Contrast enhancement in clearly spoken Mandarin sibilants

Ivy Hauser, Xinwen Zhang, and Shang-Yune Tang
University of Texas Arlington

In clear speech, talkers adapt a hyperarticulated speaking style which includes global signal *enhancement* as well as *contrast enhancement* of language-specific cues (e.g. Smiljanic, 2021). This study examines acoustics of clearly spoken sibilants in Mandarin, with three main questions: (1) What are the acoustic characteristics of clearly spoken Mandarin sibilants? (2) Which characteristics are due to *signal enhancement* versus *contrast enhancement*? (3) What implications does clear contrast enhancement have for phonological analyses of Mandarin alveolopalatal sequences?

Standard Mandarin exhibits a three-way place contrast between alveolar, retroflex, and alveolopalatal sibilants /s ʂ ɕ/. The primary cues for place distinction are spectral center of gravity (CoG) for /s/ vs. /ʂ/ (e.g. Lee-Kim, 2011) and onset of the second formant of the follow vowel (F2) for /ɕ/ vs. /s ʂ/ (e.g. Li, 2008). There is disagreement on the nature of the raised F2 in vowels following [ɕ], in particular, whether it is due to phonetic coarticulation/offglide (e.g. Duanmu, 2007), phonological assimilation (e.g. Hauser, 2020), or surface diphthongs (e.g. Lee & Zee, 2003). This project aims to determine whether speakers actively control this vowel fronting with contrast enhancement in clear speech.

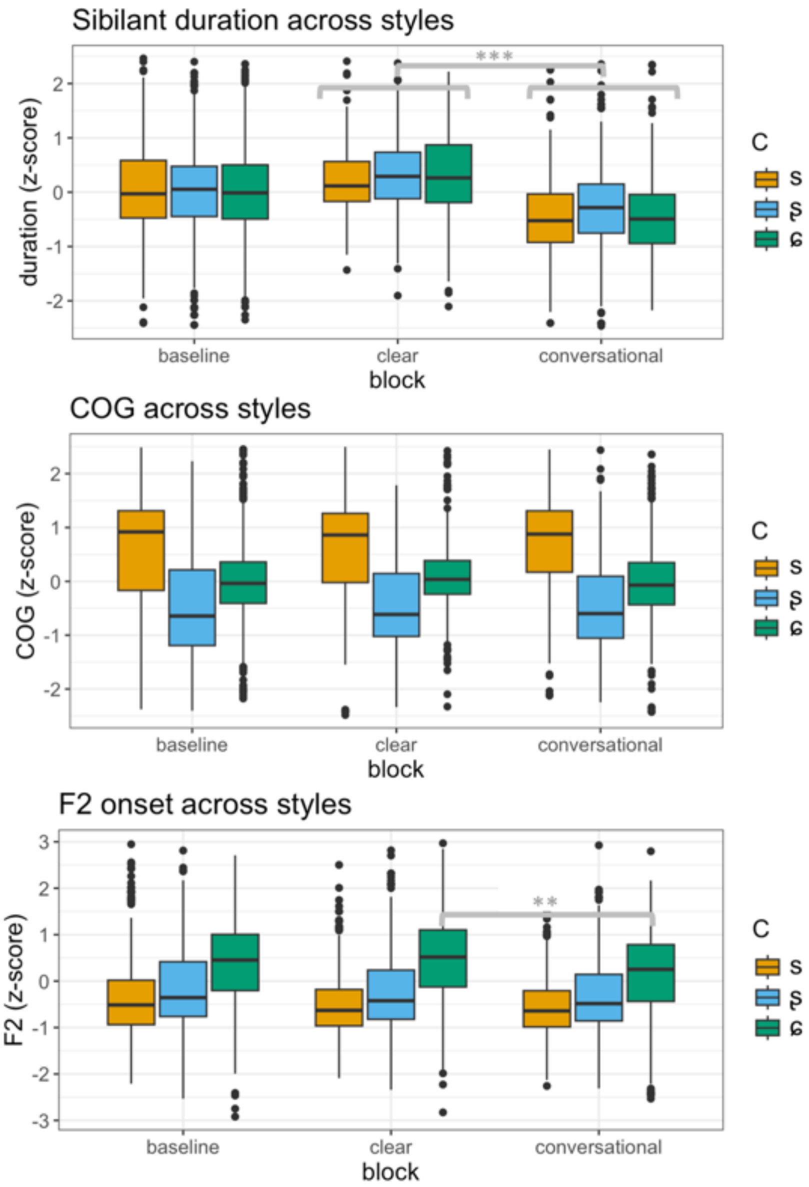
Sibilant-initial words were elicited in a laboratory task and were crossed according to: sibilant (s, ʂ, ɕ), following vowel (a u), frequency, number of syllables, and tone. There were 130 sibilant-initial target words and 57 fillers. The participants were 12 native speakers of standard Mandarin. They read the word list aloud in the carrier phrase “wǒ bǎ X dú yī biàn” (I read X once) with a different randomization in each block. The first baseline block had no instructions on speech style. The second and third blocks contained instructions on speech style and their block order was balanced across participants. The clear block instructed participants to speak clearly 'as if talking to someone hard of hearing’. The conversational block instructed participants to speak conversationally 'as if casually talking to a friend’.

The sound files were force aligned using the Montreal Forced Aligner with the pre-trained Mandarin model (McAuliffe et al., 2017). Praat scripts were used to extract: duration and amplitude of sibilants and following vowels, spectral moments of sibilants, and formants of following vowels at vowel onset (10ms into the vowel). Separate linear mixed effect models were constructed for each acoustic measure as dependent variable (duration in ms, CoG in Hz, F2 in Hz) with style, block order, sibilant, and following vowel as fixed effects and speaker and word as random effects. Figures are shown with all dependent variables as within-speaker z-scores for visual comparison.

Speakers significantly enhanced duration and amplitude of all sibilants and vowels in the clear block (Figure 1). CoG and other spectral moments were not significantly enhanced in the clear block for any sibilants (Figure 2). However, F2 following [ɕ] was significantly higher in the clear block relative to the conversational block (Figure 3).

Overall, clearly spoken Mandarin sibilants exhibited global signal enhancement through increased duration and amplitude. Departing from previous work on English, there was no enhancement in spectral moments for any sibilants, which may be due to differences in type of task (e.g. explicit instruction to speak clearly vs. inducing clear speech through contrastive focus contexts). We did observe enhancement of F2 following [ɕ], demonstrating that extrinsic primary cues can be enhanced in clear speech. This enhancement cannot be attributed to global signal enhancement of vowel formants as F2 was not significantly enhanced following other sibilants. We argue that the F2 enhancement is contrast enhancement, showing that speakers actively control vowel fronting following [ɕ]. Such active control would be expected with phonological approaches to Mandarin alveolopalatal sequences where the vowel fronting is grammaticalized in some way (e.g. Hauser, 2020; Lee & Zee, 2003).

Figures 1-3. Duration, CoG, and F2 onset across the different styles elicited (duration in ms, CoG in Hz, F2 in Hz all shown with within-speaker z-scores in graphs)



References

[1] Duanmu, S. (2007). The phonology of standard Chinese. Oxford University Press.

[2] Hauser, I. (2020). Coarticulation with alveopalatal sibilants in Mandarin and Polish: Phonetics or phonology? In *Proceedings of the Annual Meetings on Phonology*.

[3] Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112.

[4] Lee-Kim, S.-I. (2011). Spectral analysis of Mandarin Chinese sibilant fricatives. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1178–1181).

[5] Li, F. (2008). The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese. Ohio State University.

[6] McAuliffe, M., et al. (2017). *Montreal forced aligner* (Version 0.9.0).

[7] Smiljanic, R. (2021). Clear speech perception: Linguistic and cognitive benefits. *The handbook of speech perception*, 177-205.