Articulation of Mandarin Sibilants: a multi-plane realtime MRI study

Michael Proctor, Li Hsuan Lu, Yinghua Zhu, Louis Goldstein, Shrikanth Narayanan*

Speech Production and Articulation Knowledge Group
University of Southern California

http://sail.usc.edu/span michael.proctor@uws.edu.au

Abstract

Production of sibilant fricatives by four speakers of Mandarin Chinese was examined using real-time Magnetic Resonance Imaging (rtMRI). Data were reconstructed at 33.18 frames per second in the midsagittal plane, and additionally acquired in a coronal plane intersecting the vocal tract behind the alveolar ridge. Although individual speakers differed in the details of articulation, each of the three sibilants was found to be distinguished by a characteristic constriction location, groove geometry, and tongue shaping behind the primary constriction.
Index Terms: Mandarin Chinese, sibilant fricatives, MRI

1. Introduction

Standard Chinese (SC, or Mandarin) contrasts three voiceless sibilant fricatives, represented in Pinyin using the orthographs s - sh - x.[1] Each of these segments also contrasts with aspirated (c - ch - q), and non-aspirated (z - zh - j) affricates produced at the same places of articulation [2, 3].

Descriptions of the ways that these consonants are produced differ. The sibilants s - z - c are variously characterized as dental [2, 4] or alveolar [3, 5, 6]. Fricatives sh - zh - ch are most commonly described as retroflex [1, 2] or apical palato-alveolar [7]. [3] characterize this series as 'flat post-alveolar,' noting that they are produced as laminals, rather than apicals, with a different tongue shape to that observed in Dravidian or Hindi retroflex segments. The third series of sibilants is typically classified as palatal, but articulatory and phonological characterizations of these segments also vary: [8] describe x as 'palatalized post-alveolar,' while [6] concludes that "it's meaningless to characterize [ɕ] in terms of a distinct place of articulation." For many speakers of Beijing Standard Chinese, [2] reanalyzes x - j - q as palatalized allophones [s^{j}] - [ts^{j}] - [ts^{hj}] of the dentals s - z - c, rather than the separate series of palatal phonemes /ɕ/ - /tɕ/ - /tɕ^{h}/ used in other SC varieties. The situation is complicated by sociolinguistic factors, including the perception of retroflexed initials as prestigious [7], and diachronic shifts amongst the sibilant series [9].

Differences among speakers and registers are no doubt responsible for much of this variability [7, 6]. Lack of consensus on the characterization of Mandarin sibilants also appears to be due in part to a lack of articulatory data and difficulties interpreting data acquired using different modalities. The acoustic characteristics of Mandarin fricatives have been well documented [10, 11, 5], but fewer studies have examined sibilant articulation, and to the best of our knowledge, none have used large subject populations. [8] used x-ray and palatography to examine production of sibilants by three speakers of Mandarin. [6] presents EPG data from four speakers, and EMA data from two speakers of Standard Chinese; no further information about their language backgrounds is provided.

More detailed and more diverse data are required to better understand the phonetic and phonological properties of these segments. The goal of this study is to examine the dynamic articulation of Mandarin sibilants across vowel contexts – using a novel combination of imaging planes providing greater coverage of the vocal tract – to gain more insights into their goals of production.

2. Method

Two male and two female adult native speakers of Standard Chinese participated in the study (Table 1). All participants were born in mainland China and raised in a Mandarin-speaking environment to first-language speakers of Standard Chinese. Subject W1 is an L2 speaker of Cantonese and also speaks some Taiwanese. All four subjects have some proficiency in English as a second or third language, and have lived as adults in Los Angeles for up to 4 years.

ID	SEX	AGE	BORN	RAISED
M1	M	26	Nanning, GX	Nanning, GX
M2	M	24	Zhengzhou, HA	Urumuqi, XJ
W1	F	24	Shenzhen, GD	Shenzhen, GD
W2	F	23	Hangzhou, ZJ	Hangzhou, ZJ

Table 1: ***Demographics of Study Participants***.

Sibilant fricatives s - sh - x were elicited using a set of nonsense phrases presented in Pinyin (Table 2). Each target consonant was elicited word-initially in three maximally-contrastive vowel contexts. 1st Tone (*yīnpíng*: high-level) was indicated on the target word and on the labial-initial words of the elicitation phrase, to eliminate tone as a source of phonetic variation.

TARGET	HIGH FRONT	LOW	HIGH BACK
s	bī sɪ bī	bā sā bā	bū sū bū
sh	bī shī bī	bā shā bā	bū shū bū
x	bī xɪ bī	bā xā bā	bū xū bū

Table 2: ***Stimuli used in rtMRI experiment***. *Pinyin pseudo-phrases used to elicit Mandarin sibilant fricative initials.*

* Now at MARCS Institute/School of Humanities and Communication Arts, University of Western Sydney

[1] Many different sets of phonetic symbols have been used to represent these phonemes [1, 2]. Because it is not yet clear what the most appropriate phonemic representations of these segments might be, Pinyin orthographs will be used to refer to Mandarin consonants throughout this paper. IPA and articulatory characterizations of the sibilants produced by the speakers in this study are presented in the discussion.

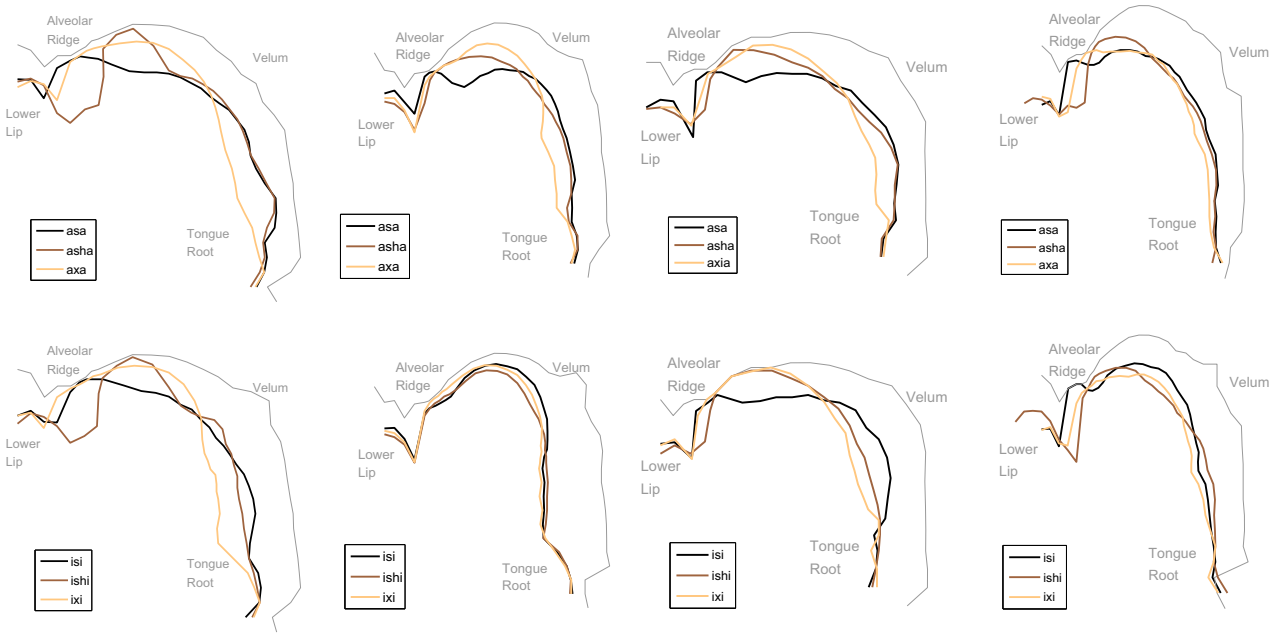

Figure 1: ***Mid-consonantal articulatory targets for Mandarin sibilants s - sh - x.*** *Lingual postures captured at point of maximal coronal constriction, compared for four speakers of Standard Chinese, from left-to-right: subject M1, M2, W1 and W2. Top row: sibilants produced in a low vowel context. Bottow row: sibilants produced in a high-front vowel context. Lingual outlines show mean posture calculated over three frames f_{i-1}, f_i, f_{i+1}, centered on the mid-consonantal frame of interest f_i.*

2.1. Image Acquisition

Data were acquired using a rtMRI protocol developed specifically for the dynamic study of speech production [12]. The subject's upper airway was imaged in the midsagittal plane with spatial resolution of 68 x 68 pixels over a 200 x 200 mm field of view. New image data were acquired at a rate of 18.43 frames per second, and reconstructed as 33.18 frames/sec. video, using a sliding window technique [13]. Audio was simultaneously recorded at 20 kHz inside the scanner, and subsequently noise-reduced [14]. The resulting companion video and audio recordings allow for dynamic visualization of the entire midsagittal plane of the subject's vocal tract during speech.

2.2. Articulatory Analysis

MRI data were loaded into a custom graphical user interface designed for the synchronization of companion audio and video recordings, and inspection and analysis of time-aligned video, audio and spectra [15, 16]. For each token, the articulatory target of the fricative was identified by locating the image frame in which maximal coronal constriction was observed. A 450 to 600 msec interval centered on the mid-consonant frame was identified, corresponding to the entire interval of fricative production: the V.CV sequence beginning at the articulatory midpoint of the preceding vowel, and ending in the middle of the tautosyllabic vowel (where the tongue dorsum was observed to be maximally static at the target vocalic constriction).

Midsagittal tongue posture in each frame was captured by automatically identifying air-tissue boundaries [15], and manually correcting the tongue outline against the MR image where the algorithm failed to locate the edges of lingual tissue with sufficient accuracy. Tongue contours and passive vocal tract structures were defined with respect to a semi-polar analysis grid superimposed on the MR images, allowing for quantification of tongue displacement and comparison of constriction degree at different points in time. Because subject's heads were stabilized throughout each scan, tongue position can be compared both within and across utterances.

3. Results

Midsagittal tongue postures comparing mid-consonantal articulatory targets for sibilants produced in two antagonistic vowel contexts are illustrated in Fig. 1. Lingual outlines show mean tongue edges calculated over a three-frame window centered on the mid-fricative frame. The data reveal speaker-specific differences in place of articulation and tongue posture, yet three broad patterns of articulation are evident across this population.

s is produced as a dental fricative by W2, and an alveolar by the other subjects, yet in each case it is articulated as a highly apical coronal consonant. For all four speakers, it is the fricative produced with the most anterior location of the critical constriction formed between the tip of the tongue and the dental or post-dental passive articulators.

sh shows the most variation in production across this group of speakers. Counter to the most common description of this segment, most of the sibilants produced by these speakers are not retroflexed, other than some of the sh tokens produced by M1 (Fig. 1, left column). However, consistent with the observations of Ladefoged & Maddieson [3], none of these tongue postures closely resemble those of Dravidian [17], North Indian [3], and Pama-Nyungan [18] retroflex consonants. Even the most retroflexed of the sibilants produced by M1 appear to use a more bunched coronal gesture than the highly articulated tongue tip gestures typical of Tamil, Arrernte, and Hindi retroflexes consonants. For the other three participants in this

study, sh tokens were all produced with a more laminal coronal gesture. The point of maximal constriction formed between the tongue and the passive articulators ranged from alveolar (M2), through post-alveolar (W1, W2), to palatal (M1), but for all speakers, this was typically the most posterior critical constriction location of the three sibilants.

Across this group of subjects, x was the most consistently realized of the three sibilants in the spatial domain: a critical constriction was invariably formed between the tongue blade and the apex of the alveolar ridge (Fig. 1), behind which a constriction of greater aperture was formed between the tongue dorsum and the palatal-alveolar region. In the high front vowel context (ixi), the tongue created an extended critical constriction throughout the entire palato-alveolar region. Although the coordination and degree of constriction of the palatal gesture varied amongst speakers and vowel contexts, the anterior part of critical constriction for x appears to be invariably located between that observed in s and sh.

3.1. Dynamic Midsagittal Sibilant Articulation

Further insights into the articulatory characteristics of Mandarin sibilants may be obtained by considering constriction formation and release over time, in addition to the target lingual postures examined above. Patterns of tongue movement for sibilants produced by these speakers were examined by superimposing tongue edges extracted from successive image frames throughout the intervals of interest.

Sequences illustrating the dynamic production by subject W1 of each the three Mandarin sibilants in low vowel contexts are shown in Figures 2 to 4. In the left panel of each figure, the superimposition of midsagittal tongue positions captured at successive 30 msec intervals shows the transition from the preceding vowel into the target fricative; in the panels on the right, change in tongue position is tracked – from the midpoint of the consonant into the following vowel.[2]

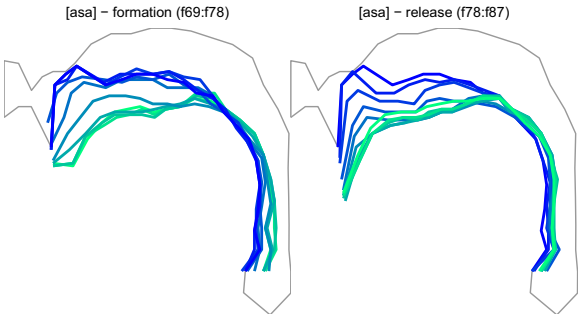

Figure 2: ***Dynamics of sibilant articulation: ā.sā.*** *(Subject W1). Left: constriction formation (7 frames, 210 msec). Right: sibilant release into following tautosyllabic vowel (8 frames, 240 msec).*

Three clear patterns of tongue movement are evident in these data. Articulation of the apical sibilant s primarily involves a coronal-alveolar approximation while the back of the tongue remains largely stationary. Articulation of the sibilant sh involves approximation of the tongue blade towards a more posterior passive articulatory target, which recruits more of the tongue body and therefore causes some forward displacement of the dorsun from the vocalic context posture. Production of x involves even more global lingual movement, as the entire anterior portion of the tongue, from the blade to the middle of the dorsum is presented to the palatal-alveolar region. Importantly, the data reveal that – for these four speakers – this is achieved in a single lingual motion, rather than as a sequence of a coronal constriction followed by a palatalization gesture.

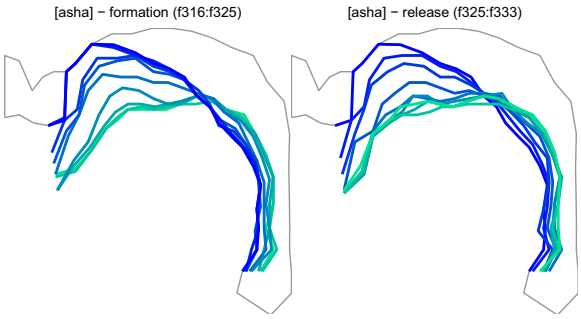

Figure 3: ***Dynamics of sibilant articulation: ā.shā.*** *(Subject W1). Left: constriction formation (7 frames, 210 msec). Right: sibilant release into following tautosyllabic vowel (8 frames, 240 msec).*

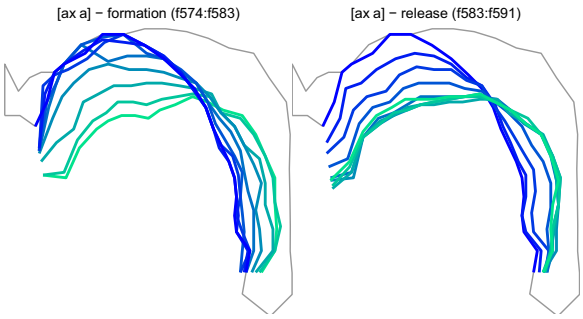

Figure 4: ***Dynamics of sibilant articulation: ā.xā.*** *(Subject W1). Left: constriction formation (7 frames, 210 msec). Right: sibilant release into following tautosyllabic vowel (8 frames, 240 msec).*

3.2. Grove Formation

Data acquired using from regions of the vocal tract beyond the midsagittal plane are another important source of information about sibilant production afforded by MR imaging. Frames acquired from coronal imaging planes intersecting the vocal tract at a fixed point in the post-alveolar region, approximately normal to the tract axis, are illustrated in Figure 5.

The data demonstrate that in all three sibilants produced by subject M2 (top row), an elongated midsagittal groove is formed in the tongue dorsum to channel the airstream into the anterior fricative constriction. The total area of the aperture formed between the tongue and the passive articulators is maximal for s (left), and decreases for sh (center) and x (right), consistent with the characterization of constriction degree in this region of the tract established from the analysis of the midsagittal data.

Coronal data acquired in the same region for subject W2 (Fig. 5, bottom row) show more pronounced differences in groove geometry: s being produced with the most shallow

[2] Data are shown for Subject W1; sibilants produced by all four subjects showed the same fundamental patterns of articulation as those described here.

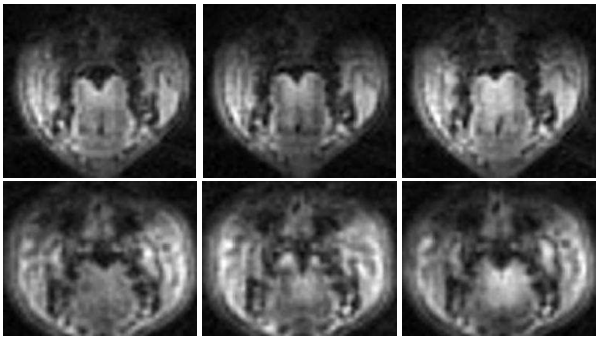

Figure 5: ***Coronal lingual articulation in Mandarin sibilants s - sh - x.*** *Lingual postures captured during mid-fricative production, using fixed tract-normal post-alveolar coronal imaging plane. Top row: subject M2; Bottom row: subject W2. Left-to-right: a.sa, a.sha, a.xa.*

tongue groove (left), x with a deeper groove of similar width (right), and sh with a much deeper, narrower groove (center) than the other sibilants produced by this subject.

3.3. Summary of Results

Articulatory characterizations of the sibilants produced by the speakers in this study are summarized in Table 3. These results are consistent with the observation that individual speaker variation might account for much of the variation in previous phonetic and phonological descriptions of these segments (§1).

SUBJECT	s	sh	x
M1	/s/	/ʂ/	/ɕ/
M2	/s/	/ʃ/	/ɕ/
W1	/s/	/ʃ̻/	/ɕ/
W2	/s̻/	/ʃ̻/	/ɕ/

Table 3: ***Phonetic characterization of sibilants produced by subjects.***

4. Discussion

More data will be needed to examine some of these issues in more depth. It is unclear to what extent vocal tract morphology influences speaker-specific articulation. MR data obtained from a greater array of imaging planes at higher spatial resolutions will lend further insights into individual details and general characteristics of sibilant production and groove geometry.

Richer sets of multi-modal data are also needed. A proper understanding of the goals of sibilant production will require a more complete treatment of the complex interactions between articulation, aerodynamics and acoustics, and the rich system of coronal fricatives and affricates in Mandarin represents an important case study with which to test theories of fricative production and phonological contrast.

More data from socially diverse populations of Chinese speakers are also needed to address the pervasive sociophonetic factors which influence production and perception of coronals. Beckman et al. [9] have observed that young female speakers of *Beijinghua* accents produce less monolithic, palatalized variants of the palatoalveolar sibilant (c.f. [2]), and Li [19, 20] points out some important implications of sociophonetic variation on phonological acquisition of SC sibilants.

5. Conclusions

The results of this study provide further insights into the mechanisms of production of sibilant consonants in Mandarin, and coronal fricative production in general. The data suggest that although individual speakers may differ in their specific realizations of different segments, the same patterns of sibilant production – formation of critical constrictions in contrasting regions of the tract – may be observed across a diverse speaker population. These data further illustrate the importance of real-time MRI as a method for studying the dynamics of coronal articulation and groove formation in fricative consonants.

6. Acknowledgements

Research supported by NIH Grants R01 DC007124-01 and R01 DC03172.

7. References

[1] S. R. Ramsey, *The languages of China.* Princeton: Princeton University Press, 1987.

[2] S. Duanmu, *The phonology of standard Chinese*. Oxford; New York: Oxford University Press, 2007.

[3] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Oxford; Cambridge, MA: Blackwell, 1996.

[4] Y.-R. Chao, *A grammar of spoken Chinese.* Berkeley: University of California Press, 1968.

[5] C.-Y. Lee, "An acoustic study of strident fricatives in Mandarin Chinese," *JASA*, vol. 113, no. 4, p. 2329, 2003.

[6] F. Hu, "The three sibilants in Standard Chinese," in *Proceedings of the 8th Intl. Seminar on Speech Production*, R. Sock, S. Fuchs, and Y. Laprie, Eds., 2008, pp. 105–108.

[7] C. W.-C. Li, "Conflicting notions of language purity: the interplay of archaising, ethnographic, reformist, elitist and xenophobic purism in the perception of Standard Chinese," *Language & Communication*, vol. 24, no. 2, pp. 97–133, 2004.

[8] P. Ladefoged and Z. Wu, "Places of articulation: An investigation of Pekingese fricatives and affricates," vol. 12, pp. 267–278, 1984.

[9] M. E. Beckman, E. J. Kong, F. Li, and J. Edwards, "Aligning the timelines of phonological acquisition and change," in *2nd Workshop on Sound Change*, M. Stevens and J. Harrington, Eds., Kloster Seeon, 2-4 May 2012.

[10] J.-O. Svantesson, "Acoustic analysis of Chinese fricatives and affricates," *J. Chinese Linguistics*, vol. 4, no. 1, pp. 53–70, Jan 1986.

[11] W.-S. Lee, "An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin," *Proc. 14th Intl. Cong. Phonetic Sciences*, pp. 413–416, 1999.

[12] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *JASA*, vol. 115, no. 4, pp. 1771–1776, 2004.

[13] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time MR Imaging [Exploratory DSP]," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 123–132, 2008.

[14] E. Bresch, J. Nielsen, K. Nayak, and S. Narayanan, "Synchronized and noise-robust audio recordings during realtime MRI scans," *JASA*, vol. 120, no. 4, pp. 1791–1794, 2006.

[15] M. I. Proctor, D. Bone, and S. S. Narayanan, "Rapid semi-automatic segmentation of rtMRImages for parametric vocal tract analysis," in *InterSpeech*, Makuhari, Japan, 2010, pp. 1576–1579.

[16] S. Narayanan, E. Bresch, P. K. Ghosh, L. Goldstein, A. Katsamanis, Y.-C. Kim, A. Lammert, M. I. Proctor, V. Ramanarayanan, and Y. Zhu, "A multimodal real-time MRI articulatory corpus for speech research," in *InterSpeech*, Florence, 2011, pp. 837–840.

[17] M. Proctor, L. Goldstein, D. Byrd, E. Bresch, and S. Narayanan, "Articulatory comparison of Tamil liquids and stops using rtMRI," *JASA*, vol. 125, no. 4, p. 2568, 2009.

[18] A. Butcher, *The Phonetics of Australian Languages*. Oxford: Oxford University Press, 1995.

[19] F. Li, "An acoustic study on feminine accent in the Beijing Dialect," in *Proc. 17th North American Conf. on Chinese Linguistics*, Q. Gao, Ed. USC Ling. Publ., 2005, pp. 219–224.

[20] ——, "The phonetic development of voiceless sibilant fricatives in children speaking English, Japanese and Mandarin Chinese," Ph.D. dissertation, Ohio State University, Dept. Linguistics, 2008.