Research Article

Shuwen Chen, Douglas H. Whalen and Peggy Pik Ki Mok* What R Mandarin Chinese /」/s? – acoustic and articulatory features of Mandarin Chinese rhotics

https://doi.org/10.1515/phon-2023-0023 Received May 23, 2023; accepted August 16, 2024; published online September 16, 2024

Abstract: Rhotic sounds are well known for their considerable phonetic variation within and across languages and their complexity in speech production. Although rhotics in many languages have been examined and documented, the phonetic features of Mandarin rhotics remain unclear, and debates about the prevocalic rhotic (the syllable-onset rhotic) persist. This paper extends the investigation of rhotic sounds by examining the articulatory and acoustic features of Mandarin Chinese rhotics in prevocalic, syllabic (the rhotacized vowel [&]), and postvocalic (r-suffix) positions. Eighteen speakers from Northern China were recorded using ultrasound imaging. Results showed that Mandarin syllabic and postvocalic rhotics can be articulated with various tongue shapes, including tongue-tip-up retroflex and tongue-tip-down bunched shapes. Different tongue shapes have no significant acoustic differences in the first three formants, demonstrating a many-to-one articulation-acoustics relationship. The prevocalic rhotics in our data were found to be articulated only with bunched tongue shapes, and were sometimes produced with frication noise at the start. In general, rhotics in all syllable positions are characterized by a close F2 and F3, though the prevocalic rhotic has a higher F2 and F3 than

The speakers and Mandarin data presented in this paper are the same as those used in Chen et al. (2024). However, this study includes data from 18 speakers, compared to the 17 speakers in Chen et al. (2024). While the current study examines the native Mandarin production of these participants, Chen et al. (2024) focused on their L2 English production.

^{*}Corresponding author: Peggy Pik Ki Mok, Department of Linguistics and Modern Languages, G/F, Leung Kau Kui Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China, E-mail: peggymok@cuhk.edu.hk

Shuwen Chen, Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China; and Key Laboratory of Linguistics, Chinese Academy of Social Sciences, Beijing, China, E-mail: chensw@cass.org.cn. https://orcid.org/0000-0002-0707-2619

Douglas H. Whalen, City University of New York, New York, USA; and Yale University, New Haven, CT, USA, E-mail: douglas.whalen@yale.edu

the syllabic and postvocalic rhotics. The effects of syllable position and vowel context are also discussed.

Keywords: Mandarin Chinese; rhotics; articulation; acoustics; ultrasound imaging

1 Introduction

Rhotic sounds, or "r-sounds", attract great attention in the field of phonetics because of their complex and diverse phonetic properties. In phonetics and phonology, most sound classes are defined based on a set of articulatory or acoustic features. Rhotics, however, are difficult to define because it is hard to identify the shared articulatory or acoustic characteristics of this class of sounds (Lindau 1985; Ladefoged and Maddieson 1996). Rhotics are usually represented orthographically by the letter "r" or its Greek counterpart *rho* (Ladefoged and Maddieson 1996). Rhotics vary drastically in their place and manner of articulation across languages: we find examples such as an alveolar trill [r], a post-alveolar approximant [1], an alveolar tap or flap [r], a uvular trill [R], a uvular fricative [B], and so on (Ladefoged and Maddieson 1996; Lindau 1985). Within a language, several rhotic variants can coexist, such as the Dutch rhotics that can be realized as either a uvular trill or a post-alveolar approximant (Scobbie and Sebregts 2010). Wide articulatory variation can be found, even within a single rhotic segment. For example, the English post-alveolar approximant /1/ can be articulated with a continuum of tongue shapes ranging from tongue-tip-up (retroflex tongue shape) to tongue-tip-down (bunched tongue shape) (Delattre and Freeman 1968; Mielke et al. 2010, 2016; Westbury et al. 1998; Zhou et al. 2008). As identifying common articulatory and acoustic features between all rhotics is challenging (Lindau 1985; Ladefoged and Maddieson 1996), Howson and Monahan (2019) proposed a perceptual basis for categorizing rhotics as a distinct class, suggesting that F1 and F2 formant frequencies and trajectories might serve as acousticperceptual correlates that explain both the perceptual similarities within the class and any clear distinctions from other sounds.

Due to the elusive definition of rhotics, describing these sounds across the world's languages has been a nontrivial issue. The phonetic features of rhotics in some languages have been examined in great detail, such as the English post-alveolar approximant /I/ (e.g., Delattre and Freeman 1968; Lindau 1985; Mielke et al. 2010, 2016; Tiede et al. 2010; Westbury et al. 1998; Zhou et al. 2008). However, the phonetic features of Mandarin Chinese rhotics have not been well-investigated to date. Mandarin Chinese, also known as Standard Chinese or Putonghua (普通话), is the national language of the People's Republic of China and is used in schools and workplaces throughout the country. It is based on Mandarin dialects and takes

Beijing pronunciation as its standard pronunciation. The term *Mandarin Chinese* is used in this study because it is a very commonly used term in the literature written in English. In *Pinyin*, the official romanization system for Mandarin Chinese used in mainland China, the rhotics are represented by an "r" (Chinese Ministry of Education 1958). In syllable-onset position (i.e., *ri* in *Pinyin*), a rhotic is often described as a post-alveolar approximant /*I*/ (Fu 1956; Lee and Zee 2003; Lin 2007), or sometimes as a voiced fricative (Duanmu 2007; Karlgren 1915-1926; Yuan 1960). When a rhotic sound occurs in a syllable-nucleus position (i.e., *er* in *Pinyin*), it is usually transcribed as a rhotacized vowel [ə-] (Duanmu 2007; Lee and Zee 2003; Zee and Lee 2001), or a syllabic post-alveolar approximant /*I*/ (Lin 2007). The Mandarin Chinese rhotic can also be a suffix and merge with the preceding vowel in r-suffixation (*er-hua* "*J*L(*K*"). In this circumstance, the rhotic is analyzed either as a rhotic feature of the preceding vowel (Lin and Wang 2013; Wang 1993), or as a postvocalic rhotic (Chao 1968; Lin 1989, 2007). The current study aims to investigate the phonetic features of Mandarin rhotics in these three circumstances using acoustic and articulatory measures.

1.1 Rhotics in Mandarin Chinese

Different phonetic symbols have been used in the literature to represent Mandarin rhotic sounds, which will be introduced in detail in the next two sections. For ease of discussion, in the present study we will use the symbol for post-alveolar approximant /1/ to represent rhotics in prevocalic and postvocalic positions, and the symbol of the syllabic approximant /1/ in syllable-nucleus position, to indicate that they are all rhotic sounds.

1.1.1 The Mandarin Chinese prevocalic rhotic

In prevocalic position, the rhotic sound has been transcribed as a post-alveolar approximant [I] (or [r] for ease of typing and printing) (Fu 1956; Lin 2007), a postalveolar approximant with a subscript indicating apical features [I] (Lee 1999; Lee and Zee 2003), or a post-alveolar voiced fricative [z] (Duanmu 2007; Karlgren 1915-1926; Yuan 1960). This is because the phonological status of the prevocalic rhotic sound is controversial. Together with the post-alveolar fricative and affricates /\$/, /t\$/ and /t\$^h/, prevocalic rhotics are usually called "retroflex consonants" in the literature and in classroom settings (Chao 1968; Duanmu 2007). This is because early accounts of these consonants described the sounds as articulated with the tongue tip curling up (Chao 1948, 1968). In prevocalic position, a rhotic can be followed by the retroflex apical vowel [η], high back vowel [u], mid vowel [γ ə], low vowel [a d], and diphthong [au ou], but not by the high front unrounded vowel [i], high front rounded vowel [y], or apical vowel [η]. We use square brackets [] here to refer to phonetic vowel qualities rather than their phonemic status because the number of vowel phonemes in Mandarin Chinese is controversial in the literature (see Mok 2013 for a summary). Mandarin Chinese has two apical vowels, [η] and [η]. These two symbols, [η] and [η], are not IPA symbols, but they are commonly used in the literature on Mandarin Chinese phonology to represent the two high front apical segments. The phonological status of the two sounds is also controversial, but it is not central to our study (see the discussion in Lee-Kim 2014). The apical vowel [η] appears after the dental affricates and fricatives /ts/, /ts^h/, and /s/, while the apical vowel [η] appears after the post-alveolar consonants /tş/, /ts^h/, /s/, and /I/.

One debate about the prevocalic rhotic is whether this sound is an approximant or a fricative. In early descriptions of the Mandarin sound inventory, Karlgren (1915– 1926) and Yuan (1960) transcribed the Mandarin prevocalic rhotic as a voiced retroflex fricative [z]. In Duanmu (2007)'s phonological analysis, he also proposed that this rhotic should be categorized as a voiced fricative /z/, based on his interpretation that the relationship between Mandarin |s| and |z| was similar to that between English /s/ and /z/. But he also admitted that this categorization would introduce the only voiced obstruent in the Mandarin sound inventory. In this line of study, the prevocalic rhotic and rhotics in other syllable positions (the syllabic rhotic and the r-suffix) are different phonemes (Duanmu 2007; Karlgren 1915–1926). Other studies, following the tradition of Chao (1948, 1968), described the prevocalic rhotic as a post-alveolar approximant, and used |J| or |J| (an approximant that is produced with the apical part of the tongue) to represent it (Fu 1956; Lee and Zee 2003; Lin 2007). A reference book on Chinese dialects treated it as a voiced fricative |z|phonologically but explained that it was a retroflex approximant [J] phonetically (Department of Chinese at Peking University 2003).

One of the central issues in this debate relates to the existence of frication noise in the prevocalic rhotic. Previous studies examining this factor, however, have reported mixed results. Smith (2010) found some frication noise in the prevocalic rhotic, while Lee (1999) reported an absence of frication noise. However, Smith (2010) only provided one spectrogram with frication noise, without mentioning how frequently the frication noise was found, and the conclusion of Lee (1999) was based on only four speakers. Based on data from 18 Beijing speakers, Xing (2021) found that the prevocalic rhotic followed by the high vowel /u/ showed the highest occurrence of frication. Individual differences have also been found in the production of the prevocalic rhotic. Liao and Shi (1987) examined the production of the Mandarin Chinese prevocalic rhotic in monosyllabic and disyllabic words spoken by four speakers. A male speaker from Beijing produced prevocalic rhotics without any frication noise. A male speaker from Heilongjiang Province exhibited slight frication noise preceding the high vowel [u] and the apical vowel [1]. The remaining two speakers, a female from Beijing and a female from Xinjiang Province, both demonstrated noticeable frication. They concluded that prevocalic rhotics could be categorized into three types: strong frication, weak frication, and no frication, with the first two variations representing voiced fricatives and the last representing an approximant. Chuang et al. (2015) examined and transcribed the production of the prevocalic rhotic in Taiwan Mandarin, a variant of Mandarin that is spoken in Taiwan and strongly influenced by Taiwan Southern Min. They found that the Mandarin prevocalic rhotic could be realized as a voiced retroflex fricative [z], retroflex approximant [J], voiced fricative [z] or a lateral [l], and they found that the predominant realization of this sound was an approximant, rather than a fricative. But this study was based only on auditory judgments. Zhu (2007) argued against categorizing the Mandarin Chinese prevocalic rhotic as a fricative, citing the rarity of voiced fricatives in languages without voiced stops. He compared 317 languages, finding only three (including Mandarin Chinese) that had voiced fricatives but no voiced stops, suggesting this is inconsistent with linguistic typology. Therefore, a more detailed examination of frication noise is warranted to resolve the debate on the prevocalic rhotic.

Acoustically, previous studies have reported that the mean frequencies of the first three formants of the prevocalic rhotic are around 291 Hz, 1,647 Hz and 2,713 Hz for male speakers and 300 Hz, 1,900 Hz, and 2,193 Hz for female speakers (Lee 1999). Articulatorily, Chao (1968) stated that Mandarin rhotic sounds involved the tongue tip curling up, but this assertion was not grounded in experimental data. A recent ultrasound study (Xing 2021) provided evidence supporting tip-up retroflex tongue shapes for Mandarin Chinese prevocalic rhotics in 8 out of 18 Beijing speakers. However, some studies have suggested that the articulation of the prevocalic rhotic does not involve tip-up retroflex tongue shapes. Lee (1999) investigated the articulation of the prevocalic rhotic in four native Beijing Mandarin speakers using palatograms and linguagrams and found no evidence of curling up of the tongue. In a more recent study, Zhu and Mok (2023) examined eight native Beijing Mandarin speakers and eight Japanese-Mandarin simultaneous bilinguals with ultrasound imaging and found only bunched gestures for the Mandarin Chinese prevocalic rhotic. Although not examining the prevocalic rhotic directly, Luo (2020) examined the other Mandarin Chinese "retroflex consonants" /s ts ts^h/ and alveolar sibilants /s ts ts^h/ using ultrasound imaging. She found various tongue shapes for /s ts ts^h/, including bunched, retroflex, and humped tongue shapes. But only one instance of the retroflex tongue shape and nine tokens of the bunched tongue shape were observed, while the major tongue shape was the humped tongue shape. The humped tongue shape refers to a semi-circular shape without alveolar constriction or tongue root retraction, and it has been included in the category of "bunched" tongue shape

DE GRUYTER

in previous studies on the English /x/ (Lawson et al. 2011, 2018; Mielke et al. 2016) or treated as a misarticulation of the English /x/ in children (Klein et al. 2013). In summary, it is still unclear how common the production of the prevocalic rhotic involving tip-up retroflex tongue shapes actually is.

1.1.2 The Mandarin Chinese syllabic and postvocalic rhotics

Unlike the prevocalic rhotic, the syllabic and postvocalic rhotics are often treated as rhotacized vowels rather than consonants in many studies (Duanmu 2007; Lin and Wang 2013; Zee and Lee 2001). When a rhotic forms a syllable on its own, it has been described as a rhotacized vowel [\Rightarrow] (Lee and Zee 2014, 2003; Duanmu 2007), a syllabic post-alveolar approximant [\mathfrak{I}] (or [\mathfrak{r}] for ease of typing and printing) together with a diacritic [\mathfrak{I}] ([\mathfrak{r}]) (Lin 2007), or sometimes as a mid-central vowel followed by a post-alveolar consonant [$\mathfrak{d}\mathfrak{I}$] (Lin 2007; Lin and Wang 2013). In this circumstance, it cannot be followed or preceded by any consonants. Therefore, Mandarin Chinese has only a few words with this syllable structure and varying tones, such as / \mathfrak{I}_{35} / 'son, child' (" $\mathfrak{I}\mathfrak{L}$ "), / \mathfrak{I}_{214} / 'ear' (" \mathfrak{P} "), and / \mathfrak{I}_{51} / 'two' (" \mathfrak{T} ").

The /I/ phoneme is not allowed in postvocalic position in its underlying form, but the /I/ sound does occur in coda position after an r-suffixation process. R-suffixation is a common feature of the Mandarin spoken in Northern China, such as Beijing, Shandong Province, and Hebei Province (Wang 2005). The r-suffix is a diminutive suffix, or is used to refer to a familiar object (Li 1996; Lin 1992; Wang 2005). Orthographically, it is represented by the word /I_35/ 'son, child' (" /L"). In Mandarin Chinese, in most cases one character represents one syllable. But for the two-character sequence that combines a word with /I_35/, the two characters are pronounced as only one syllable. That is, the syllabic /I/ undergoes syllable contraction, and merges with the preceding vowel as part of the rime (Lin 2007). Therefore the sequence of vowel and r-suffix is formed through a morphophonological process, and is usually transcribed as a single vowel together with a [I] sound, or a rhotic vowel [\mathfrak{P}], such as [\mathfrak{a} I] and [\mathfrak{a} -] (Duanmu 2007; Hu 2020; Huang et al. 2020; Lin 1989, 2007; Lee and Zee 2014), or with a diacritic [-], such as [\mathfrak{u} -] (Lee 2005).

Phonologically, the r-suffix can be combined with both simple (monophthongs) and complex (diphthongs, or a vowel followed by a nasal coda) rimes (Lin 2007). When a monophthong undergoes r-suffixation, the /J/ can be attached to a low vowel [a] (/ts^ha155/ 'cross' 叉儿), a mid vowel [s] (/kx155/ 'song' 歌儿), or a high back vowel [u] (/hu135/ 'soul' 魂儿), while a schwa [ə] will be inserted when the /J/ is attached to a high front vowel (/tciə155/ 'chicken' 鸡儿). In terms of articulation, the inserted schwa is a result of the tongue passing through a schwa-like configuration when transiting from the high front vowel to the following /J/ sound (Gick and Wilson 2006; Huang et al. 2020; Jiang et al. 2019a). When a diphthong rime undergoes r-suffixation, the

diphthong can either become a monophthong or remain the same, depending on the compatibility of the two vowels in the diphthong and the /I/ coda (e.g., [$t^{h}ou_{35}$] \rightarrow [$t^{h}ou_{35}$] 'head', [tai_{51}] \rightarrow [taI_{51}] 'bag'). As for words containing a nasal coda, the nasal will be deleted before attaching the /I/ coda (e.g., [$kwan_{55}$] \rightarrow [$kwaI_{55}$] "officer").

Acoustically, the postvocalic /1/ is marked by a low F3 (Hu 2020; Lee 2005; Lee and Zee 2014). It is controversial as to whether the articulation of the syllabic and postvocalic rhotics involves retroflexion. Using Electromagnetic Articulography (EMA) data from three Beijing speakers (one male and two females), Lee (2005) reported that there was no retroflexion in the articulation. Tip-up tongue shapes have, however, been found in some other studies (Jiang et al. 2019a; King and Liu 2017; Xing 2021). King and Liu (2017) examined the tongue shapes of the postvocalic rhotics of 12 native Mandarin speakers using ultrasound imaging. They found that the postvocalic rhotic could be articulated with various tongue shapes – tip up, front up and front bunched. Jiang et al. (2019a) also found tip-up tongue shapes in three Beijing Mandarin speakers using EMA. Xing (2021) reported that 15 out of 18 Beijing speakers consistently used retroflex tongue shapes. Dynamically, the Mandarin Chinese prevocalic and postvocalic rhotics can be produced with two active movements of the tongue – tongue-anterior raising and tongue-root backing. The tongue-root backing gesture begins earlier than the maximum displacement of the tongue-anterior raising gesture (Gick et al. 2006).

Pharyngeal retraction has been reported in some previous studies. Lee and Zee (2014) found pharyngeal retraction in the postvocalic rhotic (r-suffix) in Beijing Mandarin by comparing the tongue shapes of r-suffixed vowels with their corresponding plain vowels using one female speaker's EMA data. Xing (2021) further compared the prevocalic and postvocalic rhotics using ultrasound data, and found a more retracted tongue root in the postvocalic rhotic. Another study examining Southwestern Mandarin, however, found no pharyngealization in either the postvocalic or syllabic rhotics (rhotic schwas in their study) using ultrasound imaging (Huang et al. 2020).

There are also sub-dialectal differences in the articulation of Mandarin Chinese rhotics. Jiang et al. (2019a) compared the tongue shapes of the postvocalic rhotic in Beijing Mandarin and Northeastern Mandarin. Three Beijing Mandarin speakers and three Northeastern Mandarin speakers were examined using EMA. They found that curling-up tongue shapes could be found in Beijing Mandarin but not Northeastern Mandarin. Jiang et al. (2019b) compared the articulation of the syllabic and postvocalic rhotics in Northeastern Mandarin using EMA data from three speakers. Their data showed that, in Northeastern Mandarin, the postvocalic rhotic was produced with tip-down bunched gestures, while the syllabic rhotic was produced with a tip-up tongue shape. They also found that the production of the postvocalic rhotic in Northeastern Mandarin involved a retraction of the tongue body. Huang et al. (2020) examined the articulatory gestures of the syllabic and postvocalic rhotics in Southwestern Mandarin. Two speakers from western Hubei who spoke a variety of Chengdu-Chongqing dialects were recorded using co-registered EMA and ultrasound. They found that the syllabic and postvocalic rhotics in Southwestern Mandarin were both produced with bunched tongue shapes. They also compared the tongue shapes of the syllabic and postvocalic rhotics. When the postvocalic rhotic followed non-high vowels, the tongue shapes of the syllabic and postvocalic rhotics were similar, while in high front vowel contexts, the postvocalic rhotic had a significantly higher tongue tip.

In summary, the Mandarin syllabic and postvocalic rhotics are acoustically characterized by a low F3. Various tongue shapes in their articulation, as well as pharyngeal retraction, have been reported. Additionally, variations in tongue shapes exist across sub-dialects.

1.2 The current study

This study explores the variability in the phonetic realization of rhotics in various syllable positions in Mandarin Chinese. We aim to answer three research questions. The first question concerns the articulatory features of Mandarin Chinese rhotics. As reviewed in previous sections, mixed results have been found regarding the tongue shapes of Mandarin rhotics. It would be interesting to see if tongue-shape variation – tip-up retroflex and tip-down bunched tongue shapes – can be found in the post-alveolar rhotic in Mandarin Chinese, as such variation has been reported in post-alveolar rhotics in other languages. Ultrasound evidence will be used to categorize the various tongue shapes used in Mandarin Chinese.

The second aim of this study concerns the effects of syllable position and vowel context on the phonetic characteristics of rhotics. Mandarin Chinese rhotics in different syllable positions have different characteristics. The prevocalic rhotic is treated as a consonant (Duanmu 2007; Lin 2007; Lee and Zee 2003; Lin and Wang 2013), while the postvocalic rhotic in the r-suffix is considered either as part of a (rhotic) vowel (Duanmu 2007; Hu 2020; Lee and Zee 2003) or as a consonant (Chao 1968; Lin 1989, 2007; Wang 1993). It is necessary to investigate the effect of syllable position and the effect of preceding and following vowels on the acoustic and articulatory features of Mandarin Chinese rhotics.

The third aim of the study relates to the frication noise in Mandarin Chinese rhotics. The existence of frication noise in the prevocalic position is largely unexplored. Examining the frication noise helps us understand the nature of the rhotic sounds. Therefore, the third strand of the research will determine the frequency of rhotic realizations that have measurable frication. By providing a systematic phonetic description of Mandarin Chinese rhotics, the current study can add further phonetic evidence to the literature on rhotics in the world's languages. It can also provide phonetic evidence that may help us resolve the debate as to the precise nature of the prevocalic rhotic.

2 Method

2.1 Participants

Eighteen native Mandarin speakers (4 males and 14 females) from Northern China were recorded using ultrasound imaging. As mentioned in Section 1.2.2, r-suffixation (*er-hua*) is a common feature of Mandarin Chinese and Mandarin dialects spoken in various places in Northern China, such as Beijing, Shandong Province and Hebei Province. All the speakers in the current study were born and grew up in these three places. Besides their Mandarin dialects, they could all speak fluent Mandarin Chinese and used r-suffixation naturally in daily communication. Their average age was 23.3 years old (Range: 21–-28, SD = 1.99) at the time of the experiment. They were post-graduate students studying in Hong Kong and the United States and also spoke English as an L2. Their English proficiency ranged from IELTS 6.5 to 8.5. The details of their English proficiency are reported in Chen et al. (2024).

2.2 Stimuli

The stimuli included words containing the prevocalic rhotic followed by the [$[1 a \times u]$ vowels, the postvocalic rhotic preceded by the [$i 1 \eta y u a \gamma$] vowels, and the syllabic rhotic (see Appendix). The Mandarin Chinese low vowel/a/ has three allophones, [a], [a] and [ɛ] (Lin 2007). Due to phonotactic constraints, the prevocalic rhotic cannot be combined with [a] and [a] without a coda. Thus, [Jan] and [Jan] were used in the current study. The tone of the tested words was not controlled for because there are some accidental gaps. For example, all words with the prevocalic rhotic bear Tone 4 except [Jan₃₅] "but" 然 which has Tone 2. This is because [Jan] cannot carry Tone 4. Sonorant-initial syllables exhibit a strong tendency to co-occur with Tone 2, a pattern that is plausibly the result of diachronic sound changes. Furthermore, in Mandarin Chinese, not every word can undergo r-suffixation. By consulting native speakers, we first ensured that words with the postvocalic rhotic were natural and frequently used in daily communication, rather than keeping the tone consistent in different vowel contexts. Therefore the stimuli with the postvocalic rhotic in different vowel contexts had different tones (Tone 1, Tone 2, Tone 3). For words with the syllabic

rhotic, we included three tones (Tone 2, Tone 3, Tone 4). This allowed post hoc comparisons examining the effect of tone on tongue shape. There are no words with the syllabic rhotic and Tone 1, and thus they were not included.

There are some commonly used minimal pairs consisting of syllabic and postvocalic /1/s in Mandarin Chinese. These minimal pairs are two-character words that are identical in segmental combination, but the syllable position of the rhotic is different. One is a compound word where the rhotic is syllabic and the word is pronounced as two syllables (such as /y₃₅.J₂₁₄/ 鱼饵'fish bait'), and the other is a word with r-suffixation where the two syllables are pronounced as only one syllable (such as /y₃₅/ 鱼儿 'small fish'). In /y₃₅.J₂₁₄/, the rhotic sound is syllabic, while in /y₃₅/, the rhotic sound is a coda (r-suffix). Five such minimal pairs were included (see Appendix) to see if syllable position and syllable boundary would affect the articulatory and acoustic features of the rhotics.

The Mandarin Chinese words were produced in the carrier phrase /tsv₅₁ kv _____ pa/ "This is ____." The carrier phrases were designed to have as little a coarticulatory effect as possible. The target words were embedded between the mid vowel /v/ and the bilabial stop /p/. The word /pa/ is a sentence final particle with a neutral tone. The [p] sound does not have any lingual target and the /a/ is pronounced as [ə] because of its particle status, so it should have a lesser coarticulatory effect on the target word. The word /tsv₅₁ kv/ "this" in the carrier phrase is a function word, so the vowel [v] is reduced and its phonetic realization is also close to a schwa [ə]. The reduced vowel quality of [kə] is very different from content words such as [v₃₅] 'goose' that have a full vowel. Therefore the carrier phrases have very little coarticulatory influence on the target words. All Mandarin Chinese prompts were presented in Chinese characters. We consulted native speakers to make sure that the words used did not have multiple pronunciations. All stimuli were randomized and repeated eight times.

2.3 Procedure

Before the experiment, all the participants were briefed about the experimental procedure and ultrasound machine and signed the consent forms. They were also asked to read through the stimulus list to familiarize themselves with the words. During the experiment, participants were seated in a soundproof booth, facing a computer screen that displayed the prompts. At the beginning of each session, speakers were asked to swallow a sip of water to make the hard palate visible, following Stone (2005). They were then asked to raise their tongue tip to touch the alveolar ridge, and then move the tongue tip back along the midline of their mouth as much as possible. These two actions were used to capture the ultrasound image of the hard palate. All speakers repeated the two actions multiple times until an image of the hard palate was clearly captured.

2.4 Ultrasound data acquisition

Two ultrasound imaging systems with the same stimuli and experimental procedure were used in this experiment. One was the Siemens ACUSON X300 ultrasound system at Haskins Laboratories with blue dots head correction (Chen et al. 2017; Noiray et al. 2020; Whalen et al. 2005). The other system was an EchoB ultrasound machine together with the Articulate Assistant Advanced (AAA) software (Articulate Instruments Ltd 2012) at the Chinese University of Hong Kong. Of the 18 speakers, 6 were recorded with the Siemens system in the Haskins Laboratories, and 12 were recorded with the EchoB system at the Chinese University of Hong Kong. The compatibility of data collected with the two systems will be discussed below.

With the Siemens ACUSON X300 system, the ultrasound probe was held on a microphone stand and positioned under the participants' chins during the experiment. The probe could move freely with the jaw. The participants were asked to look at the screen in front of them where the stimuli were presented. In order to image the midsagittal plane of the tongue, the experimenter stood in front of the participants and reminded them to avoid side-to-side head movements or rotation during the recording. The participants were asked to adjust their head position and repeat the words when there were any out-of-plane movements.

The relative position between the probe and the head was not constant. To make the ultrasound images comparable across frames, the ultrasound splines from the raw images had to be corrected according to the movements of the probe and the head. Two video cameras were positioned in front and at the side of the participants to record the front and side views of the participants' faces in order to get head movement information during recording. The head movement was represented by the movement of blue dots on the participants' heads. The blue dots' movement was tracked by a tracking algorithm implemented by the in-house MATLAB procedure DotsTracking.

For bunched gestures, the frame where the gesture reached the maximum constriction in the post-alveolar region was selected as the representative frame. For retroflex gestures, the frame with an additional bright line above the tongue surface was selected.¹ There were usually one or two frames containing a bright white line in

¹ Due to the nature of ultrasound imaging, surfaces that are parallel to the ultrasound beam are imaged poorly (Stone 2005). For retroflex gestures, the front part of the tongue is often invisible in a single ultrasound image when the tongue is curled up. In ultrasound videos, however, it can be seen that during the rhotic sound, the front part of the tongue goes up, disappears for a moment when it reaches the highest point, and then become visible again when it goes down. When the front part of the tongue becomes invisible in an ultrasound video (usually in one or two ultrasound frames), a bright white line shows up above the tongue surface (see online supplemental materials for a sample video). The white line is the reflection of the retroflexion, and this is the region where the tongue tip is expected. An example of the ultrasound image can be seen in Figure 1b. The presence of this bright

the retroflex data. If there was more than one frame with a bright white line, the frame where the bright line was closest to the post-alveolar region was selected. On the representative ultrasound frame, the tongue splines were drawn with the interactive MATLAB procedure 'GetContours' (Tiede 2018). The tongue splines were exported as 100 equally spaced data points from 'GetContours' for head movement correction. The articulatory data were collected at a frame rate of 36 frame/s.

Before the tongue spline correction, synchronization of the data from the ultrasound machine with that from two video cameras was achieved through crosscorrelation, a method that measures the temporal displacement between signals. This process enabled the calculation of the time lags between the ultrasound and the cameras. The head movement was then corrected according to the blue dot positions implemented by an optimization method implemented in MATLAB (Chen et al. 2017). The correction algorithm moved and overlapped the dots as much as possible and exported head-corrected tongue splines.

With the Echo B system, the articulatory and acoustic data were collected with the Articulate Assistant Advanced (AAA) software. The ultrasound probe was stabilized under the chin with a headset made by Articulate Instruments Ltd. to make sure that the relative position of the probe and the head was largely maintained (Articulate Instruments Ltd 2008). The software recorded ultrasound videos and audio signal, and automatically synchronized the two signals. The ultrasound videos were recorded at a frame rate of 60 frame/s. The synchronized ultrasound videos were segmented and labeled manually in AAA. A key frame where the maximum constriction could be seen was selected as the target frame of typical rhotics. The tongue splines in the key frames were manually tracked, with the aid of the "autofit" function in AAA that could automatically smooth the splines based on the ultrasound images. Each spline was exported as 124 equally spaced data points.

The splines were drawn on the lower boundary of the lighter line which represents the tongue-air interface in the ultrasound images. Figure 1a demonstrates the raw ultrasound image of a bunched tongue gesture, and the fitted splines on the ultrasound images. Due to the limitations of ultrasound imaging, it is hard to draw tongue splines of retroflex tongue shapes when part of the tongue front is not imaged well. In the current study, to illustrate the tongue shapes based on all the useful information in the ultrasound images, only the visible part of the tongue in the retroflex tongue shape was drawn, as exemplified in the right panel of Figure 1b. The position of the tongue front and tongue tip, therefore, cannot be seen in the fitted splines.

white line in the ultrasound images of retroflex tongue shapes has also been documented for the American English /1/ sound, as reported by Mielke et al. (2016) and King and Ferragne (2020).



a) An ultrasound image of a bunched tongue shape (left) and the fitted spline (right)



b) An ultrasound image of a retroflex tongue shape (left) and the fitted spline (right)

Figure 1: Raw ultrasound images of bunched and retroflex tongue shapes and the fitted tongue splines. The two images in the second column are raw ultrasound images with fitted tongue splines (red). The anterior of the tongue is on the right. (Color online).

The major differences between the Siemens and Echo B systems are the frame rates and stabilization methods. The ultrasound videos collected from the Siemens system have a frame rate of 36 frame/s, while videos from the Echo B system have a frame rate of 60 frame/s. A higher frame rate means that the ultrasound machine captures more ultrasound images per second. When producing approximants, tongue movements are relatively slow, and 30 frame/s has been shown to be sufficient to capture them (Lawson et al. 2011; Mielke et al. 2016). A faster frame rate is needed for sounds like flaps, clicks and stops because the tongue moves faster when producing them (Stone 2005). In the current study, the temporal resolution of both ultrasound systems was sufficient for the purpose of examining the tongue movements of rhotics. The higher frame rate in the Echo B system resulted in some repeated ultrasound images of the same tongue shape. As for the stabilization techniques, both methods have been proven to be efficient in maintaining the relative position between the ultrasound probe and the head, or correcting for such movements (Chen et al. 2017; Scobbie et al. 2008). Therefore, differences in frame rate and stabilization method do not influence the reliability of the data. The splines from the two ultrasound systems were never compared in one statistical model. Therefore, although two systems have been used to acquire the ultrasound data, the data from the two systems are comparable for the purposes of this study, and caution has been taken to make sure that the data analysis is legitimate.

2.5 Ultrasound data analysis

The raw ultrasound data were first visually inspected and described. The tongue shapes were then categorized as bunched or retroflex based on the tongue tip position. We classify tongue shape into these two broad types for three reasons. First, there is a controversy in previous literature regarding whether the production of Mandarin rhotics involves curling or pointing up of the tongue tip. Bunched and retroflex tongue shapes differ mainly in tongue-tip position. Therefore categorizing tongue shape into these two types can properly address the debate. Second, the twoway categorization based on tongue tip position is more practical than other categorizations. Different methods of categorization have been adopted in previous studies to describe Mandarin rhotics and retroflex consonants. King and Liu (2017) categorized the postvocalic rhotic into three categories – tip up, front up and front bunched. Xing (2021) categorized rhotic tongue shapes into two broad categories (retroflex and post-alveolar) with five sub-types (curled up, tip up, front up, flat postalveolar, domed post-alveolar). They included curled up, tip up and front up in the retroflex category. Luo (2020) used a three-way categorization (bunched, retroflex, and humped tongue shape) for retroflex consonants. Although Jiang et al. (2019b) did not categorize the tongue shapes of rhotics directly, they suggested that the tongue shapes of the syllabic and postvocalic rhotics in Northeastern Mandarin are reminiscent of the bunched and retroflex contrast of the English /J/. From a practical perspective, dividing tongue shapes into many categories with slight differences could possibly lead to a decrease in reliability of classification and a lower inter-rater consistency when there are multiple raters because the tongue shapes (especially the different bunched tongue shapes) can be quite ambiguous. It would also be less convenient for cross-study comparisons or cross-dialectal comparisons. Therefore, using a two-way categorization is a practical compromise. Third, the two-way categorization of rhotic sounds has been widely adopted by studies of other languages, such as English and Dutch (Delattre and Freeman 1968; Mielke et al. 2010, 2016; Scobbie and Sebregts 2010). English rhotics can be produced with various tongue shapes, and many different ways of categorizing English rhotics have been used (Delattre and Freeman 1968; King and Ferragne 2020; Lawson et al. 2013). But the twoway categorization is the most widely adopted one in the literature (Mielke et al. 2010, 2016; Twist et al. 2007; Zhou et al. 2008) and therefore is used in the current study.

Following previous studies, two basic criteria were used in the categorization: 1) which part of the tongue was used to make the constriction, and 2) whether the tongue tip was pointing up or pointing down (Mielke et al. 2010, 2016). While it was sometimes difficult to determine the position of the tongue tip based on a single ultrasound frame, a sequence of tongue contour movements between the segments before and after the rhotic was examined. The first author and another trained phonetician experienced in ultrasound imaging did the categorization. They initially worked separately, then discussed the different judgments together. If they had the same categorization, or they agreed with each other after their discussion, the judgment of that particular token was marked as "same". If they disagreed with each other even after the discussion, the judgment was marked as "different" and that token was not analyzed. The inter-rater agreement for all tokens in this study was 93.98 %.

Generalized additive mixed models (GAMMs) were used to quantify the tongue curves tracked from the ultrasound data using the "mgcv" packages in R (Wood 2023). GAMMs have been used to model lingual movements or tongue splines in recent ultrasound studies (such as Heyne et al. 2019). We used a customized function to visualize ultrasound splines fitted by GAMMs using "itsadug", "magrittr", "plotly" and "stringr" packages.² Polar coordinates were used to model the tongue contours because it has been proposed that tongue root position is better estimated with polar coordinates instead of Cartesian coordinates (Mielke 2015). The data points exported from "GetContours" and AAA were in Cartesian coordinates. They were converted into polar coordinates to conduct GAMMs with calculated origins for each speaker, and then converted back into Cartesian coordinates for plotting. According to Mielke (2015), the origin of the polar coordinates is an approximation of the center of an imaginary circle that corresponds to the arc of the tongue traces. For the Siemens system, the x value of the origin was the midpoint of the x values of all splines of the sample tokens for each speaker. The y value was chosen at a point where the connected lines between the origin and the tongue surface both at the root and at the tip were approaching perpendicular. For the Echo B system, the x coordinate of the origin was calculated using the formula: Depth* (1+(offset/Pixperscanline)) * cos (90-(FOV/2)) (angles in degrees), and the y coordinate of the origin was 0 (personal communication with Dr Alan Wrench).

² https://github.com/eparps/US_GAMM_visualization_polar.

2.6 Acoustic data acquisition and analysis

The audio recordings from the Siemens system were extracted from the ultrasound videos. They were segmented using forced alignment (FAVE-align) and then manually adjusted (Rosenfelder et al. 2011). The audio files from the EchoB system were exported from AAA and labeled manually. The rhotic sounds and their flanking vowels were labeled in PRAAT (version 6.0.36) (Boersma and Weenink 2017). For the prevocalic and postvocalic rhotics, we did not segment the rhotic sounds and the preceding/following vowels. The first three formants of the whole /JV/ or /VJ/ sequences were tracked at ten equidistant points. The formants were measured using linear predictive coding (LPC) in PRAAT, and the maximum formant was set as 5,000 Hz and 5,500 Hz for male and female speakers respectively. The formant values where F3 was the lowest were identified as the acoustic target of the rhotic sound, and were extracted by an R algorithm. The algorithm ensured that the minimum F3 was extracted from the first half of an /_JV/ syllable or the second half of a /V_J/ syllable. The raw formant data were plotted and visually inspected to make sure that abnormal data were excluded. The first three formant values were then transformed into the Bark scale for further analysis.

To quantify the frication noise in the prevocalic rhotic, we measured the zerocrossing rate (ZCR) for all syllables containing this sound and the subsequent vowels. ZCR is defined as the number of times the speech signal crosses zero in 1 s, calculated by dividing the number of zero-crossings by the window length. It can serve as a reliable measure of frication noise intensity, where a higher ZCR indicates increased aperiodicity. Following the approach of Shao and Ridouane (2023), we obtained upward and downward zero-crossing points in a 40 ms sliding window and analyzed 50 data points in each syllable. We modeled the time-normalized ZCR in different vowel contexts using Generalized Additive Mixed Models (GAMMs).

The first three formants and the difference between F3 and F2 (F3-F2) of rhotics were examined with linear mixed-effect models using the lmer () function from the 'lme4' package (version 1.1–21) (Bates et al. 2015). In addition to examining the first three formants, F3-F2 was also examined because this measure can partially correct the differences in speakers' vocal tracts, and thus is less influenced by individual differences in age, gender and height (McAllister and Tiede 2017). In English, the primary acoustic cues for the /I/ sound are a low F3 and a small F3-F2 difference (Delattre and Freeman 1968; Lindau 1985; Westbury et al. 1998; Zhou et al. 2008). To examine the effect of syllable position, linear mixed-effects models were conducted on F1, F2, F3 and F3-F2 with Syllable position as fixed effects, Participant and Item as random effects (both Participant and Item as a random intercept, and Participant also as a random slope). The models with best fit are presented with *p* values

calculated with the 'lmerTest' package (Kuznetsova et al. 2017), and post-hoc comparisons were done with 'emmeans' package (Lenth 2018). In addition to the static formant frequency of rhotics at the lowest F3, we also used GAMMs to analyze formant trajectories of /x/ together with adjacent vowels to get a better understanding of how these sequences were coarticulated.

3 Results

3.1 Articulatory features of Mandarin Chinese rhotics

3.1.1 Tongue shapes of Mandarin Chinese rhotics

Our ultrasound data showed that Mandarin Chinese rhotics can be produced with various tongue shapes. The tongue shapes include both tip-down bunched shapes and tip-up tongue retroflex shapes. Sample ultrasound images of the tongue shapes are shown in Figure 2.

In Type 1, the dorsum of the tongue is raised towards the palate to make the constriction. The tip of the tongue stays down and the whole tongue is retracted. Luo (2020) reported a similar tongue shape in the Mandarin Chinese sibilants /s ts ts^h/ and named it "humped tongue shape". In Type 2, the blade of the tongue is raised towards the palato-velar region to make the constriction with a concave shape at the back of



Type 1: dorsal bunched



Type 4: apical retroflex



Type 2: blade bunched with a concavity









Figure 2: Sample ultrasound images of Mandarin Chinese tongue shapes. The anterior of the tongue is on the right.

the tongue. In Type 3, the blade of the tongue is raised and the constriction is made by the blade of the tongue and the palato-velar region. In Type 4, the constriction is made by the tongue tip in the post-alveolar region. The whole tongue surface is relatively flat compared to other types. In Type 5, the tongue tip is curled back towards the post-alveolar region. A white bright line can be seen in the region where the tongue tip is expected. A detailed description of this tongue shape can be seen in Footnote 2.

To summarize, the ultrasound data showed that Mandarin Chinese rhotics can be produced with various tongue shapes ranging from tip-down bunched to tip-up retroflex. The observed tongue shape variation in Mandarin Chinese rhotics is comparable to that found in English (Delattre and Freeman 1968).

In order to simplify the discussion, the five tongue shapes were categorized as bunched or retroflex. The categorizations of tongue shape are summarized in Table 1. Recall that each stimulus was repeated eight times, and the ultrasound data showed that there was no change from one category of tongue shape to another among the eight repetitions. Therefore, only one type of tongue shape (bunched or retroflex) was determined for each stimulus per speaker.

As shown in Table 1, of the 18 speakers, 10 used bunched tongue shapes in all syllable positions and 8 used both bunched and retroflex tongue shapes. There are four main patterns. First, prevocalic rhotics were articulated only with bunched

Participant	Birthplace	Prevocalic Before /լ a ೪ u/	Syllabic	Postvocalic After /i ן ן ע u a צ/
M1	Shandong	Bunched	Retroflex	Retroflex
W1	Shandong	Bunched	Retroflex	Retroflex
M2	Shandong	Bunched	Retroflex	Retroflex
W2	Shandong	Bunched	Retroflex	Retroflex
М3	Beijing	Bunched	Bunched	Bunched
W3	Beijing	Bunched	Bunched	Bunched
W4	Shandong	Bunched	Retroflex	Retroflex
W5	Shandong	Bunched	Retroflex	Retroflex
W6	Hebei	Bunched	Bunched	Bunched
W7	Beijing	Bunched	Bunched	Bunched
W8	Shandong	Bunched	Bunched	Bunched
W9	Beijing	Bunched	Retroflex	Retroflex
W10	Beijing	Bunched	Bunched	Bunched
W11	Hebei	Bunched	Bunched	Bunched
W12	Shandong	Bunched	Bunched	Bunched
W13	Shandong	Bunched	Retroflex	Retroflex
M4	Shandong	Bunched	Bunched	Bunched
W14	Shandong	Bunched	Bunched	Bunched

Table 1: Summary of the tongue shapes of Mandarin Chinese rhotics.

tongue shapes. No tip-up tongue shapes were observed in our data. But there was within-category variation among the bunched tongue shapes. Variants of bunched and retroflex tongue shapes were found in syllabic and postvocalic positions. Second, each speaker used the same gesture (either bunched or retroflex) in syllabic and postvocalic positions. Speakers did not categorically change tongue shape in syllabic and postvocalic positions where both bunched and retroflex gestures were found. Third, the number of bunched and retroflex tongue shapes in syllabic and postvocalic positions were similar (8 speakers used retroflex gestures and 10 used bunched gestures). Fourth, the tongue shape of rhotics was not influenced by the vowel context. In prevocalic position, rhotics can be followed by $[t a \propto u]$, while in postvocalic position, they can be preceded by $[i \ 1 \ 1 \ y \ u \ a \ x]$. The tongue shapes, however, are not marked for individual vowel contexts in Table 1 because the gesture was consistent in the same syllable position across all vowel contexts.

The bunched and retroflex tongue shapes were compared using GAMMs. It is important to note that these comparisons were made despite the tongue shapes being in different syllable positions. This is because speakers used the same tongue gesture - either bunched or retroflex - across syllabic and postvocalic positions. Furthermore, the prevocalic rhotic tokens in the current study were all produced with bunched tongue shapes. Therefore, it is impossible to compare bunched and retroflex tongue shapes within the same syllable position for a single speaker. To compare bunched and retroflex tongue shapes, we compared a bunched shape in a prevocalic position with a retroflex shape in a postvocalic position. In Figure 3, the GAMMs illustrate the comparison of tongue shapes in the [1] context. The GAMMs for $[a \propto u]$ vowel contexts are available in the online supplemental materials. As mentioned earlier, ultrasound imaging does not provide a clear visualization of the tongue tip when the tongue is perpendicular to the probe. As a result, the tongue splines we have drawn only represent the middle and posterior parts of the tongue. Aside from the difference in the tongue tip, we observed significant differences in the tongue back and tongue root. The tongue root was more retracted in retroflex than in bunched tongue shapes.

For the minimal pairs, as illustrated in Table 1, the syllabic and postvocalic rhotics in the minimal pairs were produced with the same type of tongue shape by each speaker (bunched or retroflex) at the time of maximum constriction. Differences in gestural timing could potentially account for distinctions between the minimal pairs. An illustration of the tongue movements of /y.i/ and /yi/ is provided in the supplementary materials, showing these distinctions.

One point worth mentioning is that lexical tone does not seem to influence the tongue-shape category of rhotics. The three tested words with the syllabic rhotic (/ μ_{35} / "son", / μ_{214} / "ear", / μ_{51} / "two") have three different tones (Tone 2, Tone 3, Tone 4), but all speakers used the same tongue-shape category for all three words.



DE GRUYTER



prevocalic.
 postvocalic.













M2









W5



Figure 3: GAMMs comparing tongue splines of the bunched tongue shapes in prevocalic position and retroflex tongue shapes in postvocalic position in the [$_1$] vowel context. The tongue shape was extracted at the frame where the maximum constriction could be seen. The shaded areas indicate significant differences. The anterior of the tongue is on the right. The unit of the *x* and *y* axes is mm. (color online).

3.1.2 Lingual movements of Mandarin Chinese rhotics

In addition to static analyses of the Mandarin Chinese /I/, the dynamic aspects of this sound are briefly described below to provide a more comprehensive picture. In general, there were two movements, a narrowing in the post-alveolar region by curling up or bunching the anterior part of the tongue, and a lowering at the back of the tongue. The tongue movements of two representative tokens of the word / I_{214} / "ear" produced by W4 and W6 are shown in Figure 4. Figure 4a demonstrates the lingual movement of a retroflex /I/ in syllabic position by Speaker W4. When the tongue tip curved up (left panel in Figure 4a). Figure 4b shows the lingual movement of a bunched /I/ in syllabic position by Speaker W6. As the tongue front bunched up, the back of the tongue was lowered (left panel in Figure 4b), and it was raised again when the tongue front was lowered (right panel in Figure 4b).





Tongue-tip falling movements

a) A retroflex tongue shape in a syllabic rhotic produced by Speaker W4.



b) A bunched tongue shape in a syllabic rhotic produced by Speaker W6.

Figure 4: The tongue movement of retroflex and bunched rhotics produced by W4 and W6. The splines show the tongue movements over the whole word $/_{214}$ / "ear". Darker colors indicate earlier and lighter colors later in time. The frame where the tongue reached the highest point (maximum constriction in the vocal tract) is plotted in both the left and right panels to serve as a reference, meaning that the lightest tongue spline in the left panel is the same spline as the darkest spline in the right panel. The anterior of the tongue is on the right. The unit of the *x* and *y* axes is mm. (color online).

3.1.3 Effects of syllable position

Another important question is whether there are significant differences in the tongue shape of rhotics in different syllable positions. First, the tongue shapes of the 10 speakers who consistently used bunched tongue shapes in all syllable positions were compared. GAMMs comparing the tongue splines of prevocalic and postvocalic rhotics when followed/preceded by the vowel [1] are shown as an example in Figure 5. Though the tongue shapes were all categorized as bunched, they were significantly different in various places, such as tongue front or tongue back. Figure 5 also shows that some speakers showed more gestural variation than others. Similarly, significant differences in tongue splines were found between prevocalic and postvocalic rhotics in [a γ u] contexts (see online supplemental materials for the GAMMs results). Considerable inter-speaker variation in tongue shape was found for rhotics in different syllable positions. In general, the tongue root was more retracted in postvocalic than prevocalic positions.

A comparison of prevocalic and postvocalic rhotics with different tongue shapes is presented in Figure 3. It involved eight speakers who produced bunched tongue shapes in prevocalic position and retroflex tongue shapes in postvocalic position (note that no retroflex tongue shape was found in prevocalic position). We observed that the tongue root was more retracted with retroflex tongue shapes in postvocalic position compared to bunched tongue shapes in prevocalic position.

To summarize, for bunched tongue shapes, there was within-category variation in tongue shape for rhotics in different syllable positions. For both bunched and retroflex tongue shapes, for most speakers the rhotics showed more retracted tongue root in postvocalic position, indicating pharyngeal retraction in the postvocalic rhotic.

3.2 Acoustic features of Mandarin Chinese rhotics

3.2.1 Frication noise in the prevocalic rhotic

Frication noise was observed in many tokens of the prevocalic rhotic, but never in syllabic or postvocalic rhotics. Figure 6 shows representative spectrograms of the prevocalic rhotic in different vowel contexts, with and without frication noise.

Table 2 summarizes the occurrence of frication noise in the prevocalic rhotic when followed by $[\eta u \gamma]$ and the two allophones of the low vowel /a/ ([a a]) for each speaker. Frication noise was more often observed when the rhotic was followed by the high vowels ([ηu]) than low vowels ([a] and [a]). This suggests a correlation between the occurrence of frication noise and the height of the following vowel in a prevocalic context. Moreover, there was large inter-speaker variation in the use of









W11







W12







W14









Figure 5: GAMMs comparing tongue splines of the bunched tongue shapes in prevocalic and postvocalic positions in the [1] vowel context. The tongue shape was extracted at the frame where the maximum constriction could be seen. The shaded areas indicate significant differences. The anterior of the tongue is on the right. The unit of the *x* and *y* axes is mm. (color online).



Figure 6: Waveforms and spectrograms of prevocalic rhotics produced with and without frication noise followed by the vowels [a ¥]u] by speaker W1 (left column) and speaker W7 (right column). The left panel shows prevocalic rhotics without frication noise, and the right panel shows prevocalic rhotics with frication noise.

Speaker	Birthplace	l	u	۲	a	α
M1	Shandong	+	+	+	+	+
W1	Shandong	Some tokens	Some tokens	-	-	-
M2	Shandong	+	+	+	+	+
W2	Shandong	+	+	+	_	-
M3	Beijing	+	+	+	+	-
W3	Beijing	+	+	+	+	+
W4	Shandong	+	+	+	-	-
W5	Shandong	Some tokens	Some tokens	+	-	Some tokens
W6	Hebei	+	-	Some tokens	_	-
W7	Beijing	+	+	+	Some tokens	Some tokens
W8	Shandong	+	+	+	+	Some tokens
W9	Beijing	+	+	+	+	+
W10	Beijing	+	+	+	+	Some tokens
W11	Hebei	-	Some tokens	+	Some tokens	-
W12	Shandong	+	+	+	_	Some tokens
W13	Shandong	+	Some tokens	Some tokens	-	-
M4	Shandong	+	Some tokens	+	+	+
W14	Shandong	+	+	+	-	Some tokens

Table 2: Summary of frication observed in the prevocalic rhotic (+ indicates the presence of frication; – indicates the absence of frication; "some tokens" means that frication was found in some repetitions).

frication noise. Four speakers (M1, M2, W3, W9) produced frication noise consistently across all vowel contexts and in all repetitions. Speakers W10 and M4 produced frication noise in most of their tokens. In contrast, Speaker W1 produced no frication noise in most of his (or her) tokens, while Speaker W6 produced consistent frication noise only when the prevocalic rhotic was adjacent to the apical vowel [η].

The ZCR of the syllables containing the prevocalic rhotic and the following vowels h u r a α / is presented in Figure 7. The ZCR exhibited higher values at the onset



Figure 7: The mean ZCR of syllables containing the prevocalic rhotic and the following vowels, along with 95 % confidence intervals. The *x* axis represents the normalized time of the syllable and the *y* axis represents the zero-crossing times per second.

of the syllable, at approximately 10 % of its duration, indicating greater aperiodicity. The elevated ZCR at the beginning of the syllable aligns with the frication noise we observed in the spectrograms. In addition, the ZCR values were higher for the vowels [α] and [a], while they were lower for the vowel/u/. These findings are consistent with previous research, such as that of Shao and Ridouane (2023), who reported ZCRs reaching approximately 1,000 times per second for the vowel/ α / and approximately 500 times per second for the vowel/ μ /. Furthermore, we observed considerable interspeaker variability in the ZCR, reflecting the variation in frication noise production among speakers. Detailed ZCR data for individual speakers are available in the online supplementary materials.

3.2.2 Formant frequencies of Mandarin Chinese rhotics

The mean formant values of rhotics in different syllable positions are summarized in Table 3. As mentioned earlier, the minimum F3 of the whole syllable (/ $_{IV}$ /, / $_{I}$ / or /V $_{I}$ /) was taken as the acoustic target of the rhotic sound, and the formants were transformed into Bark for statistical analysis. As shown in the table, Mandarin Chinese rhotics are also characterized by a close F3 and F2.

The formant frequencies of the bunched and retroflex rhotics were compared to see if rhotics produced with different tongue shapes had any acoustic differences. Linear mixed-effects models were constructed to compare the mean first three formant values and F3-F2 of bunched and retroflex rhotics in postvocalic and syllabic positions. No significant differences between bunched and retroflex tongue shapes were found in the first three formants or F3-F2. This suggests that the articulatory variants of rhotics did not have any acoustic differences in the first three formants.

The first three formants and F3-F2 of rhotics in different syllable positions were also compared (Figure 8). Linear mixed-effect models were performed on three formants and F3-F2 to examine the effects of syllable position (prevocalic, postvocalic or syllabic) and the results are summarized in Table 4. The baseline was the prevocalic position. Results showed that there was a significant effect of syllable

	Prevocalic		Syll	abic	Postvocalic	
	Hz	Bark	Hz	Bark	Hz	Bark
F1	435	4.28	603	5.76	545	5.26
F2	1705	11.83	1,450	10.85	1,438	10.78
F3	2,321	13.97	1908	12.67	1940	12.77
F3-F2	616	2.14	458	1.82	502	1.99

Table 3: Mean formant values of Mandarin Chinese rhotics in Hertz and Bark.



Figure 8: Formant values (F1, F2 and F3) and the difference between F3 and F2 (F3-F2) of rhotics in prevocalic, postvocalic, and syllabic positions. The boxes and whiskers show the median, first, and third quartiles, and $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$. Different scales were used in the four box plots to better illustrate the differences between syllable positions.

Table 4: Linear mixed-effects model results on formant frequency values of rhotics in prevocalic, post-vocalic, and syllabic positions.

		Estimate	SE	df	t value	Pr (> t)	
F1	(Intercept)	3.602	0.282	37.985	12.794	0.000	***
	SylPositionPostvocalic	2.114	0.258	202.913	8.200	0.000	***
	SylPositionSyllabic	1.290	0.379	149.458	3.403	0.001	***
F2	(Intercept)	12.246	0.249	53.134	49.116	0.000	***
	SylPositionPostvocalic	-1.700	0.214	134.646	-7.945	0.000	***
	SylPositionSyllabic	-1.472	0.311	98.826	-4.726	0.000	***
F3	(Intercept)	14.043	0.119	31.802	117.656	0.000	***
	SylPositionPostvocalic	-1.283	0.145	31.095	-8.870	0.000	***
	SylPositionSyllabic	-1.455	0.166	35.445	-8.763	0.000	***
F3-F2	(Intercept)	2.040	0.168	31.571	12.150	0.000	***

Best Imer models: F1 ~ SyllablePosition + (1+SyllablePosition|Subject) + (1| Item); F2 ~ SyllablePosition +

(1+SyllablePosition|Subject) + (1| Item); F3 \sim SyllablePosition + (1+SyllablePosition|Subject) + (1| Item); F3-F2 \sim 1 + (1+SyllablePosition|Subject) + (1| Item); The model assumptions have been met, and p values were derived by

Satterthwaite's degrees of freedom method.

position for all three formants. Posthoc analysis showed that the F3 and F2 of prevocalic rhotics were significantly higher than those of postvocalic rhotics (F3: estimate = 1.283, Std. Error = 0.145, t = 8.829, p < 0.001; F2: estimate = 1.455, Std. Error = 0.167, t = 8.716, p < 0.001) and syllabic rhotics (F3: estimate = 1.453, Std. Error = 0.111, t = 13.042, p < 0.001; F2: estimate = 1.472, Std. Error = 0.321, t = 4.589, p < 0.001). No difference was found between postvocalic and syllabic rhotics in F2 and F3. The F1 of prevocalic rhotics was significantly lower than that of postvocalic rhotics (estimate = -2.114, Std. Error = 0.263, t = -8.047, p < 0.001) and syllabic rhotics (estimate = -1.29, Std. Error = 0.389, t = -3.315, p = 0.003). The F1 of postvocalic rhotics was also significantly lower than that of syllabic rhotics (estimate = 0.824, Std. Error = 0.309, t = 2.667, p = 0.02). No significant differences were found in F3-F2 between syllable positions. Significant differences were also found in formant trajectories between prevocalic, syllabic, and postvocalic rhotics for F1, F2, and F3, as depicted in Figure 9. Detailed data from the GAMMs are reported in Table 5.

In summary, based on our data, Mandarin Chinese rhotics are characterized by a close F3 and F2. Prevocalic rhotics have a higher F3 and F2 than syllabic and post-vocalic rhotics. In addition, there is no difference in the first three formant values between bunched and retroflex rhotics.

4 Discussion

In the current study, the articulatory and acoustic features of Mandarin Chinese rhotics were examined. There are three main findings, which answer the three research questions. The first research question was concerned with the articulatory shape of the tongue for Mandarin Chinese rhotics. Our data showed that the tip-up tongue shape is one of the variants used in the production of Mandarin Chinese rhotics. This result is consistent with King and Liu (2017), Jiang et al. (2019a), and Xing (2021). Eight out of 18 speakers in the current study used tip-up retroflex tongue shapes. Our ultrasound data also showed that Mandarin rhotics can be articulated with a continuum ranging from tongue tip-up retroflex to tip-down bunched tongue shapes. The tongue shape variation between bunched and retroflex has been found in English and Dutch, two languages that are typologically different from Mandarin Chinese. Dynamically, the articulation of rhotics involves a lowering of the tongue back and the bunching or retroflexing of the tongue tip.

The second research question was concerned with the effects of syllable position and vowel contexts on the phonetic characteristics of rhotics. Our data found that the syllabic and postvocalic rhotics can be articulated with both tongue tip-up retroflex and tip-down bunched tongue shapes, while the prevocalic rhotic is produced with



Figure 9: GAMMs comparing formant trajectories of rhotics across three syllable positions.

TrackNo

6

4

8

10

F1	Parametric coefficients:								
		Estimate	Std. Error	t value	Pr(> t)				
	(Intercept)	6.118	0.250	24.506	0.000	***			
	SylPos.ordpre	-1.841	0.105	-17.519	0.000	***			
	SylPos.ordsyl	-0.420	0.101	-4.154	0.000	***			
	Approximate significance of	f smooth terms	:						
		edf	Ref.df	F	p-value				
	s(TrackNo)	6.708	7.117	8.039	0.000	***			
	s(TrackNo):SylPos.ordpre	7.408	7.794	17.564	0.000	***			
	s(TrackNo):SylPos.ordsyl	1.001	1.001	13.123	0.000291	***			
	s(Subject)	17.951	18.000	362.906	0.000	***			
	s(TrackNo, TokenNo)	169.980	215.000	210.579	0.000	***			
F2	Parametric coefficients:								
		Estimate	Std. Error	t value	Pr(> t)				
	(Intercept)	10.711	0.212	50.430	0.000	***			
	SylPos.ordpre	0.801	0.083	9.669	0.000	***			
	SylPos.ordsyl	-0.346	0.080	-4.339	0.000	***			
	Approximate significance of smooth terms:								
		edf	Ref.df	F	p-value				
	s(TrackNo)	7.176	7.451	15.390	0.000	***			
	s(TrackNo):SylPos.ordpre	6.552	7.022	24.150	0.000	***			
	s(TrackNo):SylPos.ordsyl	1.000	1.001	27.980	0.000	***			
	s(Subject)	17.978	18.000	772.300	0.000	***			
	s(TrackNo, TokenNo)	180.665	215.000	358.080	0.000	***			
F3	Parametric coefficients:								
		Estimate	Std. Error	t value	Pr(> t)				
	(Intercept)	13.733	0.114	119.990	0.000	***			
	SylPos.ordpre	1.043	0.076	13.730	0.000	***			
	SylPos.ordsyl	-0.348	0.076	-4.570	0.000	***			
	Approximate significance of	Approximate significance of smooth terms:							
		edf	Ref.df	F	p-value				
	s(TrackNo)	6.020	6.454	32.020	0.000	***			
	s(TrackNo):SylPos.ordpre	5.703	6.234	18.980	0.000	***			
	s(TrackNo):SylPos.ordsyl	3.810	4.276	10.640	0.000	***			
	s(Subject)	17.958	18.000	416.110	0.000	***			
	s(TrackNo, TokenNo)	168.987	215.000	68.110	0.000	***			

Table 5: GAMM results of formant trajectories for F1, F2 and F3.

tongue shapes uniformly categorized as the bunched gesture (with within-category variations). For the syllabic and postvocalic rhotics, 8 speakers used retroflex tongue shapes, and 10 used bunched tongue shapes. No speaker changed tongue shape categorically between syllabic and postvocalic rhotics. The variation in articulatory tongue shape in the syllabic and postvocalic rhotics is more a matter of individual preference, and might possibly be associated with anatomical differences between speakers' palates, as suggested by Dediu and Moisik (2019). This differs from Northeastern Mandarin which has been found to have tip-down tongue shapes for the postvocalic rhotic and tip-up tongue shapes for the syllabic rhotic (Jiang et al. 2019b). In addition, vowel context does not categorically change the tongue shape of the prevocalic or postvocalic rhotics. This is consistent with what King and Liu (2017) found for the Mandarin postvocalic rhotic. In terms of acoustics, rhotics in all syllable positions are characterized by a close F3 and F2. The prevocalic rhotic has a significantly higher F3 and F2 than the syllabic and postvocalic rhotics.

The third research question was concerned with whether Mandarin Chinese rhotics consistently exhibit frication noise and have consistent formant values across syllable positions. Our data showed that the prevocalic rhotic can be articulated with or without frication noise, both within and across speakers. Tokens without frication noise are less common than tokens with frication noise, but can still be found. There is large inter-speaker variation in the use of frication noise in the production of the prevocalic rhotic.

4.1 Articulatory features of Mandarin Chinese rhotics

Our articulatory and acoustic data show that Mandarin Chinese rhotics can be articulated with various tongue shapes in syllabic and postvocalic positions, and no significant difference was found between bunched and retroflex tongue shapes in the first three formants. This seems to suggest a many-to-one relationship between articulation and acoustics for Mandarin Chinese rhotics, which has also been observed for the English /I/ (Delattre and Freeman 1968; Lindau 1985; Westbury et al. 1998).

Although the tongue shape variation in Mandarin Chinese rhotics is similar to the well-established bunched-retroflex variation in English and Dutch rhotics, our data showed that Mandarin Chinese rhotics have language-specific features in their articulation. First of all, vowel context does not categorically change the tongue shape of Mandarin Chinese rhotics. Mielke et al. (2010, 2016) suggested that in English, retroflex tongue shapes are preferred when the /x/ sound is adjacent to low or back vowels due to compatibility of gestures. Retroflexion involves retraction of the tongue body, which is also required for English low back vowel production. The tongue shapes of bunched gestures involve the raising of the tongue front, which is similar to English high front vowels. Therefore, retroflex tongue shapes are more compatible with low back vowels, while bunched tongue shapes are more compatible with high front vowels in English. In Mandarin Chinese, however, tongue shape is not affected by vowel context. If a speaker chooses to use a bunched or retroflex gesture, he/ she would continue with that tongue-shape category in spite of any segmental changes. One may argue that high vowels do not prevent retroflex tongue shapes in Mandarin Chinese because the postvocalic rhotic does not directly follow high front vowels. Phonological analysis suggests that a schwa is inserted between a high front vowel and postvocalic rhotic during r-suffixation (Lin 1989). The perceived insertion of schwa is due to the coarticulatory influence of the high front vowel on the subsequent rhotic sound. In spectrograms, the F3 gradually lowers when transiting from a high front vowel to the following /1/ sound. In terms of articulation, the inserted schwa is the result of the tongue passing through a schwa-like configuration (Gick and Wilson 2006). Phonologically, in Mandarin Chinese, this phenomenon is interpreted as a rule-driven insertion of schwa. On a phonetic level, it represents a coarticulatory pattern that is comparable across both English and Mandarin Chinese languages. Hall and Hamann (2010) also suggest that the avoidance of the rhotic plus high front vocoid sequences is a cross-linguistic phenomenon likely grounded in articulatory phonetics.

One possible reason that lingual compatibility does not influence the production of Mandarin Chinese rhotics to the same extent as English rhotics might be because high vowels in Mandarin Chinese are not as high as English high vowels. The F1 of the English high front vowel /i/ is about 270 Hz and 310 Hz for male and female speakers (Peterson and Barney 1952), while the F1 of the Mandarin Chinese /i/ is about 300 Hz and 401 Hz for male and female speakers (Zee and Lee 2001). Apical vowels in Mandarin Chinese are more like mid vowels than high vowels in terms of their acoustics and articulation (Zee and Lee 2001; Lee-Kim 2014). The tongue shape of Mandarin Chinese rhotics, therefore, is affected by the raising of the tongue front to a lesser degree than in English, resulting in fewer conflicts in articulation. In other words, high vowels in Mandarin Chinese are not as elevated, allowing better compatibility with both tongue shapes. Consequently, there is no distinct preference for either bunched or retroflex tongue shapes, providing speakers with more flexibility in their choice of which to use.

A similar phenomenon has been reported in King and Liu (2017). They examined the Mandarin postvocalic rhotic of 12 Mandarin speakers using ultrasound imaging and found little coarticulatory effect of vowels on the tongue shape of the following rhotic. The cumulative results suggest that lingual compatibility may not be as important a factor in Mandarin as it is in English. Mandarin Chinese speakers can "overcome" the incompatibility between retroflex tongue shapes and high front vowels to use their preferred tongue shape in production. In syllable positions where articulatory variation is allowed (syllabic and postvocalic positions), individual speech production strategies determine tongue shape in the production of Mandarin Chinese rhotics. Another aspect worth noting is that all the tokens for each vowel context in each syllabic position were elicited by eight repetitions of one or two test items. We cannot entirely dismiss the possibility that the null effect of vowel context may stem from a lack of variability in the stimuli. In speech perception, previous studies on categorical perception have shown that stimulus repetition may introduce a bias in the experiment results (Rogers and Davis 2009). It is plausible that the repetition of stimuli could similarly influence production outcomes. Future investigations incorporating a broader range of materials would be valuable to validate this potential impact.

Another difference is the effect of syllable position. In English, tongue-shape variation can be found in all syllable positions, and retroflex tongue shapes are more frequently found in the prevocalic /1/ than in other syllable positions (Heyne et al. 2020; King and Ferragne 2020; Mielke et al. 2010, 2016). Mielke et al. (2016) proposed that the preference for a larger anterior gesture in syllable-onset position might result in more instances of the retroflex /1/ in prevocalic position. In the current study examining Mandarin Chinese, however, tongue-shape variation can only be observed in postvocalic and syllabic positions. The prevocalic rhotic was consistently found across speakers to be produced only with bunched tongue shapes. This seems to suggest that the prevocalic rhotic is different from the postvocalic and syllabic rhotics to some extent in terms of tongue-shape variation. As reviewed in Section 1.1.1, there has been a long debate on the phonological status of the prevocalic rhotic because of its special phonetic qualities. A detailed discussion on this issue based on our articulatory and acoustic data will be provided in Section 4.2. For the syllabic and postvocalic rhotics, where both bunched and retroflex tongue shapes can be found, syllable position does not affect tongue-shape categories. Each speaker consistently used the same tongue shape for syllabic and postvocalic rhotics. Unlike our study but also using ultrasound imaging, Xing (2021) reported tip-up retroflex tongue shapes in prevocalic position in 8 out of 18 Beijing speakers. One difference between the current study and Xing (2021) is the criteria for tongue shape categorization. Xing (2021) classified Mandarin Chinese tongue shapes in various syllable positions into three categories: retroflex, bunched and post-alveolar. Within the retroflex category, three subcategories were identified – curled up, tip up, and front up, while the postalveolar category included two subcategories – flat post-alveolar and domed postalveolar. In her examination of the prevocalic rhotic, Xing (2021) observed both retroflex and post-alveolar tongue shapes. In contrast, our current study employed a more conventional two-way categorization, relying on tongue tip position as the primary criterion. Specifically, if the tongue tip pointed upward, the tongue shape was categorized as retroflex; otherwise, it was classified as bunched. Consequently, in the present study, front-up tongue shapes from Xing (2021) were reclassified as bunched, which may have played a role in the differences observed between the two studies. Another possible reason for the discrepancies between the current study and Xing (2021) is that ultrasound imaging does not provide clear visualization of the tongue tip when the tongue is perpendicular to the probe. The poor imaging quality of the tongue tip could lead to controversy regarding its position. Another possible reason is that the current study included participants from Beijing, Hebei and Shandong, while the participants in Xing (2021) were all from Beijing. Although the participants all spoke Mandarin Chinese during the experiment, we cannot entirely exclude the possibility that their Mandarin dialects may have had some influence on tongue shape.

In addition, the percentages of bunched and retroflex tongue shapes are different in Mandarin Chinese and English. According to Mielke et al. (2010, 2016), bunched tongue shapes are much more prevalent in all syllable positions in English. Of the 27 American English speakers they examined, 16 used only bunched gestures, 2 used only retroflex gestures, and 9 used a mix of bunched and retroflex gestures. However, our data show that the occurrence of bunched and retroflex tongue shapes is about equal in postvocalic and syllabic positions for Mandarin Chinese rhotics (10 bunched speakers and 8 retroflex speakers).

In addition, we also found a more retracted tongue root in postvocalic rhotics compared to prevocalic rhotics. This is consistent with the results of Lee and Zee (2014) and Xing (2021). Pharyngeal retraction has been found in rhotics in many languages such as English (Delattre and Freeman 1968; Zhou et al. 2008), Upper Sorbian and Brazilian Portuguese (Howson 2018). Our data provides evidence for the suggestion in Howson (2018) that a constriction in the pharynx achieved by retracting the tongue root might be a shared characteristic of rhotics.

4.2 Nature of the Mandarin Chinese prevocalic rhotic

In the current study, we found that the Mandarin Chinese prevocalic rhotic can only be articulated with bunched tongue shapes, which is different from the articulatory variation observed in the syllabic and postvocalic rhotics. This finding is consistent with Lee (1999) and Zhu and Mok (2023). Lee (1999) examined the tongue shapes of Mandarin prevocalic rhotics in four speakers using palatograms and linguagrams, while Zhu and Mok (2023) conducted a study involving eight Beijing Mandarin speakers and eight Japanese-Mandarin simultaneous bilinguals using ultrasound imaging. Both studies did not report any instances of retroflexion of the tongue. The current study did not find retroflexion either, even with a greater number of speakers. As for the acoustic properties, the prevocalic rhotic is characterized by a low F3, around 2,321 Hz. Moreover, frication noise was frequently observed in various vowel contexts in prevocalic rhotics before a clear formant structure characterized by a low F3. Nevertheless, speakers who did not produce frication noise in prevocalic rhotics could also be found. In summary, the prevocalic rhotic is characterized by a low F3 and an optional fricated onset. The complexity of its phonetic properties might be the reason why the prevocalic rhotic has been an "enigmatic" sound which is hard to categorize. Based on acoustic and articulatory features, the prevocalic rhotic can be categorized as a post-alveolar approximant with an optional fricated onset or as a voiced post-alveolar fricative that is devoiced in some vowel contexts. We will discuss the two types of analyses and the evidence for them below. The two categories assume the same place of articulation, but differ in manner of articulation. Both involve the approximation of the tongue tip and the post-alveolar region, with the fricative having a narrower oral constriction.

The first account is that the prevocalic rhotic is underlyingly a post-alveolar approximant, but sometimes it can be realized phonetically with additional frication at the beginning. The two phonetic realizations are free variations. This account is supported by three kinds of evidence. First, the prevocalic rhotic can be found both with and without frication noise, suggesting that frication noise is not an obligatory phonetic component in the production of this sound. Both variants are perceptually acceptable for the prevocalic rhotic. Second, frication noise is more frequently observed in high vowel contexts than in low vowel contexts. It suggests that the frication noise might be the result of a relatively higher tongue position (either bunched or retroflex) in high vowel contexts compared to other vowel contexts and caused by a coarticulatory effect. The high vowel context is a contributing factor because of the narrowing between the tongue and the hard palate in terms of aerodynamics and airflow. This differs from gestural coarticulation, as observed in Mandarin Chinese rhotics, where tongue shapes remain unaffected by vowel context. Third, this sound is characterized by a low F3, which is the primary acoustic cue for the syllabic and postvocalic rhotics, and for the approximant /x/in English. Luo (2020) showed that the retroflex fricative /s/ and retroflex affricates /ts ts^h/ did not feature a low F3. The mean F3 of /s/ reported in her study was around 3,100 Hz, while the mean F3 of the prevocalic rhotic in our study was 2,321 Hz. It indicates that the prevocalic rhotic has a much lower F3 compared to the voiceless fricative /s/. In addition, Hu (2020) examined Mandarin Chinese fricatives [s s] and the prevocalic rhotic produced by four Beijing Mandarin speakers using electropalatography (EPG). The linguopalatal contact for both the prevocalic rhotic and [s] was in the post-alveolar region, but the percentage of linguopalatal contact was smaller for the prevocalic rhotic. This provides direct evidence that the tongue makes less contact with the palate in the prevocalic rhotic than in a typical fricative with the same place of articulation, suggesting that [s] is a postalveolar fricative and the prevocalic rhotic is a post-alveolar approximant.

Another account is that the prevocalic rhotic is underlyingly a voiced fricative. This is because the prevocalic rhotic differs from the postvocalic and syllabic rhotics in both acoustics and articulation. In terms of articulation, tongue-shape variation cannot be found for the prevocalic rhotic. Though our data only represent 18 speakers from Northern China, this is consistent with previous studies using other articulatory measures (Lee 1999). The lack of variation in tongue shape suggests that the prevocalic rhotic is different from the syllabic and postvocalic rhotics, which are uncontroversially approximants. However, it needs to be noted that only 18 speakers in their twenties were tested in the current study. It is possible that retroflex tongue shapes might be found in prevocalic rhotics if more speakers, or speakers from other age groups, were examined.

In terms of acoustics, frication noise is only observed in the prevocalic rhotic and not in the syllabic or postvocalic rhotics. The lack of frication noise in some prevocalic tokens can be caused by devoicing. According to Ohala (1997), the production of voiced fricatives is more complex than that of voiceless fricatives due to aerodynamics. To produce frication noise, the oral air pressure has to be as high as possible in order to produce a large velocity of air through the oral constriction. In contrast, to produce voicing, one has to maximize the pressure difference across the glottis by keeping the oral air pressure as low as possible. Therefore, "fricatives favor voicelessness" (Ohala 1997). This is why voiced fricatives are less common than voiceless fricatives in the world's languages (Ohala 1983, 1994). Consequently, a voiced fricative might sometimes lose the frication noise because speakers fail to maintain frication, as we observed in this study.

The lowered F3 might be caused by lip protrusion. It has been suggested that lip constriction can lower F3 by increasing the volume of the front cavity (Espy-Wilson et al. 2000). Using a rather small sample size (3 speakers) and older speakers, Tiede et al. (2019) also observed lip protrusion in the retroflex consonants [s ts^h ts] and the prevocalic rhotic in Taiwan Mandarin. It is possible that the prevocalic rhotic in-volves a constriction at the lips which also causes a low F3. However, the existence of lip protrusion is not the only factor that lowers F3 so this does not necessarily lead to the conclusion that this sound is a fricative. The production of English approximant /*I*/ also involves lip protrusion (Delattre and Freeman 1968; King and Ferragne 2020; Zhou et al. 2008). Furthermore, a closer examination of the formants of prevocalic rhotics in the current study shows that F2 was not lowered. It casts doubt on the lip protrusion account because lip protrusion should also lower F2, although it was observed that increased lip protrusion for the English /*I*/ had an impact on F3 and not on F2 (King and Ferragne 2019). To further clarify the question, future studies are needed to examine the existence of lip protrusion and its influence on F3 lowering.

Although the data from the current study do not allow us to draw a firm conclusion, the approximant account seems to be more compatible with the phonetic cues of the prevocalic rhotic, while the fricative account cannot be completely ruled out.

From a phonological point of view, Natvig (2020) argues that the category "rhotic" in the phonological system is an unspecified sonorant consonant. The particular surface form of rhotic is a result of its relationship to other potential liquid

phonemes and their phonological properties, as well as the computations to make all underspecified representations pronounceable. Sebregts (2015), in contrast, advocates for characterizing rhotic variants through "family relationships". Sebregts (2015) argues for establishing diachronic links between different rhotic variants, exploring historical and geographical variations in the data. By closely examining the phonetic details of rhotics in a large corpus, particularly urban accent data, Sebregts (2015) aimed to identify the origins of specific variants, often tied to casual speech processes like lenition. Chabot (2019) proposed that the membership of rhotics is arbitrary. The cross-linguistic commonality of rhotics is procedural stability and diachronic stability. Procedural stability implies that rhotics implicated in phonological processes can vary phonetically without affecting the process itself, while diachronic stability suggests that the phonetics of rhotics can change over time without impacting their phonotactics.

In light of the perspectives outlined in these studies, it is evident that relying solely on the phonetic characteristics of Mandarin Chinese is insufficient to address the phonological status of rhotics. Investigations on diachronic and geographical variations are needed for a comprehensive understanding. Therefore, future studies could approach this question from the following perspectives to better resolve the debate. The first possibility is to check the historical development of words containing the prevocalic rhotic. The prevocalic rhotic in Mandarin originated from a nasal sound in Middle Chinese (Chinese spoken from the 6th to 10th centuries in the Southern, Northern, Sui, Tang, and Song dynasties), but it is less clear how it became oralized and gained frication (Karlgren 1915–1926; Hu 2020). The historical origins of the prevocalic rhotic seem to support the approximant account because changing from a nasal to an approximant is more natural than changing from a nasal to a fricative. Another possible line of inquiry is to examine how the prevocalic rhotic is realized in other Chinese dialects. A few studies have investigated the r-suffix in some Chinese dialects such as Liaoning dialect (Jiang et al. 2019b) and Hangzhou Wu (Yue and Hu 2019), but the prevocalic rhotic in other dialects is yet to be examined. Finally, examining this issue from a phonological perspective may bring further insights. Xing (2021) asked speakers from Beijing to produce a sequence of Vr#rV and found that speakers assimilated the prevocalic rhotic to the postvocalic rhotic in the preceding syllable. She proposed that Mandarin Chinese prevocalic and postvocalic rhotics share the same [(+)rhotic] feature.

5 Conclusions

In summary, in this study, we presented ultrasound images examining the articulatory and associated acoustic characteristics of Mandarin Chinese rhotics. As the best approximation of a definition, we examined the phonetic properties of Mandarin Chinese sounds represented by an "r" in romanization (cf. Ladefoged and Maddieson 1996). We included prevocalic, syllabic (the rhotacized vowel) and postvocalic (r-suffix) rhotics, though their phonological status might be controversial in some syllable positions. Our data showed that, in syllabic and postvocalic positions, rhotics featured variation in tongue shapes ranging from tip-up retroflex to tip-down bunched. The prevocalic rhotic was found to be produced with only bunched tongue shapes. Acoustically, rhotics were signaled with a close F2 and F3 in all syllable positions. The prevocalic rhotic was produced with optional frication noise, mixing the phonetic features of approximants and fricatives. Based on acoustic and articulatory data, the common feature of rhotics in different syllable positions was a close F2 and F3.

Acknowledgements: We would like to thank Dr. Mark Tiede and Dr. Wei-rong Chen for helping with the ultrasound imaging and ultrasound data processing at Haskins Laboratories. We also want to thank our participants for making our experiments possible.

Research funding: The study was supported by a graduate studentship and a Global Scholarship Programme for Research Excellence scholarship from the Chinese University of Hong Kong, as well as a Youth Development Program (YDP) grant (2024QQJH025) from the Chinese Academy of Social Sciences awarded to the first author. The experiment conducted at Haskins was supported by NIH grant DC-002717 to Haskins Laboratories. The Siemens ACUSON X300 system at Haskins Laboratories was available due to a generous loan agreement with Siemens Medical Solutions USA, Inc.

Author contributions: Shuwen Chen designed and conducted the experiment under the supervision of Peggy Pik Ki Mok and Douglas H. Whalen. Shuwen Chen also did the statistical analysis, and took the lead in writing and revising the manuscript. Douglas H. Whalen supervised the ultrasound experiment conducted at Haskins Laboratories, and also contributed to the writing and revision of all sections. Peggy Pik Ki Mok contributed to the conception of this research, the design of the stimuli, supervised the ultrasound experiment conducted in Hong Kong, and the writing and revision of this paper.

Conflict of interest statement: The authors have no conflicts of interest to declare. **Ethics Statement**: Signed consent forms were obtained from all participants recruited in the United States and in Hong Kong. The experiment was approved by the Yale University Human Investigation Safeguards (HIC) (No. 0706002750) and The Chinese University of Hong Kong Survey and Behavioural Research Ethics Committee.

Appendix: The stimuli

Table A1: Words with the prevocalic rhotic.

Vowel context	Word	Meaning	Chinese character
1	J [51	Sun	日
Y	J¥51	Hot	热
u	JU ₅₁	Enter	Л
a	Jan ₃₅	But	然
α	Jaŋ ₅₁	Allow	让

*/Ja/is phonotactically illegal in Mandarin Chinese.

Table A2: Words with the syllabic rhotic (rhotacized vowel).

Word	Meaning	Chinese character
ļ35	Son	儿
J 214	Ear	耳
451	Тwo	$\stackrel{\frown}{\rightarrow}$

Table A3: Words with the postvocalic rhotic (r-suffix).

Vowel context	Word	Meaning	Chinese character
i	tei155	Chicken	鸡儿
1	sאנ ₂₅	Thread	丝儿
l	t ุรุ1,4 ₅₅	Branch	枝儿
у	y1 ₅₅	Fish	鱼儿
u	huរ ₃₅	Soul	魂儿
a	pa155	Handle	把儿
Y	ky155	Song	歌儿

Table A4: Words comparing the syllabic rhotic with the diminutive suffix.

Vowel		Word (syllabic rhot	Word (syllabic rhotic)		Word (diminutive suffix)		
	IPA	Meaning	Chinese character	IPA	Meaning	Chinese character	
i	tɕ ^ʰ i₅1.斗 ₃₅	'Abandoned children'	弃儿	ts ^h iJ ₅₁	'Breath'	气儿	
у	y 35.4214	'Fish bait'	鱼饵	y1 ₅₅	'Small fish'	鱼儿	
u	tşu ₅₅ .4 ₂₁₄	'Pig's ear'	猪耳	tşuរ ₅₅	'Pearl'	珠儿	
۲	tşγ ₅₁ .μ ₂₁₄	Zhe'er (A proper name)	浙尔	tşɣɹ ₅₁	'Here'	这儿	
	X¥35.451	He'er (A person's name)	何二	XX1 ³²	'Small boxes'	盒儿	
a	tş ^h a ₅₅ .µ ₅₁	'Missing two (of something)'	差二	tş ^h aı ₅₅	'Cross'	叉儿	

References

- Articulate Instruments Ltd. 2008. Ultrasound stabilisation headset users manual: Revision 1.4. Edinburgh, UK: Articulate Instruments Ltd.
- Articulate Instruments Ltd. 2012. Articulate assistant advanced user guide: Version 2.14. Edinburgh: Articulate Instruments Ltd.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using Ime4. *Journal of Statistical Software* 67(1). 1–48.
- Boersma, Paul & David Weenink. 2017. *PRAAT: Doing phonetics by computer. Version 6.0.36*, http://www. praat.org.(accessed November 2017).

Chao, Yuen Ren. 1948. Mandarin primer. Cambridge, Massachusetts: Harvard University Press.

- Chao, Yuen Ren. 1968. *A Grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.
- Chabot, Alex. 2019. What's wrong with being a rhotic? Glossa: A Journal of General Linguistics 4(1). 1-24.
- Chen, Wei-Rong, Mark Tiede & Shuwen Chen. 2017. An optimization method for correction of ultrasound probe-related contours to head-centric coordinates. In *Paper presented at ultrafest VIII*. Potsdam: University of Potsdam, 4-6 October.
- Chen, Shuwen, D. H. Whalen & Peggy Pik Ki Mok. 2024. Production of the English/J/by Mandarin–English bilingual speakers. *Language and Speech* 0(0). https://doi.org/10.1177/00238309241230895.
- Chinese Ministry of Education. 1958. Scheme for the Chinese phonetic alphabet. http://www.moe.gov.cn/ ewebeditor/uploadfile/2015/03/02/20150302165814246.pdf (accessed 28 July 2021).
- Chuang, Yu-Ying, Sheng-Fu Wang & Janice Fon. 2015. Cross-linguistic interaction between two voiced fricatives in Mandarin-Min simultaneous bilinguals. In *The scottish Consortium for ICPhS 2015 (ed.), Proceedings of the 18th international Congress of phonetic sciences*. Glasgow, UK: the University of Glasgow.
- Dediu, Dan. & Scott R. Moisik. 2019. Pushes and pulls from below: Anatomical variation, articulation and sound change. *Glossa: a journal of general linguistics* 4(1). 7.
- Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American R's by X-ray motion picture. *Linguistics* 6. 29–68.
- Department of Chinese at Peking University. 2003. *Hanyu Fangyin Zihui [A collection of the sounds in Chinese dialects]*. Beijing: Language and Culture Press.
- Duanmu, San. 2007. The Phonology of standard Chinese. Oxford: Oxford University Press.
- 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.
- Fu, Maoji. 1956. 北京话的音位和拼音字母 [Phonemes and Pinyin symbols in the Beijing speech]. 中国 语文[Zhongguo Yuwen] 5. 309–326.
- Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34(1). 49–72.
- Gick, Bryan & Ian Wilson. 2006. Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.), *Phonology and phonetics*. Berlin, New York: Mouton de Gruyter.
- Hall, Tracy Alan & Silke Hamann. 2010. On the cross-linguistic avoidance of rhotic plus high front vocoid sequences. *Lingua* 120(7). 1821–1844.

- 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 Dorreen & Kevin Watson. 2020. The articulation of / J/ in New Zealand English. Journal of the International Phonetic Association 50(3). 366–388.
- Howson, Phil J. 2018. A phonetic examination of rhotics: Gestural representation accounts for phonological behaviour. Toronto: University of Toronto PhD dissertation.
- Howson, Phil J. & Philip J. Monahan. 2019. Perceptual motivation for rhotics as a class. Speech Communication 115. 15–28.
- Hu, Fang. 2020. 元音研究 [The vowel: A general introduction with reference to Chinese data]. Beijing: Foreign Langauge Teaching and Research Press.
- Huang, Jing, Feng-fan Hsieh & Yueh-chin Chang. 2020. Er-suffixation in southwestern Mandarin: An EMA and ultrasound study. *Interspeech*. 661–665.
- Jiang, Song, Yueh-chin Chang & Feng-fan Hsieh. 2019a. A cross-dialectal comparison of Er-suffixation in Beijing Mandarin and northeastern Mandarin: An electromagnetic Articulography study. In Paper presented at HISPhonCog 2019: Hanyang international Symposium on phonetics & cognitive Sciences of language, 24–25. Seoul, Korea: Hanyang University.
- Jiang, Song, Yueh-chin Chang & Feng-fan Hsieh. 2019b. An EMA study of er-suffixation in Northeastern Mandarin monophthongs. In Sasha Calhoun, Paola Escudero, Marija Tabian & Paul Warren (eds.). Proceedings of the 19th international Congress of phonetic sciences, 3617–3621. Melbourne, Australia.

Karlgren, Bernhard. 1915-1926. *Etudes sur la phonologie chinoise*. Upsala: K.W.Appelberg.

- King, Hannah & Emmanuel Ferragne. 2019. The contribution of lip protrusion to Anglo-English/r/: Evidence from hyper- and non-hyperarticulated speech. In *Proceedings of interspeech*, 3322–3326. Graz, Austria: Interspeech.
- 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. 1–19.
- King, Hannah & Anqi Liu. 2017. An ultrasound and acoustic study of the rhotic suffix in Mandarin. In *Paper* presented at ultrafest VIII. Potsdam: University of Potsdam, 4-6 October.
- Klein, Harriet B., Tara McAllister Byun, Lisa Davidson & Maria I. Grigos. 2013. A multidimensional investigation of children's/r/productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology* 22(3). 540–553.
- Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

Ladefoged, Peter & Ian Maddieson. 1996. The sounds of the world's languages. Oxford: Blackwell.

- Lawson, Eleanor, James M. Scobbie & Jane Stuart-Smith. 2011. The social stratification of tongue shape for postvocalic/r/in Scottish English. *Journal of Sociolinguistics* 15(2). 256–268.
- Lawson, Eleanor, James M. Scobbie & Jane Stuart-Smith. 2013. Bunched/r/promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics* 41(3). 198–210.
- Lawson, Eleanor, Jane Stuart-Smith & James M. Scobbie. 2018. The role of gesture delay in coda/r/ weakening: An articulatory, auditory and acoustic study. *The Journal of the Acoustical Society of America* 143(3). 1646–1657.
- Lee, Wai-Sum. 1999. An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin. In *Proceedings of the 14th international congress of phonetic sciences*, 413–416. San Francisco, CA, USA: The Regents of the University of California.
- Lee, Wai-Sum. 2005. A phonetic study of the 'er-hua' rimes in Beijing Mandarin. In *Proceedings of the 9th European Conference on speech Communication and technology*, 1093–1096. Lisbon, Portugal.

- 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 C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The Handbook of Chinese linguistics*, 369–399. Hoboken: John Wiley & Sons, Inc.
- Lee-Kim, Sang-Im. 2014. Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3). 261–282.
- Lenth, Russell. V. 2018. Emmeans: Estimated marginal means, aka LeastSquares means. R package version 1.1 https://CRAN.R-project.org/package=emmeans.
- Liao, Rongrong & Feng Shi. 1987. An experimental study on the sound quality of the r consonant of Mandarin Chinese. *Language Research* 02. 146–160. [汉语普通话r声母音质的实验研究. 语言研 究(02),146-160].
- Li, Yanrui. 1996. 论普通话儿化韵及儿化音位. [Discussion on Mandarin Er-hua and its phoneme]. 语文 研究 [Yuwen Yanjiu] 59. 21-26.
- Lin, Baoqin. 1992. 普通话的儿化 [Er-hua in Mandarin Chinese]. 语言文字应用[Yuyan Wenzi Yingyong] 4. 91–94.
- Lin, Xi & Lijia Wang. 2013. 语音学教程(增订版) [A Course in Phonetics (2nd edition)]. Beijing: Peking University Press.
- Lin, Yen-Hwei. 1989. *Autosegmental treatment of segmental process in Chinese phonology*. Austin: University of Texas PhD dissertation.
- Lin, Yen-Hwei. 2007. *The Sounds of Chinese with audio CD*, 1st ed. Cambridge, UK: Cambridge University Press.
- Lindau, Mona. 1985. The Story of/r. In Victoria A. Fromkin (ed.), *Phonetic linguistics: Essays in honor of peter ladefoged*, 157–168. New York: Academic Press.
- Luo, Shan. 2020. Articulatory tongue shape analysis of Mandarin alveolar–retroflex contrast. *The Journal of the Acoustical Society of America* 148(4). 1961–1977.
- McAllister, Byun & Tara & Mark Tiede. 2017. Perception-production relations in later development of American English rhotics. *PLoS One* 12(2). e0172022.
- Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America* 137(5). 2858–2869.
- Mielke, Jeff, Adam Baker & Archangeli Diana. 2010. Variability and homogeneity in American English/r/ allophony and/s/retraction. *Laboratory Phonology* 10. 699–730.
- Mielke, Jeff, Adam Baker & Archangeli Diana. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex/*J*. *Language* 92(1). 101–140.
- Mok, Peggy P. K. 2013. Does vowel inventory density affect vowel-to-vowel coarticulation? Language and Speech 56(2). 191–209.
- Natvig, David. 2020. Rhotic underspecification: Deriving variability and arbitrariness through phonological representations. *Glossa:A Journal of Generallinguistics* 5(1). 48.1–28.
- Noiray, Aude, Jan Ries, Mark Tiede, Elina Rubertus, Catherine Laporte & Lucie Ménard. 2020. Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & Optical Linguo-Labial Articulation Recording system. *Laboratory Phonology* 11. 14.
- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.). *The Production of speech*, 189–216. New York: Springer-Verlag.
- Ohala, John J. 1994. Speech aerodynamics. In Ron E. Asher & J. M. Y. Simpson (eds.). The Encyclopedia of Language and linguistics, 4144–4148. Oxford: Pergamon.
- Ohala, John J. 1997. Aerodynamics of phonology. In *Proceedings of 4th seoul international conference on linguistics*, 92–97. Seoul: Linguistic Society of Korea.

- Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24(2). 175–184.
- Rogers, Jack. C. & Matthew H. Davis. 2009. Categorical perception of speech without stimulus repetition. In *Proceedings of interspeech 2009*, 376–379. Brighton, UK: Interspeech.
- Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Yuan Jiahong. 2011. FAVE (forced alignment and vowel extraction) program suite.
- Scobbie, James M. & Koen Sebregts. 2010. Acoustic, articulatory, and phonological perspectives on allophonic variation of/r/in Dutch. In Raffaella Folli & Christiane Ulbrich (eds.), *Interfaces in linguistics: New research perspectives*, 257–277. Oxford: Oxford University Press.
- Scobbie, James M., Alan A. Wrench & Marietta L. van der Linden. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In 8th international Seminar on speech production, 373–376. Strasbourg: INRIA.
- Sebregts, Koen. 2015. *The sociophonetics and phonology of Dutch r*. Utrecht: Utrecht University PhD dissertation.
- Shao, Bowei & Rachid Ridouane. 2023. On the nature of apical vowel in Jixi-Hui Chinese: Acoustic and articulatory data. *Journal of the International Phonetic Association*. 1–26. https://doi.org/10.1017/ s0025100322000196.
- Smith, James Gordon. 2010. Acoustic properties of English/l/and/J/produced by Mandarin Chinese speakers. Toronto: University of Toronto MA thesis.
- Stone, Maureen. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7). 455–501.
- Tiede, Mark K. 2018. GetContours (version 1.3). Retrieved from https://github.com/mktiede/GetContours.
- Tiede, Mark K., Wei-rong Chen & D. H. Whalen. 2019. Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and ultrasound. In Sasha Calhoun, Paola Escudero, Marija Tabian & Paul Warren (eds.). *Proceedings of the 19th international Congress of phonetic sciences*, 427–431. Melbourne, Australia: Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Tiede, Mark K., Suzanne E. Boyce, Carol Y. Espy-Wilson & Vincent L. Gracco. 2010. Variability of North American English/r/production in response to palatal perturbation. In Ben Maassen & Pascal van Lieshout (eds.). Speech motor control: New developments in basic and applied research, 53–68. Oxford: Oxford University Press.
- Twist, Alina, Adam Baker, Jeff Mielke & Archangeli Diana. 2007. Are 'covert'/J/allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics* 13(2). 207–216.
- Wang, Jiali. 2005. 儿化规范综论 [An integrated discussion of er-hua]. 语言文字应用[Yuyan Wezi Yingyong] 3. 46–54.
- Wang, Zhijie. 1993. *The geometry of segmental features in Beijing Mandarin*. Newark: University of Delaware PhD dissertation.
- Westbury, John R., Michiko Hashi & Mary J. Lindstrom. 1998. Differences among speakers in lingual articulation for American English/r. *Speech Communication* 26. 203–226.
- Whalen, D. H., Khalil Iskarous, Mark K. Tiede, David J. Ostry, LeHouillier Heike Lehnert, Bateson Eric Vatikiotis & Donald S. Hailey. 2005. The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research* 48(3). 543–553.
- Wood, Simon. 2023. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation (R package version 1.8-42). https://cran.r-project.org/web/packages/mgcv/index.html.
- Xing, Kaiyue. 2021. *Phonetic and phonological perspectives on rhoticity in Mandarin*. Manchester: University of Manchester PhD dissertation.
- Yuan, Jiaye. 1960. 汉语方言概要 [An overview of the Chinese dialects]. Beijing: Wenzi Gaige Press.

- Yue, Yang & Fang Hu. 2019. Phonetics and phonology of the -er suffix in the Hanzhou Chinese dialect. In Sasha Calhoun, Paola Escudero, Marija Tabian & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 2056–2060. Melbourne, Australia: Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Zee, Eric & Wai-Sum Lee. 2001. An acoustical analysis of the vowels in Beijing Mandarin. In Pual Dalsgaard, Børge Lindberg, Henrik Benner & Zheng-hua Tan (eds.), *EUROSPEECH 2001 scandinavia: 7th European Conference on speech Communication and technology*, 643–646. Aalborg: Eurospeech.
- Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland & Choe Ann. 2008. A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English/r. *The Journal of the Acoustical Society of America* 123(6). 4466–4481.
- Zhu, Xiaonong. 2007. 近音——附论普通话话日母 [About approximant A supplementary discussion on the r-initial in Standard Chinese]. 方言[Dialect] 01. 2–9.
- Zhu, Zhiqiang & Peggy Pik Ki Mok. 2023. The production of Mandarin/r/by early and late Japanese-Mandarin bilinguals: Articulatory and acoustic findings. In Radek Skarnitzl & Jan Volín (eds.), Proceedings of the 20th international Congress of phonetic sciences, 2850–2854. Prague: International Phonetic Association.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/phon-2023-0023).