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

Li, M and Zhang, J 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, pp. 1–27, DOI: https://doi.org/10.5334/labphon.27

JOURNAL ARTICLE

Perceptual distinctiveness between dental and palatal sibilants in different vowel contexts and its implications for phonological contrasts

Mingxing Li and Jie Zhang
Department of Linguistics, University of Kansas, 1541 Lilac Lane, Lawrence, US
Corresponding author: Mingxing Li (mxli@ku.edu)

Mandarin Chinese has dental, palatal, and retroflex sibilants, but their contrasts before [_i] are avoided: The palatals appear before [i] while the dentals and retroflexes appear before homorganic syllabic approximants (a.k.a. apical vowels). An enhancement view regards the apical vowels as a way to avoid the weak contrast /si-ɕi-ʂi/. We focus on the dental vs. palatal contrast in this study and test the enhancement-based hypothesis that the dental and palatal sibilants are perceptually less distinct in the [_i] context than in other vowel contexts. This hypothesis is supported by a typological survey of 155 Chinese dialects, which showed that contrastive [si, tsi, tsʰi] and [ɕi, tɕi, tɕʰi] tend to be avoided even when there are no retroflexes in the sound system. We also conducted a speeded-AX discrimination experiment with 20 English listeners and 10 Chinese listeners to examine the effect of vowels ([_i], [_a], [_ou]) on the perceived distinctiveness of sibilant contrasts ([s-ɕ], [ts-tɕ], [tsʰ-tɕʰ]). The results showed that the [_i] context introduced a longer response time, thus reduced distinctiveness, than other vowels, confirming our hypothesis. Moreover, the general lack of difference between the two groups of listeners indicates that the vowel effect is language-independent.

Keywords: perceptual distinctiveness; sibilant contrast; vowel context; Chinese; speeded-AX discrimination

1. Introduction

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 & Lindblom, 1972; Lindblom, 1986, 1990; Flemming, 2002, 2004, 2005; Boersma & 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 & Hamann, 2008, among others), and this is borne out in the cross-linguistic typology of vowel systems in that nearly all languages have these three vowels (Maddieson, 1984, p. 134). The preference for more distinct contrast is also observed in diachronic sound change. During the 16th century, for example, the less distinct contrast [ɕ-ʃʲ] in Polish developed into the more distinct contrast [ɕ-ʂ] (Żygis & Padgett, 2010).

In speech, the articulatory gestures for a particular segment usually overlap with those of neighboring segments; therefore, the acoustic properties and perceptual cues of a segment are influenced by its contexts (Liberman et al., 1967, among others). For instance, the perceptual information of a consonant often partially lies on its neighboring vowels. Furthermore, the quality of the perceptual cues for a segment may differ depending on context, and this is reflected in the general observation that a particular phonological contrast may be better licensed in some contexts than others (Steriade, 1997, 1999, 2001, 2008).

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

(1) 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 (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)

Lee-Kim (2014a) attributed the neutralization 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.

Mandarin Chinese has dental, palatal, and retroflex sibilants, and their surface contrasts in the [_i] context are also avoided by realizing the /i/ as homorganic syllabic approximants after the dental and retroflex sibilants (See 2.1 below for detail). Lee-Kim (2014a) referred to the avoidance of sibilant contrast before [_i] in Mandarin as *full contrast plus full enhancement*, i.e., the underlying sibilant contrasts are preserved with a complete change of the quality of /i/ after the dental and retroflex sibilants. In this study, we first examined the typology of sibilant contrasts across Chinese dialects, focusing on the place contrasts in the /_i/ context, and we observed that (i) a three-way sibilant contrast in the [_i] context is almost always avoided and (ii) a two-way contrast between dental and palatal sibilants in the [_i] context also tends to be avoided (See 2.2 for detail). We then specifically tested a hypothesis regarding the second observation—the dental and palatal sibilants are perceptually less distinct in the [_i] context than in other vowel contexts—using a crosslinguistic speeded AX discrimination experiment.

2. Sibilant contrasts across Chinese dialects

2.1. Mandarin sibilants and apical vowels

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 & Wu, 1984; Cao, 1990).[1] 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 & Lin, 1989, p. 133). The fricative components of the affricates are in general similar to the corresponding fricatives (Svantesson, 1986).

[1] 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 & Wu, 1984). In this paper, we follow the convention in the literature and use the IPA symbols [ʂ tʂ tʂʰ] for these sounds.

In Mandarin, the vowel [i] follows the palatal sibilants and most other consonants but not the dental and retroflex sibilants. As shown in (2), the dental and retroflex sibilants are followed by the homorganic syllabic approximants [ɹ̩] and [ɻ̩] (Lee & Zee, 2003, Lee-Kim, 2014b), which are often referred to as the 'apical vowels.'[2] The acoustic analysis 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. That is, the apical vowels are the 'vocalized prolongation' of their preceding consonants (Chao, 1934, p. 374). The two apical vowels [ɹ̩] [ɻ̩] and the vowel [i] differ in their first three formants, particularly F2 (Zee & Lee, 2001; Lee & Li, 2003; Cheung, 2004; Zee & Lee 2004), and there is no obvious consonant-vowel formant transition in the syllables like [sɹ̩], [ɕi], and [ʂɻ̩] due to the homorganic relation between the sibilants and the vocalic segments.

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

a. ɹ̩ after dentals		b. i after palatals (and other consonants)		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'

With regard to the formation of the apical vowel, there are two competing views: A place assimilation view and a contrast enhancement view. The place assimilation account treats the apical vowel as the result of the dental/retroflex sibilants spreading their place features to the following /i/. For example, Wang (1985) noted that it is a natural tendency for /i/ to change into [ɹ̩] or [ɻ̩] under the influence of their preceding dental sibilants [s ts tsʰ] or retroflex sibilants [ʂ tʂ tʂʰ]. A similar position was taken in phonological analyses that treat the post-dental or post-retroflex apical vowel as a vocalic slot unspecified for place, and the place value is filled in by feature spreading from the onset sibilants (Lin, 1989; Wang, 1993, 1999; Wiese, 1997).

The contrast enhancement account, on the other hand, regards the apical vowels as an enhancement of contrast, whereby the less distinct contrast /si, ɕi, ʂi/ turns into the more distinct [sɹ̩, ɕi, ʂɻ̩] (Stevens et al., 2004; Lee & Li, 2003, Lee-Kim, 2014a, among others). More specifically, the enhancement analysis assumes 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, 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 vowel [i] (Lee & Li, 2003). With the contrast at risk, an enhancement gesture is introduced to make /s/ and /ʂ/ more distinct from /ɕ/. That is, the apical vowels [ɹ̩] and [ɻ̩] are the continuation of the enhancing gesture on /s/ and /ʂ/ throughout the vowel /i/ (Stevens et al., 2004; Keyser & Stevens, 2006), whereby a shift of vowel helps preserve the contrast /si, ɕi, ʂi/ (Lee & Li, 2003, Lee-Kim, 2014a). As the fricative components of affricates are similar to their fricative counterparts (Svantesson, 1986), the enhancement analysis for Mandarin fricatives could be easily extended to the affricates.

Although the place assimilation account and the contrast enhancement account both work for Mandarin apical vowels, they make different predictions about language typology. The assimilation account assumes sibilant-to-/i/ spreading to be the cause of

[2] In 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, p. 34), fricative vowels (Ladefoged & Maddieson, 1996, p. 314), and syllabic approximants (Lee & Zee, 2003; Lee-Kim, 2014b).

apical vowel formation. Thus, it predicts that a language may have an apical vowel as long as it has the dental/retroflex sibilants followed by /i/. For example, when there are /si, tsi, ts^hi/ syllables, the place feature of the dental sibilants should be able to spread to the following /i/, presumably for the conservation of articulatory effort as assumed for assimilation in general (Abercrombie, 1967, p. 139; Janson, 1986, among others). In contrast, the enhancement account assumes reduced contrast distinctiveness, e.g., in /si ɕi ʂi/, to be the motivation of apical vowel formation. Thus, it predicts that apical vowels should only emerge in a language when it has phonological contrasts between 'sibilant + i' sequences, e.g., /si-ɕi/, /tsi-tɕi/, or /ts^hi-$tɕ^h$i/. That is, the apical [ɹ̩] and [ɻ̩] should not appear in a language if there are no such contrasts.

2.2. The typology of sibilant contrasts across Chinese dialects

To evaluate the two analyses above, we conducted a typological survey of apical vowels by collecting the syllabic inventories of Chinese dialects. We used *Fangyan* 'Dialects,' a Chinese journal specializing in the description of Chinese dialects, as the basis of our survey. We looked through all articles in *Fangyan* from 1979 to 2012 and identified 155 articles, each describing the syllabic inventory of a Chinese dialect. About two-thirds of the 155 dialects were explicitly described in the original article as belonging to a particular dialect group, i.e., Mandarin (or the northern dialects), Xiang, Yue, Min, Gan, Hakka, Jin, etc. There was no information of dialect grouping of the other dialects, though it is clear that they are distributed across regions of different dialect groups and geographically across different areas of mainland China.

Our focus is on the apical vowels [ɹ̩], [ɻ̩] and the vowel [i], and the relevant consonants are the dental [s ts ts^h], the retroflex [ʂ tʂ $tʂ^h$], and the palatal [ɕ tɕ $tɕ^h$]. We divided the dialects by the number of sibilant places and we observed that 31 of the 155 dialects have sibilants at one place, 78 dialects have sibilants at two places, and 46 dialects have sibilants at three places. The typology of sibilant inventory is listed in **Table 1**. **Table 1** showed that, for the 31 dialects with sibilants at one place, the majority (27/31) have dental sibilants, for the 78 dialects with sibilants at two places, the majority (76/78) have dental vs. palatal sibilants, and all of the 46 dialects with sibilants at three places have dental vs. palatal vs. retroflex sibilants. Within each group, the dialects generally belong to different dialect groups and are geographically distributed across different areas in mainland China.

We checked all of the CV syllables in the 155 dialects whose onsets are the dental, palatal, or retroflex sibilants *and* whose vowels are [i], [ɹ̩], or [ɻ̩]. That is, we checked

Table 1: Typology of sibilant inventory across 155 Chinese dialects.

No. of sibilant place	*No. of dialects*	*Example*
1 Place	31 total	
Dental	27	*Xiamen* (Zhou, 1991)
Palatal	4	*Lianzhou* (Cai, 1987)
2 Places	78 total	
Dental vs. Palatal	76	*Jiangyong* (Huang, 1988)
Dental vs. Retroflex	1	*Jinggangshan Hakka* (Lu, 1995)
Palatal vs. Retroflex	1	*Haizhou* (Su, 1990)
3 Places	46 total	
Dental vs. Palatal vs. Retroflex	46	*Harbin* (Yin, 1995)

five types of syllables: [sɹ̩ tsɹ̩ tsʰɹ̩], [si tsi tsʰi], [ɕi tɕi tɕʰi], [ʂɻ̩ tʂɻ̩ tʂʰɻ̩], and [ʂi tʂi tʂʰi].[3] Each of the 155 dialects has at least one of the five types of syllables. For example, *Xiamen Chinese* (Zhou, 1991) has [si tsi tsʰi] only and no other types, while Mandarin have three types of these syllables, as in (2). It is observed from the 155 dialects that the apical vowel appears in a dialect only when there are phonological contrasts between 'sibilant + i' sequences. That is, there is no dialect where apical vowels [ɹ̩] and [ɻ̩] appear after the sibilants in the absence of any contrast. In particular, this holds true in the 27 dialects that are reported to have the dental sibilants only like *Xiamen* (Zhou, 1991). Significantly, none has the place assimilation pattern as shown in (3).[4] The absence of cases like (3) is predicted by the contrast enhancement analysis, which takes the existence of potentially weak contrast (e.g., /si, ɕi, ʂi/) to be the cause of apical vowel formation. The place assimilation account, however, predicts the presence of (3), as it assumes that sibilant-to-[i] assimilation should apply without referring to any phonological contrast. The typological survey, therefore, supports the contrast enhancement analysis.

(3) Nonexistent dialect: Place assimilation in the absence of any contrast

/si/		[sɹ̩]
/tsi/	-- *'place assimilation'*→	[tsɹ̩]
/tsʰi/		[tsʰɹ̩]

The enhancement analysis of apical vowels (Stevens et al., 2004, Lee & Li, 2003, Lee-Kim, 2014a) was proposed with regard to the sibilant inventory of Mandarin. That is, apical vowels [ɹ̩] and [ɻ̩] are assumed to be formed to enhance the three-way fricative distinction in /si, ɕi, ʂi/. This raises the question of whether apical vowels are *only* formed when there are three sibilant places in the sound inventory, i.e., in a more crowded sibilant space.[5] To investigate this issue, we further checked separately the dialects that have three sibilant places and those that have only two sibilant places. Among the 155 dialects in the survey, there were 46 with sibilants at three places (dental vs. palatal vs. retroflex) and 45 of them avoided the three-way contrast in the [_i] context with the introduction of apical vowels, which is consistent with the enhancement analysis (Stevens et al., 2004, Lee & Li, 2003) and the cross-linguistic typology (Lee-Kim, 2014a). In the 76 dialects with the two-way dental vs. palatal sibilant contrast, contrastive dental-[i] vs. palatal-[i] is only allowed in 22 (≈29%), e.g., *Jiangyong* (4a), and avoided in 54 (≈71%), e.g., *Dayü* (4b) and *Shibei* (4c). Moreover, in the 54 dialects that avoid the dental-[i] vs. palatal-[i] contrasts, 52 introduced the apical vowel [ɹ̩] after the dental sibilants, as in *Dayü* (4b); the other two dialects avoided the contrasts with the non-combination of dental sibilants and the vowel [_i], as in *Shibei* (4c). Finally, in the two other dialects with two sibilant places—*Jinggangshan* (dentals vs. retroflexes; Lu, 1995) and *Haizhou* (palatals vs. retroflexes; Su, 1990)—the apical vowel [ɻ̩] has been developed after the retroflexes, and therefore, the two-way place distinction in the [_i] context is also avoided.

---

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

[4] Most of the 27 dialects were reported to belong to the Min, Yue, and Hakka dialect groups, and a few of them were not explicitly described as belonging to a group in the original sources.

[5] Thanks to an anonymous reviewer for raising this point.

(4) Examples of dialects with two place sibilants

a. *Jiangyong* (Huang, 1988)

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. *Dayü* (Liu, 1995)

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

c. *Shibei* (Hiroyuki, 2004)

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

Each legitimate syllable bears a high-falling tone.

In general, the typological survey shows that a three-way sibilant contrast in the [_i] context is virtually always avoided via the introduction of apical vowels, in support of the contrast enhancement view of apical vowels (Stevens et al., 2004, Lee & Li, 2003). In addition, the typology also shows that a two-way sibilant contrast between dentals and palatals in the [_i] context also tends to be avoided with the introduction of the dental apical vowel [ɹ̩].

2.3. Distinctiveness between dental vs. palatal sibilants

The contrast enhancement analysis (Stevens et al., 2004; Lee & Li, 2003) was proposed primarily on the basis of the acoustic properties of Mandarin fricatives and an essential component of this view is that sibilant place contrasts have reduced distinctiveness in the [_i] context. There have been few studies that directly tested the effect of vowel contexts, e.g., [_i] vs. other vowels, on the perceptual distinctiveness of Mandarin sibilants. In this study, we conducted a perceptual experiment on the distinctiveness of Mandarin sibilants. Given that the typological survey shows that in the [_i] context, the dental vs. palatal sibilants tend to be avoided even when there are no retroflex sibilants in the sound system, we focus on the perceptual distinctiveness of the dental and palatal sibilants in this study. We reserve the distinctiveness among three sibilant places for future research.

Following the enhancement view (Stevens et al., 2004; Lee & Li, 2003), we hypothesize that the dental and palatal sibilants form perceptually 'weak contrasts' in the [_i] context, e.g., [si tsi tsʰi] vs. [ɕi tɕi tɕʰi]. As a baseline for comparison, we hypothesize that dental vs. palatal sibilants form more distinct contrast in the other vowel contexts (e.g., [_a] and [_ou]) than in the [_i] context. More specifically, we predict that the sound pair [si-ɕi] is less distinct than (represented by '<') the pairs [sa-ɕa] and [sou-ɕou] and that the same holds for the affricate pairs:

[si-ɕi]	<	[sa-ɕa], [sou-ɕou];
[tsi-tɕi]	<	[tsa-tɕa], [tsou-tɕou];
[tsʰi-tɕʰi]	<	[tsʰa-tɕʰa], [tsʰou-tɕʰou].

We also hypothesize that the introduction of the apical vowel enhances the contrast. That is, we predict that pairs like [si-ɕi] are less distinct than pairs like [sɹ̩-ɕi]:

[si-ɕi] < [sɹ̩-ɕi];
[tsi-tɕi] < [tsɹ̩-tɕi];
[ts^{h}i-tɕ^{h}i] < [ts^{h}ɹ̩-tɕ^{h}i].

The hypotheses above refer to the sibilants in Chinese dialects, yet the vowel effect is by no means assumed to be specific to Chinese, as the avoidance of sibilant place contrast in the [_i] context has been observed crosslinguistically (Lee-Kim, 2014a).

3. Speeded AX discrimination: Vowel effect on sibilant distinctiveness

The evaluation of perceptual distinctiveness between sound pairs can be achieved with various experiments, e.g., similarity rating (Greenberg & Jenkins, 1964; Mohr & Wang, 1968) and AX discrimination (Pisoni, 1973; Johnson & Babel, 2010; Babel & 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 & Logan, 1985; Johnson & Babel, 2010) and the contrast and allophony in the listeners' native language (Gandour, 1983; Kuhl et al., 1992; Flege et al., 1996; Dupoux et al., 1999; Best et al., 2001, Hume & Johnson, 2001; Boomershine et al., 2008, etc.). In AX discrimination tasks, the listeners have been shown to access low-level acoustic information about a speech stimulus (Pisoni, 1973; Pisoni & Tash, 1974, Werker & 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 on AX discrimination. For example, Boomershine et al. (2008) tested the discrimination of [ð], [d], and [ɾ] by native listeners of English and Spanish, whereby [d-ɾ] are allophonic in English, phonemic in Spanish, and [ð-d] are allophonic in Spanish, 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 500ms 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. Although the phonemic systems of these voiceless fricatives for English and 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 & Babel, 2010), Johnson and Babel observed no effect of the listeners' native languages on the response time, which indicates that the speeded nature of the experiment has bypassed the influence from the listeners' L1 phonology.

The current paper hyothesizes that, for the dental vs. palatal sibilants, the [_i] context would induce less distinctiveness as compared to the other vowels. As shown in the typology of Chinese dialects, the avoidance of dental vs. palatal sibilants in the [_i] context is robust across different dialects. Thus, we assume that this generalization must be language-independent and the reduced distinctiveness in the [_i] context must be psychoacoustic in nature. Therefore, it is desirable to adopt a method of assessing perceptual distinctiveness that is minimally affected by the L1 background of the listeners. Following Johnson and Babel (2010) and Babel and Johnson (2010), we adopted the speeded AX discrimination method to investigate the effect of vowel contexts on the perceptual distinctiveness of dental and palatal sibilants. To check whether the results reflect psychoacoustic perception, listeners from two language backgrounds were recruited: Chinese listeners, whose L1 phonology has the dental and palatal sibilants, and English listeners, whose L1 does not.

3.1. Method

3.1.1. Participants

Two groups of subjects were recruited for this experiment: 20 native English listeners who have no previous exposure to Chinese (Mandarin or other dialects) and 10 native Mandarin listeners. The English listeners were undergraduates and graduates at the University of Kansas who received extra course credit for participation and the Chinese listeners were volunteer graduates and undergraduates. The participants completed a consent form (University of Kansas Human Subjects Committee Approval #20892) and none of the subjects reported hearing impairment.

3.1.2. Stimuli

The stimuli we used in the discrimination task were CV pairs whose onsets were dental vs. palatal sibilants and whose vowels were [_i] [_a] or [_ou], e.g., [si-ɕi], [sa-ɕa], and [sou-ɕou]. Mandarin has the dental and the palatal sibilants in the [_a] and [_ou] contexts, but in the [_i] context it has the palatals only (i.e., [ɕi tɕi tɕ^{h}i], but *[si tsi ts^{h}i]), as in (2). To obtain natural production of the contrastive CV pairs in the [_i] context, we need a language that keeps the contrasts between [si tsi ts^{h}i] and [ɕi tɕi tɕ^{h}i]. While such contrasts are absent in most of Chinese dialects, they are preserved in the speech and singing of Peking opera, a traditional Chinese vocal performance. We asked a trained male actor of Peking opera, who is also a native Mandarin speaker, to produce the syllables in (5).

(5) Stimulus syllables produced by the speaker

	A. sibilant-[i]	*B.* sibilant + **[ɹ̩]**	*C.* sibilant-[a]	*D.* sibilant-[ou]
Fricatives	**si** 西 'west' **ɕi** 兮 particle	**sɹ̩** 思 'to think'	sa 撒 'to release' ɕa 瞎 'blind'	sou 搜 'to search' ɕou 修 'to fix'
Unasp. Affricates	**tsi** 齑 'fragment' **tɕi** 鸡 'rooster'	**tsɹ̩** 资 'capital'	tsa 咂 'to smack lips' tɕa 佳 'good'	tsou 邹 surname tɕou 揪 'to clutch'
Asp. Affricates	**ts^{h}i** 七 'seven' **tɕ^{h}i** 欺 'to cheat'	**ts^{h}ɹ̩** 差 'uneven'	ts^{h}a 擦 'to wipe' tɕ^{h}a 掐 'to pinch'	ts^{h}ou 凑 'to assemble' tɕ^{h}ou 秋 'autumn'

All syllables bear a high-level tone except ts^{h}ou 凑'gather,' which has a falling tone. The speaker was asked to produce ts^{h}ou with a high-level tone. Boldface marks CV syllables produced as speech in Peking opera.

The speaker read the Chinese characters in columns C and D in (5) in normal speech in Mandarin and the characters in column A and B were read as they would be pronounced in the speech of Peking opera. 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 speaker produced six tokens of each syllable in (5) and, for each syllable, the token whose sibilant intensity was the closest to the mean of the six tokens was selected as the stimulus syllable (to be further manipulated). The acoustic information of the selected tokens (before manipulation) is given in **Table 2**.

Table 2: Acoustic measurements of the selected tokens (before manipulation).

	Onset		***Vowel***		
Syllable	***Duration (ms)***	***Intensity (dB)***	***Duration (ms)***	***Intensity (dB)***	***F0 onset-offset (Hz)***
si	171	57.45	345	70.63	296 – 271
sa	168	54.82	249	72.93	170 – 177
sou	164	57.14	246	69.15	181 – 177
sɹ̩	155	60.71	377	69.80	271 – 275
ɕi	168	60.07	315	73.06	278 – 272
ɕa	167	57.94	243	71.03	178 – 176
ɕou	203	60.44	230	67.94	191 – 186
tsi	98	58.28	383	68.83	274 – 258
tsa	58	59.24	285	72.76	173 – 176
tsou	60	59.15	226	69.15	190 – 187
tsɹ̩	97	54.69	409	71.36	272 – 278
tɕi	68	62.65	302	74.94	277 – 267
tɕa	56	57.14	254	70.53	174 – 175
tɕ^{j}ou	69	54.79	262	68.10	192 – 186
ts^{h}i	186	59.17	364	71.25	289 – 274
ts^{h}a	157	56.71	264	73.93	183 – 174
ts^{h}ou	168	55.98	235	68.23	178 – 170
ts^{h}ɹ̩	164	59.22	296	68.79	295 – 285
tɕ^{h}i	148	61.44	286	73.09	286 – 270
tɕ^{h}a	161	58.43	232	71.06	182 – 179
tɕ^{h}ou	158	60.50	223	67.22	189 – 187

It should be noted that, in the columns C and D of (5), the rime of the syllables with the palatal onsets is typically represented as ‘ia’ in the Chinese Pinyin orthography and the relevant syllables are sometimes transcribed as [ɕ^{j}a tɕ^{j}a tɕ^{hj}a] and [ɕ^{j}ou tɕ^{j}ou tɕ^{hj}ou] in the phonological literature. However, as noted by Ladefoged and Maddieson (1996, p. 150), the alleged onglide [j] involves “nothing other than a normal transition between the initial consonant and the following vowel in all these cases.” Therefore, we referred to such syllables as [ɕa tɕa tɕ^{h}a] and [ɕou tɕou tɕ^{h}ou].

In the speeded AX discrimination, the duration of the stimulus syllable must be controlled to facilitate the comparison of response time across vowel contexts, e.g., [si-ɕi] vs. [sa-ɕa]. Therefore, several steps of manipulation were performed on the naturally produced syllables in (5) before the formation of the CV pairs.

First, the durations of the onsets were normalized to lengths typical of the sibilants in normal speech. Based on Feng’s study of Mandarin consonants (1985), we used 125 ms as the target duration for the fricatives [s, ɕ], 50 ms for the unaspirated affricates [ts, tɕ], and 100 ms for the aspirated affricates [ts^h, tɕ^h]. As shown in **Table 2**, the durations of the naturally produced onset sibilants were generally longer than their target durations, presumably because these syllables were produced in the focus position. We shortened the sibilant onsets to the target durations using the Manipulation function in Praat (Boersma, 2001). It should be noted that in Feng (1985), the dental and palatal sibilants have slightly different durations—136 ms for [s] vs. 145 ms for [ɕ] word-initially; 110 ms for [s] vs. 122 ms for [ɕ] word-medially. However, a comparison of the dental and palatal sibilants across Chinese dialects shows no consistent pattern of duration difference: The dentals were reported to be longer than the palatals in some dialects but the reverse was reported in others (Ran, 2005; Liu, 2010; Pan, 2010). Therefore, in the current study, we used the same duration for the dental and palatal sibilants.

Second, the vocalic portion of each CV syllable was normalized to 120 ms. More specifically, we normalized the consonant-vowel transition portion to 50 ms as Delattre et al. (1955) showed that a formant transition of 50 ms is sufficient for the perception of consonant place. We normalized the steady vowel portion (i.e., the steady vowel formants) to 70 ms to get a total duration of 120 ms to match the duration of the vocalic portion in natural speech (Feng, 1985). As shown in **Table 2**, the vowel durations in the natural production were typically longer than 120 ms and their durations were shortened. In the manipulation of [ɕa], for example, we marked the interval between the start of the vocalic part and the start of the steady formants for [a], as in **Figure 1(a)**. Across all the stimulus syllables, the duration of this interval (i.e., the CV formant transition) was generally between 70 and 90 ms, and we shortened this interval to 50 ms following Delattre et al. (1955), as in **Figure 1(b)**. The shortening was performed in Praat, which adopts the PSOLA (Pitch-Synchronous Overlap-Add) technique (Moulines & Charpentier, 1990). That is, for

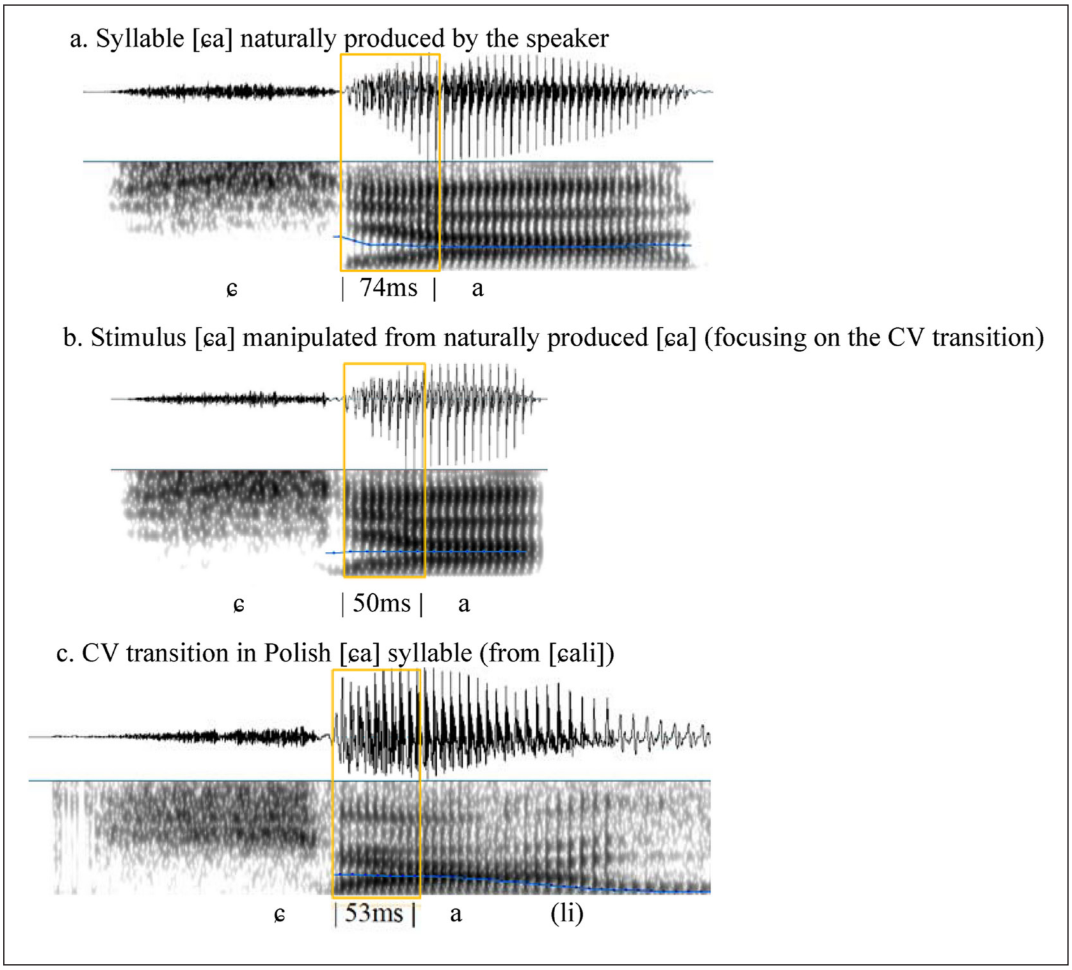

Figure 1: Manipulation of the stimulus syllable.

Note: Polish sound file was obtained from http://www.phonetics.ucla.edu/course/chapter7/polish/polish.html.

an interval, 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 production. Then a waveform was resynthesized by overlapping and adding the remaining frames. Similar manipulation applied to the syllables [tɕa tɕʰa ɕou tɕou tɕʰou] respectively. We compared the resultant syllables like [ɕa] with the naturally produced syllable [ɕa] in Polish in **Figure 1(c)** and the CV transition in our resultant CV syllable was close to the natural CV transition in Polish [ɕa] in duration and F2 onset. In general, the manipulated syllables all sounded natural. 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). Yet, to facilitate the comparison of response time across vowel contexts, we normalized the vowels [_i], [_a] and [_ou] all to 120 ms (50 ms CV transition plus 70 ms steady vowel formants).

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 & Charpentier, 1990) as described above.

Fourth, the root-mean-square intensity of the vocalic portion was normalized to 70 dB in Praat, and the amplitude of the vowel faded out to zero within the last 20 ms. The intensity of the onset sibilants was left intact from the naturally produced tokens as intensity could potentially be a place cue for the sibilant contrast. Moreover, the stimulus syllable was selected such that the sibilant intensity was the closest to the mean of the six repetitions of the syllable. The spectrograms of the syllables after manipulation are given in **Figure 2**.

After the manipulation, the CV pairs in (6) were formed where the dental and palatal sibilants contrast in the vowel contexts [_i], [_a], and [_ou] (columns A, B, and C). We also formed the pairs in column D, e.g., [sɹ̩-ɕi], with the dental sibilants followed by the apical vowel [ɹ̩] and the palatals followed by [i]. These pairs are the actual contrasts in Mandarin and many other Chinese dialects, and they were included to compare with pairs like [si-ɕi].

(6) Stimulus pairs for the perceptual experiment

	A. [_i]	*B.* [_a]	*C.* [_ou]	*D.* [_ɹ̩_i]
Fricatives	si-ɕi	sa-ɕa	sou-ɕou	sɹ̩-ɕi
Unasp. Affricates	tsi-tɕi	tsa-tɕa	tsou-tɕou	tsɹ̩-tɕi
Asp. Affricates	tsʰi-tɕʰi	tsʰa-tɕʰa	tsʰou-tɕʰou	tsʰɹ̩-tɕʰi

Each cell in (6) resulted in 4 stimulus pairs. For example, [si-ɕi] and [ɕi-si] were formed as different pairs and [si-si] and [ɕi-ɕi] as identical pairs. Thus, the 12 cells in (6) led to 48 stimulus pairs. Apart from these pairs, 16 filler pairs were added, e.g., [tu-ti], [ti-tu], [tu-tu], [ti-ti]. 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 & Logan, 1985; Johnson & Babel, 2010). An additional 50 ms was added between the pairs whose onsets were [ts tɕ tsʰ tɕʰ] to compensate for the duration of oral closure before the release of an affricate. The same settings applied to the fillers.

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.

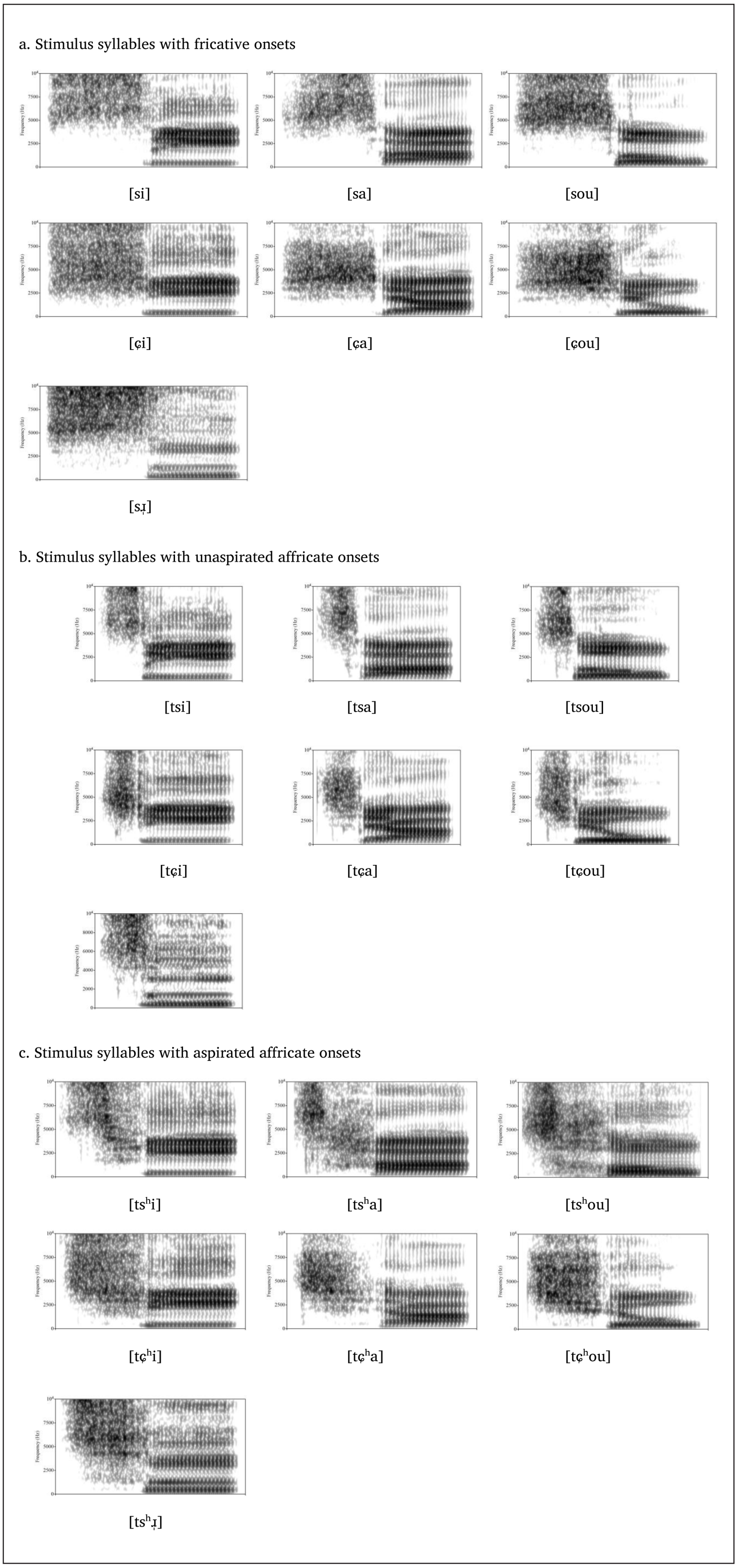

Figure 2: Spectrograms of manipulated stimulus syllables.

The listeners were instructed to respond as quickly as possible, with a response time goal of 500 ms, following Johnson and Babel (2010) and Babel and Johnson (2010). After each trial, feedback was presented on the screen about their response time (longer than 500 ms or not) and accuracy (correct or incorrect) on the pair just heard, as well as the overall accuracy of their judgments up to the current pair.[6] The listeners had 1.5 seconds to respond. If they did not respond within 1.5 seconds, the next pair would start automatically. The main experiment was preceded by a practice session, which included half of the pairs in (6). In the main experiment, each stimulus pair was repeated four times. Thus, one subject gave 192 responses (12 comparisons × 4 pairs × 4 repetitions) excluding the fillers. The whole experiment lasted about 35 minutes.

3.2. Predictions

It was hypothesized that, for dental vs. palatal sibilants, the [_i] context would introduce reduced perceptual distinctiveness compared with other vowels. For a speeded AX-discrimination task, we predicted that the [_i] context will lead to a longer response time than other vowels. We also hypothesized that speeded AX discrimination is able to bypass the influence of L1 phonology and elicit psychoacoustic perception. Therefore, the English and Chinese listeners should have no difference in their responses, even though Chinese and English differ in their sibilant inventory.

3.3. Results

For each stimulus pair, we calculated the response time from the onset of the sibilant in the second stimulus, e.g., from the start of the frication noise of [ɕi] in the pair [si-ɕi]. The raw response time was transformed into Log Response Time (LogRT) and we analyzed the listeners' 'different' responses to phonetically different pairs (i.e., the correct responses to different pairs). For each onset pair per listener, the data points outside two standard deviations from the mean of LogRT were trimmed off to exclude outliers. That is, for each listener, we trimmed off the outliers outside two standard deviations separately for the stimulus pairs whose onsets were [s-ɕ], [ts-tɕ], and [tsʰ-tɕʰ]. This was necessary because the onset pairs intrinsically differ in their durations ([s-ɕ] = 125 ms, [t͡s-t͡ɕ] = 50 ms, and [tsʰ-tɕʰ] = 100 ms) and the onset duration was included in the response time. The 30 listeners gave 2865 responses to phonetically different pairs and 2701 (=94%) were correct ones. Out of the 2701 correct responses, a total of 120 tokens (=4.4%) were excluded as outliers, including 38 tokens below two standard deviations and 82 tokens above two standard deviations. The statistical analysis applied to the remaining 2581 tokens. The mean LogRTs for each CV pair are plotted in **Figure 3** with separate graphs for Chinese and English listeners.

The whole data set (=the 2581 tokens) was analyzed in the Linear Mixed Effects Models using the *lmer* function in the R package lme4 (Bates et al., 2015a, b) and the *p*-values were determined by the R package lmerTest (Kuznetsova et al., 2015). LogRT was taken as the dependent variable, and the predicting variables were Vowel (the four vowel contexts, [_i], [_a], [_ou] and [_ɻ̩_i]), Onset (the three onset pairs, [s-ɕ], [ts-tɕ], [tsʰ-tɕʰ]), and NativeLg (the listeners' native languages as English vs. Chinese). The baseline for Vowel was [_i] and that for Onset was [s-ɕ] and we used contrastive coding for NativeLg (Chinese = –0.5 and English = 0.5). The random factors were Subject, Subject:Onset, and Subject:Vowel.

The fixed effects in the final model are presented in **Table 3** and several steps were taken before arriving at this model. First, a null model with Subject, Subject:Onset, and Subject:Vowel as the random factors was compared separately with three superset models

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

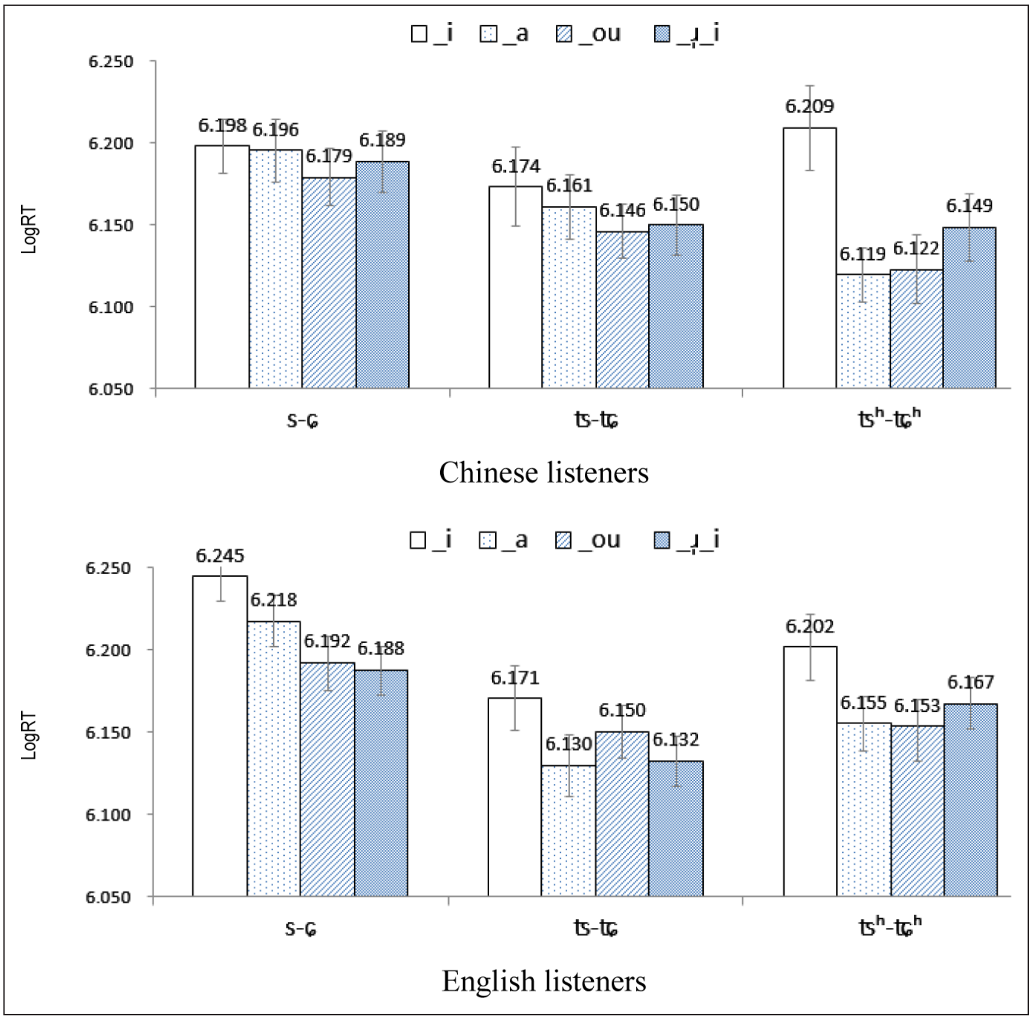

Figure 3: Mean LogRT for Chinese listeners (upper) and English listeners (lower). The response time was counted from the onset of the second stimulus and transformed into LogRT. (6.10 = 446 ms, 6.15 = 469 ms, 6.20 = 493 ms, 6.25 = 518 ms.) The error bars indicate the standard errors of the mean values.

Table 3: Fixed effects in the mixed-effect linear regression for LogRT.

	Estimate	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	6.232	0.019	45.08	335.508	< 0.001***
Vowel(_a)	−0.035	0.010	93.71	−3.298	0.001**
Vowel(_ou)	−0.039	0.010	93.25	−3.725	< 0.001***
Vowel(_ɹ̩_i)	−0.036	0.010	93.09	−3.460	< 0.001***
Onset(ts-tɕ)	−0.055	0.009	57.16	−6.080	< 0.001***
Onset(tsʰ-tɕʰ)	−0.041	0.009	57.03	−4.532	< 0.001***

Model: LogRT ~ Onset + Vowel + (1|Subject) + (1|Subject:Onset) + (1|Subject:Vowel).
Baselines: Onset = [s-ɕ] and Vowel = [_i].
Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

with Vowel, Onset, or NativeLg as the predicting variables. The addition of Vowel and Onset both significantly improved the model (Vowel: $X^2 = 20.055$, $df = 3$, $p < .001$; Onset: $X^2 = 46.708$, $df = 2$, $p < .001$) and the addition of NativeLg did not significantly improve the model ($X^2 = 0.086$, $df = 1$, $p = .769$). Second, a model with Vowel, Onset, and NativeLg as the predicting variables and Subject, Subject:Onset, and Subject:Vowel as the random factors was compared separately with three superset models each including a two-way interaction, e.g., Vowel*Onset, Vowel*NativeLg, or Onset*NativeLg. It turned out that the addition of none of these interactions significantly improved the model. Finally, a model with all the two-way interactions, i.e., Vowel*Onset, Vowel*NativeLg, and Onset*NativeLg, was compared with a superset model with the three-way interaction Onset*Vowel*NativeLg. The addition of the three-way interaction did not significantly improve the model. Therefore, the final model in **Table 3** included Vowel and Onset as the fixed effects and Subject, Subject:Onset, and Subject:Vowel as the random factors.[7]

From **Figure 3**, we can see that the effect of Vowel was due to the fact that, for each onset pair, the [_i] context had a longer RT than the contexts [_a], [_ou], and [_ɹ̩_i]; this was also shown by the coefficient estimates in **Table 3**. There is no evidence for a difference between Chinese and English listeners in terms of the Vowel effect since there was no Vowel*NativeLg interaction or Vowel*Onset*NativeLg interaction. We further checked the difference among the four vowel contexts [_i], [_a], [_ou], and [_ɹ̩_i] using [_i], [_a], [_ou] alternatively as the baseline. There were significant differences between [_i] and the other three vowel contexts and no significant differences among the three contexts [_a], [_ou], and [_ɹ̩_i], as summarized in **Table 4**.

The effect of Onset was not of interest as the onset pairs had intrinsically different durations ([s-ɕ] = 125 ms, [ʦ-ʨ] = 50 ms, [ʦʰ-ʨʰ] = 100 ms) and the response time was calculated from the onset of the second syllable in each pair. We do not discuss Onset further since there was no interaction between Onset and the other predicting variables.

The overall error rate of the responses for different pairs was 6% and the 30 listeners made a total of 164 discrimination errors (i.e., phonetically different pairs judged as being 'the same'): 54 from the 10 Chinese listeners and 110 from the English listeners. **Figure 4** below provides the number of errors in each CV pair for Chinese and English listeners. Generally, the [_i] context induced a larger number of errors compared with other vowel contexts for both groups of listeners. Due to the small number of data points, no statistics was run on the accuracy data.

3.4. Discussion

In the response time data, the main effect of vowel context came from the fact that the [_i] context introduced a significantly longer response time than other vowels. Assuming that a longer RT indicates less distinctiveness, the result confirms the hypothesis that the dental vs. palatal sibilants are less distinct in the [_i] context than in other vowel contexts. The absence of Vowel*Onset interaction indicated that the same vowel effect held for all

Table 4: Differences between vowel contexts: *t* value (*p* value).

	[_a]	**[_ou]**	**[_ɹ̩_i]**
[_i]	−3.358 (= **0.001****)	−3.552 (<**0.001*****)	−3.768 (<**0.001*****)
[_a]		−0.415 (0.679)	−0.190 (0.849)
[_ou]			0.226 (0.822)

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

[7] The same result was obtained from a top-down procedure of model fitting, i.e., starting with the full model and gradually removing non-significant predicting variables.

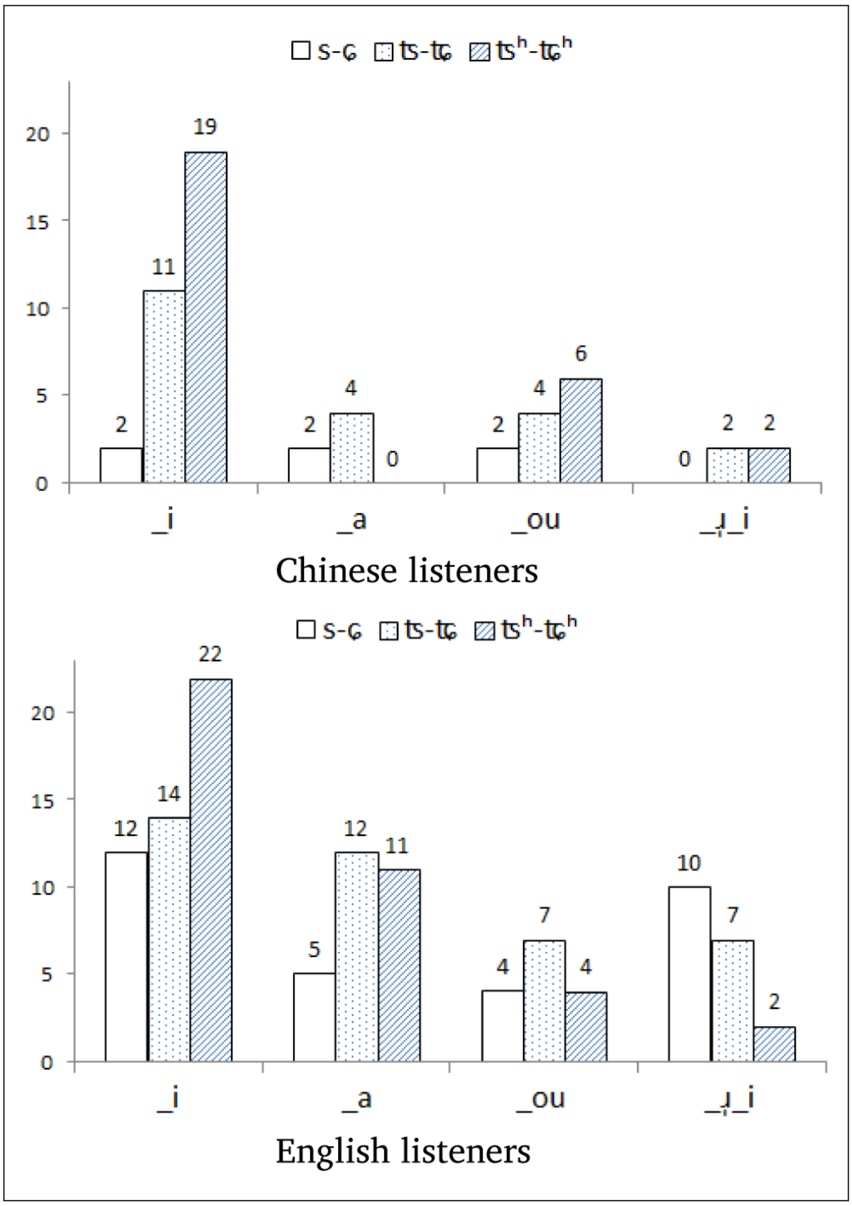

Figure 4: Number of discrimination errors by Chinese listeners (upper) and English listeners (lower).

onset pairs. That is, the [_i] context generally induces less distinctiveness to the contrasts [s-ɕ], [ts-tɕ], and [tsʰ-tɕʰ]. Moreover, the lack of a significant effect of NativeLg and its interactions with Vowel and Onset supported the idea that English and Chinese listeners did not differ in their responses, providing evidence for a language-independent effect of vowel context on the perceived distinctiveness of sibilant contrast. This result is consistent with the typological pattern that contrastive dental vs. palatal sibilants in the [_i] context tend to be avoided across Chinese dialects even when there are only two sibilant places in the sound inventory. Moreover, the dental vs. palatal sibilants are less distinct in the [_i] context than in the [_ɿ_i] context (i.e., dentals followed by apical vowel and palatals followed by [i]) and this is consistent with the enhancement account of Mandarin apical vowels (Stevens et al., 2004, Lee & Li, 2003, Lee-Kim, 2014a).

For the English subjects, the directionality of the vowel effect cannot be attributed to their L1 phonology. English has the vowels [_ɑ], [_oʊ] and [_i] but no word-initial contrast between the dental and palatal sibilants. Thus, without the dental vs. palatal contrast in any vowel context, L1 phonology should not have biased English listeners' perception. On the other hand, even if the English listeners had tried to map the dental vs. palatal contrast to L1 contrast, the observed results still cannot be accounted for by their L1 phonology. More specifically, English has the alveolar vs. postalveolar fricatives ([s] vs. [ʃ]) in the vowel contexts [_ɑ], [_oʊ], and [_i], as in (7).

(7) Alveolar vs. postalveolar fricatives in English

	a.C + [ɑ]		b.C + [oʊ]		c.C + [i]	
Fricatives [s] vs. [ʃ]	/**sɑ**k/ 'sock'	/ʃɑk/ 'shock'	/**soʊ**/ 'so'	/ʃoʊ/ 'show'	/**si**/ 'sea'	/ʃi/ 'she'

The English listeners may have tried to map the non-native pair [s-ɕ] to the L1 [s-ʃ], as in **Figure 5(a)**, i.e., a 'two-category assimilation' in the Perceptual Assimilation Model (PAM, Best, 1995; Best et al., 2001); alternatively, they may have left the non-native [ɕ] uncategorized, as in **Figure 5(b)**, i.e., an uncategorized-categorized assimilation in PAM. In either case, the mapping should not bias the perceived distinctiveness towards any vowel context, since the English [s-ʃ] contrast is observed in [_ɑ], [_oʊ] as well as [_i], as in (7). Therefore, the observed vowel effect ([_i] being less distinct) must come from factors other than L1 phonology, i.e., the psychoacoustic perception of sibilant distinctiveness in the [_i] context as compared with the other vowel contexts.

For Chinese listeners, the reduced distinctiveness in the [_i] context agrees with the phonotactics of Mandarin, as Mandarin has the dental vs. palatal contrast in the [_a], [_ou], and [_ɹ̩_i] (i.e., dentals followed by apical vowel and palatals followed by [i]) contexts, but not the [_i] context. But the fact that English listeners did not behave differently from the Mandarin listeners indicates that the observed vowel effect cannot be based on the language-specific phonotactics of Mandarin.

4. General discussion

4.1. Phonetic basis of the vowel effect

There are three possibilities for the acoustic basis of the vowel effect observed in our experimental results. First, the vowel effect may have resulted from the acoustic properties of the sibilant consonants. For example, it could be that the dental vs. palatal sibilants were acoustically more similar in the [_i] context, due to coarticulation or palatalization from the vowel [_i] (Lee & Li, 2003). To examine this possibility, we measured the center of gravity (COG) and the intensity of the sibilants. More specifically, COG was measured over the center 80% of the sibilants for the frequency range 0–10 kHz and intenstity was measured over the entire consonant. For the aspirated affricates [ts^h] and [tɕ^h], the COG was measured on the turbulent noise before the aspiration portion. The results were listed in **Table 5**. Note that the [_ɹ̩_i] context was not included because the stimulus syllables, e.g., [sɹ̩-ɕi], differ in both consonants and vowels and the comparison of COG and intensity would not capture the acoustic difference between the relevant stimulus pairs.

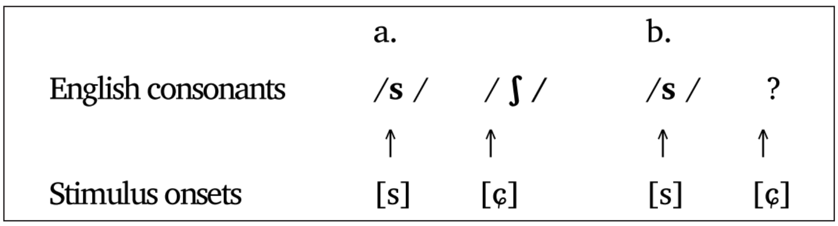

Figure 5: Mapping of stimuli onsets to English consonants.

Table 5: Acoustic difference between the sibilants in the stimulus pairs.

a. Center of gravity (Hz).

Vowel	[s]	[ɕ]	*ΔCOG*	[ts]	[tɕ]	*ΔCOG*	[tsʰ]	[tɕʰ]	*ΔCOG*
[_i]	8036	4888	***3148***	8153	4884	***3269***	6821	4811	***2010***
[_a]	6496	4441	*2055*	6888	5477	*1411*	6184	4825	*1359*
[_ou]	6150	4463	*1687*	6179	4586	*1593*	5420	3895	*1525*

b. Intensity (dB).

Vowel	[s]	[ɕ]	*ΔIntensity*	[ts]	[tɕ]	*ΔIntensity*	[tsʰ]	[tɕʰ]	*ΔIntensity*
[_i]	55.1	59.3	*4.2*	63.1	64.1	*1*	60.2	61.1	*0.9*
[_a]	52.9	57.1	*4.2*	57.9	57.9	*0*	56.1	56.9	*0.8*
[_ou]	57.5	60.1	*2.6*	65.2	60.3	*4.9*	56.5	60.5	*4.0*

Note: Δ*COG* and Δ*Intensity* indicates the difference between COG and Intensity of the two sibilants on the left.

As shown in **Table 5(a)**, the COG differences (ΔCOG) between the dental and palatal sibilants were larger in the [_i] context than in the [_a] and [_ou] contexts. In **Table 5(a)**, for example, the sibilant COG difference of [si-ɕi] was larger than that of [sa-ɕa], which is opposite to the assumption that the reduced perceptual distinctiveness in the [_i] context came from a smaller acoustic difference between the dentals and the palatals in the [_i] context. There was also no systematic pattern of intensity difference corresponding to the observed reduced distinctiveness in the [_i] context. In **Table 5(b)**, for example, the intensity difference (Δ*Intensity*) of [si-ɕi] was larger than that of [sou-ɕou] whereas the intensity difference of [tsi-tɕi] was smaller than that of [tsou-tɕou]. Therefore, it is unlikely that the vowel effect ([_i] being less distinct) was rooted in the acoustic similarity of the onset sibilants.

Second, it is possible that the vowel effect came from the acoustic properties of the vowels. That is, it may be that in the stimulus pairs the vowel [_i]s (e.g., in [tsi] vs. [tɕi]) were more similar to each other than the vowel [_a]s (e.g., in [tsa] vs. [tɕa]). To test this possibility, we measured F1 and F2 at the mid point of the steady formant portion of each CV syllable. The results were listed in **Table 6**. To evaluate the formant difference, ΔF1 and ΔF2 were included in **Table 6**, which were the differences between the steady formant midpoint values in a stimulus pair. For example, for the pair [si-ɕi], ΔF2 = 2578–2549 = 29.

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 & Karo, 1968; Mermelstein, 1978; Nord & Sventelius, 1979), but JND as low as 1.5% has also been reported (Kewley-Port & Watson, 1994). For vowels in consonantal contexts, Mermelstein (1978) reported mean difference limens of 60 Hz for F1 and 176 Hz for F2. To examine if the formant difference is perceivable, a ratio (in parentheses in **Table 6**) was calculated by dividing the formant difference value (ΔF) by the higher formant value on the left. For example, in the first cell, ΔF2 was 29Hz (i.e., 2578–2549) and the reference F2 was that of [si] (2578Hz). Thus, the ratio was 29/2578 = 1%. As shown in **Table 6**, most of the ratio values were below or close to 5% and only the ΔF1 of [sou-ɕou] and [tsou-tɕou] were above 10%. Thus, most of the formant differnces should not have led to salient perceptual differences. Put simply, in each stimulus pair, the two

Table 6: Acoustic difference between vowels in the stimulus pairs: Steady F1 and F2 (Hz).

Vowel		[s]	[ɕ]	Δ*F*	[ts]	[tɕ]	Δ*F*	[tsʰ]	[tɕʰ]	Δ*F*
[_i]	*F2*	2578	2549	*29* (1%)	2573	2555	*18* (1%)	2636	2578	*58* (2%)
	F1	395	377	*18* (5%)	371	377	*−6* (2%)	371	377	*−6* (2%)
[_a]	*F2*	1474	1445	*29* (2%)	1370	1435	*−65* (5%)	1358	1416	*−58* (4%)
	F1	1039	951	*88* (8%)	1020	998	*22* (2%)	951	893	*58* (6%)
[_ou]	*F2*	1068	980	*88* (8%)	957	976	*−19* (2%)	885	957	*−72* (8%)
	F1	574	487	*87* (15%)	504	449	*55* (11%)	558	540	*18* (3%)

Note: Δ*F* indicates the difference between the vowel formant values (F2 or F1). The percentage in parentheses indicates the value of ΔF divided by the higher formant value on the left.

vowels are close to each other in terms of F1 and F2, and it is unlikely for the observed vowel effect to have come from the difference in the steady-state vowels.

Third, the vowel effect might have come from the properties of consonant-vowel formant transition in different vowels. Formant transition has been shown to be important in the place identification of consonants (Delattre et al., 1955; Whalen, 1981, 1991; Nowak, 2006; Babel & 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. Regarding 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 smaller palatal transition and thus a less robust perceptual cue (Lee-Kim, 2014a). To investigate this possibility, we evaluated the transitonal difference between the dental and palatal sibilants in each CV pair, e.g., [si-ɕi], and compared this difference across the three vowel contexts. In **Table 7** below, $F2_{onset}$ 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 the following vowels.[8] We calculated $\Delta F2_{onset}$, which was 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 held for $\Delta F2_{offset}$.

The values of $\Delta F2_{offset}$ were generally small and therefore the transitional difference between the two syllables in a stimulus pair was mostly determined by the value of $\Delta F2_{onset}$. As shown in **Table 7**, $\Delta F2_{onset}$ for the same sibilant pair was always smaller in the [_i] context than in the [_a] and [_ou] contexts. In other words, the transitional difference between the dentals and palatals was the smallest in the [_i] context, which is consistent with the observation that the [_i] context induced reduced distinctiveness between the dental and palatal sibilants compared with the other vowel contexts.

To summarize, based on the measurements of onset COG and intensity, the vowel effect is unlikely to be rooted in the acoustic differences in the sibilants, nor is it likely to come from the acoustic difference in the steady vowel formants. Rather, the vowel effect is most likely to come from the fact that the formant transitions of the dentals and palatals are acoustically more similar in the [_i] context than in other vowel contexts. This is schematized in **Figure 6**. Given that the COG difference in the sibilants is in fact larger

[8] $F2_{offset}$ was a different measurement from F2 steady state. More specifically, $F2_{offset}$ was measured at the end of the CV transition and F2 steady state was measured at the midpoint of the steady vowel. The decision on the end of the CV transition was made referring to both F1 and F2, and it is possible that the $F2_{offset}$ value differs from the F2 steady state value. In our stimuli, the differences were generally small and below the JND of F2 reported in Mermelstein (1978).

Table 7: Acoustic difference between formant transitions in the stimulus pairs (Hz).[9]

Vowel		[s]	[ɕ]	$\Delta F2_{ons, off}$	[ts]	[tɕ]	$\Delta F2_{ons, off}$	[ts^h]	[tɕ^h]	$\Delta F2_{ons, off}$
[_i]	$F2_{onset}$	1916	2440	***524***	1785	2331	***546***	2593	2244	***–349***
	$F2_{offset}$	2615	2528	*–87*	2484	2571	*87*	2659	2550	*109*
[_a]	$F2_{onset}$	1375	2065	***690***	1248	1974	***726***	1304	1828	***524***
	$F2_{offset}$	1466	1539	*73*	1393	1539	*146*	1348	1435	*87*
[_ou]	$F2_{onset}$	1103	2010	***907***	1194	2010	***816***	939	1756	***817***
	$F2_{offset}$	1139	1030	*–109*	1176	1121	*–55*	994	1030	*36*

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.
$F2_{offset}$ indicated 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.

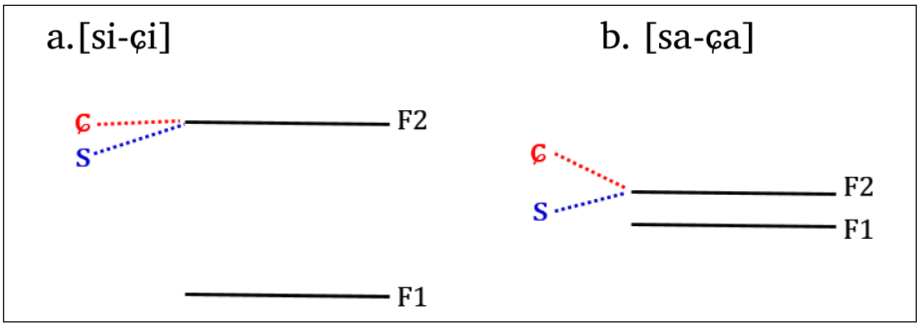

Figure 6: Schematic illustration of the vowel effect on the consonant place distinction: [_i] vs. [_a].

in the [_i] context, our results also suggest that in the listeners' discrimination of the CV pairs, the difference in the CV transition has generally overidden the difference in the COG of the sibilants. However, we must recognize the following caveats to our conclusion: (a) Our sibilant measurements were restricted to COG and intensity, and it is possible that the onset pairs were more similar (or more distinct) in other acoustic aspects; and (b) these conclusions were drawn from the specific stimuli use in our experiment, and it is possible that the realization of the dental-i sequence in a different language is different (e.g., with more palatalization on the sibilant).

4.2. *Contrast distinctiveness and sound systems*

The linguistic literature generally agrees that human languages prefer sounds that are more distinct from each other. The 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

[9] This table shows that, in the [_i] context, the $F2_{onset}$ values were higher after the [ɕ] [tɕ] than after [s] [ts]. Yet, the reverse pattern was observed for [ts^h] and [tɕ^h]: The $F2_{onset}$ value after [ts^h] was higher than that after [tɕ^h], which can also be seen in the spectrograms of the stimulus syllables [ts^{h}i] and [tɕ^{h}i] in Figure 2(c). It is unclear why the patterns of the aspirated affricates differ from those of the fricatives and the unaspirated affricates. A possible explanation might be that, in [ts^{h}i], the aspiration portion masks the energy at lower frequency, which makes it difficult to detect the formant transition at lower frequency at the very beginning of the vocalic portion. Thus when meaning $F2_{onset}$ in [ts^{h}i], energy concentration at higher frequencies might have been detected as F2, which led to higher $F2_{onset}$ value after [ts^h]. On the other hand, for the palatal aspirated affricate in [tɕ^{h}i], the aspiration did not affect the formation transition, as there was literally no formant transition in terms of F2 due to the shared place of articulation between [tɕ^h] and [i].

(Steriade, 1997, 1999, 2001, 2008). Our study falls into the second category and the experimental results suggest that the dental and palatal sibilants are less distinct in the [_i] context than in the [_a] and [_ou] contexts where the place distinction can be better cued by larger formant transition differences between the dental and palatal sounds. This suggests that CV combination could potentially be taken as a unit on 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 et al., 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 experimental study in this paper showed that, regarding 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. This is similar to the observation in Lee-Kim (2014a) that crosslinguistically, sibilant place contrasts in the [_i] context tend to be avoided. That is, the less recoverability of place cues in the [_i] context showed its effect in language typology.

Regarding the sound system of a particular language, evaluating distinctiveness in a unit larger than a 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., [tsɨ̞-tɕi]).[10] 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 in 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). In other words, the phonological change may be deemed as an enhancment 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. Conclusions

The typological survey across Chinese dialects supports a contrast enhancement view of the formation of apical vowels. In accord with this view, we hypothesized that the dental vs. palatal sibilant contrast is perceptually less distinct in the [_i] context than in other vowel contexts. To test this hypothesis, a speeded-AX discrimination task was conducted with English and Chinese listeners, the results of which showed reduced perceptual distinctiveness in [_i] context compared with other vowel contexts, confirming the hypothesis. Acoustic examination of the stimulus pairs further suggested that the vowel effect was more likely to be rooted in the less salient formant transition difference of the [_i] context rather than in the acoustic properties of the onsets or the steady-state

[10] This change represents a true historical sound change in Mandarin from the 7th century to the 11th century (Chen, 1976; Wang, 1985; Li & Zhou, 1993).

vowels. The vowel effect also suggests that it may be useful to adopt units larger than segments in the evaluation of contrast distinctiveness of a sound system.

We have shown in this paper that sibilant contrasts in the [_i] context are less distinct than in other vowel contexts. However, it is not clear if and how this effect would apply to other consonant types, e.g., stops, nasals, or liquids. Moreover, our experimental results make no clear predictions on how a language will avoid a less distinct contrast like [tsʰi-tɕʰi], e.g., when to introduce the apical vowel and when to avoid the contrast via consonant neutralization. In addition, as observed in the typological survey, a number of Chinese dialects allow contrastive dental vs. palatal sibilants in the [_i] context. Further studies need to investigate how the contrasts are realized in those dialects and whether there are sound changes to avoid the contrasts. More generally, more empirical studies are necessary with regard to the relationship between psychoacoustic distinctiveness and phonological contrast in a sound system.

Acknowledgements

We thank Profs. Allard Jongman, Joan Sereno, Annie Tremblay, and our colleagues at The University of Kansas Phonetics and Psycholinguistics Laboratory for their suggestions and help. Thanks to the audience at *The 19th Meeting of the Mid-Continental Phonetics and Phonology Conference*, *The 168th Meeting of the Acoustical Society of America*, and *The 27th North American Conference on Chinese Linguistics* for their feedback. To Profs. Ping Jiang-King, Peggy Mok, Thomas Hun-tak Lee, William S-Y. Wang, Xiaonong Zhu, Lian Hee Wee, Marzena Źygis, Paul Kiparsky, Alan Yu, Edward Flemming, Gang Peng, Ms. Winnie Cheung, among others, for their advice on earlier versions of this study. Special thanks to the anonynous reviwers of *Laboratory Phonology* and Prof. Lisa Davidson for their vaulable comments.

Competing Interests

The authors have no competing interests to declare.

References

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Babel, M. and Johnson, K. 2010. Accessing psycho-acoustic perception with speech sounds. *Journal of Laboratory Phonology, 1*, 179–205. DOI: https://doi.org/10.1515/labphon.2010.009

Babel, M. and McGuire, G. 2013. Listener expectations and gender bias in nonsibilant fricative perception. *Phonetica, 70*, 117–151. DOI: https://doi.org/10.1159/000354644

Bates, D., Maechler, M., Bolker, B. and 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. and 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.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 171–204. Baltimore, MD: York Press.

Best, C. T., McRoberts, G. W. and 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. DOI: https://doi.org/10.1121/1.1332378

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International 5*(9/10), 341–345.
Boersma, P. and Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology, 25*, 217–270. DOI: https://doi.org/10.1017/S0952675708001474
Booij, G. 1999. *The Phonology of Dutch*. New York: Oxford University Press.
Boomershine, A., Hall, K. C., Hume, E. and Johnson, K. 2008. The impact of allophony versus contrast on speech perception. In: Avery, P., Dresher, E. and Rice, K. (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition*, 145–171. New York: Mouton de Gruyter.
Cai, Q. 1987. Guangxi Lianzhou fangyan yinxi [Phonology of Lianzhou dialect in Guangxi Autonomous Region]. *Fangyan, 1*, 49–57.
Cao, J. 1990. *Xiandai Yuyin Zhishi* [*Fundamentals of Modern Phonetics*]. Beijing: Renmin Jiaoyu Press (People's Education Press).
Chao, Y. 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 Joos, M. (Ed.), *Readings in Linguistics* (3rd Edition, 1963), 38–54. New York: American Council of Learned Societies.
Chen, M. Y. 1976. From Middle Chinese to Modern Peking. *Journal of Chinese Linguistics 2*, 113–227.
Cheung, Y. M. 2004. A perceptual analysis of the apical vowels in Beijing Mandarin (MPhil. thesis). Hong Kong: City University of Hong Kong.
Delattre, P. C., Liberman, A. M. and Cooper, F. S. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*, 769–773. DOI: https://doi.org/10.1121/1.1908024
Duanmu, S. 2007. *The Phonology of Standard Chinese*, 2nd Edition. Oxford: Oxford University Press.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. and Mehler, J. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568–1578. DOI: https://doi.org/10.1037/0096-1523.25.6.1568
Feng, L. 1985. Beijinghua yuliu zhong shengyundiao de shichang [Duration of consonants, vowels and tones in colloquial Beijing Mandrain]. In: Lin, T. and Wang, L. (Eds.), *Beijing Yuyin Shiyan Lu [Experimental Studies in the Sounds of 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. DOI: https://doi.org/10.1121/1.1907979
Flege, J. E., Takagi, N. and 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. DOI: https://doi.org/10.1121/1.414884
Flemming, E. 2002. *Auditory Representation in Phonology*. London & New York: Routledge.
Flemming, E. 2004. Contrast and perceptual distinctiveness. In: Hayes, B., Kirchner, R. and Steriade, D. (Eds.), *Phonetically Based Phonology*, 232–276. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511486401.008
Flemming, E. 2005. Speech perception and phonological contrast. In: Pisoni, D. B. and Remez, R. E. (Eds.), *The Handbook of Speech Perception*, 156–181. Oxford: Blackwell. DOI: https://doi.org/10.1002/9780470757024.ch7
Gandour, J. T. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics, 11*. 149–176.

Greenberg, J. H. and Jenkins, J. J. 1964. Studies in the psychological correlates of the sound system of American English. *Word, 20*, 157–177. DOI: https://doi.org/10.1080/00437956.1964.11659816

Hiroyuki, A. 2004. Fujian Shibei fangyan yinxi [Phonology of Shibei dialect in Fujian Province]. *Fangyan, 1*, 76–91.

Huang, X. 1988. Hunan Jiangyong fangyan yinxi [Phonology of Jiangyong dialect in Hunan Province]. *Fangyan, 3*, 161–176.

Hume, E. and Johnson, K. 2001. A model of the interplay of speech perception and phonology. In: Hume, E. and Johnson, K. (Eds.), *The Role of Speech Perception in Phonology*, 3–26. New York: Academic Press.

Janson, T. 1986. Cross-linguistic trends in the frequency of CV sequences. *Phonology Yearbook, 3*, 179–195. DOI: https://doi.org/10.1017/S0952675700000634

Johnson, K. and Babel, M. 2010. On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics, 38*, 127–136. DOI: https://doi.org/10.1016/j.wocn.2009.11.001

Kakusho, O. and Kato, K. (1968). Just discriminable change and matching range of acoustic parameters of vowels. *Acustica, 20*, 46–54.

Kensinger, K. M. 1963. The phonological hierarchy of Cashinahua. In: Elson, B. F. (Ed.) *Studies in Peruvian Indian Languages*, 207–217. Norman: Summer institute of Linguistics.

Kewley-Port, D. and Watson, C. S. 1994. Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America, 95*, 485–496. DOI: https://doi.org/10.1121/1.410024

Keyser, S. J. and Stevens, K. 2006. Enhancement and overlap in the speech chain. *Language, 82*, 33–63. DOI: https://doi.org/10.1353/lan.2006.0051

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. and Lindblom, B. 1992. Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science, 255*, 606–608. DOI: https://doi.org/10.1126/science.1736364

Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. 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, P. and Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford Cambridge, MA: Blackwell Publishers Ltd.

Ladefoged, P. and Wu, Z. 1984. Place of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics, 12*, 267–278.

Lee, C. and Li, Z. 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-Kim, S. 2014a. Contrast neutralization and enhancement in phoneme inventories: Evidence from sibilant place contrast and typology (Doctoral dissertation). New York: New York University.

Lee-Kim, S. 2014b. Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association, 44*, 261–282. DOI: https://doi.org/10.1017/S0025100314000267

Lee, W. and Zee, E. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association, 33*(1), 109–122. DOI: https://doi.org/10.1017/S0025100303001208

Lehiste, I. 1970. *Suprasegmentals.* Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review, 74*, 431–461. DOI: https://doi.org/10.1037/h0020279

Liljencrants, J. and Lindblom, B. 1972. Numerical simulation of vowel quality contrasts: The role of perceptual contrast. *Language, 48*, 839–62. DOI: https://doi.org/10.2307/411991
Lin, Y. 1989. Autosegmental treatment of segmental processes in Chinese phonology (Doctoral dissertation). Austin, TX: The University of Texas at Austin.
Lindblom, B. 1986. Phonetic universals in vowel systems. In: Ohala, J. and Jaeger, J. J. (Eds.), *Experimental Phonology*, 13–44. Orlando, Fl: Academic Press.
Lindblom, B. 1990. Phonetic content in phonology. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS), 11*, 101–118.
Lisker, L. 1963. *Introduction to Spoken Telugu*. New York: American Council of Learned Societies.
Liu, C. 2010. Hanyu fangyan cayin shengxue shiyan yanjiu [An acoustic study of fricatives in Chinese dialects] (Master thesis). Nanjing: Nanjing Normal University.
Liu, L. 1995. Jiangxisheng Dayü (Nan'an) fangyan yinxi [Phonology of Dayü (Nan'an) dialect in Jiangxi Province]. *Fangyan, 1*, 63–69.
Li, Z. and Zhou, C. 1993. *Hanzi Gujin Yinbiao [Historical Pronunciation of Chinese Characters]*. Beijing: Zhonghua Shuju Press (Zhonghua Book Company).
Lu, S. 1995. Jinggangshan Kejiahua yinxi [Phonology of Hakka dialect in Jinggangshan]. *Fangyan, 2*, 121–127.
Maddieson, I. 1984. *Patterns of Sound*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511753459
Martinet, A. 1952. Function, structure, and sound change. *Word, 8*, 1–32. DOI: https://doi.org/10.1080/00437956.1952.11659416
McGuire, G. 2010. A brief primer on experimental designs for speech perception research (Unpublished manuscript). Santa Cruz, CA: UC Santa Cruz.
Mermelstein, P. 1978. Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America, 63*, 572–580. DOI: https://doi.org/10.1121/1.381756
Miller, W. R. 1965. *Acoma Grammar and Texts*. Berkeley and Los Angeles, CA: University of California Press.
Mohr, B. and Wang, W. S. 1968. Perceptual distance and the specification of phonological features. *Phonetica, 18*, 31–45. DOI: https://doi.org/10.1159/000258597
Moulines, E. and Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*(5), 453–467. DOI: https://doi.org/10.1016/0167-6393(90)90021-Z
Nord, L. and 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, P. M. 2006. The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics, 34*, 139–152. DOI: https://doi.org/10.1016/j.wocn.2005.03.001
Pan, Y. 2010. Hanyu fangyan secayin shengxue shiyan yanjiu [An acoustic study of affricates in Chinese dialects] (Master thesis). Nanjing: Nanjing Normal University.
Perception Research Systems. 2007. Paradigm Stimulus Presentation, Retrieved from: http://www.paradigmexperiments.com.
Pisoni, D. 1973. Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13*, 253–260. DOI: https://doi.org/10.3758/BF03214136
Pisoni, D. and Tash, J. 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15*, 285–290. DOI: https://doi.org/10.3758/BF03213946

Prost, G. 1967. Phonemes of the Chacobo language. *Linguistics, 35*, 61–65. DOI: https://doi.org/10.1515/ling.1967.5.35.61

Ran, Q. 2005. Jiyu Putonghua de Hanyu zuse fuyin shiyan yanjiu [An experimental study of Chinese sibilants with an emphasis on Mandarin] (Doctoral dissertation). Tianjin: Nankai University.

Sereno, J. A., Baum, S. R., Marean, G. C. and Lieberman, P. 1987. Acoustic analyses and perceptual data on anticipatory coarticulation in adults and children. *Journal of the Acoustical Society of America, 81*, 512–519. DOI: https://doi.org/10.1121/1.394917

Steriade, D. 1997. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Phonetics, 3*, 25–146.

Steriade, D. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In: Fujimura, O., Joseph, B. D. and Palek, B. (Eds.), *Proceedings of LP '98: Item Order in Language and Speech*. Prague: Charles University and Karolinum Press.

Steriade, D. 2001. Directional asymmetries in place assimilation: A perceptual account. In: Hume, E. and Johnson, K. (Eds.), *The Role of Speech Perception in Phonology*, 219–250. San Diego: Academic Press.

Steriade, D. 2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In: Hanson, K. and Inkelas, S. (Eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*, 151–180. Cambridge, MA: The MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262083799.003.0007

Stevens, K., Li, Z., Lee, C. and Keyser, S. J. 2004. A note on Mandarin fricatives and enhancement. In: Fant, G., Fujisaki, H., Cao, J. and Xu, Y. (Eds.), *From Traditional Phonology to Modern Speech Processing*, 393–404. Beijing: Foreign Language Teaching and Research Press.

Su, X. 1990. Haizhou fangyan tongyinzihui [The homophonous syllabary of Haizhou dialect]. *Fangyan, 2*, 87–99.

Svantesson, J. 1986. Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics, 14*, 53–70.

Wang, J. Z. 1993. The geometry of segmental features in Beijing Mandarin (Doctoral dissertation). Newark, DE: University of Delaware.

Wang, J. Z. 1999. Beijinghua de yinjie yu yinxi [Syllable and Phonology of Beijing Mandarin]. In: Xu, L. (Ed.) *Gongxing yu Gexing: Hanyu Yuyanxue zhongde Zhengyi* [*Universals and Specificity: Controversial Issues in Chinese Linguistics*]. Beijing: Beijing Yuyan Wenhua Daxue Press (Beijing Language and Culture University Press).

Wang, L. 1985. *Hanyu Yuyin Shi [Historical Phonology of Chinese]*. Beijing: Zhongguo Shehui Kexue Press (China Society Science Publishing House).

Wang, W. S. 1968. The basis of speech. *Project on Linguistic Analysis Reports,* University of California at Berkeley.

Werker, J. F. and Logan, J. S. 1985. Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*, 35–44. DOI: https://doi.org/10.3758/BF03207136

Whalen, D. 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. DOI: https://doi.org/10.1121/1.385348

Whalen, D. 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. DOI: https://doi.org/10.1121/1.401658

Wiese, R. 1997. Underspecification and the description of Chinese vowels. In: Wang, J. and Smith, N. (Eds.), *Studies in Chinese Phonology*, 219–249. Berlin/New York: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110822014.219

Wu, Z. and Lin, M. 1989. *Shiyan Yuyinxue Gaiyao [An Outline of Experimental Phonetics]*. Beijing: Gaodeng Jiaoyu Chubanshe (Higher Education Press).

Yin, S. 1995. *Ha'erbin Fanyang Cidian* yinlun [The Introductory Chapter in *A Dictionary of Harbin Dialect*]. *Fangyan,1*, 17–25.

Zee, E. and Lee, W. 2001. An acoustical analysis of the vowels in Beijing Mandarin. *Proceedings of Eurospeech, 1*, 643–646.

Zee, E. and Lee, W. 2004. The apical sounds in Beijing Mandarin. In: Lu, J. and Wang, J. (Eds.) *Xiandai Yuyinxue yu Yinxixue Yanjiu [Studies on Modern Phonetics and Phonology]*. Tiajin: Tianjin Shehui Kexue Yuan Press (Tianjin Society Science Publishing House).

Zeng, Y. 1993. Xiangtan fangyan tongyinzihui [The homophonous syllabary of Xiangtan dialect]. *Fangyan, 4*, 295–305.

Zhou, C. 1991. Xiamen fangyan tongyinzihui [The homophonous syllabary of Xiamen dialect]. *Fangyan, 2*, 99–118.

Źygis, M. and Padgett, J. 2010. A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics, 38*, 207–226. DOI: https://doi.org/10.1016/j.wocn.2009.10.003

How to cite this article: Li, M. and Zhang, J. 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, pp. 1–27, DOI: https://doi.org/10.5334/labphon.27

Submitted: 26 May 2016 **Accepted:** 04 April 2017 **Published:** 18 July 2017

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS