Relating frication to articulation in Standard Mandarin apical vowels

Sean Foley[1], Bowei Shao[2], Matthew Faytak[3]

[1]*University of Southern California*
[2]*École Normale Supérieure-Université PSL*
[3]*University at Buffalo*
seanfole@usc.edu, bowei.shao@ens.psl.eu, faytak@buffalo.edu

Abstract

Sibilants are characterized by the production of turbulent airflow, which involves both a narrow constriction in the vocal tract and a certain volume velocity of the airflow. Despite both conditions being necessary for sibilant production, studies of constriction degree predominate in the literature. Using acoustic and articulatory data, we show that in certain sequences Standard Mandarin apical vowels exhibit minimal lingual adjustment compared to adjacent sibilants, while also exhibiting a considerable drop in frication noise. The same result was found for the vowel /i/. We hypothesize that the change in frication noise could be due to a number of different non-lingual factors and discuss the potential implications for models of sibilants.
Index Terms: sibilants, apical vowels, Standard Mandarin, ultrasound

1. Introduction

Sibilants are sounds characterized by the production of audible turbulent airflow (Stevens 1998). Mechanical models of sibilants dictate that the production of turbulent airflow requires both the formation of a narrow constriction in the vocal tract and air projected at a certain velocity through this constriction (Shadle 1990; Catford et al. 1977). These aerodynamic principles suggest that in connected speech the production of frication noise rests on a certain balance being struck between these two factors, e.g. a larger constriction necessitates greater volume velocity and vice versa (Yoshinaga, Nozaki, and Wada 2019). We investigate this relationship between lingual constrictions and aerodynamics in Standard Mandarin apical vowels using both articulatory and acoustic data.

In Standard Mandarin (SM), there is a three-way place contrast among sibilants, with the language contrasting dental, alveolo-palatal and retroflex sibilants, e.g. /s ɕ ʂ/. One consequence of this three-way place contrast is the co-occurrence restriction on the high front vowel /i/ following dental and retroflex sibilants, e.g. *si *ʂi. In these contexts, in place of the high front vowel, there occurs two apical segments, [ɿ] and [ʅ], which occur only after sibilants they are homorganic with, e.g. [sɿ] and [ʂʅ] (Duanmu 2007). [ɿ] will be referred to as the "dental apical vowel" and [ʅ] as the "retroflex apical vowel" in keeping with the previous literature.

Two key characteristics of the apical vowels are the focus of the current study. First, previous research has shown that both apical vowels are produced with a lingual configuration that closely resembles their onsets (Lee-Kim 2014; Faytak and Lin 2015; Shao and Ridouane 2023), though questions remain on the exact nature of the lingual transition from the onset sibilant to apical vowel. While studies have reported a range of adjustments, it is difficult to rule out if any observed differences between the onset and apical vowel were due to coarticulatory effects from the segment following the apical vowel (Foley 2023). Second, there is some debate on whether the apical vowels have frication noise targets (Lee-Kim 2014; Duanmu 2007; Yu 1999; Shao and Ridouane 2023), with few studies firmly quantifying the rate of turbulent airflow during these segments (Shao and Ridouane 2023). While Lee-Kim (2014) concluded that the segments lack frication and termed them "syllabic approximants", Yu (1999) concluded that they are "syllabic sibilants", with both studies using impressionistic inspection of spectrograms and waveforms as evidence.

To further explore the mechanics of the SM apical vowels, we looked at sequences where each segment occurs adjacent to the sibilant they are homorganic with on *both* sides. Given previous research, there are a number of potential hypotheses of what would occur in such sequences. If both apical vowels have frication noise targets, we would likely see no lingual adjustment as well as little to no change in frication noise during the entire sequence. If both segments lack frication noise targets, we should see a sizeable drop in frication during the apical vowels, comparable to that of other vowels. The general expectation is that such a drop should be accompanied by an increase in the channel size, i.e. tongue tip lowering, though a non-lingual adjustment is also possible, e.g. manipulation of the volume velocity or cavity expansion.

2. Methods

2.1. Ultrasound experiment

Seven speakers of SM with no history of speech or hearing disorders took part in the study. Data from two speakers was excluded due to errors in the placement of the ultrasound probe. The five remaining speakers were all aged 18-25 years old; three speakers were from northern provinces (Liaoning, Shandong, Shaanxi) and two were from central/southern provinces of China (Henan, Jiangsu).

Stimuli consisted of disyllabic pseudo-words. The target segments in the first syllable are [ɿ ʅ i u], paired with three different onsets [s ʂ ɕ]. Due to phonotactic restrictions, each apical vowel occurs only after homorganic sibilants, [i] occurs only after [ɕ], and [u] occurs only after [s] and [ʂ]. The second syllable is one of [sa ʂa ɕa]. Target sequences are those containing the apical vowels flanked on both sides by a homorganic sibilant, i.e. [sɿ.sa] and [ʂʅ.ʂa], with other sequences containing [i u] in the first syllable used for comparison. All syllables were produced with a high level tone. Sixteen disyllabic filler items were also presented. Stimuli were presented in blocks of five, randomized so that each target phrase was seen a total of five

times across all blocks. The stimuli were presented as simplified Chinese characters in the following carrier phrase: 我觉得___ 很好[wə^{21} tɕyei^{35} də_xən^{35} xau^{213}] "I think _ is very good".

Ultrasound video and audio were co-recorded in a sound-attenuated booth using the Articulate Assistant Advanced (AAA) software. Ultrasound was recorded using a Telemed MicrUs and two different probes, a Telemed MC10 microconvex probe for speakers SP_06 and SP_07 and a Telemed MC4 microconvex probe for all other speakers. Probes were stabilized with a metallic Articulate Instruments stabilization headset. Audio files were analyzed in Praat and segmented using the Montreal Forced Aligner (McAuliffe et al. 2017) with manual corrections as needed.

2.2. Analysis

Zero-crossing rate (ZCR) was used to measure the time course of frication during target sequences. ZCR measures the number of crossings of zero dB per second in the waveform without relying on voicing or pitch, and has been used to gauge frication levels in similar segments (Shao 2020; Shao and Ridouane 2023). Generalized Additive Mixed Models (GAMMs) (Wood 2011) were constructed to model the dynamics of z-scored ZCR in target sequences, using mgcv v1.8-40 (Wood 2011). We constructed a single model to model all sequences and reported the estimated differences in separate difference figures. In the model, ZCR of [C_1{ɿ, ʅ, i, u}C_2a] sequences was estimated over time, with factor smooths for speaker. Because ZCR has a left-skewed, long-tailed distriubtion, Tweedie distributions were used in the GAMM models. Results were visualized using tidyverse v1.3.2 (Wickham et al. 2019) and tidymv v3.3.2 [1].

Ultrasound frames recorded during the acoustic duration of the target segments were processed in Articulate Assistant Advanced (AAA). Tongue contours were estimated using speaker templates, hand-corrected as necessary, and exported in polar and Cartesian coordinates. To visualize tongue posture over the duration of the target [C_1{ɿ, ʅ}C_2a] items and the comparable [C_1{i, u}C_2a] items, smoothing-spline ANOVAs (SSANOVAs) were generated in polar coordinates comparing the midpoints of the first homorganic fricative (C_1), apical vowel, second homorganic fricative (C_2), and final [a] (Davidson 2006; Gu 2014). The resulting splines and 95% confidence intervals were visualized using tidyverse v1.3.2 (Wickham et al. 2019). The SSANOVAs serve to confirm whether there are any broad adjustments to tongue posture in the transitions between the apical vowels and their flanking homorganic fricatives, and to compare this adjustment to comparison items containing [i u] flanked by the same fricatives.

Additionally, constriction degree (CD) was calculated in AAA using a fiducial line drawn from the probe origin through the alveolar or postalveolar area depending on the constriction at issue. CD was calculated as the distance between the intersections of the fiducial line with the tongue contour and the palate trace. All values were z-scored across speakers. GAMMs were also fit on CD data to model change over the target sequences. The model design was the same as the ZCR models, but the CD models were fit using a Gaussian distribution. Both the ZCR and CD GAMMs were fit using by-phrase relativized time, calculated using $\mathbf{t}_i^{rel} = \mathbf{t}_i - \min(\mathbf{t})/\max(\mathbf{t}) - \min(\mathbf{t})$, where $\mathbf{t}_i$ is a single timepoint.

[1] https://stefanocoretta.github.io/tidymv

3. Results

3.1. SSANOVAs

The SSANOVA splines in Figure 1 summarize the typical posture for the imaged portion of the tongue at the midpoint of each segment in target [C_1{ɿ, ʅ}C_2a] items, with [ɕi.ʂa] shown for comparison. In the apical vowel targets, the tongue blade does not visibly differ in position between the first onset fricative, the apical vowel, and the second onset fricative. Some slight variation in tongue dorsum and blade position between the apical vowel and the second onset consonant can be attributed to anticipation of the upcoming low vowel [a]. Unexpectedly, the tongue blade is also raised at the midpoint of [i] for all speakers, not appreciably differing from the raising observed for [ɕ]; in fact, the tongue postures of [ɕ] and [i] are essentially the same, except for speakers SP_02 and SP_05 who show somewhat more dorsum raising during [i].

3.2. Zero-crossing rate

Figure 2 shows the results from the time-aligned ZCR (bottom) and CD (top) GAMMs. Constants were added to both sets of values for visualization. Two clear peaks in ZCR corresponding to the sibilants [s ɕ ʂ] are visible in the targets, as well as two valleys corresponding to the nuclei [a u ɿ ʅ]. The ZCR values in V_1 position are consistently much lower compared to the two flanking peaks, suggesting that V_1 has reduced aperiodicity compared to [s ɕ ʂ]. Crucially, while we can see clear differences in ZCR between the four phrases during the two flanking sibilants, the ZCR trajectories all converge to a common minimum near the V_1 midpoint.

The GAMM estimates of difference in ZCR are shown in Figure 3, where shaded red regions show intervals during which this comparative difference is significant. In most cases, the difference in aperiodic noise is significantly different during the two sibilants, i.e. C_1 and C_2, with a gap in this difference during the midpoint of V_1. Interestingly, the former is true even in the comparison when both phrases have the same sibilants (top panel). This suggests that the different nuclei in each phrase have a direct impact on how favorable the conditions are for the production of turbulent airflow, with the dental apical vowel creating a more favorable context. This is likely related to the homorganicity between the onset [s] and apical vowel [ɿ].

3.3. Constriction degree

GAMMs fit on the CD data are shown in Figure 2 (top) for the target phrases. It can be seen clearly that for all of the phrases with homorganic sequences, i.e. those with the apical vowels and [i], a consistent CD is maintained during the first sibilant and V_1, with the constriction being released during the C_2 in anticipation of the following [a]. For the other phrase containing [u] as V_1, a sizeable dip in CD occurs in accord with the V_1 onset, only for a subsequent constriction to be formed for the second sibilant. This is indicative of a slight release in the tongue front constriction during the production of the vowel [u] in this phrase. This slight release in constriction coincides with the drop in frication seen in the ZCR GAMM, while for the other phrases, there is no perceptible change in CD that coincides with the change in aperiodic noise.

The GAMM estimates of difference in CD are shown in Figure 4. In the comparisons in panels 1 and 3 (1 being the top panel), where the phrases with the apical vowels are compared to those containing the phrase with [u], we see a period of significant difference during the drop in CD that occurs during the

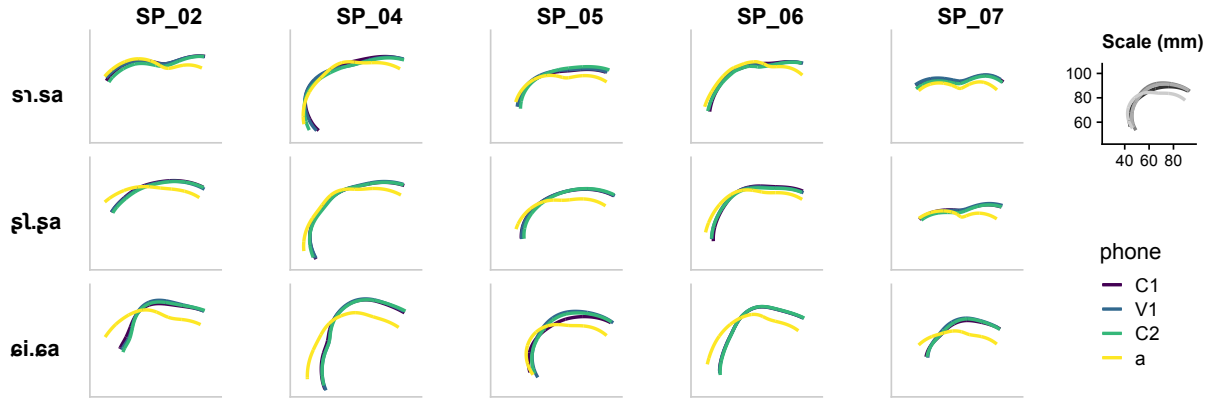

Figure 1: Tongue surface SSANOVA splines for segment midpoints in target [$C_1V_1C_2$a] items. Anterior is right in each figure.

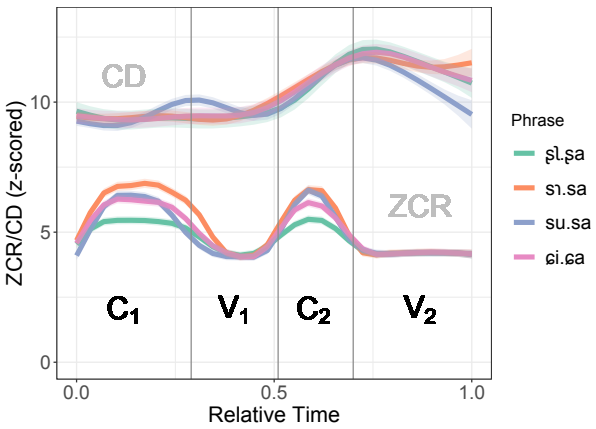

Figure 2: ZCR (bottom) and CD (top) GAMMs for all target sequences. Grey vertical lines indicate phone boundaries.

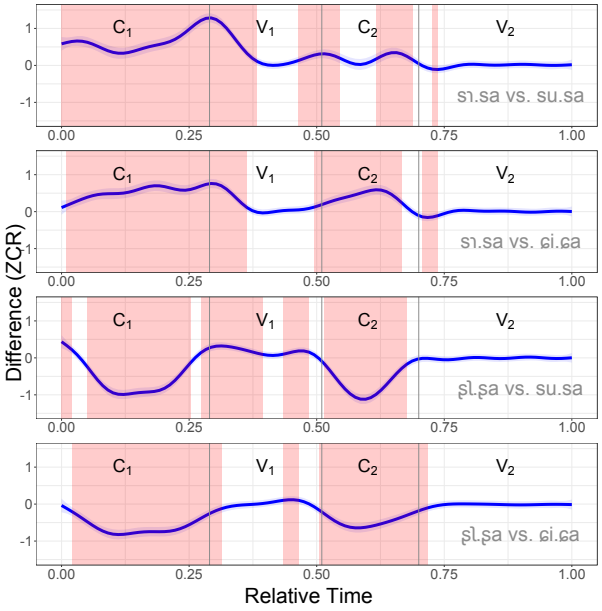

Figure 3: GAMM difference smooths for ZCR. Grey vertical lines indicate phone boundaries. Red shaded regions indicate regions of statistically significant difference.

onset of [u]. Interestingly, in the [sɻ.sa] versus [su.sa] comparison, there is also a period of difference during the formation of the second sibilant, with a more narrow constriction formed during the latter phrase. In the two comparisons between homorganic sequences, i.e. panels 2 and 4, the differences are near zero for the entirety of the duration, indicating that the changes in CD during these phrases follow very similar trajectories.

4. Discussion

To our knowledge, this study presents the first analysis of time-aligned CD and frication measures in apical vowel sequences, highlighting the complex interplay between constriction, frication, and aerodynamics in such sequences. Two major findings are evident in the results. First, during the target [$C_1V_1C_2$a] items, a considerable drop in frication occurs during apical vowels in V_1 position, following nearly the same trajectory as the other vowels examined. Second, no change in CD occurs during the apical vowels in the target sequences, as confirmed by examination of tongue posture at segment midpoints and kinematic analysis over the whole duration of the items. Interestingly, this same result occurred for the vowel [i]. These findings are surprising, starting from the expectation that such a drop in frication should be due to some lingual adjustment, perhaps an increase in channel size.

During the target sequences, speakers may turn towards some non-lingual adjustment to suppress frication during V_1 so as not to significantly interrupt the current arrangement of the articulators in anticipation of the following sibilant. Sibilants are known for requiring a precise arrangement of the articulators, with constraints put on both the tongue body and tongue front (Iskarous, Shadle, and Proctor 2011; Recasens, Pallarès, and Fontdevila 1997). One potential hypothesis is that speakers are directly manipulating the rate of airflow in the vocal tract during the apical vowels in V_1 positions. This would indicate the presence of airflow velocity targets separate from constriction degree targets, suggesting that gestural approaches to phonology that only incorporate constriction degree targets are overly simplistic (Iskarous, Shadle, and Proctor 2011; Browman and Goldstein 1989).

Alternatively, one could argue that the drop in frication dur-

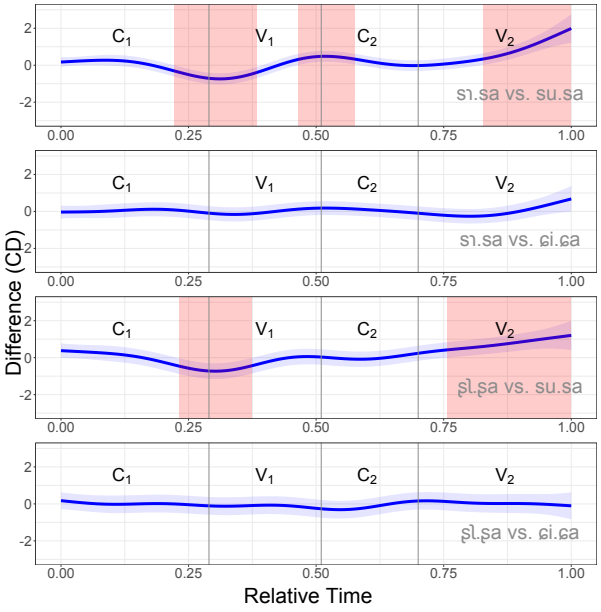

Figure 4: GAMM difference smooths for CD. Grey vertical lines indicate phone boundaries. Red shaded regions indicate regions of statistically significant difference.

ing the apical vowels and [i] is merely due to the onset of voicing. The antagonistic relationship between voicing and frication could potentially lead to a drop in the rate of turbulent airflow during the apical vowels (Ohala and Solé 2010). However, to maintain the position that the apical vowels have frication noise targets, this predicts that the overall rate of frication during the apical vowels should be *higher* than that of other vowels, as reported for the Jixi apical vowel (Shao and Ridouane 2023). The current results show no significant difference in the trajectory of frication noise during the apical vowel sequences compared to that of the other vowels. Incorporating the voiced fricative [z̩] before the apical vowel [ɻ] into the stimuli would allow for testing this hypothesis (e.g. [z̩ɻ.z̩u]). If the trajectory of frication during these sequences does not differ from those observed here, that would suggest other mechanisms are at play here.

In conclusion, this study looked at the trajectory of frication and CD during sequences containing SM apical vowels and sibilants they are homorganic with in comparison to other vowels in the same sequences. The results showed little to no adjustment in CD during the apical vowels in these sequences, with a considerable drop in frication during this same period. Given that turbulency requires both a certain channel size and airflow velocity, we hypothesize that some adjustment is suppressing the rate of airflow during these sequences. This leaves open the possiblity that speakers are directly manipulating the rate of airflow, though other adjustments are possible. Further investigation is needed in these regards.

5. Acknowledgments

This work was supported by NIH grant T32 DC009975 (Foley).

6. References

Browman, Catherine P and Louis Goldstein (1989). "Articulatory gestures as phonological units". In: *Phonology* 6.2, pp. 201–251.

Catford, John Cunnison et al. (1977). *Fundamental problems in phonetics*. Midland Books.

Davidson, Lisa (2006). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance". In: *The Journal of the Acoustical Society of America* 120.1, pp. 407–415.

Duanmu, San (2007). *The phonology of standard Chinese*. OUP Oxford.

Faytak, Matthew and Susan Lin (2015). "Articulatory variability and fricative noise in apical vowels." In: *ICPhS*.

Foley, Sean (2023). "The coarticulatory behavior of Standard Mandarin apical vowels". In: *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*.

Gu, Chong (2014). "Smoothing Spline ANOVA Models: R Package gss". In: *Journal of Statistical Software* 58.5, pp. 1–25. URL: https://www.jstatsoft.org/v58/i05/.

Iskarous, Khalil, Christine H Shadle, and Michael I Proctor (2011). "Articulatory–acoustic kinematics: The production of American English/s". In: *The Journal of the Acoustical Society of America* 129.2, pp. 944–954.

Lee-Kim, Sang-Im (2014). "Revisiting Mandarin 'apical vowels': An articulatory and acoustic study". In: *Journal of the International Phonetic Association* 44.3, pp. 261–282.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger (2017). "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi." In: *Interspeech*. Vol. 2017, pp. 498–502.

Ohala, John J and Maria-Josep Solé (2010). "Turbulence and phonology". In: *Turbulent sounds: An interdisciplinary guide*, pp. 37–97.

Recasens, Daniel, Maria Dolors Pallarès, and Jordi Fontdevila (1997). "A model of lingual coarticulation based on articulatory constraints". In: *The Journal of the Acoustical Society of America* 102.1, pp. 544–561.

Shadle, Christine H (1990). "Articulatory-acoustic relationships in fricative consonants". In: *Speech production and speech modelling* 55, pp. 187–209.

Shao, Bowei (2020). "The apical vowel in Jixi-Hui Chinese: phonology and phonetics". PhD thesis. Université Sorbonne Nouvelle.

Shao, Bowei and Rachid Ridouane (2023). "On the nature of apical vowel in Jixi-Hui Chinese: Acoustic and articulatory data". In: *Journal of the International Phonetic Association*, pp. 1–26.

Stevens, Kenneth N (1998). *Acoustic phonetics*. MIT press.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: 10.21105/joss.01686.

Wood, S. N. (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models". In: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36.

Yoshinaga, Tsukasa, Kazunori Nozaki, and Shigeo Wada (2019). "A simplified vocal tract model for articulation of [s]: The effect of tongue tip elevation on [s]". In: *PloS one* 14.10, e0223382.

Yu, Alan CL (1999). "Aerodynamic constraints on sound change: The case of syllabic sibilants". In: *The Journal of the Acoustical Society of America* 105.2, pp. 1096–1097.