THE COARTICULATORY BEHAVIOR OF STANDARD MANDARIN APICAL VOWELS

Sean Foley

University of Southern California
seanfole@usc.edu

ABSTRACT

Although not uncommon, syllabic consonants remain understudied. This study looks at gestural overlap in the apical vowels of Standard Mandarin, which have been analyzed as "syllabic approximants" and "syllabic fricatives". Using ultrasound tongue imaging, coarticulatory effects from adjacent consonants were quantified as a measure of coarticulatory resistance for the two apical vowels in comparison to the three corner vowels of the language. The results show that both apical vowels are considerably less resistant to coarticulation than all three corner vowels, while showing particular susceptibility to these effects in the tongue dorsum. The lack of tongue body stability suggests both segments lack frication noise targets, with their primary tasks involving the tongue tip or tongue blade. The low resistance to coarticulatory effects in both segments stands in contrast to previous studies on gestural overlap in syllabic consonants and is attributed to the nature of their constrictions and their low informativity.

Keywords: coarticulation, apical vowels, Standard Mandarin, ultrasound

1. INTRODUCTION

1.1. Syllabic consonants and gestural overlap

Despite not being particularly rare [1], the phonetics of syllabic consonants is not very well understood, with most studies coming from a few languages, e.g. Tashlhity Berber [2, 3] and Slovak [4]. One of the core claims of models of syllabic structure is that CV syllables are the most stable, as they allow for in-phase coordination between the onset and nucleus [5, 4, 6]. As such, syllables occupied solely by consonants should be dispreferred, leading some to hypothesize that syllabic consonants may become more vowel-like when occupying the nucleus [4, 7].

However, studies on articulatory timing and gestural overlap in syllabic consonants in comparison to onset and coda consonants have generally come to the same conclusion: position in the syllable has little effect on the relative timing and gestural coordination of a consonant – phonetically nucleic consonants are indistinguishable from their coda and onset counterparts [3, 4]. Furthermore, these studies have shown that less gestural overlap occurs between consonants when the nucleus is filled by a consonant, as opposed to onset and coda clusters [4, 3, 8]. [7] posited that the lack of gestural overlap in these cases occurs to allow for recoverability of the consonantal gestures.

In one of the few typologies of syllabic consonants, [1] states that diachronically the source of syllabic consonants is almost always a vowel, whether its due to vowel deletion or vowel "consontantalization". According to [1], this latter process is particularly rare, occurring only in Sino-Tibetan languages or Lendu (Central Sudanic) (See [9] for other potential examples). This difference suggests that syllabic consonants that arise from vowel "consonantalization" may differ phonetically from those due to vowel deletion. This study looks at the apical vowels of Standard Mandarin, which are believed to have their origin in high front vowel "apicalization" due to diachronic pressures to maintain sibilant contrasts [10, 11].

1.2. Standard Mandarin apical vowels

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] [12]. Therefore, many analyses consider these segments to be allophones of /i/ [11]. Although "apico-dental/alveolar" and "apico-postalveolar" may be potentially more accurate phonetic descriptions of the segments [9], [ɿ] will be referred to as the "dental apical vowel" and [ʅ] as the "retroflex apical vowel" in keeping with the previous literature.

Previous research has shown that both apical vowels are produced with a lingual configuration that closely resembles their onsets [11, 13], though questions remain whether this is due to an assimilatory effect or is inherent to the segments themselves. Despite some impressionist analyses that have analyzed them as "syllabic fricatives" [12, 14], recent evidence suggests that frication noise is present only during the onset of the segments and is likely attributable to the preceding sibilant [15]. As argued in [11], these facts suggest that both segments could best be analyzed as "syllabic approximants". Therefore, given previous research on syllabic consonants, sibilant plus apical vowel sequences could be seen as CC clusters, with a sequence such as [sɿ sa] analyzed as CC.CV. Given these assumptions, the aim of this study is to investigate the coordinating relations and coarticulatory behavior of both apical vowels in comparison to the three corner vowels [i a u] to assess whether or not they pattern with typical vowels in these respects and to assess the nature of their gestural targets.

Assuming that coarticulation is constrained by both the nature of the linguistic task [16] and the hydrostatic properties of the tongue [17], the coarticulatory behavior of the apical vowels should provide insight into their phonetic targets. Viewing the segments as "syllabic fricatives" predicts high resistance to coarticulatory effects in both the tongue tip and tongue body, given the strict demands on both articulators to produce frication noise [18, 19]. In contrast, the "syllabic approximant" view only predicts local resistance in the primary articulator, i.e. the tongue tip or tongue blade.

2. METHOD

2.1. Ultrasound experiment

Seven native speakers of SM took part in the study, all of whom reported no history of hearing or speech disorders. Data from two speakers was excluded due to errors in the placement of the ultrasound probe (SP_03 and SP_06). 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).

The stimuli used in the study consist of 48 disyllabic nonce words. The use of nonce words was warranted due to the lack of disyllabic words containing the target sequences. The target segments in the first syllable are the four vocalic segments [ɿ ʅ i a u], with three different onsets [s ʂ ɕ]. Due to phonotactic restrictions, each apical vowel occurs only after a homorganic sibilant, while [i] occurs only after [ɕ]. Furthermore, [u] occurs only after [s] and [ʂ]. The second syllable consists of the consonants [t^h s ʂ ɕ x k^h] with [a] serving as the nucleus. These six consonants were chosen to represent a range of both manners and places of articulation. The total disyllabic sequences includes all combinations of the target first syllable and second syllable. All syllables had Tone 1 (high level). 16 disyllabic fillers were also used, containing syllables and tones not present in the target stimuli.

Participants were recorded in a sound-attenuated booth. The stimuli were presented in blocks of five and were randomized so that each target phrase was seen a total of five times across all blocks. Different randomizations were used across all speakers , with only SP_02 and SP_07 seeing the same randomization. The stimuli were presented in the following carrier phrase：我覺得 __ 很好 [$wə^{21}$ $tɕyei^{35}$ də__$xən^{35}$ xau^{213}] "I think__(is) very good". Simultaneous ultrasound video and audio were recorded for each participant. Ultrasound images were recorded using an Articulate Instruments Micro portable ultrasound machine.

2.2. Data analysis

Audio files were analyzed in Praat [20] and segmented using the Montreal Forced Aligner [21]. All alignments were manually checked and corrected. Stop closure and bursts were not manually segmented by the forced aligner and were segmented by hand using the offset of the V_1 formants where closure starts to the C_2 offset just before the stop burst as landmarks. Ultrasound frames corresponding to the target segments were processed in Articulate Assistant Advanced (AAA) [22].

Frames closest to the midpoint of the target C_2 and V_1 tokens were analyzed using Smoothing spline (SS) ANOVAs [23] (Figure 1). The tongue root (TR), tongue dorsum (TD), and tongue tip (TT) were also tracked for all target segments using AAA (Figure 2). This was done by specifying radii in AAA that correspond to each of these parts of the tongue from the 42 radii that compose the ultrasound fan. This was done independently by speaker by observing constrictions of these regions in the ultrasound video. From these five repetitions the mean distance distane from the origin was taken at the midpoint of each target segment.

Coarticulatory resistance (CR) was measured as the degree of variability in tongue contours for a segment across all C_2 contexts [19]. The two apical vowels [ɿ ʅ] were compared to the three peripheral vowels [i a u]. This was quantified for overall variability in lingual posture and variability in the three tongue regions. Overall variability exhibited by each vocalic segment was measured using the polar coordinates that were the output of the SSANOVAs and by taking the mean variance (σ^2) of the distances across all angles, i.e. six distances per angle from six SSANOVA splines (Figure 1). This measure was found to be highly correlated with taking the mean Euclidean distance of the min and max distance across all angles [24] and fitting a polygon over the SSANOVAs and getting the area of that polygon [19]. Variability by tongue region was quantified by taking the variance of the distances for each region across all contexts. In sum, there is one global measure and three local measures of CR (i.e. for the TR, TD, and TT).

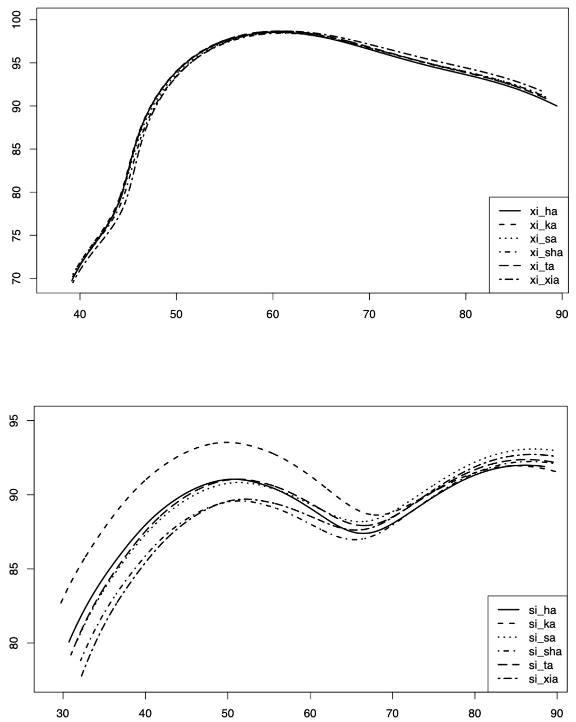

Figure 1: Midsagittal SSANOVA splines for vowels [i] (top) and [ɿ] (bottom) at their midpoints across all target consonant contexts for SP_02. Anterior is to the right.

3. RESULTS

Results are separated into overall tongue variability and variability by the three tongue regions – TR, TD, and TT. To account for across-speaker variability, all results are shown by speaker. As the overall

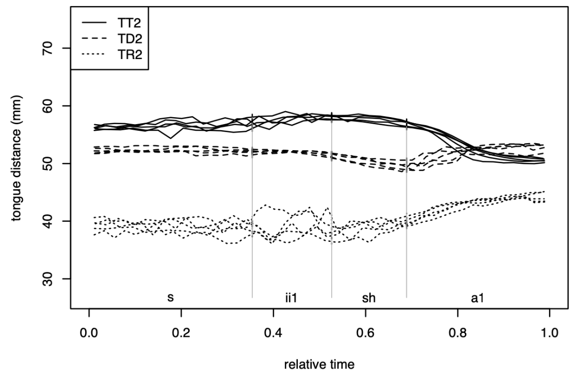

Figure 2: TR, TD, and TT trajectories during "si sha" for SP_04.

and by tongue region variability was measured using variance (σ^2), the ratio of two variances (i.e. one for each segment in a two-way comparison) was taken to get an F-statistic following an F-distribution. F-tests were computed by speaker for all pairwise combinations of the target segments grouped by onset.

3.1. Overall

Figure (3) shows the overall variance for each target vowel segment, separated by the four onset groups, with [i ɿ ʅ] grouped as "homorganic onset" due to each being homorganic with their onsets, i.e. [ɕ s ʂ] respectively. Taller bars indicate greater variance (lower CR) and shorter bars indicate less variance (higher CR). It can be seen for four of the five speakers that the apical vowels consistently show significantly more variability than the three corner vowels ($p < 0.05$). SP_05 behaved differently from the others in this regard, and is indicated as *SP_05. Furthermore, between the two apical vowels, the retroflex segment was consistently more resistant to coarticulatory effects than the dental segment.

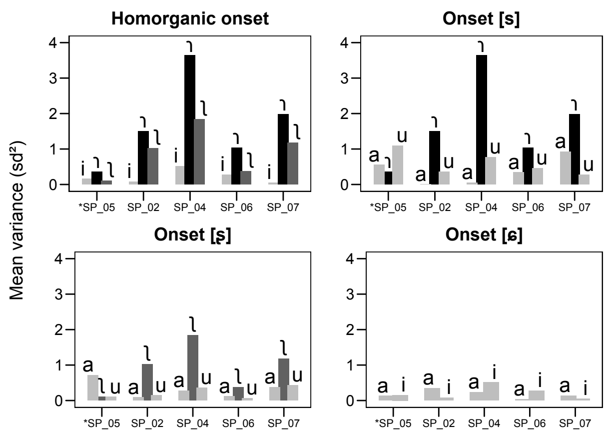

Figure 3: Overall CR for all target segments, organized by shared onset and by speaker.

3.2. By tongue region

Figure (4) shows CR plotted by speaker and by the three tongue regions: TR, TD, and TT, and grouped by onset. A number of observations can be made here. First, four of the five speakers show a tendency for both apical vowels to exhibit greater variability than the three corner vowels in both the TD and TR, with both apical vowels showing much greater variability in the TD than the other two tongue regions. Second, for most speakers, the dental segment [ɿ] has significantly greater TD variance than [i], but this is not always true for [ʅ]. TR variability in both apical vowels is often significantly greater than all three corner vowels, but this is not consistent across speakers. Third, between the two apical vowels, the retroflex segment was consistently more resistant to coarticulatory effects than the dental segment in all three regions. Lastly, linear regression models also showed that for three out of five speakers, the relationship between TT and TD variability is significantly different for the apical vowels compared to the three corner vowels ($p < 0.05$).

4. DISCUSSION

The results show that both apical vowels are considerably less resistant to coarticulatory effects than all three corner vowels. This holds both for overall tongue variability as well as variability in both the TD and TR. While both segments are much more resistant to coarticulatory effects in the TT than other regions, across speakers they were not consistently less resistant to effects in the TT than all three corner vowels. Together, these results suggest that the primary task for both apical vowels is a tongue tip or tongue blade raising gesture, with the lack of tongue body stability in their productions being indicative of both segments not having frication noise targets. This supports previous analyses of the segments as "syllabic approximants".

Previous studies of syllabic consonants have suggested that syllabic consonants permit less gestural overlap to allow for full recoverability of the consonantal gestures [4, 25]. One potential reason for the lack of resistance in the apical vowels is that the information conveyed by the apical vowels is essentially redundant as most of the cues they serve to enhance are already present in the preceding sibilant [11]. In addition, compared to the syllabic consonants of other studies, e.g. /l/ and /r/ in Slovak [4], the constriction tasks of the apical vowels likely require less tongue body activation [26].

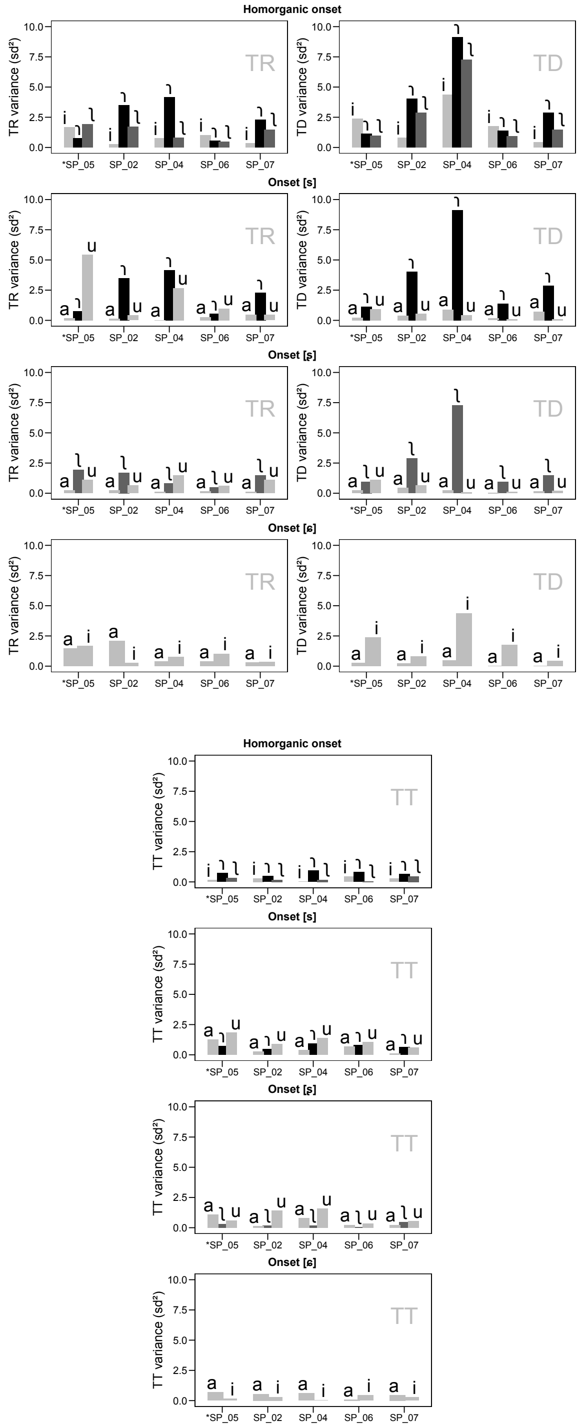

Figure 4: CR by tongue region for all target segments, organized by shared onset and by speaker.

The results are consistent with the view that coarticulatory resistance is highly local [27]. This suggests that rather than viewing segments as "resistant", resistance exists at regions in the vocal tract essential to the task, thus permitting variability in non-task regions [28].

5. ACKNOWLEDGEMENTS

Thanks are due to Jeff Mielke for technical assistance and Elliott Moreton, Jennifer L. Smith, Louis Goldstein, Khalil Iskarous, and members of UNC-Chapel Hill P-Side and USC PhonLunch for helpful comments and discussion.

6. REFERENCES

[1] A. Bell, "Syllabic consonants," in *Universals of human language*, J. Greenberg, Ed. Stanford: Stanford University Press, 1978, vol. 2: Phonology, pp. 153–201.

[2] R. Ridouane, "Syllables without vowels: Phonetic and phonological evidence from tashlhiyt berber," *Phonology*, vol. 25, no. 2, pp. 321–359, 2008.

[3] C. Fougeron and R. Ridouane, "On the phonetic implementation of syllabic consonants and vowel-less syllables in tashlhiyt," *Estudios de Fonética Experimental*, pp. 140–175, 2008.

[4] M. Pouplier and Š. Beňuš, "On the phonetic status of syllabic consonants: Evidence from slovak," *Laboratory phonology*, vol. 2, no. 2, pp. 243–273, 2011.

[5] L. Goldstein, D. Byrd, and E. Saltzman, "The role of vocal tract gestural action units in understanding the evolution of phonology," *Action to language via the mirror neuron system*, pp. 215–249, 2006.

[6] H. Nam, L. Goldstein, and E. Saltzman, "Self-organization of syllable structure: A coupled oscillator model," *Approaches to phonological complexity*, vol. 16, pp. 299–328, 2009.

[7] P. Hoole, M. Pouplier, S. Benus, and L. Bombien, "Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: Data and modelling," *Rhotics. New data and perspectives*, pp. 81–98, 2013.

[8] R. A. W. Bladon and A. Al-Bamerni, "Coarticulation resistance in english/l," *Journal of Phonetics*, vol. 4, no. 2, pp. 137–150, 1976.

[9] M. Faytak, "Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese," Ph.D. dissertation, University of California, Berkeley, 2018.

[10] S. Baron, "On the tip of many tongues: Apical vowels across sino-tibetan," in *7th International Conference on Sino-Tibetan Language and Linguistic Studies*, 1974.

[11] S.-I. Lee-Kim, "Revisiting mandarin 'apical vowels' : An articulatory and acoustic study," *Journal of the International Phonetic Association*, vol. 44, no. 3, pp. 261–282, 2014.

[12] S. Duanmu, *The phonology of standard Chinese*. OUP Oxford, 2007.

[13] M. Faytak and S. Lin, "Articulatory variability and fricative noise in apical vowels." in *ICPhS*, 2015.

[14] A. C. Yu, "Aerodynamic constraints on sound change: The case of syllabic sibilants," *The Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 1096–1097, 1999.

[15] M. Faytak and B. Shao, "Degrees of frication in three chinese varieties' fricative vowels," 2022, paper presented at the *LSA Annual Meeting*.

[16] P. A. Keating *et al.*, "The window model of coarticulation: articulatory evidence," *Papers in laboratory phonology I*, vol. 26, pp. 451–470, 1990.

[17] W. M. Kier and K. K. Smith, "Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats," *Zoological journal of the Linnean Society*, vol. 83, no. 4, pp. 307–324, 1985.

[18] M. Tabain, "Variability in fricative production and spectra: Implications for the hyper-and hypo-and quantal theories of speech production," *Language and speech*, vol. 44, no. 1, pp. 57–93, 2001.

[19] D. Recasens and C. Rodríguez, "A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound," *Journal of Phonetics*, vol. 59, pp. 58–75, 2016.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]," *http://www. praat. org/*, 2021.

[21] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: Trainable text-speech alignment using kaldi." in *Interspeech*, vol. 2017, 2017, pp. 498–502.

[22] A. Wrench, "Articulate assistant advanced user guide," *[Computer Program]. Version*, vol. 2, p. 12, 2012.

[23] L. Davidson, "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 407–415, 2006.

[24] N. Zharkova and N. Hewlett, "Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using english /t/ and /a/," *Journal of Phonetics*, vol. 37, no. 2, pp. 248–256, 2009.

[25] I. Chitoran, L. Goldstein, and D. Byrd, "Gestural overlap and recoverability: Articulatory evidence from georgian," *Laboratory phonology*, vol. 7, no. 4-1, pp. 419–447, 2002.

[26] D. Recasens, M. D. Pallarès, and J. Fontdevila, "A model of lingual coarticulation based on articulatory constraints," *The Journal of the Acoustical Society of America*, vol. 102, no. 1, pp. 544–561, 1997.

[27] K. Iskarous, C. Mooshammer, P. Hoole, D. Recasens, C. H. Shadle, E. Saltzman, and D. H. Whalen, "The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance," *The Journal of the acoustical society of America*, vol. 134, no. 2, pp. 1271–1282, 2013.

[28] P. J. Jackson and V. D. Singampalli, "Statistical identification of articulation constraints in the production of speech," *Speech Communication*, vol. 51, no. 8, pp. 695–710, 2009.