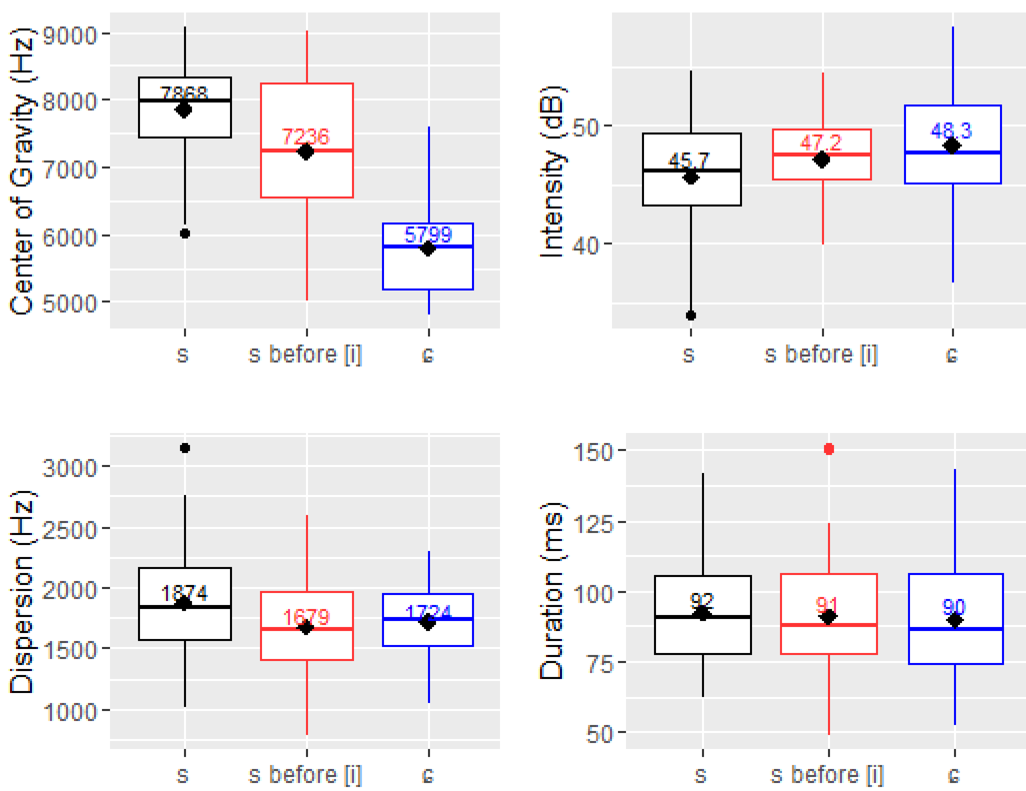


Figure 12 Acoustic properties of dental and palatal fricative across the 11 female speakers.
COG (Hz), Dispersion (Hz), Intensity (dB), and Duration (ms).
The bar and the number in the middle of each box indicates the mean value.

To examine the difference among the three types of fricatives, a series of statistical analyses was run separately on the COG, dispersion, intensity, and duration measurements.

A. COG

The COG of fricatives were analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (33).

(33) Variables in the analysis of fricatives – COG

Variable names	*Notes and abbreviation*
Dependent variable	
COG	The acoustic measurements of COG in Hz
Predicting variables	
Onset	s (Canonical /s/), si (/s/ in the [_i] context), ɕ (Canonical /ɕ/)
	(Baseline = s (Canonical /s/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 30. This model was obtained by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 197.22, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 30 Fixed effects in the mixed-effect linear regression for fricative – COG

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	7867.92	169.48	13.84	46.423	< .001***
Onset(si)	–631.61	113.97	194.67	–5.542	< .001***
Onset(ɕ)	–2068.74	113.97	194.67	–18.152	< .001***

Model: COG~ Onset + (1|Speaker)
Baseline: Onset = s (Canonical /s/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among s (Canonical /s/), si (/s/ in the [_i] context), and ɕ (Canonical /ɕ/) were further checked using s (Canonical /s/) and si (/s/ in the [_i] context) alternatively as the baseline. The results are reported in Table 31 in terms of t values (p values). It turned out that the COG of three sibilant onsets were all significantly different from one another. More spefically, s (Canonical /s/) had a higher COG than si (/s/ in the [_i] context), which in turn had a higher COG than ɕ (Canonical /ɕ/).

Table 31 Acoustics of fricative COG – Onset difference in *t* value (*p* value)

	si (/s/ in the [_i] context)	ɕ (Canonical /ɕ/)
s (Canonical /s/)	–5.542 (< **.001*****)	–18.152 (<**.001*****)
si (/s/ in the [_i] context)		–12.610 (<**.001*****)

P values appear in parentheses and boldface marks those that reached significance (.05).

B. Dispersion

The measurements of the dispersion in the fricatives were analyzed using Linear Mixed Effects Models too, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (34).

(34) Variables in the analysis of fricatives – Dispersion

Variable names	*Notes and abbreviation*
Dependent variable	
Dispersion	The acoustic measurements of Dispersion in Hz
Predicting variables	
Onset	s (Canonical /s/), si (/s/ in the [_i] context), ɕ (Canonical /ɕ/) (Baseline = s (Canonical /s/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 32. This model was obtained by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 16.831, *df*= 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 32 Fixed effects in the mixed-effect linear regression for fricative – Dispersion

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	1873.82	87.50	12.45	21.415	< .001***
Onset(si)	–194.77	48.87	185.33	–3.986	< .001***
Onset(ɕ)	–149.89	48.87	185.33	–3.067	.002**

Model: Dispersion ~ Onset + (1|Speaker)
Baseline: Onset = s (Canonical /s/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among s (Canonical /s/), si (/s/ in the [_i] context), and ɕ (Canonical /ɕ/) were further checked using s (Canonical /s/) and si (/s/ in the [_i] context) alternatively as the baseline. The results are reported in Table 33 in terms of t values (p values). It turned out that s (Canonical /s/) had a significantly smaller Dispersion value than that of si (/s/ in the [_i] context) and ɕ (Canonical /ɕ/), and there was no significant difference between the latter two onsets.

Table 33 Acoustics of Fricative Dispersion – Onset difference in *t* value (*p* value)

	si (/s/ in the [_i] context)	ɕ (Canonical /ɕ/)
s (Canonical /s/)	–3.986 (**< .001*****)	–3.067 (**.002****)
si (/s/ in the [_i] context)		0.918 (.36)

P values appear in parentheses and boldface marks those that reached significance (.05).

C. Intensitiy

The measurements of the Intensity in the fricatives were also analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variables are listed in (35).

(35) Variables in the analysis of fricatives – Intensity

Variable names	*Notes and abbreviation*
Dependent variable	
Intensity	The acoustic measurements of Intensity in dB
Predicting variables	
Onset	s (Canonical /s/), si (/s/ in the [_i] context), ɕ (Canonical /ɕ/)
	(Baseline = s (Canonical /s/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 34. This model was obtained by comparing a null model with Speaker as the random factor with a superset model adding Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 16.762, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 34 Fixed effects in the mixed-effect linear regression for fricative – Intensity

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	45.6864	0.8563	14.82	53.351	< .001***
Onset(si)	1.4788	0.6286	185	2.352	.020*
Onset(ɕ)	2.6106	0.6286	185	4.153	< .001***

Model: Intensity ~ Onset + (1|Speaker)
Baseline: Onset = s (Canonical /s/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among s (Canonical /s/), si (/s/ in the [_i] context), and ɕ (Canonical /ɕ/) were further checked using s (Canonical /s/) and si (/s/ in the [_i] context) alternatively as the baseline. The results are reported in Table 35 in terms of t values (p values). It turned out that the Intensity of s (Canonical /s/) was significantly smaller than that of si (/s/ in the [_i] context) and ɕ (Canonical /ɕ/), without a significant difference between the latter two onsets.

Table 35 Acoustics of fricative intensity – Onset difference in *t* value (*p* value)

	si (/s/ in the [_i] context)	ɕ (Canonical /ɕ/)
s (Canonical /s/)	2.352 (**.020***)	4.153 (< **.001*****)
si (/s/ in the [_i] context)		1.800 (.073.)

P values appear in parentheses and boldface marks those that reached significance (.05).

D. *Duration*

The measurements of the duration of the fricatives were analyzed using Linear Mixed Effects Models as well, for which the dependent variable, the predicting variable and its baseline, and random variables are listed in (36).

(36) Variables in the analysis of fricatives – Duration

Variable names	*Notes and abbreviation*
Dependent variable	
Duration	The acoustic measurements of Duration in ms
Predicting variables	
Onset	s (Canonical /s/), si (/s/ in the [_i] context), ɕ (Canonical /ɕ/) (Baseline = s (Canonical /s/))
Random variable	
Speaker	11 speakers

Model comparison was adopted to determine if Onset made a difference in the duration of s (Canonical /s/), si (/s/ in the [_i] context), ɕ (Canonical /ɕ/). More specifically, a null model with Speaker as the random factor was compared with a superset model with Onset as the predicting variable. It turned out that the addition of Onset did not significantly improve the model. Therefore, Onset was not a significant predictor of duration.

A summary of the effect of Onset in the four acoustic measurements is given in (37) below, focusing on whether Onset has a significant effect and the directionality of the effect, where '>' and '<' indicate 'have a significantly higher/lower value than'.

(37) Fricative: Onset effect in model comparisons and the directionality of onset effect

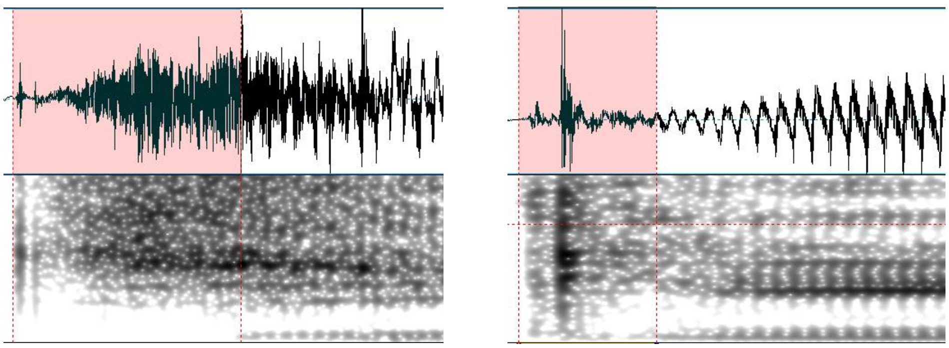
	Onset effect	*Directionality*
COG	✓	s > s before [i] > ɕ
Dispersion	✓	s > s before [i] & ɕ
Intensity	✓	s < s before [i] & ɕ
Duration	✗	

4.2.4.2. Unaspirated affricates

For the unaspirated affricates, the three types of sibilants are the canonical dental /ts/ in [tsɹ̩], the pre-[i] /ts/ in [tsi], and the canonical palatal /tɕ/ in [tɕi].

When manually checking the data, it was found that four speakers' production of the /tɕi/ syllables involved variants whose onsets are the unaspirated velar stop [k]. As shown in Figure 13, the onset sibilant of the /tɕi/ in F01's production was the affricate [tɕ], with a clear rise of the amplitude of frication noise. The onset sibilant in F06's production, however, was a typical stop, with a release burst followed by a brief period of frication noise. Among the 11 speakers, velar stops were produced by four speakers: F06 and F09 produced all their /tɕi/ syllables as [ki], while F10 and F11 produced some of their /tɕ/ syllables as [ki].[24] In the acoustic measurements of the canonical palatal /tɕ/, the data from these 4 speakers were excluded. The measurements of the canonical /ts/ and the pre-[i] /ts/ were performed over all 11 speakers because the manual check showed no irregularity of the production of these two types of onset sibilants.

a. F01_tci_1 b. F06_tci_1



[tɕ] [k]

Figure 13 Palatal affricate vs. velar stop in the production of /tɕi/.

The acoustic properties of the three unaspirated affricates are plotted in Figure 14, where 'ts' refers to canonical /ts/ in [tsɿ], 'ts before [_i]' refers to the onset in [tsi], and 'tɕ' refers to canonical palatal /tɕ/ in [tɕi].

---

[24] It is unclear why these four speakers produced the palatal /tɕ/ differently from the other speakers. A potential reason is that these speakers have been influenced by speakers of another Chinese dialect. In Liuyang Chinese, a dialect spoken 100km from Xiangtan, for example, [ki] and [k^hi] are reported to be the pronunciations of the lexical items 鸡 'rooster' and 欺 'to cheat' (Xia 1983), which are [tɕi] and [$tɕ^h$i] in Xiangtan. Further research is needed to investigate the stop variants of the affricates and whether this occurred in other vowel contexts.

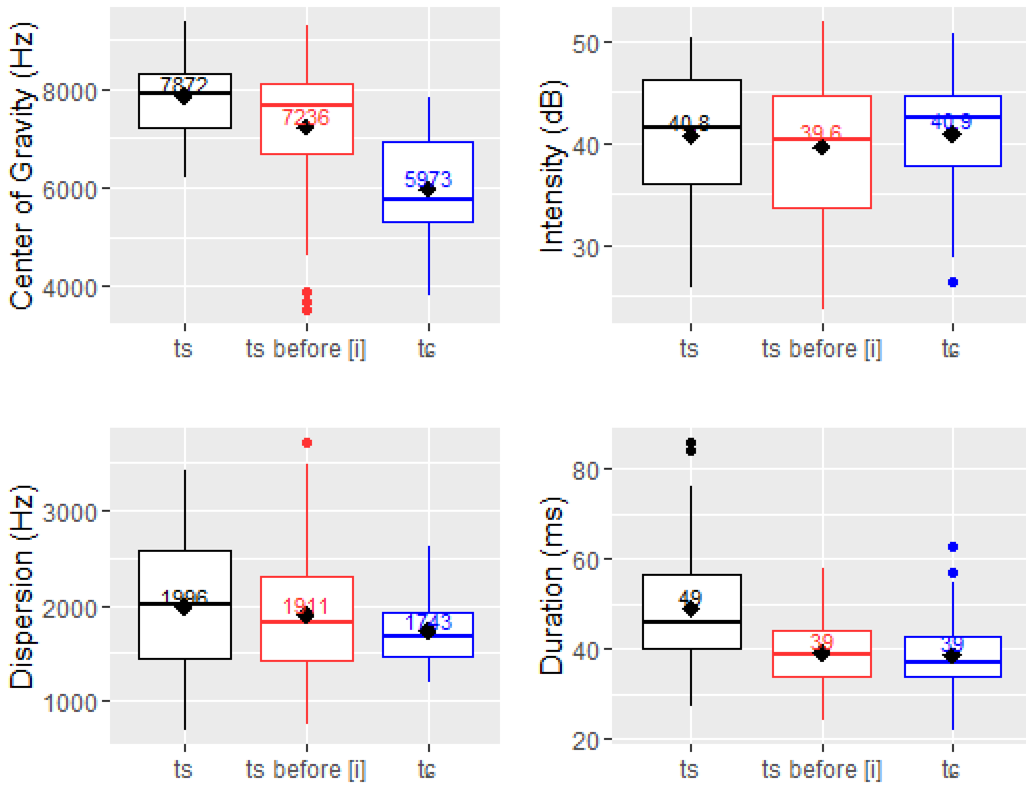


Figure 14 Acoustic properties of dental and palatal unaspirated affricates.
COG (Hz), Dispersion (Hz), Intensity (dB), and Duration (ms).
The bar and the number in the middle of each box indicates the mean value.

To examine if the three types of onsets significantly differ in the acoustic meansurements in Figure 14, a series of statistical analyses was performed separately on the measurements of COG, Dispersion, Intensity, and Duration.

A. COG

The COG measurements of the unaspirated affricates were analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (38).

(38) Variables in the analysis of unaspirated affricates – COG

Variable names	*Notes and abbreviation*
Dependent variable	
COG	The acoustic measurements of COG in Hz
Predicting variables	
Onset	ts (Canonical /ts/), tsi (/ts/ in the [_i] context), tɕ (Canonical /tɕ/) (Baseline = ts (Canonical /ts/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 36. This model was obtained by comparing a null model with Speaker as the random factor with a superset model adding Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 93.064, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 36 Fixed effects in the mixed-effect linear regression for unaspirated affricates – COG

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	7871.73	258.96	12.09	30.397	< .001***
Onset(tsi)	–636.18	139.53	165.38	–4.559	< .001***
Onset(tɕ)	–1840.07	166.12	167.79	–11.077	< .001***

Model: COG~ Onset + (1|Speaker)
Baseline: Onset = ts (Canonical /ts/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among ts (Canonical /ts/), tsi (/ts/ in the [_i] context), and tɕ (Canonical /tɕ/) were further checked using ts (Canonical /ts/) and tsi (/ts/ in the [_i] context) alternatively as the baseline. The results are reported in Table 37 in terms of t values (p values). It turned out that the COG of three sibilant onsets were all significantly different from one another. More spefically, ts (Canonical /ts/) had a higher COG than tsi (/ts/ in the [_i] context), which in turn had a higher COG than tɕ (Canonical /tɕ/).

Table 37 Acoustics of unaspirated affricate COG – Onset difference in *t* value (*p* value)

	si (/s/ in the [_i] context)	ɕ (Canonical /ɕ/)
s (Canonical /s/)	–4.559 (**<.001*****)	–11.077 (<**.001*****)
si (/s/ in the [_i] context)		–7.247 (<**.001*****)

P values appear in parentheses and boldface marks those that reached significance (.05).

B. Dispersion

The measurements of the dispersion in the unaspirated affricates were analyzed with Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (39).

(39) Variables in the analysis of unaspirated affricates – Dispersion

Variable names	*Notes and abbreviation*
Dependent variable	
Dispersion	The acoustic measurements of Dispersion in Hz
Predicting variables	
Onset	ts (Canonical /ts/), tsi (/ts/ in the [_i] context), tɕ (Canonical /tɕ/) (Baseline = ts (Canonical /ts/))
Random variable	
Speaker	11 speakers

Model comparison was adopted to determine if Onset made a difference in the Dispersion of ts (Canonical /ts/), tsi (/ts/ in the [_i] context), tɕ (Canonical /tɕ/) by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. It turned out that the addition of Onset did not significantly improve the model. Therefore, Onset was not a significant predictor of Dispersion.

C. Intensitiy

The measurements of the intensity in the unaspirated fricatives were analyzed using Linear Mixed Effects Models too, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (40).

(40) Variables in the analysis of unaspirated affricates – Intensity

Variable names	*Notes and abbreviation*
Dependent variable	
Intensity	The acoustic measurements of Intensity in dB
Predicting variables	
Onset	ts (Canonical /ts/), tsi (/ts/ in the [_i] context), tɕ (Canonical /tɕ/) (Baseline = ts (Canonical /ts/))
Random variable	
Speaker	11 speakers

Model comparison was adopted to determine if Onset made a difference in the Intensity of ts (Canonical /s/), tsi (/s/ in the [_i] context), tɕ (Canonical /ɕ/) by comparing a null model with Speaker as the random factor with a superset model adding Onset as the predicting variable. It turned out that the addition of Onset did not significantly improve the model. Therefore, Onset was not a significant predictor of Intensity.

D. Duration

The measurements of the duration in the unaspirated affricates were analyzed using Linear Mixed Effects Models as well, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (41).

(41) Variables in the analysis of unaspirated affricates – Duration

Variable names	*Notes and abbreviation*
Dependent variable	
Duration	The acoustic measurements of Duration in ms
Predicting variables	
Onset	ts (Canonical /ts/), tsi (/ts/ in the [_i] context), tɕ (Canonical /tɕ/)
	(Baseline = ts (Canonical /ts/))
Random variable	
Speaker	11 speakers

Model comparison was adopted to determine if Onset made a difference in the duration of ts (Canonical /ts/), tsi (/ts/ in the [_i] context), and tɕ (Canonical /tɕ/) by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. The addition of Onset turned out to significantly improve the model (X^2 = 40.622, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor, as in Table 38.

Table 38 Fixed effects in the mixed-effect linear regression for unaspirated affricates – Duration

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	48.985	1.877	16.42	26.092	< .001***
Onset(tsi)	–9.773	1.629	160.85	–5.998	< .001***
Onset(tɕ)	–10.246	1.925	167.04	–5.324	< .001***

Model: Duration~ Onset + (1|Speaker)
Baseline: Onset = ts (Canonical /ts/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among ts (Canonical /ts/), tsi (/ts/ in the [_i] context), and tɕ (Canonical /tɕ/) were further checked using ts (Canonical /ts/) and tsi (/ts/ in the [_i] context) alternatively as the baseline. The results are reported in Table 39 in t values (p values), which showed that the duration of ts (Canonical /s/) was significantly longer than that of tsi (/s/ in the [_i] context) and tɕ (Canonical /tɕ/), while the latter two had no significant difference.

Table 39 Acoustics of unaspirated affricates – Onset difference in *t* value (*p* value)

	tsi (/ts/ in the [_i] context)	tɕ (Canonical /tɕ/)
ts (Canonical /ts/)	–5.998 (< **.001*****)	–5.324 (<**.001*****)
tsi (/ts/ in the [_i] context)		–0.246 (.806)

P values appear in parentheses and boldface marks those that reached significance (.05).

The Onset effect in the four acoustic measurements of are summarized in (42) below, where '>' indicates 'a significantly higher value than'.

(42) Unaspirated affricates: Onset effect in model comparisons and the directionality of onset effect

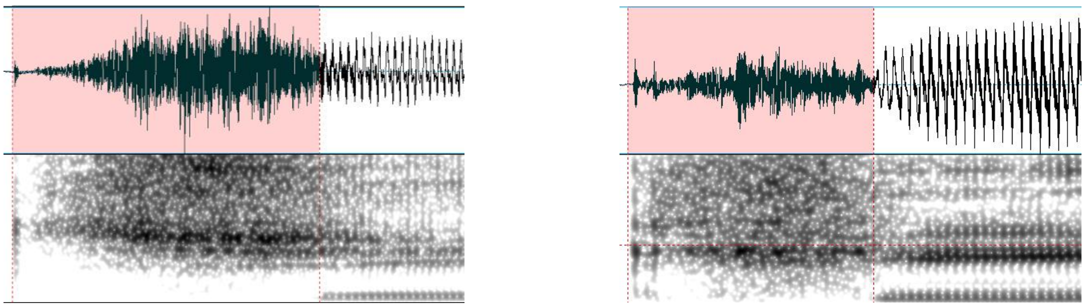
	Onset effect	*Directionality*
COG	✓	ts > ts before [i] > tɕ
Dispersion	×	
Intensity	×	
Duration	✓	ts > ts before [i] & tɕ

4.2.4.3. Aspirated affricates

For the aspirated affricates, the three types of sibilants are the canonical dental /tsʰ/ in [tsʰɹ̩], the pre-[i] /tsʰ/ in [tsʰi], and the canonical palatal /tɕʰ/ in [tɕʰi].

A manual check of the data revealed that the same four speakers who variably produced /tɕi/ as [ki] also produced some or all the /tɕʰi/ syllables as [kʰi]. As Figure 15 shows, the sibilant onset of /tɕʰ/ produced by in F01 was a [tɕʰ], with a clear rise of the amplitude in the frication noise, whereas the one produced by F06 was a clear [kʰ] with double release bursts and aspiration noise afterward. Of these four speakers, F06 produced all her /tɕʰi/ syllables as [kʰi] while F09, F10, and F11 produced some of their /tɕʰi/ syllables as [kʰi]. These four speakers' production of /tɕʰi/ was thus not included in the acoustic measurements.

a. F01_tchi_1 b. F06_tchi_1



[tɕ^h] [k^h]

Figure 15 Palatal affricate vs. velar stop in the production of /tɕ^{h}i/.

The measurements of the canonical /ts^h/ and the pre-[i] /ts^h/ were performed over all 11 speakers. The acoustic properties of the three aspirated affricates are plotted in Figure 16, where ‘ts^h’ is the canonical /ts^h/ in [ts^{h}ɹ̩], ‘ts^h before [_i]’ is the onset in [ts^{h}i], and ‘tɕ^h’ is the canonical palatal /tɕ^h/ in [tɕ^{h}i].

Similar to the analyses above, a series of statistical analyses were performed separately on the measurements of COG, Dispersion, Intensity, and Duration, to examine whether there was a significant difference among the three types of sibilants.

A. COG

The COG of the aspirated affricates were analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (43).

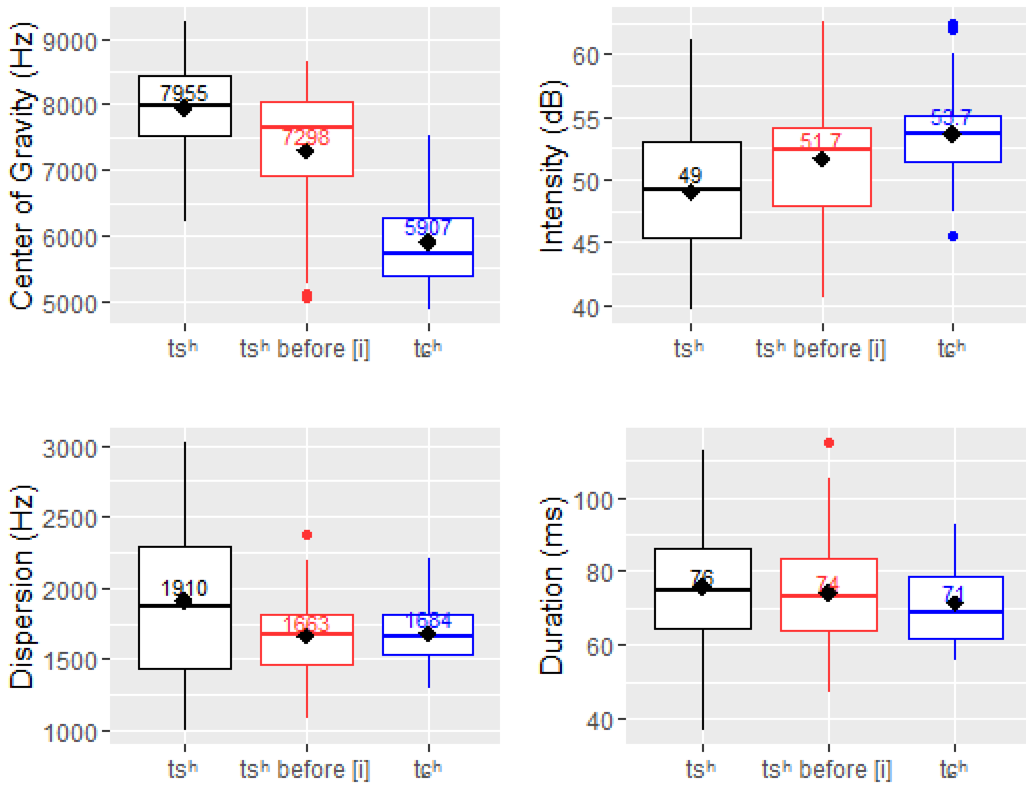


Figure 16 Acoustic properties of dental and palatal aspirated affricates.
COG (Hz), Dispersion (Hz), Intensity (dB), and Duration (ms).
The bar and the number in the middle of each box indicates the mean value.

(43) Variables in the analysis of aspirated affricates – COG

Variable names	*Notes and abbreviation*
Dependent variable	
COG	The acoustic measurements of COG in Hz
Predicting variables	
Onset	tsʰ (Canonical /tsʰ/), tsʰi (/tsʰ/ in the [_i] context), tɕʰ (Canonical /tɕʰ/) (Baseline = tsʰ (Canonical /tsʰ/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 40. This model was obtained by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 128.3, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 40 Fixed effects in the mixed-effect linear regression for aspirated affricates – COG

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	7955.23	178.02	13.19	44.688	< .001***
Onset(tsʰi)	–656.73	115.41	159.97	–5.691	< .001***
Onset(tɕʰ)	–1877.25	137.12	163.36	–13.690	< .001***

Model: COG~ Onset + (1|Speaker)
Baseline: Onset = tsʰ (Canonical /tsʰ/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among tsʰ (Canonical /tsʰ/), tsʰi (/tsʰ/ in the [_i] context), and tɕʰ (Canonical /tɕʰ/) were further checked using tsʰ (Canonical /tsʰ/) and tsʰi (/tsʰ/ in the [_i] context) alternatively as the baseline. The results are reported in Table 41 in terms of t values (p values). It turned out that the COG of the three sibilant onsets were all significantly different from one another. More specifically, tsʰ (Canonical /tsʰ/) had a higher COG than tsʰi (/tsʰ/ in the [_i] context), which in turn had a higher COG than tɕʰ (Canonical /tɕʰ/).

Table 41 Acoustics of aspirated affricate COG – Onset difference in *t* value (*p* value)

	tsʰi (/tsʰ/ in the [_i] context)	tɕʰ (Canonical /tɕʰ/)
tsʰ (Canonical /tsʰ/)	–5.691 (**< .001*****)	–13.690 (<**.001*****)
tsʰi (/tsʰ/ in the [_i] context)		–8.901 (<**.001*****)

P values appear in parentheses and boldface marks those that reached significance (.05).

B. Dispersion

The measurements of the dispersion of the aspirated affricates were analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (44).

(44) Variables in the analysis of aspirated affricates – Dispersion

Variable names	*Notes and abbreviation*
Dependent variable	
Dispersion	The acoustic measurements of Dispersion in Hz
Predicting variables	
Onset	ts^h (Canonical /ts^h/), ts^hi (/ts^h/ in the [_i] context), $tɕ^h$ (Canonical /$tɕ^h$/)
	(Baseline = ts^h (Canonical /ts^h/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 42. This model was obtained by comparing a null model with Speaker as the random factor with a superset model adding Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 23.682, *df*= 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 42 Fixed effects in the mixed-effect linear regression for aspirated affricates – Dispersion

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	1910.45	99.12	11.57	19.275	< .001***
Onset(ts^hi)	–247.23	50.65	161.16	–4.881	< .001***
Onset($tɕ^h$)	–186.63	60.33	13.33	–3.094	.002**

Model: Dispersion~ Onset + (1|Speaker)
Baseline: Onset = ts^h (Canonical /ts^h/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among tsʰ (Canonical /tsʰ/), tsi (/tsʰ/ in the [_i] context), and tɕʰ (Canonical /tɕʰ/) were further checked using tsʰ (Canonical /tsʰ/) and tsʰi (/tsʰ/ in the [_i] context) alternatively as the baseline. The results are reported in Table 43 in terms of t values (p values). It turned out that the three sibilant onsets were all significantly different from one another in terms of Dispersion. More spefically, tsʰ (Canonical /tsʰ/) had a larger dispersion than tsʰi (/tsʰ/ in the [_i] context), which in turn had a larger dispersion than tɕʰ (Canonical /tɕʰ/).

Table 43 Acoustics of unaspirated affricate Dispersion – Onset difference in *t* value (*p* value)

	tsʰi (/tsʰ/ in the [_i] context)	tɕʰ (Canonical /tɕʰ/)
tsʰ (Canonical /tsʰ/)	–4.881 (< **.001*****)	–3.094 (**.002****)
tsʰi (/tsʰ/ in the [_i] context)		1.004 (**.317**)

P values appear in parentheses and boldface marks those that reached significance (.05).

C. Intensitiy

The measurements of the intensity in the aspirated fricatives were analyzed using Linear Mixed Effects Models too, for which the dependent variable, the predicting variable and its baseline, and the random variable are listed in (45).

(45) Variables in the analysis of aspirated affricates – Intensity

Variable names	*Notes and abbreviation*
Dependent variable	
Intensity	The acoustic measurements of Intensity in dB
Predicting variables	
Onset	tsʰ (Canonical /tsʰ/), tsʰi (/tsʰ/ in the [_i] context), tɕʰ (Canonical /tɕʰ/) (Baseline = tsʰ (Canonical /tsʰ/))
Random variable	
Speaker	11 speakers

The fixed effects in the final model are presented in Table 44. This model was obtained by comparing a null model with Speaker as the random factor with a superset model adding Onset as the predicting variable. The addition of Onset significantly improved the model (X^2 = 51.842, *df* = 2, p < .001). Therefore, the final model included Onset as the predicting variable and Speaker as the random factor.

Table 44 Fixed effects in the mixed-effect linear regression for aspirated affricates – Intensity

	Estimate	*Std.Error*	*df.*	*t value*	*Pr(>|t|)*
(Intercept)	49.0182	1.1644	11.31	42.098	< .001***
Onset(tsʰi)	2.6985	0.5359	160.92	5.035	< .001***
Onset(tɕʰ)	4.7746	0.6388	162.65	7.475	< .001***

Model: Intensity ~ Onset + (1|Speaker)
Baseline: Onset = tsʰ (Canonical /tsʰ/)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

For the significant effect of Onset, the differences among tsʰ (Canonical /tsʰ/), tsi (/tsʰ/ in the [_i] context), and tɕʰ (Canonical /tɕʰ/) were further checked using tsʰ (Canonical /tsʰ/) and tsʰi (/tsʰ/ in the [_i] context) alternatively as the baseline. The results are reported in Table 45 in terms of t values (p values). It turned out that the three sibilant onsets were all significantly different from one another in Dispersion. More spefically, tsʰ (Canonical /tsʰ/) had a smaller intensity than tsʰi (/tsʰ/ in the [_i] context), which in turn had a smaller intensity than tɕʰ (Canonical /tɕʰ/).

Table 45 Acoustics of unaspirated affricate intensity – Onset difference in *t* value (*p* value)

	tsʰi (/tsʰ/ in the [_i] context)	tɕʰ (Canonical /tɕʰ/)
tsʰ (Canonical /tsʰ/)	5.035 (**< .001*****)	7.475 (**001*****)
tsʰi (/tsʰ/ in the [_i] context)		3.250 (**.001****)

P values appear in parentheses and boldface marks those that reached significance (.05).

D. Duration

The measurements of the duration in the aspirated affricate were analyzed using Linear Mixed Effects Models, for which the dependent variable, the predicting variable and its baseline, and random variables are listed in (46).

(46) Variables in the analysis of fricatives – Duration

Variable names	*Notes and abbreviation*
Dependent variable	
Duration	The acoustic measurements of Duration in ms
Predicting variables	
Onset	ts^h (Canonical /ts^h/), ts^{h}i (/ts^h/ in the [_i] context), tɕ^h (Canonical /tɕ^h/) (Baseline = ts^h (Canonical /ts^h/))
Random variable	
Speaker	11 speakers

Model comparison was adopted to determine if Onset made a difference in the duration of ts^h (Canonical /ts^h/), ts^{h}i (/ts^h/ in the [_i] context), tɕ^h (Canonical /tɕ^h/) by comparing a null model with Speaker as the random factor with a superset model with Onset as the predicting variable. It turned out that the addition of Onset did not significantly improve the model. Therefore, Onset was not a significant predictor of duration.

A summary of the effect of Onset in the four acoustic measurements of aspirated affricates is given in (47) below, focusing on whether Onset has a significant effect and the directionality of the effect, where ‘>’ and ‘<’ indicate ‘have a significantly higher/lower value than’.

(47) Aspirated affricates: Onset effect in model comparisons and the directionality of onset effect

	Onset effect	*Directionality*
COG	✓	ts^h > ts^h before [i] > $tɕ^h$
Dispersion	✓	ts^h > ts^h before [i] & $tɕ^h$
Intensity	✓	ts^h < ts^h before [i] < $tɕ^h$
Duration	×	

The results of the acoustic measurments, as listed in (37) (42) and (47), show that the dental sibilants in the [_i] context in general fall between the canonical dental sibilants and the palatal sibilants. The focus of the study in Xiangtan sibilants is if in the [_i] context, there is a sign of neutralization between the dental and palatal sibilants. It is predicted that there should be a sign of neutralization, by which the dental sibilants in the [_i] context is indistinguishable from the canonical palatal sibilants. To test this prediction, the acoustic measurements of the sibilants in Xiangtan Chinese were submitted to a linear discrimination analysis, as detailed in 4.2.5.

4.2.5 Discriminant analyses

To examine whether there is a sign of neutralization between the dental and palatal sibilants in the [_i] context, the measurements in 4.2.4 were submitted to a series of linear discriminant analyses (LDA) — a statistical analysis using a discriminant function to assign data points to one of two or more groups.[25] In an LDA analysis, a classifer was first built from a training set that have two (or more) groups based on a number of measurements. The classifer was trained to distinguish the members of the two (or more) groups based on the measurements. A typical function of a built classifer is to predict the group affiliation of a new number that have the same

[25] There are two major types of discriminant analyses: Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). LDA was adopted in this study due to the relatively small number of data points. An analysis using QDA was also conducted and it returned very similar results to those of the LDA.

measurements as the ones in the training set.

In this study, LDA would be adopted to recognize a pre-[i] dental sibilant as being dental or palatal. More specifically, the two groups to be recognized as are dental sibilants and palatal sibilants. For these two groups, the training set included the acoustic measurements of canonical dental sibilants and those of canonical palatal sibilants. Then, a classifer would be built based on the acoustic measurements of dental vs. palatal sibilants, i.e., COG, Dispersion, Intensity, Duration, etc. The build classifer would then be adopted to predict whether a dental sibilant in the [_i] context should be recognized as a dental sibilant or a palatal sibilant. Below, separate analyses were performed on the fricatives, unaspirated affricates, and aspirated affricates.

4.2.5.1. Fricatives

The acoustic measurements of fricatives (4.2.2.1) were divided into two subsets:

(a) A *training* set including the measurements of COG, Dispersion, Intensity, and Duration for the 66 canonical dental /s/ in [sɻ̩] and the 66 canonical palatal /ɕ/ in [ɕi];

(b) A *test* set including the measurements of COG, Dispersion, Intensity, and Duration for the 66 pre-[i] dental /s/ in [si].

In both datasets, the original acoustic measurements were centered and z-scored. In the linear discriminant analysis (LDA), the classifier was trained on the training set to distinguish dental vs. palatal sibilants using the acoustic measurements of COG, Dispersion, Intensity, and Duration.[26] Then, the trained classifier was adopted to categorize the pre-[i] dental /s/ in the test set as being

[26] A linear discriminant analysis creates one or more linear combinations of predictors to differentiate two (or more) groups. As to be shown later, an acoustic measurement may involve a significant difference between the dental vs. palatal sibilants, yet it may not be a significant predictor in the classifer built in the linear discrimination analysis.

dental or palatal. If palatalization has applied to the pre-[i] dental sibilants in Xiangtan, then it is predicted that some of the pre-[i] /s/ tokens would be classified as palatals.

In the training phase, a discriminant analysis was performed using the *greedy.wilks* function in the r package klaR (Weih, Ligges, Luebke, Raabe 2005) with stepwise variable selection. The significance cutoff point was set as .05 for the p-value of the F-statistic. Information of the classifier function obtained from the training is given in (48) below. The classifer in (48a) shows that, in distinguishing /s/ vs. /ɕ/, COG is the most important predictor, followed by Intensity, and then Disperion. More specifically, /ɕ/ is characterized by a lower COG, higher intensity, and lower dispersion compared with /s/. The overall accuracy of the classifier on the training set was 96%, and the predictions of dentals and palatals were equally good, as shown in (48b).

(48) Classifier function obtained from the training set: Fricatives

a. Classifier:

Place ~ (-1.78)*COG + 0.36*Intensity + (-0.24)*Dispersion

b. Performance of discrimination on the training set:

	Dental token	Palatal token
Classified as dental	64	3
Classified as palatal	2	63

c. Overall accuracy of discrimination on the training set:

96%

The classifier function was then used to predict the place of the tokens in the test set, i.e., the pre-[i] /s/ sounds in [si]. The 11 native speakers of Xiangtan Chinese each produced 6 tokens of /si/, thus a total of 66 tokens were categorized by the classifier as either dental or palatal. It turned out that of the 66 /s/ tokens in the [_i] context, 39 were classified as dentals and 27 as palatals. That is, 41% (= 27/66) of the pre-[i] /s/s were classified as palatal sibilants.

The results of classification for the 11 speakers are plotted in Figure 17, where the x-axis indicates the 11 speakers and the y-axis indicates the number of tokens for each speaker. In Figure 17, dark color marks the tokens classified as a dental sibilant and light color marks tokens classified as a palatal sibilant.

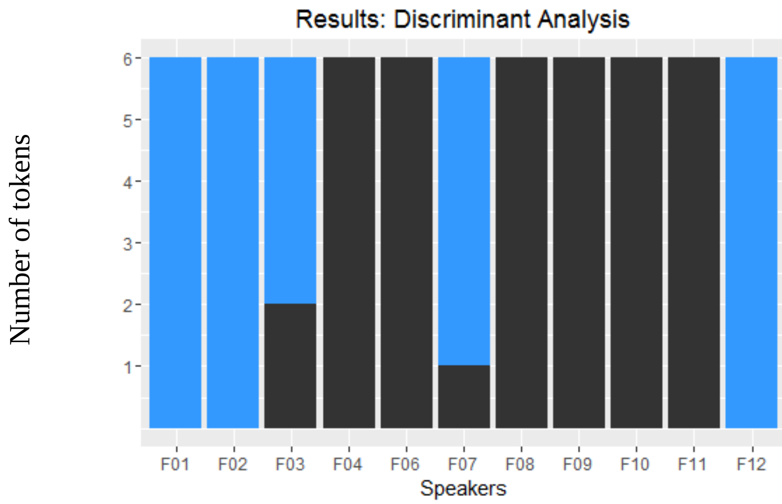


Figure 17 Predicted place of the pre-[i] sibilant: Fricatives.
(Dark = predicted as dental; light/blue = predicted as palatal)

The results in Figure 17 show that there are three types of speakers: First, those who did not palatalized their /s/ in the [_i] context, like F04; second, those who have palatalized some of their /s/ in the [_i] context, like F03; and third, those who have palatalized all their /s/ in the [_i] context, like F01. Note that all these three types of speakers produce /ɕ/ in the [_i] context as palatal sibilants. Focusing on the contrasts between detal vs. palatal sibilants in the [_i] context, the three groups of speakers were named as below in (49): *Contrast preservers* like F04, who did not palatalize any of her /s/s in the [_i] context and thus preserved the dental vs. palatal contrast in this context; *Partial palatalizers* like F03, who palatalized some of her /s/s in the [_i] context, thus were in the process of neutralizing the dental vs. palatal contrast in this context; *Complete* *palatalizers* like F01, who palatalized all of the /s/s in the [_i] context, thus had completed the neutralization.

(49) Difference in the sibilant onsets: Fricative

Speaker types	Phonetic forms of /si/ syllables	Phonetic forms of /ɕi/ syllables	Speakers
Contrast preserver	si	ɕi	F04, F06, F08, F09, F10, F11
Partial neutralizer	si~ ɕi	ɕi	F03, F07
Complete neutralizer	ɕi	ɕi	F01, F02, F12

Overall, the patterns in (49) showed that, although not all speakers showed a sign of neutralizing /s/ and /ɕ/ in the [_i] context, some speakers did, which confirms the prediction that a sound change neutralizing the pre-[i] dental /s/ and the palatal [ɕ] is in progress.

4.2.5.2. Unaspirated affricates

The acoustic measurements of the unaspirated affricates (4.2.2.2) were divided into two subsets:

(a) A *training* set including the measurements of COG, Dispersion, Intensity, and Duration for the 66 canonical /ts/s in [tsɹ̩] and the 42 canonical /tɕ/s in [tɕi]. Note that 24 /tɕi/ tokens were excluded, i.e., the ones from the speakers F06, F09, F10, and F11, who produced /ki/ variants for the /tɕi/ syllables (c.f. 4.2.2.2).[27]

(b) A *test* set including the measurements of COG, Dispersion, Intensity, and Duration of the 66 pre-[i] dental /ts/s in [tsi].

[27] As noted in 4.2.4.2, F06 and F09 produced all /tɕi/ as [ki], while F10 and F11 produced some of /tɕi/s as [ki]. It is unclear, for F10 and F11, whether the [tɕi] they produced can be considered as canonical or not. To avoid arbitrariness in the decision, all 24 tokens of /tɕi/ from these speakers were excluded in the LDA analysis.

Centering and scaling were performed on the original acoustic measurements before the linear discriminant analysis (LDA). Similar to the LDA analysis of the fricatives above, the classifier was trained on the training set to distinguish dental vs. palatal sibilants using COG, Dispersion, Intensity, and Duration. The details of the classifier are given in (50). Below, (50a) shows that, in distinguishing /ts/ vs. /tɕ/, COG is the most important predictor, followed by Disperion. More specifically, /tɕ/ is characterized by a lower COG and a lower dispersion compared with /ts/.

(50) Classifier function obtained from the training set: Unaspirated affricates

a. Classifier:

Place ~ (-1.56)*COG* + (-0.60)**Dispersion*

b. Performance of discrimination on the training set:

	Dental token	Palatal token
Classified as dental	64	12
Classified as palatal	2	30

c. Overall accuracy of discrimination on the training set:

87%

As (50) shows, the overall accuracy of the classifier on the training set was 87%. A closer look shows that the classification for canonical /ts/ was very acccurate (64/66 = 97%) while that for canonical /tɕ/ was less so (30/42 = 71%). To examine the lower accuracy in the prediction of palatal sibilants, the acoustic properties of the canonical /ts/ tokens and canonical /tɕ/ tokens were examined. Note that the classifier in (50a) has COG and Dispersion as the two predictors, therefore these two acoustic measurements were examined for dental vs. palatal sibilants, as plotted in Figure 18. In Figure 18, the raw measurements of COG and Dipersion were scaled and centered, and D and P represented ‘Dental’ and ‘Palatal’ respectively, i.e., the membership of a token as a canonical dental sibilant (D) or a canonical palatal sibilant (P). The upper light/blue region is what the classifer recognized as ‘dentals’ in the two-dimensional space of COG and Dispersion; the lower dark/purple region is what the classifier recognized as ‘palatals’. Close to the boundary in the middle, it can be seen that a number of canonical palatal sibilants (P-tokens in red-color) were categorized as dental sibilants based on their COG and Dispersion.

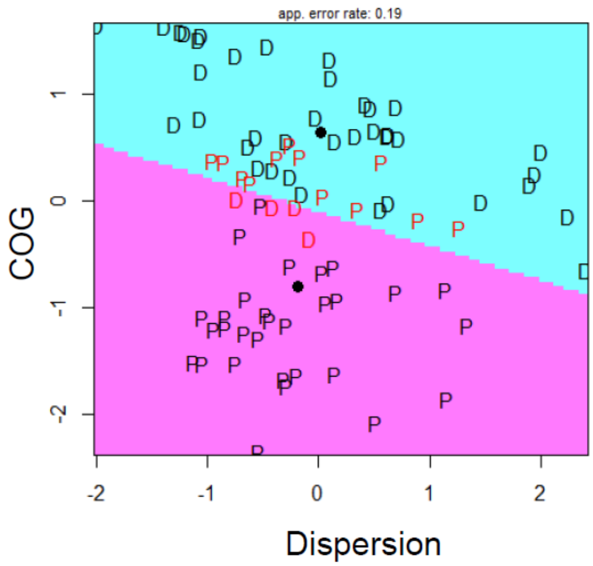


Figure 18 Classification of canonical /ts/ and /tɕ/ in the training set.

The classifier function in (50a) was then used to predict the place of the 66 tokens of the pre-[i] /ts/. Out of the 66 tokens /ts/ in the [_i] context, 14 tokens (= 21%) were classified as a palatal sibilant. The classification results for the 11 speakers are plotted in Figure 19, where the x-axis indicates the 11 speakers and the y-axis indicates the number of tokens for each speaker; dark color marks the tokens classified as dental and light color marks tokens classified as palatal.

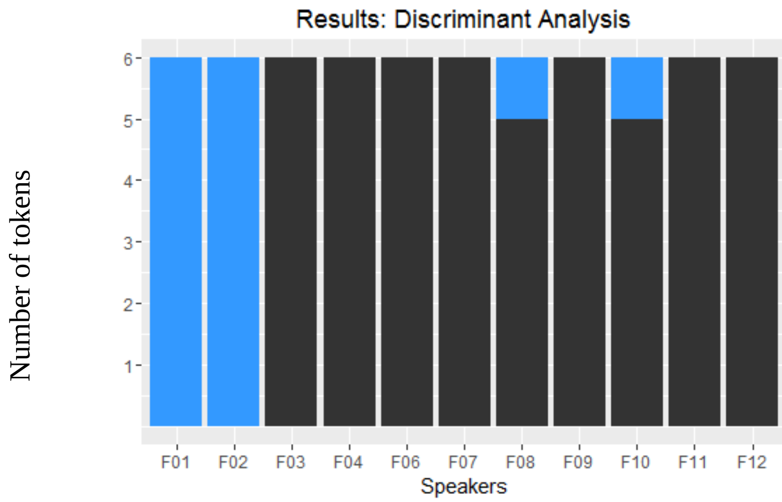


Figure 19 Predicted place of the pre-[i] sibilants: Unaspirated affricates.
(Dark = predicted as dental; light/blue = predicted as palatal)

Out of the 66 tokens of pre-[i] /ts/, 52 were classified as dentals and 14 as palatals.[28] A quick look at Figure 19 shows the same three types of speakers as for the fricatives in (49), i.e., *Contrast preservers* like F03, *Partial neutralizers* like F08, and *Complete neutralizers* like F01. However, as noted in 4.2.4.2, there were four speakers who produced velar stop variants for the palatal sibilants, F06, F09, F10, and F11. Taking this into consideration, there turned out to be a more nuanced pattern across the speakers in the contrast between dental and palatal unaspirated affricates in the [_i] context, as summarized in (51). The four speakers with velar variants of the palatal sibilants fell into three types: F06 and F09 produced all their palatal sibilants as velar stops and all their pre-[i] dentals as dentals; F11 had variation between velar stops and palatal sibilants and produced all her pre-[i] dentals as dentals; F10 had variation between velar stops

[28] As shown above, the classifier was more accurate in predicting dentals than in predicting palatals, i.e., the classifier has a 'dental bias'. Therefore, an unbiased classifier (presumably obtainable from more training data) may have recognized more tokens in the test set as palatals than reported in this study.

and palatal sibilants and variation between dental and palatal sibilants.[29] The other 7 speakers fell into the same three types as in (49), i.e., *Contrast preservers*, *Partial neutralizers,* and *Complete neutralizers*, in the [_i] context. Therefore, there are six types of speakers in (51).

(51) Difference in the sibilant onsets: Unaspirated affricates

Speaker types	Phonetic forms of /tsi/ syllables	Phonetic forms of /tɕi/ syllables	Speakers
No palatal speaker	tsi	ki	F06, F09
Variant palatal speaker 1	tsi	tɕi ~ ki	F11
Variant palatal speaker 2	tsi~ tɕi	tɕi ~ ki	F10
Contrast preserver	tsi	tɕi	F03, F04, F07, F12
Partial neutralizer	tsi~ tɕi	tɕi	F08
Complete neutralizer	tɕi	tɕi	F01, F02

Generally, the pattern in (51) shows that there is a sign of the neutralization between the dental /ts/ and the palatal /tɕ/ in the [_i] context, as shown by the results of the *Variant palatal speaker 2* (F10), *Partial neutralizers* (F08), and *Complete neutralizers* (F01, F02). On the other hand, palatalization of the dental /ts/ was not observed in *No palatal speakers* (F06, F09), *Variant palatal speaker 1* (F11), and *Contrast preservers* (F03, F04, F07, F12).

4.2.5.3. Aspirated affricates

Similar to the analysis above, the measurements of the aspirated affricates (4.2.2.3) were divided into the following two subsets:

[29] For the four speakers F06, F09, F10 and F11, the onsets in the /tsi/ syllables are all recognized as the dental sibilant, except that one token from F10 was recognized as a palatal sibilant by the classifier. Condidering that some of their /tɕi/ tokens are produced as /ki/, the four speakers are in fact preserving the contrasts of a dental sibilant /ts/ vs. a velar stop /k/.

(a) A *training* set including the measurements of COG, Dispersion, Intensity, and Duration for the 66 canonical /tsʰ/ in [tsʰɹ̩], and the 42 canonical /tɕʰ/ in [tɕʰi], excluding the 24 /tɕʰi/ tokens from F06, F09, F10, and F11, who produced /kʰi/ variants for /tɕʰi/ (c.f. 4.2.2.3).[30]

(b) A *test* set including the measurements of COG, Dispersion, Intensity, and Duration of the 66 pre-[i] dental /tsʰ/ in [tsʰi].

Centered and z-scored measurements were submitted to the linear discriminant analysis (LDA). The details of the classifier are given in (52). Below, (52a) shows that, in distinguishing /tsʰ/ vs. /tɕʰ/, COG is the most important predictor, followed by Disperion. More specifically, /tɕʰ/ is characterized by a lower COG and a lower dispersion than /tsʰ/.

(52) Classifier function obtained from the training set: Aspirated affricates

a. Classifier:

Place ~ (-1.79)**COG* + (-0.50)**Dispersion*

b. Performance of discrimination on the training set:

	Dental token	Palatal token
Classified as dental	64	6
Classified as palatal	2	36

c. Overall accuracy of discrimination on the training set:

93%

As shown in (52), the overall accuracy of the classifier on the training set was 93%. The classification of the canonical dental /tsʰi/ was highly accurate (64/66= 97%) while that of the canonical palatal /tɕʰi/ was somewhat less so (30/42= 86%), though more accurate than the classifier function built to predict /tɕ/ in 4.2.5.2.

[30] As noted in 4.2.4.3, F06 produced all /tɕʰi/ as [kʰi], while and F09, F10 and F11 produced some of their /tɕʰi/s as [kʰi]. For similar consideration in the last footnote, all the 24 tokens of /tɕʰi/ were excluded in the analysis.

Again, the classifier function was used to predict the place of the 66 tokens of the pre-[i] /tsʰ/. Out of the 66 tokens /tsʰ/ in the [_i] context, 18 tokens (= 27%) were classified as a palatal sibilants. The overall results for the 11 speakers are plotted in Figure 20, where the x-axis indicates the 11 speakers and the y-axis indicates the number of tokens for each speaker.

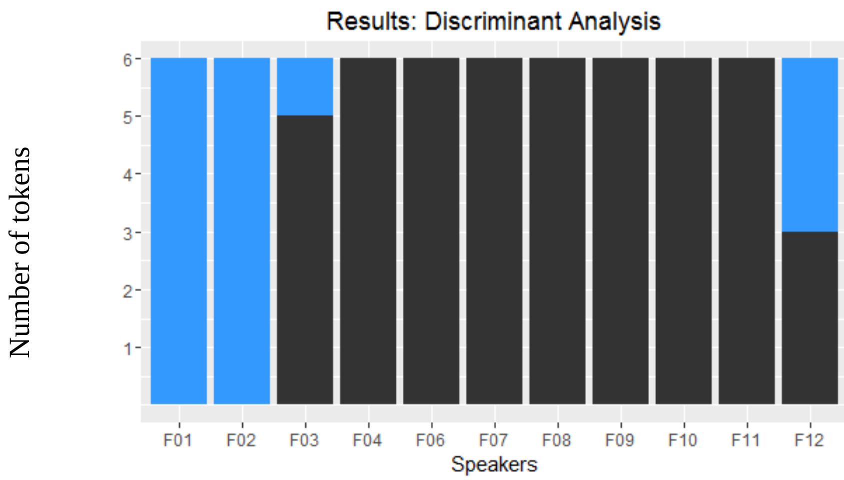


Figure 20 Predicted place of the pre-[i] sibilant: Aspirated affricates.
(Dark = predicted as dental; light/blue = predicted as palatal)

Out of the 66 tokens of pre-[i] /tsʰ/, 50 were classified as dentals and 16 as palatals. Note again that, as stated in 4.2.4.3, four speakers' production of palatal sibilants involved variants as velar stops, namely F06, F09, F10, and F11. Taking their details into consideration, the 11 speakers' pattern is summarized in (53). There turned out to be five types of speakers: *No palatal speakers* like F06 produced all her palatal sibilants as velar stops and all her pre-[i] dentals as dentals; *Variant palatal speakers* like F09 had variation between velar stops and palatal affricates and produced all her pre-[i] dentals as dentals; *Contrast preservers* like F04 preserved the contrastive dental vs. palatal sibilants in the [_i] context; *Partial neutralizers* like F03 had palatal sibilants and produced some of their pre-[i] dentals as palatals; *Complete neutralizers* like

F01 produced both pre-[i] dentals and palatals as palatals. Therefore, the pattern in (53) generally confirmed the prediction that a sound change neutralizing the dental /ts/ and the palatal [tɕ] in the [_i] context is in progress.

(53) Difference in the sibilant onsets: Aspirated affricates

Speaker types	Phonetic forms of /tsʰi/ syllables	Phonetic forms of /tɕʰi/ syllables	Speakers
No palatal speaker	tsʰi	kʰi	F06
Variant palatal speaker	tsʰi	tɕʰi ~ kʰi	F09, F10, F11
Contrast preserver	tsʰi	tɕʰi	F04, F05, F07, F08
Partial neutralizer	tsʰi~ tɕʰi	tɕʰi	F03, F12
Complete neutralizer	tɕʰi	tɕʰi	F01, F02

Generally, the pattern in (53) shows that there is a sign of the neutralization between /tsʰ/ and /tɕʰ/ in the [_i] context, as shown by the results of the *Partial neutralizer* (F03, F12) and *Complete neutralizers* (F01, F02).

For /s/, /ts/, and /tsʰ/ in the [_i] context, the discrimination analyses above showed that across the three manners of articulation, there is generally a sign of neutralization between the dental and palatal sibilants. This supports the prediction that the pre-[i] dental sibilants are in the process of being palatalized in Xiangtan Chinese. Variation in palatalization was observed in the production of dental sibilants in the [_i] context across different speakers, e.g., F01 was a complete neutralizer and F04 was a contrast preserver. On the other hand, variation in palatalization was also observed within speakers, e.g., F03's production of /s/ and /tsʰ/ in the [_i] context showed variation between dental and palatal sibilants.

4.3 Perceptual experiment: Sibilant identification

The acoustic measurements and the linear discrimination analyses have shown that some of the dental sibilants in the [_i] context were indistinguishable from palatal sibilants, based on their acoustic properties (i.e., COG, dispersion, intensity, and duration). In this section, a forced-choice identification experiment was conducted to examine the perceptual properties of the [s ts tsʰ] extracted from the [_i] context. In particular, the perceptual experiment aims to examine whether some of the dental sibilants in the [_i] context are perceptually equivalent to palatal sibilants and to verify the results of the discriminant analyses.

4.3.1 Method

4.3.1.1 Listeners

The aim of the identification experiment was to examine whether the dental sibilants in the [_i] context in Xiangtan are perceived as palatal sibilants. In this study, Mandarin native listeners were recruited as the participants based on a number of considerations:

(a) Mandarin Chinese has the same sibilant inventory as Xiangtan Chinese, i.e., both languages have dental sibilants [s ts tsʰ], palatal sibilants [ɕ tɕ tɕʰ], and retroflex sibilants [ʂ tʂ tʂʰ] in their phonetic inventory.

(b) Dental vs. palatal sibilants in Xiangtan Chinese and Mandarin are acoustically similar in terms of COG. In Mandarin, the COG of dental [s ts tsʰ] is typically around 6000–7000 Hz and that of palatal [ɕ tɕ tɕʰ] is around 5000–6000 Hz (Svantesson 1986). As shown in Figure 12, Figure 14, and Figure 16, the dental and palatal sibilants in Xiangtan Chiense generally differ the same way as their counterparts in Mandarin.

(c) Mandarin listeners have been shown to treat dental vs. palatal sibilants as different categories (Lu 2014), even though they are not contrastive in the [_i] context. In the identification experiment, the listeners heard only the first half of the sibilant, so the vowel context should not matter to Mandarin listeners.

Taking these considerations into acount, the use of Mandarin listeners should give a reliable identification of the place of articulation for a particular sibilant given the lack of vowel context in the stimuli and the acoustic similarity of the sibilants in Xiangtan and Mandarin.

Twenty native Mandarin speakers were recruited as the listeners in the identification experiment. These 20 listeners were undergraduate students at the University of Kansas and all of them knew English. Each listener completed a language background questionaire. All listeners reported that they started speaking Mandarin in early childhood, though they were from different geographic regions in China.[31] No hearing impairment was reported among these 20 listeners.

4.3.1.2 Stimuli

The stimuli in this experiment were the dental vs. palatal sibilants in the [sɹ̩ tsɹ̩ tsʰɹ̩], [ɕi tɕi tɕʰi], and [si tsi tsʰi] syllables in Xiangtan Chinese, shown in (54). More specifically, the stimuli consisted of two sets: (i) the *screening stimuli* of 18 canonical dental sibilants extracted from [sɹ̩ tsɹ̩ tsʰɹ̩] and 18 canonical palatal sibilants extracted from [ɕi tɕi tɕʰi], which were randomly selected from the caononical dental and palatal tokens from three female speakers; (ii) the *test stimuli* included 198 pre-dental sibilants extracted from [si tsi tsʰi]. The screening stimuli were

[31] In addition to the questionaire, each listener was also asked the produce all of the Mandarin sibilants in [sɹ̩ tsɹ̩ tsʰɹ̩], [ɕi tɕi tɕʰi], and [ʂɻ̩ tʂɻ̩ tʂʰɻ̩] to further verify his/her ability to produce the sibilants at different places. As a native speaker of Mandarin, the author of this dissertation observed that all the 20 listeners could produce the dental, palatal, and retroflex sibilants in Mandarin correctly.

used to test the listener's ability to accurately identify the dentals vs. palatals, and the test stimuli were used to test the perceived place (dental vs. palatal) of the pre-[i] dental sibilants. To be consistent with the acoustic measurements and the linear discrimination analyses, only the first half of the sibilants was used as the stimuli in the identification experiment.[32]

(54) Stimulus sibilant (first half) used in the identification experiment

	Sibilants extracted from the syllables	# of tokens
Screening stimuli	[s]ɹ̩ [ts]ɹ̩ [tsʰ]ɹ̩	6 tokens × 3 sibilants
	[ɕ]i [tɕ]i [tɕʰ]i	6 tokens × 3 sibilants
Test stimuli	[s]i [ts]i [tsʰ]i	66 tokens × 3 sibilants

4.3.1.3 Procedure

The experiment was conducted using Paradigm (Perception Research Systems 2007). The listeners were told that they would listen to speech sounds and they should identify a sound as a Mandarin onset consonant by choosing from two options. More specifically, on hearing a stimulus, a listener was presented with two choices on the computer screen, as shown in Figure 21. The two choices were in Chinese Pinyin with accompanying Chinese characters, one corresponding to a dental sibilant and another corresponding to a palatal sibilant. In Figure 21, for example, 's' is the Pinyin form of the dental sibilant [s] and '思' is a Chinese word with the pronunciation [sɹ̩]; 'x' is the the Pinyin form of the palatal sibilant [ɕ] and '西' is a Chinese word with the pronunciation [ɕi]. Color coding was used for dental vs. palatal choices: The choice 's (思)' appeared in blue color and the choice 'x (西)' appeared in red color.

[32] The stimulus sounds were the first half of the sibilants as they were extracted and thus all the stimulus sounds ended abruptly without a fade-out phase.

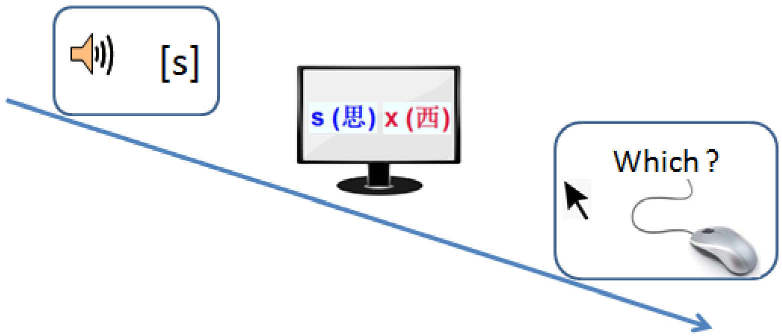


Figure 21 Procedure of the identification experiment.

The experiment consisted of two parts, a *screening test* and the main *identification test*. In the screening test, the stimulus sibilants were the screening stimuli in (54). To ensure the accuracy of the place identification in the main experiment, a listener's data were excluded if his/her accuracy in the screening test was lower than 90% (i.e., if 4 or more errors were made for the 36 tokens). In the main identification test, the stimuli were the test stimuli in (54).

The main experiment consisted of three blocks and the stimuli were [s] [ts] [ts^h] respectively in each block. The listeners had the choice to take a short break between blocks. The whole experiment, including a questionnaire on language background, lasted for about 20 minutes.

4.3.2 Results

4.3.2.1 Screening test

Holding to an accuracy threshold of 90%, a total of 10 listeners passed the screening test. In other words, 10 out the 20 listeners served as the listeners in the subsequent identification task. One potential explanation for the relatively low passing rate in the screening is that the task was difficult, as the listeners needed to give the correct identification of the place of a sibilant when they only heard the first half of the sibilant. The mean accuracy for the 10 listeners was 95.6%

and the lowest accuracy was 93%.

A closer examination of the 10 listeners showed that the listeners generally had a higher accuracy for dental sibilants than for the palatal sibilants in the screening test. All the dental sibilants were correctly identified by the 10 listeners, except for 1 error made by listener LF15; almost all the errors were made in the identification of the palatal sibilants. Therefore, despite the relatively small number of tokens, there was a tendency for a ‘dental bias’ in the 10 listeners’ identification.

One explanation for this ‘dental bias’ can be found in the variability of COG in Mandarin sibilants. For example, for three of the four male Mandarin speakers in Svantesson (1986), the dental vs. palatal sibilants were clearly separated by COG, with dentals around 6000–7000Hz and palatals around 5000–6000Hz; for the other speaker, the COG of the dental sibilants was around 5000Hz, overlapping with that of the palatal sibilants. For female Mandarin speakers in particular, S. Lee (2011) reported COG values around 8000–10000Hz. In other words, Mandarin dental sibilants may have a larger COG variance (5000–10000Hz) compared with the palatal sibilants. Therefore, a native Mandarin speaker aware of this variance may be biased to identify a palatal sibilant as a dental one if its COG is around 6000Hz.[33]

[33] Interestingly, the dental bias in the 10 listeners’ responses is inconsistent with the phoneme probability of dental vs. palatal sibilants in Mandarin. For example, F. Li (2008) reported that Mandarin dental /s/ is lower in phoneme frequency than the palatal /ɕ/.

4.3.2.2 Main identification task

For each token, the 10 listeners each gave a response as either dental or palatal, thus the 10 listeners gave a total of 1980 responses to the dental sibilants [s] [ts] [ts^h] extracted from the 11 speakers' production of [si] [tsi] [ts^hi]. This led to a total of 60 responses respectively for the [s] [ts] [ts^h] produced by one speaker of Xiangtan Chinese. For example, speaker F01 produced 6 tokens of [s] and the 10 listeners each gave a response to these 6 tokens. For each speaker then, the rate of 'dental responses' was calculated respectively for [si] [tsi] [ts^hi], by dividing the total number of dental response by 60 (i.e., the responses as dental and palatal together). Out of the 660 tokens of /s/ in the [_i] context, a total of 274 tokens (= 41.5%) were identified as palatal; out of the 660 /ts/ tokens in the [_i] context, 234 tokens (= 35.5%) were identified as palatal; out of the 660 /ts^h/ tokens in the [_i] context, 356 tokens (= 54.9%) were identified as palatal. The results of the mean rates of dental responses are plotted in Figure 22, where dark color marks the rate of dental responses and light color corresponds to the rate of palatal responses.

As shown in Figure 22, for the dental sibilants in the [_i] context, the Mandarin listeners identified most of them as dentals and some of them as palatals. Across-speaker variation was observed: some speakers' pre-[i] dentals were all perceived as palatals, e.g. F01, but the same was not true for other speakers. Within-speaker variation was also observed: for some speakers, some of the produced tokens were generally perceived as dentals, but the same was not true for other speakers.

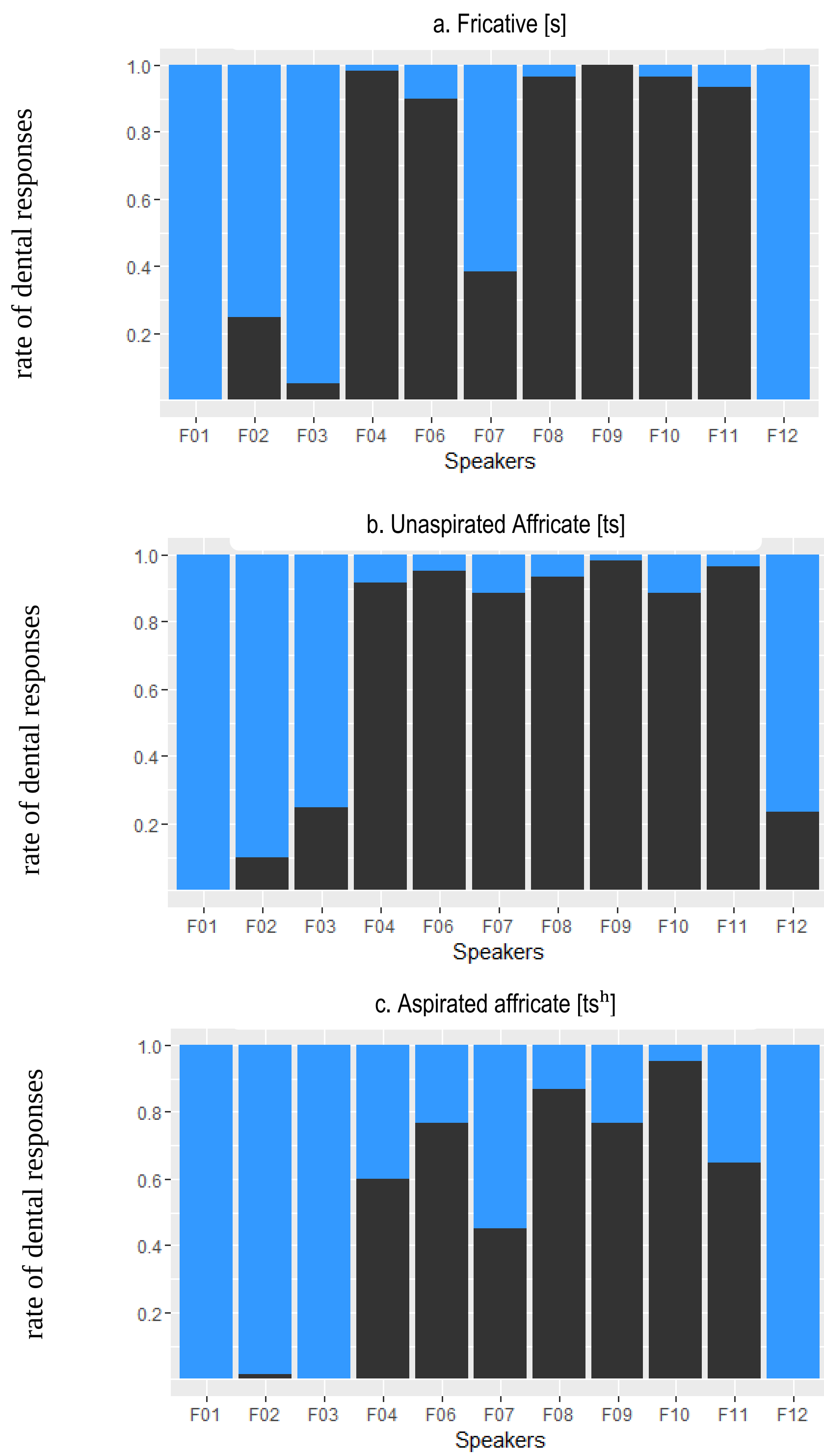


Figure 22 Rate of dental responses in the identification experiment.
(Dark = dental; light/blue = palatal)

4.3.3 Identification and acoustic properties

The results of the listeners' identification are similar to those in the linear discriminant analyses, which are based on the acoustic properties of the sibilants, in that both analyses showed that some of the dental sibilants in the [_i] context are recognized as palatal sibilants. In this section, the relation between the acoustic properties of sibilants and the listeners' identification is examined. A series of Mixed Effects Logistic Regression was run separately for the identification of [s], [ts], and [ts^h] in the [_i] context, using the *glmer* function in the r package lme4 (Bates, Maechler, Bolker and Walker 2015a, b). For each analysis, the dependent variable, predicting variables and random variables are listed in (55). For each predicting variable in (55), the original measurements were centered and z-scored.

(55) Variables in the analysis of the listeners' identification

Variable names	*Notes*
Dependent variable	
Identification	0 = Dental, 1 = Palatal as identified by a listener
Predicting variables	
COG	Center of gravity (spectral mean) over the first half of the sibilants
Dispersion	Spectral variance over the first half of the sibilants
Intensity	Intensity over the first half of the sibilants
Duration	Duration of the first half of the sibilants
Random variable	
Listener	10 listeners

Backward elimination of non-significant effects was performed in the analysis of each set of data. More specifically, a full model with COG, Dispersion, Intensity, and Duration as the predicting variables and Listener as the random variable was compared separately with four subset models each eliminating one predicting variable.[34] A predicting variable was kept in the final model if there is a significant difference between the superset model and the corresponding subset model. For the identification of [s], for example, it turned out that COG and Duration were both significant predictors (COG: $X^2 = 516.11$, $df = 1$, $p < .001$; Duration: $X^2 = 40.983$, $df = 1$, $p < .001$) but not Dispersion or Intensity. The final model then included COG and Duration as the predicting variable and Listener as the random variable.

The final models for the identification of [s], [ts], and [ts^h] are listed in Table 46 (dental responses were encoded as 0 and palatal responses were encoded as 1). As shown in Table 46, palatal responses are more likely to be given to the fricative tokens with a lower COG and a longer duration, to the unaspirated affricate tokens with a lower COG, narrower dispersion, higher intensity, and shorter duration, and to the aspirated affricate tokens with a lower COG, higher intensity, and longer duration.

[34] To match the linear discriminat analyses, the statistical analysis did not involve interaction terms between any predicting variables.

Table 46 Fixed effects in the final models for the listeners' identification

a. Fricative (/s/i)

	Estimate	*Std.Error*	*z value*	*Pr(>|t|)*
(Intercept)	–8.86	0.194	–4.422	< .001***
COG(z score)	–4.26	0.369	–11.546	< .001***
Duration(z score)	1.33	0.193	6.914	< .001***

Model: Identification ~ (-4.26)*COG + 1.33*Duration + (1|Listener)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

b. Unaspirated affricate (/ts/i)

	Estimate	*Std.Error*	*z value*	*Pr(>|t|)*
(Intercept)	–0.83	0.160	–5.157	< .001***
COG(z score)	–1.71	0.147	–11.654	< .001***
Dispersion(z score)	–0.87	0.151	–5.719	< .001***
Intensity(z score)	0.76	0.145	5.277	< .001***
Duration(z score)	–0.24	0.117	–2.080	.038**

Model: Identification ~ (-1.71)*COG + (-0.87)*Dispersion + (0.76)*Intensity + (-0.24)*Duration + (1|Listener)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

c. Aspirated affricate (/ts^h/i)

	Estimate	*Std.Error*	*z value*	*Pr(>|t|)*
(Intercept)	1.28	0.296	4.307	< .001***
COG(z score)	–3.66	0.338	–10.803	< .001***
Intensity(z score)	0.50	0.118	4.232	< .001***
Duration(z score)	0.41	0.124	3.319	< .001***

Model: Identification ~ (-3.66)*COG + 0.50*Intensity + 0.41*Duration + (1|Listener)
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05

4.3.4 Linear discriminant analysis vs. listeners' identification

The linear discriminant analysis and the listeners' identification both focused on the [s ts ts^h] extracted from [si tsi ts^{h}i]. To examine the correlation between the identification by the Mandarin listeners and the results from the linear discrimination analysis, the correlation between two measurements are examined for each token:[35]

[35] An alternative way to examine the relation between the results of the discriminant analysis and the listeners' identification is to use the results of the discriminant analysis as a predictor for the listeners' identification. But given that the results of both analyses were directly related to the acoustic properties of the sibilants, this section examines the correlation between the results of these two analyses.

(a) The probability of being recognized as a palatal sound in the results of the linear discriminant analysis. For each token, the linear discriminant analysis gives a categorization (e.g., the token 'F02_si_1' was classified as a palatal sound) as well as a group probability of the classification (e.g., for the token 'F02_si_1', the probability of being a palatal sound was 0.9794 and that of being a dental sound was 0.0206).

(b) The rate of palatal responses from the 10 listeners' identification. For each token, this rate was calculated by dividing the number of 'palatal' responses from the 10 listeners' identification by the total numbe of responses (e.g., for the token 'F02_si_1', 9 out of the 10 listeners identified it as a palatal sound, thus the rate was 0.90.)

Kendall rank correlations were calculated between these two measurements for each token, and separate calculation was conducted for the sibilants [s], [ts], and [ts^h]. For each dataset, the correlation turned out to be significant and positive, as shown in (56), which indicates that the results of the linear discriminant analysis are generally consistent with those from the listeners' identification.

(56) Kendall rank correlation coefficients of the palatal probability from linear discriminant analysis and the rate of palatal responses in the listeners' identification

a. [s]: $\tau = .714, p < .001$
b. [ts]: $\tau = .546, p < .001$
c. [ts^h]: $\tau = .662, p < .001$

A close comparison between the two analyses also revealed some detailed differences. Even though COG was the most important factor in both analyses, the classifier in the discriminant analysis differs from the model for the listeners' classification in terms of other factors. For example, for [s], the classifier obtained from the discriminant analysis was 'Place ~ (-1.78)*COG

+ 0.36*Intensity + (-0.24)*Dispersion’ (48a), while the model for the listeners’ identification was ‘Identification ~ (-4.26)*COG + 1.33*Duration’ (56a). Intensity and Dispersion were significant factors in the classifier but not in the model of the listeners’ identification while Duration was a significant factor in the the listeners’ model but not in the classifier. It is unclear why the two analyses differ in the contribution of these factors, given the similarity between Xiangtan and Mandarin in terms of the sibilant inventory and the acoustic properties of dental and palatal sibilants (4.3.1.1). A possible reason for the difference in the models of discrimination and lidentification could be that, in doing the identification task, the Mandrin listeners might have referred to their knowledge about the acoustic properties of Mandarin sibilants. For example, Feng (1985) reported that in normal speech the mean duration of Mandarin [s] (123 ms) is shorter than that of [ɕ] (134 ms), which is consistent with the observation that in the listeners’ identification, longer duration introduced more palatal responses. On the other hand, it is also possible that the listeners may pay attention to acoustic properties other than the four acoustic measurements in this dissertation, e.g., spectral peak location, skewness, and kurtosis (Jongman, Wayland, and Wong 2000, McMurray and Jongman 2011, among others). Future research is needed to investigate this issue.

4.4 Discussion

4.4.1 Neutralization of dental vs. palatal sibilants

The typological survey of sibilant place contrasts, the historical development of Mandarin sibilants, and the perceptual experiment reported in Chapter 3 all suggest that dental vs. palatal sibilants in the [_i] context are unstable contrasts and thus tend to be avoided in a sound system. Based on the historical developments of Mandarin Chinese, it was predicted that for a sound system like Xiangtan Chinese, the dental vs. palatal sibilants in the [_i] context would likely be neutralized with the dental sibilants [s ts ts^h] being palatalized into [ɕ tɕ tɕ^h].

This prediction is borne out in both the linear discriminant analysis, which was based on the acoustic measurements of the dental vs. palatal sibilants, and the perceptual identification experiment, by which Mandarin listeners identified the place of the pre-[i] dental sibilants. In both, some of the [s ts ts^h] extracted from [si tsi ts^{h}i] were recognized as being palatal sibilants; the palatalized /s ts ts^h/ were observed across different speakers and within the production of the same speaker. This suggests that a sound change is in progress towards the neutralization of the dental vs. palatal sibilants in the [_i] context, duplicating the sound change of Mandarin between the 18th century and the 20th century.

4.4.2 Manner difference in the listeners' identification

As shown in Figure 22, the listeners recognized some of the dental sibilants as palatals, and a closer examination showed that the aspirated affricates introduced more palatal responses than the fricatives and the unaspirated affricates. This differs from the results of the linear discriminant analysis in Figures 18 and 19, which showed similar rates of palatalization between the unaspirated affricates and the aspirated affricates. The larger proportion of palatal responses for the aspirated affricates in the listeners’ identification can be accounted for as an effect of the listeners’ knowledge of the acoustic properties of the aspirated affricates and the nature of the stimuli used in the identification experiment.[36]

In the natural speech of Mandarin, the aspirated [ts^{h}] and [tɕ^{h}] both have two portions: the frication noise and the aspiration, as illustrated in Figure 23. In terms of frication noise, the COG of a [ts^{h}] sound was higher than that of a [tɕ^{h}] sound. Within a [ts^{h}] sound then, there is an abrupt drop of COG from the frication noise to the aspiration portion, as illustrated in Figure 23a; within a [tɕ^{h}] sound, the COG drop was less abrupt, as illustrated in Figure 23c. Thus, a native Mandarin listener may use COG drop as a perceptual cue to distinguish dental sibilants (a higher COG at the start followed by a more abrupt drop) and palatal sibilants (a lower COG at the start followed by a less abrupt drop) in addition to other perceptual cues.

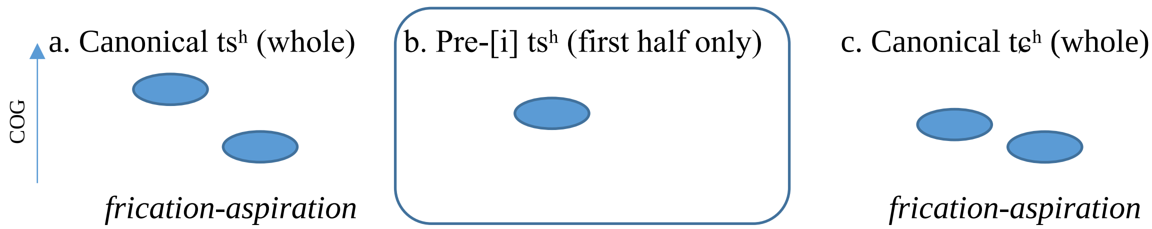


Figure 23 COG patterns in canonical ts^{h}, pre-[i] ts^{h} (first half), and canonical tɕ^{h}.

In the identification experiment, the stimuli only included the first half of the sibilants, mostly the frication noise of [ts^{h}] as extracted from [ts^{h}i] in Xiangtan Chinese, without the aspiration. As shown in the acoustic measurements, the COG of the [ts^{h}] before [_i] was generally lower than the COG of the canonical [ts^{h}] in [ts^{h}ɹ̩]. Therefore, compared with a whole

[36] Thanks to Charles Redmon for suggestions on the following analysis.

[tsʰ] sound, the pre-[i] [tsʰ] presented to the Mandarin listeners has two properties:

(a) The COG was lower than the COG of a canonical dental [tsʰ];

(b) There was no significant COG drop across the whole stimuli.

These two properties may have biased the listeners towards a palatal response in that both properties are cues to a palatal [tɕʰ], particularly when the listeners assumed that they were listening to a whole segment. Note that in the experiment, the listeners were not informed that the stimuli were the first half of the sibilants, thus they may have simply assumed that they were listening to a whole sibilant (presumably at a fast speech rate since the duration was short). In other words, in addition to (a) and (b) above, an additional assumption by the listeners was likely to be:

(c) I am listening to a whole consonant without a following vowel (in fast speech)

Therefore, when asked to judge whether the first half of [tsʰ] was an aspirated dental or palatal affricate, the listeners were more likely to be biased towards a palatal response because it would sound more likely to be an aspirated palatal [tɕʰ] than an aspirated dental [tsʰ]. Simply put, the bias above may have arisen from the listeners' knowledge of the acoustic properties of the canonical [tsʰ] vs. canonical [tɕʰ] in Mandarin.

Note that the bias sketched above should be assumed to be specific to the identification of [tsʰ] in the [_i] context, and the same effect may not be expected for the fricatives [s-ɕ] and the unaspirated affricates [ts-tɕ]. In the stimulus words, the COG values of the dental sibilants in the [_i] context were comparable across the three manners of articulations. For example, in the [_i] context, the dental [s ts ts^h] generally have a mean COG around 7000Hz, which are lower than the mean COG of canonical dental sibilants, i.e., 7000 – 10000Hz as reported for female Mandarin speakers (S. Lee 2011). However, the perception of canonical Mandarin [s] and [ts] does not involve an additional cue like COG-drop, which is available for the perception of canonical Mandarin [ts^h]. Therefore, when Mandarin listeners were listening to the pre-[i] /s/ and the pre-[i] /ts/ in the identification experiment, they may have assumptions (a) and (c), however, (b) would not be relevant the same way as when they were listening to the pre-[i] [ts^h]. In other words, the relevance vs. irrelevance of (b) may have contributed to the difference between the Mandarin listeners' identification of [s] and [ts] vs. that of [ts^h]. Simply put, with the additional cue (b), the Mandarin listeners were more biased towards a palatal response when listening to the pre-[i] /ts^h/ compared to when listening to the pre-[i] /s/ and the pre-[i] /ts/.[37]

4.5 Summary of the phonetic studies on Xiangtan sibilant contrasts

Based on the results of the perceptual experiment and the historical development of Mandarin sibilants, it was predicted that contrastive dental vs. palatal sibilants in the [_i] context form weak contrasts and are likely to be avoided in sound change. To test this prediction, the dental vs. palatal sibilants in Xiangtan Chinese, which were reported to fully contrast in the [_i] context, were investigated to examine if there is any sign of neutralization between dental and palatal sibilants in the [_i] context. More specifically, the acoustic measurements were made on the canonical dental sibilant, canonical palatal sibilants, and the dental sibilants in the [_i] context,

[37] The observed manner difference in the identification experiment (i.e. more palatal responses for [ts^h] than for [s] and [ts]) was thus specific to the design of the identification experiment in this study. It should not be taken as evidence as to whether [ts^h] is more likely to be palatalized than [s].

and then a series of linear discriminant analyses were conducted to investigate whether the pre-[i] dental sibilants were distinguishable from the canonical palatal sibilants. The results from the linear discriminant analyses showed that some of the dental sibilants in the [_i] context were indistinguishable from the canonical palatal sibilants based on the acoustic measurements of COG, Dispersin, Intensity, and Duration. The results were further verified by an identification experiment, in which Mandarin listeners perceived some of the pre-[i] dental sibilants as palatal sibilants. In short, the phonetic studies of Xiangtan Chinese showed that there is a sign of neutralization between the dental vs. palatal sibilants that were reported to fully contrast in the [_i] context. These results further support the contention that dental vs. palatal sibilants form weak contrasts in the [_i] context, and these contrasts are likely to be avoided in a sound system through sound changes.

Chapter 5. Contrast distinctiveness in sound systems

5.1 Summary of the two studies

The cross-linguistic typology in Lee-Kim (2014a) has shown that sibilant place contrasts in the [_i] context tends to be avoided, based on the survey of languages that have sibilants at three places of articulation. This typological generalization was examined in this dissertation across 170 Chinese dialects in two aspects:

(a) Sibilant place contrasts in the [_i] context in Chinese dialects that have sibilants at three places and in those that have sibilants at two places;

(b) Sibilant place contrasts across different manners of articulation: fricatives, unaspirated affricates, and aspirated affricates.

The typological survey across the 170 Chinese dialects showed that:

(a) Sibilant place contrasts in the [_i] context tend to be avoided in dialects that have sibilant at three places as well as in those that have only dental and palatal sibilants;

(b) For dental vs. palatal sibilants in the [_i] context, the place contrast of affricates implies that of fricatives.

The historical developments of sibilants in Mandarin Chinese verified the observation (a) above, and further showed that:

(c) The dental vs. palatal sibilant contrasts in the [_i] context once developed into the dental vs. retroflex contrasts.

Focusing on the perceptual properties of sibilant contrasts, three hypotheses were made based on the three observations above:

(a) *Hypothesis I* (Vowel effect): Dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts;

(b) *Hypothesis II* (Place effect): In the [_i] context, the place contrast of dental vs. retroflex sibilants is more distinct than the place contrast of dental vs. palatal sibilants;

(c) *Hypothesis III* (Manner effect): For dental vs. palatal sibilants in the [_i] context, the place contrast of affricates is less distinct than the place contrast of fricatives.

These hypotheses were tested in a speeded-AX discrimination experiment, the results of which confirmed the three hypotheses:

(a) For dental vs. palatal sibilants, the [_i] context generally introduced longer response times and more discrimination errors, thus reduced perceptual distinctiveness, than other vowel contexts;

(b) For the [_i] context, the place contrast of dental vs. palatal sibilants introduced longer response times and more discrimination errors, thus less distinctiveness, than the place contrast of dental vs. retroflex sibilants;

(c) For contrastive dental vs. palatal sibilants in the [_i] context, the place contrast of affricates introduced more discrimination errors, thus less distinctiveness, compared with the place contrast of fricatives.

Out of these conclusions from the perceptual experiments, the most important one is the vowel effect on the perceptual distinctiveness of dental vs. palatal sibilants. Based on the results of the perceptual experiment, the typology of sibilant place contrasts, and the historical development of Chinese sibilants, it is predicted that dental vs. palatal sibilants in the [_i] context form weak contrasts, which is likely to be avoided in a sound system. A phonetic study was conducted on the dental vs. palatal sibilants in Xiangtan Chinese, which were reported to fully contrast in the [_i] context, to examine if there is a sign of avoiding contrastive dental vs. palatal sibilants in the [_i] context. Acoustic measurements were made on the canonical dental sibilant, canonical palatal sibilants, and the dental sibilants in the [_i] context, and then the acoustic meansurements were submitted to a series of linear discriminant analyses to determine whether there is neutralization between the dental vs. palatal sibilants in the [_i] context. The results showed that some of the pre-[i] dental sibilants were in fact indistinguishable from the canonical palatal sibilants, and this observation was further verified by the results from an identification experiment by Mandarin listeners. Generally, the phonetic studies of dental vs. palatal sibilants in Xiangtan Chinese confirmed the prediction that the place contrasts between dental and palatal sibilants in the [_i] context are likely to be avoided, which further supports the contention that contrastive dental vs. palatal sibilants form weak contrasts in the [_i] context.

5.2 Contrast distinctiveness, typology, and sound change

In this dissertation, a striking connection was observed between the results of the perceptual study, the results of the production study, the patterns observed in the cross-linguistic typology, and the historical sound changes. In other words, there is a constraint against the existence of contrastive dental vs. palatal sibilants in the [_i] context. This constraint is grounded in the psychoacoustic similarity of the contrastive elements and it is manifested in the cross-linguistic typology and historical sound changes.

The connection between contrast distinctiveness and linguistic typology has long been noticed in the linguistic literature. It has been observed that languages may apply diverse phonological processes to avoid a perceptually weak contrast. For example, the voicing contrast between /k/ and /ɡ/ is avoided with the palatalization of /ɡ/ in Arabic, the nasalization of /ɡ/ in Japanese, the spirantization of /ɡ/ in Low German, and the spirantization plus pharyngealization of /ɡ/ in Czech, Slovak, Ukrainian (Boersma 1998:384-386). The typological survey in this study provides further illustration of this observation in that different Chinese dialects adopt different strategies to avoid contrastive dental vs. palatal sibilants in the [_i] context, e.g., vowel enhancement with the introduction of apical vowels after the dental sibilants, e.g., [sɹ̩-ɕi] in *Dayü* (L. Liu 1995), and phonotactic gap whereby the dental sibilants do not combine with a following [_i], e.g. *[si-ɕi] in *Shibei* (Hiroyuki 2004).

Studies in historical phonology have also shown that different sound changes in a language may achieve the same goal. The historical development of Mandarin contributes to the literature an additional case: The constraint against weak contrasts manifests itself through diverse sound changes across different historical stages to avoid contrastive dental vs. palatal sibilants in the [_i] context, e.g., the formation of apical vowel around the 7th century, the shift of sibilants around the 11th century, and the sibilant neutralization around the 18th century. Thus, the study in this dissertation strengthens the empirical connection between perceptual distinctiveness and historical phonology.

5.3 Perceptual vs. articulatory factors in language typology

Lee-Kim's (2014a) typological study centers on the place contrasts of sibilants at different places in the [_i] context. Her typological generalization was primarily drawn from a survey of the languages with three-way sibilant place distinction. The languages surveyed in her typology typically have a distinction among dental, palatal, and retroflex sibilants, and her typology does not examine languages with only a two-way sibilant place distinction. The typological survey in this dissertation differs from Lee-Kim's study in that:

(a) Two-way sibilant place contrasts in the [_i] contexts were examined as well as three-way sibilant place contrasts;

(b) Pair-wise place contrasts, e.g., dental vs. palatal, dental vs. retroflex, were investigated in languages with two sibilant places and three sibilant places alike.

5.3.1 Vowel effect as place-specific

The typological survey in this study showed that for dialects with a two-way sibilant place distinction, the contrasts are mostly between dental and palatal sibilants. When it comes to dental vs. palatal sibilants, there is a tendency to avoid contrastive dental vs. palatal sibilants in the [_i] context even when there are no retroflex sibilants in a sound system. As noted in 5.1, this typology is consistent with the results from the perceptual experiment, the historical development of Mandarin, and the phonetic studies of Xiangtan Chinese.

The typology of dental vs. palatal sibilants in the [_i] context to the exclusion of retroflex sibilants suggested that the reduced distinctiveness of sibilant place contrast in the [_i] context should be specific to sibilants contrasts between particular places. First, as shown in the results of the perceptual experiment, the [_i] context leads to reduced perceptual distinctiveness for dental vs. palatal sibilants, but not for the place contrasts of [s-ʂ] and [ts^h-tʂ^h]. Second, the tendency to avoid contrastive dental vs. palatal sibilants in the [_i] context is independent of whether retroflex sibilants exist in a sound system. In other words, the avoidance of contrastive dental vs. palatal sibilants in the [_i] context is primarily based on the psychoacoustic similarity between the two contrastive elements, oblivious of whether the language has a two-way or a three-way place distinction for sibilants.

5.3.2 Retroflex sibilants in the typology

The typological survey in this study investigated the patterns of three-way place contrasts of sibilants across 51 Chinese dialects (2.2.2). The general pattern is summarized in (57), where the dialects are grouped in terms of the presence/absence of place contrasts in the [_i] context. A closer look at (57) showed that a prominent factor contributes to the avoidance of three-way place contrasts, that is, retroflex sibilants generally do not combine with [_i]. Out of the 51 dialects in (57), for example, only 3 dialects allow the combination of retroflex with [_i], namely Qimen (T. Shen 1989), Datan (Y. Chen 2015), and Xuancheng (Shen and Huang 2015). Among the 17 dialects of Group 2 in (57), 16 dialects avoid the three-way contrasts exclusively with the non-combination of retroflex sibilants and [_i], e.g., ʂɻ̩ in Nanjing (D. Liu 1994) and *ʂi in Changshu (D. Yuan 2010). A similar pattern is observed in Lee-Kim's (2014a) generalization in (2), repeated as (58) below, which showed that retroflex sibilants tend not to be followed by a following vowel [_i]. For example, in the avoidance of three-way place contrast, most languages enhanced or avoided the combinations like ʂi, except Telugu.

(57) Typology of three-way sibilant place contrasts across 51 Chinese dialects

Group	*Contrast Pattern*	*Example dialect*
1. Three-way contrasts present (2)	sɹ̩ **si ɕi ʂi**	Qimen (T. Shen 1989)
2. Two-way contrasts present (17)		
a. Dental vs. palatal contrasts present (*16*)		
a1) Retroflex with vowel enhancement (*<u>13</u>*)	sɹ̩ **si ɕi** ʂɻ̩	*Nanjing* (D. Liu 1994)
a2) Phonotactic gap at the retroflex (*<u>3</u>*)	sɹ̩ **si ɕi** *ʂi	*Changshu* (D. Yuan 2010)
b. Palatal vs. retroflex contrast present (*1*)	sɹ̩ **ɕi ʃi**[38]	*Xuancheng* (Shen and Huang 2015)
3. Contrast avoided/enhanced (32)		
a. Dental/retroflex with vowel enhancement (*28*)	sɹ̩ ɕi ʂɻ̩	*Anqing* (H. Bao 2012)
b. Dental with vowel enhancement + Phonotactic gap at the retroflex (*2*)	sɹ̩ ɕi *ʃi[39]	*Dangtu* (Zhang, Yuan, and Shen 2012)
c. Dental with vowel enhancement + Phonotactic gap at the retroflex (*2*)	sɹ̩ *ɕi ʂɻ̩	*Ledu* (Cao and Shao 2001)

Notes: i) In the list above, the number in parentheses indicates the number of dialects.
ii) Most Chinese dialects have a three-way distinction of fricative, unaspirated affricates and aspirated affricates. A fricative above thus generally represents the three relevant sibilants.
iii) As shown in Table 5, there are seven dialects like *Daoxian* (K. He 2003) and *Datan* (Y. Chen 2015), where the affricates do not behave the same as their fricative counterparts in terms of their place contrasts in the [_i] context. This detail is not reflected in (57).

(58) Recall the reduction of sibilant contrast in the /_i/ context – Repeated from (2) (Lee-Kim 2014a)

Language (Family)	*Sibilants*	*Contrast before /i/*	*Phonological pattern*
Acoma (Kersan)	s-ɕ-ʂ	s-ɕ	ʂ→s or ɕ (more) / _i
Chacobo (Panoan)	s-ʃ-ʂ	s-ʃ	ʂ→ ʃ / _i
Cashinahua (Panoan)	s-ʃ-ʂ	s-ʃ	*ʂi
Telugu (Dravidian)	s-ɕ-ʂ	ɕ-ʂ	s→ɕ/_i/e (optional)

For the non-combination of a retroflex consonant and a following vowel [i], there have been a number of analyses in terms of articulatory incompatibility and perceptual salience. In terms of articulatory gesture, Pulleyblank (1984:25-26) suggested that a retroflex consonant involves a

[38] Shen and Huang (2015) used the symbol [ʃ] and they did not explicitly state whether [ʃ] refers to a retroflex sibilant. Considering the existence of the palatal [ɕ] in the consonant system of *Xuancheng*, it is likely that the [ʃ] is at least close to a retroflex sibilant.
[39] Similar to the footnote above, Zhang, Yuan, and Shen (2012) did not explicitly state whether [ʃ] is a retroflex sibilant, though it is likely that it is due to the existence of [ɕ] in the sound system.

non-front constriction and is thus articulatorily incompatible with a following [_i], which may lead to a tendency to eliminate high front vowel after the retroflex consonants. Similarly, Hamann (2003:95-96) noted that a large number of languages avoid retroflexes in the front vowel context, even though this might not be a universal principle. In terms of perceptual cue, the retroflex-to-vowel transitions are similar or identical to the apicoalveolar-to-vowel transitions (Stevens and Blumstein 1975), and in a word-initial prevocalic position, the primary perceptual cues for a retroflex consonant (the lowered F3 and F4) are decreased as compared with a post-vocalic position (i.e., vowel-to-retroflex transition) (Steriade 1999). Thus, in terms of transitional cues alone, it would be hard to maintain a distinct contrast between a retroflex consonant and an apical-alveolar consonant in the prevocalic position.

The inherent incompatibility between a retroflex consonant and a following [_i] indicates that, in the typology of sibilant place contrasts in (2) or (57), the avoidance of the three-way contrasts of sibilant places in the [_i] context could be broken down into two different patterns:

(a) The avoidance of contrastive dental vs. palatal sibilants in the [_i] context;

(b) The avoidance of a combination of a retroflex consonants and a following vowel [_i].

The former pattern is likely to be rooted in the perceptual similarity of the contrasts, as evidenced by the results of the perceptual experiment in this study. The latter pattern, however, is likely to be rooted in the articulatory configuration of a sound sequence (Stevens and Blumstein 1975) and the difficulty to perceptually maintain a place contrast between a retroflex and another consonant (Steriade 1999). Further research is needed to tease apart these two possibilities.

5.3.3 Distinctiveness difference between manners of articulation

In linguistic typology, implicational relations have been observed to hold among different manners of articulation in terms of consonant inventory, place contrasts, and phonological processes. For the consonant inventory, the typological survey in Maddieson (1984:13-14) showed that the existence of nasal consonants at a particular place implies that of stops or affricates at the same place of articulation. For place contrasts, Hockett (1966) observed that, in a particular language, the place contrast of nasals implies that of stops. For place assimilation procceses, it has been observed that if stops are the targets of place assimilation, so are nasals (Jun 1995).

The typological survey in this dissertation showed that among the 41 dialects with contrastive dental vs. palatal sibilants in the [_i] context, the place contrast of affricates (i.e., [tsi-tɕi] [ts^{h}i-tɕ^{h}i]) implies the place contrast of fricatives (i.e., [si-ɕi]). The observation of the implicational relation between fricatives and affricates thus supplies to the literature a case of nuanced manner hierarchy in terms of consonant place contrasts.

Previous studies have made attempts to connect perceptual distinctiveness and linguistic typology. For example, Mohr and Wang (1968) observed from their perceptual experiment that voiced stops (b, d, g) are generally perceived by the listeners as less distinct from each other compared with voiceless stops (p, t, k). In other words, in terms of the perceptual distinctiveness, the place contrasts among voiced stops are found to be less distinct than the place contrasts among voiceless stops. Mohr and Wang (1968) connected their observation to the notion that voiced stops are marked and voiceless stops are unmarked in linguistic typology. In their view, the markedness of a phonetic category is related to the lower perceptual distinctiveness among its members, as compared with the higher distinctiveness among the members in the corresponding unmarked category.

In this dissertation, a similar connection is established between the manner differences in phonological contrasts and the results of perceptual experiments. Focusing on the perceptual distinctiveness of contrastive elements, it was hypothesized in this dissertation that a [si-ɕi] contrast is perceptually more distinct than a [tsi-tɕi] or [tsʰi-tɕʰi] contrast. This hypothesis was confirmed by the results of the speeded AX discrimination experiment, which showed that a [si-ɕi] pair is significantly less likely to introduce discrimination errors (i.e., different sounds judged as the same) than a [tsi-tɕi] pair or a [tsʰi-tɕʰi] pair. This result is thus consistent with the unpopularity of the affricate place contrasts in the [_i] context compared to that of fricatives. This connection between linguistic typology and the results of the perceptual experiment therefore supplied another case study in the perceptual explanation of language typology.

5.4 Contrast distinctiveness in sound change

5.4.1 The unit by which to evaluate contrast distinctiveness

Contrast distinctiveness has been discussed in terms of consonant and vowel inventories (Martinet 1952; Lindblom 1986, 1990; Flemming 2002, 2004, 2005, etc.) as well as the effect of neighboring sounds on the perception of consonant contrasts (Steriade 1997, 1999, 2001, 2008). This study falls into the second category when evaluating the distinctiveness of contrastive dental vs. palatal sibilants in different vowel contexts.

Hume & Mailhot (2013), for example, argued that perceptual distinctiveness can be modeled as a function of miscategorization probability of a contrast, by which a high degree of overlap is correlated with poor perceptual distinctiveness and high confusability. The results of response time in the perceptual experiments suggest that the dental and palatal sibilants are perceptually less distinct (i.e., more confusable) in the [_i] context than in the [_a] context, where the place distinction can be better cued by larger formant transitional differences between the dental and palatal consonants. This observation suggests that CV combination could potentially be taken as a unit by which a language configures its contrast distinctiveness. The motivation to consider units larger than segments is that the perceptual information for a segment is usually distributed over its neighboring segments (Liberman et al. 1967; Sereno, Baum, Marean and Lieberman 1987). While it is certainly possible to discuss contrast distinctiveness in a context-neutral way (e.g., a vowel system like /i-a-u/ is generally preferred cross-linguistically), taking into account the following vowel allows a more nuanced understanding of the perceptual distinctiveness between consonants.

This perspective is compatible with the proposal of Licensing by Cue (Steriade 1997, 1999) or P-Map (Steriade 2001, 2008), which posits a greater likelihood of contrast in the phonetic environments where the contrasting cues can be better recovered by the listener. The results of the perceptual experiment in this study showed that, in regard to consonant place contrast, different vocalic contexts may differ in cue recoverability. For the dental vs. palatal sibilant contrast, the transitional cues in the [_i] context tend to be less recoverable due to the smaller transition difference in this context. Such contrasts are shown to be dispreferred in the typology of Chinese dialects, similar to the observation in Lee-Kim (2014a). Simply put, the weak recoverability of place cues in the [_i] context shows its effect in language typology.

For the sound system of a particular language, evaluating distinctiveness by a unit larger than the segment would make different claims about the sound inventory. In the case of Mandarin, for example, the less distinct contrasts between the dentals and the palatals in the [_i] context (e.g., [tsi-tɕi]) are avoided with the introduction of an apical vowel after the dental sibilants (e.g., [tsɹ̩-tɕi]). This introduced one more vocalic sound into the vowel system, and it will inevitably make the vowel space 'more crowded' under a theory that evaluates the density of the vowel space with the number of vowels on the F1 and F2 dimensions (e.g., the introduction of [ɹ̩] as an allophone of /i/ makes the /i-ɤ/ contrast more crowded in the dental context). For example, a previous [tsi-tsɤ] contrast may develop into a [tsɹ̩-tsɤ] contrast, in which the phonological change may be deemed as an enhancement of a consonantal contrast at the cost of undermining vowel distinctiveness. However, if the CV combination is adopted as a unit to evaluate distinctiveness, it is in principle possible to compare the perceptual distinctiveness among these larger units in a unified space (e.g., [tsi-tɕi-tsɤ] vs. [tsɹ̩-tɕi-tsɤ]).

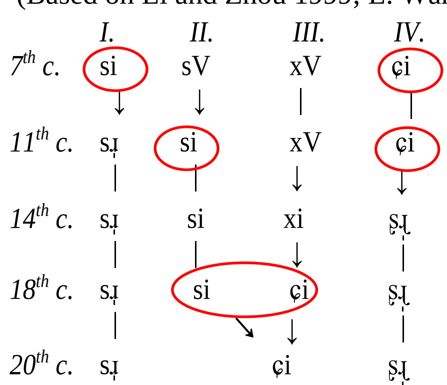
5.4.2 The emergence of weak contrasts

A key notion in the current study is that a weak contrast in a sound system tends to be enhanced or avoided in sound change, which is consistent with the position in a number of previous studies on the enhancement of contrasts in diachronic sound change (Penny 2002, Zampaulo 2013, Padgett and Źygis 2007; Źygis and Padgett 2010, among others). In the historical development of Polish, for example, the [ɕ-ʃʲ] contrast developed into the [ɕ-ʂ] contrast and the perceptual experiment by Zygis and Padgett (2010) showed that the latter pair is perceptually more distinct than the former pair. Although the notion of contrast enhancement in sound changes seems straightforward, an important and difficult question, as Zygis and Padgett (2010) put it, is why a less distinct contrast was created in a sound system in the first place, if it was to be avoided in the sound system later?

The current study is not in a position to fully answer this question. Yet, the development of sibilants and contrasts in Mandarin does suggest a possible direction to explain the emergence of weak contrasts in a sound system: less distinct contrasts can be created in a sound system by independently-motivated and unrelated sound changes that accidentally bring the contrasts into the sound system. In the historical development of Mandarin sibilants in (14), repeated as (59) below, weak contrasts like [si-ɕi] were created by unrelated sound changes across different historical stages. Prior to the 7th century, dental vs. palatal sibilants existed in the sound system but there was no rime [_i] (L. Wang 1985). In other words, dental vs. palatal sibilants do not form weak contrasts like [si-ɕi], even though they may contrast in other vowel contexts.

(59) Recall of the historical development of Mandarin in (14)

(Based on Li and Zhou 1999; L. Wang 1985)



Between the 7th century and the 11th century, the weak contrasts between dental and palatal sibilants in the [_i] context were created three times, through independent sound changes:

(a) Around the 7th century, contrasts like [si-ɕi] were created by a comprehensive rime merger, by which various rimes merged into [_i]. Therefore, for the lexical items whose onsets were dental and palatal sibilants and whose rimes were relevant to this rime merger, the phonetic contrasts turned into contrastive dental vs. palatal sibilants in the [_i] context. As shown in (59), these phonetic contrasts were avoided with the formation of apical vowels in Column I (7th c. – 11th c.).

(b) Between the 11th century and the 14th century, the rime merger created syllables like [si] again (Column II), which turned out to contrast with syllables like [ɕi]. The contrasts were subsequently avoided with the retroflexion of the palatal sibilants in the 11th century.

(c) Around the 18th century, the velar obstruents (Column III in (59)) were palatalized, e.g., [kʰi] → [tɕʰi], and this caused the contrasts of dental vs. palatal sibilants to be formed in the [_i] context again. As an independent sound change, the palatalization of velar obstruents before a high front vowel, for example, is crosslinguistically common (Bhat 1978) and well-motivated in terms of articulation (Keating and Lahiri 1993, Recasens and Espinosa 2009), acoustics (Guion 1998, Wilson 2006), and perception (Clements 1999, Chang, Plauché, and Ohala 2001).

Simply put, the weak contrasts of dental vs. palatal sibilants in the [_i] context were created three times in the development of (59), each time by an independently motivated sound change. Back to the question ‘why weak contrasts exist in a sound system in the first place’, it is thus suggested from the development of Mandarin sibilants that weak contrasts may come into existence in a phonological system as the results of different sound changes that may not be directly related to the pre-existing contrasts in a sound system. Ultimately, the weak contrasts tend to be resolved by additional sound changes.

5.4.3 The enhancement of weak contrasts

The perceptual experiment in Chapter 2 has established that contrasts like [si-ɕi], [tsi-tɕi], [tsʰi-tɕʰi] are perceptually less distinct than contrasts like [sɹ̩-ɕi], [tsɹ̩-tɕi], [tsʰɹ̩-tɕʰi], where the dental sibilants are followed by apical vowels. This is consistent with the observations that, across Chinese dialects, the introduction of apical vowels is the most common way to avoid contrastive dental vs. palatal sibilants in the [_i] context. This synchronic pattern is mirrored in the sound change of Mandarin as well. Between the 7^{th} century and the 11^{th} century, apical vowels were developed in the syllables [si tsi tsʰi] (Column I in (59)), which led to the avoidance of contrastive dental vs. palatal sibilants in the [_i] context. The results of the perceptual experiment in this dissertation suggest an answer as to *why* the apical vowels are formed, i.e., to enhance the weak contrasts of dental vs. palatal sibilants in the [_i] context. What remains unclear is *how* this sound change (e.g., [si]→ [sɹ̩]) was completed.

Contrast-enhancing sound changes occurred in languages like Trique, where the less distinct contrast of [uga] vs. [uda] developed into the more distinct contrast of [ugʷa] vs. [uda] (Silverman 2006a). Silverman (2006a:133–143, 2006b), for example, attributed the relevant sound change (i.e., [uga] → [ugʷa]) to (i) the inherent gradience of speech production, by which [uga] have variant forms like [ugʷa] in production and (ii) the preference for tokens that may reduce the probability of coomunication confusion, by which [ugʷa] stands out to be more distinct from [uda].[40]

The formation of apical vowels in Mandarin is a case where a *vowel* underwent a complete change to enhance a lexical contrast that used to surface as a *sibilant* place contrast in the [_i] context. The mechanism laid out in Silverman (2006a), by which the variant of a sound is established as a new category, may not be adequate to fully account for apical vowel formantion (e.g., [si]→ [sɹ̩]). While this dissertation is not in a position to propose a full model for the formation of apical vowel, it is possible to sketch a scenario of this sound change, based on previous studies on hyperarticulation (Lindblom 1990b; Johnson, Flemming, and Wright 1993), listener's reinterpretation of speech signals (Ohala 1981, Beddor 2009), and presumably the mechanism laid out in Silverman (2006a) as well, as shown below:

(a) As the results of independent sound changes, contrasts like [si-ɕi] emerged in a sound system. These contrasts are perceptually weak and may cause a higher possibility of

[40] A similar position is taken by Pierrehumbert (2002), among others, who explained the hyperarticulation of words by referring to their higher likelihood to be categorized correctly. Under such an approach, contrast enhancement is non-teleological in the sense that 'enhancement' is in fact the overrepresentation of 'exaggerated' tokens in the exemplar cloud.

confusion between the relevant lexical items.

(b) Faced with (a), the speakers may attempt to hyper-articulate the contrasts (Lindblom 1990b; Johnson, Flemming, and Wright 1993, Cho, Lee, and Kim 2011, Wedel, Nelson, and Sharp 2016, among others) to lower the chance of the confusion among lexical items. For example, they may hyperarticulate a lexical item corresponding to /si/ to make it acoustically and perceptually more distinct from [ɕi].

(c) One likely way to hyperarticulate a syllable like /si/ is to strengthen the co-articulation between the consonant and the vowel, e.g., by lengthening the duration of the coarticulation to provide a better cue for the place of the consonants (Scarborough 2004, Pycha 2015).[41] This is consistent with the suggestion from the perceptual experiment that the reduced distinctiveness of sibilant place contrasts in the [_i] context is likely to be rooted in the similar transitional properties related to the dental and palatal sibilants in the [_i] context.[42]

(d) The hyper-articulated /si/ tokens (represented as [sHi] below) may be favored by the speaker-listeners in that they lower the possibility of confusion between [si] and [ɕi] and

[41] In the perceptual experiment, the duration of the CV transition was manipulated (= shortened) to 50ms across all stimulus syllables to facilitate the comparison of response times. Note that in real speech, it is possible to lengthen the CV transition to enhance the perceptual recoverability of the place of the consonants (Scarborough 2004, Pycha 2015).

[42] Hyper-articulation on the sibilants is conceivable; however, it might not be as helpful as the hyper-articulation of the CV coarticulation. For the stimuli used in the perceptual experiment, for example, it has been shown that the acoustic difference of the dental vs. palatal sibilants was in fact greater in the [_i] context than in the [_a] context. Yet, the smaller transitional difference in the [_i] context overrides the larger acoustic difference in the onset sibilants and leads to the reduced perceptual distinctiveness of the sibilant place contrast in the [_i] context. Therefore, hyper-articulation on the consonant vowel coarticulation might be more effective than one on the sibilant onsets. Of course, there remains the possibility of hyper-articulation in both sibilants and the consonant-vowel coarticulation.

thus facilitate communication. Over time, the hyper-articulated exemplars of [sʜi] may be retained by the speaker-listener.

(e) In the speaker-listener interaction, the hyper-articulated CV coarticulation, as in [sʜi], might be reinterpreted as a property of the vowel rather than simply consonant-vowel coarticulation (Ohala 1981, Beddor 2009). Beddor's (2009) study of vowel nasalization, for instance, showed that in a ṼN combination, the velum lowering associated with the consonant constriction can be reinterpreted as being associated with the vowel, thus leading to the sound change ṼN→ Ṽ. In case of the hyperarticulated [sʜi], similarly, the coarticulation could be reinterpreted as a property of the vocalic portion, e.g., [-ɹ̩i], rather than simply a transition between the consonant and the vowel.

(f) With the reinterpretation assumed above, the listener-speaker interaction would create [sʜi] tokens that are phonetically even more similar to [sɹ̩]. These exemplars would be retained by the speakers and finally lead to the establishment of [sɹ̩] in the language. In the end, a sound change like [si] → [sɹ̩] is completed and the probability of confusion between the original contrasts of /si/ vs. /ɕi/ is lowered.[43]

The scenario above of course remains hypothetical and empirical evidence is needed to verify its validity. In general, the formation of apical vowel is likely to be accounted for by a model that reconciles hyperacticulation of contrastive elements (Lindblom 1990b; Johnson, Flemming, and Wright 1993) and the overrepresentation of exaggerated tokens (Silverman 2006a).

[43] Another possible factor that may facilitate the establishment of [sɹ̩] can be its ease of articulation compared with [si]: In [sɹ̩], the consonant and vowel share the same place of articulation and thus fewer articulatory gestures are involved compared to a [si] syllable.

5.5 Enhancement and neutralization in sound change

5.5.1 Enhancement vs. neutralization

As Hume and Johnson (2001:8) noted, weak contrasts may be avoided by enhancing the contrast or neutralizing it. For contrast enhancement, the typological survey in this dissertation has shown that vowel enhancement is the most common way of avoiding the weak contrasts between dental and palatal sibilants in the [_i] context. Neutralization of contrasts, on the other hand, seems not to be as preferred as enhancement in a sound change. Silverman (2010), for example, examined a range of consonant neutralization processes in Korean and observed that only a very low level of homophony results from these processes. In other words, even when there is a natural phonetic tendency toward a phoneme-neutralizing sound change, it is likely that a language will not evolve toward the neutralization of lexical contrasts (Silverman 2010).[44]

The diachronic development of Mandarin sibilants involved three separate sound changes that turned out to avoid contrastive dental vs. palatal sibilants in the [_i] context, as illustrated in (60), where bold face marks the contrasting forms before and after the sound changes. Out of the three sound changes, the first two preserved the lexical contrasts, i.e., the formation of apical vowel in (60a) and the retroflexion of palatal sibilants in (60b). This is consistent with the contention that the neutralization of lexical contrasts is not preferred in sound change (Silverman 2010, among others).

[44] Studies by Wedel, Kaplan, and Jackson (2013) and Silverman (2010) showed different aspects of the neutralization of lexical contrasts: Wedel et al. (2013) showed that contrasts with higher functional load are less likely to be neutralized while Silerman (2010) showed that, in the phonemic neutralization in Korean, the actually neutralized lexical contrasts are relatively rare.

(60) The avoidance of dental vs. palatal sibilants in the development of Mandarin

a. Apical vowel formation (7th – 11th century)

	si	**ɕi**
	↙	|
sɹ̩		**ɕi**

b. Retroflexion of palatal sibilants (11th – 14th century)

sɹ̩	**si**	**ɕi**	ʂ_(*ʂi)
|	|	↘	
sɹ̩	**si**		**ʂi**

c. Sibilant neutralization (18th – 20th century)

sɹ̩	**si**	**ɕi**	ʂɻ̩
|	↘	|	|
sɹ̩		**ɕi**	ʂɻ̩

As stated in 4.1.2, the sibilant neutralization in (60c) turned out to be a case where the weak contrasts (i.e., dental vs. palatal sibilants in the [_i] context) needed to be avoided and the existing syllabic inventory provided no room for vowel or consonant enhancement, that is, the sibilant neutralization was left as the only option to avoid the weak contrasts. The sound change in (60c), thus, showed that the need to avoid weak phonetic contrasts may override the force to preserve lexical contrasts, even though the sound change may lead to an increase of the level of homophones in the language.

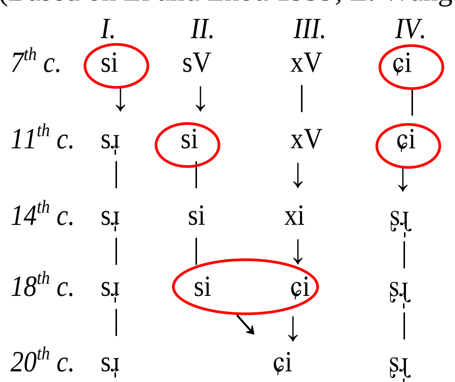
A closer examination of (60) also showed that the retroflexion of palatal sibilants (60b) and the sibilant neutralization (60c) both involved phonological neutralization. The palatal retroflexion in (60b) involved the neutralization of the palatal sibilants and the pre-existing retroflex sibilants, where the retroflex sibilants used not to combine with a following vowel [_i]. Therefore, the sound change in (60b) involved the neutralization of two sound categories (i.e., palatal vs. retroflex sibilants) but no neutralization of lexical items. On the other hand, the neutralization in (60c) involved the neutralization of lexical contrasts. This difference suggested that, in the evaluation of neutralization of a sound system, a distinction needs to be made between a neutralization involving phonemic categories and a neutralization involving lexical items, as Silverman (2010) did in the analysis of consonant neutralization in Korean.

5.5.2 Syllable inventory, enhancement, and neutralization

A natural question related to the avoidance of a weak contrast, e.g., dental vs. palatal sibilants in the [_i] context, is what determines whether a sound system chooses to enhance the contrasts (e.g., with the introduction of an apical vowel) or to neutralize the contrasts (e.g., with the palatalization of dental sibilants in the [_i] context). Focusing on linguistics-internal factors, the historical developments of Mandarin sibilants suggested that the existing syllabic inventory in a sound system may influence whether a language choose to enhance the contrasts or neutralize the contrasts. The historical sound changes of Mandarin sibilant, as in (14), is repeated as (61).

(61) Sibilant contrasts in Mandarin sound change repeated from (14)

(Based on Li and Zhou 1999; L. Wang 1985)



As shown in (61), the contrasts of dental vs. palatal sibilants came into the sound system around the 7th century, the 11th century, and the 18th century. Contrast enhancement applied in the first two cases and sibilant neutralization applied in the third case. An examination of the syllabic inventory showed that, in the 7th and the 11th century, there were fewer syllables in the form of 'sibilant+[+high, -back]vowel' than in the 18th century. As discussed in 4.1.2, the neutralization of dental vs. palatal sibilants in the 18th century seems to be a case where there was no room left for vowel enhancement, e.g., [si-ɕi] → [sɹ̩-ɕi], or sibilant enhancement, e.g., [si-ɕi] → [si-ʂi]. Therefore, for the enhancement vs. neutralization of sibilant place contrasts in the [_i] context, it is reasonable to assume that:

(a) A sound system with fewer types of syllables in the form of 'sibilant+[+high, -back] vowel' is more likely to choose contrast enhancement and less likely to choose sibilant neutralization;

(b) A sound system with more types of syllables in the form of 'sibilant+[+high, -back] vowel' is less likely to choose contrast enhancement and more likely to choose sibilant neutralization.

The validity of these assumptions needs to be tested with acoustic measurements of sibilants and vowels in languages that have different numbers of syllables in the form of 'sibilant+[+high, -back] vowel'.

5.6 Contrast distinctiveness and usage factors in sound change

For sibilant place contrasts, the focus of this dissertation is the perceptual distinctiveness of the contrastive elements. It is obvious that multiple factors may influence the change in a sound system, including phonetic factors (e.g., perceptual distinctiveness, articulatory effort, etc.), usage factors (e.g., functional load, frequency), and processing factors (e.g., structural complexity). This section attempts to give a brief discussion of the relation between contrast distinctiveness and usage factors in language.

5.6.1 Perceptual distinctiveness and lexical contrasts

The existence of lexical contrasts or minimal pairs has been shown to be relevant to hyperarticulation (Ohala 1994; Stent, Huffman, and Brennan 2008; Maniwa, Jongman, and Wade 2008, 2009; Schertz 2013; Seyfarth, Buz, and Jaeger 2016), by which phonetic cues are hyperarticulated as a response to the competition between lexical items that are phonetically similar. For example, Wedel, Nelson, and Sharp (2016) observed that, in natural speech data, the existence of minimal pairs is correlated with an enhancement of phonetic cues in creating a greater perceptual distance between the relevant lexical competitors.

An important characteristic of Mandarin and other Chinese dialects is that a syllable typically corresponds to a lexical item or a morpheme, e.g., 思 [sɹ̩] 'think' vs. 西 [ɕi] 'west'. Around the 7th century, the general rime merger into [_i] in Mandarin created a situation where the phonetic forms of different lexical items were contrastive by the onset sibilants only. For example, the two lexical items 思 'to think' and 赐 'to give' were phonetically [sĭə] vs. [ɕĭe] prior to the rime merger in the 7th century (Li and Zhou 1993). With the general rime merger, the phonetic forms of the same lexical contrast became [si] 'to think' vs. [ɕi] 'to give', where the onsets took the role of distinguishing the relevant lexical items. Therefore, the emergence of minimal pairs around 7th century Mandarin, e.g., [si] 'to think' vs. [ɕi] 'to give', was likely to be relevant to the hyperarcitulation of the contrast, as discussed in 5.4.3.

Compared with the lexical contrasts dicussed in Wedel, Nelson, and Sharp (2016), the contrast in 7th century Mandarin has an additional property, i.e., the contrastive dental vs. palatal sibilants in the [_i] conext are shown to form weak contrasts as evidenced by the results of the perceptual experiment. Therefore, it is conceivable that the perceptual similarity in contrasts like [si-ɕi] may further call for hyperarticulation of the contrasts.

5.6.2 Perceptual distinctiveness and functional load

Between the 18th century and the 20th century, the contrastive dental vs. palatal sibilants in the [_i] context were avoided with the neutralization of dental vs. palatal sibilants, illustrated as (62a) below. As discussed in 4.1.2, the avoidance of contrastive dental vs. palatal sibilants in the [_i] context can be avoided with the neutralization of contrasts like [si-ɕi] in (62a) as well as the hypothetical neutralization of contrasts like [sɹ̩-si] in (62b). In both cases, the neutralization of lexical contrasts is involved. The choice of (62a) over (62b) turned out to be a neutralization of two elements that are perceptually more similar, e.g., [si-ɕi] in stead of [sɹ̩-si].

(62) Neutralization of dental vs. palatal sibilants (a partial recall of (31)):

a. Neutralization of dental vs. palatal sibilants (18th c. – 20th c.)

sɹ̩ **si** ɕi ʂɻ̩
| ↘ | |
sɹ̩ ɕi ʂɻ̩

b. Hypothetical apical vowel formation (18th c. – 20th c.)

sɹ̩ **si** ɕi ʂɻ̩
| ↙ | |
sɹ̩ ɕi ʂɻ̩

Perceptual similarity is by no means the only factor that can influence the neutralization of phonological contrasts. Previous studies have shown that the more ‘work’ a sound pair does in the communication of a language, the less likely their contrast is to be lost (de Courtenay 1895; Hockett 1969; Wang 1969; Silverman 2006a, 2006b, 2010; Kaplan 2011; Wedel, Kaplan, and Jackson 2013; Wedel, Jackson, and Kaplan 2013, among others). More specifically, the more minimal pairs a sound pair involves, the less likely the contrast will be neutralized (Wedel, Kaplan, and Jackson 2013; Kaplan 2011).

A more complete account of the the neutralization in (62a) needs to incorporate the functional factors such as the counts of minimal pairs. However, an analysis of the functional load of a sibilant contrast in Chinese is challenging due to the existence of a high level of homophony among morphemes. In Mandarin, for example, the phonetic form [sɹ̩] with a high-level tone corresponds to lexical items like 思 'to think', 撕 'to tear', 丝 'silk', etc. One possible way to evaluate the role of functional load in diachronic sound change is to evaluate 'created homophones', following Kaplan (2011), i.e., to compare the level of homophony before and after a sound change. For example, it could be that one neutralization process creates a smaller number of homophones than its alternatives. Future research is needed to evaluate the role of functional factors in the sound change of Mandarin.

5.8 Conclusions

This dissertation showed that the perceptual distinctiveness of dental vs. palatal sibilants is reduced in the [_i] context compared with other vowel contexts. The reduced distinctiveness of sibilant place contrasts in the [_i] context turns out to function like a constraint in linguistic typology and in sound change, as evidenced in a survey of the typology of Chinese dialects and the historical development of Mandarin sibilants. Similarly, the relative distinctiveness of sibilant contrasts among different manners, as shown in the results of the perceptual experiment, is shown to be correlated with the patterns observed in the crosslinguistic typology as well.

For the avoidance of sibilant place contrast in the [_i] context, the study in this dissertation suggests the need to disentangle perceptual and articulatory factors in the explanation of language typology. The experimental study in this dissertation supplied evidence for the perceptual aspect and more work is needed to explore the articulatory aspect. In terms of sound change, particularly for the enhancement vs. neutralization of a weak contrast, this dissertation supplied empirical evidence on language-internal factors such as the relative perceptual distinctiveness of sibilant place contrast in different vowel contexts. For a more complete model of sound change, future studies are needed to investigate the interaction of linguistic factors like contrast distinctiveness and usage factors like functional load.

Appendices

Appendix I. Typology of sibilant contrasts across Chinese dialects: An overview

The following data are drawn from *Fangyan* [*Dialects*], a Chinese journal that focuses on the description of Chinese dialects. This list includes an overview of pattern of sibilant place contrasts across the 170 Chinese dialects. The focus is on the CV syllables whose onsets were sibilants and whose vowels are [i], or apical vowels [ɿ/ʅ].

1. Sibilants at 1 Place

1.1 Sibilants at *one* place - No apical vowel: 28 dialects

1.1.1 Sibilants at *one* place – Dental - No apical vowel: 22 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Chenghai	pi p^{h}i bi	ti t^{h}i			ki k^{h}i gi
(L. Lin 1994)		tsi ts^{h}i dzi			
澄海[廣東]		si			hi
	mi	ni li	i		ŋi

1.1.2 Sibilants at *one* place – **Palatal -** No apical vowel: 6 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Cangwu	pi p^{h}i ɓi	ti t^{h}i *ɗi			ki k^{h}i *kwi *k^{h}wi
(Z. Zhong 2015)			tʃi tʃ^{h}i		
蒼梧(石橋鎮)[廣西]	fi		ʃi		hi
粤语-勾漏片	mi	ni li	i		ŋi wi

1.2 Sibilants at *one* place - *Contrastive* ɿ and i: 7 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Huiyang	pi p^{h}i	ti t^{h}i			ki k^{h}i
(R. Zhou 1987)		tsɿ-tsi ts^{h}ɿ-ts^{h}i			
惠陽[廣東]	*fi *vi	sɿ-si			hi
客家話	mi	ni li	i		ŋi

2 Sibilants at 2 Places

2.1 Sibilants at *2* places– Dental vs. Palatal

2.1.1 Sibilants at *2* places– Dental vs. Palatal – ***Contrastive***

2.1.1.1 Sibilants at *2* places- Dental vs. Palatal – ***Contrastive*** – **si-ɕi**: 3 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jiangyong	pi p^{h}i	ti t^{h}i			(ki) (k^{h}i)
(X. Hunag 1988)		tsi ts^{h}i	tɕi tɕ^{h}i		
江永土話[湖南]	fi (vi)	si	ɕi		(hi)
土话	mi	ni li	ȵi i		*ŋi

2.1.1.2 Sibilants at *2* places- Dental vs. Palatal – ***Contrastive*** – **sɿ-si-ɕi**: 20 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Changle	pi pʰi bi	ti tʰi di			*ki *kʰi *gi ʔi
(Z. Qian 2003)		tsɿ-tsi tsʰɿ-tsʰi dzɿ-dzi	tɕi tɕʰi dʑi		
長樂(老派)[浙江]	fi vi	sɿ-si zɿ-zi	ɕi *ʑi		*hi *ɦi
吳語-太湖片(臨紹)	mi	ni li	ȵi i		*ŋi

2.1.2 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive***

2.1.2.1 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – **sɿ-ɕi**: 56 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Anxiang	pi pʰi	ti tʰi			*ki *kʰi
(Y. Ying 1988)		tsɿ tsʰɿ	tɕi tɕʰi		
安鄉[湖南]	*fi	sɿ	ɕi	*ʐi	*xi
湘語/西南官話	mi	li	i		*ŋi

2.1.2.2 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – ***si-ɕi** : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Shibei	pi pʰi bi	ti tʰi di			ki kʰi gi *ʔi
(A. Hiroyuki 2004)		*tsi *tsʰi *dzi	tɕi tɕʰi dʑi		
石陂[福建]		*si	ɕi		xi ɦi
閩語-閩北區	mi	ni li	i		ŋi

2.1.2.3 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – **(sɿ-) si-*ɕi** : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jixi	pi-pɿ *pʰi-pʰɿ	ti tʰi			ki kʰi
(R. Zhao 1989)		tsɿ-tsi tsʰɿ-tsʰi	*tɕi *tɕʰi		
績溪[安徽]	fi vi	sɿ-si zɿ-*zi	*ɕi		xi
	mi-mɿ	ni-nɿ	*ȵi *i		*ŋi

2.2 Sibilants at *2* places– Dental vs. Retroflex

2.2.1 Sibilants at *2* places– Dental vs. Retroflex – Contrastive –**sɿ-si-ʂi:** 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal(No onset)*	*Retroflex*	*Velar (Glottal)*
Wuhua	pi pʰi	ti tʰi			ki kʰi
(Y. Wei 1997)		tsɿ-tsi tsʰɿ-tsʰi		tʂi tʂʰi	
五華[廣東]	fi vi	sɿ-si		ʂi	hi
客家話	mi	ni li	i		*ŋi

2.2.2 Sibilants at *2* places– Dental vs. Retroflex – Non-contrastive –**sɿ-si-ʃʅ**: 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jinggangshan	pi pʰi	ti tʰi			ki kʰi
(S. Lu 1995)		tsi-tsɿ tsʰi-tsʰɿ		tʃʅ tʃʰʅ	
井岡山[江西]	fi vi	si-sɿ		ʃʅ	xi
客家話	mi	ni li		i	ŋi

2.3 Sibilants at *2* places– Palatal vs. Retroflex

Sibilants at *2* places– Palatal vs. Retroflex – Non-contrastive – **ɕi-ʂʅ**: 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Haizhou	pi *pʰi	ti tʰi			*ki *kʰi
(X. Su 1990)			tɕi tɕʰi	tʂʅ tʂʰʅ	
海州[江蘇]	*fi		ɕi	ʂʅ *ʐʅ	*xi
	mi	li	i		

3 Sibilants at 3 or more Places

3.1 Sibilants at *3* places– Dental vs. Palatal vs. Retroflex

3.1.1 Sibilants at *3* places– ***3-place Contrasts*** **–** (sɿ-)**si-ɕi-ʂi:** 2 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Qimen	pi pʰi	ti tʰi			*ki *kʰi
(T. Shen 1989)		tsɿ-*tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂi tʂʰi	
祁門[安徽]	fi	sɿ-si	ɕi	ʂi	xi
	mi	(ni) li	ȵi i		*ŋi

3.1.2 Sibilants at *3* places– 2-place contrasts – Contrastive dental vs. palatal

3.1.2.1 Sibilants at *3* places**–*2-place Contrasts -1-*** (sɿ-)**si-ɕi-ʂʅ*:*** 12 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Binxian	pi pʰi	ti *tʰi		*ʈi *ʈʰi	*ki *kʰi
(Qiao & Chao 2002)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
彬縣[陝西咸陽]	*fi	sɿ-si zɿ-*zi	ɕi	ʂʅ ʐʅ	*xi
官話-中原-關中片	mi	li	ȵi i		*ŋi

3.1.2.1 Sibilants at *3* places**–*2-place Contrasts -2-* (sɿ-)si-ɕi-*ʂi*:*** 3 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Changshu (Meili)	pi pʰi bi	ti tʰi di			*ki *kʰi *gi
(D. Yuan 2010)		tsɿ-tsi tsʰɿ-tsʰi dzɿ-dzi	tɕi tɕʰi dʑi	*tʂi *tʂʰi *dʐi	
常熟(梅李) [江蘇]	fi vi	sɿ-si zɿ-zi	ɕi	*ʂi *ʐi	*hi *ɦi
吳語太湖片(蘇滬嘉)	mi	*ni li	ȵi ji i		*ŋi

3.1.3 Sibilants at *3* places– 2-place contrasts – Contrastive Palatal vs. Retroflex : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex(?)*	*Velar (Glottal)*
Xuancheng	pi pʰi pfɦi	ti tʰi tɦi tɦɿ			*ki *kʰi
(Shen & Huang 2015)	pɿ pʰɿ	tsɿ tsʰɿ *tsɦɿ	tɕi tɕʰi tɕɦi	tʃi tʃʰi *tʃɦi	
宣城(雁翅)[安徽]	*fi *vi vɿ	sɿ zɿ	ɕi	ʃi ʒi	*xi *hɦi
吳語-宣州片-銅涇小片	mi mɿ	ni li	i		*ŋi

3.1.4 Sibilants at *3* places– ***Non-contrastive***

3.1.4.1 Sibilants at *3* places– ***Non-contrastive-1*- sɿ-ɕi-ʂʅ -A**: 20 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Anqing	pi pʰi	ti tʰi			*ki *kʰi
(H. Bao 2012)		tsɿ tsʰɿ	tɕi tɕʰi	tʂʅ tʂʰʅ	
安慶[安徽]	*fi	sɿ	ɕi	ʂʅ ɻʅ	*xi
官話-江淮-洪巢片	mi	ni	i		*ŋi

3.1.4.2 Sibilants at *3* places– ***Non-contrastive-1*- sɿ-ɕi-ʂʅ -B**: **(with-labial affricates)**: 8 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jiangxian	pi pʰi	ti tʰi			ki *kʰi
(L. Wang 2014)	pfi pfʰi	tsɿ tsʰɿ	tɕi tɕʰi	tʂʅ tʂʰʅ	
绛縣[山西]	fi *vi	sɿ	ɕi	ʂʅ zʅ	*xi *ɣi
官話-中原-汾河片	mi	*ni li	ȵi i		*ŋi

3.1.4.3 Sibilants at *3* places– ***Non-contrastive-1*- sɿ-ɕi-*ʂi**: 2 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Dangtu (Huyang)	pi pʰi bi	ti tʰi ɾi			*ki *kʰi *gi
(Zheng *et al.* 2012)		tsɿ tsʰɿ dzɿ	tɕi tɕʰi *dʑi	*tʃi *tʃʰi	
當涂(湖陽)[安徽]	*fi	sɿ zɿ	ɕi ʑi	*ʃi *ʒi	*xi *ɣi *ɦi
吳語-宣州片(太高)	mi	*ni *li ɿ	ȵi i	*ɻi	*ŋi

3.1.4.4 Sibilants at *3* places– ***Non-contrastive-1*- sɿ-*ɕi-ʂʅ**:: 2 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Ledu	pɿ pʰɿ	*ti *tʰi			*ki *kʰi
(Cao & Shao 2001)		tsɿ tsʰɿ	*tɕi *tɕʰi	tʂʅ tʂʰʅ	
樂都[青海]	*fi *vi	sɿ zɿ	*ɕi	ʂʅ zʅ	*hi
	mɿ	ni lɿ			

3.2 Sibilants at *4* places– Dental vs. Palatal vs. Retroflex: 1 dialect

Dialects	*Labial/Dental*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jüxian	pi pʰi	ti tʰi			*ki *kʰi
(M. Shi 1987)	tθɿ tθʰɿ	tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
莒縣[山東]	θɿ *fi	sɿ-si	ɕi	ʂʅ	*xi
官話-北方	mi	ni li	i		*ŋi

Appendix II. Typology of sibilant contrasts across Chinese dialects: A complete list

Notes:

1. In the tables below, the dialects are ordered first by the number of sibilant places, then by the pattern of sibilant contrast (if applicable), and lastly by alphabetical order within each sublist.
2. For each dialect, the columns follow the order of place of articulation; the rows follow the order of stop, affricate, fricative, approximant, etc. In the original resources, most of the aspirated consonants are expressed like p‘, tɕ‘, etc., in accordance with the convention in Chinese literature. Following the IPA style, such consonants are transcribed into p^h, $tɕ^h$, etc. in the following tables.
3. For each dialect, symbols like “*”, “( )”, and “/” are used in the following way:

   pi = the dialect has the onset [p], and the combination [pi];

   *fi = the dialect has the onset [f] and [i], but [f] does not combine with the rime [i];

   (zʅ) = the dialect has [zʅ] only colloquially, without written forms;

   si-sʅ = the dialect has contrastive [si] vs.[sʅ].

1 Sibilants at 1 Place

1.1 Sibilants at *one* place - No apical vowel: 28 dialects

1.1.1 Sibilants at *one* place – Dental - No apical vowel: 22 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Binyang (H. Mo 2014) 賓陽(王靈鎮)[廣西] 平話-桂南	pi p^{h}i *fi *vi mi	ti *t^{h}i tsi ts^{h}i si ni li ɬi	 ɲi ji		*ki k^{h}i *kwi *k^{h}wi *hi *ŋi *ŋwi
Chenghai (L. Lin 1994) 澄海[廣東]	pi p^{h}i bi mi	ti t^{h}i tsi ts^{h}i dzi si ni li	 i		ki k^{h}i gi hi ŋi
Danzhou (Y. Wu 1988) 儋州[海南]	ʔbi ɸi vi mi	ti ʔdi tsi si zi ni li	 i		ki xi hi ŋi
Dongguan (Zhan & Chen 1995) 東莞[廣東] 粵語	*pi *p^{h}i fi vi *mi	(ti) *t^{h}i tsi ts^{h}i si zi *ni	 i		*ki (k^{h}i) kui k^{h}ui *hi (ŋi)
Fuqing (A. Feng 1988) 福清[福建福州] 閩語-閩東區-侯官片	pi p^{h}i mi	ti t^{h}i tsi ts^{h}i si ni li	 i		ki k^{h}i hi ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Haikang (Z. Zhang 1987) 海康[廣東湛江] 闽语	pi pʰi bi mi	ti tʰi tsi tsʰi si ni li	 i		ki kʰi hi ŋi
Haikou (Y. Du 2007) 海口(老派)[海南]	ɓi fi vi mi	ti ɗi tsi si zi ni li	 i		ki xi hi ŋi
Huaiji (Xiafang) (B. Yang 2012) 懷集(下坊)[廣東]	*pi *pʰi *fi *mi	ti *tʰi tsi tsʰi θi ni li	 i		ki kʰi ʔi hi *ŋi
Lechang (Tatouba) (C. Zhuang 1996) 樂昌(塔頭壩)[廣東] 闽语	pi pʰi *mi	ti tʰi tsi tsʰi si *ni li	 i		ki kʰi hi *ŋi
Lianjiang (C. Lian 2002) 廉江石角[廣東] 客家話	pi pʰi *fi vi mi	ti tʰi tsi tsʰi si *ni li	 i		ki kʰi hi ŋi
Nanhai Shatou (X. Peng 1990) 南海(沙頭)[廣東]	*pi *pʰi *fi *mi	ti *tʰi *ȶi tsi tsʰi si ni li	 i		ki kʰi *kui *kʰui hi ŋi *ui
Nanning (Ping) (Y. Tan 1996) 南寧(平話)[廣西] 平話	pi pʰi fi *βi mi	ti *tʰi tsi tsʰi si ni li ɬi	 ɲi i		ki kʰi *kwi *kʰwi hi *ŋi
Ningde (P. Sha 1999) 宁德[福建] 閩語-閩東方言	pi pʰi mi	ti tʰi tsi tsʰi si ni li	 i		ki kʰi hi ŋi
Rongxian (X. Chen 1999) 容縣(縣底村)[廣西] 客家話	pi pʰi fi *vi mi	ti tʰi tsi tsʰi zi ni li ɬi	 i		ki kʰi hi ŋi
Sanya (Mai) (Jiang *et al*. 2007) 三亞邁話[海南]	pʰi bi vi mi	ti tʰi di tsi tsʰi θi zi ni li			ki kʰi ʔi hi ŋi
Wenchang (Y. Liang 1986) 文昌[海南]	ʔbi bi ɸi mi	ti ʔdi (diʔ) tsi si zi ni li	 i		ki *gi xi hi ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Wuchuan (Z. Zhang 1992) 吴川[廣東] 粵語(吳陽白話)	*pʰi *bi *fi *vi *mi	*ti *tʰi *di tsi tsʰi si *ni *li *ɬi	 i		*ki *kʰi *hi ŋi
Xiamen (C. Zhou 1991) 厦門[福建] 閩語	pi pʰi bi *mi	ti tʰi tsi tsʰi si *ni li	 i		ki kʰi gi hi *ŋi
Xinfeng (R. Zhou 1992) 新豐(客家)[廣東] 客家话	pi pʰi (fi) *vi mi	ti tʰi tsi tsʰi si li	 i		ki kʰi (hi) ŋi
Zengcheng (W. He 1990) 增城[廣東]	pi *pʰi *fi mi	ti *tʰi tsi tsʰi si ni *li	 i		ki *kʰi (kui) *kʰui *hi (ui) ŋi
Zhangzhou (Z. Ma 1993) 漳州[福建]	pi pʰi bi *mi	ti tʰi tsi tsʰi dzi si *ni li	 i		ki kʰi gi hi *ŋi
Zhaoping (Q. Huang 2006) 昭平[廣西] 土白話	pi pʰi fi mi	ti tʰi tsi tsʰi si ni li	 ȵi i		ki kʰi hi *ŋi

1.1.2 Sibilants at *one* place – **Palatal** - No apical vowel: 6 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Cangwu (Z. Zhong 2015) 蒼梧(石橋鎮)[廣西] 粵语-勾漏片	pi pʰi ɓi fi mi	ti tʰi *ɗi ni li	 tʃi tʃʰi ʃi i		ki kʰi *kwi *kʰwi hi ŋi wi
Hezhou (Babu) (X. Chen 2012) 賀州八步[廣西]	pi pʰi fi mi	ti tʰi θi ni li	 tʃi tʃʰi ʃi ȵi i		ki *kʰi hi ŋi
Lianzhou (Q. Cai 1987) 廉州[廣西] 粵语	pi pʰi fi *ʋi mi	*ti tʰi ni li ɬi	 tʃi tʃʰi ʃi ȵi ji		ki kʰi *kwi *kʰwi xi *ŋi
Mengjiang (S. Yang 2013) 濛江[廣西] 粵语-勾漏片	pi pʰi fi *vi mi	ti tʰi θi ni li	 tʃi tʃʰi ʃi ȵi ji i		ki kʰi *kui *kʰui xi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Yongnin (Sitang) (C. Zhou 1991) 邕寧四塘[福建] 平話	pi pʰi fi mi	ti tʰi ni li ɬi	 tʃi tʃʰi ȵi ji i		ki *kʰi *kwi *kʰwi hi *ŋi *wi
Zhongshan (Y. Deng 2000) 鐘山[廣西]	pi pʰi fi vi mi	ti tʰi θi ni li	 tʃi tʃʰi ʃi ȵi ji		ki kʰi hi *ŋi

1.2 Sibilants at *one* place - *Contrastive* ɿ and i: 7 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Dongfang(Sigeng) (Liu & Ou 2005) 東方四更(付馬)[海南]	pʰi ɓi tθi *fi *vi mi	tʰi ɗi *tsɿ-tsi tsʰɿ-tsʰi sɿ−si ni li ɹi	 i		ki kʰi hi *ŋi
Gongcheng (Y. Guan 2005) 恭城(直話)[廣西]	pi pʰi bi *ɸi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi *dzɿ-*dzi sɿ−si zɿ−zi ni li	 i		ki kʰi gi xi *ɣi *ŋi
Huiyang (R. Zhou 1987) 惠陽[廣東] 客家話	pi pʰi *fi *vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 i		ki kʰi hi ŋi
Lishi (D. Li 2013) 犁市土話[廣東] 客家話	pi pʰi bi fi vi mi	ti tʰi tsɿ-tsi tsʰɿ sɿ li	 ji		ki kʰi gi hi
Meixian Hakka (X.Hunag 1992) 梅縣[廣東] 客家話	pi pʰi fi vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 i		ki kʰi hi ŋi
Nanning (Bai) (J. Xie 1994) 南寧(白話)[廣西]	pi pʰi fi mi	ti *tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li ɬi	 ji		ki kʰi *ʔi *kui *kʰui hi *wi *ŋi
Wuping Zhongshan (Y. Liang 1990) 武平(中山)[福建] ??軍家話	pi pʰi fi vi mi	ti *tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ−si ni li	 i		ki (kʰi) *xi ŋi

2 Sibilants at 2 Places

2.1 Sibilants at *2* places– Dental vs. Palatal

2.1.1 Sibilants at *2* places– Dental vs. Palatal – ***Contrastive***

2.1.1.1 Sibilants at *2* places- Dental vs. Palatal – ***Contrastive*** – **si-ɕi**: 3 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Jiangyong (X. Hunag 1988) 江永土話[湖南]	pi pʰi fi (vi) mi	ti tʰi tsi tsʰi si ni li	 tɕi tɕʰi ɕi ȵi i		(ki) (kʰi) (hi) *ŋi
Lingui Liangjiang (J. Liang 1996) 臨桂兩江[廣西]	pi pʰi *fi mi	ti tʰi tsi tsʰi si ni li	 tʃi tʃʰi ʃi ȵi i		ki kʰi *kwi *kʰwi hi *ŋi *wi
Lingui Sitang (M. Luo 1996) 臨桂四塘[廣西]	pi pʰi fi mi	ti tʰi tsi tsʰi si ni li	 tʃi tʃʰi ʃi i		ki *kʰi xi *ŋi

2.1.1.2 Sibilants at *2* places- Dental vs. Palatal – ***Contrastive*** – **sɿ-si-ɕi**: 20 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Changle (Z. Qian 2003) 長樂(老派)[浙江] 吳語-太湖片(臨紹)	pi pʰi bi fi vi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi dzɿ-dzi sɿ-si zɿ-zi ni li	 tɕi tɕʰi dʑi ɕi *ʑi ȵi i		*ki *kʰi *gi ʔi *hi *ɦi *ŋi
Daoxian Shouyan (K. He 2003) 道縣壽雁[湖南]	pi pʰi mi	ti tʰi tsɿ tsʰɿ sɿ-si ni li	 tɕi tɕʰi ɕi i		ki kʰi xi ŋi
Guiyang (J. Fan 2000) 桂陽敖全土話[湖南]	pi pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	*ȶi *ȶʰi tʃi tʃʰi ʃi ȵi i		ki *kʰi *hi *ŋi
Hexian (Y. Deng 1996) 賀縣(蓮塘)[廣西] 客家話	pi pʰi *fi *vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tʃi tʃʰi ʃi ȵi ji		ki kʰi *hi *ŋi
Jinhua (Z. Cao 1994) 金華[浙江] 吳語-婺州片	pi pʰi bi fi vi mi	ti tʰi di tsɿ−tsi tsʰɿ−tsʰi dzɿ−dzi sɿ−si zɿ−zi ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *hi *ɦi *ŋi
Jinhua Shanhu (Z. Cao 2004) 金華珊瑚[浙江] 客家話	pi pʰi fi vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si *ni li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *x *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (Onsetless)*	*Retroflex*	*Velar (Glottal)*
Leqing (R. Cai 1999) 樂清[浙江]	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ-si zɿ-zi ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *hi *ɦi *ŋi
Linhai (X. Huang 2007) 臨海[浙江] 吴語-台州片	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ–si zɿ li *ni	ci cʰi ɟi tɕi tɕʰi *dʑi ɕi ʑi çi ɲi i		*ki *kʰi *gi ɦi *hi *ŋi
Linwu (H. Chen 2002) 臨武(麥市)[湖南]	pi pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tʃi tʃʰi ʃi i		*ki *kʰi *xi *ŋi
Loudi (Li *et al*. 1987) 婁底[湖南]	pi pʰi bi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi dzɿ-*dʑi sɿ-si ni li	 tɕi tɕʰi dʑi ɕi *ʑi i		*ki *kʰi ɣi *xi *ŋi
Longchuan (X. Hou 2008) 龍川(佗城)[廣東] 粵語-粵中片	pi pʰi *fi *vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tʃi tʃʰi ʃi ȵi i		ki kʰi *hi *ŋi
Pingdiyao (X. Li 2015) 平地瑤(七都)[湖南]	pi *pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si *ni li	 tʃi tʃʰi ʃi *ʒi i		ki kʰi xi ŋi
Pingyang (C. Chen 1979) 平陽[浙江] 吴語	pi pʰi bi fi vi mi	ti tʰi di tsɿ–tsi tsʰɿ–tsʰi dzɿ–dzi sɿ–si zɿ–zi *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki kʰi gi ʔi *hi *ɦi *ŋi
Shanghai (Xu &Tao 1995) 上海(老派)[上海] 吴語	pi pʰi bi ɸi βi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi sɿ-si zɿ-zi ni li	 tɕi tɕʰi dʑi ɕi ȵi i		*ki *kʰi *gi ɦi *hi *ŋi
Suzhou (X. Ye 1992) 蘇州(老派)[江蘇]	pi pʰi bi fi vi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi sɿ-si zɿ-zi ni li	 tɕi tɕʰi dʑi ɕi ȵi i		*ki *kʰi *gi ɦi *hi *ŋi
Taishun (Siqian) (Z. Liu 2007) 泰順司前[浙江]	pi pʰi *fi mi	ti tʰi tsɿ-tsi tsʰɿ-*tsʰi sɿ-si *ni li	 tɕi tɕʰi ɕi ȵi i		ki kʰi *hi *ŋi
Taojiang (Zhang *et al*. 1988) 桃江(高橋)[湖南]	pi pʰi *fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si zɿ-zi li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal* *(No onset)*	*Retroflex*	*Velar* *(Glottal)*
Wannan-Chuanshang (X. Huang 2013) 皖南(船上)[安徽]	pi pʰi fi vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tɕi tɕʰi ɕi ȵi i		ki *kʰi *xi *ŋi
Xiuning (S. Hirata 1982) 休寧[安徽]	pi pʰi fi vi mi-mɿ	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si li-lɿ	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Yongzhou (X. Li 2003) 永州嵐角山[湖南]	pi pʰi fi vi mi	ti tʰi di tsɿ tsʰɿ sɿ−si zɿ−zi *ni li	 tɕi tɕʰi ɕi *ʑi ȵi i		*ki *kʰi *xi *ɣi *ŋi

2.1.2 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive***

2.1.2.1 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – **sɿ-ɕi**: 56 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Anxiang (Y. Ying 1988) 安鄉[湖南] ?湘語/西南官話	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi i	 *ʐi	*ki *kʰi *xi *ŋi
Babu (Ertang) (Y. Mai 2008) 八步(鵝塘)［廣西］	*pi *pʰi *fi *mi	ti *tʰi tsɿ tsʰɿ sɿ *ni *li	 tʃi tʃʰi ʃi ji	 ɲi	ki kʰi *kʷi *kʷʰi xi *ŋi *ŋʷi *wi
Chengbu (H. Bao 1993) 城步(儒林)[湖南] 湘語-婁邵片	pi pʰi bi *fi *vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *xi *ɣi *ŋi
Chengdu (D. Liang 1993) 成都[四川]	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Danyang (G. Cai 1994) 丹陽[安徽] 吳語-常州/江淮官話	pi pʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ *ni li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Danzhai (L. Li 1994) 丹寨[貴州]	pi pʰi vi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ɣi *ŋi
Daoxian (Xiaojia) (X. Zhou 1994) 道縣(小甲)[湖南]	pi pʰi fi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i		ki kʰi *xi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal* *(No onset)*	*Retroflex*	*Velar (Glottal)*
Dayü (L. Liu 1995) 大餘(南安)[江西]	pi pʰi fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *li	 tɕi tɕʰi ɕi ȵi *i		*ki *kʰi *hi *ŋi
Fenghuang (Q. Li 2011) 鳳凰[湖南] 官話-西南-黔北小片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi
Fuyang (Q. Wang 2012) 阜陽[安徽] 官話-中原-鄭曹片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ɣi
Guangfeng (Hiroyuki 2000) 廣豐[江西] 吳語-處衢片	pi pʰi bi fi mi	ti tʰi di tsɿ *tsʰɿ *dzɿ sɿ *ni li	 tɕi tɕʰi dʑi ɕi ȵi i		ki kʰi gi xi *ŋi
Guanyang (Guanyinge) (Q. Huang 1994) 灌陽(觀音閣)[廣西]	pi pʰi *ɸi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi
Guidong (Z.Cui 1997) 桂東[湖南]	pi pʰi fi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Hengshan (B. Mao 1995) 衡山[湖南] 湘語-長益片	pi pʰi *fi *mi	ti tʰi tsɿ tsʰɿ sɿ li	*ȶi *ȶʰi tɕi tɕʰi ɕi *ȵi i		*ki *kʰi *xi *ŋi
Jiangjin (W. Zhong 2002) 江津[四川] 官話-西南	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *zɿ ni	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi
Jiaxing (G. Yu 1988) 嘉興(老派)[浙江]	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ sɿ zɿ *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *ʔi *hi ɦi *ŋi
Lanxian (M. Shen 2009) 嵐縣[山西] 晉語-呂梁片-汾州小片	pi pʰi *fi mi	ti *tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i	 *ʐi	*ki *kʰi *xi *ŋi
Leiyang (L. Zhong 1987) 耒陽[湖南] 贛語	pi pʰi fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ li	*ȶi *ȶʰi tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Lezhi (R. Cui 1988) 樂至(靖州腔)[四川] 湘語(新湘語)	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni	 tɕi tɕʰi ɕi ȵi i		ki kʰi *xi *ŋi
Lianshui (S. Hu 1989) 漣水[江蘇]	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ *zɿ ni	 tɕi tɕʰi ɕi i		*ki *kʰi *xi
Linwu (K. Yin 2014) 臨武(楚江)[湖南]	pi pʰi fi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i		ki kʰi xi *ŋi
Liuyang *(Nanxiang)* (J. Xia 1983) 瀏陽南鄉[湖南]	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi i		ki kʰi *xi ŋi *ui
Luxi (Baisha) (J. Qu 2008) 瀘溪(白沙)[湖南]	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *ɡi *xi *ɣi *ŋi
Lüsi (J. Lu 1986) 呂四[江蘇] 吳語	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *ɡi *xi *ɣi *ŋi
Mouping (F. Luo 1995) 牟平[山東] 官話-膠遼-登連片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	ci cʰi tɕi tɕʰi ɕi çi ȵi i		*ki *kʰi *xi
Nanchang (Z. Xiong 1994) 南昌[江西]	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *hi *ŋi
Nancheng (S. Qiu 1991) 南城[江西] 贛語-撫廣片	pi pʰi *fi mi	ti tʰi tsɿ sɿ ni (li)	 tɕi tɕʰi ɕi i		*ki *kʰi hi *ŋi
Nanping (H. Su 1994) 南平[福建] 官話-北方(方言島)	pi pʰi mi	ti tʰi tsɿ tsʰɿ sɿ li	 *tɕi *tɕʰi ɕi i		ki kʰi xi *ŋi
Neihuang (X. Li 2012) 內黃[河南] 官話-中原-鄭曹片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		ki *kʰi *xi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Qingshuipin (Xianghua) (Zhao & Li 2014) 清水坪(沅陵)[湖南]	pi pʰi bi *fi *vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ ni	 tɕi tɕʰi dʑi ɕi ȵi i		*ki *kʰi *gi *xi *ŋi
Qingyuan (Hiroyuki & Cao 1998) 慶元[浙江]	pi pʰi ʔbi fi *mi	ti *tʰi ʔdi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi *ȵi i		*ki *kʰi *xi *ŋi
Shaoxing (F. Wang 2008) 紹興[浙江] 吳語	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *hi ɦi *ŋi
Shaoyang (H. Bao 1989) 邵陽[湖南] 湘語-婁邵片	pi pʰi bi *fi *vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ ni	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *xi *ɣi *ŋi
Shangcheng(Nansi) (Y. Yang 2008) 商城[河南]	pi pʰi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi
Suiping (S. Ding 1989) 遂平[河南]	pi pʰi fi mi	ti tʰ tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi
Taihu (S. Huang 2011) 太湖[安徽] 贛語-懷岳片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ ni li ɿ	 tɕi tɕʰi ɕi *ȵi i		*ki *kʰi *xi *ŋi
Taiping (Xianyuan) (S. Zhang 1991) 太平(仙源)[安徽] 吳語-宣州片	pi pʰi bi *fi *vi mi	ti tʰi di tsɿ tsʰɿ sɿ zɿ *li	 tɕi tɕʰi ɕi ʑi ȵi i		ki kʰi gi *xi *ɣi
Taixing (Q. Gu 1990) 泰興[江蘇] 官話-江淮	pi pʰi *fi *vi mi	*ti *tʰi tsɿ tsʰɿ sɿ *zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi
Taiyuan (M. Shen 1993) 太原[山西]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ɣi
Taizhou (Y. Yang 1991) 泰州(老派)[江蘇] 官話-江淮-泰如片	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *zɿ ni	 tɕi tɕʰi ɕi i		*ki *kʰi *xi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Tongling (T. Wang 1983) 銅陵[安徽] 吳語(方言島)	pi pʰi *fi vi mi	ti tʰi ɾi tsɿ tsʰɿ sɿ zɿ *ni li	 tɕi tɕʰi ɕi ʑi ȵi i		*ki *kʰi *xi *ɣi *ŋi
Wuning (Lixi) (M. Zhong 2004) 武寧(禮溪)[江西] 贛語	pi pʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi i		ki kʰi xi *ŋi
Xiaoshan (J. Zhang 1997) 蕭山[浙江] 吳語-太湖片(臨紹)	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *hi ɦi *ŋi
Xinghe (D. Li 1986) 興和[內蒙古] 晉語	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi *ŋi
Xintian (Nanxiang) (Q. Xie 2004) 新田南鄉土話[湖南]	pi pʰi fi *mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Xinyü (X. Wang 2010) 新餘[江西] 官話-江淮	pi pʰi *θi mi	*ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi-ȵɿ i		*ki *kʰi *hi *ŋi
Yancheng (X. Su 1993) 鹽城[江蘇] 官話-江淮	pi pʰi *fi mi	ti tʰi tsɿ tsʰ ɿ sɿ ni li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi
Yangshuo (L. Li 2015) 陽朔[廣西]	pi pʰi pɦi *fi *fɦi mi mɦi	ti tʰi tɦi tsɿ tsʰɿ tsɦɿ sɿ *sɦɿ ni nɦi li lɦi	 tɕi tɕʰi tɕɦi ɕi ɕɦi *ʑi i		ki kʰi kɦi xi *xɦi ŋi ŋɦi
Yangzhou (Wang & Huang 1993) 揚州[江蘇]	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi i		*ki *kʰi *xi
Yinxian (Z. Chen 1990) 鄞縣[浙江] 吳語-太湖片	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ li *ni	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi *hi ɦi *ŋi
Yixing (Ye & Guo 1991) 宜興[江蘇] 吳語-太湖片(毗陵)	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ *ni li	 tɕi tɕʰi dʑi ɕi ʑi ȵi i		*ki *kʰi *gi ɦi *hi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Yiyang (Y. Zeng 1995) 益陽[湖南]	pi pʰi bi *fi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ li	 tɕi tɕʰi dʑi ɕi ȵi i		*ki *kʰi *gi *hi *ŋi
Yongxing (S. Hu 2009) 永興(老派)[湖南] 贛語	pi-pɿ pʰi-pʰɿ fi mi−mɿ	ti tʰi tsɿ tsʰɿ sɿ li-lɿ	 tʃi tʃʰi ʃi ʒi ȵi i		*ki *kʰi *xi
Youxian (Z. Dong 1990) 攸縣[湖南] 贛語	pi pʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi i		*ki *kʰi *xi *ŋi
Yunhe (Ota & Cai 1998) 雲和[浙江]	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ *ni li	 tʃi *tʃʰi *dʒi *ʃi *ʒi ɲi i		ki kʰi gi *xi *ŋi
Zhaji (Liu & Chen 2015) (涇縣)查濟[安徽] 吳語-宣州片-銅涇小片	pi pʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi i		ki kʰi *xi *ŋi

2.1.2.2 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – ***si-ɕi** : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Shibei (A. Hiroyuki 2004) 石陂[福建] 閩語-閩北區	pi pʰi bi mi	ti tʰi di *tsi *tsʰi *dzi *si ni li	 tɕi tɕʰi dʑi ɕi i		ki kʰi gi *ʔi xi ɦi ŋi

2.1.2.3 Sibilants at *2* places– Dental vs. Palatal – ***Non-Contrastive*** – **(sɿ-) si-*ɕi** : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jixi (R. Zhao 1989) 績溪[安徽]	pi-pɿ *pʰi-pʰɿ fi vi mi-mɿ	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si zɿ-*zi ni-mɿ	 *tɕi *tɕʰi *ɕi *ȵi *i		ki kʰi xi *ŋi

2.2 Sibilants at *2* places– Dental vs. Retroflex

2.2.1 Sibilants at *2* places– Dental vs. Retroflex – Contrastive –**sɿ-si-ʂi:** 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal(No onset)*	*Retroflex*	*Velar (Glottal)*
Wuhua	pi pʰi	ti tʰi			ki kʰi
(Y. Wei 1997)		tsɿ-tsi tsʰɿ-tsʰi		tʂi tʂʰi	
五華[廣東]	fi vi	sɿ-si		ʂi	hi
客家話	mi	ni li	i		*ŋi

2.2.2 Sibilants at *2* places– Dental vs. Retroflex – Non-contrastive –**sɿ-si-ʃʅ**: 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jinggangshan	pi pʰi	ti tʰi			ki kʰi
(S. Lu 1995)		tsi-tsɿ tsʰi-tsʰɿ		tʃʅ tʃʰʅ	
井岡山[江西]	fi vi	si-sɿ		ʃʅ	xi
客家話	mi	ni li		i	ŋi

2.3 Sibilants at *2* places– Palatal vs. Retroflex

Sibilants at *2* places– Palatal vs. Retroflex – Non-contrastive – **ɕi-ʂʅ**: 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Haizhou	pi *pʰi	ti tʰi			*ki *kʰi
(X. Su 1990)			tɕi tɕʰi	tʂʅ tʂʰʅ	
海州[江蘇]	*fi		ɕi	ʂʅ *ʐʅ	*xi
	mi	li	i		

3 Sibilants at 3 or more Places

3.1 Sibilants at *3* places– Dental vs. Palatal vs. Retroflex

3.1.1 Sibilants at *3* places– ***3-place Contrasts* – (**sɿ-)**si-ɕi-ʂi:** 2 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Datan	pi pʰi	ti tʰi			*ki *kʰi
(Y. Chen 2015)		tsɿ-*tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂi tʂʰi	
大坦[安徽]	fi	sɿ-si	ɕi	*ʂi	*xi
徽語-祁德片	mi	*ni li	i		*ŋi
Qimen	pi pʰi	ti tʰi			*ki *kʰi
(T. Shen 1989)		tsɿ-*tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂi tʂʰi	
祁門[安徽]	fi	sɿ-si	ɕi	ʂi	xi
	mi	(ni) li	ȵi i		*ŋi

3.1.2 Sibilants at *3* places– 2-place contrasts – Contrastive dental vs. palatal

3.1.2.1 Sibilants at *3* places–***2-place Contrasts -1-* (**sɿ-)**si-ɕi-ʂʅ:** 12 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Binxian	pi pʰi	ti *tʰi		*ʈi *ʈʰi	*ki *kʰi
(Qiao & Chao 2002)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
彬縣[陝西咸陽]	*fi	sɿ-si zɿ-*zi	ɕi	ʂʅ ʐʅ	*xi
官話-中原-關中片	mi	li	ȵi i		*ŋi
Fufeng	pi pʰi	*ti *tʰi			*ki *kʰi
(X. Wu 1997)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
扶風[陝西]	*fi	sɿ-si	ɕi	ʂʅ ʐʅ	*xi
	mi	li	ȵi i		*ŋi
Ganyü	pi pʰi	ti tʰi	ci cʰi		*ki *kʰi
(B. Liu 1990)		tsɿ-tsi tsʰɿ-tsʰi	tʃi tʃʰi	tʂʅ tʂʰʅ	
贛榆(劉溝)[江蘇]	*fi *ʋi	sɿ-si	ʃi çi	ʂʅ	*xi
官話-中原-鄭曹片	mi	ni li	i		*ŋi
Kaifeng	pi pʰi	ti tʰi			*ki *kʰi
(D. Liu 1997)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
開封(老派)[河南]	fi	sɿ-si	ɕi	ʂʅ ʐʅ	*xi
官話-中原	mi	ni li	i		
Luoyang	pi pʰi	ti tʰi			*ki *kʰi
(W. He 1984)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
洛陽(老城區)[河南]	fi vi	sɿ-si	ɕi	ʂʅ ʐʅ	*xi
	mi	ni li	i		
Miluo	pi pʰi	ti tʰi		*ʈi *ʈʰi	ki kʰi
(S. Chen 2006)		tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	
汨羅長樂鎮[湖南]	*fi	sɿ-si	ɕi	ʂʅ	*xi
湘語-長益片	mi	li	ȵi i		ŋi
Nanjing	pi pʰi	ti tʰi			ki kʰi
(D. Liu 1994)	*fi	tsɿ-tsi tsʰɿ-tsʰi	tɕi tɕʰi	tʂʅ tʂʰʅ	*xi
南京(老派)[江蘇]		sɿ-si	ɕi	ʂʅ ʐʅ	
吳語	mi	li	i		

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Ningling (Y. Duan 2013) 宁陵[河南] 官話-中原-鄭曹片	pi pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ	*ki *kʰi *xi *ŋi
Pingxiang (G. Wei 1995) 萍鄉[江西]	pi pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ *ɳi	*ki *kʰi *hi *ŋi
Xiangcheng (S. Liu 1993) 襄城[河南] 官話-中原-鄭曹片	pi pʰi fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ	*ki *kʰi *xi
Xiangtan (Y. Zeng 1993) 湘潭(老派)[湖南] 湘語-長衡片	pi pʰi bi *ɸi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi dzɿ-dzi sɿ-si ni	 tɕi tɕʰi dʑi ɕi ȵi i	 tʂʅ tʂʰʅ dʐʅ ʂʅ	*ki *kʰi *gi *hi *ɦi *ŋi
Yüdu (L. Xie 1997) 于都(老派)[江西] 客家話-于桂片	pi pʰi fi vi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ	*ki *kʰi *hi *ŋi

3.1.2.2 Sibilants at *3* places–**2-place Contrasts -2- (sɿ-)si-ɕi-*ʂi:** 3 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Changshu (Meili) (D. Yuan 2010) 常熟(梅李) [江蘇] 吳語太湖片(蘇滬嘉)	pi pʰi bi fi vi mi	ti tʰi di tsɿ-tsi tsʰɿ-tsʰi dzɿ-dzi sɿ-si zɿ-zi *ni li	 tɕi tɕʰi dʑi ɕi ȵi ji i	 *tʂi *tʂʰi *dʐi *ʂi *ʐi	*ki *kʰi *gi *hi *ɦi *ŋi
Fuliang (Jiuchengcun) (L. Xie 2011) 浮梁(舊城村)[江西] 徽語-祁德片	pi pʰi *fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni *li	 tɕi tɕʰi ɕi i	 *tʂi *tʂʰi *ʂi	*ki *kʰi *xi *ŋi
Yügan (C. Chen 1990) 餘幹[江西]	pi pʰi *ɸi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si li	 tɕi tɕʰi ɕi ȵi i	 *tʃi *tʃʰi *ʃi *ɳi	*ki *kʰi *hi *ŋi

3.1.3 Sibilants at *3* places– 2-place contrasts – Contrastive Palatal vs. Retroflex : 1 dialect

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex(?)*	*Velar (Glottal)*
Xuancheng (Shen & Huang 2015) 宣城(雁翅)[安徽] 吳語-宣州片-銅涇小片	pi pʰi pfɦi pɿ pʰɿ *fi *vi vɿ mi mɿ	ti tʰi tɦi tɦɿ tsɿ tsʰɿ *tsɦɿ sɿ zɿ ni li	 tɕi tɕʰi tɕɦi ɕi i	 tʃi tʃʰi *tʃɦi ʃi ʒi	*ki *kʰi *xi *hɦi *ŋi

3.1.4 Sibilants at *3* places– ***Non-contrastive***

3.1.4.1 Sibilants at *3* places– ***Non-contrastive-1-* sɿ-ɕi-ʂʅ -A**: 20 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Anqing (H. Bao 2012) 安慶[安徽] 官話-江淮-洪巢片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ ni	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ɻʅ 	*ki *kʰi *xi *ŋi
Chenxi (B. Xie 2010) 辰溪[湖南] 湘語-吉溆片	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ ni *li	 tɕi tɕʰi dʑi ɕi i	 tʂʅ tʂʰʅ dʐʅ ʂʅ ʐʅ 	ki *kʰi *gi *xi *ŋi
Harbin (S.Yin 1995) 哈爾濱[黑龍江] 	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi
Huangyuan (L. Lu 2011) 湟源[青海] 官話-中原-秦隴片	pi-pɿ pʰi-pʰɿ *fi mɿ	*ti tʰi tsɿ tsʰɿ sɿ *zɿ ni *li ɿ	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *hi
Jimsar (L. Zhou 1991) 吉木薩爾[新疆] 官話-蘭銀-北疆片	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Ji'nan (Z. Qian 1995) 濟南[山東] 	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Linyi (Y. Cao (2005) 臨邑[山東] 官話-冀魯-滄惠片	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Luxi (J. Qu 2005) 瀘溪(浦市)[湖南] 湘语-吉溆片	pi pʰi bi fi vi mi	ti tʰi di tsɿ tsʰɿ dzɿ sɿ zɿ ni	 tɕi tɕʰi dʑi ɕi *ʑi i	 tʂʅ tʂʰʅ dʐʅ ʂʅ ʐʅ 	*ki *kʰi *gi *xi *ŋi
Minqin (K. Wu 2009) 民勤[甘肅] 	pi pʰi *fi vi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni *li	 tɕi tɕʰi ɕi ʑi *i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ɣi
Pingli (Luohe) (Z. Zhou 2005) 平利洛河[陝西] 	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ *ɳi	*ki *kʰi *xi *ŋi

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Susong (Y. Sun 2002) 宿松[安徽] 贛語-懷岳片	pi pʰi *fi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i	 tʃʅ tʃʰʅ ʃʅ *ʒʅ 	*ki *kʰi *hi *ŋi
Turpan (L. Zhou 1998) 吐魯番[新疆]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi
Urumchi (L. Zhou 1994) 烏魯木齊[新疆] 官話-蘭銀-北疆片	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Wubaodong (X. Xing 2011) 吳堡東(王家山)[陝西] 晉語-呂梁片	pi pʰi *fi mi	*ti *tʰi tsɿ tsʰɿ sɿ *zɿ ni *li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ *ʂʅ *ʐʅ 	*ki *kʰi *xi *ŋi
Xinzhou (Wen & Zhang 1994) 忻州[山西]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ (zɿ) ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ *ʐʅ 	*ki *kʰi *xi *ŋi
Xingxian (Shi & Zhang 2014) 興縣[山西] 晉語-呂梁片-汾州小片	pi pʰi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ *ʐʅ 	*ki *kʰi *xi *ŋi
Xuzhou (Su & Lü 1994) 徐州[江蘇]	pi pʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi
Yanqi (L. Liu 1988) 焉耆[新疆]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Yibin (Wangchang) (F. Zuo 1995) 宜賓王場[四川]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Yinchuan (Li & Zhang 1995) 銀川[寧夏]	pi pʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi

3.1.4.2 Sibilants at *3* places– ***Non-contrastive-1-* sɿ-ɕi-ʂʅ -B**: **(with-labial affricates)**: 8 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jiangxian (L. Wang 2014) 绛縣[山西] 官話-中原-汾河片	pi pʰi pfi pfʰi fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	ki *kʰi *xi *ɣi *ŋi
Lanzhou (B. Gao 1980) 蘭州[甘肅] 	pi pʰi *pfi *pfʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ (zɿ) *ni	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	ki *kʰi *xi
Linyi (Shanxi) (L. Wang 2003) 臨猗[山西] 官話-中原-汾河片	pi pʰi *pfi *pfʰi *fi *vi mi	ti tʰi tsɿ tsʰɿ sɿ zɿ li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Loufan (Y. Zhai 1989) 婁煩(老派)[山西] 	pi pʰi *pfi *pfʰi *fi *vi mi	ti *tʰi tsɿ tsʰɿ sɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ *ʐʅ 	*ki *kʰi *xi *ŋi
Wanrong (J. Wu 1996) 萬榮[山西] 	*pi *pʰi *pfi *pfʰi *fi *vi *mi	ti tʰi tsɿ tsʰɿ sɿ zɿ *ni li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Wudu (J. Shi 1992) 五都[甘肅] 	pi pʰi *tsfi *tsfʰi *fi *vi *sfi *zfi mi	ti tʰi tsɿ tsʰɿ sɿ *zɿ li	 tɕi tɕʰi ɕi ȵi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ 	*ki *kʰi *xi *ŋi
Xi'an (J. Wang 1995) 西安[陝西] 官話-中原-關中片	pi pʰi *pfi *pfʰi fi vi mi	ti tʰi tsɿ tsʰɿ sɿ ni li	 tɕi tɕʰi ɕi *ȵi i	 tʂʅ tʂʰʅ ʂʅ *ʐʅ 	*ki *kʰi *xi *ŋi
Zhangye (D. Huang 2009) 張掖[甘肅] 官話-蘭銀	pi pʰi *pfi *pfʰi *fi *vi mi	*ti *tʰi tsɿ tsʰɿ sɿ *ni li	ȶi ȶʰi tɕi tɕʰi ɕi ʑi ȵi i	 tʂʅ tʂʰʅ ʂʅ *ʐʅ 	*ki *kʰi *ʀi *xi *ŋi

3.1.4.3 Sibilants at *3* places– ***Non-contrastive-1***- **sɿ-ɕi-*ʂi**: 2 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Dangtu (Huyang) (Zheng *et al.* 2012) 當涂(湖陽)[安徽] 吳語-宣州片(太高)	pi pʰi bi *fi mi	ti tʰi ɾi tsɿ tsʰɿ dzɿ sɿ zɿ *ni *li ɿ	 tɕi tɕʰi *dʑi ɕi ʑi ȵi i	 *tʃi *tʃʰi *ʃi *ʒi *ɻi	*ki *kʰi *gi *xi *ɣi *ɦi *ŋi
Taizhou (X. Lin 2012) 台州[浙江] 吳語-台州片	pi pʰi pɦi fi fɦi mi mɦi	ti tʰi tɦi tsɿ tsʰɿ tsɦɿ sɿ sɦɿ *ni *nɦi li lɦi	 tɕi tɕʰi tɕɦi ɕi ɕɦi ȵi ȵɦi i	 *tʃi *tʃʰi *tʃɦi *ʃi *ʃɦi	*ki *kʰi *kɦi *hi ɦi *ŋi ŋɦi

3.1.4.4 Sibilants at *3* places– ***Non-contrastive-1***- **sɿ-*ɕi-ʂʅ**: 2 dialects

Dialects	*Labial*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Ledu (Cao & Shao 2001) 樂都[青海]	pɿ pʰɿ *fi *vi mɿ	*ti *tʰi tsɿ tsʰɿ sɿ zɿ ni lɿ	 *tɕi *tɕʰi *ɕi	 tʂʅ tʂʰʅ ʂʅ ʐʅ	*ki *kʰi *hi
Lujiang (Y. Zhou 2001) 廬江[安徽] 官話-江淮	pɿ pʰɿ *fi *vi mɿ	ti tʰi tsɿ tsʰɿ sɿ zɿ ni-mɿ *li	 *tɕi *tɕʰi *ɕi i	 tʂʅ tʂʰʅ ʂʅ ʐʅ	ki *kʰi *xi *ŋi

3.2 Sibilants at *4* places– Dental vs. Palatal vs. Retroflex: 1 dialect

Dialects	*Labial/Dental*	*Dental (Alveolar)*	*Palatal (No onset)*	*Retroflex*	*Velar (Glottal)*
Jüxian (M. Shi 1987) 莒縣[山東] 官話-北方	pi pʰi tθɿ tθʰɿ θɿ *fi mi	ti tʰi tsɿ-tsi tsʰɿ-tsʰi sɿ-si ni li	 tɕi tɕʰi ɕi i	 tʂʅ tʂʰʅ ʂʅ	*ki *kʰi *xi *ŋi

Appendix III. Chinese characters/syllables examined for the historical development

Notes:

1. The Chinese characters/syllables in the lists are those
   (a) whose pronounciations are [sɹ̩ ɕi ʂɻ̩], [tsɹ̩ tɕi tʂɻ̩], or [tsʰɹ̩ tɕʰi tʂʰɻ̩] in Mandarin, *and*
   (b) which appeared in the list in Li and Zhou (1999).
2. Below, separate lists are given to Mandarin syllables whose onsets are fricative (II-1), aspirated affricates (II-2), and unaspirated affricates (II-3).
   (a) Within each part, the Mandarin syllables in the first column are generally ordered by the Pinyin transcription and tonal marks, e.g., shī (施), shí(時), shǐ (矢), shì(市).
   (b) Whenever possible, the syllables sharing the same rime in Middle Chinese (MC) are listed together, e.g., sī(私) sǐ(死) sì(四)
3. The phonetic forms in the lists below are obtained from the reconstructions in Li and Zhou (1999), who followed the reconstructions in Wang (1980, 1985) and Ning (1985).
4. Within each list, the following abbreviations are used

   MC= Middle Chinese;

   EM= Early Mandarin;

   Mdr= Mandarin .
5. In Li and Zhou (1999), the following symbols are used:

   ĭ = a short high front vowel;

   ï = a cover symbol for apical vowels in Early Mandarin (around 18th century);

   ʅ = the dental apical vowel [ɹ̩] in Mandarin;

   ʅ = the retroflex apical vowel [ɻ̩] in Mandarin.

   These symbols are kept as they are in Li and Zhang (1999).

List II-1. Fricative-initial syllables

Pinyin &Character	*MC*	*EM*	*Mdr*
sī(思司)	sĭə	sï	sɿ
sī(斯廝)	sĭe	sï	sɿ
sī(私) sǐ(死) sì(四)	si	sï	sɿ
sì(似祀 寺嗣飼)	zĭə	sï	sɿ
sì(兕)	zi	sï	sɿ
sì(厮)	tʃʰĭə	tsʰï	sɿ
xī(僖嘻熹) xǐ(喜)	hĭə	hi	ɕi
xī(昔惜)	sĭɛk	si	ɕi
xī(羲)	hĭe	hi	ɕi
xī(希稀)	hĭəi	hi	ɕi
xī(醯)	hiei	hi	ɕi
xī(吸噏)	hĭĕp	hi	ɕi
xī(溪)	kʰiei	kʰi	ɕi
xī(奚兮) xì(系)	ɣiei	hi	ɕi
xī(西犀) xì(細)	siei	si	ɕi
xī (息熄)	sĭək	si	ɕi
xī(錫析)	siek	si	ɕi
xí(席蓆)	zĭɛk	si	ɕi
xí(習襲)	zĭĕp	si	ɕi
xǐ(徙)	sĭe	si	ɕi
xǐ (洗)	siei	si	ɕi
xǐ(璽)	sĭe	si	ɕi
xǐ(屣)	ʃĭe	si	ɕi
xì(戲)	hĭe	hi	ɕi
shī(濕/溼)	ɕĭĕp	ʂi	ʂʅ
shī(尸屍) shǐ (矢)	ɕi	ʂï	ʂʅ
shī(施) shǐ(弛)	ɕĭe	ʂï	ʂʅ
shī(詩) shǐ(始) shì(試弑)	ɕĭə	ʂï	ʂʅ
shī(失) shì(室)	ɕĭĕt	ʂi	ʂʅ
shí(實)	dʑĭĕt	ʂi	ʂʅ
shí(識)	ɕĭək	ʂi	ʂʅ
shí(石)	ʑĭɛk	ʂi	ʂʅ
shí(十什)	ʑĭĕp	ʂi	ʂʅ
shí(食蝕)	dʑĭək	ʂi	ʂʅ
shí(時) shì(市恃侍蒔)	ʑĭə	ʂï	ʂʅ
shì(嗜)	ʑi	ʂï	ʂʅ
shì(是氏)	ʑĭe	ʂï	ʂʅ
shì(示)	dʑi	ʂï	ʂʅ
shì(師獅)	ʃi	ʂï	ʂʅ
shì(史使)	ʃĭə	ʂï	ʂʅ
shì(士仕柿)	dʒĭə	ʂï	ʂʅ
shì(世)	ɕĭɛi	ʂi	ʂʅ
shì(釋)	ɕĭɛk	ʂi	ʂʅ
shì(適)	ɕĭɛk	ʂi	ʂʅ
shì(逝誓)	ʑĭɛi	ʂi	ʂʅ

List II-2. Aspirated-affricate-initial syllables

Pinyin &Character	*MC*	*EM*	*Mdr*
cī(差)	tʃʰĭe	tʂʰï	tsʰɿ
cī(疵)	dzĭe	tsʰï	tsʰɿ
cí(詞祠)	zĭə	sï	tsʰɿ
cí(慈鷀)	dzĭə	tʂʰï	tsʰɿ
cí(茨瓷)	dzi	tsʰï	tsʰɿ
cí(雌) cǐ(此)	tsʰĭe	tsʰï	tsʰɿ
cì(次)	tsʰi	tsʰï	tsʰɿ
cì(賜)	sĭe	sï	tsʰɿ
qī(戚)	tsʰiek	tsʰi	tɕʰi
qī(七漆)	tsʰĭĕt	tsʰi	tɕʰi
qī(棲)	siei	tsʰi	tɕʰi
qī(欺) qǐ(起杞)	kʰĭə	kʰi	tɕʰi
qī(妻淒) qì(砌)	tsʰiei	tsʰi	tɕʰi
qí(耆祁)	gi	kʰi	tɕʰi
qí(其旗)	gĭə	kʰi	tɕʰi
qí(奇騎)	gĭe	kʰi	tɕʰi
qí(祇芪軝)	gĭe	kʰi	tɕʰi
qí(祈頎)	gĭəi	kʰi	tɕʰi
qí(齊臍)	dziei	tsʰi	tɕʰi
qǐ(綺)	kʰĭe	kʰi	tɕʰi
qǐ(啟)	kʰiei	kʰi	tɕʰi
qǐ(乞)	kʰĭət	kʰi	tɕʰi
qǐ(豈) qì(气氣)	kʰĭəi	kʰi	tɕʰi
qì(棄器)	kʰi	kʰi	tɕʰi
qì(契)	kʰiei	kʰi	tɕʰi
qì(憩)	kʰĭɛi	kʰi	tɕʰi
qì(泣)	kʰĭĕp	kʰi	tɕʰi
qì(訖)	kĭət	kʰi	tɕʰi
chī(鴟)	tɕʰi	tʂʰi	tʂʰʅ
chī(蚩吃)	tɕʰĭə	tʂʰi	tʂʰʅ
chī(郗絺)	ȶʰi	tʂʰi	tʂʰʅ
chī(螭魑眵)	ȶʰĭe	tʂʰi	tʂʰʅ
chī(癡笞) chí(持) chǐ(耻)	ȶʰĭə	tʂʰi	tʂʰʅ
chí(馳池)	ȡĭe	tʂʰi	tʂʰʅ
chí(墀遲)	ȡi	tʂʰi	tʂʰʅ
chí(匙) chǐ(豉)	ʑĭe	ʂï	tʂʰʅ
chǐ(侈)	tɕʰĭe	tʂʰi	tʂʰʅ
chǐ(齒)	tɕʰĭə	tʂʰï	tʂʰʅ
chǐ(尺) chì(赤斥)	tɕʰĭɛk	tʂʰi	tʂʰʅ
chì(叱)	tɕʰĭĕt	tʂʰi	tʂʰʅ
chì(勅飭敕)	ȶʰĭək	tʂʰi	tʂʰʅ
chì(翅)	ɕĭe	tʂʰï	tʂʰʅ
chì(熾)	tɕʰĭə	tʂi	tʂʰʅ

List II-3. Unaspirated-affricate-initial syllables

Pinyin &Character	*MC*	*EM*	*Mdr*
zī(茲滋孳) zǐ(子)	tsĭə	tsï	tsɿ
zī(貲訾) zǐ(紫)	tsĭe	tsï	tsɿ
zī(咨資姿) zǐ(姊) zì(恣)	tsi	tsï	tsɿ
zì(自)	dzi	tsï	tsɿ
zì(漬)	dzĭe	tsï	tsɿ
zì(淄輜)	tʃĭə	tsï	tsɿ
jī(飢肌)	ki	ki	tɕi
jī(雞稽)	kiei	ki	tɕi
jī(機)	kĭəi	ki	tɕi
jī(激擊)	kiek	ki	tɕi
jī(敧)	kʰĭe	kʰi	tɕi
jī(積績) jǐ(脊)	tsĭɛk	tsi	tɕi
jī(齏躋) jǐ(濟水) jì(霽濟)	tsiei	tsi	tɕi
jī(姬基箕) jì(紀記)	kĭə	ki	tɕi
jī(羈羇) jì(寄)	kĭe	ki	tɕi
jí(急)	kĭĕp	ki	tɕi
jí(吉)	kĭĕt	ki	tɕi
jí(棘)	kĭək	ki	tɕi
jí(及笈)	gĭĕp	ki	tɕi
jí(極)	gĭək	ki	tɕi
jí(集輯)	dzĭĕp	tsi	tɕi
jí(疾嫉)	dzĭĕt	tsi	tɕi
jǐ(几)	ki	ki	tɕi
jǐ(蟣幾)　jì(既)	kĭəi	ki	tɕi
jì(冀驥)	ki	ki	tɕi
jì(計)	kiei	ki	tɕi
jì(薊)	kiei	ki	tɕi
jì(忌)	gĭə	ki	tɕi
jì(技妓騎)	gĭe	ki	tɕi
jì(偈)	giɛi	ki	tɕi
jì(悸痵)	gwi	kui	tɕi
jì(際祭)	tsĭɛi	tsi	tɕi
jì(稷)	tsĭək	tsi	tɕi
jì(劑)	dziei	tsi	tɕi
jì(寂)	dziek	tsi	tɕi
zhī(隻)	tɕĭɛk	tṣi	tṣʅ
zhī(汁)	tɕĭĕp	tṣi	tṣʅ
zhī(織)	tɕĭək	tṣi	tṣʅ
zhī(之芝)	tɕĭə	tṣï	tṣʅ
zhī(支枝肢)	tɕĭe	tṣï	tṣʅ
zhī(胝)	ʈi	tṣï	tṣʅ
zhī(脂) zhǐ(旨指)	tɕi	tṣï	tṣʅ
zhī(知) zhì(智)	ʈĭe	tṣi	tṣʅ
zhí(直)	ɖĭək	tṣi	tṣʅ
zhǐ(徵)	ʈĭə	tṣï	tṣʅ
zhǐ(祉)	ʈʰĭə	tṣï	tṣʅ
zhǐ(紙)	tɕĭe	tṣï	tṣʅ
zhǐ(止)	tɕĭə	tṣï	tṣʅ
zhì(制製)	tɕĭɛi	tṣi	tṣʅ
zhì(質)	tɕĭĕt	tṣi	tṣʅ
zhì(致)	ʈi	tṣi	tṣʅ
zhì(雉　稚)	ɖi	tṣi	tṣʅ
zhì(彘)	ɖĭɛi	tṣi	tṣʅ
zhì(滯)	ɖĭɛi	tṣi	tṣʅ
zhì(秩)	ɖĭĕt	tṣi	tṣʅ
zhì(志)	tɕĭə	tṣï	tṣʅ
zhì(擲)	ɖĭɛk	tṣi	tṣʅ

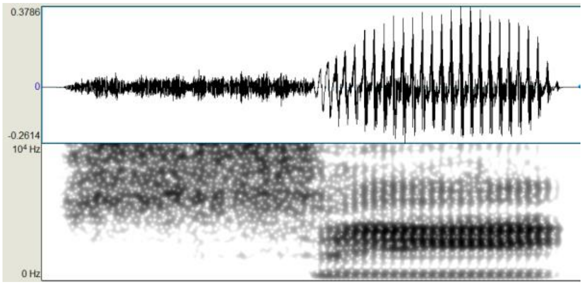
Appendix IV. Acoustic properties (mean values) of the naturally produced CV syllables

Notes: For vowel, the F0, F1, F2, F3 are measured at the mid-points.

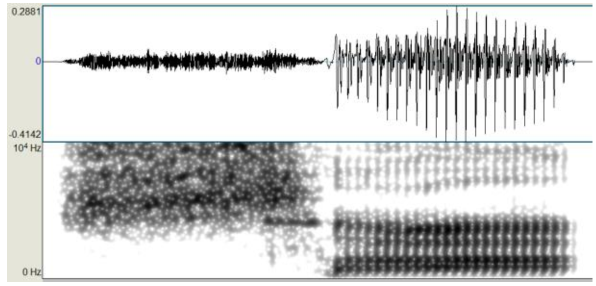
Syllable	Onset	Vowel					Transition
	Duration (ms)	Duration(ms)	F0(Hz)	F1(Hz)	F2(Hz)	F3(Hz)	Duration(ms)
si	176	355	295	311	2494	2894	87
ɕi	169	331	282	322	2586	2919	60
ʂi	195	321	283	315	2468	2938	125
sa	147	266	171	907	1358	2628	46
ɕa	169	283	166	916	1390	2522	86
ʂa	206	302	162	958	1369	2659	55
sɹ̩	157	328	304	420	1295	3189	
ʂɻ̩	176	278	287	344	1775	2551	
tsi	81	342	284	332	2564	2868	70
tɕi	63	315	285	317	2558	2940	53
tʂi	77	317	283	313	2471	2936	122
tsa	45	302	169	927	1315	2726	49
tɕa	72	315	160	929	1404	2483	106
tʂa	56	312	163	951	1373	2638	48
tsɹ̩	79	366	289	354	1343	3145	
tʂɻ̩	74	320	275	342	1888	2766	
tsʰi	165	325	288	320	2573	2980	70
tɕʰi	141	309	283	319	2544	2938	63
tʂʰi	156	306	279	322	2566	2943	101
tsʰa	152	268	168	942	1300	2670	48
tɕʰa	151	278	165	929	1395	2511	86
tʂʰa	144	286	167	927	1337	2661	52
tsʰɹ̩	154	307	298	394	1276	3214	
tʂʰɻ̩	145	275	281	454	1915	2784	

Appendix V. Waveforms and Spectrograms of the Stimulus Syllables

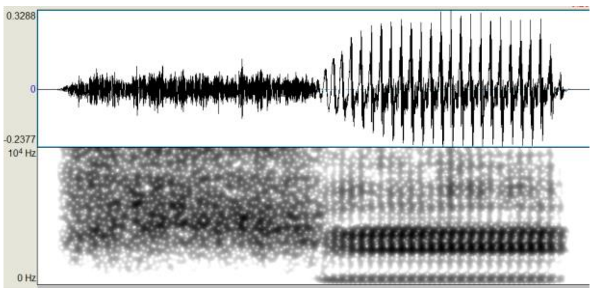
V-1 Stimulus syllables with fricative onsets



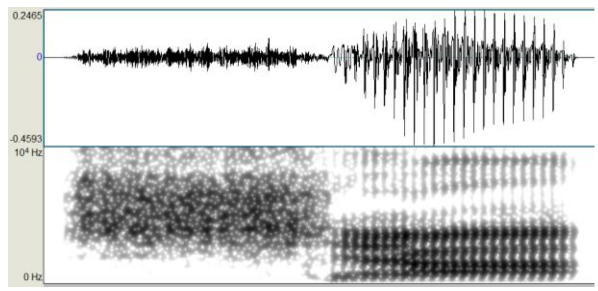
[si]



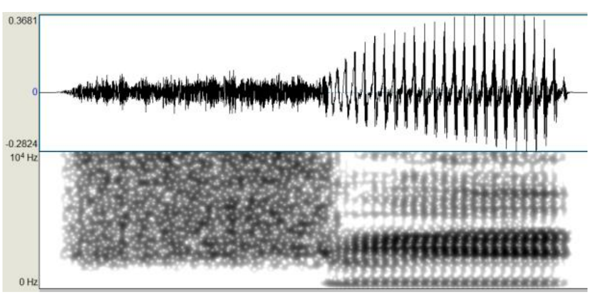
[sa]



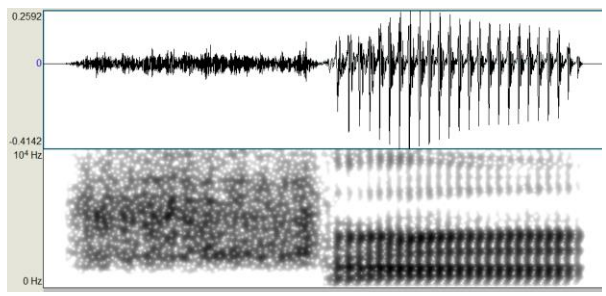
[ɕi]



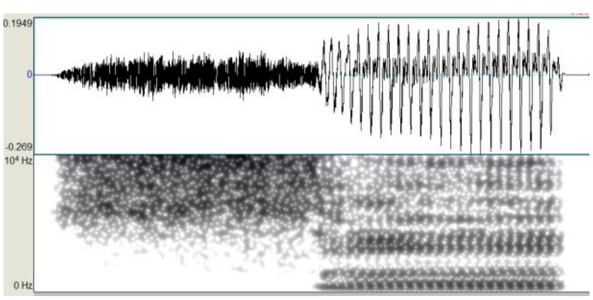
[ɕa]



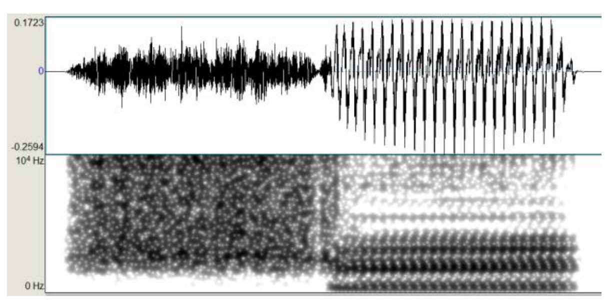
[ʂi]



[ʂa]

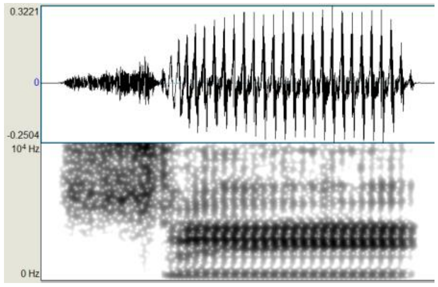


[sɹ̩]

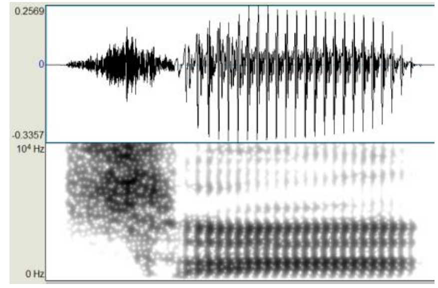


[ʂɻ̩]

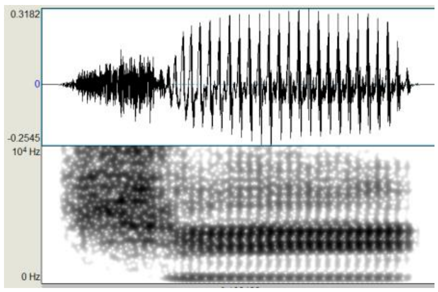
V-2 Stimulus syllables with unaspirated affricate onsets



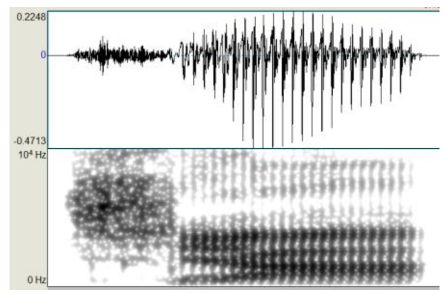
[tsi]



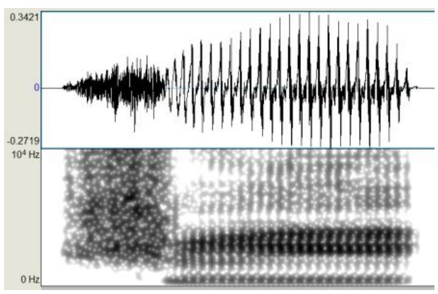
[tsa]



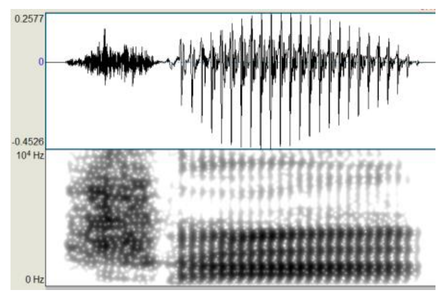
[tɕi]



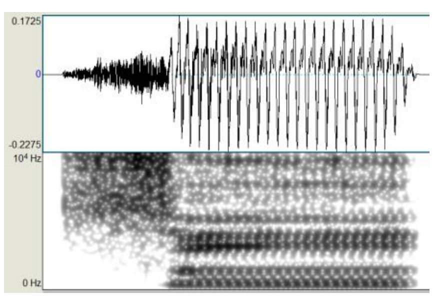
[tɕa]



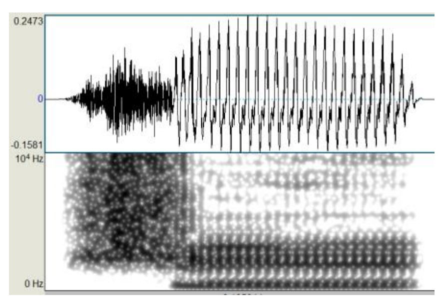
[tʂi]



[tʂa]

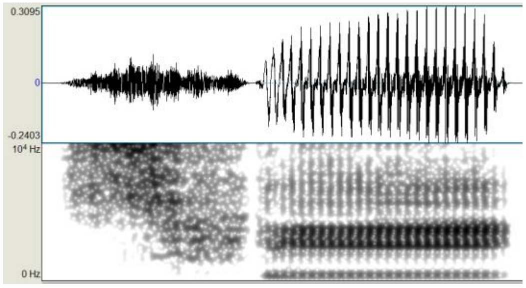


[tsɹ̩]

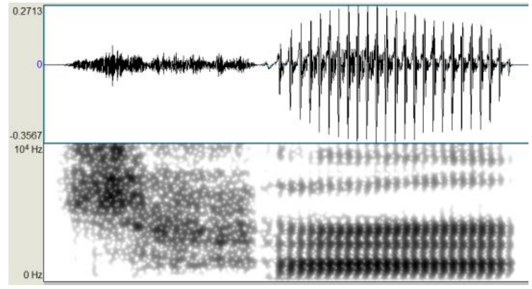


[tʂɻ̩]

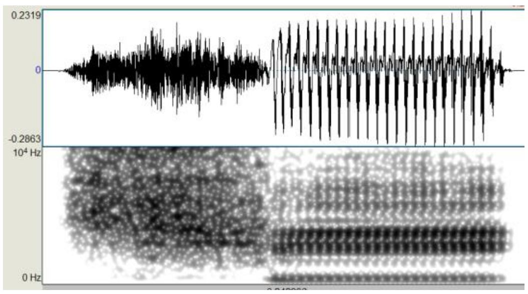
V-3 Stimulus syllables with aspirated affricate onsets



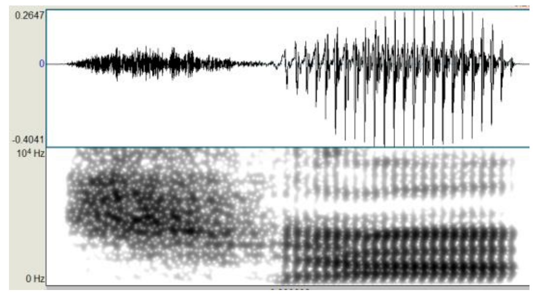
[tsʰi]



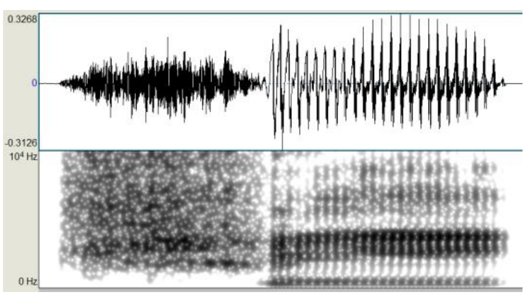
[tsʰa]



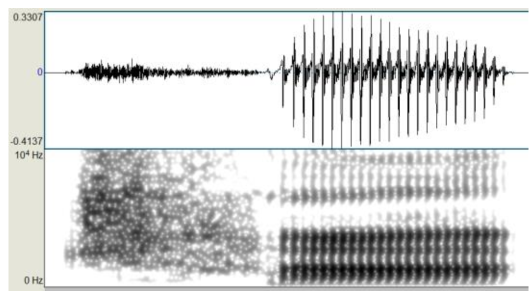
[tɕʰi]



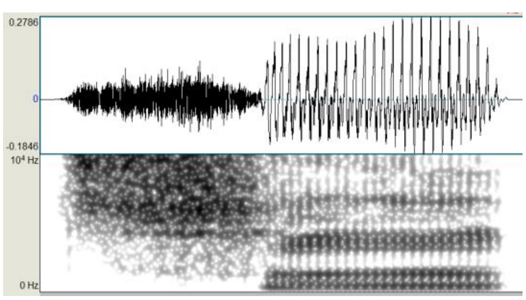
[tɕʰa]



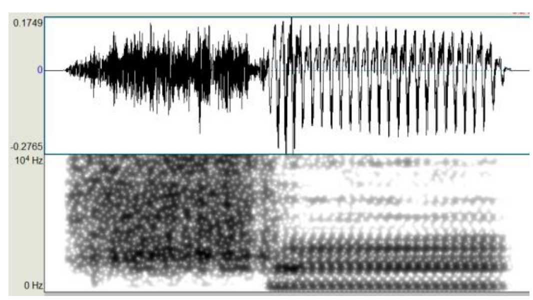
[tʂʰi]



[tʂʰa]



[tsʰɹ̩]



[tʂʰɻ̩]

References

Abercrombie, David. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Babel, Molly; Johnson, Keith. (2010). Accessing psycho-acoustic perception with speech sounds. *Journal of Laboratory Phonology* 1. 179-205.

Babel, Molly; McGuire, Grant. (2013). Listener expectations and gender bias in nonsibilant fricative perception. *Phonetica* 70. 117–151.

Bao, Hong. (2012). Anhui Anqing fangyan tongyi zihui [The homophonous syllabary of Anqing dialect, Anhui Province]. *Fangyan* 3:231-241.

Bao, Houxing. (1989). Hunan Shaoyang fangyan yinxi [Phonology of Shaoyang dialect in Hu'nan Province]. *Fangyan* 3:196-207.

Bao, Houxing. (1993). Hu'nan Chengbu (Rulin) fangyan yinxi [Phonology of Chengbu (Rulin) dialect in Hu'nan Province]. *Fangyan* 1:31-41.

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. (2015a). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9, https://CRAN.R-project.org/package=lme4.

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. (2015b). Fitting Linear Mixed-Effects Models using lme4. ArXiv e-print; in press, *Journal of Statistical Software*, http://arxiv.org/abs/1406.5823.

Beddor, Patrice Speeter. (2009). A coarticulatory path to sound change. *Language* 85(4): 785-821.

Best, Catherine T. (1995). A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 171–204. Baltimore, MD: York Press.

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

Bhat, D. N. S. (1978). A general study of palatalization. In Greenberg, Joseph H. (ed.) *Universals of Human Language*, 47-92. Stanford University Press.

Blevins, Juliette. (2016). Evolutionary phonology: A holistic approach to sound change typology. In Honeybone, Patrick; Salmons, Joseph (eds.) *The Oxford Handbook of Historical Phonology*. Oxford University Press.

Blumstein, S. E.; Stevens K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America* 66, 1001-1017.

Boersma, Paul. (1998). Functional phonology: Formalizing the interactions between articulatory and perceptual drives. Doctoral thesis, University of Amsterdam.

Boersma, Paul. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Boersma, Paul; Hamann, Silke. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217-270.

Booij, Geert. (1999). *The Phonology of Dutch*. New York: Oxford University Press.

Boomershine, A.; Hall, K. C.; Hume, E.; Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, E. Dresher & K. Rice (eds.), *Contrast in Phonology: Theory, Perception, Acquisition*, 145–171. New York: Mouton de Gruyter.

Bybee, Joan. (2001). *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.

Bybee, Joan. (2006). *Frequency of Use and the Organization of Language*. Oxford, UK: Oxford University Press.

Cai, Guolu. (1994). *Danyang Fangyan Cidian* yinlun [The introductory chapter in *A Dictionary of Danyang Dialect*]. *Fangyan* 2:103-118.

Cai, Quan. (1987). Guangxi Lianzhou fangyan yinxi [Phonology of Lianzhou dialect in Guangxi Autonomous Region]. *Fangyan* 1:49-57.

Cai, Rong. (1999). Zhejiang Leqing fangyan yinxi [Phonology of Leqing dialect in Zhejian Pronvince] *Fangyan* 4:266-276.

Cao, Jianfen. (1990). *Xiandai Yuyin Zhishi* [*Fundamentals of Modern Phonetics*]. Beijing: Renmin Jiaoyu Press (People's Education Press).

Cao, Yanjie. (2005). Shandong Linyi fangyan yinxi [Phonology of the Lingyi dialect in Shandong Province]. *Fangyan* 1:36-49.

Cao, Zhiyun. (1994). *Jinhua Fangyan Cidian* yinlun [The introductory chapter in *A Dictionary of Jinhua Dialect*]. *Fangyan* 4:252-267

Cao, Zhiyun. (2004). Zhejiang Jinhua Shanhu kejiahua yinxi [Phonology of the Shahu Hakka in Jinhua City, Zhejiang Pronince]. *Fangyan* 3:200-211.

Cao, Zhiyun; Shao, Zhaoyang. (2001). Qinghai Ledu fangyan yinxi [Phonology of Ledu dialect in Qinghai Province]. *Fangyan* 4:373-383.

Chang, Steve; Plauché, Madelaine C.; Ohala, John J. (2001). Markedness and consonant confusion asymmetries. In Hume, Elizabeth; Johnson, Keith (eds.) *The Role of Speech Perception in Phonology*. San Diego: Academic Press.

Chao, Yuen-Ren. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute for History and Philology* (*Academia Sinica*) 4. 363–397. Reprinted in Martin Joos (ed.), *Readings in Linguistics* (*3rd Edition*, 1963), 38–54. New York: American Council of Learned Societies.

Chen, Matthew Y. (1976). From Middle Chinese to Modern Peking. *Journal of Chinese Linguistics* 2. 113-227.

Chen, Changyi. (1990) Yügan fangyan tongyinzihui [Homophonic syllable inventory of Yügan dialect]. *Fangyan* 3:180-191.

Chen, Chengrong. (1979). Pingyan fangyan jilüe [The homophonous syllabary of Pingyang dialect]. *Fangyan* 1:47-74.

Chen, Hui. (2002). Hu'nan Lingwu(Maishi) Tuhua yüyin fenxi [The phonological analysis of the Linwu(Maishi) Tuhua in Hu'nan Province]. *Fangyan* 2:151-165.

Chen, Shanqing. (2006). Hu'nan Miluo Changle fangyan yinxi [Phonology of Changle dialect in Miluo County, Hunan Province]. *Fangyan* 1:56-71.

Chen, Xiaojin. (1999). Guangxi Rongxian Kejia fangyandao diaocha ji [An investigation of the Hakka dialect island in Rongxian, Guangxi Autonomous Region]. *Fangyan* 3:205-214.

Chen, Xiaoyan. (2009). Guangxi Hezhou Babu (Guiling) bendihua yinxi [Phonology of the local dialect of Babu (Guiling), in Hezhou, Guangxi Automomous Reigion]. *Fangyan* 1:53-71.

Chen, Yao. (2015). Anhui Huangshan Qimen Datan hua tongyinzihui [The homophonous syllabary of datan dialect in Qimen, Huangshang, Anhui]. *Fangyan* 4:367-377.

Chen, Zhongmin. (1990). Yinxian fangyan tongyinzihui [The homophonous syllabary of Yinxian dialect]. *Fangyan* 1:32-41.

Cheung, Yuk Man. (2004). A perceptual analysis of the apical vowels in Beijing Mandarin. MPhil. Thesis, City University of Hong Kong.

Cho, Taehong; Lee, Yoonjeong; Kim, Sahyang. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics* 39:344–361.

Clements, G. N. (1999). Affricates as noncontoured stops. In Fujimura, Osamu; Joseph, Brian D.; Palek, Bohumil (eds.) *Proceedings of LP '98: Item Order in Language and Speech*, 271–299. Prague: The Karolinum Press.

Croft, William. (2003). *Typology and Universals*, 2nd edition. Cambridge University Press.

Cui, Rongchang. (1988). Sichuan Lezhixian "Jingzhou qiang" yinxi [Phonology of Jingzhou accent in Lezhi County, Sichuan Province]. *Fangyan* 1:42-51.

Delattre, P. C.; Liberman, A. M.; Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27, 769-773.

Deng, Yürong. (1996). Guangxi Hexian (Liantang) Kejiahua yinxi [Phonology of the Hakka dialect in Hexian (Liantan), Guangxi Autonomous Region]. *Fangyan* 4:276-289.

Deng, Yürong. (2000). Guangxi Zhongshan fangyan yinxi [The phonology of the Zhongshan dialect in Guangxi Autonomous Region]. *Fangyan* 4:326-341.

DiCanio, Christian. (2013). Spectral moments of fricative spectra script in Praat. Retrieved from http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives.praat. June

2016.
Ding, Shengshu. (1989). He'nansheng Suiping fangyan jilüe [The homophonous syllabary of Suiping dialect in He’nan Province]. *Fangyan* 2:81-90.
Dong, Zhengyi. (1990). Hu’nansheng Youxian fangyan jilüe [The homophonous syllabary of Youxian dialect in Hu’nan Province]. *Fangyan* 3:169-179.
Du, Yiqing. (2007). Haikou fangyan (laopai) tongyinzihui [The The homophonous syllabary of the old-styled Haikou dialect, Hainan Province]. *Fangyan* 2:147-162.
Duan, Yaguang. (2013). He’nan Ningling fangyan yinxi [Phonology of Ningling dialect in He’na Province]. *Fangyan* 2:168-178.
Duanmu, San. (2007). *The Phonology of Standard Chinese*, 2nd Edition. Oxford University Press.
Dupoux, E.; Kakehi, K.; Hirose, Y.; Pallier, C.; Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568-1578.
Fan, Junjun. (2000). Hu'nan Guiyangxian Aoquan Tuhua tongyinzihui [A The homophonous syllabary of homophones of Aoquan Tuhua in Guiyang County, Hu’nan Province]. *Fangyan* 1:80-88.
Feng, Aizhen. (1988). Fujiansheng Fuqing fangyan de yüyin xitong [The homophonous syllabary of Fuqing dialect in Fujian Province] *Fangyan* 4:287-300.
Feng, Long. (1985). Beijinghua yuliu zhong shengyundiao de shichang [Duration of consonants, vowels and tones in colloquial Beijing Mandrain]. In Tao Lin & Lijia Wang (eds.), *Beijing Yuyin Shiyan Lu* [*Experimental Studies in Beijing Mandarin*]. Beijing: Peking University Press.
Flanagan, J. L. (1955). A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America* 27: 613-617.
Flege, J. E.; Takagi, N.; Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults’ perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99. 1161-1173.
Flemming, Edward. (2002). *Auditory Representation in Phonology*. London & New York: Routledge.
Flemming, Edward. (2004). Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically Based Phonology*, 232–276. Cambridge: Cambridge University Press.
Flemming, Edward. (2005). Speech perception and phonological contrast. In David B. Pisoni & Robert E. Remez (eds.), *The Handbook of Speech Perceptio*n, 156–181. Oxford: Blackwell.
Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics* 11.

149-176.

Gao, Baotai. (1980). Lanzhou yinxi shuolüe [A brief phonology of Lanzhou dialect]. *Fangyan* 3:224-231.

Greenberg, Joseph H.; Jenkins, James J. (1964). Studies in the psychological correlates of the sound system of American English. *Word* 20. 157-177.

Gu, Qian. (1990). Taixing fangyan tongyinzihui [The homophonous syllabary of Taixing dialect] *Fangyan* 4:284-292.

Guan, Yingwei. (2005). Guangxi Gongcheng Zhihua yinxi [Phonology of Zhihua dialect in Gongcheng County of Northern Guangxi Autonomous Region]. *Fangyan* 3:232-241.

Guion, Susan Guignard. (1998). The role of perception in the sound change of velar palatalization. *Phonetica* 55:18–52.

Guo, Xiliang. (1986). *Hanzi Guyin Shouce* [*Handbook of the Ancient Pronunciation of Chinese Characters*]. Beijing: Peking University Press.

Gussmann, Edmund. (2007). *The Phonology of Polish*. Oxford: Oxford University Press.

Hamann, Silke. (2003). *The Phonetics and Phonology of Retroflexes*. Utrecht: LOT Press.

Hansson, Gunnar Ólafur. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2/5: 859–893,

He, Kailin. (2003). Hu'nan Daoxian Shouyan Pinghua yinxi [Phonology of the Daoxian (Shouyan) Pinghua in Hu'nan Province]. *Fangyan* 1:78-87.

He, Wei. (1984). Luoyang fangyan jilüe [A brief description of Luoyan dialect]. *Fangyan* 4:278-299.

He, Weitang. (1990). Guangdong Zengcheng fangyan tongyinzihui [The homophonous syllabary of Zengchen dialect in Guangdong Province]. *Fangyan* 4:270-283.

Heinz, J. M.; Stevens, K. N. (1961). On the properties of the voiceless fricative consonants. *Journal of the Acoustical Society of America* 33: 589–96.

Hirata, Shoji. (1982). Xiuning yinxi jianjie [A brief phonology of Xiuning dialect]. *Fangyan* 4:276-284.

Hiroyuki, Akitani. (2000). Jiangxi Guangfeng fangyan yinxi [Phonology of the Guangfeng dialect, Jiangxi Pronince]. *Fangyan* 3:260-275.

Hiroyuki, Akitani. (2004). Fujian Shibei fangyan yinxi [Phonology of the Shibei dialect, Fujian Province]. *Fangyan* 1:76-91.

Hiroyuki, Akitani; Cao, Zhiyun. (1998). Zhejiang Qingyuan fangyan yinxi [Phonology of Qingyuan dialect, Zhejiang Province]. *Fangyan* 1:23-36.

Hockett, Charles F. (1966). The problem of universals in language. In Greenberg, Joseph H. (ed.) *Universals of Language* (Second Edition), 1-29. The MIT Press.

Hou, Xiaoying. (2008). Guangdong Longchuanxian Tuocheng Kejia fangyan yinxi [Phonology of Hakka language in Tuocheng, Longchuan County, Guangdong Province]. *Fangyan* 2:147-160.

Hu, Shiyun. (1989). Lianshui fangyan tongyizihui [The homophonous syllabary of Lianshui dialect]. *Fangyan* 2:131-143.

Hu, Sike. (2009). Hu'nan Yongxing Ganfangyan tongyizihui [The homophonous syllabary of Gan dialect in Yongxing, Hu'nan Province]. *Fangyan 3*: 271-282.

Huang, Daxiang. (2009). Gansu Zhangye fangyan tongyizihui [The homophonous syllabary of Zhangye dialect, Gansu Province]. *Fangyan 4*: 342-352.

Huang, Qiliang. (2008). Guangxi Guanyang Guanyinge tuhua yinxi [Phonology of Guangyinge dialect, Guanyang, Guangxi Automomous Region]. *Fangyan 3*:272-280.

Huang, Qun. (2006). Guangxi Zhaoping fangyan yinxi [Phonology of Zhaoping dialect in Guangxi Autonomous Region]. *Fangyan* 2:149-167.

Huang, Shiquan. (2011). Anhui Taihu fangyan yinxi [Phonology of Taihu dialect, Anhui Province]. *Fangyan* 1: 80-90.

Huang, Xiaodong. (2007). Zhejiang Linhai Fangyan yinxi [Phonology of Lianhai dialect in Zhejiang Province]. *Fangyan* 1:35-51.

Huang, Xiaodong. (2013). Wannan jiuxing yumin fangyang yinxi [Phonology of nine-surname fishermen in South Anhui]. *Fangyan* 2:179-191.

Huang, Xuezhen. (1988). Hu'nan Jiangyong fangyan yinxi [Phonology of Jiangyong dialect in Hu'nan Province]. *Fangyan* 3:161-176.

Huang, Xuezhen. (1992). Meixian Kejiahua de yüyin tedian [Phonetic properties of Hakka dialect in Meixian County]. *Fangyan* 4:275-289.

Hume, Elizabeth; Johnson, Keith. (2001). A model of the interplay of speech perception and phonology. In Hume, Elizabeth; Johnson, Keith (eds.), *The Role of Speech Perception in Phonology*, 3–26. New York: Academic Press.

Hume, Elizabeth; Mailhot, Frederic. (2013). The role of entropy and surprisal in phonologization and language change. In Yu, Alan (ed.) *Origins of Sound Patterns: Approaches to Phonologization*, 29-47, Oxford University Press.

Iskarous, Khalil; Shadle, Christine H.; Proctor, Michael I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *Journal of the Acoustical Society of America* 129(2): 944-954.

Janson, Tore. (1986). Cross-linguistic trends in the frequency of CV sequences. *Phonology Yearbook* 3. 179-195.

Jiang, Di; Ouyang, Jueya; Zou, Jiayan. (2007). Hainansheng Sanyashi Maihua yinxi [Phonology of Maihua in Sanya City, Hainan Province]. *Fangyan* 1:23-34.

Johnson, Keith; Babel, Molly. (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, Keith; Flemming, Edward; Wright, Richard. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69(3): 505-528.

Jongman, Allard, Wayland, Ratree; Wong, Serena. (2000). Acoustic characteristics of English fricatives, *Journal of the Acoustical Society of America* 106: 1252-1263.

Jun, Jongho. (1995). Perceptual and articulatory factors in place assimilation: an optimality theoretic approach. PhD Dissertation, UCLA.

Kakusho, O.; Kato, K. (1968). Just discriminable change and matching range of acoustic parameters of vowels. *Acustica* 20: 46-54.

Karlgren, Bernhard. (1917). *Etudes sur la Phonologie Chinoise* [*Studies on Chinese Phonology*]. Stockholm: Boktryckeri.

Keating, Patricia; Lahiri, Aditi. (1993). Fronted velars, palatalized velars, and palatals. *Phonetica* 50:73–101.

Kensinger, Kenneth M. (1963). The phonological hierarchy of Cashinahua. In ed. Benjamin F. Elson (ed.) *Studies in Peruvian Indian Languages*, 207–217. Norman: Summer institute of Linguistics.

Kewley-Port, D.; Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America* 95. 485-496.

Keyser, Samuel Jay; Stevens, Kenneth. (2006). Enhancement and overlap in the speech chain. *Language* 82. 33-63.

Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K. N.; Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science* 255. 606-608.

Kuznetsova, Alexandra; Brockhoff, Per Bruun; Christensen, Rune Haubo Bojesen. (2015). lmerTest: tests in linear mixed effects models. R package version 2.0-20. https://cran.r-project.org/web/packages/lmerTest/index.html.

Ladefoged, Peter; Maddieson, Ian. (1996). *The Sounds of the World's Languages*. Blackwell Publishers Ltd.

Ladefoged, Peter; Wu, Zongji. (1984). Place of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12. 267-278.

Lee, Chao-Yang; Li, Zhiqiang. (2003). Enhancement of phonological contrast: Acoustics of apical and retroflex vowels in Mandarin Chinese. Poster Presentation on *34th Annual Meeting of the North Eastern Linguistic Society*. Stony Brook, NY.

Lee, Sang-Im. (2011). Spectral analysis of Mandarin Chinese sibilant fricatives. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, Aug. 17-21, 2011.

Lee, Wai-Sum; Zee, Eric. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33.1, 109–122.

Lee, Wai-Sum; Zee, Eric. (2015). Chinese phonetics. In Huang, C.-T. James; Li, Y.-H. Audrey; Simpson， Andrew (eds.), *The Handbook of Chinese Linguistics*. Malden, MA, USA, Wiley-Blackwell.

Lee-Kim, Sang-Im. (2014a). Contrast neutralization and enhancement in phoneme inventories: Evidence from sibilant place contrast and typology. Ph.D. dissertation, New York University.

Lee-Kim, Sang-Im. (2014b). Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study. *Journal of the International Phonetic Association* 44. 261-282.

Lehiste, Ilse. (1970). *Suprasegmentals.* Cambridge, Mass.: MIT Press.

Lennes, Mietta. (2002). Praat script: Save intervals to wav sound files. Retrieved Jan 2016 from http://www.helsinki.fi/~lennes/praat-scripts/.

Li, Daye. (1986). Xinhe yinxi shuolüe [A brief phonology of Xinhe dialect]. *Fangyan* 4:277-282.

Li, Dongxiang. (2013). Yuebei Lishi tuhua yinxi [The phonology of Lishi dialect in Northern Guangdong]. *Fangyan* 4:369-384.

Li, Fangfang. (2008). The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese. Ph.D. dissertation, The Ohio State University.

Li, Jiyuan; Liu, Lihua; Yan, Qinghui. (1987). Hu’nan Loudi fangyan tongyinzihui [The homophonous syllabary of Loudi dialect, Hu’nan Province]. *Fangyan* 4:294-305.

Li, Lan. (1994). Guizhou Danzhai fangyan yinxi [Phonology of Danzhai dialect, Guizhou Province]. *Fangyan* 1:20-30.

Li, Lan. (2015). Guangxi Yangshuo chengguan tuhua yinyun yanjiu [Phonology of the dialect outside Yangshuo city in Guangxi Zhuang Autonomous Region]. *Fangyan* 3:193-210.

Li, Mingxing. (2011). A perceptual study of Mandarin apical vowels and sibilants. Master's thesis, the Chinese University of Hong Kong.

Li, Mingxing; Zhang, Jie. (2017). Perceptual distinctiveness between dental and palatal sibilants in different vowel contexts and its implications for phonological contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1), 18. DOI: https://doi.org/10.5334/labphon.27

Li, Qiqun. (2011). Hu’nan Fenghuang fangyan tongyinzihui [The homophonous syllabary of Fenghuang dialect, Hu’nan Province]. *Fangyan* 4: 341-353.

Li, Rong. (1956). *Qieyun Yinxi* [*The Phonological System of Qieyun*]. Peking: Kexue Chubanshe.

Li, Shanping; Gu, Wentao. (2015). Acoustic analysis of Mandarin affricates. *Interspeech* 1680-1684

Li, Shuyan; Zhang, Ansheng. (1995). *Yinchuan Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Yinchuan Dialect*]. *Fangyan* 2:94-106.

Li, Xinghui. (2003). Hu'nan Yongzhou Lanjiaoshan Tuhua yinxi [Phonology of Lanjiaoshan dialect in Yongzhou County, Hu'nan Pronince]. *Fangyan* 1:67-77.

Li, Xinghui. (2015). Hu'nan Jianghua Pingdiyao Qiduhua yinxi [Phonology of Pingdiyao Qidu dialect in Jianghua, Hu'nan Pronince]. *Fangyan* 4:356-366.

Li, Xinkui. (1983). *Zhongyuan Yinyun Yinxi Yanjiu* [*The Sound System of Zhongyuan Yinyun*]. Zhongzhou Shuhua She.

Li, Xuejun. (2012). He'nan Neihuang fangyan yinxi [Phonology of Neihuang dialect, He'nan Province]. *Fangyan* 1: 67-78.

Li, Zhenhua; Zhou, Changji. (1999). *Hanzi Gujin Yinbiao* [*A List of the Historical Pronunciation of Chinese Characters*] (Revised Edition). Beijing: Zhonghua Shuju Press.

Lian, Chunzhao. (2002). Yüexi Lianjiang Shijia Kejia fangyan yinxi [Phonology of the Hakka dialect in Shijiao, Lianjiang of Western Guangdong]. *Fangyan* 3:272-288.

Liang, Deman. (1993). *Chengdu Fangyan Cidian* yinlu [The Introductory Chapter in *A Dictionary of Chengdu Dialect*]. *Fangyan* 1:2-13.

Liang, Jinrong. (1996). Lingui Liangjian Pinghua tongyinzihui [The homophonous syllabary of Liangjiang Pinghua in Lingui] *Fangyan* 3:180-189.

Liang, Yougang. (1986). Hainandao Wenchang fangyan yinxi [Phonology of Wenchang dialect in Hainandao Island]. *Fangyan* 2:123-132.

Liang, Yüzhang. (1990). Wupingxian Zhongshanzhen Junjiahua [The "Junjiahua" dialect in Zhongshanzhen, Wuping County]. *Fangyan* 3:192-203.

Liberman, A. M.; Cooper, F. S.; Shankweiler, D. P.; Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74. 431-461.

Liljencrants, Johan; Lindblom, Björn. (1972). Numerical simulation of vowel quality contrasts: the role of perceptual contrast. *Language* 48: 839-62.

Lin, Lunlun. (1994). Guangdongsheng Chenghai fangyan tongyinzihui [The homophonous syllabary of Chenhai dialect in Guangdong Province]. *Fangyan* 2:128-142.

Lin, Xiaoxiao. (2012). Zhejiang Taizhou Luqiao fangyan tongyizihui [The homophonous syllabary of Luqiao dialect, Taizhou, Zhejiang Province]. *Fangyan 2:* 142-154.

Lin, Yen-Hwei. (1989). Autosegmental treatment of segmental processes in Chinese phonology. Ph.D. dissertation. The University of Texas at Austin.

Lin, Yi; Yu, Jin. (2009) Guangxi Yongning Sitang Pinghua tongyinzihui [The homophonous syllabary of Sitang Pinghua, Yongning, Guangxi Zhuang Autonomous Region]. *Fangyan* 3: 257-270.

Lindblom, Björn. (1986). Phonetic universals in vowel systems. In John Ohala & J. J. Jaeger (eds.), *Experimental Phonology*, 13–44. Orlando, Fl: Academic Press.

Lindblom, Björn. (1990a). Phonetic content in phonology. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* (*PERILUS*) 11. 101-118.

Lindblom, Björn. (1990b). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, William J.; Warchal, A. (ed.) *Speech Production and Speech Modeling*, 403-439. Dordrecht: Kluwer.

Lisker, Leigh. (1963). *Introduction to Spoken Telugu*. New York: American Council of Learned Societies.

Liu, Bin. (1990). Zhangyü (Liugou) fangyan tongyinzihui [Syllabic inventoy of Zhangyü (Liugou) dialect]. *Fangyan* 1:11-20.

Liu, Boxiang; Chen, Li. (2015). Anhui Jingxian Zhaji fangyan tongyinzihui [The homophonous syllabary of Zhaji dialect in Jing County, Anhui Province]. *Fangyan* 3:211-220.

Liu, Chao. (2010). Hanyu fangyan seyin shengxue shiyan yanjiu [An acoustic study of fricatives in Chinese dialects]. Master thesis, Nanjing Normal University, Nanjing, China.

Liu, Danqing. (1994). *Nanjing Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Nanjing Dialect*]. *Fangyan* 2:81-102.

Liu, Dongbing. (1997). Kaifeng fangyan jilüe [A brief description of Kaifeng dialect]. *Fangyan* 4:271-285.

Liu, Lili. (1988). Yanqi yinxi jilüe [A brief phonology of Yanqi dialect]. *Fangyan* 1:31-41.

Liu, Lunxin. (1995). Jiangxisheng Dayü (Nan'an) fangyan yinxi [Phonology of Dayü (Nan'an) dialect in Jiangxi Province]. *Fangyan* 1:63-69.

Liu, Songhao. (1993). Xiangcheng fangyan yinxi [Phonology of Xiangcheng dialect]. *Fangyan* 4:284-294.

Liu, Xinzhong; Ou, Jing. (2010). Hainan Dongfangshi Sugengzhen Fumahua tongyinzihui [The homophonous syllabary of Fuma dialect, Sigeng, Dongfang City, Hainan Province]. *Fangyan* 1:30-44.

Liu, Zemin. (2007). Zhejiang Taishun Siqian Shehua yinxi [Phonology of Siqian Shehua, Taishun, Zhejiang Province]. *Fangyan 4*: 327-338.

Lu, Jinyuan. (1986). Lüsi fangyan jilüe [A brief description of Lüsi dialect]. *Fangyan* 1:52-70.

Lu, Lanhua. (2011). Qinghai Huangyuan fangyan yinxi [Phonology of Huangyuan dialect in Qinghai Province]. *Fangyan 1*:68-79.

Lu, Shaohao. (1995). Jinggangshan Kejiahua yinxi [Phonology of Hakka dialect in Jinggangshan]. *Fangyan* 2:121-127.

Lu, Yu-An. (2014). Mandarin fricatives redux: the psychological reality of phonological representations. *Journal of East Asian Linguistics* 23: 43–69.

Lu, Zhiwei. (1947). *Guyin Shuolüe* [*A Brief Treatise on Ancient Sounds*]. Harvard-Yenching

Institute.
Luo, Futeng. (1995). *Mouping Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Mouping Dialect*]. *Fangyan* 1:1-16.
Ma, Zhongqi. (1993). Zhangzhou fangyan tongyinzihui [The homophonous syllabary of Zhangzhou dialect]. *Fangyan* 3:199-217.
Maddieson, Ian. (1984). *Patterns of Sound*. Cambridge University Press.
Maniwa, Kazumi; Jongman, Allard; Wade, Travis. (2008). Perception of clear English fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America* 123, 1114–1125.
Maniwa, Kazumi; Jongman, Allard; Wade, Travis. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* 125(6), 3962–3973.
Martinet, Andre. (1952). Function, structure, and sound change. *Word* 8. 1-32.
Mao, Bingsheng. (1995). Hu'nan Hengshan fangyan yinxi [Phonology of Henshan dialect in Hu'nan Province]. *Fangyan* 2:112-120.
McGuire, Grant. (2010). A brief primer on experimental designs for speech perception research. Manuscript. UC Santa Cruz.
McMurray, Bob; Jongman, Allard. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118. 219-246.
Mermelstein, Paul. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America* 63: 572-580.
Miller, Wick R. (1965). *Acoma Grammar and Texts*. Berkeley and Los Angeles, CA: University of California Press.
Mo, Haiwen. (2014). Guangxi Bingyan Wanglingzhen Pinghua tongyinzihui [The homophonous syllabary of the Wanglingzhen dialect in Binyang, Guangxi Zhuang Autonomous Region]. *Fangyan* 1:73-89.
Mohr, Burckhard; Wang, William S-Y. (1968). Perceptual distance and the specification of phonological features. *Phonetica* 18. 31-45.
Moulines, Eric; Charpentier, Francis. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9.5. 453 – 467.
Ning, Jifu. (1985). *Zhongyuan Yinyun Biaogao* [*Draft Tables of Phonology of the Central Plains*]. Jilin: Jilin Wenshi chubanshe.
Nord, L.; Sventelius, E. (1979). Analysis and prediction of difference limen data for formant frequencies. *Quarterly Progress Status Report* 3-4. 60-72. Stockholm, Sweden: Speech Transmission Laboratory, Royal Institute of Technology.

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

Ohala, John J. (1981). The listener as a source of sound change. In Masek, Carrie S.; Hendrick, Roberta A.; Miller, Mary Frances (eds.) *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society*, 178 – 203. Chicago Linguistic Society.

Ota, Itsuku; Cao, Zhiyun. (1998). Zhejiang Yunhe fangyan yinxi [Phonology of Yunhe dialect in Zhejiang Province]. *Fangyan* 4:290-303.

Packard, Jerome L. (2015). Morphology: Morphemes in Chinese. In Wang, William S-Y.; Sun, Chaofen (ed.) *The Oxford Handbook of Chinese Linguistics*, 263 - 273. Oxford University Press.

Padgett, Jaye. (2001). Contrast dispersion and Russian palatalization. In Hume, Elizabeth; Johnson, Keith (eds.), *The Role of Speech Perception in Phonology*. Academic Press.

Padgett, Jaye. (2003). Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21: 39–87.

Padgett, Jaye; Źygis, Marzena. (2007). The evolution of sibilants in Polish and Russian. *Journal of Slavic linguistics* 15.2, 291-324.

Pan, Wuyun. (2000). *Hanyu Lishi Yinyunxue* [*Chinese Historical Phonology*]. Shanghai: Shanghai Jiaoyu Chubanse.

Pan, Ying. (2010). Hanyu fangyan secayin shengxue shiyan yanjiu [An acoustic study of affricates in Chinese dialects]. Master thesis, Nanjing Normal University, Nanjing, China.

Peng, Xiaochuan. (1990). Guangdong Nanhai (Shatou) fangyan yinxi [Phonology of Nanhai (Shatou) dialect, Guangdong Province]. *Fangyan* 1:21-31.

Penny, Ralph. (2002). *A History of the Spanish Language*. Cambridge: Cambridge University Press.

Perception Research Systems. (2007). Paradigm Stimulus Presentation, Retrieved from http://www.paradigmexperiments.com.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Gussenhoven, Carlos; Warner, Natasha (eds.) *Papers in laboratory phonology VII*, 101–39. Berlin, Germany: Mouton de Gruyter.

Pisoni, David. (1973). Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 13. 253-260.

Pisoni, David; Tash, Jeffrey. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15. 285-290.

Prost, Gilbert. (1967). Phonems of the Chacobo language. *Linguistics* 35: 61–65.

Pulleyblank, Edwin G. (1984). *Middle Chinese: A Study in Historical Phonology*. The University of British Columbia Press.

Pycha, Anne. (2015). Co-articulatory cues for communication: An investigation of five environments. *Language and Speech* 59(3): 1–23.

Qian, Zengyi (1995). *Ji'nan Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Ji'nan Dialect*]. *Fangyan* 4:242-256.

Qian, Zengyi (2003). Changlehua yinxi [Phonology of Changle dialect in Zhejiang Province]. *Fangyan* 4:299-313.

Qiao, Guanming; Chao, Baotong. (2002). Binxian fangyan tongyinzihui [The homophonous syllabary of the Binxian dialect in Shaanxi Province]. *Fangyan* 3:260-271.

Qiu, Shangren. (1991). Nancheng fangyan de yüyin xitong [Phonology of Nancheng dialect]. *Fangyan* 1:30-39.

Qu, Jianhui. (2005). Hu'nansheng Luxi (Pushi) fangyan yinxi [Phonology of the Luxi(Pushi) dialect in Hu'nan Province]. *Fangyan* 1:50-63.

Qu, Jianhui. (2008). Hu'nansheng Luxi (Baisha) xianghua yinxi [Phonology of the Luxi (Baisha) dialect in Hu'nan Province]. *Fangyan* 2: 161-172.

Ran, Qibin. (2005). Jiyu Putonghua de Hanyu zuse fuyin shiyan yanjiu [An experimental study of Chinese sibilants with an emphasis on Mandarin]. Ph.D. dissertation, Nankai University, Tianjin, China.

Raphael, Lawrence. (2005). Acoustic cues to the perception of segmental phonemes. In Pisoni, David B.; Remez, Robert E. (eds.), *The Handbook of Speech Perception*, pp.182-206.Blackwell Publishing Ltd.

Recasens, Daniel; Espinosa, Aina. (2009). Acoustics and perception of velar softening for unaspirated stops. *Journal of Phonetics* 37:189–211.

Scarborough, Rebecca. (2004). Coarticulation and the structure of the lexicon. PhD. dissertation. UCLA.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3):249–263.

Sereno, Joan. A.; Baum, Shari R.; Marean, G. Cameron; Lieberman, Philip. (1987). Acoustic analyses and perceptual data on anticipatory coarticulation in adults and children. *Journal of the Acoustical Society of America* 81. 512-519.

Seyfarth, Scott; Buz, Esteban; Jaeger, T. Florian. (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, 139(2):EL31–EL37.

Sha, Ping (1999). Fujiansheng Ningde fangyan tongyinzihui [The homophonous syllabary of Ningde dialect in Fujian Province]. *Fangyan* 4:282-295.

Shen, Ming (1993). *Taiyuan Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Taiyuan Dialect*]. *Fangyan* 4:241-252.

Shen, Ming (2009). Shanxin Lanxian fangyan yinxi [Phonology of Lanxian dialect in Shanxin Province]. *Fangyan* 4: 326-341.

Shen, Ming; Huang, Jing'ai. (2015). Anhui Xuancheng Yanchi fangyan yinxi [Phonology of Yanchi dialect in Xuancheng, Anhui Province]. *Fangyan* 1: 58-69.

Shen, Tong. (1989). Qimen fangyan de yüyin tedian [Phonetic properties of Qimen dialect]. *Fangyan* 1:30-39.

Shi, Jianguo. (1992). Gansusheng Wudu fangyan tongyinzihui [The homophonous syllabary of Wudu dialect in Gansu Province]. *Fangyan* 1:45-53.

Shi, Mingyuan. (1987). Shandongsheng Jüxian fangyan yinxi [Phonology of Jüxian dialect in Shandong Province]. *Fangyan* 3:179-189.

Shi, Xiujü; Zhang, Li. (2014). Shanxi Xinxian fangyan tongyinzihui [The homophonous syllabary of Xingxian dialect in Shanxi Province]. *Fangyan* 4: 343-355.

Silverman, Daniel. (2006a). *A Critical Introduction to Phonology: Of sound, Mind, and Body*. London, UK: Continuum.

Silverman, Daniel. (2006b). The diachrony of labiality in Trique, and the functional relevance of gradience and variation. In Goldstein, Louis M.; Whalen, Douglas H.; Best, Catherine T. (eds.) *Papers in laboratory phonology VIII*, 135–54. Berlin, Germany: Mouton de Gruyter.

Silverman, Daniel. (2010). Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46: 453–482.

Stent, Amanda J.; Huffman, Marie K.; Brennan, Susan E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 50(3):163–178.

Steriade, Donca. (1997). Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Phonetics* 3. 25-146.

Steriade, Donca. (1999). Alternatives to syllable-based accounts of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.), *Proceedings of LP '98: Item Order in Language and Speech*. Prague: Charles University and Karolinum Press.

Steriade, Donca. (2001). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The Role of Speech Perception in Phonology*, 219–250. San Diego: Academic Press.

Steriade, Donca. (2008). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Hanson, Kristin; Inkelas, Sharon (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*, 151–180. Cambridge, Mass.: The MIT Press.

Stevens, Kenneth; Blumstein, Sheila. (1975). Quantal aspects of consonant production and perception: A study of retroflex consonants. *Journal of Phonetics* 3: 215-234.

Stevens, Kenneth; Li, Zhiqiang; Lee, Chao-Yang; Keyser, Samuel Jay. (2004). A note on Mandarin fricatives and enhancement. In Gunnar Fant, Hiroya Fujisaki, Jianfen Cao & Yi Xu (eds.), *From Traditional Phonology to Modern Speech Processing*, 393–404. Beijing: Foreign Language Teaching and Research Press.

Su, Hua. (1994). Fujian Nanping fangyan tongyinzihui [The homophonous syllabary of Nanping dialect in Fujian Province]. *Fangyan* 1:37-45.

Su, Xiaoqing. (1990). Haizhou fangyan tongyinzihui [The homophonous syllabary of Haizhou dialect]. *Fangyan* 2:87-99.

Su, Xiaoqing. (1993). Jiangsusheng Yancheng fangyan de yüyin [Phonetics of Yancheng dialect in Jiangsu Province]. *Fangyan* 2:121-128.

Su, Xiaoqing; Lü, Yongwei. (1994). *Xuzhou Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Xuzhou Dialect*]. *Fangyan* :184-197.

Sun, Yizhi. (2002). Anhui Susong fangyan tongyinzihui [The homophonous syllabary of Susong dialect in Anhui Province]. *Fangyan* 4:368-379.

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

Tan, Yuanxiong. (1996). *Nanning Pinghua Cidian* yinlun [The Introductory Chapter in *A Dictionary of Nanning Pinghua*]. *Fangyan* 3:161-179.

Tung, T'ung-Ho. (1954). *Zhongguo Yuyin Shi* [*History of Chinese Phonology*]. Taibei: Zhonghua Wenhua Chuban Shiye Weiyuanhui.

Wang, Futang. (2008). Shaoxing fangyan tongyizihui [The homophonous syllabary of Shaoxing dialect]. *Fangyan* 1: 1-17.

Wang, Jenny Zhijie. (1993). The geometry of segmental features in Beijing Mandarin. Ph.D. dissertation, University of Delaware.

Wang, Jenny Zhijie. (1999). Beijinghua de yinjie yu yinxi [Syllable and Phonology of Beijing Mandarin]. In Xu, Liejiong (ed.) *Gongxing yu Gexing: Hanyu Yuyanxue zhongde Zhengyi* [*Universals and Specificity: Controversial Issues in Chinese Linguistics*]. Beijing: Beijing Yuyan Wenhua Daxue Press.

Wang, Junhu. (1995). *Xi'an Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Xi'an Dialect*]. *Fangyan* 2:81-93.

Wang, Li. (1980). *Hanyu Shigao* [*A Draft of the History of Chinese*]. Zhongguo Shuju Press.

Wang, Li. (1985). *Hanyu Yuyin Shi* [*Historical Phonology of Chinese*]. Zhongguo Shehui Kexue Press.

Wang, Linhui. (2003). Shanxi Linyi fangyan tongyinzihui [The homophonous syllabary of Linyin dialect, Shanxi Province]. *Fangyan* 3:277-288.

Wang, Linhui. (2014). Shanxi Jiangxian fangyan tongyinzihui [The homophonous syllabary of Jiangxian dialect in Shanxi Province]. *Fangyan* 2: 170-179.

Wang, Qin. (2012). Wanbei Fuyang fangyan tongyinzihui [The homophonous syllabary of Fuyang dialect, Northen Anhui Province]. *Fangyan* 3: 242-255.

Wang, Shihua; Huang, Jilin. (1993). *Yangzhou Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Yangzhou Dialect*]. *Fangyan* 3:161-175.

Wang, Taiqing. (1983). Tongling fangyan jilüe [A brief description of Tongling dialect]. *Fangyan* 2:99-119.

Wang, William S.-Y. (1968). The basis of speech. *Project on Linguistic Analysis Reports*, University of California at Berkeley.

Wang, Xiaojun. (2010). Jiangxi Xinyü Gan fangyan yinxi [Phonology of Gan dialect in Xinyü, Jiangxi Province]. *Fangyan* 3: 259-272.

Wedel, Andrew; Jackson, Scott; Kaplan, Abby. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech* 56(3): 395– 417.

Wedel, Andrew; Nelson, Noah; Sharp, Rebecca. (2016). The phonetic specificity of contrastive hyperarticulation in natural speech. MS. University of Arizona, Tucson.

Wei, Gangqiang. (1995). *Pingxiang Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Pingxiang Dialect*]. *Fangyan* 3:161-174.

Wei, Yüwen. (1997). Wuhua fangyan tongyinzihui [The homophonous syllabary of Wuhua dialect]. *Fangyan* 3:206-219.

Weihs, Claus; Ligges, Uwe; Luebke, Karsten; Raabe, Nils. (2005). klaR analyzing German business cycles. In Baier, D., Decker, R. and Schmidt-Thieme, L. (eds.). *Data Analysis and Decision Support*, 335-343, Springer-Verlag, Berlin.

Wen, Duanzheng; Zhang, Guangming. (1994). *Xinzhou Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Xinzhou Dialect*]. *Fangyan* 1:1-12.

Werker, Janet F.; Logan, John S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37. 35-44.

Whalen, Douglas H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[ś] boundary. *Journal of the Acoustical Society of America* 69. 275–282.

Whalen, Douglas H. (1991). Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America* 90. 1776–1785.

Wilson, Colin. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30: 945–982.

Wiese, Richard. (1997). Underspecification and the description of Chinese vowels. In Jialing Wang & Norval Smith (eds.), *Studies in Chinese Phonology*, 219–249. Berlin/New York: Mouton de Gruyter.

Wu, Jiansheng. (1996). *Wanrong Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Wanrong Dialect*]. *Fangyan* 4:241-256.

Wu, Kaihua. (2009). Gansu Minqin fangyan yinxi [Phonlogy of Minqin dialect in Gansu Province]. *Fangyan* 1:40-52.

Wu, Xiaozhi. (1997). Shanxisheng Fufeng fangyan tonyinzihui [The homophonous syllabary of Fufeng dialect, Shaanxi Province]. *Fangyan* 3:192-205.

Wu, Yingjun. (1988) Hainansheng Danzhou fangyan danyinzibiao [The homophonous syllabary of Danzhou dialect, Hainan Province]. *Fangyan* 2:113-122.

Wu, Zongji; Lin, Maocan. (1989). *Shiyan Yuyinxue Gaiyao* [*An Outline of Experimental Phonetics*]. Beijing: Gaodeng Jiaoyu Chubanshe [Higher Education Press].

Xia, Jianqin. (1983). Liuyang Nanxiang fangyan jilüe [A brief description of Nanxiang dialect in Liuyang]. *Fangyan* 1:47-58.

Xie, Boduan. (2010). Hu'nan Chenxi fangyan yinxi [Phonology of Chenxi dialect in Hu'nan Province]. *Fangyan* 2: 179-190.

Xie, Jianyou. (1994). Nanning Baihua tongyinzihui [The homophonous syllabary of Nanning Baihua]. *Fangyan* 4:286-303.

Xie, Liuwen. (1997). *Yüdu Fangyan Cidia*n yinlun [The Introductory Chapter in *A Dictionary of Yüdu Dialect*]. *Fangyan* 1:14-26.

Xie, Liuwen. (2011). Jiangxi Fuliang (Jiuchengcun) fangyan tongyinzihui [The homophonous syllabary of Fuliang (Jiuchengcun) dialect, Jiangxi Province]. *Fangyan* 2: 117-131.

Xie, Qiyong. (2004). Hu'nan Xintian Nanxiang Tuhua tongyinzihui [The homophonous syllabary of the South Rural Tuhua in Xintian County, Southern Hu'nan Province]. *Fangyan* 2:158-167.

Xing, Xiangdong. (2011). Shanbei Wubaodong Wangjiashan fangyan tongyinzihui [The homophonous syllabary of Wangjiashan dialect, Wubaodong in northern Shanxi Province]. *Fangyan* 3:200-214.

Xiong, Zhenghui. (1994). *Nanchang Fanyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Nanchang Dialect*]. *Fangyan* 3:170-183.

Xu, Baohua; Tao, Huan. (1995). *Shanhai Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Shanghai Dialect*]. *Fangyan* 4:257-275.

Yang, Biwan. (2012). Guangdong Huaiji Xiafanghua tongyinzihui [The homophonous syllabary of Xiafang dialect, Huaiji, Guangdong Province]. *Fangyan* 4:326-343.

Yang, Shiwen. (2013). Guangxi Tengxian Mengjiang fangyan yinxi [Phonology of Mengjiang dialect, Tengxian, Guangxi Zhuang Autonomous Region]. *Fangyan* 1:71-85.

Yang, Yonglong. (2008). He'nan Shangcheng (Nansi) fangyan yinxi [Phonology of Shangcheng (Nansi) dialect in He'nan Province]. *Fangyan* 2: 133-146.

Yuan, Dan. (2010). Jiangsu Changshu Meili fangyan tongyinzihui [The homophonous syllabary of Meili dialect, Changshu, Jiangsu Province]. *Fangyan* 4: 325-337.

Ye, Baokui. (2001). Mingqing Guanhua Yinxi [Phonology of Ming-Qing Mandarin].

Ye, Xiangling. (1992). *Suzhou Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Suzhou Dialect*]. *Fangyan* 4:255-271.

Ye, Xiangling; Guo, Zongjun. (1991). Yixing fangyan tongyinzihui [The homophonous syllabary of Yinxing dialect]. *Fangyan* 2:88-98.

Yin, Kai. (2014). Hu'nan Linwu Chujiang tuhua tongyinzihui [The homophonous syllabary of Chujiang, Linwu, Hu'nan Province]. *Fangyan* 4: 277-289.

Yin, Shichao. (1995). *Ha'erbin Fanyang Cidian* yinlun [The Introductory Chapter in *A Dictionary of Harbin Dialect*]. *Fangyan*1:17-25.

Ying, Yutian. (1988). Hu'nan Anxiang fangyan jilue [A brief description of the Anxiang dialect in Hu'nan Province]. *Fangyan* 1: 52-67.

Yü, Guangzhong. (1988). Jiaxing fangyan tongyinzihui [The homophonous syllabary of Jiaxing dialect]. *Fangyan* 3:195-208.

Yü, Yang. (1991). Taizhou fangyan tongyinzihui [The homophonous syllabary of Taizhou dialect]. *Fangyan* 4:259-274.

Zampaulo, André. (2013). Sibilant dissimilation in the history of Spanish: An information-theoretic approach. In Howe, Chad; Blackwell, Sarah E.; Quesada, Margaret Lubbers (eds), *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, 172-178. Somerville, MA: Cascadilla Proceedings Project.

Zee, Eric; Lee, Wai-Sum. (2001). An acoustical analysis of the vowels in Beijing Mandarin. *Proceedings of Eurospeech* 1:643-646.

Zee, Eric; Lee, Wai-Sum. (2004). The apical sounds in Beijing Mandarin. In Lu, Jilun & Jialing Wang (eds.) *Xiandai Yuyinxue yu Yinxixue Yanjiu* [*Studies on Modern Phonetics and Phonology*]. Tianjin Shehui Kexue Yuan Press.

Zeng, Yümei. (1993). Xiangtan fangyan tongyinzihui [The homophonous syllabary of Xiangtan dialect]. *Fangyan* 4:295-305.

Zeng, Yümei. (1995). Hu'nan Yiyan fangyan tongyinzihui [The homophonous syllabary of Yiyan dialect, Hu'nan Pronince]. *Fangyan* 4:276-289.

Zhai, Yingyi. (1989). Shanxin Loufan fangyan tongyinzihui [The homophonous syllabary of Loufan dialect, Shanxin Province]. *Fangyan* 1:11-19.

Zhan, Bohui; Chen, Xiaojin. (1995). *Dongguan Fangyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Dongguan Dialect*]. *Fangyan* 3:175-187.

Zhang, Jie. (1997) Xiaoshang fangyan tongyinzihui [The homophonous syllabary of Xiaoshan dialect]. *Fangyan* 2:138-150.

Zhang, Shengyü. (1991). Taiping (Xianyuan) fangyan tongyinzihui [The homophonous syllabary of Taiping (Xianyuan) dialect]. *Fangyan* 3:188-199.

Zhang, Shengyü; Wang, Ping; Shen, Tong. (1988). Hu'nan Taojian (Gaoqiao) fangyan tongyinzihui [The homophonous syllabary of Taojiang (Gaoqiao) dialect in Hu'nan Province]. *Fangyan* 4:270-286.

Zhang, Zhenxing. (1987). Guangdong Haikang fangyan jilüe [A brief description of Haikang dialect in Guangdong Province]. *Fangyan* 4:264-282.

Zhang, Zhenxing. (1992). Guangdongsheng Wuchuan fangyan jilüe [A brief description of Wuchuan dialect in Guangdong Province]. *Fangyan* 3:195-213.

Zhao, Rixin. (1989). Anhui Jixi fangyan yüyin tedian [Phonetic properties of Jixi dialect, Anhui Province]. *Fangyan* 2:125-130.

Zhao, Rixin; Li, Jiaolei. (2014). Hu'nan Yuanling Qingshuiping Xianghua tongyinzihui [The homophonous syllabary of Qingshuiping dialect, Yuanling, Hu'nan Province]. *Fangyan* 2:155-169.

Zheng, Wei; Yuan, Dan; Shen, Ruiqing. (2012). Anhui Dangtu Huyang Wuyu tongyin zihui [The homophonous syllabary of Huyang dialect, Dangtu, Anhui Province]. *Fangyan* 4:314-325.

Zhong, Liming. (2004). Jiangxi Wuning Lixihua yinxi [Phonology of Lixi dialect in Wuning County, Jiangxi Province]. *Fangyan* 4:371-382.

Zhong, Longlin. (1987). Hu'nansheng Leiyang fangyan jilüe [A brief description of Leiyang dialect, Hu'nan Province]. *Fangyan* 3:215-231.

Zhong, Weike. (2002). Jiangjin fangyan tongyinzihui [The homophonous syllabary of Jiangjin dialect in Chongqing]. *Fangyan* 2:183-192.

Zhong, Ziqiang. (2015). Guangxi Canwu bendihua yinxi [Phonology of local dialect of Cangwu in Guangxi Zhuang Automamous Region]. *Fangyan* 2:177-192.

Zhou, Changji. (1991). Xiamen fangyan tongyinzihui [The homophonous syllabary of Xiamen dialect]. *Fangyan* 2:99-118.

Zhou, Lei. (1991). Jimusa'er fangyan tongyinzihui [The homophonous syllabary of Jimusa'er dialect]. *Fangyan* 1:40-49.

Zhou, Lei. (1994). *Wulumuqi Fanyan Cidian* yinlun [The Introductory Chapter in *A Dictionary of Urumqi Dialect*]. *Fangyan* 4:241-151.

Zhou, Lei. (1998). Tulufan hanyu fangyan yinxi [Phonology of Turpan Mandarin]. *Fangyan* 2:122-131.

Zhou, Rijian. (1987). Guangdongsheng Huiyang Kejiahua yinxi [Phonology of Hakka dialect in Huiyan, Guangdong Province]. *Fangyan* 3:232-237.

Zhou, Rijian. (1992). Guangdong Xinfeng Kejia fangyan jilüe [A brief description of Xinfeng Hakka in Guangdong Province]. *Fangyan* 1:31-44.

Zhou, Xianyi. (1994). Hunan Daoxian (Xiaojia) Tuhua tongyinzihui [The homophonous syllabary of Daoxian (Xiaojia) Tuhua in Hu'nan Province]. *Fangyan* 3:201-207.

Zhou, Yunalin. (2001). Anhui Lujiang fangyan yinxi [Phonology of Lujiang dialect in Anhui Province]. *Fangyan* 3:263-280.

Zhou, Zheng. (2005). Shanxi Pingli Luohe fangyan tongyinzihui [The homophonous syllabary of Luohe dialect in Pingli County, Shaanxi Province]. *Fangyan* 4:314-355.

Zhuang, Chusheng. (1996). Lechangshi Tatouba fangyan yinxi [Phonology of Tatouba dialect in Lechang city]. *Fangyan* 3:208-215.

Zuo, Fuguang. (1995). Sichuansheng Yibin Wangchang fangyan jilüe [A brief description of Wangchang dialect in Yibin, Sichuan Province]. *Fangyan* 1:51-62.

Źygis, Marzena; Padgett, Jaye. (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics* 38. 207–226.