On the nature of apical vowel in Jixi-Hui Chinese: Acoustic and articulatory data

Bowei Shao
Laboratoire de Phonétique et Phonologie (UMR 7018),
Université Sorbonne Nouvelle – CNRS, France
bowei.shao@sorbonne-nouvelle.fr

Rachid Ridouane
Laboratoire de Phonétique et Phonologie (UMR 7018),
Université Sorbonne Nouvelle – CNRS, France
rachid.ridouane@sorbonne-nouvelle.fr

Chinese languages have a set of segments known as apical vowels, which have been analysed in previous studies as either genuine vowels, fricative vowels, fricative consonants, or approximants. This study is concerned with the apical vowel attested in Jixi-Hui Chinese. We examine this segment from acoustic and articulatory perspectives and argue that it is best defined as a fricative /z/. Phonologically, Jixi-Hui Chinese /z/ is a distinct phoneme that is exclusively attested in syllable nucleus position where it constitutes a tone-bearing unit and which can undergo tonal sandhi processes. It can appear not only after coronal sibilants /s ts ts^h/, but also after bilabials /p p^h/ and nasals /m n/. Acoustically, we show that this segment contains frication noise in its initial phase in the majority of cases, with a formant structure towards its end. The analysis of the zero-crossing rate confirms this significant presence of noise, clearly distinguishing this segment from genuine vowels. Furthermore, articulatory analyses of ultrasound data show that /z/ has a near-identical tongue shape to fricative /s/ on both mid-sagittal and coronal planes, in both sibilant and non-sibilant contexts. These findings are viewed in light of the variability in the way /z/ is phonetically implemented in Jixi-Hui Chinese.

1 Introduction

Chinese languages have a specific set of segments, known as apical vowels. To date, their exact nature remains a source of debate: Some view these segments as genuine vowels, while others consider them to be consonants (fricatives or approximants). In the current study we explore the apical vowel in Jixi-Hui Chinese (JHC) 绩溪话. The apical vowel in JHC is of a special interest in contributing to a better understanding of the nature of these segments, since it displays two structural properties that make it different from the most studied variants in other Chinese languages: It is a separate phoneme contrastive to /i/ and to other vowels and it occurs not only following alveolar sibilants but also after bilabial plosives /p p^h/, bilabial nasal /m/ and alveolar nasal /n/.

Journal of the International Phonetic Association (2023) 53/3

doi:10.1017/S0025100322000196

This article is organised as follows. We first present a brief review of the literature on apical vowels in Chinese languages. This is followed by a description of JHC apical vowel based on common principles of phonological analyses: lexical distribution, phonemic contrast, and function within the syllable. We then report results from two experiments to determine the phonetic characteristics of this segment, based on acoustic and ultrasound data. We conclude with discussion of the fricative/approximant nature of JHC apical vowel.

1.1 Apical vowels in Chinese languages: A brief review[1]

Most previous studies on apical vowels have focused on Standard Chinese or Beijing Mandarin, but apical vowels are also attested in other Chinese languages (Wu 1995, Wang 2006, Hu 2007, Hou 2009), and some non-Chinese Sino-Tibetan languages (Baron 1974, Michaud 2008, Wang 2010). The presence of apical vowels in all these languages has always been related to a historical /i/ at some stage of the evolution (Zhu 2004, Zhao 2007, Jacques & Michaud 2011, Gong 2016).

1.1.1 Apical vowels in Standard Chinese

The terminology 'apical vowels' and the non-IPA symbols [ɿ ʅ] used to transcribe them date back to Karlgren (1915)'s study of Standard Chinese (SC), and have been widely used since then among researchers working on the phonetics and phonology of Chinese (R. Cheng 1966, Trubetzkoy 1969, C. Cheng 1973, Howie 1976, Svantesson 1984, Ladefoged & Maddieson 1996, Zee & Lee 2007, Faytak & Lin 2015, Shi, Peng & Liu 2015, Faytak 2018). Although these symbols are widely used in the transcription of Chinese (e.g. for the minimal pair [sɿ^{55}] 丝 'silk' and [ṣʅ^{55}] 诗 'poem'), they are not accepted by the IPA. The 1949 Principles of IPA (IPA 2010) consider [ɿ] as a 'ɯ-type sound accompanying friction' thus 'resembling z', and [ʅ] 'accompanying friction and resembles ʒ'. Pullum & Ladusaw (1996) think that [ɿ] is 'essentially a syllabic [z]' and [ʅ] 'essentially a syllabic [ẓ] or maybe [ʒ]'.

Trubetzkoy (1969: 171) described apical vowels as 'a type of vowel with a much lesser degree of aperture and with a much more fronted position of articulation than, for example, i, so that a friction like noise resembling a humming is audible in its production'. Several reasons have been put forth to argue that these segments are vowels (R. Cheng 1966, C. Cheng 1973, Duanmu 2007). First, they are allophonic variants of the vowel /i/; [ɿ] occurs after dental sibilants, [ʅ] occurs after retroflex sibilants, and [i] occurs elsewhere. Second, they behave as vowels in that they are syllable nuclei and can be tone-bearing units. Third, they sometimes have phonetic characteristics of a vowel: they have formants (Howie 1976, Svantesson 1984, Shi et al. 2015) and may be produced with a raised tongue body (Zhou & Wu 1963, C. Cheng 1973).

There are also reasons, however, to consider these segments as fricatives. First, they contain frication noise (Ladefoged & Maddieson 1996, Yu 1999, Faytak & Lin 2015). Second, they are always homorganic with the preceding sibilant consonants (Chao 1961, 1968). Third, they have the same tongue gesture and alveolar or post-alveolar constriction as the fricatives /s/ and /ṣ/ (Zhou & Wu 1963). From a phonological point of view, Dell (1994) considered these segments as syllabic fricatives and interpreted them as the voiced prolongation of the preceding sibilant consonants. Wiese (1997) and Duanmu (2007) analysed these segments as

---

[1] Multiple references with different transcription conventions are cited in this study, and the phonemic status of the segments involved is not always clear. In Section 1, forms enclosed between slashes are phonemic transcriptions. Those enclosed between square brackets are intended to represent phonetic representations. In Sections 2, 3 and 4, we report and discuss the phonetic shapes of the concerned segments and consistently use brackets for the sake of clarity.

empty syllable nuclei. For Duanmu (2007), the sibilant property of apical vowels is triggered by the spreading of the feature [+fricative] from the preceding onset consonant.

The analysis of apical vowels as fricatives has been questioned by some recent studies, while still arguing that these segments are consonants (Lee & Zee 2003, Lee-Kim 2014). In their IPA description of SC, Lee & Zee (2003), who use the same symbol [ɹ̩] to transcribe the two apical vowels [ɿ ʅ], describe them as 'syllabic apical post-alveolar approximant' and 'syllabic apico-laminal or laminal dento-alveolar approximant'. For Lee-Kim (2014), [ɿ] and [ʅ] are syllabic approximant counterparts of the dental and retroflex sibilants, respectively. Her acoustic and articulatory data showed that these segments display very short durations of frication noise and are homorganic to the sibilant onsets, with a slightly retracted tongue root for [ɿ] and a slightly lowered tongue body for [ʅ]. The short frication noise observed is, according to Lee-Kim, a consequence of gestural overlap with the preceding sibilants.

The different definitions of apical vowels in SC (i.e. vowel, fricative or approximant) depend on the weight assigned to phonetic or phonological criteria. On the one hand, the analysis of this segment as a vowel is essentially based on THE PHONOLOGICAL PATTERNING of apical vowels: (i) they are allophonic to vowel /i/; (ii) they function as syllable nuclei; and (iii) they can be tone-bearing units. This point of view is phonologically convenient since it complies with the canonical syllable structure of SC, in that the nucleus of a syllable should contain a vowel.[2] In this view, the acoustic presence of a formant structure in apical vowels is considered as supplementary evidence for a vowel analysis. On the other hand, the analysis of apical vowels as fricatives is based mainly on PHONETIC OBSERVATIONS, i.e. the acoustic presence of frication noise and the same tongue configuration as for /s/ and /ʂ/. This analysis is however phonologically unnatural since it assumes that an underlying vowel (here /i/) has consonants as allophonic variants (Wiese 1997). It also assumes that the fricative consonants in [sz ʂʐ] syllables are tone-bearing units, a striking exception to the behaviour of other obstruents in the language.

The approximant analysis is also based on PHONETIC OBSERVATIONS. It captures the fact that the tongue shape is slightly different between the sibilant onsets and the apical vowel, and thus explains the absence of frication noise in some cases. The absence of frication noise is not consistent across all studies, however. While Lee-Kim (2014) reports no or little frication noise on apical vowels, other researchers report important interspeaker variation (Faytak & Lin 2015).

The difficulty encountered in the analysis of apical vowels in SC is that their phonetic implementation does not match their phonological behaviour. The former provides evidence for a consonant analysis while the latter provides arguments for a vowel analysis. The same dilemma, as we show below, has been faced by researchers working on apical vowels in other Chinese languages.

1.1.2 Apical vowels in other Chinese languages

Hefei-Mandarin Chinese 合肥话 (HMC) is a Mandarin dialect of Jianghuai-Mandarin 江淮官话 group (Li, Xiong & Zhang 1987). It has in addition to [ɿ ʅ] a third apical vowel [ʮ], which represents the rounded version of [ɿ] (Karlgren 1915: 297; Kong, Wu & Li 2019). Unlike in SC, the apical vowel [ɿ] in HMC can be preceded not only by a homorganic onset (i.e. [ts tsʰ s z]) but also by [p pʰ m] (Kong et al. 2019). At the phonetic level, the study from Hou (2009) reports that the HMC apical vowel [ɿ] has strong high-frequency frication noise in the 3000–5000 Hz region, but the frication noise does not extend to the end of the segment. Hou (2009) considers the apical vowels as fricative vowels 带擦元音 but does not provide further discussion on its phonological patterning. Kong et al. (2019) also report that the three apical vowels display frication noise, with male speakers having more frication than female

---

[2] The nucleus of a syllable in SC is almost always a vowel, only very marginal cases of syllabic consonants such as [m n ŋ] have been reported (Duanmu 2007: 34)

speakers. Following Ladefoged & Maddieson (1996), they treat this frication as a secondary vowel feature, introduced to enhance the perceptual saliency of the apical vowels.

Qinghai-Mandarin Chinese (QMC) 青海方言 is a Mandarin dialect of Lanyin-Mandarin 兰银官话 group (Li et al. 1987). It has one apical vowel [ɿ] which is phonologically analysed as a free variant of /i/ (Wang 2006). Similar to HMC, [ɿ] in QMC is not always homorganic to the preceding onset consonant, as it can be preceded by /p p^{h} m l s ts ts^{h}/. To our knowledge, there has been no published study to date describing the phonetic characteristics of this apical vowel.

An interesting case has been observed in Suzhou-Wu Chinese 苏州话 (SWC), a Wu 吴 group Chinese language (Li et al. 1987). This language has two apical vowels, a rounded /ʮ/ and an unrounded /ɿ/, both occurring after sibilant onsets /ts ts^{h} s z/ (Ye 1993). Interestingly, these apical vowels are independent phonemes which contrast with /i/, as shown by the following minimal triplet: /sɿ^{315}/ 四 'four' vs. /sʮ^{315}/ 世 'world' vs. /si^{315}/ 细 'small'. For Faytak (2018: 45) these segments 'have an apico-alveolar constriction similar to a /z/ and could be transcribed as syllabic rounded and unrounded alveolar fricatives with a loose degree of constriction, i.e. syllabic, lowered [z z^{w}]; both exhibit noticeable strident frication with a [z]-like quality'. Ling's (2009) acoustic and articulatory (EMA) analyses show that the two apical vowels display similar F1 and F2 values, present frication noise in the 3000–8000 Hz region, and have a flat or concave tongue shape and an apico-alveolar constriction.

As can be inferred from the studies above, apical vowels do not display similar structure and behaviour across Chinese languages. Depending on the language, they can have a more or less restricted distribution and can be phonemic or allophonic. Depending also on the language, and sometimes on the speaker, apical vowels may exhibit a more or less salient frication noise. Our study builds on the works reviewed above and provides a novel contribution by examining the nature of the apical vowel in an understudied Chinese language, Jixi-Hui Chinese. We first introduce some phonological characteristics of the apical vowel in this language, before reporting two production experiments designed to examine its acoustic and articulatory characteristics. We hereafter adopt the symbol /z/ to transcribe this segment, following the IPA guidelines (see also Chao 1961, 1968; Dell 1994; Wiese 1997; Yu 1999; Duanmu 2007).

1.2 The apical vowel in Jixi-Hui Chinese

Jixi-Hui Chinese is a Hui 徽 group language, spoken in the Jixi county, located 280 km southwest of Shanghai, in Anhui province 安徽省. It has two major variants, the Lingnan 岭南 variant (Luo 1936, Zhao 1989, Hirata 1998, Zhao 2003), which is the variant examined in this study, and the Lingbei 岭北 variant (Chao 1962, Chao & Yang 1965). The administrative centre is the town of Huayang 华阳镇, where the Lingnan variant is spoken. According to the most recent description (Zhao 2003), the vocalic system of JHC consists of eight monophthong phonemes (excluding the apical vowel), as shown in Figure 1.

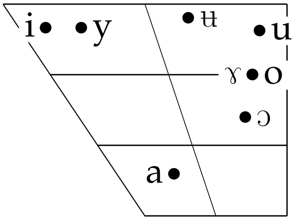

Figure 1 Monophthong phonemes of JHC (adapted from Zhao 2003). The apical vowel, represented by /ɿ/ in Zhao's description, is not shown in this figure. The /ʉ/ vowel is transcribed as /ɵ/ in Zhao's description, but our acoustic and ultrasonic data show that it is best transcribed as /ʉ/.

The consonants of JHC are presented in Table 1 (based on Zhao 2003). Among the consonants, three can function as syllabic nuclei: [v m n]. Items with syllabic nasals [m n] are not lexically frequent, but they are not limited to interjections as in SC, as the examples in (1) show:

(1) v^{213} 五 'five'
n^{213} 你 'you'
m^{213} 母 'female' or 'mother'

The syllabic fricative [v], which is a contextual variant of /u/, is more frequent (Zhao 2003). JHC has six tones including a checked tone. These are presented in Table 2. Syllables containing the apical vowel [z] occur on all five non-checked tones.

Table 1 Consonants of JHC according to Zhao (2003).

	Bilabial	Labio-dental	Alveolar	Alveolo-palatal	Velar	Glottal
Plosive	p pʰ		t tʰ		k kʰ	ʔ
Affricate			ts tsʰ	tɕ tɕʰ		
Nasal	m		n	ɲ	ŋ	
Fricative		f v	s	ɕ	x	

Table 2 Tones of JHC according to Zhao (2003), represented in tone letters (Chao 1930) and in IPA tone symbols.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6 (checked)
Tone letters	31	44	213	35	22	32
IPA symbols	˧˩	˦˦	˨˩˧	˧˥	˨˨	˧˨

1.2.1 The phonological patterning of JHC apical vowel

Syllables containing /z/ in JHC are not uncommon, accounting for 7.2% of the monosyllabic entries of Zhao's (2003) dictionary. This segment behaves in a different way compared to the apical vowels in SC. First, /z/ is a distinct phoneme, which can contrast with /i/ and other vowels, as shown by the minimal pairs and triplets in (2):[3]

(2) sz^{213} 死 'die' vs. si^{213} 洗 'wash'
tsz^{213} 紫 'purple' vs. tsi^{213} 走 'walk'
tsʰz^{213} 此 'here' vs. tsʰi^{213} 丑 'ugly'
pz^{31} 屄 'female genitals' vs. pi^{31} 碑 'stela'

---

[3] The pairs [pz^{31} mz^{213}] and [pi^{31} mi^{213}] were reported by Zhao (1989, 2003). However, the speakers recorded in this study did not pronounce [pi^{31} mi^{213}] but rather produced [pa^{31} mẽ^{213}] respectively. The forms [pi^{31} mi^{213}] are reported to be absent in the 'city (i.e. Huayang county) accent'.

mz^{213}	米 ‘rice’	vs.	mi^{213}	美 ‘beautiful’			
pz^{35}	闭 ‘close’	vs.	pa^{35}	背 ‘back’	vs.	pʉ^{35}	簸 ‘winnow’
p^{h}z^{35}	屁 ‘fart’	vs.	p^{h}a^{35}	配 ‘match’	vs.	p^{h}ʉ^{35}	破 ‘broken’
mz^{213}	米 ‘rice’	vs.	ma^{213}	每 ‘every’	vs.	mʉ^{213}	某 ‘someone’
nz^{33}	泥 ‘dirt’	vs.	na^{33}	来 ‘come’	vs.	nʉ^{33}	罗 ‘sift’

Second, it can be preceded by /p p^{h} m n/ in addition to /ts ts^{h} s/, showing that it is not always homorganic with the preceding onset.[4] As shown in (3), /z/ can also stand for a syllable on its own:

(3)	pz^{213}	比 ‘compare’	p^{h}z^{213}	被 ‘quilt’	mz^{213}	米 ‘rice’		
	nz^{213}	里 ‘in’	sz^{213}	死 ‘die’	tsz^{213}	紫 ‘purple’	ts^{h}z^{213}	此 ‘here’
	z^{213}	椅 ‘chair’						

Third, similar to SC apical vowels, /z/ in JHC can be a tone-bearing unit and can undergo tone sandhi processes. This is shown in (4):

(4)	/sz^{31}/	‘western’	+/tsõ^{31}/	‘clothes’	→	[sz^{33}tsõ^{31}]	西装	‘suit’
	/ts^{h}z^{35}/	‘gas’	+/tɕ^{h}ɔ^{31}/	‘vehicle’	→	[ts^{h}z^{53}tɕ^{h}ɔ^{31}]	汽车	‘car’
	/p^{h}z^{22}/	‘prepare’	+/t^{h}a^{31}/	‘tyre’	→	[p^{h}z^{53}t^{h}a^{31}]	备胎	‘spare tyre’

To sum up, the phonological behaviour of /z/ in JHC is similar to the behaviour of a vowel. It can be the nucleus of a syllable and a tone-bearing unit, and can undergo tone sandhi processes. However, as we will show below, acoustic and ultrasound data provide evidence that /z/ has the phonetic characteristics of a consonant.

2 Production experiment I: Acoustic study

This experiment aims to determine whether [z] displays acoustic characteristics of a vowel or a consonant. We analyse the formant structure of the apical vowel, and provide a detailed analysis of the frication noise that accompanies this segment.

2.1 Speakers

Speakers of JHC were recruited according to strict criteria to limit possible dialectal variation. They had to be born and raised in the town of Huayang, with both parents also born in the same town. Further criteria were that they had to live in the town of Huayang and speak JHC on a daily basis, and their age should be around 50 years old. Since JHC speakers all understand and speak SC, as it is common for Chinese people living in a city, we selected only those who speak JHC in both professional and non-professional contexts to limit the influence of SC. Five female speakers (FS1–5) and five male speakers (MS1–5) satisfying

---

[4] It is reported that outside of the Huayang county but inside the Lingnan variant, [ts^{h}z] could also be realised as [t^{h}z] (Chao 1962). This has not been observed among the speakers recorded for this study.

these criteria were chosen to participate in the study. The mean age of the speakers was 49 years (±3.8). None of them reported to have speech-related anomalies, and all of them considered themselves as native speakers of JHC with no accent.

2.2 Materials

Acoustic data were recorded with a hypercardioid headset microphone (AKG C520), an external sound card (Edirol UA25) and Audacity (v. 2.1.0) on a portable computer. The sampling rate was set to 44100 Hz. We had access to the sound attenuated studio of the local television channel for our recording sessions. Our speakers sat in a chair and read a word list embodied in a carrier sentence at a normal speech rate. The word list (see Table 3) was constructed with the segments /i a u ʉ z/ occupying the nuclei of monosyllabic words starting with /p p^{h} m n ts ts^{h} s/. With different tones, they form single noun phrases, presented to the speakers in Chinese characters. Each word was produced within the carrier phrase [ki^{44}ɕɔ^{213}_ ɕɔ^{213}sɔ^{44}fa^{44}] ‘He writes _ three times’. The entire list was repeated five times per speaker, yielding 2150 tokens ([z]: 550, [i]: 400, [u]: 450, [a]: 350, [ʉ]: 400).[5]

Table 3 The word list used in data acquisition for both acoustic and ultrasound studies.

Onset type	Phonetic transcription	Orthographic transcription	Gloss	Acoustic study	Ultrasound study	Notes
Labial	pz^{31}	屄	‘female genital’	√	√	
	pz^{213}	比	‘compare’	√	√	
	p^{h}z^{44}	皮	‘skin’	√	√	
	p^{h}z^{213}	被	‘quilt’	√	√	
	mz^{213}	米	‘rice’	√	√	
Coronal	nz^{44}	泥	‘mud’	√	√	
	nz^{213}	里	‘in’	√	√	
	tsz^{213}	紫	‘purple’	√	√	
	ts^{h}z^{213}	弟	‘younger brother’	√	√	
	sz^{213}	死	‘die’	√	√	
	sz^{35}	四	‘four’	√	√	
Labial	pa^{31}	杯	‘cup’	√		
	p^{h}a^{44}	赔	‘compensate’	√		
	ma^{213}	每	‘every’	√		
Coronal	la^{44}	来	‘come’	√		
	tsa^{213}	宰	‘slaughter’	√	√	
	ts^{h}a^{213}	踩	‘step on’	√	√	
	sa^{35}	碎	‘smash into pieces’	√	√	
Labial	pi^{31}	碑	‘stela’	√		produced as [pa^{31}]
	mi^{213}	美	‘beautiful’	√		produced as [mɛĩ^{213}]
Coronal	li^{44}	留	‘stay’	√		
	li^{213}	柳	‘willow’	√		
	tsi^{213}	走	‘walk’	√	√	

---

[5] The uneven numbers of syllables are due to the lack of actual JHC real words with the relevant structure.

Table 3 Continued.

Onset type	Phonetic transcription	Orthographic transcription	Gloss	Acoustic study	Ultrasound study	Notes
	ts^{h}i^{213}	丑	'ugly'	✓	✓	
	si^{213}	洗	'wash'	✓	✓	
	si^{35}	瘦	'thin'	✓	✓	
Labial	pu^{213}	补	'repair'	✓		
	p^{h}u^{44}	葡	'grape'	✓		
	p^{h}u^{213}	普	'general'	✓		
Coronal	lu^{44}	炉	'stove'	✓		
	lu^{213}	鲁	'a family name'	✓		
	tsu^{213}	组	'group'	✓	✓	
	ts^{h}u^{213}	础	'foundation'	✓	✓	
	su^{213}	竖	'vertical, erect'	✓	✓	
	su^{35}	漱	'rinse'	✓	✓	
Labial	pʉ^{31}	波	'wave'	✓	✓	
	p^{h}ʉ^{44}	婆	'elderly woman'	✓	✓	
	p^{h}ʉ^{213}	叵	'impossibly'	✓	✓	
	mʉ^{213}	某	'someone, something'	✓	✓	
Coronal	lʉ^{44}	揉	'knead'	✓		
	tsʉ^{213}	左	'left'	✓	✓	
	tsʉ^{35}	做	'make'	✓	✓	
	sʉ^{213}	锁	'lock'	✓	✓	

2.3 Data analysis

The recorded acoustic data were segmented and annotated using Praat (Boersma & Weenink 2018). It was challenging to determine a clear boundary between a sibilant onset and the following apical vowel (for example in [tsz]). We labelled our data by taking the first voicing pulse detected in Praat as the beginning of the apical vowel. This is illustrated in Figure 2.

Measurement criteria were established after visual inspection of both spectrograms and waveforms. The following parameters were considered:

(i) FRICATION IN APICAL VOWELS. This aspect was obtained through visual inspection of the acoustic waveforms and spectrograms for all utterances. Fricated apical vowels were those exhibiting a degree of turbulent airflow on the spectrogram in the high and mid frequencies.

(ii) FORMANT VALUES OF [a i u ʉ] AND [z]. A Praat script trisected all vocalic segments and calculated mean F1, F2 and F3 values at the middle third of the target segments. The maximum frequency of formant calculation was set to 5500 Hz for female speakers and 5000 Hz for male speakers.

(iii) ZERO-CROSSING RATE (ZCR) OF [a i u ʉ] AND [z]. The upward and downward zero-crossing points within a 40 ms sliding-window on the acoustic signals were obtained as a PointProcess object in Praat. The zero-crossing rate was calculated as the number of zero-crossings divided by the length of the window. Fifty-two data points (the onset and offset of [a i u ʉ z] and 50 data points evenly spaced throughout each segment) were extracted and analysed.

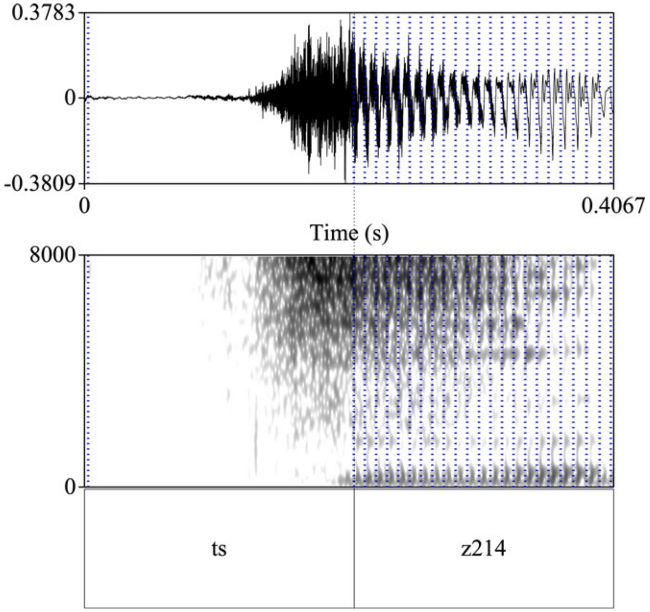

Figure 2 Illustration of the segmentation used to delimit a sibilant onset and an apical vowel within the word [tsz^{214}] as pronounced by FS3. The dotted lines on the signal represent voicing pulses detected by Praat.

2.4 Results

2.4.1 Frication in apical vowels

The examination of the waveforms and spectrograms showed that 88% of the realisations of [z] (472 out of 537 tokens)[6] were produced with frication noise. As shown in Table 4, frication noise was observed for all speakers and in all contexts. The amount of the frication noise varied greatly, with the majority of cases displaying frication in less than half of the [z] duration. Some illustrative examples are shown in Figure 3, with apical vowels displaying frication following [p m n].

The presence of frication in apical vowels was not dependent on the nature of the preceding consonant, as it was implemented following either sibilants or non-sibilants. Given that the labial stop [p] and the nasals [m n] had clearly no frication to spread, the frication noise displayed by [z] in these contexts suggests that it is an inherent property of this segment. It is important to note, however, that frication noise never extended until the end of apical vowel. When frication diminished, periodic waveforms became clearer, and formant structures were visible in the spectrograms. Note also that this frication was systematically superposed on voicing, showing that frication does not necessarily inhibit the voicing of the apical vowel. A more detailed analysis concerning this parameter in different contexts is presented in Section 2.4.3.

The strict segmentation criteria applied in this study ensured that only the voiced portion of [z] was taken into account. A probable consequence of this segmentation is that [z] displayed shorter duration after [s ts ts^{h} p^{h}][7] than after [p m n] (Table 5). Given that [z] is homorganic to the coronal sibilant onsets, it could be the case that part of the frication noise

---

[6] Thirteen data points were excluded when the speaker made an error, or when disfluencies occurred.

[7] The reason why [p^{h}] patterns with sibilants rather than with labials was reported in Shao & Ridouane (2021).

Table 4 Qualitative observation of the presence of frication noise on [z] in JHC. Presence of frication noise in each context is classified according to the quantity of observable noise on each [z] segment. < indicates that frication noise is observable over less than half the duration of [z]; > indicates that it is observable over more than half the duration of [z]; ≈ indicates that it is observable over roughly half the duration of [z]; two symbols in the same context indicate that both cases are observed for the same speaker.

	p	p^h	m	n	ts	ts^h	s
FS1	<	<	<	<	<≈	<≈	<≈
FS2	<≈	<	<	<	<	<	<
FS3	<	<	<	<	<≈	<≈	<≈
FS4	<	<	<≈	<≈	<≈	<≈	<≈
FS5	<	<	<	<	<	<	<
MS1	<	<	<	<≈	<≈	<≈	<≈
MS2	<	<	>	>	<≈	<≈	<≈
MS3	<	>	>	>	>	>	>
MS4	<	<	<	<	<	<	<
MS5	<	<	<	<	<≈	<≈	<≈

FS = female speaker; MS = male speaker

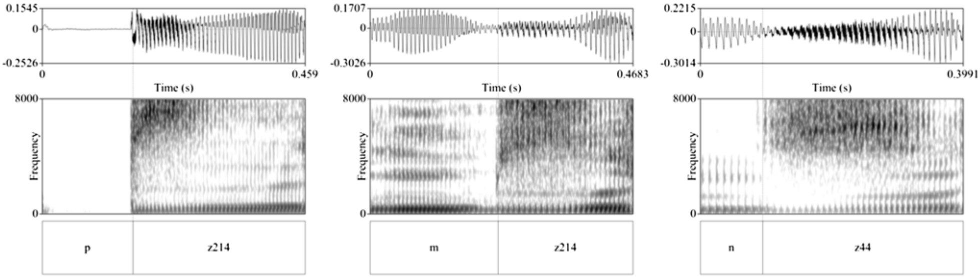

Figure 3 The waveforms and spectrograms of JHC apical vowels containing frication noise. The productions of [pz^{214} nz^{44} mz^{214}] come from speakers FS1, MS3 and MS2, respectively.

at the offset of the sibilant onsets was, in fact, the onset of the frication noise of [z] (suggesting that [z] was probably longer than what the segmentation criteria used indicated).

Table 5 Mean durations (ms) of [z] after different onsets, with standard deviations in parentheses.

	s	ts	ts^h	p^h	p	m	n
Mean duration of [z]	192 (47.3)	180 (56.1)	176 (50.3)	191 (49.2)	234 (57.6)	241 (53.0)	266 (48.4)

2.4.2 Formant analysis

We calculated mean formant values at the middle third of [z], where the frication noise started diminishing and a clear formant structure became visible. The values obtained were compared to the formant values of the vowels [a i u ʉ]. The results for F1 and F2 are presented in Figure 4.[8] They show that the values of [z] overlapped those of [ʉ] for female speakers and those of [ʉ] and [u] for male speakers.

---

[8] The results for F1 and F2 were used to adjust the vowel positions in Figure 1. The structure of F1, F2 and F3 values of [i u a ʉ z] is shown in appendix Figure A1.

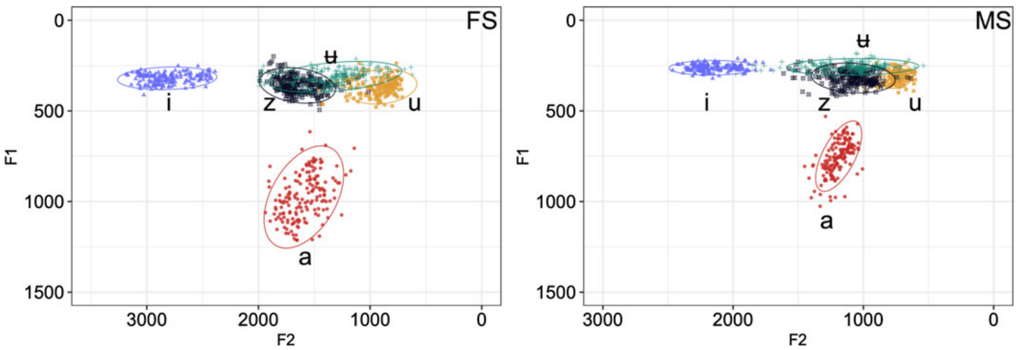

Figure 4 (Colour online) Scatter plot of formant values in Hz of [i u a ʉ z] of JHC. Data points represent mean values of the middle third of each segment, with 95% confidence ellipses. Female speakers (FS) are on the left and male speakers (MS) on the right.

Table 6 Statistical results of linear mixed-effects analyses conducted on formant values (Bark-scaled) of [z u ʉ] in JHC; standard deviations in parentheses.

			Fixed effects estimates (Bark)			*p*-values		
	N of obs.		Intercept[ʉ] (se)	Segment [u] (se)	Segment [z] (se)	u - ʉ	z - ʉ	z - u
FS	633	F1	3.14 (0.12)	0.53 (0.15)	0.45 (0.12)	*	*	n.s.
		F2	10.24 (0.46)	−2.13 (0.20)	1.41 (0.55)	***	n.s.	***
		F3	15.5 (0.15)	−0.06 (0.08)	0.64 (0.17)	n.s.	*	**
MS	619	F1	2.56 (0.06)	0.63 (0.08)	0.66 (0.09)	**	**	n.s.
		F2	9.05 (0.16)	−2.18 (0.15)	−0.01 (0.31)	***	n.s.	**
		F3	13.90 (0.22)	0.29 (0.17)	0.84 (0.12)	n.s.	**	**

FS = female speakers; MS = male speakers

* $p < .05$; ** $p < .01$; *** $p < .001$; n.s. = non-significant *p*-values

We conducted linear mixed-effects analyses to compare the differences in F1, F2 and F3 across the three segments [z u ʉ]. The formant values were converted to the Bark scale (Traunmüller 1990) to better capture their auditory properties. We used R (R Core Team 2017) and *lme4* (Bates et al. 2015) to perform linear mixed-effects analyses of the relationship between formants and the three segments. In the models, we entered Bark-scaled formants as dependant variables, and we entered segments as fixed effects. As random effects, we had by-speaker random slopes for the effect of segments. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The results are reported in Table 6.[9] F1 values of [u] and [z] did not differ significantly, with both having higher F1 values than [ʉ]. The results further confirmed that for all speakers, F2 values of [ʉ] and [z] were not significantly different. F3 values were significantly lower in [u ʉ] than in [z], reflecting the fact that [u ʉ] are rounded vowels whereas [z] does not involve lip rounding.

2.4.3 Zero-crossing-rate of syllable nuclei

ZCR is defined as the number of times in a given time-interval the speech signal passes zero. It measures the times of zero-crossings in a given time-interval, without involving the

---

[9] The *p*-values of the pair [z]–[u] were obtained by relevelling the segments to have the segment [u] as intercept. The results of the releveled models are shown in appendix Table A1.

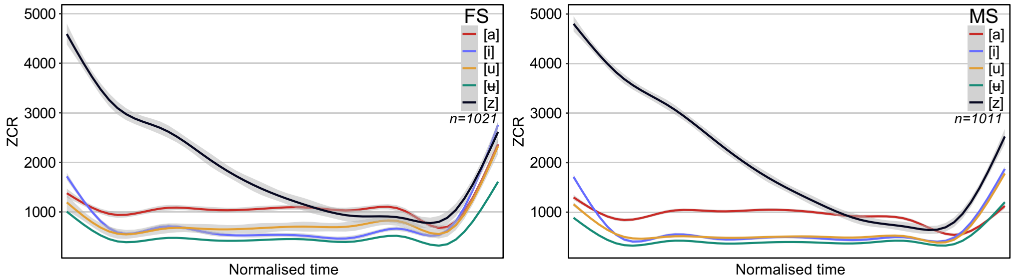

Figure 5 (Colour online) Zero-crossing rate (ZCR) of [a i u ʉ z] in JHC. The curves were generated using the loess smoothing method. The x-axis represents normalised time of the vocalic segments and the y-axis represents the zero-crossing times per second. Female speakers (FS) are shown on the left and male speakers (MS) are on the right.

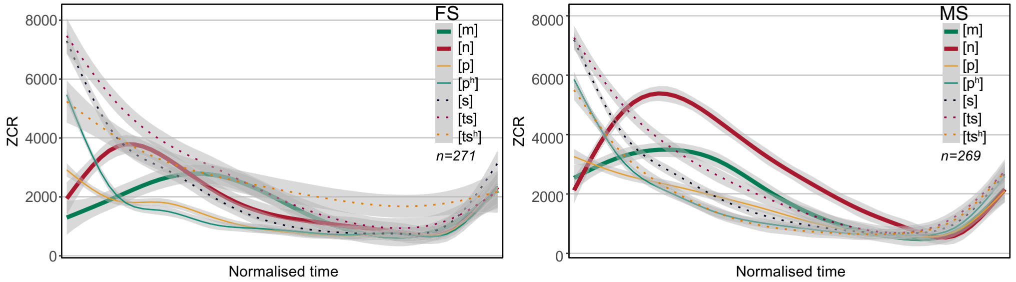

Figure 6 (Colour online) Zero-crossing rate of [z] in different consonantal contexts in JHC. The curves were generated using the loess smoothing method, x-axis represents normalised time of the [z] segments and the y-axis represents the zero-crossing times per second. Female speakers (FS) are shown on the left and male speakers (MS) are on the right.

detection of voicing or pitch. It is considered to be a reliable measurement of the intensity of frication noise (Shosted 2006), with higher ZCR indicating higher aperiodicity. ZCR can be used for all syllable nuclei regardless of their phonetic nature, and it is less speaker-dependent than spectral analyses (Ito & Donaldson 1971). This measurement has been applied for fricatives and vowels (Ito & Donaldson 1971), nasalised fricatives (Bombien 2006), or aspirated vowels (Gordeeva & Scobbie 2010). We present ZCR as a variable over the normalised duration of the concerned segments.

ZCR values of [a i u ʉ] and [z] are presented in Figure 5. As the figure shows, [z] behaved differently from the vowels. It started with a very high ZCR (well above 1000 times per second), corresponding to the frication noise observed at the beginning of the segment, then ZCR lowered, which corresponded to the slow disappearance of this noise.

Figure 6 shows the relation between ZCR values of [z] and the nature of the preceding consonant. All [z] segments displayed a high ZCR rate (i.e. above 1000 times per second), regardless of whether the preceding consonant was a sibilant or not (at least during the first half). The contexts where [z] was preceded by [m n] displayed a noticeable divergence from the other contexts. ZCR started at a lower level and increased to achieve a rather high level before the final falling phase. This lower ZCR at the starting point could be attributed to the nasality of the preceding consonants, as nasal consonants require an open nasal cavity that prevents high intraoral pressure. After the release of the nasal consonant, the nasal cavity closes, and intraoral pressure increases gradually. The gradual closure of the nasal cavity led to a gradual increase of the intraoral pressure, which in turn led to the appearance of the frication noise, as indicated by an increase in ZCR. The difference between [z] preceded by [m] and [z] preceded by [n] could be attributed to the alveolar constriction involved in [n], but not in [m], suggesting that in [mz], the alveolar constriction was achieved later and was less constricted.

2.5 Interim summary

The analysis of the acoustic data showed that the overwhelming majority of the apical vowel productions contained frication noise in at least the first half of their durations. This frication was observed following both sibilants and non-sibilants, suggesting that it could not be attributed to the coarticulation of preceding consonants, at least not in the context of non-sibilants /p p^{h} m n/. Results from ZCR provided additional evidence that [z] displayed characteristics of a fricative sound, making it acoustically different from genuine vowels. Frication noise, however, never extended until the end of the apical [z]. When it diminished, periodic waveforms appeared, and a clear formant structure was visible. The analysis of this formant structure indicated that [z] had similar F1–F2 structure to [ʉ], shown by the significant overlap in the vocalic space.

3 Production experiment II: Articulatory study

The ultrasound experiment examined the tongue shape configuration of [z] in different contexts, both at the mid-sagittal and coronal planes. It specifically sought to determine to what extent the tongue shape of this segment differed from or resembled that of the fricative consonant [s].

3.1 Participants and materials

The ultrasound data were recorded in the same sound-attenuated room as for experiment I. Six of the ten speakers recorded for the acoustic experiment participated in the data acquisition (female speakers FS1, FS3, FS5 and male speakers MS2, MS3, MS5).[10] The word list was presented in Table 3. The subjects read the words in the same frame sentence [ki^{44}çɔ^{213}_ çɔ^{213}sɔ^{44}fa^{44}] ‘He/She writes _ three times’, repeated three times for the mid-sagittal plane and three times for the coronal plane (1218 tokens; [z]: 462, [i]: 168, [u]: 168, [a]: 126, [ʉ]: 294).

3.2 Data acquisition and analysis

The ultrasound data were recorded with the Ultrasound Stabilisation Headset (Articulate Instruments Ltd. 2008) and the Articulate Assistant Advanced software (AAA, V217.03) (Articulate Instruments Ltd. 2012). The probe used in data acquisition was a portable micro-convex ultrasonic probe with a diameter of 40 mm. The speakers sat in a comfortable chair; the headset was then adjusted to the anatomical specificities of each speaker, so that it did not move during the recording, nor did it cause discomfort. Data were recorded first in the mid-sagittal plane, and then in the coronal plane, with a small pause between two sessions during which the headset was removed.

The headset kept the ultrasound probe in the mid-sagittal plane and the probe was pointed to the anterior of the tongue in order to have the best image of the tongue tip. The coronal recording was obtained by turning the probe in a 90° angle and pointing the probe to the anterior part of the tongue. The probe was adjusted in a way that the medial grooving of the fricative [s] was easily observable as had been shown in Stone (1992). The overall angle was

---

[10] Seven speakers were initially recorded, but data from MS1 were excluded from analysis due to technical reasons.

similar across speakers. However, the direction of the ultrasound probe was not controlled precisely as in Stone et al. (1988) or in Stone & Lele (1992), mainly due to the morphological differences across speakers.[11] The ultrasound probe had a field of view of 92°; the depth was adjusted to have a maximum view of the tongue, necessitating the use of different frame rates for female and male speakers (82.1 fps for female speakers and 81.5 fps for male speakers).

The ultrasound data recorded were segmented manually in AAA using the corresponding audio, and traced manually with the help of the built-in tracing algorithm (Articulate

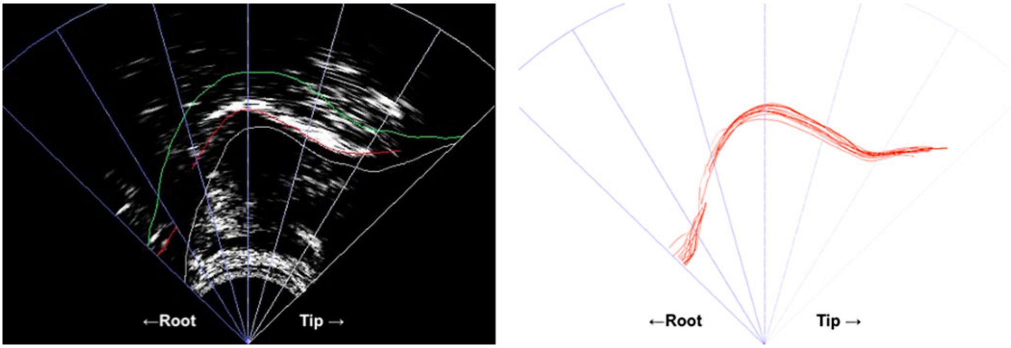

Figure 7 (Colour online) The left image shows the nearest midpoint mid-sagittal ultrasound image of one representative realisation of [z] in [pz^{31}], produced by FS3. The green line represents the palate and the grey line represents the 'min-tongue' in the built-in tracing algorithm of AAA. The red line represents the traced tongue surface. The right image presents mid-sagittal tongue traces of the nearest midpoint images from 11 tokens of [z] in [pz^{31} pz^{213} p^{h}z^{35} p^{h}z^{213} mz^{213} nz^{33} nz^{213} tsz^{213} ts^{h}z^{213} sz^{213} sz^{35}] produced by FS3, shown in the Spline Workspace of AAA. The tongue tip is on the right and the tongue root is on the left.

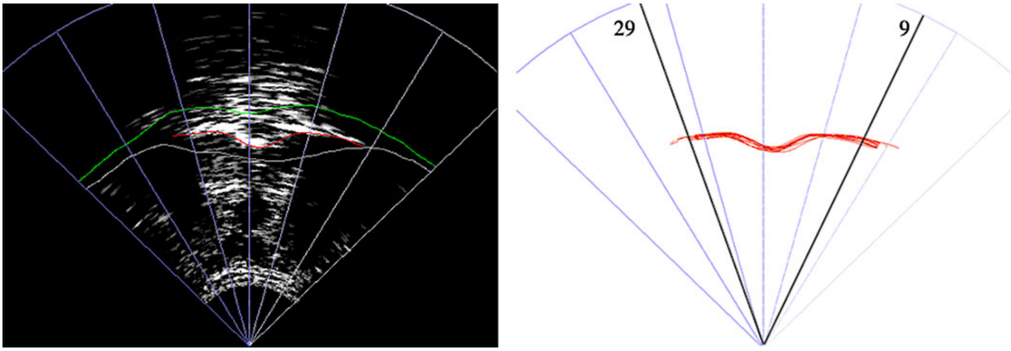

Figure 8 (Colour online) The left image shows the nearest midpoint coronal ultrasound image of one representative realisation of [z] in [pz^{31}], produced by FS3. The green line represents the palate and the grey line represents the 'min-tongue' in the built-in tracing algorithm of AAA. The red line represents the traced tongue surface. The right image presents coronal ultrasound tongue traces of the nearest midpoint image from 11 tokens of [z] in [pz^{31} pz^{213} p^{h}z^{35} p^{h}z^{213} mz^{213} nz^{33} nz^{213} tsz^{213} ts^{h}z^{213} sz^{213} sz^{35}] produced by FS3, shown in the Spline Workspace of AAA. Fan lines 9 and 29 are drawn to show the range of the data points considered as reliable.

---

[11] Stone et al. (1988) and Stone & Lele (1992) recorded one speaker in each of the two studies. Their setup had an ultrasound probe holder which had a perpendicular base. This holder is an inversed L-shaped metal shelf fixed on the floor with the ultrasound probe attached on the top. They were able to adjust the angle of the coronal plane by adjusting the direction of the probe with goniometers relative to the perpendicular metal shelf.

Instruments Ltd. 2012). Unless indicated otherwise, the tongue contours for [z] reported in this study were generated from all occurrences of this segment (i.e. from [pz^{31} pz^{213} p^{h}z^{35} p^{h}z^{213} mz^{213} nz^{33} nz^{213} tsz^{213} ts^{h}z^{213} sz^{213} sz^{35}]). The contours, corresponding to the nearest mid-point image of the segment based on the acoustic signals, were exported as polar coordinates. These polar coordinates were then analysed using smoothing-spline analysis of variance (SS ANOVA) in R using the *gss* package (Gu 2014). The results were analysed by speaker and presented in individual 92° fan diagrams.

For the SS ANOVA analysis of mid-sagittal tongue shapes, the entire visible tongue was included. One example is given in Figure 7. For the analysis of coronal tongue shapes, there was no anatomical reference to define a reliable range of the tracing, making the actual lateral edges of the tongue hard to determine. We considered the area within the junction of all traces of all segments to be reliable (e.g. for FS3, between the fan lines 9 and 29 in Figure 8, the angle between the two fan lines was approximatively 46°). This junction area could be easily identified as the tongue shapes of [i u ʉ a z] all passed through. We used traces within this area as this was the only reliable way to generalise comparisons between segments. These areas were defined individually based on data from each speaker.[12] The palatal traces were obtained by averaging six separate water-swallow trials, which were performed once at the beginning and once at the end of each session on both mid-sagittal and coronal planes.

3.3 Results

The generalisations based on the SS ANOVA in polar coordinates proved to be representative as the speakers had highly consistent tongue contours. We analysed the tongue shape for [z] and compared it to the tongue shapes for the fricative [s] and for the high vowels [i u ʉ]. The effect of context on the tongue configuration of [z] was shown by comparing labial and coronal contexts.

3.3.1 Comparing [z] to [s] and to [i u ʉ]

Figure 9 displays the mid-sagittal tongue contours for [z] compared to [s] and [i u ʉ]. For all six speakers, tongue configurations for [z] and [s] were similar to each other, while being different from those for the high vowels [i u ʉ].

The difference in tongue shapes between [z] and the high vowels was more marked at the tongue dorsum, the basic gesture involved in the production of vowels. The tongue dorsum has a convex shape for [i u ʉ], and a relatively flat shape for [z] and [s]. The tongue front displayed variable configurations depending on speakers. Recall however that the tip of the tongue, generally considered neutral for vowels, was often invisible in ultrasound tongue images because of the mandible shadow, even with the ultrasound probe pointed towards the front of the tongue.

A slight difference in the tongue configurations for [s] and [z] could be observed for some speakers (e.g. FS1, MS2 and MS5), who displayed a lower tongue body for [z] than for [s]. As was shown in the acoustic experiment, most of the apical vowels had frication noise in the first half of their duration, and the frication noise decreased while the formant structure became clearer. The lowered tongue body for [z] observed here may be related to this reduction of frication noise: As the tongue body became lower, the narrowed air channel was widened, and frication noise diminished. The tongue dorsum lowering may also be explained by the laryngeal contrast between voiceless [s] and voiced [z]. In English and Portuguese, for example, tongue root advancement/tongue body lowering was found for the voiced plosives /b d g/ compared to the voiceless counterparts /p t k/ (Westbury 1983, Ahn 2018).

---

[12] The areas defined by fan lines and the approximative field-of-views are listed hereafter. FS1: fan lines 10 and 29 (43.7°); FS5: fan lines 7 and 30 (52.9°); MS2: fan lines 11 and 30 (43.7°); MS3: fan lines 9 and 30 (48.3°); MS5: fan lines 11 and 29 (41.4°).

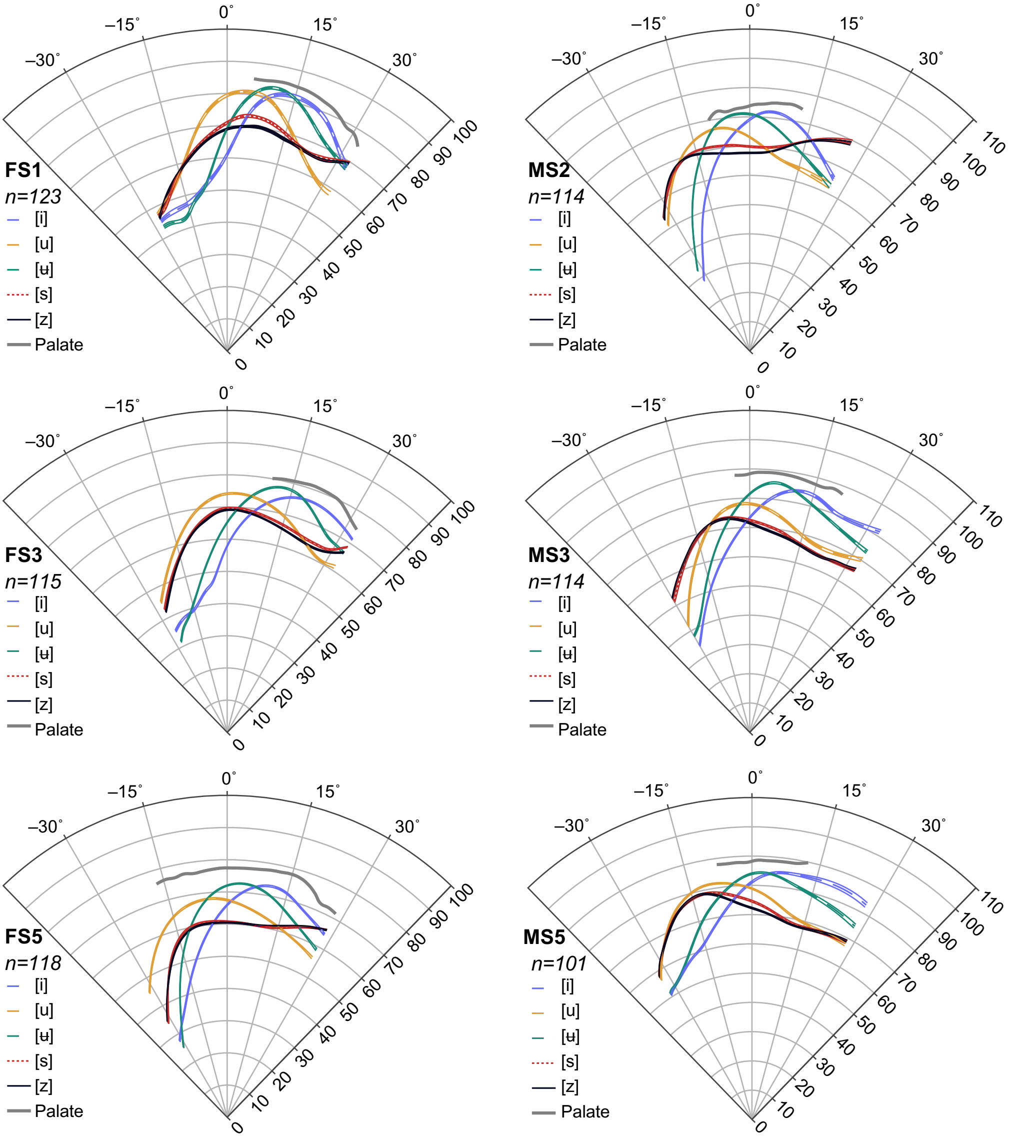

Figure 9 (Colour online) SS ANOVA splines of mid-sagittal ultrasound tongue contours of the six speakers, extracted in polar coordinates at the nearest midpoint image of each segment (tongue front on the right and tongue root on the left). Each speaker is presented in an individual fan diagram, female speakers on the left and male speakers on the right (*n* = number of tokens per speaker). The tongue contours are presented with 95% Bayesian confidence intervals. The total range of data shown is 92° (from −46° to +46°), corresponding to the field of view of the ultrasound probe. The radius axis is in millimetres.

This adjustment served to maintain the voicing of /b d g/ by enlarging the supraglottal cavity volume. In the case of the apical vowel [z], given that the same principle could be applied to voiced fricatives (Proctor, Shadle & Iskarous 2010), this articulatory adjustment could contribute to the following scenario: The tongue body lowering could facilitate the voicing of [z], which in turn could be responsible for the decreasing frication noise.

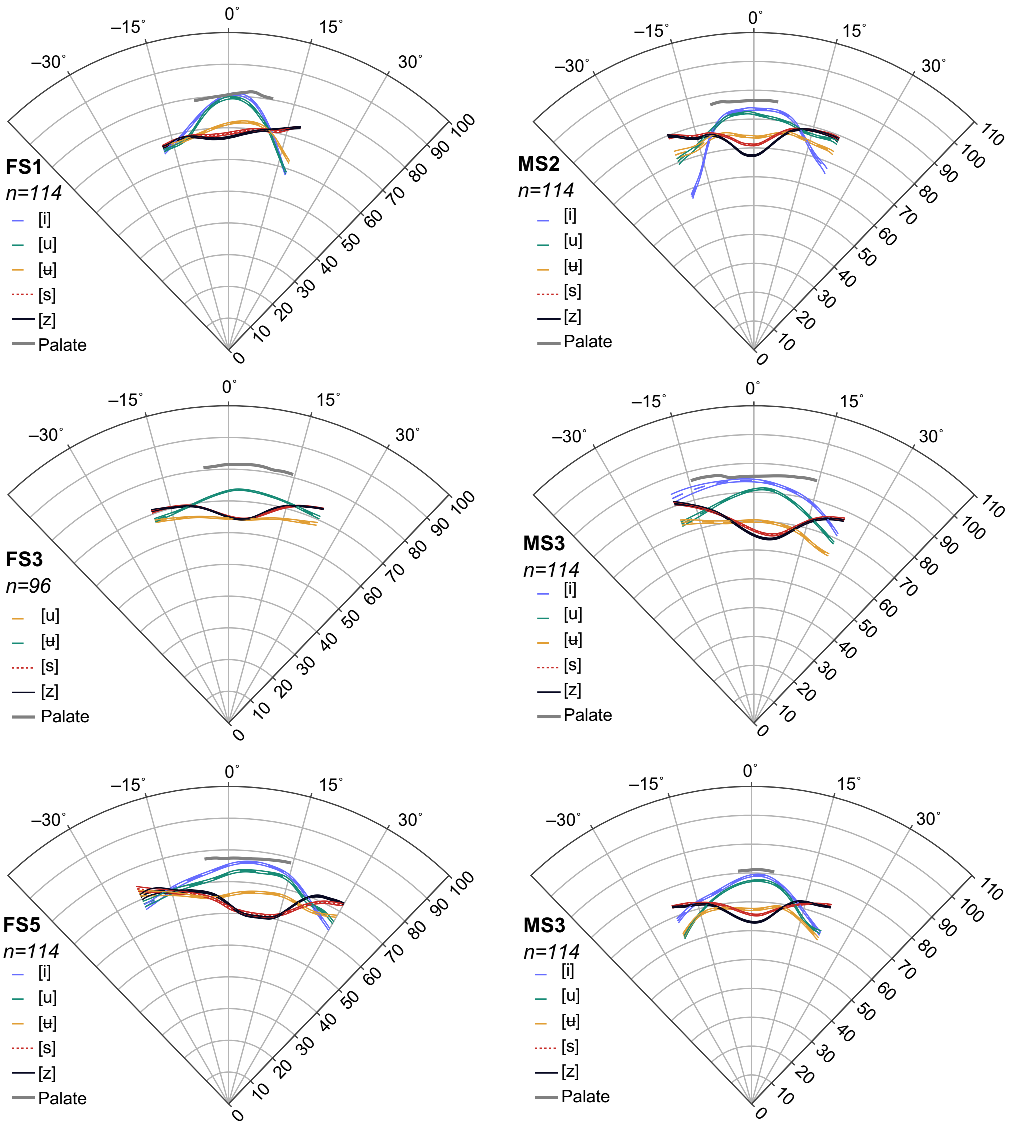

Figure 10 (Colour online) SS ANOVA splines of coronal ultrasound tongue contours of the six speakers, extracted in polar coordinates at the nearest midpoint image of each segment. Each speaker is presented in an individual fan diagram, female speakers on the left and male speakers on the right ($n =$ number of tokens per speaker). The tongue contours are presented with 95% Bayesian confidence intervals. The total range of data shown is 92° (from −46° to +46°), corresponding to the field of view of the ultrasound probe. The radius axis is in millimetres. The tongue shape of the vowel [i] from FS3 is not traced due to lack of consistent visibility in the data.

Figure 10 displays the coronal tongue contours for [z] compared to [s] and [i u ʉ]. Here too, the six speakers had virtually identical tongue shapes for [z] and [s], while both being different from those for the high vowels [i u ʉ]. Similar to [s], [z] displayed a medial-grooved tongue shape, with the medial part of the tongue being much lower than the lateral edges, signalling the presence of a narrowed air channel typical of fricatives. The two contours for [z s] could appear to be virtually identical as for FS3, or slightly different as for

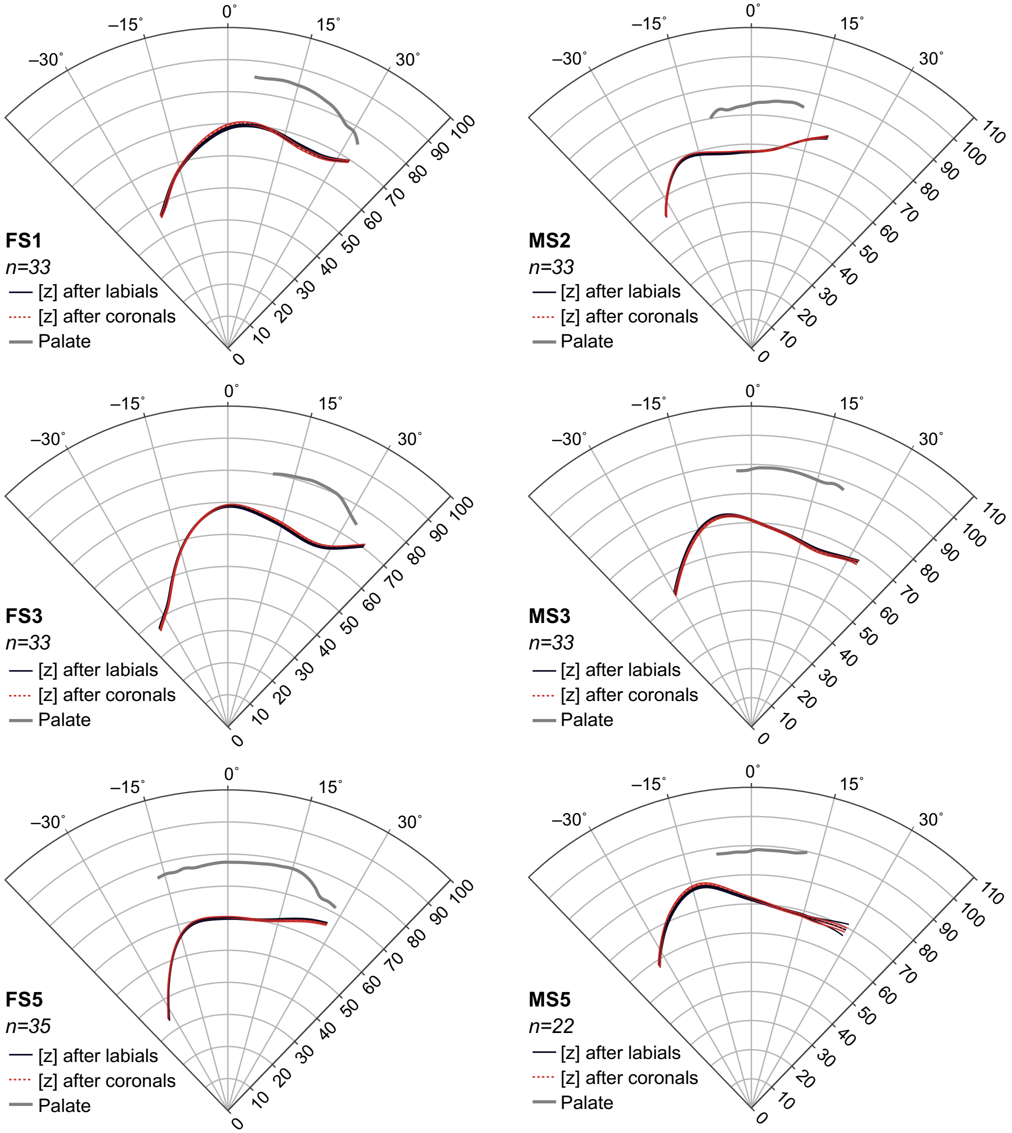

Figure 11 (Colour online) SS ANOVA splines of mid-sagittal ultrasound tongue contours of the six speakers, extracted in polar coordinates at the nearest midpoint image of apical vowel [z], grouped by labial or coronal onsets. Each speaker is presented in an individual fan diagram, female speakers on the left and male speakers on the right (n = number of tokens per speaker). The tongue contours are presented with 95% Bayesian confidence intervals. The total range of data shown is 92° (from −46° to +46°), corresponding to the field of view of the ultrasound probe. The radius axis is in millimetres.

MS2 and MS5. For the latter two speakers, the medial-grooving was deeper in [z] than in [s]. This medial-groove deepening is most probably related to the tongue dorsum lowering exhibited by the same speakers and could be related to the disappearance of frication noise.

FS1 was a notable exception to this general pattern. Interestingly, this speaker did not produce a significant medially-grooved tongue shape for fricative [s] either (see also Figure 12

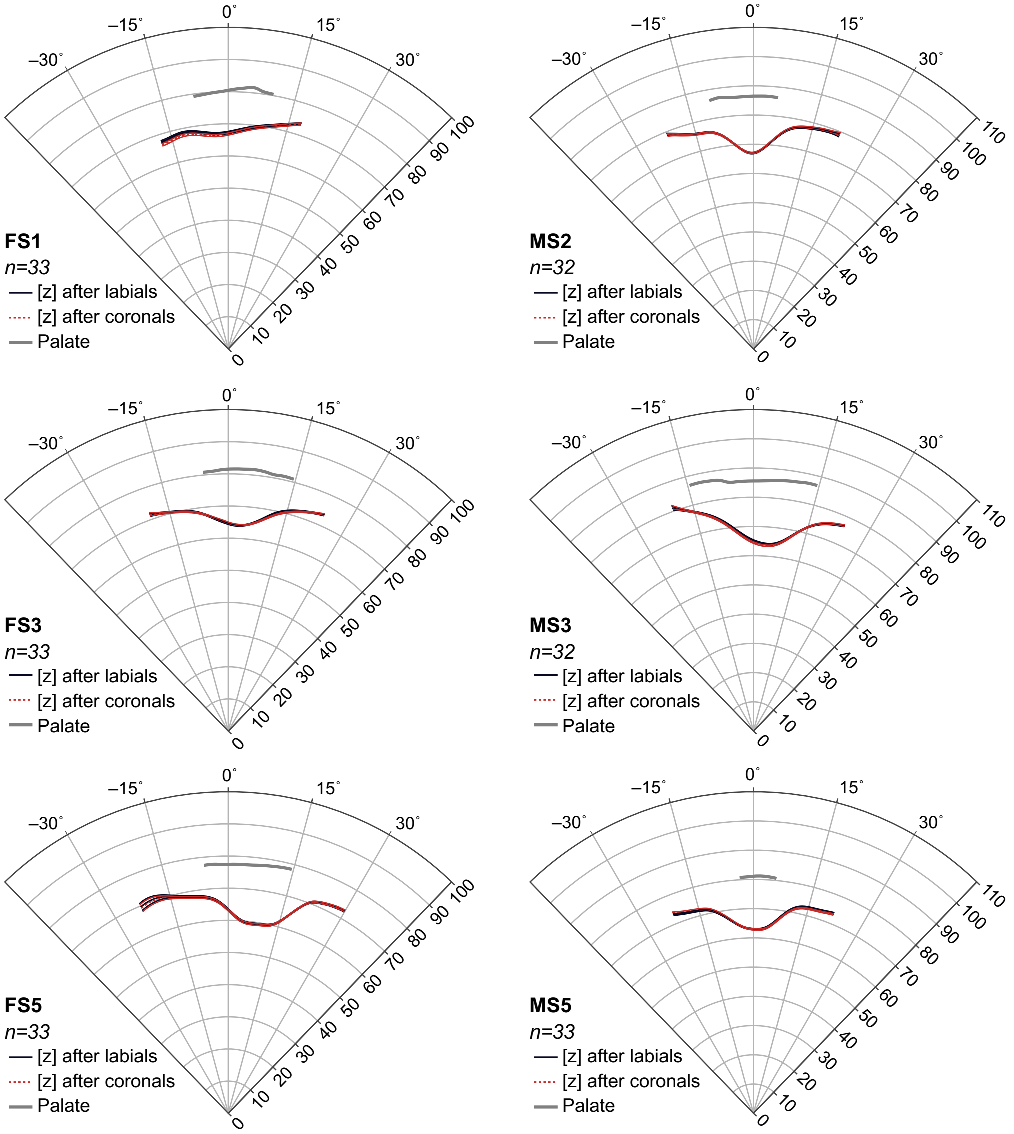

Figure 12 (Colour online) SS ANOVA splines of coronal ultrasound tongue contours of the six speakers, extracted in polar coordinates at the nearest midpoint image of apical vowel [z], grouped by labial or coronal onsets. Each speaker is presented in an individual fan diagram, female speakers on the left and male speakers on the right ($n =$ number of tokens per speaker). The tongue contours are presented with 95% Bayesian confidence intervals. The total range of data shown is 92° (from -46^{0} to $+46^{0}$), corresponding to the field of view of the ultrasound probe. The radius axis is in millimetres.

below).[13] Figure 10 also shows that [u] had a slightly medially-grooved tongue shape (namely for MS2). This slight grooving, also reported by Stone et al. (1988) and Stone & Lele (1992), was much shallower than for [z], and probably reflected the natural grooving of the tongue rather than an intended tongue gesture.

---

[13] A possible explanation for this could be that the ultrasound probe was pointed towards the dorsum of the tongue, rather than the front, suggesting that FS1 could have medial-grooving at tongue tip and tongue front, which was not captured.

As reported in previous studies (e.g. Dixit & Hoffman 2004, Liker, Horga & Mildner 2012), the lingual–palatal contact for [z] and [s] was not always symmetrical (FS3 being an exception). In addition, the lateral edges for [z] seemed to be more raised than for [s] (for FS5 and MS5). This could be related to the articulatory difference between the two segments (Stone et al. 1992, Dagenais 1994, Liker et al. 2012, Skarnitzl, Sturm & Machac 2013, Kochetov 2014): [z] has more anterior lingual–palatal contact than [s], and the anterior medial groove is narrower in [z]. The lateral edges of the tongue were not symmetrical during the articulation of high vowels. For FS1, FS5 and MS3, one side of the lateral edges seemed to move less than the other side, which could be a case of lateral bracing that had been reported for other languages (Cheng, Schellenberg & Gick 2017, Gick et al. 2017).

3.3.2 [z] preceded by labial and coronal consonants

We compared the smoothing splines of [z] in two contexts: after labials [p p^h m] and after coronal sibilants [s ts ts^h]. The aim was to determine whether the [s]-like tongue configuration for [z] was observed regardless of the nature of the preceding consonant. Results, presented in Figure 11 for the mid-sagittal plane, showed that the [s]-like tongue shape for [z] was implemented when [z] was preceded by both sibilants and non-sibilants. The same comparisons were conducted in the coronal plane. The results are shown in Figure 12. Here too, the medial-grooved [s]-like tongue shape for [z] was consistent regardless of the nature of the preceding consonant. The fact that [z] maintained the same tongue configuration, whether preceded by sibilants or not, is mirroring the results obtained in the acoustic study, strengthening the argument that this tongue shape is an inherent property of [z] in JHC.

3.4 Interim summary

The analysis of the ultrasound data showed that the tongue shape for [z] was virtually identical to that of the alveolar sibilant [s]. This similarity was observed in both mid-sagittal and coronal planes. The medially-grooved tongue shape of [z] was particularly important, as it was a fundamental indicator of a narrowed air channel which is typical for a fricative gesture. The fricative-like tongue shape of [z] was observed in both sibilant and non-sibilant contexts. This was another important finding since it showed that the tongue configuration of [z] was not a mere consequence of gestural overlap between homorganic sibilants and apical vowels. While the tongue configuration of [z] was similar to that of [s], it was very different from the configuration of high vowels. This difference was observed on the mid-sagittal plane, most notably regarding the arching of the dorsum typical of vowel articulation. It was even more evident on the coronal plane, with [z] displaying a medial-grooved configuration unlike genuine vowels.

It is important, however, to note that the tongue configurations of [z] and [s] displayed some differences. On the mid-sagittal plane, the tongue dorsum was positioned lower in [z] than in [s] for almost all speakers. And on the coronal plane, a deeper medial grooving for [z] was observed for some speakers. The medial-groove deepening and the tongue dorsum lowering indicated the enlargement of the narrowed air channel that could be responsible for the disappearance of the frication noise reported in the acoustic study.

4 General discussion and conclusion

This study investigated the acoustics and articulation of the apical vowel in Jixi-Hui Chinese (JHC). The results obtained showed that this segment had characteristics of an alveolar consonant. At the acoustic level, it exhibited frication noise for all the speakers recorded. This frication was systematically superposed on voicing. The presence of frication was independent of the nature of the preceding consonant, as it occurred when preceded by sibilants as well as by non-sibilants. The ultrasound data showed that the tongue configuration of [z] resembled that of the alveolar sibilant [s]. This similarity was observed in mid-sagittal planes with the two segments displaying similar (though not entirely identical) tongue configurations. Comparably, on the coronal plane, [z] displayed a medial-grooved tongue shape similar to the fricative [s]. The fricative-like tongue shape of [z] was also observed independently of the nature of the preceding segment.

While having acoustic and articulatory characteristics of a fricative consonant, the JHC [z] almost never displayed frication throughout its entire duration. Although present in 88% of the tokens, frication was implemented variably, from less than half to more than half of the segment. The gradual disappearance of noise gave way to a more and more visible formant structure. The dynamic behaviour of [z] during its time course made it often realised as a hybrid segment with the first part being fricative-like and the second part being more approximant-like. This ultimately raises the question of how to best define this segment: is it a fricative or an approximant? In the rest of this article, we briefly discuss the two alternatives.

An approximant is phonetically defined as a segment lacking turbulent noise (IPA 1999: 8; Martínez-Celdrán 2004; Ladefoged & Johnson 2011: 15). Analysing the JHC apical vowel as an approximant requires an explanation for the presence of frication noise, akin to what has been proposed for Standard Chinese (SC) apical vowels. Given that in SC, apical vowels occur only after sibilants, the presence of frication noise can be readily explained by gestural overlap (Lee-Kim 2014). The JHC case, however, is different: [z] appears after sibilants as well as after non-sibilants [p p^h m n]. These non-sibilants have objectively no frication noise to propagate into the following segment. This being said, analysing JHC [z] as a fricative requires an account for its hybrid configuration. Our argument is that this is a consequence of two interacting constraints: a physical one related to the incompatibility of voicing and frication, and a structural one related to the role [z] plays within syllable structure.

The phonetic implementation of fricative [z] requires a compromise between voicing and frication. This is because of the incompatibility between two aerodynamic requirements: an increase in the intraoral pressure for frication and a certain transglottal pressure difference to maintain voicing. As Ohala (1983: 201–202) put it: ‘To the extent that the segment retains voicing, it may be less of a fricative, and if it is a good fricative it runs the risk of being devoiced’. Considering this aerodynamic voicing constraint, the JHC [z], and voiced fricatives in general, should imply a dichotomy between a voiceless fricative and a voiced approximant. In many languages, this incompatibility is achieved at the expense of voicing. This is the case in Tashlhiyt where geminate fricatives rarely maintain voicing throughout (Ridouane 2007), or in English (Smith 1997) and Hungarian (Bárkányi & Kiss 2010) where /z/ and /v/ tend to devoice. In French, the variation of /ʁ/ leads to a continuum between unvoiced fricative and voiced approximant (Gendrot, Kühnert & Demolin 2015). In JHC, the compromise is never achieved at the expense of voicing. A possible reason for this could be related to the role [z] plays within the syllable structure. As a tone-bearing unit, [z] has to be voiced throughout to carry the tone. Being obligatorily voiced, it can hardly display strong frication noise throughout its full duration. This is all the more so given that [z], similar to other syllable nuclei, is particularly long (see Table 5 above). But why should frication be maintained? Why not produce a voiced approximant? A plausible answer, grounded on both structural and acoustic-auditory properties, is that the JHC [z] needs to differentiate itself from the close central vowel [ʉ]. Since the close central vowel [ʉ] has a formant structure which is similar to [z], the loss of frication noise may put the distinction between the two segments at risk.

While it favours a fricative account, our analysis of [z] in JHC does not totally rule out an approximant account. Indeed, one may object that it is more natural to have an approximant as a tone bearing unit, rather than a voiced fricative. Future studies will have to complement the present work in order to evaluate the proposed analyses and to increase our understanding of the nature of apical vowels in JHC in particular and in Chinese languages in general. One avenue for future studies concerns the perceptual cues to apical vowels and the role played by frication noise. Related to this is the assumption that frication noise may serve as an enhancing feature in some Chinese languages (Kong et al. 2019). A possible scenario could be that this frication was first introduced as an enhancing attribute to maximise the contrast between sibilants. Depending on the phonemic inventory this attribute took on a distinctive function in Chinese languages where [z] occurs not only following sibilants. This is clearly speculative although cases where enhancement gestures become primary (when the defining gestures are weakened or lost) have already been reported in literature. In many Bantu languages, the distinction between upper and lower vowels has disappeared and assibilation took the distinctive role (Mpiranya 1997, Clements & Ridouane 2006). Assibilation, which was an enhancing gesture in the historical development of these vowels, preserves the distinction between words with earlier upper high vowels and those with earlier lower high vowels. A similar trajectory may have been followed in the historical sound change of apical vowels in Chinese languages.

Acknowledgements

The authors are grateful to the participants of this study. We would like to thank the Associate Editor Dr Alexei Kochetov and two anonymous reviewers for their valuable comments and careful proofreading, which greatly improved the quality of the manuscript. All errors remain our own. This work has partially benefited from a government grant managed by the Agence Nationale de la Recherche under the Investissements d'Avenir programme with the reference ANR-10-LABX-0083 (contributing to the IdEx University of Paris – ANR-18-IDEX-0001, LABEX-EFL).

Appendix

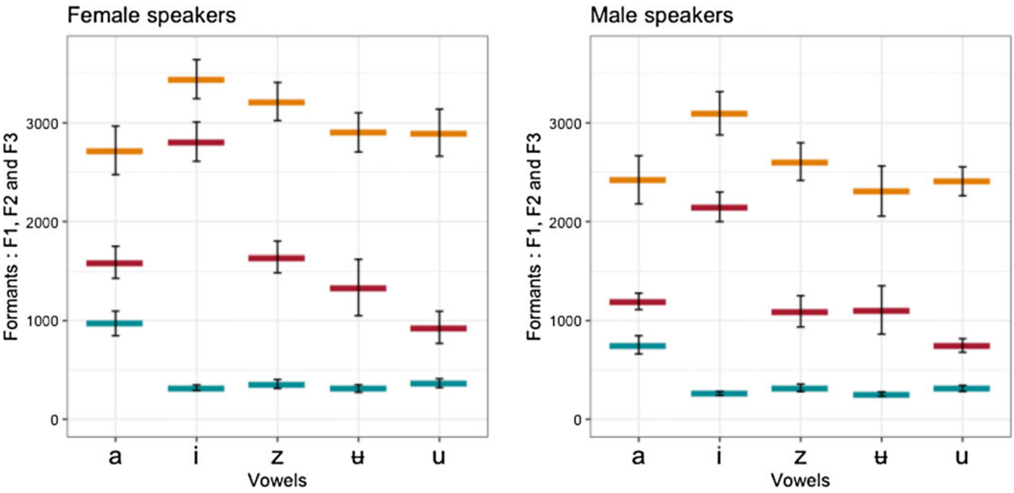

Figure A1 (Colour online) The structure of F1, F2 and F3 values (in Hz) of [i u a ʉ z] in JHC.

Table A1 Statistical results of linear mixed-effects analyses conducted on formant values (Bark-scaled) of [z u ʉ] in JHC; standard deviations in parentheses.

	N of observations		Fixed effects estimates (Bark)		
			Intercept[u] (se)	Segment[ʉ] (se)	Segment[z] (se)
FS	633	F1	3.67 (0.12)	−0.53 (0.15)	−0.08 (0.15)
		F2	8.10 (0.32)	2.14 (0.20)	3.56 (0.40)
		F3	15.40 (0.14)	0.06 (0.08)	0.70 (0.12)
MS	619	F1	3.19 (0.12)	−0.63 (0.08)	0.03 (0.06)
		F2	6.87 (0.07)	2.18 (0.15)	2.17 (0.25)
		F3	14.20 (0.10)	−0.29 (0.17)	0.55 (0.08)

FS = female speakers; MS = male speakers

References

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71, 451–467.

Articulate Instruments Ltd. 2008. *Ultrasound Stabilisation Headset User's Manual: Revision 1.5*. Edinburgh: Articulate Instruments Ltd.

Articulate Instruments Ltd. 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh: Articulate Instruments Ltd.

Bárkányi, Zsuzsanna & Zoltán Kiss. 2010. A phonetic approach to the phonology of v: A case study from Hungarian and Slovak. In Fuchs et al. (eds.), 103–142.

Baron, Stephen. 1974. On the tip of many tongues: Apical vowels across Sino-Tibetan. *7th International Conference on Sino-Tibetan Language and Linguistic Studies*, Atlanta, GA.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using *lme4*. *Journal of Statistical Software* 67(1), 1–48.

Boersma, Paul & David Weenink. 2018. PRAAT: Doing phonetics by computer. Version 6.0.43, http://www.praat.org (accessed March 2018).

Bombien, Lasse. 2006. Voicing alterations in Icelandic sonorants: A photoglottographic and acoustic analysis. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 37, 63–82.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le Maître phonétique* 45, 24–27.

Chao, Yuen-Ren. 1961. *Mandarin primer: An intensive course in spoken Chinese*. Oxford: Harvard University Press.

Chao, Yuen-Ren. 1962. Jìxī Lǐngběi Yīnxì [The phonology of Jixi Lingbei]. *Bulletin of the Institute of History and Philology* 34(1), 27–30.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California

Chao, Yuen-Ren & Shifeng Yang. 1965. Jìxī Lǐngběi Fāngyán [The dialect of Jixi Lingbei]. *Bulletin of the Institute of History and Philology* 36(1), 11–113.

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

Cheng, Lauretta, Murray Schellenberg & Bryan Gick. 2017. Cross-linguistic bracing: A lingual ultrasound study of six languages. *Canadian Acoustics* 45(3), 186–187.

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

Clements, Georges Nick & Rachid Ridouane. 2006. Distinctive feature enhancement: A review. *ITRW on Experimental Linguistics*, 97–100. Athens, Greece.

Dagenais, Paul A., Leah Lorendo & Martin McCutcheon. 1994. A study of voicing and context effects upon consonant linguapalatal contact patterns. *Journal of Phonetics* 22(3), 225–238.

Dell, François. 1994. Consonnes à prolongement syllabique en Chine. *Cahiers de linguistique – Asie orientale* 23(1), 87–94.

Dixit, R. Prakash & Paul R. Hoffman. 2004. Articulatory characteristics of fricatives and affricates in Hindi: An electropalatographic study. *Journal of the International Phonetic Association* 34(2), 141–159.

Duanmu, San. 2007. *The phonology of standard Chinese*, 2nd edn. New York: Oxford University Press.

Faytak, Matthew. 2018. *Articulatory uniformity through articulatory reuse: Insights from an ultrasound study of Sūzhōu Chinese*. Ph.D. dissertation, University of California, Berkeley.

Faytak, Matthew & Susan S. Lin. 2015. Articulatory variability and fricative noise in apical vowels. *18th International Congress of Phonetic Sciences* (ICPhS XVIII), Glasgow, UK.

Fuchs, Susanne, Martine Toda & Marzena Żygis (eds.). 2010. *Turbulent sounds: An interdisciplinary guide*, vol. 21. Berlin & New York: Walter de Gruyter.

Gendrot, Cédric, Barbara Kühnert & Didier Demolin. 2015. Aerodynamic, articulatory and acoustic realization of French /ʁ/. *18th International Congress of Phonetic Sciences* (ICPhS XVIII), Glasgow, UK.

Gick, Bryan, Blake Allen, François Roewer-Després & Ian Stavness. 2017. Speaking tongues are actively braced. *Journal of Speech, Language, and Hearing Research* 60(3), 494–506.

Gong, Xun. 2016. A phonological history of Amdo Tibetan rhymes. *Bulletin of the School of Oriental and African Studies* 79, 347–374.

Gordeeva, Olga & James Scobbie. 2010. Preaspiration as a correlate of word-final voice in Scottish English fricatives. In Fuchs et al. (eds.), 167–207.

Gu, Chong. 2014. Smoothing Spline ANOVA Models: R Package gss. *Journal of Statistical Software* 58(5), 1–25.

Hirata, Shoji. 1998. *Huīzhōu fāngyán yánjiū* [Study on Huizhou dialects]. Tokyo: Kohbun Press.

Hou, Chao. 2009. Héféi fāngyán dàicāyuányīn de shíyàn yánjiū [An experimental study on [ʅ] in Hefei dialect]. *Journal of School of Chinese Language and Culture* (Nanjing Normal University) 1, 172–177.

Howie, John M. 1976. *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.

Hu, Fang. 2007. Lùn Níngbō fāngyán hé Sūzhōu fāngyán qiángāo yuányīn de qūbié tèzhēng – Jiāntán gāoyuányīn jìxù gāohuà xiànxiàng [On the distinctive features of the high front vowels in Ningbo and Suzhou Wu dialects with reference to sound changes of high vowels]. *Zhongguo Yuwen* 5, 455–465.

IPA [International Phonetic Association & International Phonetic Association Staff]. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

IPA [International Phonetic Association]. 2010. The principles of the International Phonetic Association 1949. *Journal of the International Phonetic Association* 40(3), 299–358.

Ito, Mabo R. & Robert W. Donaldson. 1971. Zero-crossing measurements for analysis and recognition of speech sounds. *IEEE Transactions on Audio and Electroacoustics* 19(3), 235–242.

Jacques, Guillaume & Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4), 468–498.

Karlgren, Bernard. 1915. *Études sur la phonologie chinoise*. Uppsala: K. W. Appelberg.

Kochetov, Alexei. 2014. Voicing and tongue-palate contact differences in Japanese. *Journal of the Phonetic Society of Japan* 18(2), 63–76.

Kong, Huifang, Shengyi Wu & Mingxing Li. 2019. Héféi Huà Shéjiān Yuányīn de Mócā Xìngzhi jí Gǎnzhī Zēngqíang Lǐlùn Jiědú [The frication property of apcial vowels in Hefei Chinese and its perceptual enhancement account]. *Yuyan Yanjiu* 39(1), 23–33.

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics*, 6th edn. Boston, MA: Wadsworth.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1), 109–112.

Lee-Kim, Sang-Im. 2014. Revisiting Mandarin apical vowels: An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3), 261–282.

Li, Rong, Zhenghui Xiong & Zhenxing Zhang. 1987. *The language atlas of China*. Hong Kong: Longman.

Liker, Marko, Damir Horga & Vesna Mildner. 2012. Electropalatographic specification of Croatian fricatives /s/ and /z/. *Clinical Linguistics & Phonetics* 26(3), 199–215.

Ling, Feng. 2009. *A phonetic study of the vowel system in Suzhou Chinese*. Ph.D. dissertation. City University of Hong Kong.

Luo, Changpei. 1936. Jìxī fāngyīn Shùlüè [Brief introduction to Jixi dialect]. In 2008 reprint: Gao Gengsheng & Sun Hongkai (eds.), *Luó Chángpéi Wénjí* [The collected linguistic works of Luo Changpei], vol. 9, 175–204. Jinan: Shandong jiaoyu chubanshe.

Martínez-Celdrán, Eugenio. 2004. Problems in the classification of approximants. *Journal of the International Phonetic Association* 34(2), 201–210.

Michaud, Alexis. 2008. Phonemic and tonal analysis of Yongning Na. *Cahiers de linguistique – Asie orientale* 37(2), 159–196.

Mpiranya, Fidèle. 1997. Spirantisation et fusion vocalique en bantu: essai d'interprétation fonctionnelle. *Linguistique Africaine* 18, 51–77.

Ohala, John. J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer.

Proctor, Michael I., Christine H. Shadle & Khalil Iskarous. 2010. Pharyngeal articulation in the production of voiced and voiceless fricatives. *The Journal of the Acoustical Society of America* 127(3), 1507–1518.

Pullum, Geoffrey K. & William A. Ladusaw. 1996. *Phonetic symbol guide (Second Edition)*. Chicago, IL: The University of Chicago Press.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ridouane, Rachid. 2007. Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association* 37(2), 119–142.

Shao, Bowei & Rachid Ridouane. 2021. 'Apical vowels' are not vowels: Acoustic and ultrasound evidence from Jixi-Hui Chinese. *Proceedings of the 12th International Seminar on Speech Production*, 110–113.

Shi, Feng, Gang Peng & Yi Liu. 2015. Vowel distribution in isolated and continuous speech. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 459–473. Oxford: Oxford University Press.

Shosted, Ryan K. 2006. *The aeroacoustics of nasalized fricatives*. Ph.D. dissertation, University of California, Berkeley.

Skarnitzl, Radek, Pavel Sturm & Pavel Machac. 2013. The phonological voicing contrast in Czech: An EPG study of phonated and whispered fricatives. *14th Annual Conference of the International Speech Communication Association*, Lyon, France, 3191–3195.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25(4), 471–500.

Stone, Maureen. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics* 20, 253–270.

Stone, Maureen & Subhash Lele. 1992. Representing the tongue surface with curve fits. *Second International Conference on Spoken Language Processing*, Banff, Alberta, Canada.

Stone, Maureen, Thomas H. Shawker, Thomas L. Talbot & Alan H. Rich. 1988. Cross-sectional tongue shape during the production of vowels. *The Journal of the Acoustical Society of America* 83(4), 1586–1596.

Svantesson, Jan-Olof. 1984. Vowels and diphthongs in Standard Chinese. *Working Papers of Lund University, Department of Linguistics* 27, 209–235.

Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88(1), 97–100.

Trubetzkoy, Nicolas. S. 1969. *Principles of phonology*. Berkley, CA & Los Angeles, CA: University of California Press.

Wang, Shuangcheng. 2006. Qīnghǎi fāngyán yuányīn [i] de shéjiānhuà yīnbiàn [Apicalisation sound change of the Qinghai dialects' [i] vowel]. *Zhongguo Yuwen* 4, 359–363.

Wang, Shuangcheng. 2010. Ānduō zàngyǔ [i] de shéjiānhuà jí qí lèixíngxué yìyì [Sound Change of [i] in Amdo Tibetan and its typological meaning]. *Studies in Language and Linguistics* 30(2), 122–127.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America* 73(4), 1322–1336.

Wiese, Richard. 1997. Underspecification and the description of Chinese vowels. *Studies in Chinese phonology*, 219–249.

Wu, Wei. 1995. Héféi huà "-i", "-y" yīnjié shēngyùnmǔ qiánhuà tàntǎo [Exploration of the fronting of '-i' and '-y' rhymes in the Hefei dialect]. *Yuyan yanjiu* 3, 58–60.

Ye, Xiangling. 1993. *Sūzhōu fāngyán cídiǎn* [Dictionary of Suzhou dialect]. Nanjing: Jiangsu jiaoyu chubanshe.

Yu, Alan. 1999. Aerodynamic constraints on sound change: The case of syllabic sibilants. *The Journal of the Acoustical Society of America* 105(2), 1096–1097.

Zee, Eric & Wai-Sum Lee. 2007. Vowel typology in Chinese. *16th International Congress of Phonetic Sciences* (ICPhS XVI), Saarbrücken, Germany, 1429–1432.

Zhao, Rixin. 1989. Ānhuī Jìxī fāngyán yīnxì tèdiǎn [Phonological characters of Jixi dialect from Anhui Province]. *Fangyan* 2, 125–130.

Zhao, Rixin. 2003. *Jìxī fāngyán cídiǎn* [Dictionary of Jixi dialect]. Nanjing: Jiangsu jiaoyu chubanshe.

Zhao, Rixin. 2007. Hànyǔ fāngyán zhōng de [i] > [ʅ] [Sound change of [i] > [ʅ] in Chinese dialects]. *Zhongguo yuwen* 1, 46–54.

Zhou, Dianfu & Zongji Wu. 1963. *Pǔtōnghuà fāyīn túpǔ* [Articulatory diagrams of Standard Chinese]. Beijing: Shangwu yinshuguan.

Zhu, Xiaonong, 2004. Hànyǔ yuányīn de gāodǐngchūwèi [Sound changes of high vowels in Chinese dialects]. *Zhongguo yuwen* 5, 440–451.