RIME CHANGES AND MERGERS IN BEIJING RETROFLEX SUFFIXATION: AN ACOUSTIC STUDY

Yutong Wang[1], Mitko Sabev[2]

[1]University of Cambridge, UK, [2]Saarland University, Germany
yw590@cantab.ac.uk, msabev@lst.uni-saarland.de

ABSTRACT

While Mandarin retroflex suffixation and resultant rime mergers have been approached from phonological and impressionistic phonetic angles, empirical studies remain scarce and inconclusive. This investigation is a step towards filling that gap. 10 Beijing Mandarin speakers were recorded and MANOVAs were performed for the effects of suffixation and rime identity on F_1 and F_2 frequencies of nuclear vowels. Results are reported for rimes that were previously claimed to merge: those with high, mid unrounded, and low nuclei. Overall, suffixed nuclei are retracted, while high vowels also tend to be lowered and low vowels to be raised. Our findings confirm complete acoustic neutralisation of the nuclear [a] rimes, where monophthongal rimes merge with those ending in [i] and [n]. On the other hand, suffixed [ie, y, u, ei] remain acoustically distinct within their merger groups. The differences, however, are small and likely to be perceptually undetectable, pointing to incomplete neutralisation.

Keywords: Beijing Mandarin, retroflex suffixation, vowel merger, incomplete neutralisation

1. INTRODUCTION

A body of phonological literature is devoted to Beijing retroflex suffixation (BRS) and the intricate rime mergers it leads to [1–4]. Generally, the retroflex suffix, [ɚ] or [ɻ], is directly attached to simple monophthongal rimes, and [ə] is inserted after front vowels [i, y, e]. In complex rimes (diphthongs and rimes with nasal codas), the second element of the rime is deleted if it is front or alveolar ([i] or [n]). If it is back (rounded) or velar ([u] or [ŋ]), it is not deleted completely, but leaves a trace of rounding or nasalisation, respectively. Within feature models such as [5, 6], this would mean that coronal postnuclear segments are completely deleted, while dorsal ones are not.

There have been few empirical studies of the acoustic properties of nuclear vowels in suffixed vs unsuffixed rimes. [7] first visualised and compared the dynamic trajectories of F_1 and F_2 frequencies in different rimes, and their perceptual discrimination. The study confirmed that rimes in triplets such as [a - ai - an] merged after suffixation, and also found that suffixed [i] and [ie] did not merge while [ɤ] and [ei] did. However, the stimuli did not cover the full range of rimes, as only highly frequent suffixable lexemes were selected. Moreover, the tokens carried different tones, the effects of which on BRS realisation were not considered. [8]'s acoustic and articulatory study reached the same conclusion on the merger of triplets. However, here the results suggested that the front [e] in [ie] and [ye] did not contribute to maintaining contrast with the simple rimes [i] and [y], respectively, as it was dropped in suffixation. What is also different from [7]'s earlier findings is that [8] reports that [ɤ] did not merge with [ei] and [ən].

[9] postulated a 'Simultaneity of Compatible Articulations' principle which eliminates all segments incompatible with [+retroflex], although some features can be preserved. The loss of certain features, together with rime coalescence, results in various cases of rime merger, as in the triplet [a - ai - an], which in suffixation retain only the nuclear vowel of the rime. The phonological literature does not as a rule discuss the degree of overlap between different suffixed rimes, but rather assumes complete neutralisation and addresses issues such as whether suffixation results in overall rime rhotacisation or rhotic attachment with an intervening [ə].

In this paper, we report the results of a production experiment and address spectral changes in suffixed rimes, as well as their resultant acoustic neutralisation, from a quantitative perspective. Here we focus only on rimes that potentially merge with other rimes, and leave out of the discussion [uo] and rimes ending in [ŋ, u], which remain contrastive.

2. METHODS

2.1. Participants, procedure and recording

Ten participants (three male and seven female, aged 19–26), volunteered to participate in the production task in return for nominal payment. The test items were displayed to the participants in carrier sentences on a computer screen, one pair at a time, in Chinese characters with a parallel *pinyin* transcription, in a quasi-random order. A MixPre-6M multitrack recorder with a Sennheiser ME64/K6 condenser microphone was used, recordings were digitised at a sampling rate of 48,000 Hz and stored as uncompressed single-channel WAV files.

2.2. Test items

Beijing Mandarin phonology allows 34 types of rime (excluding the suffix itself, [ɚ]). In a pilot production test, a native speaker (aged 23) produced all possible onset-rime combinations (398 in total), which were analysed acoustically to facilitate the selection of test items to be used in the actual experiment. To control for lexical tone, all selected items were in tone 55, which is a high level tone.

The test items consisted of 47 actual monosyllabic words and 89 phonologically well-formed monosyllabic non-words. The test items were shown in one of two alternative carrier sentences: (a) [wo^{214} tʂʐ̩^{55}tau ____ ʂʐ̩^{51} ʂən^{35}mɤ i^{51}ʂʐ̩] 'I know what ____ means', or (b) [wo^{214} pu^{51} tʂʐ̩^{55}tau ____ ʂʐ̩^{51} ʂən^{35}mɤ i^{51}ʂʐ̩] 'I don't know what ____ means'. Actual monosyllabic words were embedded in sentence (a), and non-words in sentence (b). To stimulate natural production, a question was visually presented, which was to be answered with sentence (a) or (b): 'Do you know what ____ means?'. The subjects were encouraged to speak in a casual style.

2.3. Analysis

Target word rimes were manually segmented in [10] based on the synchronised wideband spectrogram (Gaussian window shape, 0.004 s window length, 0.0001 s time step), waveform and auditory inspection; boundaries were determined by the presence of clear formant structure and a sharp change in intensity. In total, 2960 rime tokens were analysed. For each vowel, the first two formant frequencies were measured at the point of maximum F_1 frequency, using a Praat script [11].

Formant frequencies were then normalised using Lobanov's formant-intrinsic, vowel-extrinsic, speaker-intrinsic method [12]. Outliers, defined as values outside IQR by 1.5 times IQR, were removed (6.89% for F_1', 5.71% for F_2').

Vowel space areas and centroids in Figure 1 were computed using R package [13]. MANOVAs were performed on each unsuffixed–suffixed pair to measure the effect of suffixation on vowel quality. A second series of MANOVAs was used to quantify contrastiveness in potentially merging rime groups. In groups where significant differences emerged, additional comparisons were computed on subsets of the rimes, with Bonferroni correction for multiple testing applied to the significance level α. Pillai's trace (PT) was used as a metric of overall spectral difference. It is a common MANOVA statistic that has been used to quantify vowel merger in a growing body of work [14–24]. PT ranges from 0 to 1, with higher values indicating greater separation, and lower values greater overlap. Symmetric differences between the probability density functions (PDF) of each formant frequency ($\Delta F_1', \Delta F_2'$) in suffixed vs unsuffixed rimes were computed using R package [25].

3. RESULTS

An initial MANOVA was performed on all tokens with the normalised F_1', F_2' frequencies as response variables, and rime, suffixation, carrier sentence type (a or b), and speaker as predictor variables. Rime and suffixation both had a significant effect ($p < 0.0001$), while sentence type ($p = 0.7366$) and speaker ($p = 1$) did not, indicating that the use of alternative carrier sentences did not affect the results, and that the normalisation was highly effective.

Figure 1 shows the mean formant frequencies of nuclear vowels in the rimes under investigation. The area of the vowel space, defined as the area of the convex hulls enclosing the nuclear vowels in the $F_1' \times F_2'$ frequency space, was found to be 3.35 for unsuffixed and 1.16 for suffixed rimes, meaning that the suffixed space area is just over a third of the area of the unsuffixed space. Suffixation therefore results

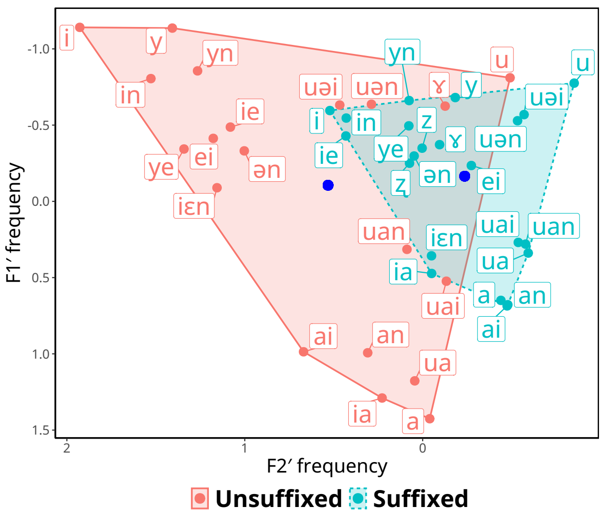

Figure 1: Mean F_1' and F_2' frequencies of rime nuclei; blue dots: vowel space centroids.

in a substantial contraction of the vowel space. The centroids of the two vowel spaces, plotted as blue dots, indicate that suffixation causes overall retraction, which is also observed in individual rimes. In addition, high unsuffixed nuclei are lowered, while low nuclei are raised with suffixation.

Table 1 reports the results of MANOVA comparisons, as well as symmetric differences between PDFs of the F_1' and F_2' frequencies of suffixed and unsuffixed nuclei. MANOVAs were significant for all nuclei except [u]. The overall change is small in [ɤ, ie, uai], with Pillai's trace values under 0.30. Moderate change is found in [uəi, uən, a, ia, ua], where PT ranges between 0.30 and 0.59. The remaining eleven nuclei all exhibit strong change, with PT scores above 0.60. Within the high front nuclei, F_1' frequency increases while F_2' decreases, thus resulting in mid-centralisation; for [yn] the change in F_2' is considerably stronger. In all mid nuclei except [ɤ] there is substantial decrease in F_2' (less pronounced in [ie]), whereas changes in F_1' are smaller. The monophthong [ɤ] shows appreciable increase in F_1' and little change in F_2' frequency. The low nuclei undergo reduction in both formant frequencies, with a stronger change in F_2' in most cases (not in [a, ia]).

The merger patterns of suffixed rimes are displayed as two-dimensional $F_1' \times F_2'$ PDFs in Figures 2 and 3. The results of MANOVAs comparing the nuclei of merging rimes are given in Table 2.

Rime	p	PT	$\Delta F_1'$	$\Delta F_2'$
i	0.0000	0.78	0.91	0.74
in	0.0000	0.68	0.64	0.77
y	0.0000	0.88	0.71	0.96
yn	0.0000	0.68	0.32	0.75
u	0.6553			
ɤ	0.0000	0.26	0.44	0.07
ie	0.0002	0.22	0.18	0.42
ye	0.0000	0.62	0.16	0.76
ei	0.0000	0.74	0.28	0.85
ən	0.0000	0.70	0.19	0.73
uəi	0.0000	0.33	0.22	0.59
uən	0.0000	0.31	0.20	0.58
a	0.0000	0.36	0.46	0.43
ia	0.0000	0.45	0.56	0.35
ua	0.0000	0.46	0.53	0.62
ai	0.0000	0.75	0.30	0.92
an	0.0000	0.62	0.23	0.76
iɛn	0.0000	0.73	0.39	0.64
uai	0.0013	0.17	0.18	0.50
uan	0.0000	0.69	0.14	0.79

Table 1: MANOVA comparisons of suffixed and unsuffixed nuclei; PT: Pillai's trace; Δ: symmetric difference between PDFs.

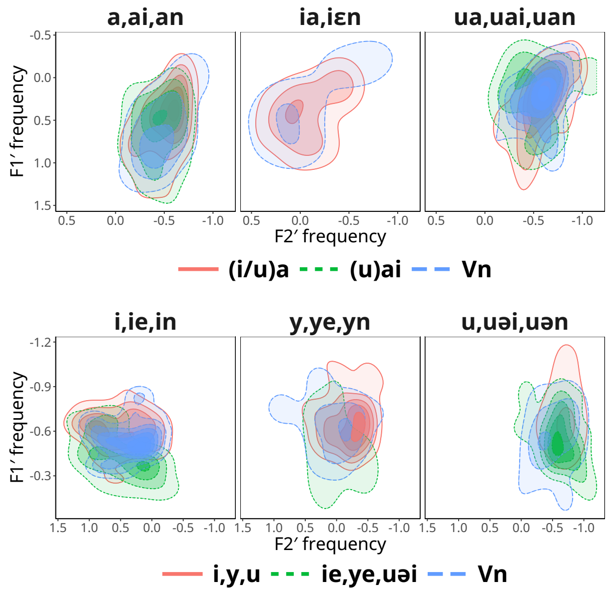

Figure 2: Suffixed rime mergers ($F_1' \times F_2'$ PDFs).

The top panels of Figure 2 show nearly total overlap in the rime groups with nuclear [a], and indeed the MANOVAs (Table 2) identified no significant differences in any of these groups: each of suffixed [a - ai - an], [ia - iɛn], and [ua - uai - uan] are merged. A somewhat lesser degree of overlap is visible in the bottom panels of Figure 2, and the corresponding MANOVAs indicate that one rime in each group is significantly different from the rest: [ie] in [i - ie - in], [y] in [y - ye - yn] and [u] in [u - uəi - uən]. There is strong overlap in the rime group in Figure 3, except that suffixed [ei] spans only about half of the F_2' range for the group. Indeed, the MANOVAs show that only [ei] is significantly distinct in this group.

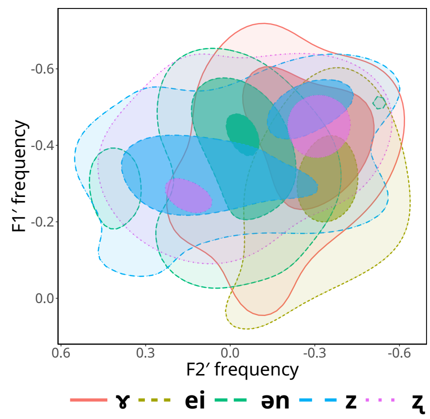

Figure 3: Suffixed [ɤ, ei, ən, z̩, ʐ̩] (PDFs).

Comparison	p	$\gtrless$	α	PT
a, ai, an	0.8982	>	0.050	
ia, iɛn	0.2684	>	0.050	
ua, uai, uan	0.6489	>	0.050	
i, ie, in	0.0000	<	0.050	0.23
i, in	0.1410	>	0.025	
ie, {i+in}	0.0000	<	0.025	0.21
y, ye, yn	0.0009	<	0.050	0.17
ye, yn	0.1526	>	0.025	
y, {ye+yn}	0.0009	<	0.025	0.12
u, uəi, uən	0.0001	<	0.050	0.22
uəi, uən	0.6042	>	0.025	
u, {uəi+uən}	0.0000	<	0.025	0.22
ɤ, ei, ən, z̩, ʐ̩	0.0004	<	0.050	0.16
ɤ, ən, z̩, ʐ̩	0.3519	>	0.025	
ei, {ɤ+ən+z̩+ʐ̩}	0.0000	<	0.025	0.12

Table 2: MANOVA comparisons of suffixed rimes; $p \leq \alpha$: significant; PT: Pillai's trace.

All significant differences, however, are small (PT $\leq$ 23) and thus may not be perceptually salient.

4. DISCUSSION AND CONCLUSIONS

Our results demonstrate that Beijing Mandarin rimes undergo a variety of significant spectral changes in retroflex suffixation, and that certain rime groups are completely merged after suffixation, while in other instances neutralisation is incomplete.

In most cases our findings are consistent with earlier phonetic treatments of BRS. In the high front nuclear vowels, viz. [i] and [y], suffixation brings about substantial changes in F_1' and F_2' frequencies in the direction of mid-centralisation. In mid (unrounded) nuclei, on the other hand, spectral changes are considerably smaller. In [uəi, uən], the pre-nuclear glide [u] appears to have an effect on the nuclei, which have lower F_2' frequencies than [əi, ən] in general, while suffixation results in additional retraction. A range of F_2' frequencies are found in the suffixed mid range of the vowel space (cf. [ie, in] vs [uəi, uən] in Figure 1), which refutes the frequent assumption of many phonological accounts that all non-low nuclei merge in an undifferentiated [ə]. In the low nuclei, we find considerable raising in most rimes, as well as comparable degrees of retraction. Pillai's trace for suffixed vs unsuffixed [uai] is exceptionally small for a low nucleus (0.18). Overall, the comparison of unsuffixed vs suffixed rimes reveals a pattern of global retraction, as well as moderate lowering and raising, respectively, for high and low nuclei. Suffixation also leads to dramatic shrinkage of the spectral vowel space, in terms of both formant frequencies.

As regards the neutralisation of contrast between rimes, our findings confirm complete merger in the rime groups with low nuclei, as MANOVAs yielded non-significant results.

The rime groups [i - ie - in], [y - ye - yn] and [u - uəi - uən] also tend to merge, although there is one rime in each which remains statistically distinct from the rest, namely [ie, y, u]. [ie] has a slightly higher F_1' frequency than the rest of its group when suffixed, while [y] and [u] have lower F_2' frequencies than the rest, which may be caused by greater retraction or rounding; suffixed [u] also has lower F_1' frequency than [uəi, uən]. These findings are at odds with the earlier claim that these groups are also fully merged [8]. Empirical studies such as [7, 26] found [ie, ye] to be distinct, while our results corroborate this only for the former rime.

A long-standing dispute concerning suffixed mid unrounded nuclei is whether [ɤ] merges with the rest of the group, [ei - ən - z̩ - ʐ̩]. Our results indicate that it does, which is consistent with [7]. We also found that, in this group, suffixed [ei] remains distinct from the rest, which is at variance with what [8] reported.

It should be noted, however, that since Pillai's traces for the rimes that were identified as distinct within their groups are relatively low, ranging from 0.12 to 0.22, these contrasts are likely to be perceptually undetectable: in all likelihood, we are dealing with cases of incomplete neutralisation, where statistically significant acoustic differences are not large enough to be perceptible.

Participants in this production experiment were asked to read out items in both their strongly contrastive (unsuffixed) and potentially merged (suffixed) forms, which may lead one to speculate whether the elicitation design could have given rise to untypically hyperarticulated renditions of the suffixed rimes. Such hyperarticulation, however, appears to be unlikely, for only four of the rimes emerged as significantly different in their groups, while the majority were clearly merged.

While the experimental results reported here contribute to closing a number of gaps in the literature on Beijing retroflex suffixation, many aspects of the process remain to be examined in the future. These include the effect of lexical tone, utterance-level prosody (e.g. focus, position in a phrase), word frequency, gender, speaking style, interpersonal variation, as well as how acoustic results are related to perception. [7], for example, pointed out that suffixed [i] and [ie] are more accurately perceptually distinguished in some tones than in others. The reasons for such interactions between lexical tone and rime merger remain largely understudied [4].

5. ACKNOWLEDGEMENTS

We are grateful to Beeke Muhlack for writing the *F1-max* Praat script [11], and to our two anonymous reviewers for helpful questions and suggestions.

6. REFERENCES

[1] Y. Lin, "Autosegmental treatment of segmental processes in Chinese phonology," Ph.D. dissertation, University of Texas at Austin, 1989.
[2] S. Duanmu, "A formal study of syllables, tones, stress and domain in Chinese language," Ph.D. dissertation, MIT, 1990.
[3] ——, *The Phonology of Standard Chinese (Phonology of the world's languages)*, 1st ed. Oxford: Oxford University Press, 2000.
[4] ——, *The Phonology of Standard Chinese (Phonology of the world's languages).*, 2nd ed. Oxford: Oxford University Press, 2007.
[5] G. N. Clements and E. V. Hume, "The Internal Organization of Speech Sounds," in *The Handbook of Phonological Theory*, J. A. Goldsmith, Ed. Oxford: Blackwell, 1995, pp. 245–306.
[6] A. Lahiri and H. Reetz, "Distinctive features: Phonological underspecification in representation and processing," *Journal of Phonetics*, vol. 38, pp. 44–59, 2010.
[7] L. Wang and N. He, "*Beijing Erhuayun de Tingbian Shiyan he Shengxue Fenxi* (the perception experiments and acoustic analyses on Beijing erhua)," in *Beijing Yuyin Shiyanlu (The Handbook of Beijing Phonetic Experiments)*, T. Lin and L. Wang, Eds. Beijing: Peking University Press, 1985, pp. 27–72.
[8] W. Lee, "A phonetic study of the erhua rimes in Beijing Mandarin," in *Proceedings of Eurospeech, 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2005, pp. 1093–1096.
[9] Y. Chao, *A Grammar of Spoken Chinese*. Berkeley, Calif.; London: University of California Press, 1986.
[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," http://www.praat.org/ Computerprogramversion6.0., 2015.
[11] B. Muhlack, "F1-max," Praat script, 2022.
[12] B. Lobanov, "Classifications of Russian vowels spoken by different listeners," *Journal of the Acoustical Society of American*, vol. 49, pp. 606–608, 1971.
[13] R. Hijmans, C. Karney, E. Williams, and C. Vennes, "*geosphere*: Spherical Trigonometry," https://CRAN.R-project.org/package=geosphere, 2022.
[14] J. Hay, P. Warren, and K. Drager, "Factors influencing speech perception in the context of a merger-in-progress," *Journal of Phonetics*, vol. 34, pp. 458–484, 2006.
[15] L. Hall-Lew, "Improved representation of variance in measures of vowel merger," *Proceedings of Meetings on Acoustics*, vol. 9, no. 1, pp. 1–10, 2010.
[16] M. Sloos, "The reversal of the Bären–Beeren merger in Austrian Standard German," *Mental Lexicon*, vol. 8, no. 3, pp. 353–371, 2013.
[17] J. Nycz and L. Hall-Lew, "Best practices in measuring vowel merger," *Proceedings of Meetings on Acoustics: The Acoustical Society of America through the American Institute of Physics*, vol. 20, no. 1, pp. 1–19, 2014.
[18] M. Amengual and P. Chamorro, "The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts," *Phonetica*, vol. 72, no. 4, pp. 207–236, 2015.
[19] M. E. L. Renwick and R. M. Olsen, "Analyzing dialect variation in historical speech corpora," *Journal of the Acoustical Society of America*, vol. 142, no. 1, pp. 406–421, 2017.
[20] M. Sabev and E. Payne, "A cross-varietal continuum of unstressed vowel reduction: evidence from Bulgarian and Turkish," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Canberra, Australia: Australasian Speech Science and Technology Association Inc, 2019, pp. 1164–1168.
[21] C. Diskin, D. Loakes, R. Billington, H. Stoakes, S. Gonzalez, and S. Kirkham, "The /el–æl/ merger in Australian English: acoustic and articulatory insights," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Canberra, Australia: Australasian Speech Science and Technology Association Inc, 2019, pp. 1764–1768.
[22] M. Sabev, "Spectral and Durational Unstressed Vowel Reduction: an acoustic study of monolingual and bilingual speakers of Bulgarian and Turkish," Ph.D. dissertation, University of Oxford, 2020.
[23] ——, "Unstressed vowel reduction and contrast neutralisation in Western and Eastern Bulgarian: a current appraisal," *Journal of Phonetics*, forthcoming.
[24] J. A. Stanley and B. Sneller, "Sample size matters in calculating Pillai scores," *Journal of the Acoustical Society of America*, vol. 153, no. 1, pp. 54–67, 2023.
[25] M. Pastore, "Overlapping: a R package for estimating overlapping in empirical distributions," *The Journal of Open Source Software*, vol. 3(32), p. 1023, 2018.
[26] Y. Wang, "Contrast neutralisation in Beijing retroflex suffixation: an acoustic study," MPhil dissertation, University of Cambridge, 2022.