INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A cross-dialectal comparison of apical vowels in Beijing Mandarin, Northeastern Mandarin and Southwestern Mandarin: An EMA and ultrasound study

Jing Huang, Feng-fan Hsieh, Yueh-chin Chang

Institute of Linguistics, National Tsing Hua University, Taiwan, R.O.C.
captainmolly94@gmail.com, siahonghoan@faculty.hss.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Abstract

This paper is a comparative study of the articulation of the "apical vowels" in three Mandarin dialects: Beijing Mandarin (BJM), Northeastern Mandarin (NEM), and Southwestern Mandarin (SWM), using co-registered EMA and ultrasound. Data from 5 BJM speakers, 5 NEM speakers and 4 SWM speakers in their twenties were analyzed and discussed. Our recording materials include the dental and retroflex apical vowels, and their *er*-suffixed forms. Results suggest that distinct lingual configurations are found among the three dialects of Mandarin, even though these apical vowels are not perceptually distinguishable. Specifically, the dental apical vowel [ɿ] has a grooved tongue shape in BJM, a retracted tongue dorsum in NEM, and a relatively flat tongue shape in SWM. The retroflex apical vowel [ʅ] has a domed tongue shape as well as a bunched tongue body in NEM, while a slightly domed tongue posture is found in SWM. Moreover, the retroflex apical vowel [ʅ] is, articulatorily speaking, very similar to the *er*-suffix in BJM (cf. [10]). In sum, we observed yet another instance of the articulatory-acoustic mismatch.

Index Terms: EMA, ultrasound, articulation, apical vowel, Beijing Mandarin, Northeastern Mandarin, Southwestern Mandarin

1. Introduction

It is well-known that Beijing Mandarin (BJM) has the so-called "apical vowels," which only appear after the dental and retroflex sibilants (i.e., [ɿ], [ʅ] respectively). Diachronically speaking, it is not controversial that the apical vowels were derived from the high front vowel /i/ in Middle Chinese. Synchronically, nevertheless, the consonantal gesture continues throughout the following vocalic portion, resulting in a homorganic CV combination, according to [1]'s ultrasound data from one BJM speaker. In the literature, various descriptions were proposed for these sounds; for example, they may be "apical vowels" ([2]), "syllabic fricatives" ([3] and [4]), "fricative vowels" ([5]), or "syllabic approximants" ([6]). Of particular importance is that [1], [2], [7], [8] and [9] all observed that there are clear formant structures in these sounds and the tongue dorsum is raised to form a vowel articulation. In addition, the apical vowels are in allophonic variation with the high front vowel /i/. Consequently, the general consensus is that these sounds are better treated as a (special) type of vowel. In contrast, [4] contends that these sounds should be described as syllabic fricatives because the unique gesture of these sounds is completely different from typical vowel articulation. These sounds can be created only via feature spreading from the sibilant onset to the (empty) nucleus. In the present study, following [1] and [6], we treat these typologically rare sounds as syllabic alveolar and retroflex approximants ([ɹ̩], [ɻ̩], respectively), primarily because of the absence of frication noise in them. Nevertheless, we continue to use the more familiar term "apical vowels" throughout this paper.

The main research question of this study is to examine the articulatory characteristics of the apical vowels in other dialects of Mandarin Chinese in addition to Beijing Mandarin (BJM): namely, Northeastern Mandarin (NEM) and Southwestern Mandarin (SWM). NEM, SWM and BJM are three distinct subdialects under the eight major groups of Mandarin Chinese. SWM is mainly spoken in Sichuan, Chongqing, Yunnan, Guizhou, and western Hubei, while NEM is mostly spoken in Heilongjian, Jilin, Liaoning, and eastern parts of Neimenggu. Previous studies are predominantly concerned with the case of BJM (or, Standard Chinese), while the apical vowels of the other dialects of Mandarin Chinese are largely understudied. This study helps to fill that gap.

There is good evidence indicating that similar sounds may be articulatorily distinct among different varieties of Mandarin Chinese, even though impressionistically speaking they are not distinguishable at all. As a matter of the fact, [10] reports that BJM and NEM are different in the articulations of the *er*-suffixation (i.e., diminutives); specifically, the *er*-suffix in BJM is retroflex whereas the same suffix is realized with a bunched tongue shape in NEM, even though these rhotic suffixes are perceptually indistinguishable. In light of this, the present study aims at exploring whether there are articulatory and acoustic differences among the apical vowels in these Mandarin dialects. In addition, since both alveolar and retroflex apical vowels may be *er*-suffixed if in a stem, we will present preliminary results of these *er*-suffixed forms as a follow-up study of [10]. To this end, the present study also helps us better understand potential diversity of articulatory properties of both apical vowels and *er*-suffixation across different Mandarin dialects.

2. Experiment

The present study is primarily concerned with a cross-dialectal comparison of the apical vowels in NEM, SWM, and BJM. Five native speakers of BJM (BJ01, BJ02, BJ03, BJ04, BJ05) and NEM (NE01, NE02, NE03, NE04, NE05), and four native speakers of SWM (SW01, SW02, SW03, SW04) participated in the experiments. All participants (in their twenties) were born and raised in their respective dialect speaking areas and have no reported history of speech and/or hearing disorders.

The recording materials are comprised of both the apical vowels (i.e., [sɿ], [ʂʅ]; *si*, *shi* in Pinyin) and their *er*-suffixed forms (i.e., [sɿɚ], [ʂʅɚ]; *sir*, *shir* in Pinyin). The participants were asked to read a randomized list of the target words from a computer screen at a normal speech rate at the sound-proof recording booth in the Phonetics Laboratory at the National Tsing Hua University. For BJM and NEM speakers, the target

Copyright © 2021 ISCA

http://dx.doi.org/10.21437/Interspeech.2021-1326

words were embedded in the carrier phrase, “__, *mà* __ *ba*”, meaning “__, scold __ (Sentence Final Particle)”. For SWM speakers, the target words were embedded in the carrier phrase, “__, *ba* ___ *ba*”, meaning “___, give ___(Sentence Final Particle)” in SWM. For BJM and NEM, five repetitions were collected for each token, and for SWM, six repetitions were collected for each token. Only the second occurrences of the stimuli in a carrier phrase were analyzed. For BJM and NEM speakers, the EMA and ultrasound experiments are conducted separately, while the articulatory data of SWM were collected using co-registered EMA and ultrasound. We used an NDI Wave to track flesh-point movements in time and space and a Micro system from Articulate Instruments Ltd. to capture holistic tongue shape configurations. Acoustic data was recorded simultaneously with a sampling rate of 24 kHz, using a Sennhesier unidirectional shotgun microphone and a Sound Devices Mixpre mixer. The EMA data were sampled at a rate of 200 Hz. The ultrasound data were collected using a transducer with a 92° field of view, set at a depth of 120mm. The frame rate was 65 f.p.s. The participants were asked to wear an all-plastic Ultrafit probe stabilizer and blue dots were put on the probe, the forehead, chin, etc. for the subsequent processes of head correction.

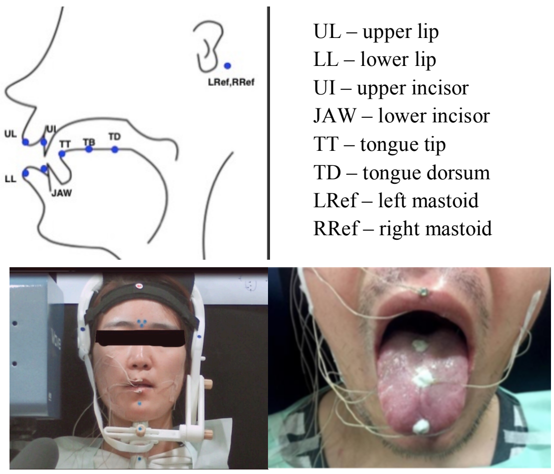

Figure 1. *Experiment setup: Ultrasound probe stabilizer and positions of the EMA sensors*

In this study, the EMA data were analyzed with the help of Mview [11] and the ultrasound images were annotated and analyzed by GetContours [12]. Praat [13] was used for acoustical analysis.

3. Results

3.1. Ultrasound imaging results

In this section, we demonstrate ultrasound imaging results from representative speakers in each variety. Fig. 2 illustrates a comparison of *si* and *shi* with respect to the temporal changes of the lingual configurations in BJM, NEM and SWM. Recall from section 1 that the apical vowels are perceptually not distinguishable in the three dialects. As shown in Fig. 2, it is obvious that both of the apical vowels exhibit distinct tongue configurations across the three Mandarin dialects. Regarding *si*, we witness a retracted tongue dorsum in NEM, a lowered tongue tip and a bunched tongue body in BJM, and a relatively flat tongue shape in SWM. As for *shi*, there is a pitting of the lower part of the tongue in NEM, a domed tongue shape in BJM, and a slightly domed tongue shape in SWM. In each dialect, we can see that a contrast of the tongue shapes is maintained between *si* and *shi*, although the distinction is implemented in a different fashion.

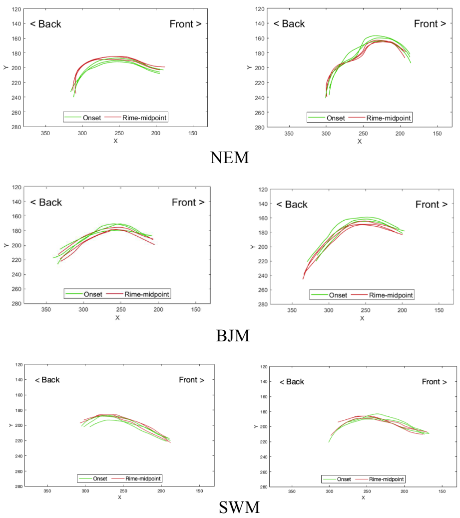

Figure 2. *Temporal changes of lingual configurations of si vs. shi (The participants are facing right; green line = starting point and red line = midpoint).*

Fig. 3 illustrates the *er*-suffixed *si* and *shi*. For *sir*, the tongue body/blade is raised in NEM, the tongue tip is raised in BJM, and a raising of the tongue tip and the tongue body/blade is involved in SWM. As for the retroflex apical vowel *shi*, there appears a grooved tongue shape towards the end of the rime in NEM, and there seems to be no significant tongue movement in both BJM and SWM, though the tongue shapes are quite different between the BJM and SWM speakers. In NEM, BJM and SWM, we observe a relatively higher tongue tip for *sir*, if compared with *shir*. In other words, we can say that the tongue tip is consistently raised across the three dialects in the production of *sir*.

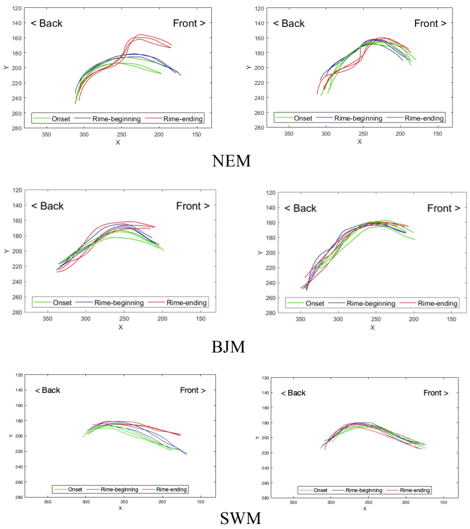

Figure 3. *Temporal changes of lingual configurations of the er-suffixed si and shi {sir, shir} .(Where green line = starting point, bule line = midpoint and red line = endpoint).*

Fig. 4 gives a direct comparison of *si* and *shi* with their respective *er*-suffixed forms. As we can see, the present data from the three dialects confirm that *er*-suffixed *sir* involves a tongue tip raising across the board. In contrast, comparing *shi* with *shir*, we do not observe significant tongue movement in any of the data. In particular, the tongue shapes of *shi* and *shir* are almost overlapped (i.e., the red lines and dark blue lines in Fig. 4), indicating that the retroflex apical vowel may share the same articulatory posture with the *er*-suffix. As a matter of fact, this homorganicity has already been described in [14]'s transcription, according to which the *er*-suffix in BJM is a retroflex apical vowel /ʅ/, which undergoes the so-called fusional process once it is suffixed to a non-front vowel. In SWM, however, we observe a tongue dorsum retraction, when the retroflex apical vowel and its er-suffixed from are compared.

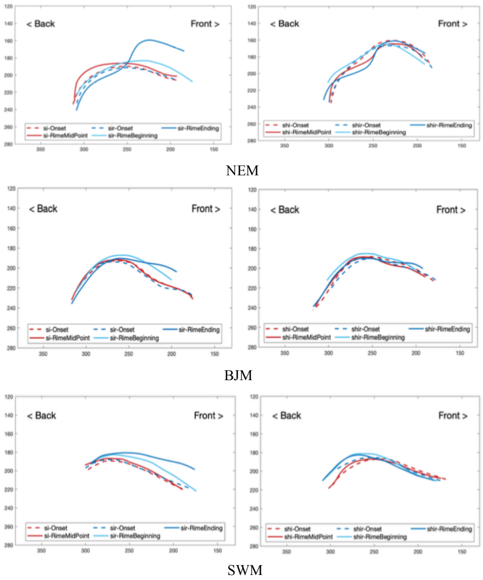

Figure 4. *The lingual configurations of apical vowels and er-suffixed forms {si, sir, shi, shir}. (Where red dashed line = starting point of apical vowel; red solid line = midpoint of apical vowel; blue dashed line = starting point of suffixed form; light blue solid line = midpoint of suffixed form; dark blue solid line = endpoint of suffixed form)*

In sum, the ultrasound data reported in this section confirm that both the dental and retroflex apical vowels exhibit distinct tongue configurations among NEM, BJM, and SWM, even though they are, to a great extent, highly similar in perception.

3.2 EMA results

In addition to the ultrasound data was collected in this study, EMA experiments were conducted separately for BJM and NEM, while the EMA data of SWM were collected using co-registered EMA and ultrasound. The EMA technology is advantageous in that the kinematics of the flesh points (or, articulators, e.g., TT (Tongue tip), TB (Tongue body), and TD (Tongue dorsum) and so on), can be accurately tracked and analyzed accordingly. Consequently, the EMA data are examined in this section for the sake of cross-validation of the ultrasound results in section 3.1.

It is well-known that the *er*-suffix is realized differently when suffixed to different vowel stems in BJM (or, Standard Chinese) (e.g., [15]), and that its counterparts in NEM are quite similar (see, e.g., [10]) as well. Moreover, [16]'s EMA study of the *er*-suffixation in SWM indicates that only the high vocoids, including the apical vowels, are not deleted after *er*-suffixation occurs. That being the case, Figs. 5-6 illustrate the temporal lingual configurations of the *er*-suffixed forms of the dental apical vowel *sir* in the three varieties. Regarding the suffixed form *sir*, the SWM and NEM speakers share a similar pattern, even though the tongue movement is not so significant among the SWM speakers. Specifically, we can see in Figs. 5, 6 that a retracted TD and a raising TT are attested in both NEM and SWM, plus a bunched tongue shape (having positive Retroflex Angle (RA), a measure of degree of retroflexion proposed in [17]). In sharp contrast to this, BJM exhibits a completely different pattern, whereby a retroflex tongue shape is attested (whose RA is negative). As for the dental apical vowel *si*, along with the identical observations made in the ultrasound data, the tongue configurations varied between the three Mandarin dialects. Furthermore, intra-speaker variation is also found (which is not illustrated due to the limitation of space).

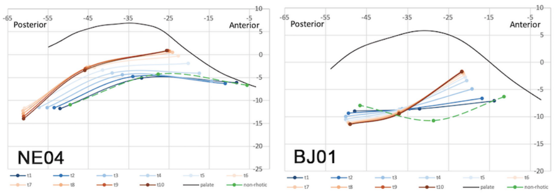

Figure 5. *The temporal lingual configurations of the dental apical vowel and it er-suffixed forms {sir} of NEM and BJM speakers (The participants are facing right; blue line = vowel onset of sir, red line is the endpoint of sir, and the green dash line is the non-rhotic s).*

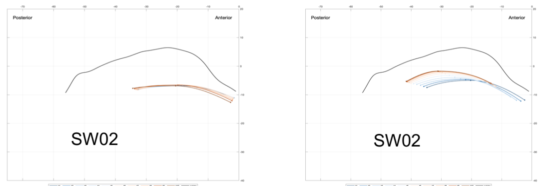

Figure 6. *The temporal lingual configurations of the dental apical vowel {si} (in the left) and the dental apical vowel's er-suffixed forms {sir} (in the right) of SWM speakers.*

Regarding Figs. 7-8, for BJM and NEM, we can see that the green dashed lines are largely overlapped with the blue lines in NEM and with the red lines in BJM. The results confirmed the ultrasound results reported in section 3.1, namely that the *er*-suffixation may share an identical articulatory posture with the retroflex apical vowel. Furthermore, both BJM and SWM speakers show a TT raising in *er*-suffixation.

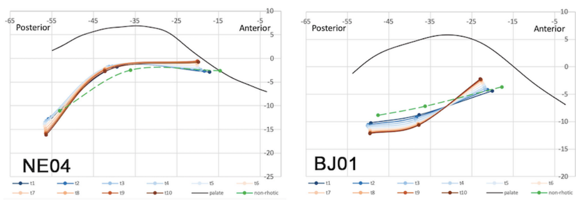

Figure 7. *The temporal lingual configurations of the retroflex apical vowel and its er-suffixed forms {shir} of NEM and BJM speakers.*

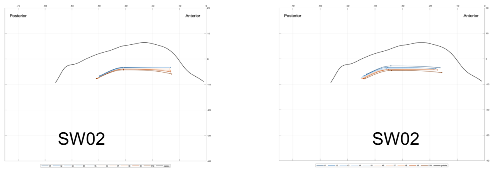

Figure 8. *The temporal lingual configurations of the retroflex apical vowel {shi} (in the left) and the retroflex apical vowel's er-suffix form {shir} (in the right) of SWM speakers.*

3.3 Quantifying differences in articulatory trajectories using Generalized Additive Mixed Models (GAMMs)

Following [18], we used the *bam* function from the *mgcv* package [19] in R [20], which allows for the inclusion of factor smooths to represent full random effects, to construct the GAMMs of our experimental data to compare the differences among articulatory trajectories of the EMA sensors at a confidence level of 95%. As a sample illustration of the GAMM results, in Fig. 14, the trajectories of the Tongue Tip (TT), the Tongue Body (TB), and the Tongue Dorsum (TD) are compared, where *x* stands for the front-back dimension and *z* stands for the up-down dimension. We can see in Fig. 9 that the trajectories of TTx, TBx, and TDx are significantly different in /si/ 'to tear' vs. /(si.)si-er/ 'gracile' in SWM. Note that the GAMM analyses are based on *z*-score transformed kinematic data and an *x*-axis in red means a statistically significant difference.

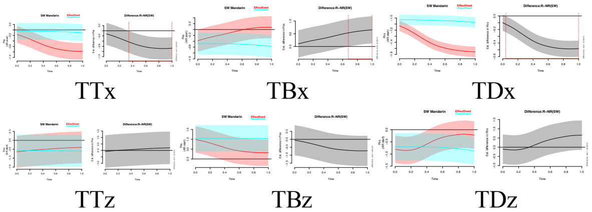

Figure 9. *GAMM models of the EMA sensor trajectories of the dental apical vowel vs. the er-suffix form /sir/ in SWM.*

The overall results for the apical and *er*-suffixed forms are summarized in the Tables 1-3.

Table 1. *GAMMs results: NEM*

Unsuffixed vs. *er*-suffixed form	**TT**	**TB**	**TD**
The dental apical vowel	TTx/z	TBx/z	TDx
The retroflex apical vowel	n.s.	TBx	n.s.

Table 2. *GAMMs results: BJM*

Unsuffixed vs. *er*-suffixed form	**TT**	**TB**	**TD**
The dental apical vowel	TTx/z	TBx	TDx
The retroflex apical vowel	n.s.	n.s.	n.s.

Table 3. *GAMMs results: SWM*

Unsuffixed vs. *er*-suffixed form	**TT**	**TB**	**TD**
The dental apical vowel	TTx	TBx	TDx
The retroflex apical vowel	n.s.	n.s.	n.s.

(Note that when an articulator appears in a cell, that means the unsuffixed form is higher (z) or more fronted (x) than that of the *er*-suffixed form.)

3.4 Acoustical data

As revealed in both EMA and ultrasound studies, [ɿ] and [ʅ] are articulatorily distinct among different varieties of Mandarin Chinese, even though impressionistically speaking they are not distinguishable at all. Here we show the acoustical data to see if there is any difference of apical vowels in the three Mandarin varieties. Table 4 illustrates the averaged formant values (F1, F2, and F3) at the midpoint of a steady state. The Linear Mixed Model (LMM) results indicate that there is no significant difference between F1, F2, and F3 of both the dental and retroflex apical vowels in these three varieties of Mandarin.

Table 4. *Acoustical results: BJM, NEM and SWM ($\bar{X}$: mean value in Hertz; SD: Standard deviation)*

		F1	**F2**	**F3**
		$\bar{X}$(SD)	**$\bar{X}$(SD)**	**$\bar{X}$(SD)**
[ɿ]	BJM	436 (45.04)	1542 (152.9)	3124 (179.9)
	NEM	386 (25.03)	1481(90.99)	2757 (178.8)
	SWM	504 (29.83)	1626 (70.36)	3128 (158.9)
[ʅ]	BJM	456 (34.30)	1825 (166.3)	2569 (287.0)
	NEM	393 (19.78)	1663 (95.43)	2278 (337.3)
	SWM	544 (28.49)	1818 (63.17)	2865 (277.1)

4. Discussion

The major finding of this study is that the tongue configurations are quite different among the three Mandarin dialects, even though these apical vowels are quite similar acoustically. The dental apical vowel [ɿ] has a grooved tongue shape in BJM, a retracted tongue dorsum in NEM, and a relatively plain tongue shape in SWM. The retroflex apical vowel [ʅ] has a domed tongue shape in BJM, a clearly bunched tongue body in NEM, and a slightly domed tongue posture in SWM. To this end, our results seem to suggest that there is a relatively stable perceptual goal for the apical vowels despite the fact that distinct articulatory postures are implemented in these closely related varieties of Mandarin.

5. Conclusion

This paper is a comparative study of the articulation of the "apical vowels" in three Mandarin dialects, of which two dialects(NEM and SWM) belong to relatively understudied dialectal groups of Mandarin Chinese. With the help of EMA and ultrasound technologies, our experimental results reveal that the lingual configurations of the apical vowels are quite distinct among these dialects, even though there is no acoustical difference. Of particular interest is the discovery that, just like in the case of English /r/ sound ([21], [22], [23]), the apical vowels in Mandarin dialects also share a stable acoustic goal, albeit with distinct articulatory variations.

6. Acknowledgments

This study was supported by research grants (MOST 103-2410-H-007-036-MY3 and 109-2410-H-007-061-), to which we are grateful. This study has been approved by the National Tsing Hua University's Research Ethics Committee (#10712HS104 and #10612HS089).

7. References

[1] 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

[2] B. Karlgren, Etudes sur la phonologie chinoise. Uppsala: K. W. Appelberg, 1915-1916

[3] R. Wiese, Underspecification and the description of Chinese vowels. *Studies in Chinese phonology*, pp, 219-249, 1997

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

[5] P. Ladefoged, & I. Maddieson, *The sounds of the world's languages*. Oxford: Blackwell, vol. 1012, 1996

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

[7] E. Zee, & W.-S. Lee, An acoustical analysis of the vowels in Beijing Mandarin. In *Seventh European Conference on Speech Communication and Technology*, 2001

[8] F. F. S. Hsueh, *Beijing Yinxi Jiexi [Analysis of the Beijing dialect sound system]*. Beijng: Beijing Yuyan Xueyuan Chubanshe, 1986.

[9] Y. W. Wu, Mandarin segmental phonology. Ph.D. dissertation, University of Toronto, 1994

[10] S. Jiang, Y.-C. Chang, and F.-F. Hsieh, "An EMA study of er-suffixation in Northeastern Mandarin monophthongs," In *the Proceedings of International Congress of Phonetic Sciences 2019, Melbourne: Australasian Speech Science and Technology Association Inc*, 2019, pp. 3617-3621.

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

[12] M. Tiede, and D. H. Whalen, "Getcontours: An interactive tongue surface extraction tool," *Proceedings of Ultrafest VII*, 2015.

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

[14] S. J. Li, *Hanyu 'Er' /ɚ/ Yinshi Yanjiu [A historical analysis of Chinese /ɚ/]*, Beijing: The Commercial Press, 1986.

[15] W.-S. Lee, A phonetic study of the "er-hua" rimes in Beijing Mandarin. In *Ninth European conference on speech communication and technology*, 2005

[16] J. Huang, F.-F. Hsieh, & Y.-C. Chang, *Er*-suffixation in Southwestern Mandarin: An EMA and ultrasound study. *Proc. Interspeech 2020*, pp, 661-665, 2020

[17] M. Tiede, W. R. Chen, and D. H. Whalen, Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and Ultrasound. *In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.)Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.* Canberra, Australia: Australasian Speech Science and Technology Association Inc, 2019

[18] 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.

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

[20] 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.

[21] J. R. Westbury, M. Hashi, and M. J. Lindstrom, Differences among speakers in lingual articulation for American English/ɹ. *Speech Communication*, vol.26(3), pp.203-226, 1998.

[22] 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.

[23] S. Boyce, M. Tiede, C. Y. Espy-Wilson, and K. Groves-Wright, Diversity of tongue shapes for the American English rhotic liquid. *ICPhS*, 2015.