



Er-suffixation in Southwestern Mandarin: An EMA and ultrasound study

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

Institute of Linguistics, National Tsing Hua University

s107044871@m107.nthu.edu.tw, siahonghoan@faculty.hss.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Abstract

This paper is an articulatory study of the *er*-suffixation (a.k.a. *erhua*) in Southwestern Mandarin (SWM), using co-registered EMA and ultrasound. Data from two female speakers in their twenties were analyzed and discussed. Our recording materials contain unaffixed stems, *er*-affixed forms and the rhotic schwa [ə], a phonemic vowel in its own right. Results suggest that the *er*-suffixation in SWM involves suffixing a rhotic schwa [ə] to the stem, unlike its counterpart in Beijing and Northeastern Mandarin [5]. Specifically, an entire rime will be replaced with the *er*-suffix if the nucleus vowel is non-high; only high vocoids will be preserved after the *er*-suffixation. The “rhoticity” is primarily realized as a *bunched* tongue shape configuration (i.e. a domed tongue body), while the Tongue Tip gesture plays a more limited role in SWM. A phonological analysis is accordingly proposed for the *er*-suffixation in SWM.

Index Terms: EMA, ultrasound, articulation, *er*-suffixation, rhotic sounds, Southwestern Mandarin

1. Introduction

The *er*-suffixation (a.k.a. *erhua*) is perhaps one of the most famous morphophonological processes in Mandarin Chinese. Diachronically speaking, this suffix was primarily derived by means of attaching *er* ‘son’ to a stem to form diminutives of nouns, although in contemporary Standard Chinese, *erhua* has been often, if not always, semantically bleached to a certain extent. Previous studies have overwhelmingly focused on the *er*-suffixation in Standard Chinese (or, Beijing Mandarin). For example, [1]’s seminal work on modern Chinese grammar describes the *er*-suffix in Beijing Mandarin as a “retroflex morpheme” and transcribes *erhua* as /l/. Later, [2] holds that *erhua* is actually /l/ (i.e. a retroflex apical vowel) and proposes that there are five exponents of it, namely, {əl, əl, ə¹, ə², ə³}. [3] and [4] claim that the *er*-suffix in Standard Mandarin is a (subsyllabic) rhotic schwa [ə] and may be featurally represented as coronal-[+retroflex] ([4]). More recently, an electromagnetic articulography (EMA) study of the *er*-suffixation in both Beijing and Northeastern Mandarin reveals that the *er*-suffix may be, details aside, composed of two gestures, namely the anterior of lingual gestures (i.e. Tongue Tip and/or Tongue Body) and the Tongue Dorsum gesture ([5]). However, little is known about the *er*-suffixation in other varieties of Mandarin Chinese. As has been discussed in [6], it is likely the *er*-suffixation may not be uniformly implemented across different dialects of Mandarin Chinese (cf. the /l/ sound in English). The present study helps fill the gap by investigating articulatory characteristics of the *er*-suffixation in Southwestern Mandarin (hereafter SWM).

SWM is one of the eight subgroups of Mandarin Chinese, mainly spoken in Sichuan, Chongqing, Yunnan, Guizhou, and western Hubei. It has been noted in [7] and [8] that the *er*-suffixation involves much fewer alternations in SWM than in

Beijing Mandarin. The generalization is quite straightforward, at first sight. If the stem vowel/prenuclear glide is [+high], suffix a subsyllabic [ə] to the stem (after deleting the nasal coda, if any). Otherwise, delete the entire rime and replace it with a subsyllabic [ə] to derive an *er*-affixed form. For example, /pa/ ‘give’ → pa.pa → [pa.pə] ‘handle’ vs. /p^{hi}/ ‘peel’ → p^{hi}.p^{hi} → [p^{hi}.p^{hi}ə] ‘skin’. The starting point of this study is, thus, to investigate whether previous impressionistic transcriptions can be instrumentally confirmed. In addition, the present study further helps us better understand the potential diversity of articulatory properties across Mandarin dialects with respect to the *er*-suffixation.

2. Experiment

The present study is concerned with the *er*-suffixation in a variety of the Chengdu-Chongqing dialect group (or, “Cheng-yu” group in Chinese dialectological terminology) spoken in western Hubei. Two (2) female native speakers (S1, S2) of SWM in their twenties participated in the ultrasound and EMA study. Both of them were born and raised in Hubei and have no reported history of speech or hearing disorders.

The recording materials are comprised of 91 meaningful words, including (i) unaffixed forms, (ii) *er*-affixed forms and (iii) the phonemic rhotic schwa. Note that the *er*-affixed forms we investigated in the experiment are derived via total reduplication from a monosyllabic base and subsequent *er*-suffixation, e.g. /pa/ ‘give’ → pa.pa → pa.pa.er → [pa.per] ‘handle’. The syllable structures of the stimuli include CV, CGV, CVG and C(G)VN, where C = {p, t, k, t^h}, G = {j, w, ɥ}, V = {a, i, u, e, ə, y} and N = {n, ŋ}. In addition, six tokens for the phonemic rhotic schwa, e.g. /ə.two/ ‘ear’ are also included in order to conduct a comparison among different rhotic sounds in SWM.

The participants were asked to read a randomized list of the target words from a computer screen at normal speech rate at the sound-proof recording booth at the Phonetics Lab, National Tsing Hua University. The target words were embedded in the carrier phrase, “__, pa __ pa”, meaning “__, give __ Sentence Final Particle” in SWM. Six repetitions were collected for each token and only the second occurrences of a stimulus in the carrier phrase were analyzed (as they are in phrase-medial position). The articulatory data were collected using co-registered EMA and ultrasound. We used an NDI wave system 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.

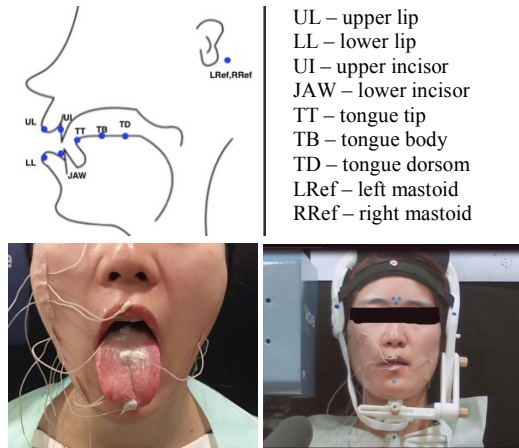


Figure 1. *Experimnt setup: Positions of EMA sensors and Ultrasound probe stabilizer*

In this study, the EMA data were analyzed with the help of Mview [9] and the ultrasound images were annotated and analyzed by GetContour [10]. Praat [11] was used for acoustical analysis.

3. Results

3.1. EMA results

3.1.1 *Er*-suffixed vs. unsuffixed forms

Figure 2. illustrates a comparison of unsuffixed forms and *er*-suffixed forms with respect to the temporal changes of the lingual configurations. The *er*-suffixed forms here include {pa, pai, pan} in the stems and the unsuffixed form is a monosyllabic {pa}. Recall from section 1 that the *er*-suffixed forms involve total reduplication plus *er*-suffixation. Firstly, the data in Fig. 2 indicate that these *er*-suffixed syllables are not distinguishable from one another regarding the tongue postures, suggesting that complete neutralization occurs in these *er*-suffixed forms (see 3.3 for statistical results). Secondly, inter-speaker variation can be found, too. There is significant Tongue Tip (TT) raising for S2 but that is not the case for S1 (as confirmed by the GAMM results). Thirdly, the data confirm that the stem vowels in {pa, pai, pan} are totally replaced with the *er*-suffix because it is obvious that the rimes {a, ai, an} are absent in the *er*-suffixation.

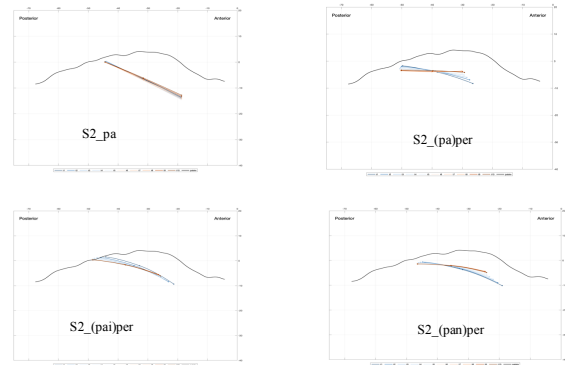
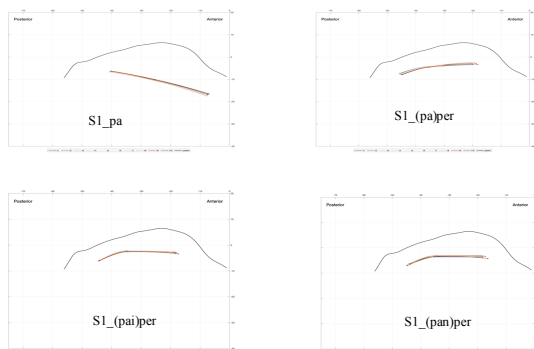


Figure 2. *Temporal changes of lingual configurations of suffixed vs. unsuffixed forms (The participants are facing right; blue line = starting point and red line = endpoint).*

Likewise, similar results can be observed in Fig. 3. Firstly, the stem vowel /e/ is completely deleted (ditto for /ɔ/, not shown here). Secondly, the *er*-suffixed forms of {p^he, pei, pen} have similar tongue postures. Data from S2 are omitted from now on if there is no substantial inter-speaker variation.

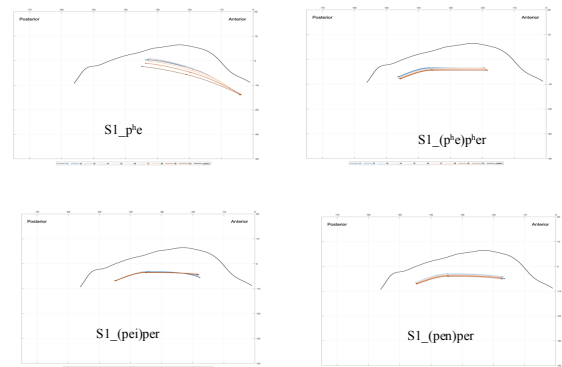


Figure 3. *Temporal changes of lingual configurations of unsuffixed and er-suffixed forms {p^he, pei, pen}.*

In contrast, we can see in Figs. 4-6 that a gliding motion of the tongue occurs if the rime has a high vowel {i, y, u} or a prenuclear glide/onglide. More specifically, only high vocoids, including prenuclear glides/onglides {j, w, ɥ}, survive in the *er*-suffixation, resulting in a noticeable gliding motion from front to back position.

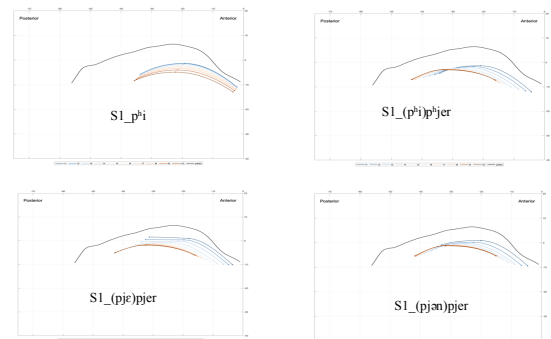


Figure 4. *Temporal changes of lingual configurations of unsuffixed and er-suffixed forms {p^hi, pje, pjən}.*

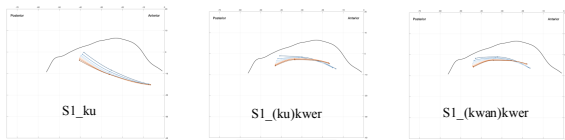


Figure 5. Temporal changes of lingual configurations of unsuffixed and suffixed forms {ku, kwan}.

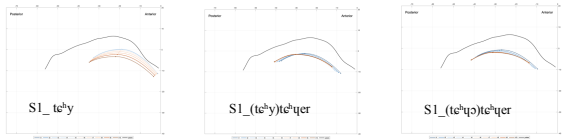


Figure 6. Temporal changes of lingual configurations of unsuffixed and suffixed forms {te^hy, te^hy ə}.

In sum, the EMA data reported in this section confirm the impressionistic transcriptions in previous studies ([7], [8]). The *er*-suffix may be transcribed as a rhotic schwa [ə] and the underlying rimes are deleted in *er*-suffixation, except for high voicoids such as {i, y, u} and prenuclear glides {j, w, ɥ}. Finally, inter-speaker variation is found, too.

3.1.2 The phonemic rhotic schwa versus the *er*-suffix

Given that the *er*-suffix is a rhotic schwa, we take a further step in this section to describe the articulatory characteristic of the phonemic rhotic schwa [ə_p] (where p is for “phonemic”). The comparisons are given in Fig. 7. At first blush, the *er*-suffix and the phonemic rhotic schwa are quite similar. In section 3.3, nevertheless, statistical results reveal that the *er*-suffix and the phonemic rhotic schwa primarily differ in the Tongue Tip (TT) gesture along both the horizontal and vertical dimensions.

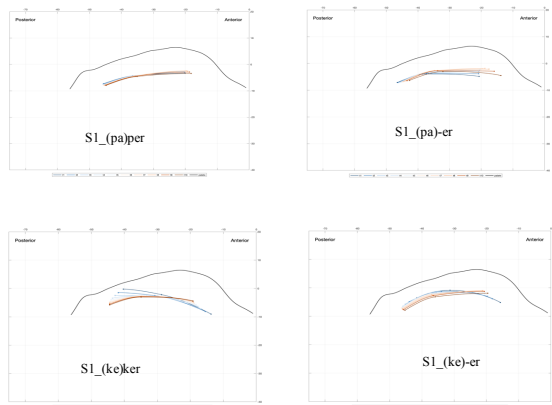


Figure 7. Temporal changes of lingual configurations of the *er*-suffix (left) and the phonemic [ə_p] (right) for S1.

3.2 Ultrasound results

In this study, the kinematic data were collected using co-registered EMA and ultrasound. The ultrasound imaging is advantageous in that the entire tongue shape may be captured for analysis, rather than discontinuous fleshpoints in EMA. In other words, ultrasound technology helps us conduct a holistic comparison among tongue postures in various contexts. Returning to the triplet {pa, pai, pan} (cf. Fig. 2), the ultrasound results in Fig. 8 confirm that the *er*-suffix exhibits a *bunched* tongue shape configuration. Comparing Figs.8 and 9, we can

see more robust gliding motions in Fig. 9, which is attested for both speakers (S1 and S2), along with the identical observations made in the EMA study in section 3.1.1.

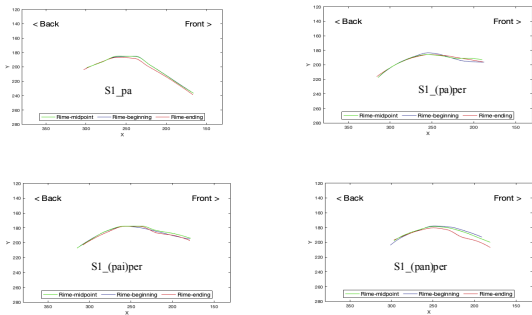


Figure 8. Ultrasound lingual configurations of unsuffixed vs. *er*-suffixed forms {pa, pai, pan} for S1 (The participants are facing right; blue line = vowel onset, green line = midpoint and red line = endpoint of a rime).

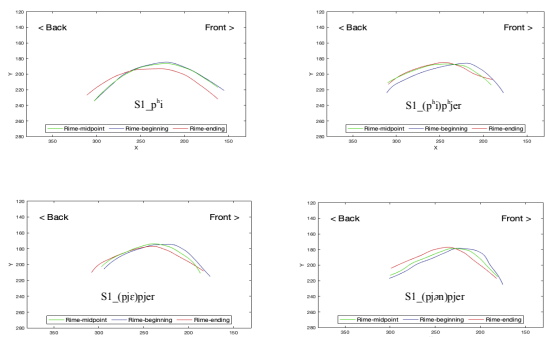


Figure 9. Ultrasound lingual configurations of unsuffixed vs. *er*-suffixed forms {p^hi, pje, pjan} for S1.

A comparison of the *er*-suffix and the phonemic rhotic schwa is shown in Fig. 10. Again, the ultrasound results here are quite similar to the EMA results reported in section 3.1.2, although it is much clearer from the present ultrasound data that both have the bunched tongue shape configurations only.

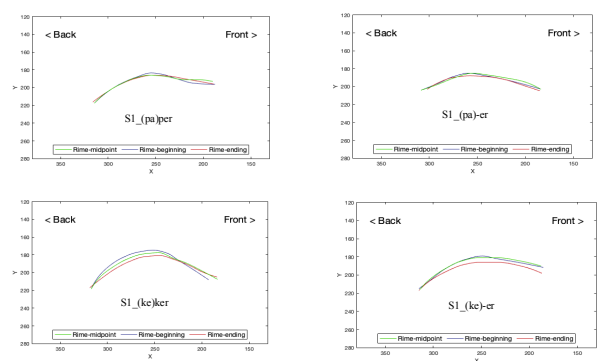


Figure 10. Ultrasound lingual configurations of the *er*-suffix (left) and the phonemic [ə_p] (right) for S1.

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

Following [12], the *bam* function from the *mgcv* package [14] in R [15], which allows for the inclusion of factor smooths to represent full random effects, is used to construct the GAMMs of our data to track the statistical differences in articulatory trajectories of the EMA sensors at a confidence level of 95%. The GAMM analyses are based on *z*-transformed kinematic data. The main effect is set to be the suffixed vs. unsuffixed forms and dependent variable the normalized positions for each EMA sensor. As a sample illustration of the GAMM results, in Fig 11, the trajectories of the sensors for 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 vertical dimension. Note that an *x*-axis in red means a statistically significant difference. We can see in Fig. 11 that the trajectories of TT*x*, TB*x*, TD*x*, TT*z* and TB*z* are significantly different in /pa/ ‘give’ vs. /pa.pa-er [pə-] ‘handle’

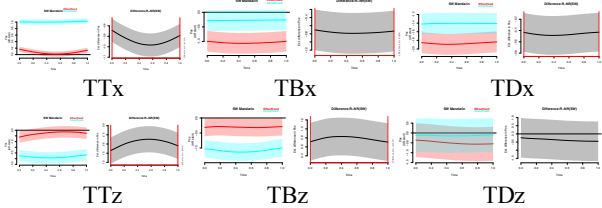


Figure 11. GAMM models of EMA sensor trajectories of the unsuffixed /pa/ ‘give’ and the *er*-suffix in /pa.pa-er [pə-] ‘handle’.

The overall results for the suffixed vs. *er*-suffixed forms are summarized in Table 1.

Table 1 Unsuffixed vs. *er*-suffixed vowels: GAMMs results

Unsuffixed vs.	TT	TB	TD
<i>Er</i> -suffixed Non-high V	TT <i>x/z</i>	TB <i>x/z</i>	TD <i>x</i>
<i>Er</i> -suffixed High V	TT <i>x</i>	TB <i>x</i>	TD <i>x</i>

Regarding the phonemic rhotic schwa [ə_r], our results show that [ə_r] is basically not distinguishable from the *er*-suffixed forms {pa.per, ke.ker}, suggesting that these two rhotic sounds are essentially identical in articulation if the stem vowel is non-high. In contrast, if the vocoids in the stem are high, then the GAMM results indicate that there are significant differences in the trajectories of TT*x* and TT*z*, suggesting that in particular, TT-raising occurs in the environments of *er*-suffixed high vocoids in SWM.

Table 2 Rhotic schwa vs. the *er*-suffix: GAMMs results

Rhotic schwa vs.	TT	TB	TD
<i>Er</i> -suffixed Non-high V	n.s.	n.s.	n.s.
<i>Er</i> -suffixed High V	TT <i>x/z</i>	n.s.	n.s.

4. Discussion

Our findings largely corroborate previous results reported in the impressionistic studies ([7], [8]). If the stem vowel is non-high, then the *er*-suffixation involves (i) deleting the rime and (ii) suffixing a [ə] to the stem. In other words, these rimes are replaced with a [ə]. As we have witnessed in sections 2-3, there is no gliding motion of the tongue postures between the unsuffixed forms and *er*-suffixed forms, suggesting that non-high vowels in the stem are completely deleted due to the *er*-

suffixation. This observation is, to a great extent, not surprising, since our experimental results have already confirmed that the *er*-suffix is a rhotic schwa, a mid central vowel in its own right. So if the *er*-suffix is underlyingly a full-fledged vowel, the stem would become “oversized,” if [ə] could be directly suffixed to a stem, e.g. [pa:] ‘give’ → [pa:ə] ‘Ba.er (personal name)’ (i.e. a superheavy syllable, which is presumably illegal in SWM or any dialect of Mandarin Chinese).

On the other hand, at least impressionistically speaking, *er*-suffixed forms such as [p^hi.p^hjə] ‘skin’ or [ku.kwə] ‘swelling’ (see Figs. 5 and 6, among others) are “diphthongized”. More precisely, an excrescent schwa can be perceived in the contexts of the *er*-suffixed high front vocoids (cf. [15]’s analysis of the excrescent schwa in several languages, which is based on articulatory gestural conflicts of adjacent gestures). So, if the vocoids (including onglides) are high in a stem, the high vocoid will be preserved; non-high elements will be deleted, and, subsequently, [ə] will be suffixed. In these environments, we argue that a stem will not become oversized (or, a superheavy syllable), simply because glide formation is possible for high vowels, e.g. /p^hi/ ‘peel’ → p^hi.p^hi → [p^hi.p^hjə] ‘skin’. Given that onglides normally do not contribute to phonological weight (but see [16] for a potential counterexample), it follows that the gliding of high vowels serves as a strategy to accommodate a full-fledged vowel [ə] in the *er*-suffixation. In contrast, it is difficult for nonhigh vowels in the stem to undergo glide formation in order to avoid a superheavy syllable. Consequently, that explains why nonhigh vowels do not survive in the *er*-suffixation in SWM.

The phonemic rhotic schwa and the *er*-suffix are more or less identical when the stem vowel is non-high, whereas more significant TT-raising is found in the *er*-suffixed high vocoids than in the phonemic rhotic schwa. A similar conclusion can be made for its counterparts in Beijing and Northeastern Mandarin [5]. In other words, the present study provides additional empirical evidence that the phonemic rhotic schwa is not equivalent to the *er*-suffix across Mandarin dialects, at least in articulation. This is, again, not so surprising because the *er*-suffix, albeit primarily derived via *er* ‘son’, has undergone substantial grammaticalization in the history of Mandarin Chinese. It is not impossible to see that the processes of grammaticalization have a significant bearing on the phonetic realization of a morpheme.

5. Conclusion

This paper is an articulatory study of the *er*-suffixation in a relatively understudied dialect group of Mandarin Chinese, i.e. Southwestern Mandarin. With the help of EMA and ultrasound technologies, our experimental results reveal a substantially different picture of this morphophonological process from those found in more familiar varieties of Beijing and Northeastern Mandarin. It is safe to say that the cross-dialectal differences may be much greater than previously thought, which, in turn, calls for more comprehensive studies in the future. The present study is simply an attempt to show only the tiny tip of the iceberg. Articulatory and acoustical data from more speakers are definitely needed for a more comprehensive understanding of the *er*-suffixation in SWM.

6. Acknowledgements

This study has been approved by the National Tsing Hua University’s Research Ethics Committee (#10712HS104 and #10612HS089).

7. References

- [1] Y. R. Chao, *A Grammar of Spoken Chinese*, Beijing: The Commercial Press, 1979.
- [2] S. J. Li, *Hanyu 'Er' /ə-/ Yinshi Yanjiu [A historical analysis of Chinese /ə-/]*, Beijing: The Commercial Press, 1986.
- [3] Z. J. Wang, "Erhua de Tezheng Jiegou [The featural geometry of r-rhymes]," *Zhongguo Yuwen*, vol. 1, pp. 2-10, 1997
- [4] S. Duanmu, *The Phonology of Standard Chinese*, OUP Oxford, 2007.
- [5] 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.
- [6] E. X. Lee, *Chinese national among 'overseas Chinese' in Singapore: the sociolinguistic authentication of mainland Chinese identities*. Diss, 2007.
- [7] Y. Y. Zheng, "Beijing hua he Chengdu hua, Chongqing hua de bijiao [The comparison of er-suffixation in Beijing dialect, Chengdu dialect and Chongqing dialect]," *Journal of Chongqing Normal University*, vol. 2, pp. 67-71, 1987
- [8] S. L. Yang, "Chengduhua yu Putonghua erhuayun zhi bijiao [The comparison of er-suffixation in Chengdu dialect and Standard Mandarin]," *Journal of Chengdu Teachers College*, vol. 1, pp. 77-81, 2002
- [9] M. Tiede, "MVIEW: software for visualization and analysis of concurrently recorded movement data," *New Haven, CT: Haskins Laboratories*, 2005.
- [10] M. Tiede, and D. H. Whalen, "Getcontours: An interactive tongue surface extraction tool," *Proceedings of Ultrafest VII*, 2015.
- [11] P. Boersma, "Praat: doing phonetics by computer (Version 5.1.05)," <http://www.praat.org/>, 2009.
- [12] 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.
- [13] Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. (2019), Wood, S. Available: <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
- [14] 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.
- [15] B. Gick, and L. Wilson, "Excescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets," *Laboratory phonology*, vol. 8, pp. 635-659, 2006.
- [16] F. F. Hsieh, "Shengmu canzhong" de Yuuin ji Yinxi Lijü – Jianlun Hanyu Jeyin nengfou canzhong [A critical review of recent approaches to onset weight: with special reference to the "medial" (onglides) in Chinese]," *Studies in Prosodic Grammar*, vol. 2, no. 2, pp. 32-51, 2017.