Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Articulatory variations in Apical Vowels in Southwestern Mandarin

Jing Huang[1], *Feng-fan Hsieh*[2], *Yueh-chin Chang*[2]

[1]City University of Macau
[2]National Tsing Hua University
jinghaung@cityu.edu.mo, siahonghoan@faculty.hss.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Abstract

This study examines the acoustic and articulatory characteristics of apical vowels in Southwestern Mandarin (SWM) using electromagnetic articulography (EMA). Data were collected from seven speakers in their twenties, focusing apical vowels, the high vowel [i], and their *er*-suffixed forms. Two distinct patterns emerged: Type A speakers demonstrated differential tongue configurations for apical vowels (*si* with dental/alveolar position, *shi* with alveolo-palatal position), while Type B speakers showed similar configurations for both vowels. Type A findings align with previous research on Standard Chinese, confirming that apical vowels share place features with preceding sibilants and showing higher F2 values for *shi*. However, Type B revealed an intriguing articulatory-acoustic mismatch: despite showing significant differences in horizontal Tongue Tip and Tongue Body positions, these vowels exhibited no significant formant differences. This unexpected pattern suggests a more complex relationship between articulatory configuration and acoustic output in apical vowel production than previously documented.

Index Terms: EMA, ultrasound, apical vowels, Southwestern Mandarin

1. Introduction

In Mandarin Chinese, the high vowel [i] does not occur after the dental and retroflex sibilants. Instead, these sibilants are followed by "apical vowels," which share their places of articulation ([1, 2, 3]) and are widely distributed across Chinese languages ([4]). As a central topic in Chinese phonetics and phonology ([4]), "apical vowels" have been extensively studied through both acoustic analysis and articulatory investigations. Formant measurements by [5, 6] revealed that the dental segment consistently exhibits a lowered F2 compared to its retroflex counterpart, indicating a mismatch between acoustic and articulatory properties. Subsequently, an ultrasound and acoustic study by [7] demonstrated that the F2 pattern alone cannot explain the apical position, as F2 is influenced by the cavity behind the apical constriction. The study further established that these segments are homorganic with the preceding dental and retroflex sibilants and can be transcribed as "dental approximant [ɹ̩]" and "retroflex approximant [ɻ̩]" respectively. While previous studies have primarily focused on Beijing Mandarin or Standard Chinese ([1, 2, 5, 6, 7]), recent studies have expanded to other Sinitic languages, including Shanghai Wu Chinese, Hefei Mandarin, and Jixi-Hui Chinese ([3, 4, 8, 9]). These investigations have shown that the acoustic and articulatory characteristics of these segments vary across Sinitic languages ([10]). The present study aims to investigate the acoustic and articulatory properties of the segments in Southwestern Mandarin. While various terms exist for these segments (including "fricative vowels" [2], "syllabic fricatives" [11], and "syllabic approximants" [12]), the present study treats them as syllabic alveolar and retroflex approximants ([ɹ̩], [[ɻ̩]) while using the conventional term "apical vowels ([ɿ], [ʅ])" throughout.

Southwestern Mandarin (SWM), with over 250 million native speakers, is the most widely spoken variety of Mandarin Chinese ([13]). As one of the eight groups of Mandarin Chinese, it is predominantly spoken across nine provinces in Southwest China. According to [14], the "alveolar vs. retroflex" contrast in Mandarin sibilants has been lost in most SWM varieties, resulting in no perceptible distinction between [ɿ] and [ʅ]. However, some SWM varieties maintain this contrast. This study addresses two research questions: first, it examines the acoustic and articulatory characteristics of apical vowels in these two contrasting patterns within SWM varieties (dubbed Types A and B). Second, it investigates the apical vowels in varieties without the "alveolar vs. retroflex" contrast, specifically exploring whether articulatory differences exist between these vowels despite the absence of perceived distinctions.

The perception of apical vowels across Mandarin Chinese dialects presents an intriguing paradox: while these sounds may be impressionistically indistinguishable, substantial evidence reveals distinct articulatory patterns across various dialects. Research has demonstrated that Beijing Mandarin (BJM), Northeastern Mandarin (NEM), and Southwestern Mandarin (SWM) exhibit notable differences in their articulation of apical vowels ([9]). Specifically, while SWM and NEM share a bunched tongue configuration for retroflex apical vowels, BJM employs a retroflexed, tip-up tongue shape. The variation is even more pronounced in dental apical vowels, where BJM demonstrates a grooved tongue shape, NEM shows a retracted tongue dorsum, and SWM maintains a relatively flat tongue configuration.

Building on these findings, this research examines articulatory and acoustic characteristics of apical vowels across distinct SWM varieties. A key focus is the detailed articulatory analysis of two apical vowels in the SWM variety that lacks the conventional "alveolar versus retroflex" contrast. This investigation serves two important purposes: it expands our understanding of how apical vowels are articulated across SWM varieties, while also revealing how apical vowels differ articulatorily in varieties where the typical contrast is absent. The findings will contribute to our broader understanding of

10.21437/Interspeech.2025-1810

phonological variation and articulatory patterns in Chinese dialectology.

2. Experiment

This study examines apical vowels in a Southwestern Mandarin (SWM) variety from the Chengdu-Chongqing dialect group (known as "Cheng-yu" in Chinese dialectological terminology) spoken in western Hubei. Seven native speakers in their twenties participated in the electromagnetic articulography (EMA) study. Five speakers (S01-S05) maintain the "alveolar versus retroflex" sibilant distinction, while two speakers (P01, P02) do not. All participants were native to Hubei with no reported speech or hearing disorders.

The experiment collected both CV and *er*-suffixed syllables, comparing apical vowels with the high vowel [i]. The recording materials comprised monosyllables and *er*-suffixed disyllables: *si* [sɿ], *shi* [ʂʅ], *pi* [p^{h}i], *(si)sier* [sɿɚ], *shier* [ʂʅɚ], and *(pi)pier* [p^{h}iɚ]. Participants read a randomized list of target words embedded in the carrier phrase "__, pa __ pa," meaning "__, give __ Sentence Final Particle" at a normal speech pace from a computer screen in a soundproof recording room.

For each token, six repetitions were recorded, with analysis focusing on the second instance of each stimulus in the carrier phrase. Articulatory data were collected using the NDI wave system at a 200 Hz sampling rate to track lip and tongue movements in time and space. Acoustic data were simultaneously recorded using a Sound Devices Mixpre mixer and a Sennheiser unidirectional shotgun microphone at a 24 kHz sampling rate.

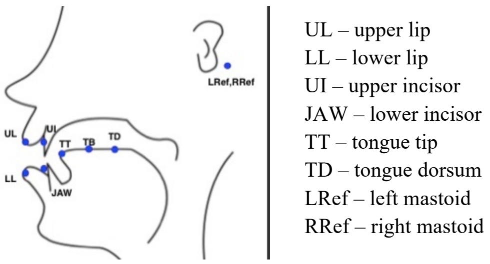


Figure 1: *Experiment setup: the placement of the EMA sensors.*

In the present study, EMA data were analysed using Mview [15], while acoustic data were annotated by Praat [16]. Generalized Additive Mixed Models (GAMMs) were applied for statistical analysis ([17]).

3. Results

3.1. Comparing Apical Vowels Across Two SWM Varieties

In this section, we present the articulatory, acoustic, and statistical results of both varieties: one that distinguishes between "alveolar" and "retroflex" sibilants, and one that does not. We also conduct a comparison of the apical vowels in these two patterns.

3.1.1. EMA results

We demonstrate tongue movements from EMA results of representative speakers in each variety. Figure 2 illustrates a comparison of apical vowels and high vowel [i] (*si*, *shi,* and *pi*) examining the temporal changes in lingual configuration for native speakers of the two varieties with different apical vowel patterns. S01 and P01 are the representative speakers of their respective varieties. As shown in Figure 2, it is evident that the apical vowels exhibit distinct tongue shapes and positions in S01. Specifically, *si* shows a closer tongue tip to the dental or alveolar position, while *shi* has a tongue tip position closer to the alveolo-palatal region. This clear distinction of the tongue position confirms the previous findings that the apical vowels share the same place of articulation with their preceding sibilants ([1, 2, 7]). In contrast, *si* and *shi* share similar tongue shape and position in P01, the variety that does not distinguish between "alveolar vs. retroflex" sibilants. Additionally, both apical vowels show less tongue movements compared to the high vowel [i] in both varieties.

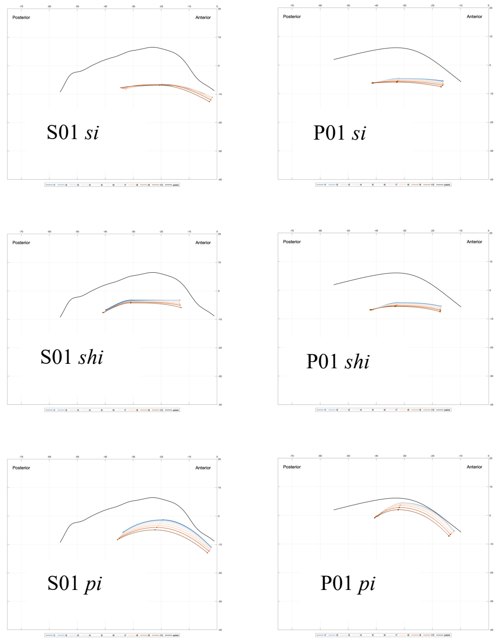


Figure 2: *Temporal changes in lingual configuration during the production of the rimes of si, shi, and pi by speakers S01 and P01. Subject orientation is rightward, with blue lines representing articulation onset and red lines representing endpoint.*

In addition to the visualized lingual movement, we also measured the quantitative differences in articulatory trajectories of EMA data using Generalized Additive Mixed Models (GAMMs). Following [17], we compared differences among articulatory trajectories of EMA sensors by applying the *bam* function from the *mgcv* package [18] in *R* [19], with EMA sensors analysed at a 95% confidence level. As shown in Figure 3, which serves as a sample of the GAMM results, the trajectories of Tongue Tip (TT), the Tongue Body (TB), and the Tongue Dorsum (TD) in both the horizontal (x, front-back) and vertical (z, up-down) dimensions are compared. Figure 3 illustrates that the trajectories of TTx, TTz, and TBx are significantly different in *si* vs. *shi* in Type A. Note that the GAMM analyses are based on normalized data, specifically z-score transformed kinematic data, and the statistically significant difference is indicated by an x-axis in red.

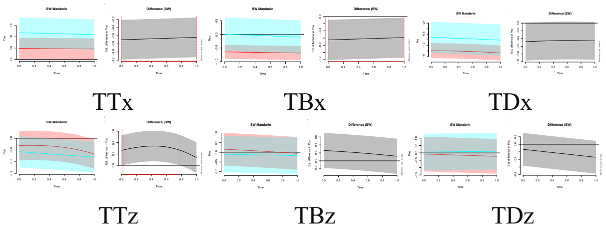


Figure 3: *GAMM models of the EMA sensor trajectories of the dental apical vowel si vs. the retroflex apical vowel shi in SWM Type A.*

The overall results for SWM Type A are summarized in the Tables 1 and 2. The primary difference between *si* and *shi* occurs in the horizontal dimension, particularly in the positioning of the Tongue Tip (TT) and Tongue Body (TB). In contrast, the difference between apical vowels and the high vowel [i] is mostly in the vertical dimension. When comparing the results in Tables 1 and 2, the unsuffixed and suffixed form exhibit similar patterns in the comparison, suggesting that the position has minimal impact on the quality of apical vowels ([20]).

Table 1: *GAMMs results of SWM Type A.*

Type A	**TT**	**TB**	**TD**
si vs. *shi*	TTx/z	TBx	n.s.
pi vs. *si*	TTz	TBz	TDz
pi vs. *shi*	TTz	TBx	TDx/z

Table 2: *GAMMs results of the er-suffixed form in SWM Type A.*

Type A	**TT**	**TB**	**TD**
(si)sier vs. *shier*	TTx/z	TBx	TDx/z
(pi)pier vs. *(si)sier*	TTz	n.s.	TDz
(pi)pier vs. *shier*	TTz	TBz	n.s.

The overall results comparing both types are summarized in the Table 3. While no significant differences were observed in the dental apical vowel between Types A and B. However, the other apical vowel showed significant differences across all three sensors in the horizontal dimension.

Table 3: *GAMMs results of the comparison of Type A and B.*

Type A vs. B	**TT**	**TB**	**TD**
si	n.s.	n.s.	n.s.
shi	TTx	TBx	TDx

3.1.2. Acoustic results

For the acoustic data, we applied GAMMs to compare the formants of each pair. Figure 4 illustrates the formant comparison between *si* and *shi* in Type A. In Figure 4, *shi* consistently exhibits significantly higher F2 than *si*, corresponding to acoustic patterns previously reported for Standard Chinese apical vowels ([6, 7]).

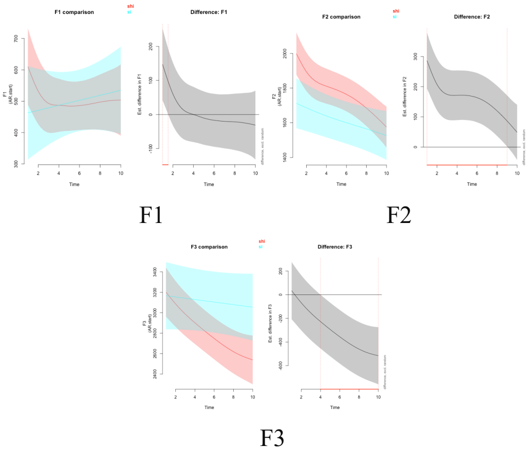


Figure 4: *GAMM models of the three formants of the dental apical vowel si vs. the retroflex apical vowel shi in SWM Type A.*

The overall results comparing both types are summarized in the Table 4. *Shi* shows significant differences in F2, whereas *si* does not. This finding aligns with the articulatory trajectory data, where *shi* demonstrates significant difference in tongue position along the horizontal (front-back) dimension between Types A and B.

Table 4: *GAMMs results of the acoustic comparison of Type A and B.*

Type A vs. B	**F1**	**F2**	**F3**
si	√	n.s.	n.s.
shi	n.s.	√	√

3.2. Apical Vowel Patterns in “Non-Contrasting” SWM Varieties

In this section, we examine the articulatory and acoustic patterns of apical vowels in Type B, the variety that lacks the “alveolar” and “retroflex” sibilant contrast.

3.2.1. EMA results

In section 3.1.1, we observed that *si* and *shi* share similar tongue positions and configurations. However, the overall results for SWM Type B, summarized in Tables 4-5, reveal an unexpected finding: despite the impressionistic observations noted earlier, significant articulatory differences exist between the two apical vowels in Type B.

Table 4: *GAMMs results of SWM Type B*

Type B	**TT**	**TB**	**TD**
si vs. *shi*	TTx/z	TBx	n.s.
pi vs. *si*	TTz	TBz	TDz
pi vs. *shi*	TTz	TBx	TDx/z

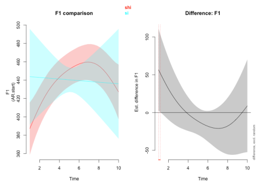
Table 5: *GAMMs results of the er-suffixed from in SWM Type B*

Type B	**TT**	**TB**	**TD**
(si)sier vs. *shier*	TTx	n.s.	n.s.

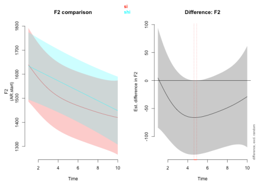
(pi)pier vs. *(si)sier*	TTz	TBz	TDz
(pi)pier vs. *shier*	n.s.	TBx/z	TDz

3.2.2. Acoustic results

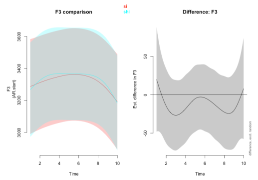
As illustrated in Figure 5, no significant differences appeared across the three formants between the two apical vowels in Type B. Unlike Type A, *si* and *shi* show no significant difference in the second formant values. However, Figure 5 reveals that *shi* maintains slightly higher F2 values than *si*.



F1



F2



F3

Figure 5: *GAMM models of the three formants of the dental apical vowel si vs. the retroflex apical vowel shi in SWM Type B.*

4. Discussion

4.1.1. Acoustic and articulatory patterns of apical vowels across varieties

For Type A, which maintains the alveolar-retroflex contrast, our findings largely corroborate previous results reported for Standard Chinese ([6, 7]). The mismatch between acoustic and articulatory patterns in *si* and *shi* stems from *shi* having a more posterior tongue position than *si*, while simultaneously showing consistently higher second formant (F2) values. As demonstrated in section 3.1, significant articulatory differences between *si* and *shi* occur along the horizontal (front-back) dimension in both Tongue Tip (TT) and Tongue Body (TB). Compared to the high vowel [i], *si* differs primarily in the horizontal dimension, while *shi* shows differences in both horizontal (front-back) and vertical (up-down) dimensions. Results for suffixed forms parallel those of unsuffixed forms, suggesting position has minimal impact on apical vowel quality.

When comparing corresponding apical vowels across Types A and B, the dental apical vowel *si* shows similar articulatory and acoustic characteristics in both varieties, with no significant differences observed. However, *shi* in Type B differs significantly from its Type A counterpart both articulatorily and acoustically. These differences primarily occur along the horizontal dimension, with Type A *shi* occupying a more posterior position than its Type B counterpart.

4.1.2. Comparison of apical vowels within Type B

The results reveal a mismatch between articulatory and acoustic characteristics when comparing apical vowels in Type B. While *si* and *shi* show significant differences in Tongue Tip (TT) and Tongue Body (TB) positioning, their formant patterns show no significant differences. This finding is particularly interesting given the impressionistic observation that *si* and *shi* are indistinguishable in Type B, though our acoustic analysis does show higher F2 values for *shi*. Such acoustic-articulatory mismatches are well-documented in phonetics, as distinct articulatory configurations can achieve similar acoustic outcomes, as demonstrated in studies of English /ɹ/ ([20, 21]).

Our findings support previous research showing that seemingly similar sounds may employ distinct articulatory strategies despite being perceptually indistinguishable. Furthermore, the comparative analysis across varieties reveals that while *si* maintains similar patterns in both Type A and B, the tongue position for *shi* in Type B appears to occupy an intermediate position between Type A's *si* and *shi*. This positioning suggests that the articulatory distinctions between Type B's apical vowels do not manifest acoustically to the same degree as in Type A.

5. Conclusion

This paper presents a comparative study of apical vowel articulation in two Southwestern Mandarin varieties using electromagnetic articulography (EMA). The experimental results reveal distinct lingual movement patterns for apical vowels across both varieties. In Type A, speakers demonstrate different tongue configurations for apical vowels, with *si* showing a dental/alveolar tongue tip position and *shi* an alveolo-palatal position. In contrast, Type B speakers display similar tongue configurations for both vowels. Both varieties exhibit reduced tongue movement for apical vowels compared to the vowel [i].

GAMM analysis of EMA sensor trajectories reveals significant cross-varietal differences for *shi* in the horizontal dimension across Tongue Tip, Tongue Body, and Tongue Dorsum positions. Acoustically, Type A shows significantly higher F2 values for *shi* compared to *si*, while Type B exhibits no significant F2 differences between apical vowels. Notably, Type B demonstrates an articulatory-acoustic mismatch: significant differences in horizontal Tongue Tip and Tongue Body positions exist despite the absence of significant formant differences.

Our findings for Type A align with previous research on Standard Chinese apical vowels, confirming that apical vowels share place of articulation with their preceding sibilants. The position of *shi* in Type B appears to occupy an intermediate position between Type A's *si* and *shi*, suggesting a unique articulatory strategy in varieties lacking the alveolar-retroflex contrast. Furthermore, our analysis of *er*-suffixed forms demonstrates that syllable position has minimal impact on apical vowel quality, as similar patterns persist across both morphological contexts. These results contribute to our understanding of how articulatory variations can maintain acoustic stability, even when canonical phonological contrasts are neutralized.

6. Acknowledgements

This study was partially funded by MOST Grant 109-2410-H007-061, awarded to the third author and was approved by the Research Ethics Committee of the National Tsing Hua University (10712HS104 and 10612HS089).

7. References

[1] Y.-R. Chao, *The non-uniqueness of phonemic solutions of phonetic system, Readings in linguistics*. New York: American Council of Learned Societies, 1963.

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

[3] H. Kong, S. Wu, M. Li, and X. Shen, "The articulatory properties of apical vowels in Hefei Mandarin," *Journal of the International Phonetic Association*, vol. 54, no. 21, pp.165-188, 2024.

[4] F. Hu and L. Feng, "Fricative vowels as an intermediate stage of vowel apicalization," *Language and Linguistics*, vol. 20, no.1, pp.1-14, 2019.

[5] J. M. Howie, *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press, 1976.

[6] E. Zee and W.-S. Lee, "An acoustical analysis of the vowels in Beijing Mandarin," *7th European Conference on Speech Communication and Technology*, Aaborg, Denmark, 2001, pp. 643–646.

[7] S.-I. Lee-Kim, "Revisiting Mandarin 'apical vowels' An articulatory and acoustic study," *Journal of the International Phonetic Association*, vol. 44(3), pp, 261-282, 2014.

[8] B. Shao and R. Ridouane, "Apical vowels are not vowels: acoustic and ultrasound evidence from Jixi-Hui Chinese," *In: 12th International Seminar on Speech Production*, 2021.

[9] J. Huang, F. F. Hsieh and Y. C. Chang, "A Cross-Dialectal Comparison of Apical Vowels in Beijing Mandarin, Northeastern Mandarin and Southwestern Mandarin: An EMA and Ultrasound Study," in *Proc. INTERSPEECH 2021 – 22rd Annual Conference of the International Speech Communication Association*, Brno, Czechia, Sep. 2021, pp. 3989-3993.

[10] B. Shao, *The apical vowel in Jixi-Hui Chinese: phonology and phonetics*. Diss. Université de la Sorbonne nouvelle-Paris III, 2020.

[11] S. Duanmu, *The Phonology of Standard Chinese*, OUP Oxford, 2007.

[12] W.-S. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, vol. 33, no.1, pp. 109-112, 2003.

[13] L. Li, "Liushi nian lai Xinanguanhua de diaocha yu yanjiu [The survey and research on Southwestern Mandarin during the last sixty years]," *Fangyan*, vol. 4, pp. 249–257, 1997.

[14] J. Huang, F. F. Hsieh, Y. C. Chang, and M. Tiede, On the two rhotic schwas in Southwestern Mandarin: when homophony meets morphology in articulation. *Phonetica*, vol. 81, no. 1, pp. 43-80, 2024.

[15] M. Tiede, "MVIEW: software for visualization and analysis of concurrently recorded movement data," *New Haven, CT: Haskins Laboratories*, 2005.

[16] P. Boersma, "Praat: doing phonetics by computer (Version 5.1.05)," *http://www. praat. org/*, 2009.

[17] M. Wieling, "Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English," *Journal of Phonetics*, vol. 70, pp. 86-116, 2018.

[18] S. Wood, Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. Available: https://cran.r-project.org/web/packages/mgcv/mgcv.pdf, 2019.

[19] R. Ihaka and G. Robert, "R: a language for data analysis and graphics," *Journal of computational and graphical statistics*, vol. 5, no. 3, pp. 299-314, 1996.

[20] J. R. Westbury, M. Hashi, and M. J. Lindstrom, Differences among speakers in lingual articulation for American English/ɹ. *Speech Communication*, vol. 26, no. 3, pp.203-226, 1998.

[21] J. Mielke, A. Baker, and D. Archangeli, Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In *Laboratory Phonology*, De Gruyter Mouton, vol.10, pp. 699-730, 2010.