SPECTRAL ANALYSIS OF MANDARIN CHINESE SIBILANT FRICATIVES

Sang-Im Lee

New York University, USA
sangim@nyu.edu

ABSTRACT

This paper investigates the spectral aspects of Mandarin Chinese sibilant [s̪, ʂ, ɕ] in three vowel context /a, i, u/. Three male and three female speakers of Mandarin Chinese were recorded. To ensure good spectral estimates, frication noise was analyzed using multitaper analysis. For comparison, one female speaker of Polish was analyzed in the same way. A strict ordering was found among Mandarin sibilants in the first spectral moment; [s̪] has the highest spectral moment, [ɕ] intermediate, and [ʂ] lowest. In contrast, the spectral moments of Polish [ɕ] and [ʂ] presented very similar values. This study also found a strong acoustic effect of the rounded vowel. The noise spectra in the [u] context appear to have a distinct narrow-bandwidth peak at a higher frequency that sometimes increases the first spectral moment.

Keywords: Mandarin Chinese, spectral analysis, sibilant fricatives, multitaper method

1. INTRODUCTION

1.1. Background

Mandarin Chinese has a three-way distinction in place of articulation for sibilant fricatives; dental [s̪], retroflex [ʂ], and palatal [ʂ]. The dental [s̪] has a constriction located close to the teeth, but the exact constriction site is known to be highly variable depending on the speaker [5]. A number of articulatory studies agree that the so-called retroflex [ʂ] of Mandarin Chinese is not a typical retroflex made with the tip of the tongue curled back, but rather it is made with the upper surface of tongue tip or blade against the post-alveolar region [5, 6, 11]. A palatogram study found that the constriction of palatal [ɕ] is made approximately in the same region with [ʂ], but an MRI study showed that it is slightly front than the retroflex [5, 11] (respectively).

Despite the similarity in the constriction location of [ʂ] and [ɕ], acoustic studies found that the centroid frequency of noise spectra of [ʂ] is lower than [ɕ] [10]. One explanation regarding the lowering of the centroid frequency of [ʂ] is concerned with the large sublingual cavity that retroflexes typically involve. Another account relies on the relatively short and slack constriction of [ʂ] that enables coupling of the back cavity with the front cavity [9], which in turn creates the concentration of energy at a lower frequency region in the noise spectra.

A fairly similar set of sibilants is found in Polish. In particular, Ladefoged and Maddieson [4] described each sibilant of Polish as nearly comparable to that of Mandarin Chinese in terms of articulation. However, no clear evidence is found in the acoustic studies of Polish sibilants that the spectral shape and the centroid of the noise spectra are significantly different between [ʂ] and [ɕ] as they are in Mandarin Chinese [3, 7]. This is further supported by the fact that the vocalic transition is important to correctly distinguish [ʂ] from [ɕ] for Polish speakers [7].

1.2. The current study

This paper investigates three aspects of the sibilants of Mandarin Chinese. First, this paper exploits a recent method for spectral analysis, namely the multitaper method [1], to ensure good spectral shapes and estimates. Considering the dynamic property of frication noise in a running speech, multitaper analysis is particularly suitable, since this method does not rely on the assumption of stationarity or ergodicity of frication noise and thus ensures a small error with good time and frequency resolution [1]. This allows a better comparison among sibilant fricatives of Mandarin Chinese.

This paper also examines the effect of vowel contexts on the frication noise. A strong acoustic effect of the rounded vowel has been discussed in the literature [8]. In particular, in the /s/ spectra of one French speaker, a wide-bandwidth main peak was observed in both /a_a/ and /i_i/ contexts, but a considerably different spectra was found in the /u_u/ context; much lower amplitude overall and one narrow-bandwidth peak located at a higher frequency. The presence of the narrow peak at a higher frequency was analyzed as resulting from the whistle-like source mechanism of that particular speaker. However, when many of speakers were analyzed, this effect seemed to be averaged out, and a rather general effect of rounded vowel, lowering of the first spectral moment due to the labial anticipation, emerged [2]. This paper examines three vowel /a, i, u/ contexts following [s̪, ʂ, ɕ] with an expectation of similar contextual effects of vowel both on the spectra and the spectral moments.

Finally, Mandarin Chinese sibilants are compared with Polish sibilants. As noted earlier, Polish sibilants are known to be articulated similar way with those of Mandarin Chinese. It is hoped that the application of the same methodology to two sibilant systems will better illuminate their similarities and differences.

2. METHOD

The stimuli included nine disyllabic words with nine filler words. The target was on the first syllable with three fricatives [s̪, ʂ, ɕ] in three different vowel contexts [i(ɯ, ɨ), a, u]. The tone of the target was fixed as the high level tone: X^{55}. The second syllable of the target words started with [xa].

Table 1: Mandarin stimuli <u>CV</u>^{55}. xaC

	s̪	ʂ	ɕ
a	s̪a^{55}.xuaŋ^{214}	ʂa^{55}.xai^{53}	ɕa^{55}.xan^{214}
i	s̪ɯ^{55}.xao^{35}	ʂɨ^{55}.xan^{35}	ɕi^{55}.xan^{53}
u	s̪u^{55}.xaŋ^{35}	ʂu^{55}.xan^{35}	ɕu^{55}.xao^{214}

Three male and three female speakers of Mandarin Chinese were recorded. All subjects spoke no Chinese dialects other than Standard Mandarin. The acoustic recording was conducted at sound attenuated recording booth at the Phonetics and Experimental Phonology Laboratory at New York University and the Acquisition Laboratory of the Chinese University of Hong Kong. A list of nine target words and nine fillers were presented 8-10 times in random order. Each sheet was presented on a music stand, and subjects were asked to read the target words in a carrier phrase "Wo shuo __ zhe ge ci" (I say __ this word). The recording was made using an Audio Technica AT 2010 cardioid condenser vocal microphone with the sampling rate of the recorder at 44.1 kHz.

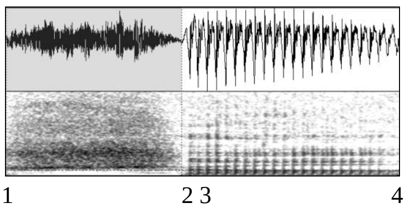
For the Polish fricatives, one female speaker of Polish was recorded. Most target words were disyllabic real words, except for [s̪iməns] and [ʃita], which were said to be loanwords. Each target word was produced in a carrier phrase "Powiedziałam ___ do ciebie" (I said __ to you). Other procedures remained the same as in the Chinese recording.

Table 2: Polish stimuli.

	s̪	ʂ	ɕ
a	s̪a.mo	ɕa.ta	ɕa.ta
i	s̪i.məns	ʃi.ta	ɕi.to
u	s̪u.ma	ʂu.ba	ɕu.ski

The acoustic segmentation was done manually following the method in [2]. Four events in the target syllable were labeled using Praat: (1) frication onset, (2) frication offset, (3) vowel onset, and (4) vowel offset.

Figure 1: A Praat screenshot of segmentation of [ʂ - a]. Frication noise:1 - 2, acoustic vowel duration: 3-4.



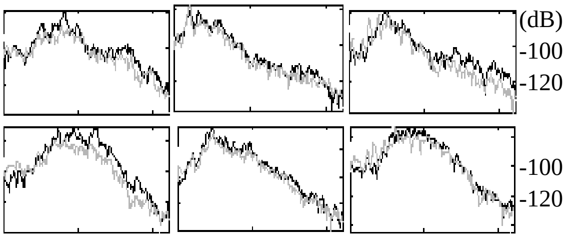
Three multitaper spectra were computed at the beginning, middle, and end of the segmented fricative (1-2). Treating the spectra as probability distributions, the first spectral moment, corresponding to the mean value of the distribution, was computed from these multitaper spectra.

3. RESULTS

3.1. Spectral shape and M1 by fricatives

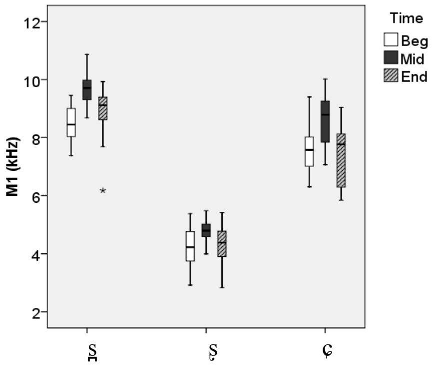
Figure 2 presents the representative spectra of one male and one female speaker in the vowel [a] context. Spectral shapes certainly vary token by token, but at least some generalizations can be made. First, the dental fricative [s̪] shows several prominent peaks in the very high frequency region: around 8 kHz for the male speaker and 10 kHz for the female speaker. In contrast, spectral peaks occurred at a considerably lower frequency region for the retroflex [ʂ]; the most prominent peak was at 1.5 kHz and there was a relatively flat main peak up until 6 kHz for the male speaker, and a relatively sharp main peak at 5 kHz for the female speaker. The spectral peak of palatal [ɕ] was intermediate between [s̪] and [ʂ] for both female and male speakers.

Figure 2: Representative spectra in the vowel context of [a]; the Mid phase is presented in black, and the End phase in gray; one male speaker (top) and one female speaker (bottom).



The spectral properties described above were well reflected in the first moment calculated by spectral normalization. Figure 3 shows the first moment in all vowel contexts by Beginning, Mid, and End of each fricative.

Figure 3: M1 of one female speaker of Mandarin Chinese in all vowel contexts. The distribution of M1s was summarized in boxplots with quartiles.



Assuming place of articulation of [ʂ] and [ɕ] is about the same [5], the difference of M1 can be attributed to the lowering of M1 for [ʂ] probably due to the expanded sublingual cavity (or the coupling with the back cavity [9]). Despite the speaker variation, it seems the relative ordering of M1 remains invariable for every speaker; [s̪] the highest, [ɕ] intermediate and [ʂ] the lowest, as can be seen in Table 3.

In addition, the fact that only the ordering was strictly maintained across speakers reveals the non-linear aspect of speech perception; listeners can effectively abstract away the variation due to physiological differences between speakers and correctly identify distinct fricatives.

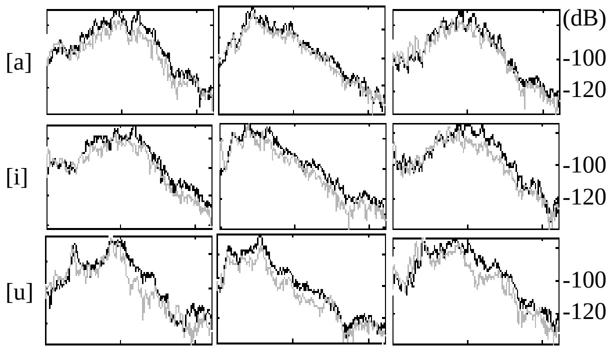
Table 3: Means (SD) of M1 of each speaker (in a row) at Mid of each fricative (in Hz)

	s̪	ʂ	ɕ
Male speakers	7,943 (570)	3,256 (250)	5,592 (300)
	7,422 (1,025)	5,469 (277)	6,237 (308)
	9,444 (807)	4,696 (376)	5,479 (480)
Females speakers	9,662 (487)	4,781 (362)	8,599 (844)
	10,980 (483)	4,629 (364)	9,626 (320)
	8,671 (532)	4,946 (450)	6,523 (551)

3.2. The effect of vowel contexts

This study also tested the effect of vowel contexts on noise spectra. Figure 4 demonstrates a clear contrast between the rounded vowel context and others. In the [u] vowel context, an additional peak was amplified at a lower frequency, which was considered to be responsible for the drop of M1.

Figure 4: Representative spectra of one female speaker in all vowel contexts; the Mid phase was presented in black, and the End phase in gray.

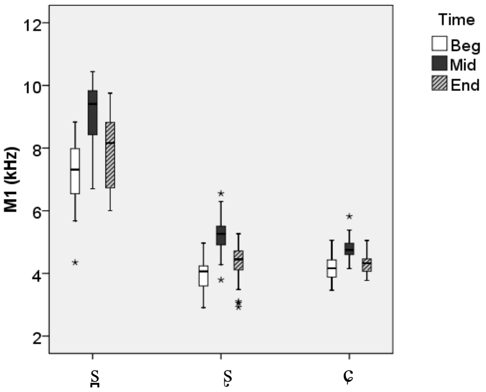


However, the M1 drop in the [u] context was not uniformly true for all fricatives. In particular, most speakers showed rather higher M1 than other contexts for the [ʂu] sequence, and three male speakers showed the same pattern for the [s̪u] sequence. This seems to be suggestive of a whistle-like source mechanism that excites the higher frequency resonance [8], which possibly results in the rising of M1. This alternative is particularly plausible considering the most robust M1 drop in the [ʂu] sequence; the slight lip protrusion and rounding of the retroflex and the labial anticipation for [u] might effectively trigger the change in source mechanism. A more comprehensive aerodynamic and articulatory study should be followed up to confirm this alternative.

3.3. A comparison with Polish retroflex

This study compares the result of Mandarin Chinese to the Polish sibilants. A clear contrast in the global distribution of M1 was found; Mandarin Chinese sibilants has a clear ordering in terms of M1 as [s̪> ɕ> ʂ] regardless of the absolute values, but this strict relation did not hold true for Polish. The M1 of [ʂ] is higher than or equal to that of [ɕ].

Figure 5: M1s of Polish fricatives produced by one female speaker in all vowel contexts. The distribution of M1s was summarized in boxplots with quartiles.



Though only one Polish speaker was tested in this study, it is consistent with the previous studies [3, 7] in that there was no significant spectral difference between [ʂ] and [ɕ]. Assuming this is mainly due to the lowering of M1 of [ɕ], this seems to suggest that the constriction of Polish [ɕ] is made further back than Mandarin [ɕ]. One implication of this finding for perception might be that the vocalic information would be quite redundant for Mandarin speakers but not for Polish speakers [7].

4. CONCLUSION

In this study, spectral aspects of sibilant fricative [s̪, ʂ, ɕ] of Mandarin Chinese were examined in three vowel contexts /a, i, u/. The results showed that [s̪] had the highest spectral peak and amplitude at higher frequencies, [ɕ] intermediate, and [ʂ] the lowest. The noise in the rounded vowel context tends to show one narrow spectral peak amplified at a higher frequency, and M1 rather increases when this peak was significantly enhanced. Yet, otherwise the rounded vowel lowered M1 in general. Finally, a brief comparison was made on the global spectral moment between Mandarin Chinese and Polish sibilants. A major difference of the two languages was the significantly lower M1 of Polish [ɕ].

5. ACKNOWLEDGEMENTS

Special thanks to Christine Shadle for her sincere supervision on this project. Thanks also to Sean Martin for helpful comments, and Chinese and Polish speakers for the recordings. This work was partially supported by the National Science Foundation grant BCS-0449560 to PI Lisa Davidson.

6. REFERENCES

[1] Blacklock, O.S.B. 2004. *Characteristics of Variation in Production of Normal and Disordered Fricatives Using Reduced-Variance Spectral Methods.* Ph.D. Thesis, School Electr. & Comp. Sci., U. Southampton.

[2] Iskarous, K., Shadle, C.H., Proctor, M.I. 2011. The articulatory-acoustic kinematics: The production of American English /s/. *J. Acoust. Soc. Am.* 129(2), 944-954.

[3] Jassem, W. 1965. The Formants of Fricative Consonants, *Language and Speech* 8(1), 1-16.

[4] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell.

[5] Ladefoged, P., Wu, Z. 1984. Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12(3), 267-278.

[6] Lee, W.S., Zee, E. 2003. Standard Chinese (Beijing), *Journal of International Phonetic Association* 33, 109-112.

[7] Nowak, P.M. 2006. The role of vowel transitions and frication noise in the perception of polish sibilants. *Journal of Phonetics* 34(2), 139-152.

[8] Shadle, C.H., Scully, C. 1995. An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics* 23, 53-66.

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

[10] Svantesson, J.O. 1986. Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics* 14(1), 52-70.

[11] Toda, M., Honda, K. 2003. An MRI-based cross-linguistic study of sibilant fricatives. *Proc. 6th International Seminar on Speech Production* Manly, Australia.