CrossMark

Revisiting Mandarin 'apical vowels': An articulatory and acoustic study

Sang-Im Lee-Kim
New York University
sangim@nyu.edu

The present study investigates the articulatory and acoustic properties of the unique apical segments in Mandarin Chinese that occur after dental and retroflex sibilants instead of the high front vowel [i]. An ultrasound study showed that the segments are homorganic with the preceding dental and retroflex sibilants. However, an acoustic study showed that they have a periodic waveform and clear formant structures with no inherent frication noise, indicating that they are not 'voiced fricatives'. The results also suggest that the observed F2 pattern can only be explained with an acoustic model of a sonorant consonant, wherein F2 is attributed to the cavity behind the apical constriction. Based on this, it is argued that the segments can be best seen as 'dental approximant [ɹ̪]' and 'retroflex approximant [ɻ]'. A phonological implication of the pattern is also discussed: the co-occurrence restriction with the high front vowel eliminates a potential chance of palatalization of the dental and retroflex sibilants that may lead to neutralization of the place contrast. The tongue front gesture in the following approximants seems to provide an additional cue to the place of the preceding consonants; the low F3 of [ɻ], for example, enhances cues to the place of the preceding retroflex sibilant.

1 Introduction

In Mandarin Chinese, the high front vowel [i] does not appear after the dental and retroflex sibilants. Instead of the sequences [si tsi tsʰi ʂi tʂi tʂʰi], the consonantal gesture continues throughout the following voiced period, creating a sequence of a sibilant followed by a homorganic voiced segment. Despite a few articulatory and acoustic studies, debate about the phonetic and phonological aspects of the segments following these sibilants has not been settled. The controversies can be partly seen in the variety of names given to those segments by different scholars: syllabic fricatives (Wiese 1997, Duanmu 2000), apical vowels (Karlgren 1915–16), fricative vowels (Ladefoged & Maddieson 1996, Ao 1997), and syllabic approximants (Lee & Zee 2003). The current study discusses some problems with the two dominant proposals, namely the syllabic fricative and apical vowel argument, by examining the articulatory and acoustic aspects of these segments. It will be shown that the results are most compatible with the analysis that these segments are corresponding syllabic approximants [ɹ̪] and [ɻ] of the dental and retroflex sibilants, respectively.

1.1 Phonetics and phonology of Mandarin 'apical vowels'

Mandarin Chinese has an unusually rich inventory of sibilant sounds: a three-way contrast in place of articulation for fricatives, and aspirated and unaspirated affricates. This is summarized

Journal of the International Phonetic Association (2014) 44/3
doi:10.1017/S0025100314000267
© International Phonetic Association
https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

in Table 1. Note that this paper retains the traditional retroflex transcriptions [ʂ tʂ tʂʰ (ɻ)] instead of apical post-alveolars [ʃ̺ tʃ̺ tʃ̺ʰ (ɹ̺)] provided by Lee & Zee (2003).[1]

Table 1 Mandarin Chinese sibilant system.

	Dental	Retroflex	Alveolo–palatal
Fricative	s	ʂ	ɕ
Affricate	ts, tsʰ	tʂ, tʂʰ	tɕ, tɕʰ

It is generally assumed that Mandarin has five underlying vowels /i y u ə a/ (Duanmu 2000, 2007). While all the sibilants can freely co-occur with the vowels [a] and [əu] (Duanmu 2007: 326–327; Li 2008: 17),[2] there is an interesting co-occurrence restriction between dental and retroflex sibilants and the high front vowel [i]. The pattern summarized with an example of fricatives in (1) holds true for the affricates.

(1) *Co-occurrence restrictions with the high front vowel*

sa	ʂa	ɕa
səu	ʂəu	ɕəu
***si**	***ʂi**	ɕi

Those two sibilants are instead followed by nuclei which are known to be made with the tongue tip/blade. As shown in Figure 1, X-ray images taken by Zhou & Wu (1963) demonstrate that the tongue tip/blade gesture inherited from the preceding dental and retroflex consonants remains nearly unchanged for the following voiced period. Since the constriction of the tongue tip/blade is more front for the segment after the dentals than after retroflexes, Zhou & Wu (1963) described the former as 'apical front vowel' and the latter as 'apical back vowel'.

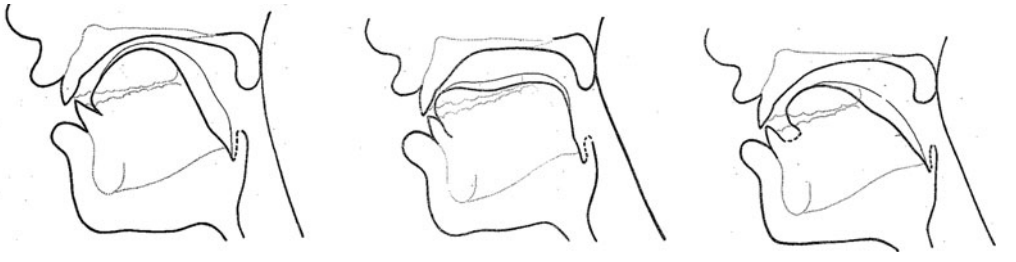


Figure 1 X-ray images of one Bejing speaker for the articulation of the high front vowel [i] (left), 'apical front vowel' (middle), and 'apical back vowel' (right) in Mandarin Chinese. These images are reproduced from Figures 27, 31, and 32, respectively, in Zhou & Wu (1963: 34–35). The solid line indicates the midsagittal images of the tongue from X-rays, and the lighter line indicates the positions of the sides of the tongue from palatograms.

[1] Lee & Zee (2003) appears to follow a narrow definition of retroflexes that involves the tongue tip curled back toward the post-alveolar region as in some Dravidian languages (Ladefoged & Maddieson 1996). However, this paper retains the retroflex transcriptions for the following reasons. First, not all retroflex consonants involve extreme displacement of the tongue tip. Rather, the large sublingual cavity due to the raised tongue tip is more likely to be common among retroflex consonants in other languages (Hamann 2003), and Mandarin retroflexes exhibit this property very clearly. Second, the incompatibility with the high front vowel is one of the common characteristics of the retroflex consonants (Hamann 2003), and this is indeed the case for Mandarin: there is a co-occurrence restriction between retroflexes and the high front vowel.

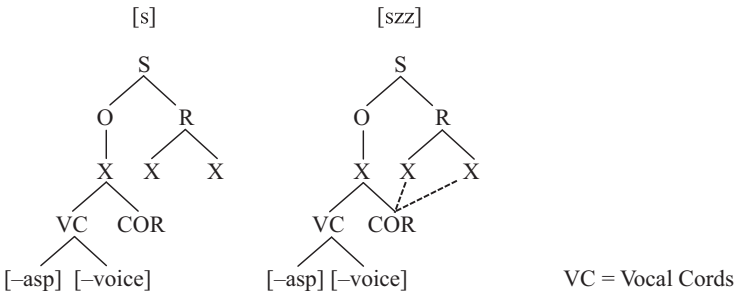
[2] The sequence [ɕa] was traditionally described as alveolo-palatal [ɕ] followed by diphthongs [ia] as in the Chinese orthographic system Pinyin. However, Ladefoged & Maddieson (1996: 150) posited that 'there is nothing other than a normal transition between the initial consonant and the following vowel'. Since there is no independent evidence that they are diphthongs, this paper follows the analysis of Ladefoged & Maddieson (1996).

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

1.2 Previous analyses: Syllabic fricatives vs. apical vowels

Despite a general agreement on the articulation of the apical segments, phonological analyses are extremely diverse and contentious. Abstracting away from the details of each theory, two dominant arguments on the status of these segments emerge: syllabic fricatives vs. apical vowels. The proponents of the syllabic fricatives argued that these segments are voiced counterparts of the corresponding sibilants, transcribing them as [z] and [z̩] (Chao 1934, 1968; Hartman 1944; Pulleyblank 1984; Ramsey 1987; Lin 1989; Wiese 1997; Duanmu 2000, 2007). Since the segments in question function as the nucleus of each syllable, they have long been called 'syllabic fricatives' among those scholars. Acknowledging that it is not phonologically preferable to posit independent phonemes for those voiced fricatives, some scholars have proposed a formal analysis utilizing feature spreading (Wiese 1997; Duanmu 2000, 2007), as shown in (2). Specifically, the coronal feature of the onset consonant is spread into the rhyme which was empty for featural specification.

(2) *Feature spreading analysis for the syllabic fricatives* (Duanmu 2000: 47)



It is worth pointing out that this feature spreading approach departs from the view that these segments are derived from the underlying high front vowel [i]. There are mainly two arguments for not deriving the apical segments from /i/. First, the unique apical gesture of these segments is radically different from the typical vowel articulation which uses the tongue body as a main articulator. Second, it is particularly unnatural to derive voiced fricatives ([z] and [z̩]) from the underlying high front vowel /i/ through featural changes. This process, if possible, would be as arbitrary as a change from an underlying /i/ to [f] or [k], as noted in Wiese (1997).

Now we consider another competing argument in the field: the apical vowel argument. Despite the apicality characteristic of the consonantal articulation, there has been a long tradition of analyzing these segments as vowels particularly in Chinese linguistics (Karlgren 1915–16, R. Cheng 1966, C. Cheng 1973, Howie 1976, Hsueh 1986, Wu 1994, Zee & Lee 2001). The phonetic grounding of this approach includes (i) the presence of clear formant structures, (ii) raised tongue back as for vowel articulation. A strong phonological advantage comes from the consideration of the distribution of these segments; they are in complementary distribution with the high front vowel, and they occur exclusively in these specific environments. Analyses assuming that the segments following dental and retroflex sibilants are vowels differ in their exact implementation. For example, R. Cheng (1966) assumes that the two apical segments are derived from the underlying high front vowel whereas C. Cheng (1973), who noticed the retracted tongue body posture, posits an abstract underlying segment [ɨ], independent from [i].

1.3 Current proposal: 'Apical vowels' as syllabic approximants [ɹ̩] and [ɻ̩]

The present study discusses problems with both approaches and provides evidence for analyzing these segments as approximants. To begin with, the syllabic fricative approach is not quite right considering the absence or at least the tenuous presence of frication noise during the voiced segments. Although having argued for the 'fricative vowel' notation

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

(Ladefoged & Maddieson 1996: 365), Ladefoged himself admits that 'it is certainly true that the fricative noise carries over into the beginning of the vowel. But the greater part of each of these vowels does not have any fricative turbulence', as seen in the UCLA phonetics lab data (http://www.phonetics.ucla.edu/appendix/languages/chinese/chinese.html). As much as it is not accurate phonetically, the syllabic fricative analysis is also not preferable from a phonological perspective. Although a feature spreading analysis can get away with positing independent phonemes for those unique segments, it is a stipulation that the rhymes of [s] and [ʂ] are empty in the first place whereas the rhyme of [ɕ] is filled with the high front vowel. Secondly, it does not address the more fundamental question of why the co-occurrence restriction exists only between dental/retroflex sibilants and the high front vowel.

Although it correctly captures the distributional property, the apical vowel argument creates an obvious conflict between articulation and acoustics of these segments. Based on their relative tongue tip/blade position in the alveolar region, Zhou & Wu (1963) described the dental segment as an 'apical FRONT vowel' and the retroflex segment as an 'apical BACK vowel'. Along the same vein, Wu (1994) transcribed the former as a 'high CENTRAL vowel [ɨ]', and the latter as a 'high BACK vowel [ɯ]'. Both approaches attempt to distinguish the anterior alveolar from the posterior one by borrowing the notion of 'backness' that is used for the vowel system. However, this necessarily makes a wrong prediction about the acoustics of the nuclei. If they are indeed vowels, the front vowel, i.e. the dental segment, should exhibit higher F2 values than the back vowel, i.e. the retroflex segment. Some formant measures (Howie 1976, Zee & Lee 2001), however, demonstrated that the opposite is the case: the retroflex segment has consistently higher F2 than the dental one. This mismatch in the articulatory–acoustic relation presents a puzzle for the apical vowel argument. Second, as will be shown later, the articulation of these segments is homorganic with the preceding sibilants, and therefore the consonantal gestures involved in these segments are radically different from the standard vowel articulation. Consequently, this makes it hard to find appropriate IPA vowel symbols. Although Sinologists traditionally used [ɿ] and [ʅ] symbols, these are now obsolete in the IPA and not informative for non-Chinese linguists. One might use the regular vowel symbols with diacritics considering their relative F2 values, e.g. a dentalized [ɯ̟] and retroflexed [ɨ̠], but the articulation of these segments are sufficiently different from the possible tongue gesture of [ɯ] or [ɨ]. Moreover, the vowel notations [ɯ] or [ɨ] do not capture the homorganicity of the consonants and voiced segments. That is, there is no necessary connection between [s] and [ɯ] or between [ʂ] and [ɨ].

Departing from these two dominant views, this paper argues that the segments in question are corresponding approximants of the preceding sibilants, as shown in (3).

(3) *Phonetic transcription of the apical segments*
[ɹ̩]/s__ (syllabic) dental approximant
[ɻ̩]/ʂ__ (syllabic) retroflex approximant

While the dental approximant is not a rhotic sound, the retroflex approximant is. Because the International Phonetic Alphabet does not have an appropriate dental approximant symbol, the transcriptions below are intended to reflect the homorganicity of these approximants with their corresponding sibilants. Phonologically, these segments are syllabic since they serve as nuclei of a syllable. The syllabic diacritic (e.g. [x̩]), however, is not used for the sake of simplicity.

While Lee & Zee (2003) used the symbol [ɹ̩] for the transcription of both apical segments in the IPA Illustration of Beijing Mandarin, this paper proposes independent IPA symbols for each approximant. To provide evidence for the current proposal, ultrasound imaging is employed, tracking the tongue movement during the articulation of the entire apical syllable for one female Beijing speaker. Ultrasound imaging, when used along with statistical techniques, is particularly useful to examine the relative tongue shape of the segments in question in relation to other segments. While Zhou & Wu's (1963) X-ray images present unanalyzed

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

data (Figure 1), the current paper presents an extensive study of the articulation of these approximants in comparison with the preceding consonants and to other point vowels.

Subsequently, an investigation of the acoustic properties of the apical segments is conducted on the recordings of a number of speakers. The acoustic study focuses mainly on two aspects. First, the presence of the alleged frication during the dental and retroflex sequences is closely examined via visual inspection on the waveform and spectrograms. Despite the simplicity of this technique, this process is important, since the presence of frication noise has served as primary evidence for the syllabic fricative argument. Second, formant measures provide the basis for an explicit acoustic model that explains the articulatory-acoustic mechanism involved in the production of the apical segments. Note that simply taking the vowel-like property as evidence for the apical vowel analysis has created a conflict between articulation and acoustics. A correct acoustic model not only provides a proper acoustic model without any conflict with articulation but it also sheds light on the phonetic nature of these segments as approximants.

Finally, this paper provides a possible phonological explanation for the co-occurrence restriction between the dental and retroflex sibilants and the high front vowel. The advantage of the approximants over the high front vowel is discussed in terms of the preservation of phonological contrast.

2 Experiment I: An ultrasound study

2.1 Participants

One female (WY) graduate student of New York University was recruited for the ultrasound imaging. She was a native speaker of Mandarin Chinese who was born and raised in Beijing in China until she moved to the United States for graduate school. She spoke standard northern Mandarin which retains the distinction between the retroflex and dental sibilants. She reported no history of speech or hearing disorders. The participant was paid for her time.

2.2 Materials

The stimuli for the ultrasound study included eight disyllabic sequences with ten filler words. The target was the first syllable composed of three fricatives [s ʂ ɕ] in three different vowel contexts [i(ɹ̩ ɻ̩) a u]. Since [ɕ] cannot be followed by the vowel [u] in Mandarin Chinese, this vowel was preceded only by [s] and [ʂ]. The tone of the target syllable was fixed with a high level tone (X^{55}).[3] The second non-target syllable started with a sequence [xa] except for one ([saxuaŋ] 'lie'), so the entire sequence was mostly of the form of $\underline{CV}^{55}$.xaC/V (where underline marks the target syllable). Table 2 includes the stimuli used in experiment.

Table 2 Target stimuli used in experiment.

	s̻	ʂ	ɕ
/a/	$sa^{55}.xuaŋ^{214}$	$ʂa^{55}.xai^{53}$	$ɕa^{55}.xan^{214}$
	spill.lie	*murder.impair*	*foolishly.shout*
	'lie'	'murder'	'shout foolishly'
/u/	$su^{55}.xaŋ^{35}$	$ʂu^{55}.xan^{35}$	—
	Su.Hang	*book.letter*	
	'City Su and Hang'	'letters'	
/i/	$s̻ɹ̩^{55}.xao^{35}$	$ʂɻ̩^{55}.xan^{35}$	$ɕi^{55}.xan^{53}$
	tear.good	*wet.cold*	*west.Han*
	'tear well'	'wet and cold'	'Western Han dynasty'

[3] Mandarin Chinese has four lexical tones: high level (X^{55}), rising (X^{35}), low-dipping (X^{214}), and falling tone (X^{51}) (Chen 2000: 364).

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

In order to maximally control the phonetic environments, the target disyllabic words were different in their grammatical units and the internal morphological structures. Despite the non-uniform grammatical status of each sequence, the prosodic parsing was expected to be comparable to one another, since in Chinese any words or phrases can function as a single noun phrase when embedded in a carrier phrase.

2.3 Procedure

Before the recording started, the participant was familiarized with the target and filler words written in Chinese, and she found all the stimuli sequences natural. The participant was first asked to swallow water for the palate image to be extracted later (Epstein & Stone 2005). A list of eight target words and ten fillers were presented five times in random order, and the participant was asked to read the target words in a carrier phrase *wo shuo zhe ge ci* [wo ʂuo __ tʂɤ-kə tsʰɻ̩] 'I say __ this word'.

During the recording session, the participant was seated in a comfortable chair and instructed to find the most comfortable pose for herself. The participant's head was then fit into a moldable head stabilizer on the wall and the head was further stabilized with Velcro (Davidson 2006). This method, when used along with the ultrasound transducer fixed with a Bogen–Manfrotto Magic Arm, ensures that the same slice of tongue is imaged (Davidson & De Decker 2005). The transducer was placed under the speaker's chin and adjusted until clear midsagittal images were obtained.

The ultrasound images were recorded using a Sonosite Titan portable ultrasound in a sound-attenuated booth at New York University. A 5–8 MHz Sonosite C-11 transducer with a 90º field of view was set at a depth of 8.2 cm. The audio was collected using an Audio Technica AT-813 microphone, and was synchronized in an *avi* movie file with the video signal from the ultrasound machine using a Canopus ADVC-1394 capture card and Adobe Premiere. The wave files were then extracted from the movie file, and submitted to PRAAT 5.2.19 (Boersma & Weenink 2011) in order to segment the target sequences. Following previous studies (e.g. Iskarous, Shadle & Proctor 2011), four acoustic events were identified using PRAAT: onset and offset of both frication and the following vocalic period. The boundaries of fricatives were identified at the beginning and ending of aperiodic noise in the waveform. The presence of F1 and F2 were used to approximate the onset and offset of the vocalic segments. A significantly long lag between the offset of the fricatives and the onset of the following vowels [a] and [u] was often observed, as shown in Figure 2.[4]

Among the frames corresponding to the voiced period, the middle frame was selected for further analysis, assuming that this frame represented the most genuine tongue shape for vocalic segments. When there was an even number of frames, the first frame after the middle was selected. Those selected frames were then submitted to EdgeTrak (Li, Kambhamettu & Stone 2005) which automated the tracking of tongue contours by extracting x–y coordinates of the target region from the lower edge of the tongue. One hundred individual values were extracted for the statistical analysis in the next step. Screenshots of the ultrasound image and tongue curve tracking in EdgeTrak are shown in Figures 3a and 3b, respectively.

To assess the tongue shape of the apical segments relative to other vowels, a smoothing spline (SS) ANOVA was employed. SS ANOVA is a statistical procedure for investigating similarities and differences of curve shapes (Gu 2002, Davidson 2006, Wang 2011). Using this method, smoothing splines which best fit the five repetitions for the stimuli were obtained. Sample smoothing splines of two vocalic segments in [sɻ̩] vs. [ɕi] are presented in Figure 4b along with the raw data points of five repetitions in Figure 4a.

[4] The ultrasound machine with the frame rate of approximately 30 frames/s is expected to take about 11 frames between the time point 2.377 s and 2.723 s, but the Matlab script automatically rounds down the start and rounds up the end of the time points of acoustic events, so it may generate one or two more frames than the predicted number of frames.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

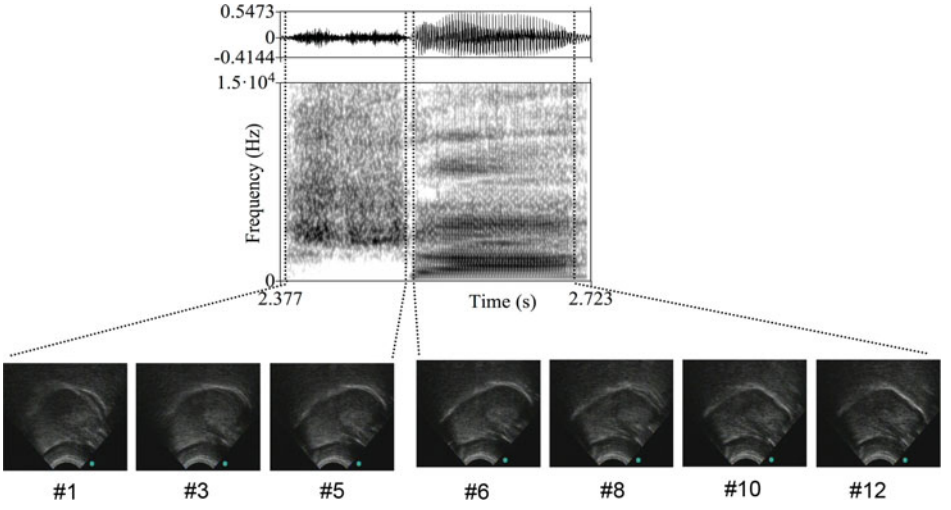


Figure 2 (Colour online) A sample waveform and spectrogram of [ʂa]. Five ultrasound frames (Frames 1–5) were obtained during the acoustic consonant period, and seven frames (Frames 6–12) were obtained during the acoustic vocalic period. The frame rate acquired by the ultrasound machine is approximately 30 frames/s. The tongue root is on the left, and the tongue front is on the right.

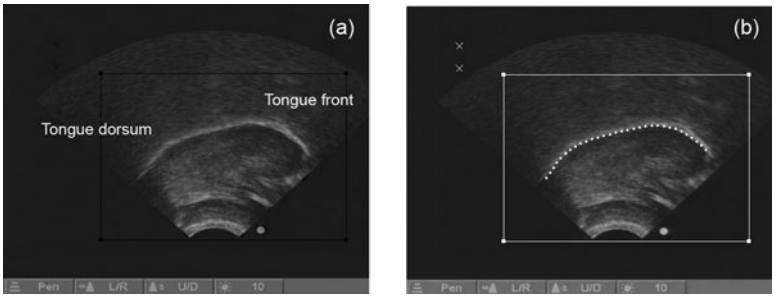


Figure 3 (a) A screenshot of the ultrasound image of the 3rd frame among five frames extracted during the articulation of the retroflex approximant [ɻ] following the retroflex [ʂ]. It illustrates the tongue image when the vocalic articulation reached its maximal constriction. (b) A screenshot of EdgeTrak. The grey dots are the edge of the tongue created by the automated edge-tracking algorithm. The tongue tip is on the right and the tongue root is on the left.

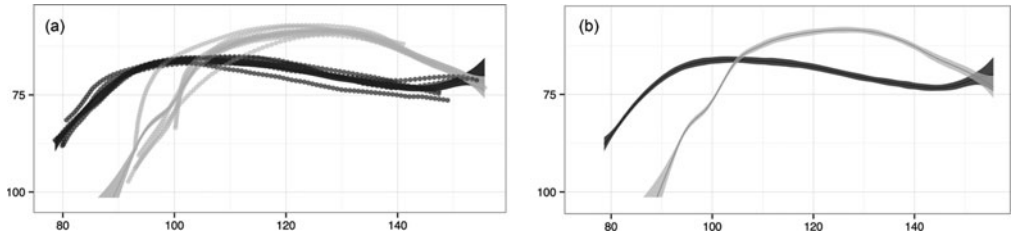


Figure 4 (a) Raw data points from five repetitions and the smoothing spline estimate (solid lines) for comparison of the tongue shape for [sɻ̈] (in black) and [ɕi] (in grey). (b) Smoothing spline estimate and 95% Bayesian confidence interval. As in the ultrasound images, the tongue tip is on the right and the tongue root is on the left. The axes are in millimeters, corresponding to the boxed-in region shown in Figure 3b above.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

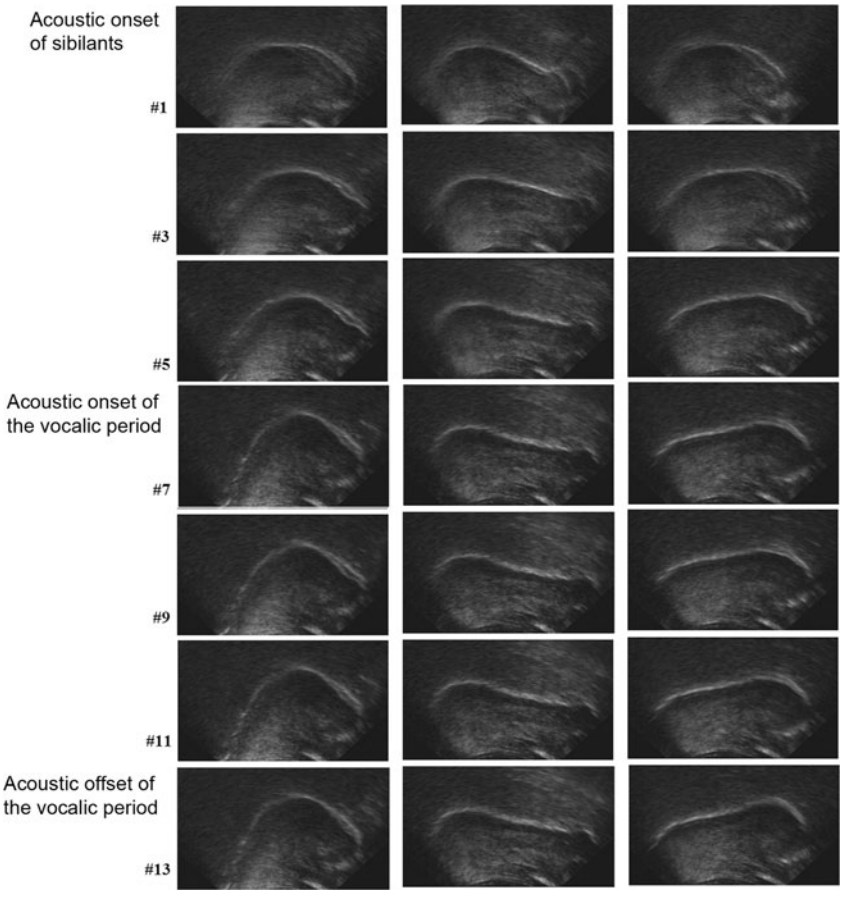


Figure 5 Ultrasound images of [ɕi], [sɹ̟̈] and [ʂɻ̊] of the female speaker WY. The tongue tip is on the right and tongue body is on the left. Frame 1 is the acoustic onset of each consonant, and Frame 7 is approximately the acoustic onset of each vocalic period. Frame 13 is the last frame of each syllable.

2.4 Results

Figure 5 illustrates how the tongue shape proceeds over time during the acoustic realization of each syllable [ɕi], [sɹ̟̈] and [ʂɻ̊]. The dental and retroflex sequences lasted approximately the same amount of time, and slightly less for the palatal sequence; between five and seven frames were obtained during the consonantal articulation, and five or six frames were obtained during the vocalic articulation.

The images on the leftmost panel present the articulation of the palatal sibilant [ɕ] and the high front vowel [i]. The fronting of the tongue body is evident in the middle of the consonantal articulation (Frames 3–5), and it reaches maximal constriction toward the end of the acoustic [ɕ] (Frames 5–7). As the acoustic vocalic period begins, the tongue root continues to move forward and the highest point of the tongue becomes higher (Frames 9–11). However, the basic tongue shape stays nearly the same until the end of the acoustic [i] due to the homorganic property of the consonant and the vowel.

The two syllables containing apical segments proceed very differently. The main articulator of the dental [s], the tongue tip, is approaching its target at around Frame 5 indicated by the little bump in the tongue front. At the same time, the tongue body and root retract and stay far back. More interestingly, both the tongue tip raising and tongue back

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

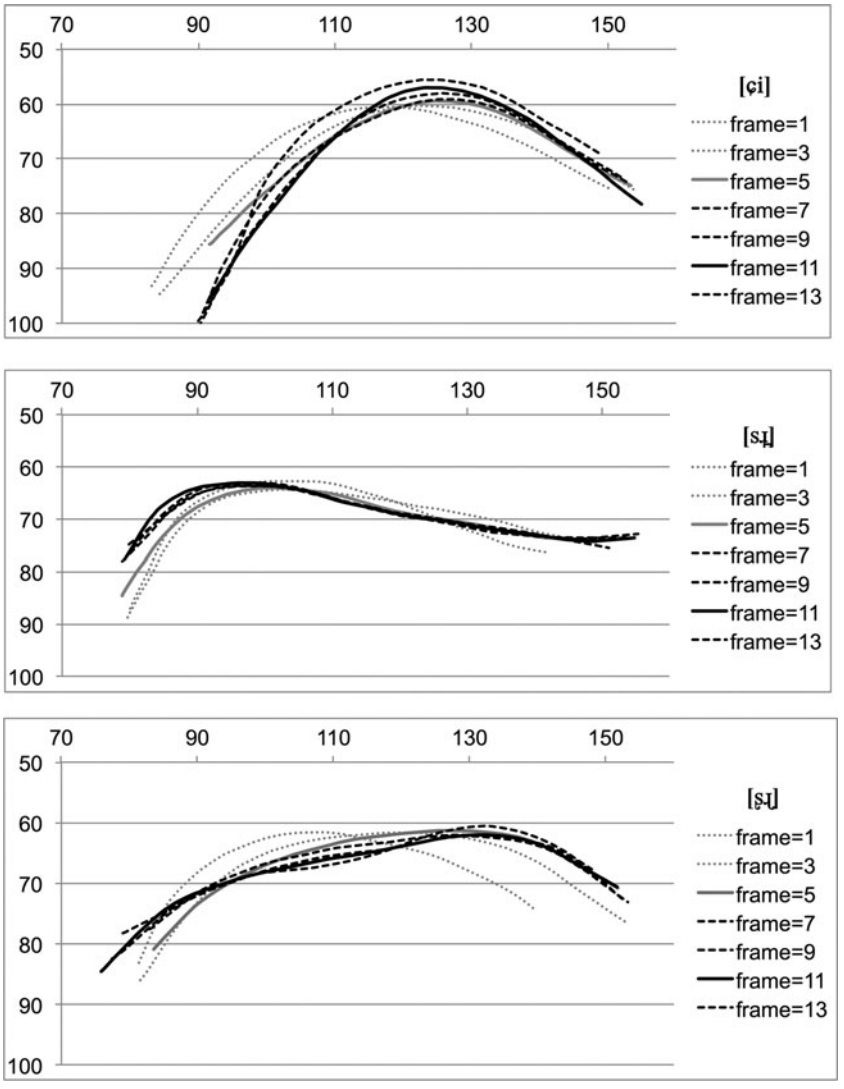


Figure 6 Time course of the articulation of [ɕi], [sɻ] and [ʂɻ] based on the ultrasound images in Figure 5 above. The consonantal gestures are represented in grey and the vocalic gestures are represented in black.

retraction are maintained throughout the entire syllable (Frames 7–13). Likewise, the tongue back is not fronted for the articulation of the retroflex and the following segment, but rather it retracts and stays far back. In accordance with the previous literature (Ladefoged & Wu 1984, Lee & Zee 2003), the retroflex consonant of this speaker is not produced with the tongue tip curled back, but a larger chunk of the tongue blade makes a constriction far behind the dental constriction (Frames 5–7). As with the other apical segment, [ɻ], the tongue shape of the consonant does not change significantly during the articulation of [ɻ]. Rather, the raised tongue blade and the retracted tongue back are maintained throughout the entire syllable (Frames 9–13). Interestingly, raising of the tongue blade is more evident toward the end of the syllable, so the final frame (Frame 13) shows the most raised tongue front gesture.

Figure 6 shows the time course of the change in articulation throughout the entire syllable. The edge of the tongue shape in Figure 5 was extracted using EdgeTrack, and was placed in

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

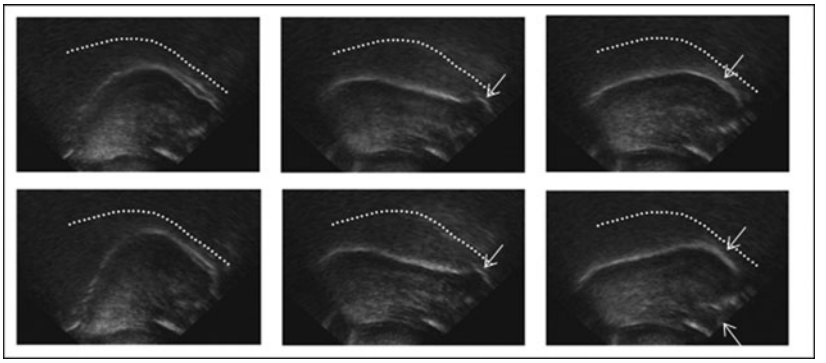


Figure 7 The tongue shapes of Frame 5 (top row) and Frame 11 (bottom row) are reproduced with the palate trace loaded (shown by the dotted line) on top of each figure. The figures in the top row represent the tongue shape of consonants [ɕ], [s] and [ʂ], respectively, whereas those in the bottom row represent the tongue shape of the corresponding high front vowel [i] and approximants [ɹ̟] and [ɻ]. The dental and retroflex constrictions are maintained in the voiced period as indicated by an arrow. A sublingual cavity for [ʂ] and [ɻ] is indicated by a dotted arrow.

one plot. Although the very fine details such as the little tongue tip gesture found in [sɹ̟] were omitted, these figures enable a thorough examination of the relative tongue position between consonants and the following vocalic gestures.

Overall, it is observed that the tongue gestures do not change significantly once it reaches the consonantal target (represented in solid grey), confirming the homorganic property between consonantal and vocalic gestures. However, it is noteworthy that there are still some changes between the consonantal (in solid grey) and vocalic (in solid black) target gestures for all three sequences. The tongue root continues to move forward and the tongue body moves upward for [i], whereas the tongue root continues to retract for the dental and retroflex segments. In addition, the tongue body continues to lower slightly for the retroflex segment, presumably to make the tongue tip/blade more raised. It seems that these continuing gestures are to achieve their maximal articulatory targets of each homorganic sequence due to a relatively long timespan allotted for each place of articulation.

Figure 7 reproduces the tongue shapes at Frame 5 and Frame 11 in Figure 5, which represent the typical consonant and the following apical segments of each syllable. The palate image taken separately was loaded on top of each figure to indicate better the relative position of the tongue.

Again, despite some differences, for each sibilant, the tongue configuration appears to be nearly the same for the sibilant and the following vocalic segments, confirming that the apical segments are homorganic with the preceding consonants. Specifically, the tongue tip/blade gesture indicated by an arrow is prominent for the apical segments, and the raised tongue blade for the retroflex becomes even more evident. In contrast with the general description of the retroflex (Hamann 2003), the tongue middle of the Mandarin retroflex is not convex, but it is rather flat. However, it shares a common property with retroflexes in other languages as it has a relatively large sublingual cavity indicated by a dotted arrow.

Figure 8 presents the smoothing splines based on five repetitions of the two apical segments, [ɹ̟] and [ɻ] (dark lines) with 95% Bayesian confidence intervals marked by the grey shading that is superimposed on the smoothing spline. Three point vowels [i], [a] and [u] (light lines) are also plotted to illustrate better the relative tongue position of the two approximants in the vocal tract.[5]

[5] Specifically, the smoothing spline of [i] was fit using five repetitions of [i] from the sequence [ɕi]. The smoothing spline of [a] incorporates six repetitions, two of which were drawn from [sa], [ʂa] and [ɕa],

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

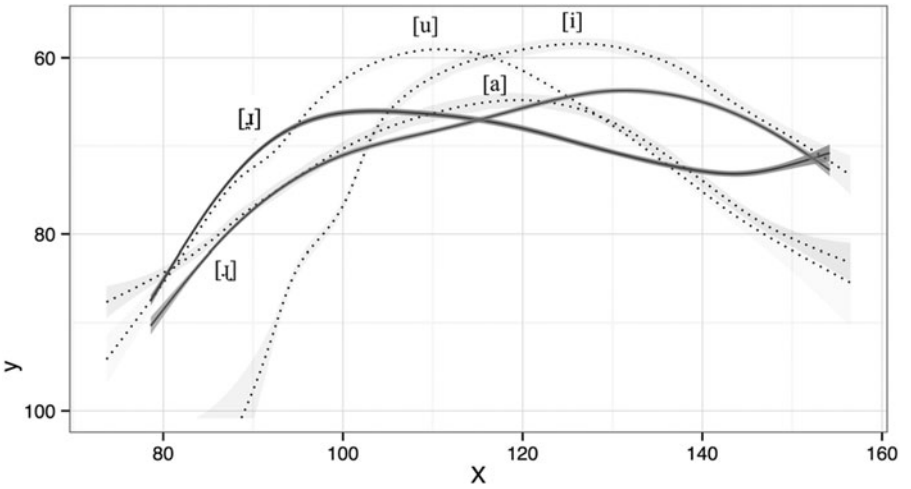


Figure 8 Smoothing spline estimates and 95% Bayesian confidence intervals for comparison of the mean curves of three point vowels [a], [i] and [u] (in light grey) and two approximants [ɹ̩] and [ɻ̩] (in dark grey). As in the ultrasound images, the tongue tip is on the right and the tongue root is on the left.

While the highest points of the vowels [a], [i] and [u] form a typical triangle shape, the approximants form very different configurations. Specifically, the dental segment [ɹ̩] is as low as the vowel [a] in terms of the tongue body,[6] but it is considerably more retracted. It might be that retraction of the tongue back for the dental segment makes a narrow constriction against the upper pharynx, although in the absence of information about the pharyngeal wall, this is not conclusive. In comparison, the tongue body of the retroflex segment [ɻ̩] is less prominent; it is lower and less retracted than that of the dental segment.

2.5 Articulatory findings and acoustic implications

The current ultrasound study is consistent with the previous X-ray images (Zhou & Wu 1963) in that the dental and retroflex segments have tongue configurations similar to the preceding sibilants. However, it is important to note that the tongue body postures of the two apical segments seen in Figure 8 are inherent to the preceding sibilants; the dental and retroflex sibilants have retracted tongue body with the relative degree of raising (see Figure 7 above). Crucially, this argues against the apical vowel analysis wherein the raised tongue body is taken to be evidence for the vowel property of these segments (C. Cheng 1973). In fact, there is a discussion in the literature about the association between dental consonants and a retracted tongue body. For example, in Malayalam, F2 in the surrounding vowels is significantly lower for dentals than for plain alveolars, implying a more retracted tongue body for the dentals (Stevens, Keyser & Kawasaki 1986). Stevens et al. conjecture that the dental constriction, which requires a flat tongue front, can be achieved more easily when the tongue back is retracted.

The association between retroflexes and a retracted tongue back is also well documented (Spencer 1984, Hamann 2003). The articulatory basis of this correlation is that the tongue

respectively, while that of [u] includes data from six repetitions, three of which were drawn from [su] and [ʂu], respectively.

[6] Note that although these secondary constrictions made by the tongue body for the approximants may be more open than vowels, the main constriction made by the tongue tip/blade is significantly narrow to be qualified as an approximant.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

body should be stretched and pulled backward in order to displace the tongue front upward and back toward the post-alveolar region. Spencer (1984) thus proposed the phonological feature [+back] along with other features to properly specify the retroflex consonant. The results of the current study do not refute this proposal, but in this data, retraction of the tongue body is in fact more evident for the dental consonant in Mandarin.

The tongue configuration in the back is of trivial acoustic importance for the sibilant consonants, considering the noise source mechanism whereby the spectral properties of the turbulence noise are mainly shaped by the front cavity (Stevens 1998). However, it can be influential when the apical constriction is open and the back cavity starts to create an acoustic signal as a legitimate resonator. Alwan, Narayanan & Haker (1997) proposed a possible three-cavity acoustic model for the American English [ɹ] based on the fact that it often has a pharyngeal constriction. In this case, the F2 can be attributed to the Helmholtz resonator formed by a narrow pharyngeal constriction and the cavity behind it. Given the prominent tongue back gesture in the Mandarin dental approximant, it might be worth considering this possibility as well as the simple two-cavity model. In the following section, the general acoustic properties of the apical segments are investigated, and the results are discussed in relation to the articulatory findings.

3 Experiment II: An acoustic study

The acoustic study focused on a few aspects of the apical segments. Given the persistent assumption of frication noise during the voiced period, the presence or absence of frication was examined by close visual inspection on the waveform and spectrogram. The formant values were measured and the results were interpreted under an acoustic model of a sonorant consonant that accounts for both articulatory and acoustic aspects of the apical segments.

3.1 Participants

Three female and three male speakers were recruited for the acoustic recordings. The female speaker who participated in the ultrasound imaging was one of the female speakers of this acoustic study. Participants were college or graduate students of the Chinese University of Hong Kong and New York University. They were all native speakers of Mandarin Chinese who were born and raised in mainland China until they started college or graduate school in Hong Kong or the United States. All the female and two of the male speakers were from Beijing in China, and one male speaker was from Harbin, northeast China. None of them reported a history of speech or hearing disorders. The participants were paid for their time.

3.2 Procedure

The stimuli for the acoustic recording were the same as in the ultrasound study (see Table 2 above). The recording procedure also remained mostly the same, except that a list of target words and fillers was presented 10 times in random order, and the subjects were asked to read the target words in a carrier phrase *wo shuo* __ *zhe ge ci* [wo ʂuo __ tʂɤ-kə $ts^{h}ɻ$] 'I say __ this word'. The acoustic recordings were made using an Audio Technica AT 2010 cardioid condenser vocal microphone with a sampling rate of 44.1 kHz in a sound-attenuated booth at the Chinese University of Hong Kong and New York University. The speech signal was submitted to PRAAT 5.2.19 (Boersma & Weenink 2011) and segmented using the same criteria as in the ultrasound study. The first three formant values of the following vocalic period were measured by creating a 25 ms PRAAT-default Gaussian window centered at the midpoint of the vocalic intervals. The maximum frequency for formant calculation was set to 5 kHz for male speakers and 5.5 kHz for female speakers.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

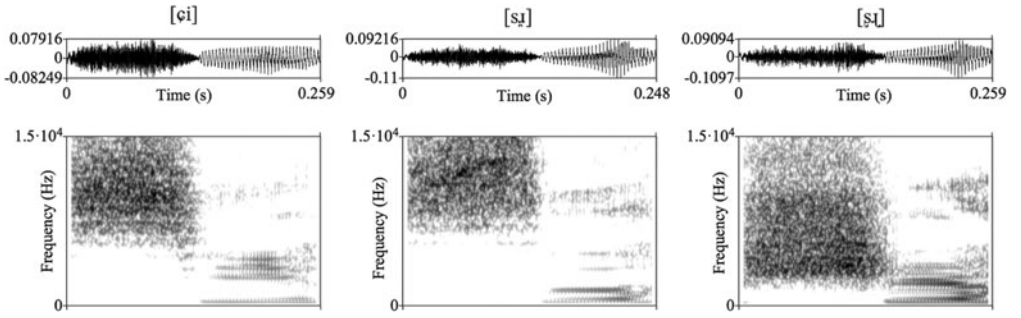


Figure 9 Spectrograms of [ɕi], [sɹ̩] and [ʂɻ̩] of a female speaker (RJ). There is no frication noise in the following periodic segments for all three syllables.

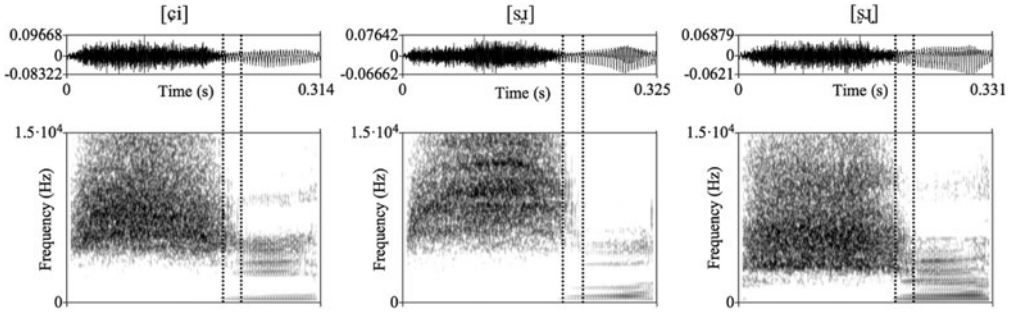


Figure 10 Spectrograms of [ɕi], [sɹ̩] and [ʂɻ̩] of a female speaker (WY). There is a short period of frication noise indicated with dotted vertical lines in the following periodic segments for all three sequences. Frication in the vocalic period is represented with dotted vertical lines.

3.3 Acoustic results

3.3.1 Sample illustration of spectrograms: Investigation of frication noise

Since it has been controversial whether or not the apical segments contain frication noise, this paper closely examines the presence or absence of frication by visual inspection of the waveforms and spectrograms of all speakers.

Figures 9–11 show representative spectrograms of the relevant syllables produced by three different speakers. The upper limit of the spectrogram was set to 15 kHz to show frication noise at higher frequencies. To begin with, Figure 9 illustrates the majority pattern demonstrated by three speakers (RJ, PY and HZ). In the speech signal of these speakers, the apical segments behaved similarly to other vowels, presenting a periodic waveform and a nearly clear separation of the frication noise from the following vocalic period.

By contrast, in two female speakers' (WY, ZD) speech, a small portion at the beginning of the vocalic period (around 20 ms, indicated with dotted vertical lines in Figures 10–11) was consistently affected by frication noise, so that formant frequencies higher than the third were not visible. However, this was not only limited to the apical segments, but also holds true for the sequence of the alveolo-palatal sibilant and the high front vowel (see the leftmost part of Figure 10). Unlike the low vowel [a], where there is often a significant gap between the sibilant and the following vowel (see the spectrogram in Figure 2 above), the high vowel tends to show some frication carried over from the preceding consonant, presumably due to its homorganic property. Thus, frication noise of 20 ms observed in apical segments can be considered to be a mere gestural overlap with the preceding homorganic sibilants.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

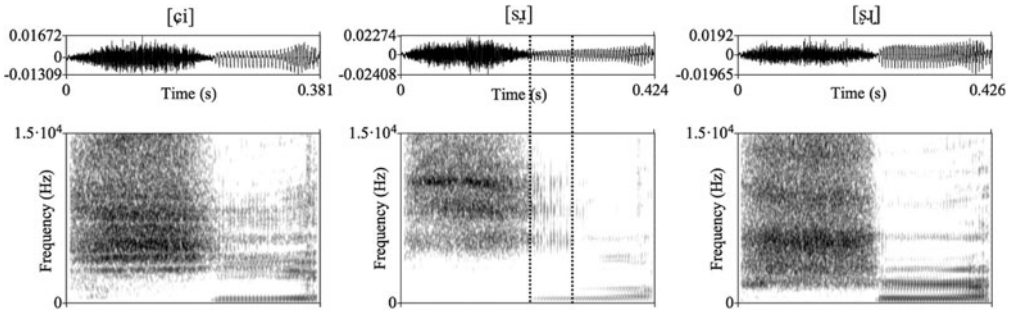


Figure 11 Spectrograms of [ɕi], [sɹ̈] and [ʂɻ] of a male speaker (HW). There is some frication noise into the following vocalic segment for [sɹ̈] (indicated with dotted vertical lines).

Finally, one male speaker, HW, showed a unique pattern with respect to frication noise. The vocalic segment following the dental sibilant in the middle part of Figure 11 was fricated up to one-third or one-half of the vocalic period (indicated with dotted vertical lines) depending on individual tokens. In contrast, the vocalic segment after the retroflex sibilant was never fricated, as seen in Figure 11 (rightmost part). The vowel after the alveolo-palatal sibilant of this speaker, however, was a bit noisy. However, the pattern is distinct from the frication seen in Figure 11 (middle): the noise in the leftmost part of Figure 11 seems to be attributed to the energy in higher frequencies being more intense, whereas the noise in the middle part of this figure is an extension of the preceding frication noise. This suggests that frication, if any, is not limited to the apical segments, but can occur as a result of variable phonetic implementation of a vowel that has relatively narrow constriction with the palate.

The presence of frication noise in the vocalic period for all six speakers is summarized in Table 3. Overall, frication noise of the preceding fricatives never spread through the 'entire' periodic segments as it would if they were truly voiced fricatives. Rather, frication noise, if any, was found at the very beginning of the following segments only for some speakers for a very short period. Moreover, this carryover frication noise was also regularly found for the plain alveolo-palatal sequence for those speakers, suggesting that frication noise is likely to be a mere gestural overlap between sibilants and the homorganic apical segments consistent within a speaker. This finding provides evidence against syllabic fricative analyses, which have overestimated the significance of occasional frication noise.

Table 3 The presence of the frication noise.

Subject (Gender)	[ɕi]	[sɹ̈]	[ʂɻ]	Note
HW (M)	none	half	none	
PY (M)	none	none	none	A little for [ʂɻ] only for a few repetitions.
HZ (M)	none	none	none	
ZD (F)	a little	a little	a little	Around 20 ms of frication for all three syllables.
RJ (F)	none	none	none	
WY (F)	a little	a little	a little	Around 20 ms of frication for all three syllables.

3.3.2 Formant measures

Figure 12 illustrates the waveform and spectrogram of a male speaker, PY. Again, the vocalic segments show a regular periodic waveform and clear formant structures as in the plain high

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

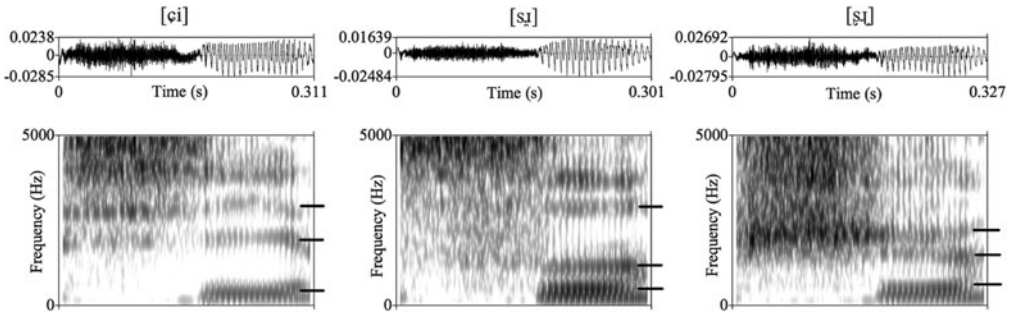


Figure 12 Spectrograms and formants of [ɕi], [sɹ̟] and [ʂɻ̟] of a male speaker (PY, first repetition). The first three formants are indicated with bars on each figure. The second formant values show a hierarchical order among three vowels: the high front vowel is the highest (on the leftmost), the approximant after retroflex is intermediate (on the rightmost), followed by the approximant after the dental sibilant (in the middle). The third formant of the retroflex segment is considerably lowered, nearly merging with the second formant.

front vowel [i]. Two interesting aspects of the apical segments stand out most clearly. First, the second formant of the apical segments is much lower than that of the high front vowel. Moreover, F2 of the dental segment is still lower than that of the retroflex. Second, F3 of the retroflex segment is quite low, nearly merging with the second formant.

The pattern was confirmed with several speakers. Table 4 presents the average formant values of female and male speakers, respectively. One-way analysis of variance verified its statistical significance. The F2 values averaged over female and male speakers were significantly different depending on the type of the following segment (i, ɻ̟, ɹ̟) ($F(2,174) = 434.9$, $p < .001$); Student–Newman–Keuls post-hoc tests indicated this was due to the following pattern: F2(i) = 2362 Hz > F2(ɻ̟) = 1694 Hz > F2(ɹ̟) = 1295 Hz. The F3 values were also significantly different (i, ɻ̟, ɹ̟) ($F(2,174) = 71.4$, $p < .001$); Student–Newman–Keuls post-hoc tests indicated this was due to the following pattern: F3(i) = 3216 Hz, F3(ɹ̟) = 3105 Hz > F3(ɻ̟) = 2476 Hz.

The results of the formant measures are consistent with the previous studies; the dental segment is characterized by low F2 and high F3, whereas the retroflex segment is characterized by mid F2 and low F3 (Howie 1976, Zee & Lee 2001). The striking difference in F3 values seems to lie in the distinct rhotic property: the retroflex approximant [ɻ̟] is a rhotic sound, whereas the dental approximant [ɹ̟] is not. Impressionistically, this is consistent with the percept of two sounds: the former gives an *r*-colored percept, whereas the latter does not. Although the symbol [ɹ] is generally used to represent a rhotic sound, this paper distinguishes between a dental approximant notated by [ɹ̟] and a retroflex approximant [ɻ̟]

Table 4 Mean formant values of three female and three male speakers. For both groups, the dental segment shows the lowest F2 value, the retroflex shows the intermediate and the high front vowel the highest. F3 of the retroflex segment is much lower than other segments.

	Female			Male		
(Hz)	(ɕ)i	(s)ɹ̟	(ʂ)ɻ̟	(ɕ)i	(s)ɹ̟	(ʂ)ɻ̟
F3	3370	3272	2627	3057	2932	2319
F2	2565	1396	1824	2152	1190	1561
F1	349	515	506	314	416	406

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

on the basis of the articulatory-acoustic differences of these sounds. That is, the low F3, a general acoustic characteristic of rhotic sounds, results from a large front cavity (Narayanan, Byrd & Kaun 1999). The rhotic property of the retroflex [ɻ̺], i.e. the low F3, can thus be attributed to the large sublingual cavity formed by raising of the tongue blade, whereas the absence of this cavity gives rise to the non-rhotic property, i.e. the high F3, for the dental approximant [ɹ̪].

3.4 Acoustic models for apical segments

An examination of the spectrogram and waveform has shown that the alleged frication noise is illusory, presenting evidence against the syllabic fricative analysis. There is no frication noise inherent in the following voiced period, and occasional frication noise is a mere gestural overlap that is often found between consonants and homorganic voiced segments. These vowel-like acoustic properties lead to a conclusion that the narrow constriction for the sibilant is slightly released and the source mechanism has changed from a turbulent jet to a sonorant sound. This implies that the back cavity starts to resonate, generating a formant structure of its own. Considering an interaction between different cavities, it is worth investigating the source of each acoustic signal in relation to the specific vocal tract configuration.

To begin with, Stevens (1998: 535–539) provides an acoustic model that explains the association between the low F3 and the retroflex gesture. The space under the tongue blade, i.e. sublingual cavity, created by the raising of the tongue front introduces a pair of acoustic pole and zero. The pole makes a high-amplitude resonance frequency, while the zero dampens the amplitude of the higher frequency energy. These events are interleaved with the pairs of resonant frequencies generated by the back cavity. Due to the relative length difference, the back cavity typically resonates at a slightly lower frequency than the acoustic pole. As a result, the whole acoustic event gives rise to the characteristic formant structure of retroflexes: low F3 near F2 and weak higher formants. In fact, this is precisely what we observe for the Mandarin retroflex approximant [ɻ̺]. It has the raised tongue blade accompanied by a large sublingual cavity, and this introduces an acoustic pole that finally appears as a low third formant in the spectrogram.

Without the addition of a sublingual cavity, the front cavity of a dental segment which is inherently short is unlikely to generate any visible formant values. Instead, the long back cavity behind the dental constriction is predicted to create a low F2, which is consistent with the current acoustic study. Moreover, the observed relative F2 values between the dental and retroflex segments receive a straightforward explanation. That is, the dental segment has a longer back cavity than the retroflex, resulting in a lower F2 for the former than the latter.

It is worth noting though that the F2 value of the dental approximant might be attributed to the Helmholtz resonator formed between the pharyngeal constriction and the cavity behind it. Recall that [ɹ̪] has a prominent tongue back gesture making a pharyngeal constriction (Figure 8 above). Alwan et al. (1997) proposed two possibilities for the origin of the second formant in the production of American English [ɹ]. For one thing, when the pharyngeal constriction is not prominent and the entire back cavity forms a relatively uniform tube, the second formant arises from the resonance of the back cavity behind the alveolar constriction. However, in the case of a narrow pharyngeal constriction, the second formant is projected from the Helmholtz resonator formed between the pharyngeal constriction and the cavity behind it. It is not clear whether the pharyngeal constriction of the Mandarin dental approximant is narrow enough to form a Helmholtz resonator. Assuming there is relatively wide inter-speaker variation, it might be more accurate to say that either case is possible, but is subject to a specific articulation of a speaker. In any case, F2 is consistently attributable to the back cavity.

Having established the relationship between the vocal tract configuration and the acoustics, it becomes clear that the vowel analysis is less tempting. Under the general acoustic model for vowels, F2 is attributed to the front cavity due to the articulatory characteristics of vowels: constriction made by the tongue body creates a long front cavity. When the main articulator

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

is tongue tip/blade as in apical segments, however, F2 is necessarily associated with the back cavity due to the inherent lack of a long front cavity. Whereas the vowel analysis is incompatible with the general acoustic model for vowels, the current approximant analysis correctly accounts for the observed formant pattern of both apical segments in relation to their articulatory properties.

4. General discussion

The present study has shown that the apical segments that occur after the dental and retroflex sibilants in Mandarin Chinese are a 'dental approximant [ɹ̪]' and a 'retroflex approximant [ɻ]', respectively. Evidence comes from both articulatory and acoustic properties of these segments. Ultrasound images showed that these segments are indeed homorganic with the preceding consonants. However, the acoustic study revealed that the alleged frication noise is not evident, suggesting that these segments are not 'voiced fricatives'. Despite the vowel-like formant structure, it has been shown that the observed F2 pattern can be best explained with an acoustic model of a sonorant consonant wherein the F2 is attributed to the back cavity behind the apical constriction. This analysis correctly predicts the relative F2 values between two apical segments; the dental segment has a longer back cavity than the retroflex, resulting in the lower F2 values for the former than the latter. Based on the articulatory and acoustic findings that refute both the voiced fricative and the apical vowel analyses, it is proposed that they can be best seen as a dental approximant [ɹ̪] and a retroflex approximant [ɻ].

However, acknowledging the complementary distribution of the approximants with the high front vowel remains a puzzle,[7] since they are quite different from each other phonetically and phonologically. The next section thus continues an in-depth discussion about the challenge that the current data brings up. To begin with, we consider first why the high front vowel [i] is prohibited specifically after the dental and retroflex sibilants in Mandarin, while other vowels are allowed. An account based on perceptual reasons is considered (Flemming 1995, Steriade 1999). The high front vowel, which often triggers palatalization of preceding consonants, may obscure the contrastive place distinctions by making the frication noise of sibilants less distinct. Once it is understood why the high front vowel [i] is not possible after all three sibilants, the second part of the puzzle is why Mandarin has approximants in place of [i], given that there are other possible scenarios that would maintain a contrast in sibilant place before [i]. The advantage of the approximants is discussed in terms of contrast maximization: they provide additional cues to the place contrast between sibilants.

4.1 Neutralization of the phonological contrast by the high front vowel

The logic behind the Mandarin pattern becomes clear if we consider a system where all three sibilants are followed by the high front vowel: [si ʂi ɕi]. A potential problem with this system lies in palatalization of the anterior coronals triggered by [i]. Palatalization, the superimposition of an [i]-like gesture on a primary gesture is known to interfere with non-labial consonants (Ladefoged & Maddieson 1996: 363–365). Labials are unaffected by a following high vowel since the movement of the tongue body toward the palate can be

[7] It should be noted that the retroflex approximant [ɻ] is in fact an independent phoneme in Mandarin Chinese according to the current system, where the post-alveolar sibilants are transcribed as retroflexes (see Table 1). This is because the production of the retroflex approximant [ɻ] results from a slight release of the narrow constriction of the retroflex sibilant [ʂ]. There might be some phonetic difference between [ɻ] in the onset, as in [ɻən] 'person', and [ɻ] in the nucleus, as in the current case. The phonetic difference, if any, can be attributed to the difference in the phonological status of two segments: the latter is a syllabic approximant, while the former is not.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

made independently from the labial gesture. In contrast, the articulation of coronal or dorsal consonants is necessarily affected by palatalization due to the interdependency of the tongue. For the coronal consonants, for instance, it is hard to make a constriction with the tongue front around the alveolar region while the tongue middle moves forward. Therefore, palatalization often results in changes in the primary articulation of coronals.

In Korean, for example, a palato-alveolar sibilant [ʃ] occurs as an allophone of the alveolar sibilant [s] before the high front vowel [i] (Martin 1992). For languages like Korean, where there is only one coronal sibilant, the change in place by palatalization does not affect the phonological system. However, palatalization can cause phonological neutralization in languages where consonantal places are contrastive. In Acoma, retroflex consonants change to alveolo-palatals before front vowels, reducing the three-way contrast [s ʂ ɕ] to a two-way contrast [s ɕ] (Miller 1985). In Telugu, the alveolar [s] is often realized as [ɕ] before front vowels in casual speech, reducing the three-way contrast [s ʂ ɕ] to a two-way contrast [ʂ ɕ] (Lisker 1963).

The cases of Acoma and Telugu show that languages may neutralize a sibilant place contrast when the cues to the contrast are weak. Specifically, palatalization of the coronal consonants obscures the contrast in sibilant places by reducing the distinction in frication noise. However, Mandarin presents an interesting case that languages can choose to maintain the place contrast by eliminating the environment that may jeopardize the phonological contrast: the high front vowel is replaced with approximants. As the current articulatory study shows, they are homorganic approximants of the corresponding sibilants, and the tongue body of the two segments is particularly retracted. In the absence of the extreme forward movement of the tongue body of a following [i], the three sibilants [s ʂ ɕ] can maintain their primary places when followed by the approximants, creating distinct acoustic signals in frication noise.

The current discussion can thus be summarized as follows: the high front vowel is avoided to eliminate a potential chance of palatalization of the preceding coronals, which may obscure distinct cues in frication noise. However, employing approximants is not the only way to achieve this goal. An alternative to this would be to simply employ some back vowels such as [ɨ] or [ɯ], resulting in a system of [sɨ ʂɨ ɕi], for example. Being back vowels, they do not interfere with the coronal constrictions of the preceding sibilants, making the frication noises distinct from each other. Thus, the next section further discusses the potential advantage of the approximants over the simple back vowels.

4.2 Maximization of the perceptual distance via cue enhancement

The difference between approximants and back vowels is one of constriction, so the question is whether there are articulatory reasons to prefer approximant constriction to the vocalic constriction. Although the approximants are articulatorily more complicated than the simple back vowels, they might be an articulatory compromise given the environment they occur in. That is, they are corresponding approximants of the preceding sibilants, so they can be made by slightly releasing the apical constriction until the turbulence noise ceases. However, it is not clear if this is indeed articulatorily easier than the complete release of the apical constriction to back vowels. Note that the tongue body is already retracted for these segments due to the inherent preference of dental and retroflex consonants for a retracted tongue body.

In comparison, a possible advantage based on perceptual reasons appears to be more straightforward. That is, those apical segments contain additional cues to the place of the preceding dental and retroflex consonants. Recall that the dental approximant is characterized by low F2 and the retroflex approximant itself is strongly retroflexed, presenting low F3. Since the additional information on the places of articulation of the preceding consonants is present throughout the voiced period, the perceptual distinction cued by inherent frication noise can be further enhanced. In contrast, the simple back vowels are limited as a source of additional perceptual cues to the place of preceding consonants. That is, the low F2 of the back vowels are

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

Table 5 Perceptual cues to the sibilant places. The hyphen represents the cues are not present.

	Frication noise	Cues in the vocalic period
{si ~ ʂi ~ ɕi}	– (potentially obscured by palatalization)	–
{sɨ ~ ʂɨ ~ ɕi}	Distinct	Partial (low F2 in [ɨ] singles out non-palatals from the alveolo-palatal sibilant)
{sɹ̩~ ʂɻ ~ ɕi}	Distinct	Complete (in addition to F2, low F3 in [ɻ] further singles out [ʂ] from [s])

distinct from the high F2 of [i], signaling the preceding non-palatal consonants, but crucially they do not provide further cues to distinguish two non-palatal consonants: [s] and [ʂ].[8]

In summary, the advantage of the approximants seems to be two-fold. First, they do not interfere with the articulation of the preceding sibilants while the high front vowel may cause palatalization of the sibilants. This gives rise to an effect of preservation of the distinction in frication noise. Second, the additional cues to the sibilant places in the vocalic period further enlarge the perceptual distance between the three sibilant pairs. The simple back vowels do not create palatalization, but they are less preferable in that they do not carry any enhancing place cues to the sibilants. Table 5 summarizes the perceptual advantage of the approximants over the high front vowel and the simple back vowel.

4.3 The approximant account in summary

Despite the neat phonological account under the approximant analysis, it is worth noting potentially undesirable aspects of this account. First, this analysis introduces a very unique sound, the dental approximant [ɹ̪] (note that [ɻ] occurs as an independent phoneme in Mandarin), presumably unattested in any other language. This may seem undesirable at first, but the pursuit of phonetic precision has ultimately led us to consider further phonological support for the approximant analysis. As discussed in the previous section, the presence of these approximants enables us to address the more fundamental question of why Mandarin has developed these unique sounds when other languages simply neutralize the phonological contrast between sibilants before the high front vowel. Although the syllabic fricative account could also be motivated by contrast maintenance, while not introducing new segments into the inventory, it rests on the assumption that frication noise in voiced segments enhances the percept of the preceding sibilants. However, there is no observable frication in the following apical segments.

Another potential objection to the approximant account is that it creates an unusual syllable type where an approximant serves as the nucleus of a syllable for Mandarin, which typically only has vowels as the nucleus of a syllable. While the vowel account completely avoids this issue, the approximant account only creates an 'unusual' syllable type, not an unattested one. That is, there is nothing phonologically unnatural in having both syllabic vowels and approximants; this scenario is consistent with the cross-linguistic universal that less sonorous nuclei imply more sonorous nuclei. An unnatural scenario would be one where fricatives or stops may serve as a syllable nucleus but approximants do not. Rather, the presence of these unique segments in Mandarin bears on a bigger phonological question, ultimately contributing to our understanding of phonology.

The approximant account is not only preferable for a phonological perspective, but it is also superior to the apical vowel account in terms of articulation and acoustics. Articulatorily, the

[8] Stevens et al. (2004: 400) mention that the low F2 of dental and retroflex segments is 'a continuation of the enhancement gesture' for the preceding consonants. In other words, low F2 in these segments signals non-palatal properties of the preceding consonants. This is closer to the case where dental and retroflex consonants are followed by simple back vowels {sɨ ~ ʂɨ ~ ɕi}, where low F2 can contribute to distinguish the palatal consonant from the non-palatal ones.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

notations [ɹ̟] and [ɻ] capture their places of articulation straightforwardly, so the connection with the preceding sibilants [s] and [ʂ] becomes immediately available. Acoustically, the approximant analysis does not create an atypical association between the resonance cavity and formants, unlike the vowel analysis, falling within the general acoustics of rhotic and liquid sounds in other languages.

5 Conclusion

This study proposes that the apical segments exclusively occurring after the dental and retroflex sibilants in Mandarin Chinese can be best analyzed as corresponding approximants [ɹ̟] and [ɻ]. The goal of introducing these unique segments to the system seems to be two-fold. First, as homorganic consonants, these segments do not affect the primary places of the preceding consonants. This results in preservation of the frication internal cue, frication noise, for the phonological contrast among three sibilants. The high front vowel, however, may jeopardize the place contrast, since it often triggers palatalization of anterior coronals which leads to neutralization of distinct places. Second, specific formant values of each approximant provide additional cues to the distinct sibilant places. Along with the frication internal cue, the external cues in the surrounding approximants appear to contribute to maximize the perceptual distance among three sibilants.

Acknowledgements

I would like to thank Lisa Davidson, Gillian Gallagher, Christine Shadle, Carina Bauman, the NYU PEP Lab and audiences at the Acoustical Society of America for their helpful discussion. I also thank the participants in this study and Peggy Mok for allowing me to use the recording facilities at the Chinese University of Hong Kong. Finally, I thank the editor of *JIPA* and the anonymous reviewers for the careful and thorough reviews of this manuscript.

References

Alwan, Albeer, Shrikanth Narayanan & Katherine Haker. 1997. Toward articulatory–acoustic models for liquid approximants based on MRI and EPG data. Part II: The rhotics. *Journal of the Acoustical Society of America* 101(2), 1078–1089.

Ao, Xiaoping B. 1997. *Phonetics and phonology of Nantong Chinese.* Ph.D. dissertation, The Ohio State University.

Boersma, Paul & David Weenink. 2011. PRAAT: Doing phonetics by computer. Version 5.2, http://www.praat.org (accessed April 2011).

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

Chao, Yuenren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

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

Cheng, Chinchuan. 1973. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, Robert L. 1966. Mandarin phonological structure. *Journal of Linguistics* 2(2), 135–262.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1), 407–415.

Davidson, Lisa & Paul De Decker. 2005. Stabilization techniques for ultrasound imaging of speech articulations. *Journal of the Acoustical Society of America* 117(4), 2544.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

Duanmu, San. 2000. *The phonology of standard Chinese*. New York: Oxford University Press.
Duanmu, San. 2007. *The phonology of standard Chinese*, 2nd edn. New York: Oxford University Press.
Epstein, Melissa A. & Maureen Stone. 2005. The tongue stops here: Ultrasound imaging of the palate. *Journal of the Acoustical Society of America* 118(4), 2128–2131.
Flemming, Edward. 1995. *Auditory representations in phonology.* Ph.D. dissertation, University of California, Los Angeles.
Gu, Chong. 2002. *Smoothing spline ANOVA models*. New York: Springer.
Hamann, Silke R. 2003. *The phonetics and phonology of retroflexes*. Ph.D. dissertation, Utrecht University. [Utrecht: LOT Press]
Hartman, Lawton M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20(1), 28–42.
Howie, John M. 1976. *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.
Hsueh, F. F. S. 1986. *Beijing Yinxi Jiexi* [Analysis of the Beijing dialect sound system]. Beijng: Beijing Yuyan Xueyuan Chubanshe.
Iskarous, Khalil, Christine H. Shadle & Michael I. Proctor. 2011. Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America* 129(2), 944–954.
Karlgren, Bernhard. 1915–16. *Études sur la phonologie chinoise*. Uppsala: K. W. Appelberg.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.
Ladefoged, Peter & Jongji Wu. 1984. Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12, 267–278.
Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1), 109–112.
Li, Fangfang. 2008. *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*. Ph.D. dissertation, The Ohio State University.
Li, Min, Chandra Kambhamettu & Maureen Stone. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6/7), 545–554.
Lin, Yenhwei. 1989. *Autosegmental treatment of segmental processes in Chinese phonology*. Ph.D. dissertation, The University of Texas at Austin.
Lisker, Leigh. 1963. *Introduction to spoken Telugu*. New York: Americal Council of Learned Societies.
Martin, Samuel Elmo. 1992. *A reference grammar of Korean*. Tokyo: Charles E. Tuttle.
Miller, Wick R. 1985. *Acoma grammar and texts*. Berkeley & Los Angeles, CA: University of California Press.
Narayanan, Shrikanth, Dani Byrd & Abigail Kaun. 1999. Geometry, kinematics, and acoustics of Tamil liquid consonants. *Journal of the Acoustical Society of America* 106(4), 1993–2007.
Pulleyblank, Edwin G. 1984. *Middle Chinese: A study in historical phonology*. Vancouver, BC: University of British Columbia Press.
Ramsey, S. Robert. 1987. *The languages of China*. Princeton, NJ: Princeton University Press.
Spencer, Andrew. 1984. Eliminating the feature [Lateral]. *Journal of Linguistics* 20(1), 23–43.
Steriade, Donca. 1999. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics: Papers in Phonology* 3, 25–145.
Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
Stevens, Kenneth N., Samuel Jay Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.
Stevens, Kenneth N., Zhiqiang Li, Chao-Yang Lee & Samuel Jay Keyser. 2004. A note on Mandarin fricatives and enhancement. In Gunner Fant, Hiroya Fujisaki, Jianfen Cao & Yi Xu (eds.), *From traditional phonology to modern speech processing*, 393–403. Beijing: Foreign Language Teaching and Research Press.
Wang, Yuedong. 2011. *Smoothing splines: Methods and applications*. Boca Raton, FL: CRC Press.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press

Wiese, Richard. 1997. Underspecification and the description of Chinese vowels. In Jialing Wang & Norval Smith (eds.), *Studies in Chinese phonology*, 219–249. Berlin: Mouton de Gruyter.
Wu, Yuwen. 1994. *Mandarin segmental phonology*. Ph.D. dissertation, University of Toronto.
Zee, Eric & Wai-Sum Lee. 2001. An acoustical analysis of the vowels in Beijing Mandarin. *7th European Conference on Speech Communication and Technology*, Aaborg, Denmark, 643–646.
Zhou, Dianfu & Jongji Wu. 1963. *Putonghua fayin tupu* [Articulatory diagrams of Standard Chinese]. Beijing: Shangwu yinshuguan.

https://doi.org/10.1017/S0025100314000267 Published online by Cambridge University Press