UCLA
UCLA Electronic Theses and Dissertations

Title
Choosing rhotacization site in Beijing Mandarin: The role of perceptual similarity

Permalink
https://escholarship.org/uc/item/2tb7j7fj

Author
Xu, Zhongshi

Publication Date
2020

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Choosing rhotacization site in Beijing Mandarin: The role of perceptual similarity

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Arts in Linguistics

by

Zhongshi Xu

2020

© Copyright by

Zhongshi Xu

2020

ABSTRACT OF THE THESIS

Choosing rhotacization site in Beijing Mandarin: The role of perceptual similarity

by

Zhongshi Xu

Master of Arts in Linguistics

University of California, Los Angeles, 2020

Professor Bruce Hayes, Co-Chair

Professor Kie Zuraw, Co-Chair

The principle of faithfulness, proposed in the Optimality Theoretic framework of phonology, has traditionally been based on binary distinctive features and discrete sound correspondences within an input–output pair. The subsequent proposal of the P-map has inspired a different approach to faithfulness—one that allows phonological grammars to evaluate faithfulness directly using the phonetic distance between different continuous speech streams, maximally preserving the subtle phonetic difference among output candidates. This paper presents a study that aims to determine whether the distance-based approach to faithfulness can better account for gradient alternation patterns than the traditional feature-based approach can. The phenomenon this study examines is rime rhotacization in Beijing Mandarin. Results of an experiment where participants were asked to choose which rime to rhotacize in nonce disyllables reveal that speakers choose to rhotacize the rime which yields the more faithful output. The results were modeled with mixed-effects logistic regression. One model incorporated feature-based faithfulness constraints and the other distance-based ones. The models confirmed that the faithfulness of rhotacization candidates is the main deciding factor. However, two independent model comparison measures yielded contradictory results regarding which model performed

better, leaving an inclusion in regard to whether distance-based faithfulness is more capable than feature-based faithfulness.

The thesis of Zhongshi Xu is approved.

Claire Moore-Cantwell

Bruce Hayes, Committee Co-Chair

Kie Zuraw, Committee Co-Chair

University of California, Los Angeles

2020

Table of Contents

List of Figures

List of Tables

1 Introduction

The Optimality Theory (OT) framework of phonology, put forward by Prince & Smolensky 1993, proposes that phonological grammars are governed by two independent principles, i.e., markedness and faithfulness, which respectively regulate the outputs and the input–output mappings. Specifically, markedness constraints encode preferences for certain surface strings regardless of the input, whereas faithfulness constraints encode dispreferences for specific deviations from the input. The independent status of these two principles allows phonologists to explain separately what drives an alternation and why a particular alternation pattern occurs.

Traditionally, as outlined in McCarthy & Prince 1995, faithfulness constraints operate on correspondences between discrete, abstract sound categories that are defined by a finite number of binary features, such as [±voice] and [±nasal]. As a result, the grammar can only make categorical distinctions among different mappings. For instance, the mappings [u] → [ʉ] and [ũ] → [ʉ̃] would be treated by the grammar as equally faithful, since both mappings involve the same featural change in [±back]; on the other hand, the mappings [i] → [y] and [i] → [e] would be categorically different and in fact incomparable in terms of faithfulness, because one change is along the featural dimension [±round] and the other [±high]. This approach, which I call *feature-based faithfulness*, abstracts away the "non-essential" phonetic detail that would be present in output speech streams and instead relies solely on binary featural contrasts between an input and the output.

While the feature-based approach has been the industry standard for OT analyses and has been proven sufficient in accounting for mostly categorical alternation patterns, studies of variable phonological processes have often found that sometimes the grammar does care about the more gradient phonetic differences between inputs and outputs. This had led to the P-map proposal (P for *perceptibility*; Steriade 2001), which recognizes that the same featural change corresponds to differing degrees of perceptual similarity in different contexts. In the case of [u] → [ʉ] vs. [ũ] → [ʉ̃], a P-map analysis would treat the former mapping as less faithful than the latter, due to the oral vowel pair having a perceivably longer F2 distance than the nasalized vowel pair; and in the case of [i] → [y] vs. [i] → [e], they would also fall on different positions on the same perceptual similarity scale, and thus becoming directly comparable to each other as well as to all other potential sound mappings. One way of implementing the P-map, therefore, is to allow phonological grammars to directly use the raw phonetic distances to evaluate the faithfulness of input–output mappings. This approach, which I call *distance-based faithfulness*, can be seen adopted in various recent analyses of gradient alternation patterns, where it is often found that the distance-based approach can account for the data better than the feature-based one can (see McCollum 2018 for such an example).

In this paper, I present a study that tests the same hypothesis, i.e., a grammar model with distance-based faithfulness constraints can explain gradient alternation patterns better than one with feature-based ones can. This study examines Beijing Mandarin rime rhotacization in disyllabic words, where the choice of which rime to rhotacize seemingly bears on how faithful the rhotacized rime is to the original non-rhotic rime, and addresses the following three research questions: (1) do native speakers of Beijing Mandarin rely on phonological principles when choosing the rhotacization site in disyllabic words—*yes*; (2) is the principle of faithfulness the main deciding factor—*yes*; (3) does distance-based faithfulness yield a better model of the grammar—*unclear*. The organization of the paper is as follows. In §2, I introduce the rime rhotacization process in Beijing Mandarin. In §3, I outline the methods of the current study. In §4, I present and discuss two models of the survey results, one using feature-based faithfulness constraints and the other distance-based ones. In §5, I discuss the study outcome and conclude the paper.

2 Background

The term Beijing Mandarin refers to the dialect of Mandarin spoken by *Beijingers*, natives of the capital city of China. The definition of the *Beijinger* population is not straightforward—it does not simply refer to those who were born and/or have lived in Beijing; rather, it is a social identity associated with the historically-grassroot culture of urban Beijing, and shared by Beijing habitants whose family, or who themselves, settled in the urban area no latter than the mid-20th century.

Beijing Mandarin was chosen as the basis for the phonology of Standard Chinese, or *Putonghua* (普通话; literally 'common speech'), at the Symposium on the Standardization of Modern Chinese in 1955, which also saw the designation of Standard Chinese as the official language of People's Republic of China. The symposium chose to base the pronunciation of the standard language on the Beijing dialect, but also decided to exclude some of its phonological properties, such as the prevalent use of the neutral tone. Another such property that distinguishes Beijing Mandarin from Standard Chinese is the use of *erhua* (儿化 [ɚ˧˥.xua˥˩]; literally 'rhotacization') rimes in speech. In this section, I discuss the syllable rime inventory of Beijing Mandarin (§2.1), introduce the rime rhotacization process (§2.2), and examine the variation of rhotacization site in multisyllabic words (§2.3).

2.1 Syllable rimes of Beijing Mandarin

The syllable structure of Beijing Mandarin is maximally CVN, where C represents an optional consonant onset, V a mandatory vocalic nucleus, and N an optional nasal coda (Wu & Kenstowicz 2015). The rime consists of the nucleus and the coda, if it is present. In Beijing Mandarin, the nucleus can be a monophthong, a diphthong, or a triphthong, but only monophthongs and diphthongs are attested in rimes with a coda; the nasal

nucleus:	mono-	di-	tri-	mono-	di- phthong
coda:				✓	✓
	[i ɿ ʅ]	[ie]	[iou]	[in]	[iuŋ]
	[y]	[ia]	[iau]	[iŋ]	[ien]
	[u]	[ye]	[uei]	[yn]	[iaŋ]
	[ɤ]	[uo]	[uai]	[uŋ]	[yen]
	[ɚ]	[ua]		[ən]	[uən]
	[a]	[ei]		[əŋ]	[uan]
		[ou]		[an]	[uaŋ]
		[ai]		[aŋ]	
		[au]			

Table 2.1: Attested syllable rimes in Beijing Mandarin (Lee & Zee 2014).

coda can be either [n] or [ŋ]. In total, there are 34 attested syllable rimes that contrast. Among them, the unrounded high front monophthong rime, [i], also has the allophones [ɿ] and [ʅ], which occur respectively after alveolar and retroflex sibilant onsets (e.g., as in 丝 [sɿ˥] 'silk' and 诗 [ʂʅ˥] 'poem'). Table 2.1 presents the phonetic transcriptions for all 36 distinct rimes on the surface, as provided by Lee & Zee 2014. Among the diphthongs, [ei ou ai au] contain offglides, where the first vowel element is more prominent; [ie ia ye uo ua iu] contain onglides, with more prominence on the second vowel element. In all four triphthongs, the most prominent vowel element is the middle one.

Phonologically, the diphthongs and triphthongs of Beijing Mandarin are often analyzed as sequences of simple vowels (V) and glides (G). Specifically, offglide diphthongs are VG, onglides GV, and triphthongs GVG. One hotly debated question that arose under this analysis is the syllabic affiliation of prenuclear glides—namely, whether they belong to the rime or the onset. Duanmu 2007 argues that prenuclear glides are realized as secondary articulation on the onset, and therefore not affiliated with the rime. However, the opposite view was offered empirical support by Wu & Kenstowicz 2015. In the experiment, they recorded five native speakers reading a corpus of monosyllabic words of Beijing Mandarin. The corpus was balanced for syllable shape (CV, CVN, CGV, or CGVN), tone, and segment. They measured the duration of entire syllables as well as the individual segments. The results show that the nucleus, V, is proportionally longer in CV syllables than in CVN and CGV, equally long in CVN and CGV, and the shortest in CGVN. The duration hierarchy (CV > CVN = CGV > CGVN) shows that the prenuclear glide influences the duration of the nucleus vowel in the same way as the nasal coda. This suggests that prenuclear glides are affiliated with the rime.

2.2 Rime rhotacization in Beijing Mandarin

The Chinese character 儿 (*er*) represents two distinct morphemes—one is a lexical morpheme meaning 'son' and pronounced [ɚ˧˥]; the other is a stylistic marker of the Beijing dialect, often referred to as the [ɚ] suffix. Historically, the stylistic *er* is derived from the lexical morpheme, likely with an intermediate stage where it acted as a productive diminutive suffix for nouns (Duanmu 2007). However, the [ɚ] suffix has arguably lost its diminutive semantics in the modern language, where it can be found attached to adjectives and verbs alike. Thus, the denotation of the [ɚ] suffix is now purely stylistic: it signals to readers that the preceding morpheme is intended to be pronounced "in the style of Beijing Mandarin"—specifically, with rhotacization of the syllable rime. Thus, even though the [ɚ] suffix is represented in orthography as the standalone character 儿, phonologically there isn't always linearly segmentable representation for it in the speech stream; instead, it is often realized simultaneously with the "suffixed" morpheme, as rhotic coloring effects imposed on that morpheme's syllable rime. For example, 肚 [tu˨˩˦] 'tripe' plus the stylistic marker *er* is pronounced [tu˞˨˩˦], a monosyllable changing the original non-rhotic rime of [tu] to its rhotic version, [u˞]. The resulting rhotic rimes are referred to as *erhua* rimes in Chinese.

All attested rimes of Beijing Mandarin also have an *erhua* counterpart attested (except for [ɚ], which is already rhotic by itself), yielding 35 plain–*erhua* rime mappings. The mappings are categorical, i.e., for each plain rime, there is only one attested *erhua* outcome

[i ie in]	→	[iɚ]	[iŋ]	→	[ĩɚ̃]
[ien]	→	[iəɚ]			
[y ye yn]	→	[yɚ]			
[yen]	→	[yəɚ]			
[ɹ̩]	→	[ɹ̩ɚ]			
[ɻ̩]	→	[ɻ̩ɚ]			
[a ai an]	→	[aɚ]	[aŋ]	→	[ãɚ̃]
[ɤ]	→	[ɤ˞]			
[u]	→	[u˞]	[uŋ]	→	[ũ˞]
			[iuŋ]	→	[ĩũ˞]
[ei ən]	→	[əɚ]	[əŋ]	→	[ə̃ɚ̃]
[uei uən]	→	[uəɚ]			
[ia]	→	[ia˞ɚ]	[iaŋ]	→	[ĩã˞ɚ̃]
[ua uai uan]	→	[ua˞ɚ]	[uaŋ]	→	[ũã˞ɚ̃]
[au]	→	[au˞]			
[ou]	→	[ou˞]			
[uo]	→	[uo˞]			
[iau]	→	[ia˞u˞]			
[iou]	→	[io˞u˞]			

Table 2.2: Plain–*erhua* rime mappings of Beijing Mandarin (Lee 2005).

across speakers as well as lexical items. However, in terms of acoustics, different plain rimes exhibit different rhotic coloring effects when they are rhotacized. Using speech data provided by three native speakers, Lee 2005 compares formant tracks of each plain–*erhua* rime pair and outlines the various acoustic manifestations of the rhotacization process. Her resulting transcriptions for the *erhua* rimes are shown in Table 2.2, consolidated by *erhua* forms. According to Lee, for monophthong rimes, the back nuclei [u] and [ɤ] become rhotacized monophthongs [u˞] and [ɤ˞], whereas the non-back nuclei [i y a ɹ̩ ɻ̩] diphthongize into [Vɚ]. For complex rimes, the following processes are observed:

(1) coda [n] and final front vowel elements are substituted with [ɚ];

(2) non-final [e] centralizes to [ə];

(3) non-initial [u], [o], and [a] become rhotacized;

(4) [ɚ] is added after final [a];

(5) coda [ŋ] is substituted with [ɚ̃] after non-back vowel elements;

(6) after [u], coda [ŋ] disappears and [u] becomes rhotacized;

(7) rimes originally containing a coda [ŋ] become nasalized entirely.

All applicable processes apply in a given rime. For instance, in the mapping [iaŋ] → [ĩã˞ɚ̃], all of processes (3), (5), and (7) apply. Several analyses have been proposed to account for the categorical plain–*erhua* rime mappings, both in the rule-based framework (Duanmu 1990, 2007) and in the OT framework (Ma 1997, Zhang 2000, Tian 2009).

2.3 Rhotacization site variation in multisyllabic words

While the plain–*erhua* rime mappings are categorical, there is between-word variation in the native lexicon regarding which rime in a multisyllabic word becomes rhotacized. For example, Table 2.3 shows two pairs of disyllabic, dimorphemic compounds[1], each pair sharing the same second morpheme; yet, within each pair, the rhotacization site is not uniformly on the first or the second morpheme. No previous study that I am aware of has examined the variation of rhotacization site in multisyllabic words of Beijing Mandarin.

瓷	[t͡sʰɹ̩˧˥]	'porcelain'	+	砖	[ʈ͡ʂuan˥]	'brick'	→	瓷砖儿	[t͡sʰɹ̩˧˥.ʈ͡ʂua˞ɚ˥]	'tile'
板	[pan˨˩˦]	'plank'	+	砖	[ʈ͡ʂuan˥]	'brick'	→	板儿砖	[paɚ˨˩˦.ʈ͡ʂuan˥]	'brick '
油	[iou˧˥]	'oil'	+	饼	[piŋ˨˩˦]	'pastry'	→	油饼儿	[iou˧˥.pĩɚ̃˨˩˦]	'fried pancake'
馅	[ɕien˥˩]	'filling'	+	饼	[piŋ˨˩˦]	'pastry'	→	馅儿饼	[ɕiəɚ˥˩.piŋ˨˩˦]	'filled pancake'

Table 2.3: Rhotacization site variation in Beijing Mandarin disyllabic words.

An important aspect of acquiring the Beijing Mandarin lexicon is to know which words ought to be produced with *erhua* realization and in multisyllabic *erhua* words which rime should be rhotacized, meaning that the variation of rhotacization site in the native lexicon is strictly type variation—for each given multisyllabic *erhua* word, the site of rhotacization is encoded in the lexical entry, thus produced with no speaker or token variation. Therefore, one possible analysis for the phenomenon is lexical allomorph specification, where the

[1]The vast majority of morphemes in Chinese languages are monosyllabic, and therefore in multisyllabic words each syllable usually corresponds to an individual morpheme.

rhotacization process is not applied online, and in multisyllabic, multimorphemic words, the choice of rhotacization site is simply down to which component morpheme has its *erhua* allomorph encoded in the lexical entry of the word.

However, there are indications of productivity of rime rhotacization in the modern language, particularly in recent loanwords that were adapted into Standard Chinese as multisyllabic monomorphemes. One such example is 兰博基尼 [lan˧˥.puo˧˥.t͡ɕi˥.ni˧˥], from *Lamborghini*, name of the famed Italian car maker. Each of the four characters in the adaption is an individual monosyllabic morpheme—兰 [lan˧˥] 'orchid; blue,' 博 [puo˧˥] 'vast,' 基 [t͡ɕi˥] 'basic,' and 尼 [ni˧˥] 'Buddhist nun'—but the denotation 'Lamborghini' has arguably nothing to do with orchid, blueness, vastness, base, or Buddhist nuns, suggesting that the tetrasyallic loanword is indeed monomorphemic rather than composed from those four individual morphemes. When I polled eight native Beijing Mandarin speakers on which syllable rime they would choose to rhotacize 兰博基尼 (asking them which of the four possible *erhua* realizations sounds more natural), they unanimously agreed on the preferred rhotacization site—[uo], rime of the second syllable. This suggests that native speakers are able to productively apply *erhua* to words that are not lexically prescribed so, and that in such words they have principled intuitions regarding which rime should be rhotacized. And since the example above is monomorphemic, the *erhua* allomorph of 博 'vast' could not have been encoded in the lexical entry of the whole word, further refuting the allomorph specification analysis.

The fact that Beijing Mandarin speakers have shared intuition for which syllable rime in a multisyllabic monomorpheme produces the most natural *erhua* output suggests that the choice of rhotacization site in such words is governed by phonological principles, which answers *yes* to research question (1) of the current study.

3 Methods

To test which phonological factors can determine the rhotacization site in Beijing Mandarin multisyllabic words, I conducted a survey where participants were asked to choose whether rhotacizing the first or the second rime in a disyllabic nonce word yields the most natural *erhua* output.

3.1 Survey stimuli

The survey included 26 disyllabic nonce words as test items, constructed with 13 pairs of syllables and by arranging the two syllables of each pair in both orders. All the syllables that were used to construct the items are gaps in the Beijing Mandarin syllable inventory. For instance, both the onset [f] and the rime [ie] are attested individually in Beijing Mandarin, but their combination, [fie], is unattested as a syllable in the language and therefore a gap in the inventory. In other words, the syllables in the test items are unattested combinations of real Beijing Mandarin onsets and rimes. There are 13 unique rimes among the 13 pairs of syllables that were used to construct the items. The rimes are listed in Table 3.1. The onsets vary between [pʰ tʰ kʰ p m f]. In each item, the tone on the two syllables was kept the same, and is either ˧˥ or ˥˩. The list of stimuli is included in Appendix A.

[i]	(→ [iɚ])	[ia]	(→ [ia˞ɚ])	[uŋ]	(→ [ũ˞])
[ie]	(→ [iɚ])	[uai]	(→ [ua˞ɚ])	[iuŋ]	(→ [ĩũ˞])
[ei]	(→ [əɚ])	[uan]	(→ [ua˞ɚ])	[ou]	(→ [ou˞])
[ən]	(→ [əɚ])	[uaŋ]	(→ [ũã˞ɚ̃])	[iou]	(→ [io˞u˞])
[ien]	(→ [iəɚ])				

Table 3.1: Unique rimes among the survey stimuli.

The purpose of using unattested syllables in constructing the test items is to avoid the potential confound of rhotacized syllable frequencies in the Beijing Mandarin lexicon.

Studies have repeatedly found that speakers are sensitive to the application rates of variable phonological processes in different lexical items (Zymet 2018 and references therein). Therefore, it is reasonable to assume that when a native Beijing Mandarin speaker is asked to choose the most natural rhotacization site for a word consisting of attested syllables, the speaker may simply choose the most frequently rhotacized syllable as the best site. With unattested syllables, there is no lexical frequency information that participants may rely on, thus eliminating the confound.

3.2 Survey procedure

In the survey, the 26 non-rhotacized stimuli were presented to participants as individual question prompts, and for each item the two possible candidates for *erhua* realization were shown embedded in the frame 我… 了 [uo˨˩˦…lə] 'I …-ed (PFV).' The stimuli, consisting of unattested syllables and without corresponding Chinese characters, were shown in *pinyin*, the standard romanization system for Mandarin languages. In the candidates, the site of rhotacization is indicated with the character 儿, which immediately follows the syllable whose rime would be rhotacized. Table 3.2 shows two presentation examples in the survey. At the start of survey, participants were asked to choose the more natural-sounding *erhua* output for each item.

		presentation
item	[fien˧˥.tʰən˧˥]	fián tén
choice 1	[fiəɚ˧˥.tʰən˧˥]	我 fián 儿 tén 了
choice 2	[fien˧˥.tʰəɚ˧˥]	我 fián tén 儿了
item	[tʰən˥˩.fien˥˩]	tèn fiàn
choice 1	[tʰəɚ˥˩.fien˥˩]	我 tèn 儿 fiàn 了
choice 2	[tʰən˥˩.fiəɚ˥˩]	我 tèn fiàn 儿了

Table 3.2: Item and choice presentations in the survey.

The survey was conducted online with Google Forms. The link to the survey was sent out to invited participants. The invitation criteria include being born and raised in

Beijing and frequently using *erhua* in conversations at home. All invited participants are acquaintances of the author, within the age range of 20–24. The survey received 16 completed responses.

Each individual response to a question prompt (item) was recorded as *first* if the *erhua* output with rhotacization on the first of the two syllable rime was chosen, or *second*, if on the second. For each response, the participant's anonymized identification code was also recorded. The survey produced 416 data points (26 items × 16 participants). A brief inspection of the responses revealed robust by-item variation and the responses are not strongly polarized toward either candidate. Figure 3.1 shows a histogram of the rates of choosing the *first* candidate for each item, across all participants.

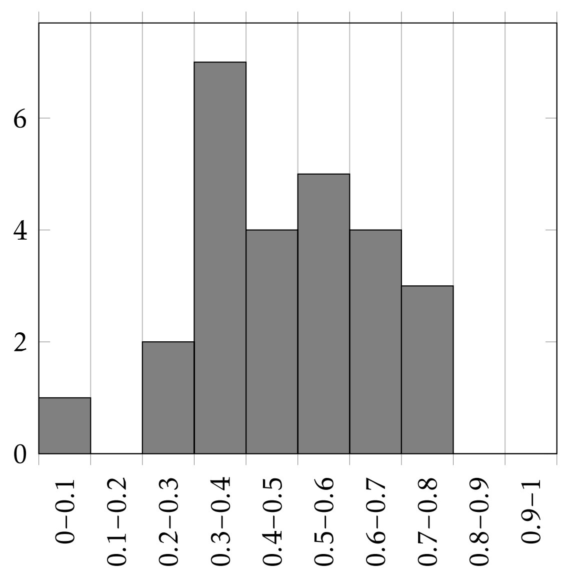

Figure 3.1: Histogram of by-item response rates for the *first* candidate.

3.3 Rime elicitation and phonetic analysis

In order to obtain the phonetic distances between the corresponding plain and *erhua* rimes for the distance-based analysis, both versions of the 13 unique rimes in the survey stimuli were elicited. For each rime, two real Standard Chinese disyllabic words where the second syllable has the target rime were included. The words were presented in Chinese characters. Each word appeared twice, once with the rhotacization suffix 儿 at the end, indicating rhotacization on the last rime, and once without. The list of elicitation items is included in Appendix B.

The speakers were two participants of the survey who were undergraduate students at UCLA. Both speakers (m) were born in Beijing and lived there until coming to study in the United States at age 18. The elicitation was conducted in the audio room of the UCLA Phonetics Laboratory, using a head-mounted microphone and Audacity® (version 2.3.1; Audacity Team 2019). The speakers were recorded one at a time, with the author also present in the room, operating the software. The elicitation items were presented on paper as two lists of words, one consisting of the unsuffixed items and the other their suffixed version. The speakers were instructed to read through the unsuffixed list twice and then the suffixed one twice, at their natural pace, with a short break between lists. The elicitation yielded eight tokens (2 speakers × 2 items × 2 iterations) for either version of each rime. The target rime tokens were extracted from the elicitation recording files and processed in Praat (version 6.0.52; Boersma & Weenink 2019).

In order to calculate the phonetic distance between a pair of rimes, the duration of the pair must be normalized. However, normalization of duration is only justifiable if the durational difference between the two rimes is not phonologically meaningful. Figure 3.2 shows a boxplot of durational distrubution of either version of each of the 13 unique rimes. It can be seen that the *erhua* rimes (represented by shaded boxes) are not consistently longer or shorter than their respective plain version (unshaded boxes), and there is no drastic durational difference among each pair. A paired t-test confirmed that the duration of *erhua* rimes is not significantly different from that of plain rimes ($P = .32$), suggesting that the token-level differences in rime duration is not phonologically meaningful and therefore rime durations can be normalized for the purpose of phonetic distance calculation.

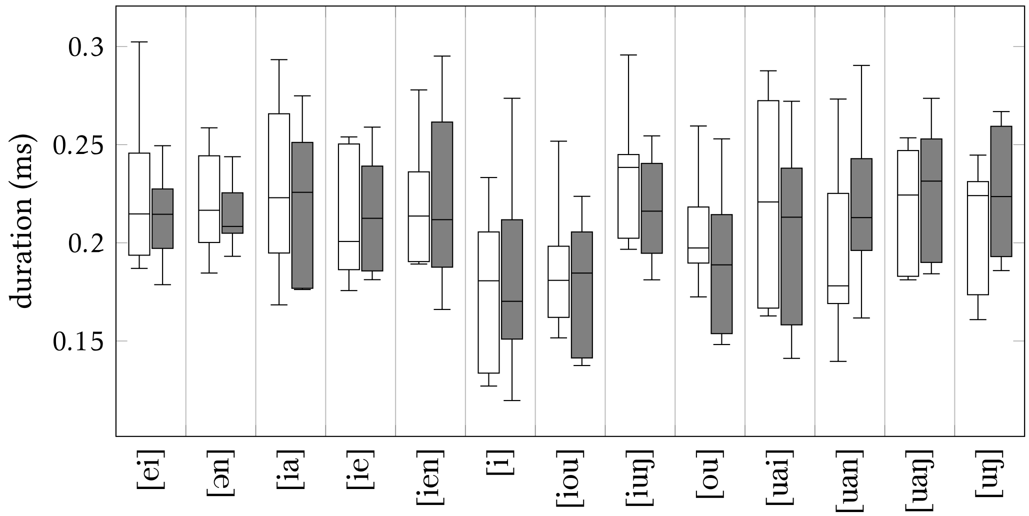

Figure 3.2: Boxplot of durational distribution of plain vs. *erhua* rimes. Unshaded boxes represent plain rimes; shaded boxes represent their *erhua* version.

3.4 Phonetic distance calculation

Given that rime duration can be normalized, the method for calculating the phonetic distance between a rime pair is as follows. First, each rime was divided into 11 slices of equal duration, resulting in 10 non-initial and non-final time points with equal intervals. At each such time point, the F1, F2, and F3 values in Bark scale were measured and recorded. Each measurement was then averaged across the eight tokens of the same rime (2 speakers × 2 items × 2 iterations), yielding one averaged value at each time point of each rime for each formant. Then, the Euclidean distance between each pair of plain rime and its *erhua* version, $r1$ and $r2$, was calculated for the three formants separately, using the following formula, as provided by Heeringa 2004:

$$\Delta(r1, r2) = \sqrt{\sum_{i=1}^{n}(r1_i - r2_i)^2} \quad (3.1)$$

In the formula above, n represents the number of time points at which formant measurements were taken ($n = 10$, in this case). This method of calculation yielded three independent distance values for each plain–*erhua* rime pair, one for each formant. Table 3.3 shows the resulting phonetic distance values for all 13 rime pairs.

	F1	F2	F3
Δ([i] , [iɚ])	6.21	8.66	8.83
Δ([ie] , [iɚ])	2.56	4.01	4.07
Δ([ei] , [əɚ])	2.17	5.45	5.61
Δ([ən] , [əɚ])	1.92	1.09	5.03
Δ([ien] , [iəɚ])	3.50	3.52	2.27
Δ([ia] , [ia˞ɚ])	1.69	2.01	1.55
Δ([uai] , [ua˞ɚ])	0.79	4.82	3.93
Δ([uan] , [ua˞ɚ])	1.72	1.17	1.69
Δ([uaŋ] , [ũã˞ɚ̃])	2.02	2.92	2.83
Δ([uŋ] , [ũ˞])	2.49	3.30	3.67
Δ([iuŋ] , [ĩũ˞])	2.62	4.04	2.42
Δ([ou] , [ou˞])	0.95	1.15	7.84
Δ([iou] , [io˞u˞])	1.19	2.41	4.47

Table 3.3: Phonetic distance values in Bark scale for all rime pairs, rounded to two decimal places.

4 Modeling

Mixed-effects logistic regression modeling was performed on the survey results, using the `lme4` package in R (R Core Team 2003). The data frame, including the survey results and stimuli specifications, were set up to mimic a maximum entropy harmonic grammar (MaxEnt HG; Goldwater & Johnson 2003) tableau, but with a few changes made to accommodate mixed-effects logistic regression.

In a traditional MaxEnt HG tableau, each output candidate of an input form occupies one row, and the observed number of tokens for each candidate, coalesced across speakers, serves as the dependent variable; and the number of times each candidate violates each constraint differentiates the candidates from each other. However, in order to include the random effect of participant in the regression models—assuming participants may have individual preferences of a rhotacization site—it is needed that the data frame preserves the behavioral difference between participants. This requires the dependent variable to be the binary individual choices of each participant, i.e., *first* or *second*, instead of the coalesced total number of tokens for each candidate. Secondly, in order to include the random effect of item—there may be intrinsic properties of certain combination of rimes that make one or the other rhotacization site more preferable—it is required that each item occupies only one row, instead of multiple rows with each candidate occupying one. Therefore, the difference between two candidates of the same input was coded not as the differentiating violation values α and β individually, but as the difference between α and β, i.e., $\beta - \alpha$. In the models reported below, the constraint violation value for each item was calculated as the violation value of the *second* candidate minus that of the *first* candidate, and the default level of the dependent variable was set to *second.* Table 4.1 shows the scheme for transforming a traditional MaxEnt HG tableau into the data frame input to mixed-effects logistic regression models.

	item	candidate	observed	CONSTRAINT 1	CONSTRAINT 2	…
MaxEnt HG tableau	A	A1	x	α_1	α_2	
	A	A2	y	β_1	β_2	
	item	participant	observed	CONSTRAINT 1	CONSTRAINT 2	…
regression data frame	A	1	A1 or A2	$\beta_1 - \alpha_1$	$\beta_2 - \alpha_2$	
	A	2	A1 or A2	$\beta_1 - \alpha_1$	$\beta_2 - \alpha_2$	
	A	…	A1 or A2	$\beta_1 - \alpha_1$	$\beta_2 - \alpha_2$	

Table 4.1: Corresponding rows in a traditional MaxEnt HG tableau and in the data frame for mixed-effects logistic regression modeling, where item A has two output candidates, A1 and A2. The values x and y are the number of times candidates A1 and A2 were respectively chosen by the pool of participants; α_n and β_n are the violation values for CONSTRAINT n, of candidates A1 and A2 respectively.

4.1 Baseline model and model selection

The baseline model, which both the featured-based and distance-based models were compared to, initially included both random effects, item and participant. However, the model did not converge until the participant random effect was removed.

Potential fixed effects independent of the two faithfulness approaches are markedness constraints that can differentiate the two *erhua* candidates of the same input. Two such markedness constraints were considered—*NONCENTRALV and *NONMIDV, which respectively penalize all front or back vowels (leaving out [ə ɚ ɚ̃ a a˞ ã˞]), and all high or low ones (leaving out [e ə ɚ ɚ̃ o o˞]). These two markedness constraints are inspired by the *PLACEV constraint proposed in Tian 2009, which penalize all non-[ə] vowels. They also reflect the cross-linguistically attested tendency for prerhotic vowels to neutralize to [ə], as in the diachronic phonology of English (Minkova 2013).

Given that there are two potential fixed effects, there are four different possible formulæ—(1) both *NONCENTRALV and *NONMIDV are included; (2) only *NONCENTRALV is included; (3) only *NONMIDV is included; and (4) neither fixed effect is included. All four versions were run with the `glmer()` function, with the item random effect included. Every two models were compared against each other, using the `anova()` function when they are nested, or basing on AIC (Akaike Information Criterion) when they are not. Comprehensive comparison of the four versions showed that only *NONMIDV significantly improved model performance. Thus, the final baseline model included the item random effect and *NONMIDV. The same model selection process was used for the featured-based model and the distance-based model as well.

The fixed effects in the final baseline model are reported in Table 4.2. The model intercept did not reach significance, indicating that overall there is no intrinsic preference for either rhotacization sites in disyllables. The markedness constraint *NONMIDV received a negative estimate, indicating that Beijing Mandarin speakers actually prefer front and back vowels than mid vowels in the output.

	estimate	std. error	z	$P(>\|z\|)$	
(intercept)	−0.1344	0.1200	−1.120	.27	
*NONMIDV	−0.8135	0.2612	−3.114	.002	**

Table 4.2: Summary of the baseline model.

4.2 Feature-based model

Traditional feature-based faithfulness constraints that are potential predictors include those in Table 4.3. The following stipulations were made in setting up sound correspondences between a plain and an *erhua* rime: (1) [i] → [ɚ] was treated as the deletion of [i] and insertion of [ɚ], thus violating MAX[i] and DEP[ɚ] but not IDENT(rhot); (2) [e] → [ɚ] was treated as [e] centralizing to [ə] and rhotacizing, thus violating IDENT(rhot) but not MAX[i] and DEP[ɚ]. The treatment of [e] → [ɚ] is motivated by the allophonic relationship between [e] and [ə] that is evidenced from their complementary distribution in the plain rime inventory of Beijing Mandarin.

However, when comparing the 31 different versions of the feature-based model (each including the item random intercept, *NONMIDV, and a different subset of the feature-based

IDENT(rhot)	(McCarthy & Prince 1995)	Do not rhotacize vowels in the plain rime (e.g., [u] → [u˞] as in [o$_1$u$_2$] → [o$_1$u˞$_2$]).
IDENT(front)	(McCarthy & Prince 1995)	Do not change the front-/back-ness specifications of the plain rime (e.g., [e] → [ə] as in [i$_1$e$_2$n$_3$] → [i$_1$ə$_2$ɚ$_3$]).
DEP[ɚ]	(McCarthy & Prince 1995)	Do not insert [ɚ] into the plain rime (e.g., [∅] → [ɚ] as in [i$_1$] → [i$_1$ɚ$_2$]).
MAX[i]	(McCarthy & Prince 1995)	Do not delete [i] from the plain rime (e.g., [i] → [∅] as in [e$_1$i$_2$] → [ə$_1$ɚ$_3$]).
MAX[+nasal]	(Zhang 2000)	Do not delete [+nasal] feature from the plain rime (e.g., [n] → [∅] as in [ə̃$_1$n$_2$] → [ə$_1$ɚ$_3$]).

Table 4.3: Feature-based faithfulness constraints.

faithfulness constraints in Table 4.3) to the baseline model, none came out significantly different from the baseline. A closer inspection at the constraint violation profile of the survey items revealed that the markedness constraint *NONMIDV and the feature-based faithfulness constraint MAX[i] are in perfect anticorrelation. Table 4.4 shows the violation profile of the only six survey items that have nonzero violation counts for either constraint. The reason for the perfect anticorrelation is that rhotacizing either of the two rimes [ei] and [uai] ([ei] → [əɚ] and [uai] → [ua˞ɚ]), [i] is deleted and [ɚ] is epenthesized, resulting in one violation of MAX[i] and one less violation of *NONMIDV, while rhotacizing [ie] ([ie] → [iɚ]) neither violates MAX[i] nor changes the violation count for *NONMIDV.

*NONMIDV		MAX[i]	
[uai + ie]	1	[uai + ie]	−1
[uai + ia]	1	[uai + ia]	−1
[ei + ie]	1	[ei + ie]	−1
[ie + ei]	−1	[ie + ei]	1
[ia + uai]	−1	[ia + uai]	1
[ie + uai]	−1	[ie + uai]	1

Table 4.4: Nonzero violation profile of survey items for *NONMIDV and MAX[i].

Due to this confound, the markedness constraint *NONMIDV was removed from all

	estimate	std. error	z	$P(>\|z\|)$	
(intercept)	−0.1344	0.1200	−1.120	.27	
Max[i]	0.8135	0.2612	3.114	.002	**

Table 4.5: Summary of the feature-based model.

versions of the feature-based model and the final feature-based model, reported in Table 4.5, includes the item random effect and Max[i]. In this model, Max[i] received a positive estimate, suggesting that Beijing Mandarin speakers prefer not to delete [i] from the original plain rimes in *erhua* productions.

4.3 Distance-based model

The change in the first three formants, F1–3, are potential predictors considered in the distance-based model. In this approach, instead of relying on traditional binary features in characterizing the difference between a plain rime and its *erhua* version, the raw phonetic distance between the two speech streams is used (See §3.3 for a detailed description of the calculation method). Therefore, for each *erhua* candidate, the violation profile for a distance-based faithfulness constraint would be the non-directional distance between an *erhua* candidate and its plain version. Due to the set-up of the data frame outlined in Table 4.1, the coding of the difference between $r1$ and $r2$, the *first* and *second erhua* candidates of item p, is therefore $\Delta(p,c2) - \Delta(p,c1)$ for each distance-based faithfulness constraint.

The same model selection method was used in determining the best-performing version of the distance-based model. The only predictor held constant in all versions was the item random effect, since the fixed effect in the baseline model turned out to be confounded with Max[i], a feature-based faithfulness constraint. The comprehensive comparison among the different versions revealed that F1 and F2 significantly improved the model performance over the model with only the item random effect. Thus, the final

	estimate	std. error	z	$P(>\|z\|)$	
(intercept)	−0.13370	0.10877	−1.229	.22	
F1	−0.30656	0.10676	−2.871	.004	**
F2	0.27178	0.06924	3.925	< .001	***

Table 4.6: Summary of the distance-based model.

model, reported in Table 4.6, includes the item random effect and the fixed effects F1, and F2. Of the two fixed effects, F2 received a positive estimate, indicating that Beijing Mandarin speakers prefer not to deviate from the plain rimes on the front-/back-ness dimension; F1, on the other hand, received a negative estimate, which suggests that speakers actually prefer changing the plain rimes along the height dimension when producing *erhua* outputs.

4.4 Model comparison

Two independent measures of model performance were used to compare the final distance-based model and the final feature-based model. AIC, which takes both model fit and model simplicity into account, estimates the information lost by a given model's representation of the actual process. A lower AIC value means less information loss and therefore better model performance, but a difference in AIC of less than two is negligible. In terms of AIC, the distance-based model achieved 562.1 and the feature-based one 565.6, indicating that the distance-based model performed marginally better than the feature-based one.

The opposite result was shown by comparing the two models' R^2 (R-squared; coefficient of determination), which is a measure of how much variance in the data was accounted for by a given model. In other words, R^2 only evaluates the goodness of fit of a model, unlike AIC, which also takes into account the number of parameters that a model uses. Since R^2 is typically used to evaluate linear regression models instead of logistic ones, I transformed the observed individual binary responses into by-item, across-participant response rates for choosing the *first erhua* candidate, e.g., if for item A, ten out of the 16 participants chose the *first* candidate, the observed *first* response rate for item A would be .625 ($10 \div 6$). Since there is no participant random effect in the models, they predicted a single probability for choosing the *first* candidate for each item, without by-participant variation, and those probabilities were taken as the predicted *first* response rates. Figures 4.1 & 4.2 were plots of those predicted vs. observed *first* response rates. As can be seen from the two scatter plots, the feature-based model achieved an R^2 of .698, much better than the distance-based model's .567.

4.5 Discussion

Recall the two yet unanswered research questions of the current study from §1: (2) is the principle of faithfulness the main deciding factor for choosing rhotacization site in Beijing Mandarin disyllabic words; and (3) does distance-based faithfulness yield a better model of the grammar. With the modeling results, (2) can now be answered—adding the potential markedness constraints *NONCENTRALV and *NONMIDV did not improve either model, suggesting that markedness is not the main deciding factor in choosing the rhotacization site in Beijing Mandarin; on the other hand, the R^2 of the two faithfulness models suggests that about 60–70% of the data can be accounted for by faithfulness alone. The answer to (2) is therefore *yes*. However, comparing the two faithfulness models did not give us a clear answer to the final research question—is distance-based faithfulness more capable than feature-based faithfulness at explaining rhotacization site preference in Beijing Mandarin. The two measures of evaluating model performance yielded contradicting results regarding which model performed better, leaving the answer to (3) unclear.

A closer inspection of the items' violation profile for the three different faithfulness constraints did not provide further insight. In terms of individual rimes mappings, the feature-based constraint MAX[i] disprefers [ei] → [əɚ˞] and [uai] → [ua˞ɚ˞] more than all others, since among the 13 unique rime mappings (Table 3.1), only [ei] → [əɚ˞] and

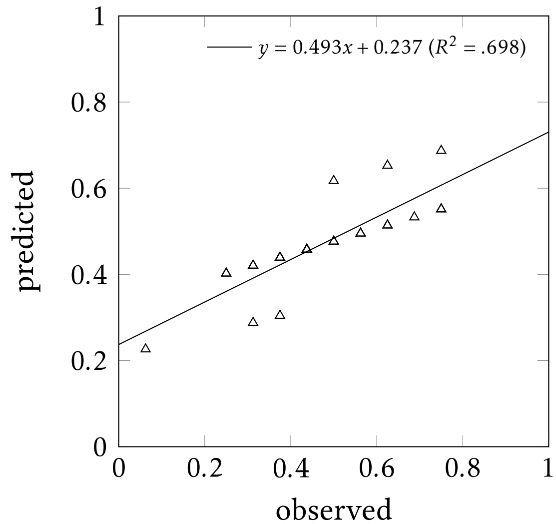

Figure 4.1: Feature-based model predicted vs. observed by-item response rate for choosing the *first* candidate.

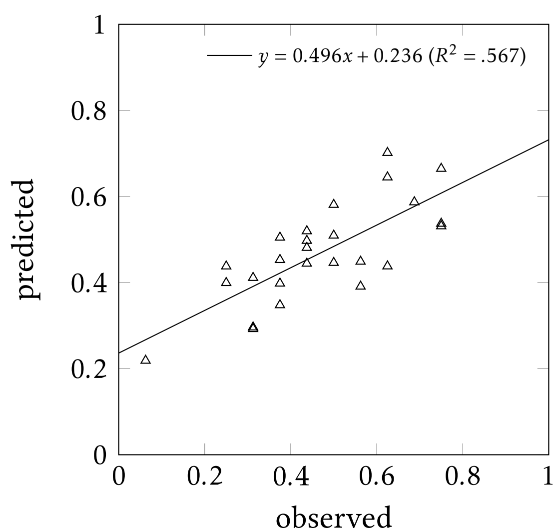

Figure 4.2: Distance-based model predicted vs. observed by-item response rate for choosing the *first* candidate.

[uai] → [ua˞ɚ] violate Max[i]. The two tiers that Max[i] distinguishes among the individual rime mappings translate to the three-tier categorical distinction among the survey items, shown in Table 4.7. In column Max[i], the items [uai+ie], [uai+ia], and [ei+ie] (top three) have a violation profile of −1; their reverse-order counterparts (bottom three) have a violation profile of 1; all other items have 0. Since the constraint Max[i] received a positive estimate, items with a violation profile of −1 are predicted to be least likely rhotacized on the first syllable rime, followed by those of 0, and those of 1 are predicted to be most likely rhotacized on the first. The distance-based faithfulness constraints, on the other hand, assign violation profiles directly basing on the raw phonetic distances, which do not yield categorical distinctions among items, but rather a continuous scale. In column F1, the items are sorted based on their F1 violation profiles, and in column F2, their F2 ones. Because the constraint F1 received a negative estimate, the higher an item's F1 violation profile value is, the more likely it is predicted to be rhotacized on the first; the opposite holds true for F2. The arrows connect the items with a Max[i] violation count of −1 or 1 to their respective position in the F1- and F2-ranked scales. It is observable that the scales ranked by the formant constraints do not support the three-tiered distinction made by the feature-based constraint, with the −1 and 1 items scattered in the middle of either scale. This suggests that the feature-based constraint Max[i] is in fact not correlated with either F1 or F2.

F1		Max[i]		F2	
[ia + i]	4.52	[uai + ie]	−1	[i + ən]	−7.57
[ən + i]	4.29	[uai + ia]	−1	[i + ia]	−6.65
[ie + i]	3.65	[ei + ie]	−1	[i + ie]	−4.65
[uai + ie]	1.77	[i + ən]	0	[uai + ia]	−2.80
[ən + ien]	1.58	[i + ia]	0	[ien + ən]	−2.43
[uai + ia]	0.90	[i + ie]	0	[ei + ie]	−1.44
[ou + ia]	0.75	[ien + ən]	0	[iou + ou]	−1.26
[uaŋ + iuŋ]	0.61	[iou + ou]	0	[iuŋ + uaŋ]	−1.12
[ei + ie]	0.39	[iuŋ + uaŋ]	0	[ia + ou]	−0.86
[ou + iou]	0.25	[ia + ou]	0	[ia + uan]	−0.84
[uan + ən]	0.21	[ia + uan]	0	[uai + ie]	−0.81
[uŋ + iuŋ]	0.13	[iuŋ + uŋ]	0	[iuŋ + uŋ]	−0.74
[ia + uan]	0.03	[uan + ən]	0	[uan + ən]	−0.08
[uan + ia]	−0.03	[ən + uan]	0	[ən + uan]	0.08
[iuŋ + uŋ]	−0.13	[uŋ + iuŋ]	0	[uŋ + iuŋ]	0.74
[ən + uan]	−0.21	[uan + ia]	0	[ie + uai]	0.81
[iou + ou]	−0.25	[ou + ia]	0	[uan + ia]	0.84
[ie + ei]	−0.39	[uaŋ + iuŋ]	0	[ou + ia]	0.86
[iuŋ + uaŋ]	−0.61	[ou + iou]	0	[uaŋ + iuŋ]	1.12
[ia + ou]	−0.75	[ən + ien]	0	[ou + iou]	1.26
[ia + uai]	−0.90	[ie + i]	0	[ie + ei]	1.44
[ien + ən]	−1.58	[ia + i]	0	[ən + ien]	2.43
[ie + uai]	−1.77	[ən + i]	0	[ia + uai]	2.80
[i + ie]	−3.65	[ie + ei]	1	[ie + i]	4.65
[i + ən]	−4.29	[ia + uai]	1	[ia + i]	6.65
[i + ia]	−4.52	[ie + uai]	1	[ən + i]	7.57

Table 4.7: Violation profile of survey items for Max[i], F1, and F2. For F1, the higher an item's violation profile value is, the less likely it is predicted to be rhotacized on the first rime; the opposite holds true for F2. For Max[i], the three possible violation profiles are −1, 0, and 1; items with a violation profile of −1 are predicted to be least likely to be rhotacized on the first rime, whereas those with 1 are most likely. Arrows connect the items with a Max[i] violation profile of −1 and 1 to their respective positions in the likelihood scales predicted by F1 and F2.

5 General discussion and conclusion

In this study, I examined Beijing Mandarin rime rhotacization in disyllabic nonce words and demonstrated that when faced with two syllable rimes as the potential site of rhotacization, native speakers of the Beijing dialect prefer to rhotacize the rime that would be more faithful to the non-rhotic version when rhotacized. It is shown in this study that (1) Beijing Mandarin speakers rely on phonological principles to determine the rhotacization site in disyllables, and (2) specifically, they prefer to rhotacize the rime whose rhotic output is more faithful to the original non-rhotic version (they rely on the faithfulness principle).

In constructing the nonce-word stimuli for the experiment, a novel method was used to eliminate the potential confound of lexical propensity. Namely, the stimuli were constructed from unattested combinations of attested Beijing Mandarin onsets and rimes—in other words, gaps in Beijing Mandarin syllable inventory. While using nonce syllables allowed clean separation of the grammar and lexical frequency patterns, in the case of Beijing Mandarin, whose native speakers had likely never been exposed to non-native, "wug" syllables, it could make the task of providing judgments for the experiment stimuli unintuitive and exceptionally difficult.

As a phenomenon that no previous study has closely examined, rime rhotacization site variation in Beijing Mandarin warrants further investigation from various perspectives. For instance, with a robust representation in the native lexicon, it can be used to test whether lexical frequency distribution is learned and generalized by native speakers. Another potential area of research this phenomenon can bear on is the effect of homophony avoidance in phonology. A subset of the Beijing Mandarin plain–*erhua* rime mappings is contrast-neutralizing (e.g., [i ie in] → [iɚ]), resulting in homophony in *erhua* forms; other mappings are not. Thus, if speakers indeed disprefer neutralization of contrasts, rimes that map to unique *erhua* forms would be more preferable than those that would result in homophony.

In the current study, Beijing Mandarin rime rhotacization site variation is used to address the following theoretical question: does distance-based faithfulness, which stems from the P-map proposal (Steriade 2001) and allows phonological grammars to evaluate faithfulness of output candidates directly using the phonetic distance between an input and its output, can account for gradient alternation patterns better than traditional feature-based faithfulness can. However, the two grammar models in the study, which respectively included distance-based faithfulness constraints and feature-based ones, performed equally well to account for the gradient variation in the data. It is likely that the variation in the data collected in this study was not robust enough to tease the two approaches apart. A follow-up study that will employ a larger and more varied pool of experiment stimuli is planned.

A Survey stimuli

rime pair		item	*pinyin*
[ei]	[ie]	[tʰei˧˥.fie˧˥] [fie˥˩.tʰei˥˩]	téi fié fiè tèi
[uai]	[ie]	[pʰuai˥˩.fie˥˩] [fie˧˥.pʰuai˧˥]	puài fiè fié puái
[uai]	[ia]	[pʰuai˧˥.fia˧˥] [fia˥˩.pʰuai˧˥]	puái fiá fià puài
[ie]	[i]	[fie˥˩.kʰi˥˩] [kʰi˧˥.fie˧˥]	fiè kì kí fié
[ien]	[ən]	[fien˧˥.tʰən˧˥] [tʰen˥˩.fien˥˩]	fián tén tèn fiàn
[uan]	[ən]	[pʰuan˥˩.tʰən˥˩] [tʰən˧˥.pʰuan˧˥]	puàn tèn tén puán
[uan]	[ia]	[pʰuan˧˥.fia˧˥] [fia˥˩.pʰuan˥˩]	puán fiá fià puàn
[ən]	[i]	[tʰən˥˩.kʰi˥˩] [kʰi˧˥.tʰən˧˥]	tèn kì kí tén
[uaŋ]	[iuŋ]	[muaŋ˧˥.fiuŋ˧˥] [miuŋ˥˩.fuaŋ˥˩]	muáng fióng miòng fuàng
[ia]	[i]	[fia˥˩.kʰi˥˩] [kʰi˧˥.fia˧˥]	fià kì kí fiá
[ia]	[ou]	[fia˧˥.pou˧˥] [pou˧˥.fia˥˩]	fiá bóu bòu fià
[uŋ]	[iuŋ]	[muŋ˥˩.fiuŋ˥˩] [miuŋ˧˥.fuŋ˧˥]	mòng fiòng mióng fóng
[ou]	[iou]	[pou˧˥.fiou˧˥] [fiou˥˩.pou˥˩]	bóu fióu fiòu bòu

B Elicitation items

target rime	word 1	word 2
[ei] (→ [əɚ])	酒杯 [t͡ɕiou˨˩˦.pei˥] 酒杯儿 [t͡ɕiou˨˩˦.pəɚ˥]	摸黑 [muo˥.xei˥] 摸黑儿 [muo˥.xəɚ˥]
[ən] (→ [əɚ])	扣分 [kʰou˥˩.fən˥] 扣分儿 [kʰou˥˩.fəɚ˥]	脸盆 [lien˨˩˦.pʰən˧˥] 脸盆儿 [lien˨˩˦.pʰəɚ˧˥]
[ia] (→ [ia˞ɚ])	老家 [lau˨˩˦.t͡ɕia˥] 老家儿 [lau˨˩˦.t͡ɕia˞ɚ˥]	两下 [liaŋ˨˩˦.ɕia˥˩] 两下儿 [liaŋ˨˩˦.ɕia˞ɚ˥˩]
[ie] (→ [iɚ])	锅贴 [kuo˥.tʰie˥] 锅贴儿 [kuo˥.tʰiɚ˥]	台阶 [tʰai˧˥.t͡ɕie˥] 台阶儿 [tʰai˧˥.t͡ɕiɚ˥]
[ien] (→ [iəɚ])	花边 [xua˥.pien˥] 花边儿 [xua˥.piəɚ˥]	唱片 [t͡ʂʰaŋ˥˩.pʰien˥] 唱片儿 [t͡ʂʰaŋ˥˩.pʰiəɚ˥]
[i] (→ [iɚ])	漏气 [lou˥˩.t͡ɕʰi˥˩] 漏气儿 [lou˥˩.t͡ɕʰiɚ˥˩]	冰皮 [piŋ˥.pʰi˧˥] 冰皮儿 [piŋ˥.pʰiɚ˧˥]
[iou] (→ [io˞u˞])	山丘 [ʂan˥.t͡ɕʰiou˥] 山丘儿 [ʂan˥.t͡ɕʰio˞u˞˥]	长袖 [t͡ʂʰaŋ˧˥.ɕiou˥˩] 长袖儿 [t͡ʂʰaŋ˧˥.ɕio˞u˞˥˩]
[iuŋ] (→ [ĩũ˞])	小熊 [ɕiau˨˩˦.ɕiuŋ˧˥] 小熊儿 [ɕiau˨˩˦.ɕĩũ˞˧˥]	词穷 [t͡sʰɹ̩˧˥.t͡ɕʰiuŋ˧˥] 词穷儿 [t͡sʰɹ̩˧˥.t͡ɕʰĩũ˞˧˥]
[ou] (→ [ou˞])	裤兜 [kʰu˥˩.tou˥] 裤兜儿 [kʰu˥˩.tou˞˥]	小偷 [ɕiau˨˩˦.tʰou˥] 小偷儿 [ɕiau˨˩˦.tʰou˞˥]
[uai] (→ [ua˞ɚ])	脚踝 [t͡ɕiau˨˩˦.xuai˧˥] 脚踝儿 [t͡ɕiau˨˩˦.xua˞ɚ˧˥]	方块 [faŋ˥.kʰuai˥˩] 方块儿 [faŋ˥.kʰua˞ɚ˥˩]
[uan] (→ [ua˞ɚ])	撸串 [lu˥.t͡ʂʰuan˥˩] 撸串儿 [lu˥.t͡ʂʰua˞ɚ˥˩]	瓦罐 [wa˨˩˦.kuan˥˩] 瓦罐儿 [wa˨˩˦.kua˞ɚ˥˩]
[uaŋ] (→ [ũã˞ɚ̃])	蛋黄 [tan˥˩.xuaŋ˧˥] 蛋黄儿 [tan˥˩.xũã˞ɚ̃˧˥]	竹筐 [t͡ʂu˧˥.kʰuaŋ˥] 竹筐儿 [t͡ʂu˧˥.kʰũã˞ɚ̃˥]
[uŋ] (→ [ũ˞])	草丛 [t͡sʰau˨˩˦.t͡sʰuŋ˧˥] 草丛儿 [t͡sʰau˨˩˦.t͡sʰũ˞˧˥]	理综 [li˨˩˦.t͡suŋ˥] 理综儿 [li˨˩˦.t͡sũ˞˥]

Bibliography

Audacity Team. 2019. *Audacity: Free audio editor and recorder* [Computer software]. Version 2.3.1 retrieved 4 May 2019 from https://audacityteam.org/.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer* [Computer software]. Version 6.0.52 retrieved 2 May 2019 from http://www.praat.org/.

Duanmu, San. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages.* Cambridge, MA: MIT dissertation.

Duanmu, San. 2007. *The phonology of Standard Chinese.* 2nd edn. Oxford: Oxford University Press.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Osten Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University, Department of Linguistics.

Heeringa, Wilbert J. 2004. *Measuring dialect pronunciation differences using Levenshtein distance.* Groningen: University of Groningen dissertation.

Lee, Wai-Sum. 2005. A phonetic study of the "er-hua" rimes in Beijing Mandarin. *Interspeech* 6. 1093–1096.

Lee, Wai-Sum & Eric Zee. 2014. Chinese phonetics. In C-T James Huang, Y-H Audrey Li & Andrew Simpson (eds.), *Handbook of Chinese linguistics*, 369–399. Oxford: Wiley-Blackwell.

Ma, Lili. 1997. Mandarin diminutive formation: an optimality analysis. *Working Papers of the Linguistics Circle of the University of Victoria* 14. 47–60.

McCarthy, John J & Alan S Prince. 1995. *Faithfulness and reduplicative identity.* Retrieved from doi:10.7282/T31R6NJ9.

McCollum, Adam G. 2018. Vowel dispersion and Kazakh labial harmony. *Phonology* 35(2). 287–326.

Minkova, Donka. 2013. *Historical phonology of English*. Edinburgh: Edinburgh University Press.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Retrieved from doi:10.7282/T34M92MV.

R Core Team. 2003. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Steriade, Donca. 2001. *The phonology of perceptibility effects: The P-map and its consequences for constraint organization*. Retrieved from doi:10.7551/mitpress/9780262083799.003.0007.

Tian, Jun. 2009. An optimality theory analysis of diminutive suffixation of Beijing Chinese. *Working Papers of the Linguistics Circle of the University of Victoria* 19. 217–232.

Wu, Fei & Michael Kenstowicz. 2015. Duration reflexes of syllable structure in Mandarin. *Lingua* 164. 87–99.

Zhang, Jie. 2000. Non-contrastive features and categorical patterning in Chinese diminutive suffixation: Max[F] or Ident[F]? *Phonology* 17(3). 427–478.

Zymet, Jesse. 2018. *Lexical propensities in phonology: Corpus and experimental evidence, grammar, and learning*. Los Angeles, CA: UCLA dissertation.