

Research Article

Jing Huang, Feng-fan Hsieh*, Yueh-chin Chang and Mark Tiede

On the two rhotic schwas in Southwestern Mandarin: when homophony meets morphology in articulation

<https://doi.org/10.1515/phon-2022-0036>

Received October 11, 2022; accepted September 22, 2023; published online November 8, 2023

Abstract: This is an acoustic and articulatory study of the two rhotic schwas in Southwestern Mandarin (SWM), i.e., the *er*-suffix (a functional morpheme) and the rhotic schwa phoneme. Electromagnetic Articulography (EMA) and ultrasound results from 10 speakers show that the two rhotic schwas were both produced exclusively with the bunching of the tongue body. No retroflex versions of the two rhotic schwas were found, nor was retraction of the tongue root into the pharynx observed. On the other hand, the *er*-suffix and the rhotic schwa, though homophonous, significantly differ in certain types of acoustic and articulatory measurements. In particular, more pronounced lip protrusion is involved in the production of the rhotic schwa phoneme than in the *er*-suffix. It is equally remarkable that contrast preservation is not an issue because the two rhotic schwas are in complementary distribution. Taken together, the present results suggest that while morphologically-induced phonetic variation can be observed in articulation, gestural economy may act to constrain articulatory variability, resulting in the absence of retroflex tongue variants in the two rhotic schwas, the only two remaining r-colored sounds in SWM.

Keywords: rhotic; *er*-suffix; articulatory uniformity; Southwestern Mandarin

1 Introduction

This work is an acoustic and articulatory study of the two rhotic schwas in an understudied dialect group of Mandarin, Southwestern Mandarin (henceforth

***Corresponding author: Feng-fan Hsieh**, Institute of Linguistics, National Tsing Hua University, 101, Section 2, Kuang-fu Road, Hsinchu 300044, Taiwan, ROC, E-mail: ffhsieh@mx.nthu.edu.tw. <https://orcid.org/0000-0002-5185-1804>

Jing Huang and Yueh-chin Chang, Institute of Linguistics, National Tsing Hua University, Hsinchu, Taiwan, ROC, E-mail: xiaokuidaren94@163.com (J. Huang), ycchang@mx.nthu.edu.tw (Y.-c. Chang)

Mark Tiede, Department of Psychiatry, Yale University, New Haven, CT, USA, E-mail: mark.tiede@yale.edu

SWM), namely the *er*-suffix and the vowel phoneme /ə/. Impressionistically speaking, the two rhotic schwas are homophonous, distinguished only in that the *er*-suffix is a functional morpheme. Therefore, the two rhotic schwas in SWM present an interesting case study of the rhotic vowels from a typological perspective. To set the stage, general descriptions of *er*-suffixation, the rhotic schwa phoneme, and the acoustic and articulatory properties of the rhotic vowels/approximants are provided in the sections that follow.

1.1 *Er*-suffixation in Mandarin

Er-suffixation (a.k.a. the *r*-suffix, the rhotic suffix, or *érhuà* ‘*er*-ize/ization’) is perhaps one of the most well-known morpho-phonological processes in Mandarin Chinese. Diachronically speaking, the documented cases of *érhuà* have at least dated back to the Ming dynasty (1368–1644 C.E.; Li 1986). Along with other sources, this suffix was primarily derived by means of attaching *er* ‘child’ to a stem to form diminutives. For obvious reasons, previous studies have overwhelmingly focused on the *er*-suffixation in Beijing Mandarin, on which Standard Chinese (*Pǔtōnghuà* ‘common speech of the Chinese language’) is based. There is no doubt that the *er*-suffix is an “*r*-like sound” (Hartman 1944: 33), although Chao’s seminal work (1970 [1968]) on modern Chinese grammar describes the *er*-suffix as a subsyllabic suffix /-l/ in Beijing Mandarin. In previous acoustic studies, it has been confirmed that F3 is lowered in *er*-suffixed vowels (or, a relatively stable small F3–F2 distance; see Huang 2010, Lee and Zee 2014, Shi 2003, Xing 2021, among others), which is conventionally taken as an indication of rhoticity (first observed in Potter et al. (1947) book, *Visible Speech*, as cited in Delattre and Freeman (1968), but see Lindau (1985)). The on-going debate, nevertheless, is whether the *er*-suffix is “segment-bound,” forming a sequence of a non-rhotic vowel plus the rhotic schwa/approximant, e.g., [paə], or realized as rhotacization throughout the whole of the rime, e.g., [a̠]. In the Chinese-language literature, Li (1986) claims that the *er*-suffix is a retroflex apical vowel (ɰ/) and proposes that the *er*-suffix may be attached to a stem, forming a diphthong, i.e., {aɰ, əɰ}, or merged into a stem, resulting in a rhotacized rime, i.e., {a̠, ə̠, u̠}. Lin and Shen (1995), among others, hold that rhoticity is almost synchronous with the vowel across the board. However, a more prevalent view, as far as we know, is that the *er*-suffix is a (subsyllabic) rhotic schwa [ə̠] (i.e., the second part of a diphthong), as described by Duanmu (2007), Wang (1997), Lee and Zee (2014), and Lin (2007), among others.

Results from articulatory studies may shed light on the debate over the phonetic realizations of the *er*-suffixation. Lee’s (2005) Electromagnetic Articulography (EMA) results from three Beijing Mandarin speakers show that *er*-suffixation is realized as a subsyllabic /ə̠/, forming a sequence of a non-rhotic vowel plus /ə̠/, when the rime

ends with a non-back vowel. On the other hand, the entire rime is rhotacized when the (unsuffixed) rime ends with a back vowel, e.g., [ʊ]. Jiang et al. (2019) report similar EMA results for the *er*-suffixed forms in Northeastern Mandarin, a closely related dialect of Beijing Mandarin. Similarly, through a qualitative exploration of the dynamics of the *er*-suffixed /au/ in Beijing Mandarin, Xing (2021: 121) remarks that “rhoticity is present from the beginning of the vowel” in her ultrasound results: [aʊ]. In sum, previous studies basically all agree that the *er*-suffix may be a subsyllabic /ə/ (i.e., the second part of a diphthong), or may lead to a rhotacized rime, depending on the context, as far as Northern Mandarin (here, Beijing and Northeastern Mandarin) is concerned.

Regarding the well-established retroflex versus bunched tongue shapes of the /ɹ/ sound in American English (Delattre and Freeman (1968), et seq.), Lee and Zee’s (2014: 386) EMA results indicate that the *er*-suffix in Beijing Mandarin “does not result in retroflexing but rhotacizing the vowels,” because “the tongue tip or tongue front is not curled up and backward, and the underside of the tongue does not touch the anterior part of the hard palate.” Jiang et al. (2019) also report that *er*-suffixation consistently involves a bunched tongue configuration in Northeastern Mandarin. On the other hand, results of ultrasound studies instead indicate that the *er*-suffix may be produced with either a retroflex or a bunched tongue shape by Mandarin speakers from Beijing (Xing 2021) or from Beijing, Hebei, and Shandong (Chen and Mok 2021). Details aside, Xing’s (2021) finding is that the retroflex variant is the dominant type (14 out of 18 participants) among Beijing Mandarin speakers, while there are more bunched variants (8 out of 12 speakers) identified in Chen and Mok (2021).

Regarding the other components in the articulation of rhotics, first, Lee and Zee (2014: 386) remark that “the tongue body is retracted towards the pharynx” during *er*-suffixation in Beijing Mandarin. Xing (2021) also makes a similar observation based on her ultrasound results. Second, it is still not clear if the production of the *er*-suffix involves lip rounding, a known characteristic of the English rhotic schwa (Delattre and Freeman (1968), et seq.).

Finally, little attention has been paid to the rhotic schwa phoneme /ə/, the stand-alone rhotic schwa in Mandarin. Jiang et al. (2019) report that the rhotic schwa is produced with tongue tip (TT) raising and involves substantial movement of tongue when gliding from initial to final vowel quality (or, diphthongization) in Northeastern Mandarin, whereas Chen and Mok (2021) find more instances of bunched tongue configurations (8 out of 12 speakers from Beijing, Hebei and Shandong) in their ultrasound results of the rhotic schwa (their syllabic /ɹ/).

The brief description above suggests that the *er*-suffix (and the rhotic schwa) in Beijing and Northeastern Mandarin may well be subject to distinct articulatory realizations, in the same way as the consonantal and syllabic /ɹ/’s in American

English (see Mielke et al. 2016 for a recent overview). In addition, tongue root retraction may also be found in the production of *er*-suffixation. Therefore, the entry point of the present study is to contribute more empirical data to a growing body of work on the (un)expected diversity of closely related languages/dialects such as the different varieties of Mandarin, by investigating the acoustic and articulatory properties of the two rhotic schwas in SWM.

1.2 The two rhotic schwas in Southwestern Mandarin

Southwestern Mandarin (SWM), with over 250 million native speakers, is the most spoken variety of Mandarin Chinese (Li 1997). SWM belongs to one of the eight groups of Mandarin Chinese and is mainly spoken in Southwest China, including Sichuan, Chongqing, Yunnan, Guizhou, most areas of Hubei, and some areas of Hunan, Shaanxi, Guangxi and Jiangxi (Li 2009; see also the colored areas in Figure 1). In the present study, our data were collected from young speakers from different sub-dialects in the representative group of SWM: the *Chéngdū-Chóngqīng* group (often abbreviated as the *Chéng-yú* dialect group in the Chinese-language literature), spoken in western Hubei, Chongqing, and eastern Sichuan (Wurm et al. 1987), indicated in yellow in Figure 1. It is widely accepted that the (sub)dialects in SWM are highly stable and homogenous in terms of phonetic and phonological patternings, since they descend from the Mandarin dialect spoken by a continuous influx of immigrants from the same neighboring provinces of Hubei, Hunan, and Jiangxi during the Ming and Qing dynasties (Li 1997).

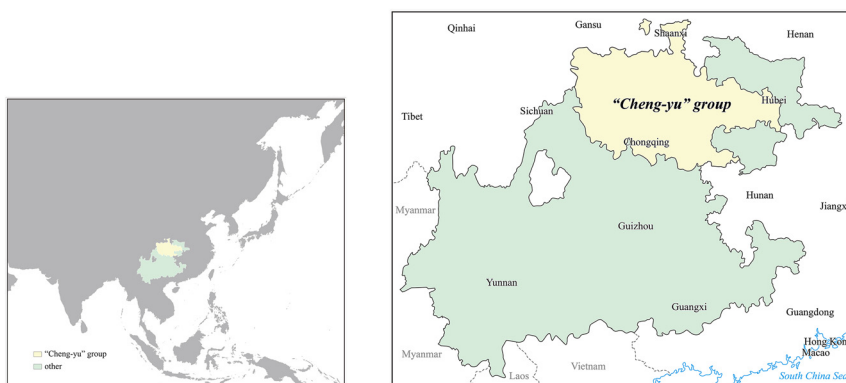


Figure 1: Geographical distribution of Southwestern Mandarin.

The *er*-suffix has been semantically bleached¹ in SWM; more importantly, unlike its counterpart in Northern Mandarin (Beijing and Northeastern alike), the *er*-suffix in SWM features the following unique characteristics: first, there are only four output forms of the *er*-suffix in SWM: {ə, jə, wə, ʋə} (Yang 2002, Zheng 1987; recall the debate over the subsyllabic /ə/ vs. rhotacized vowel in Section 1.1). Second, with some rare exceptions, the stem must be a polysyllabic word, which is typically a reduplicated disyllabic word. Third, “rime usurpation” is obligatory in *er*-suffixation in SWM (cf. Zimmermann’s (2013) analysis of mora usurpation in Yine), meaning that the *entire rime* must be completely deleted to accommodate the *er*-suffix, except for the high and rounded vocoids (more precisely, the high/rounded vowels as well as the prenuclear glides), which are preserved under glide formation. Representative examples of the four variants are provided in Table 1, where tones are omitted and √ means a lexical root.

On the other hand, SWM has a rhotic schwa phoneme: /ə/ (see also fn. 5). This phoneme cannot be combined with an onset or a coda to form a syllable, so its distribution is highly restricted in the lexicon; only a few real words/morphemes exist, e.g., ə² ‘two’, ə³ ‘bait’, ə³ √ear: lexical root for “ear” (bound morpheme)’, etc. In other words, the rhotic schwa phoneme /ə/ may be regarded as a marginal phoneme in SWM (see, e.g., Hall 2013). It is equally remarkable that there are only two rhotic/r-colored sounds in SWM, namely the /ə/ phoneme and the *er*-suffix. In contrast, Beijing Mandarin has a rhotic onset phoneme, which is represented as ‘r’ in Pinyin romanization and is transcribed as an apical post-alveolar approximant /ɻ/ in Lee and Zee (2003). This syllable-initial /ɻ/ sound is produced with a bunched tongue posture in all 12 speakers from Northern China, according to Chen and Mok (2021) and in 10 out of 18 speakers from Beijing (Xing 2021). This rhotic/r-colored phoneme

Table 1: The four variants of the *er*-suffix in Southwestern Mandarin.

[ə]	/pa/ → [pa-pə] ‘√give: handle’ (cf. [paə] ‘handle’ in B(eijing) M(andarin))
	/tau/ → [tau-tə] ‘√knife: knife’ (cf. [əau-tau] ‘small knife’ in BM)
	/keta/ → [ke.tə] ‘√lump: lump’
[jə]	/p ^h i/ → [p ^h i-p ^h jə] ‘√peel: skin’ (cf. [p ^h iə] ‘skin’ in BM)
	/pje/ → [pje-pjə] ‘√deflated: dent’
[wə]	/tu/ → [tu-twə] ‘√protruding: cheek’ (cf. [t ^h u] ‘picture’ in BM)
	/toŋ/ → [toŋ-twə] ‘√body: bare to the waist’
[ʋə]	/tə ^h y/ → [tə ^h y.tə ^h ʋə] ‘√kind of worm: cricket’ (cf. [təyə] ‘pony’ in BM)
	/tə ^h ʋo/ → [tə ^h ʋo.tə ^h ʋə] ‘√bird: bird’

1 Certain SWM dialects have developed a genuine diminutive suffix, [wə] (<wa-er ‘baby-*er* suffix’), e.g., [tau-wə] ‘knife-diminutive: a small knife’ (compare: [tau-tə] ‘knife’ in Table 1).

corresponds to a voiced alveolar fricative /z/ in SWM, presumably as the result of onset fortition, however. Furthermore, the famous three-way contrast (alveolar vs. retroflex vs. alveopalatal; see, e.g., Duanmu (2007), Lee and Zee (2003, 2014), Lin (2007) and references cited therein) in Mandarin sibilants has been lost in most varieties of SWM (in particular, the *Chéngyú* group; see Figure 1), resulting in a two-way contrast of sibilants: alveolar versus alveopalatal. Therefore, the “retroflex” apical vowel in Beijing Mandarin, always co-occurring with the “retroflex” sibilants is lost in SWM as well. Note that the “retroflex” apical vowel is transcribed as an apical post-alveolar approximant, based on Lee and Zee’s (2014) EMA results (but see Lee-Kim 2014). In sum, the variants of the *er*-suffix as well as the contrasts in the sound inventory, have been, to a significant extent, simplified in SWM, in comparison to Northern Mandarin. The significance of these cross-dialectal differences will be addressed in Section 4.1.

This section is closed with a description of the sound system of Chengdu Chinese, the representative variety of SWM (He and Rao 2014). Like Duanmu’s (2007) analysis of Standard Chinese, Chengdu Chinese also has a maximal syllable template, CGVX (where C = Consonant, G = Glide, V = Vowel, and X = Nasal coda or Glide), with the following consonant phonemes: {p, p^h, t, t^h, k, k^h, ts, ts^h, tɕ, tɕ^h, m, n/l, ɲ, ɲ, f, v, s, z, ɕ, x}, vowel phonemes: {i, u, y, a, o, e, ə} and four lexical tones in Chao’s tone notation: {T1: 45, T2: 21, T3: 42 and T4: 213}. See also Section 4.3 for description on word-level prosody in SWM.

1.3 Why the rhotic vowels in Southwestern Mandarin are “special”

Rhotic vowels are typologically rare (Maddieson 1984); nevertheless, SWM presents an interesting case of the cross-linguistic rarity of the rhotic vowels from a completely novel angle. Precisely, the phonemic /ə/ and the *er*-suffix are not, impressionistically speaking, distinguishable at all, the only difference being that the *er*-suffix is a functional morpheme. Importantly, the fact that the *er*-suffix is not a phoneme *per se* has not yet received due attention in the literature and one we believe carries significant consequences. Specifically, regarding the relationship between morphemic status and phonetic implementation of homophonous affixes and their “non-morphemic” counterparts, Plag et al. (2017) and subsequent works examine distinct acoustic realizations of the non-morphemic /s/ and /z/ versus the /s/ and /z/ morphemes (e.g., plural, genitive, etc.) in a corpus study, and suggest that morphological structures may have a bearing on surface phonetic realization. In view of this, we raise the possibility that the *er*-suffix and the /ə/ phoneme may differ in their phonetic realization as well. Support for this view follows from the

EMA results reported in Jiang et al. (2019), according to which only the rhotic schwa phoneme, not the *er*-suffix, is produced with tongue tip raising in Northeastern Mandarin. The present study is thus an attempt to distinguish between the phonetic characteristics of the *er*-suffix and the /ə/ phoneme in SWM from data collected from multiple speakers using EMA and ultrasound imaging methods. The novelty of the present study is that the contentive (the /ə/ phoneme) versus functional (the *er*-suffix) divide is systematically investigated by comparing the acoustic and articulatory measurements of the rhotic schwas in an understudied dialect group of Mandarin, SWM.

Four research questions to be addressed are listed below.

- i. Are the *er*-suffix and the rhotic schwa phoneme produced with both retroflexion and bunching variants?
- ii. “Within-group” comparisons: are the *er*-suffixes attached to different stems produced identically in acoustics and articulation?
- iii. Are the *er*-suffix and the rhotic schwa phoneme produced identically in acoustics and articulation?
- iv. Are the *er*-suffix and the rhotic schwa phoneme produced with lip rounding and/or tongue root retractions?

The paper is organized as follows. Section 2 is a description of experimental methods and data analysis. The results of our articulatory and acoustic data are presented in Section 3. Section 4 discusses the findings of the study. Finally, Section 5 concludes this paper.

2 Experimental methods

2.1 Participants

Ten native speakers (9 female) of Southwestern Mandarin participated in this study. They were undergraduate or graduate students in their twenties at the time of the experiments (average = 23.3 y.o., SD = 2.95) and were born and raised in the *Chéngyú* dialect group-speaking areas (see Figure 1; specifically, 5 from Yichang, Hubei, 3 from Enshi, Hubei, 1 from Chengdu, Sichuan, and 1 from Guang'an, Sichuan). It was confirmed via background screenings that they acquired Standard Chinese only as part of their school education. The participants had no self-reported speech or hearing problems. They all gave written informed consent and received compensation for their participation. Due to a data recording issue, we report the results of EMA data from seven participants. The ultrasound image data are based on the results of all ten participants.

2.2 Materials

The recording materials are comprised of 69 meaningful words, including (i) 33 unsuffixed monosyllabic words, (ii) 32 *er*-suffixed disyllabic forms, and (iii) 4 disyllabic words containing the /ə/ phoneme in word-final position. The syllable structures of the stimuli include CV, CGV, CVG and C(G)VN, where C = {p, t, k, tɕ^h}, G = {j, w, ɥ}, V = {i, y, e, a, o, u, ə} and N = {n, ŋ}. Tones are not controlled for primarily because tone values may not be identical across all of the subdialects under investigation. However, actual pitch values for each tone category are quite similar. Below are some representative examples (Table 2). See Appendix A for the complete wordlist.

Table 2: Pairs to be compared: some representative examples.

a.	<i>Er</i> -suffix versus the /ə/ phoneme: E.g., [p ^h u.p ^h ə] ‘a store’ versus [p ^h u.ə] ‘Pu-erh (tea)’
b.	<i>Er</i> -suffixes attached to different stems: E.g., [pa.pə] ‘a handle’ versus [pan.pə] ‘a crowd of’
c.	Unsuffixed stem versus the /ə/ phoneme: E.g., [pə] ‘thin’ versus [po.ə] ‘Second-year Ph.D. student’

2.3 Recording procedures

Prior to recording participants were asked to read a newspaper paragraph in SWM. The participants were then asked to read a randomized list of the target words from a computer screen in a sound-proof room in the phonetics lab, National Tsing Hua University. The stimuli were displayed using the Articulate Assistant Advanced (AAA, Articulate Instruments) software and each slide was shown for 4 s. The participants were asked to embed the target words in the carrier phrase “__, *pa* __ *pa*”, meaning “__, give __ Sentence Final Particle: (Speaking of) __, just give __ (to me)!” in SWM. Six repetitions were collected for each token and in order to control for outside factors, only the more naturally rendered second occurrence of a stimulus in the carrier phrase was analyzed and reported. A total of 2,989 EMA tokens (= 69 words × 6 repetitions × 7 participants) were analyzed and reported, and 4,140 tokens (= 69 words × 6 repetitions × 10 participants) were analyzed and reported for the ultrasound image results.

2.4 Apparatuses

The articulatory data were recorded concurrently using EMA (WAVE; Northern Digital Inc.) at a sampling rate of 200 Hz, and ultrasound (Micro system; Articulate

Instruments Ltd.) at 65 fps. Acoustic data were simultaneously recorded using a Sennheiser unidirectional shotgun microphone at 24 kHz. Regarding the EMA experiment, seven sensors were attached to the tongue, lips, upper incisors and lower incisors (jaw) using the instant dental adhesive α QUIN (BSA), together with the dental cement GC Fuji I. Specifically, three sensors were affixed midsagittally to the tongue: one on the tongue tip, about 0.5 cm back from the anatomical tip, one on the dorsum of the tongue, as far back as comfortable, and one midway between the tongue tip and tongue back sensor. One sensor was affixed to the lower incisors to track jaw movements and two additional sensors were placed on the vermillion border of the upper and lower lips. Three reference sensors were also placed on the left and right mastoid processes and upper incisor to correct for head movement. The occlusal plane was identified from a bite plane using a fixed triangular protractor with three sensors glued to it. A palate trace was collected using a spare sensor attached to a stir stick; participants were instructed to trace the stick from the back of the hard palate to their front teeth (Rebernik et al. 2021). The articulatory dataset produced by the EMA recordings was post-processed and analyzed using custom MATLAB scripts.

Ultrasound data were collected using a transducer with a 92° field of view, set at a depth of 120 mm. The frame rate was set to 65 fps. The participants wore an all-plastic UltraFit headset (Articulate Instruments Ltd.; see Spreafico et al. (2018) for more detail) to stabilize the probe under the chin during imaging of the midsagittal tongue profile (Wrench and Scobbie 2016).

Acoustic recordings were synchronized with the EMA and ultrasound image data by means of the WaveFront software (NDI) and the synchronization unit of the Micro system (Articulate Instruments), respectively.

2.5 Statistical analysis

For quantitative results, the articulatory and acoustic data are analyzed using generalized additive mixed modeling (GAMM) analysis (Wood 2017 [2006]). Our analysis is primarily based on the procedures and suggestions provided in Wieling (2018) as well as in Sóskuthy (2021) since the trajectories of the EMA sensors (as well as the tongue contours in ultrasound imaging and the formants) are nonlinear in nature.

2.5.1 EMA data

Regarding EMA experiments, the head-corrected data were z-transformed for subsequent GAMM analysis. We used the R package *mgcv* (Wood 2019) for model

fitting and models were constructed with the *bam()* function. For each model, “Sound” (e.g., the *er*-suffix vs. the /ə/ phoneme) was included as the main effect, and the measurement of interest was specified as the dependent variable (i.e., z-transformed positions for each EMA sensor). The models included a by-word smooth function through time to investigate articulatory changes over time, and a random smooth to account for variation between all seven SWM speakers. See also Figure 5 for a visual summary of the GAMM models fitted for EMA sensor trajectories in Section 3.2.1.

2.5.2 Ultrasound data

The ultrasound data were analyzed with the help of Articulate Assistant Advanced (AAA) software. We extracted the tongue contours at the first quartiles (25 %), midpoints (50 %), and the third quartiles (75 %) of an acoustically defined rime using the default 42 point positions exported by AAA for each tongue contour. Following Mielke’s (2015) suggestion, the extracted tongue contours were transposed into polar coordinates using AAA software. Again, we tested these predictions using Generalized Additive Mixed modeling (GAMMs; Sóskuthy 2017; Wieling 2018; Wood 2017 [2006]), with the help of the R script in Heyne et al. (2019), adapted to our data by us. We ran various models to evaluate the best fit one (e.g., no random effects, random effects, multiple predictors including Type (i.e., *er*-suffixed vs. unsuffixed, different vowels, etc.)). The model we adopted is summarized below. We modeled one variable DIST (the distance of the fitted tongue contour point from the origin), based on the following predictor variables. The tongue contours at the first quartiles, midpoints, and the third quartiles of a rime are compared using the GAMM analysis. See also Figure 6 for a visual summary of the GAMM models fitted for ultrasound splines in Section 3.2.2.

- main effect of Sound (e.g., unsuffixed vs. *er*-suffixed; *er*-suffix attached to stem /a/ vs. *er*-suffix attached to stem /an/; *er*-suffix vs. the rhotic schwa phoneme /ə/, etc.)
- smooth term for theta (the angle in relation to the origin)
- smooth term for theta by the interaction of Type and Vowel
- random by-subject smooths for theta by Vowel

2.5.3 Acoustic data

The acoustic data were analyzed using Praat (Boersma and Weenink 2007, version 6.0.30). Formant values for F1, F2, and F3 in the sonorous rimes were extracted using Praat scripts developed in the Phonetics Lab at National Tsing Hua University. The formant values subsequently were normalized using Labov’s method, as in the *Atlas of North American English* (ANAE). Labov’s ANAE method uses logarithmic means to normalize the formant values. Unlike Nearey’s methods, ANAE is speaker-extrinsic

in that it computes a single grand mean for all speakers included in this study, thereby preserving sociolinguistic variation (see Thomas and Kendall 2007 for more detail and references cited). Comparisons of the formant values were conducted using Generalized Additive Mixed modeling (Wood 2017 [2006]) as well. See Section 2.5.1 for the analytical procedures.

3 Results

3.1 Bunched configurations and the Tongue Retroflexion Angle (RA)

The first research question (i) is whether the *er*-suffix is produced with both retroflexion and bunching variants. There are no cases where an obvious Tongue Tip (TT) gesture is identified through visual inspection of the articulatory data.² However, we did find two distinct subtypes of the *er*-suffix. Consider now Figures 2 and 3, where the two distinct subtypes are illustrated. For ease of visual comparison, the temporal changes of the tongue configurations of the *er*-suffix are represented as solid lines, which refer to the different (acoustically determined) deciles of a sonorous rime, whereby the blue line refers to the onset of an *er*-suffixed rime (t1, the first decile of the rime), the brown line the offset (t10, the last decile of the rime), and so on. The positions of each EMA sensor are averaged over six repetitions for each target word and connected using a cubic spline.

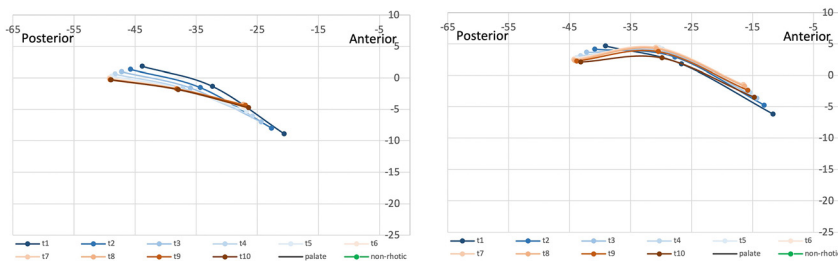


Figure 2: Two subtypes of the *er*-suffix found in [pei.pə] ‘a cup’ (Left: Type A [Subject F01/Dorsum-Up]; Right: Type B [Subject F04/Dorsum-Down]). The speakers are facing right.

² See Appendix B.1/2 for comparisons of the midpoints of the unsuffixed forms and temporal changes of lingual configurations of *er*-suffixed forms (32 pairs in total), based on the EMA data from Subject F02.

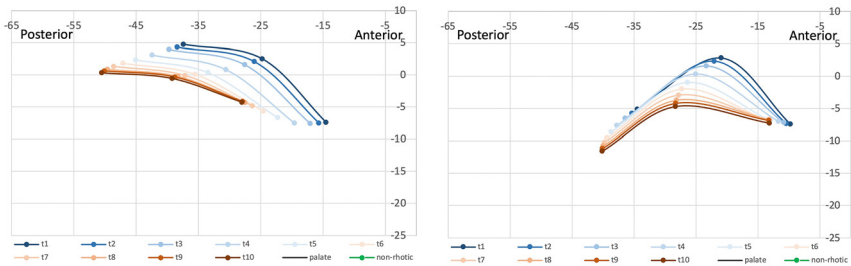


Figure 3: Two subtypes of the *er*-suffix in found [pje.pjə-] ‘a dent’ (Left: Type A [Subject F01/ Dorsum-Up]; Right: Type B [Subject F04/Dorsum-Down]). The speakers are facing right.

These tongue configurations in Figures 2 and 3 may be classified as (a) dorsum-up bunched *er* and (b) dorsum-down *er* (cf. the tip-up vs. tip-down bunched *r*’s in Espy-Wilson et al. (2000)). Dorsum-up bunched *er*’s involve (mild) tongue retraction, while dorsum-down bunched *er*’s feature a considerably more convex tongue body followed by (some) tongue retraction, especially in the presence of a prenuclear glide (Figure 3). According to our data, speakers F01, F03, F05 and F07 belong to Type A and F02, F04 and F06 Type B. The present discrepancy cannot be ascribed to sub-dialectal differences since, for example, subject F01 is from Yichang, Hubei, whereas subject F03 is from Chengdu, Sichuan, which is approximately 860 km apart as the crow flies. On the other hand, speakers F01, F04, F05 and F06 are all from Yichang, Hubei, but only speakers F04 and F06 may be classified as Type B.

As a further step, the EMA data for bunching are quantitatively analyzed by means of the Tongue Retroflexion Angle (RA), proposed in Tiede et al. (2019). Precisely, the RA is subtended by the extension of lines between TD:TB and TB:TT, as illustrated in Figure 4. A bunched tongue posture is defined (in red), if the RA is positive (measured CW) and a retroflex tongue configuration is defined as a negative RA (measured CCW; in blue).

The RA (Tongue Retroflexion Angle) values of the rhotic schwa phoneme and the *er*-suffixes attached to the six monophthongal stems {i, y, e, a, o, u} were calculated. The RA values are obtained at the offset of an *er*-suffix to minimize the potential impact from the gliding motions by the high vocoids (see Figure 3). As we shall see in Tables 3 and 4 below, the RA values are positive across the board in the current data.

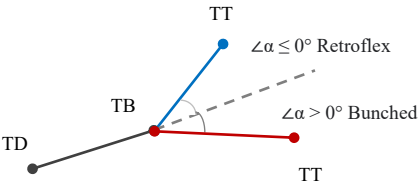


Figure 4: A schematic illustration of the measurement of Tongue Retroflexion Angle (RA).

Table 3: Tallies of the three categories of *er*-suffix (retroflex : slightly bunched : typically bunched; light gray cells; more tokens of slightly bunched *er*-suffix, dark gray cells; more tokens of typically bunched *er*-suffix).

	F01	F02	F03	F04	F05	F06	F07
[pa.pə]	0:6:0	0:2:4	0:0:6	0:5:1	0:3:3	0:0:6	0:6:0
[po.pə]	0:6:0	0:0:6	0:0:6	0:2:4	0:0:6	0:0:6	0:6:0
[tu.twə]	0:6:0	0:0:6	0:0:6	0:4:2	0:2:4	0:0:6	0:6:0
[ke.kə]	0:6:0	0:1:5	0:0:6	0:3:3	0:0:6	0:0:6	0:6:0
[pʰi.pʰjə]	0:4:2	0:0:6	0:0:6	0:3:3	0:1:5	0:0:6	0:0:6
[tɕʰy.tɕʰɥə]	0:0:6	0:0:6	0:0:6	0:4:2	0:0:6	0:0:6	0:3:3

Table 4: Tallies of the three categories of the rhotic schwa phoneme (retroflex : slightly bunched : typically bunched : more tokens of slightly bunched [ə̤], dark gray cells: more tokens of typically bunched vowel phoneme [ə̤]).

	F01	F02	F03	F04	F05	F06	F07
[pa.ə̤]	0:6:0	0:1:5	0:0:6	0:2:4	0:2:4	0:0:6	0:6:0
[po.ə̤]	0:6:0	0:3:3	0:0:6	0:3:3	0:3:3	0:0:6	0:6:0
[pʰu.ə̤]	0:6:0	0:0:6	0:0:6	0:3:3	0:0:6	0:0:6	0:5:1
[ke.ə̤]	0:6:0	0:1:5	0:0:6	0:0:6	0:1:5	0:0:6	0:4:2

For ease of discussion, we arbitrarily define the two subtypes in Figures 2 and 3 as (a) Type A: a “slightly bunched” tongue posture (whose RA is positive and is smaller than or equal to 15°) and (b) Type B: a “typically bunched” tongue posture (whose RA is greater than 15°). Consider now Tables 3 and 4, where darkly shaded cells refer to more tokens of Type B (typically bunched) and more lightly shaded tokens of Type A are lightly shaded. The tallies of the three categories of the RA values are represented as, for example, (0:6:0), meaning (0 tokens for Retroflex [$\leq 0^\circ$]: 6 tokens for Slightly Bunched [$\leq 15^\circ$]: Typically Bunched [0 tokens $> 15^\circ$]).

From the measurements of the Tongue Retroflexion Angle (RA), our finding is that there is no single instance of a typical retroflex *er*-suffix and a retroflex schwa (i.e., $RA \leq 0^\circ$) across all the participants. In sum, we can say that only bunched tongue postures were observed in this study, as far as the two rhotic schwas are concerned.

3.2 The *er*-suffix: “within-group” comparison

We now test whether these *er*-suffixes differ in tongue movements/postures, namely whether (in)complete neutralization takes place in the production of these *er*-suffixes (i.e., research question (ii)). The results are presented in this order: EMA, ultrasound, and acoustic data.

3.2.1 “Within-group” comparison: EMA results

Regarding the EMA results, the pair-wise comparisons are based on the four variants of the *er*-suffix illustrated in Table 2: {Cə versus Cə}, {Cjə versus Cjə}, {Cwə versus Cwə} and {Cɥə versus Cɥə}, where onset C’s are identical in place of articulation in each of the 27 pairs, e.g., {[pa.pə] ‘a handle’ versus [pai.pə] ‘a crowd of’}, {[pa.pə] ‘a handle’ versus [po.pə] ‘a bowl’}, {[p^{hi}.p^hjə] ‘skin’ versus [pjɛn.pjə] ‘a dent’}, etc. (see Table 5 for a complete list). The trajectories of the sensors for the Tongue Tip (TT), the Tongue Body (TB) and the Tongue Dorsum (TD) are compared along the horizontal (x) and vertical (z) dimensions, by means of the Generalized Additive Mixed Model analysis (GAMM, See Section 2.5.1). We used the R package *itsadug* (van Rij et al. 2017) for visualizing the resulting patterns. Consider now Figure 5,³ where the trajectories of TDx (Tongue Dorsum-longitudinal) and TBz (Tongue Body-vertical) of the *er*-suffixes in [tu.twə] ‘cheek’ and [tɔŋ.twə] ‘bare to the waist’ are compared.

³ R syntax for the model of Figure 5: `m7 <- bam(Pos ~ Sound + s(Time, by = Sound) + s(Time, Speaker, bs = "fs", m = 1) + s(Time, Speaker, by = SoundO, bs = "fs", m = 1), data = dat, rho = m6acf[2], AR.start = dat$start.event)`. See also Section 2.5.1 for more detail.

Table 5: Summary of GAMM results in the lingual articulators: the *er*-suffixes (TT = Tongue Tip, TB = Tongue Body, TD = Tongue Dorsum, x = front-back, z = up-down; $\sqrt{\sqrt{}}$ = significant difference greater than 80 % of the rime; $\sqrt{}$ = 80 %–50 % of the rime; blank = no difference or less than 50 % of the entire rime; see the lower panel of Figure 5).

	TTx	TTz	TBx	TBz	TDx	TDz
[pa.pə] versus [pai.pə]				$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$
[pa.pə] versus [pan.pə]	$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$
[pa.pə] versus p ^h e.pə]			$\sqrt{\sqrt{}}$	$\sqrt{}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$
[pa.pə] versus [p ^h u.p ^h ə]			$\sqrt{}$		$\sqrt{}$	
[pa.pə] versus [po.pə]				$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$
[po.pə] versus [poŋ.pə]	$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$	
[te.tə] versus [tow.tə]		$\sqrt{}$		$\sqrt{}$		$\sqrt{\sqrt{}}$
[tai.tə] versus [tan.tə]			$\sqrt{\sqrt{}}$			
[ke.tə] versus [tai.tə]			$\sqrt{\sqrt{}}$		$\sqrt{\sqrt{}}$	
[ke.tə] versus [tan.tə]			$\sqrt{\sqrt{}}$			
[pje.pjə] versus [pjen.pjə]		$\sqrt{\sqrt{}}$				
[ti.tjə] versus [tje.tjə]		$\sqrt{}$				
[tje.tjə] versus [tjen.tjə]			$\sqrt{}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	
[tu.twə] versus [toŋ.twə]	$\sqrt{}$				$\sqrt{}$	
<i>Pairs that show no differences:</i> [po.pə] versus [p ^h e.p ^h ə]/[po.pə] versus [p ^h u.pə]/[p ^h e.pə] versus [p ^h u.pə]/[ka.kə] versus [ke.kə]/[pai.pə] versus [pan.pə]/[pei.pə] versus [pen.pə]/[p ^h i.p ^h ə] versus [pje.pjə]/[p ^h i.p ^h ə] versus [pjen.pjə]/[te.tə] versus [ten.tə]/[kwa.kwə] versus [kwan.kwə]/[ku.kwə] versus [kwa.kwə]/[ku.kwə] versus [kwan.kwə]/[tɕ ^h y.tɕ ^h ɥə] versus [tɕ ^h ɥo.tɕ ^h ɥə]						

A summary of GAMM results in the lingual articulators is given in Table 5. Note that two check signs ($\sqrt{\sqrt{}}$) mean the two *er*-suffixes significantly differ along a certain dimension (Horizontal or Vertical) of a given EMA sensor (e.g., Tongue Tip, TT) throughout at least 80 % of the entire rime (see the lower panel of Figure 5); while a check sign ($\sqrt{}$) means the two trajectories are significantly different throughout at least 50 % of the entire rime. No difference or difference less than 50 % of the entire rime is left blank.

As seen in Table 5, the *er*-suffixes are not articulatorily indistinguishable in a pair-wise comparison (i.e., 15 out of 27 pairs show significant differences at least 50 % of the rime). No significantly different trajectory of any EMA sensor can be found across all the pair-wise comparisons, however. In other words, there is no consistent “within-group” difference among the *er*-suffixes attached to different stems, as far as the EMA data are concerned.

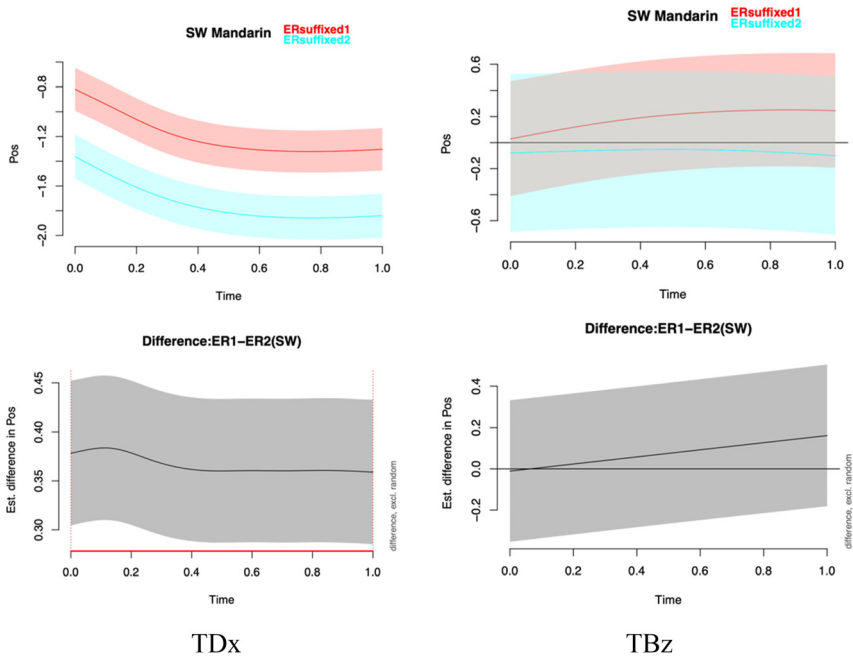


Figure 5: GAMM models of EMA sensor trajectories of the *er*-suffixes in [tu.twə] ‘cheek’ (in red: ERsuffixed1) and [toŋ.twə] ‘bare to the waist’ (in blue: ERsuffixed2), where TB = Tongue Body, TD = Tongue Dorsum, *x* = longitudinal dimension/front-back, *z* = vertical dimension/up-down. Upper panel: shaded bands represent the point-wise 95 %-confidence interval. Lower panel: when the shaded point-wise 95 %-confidence interval does not overlap with the *x*-axis (i.e., the value is significantly different from zero), this is indicated by a red line on the *x*-axis (and vertical dotted lines). Results are based on the data from 7 speakers.

3.2.2 “Within-group” comparisons: ultrasound results

The ultrasound data were concurrently collected along with the NDI Wave. For “co-referencing” purposes, the ultrasound data are used to observe holistic midsagittal tongue shapes. To begin, below is a sample illustration of how the ultrasound data are displayed in a polarscatter plot. In Figure 6, the red solid line refers to the Type 1 tongue shape (here, the unsuffixed stems), while the blue dotted line indicates the Type 2 tongue shape (here, the *er*-suffixed stems). Both were extracted from the midpoints of an acoustically defined rime. The thinner dotted lines of each color indicate the region of 95 % confidence, and an area where the background is shaded gray is where there is a statistically significant difference between the positions (or, region of significance, which was produced by the *itsadug* function

GAM smooths o.R vs. o.NR 50%

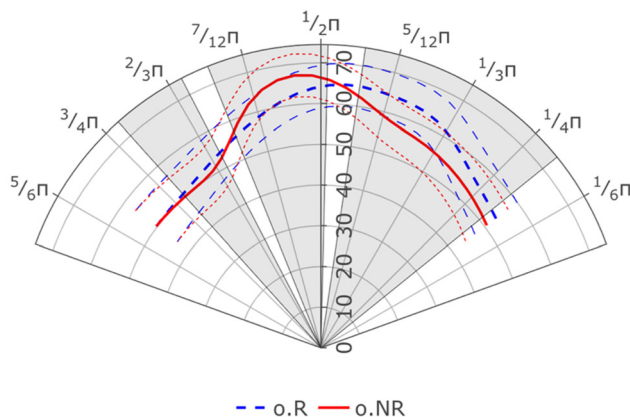


Figure 6: The fitted smoothing splines at the midpoints (50 %) for all tokens of the unsuffixed stems /o/ (o.NR) and its *er*-suffixed forms (o.R). Gray area indicates positions with a statistically significant difference. The speakers ($n = 10$) are facing right.

plot_diff. See Section 2.5.2 for a description of the model we adopted⁴). Note also that the images are shown in polar coordinates, with the right-hand side being the front of the mouth.

With this in mind, let's now move on to see if the holistic tongue configurations of the *er*-suffixes differ in any manner. All possible pairs are enumerated as follows: [Ca.Cə] versus [Cai.Cə]; [Ca.Cə] versus [Can.Cə]; [Ca.Cə] versus [Ce.Cə]; [Ca.Cə] versus [Co.Cə]; [Cai.Cə] versus [Can.Cə]; [Ci.Cjə] versus [Cje.Cjə]; [Cu.Cwə] versus [Con.Cwə]; [Cu.Cwə] versus [Cwa.Cwə]; [Cu.Cwə] versus [Cwei.Cwə]; [Cwa.Cwə] versus [Con.Cwə]; [Cwei.Cwə] versus [Con.Cwə] and [Cwa.Cwə] versus [Cwei.Cwə], where C = {p, p^h, t, k, tɛ^h}, if available. In Figures 7 and 8, the comparisons of the four representative pairs at the first quartiles (25 %), the midpoint (50 %), and the third quartiles (75 %) of the rime are illustrated. respectively. Figure 7 refers to {[Ca.Cə]

⁴ R syntax for the model of Figure 6: `bam (DIST ~ Type.Vowel + s(theta, bs = "cr", k = 10) + s(theta, bs = "cr", k = 10, by = Type.Vowel) + s(theta, subject, bs = "fs", k = 10, m = 1, by = Vowel), data = df_gam, AR.start = df_gam$start, rho = rho, discrete = TRUE, nthreads = ncores)`, where DIST is the distance of the fitted tongue contour point from the origin (of the polar coordinate), and theta is the angle in relation to the origin. The variable *Type.Vowel* encodes the interaction of *Type* (i.e., *unsuffixed* vs. *er-suffixed*) and *Vowel* (i.e., *a*, *i*, etc.). It is used as a contour adjustment for the random effect that uses *subject* ID, used to model the within-speaker variations; *AR.start* applied to tell the model the 42 points of each frame are making a tongue spline, while *k* is number of knots to control for the degree of non-linearity in the smooth.

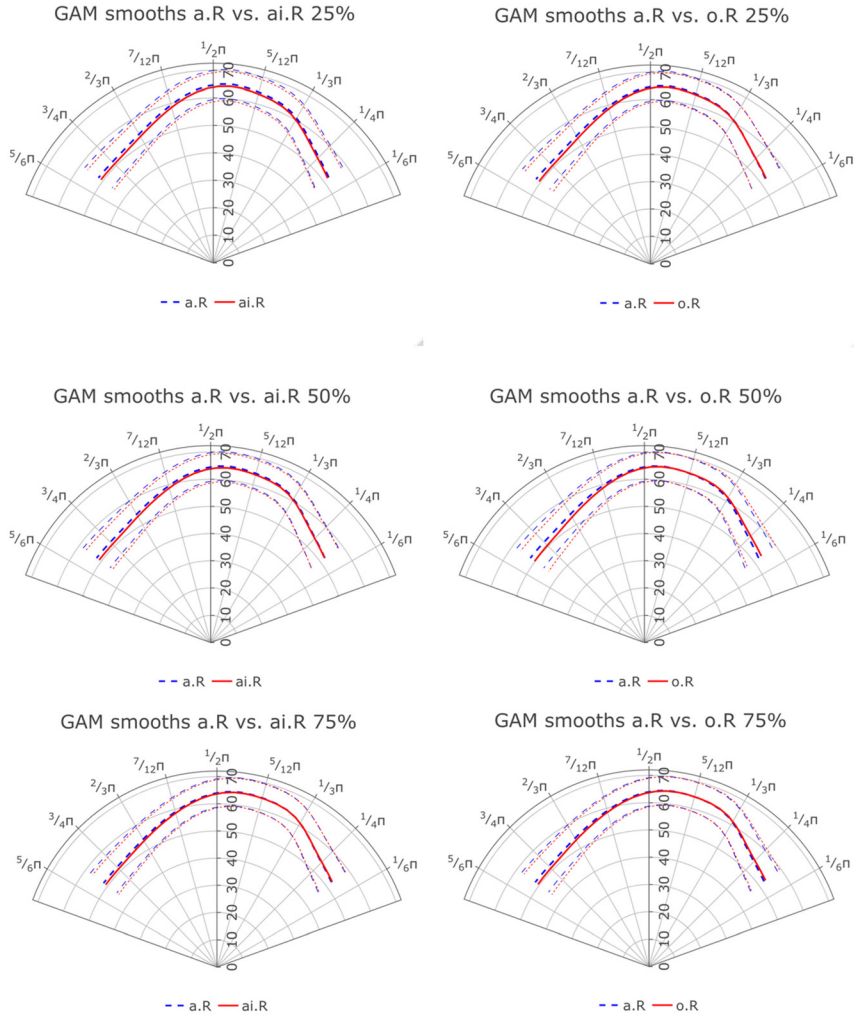


Figure 7: The fitted smoothing splines of the first quartiles (upper), midpoints (middle), and third quartiles (lower) for all tokens of the *er*-suffixed stems (a/ai.R in blue: [Ca.Cə̌] vs. [Cai.Cə̌]) versus the *er*-suffixed forms (a/o.R in red: [Ca.Cə̌] vs. [Co.Cə̌]). Gray area indicates positions with a statistically significant difference (none in this case). The speakers ($n = 10$) are facing right.

versus [Cai.Cə̌] and [Ca.Cə̌] versus [Co.Cə̌] and Figure 8 {[Ci.Cjə̌] versus [Cje.Cjə̌] and [Cu.Cwə̌] versus [Cwa.Cwə̌]}. The former *er*-suffixed stems in the pairs are represented with the dotted thick blue line and the solid red line indicates the latter *er*-suffixed stems in the pairs.

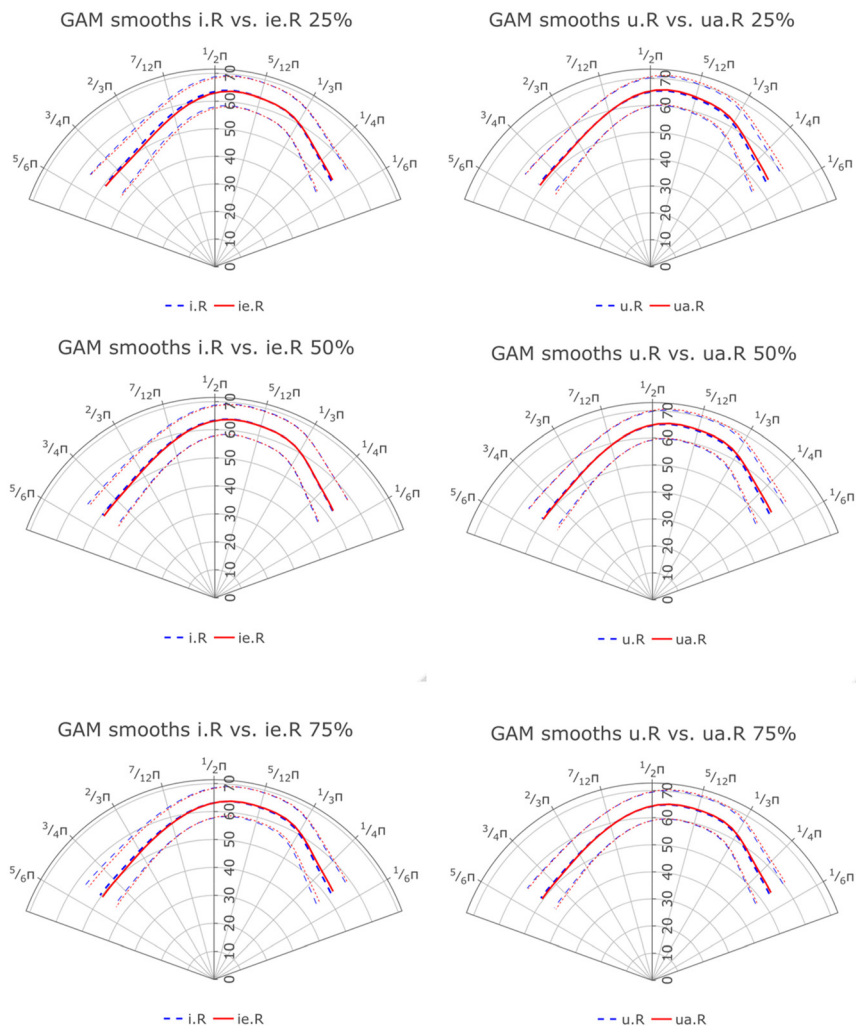


Figure 8: The fitted smoothing splines of the first quartiles (upper), midpoints (middle), and third quartiles (lower) for all tokens of the *er*-suffixed stems (i/ie.R in blue: [Ci.Ciə̌] vs. [Cje.Ciə̌]) versus the *er*-suffixed forms (u/ua.R in red: [Cu.Cwə̌] vs. [Cwa.Cwə̌]). Gray area indicates positions with a statistically significant difference (none in this case). The speakers ($n = 10$) are facing right.

We can see from Figures 7 and 8 that there is *no* significant difference between these pairs across all speakers, suggesting that these *er*-suffixes have similar tongue contours at the first quartiles (25 %), the midpoints (50 %), and the third quartiles (75 %) of an acoustically defined rime. Finally, the same conclusion may be made for the other pairs. See Appendix D for the full array of the polarscatter plots.

3.2.3 “Within-group” comparisons: acoustic results

Regarding the acoustic comparisons, a summary of GMM results of formant trajectories is given in Table 6. See also Sections 2.5.1 and 2.5.3 for analytic procedures and fn. 3 for the model adopted in this study.

Again, we can see from Table 6 that there is no perfectly consistent pattern in the pairs under comparison, either. Specifically, 10 out of 27 pairs show no significant difference across F1, F2, and F3 values.

3.2.4 Interim summary: “within-group” comparisons

The results of the pair-wise comparisons indicate that the *er*-suffixes attached to different stems are not distinguishable in the EMA and ultrasound data. In addition, there is no systematic difference found in formant trajectories, either. Taken together, the present results confirm the impressionistic transcriptions in previous

Table 6: Pair-wise comparisons between the three formants of the *er*-suffixes: GMM results ($\sqrt{}$ = difference greater than 80 % of the entire rime; $\sqrt{}$ = 80 %–50 % of the entire rime; blank = no difference or less than 50 % of the entire rime).

	F1	F2	F3
[pa.pə] versus [pai.pə]		$\sqrt{}$	
[pa.pə] versus [pan.pə]		$\sqrt{}$	
[ti.tjə] versus [tjen.tjə]	$\sqrt{}$	$\sqrt{}$	
[tje.tjə] versus [tjen.tjə]	$\sqrt{}$		$\sqrt{}$
[ka.kə] versus [ke.kə]	$\sqrt{}$		
[pei.pə] versus [pen.pə]		$\sqrt{}$	
[po.pə] versus [pon.pə]		$\sqrt{}$	
[tai.tə] versus [ke.tə]		$\sqrt{}$	
[te.tə] versus [ten.tə]	$\sqrt{}$		
[pje.pjə] versus [pjen.pjə]	$\sqrt{}$		
[ku.kwə] versus [kwa.kwə]	$\sqrt{}$		$\sqrt{}$
[ku.kwə] versus [kwan.kwə]	$\sqrt{}$		
[tɕ ^h y.tɕ ^h ɣə] versus [tɕ ^h ɣo.tɕ ^h ɣə]		$\sqrt{}$	
[pa.pə] versus [po.pə]	$\sqrt{}$		
[pa.pə] versus [p ^h e.p ^h ə]	$\sqrt{}$		$\sqrt{}$
[pa.pə] versus [p ^h u.p ^h ə]	$\sqrt{}$		
[po.pə] versus [p ^h e.p ^h ə]		$\sqrt{}$	
[te.tə] versus [tow.tə]		$\sqrt{}$	
[p ^h e.p ^h ə] versus [p ^h u.p ^h ə]	$\sqrt{}$		
<i>Pairs that show no differences:</i> [pai.pə] versus [pan.pə]/[p ^h i.p ^h ɿ] versus [pjen.pjə]/[p ^h i.p ^h ɿ] versus [pje.pjə]/[po.pə] versus [p ^h u.p ^h ɿ]/[tan.tə] versus [tai.tə]/[tan.tə] versus [ke.tə]/[kwa.kwə] versus [kwan.kwə]/[tu.twə] versus [ton.twə]			

studies (e.g., Yang 2002, Zheng 1987, among others), namely that there are only four variants of the *er*-suffix: {ə, jə, wə, ʏə} in SWM, even though it is fair to say that there is a substantial degree of incomplete neutralization both in acoustics and articulation.

3.3 The *er*-suffix and the rhotic schwa phoneme /ə/

Recall from Section 1.2 that SWM also has a rhotic schwa phoneme (/ə/), whose distributions are highly restricted, hence a marginal phoneme. Impressionistically speaking, the *er*-suffix and the rhotic schwa phoneme are not perceptibly distinctive. In this section, we compare the following pairs to see if the two rhotic schwas differ in acoustics and articulation (Table 7):

Table 7: The two rhotic schwas in Southwestern Mandarin.

[pa.pə]	‘a handle’	versus	[pa.ə]	‘slap in the face’
[po.pə]	‘a bowl’	versus	[po.ə]	‘2 nd year Ph.D. student’
[p ^h u.p ^h ə]	‘a store’	versus	[p ^h u.ə]	‘Pu-erh (tea)’
[ke.kə]	‘a cell (as in a spreadsheet)’	versus	[ke.ə]	‘Personal name’

These pairs are produced in commensurable environments since the final syllable is prosodically non-prominent in SWM (see Section 4.3). In most cases, labial onsets are used as it is assumed that labial onsets trigger the least coarticulatory carryover effects on the following vowels, especially with respect to lingual movement.

3.3.1 Comparing the two schwas: EMA results

Regarding the EMA results, a summary of GAMM results in the lingual articulators is given in Table 8. See Section 2.5.1 for analytical procedures and fn. 3 for the model adopted in this study.

The present GAMM results of the EMA recordings indicate that the rhotic schwa phoneme and the *er*-suffix mostly differ in the vertical dimension of the Tongue Dorsum (TD) sensor, with the rhotic schwa phoneme being higher than the *er*-suffix in this regard (not shown here; see Appendix C for the plots of the GAMM results of the EMA experiments).

Table 8: Summary of GAMM results in the lingual articulators: the *er*-suffix versus the rhotic schwa phoneme /ə/ (TT = Tongue Tip, TB = Tongue Body, TD = Tongue Dorsum, x = front-back, z = up-down; √√ = difference greater than 80 % of the rime; √ = greater than 50 %; blank = no difference or less than 50 % of the entire rime).

	TTx	TTz	TBx	TBz	TDx	TDz
[pa.pə] versus [pa.ə]						√
[po.pə] versus [po.ə]						√√
[p ^h u.p ^h ə] versus [p ^h u.ə]						
[ke.kə] versus [ke.ə]						√

3.3.2 Comparing the two schwas: ultrasound results

The polar scatter plots of the *er*-suffix versus the rhotic schwa phoneme are illustrated in Figure 9. See Section 2.5.2 for analytical procedures and fn. 4 for the model adopted in this study.

As shown, the *er*-suffix and the rhotic schwa phoneme /ə/ have significantly different tongue postures both at the midpoints (50 %) and at the third quartiles (75 %) in all the three pairs across all ten speakers.

3.3.3 Comparing the two schwas: acoustic results

Next consider Table 9, in which the *er*-suffix and the vowel phoneme /ə/ are compared with respect to formant values. See Sections 2.5.1 and 2.5.3 for the analytical procedures of GAMM analysis and fn. 3 for the model adopted in this study.

The acoustic results thus suggest that the /ə/ phoneme is acoustically different from the *er*-suffix along the F1 dimension in all the three pairs across all 10 speakers. It is equally remarkable that the rhotic schwa phoneme /ə/’s have higher formant values across the board (not shown here; see Appendix E for the plots of the GAMM results).

3.3.4 Interim summary: the *er*-suffix versus the rhotic schwa phoneme

In sum, the phonetic differences between the *er*-suffix and the vowel phoneme /ə/ can be recapitulated as follows:

- The rhotic schwa phoneme /ə/ is usually higher along the vertical dimension of the Tongue Dorsum (TDz) sensor (EMA results)
- The rhotic schwa phoneme /ə/ and the *er*-suffix have significantly different tongue shapes both at the midpoints (50 %) and the third quartiles (75 %) of an acoustically defined rime (Ultrasound results)

Table 9: The *er*-suffix versus the rhotic schwa phoneme: GAMM results (√/√ = the two rhotic vowels differ along a certain dimension (F1, F2, or F3) throughout 80 % or more of the entire rime; √ = 50 %–80 % of the entire rime; blank = less than 50 % of the entire rime).

	F1	F2	F3
[pa.pə] versus [pa.ə]	√	√√	
[po.pə] versus [po.ə]	√√	√√	
[p ^h u.p ^h ə] versus [p ^h u.ə]	√√		
[ke.kə] versus [ke.ə]	√√		

- The rhotic schwa phoneme /ə/ has higher F1 values across the board (Acoustic results)

3.4 Non-lingual components of the rhotics: lip rounding and pharyngealization

In this section, we move on to two more articulatory characteristics of the rhotic sounds reported in the literature, which will be addressed in turn below.

3.4.1 Tongue root retraction to the pharynx

Rhotic vowels are not produced with tongue root retraction to the pharynx (see Hussain and Mielke (2021) for a recent survey). But recall that Lee and Zee (2014: 386) report that “the tongue body is retracted towards the pharynx” during *er*-suffixation in Beijing Mandarin (see also Xing 2021). For this reason, it is necessary to examine if tongue root retractions occur in the production of the *er*-suffix and rhotic schwa phoneme in SWM. In this section, the ultrasound data are used to observe the posterior portion of the tongue dorsum, which cannot be reliably captured by flesh-point tracking systems like EMA restricted to the anterior oral tract for sensor placement.

We compare the tongue postures between the first quartiles (25 %) and the third quartiles (75 %) of the same *er*-suffixes and rhotic schwas using ultrasound imaging data. In Figures 10 and 11, the blue dashed lines refer to the tongue postures at the first quartiles (25 %) of the rime, the red lines represent the tongue postures at the third quartiles (75 %). Note further that we did not include the results of rising diphthongs (i.e., {jə, wə, ʉə} here). The representative data are provided in Figure 9, where the *er*-suffixed forms are {[Ca.Cə]; [Co.Cə]; [Ce.Cə] and [Cai.Cə]}, where C = {p,

p^h , t , k , $t\epsilon^h$ }, if available). See Appendix D.7 for the polarscatter plots of $\{[Ce.C\grave{a}]; [Cen.C\grave{a}]; [Can.C\grave{a}]\}$, whereby similar observations may be made.

We can see in Figure 10 that there is no obvious tongue root retraction in the left-hand halves of the polarscatter plots. It is also remarkable that the tongue postures differ significantly between the first and third quartiles, suggesting that the *er*-suffix is, to some extent, diphthongized (see Jiang et al. 2019 for an identical finding regarding the rhotic schwa phoneme in Northeastern Mandarin). Likewise, the same observations hold true for the rhotic schwa phoneme, too. Consider now Figure 11.

In sum, we conclude that the two rhotic schwas do not involve pharyngealization, unlike their counterparts in Beijing Mandarin (Lee and Zee 2014; Xing 2021). Moreover, SWM and Northeastern Mandarin are similar in that the rhotic schwas are both diphthongized.

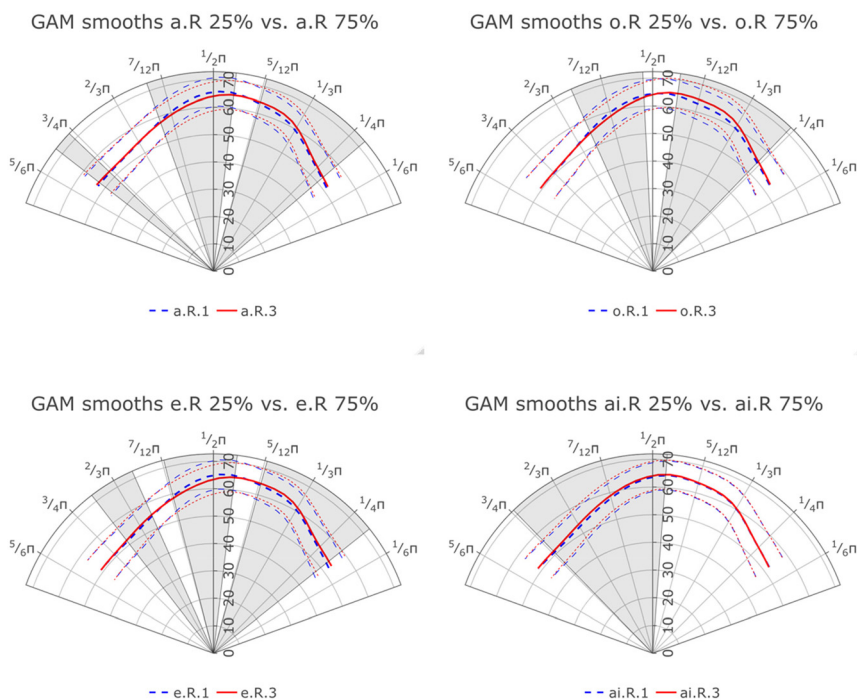


Figure 10: The fitted smoothing splines of the first quartiles (25 %: xR.1) and the third quartiles (75 %: xR.3) for all tokens of the *er*-suffixed stems ($x = a/o/e/ai$: $\{[Ca.C\grave{a}], [Co.C\grave{a}], [Ce.C\grave{a}], [Cai.C\grave{a}]\}$). Gray area indicates positions with a statistically significant difference. The speakers ($n = 10$) are facing right.

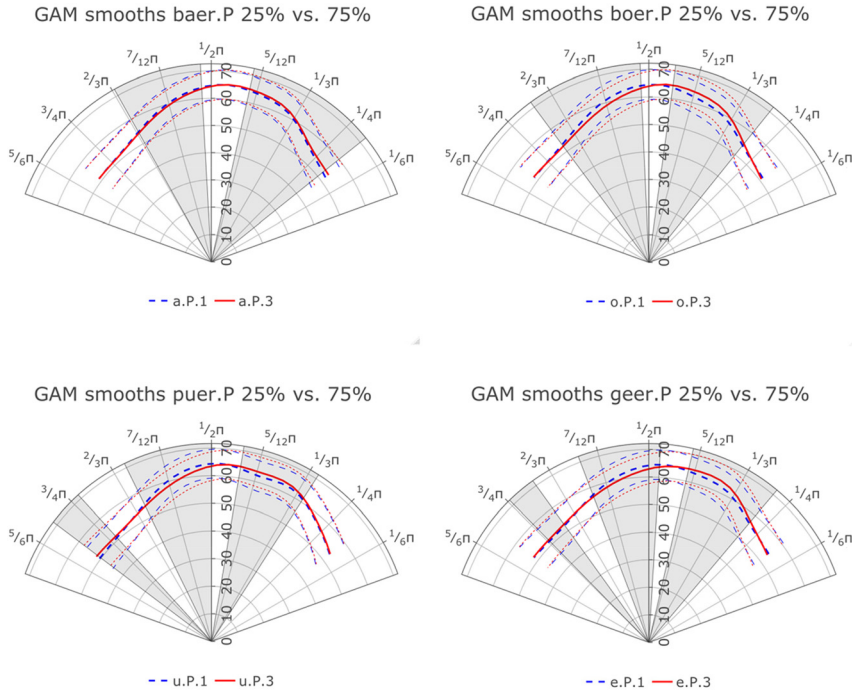


Figure 11: The fitted smoothing splines of the first quartiles (25 %: xP.1) and the third quartiles (75 %: xP.3) for all tokens of the rhotic schwa phoneme ($x = a/o/u/e$: [pa.ə], [po.ə], [p^hu.ə], [ke.ə]). Gray area indicates positions with a statistically significant difference. The speakers ($n = 10$) are facing right.

3.4.2 Lip rounding

Lip rounding is one of the key components in English rhotic sounds, especially in prevocalic positions (see King and Ferragne 2020 for a recent update and references cited therein), although Hussein and Mielke (2021: 22) remark that “vowel rhoticity does not entail lip rounding.” For the sake of thoroughness, the empirical issue to be addressed in this section is whether lip protrusion can be found in the two rhotic schwas. In this section, we compare the trajectories of the Upper Lip (UL) and Lower Lip (LL) sensors along the longitudinal (front-back) direction between the monosyllabic stems versus the *er*-suffix as well as the rhotic schwa phoneme. It is generally acknowledged that the advancement of the UL sensor corresponds to lip protrusion (Farnetani 1999, Westbury and Hashi 1997, among others), while LL may be

confounded with jaw movement (see, e.g., Fletcher and Harrington 1999). In Tables 10 and 11, both the comparisons between the UL and LL sensors are provided, again, for the sake of thoroughness. See also Appendix C for all the GAMM results.

We can see from Table 10 that only three pairs differ along the longitudinal dimension of Upper Lip (ULx), suggesting that the *er*-suffix does not frequently involve lip protrusion (4 out of 9 pairs in comparisons, whereby no difference found in {[po] vs. [po.pə]} means that this particular *er*-suffixed form also involves lip rounding; see also Table 11). The differences are more robust along the longitudinal dimension of Lower Lip (LLx), but as mentioned earlier, LLx movements may well be a passive consequence of jaw lowering. Consequently, the *er*-suffix may not be described as a rounded vowel.

In the same vein, we compare the trajectories of ULx and LLx between the monosyllabic stem [po] ‘thin’ and the rhotic schwa phoneme /ə/ since the mid rounded vowel /o/ is closest to the rhotic schwa phoneme in SWM. Consider now Table 11.

We can see from Table 11 that the rhotic schwa phoneme and the mid rounded vowel do *not* differ along the longitudinal dimensions of both Upper Lip (ULx) and Lower Lip (LLx), suggesting that the rhotic schwa phoneme is *not* different from the

Table 10: Summary of GAMM results in the labial articulators: monosyllabic stem vowels versus the *er*-suffixed stems (UL = Upper Lip, LL = Lower Lip, x = front-back; √√ = difference greater than 80 % of the rime; √ = greater than 50 %; blank = no difference or less than 50 % of the entire rime).

	ULx (Horizontal)	LLx (Horizontal)
[pa] versus [pa.pə]	√√	√√
[pai] versus [pai.pə]		√
[pan] versus [pan.pə]	√√	√
[ke] versus [ke.kə]	√√	√√
[p ^h i] versus [p ^h i.p ^h ə]		√√
[tan] versus [tan.tə]		√
<i>Pairs that show no difference:</i> po versus po.pə/te versus te.tə/ta versus ke.tə		

Table 11: Summary of GAMM results in the labial articulators: /po/ ‘thin’ versus the rhotic schwa phoneme /ə/ (UL = Upper Lip, LL = Lower Lip, x = front-back; n.s. = ‘not significant’ across the entire rime).

	ULx (Horizontal)	LLx (Horizontal)
[po] versus {[pa.ə], [po.ə] [p ^h u.ə], [ke.ə]}	n.s.	n.s.

rounded vowel /o/ with respect to lip protrusion. To this end, we may conclude that, unlike the results in Table 10, the /ə/ phoneme may be transcribed as a rounded/labialized rhotic schwa in SWM.

3.5 Summary

In this section, we have presented the results of the acoustic and articulatory experiments of the *er*-suffix and the rhotic schwa phoneme in SWM. To recapitulate, our principal findings are itemized as follows:

- Both the *er*-suffix and the rhotic schwa phoneme are invariably produced with a bunched tongue configuration (i.e., Tongue Retroflexion Angle > 0°; see Tables 3 and 4). This study attests not a single instance of retroflex/tip-up rhotic schwas.
- The *er*-suffixes may be different when attached to different stems, articulatorily (EMA/Ultrasound), acoustically (F1/F2/F3 values), or both. However, no consistent difference may be found both in acoustics and articulation. In other words, there are four variants of the *er*-suffix: {ə, jə, wə, ɤə}.
- There is no significant tongue root retraction in the production of the two rhotic schwas.
- The *er*-suffix does not involve (consistent) lip protrusion and the rhotic schwa phoneme may be described as a rounded rhotic schwa.
- The *er*-suffix and the rhotic schwa phoneme are both diphthongized (irrespective of the instances of the diphthongs in {jə, wə, ɤə}).

4 Discussion

There are two principal findings in this study. First, our quantitatively based results show that no retroflex versions of the two rhotic schwas were found. Second, the *er*-suffix and the /ə/ phoneme differ in acoustics and articulation, even though the two rhotic schwas are perceptibly indistinguishable. Finally, our discussion is closed with a note on the diachrony, synchrony, and typology of the *er*-suffixation across Sinitic languages.

4.1 Whence comes the articulatory uniformity in the production of the two rhotic schwas?

It is well-established in previous articulatory studies that there is intra- and inter-speaker variation in the tongue shape of the consonantal /ɹ/ in English (Delattre and

Freeman 1968, et seq.). In the *syllabic* context, Mielke et al. (2016), among others, note that bunching is more frequently found in /ə/ than in onset /ɹ/ in American English. Most of the speakers (23 out of 27) produced /ə/ with a bunched tongue posture. In addition, Mielke (2015) finds a very similar rate of bunching (6 out of 7) in Canadian French rhotic vowels. More recently, Hussein and Mielke (2021) report that Kalasha rhotic vowels (i.e., {i, ɪ, e, ẽ, a, â, o, õ, u, û}) are bunched for the four speakers in their ultrasound study. In Hussein and Mielke (2022: 12), the emergence of the rhotic vowels in Kalasha is hypothesized as a reflex of the diachronic loss of the retroflex approximant coda /-ɹ/, via the following evolutionary path (where “*” indicates reconstructed forms): *Vɹ (Old Indo-Aryan) → *V(ɹ)ɹ (Kalasha) → *Vɹ (Kalasha) → V (Kalasha). That being the case, it may not be surprising to see that the rhotic vowels are consistently produced with a bunched tongue shape in Kalasha because the synchronic rhotic vowels have had an identical source.⁵ In the same vein, Hussein and Mielke (2022) further entertain the possibility that the emergence of the rhotic schwa in Canadian French can be attributed to the gradual exaggeration of the lowered F3 as the result of an increase of bunching among the front rounded vowels. Returning to SWM, the two rhotic schwas seem no exception to this pattern (i.e., the bunching as the dominant tongue shape). Indeed, one of our principal findings is that the two rhotic schwas are produced exclusively with the bunching of the tongue body, since not a single token of retroflex versions of them is identified in our quantitatively-attained results (i.e., Tiede et al.’s 2019 Tongue Retroflexion Angle; see Tables 3 and 4), at least in the present study.⁶

In this limited cohort of languages discussed so far, the rhotic vowels tend to favor the bunching of the tongue body. As a matter of fact, the correlation of the (syllabic) rhotic vowels and the bunched tongue shape has been dated back to Uldall (1958), according to Mielke et al. (2016). This cross-linguistic preference is further strengthened in our study of the two rhotic schwas in SWM, a Sinitic language. To this

5 It is worth mentioning that a similar evolutionary path for the Kashala rhotic vowels may be proposed for the rhotic schwas in SWM, too. Diachronically speaking, the *er*-suffix and the rhotic schwa phoneme in Middle Chinese do have an onset, which is dubbed the *ri* (‘sun’)-initial in the jargon of traditional Chinese historical phonology. This can be seen in Sinoxenic pronunciations of *er* ‘child’, e.g., <zi> in Sino-Japanese and <nhɿ> in Sino-Vietnamese (Eom 1995). Following Hsueh (1980), we may say that the rhotic schwas are diachronically derived through these steps: */Ri/ → ... */iR/ → /ə/ (where R means the *ri*-initial, which has been reconstructed as a palatal nasal or a nasal-sibilant sequence in Middle Chinese). In other words, it is likely that the rhoticity of the rhotic schwas, too, originated from the adjacent consonant.

6 We acknowledge the possibility that our sample size may have been too low to detect a retroflex /ə/. Suppose that the incidence of retroflex production of /ɹ/ in North American English is about 1 in 10 (cf. 4 retroflex /ə/’s out of 27 /ə/’s in Mielke et al. (2016)). Assuming further for argument’s sake that something similar holds in SWM, we have about a 65.13 % probability of seeing one /ə/ in 10 experiments, and we would need 29 experiments to have a 95 % chance of seeing one retroflex /ə/.

end, it is tempting to anticipate a unified analysis for it. For example, Mielke et al. (2016) propose an OT-style constraint $*\text{CODA } \text{ɹ}$ to penalize retroflexion in coda position and this constraint could be motivated by a putative preference for larger anterior gestures in onset position. Mielke et al. (2016: 128) further remark that “[r]etroflexion is more frequent in contexts that do not place conflicting demands on the tongue tip, such as word boundaries, labial consonants, back vowels, and /l/” (see also Heyne et al. (2020) for similar results of (non-rhotic) New Zealand English; cf. the biomechanical modeling of rhotic variation set forth by Stavness et al. (2012)). On the other hand, Scobbie et al. (2015) report that it is extremely rare for Scottish English speakers to have this particular pattern of /ɹ/ allophony: bunched (B) onsets and retroflexed (R) codas, while the other patterns, RR, BB, RB are more or less evenly distributed in the corpora. Scobbie et al. (2015) speculate that the retroflexed shape is inherently more rhotic (“stronger”) than a bunched one. That being the case, the retroflexed tongue shape might be more compatible with “strong” onsets. The above-mentioned cannot be carried over to the case of the rhotic schwas in SWM, however. For one thing, the two rhotic schwas both occupy the nuclear position, not the “weak” coda position. For another thing, contextual segmental effects have been mostly if not all excluded due to our experimental design (see Appendix A for the wordlist). Here we offer one plausible explanation, based on Maddieson’s (1995: 574) proposal of *gestural economy*, according to which “there is [a tendency] to be economical in the number and nature of the distinct articulatory gestures used to construct an inventory of contrastive sounds, and it is this (rather than a more abstract featural analysis) that underlies the observed system symmetry” (see also Bybee (2001) for a similar account). In essence, gestural economy is analogous to Clements’s (2003: 287) principle of *feature economy* (namely, “languages tend to maximise the ratio of sounds over features”) operating at the phonetic level. By the same token, it is likely the reason why no rhotic schwa with the retroflexed tongue shape is found in this study is because the Tongue Tip gesture is not used in the production of vowels in SWM. In other words, if the acoustic target can be reliably achieved via the bunching of the tongue body, there seems no need to add an extra one to the repository. Note further that there is no rhotic or r-colored sound in the phoneme inventory; in particular, recall that the rhotic approximant onset /ɹ/, the “retroflex” apical vowel, and the “retroflex” sibilants in Beijing Mandarin have been lost in contemporary SWM already (see Section 1.2). We should also acknowledge the likely role of morphological differences in hard palate shape, including parasagittal shape, which may favor bunching as a strategy for achieving lowered F3. Finally, as previously noted, it is possible that we did not sample enough speakers of SWM to actually observe an instance of retroflex /ə/ production. In sum, it remains to be seen as to how and why the retroflexed tongue shapes seem cross-linguistically dispreferred in postvocalic and syllabic contexts.

4.2 Morphologically-induced contrast preservation

The other major finding in this study is that the two rhotic schwas differ in acoustics and articulation, as summarized in Section 3.5. At first glance, these differences might be simply treated as a consequence of “contrast preservation” in phonological mappings (Kiparsky 1973; Martinet 1967 [1961]; Trubetzkoy 1971 [1939]) as well as in phonetic implementation (Flemming 2004), even though it must be emphasized that the contrast between the two rhotic schwas is perceptually inconspicuous, albeit articulatorily distinct. Further scrutiny, however, reveals that the relation between the *er*-suffix and the rhotic schwa phoneme is interesting in that the two rhotic vowels are *always* in complementary distribution. Precisely, recall that the rhotic schwa phoneme cannot be combined with any syllable margin (i.e., onset and/or coda), and the *er*-suffix cannot stand alone (i.e., an onset is obligatory for the *er*-suffix). Taken together, it is obvious that the two rhotic vowels in question never endanger a contrast across the board. That being the case, the functional motivation behind contrast preservation seems irrelevant here. It is thus quite puzzling as to why reuse of phonetic targets or individual gestures across multiple speech sounds is not invoked, as has been amply documented in the literature (Chodroff and Wilson 2017; Chodroff and Wilson 2022; Faytak 2018; Fruehwald 2017; Guy and Hinskens 2016; Keating 2003; Lindblom 1983; Maddieson 1995; Ménard et al. 2008), especially when contrast preservation (or, phonological distinctiveness) is apparently not at issue here. Finally, our experimental design excluded or minimized the effects of other potential confounds as well. In particular, the target syllables under comparison are all in final position. In SWM, it has been confirmed the final syllable is prosodically weak in a disyllabic window (see Section 4.3 below for more detail). Therefore, it is fair to say that the *er*-suffix and the rhotic schwa phoneme were compared in commensurable contexts. To this end, one remaining possibility we can think of is that the grammar (or, the module of phonetic implementation) strives to distinguish between contentive and functional morphemes in phonetic realization. The present finding is thus reminiscent of Plag’s et al. (2017) findings, according to which the morphemic /s/ and /z/, for example, are significantly different from the non-morphemic /s/ and /z/ in terms of acoustic realization in English. We agree with Plag et al. (2017) that morphologically-induced phonetic variations of this sort cannot be adequately explained by both phonological theory and extant psycholinguistic models. For now, we leave the exact mechanisms for further studies in the future, by remarking that the morphologically-driven homophony avoidance in articulation, albeit imperceptible, has not been reported elsewhere, especially when contrast preservation is not at issue.

4.3 Diachrony, synchrony and typology of *Er*-suffixation

The *er*-suffix in SWM can be regarded as a half-grammaticalized suffix, if compared with its cognate suffix in Northern Mandarin. As mentioned in Section 1.1, the *er*-suffix may be realized as either a part of a diphthong or a floating feature, resulting in a rhotacized vowel in Beijing and Northeastern Mandarin (cf. the umlauting in German; see, e.g., Trommer 2021 for an updated discussion). By contrast, as witnessed in Section 3, our data have confirmed that the *er*-suffix in SWM is invariably a rhotic schwa, inducing the process of rime usurpation. One possible explanation is that the “segmental contents” of the *er*-suffix are not completely lost in SWM, or in terms of the mainstream framework of syllable weight, the *er*-suffix is underlyingly moraic. From a cross-linguistic/dialectal perspective, the diachronic evolution of the *er*-suffix may be sketched in (1), following Lin’s (2004) terms:

- (1) Full-segment as a separable affix → Full-segment incorporated into the root → Feature-sized Affix

At one extreme is the *er*-suffix in Hangzhou Chinese (a dialect of Wu Chinese, which has been extensively influenced by Pre-modern Mandarin ever since the Qing dynasty), for example. In Hangzhou Chinese, the *er*-suffix remains a separable, full-segment suffix and is transcribed as a retroflex lateral, according to Yue and Hu’s (2019) experimental results. At the other extreme, by contrast, are affixes that have lost their segmental contents and have fully grammaticalized into a floating feature. The *er*-suffix in Jiyuan Chinese (a dialect of Zhongyuan or “Central Plains” Mandarin) is a case in point (see Lin 2004 and references cited therein). Similarly, Lee (2005) and Jiang et al.’s (2019) EMA results confirm that certain output forms of the *er*-suffixation are a rhotacized rime in Beijing and Northeastern Mandarin, respectively (e.g., /ʊ-/). The SWM *er*-suffix appears to be the transitional stage between the full-sized, stand-alone *er*-suffix and the feature-sized *er*-suffix. That being the case, we propose that the rime usurpation phenomenon may be attributed to the fact that disyllabic words are prototypically trochaic in SWM because the full-toned final syllable is significantly shorter than the initial syllable in duration. Liu et al. (2022) report the results ($n = 6$) that the initial syllables are 1.5/1.35 times longer than the final syllables for disyllabic compound/monomorphemic words in Chengdu Chinese (i.e., the representative variety of SWM; see Qin 2015 for similar results), while the ratio is 0.9 for Standard Chinese, (which is based on the data from the same group of Chengdu Chinese speakers). The final syllable is slightly longer than the initial syllables in Standard Chinese, probably due to the effect of phrase-final lengthening. Liu et al. (2022) also remark that Chengdu Chinese is more likely to be a “stress-timed” language than Standard Chinese because the mean nPVI (normalized Pairwise

Variability Index) is significantly greater in Chengdu than in Standard Chinese: 55.3 and 35.4, respectively. Likewise, the same result may be found among tri-syllabic compound words, too. Recall that the *er*-suffix must be attached to a polysyllabic stem in SWM. It follows that suffixing a rhotic vowel leads to an “oversized” rime in final position, hence rime usurpation.

5 Conclusions

In this study, we have investigated the acoustic and articulatory properties of the two rhotic schwas in Southwestern Mandarin. There are two key discoveries. First, our EMA and ultrasound results from ten speakers show that the rhotic schwas are both produced exclusively with the bunched tongue body, meaning that at least in the present study, no retroflex versions of this suffix were found (i.e., Retroflexion Angle $\leq 0^\circ$; see also Tables 3 and 4), nor were tongue root retractions into the pharynx observed (cf. the Beijing Mandarin *er*-suffix described in Lee and Zee (2014) and Xing (2021)). We proposed that absence of the retroflex tongue configurations in the two rhotic schwas may be attributed to Maddieson’s (1995) gestural economy because the sound inventory and morpho-phonological alternations are substantially simplified in SWM, unlike Beijing and Northeastern Mandarin. Second, we found significant acoustic and articulatory differences between the *er*-suffix and the rhotic schwa phoneme, even though the two rhotic schwas are homophonous. We leave it for future work, by speculating that the incomplete neutralization we found here might be motivated by the contentive (the rhotic schwa phoneme) versus functional (the *er*-suffix) distinction in morphology. In a broader perspective, the present results suggest that there may be morphologically- or phonologically-driven variation and uniformity in speech production.

Research funding: This study was funded by Ministry of Science and Technology, Taiwan, ROC (MOST 109-2410-H007-061).

Author contributions: **Jing Huang:** Conceptualization, Data curation, Investigation, Visualization, Writing – original draft. **Feng-fan Hsieh:** Conceptualization, Investigation, Formal analysis, Methodology, Resources, Supervision, Writing – original draft. **Yueh-chin Chang:** Funding acquisition, Writing – review and editing. **Mark Tiede:** Software, Writing – review and editing.

Conflict of interest statement: The authors have no conflicts of interest to declare.

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

References

- Boersma, Paul & David Weenink. 2007. *Praat: Doing phonetics by computer [computer program]*. Version 6.0.30. Available at: <http://www.Praat.org>.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Chao, Yuen-Ren. 1970 [1968]. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen, Shuwen & Peggy Pik Ki Mok. 2021. Articulatory and acoustic features of Mandarin /ɹ/: A preliminary study. In *Proceedings of the 12th International Symposium on Chinese Spoken Language Processing (ISCSLP 2021)*, 1–5.
- Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47.
- Chodroff, Eleanor & Colin Wilson. 2022. Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language* 98(2). 250–289.
- Clements, George N. 2003. Feature economy in sound systems. *Phonology* 20(3). 287–333.
- Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American English r's by x-ray motion picture. *Linguistics* 44. 28–69.
- Duanmu, San. 2007. *The phonology of standard Chinese*. New York: Oxford University Press.
- Eom, Ik-sang. 1995. Some old Chinese initials in Sino-Korean and Sino-Japanese. *Studies in Language and Linguistics* 29(2). 69–84.
- Espy-Wilson, Carol Y., Suzanne E. Boyce, Michel Jackson, Shrikanth Narayanan & Abeer Alwan. 2000. Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America* 108(1). 343–356.
- Farnetani, Edda. 1999. Labial coarticulation. In William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation: theory, data and techniques (Cambridge Studies in Speech Science and Communication)*, 144–163. Cambridge: Cambridge University Press.
- Faytak, Matthew D. 2018. *Articulatory uniformity through articulatory reuse: Insights from an ultrasound study of Sūzhōu Chinese*. The University of California Berkeley Unpublished Ph.D. Dissertation.
- Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge, UK: Cambridge University Press.
- Fletcher, Janet & Johnathan Harrington. 1999. Lip and jaw coarticulation. In William Hardcastle & Nigel Hewlett (eds.), *Coarticulation: Theory, data and techniques (Cambridge Studies in Speech Science and Communication)*, 164–176. Cambridge: Cambridge University Press.
- Fruehwald, Josef. 2017. The role of phonology in phonetic change. *Annual Review of Linguistics* 3. 25–42.
- Guy, Gregory R. & Frans Hinskens. 2016. Linguistic coherence: Systems, repertoires and speech communities. *Lingua* 172(173). 1–9.
- Hall, Kathleen Currie. 2013. A typology of intermediate phonological relationships. *The Linguistic Review* 30(2). 215–275.
- Hartman, Lawton M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20(1). 28–42.
- He, Wan & Dongmei Rao. 2014. *Sichuan Chengduhua Yinxi Cihui Diaocha Yanjiu [A phonological and lexical study of Chengdu Chinese]*. Chengdu: Sichuan University Press.
- Heyne, Matthias, Donald Derrick & Jalal Al-Tamimi. 2019. Native language influence on brass instrument performance: An application of generalized additive mixed models (GAMMs) to midsagittal ultrasound images of the tongue. *Frontiers in Psychology* 10. 2597.
- Heyne, Matthias, Xuan Wang, Donald Derrick, Kieran Dorree & Levin Watson. 2020. The articulation of /ɹ/ in New Zealand English. *Journal of the International Phonetic Association* 50(3). 366–388.

- Huang, Tsan. 2010. Er-Suffixation in Chinese monophthongs: Phonological analysis and phonetic data. In *22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18)*, vol. 1, 331–334. MA: Harvard University, Cambridge.
- Hussain, Qandeel & Jeff Mielke. 2021. An acoustic and articulatory study of rhotic and rhotic-nasal vowels of Kalasha. *Journal of Phonetics* 87. 101028.
- Hussain, Qandeel & Jeff Mielke. 2022. The emergence of bunched vowels from retroflex approximants in endangered Dardic languages. *Linguistics Vanguard* 8(s5). 597–610, 20210022.
- Hsueh, Feng-Sheng. 1980. Lun “Zhi-Si yun” de xingcheng yu yanjin [On the formation and evolution of “Zhi-Si rimes”]. *Bibliography Quarterly* 14(2). 53–74.
- Jiang, Song, Yueh-chin Chang & Feng-fan Hsieh. 2019. An EMA study of er-suffixation in Northeastern Mandarin monophthongs. In *Proceedings of 19th international congress of phonetic sciences, Melbourne, Australia 2019*, 3617–3621. Canberra: Australasian Speech Science and Technology Association Inc.
- Keating, Patricia. 2003. Phonetic and other influences on voicing contrasts. In Maria-Josep Solé, Daniel Recasens & Joaquín Romero (eds.), *Proceedings of the 15th international congress of phonetic sciences*, 20–23. Barcelona, Spain.
- King, Hannah & Emmanuel Ferragne. 2020. Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics* 80. 100978.
- Kiparsky, Paul. 1973. Abstractness, opacity and global rules. In Osamu Fujimura (ed.), *Three dimensions in linguistic theory*, 57–86. Tokyo: TEC.
- Lee, Wai-Sum. 2005. A phonetic study of the “er-hua” rimes in Beijing Mandarin. In *Ninth European Conference on Speech Communication and Technology*, 1093–1096.
- Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1). 109–112.
- Lee, Wai-Sum & Eric Zee. 2014. Chinese phonetics. In James C.-T. Huang, Audrey Y.-H. Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 369–399. Chichester: Wiley-Blackwell.
- Li, Lan. 1997. Liushi nian lai Xinanguanhua de diaocha yu yanjiu [The survey and research on Southwestern Mandarin during the last sixty years]. *Fangyan* 4. 249–257.
- Li, Lan. 2009. Xinanguanhua de fenqu [The division of Southwestern Mandarin]. *Fangyan* 1. 16.
- Lee-Kim, Sang-Im. 2014. Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3). 261–282.
- Li, Sijing. 1986. *Hanyu ‘Er’ /ə/ Yinshi Yanjiu* [A historical analysis of Chinese /ə-/]. Beijing: The Commercial Press.
- Lin, Tao & Tong Shen. 1995. Beijinghua erhuayun de yuyin fenqi [Variations in the [er]-suffixed rhymes in the Beijing dialect]. *Zhongguo Yuwen* 3. 170–179.
- Lin, Yen-Hwei. 2004. Chinese affixal phonology: Some analytical and theoretical issues. *Language and Linguistics* 5(4). 1019–1046.
- Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.
- Lindau, Mona. 1985. The story of /r/. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 157–168. Orlando, Florida: Academic Press.
- Lindblom, Björn. 1983. Economy of speech gestures. In *The production of speech*, 217–245. New York, NY: Springer.
- Lindblom, Björn & Ian Maddieson. 1988. Phonetic universals in consonant systems. In Charles Li & Larry Hyman (eds.), *Language, speech and mind: studies in honour of Victoria A. Fromkin*, 62–78. London: Routledge.

- Liu, Chunhui, Kechun Li & Francis Nolan. 2022. Lun shengdiaoyuyande jiezhou yu zhongyin moshi [On the speech rhythm and stress pattern of tone languages]. *Journal of Sichuan University (Philosophy and Social Science Edition)* 240(3). 151–163.
- Maddieson, Ian. 1984. *Patterns of sounds*. New York: Cambridge University Press.
- Maddieson, Ian. 1995. Gestural economy. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th International congress of phonetic sciences*, vol. 4, 574–577. Stockholm: KTH & Stockholm University.
- Martinet, André. 1967 [1961]. *Elements of general linguistics*. Chicago: University of Chicago Press.
- Ménard, Lucie, Jean-Luc Schwartz & Jérôme Aubin. 2008. Invariance and variability in the production of the height feature in French vowels. *Speech Communication* 50(1). 14–28.
- Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137(5). 2858–2869.
- Mielke, Jeff, Adam Baker & Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–140.
- Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181–216.
- Potter, Ralph, George Kopp & Harriet Green. 1947. *Visible speech*. New York: D. Van Nostrand Co., Inc.
- Qin, Zhuxuan. 2015. Chengduhua de Liandubiandiao yu Yunlüjiegou [Tone sandhi and prosodic structure in Chengdu dialect]. *Hanyu Xuebao [Chinese Linguistics]* 50(2). 36–44.
- Rebernik, Teja, Jidde Jacobi, Roel Jonkers, Aude Noiray & Martijn Wieling. 2021. A review of data collection practices using electromagnetic articulography. *Laboratory Phonology* 12(1). 6.
- Scobbie, James M., Eleanor Lawson, Satsuki Nakai, Joanne Cleland & Jane Stuart-Smith. 2015. Onset vs. coda asymmetry in the articulation of English /r/. In *Proceeding of the 18th International congress of phonetic sciences*, 0704, 1–5. Glasgow: The University of Glasgow.
- Shi, Feng. 2003. Beijinghua erhuayun de shengxue biaoqian [Acoustic correlates of Beijing Er-rhymes]. *Nankai Yuyanxue Kan* 1. 10.
- Sóskuthy, Márton. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv:1703.05339*.
- Sóskuthy, Márton. 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84. 101017.
- Spreatico, Lorenzo, Michael Pucher & Anna Matosova. 2018. UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. In *Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India, 2–6 September 2018*, 1517–1520. International Speech Communication Association.
- Stavness, Ian, Bryan Gick, Donald Derric & Sidney Fels. 2012. Biomechanical modeling of English/r/ variants. *The Journal of the Acoustical Society of America* 131(5). EL355–EL360.
- Thomas, Erik R. & Tyler Kendall. 2007. NORM: The vowel normalization and Plotting suite. Available at: <http://lingtools.uoregon.edu/norm>.
- Tiede, Mark, Wei-rong Chen & Douglas H. Whalen. 2019. Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and ultrasound. In *Proceedings of the International Congress of Phonetic Sciences*, 427–431. Melbourne, Australia.
- Trommer, Jochen. 2021. The subsegmental structure of German plural allomorphy. *Natural Language & Linguistic Theory* 39(2). 601–656.
- Trubetzkoy, Nikolai S. 1971 [1939]. *Principles of phonology*. Berkeley: University of California Press.
- Uldall, Elizabeth. 1958. American “molar” R and “flapped” T. *Revista do Laboratório de Fonética Experimental da Faculdade de Letras da Universidade de Coimbra* 4. 103–106.
- van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2017. Itsadug: Interpreting time series and autocorrelated data using GAMMs. *R packages version 2.3*.

- Wang, Zhijie. 1997. Erhua de Tezheng Jiegou [The featural geometry of r-rhymes]. *Zhongguo Yuwen* 1. 2–10.
- Westbury, John R. & Michiko Hashi. 1997. Lip-pellet positions during vowels and labial consonants. *Journal of Phonetics* 25(4). 405–419.
- Wieling, Martijn. 2018. 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* 70. 86–116.
- Wood, Simon. 2017 [2006]. *Generalized additive models: An introduction with R*. Boca Raton: CRC—Chapman & Hall.
- Wood, Simon. 2019. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. *R-package version 1.8–31*.
- Wrench, Alan A. & James M. Scobbie. 2016. Queen Margaret University ultrasound, audio and video multichannel recording facility (2008–2016). QUM CASL Working Paper: WP(24). 1–14.
- Wurm, Stephen A., Rong Li & Theo Baumann. 1987. *Language atlas of China*. Hong Kong: Longman.
- Xing, Kaiyue. 2021. *Phonetic and phonological perspectives on rhoticity in Mandarin*. The University of Manchester Unpublished Ph.D. Dissertation.
- Yang, Shaolin. 2002. Chengduhua yu Putonghua erhuayun zhi bijiao [The comparison of er-suffixation in Chengdu dialect and Standard Chinese]. *Journal of Chengdu Teachers' College* 1. 77–81.
- Yue, Yang & Fang Hu. 2019. Phonetics and phonology of the er-suffix in the Hangzhou Wu Chinese dialect. In *Proceedings of 19th international congress of phonetic sciences, Melbourne, Australia, 2056–2060*. Canberra: Australasian Speech Science and Technology Association Inc.
- Zheng, Youyi. 1987. Beijinghua he Chengduhua, Chongqinghua de erhua bijiao [The comparison of er-suffixation in Beijing dialect, Chengdu dialect and Chongqing dialect]. *Journal of Chongqing Normal University* 2. 67–71.
- Zimmermann, Eva. 2013. Vowel deletion as mora usurpation: The case of Yine. *Phonology* 30(1). 125–163.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/phon-2022-0036>).