Language and Speech

Production of the English /ɹ/ by Mandarin–English Bilingual Speakers

Language *and* Speech
1–38

DOI: 10.1177/00238309241230895
journals.sagepub.com/home/las
Sage

Shuwen Chen
Institute of Linguistics, Chinese Academy of Social Sciences, China

D. H. Whalen
The City University of New York, USA; Yale University, USA; Haskins Laboratories, USA

Peggy Pik Ki Mok
The Chinese University of Hong Kong, China

Abstract

Rhotic sounds are some of the most challenging sounds for L2 learners to acquire. This study investigates the production of English rhotic sounds by Mandarin–English bilinguals with two English proficiency levels. The production of the English /ɹ/ by 17 Mandarin–English bilinguals was examined with ultrasound imaging and compared with the production of native English speakers. The ultrasound data show that bilinguals can produce native-like bunched and retroflex gestures, but the distributional pattern of tongue shapes in various contexts differs from that of native speakers. Acoustically, the English /ɹ/ produced by bilinguals had a higher F3 and F3–F2, as well as some frication noise in prevocalic /ɹ/, features similar to the Mandarin /ɹ/. Mandarin–English bilinguals did produce language-specific phonetic realizations for the English and Mandarin /ɹ/s. There was a positive correlation between language proficiency and English-specific characteristics of /ɹ/ by Mandarin–English bilinguals in both articulation and acoustics. Phonetic similarities facilitated rather than hindered L2 speech learning in production: Mandarin–English bilinguals showed better performance in producing the English /ɹ/ allophones that were more similar to the Mandarin /ɹ/ (syllabic and postvocalic /ɹ/s) than producing the English /ɹ/ allophone that was less similar to the Mandarin /ɹ/ (prevocalic /ɹ/). This study contributes to our understanding of the mechanism of speech production in late bilinguals.

Keywords

English /ɹ/, Mandarin–English bilinguals, articulation, ultrasound imaging, acoustics

Corresponding author:
Peggy Pik Ki Mok, Department of Linguistics and Modern Languages, Leung Kau Kui Building, The Chinese University of Hong Kong, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

Introduction

Rhotic sounds in the world's languages have diverse phonetic properties (Ladefoged & Maddison, 1996; Lindau, 1985). Rhotic sounds, such as the English /ɹ/, are complex in production and difficult to acquire for children and non-native speakers (Aoyama et al., 2004; Bradlow, 1997; Gick et al., 2007; Sheldon & Strange, 1982). In addition to cross-linguistic interference, one of the factors contributing to the difficulty of learning /ɹ/ is its complexity in articulation. The production of /ɹ/ requires more than one constriction in the vocal tract (Delattre & Freeman, 1968; Zhou et al., 2008). Moreover, various lingual tongue shapes can be adopted, from tongue tip-up to tip-down, to produce similar sounds that can be identified as the English /ɹ/. Mandarin Chinese also has the /ɹ/ phoneme, but the phonetic realization of the Mandarin /ɹ/ differs from the English /ɹ/ in both articulation and acoustics. This study aims to investigate how L2 sounds that appear phonemically in both languages but are phonetically different from L1 sounds are produced, and whether similarities between the L1 and L2 sounds would facilitate or hinder L2 production and perception.

1.1 Acoustic and articulatory features of the English /ɹ/

One of the most well-known characteristics of the English /ɹ/ sound is that it can be produced by various tongue shapes, ranging from tongue-tip-down bunched to tongue-tip-up retroflex tongue shape (Delattre & Freeman, 1968; Hagiwara, 1995; King & Ferragne, 2020; Mielke et al., 2010, 2016; Tiede et al., 2010; Westbury et al., 1998; Zhou et al., 2008). The tongue shape variation has been reported in both rhotic and non-rhotic varieties of English, such as American English, British English, Scottish English, and New Zealand English (Delattre & Freeman, 1968; Heyne et al., 2020; King & Ferragne, 2020; Lawson et al., 2011, 2013, 2018). The tongue shape variations in Scottish English are associated with different levels of rhoticity and show social stratification (Lawson et al., 2011, 2013, 2018). The continuum of tongue shapes shares the common feature that they all involve three supraglottal constrictions—a narrowing at the lips achieved by lip-rounding and protrusion, an oral constriction in the palatal region made by the tongue tip or tongue front, and a narrowing in the pharyngeal cavity made by the tongue root retracting toward the pharyngeal wall (Delattre & Freeman, 1968; Zhou et al., 2008).

The tongue shape of the American English /ɹ/ is influenced by syllable position and the flanking segments (Mielke et al., 2010, 2016; Westbury et al., 1998). The retroflex tongue shape is more common in prevocalic position than in postvocalic position, more common next to back and/or low vowels than high and/or front vowels, and also more common in labial clusters than in lingual clusters. These patterns seem to suggest that the lingual gestures of low back vowels and non-lingual labial consonants are more compatible with retroflexion than high front vowels or dorsal consonants are, because the retroflexion in the American English involves retraction of the tongue body, which is the same as the lingual gestures of low back vowels (Mielke et al., 2010). The articulatory gestures of non-lingual labial consonants do not involve any lingual movements, and would not hinder retroflexion. This pattern has been reported in New Zealand English as well (Heyne et al., 2020).

Although the American English /ɹ/ can be produced with various tongue shapes, the acoustic output is stable, showing a many-to-one articulation–acoustics relationship. The most salient acoustic feature of the English /ɹ/ is a low F3, approaching or even merging with F2 (Boyce & Espy-Wilson, 1997; Delattre & Freeman, 1968; Hagiwara, 1995; Westbury et al., 1998). The retroflex and bunched variants have similar patterns in the first three formants (Delattre & Freeman, 1968; Lindau, 1985; Westbury et al., 1998; Zhou et al., 2008), with difference only in F4 and F5

(Zhou et al., 2008). The distance between F4 and F5 was larger in /ɹ/ produced with a retroflex tongue shape than in /ɹ/ produced with a bunched tongue shape. Perceptually, bunched and retroflexed /ɹ/s are shown to be almost indistinguishable for native English listeners and Mandarin learners of English (Twist et al., 2007). Unlike American English, the British English /ɹ/ has a labiodental variant, and the labiodentalisation results in a much higher F3 than the typical British English /ɹ/ (Foulkes & Docherty, 2000).

1.2 First language acquisition of the English /ɹ/

The acquisition of the English /ɹ/ is challenging. Previous studies have shown that children failed to reach a mastery level with 90% accuracy or higher by 8-year-olds (Smit et al., 1990). Idemaru and Holt (2013) tested four groups of native English children: 4-year-olds, 4.5-year-olds, 5.5-year-olds, and 8.5-year-olds. They showed that the production of the English /ɹ/ by 4-year-olds was correctly recognized by adult listeners with 76.8% accuracy, and the accuracy reached 97.5% for the 8.5-year-old group. In addition, the production of the English /ɹ/ developed unequally in different phonological contexts. McGowan and colleagues (2004) showed that children acquire postvocalic and syllabic /ɹ/ earlier than prevocalic /ɹ/. Chung and Pollock (2021) found that rhotic sounds developed earlier in stressed positions, as in "her" or "bird," than in unstressed positions, as in "tiger" or "zipper," and earlier in /ɪɚ/ ("ear") and /ɑɚ/ ("are") than in /ɔɚ/ ("or").

The English /ɹ/ is also one of the most commonly misarticulated sounds for English-speaking children (Preston & Edwards, 2007; Shriberg & Kwiatkowski, 1994; Smit et al., 1990). Early studies on children's misarticulation of /ɹ/ usually described the errors in terms of phoneme substitution. The consonantal /ɹ/ was substituted by /w/, whereas the syllabic /ɹ/ was substituted by /ə/ and /ɔ/ (Dalston, 1975). Later studies, however, suggested that children's errors were not phonemic substitutions but de-rhotacized or distorted segments (Shriberg & Kent, 2003). Adults' production of the English /ɹ/, despite the variation in tongue shapes, involves three vocal tract constrictions: a constriction in the palatal region, a narrowing in the pharyngeal cavity, and a narrowing at the lips (Zhou et al., 2008). Gick et al. (2007) found that children's misarticulation has only one constriction, and their production improved by learning to involve two constrictions first, and finally learning to produce a canonical /ɹ/. This study also found that among the three constrictions, the tongue root constriction at the pharyngeal area is the hardest to acquire. Klein and colleagues (2013) examined the tongue shapes of the misarticulated English /ɹ/ by two native English-speaking children. They found that the tongue shape of the inaccurately articulated English /ɹ/ sounds lacked a tongue root constriction at the pharyngeal area. It is consistent with the conclusion from Boyce et al. (2011) who reviewed the ultrasound images of 37 children with persistent /ɹ/ misarticulation. Knight et al. (2007) examined the development of /ɹ/ in a speaker of Standard Southern British English (SSBE) between the ages of 3.8 and 3.11. They showed that the participant not only lowered F3 but employed various compensatory strategies to approach an adult-like /ɹ/, such as raising F2 and increasing the amplitude of F3.

1.3 Second language acquisition of the English /ɹ/

The English /ɹ/ is also one of the most challenging sounds for second language learners. The most well-studied example is Japanese learners of English (Bohn & Flege, 1992; Boyce et al., 2016; Flege, 1992; Goto, 1971; Jun & Cowie, 1994; Munro et al., 1996). Many studies have shown that Japanese speakers had great difficulties in producing the English /ɹ/ accurately (Bradlow, 1997; Bradlow et al., 1999; Goto, 1971; Sheldon & Strange, 1982). According to the Speech Learning

Model (SLM) (Flege, 1995, 2003), the difficulties in acquiring the English /ɹ/ are due to the assimilation of the English /ɹ/ and /l/ to the same Japanese category (Japanese alveolar tap /ɾ/, labio-velar approximant /w/ or Japanese high back vowel /ɯ/) as equivalent, and hence Japanese speakers produced them identically. Although it is a big challenge for Japanese speakers, they could show some improvement in producing the English /ɹ-l/ contrast after perceptual training (Bradlow, 1997; Bradlow et al., 1995, 1999).

Researchers also examined the production of the English /ɹ/ by speakers from other language backgrounds. Polish has an alveolar trill. Lyskawa (2015) examined the production of the English /ɹ/ by five Polish learners using ultrasound imaging. She found that Polish learners failed to make native-like tongue shapes when producing the English /ɹ/. Their tongue shapes were retroflex-like, and no typical bunched tongue shapes were found. Also, Harper et al. (2016) examined the production of the English /ɹ/ by French and Greek learners using real-time magnetic resonance imaging (MRI). The French (L1) rhotic sound was produced with a pharyngeal constriction that was higher than the constriction in the English /ɹ/, whereas the Greek (L1) rhotic consonant did not involve a pharyngeal constriction. Their results found that French speakers failed to produce an English-like low pharyngeal constriction when producing L2 English /ɹ/, whereas Greek speakers failed to produce any pharyngeal constriction at all in L2 English /ɹ/. As mentioned, studies examining children who learned English as their first language found misarticulation to be caused by the lack of a pharyngeal constriction. Based on acoustic data, Smith (2010) examined the production of the English /ɹ/ by Mandarin–English bilingual speakers who have lived in Canada for an average of 10.4 years. He showed that Mandarin learners had no problem producing the English /ɹ-l/ contrast, but their production of the English /ɹ/ was still significantly different from native English production, despite their immersion experience in an English-speaking country for around 10 years.

1.4 The Mandarin rhotics

The Mandarin rhotics exhibit distinct phonetic realizations in different syllable positions. The prevocalic rhotic occurs in the syllable-initial position of a syllable (e.g., /ɹɤ$_{51}$/热 “hot”), the syllabic rhotic occurs in the syllable nucleus position (e.g., /ɚ$_{35}$/儿 “son”), and the rhotic that functions as a suffix (e.g., /kɤɹ$_{55}$/歌儿“song”) occurs postvocalically.

When it is syllable-initial, which is usually called “r-initial” according to the tradition of Chinese phonology, together with the post-alveolar fricative and affricates /ʂ/, /tʂ/ and /tʂ^{h}/, the Mandarin prevocalic rhotic is usually referred to as a “retroflex consonant” in the literature and classroom settings (Chao, 1968; Duanmu, 2007). Phonetically, this sound is post-alveolar and is distinguished by a low F3 (Chen, 2020[1]; Chen & Mok, 2021; Hu, 2020; Lee, 1999). The Mandarin “r-initial” has been transcribed as a post-alveolar approximant [ɹ] (or [r] for ease of typing and printing; Fu, 1956; Lin, 2007), an apical post-alveolar approximant with a subscript indicating apical features [ɻ̩] (Lee, 1999; Lee & Zee, 2003), or a post-alveolar voiced fricative [ʐ̩] (Duanmu, 2007; Karlgren, 1915–1926; Wu & Lin, 1989; Yuan, 1960). The variation in the phonetic notation arises from the ongoing debate surrounding whether this consonant is a voiced fricative /ʐ̩/ or an approximant /ɹ/. This debate stems from the presence of frication noise in some of the r-initial tokens. The Mandarin “r-initial” variants with and without frication noise are both observed (Chen, 2020; Chuang et al., 2015; Lee, 1999; Liao & Shi, 1987), and there exists significant inter-speaker and intra-speaker variability in the production of frication.

The tongue shapes of the Mandarin prevocalic rhotic have been a subject of considerable interest. According to Chao (1968), the Mandarin rhotic sounds were believed to involve curling up of the tongue tip. However, Lee (1999) conducted articulatory analyses using palatograms and

linguograms by four native speakers of Beijing Mandarin and did not find evidence of tongue tip curling in their production of the prevocalic rhotic. Chen (2020), using ultrasound imaging on 18 Mandarin speakers from Northern China (with the data from the same speakers used in the current study, excluding one participant), also did not observe tip-up retroflex tongue shapes in the prevocalic rhotic. However, retroflex tongue shapes were observed in non-prevocalic rhotics (syllabic and postvocalic). Conversely, also using ultrasound imaging, Xing (2021) reported tip-up retroflex tongue shapes in 8 out of 18 Beijing speakers. One difference between the two studies is the divergence in classification methods and the respective categories employed. Chen (2020) used the position of the tongue tip as the primary criterion, categorizing tongue shapes as retroflex when the tip pointed upward and as bunched otherwise. In the study by Xing (2021), Mandarin tongue shapes in different syllable positions were classified into three distinct categories: retroflex (inclusive of three subcategories—Curled Up, Tip Up, and Front Up), bunched, and post-alveolar (encompassing two subcategories—flat post-alveolar and domed post-alveolar). Both retroflex and post-alveolar tongue shapes were identified in the Mandarin prevocalic rhotic in Xing's (2021) data. The Front Up tongue shapes in the study by Xing (2021) were classified as bunched in Chen's (2020) study, potentially accounting for the disparities between the two investigations. Another possible reason for the discrepancies between the studies by Chen (2020) 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 ambiguity regarding its position. It is also possible that Chen's study did not include enough speakers with retroflex tongue shapes (although both studies had 18 speakers).'

A closely related study by Luo (2020) examined other Mandarin initial "retroflex consonants" /ʂ tʂ tʂʰ/ using ultrasound imaging, although not including the prevocalic rhotic. Luo (2020) found various tongue shapes for Mandarin /ʂ tʂ tʂʰ/, but only one instance of tip-up tongue shape was found among the 162 tokens examined. In summary, evidence from articulatory studies tends to suggest that the Mandarin r-initial does not involve a tongue-tip-up gesture. However, further articulatory studies are required to reach a consensus on whether the production of Mandarin r-initial involves tip-up retroflex tongue shapes.

When the Mandarin rhotic occurs in the syllable nucleus position, it is transcribed as a rhotacized vowel [ɚ] (Duanmu, 2007; Lee & Zee, 2003; Zee & Lee, 2001), a syllabic post-alveolar approximant [ɻ̩] (Lin, 2007), or sometimes as a mid-central vowel followed by a post-alveolar rhotic approximant [əɹ] ([ər]) (Lin, 2007; Lin & Wang, 2013).

The Mandarin rhotic can also function as a suffix and merge with the preceding vowel in a process called r-suffixation or "er-hua" (儿化). This is a common feature of Mandarin dialects spoken in Northern China (Wang, 2005). The r-suffix is used as a diminutive suffix or to indicate familiarity with objects (Li, 1996; Lin, 1992). In this case, the rhotic sound is analyzed either as a rhotic feature of the preceding vowel (Lin & Wang, 2013; Wang, 1993), or as a postvocalic approximant (Lin, 1989, 2007). R-suffixation is a complex morphophonological process that involves syllable restructuring. Some rhymes undergo segmental changes after r-suffixation, such as monophthongization of diphthongs, glide insertion, or the deletion of a nasal coda (Duanmu, 2007; Lin, 2007).

The Mandarin rhotic in non-prevocalic position (syllabic and postvocalic rhotics) exhibits a characteristic low F3 (Chen, 2020; Hu, 2020; Lee, 2005). In terms of articulation, most studies, except Lee (2005), have reported that Mandarin syllabic and postvocalic rhotics can be produced with both tip-up and tip-down tongue shapes. Using electromagnetic articulography (EMA) data from three Beijing speakers (one male and two females), Lee (2005) found no evidence of retroflexion in their articulation. Conversely, other studies by Wu and Lin (1989), King and Liu (2017), Jiang et al. (2019a), Chen (2020), and Xing (2021) have consistently reported both tip-up and

tip-down tongue shapes in the production of Mandarin syllabic and postvocalic rhotics. King and Liu (2017) used ultrasound imaging to examine postvocalic rhotic production in 12 native Mandarin speakers, revealing various tongue shapes including tip-up, front-up, and front bunched configurations for Mandarin r-suffix. Jiang et al. (2019a) also observed tip-up tongue shapes in three Beijing Mandarin speakers using EMA. Chen (2020) conducted a comprehensive ultrasound imaging study involving 18 Mandarin speakers from Northern China (Beijing, Hebei, and Shandong), revealing a continuum of tongue shapes ranging from tip-up retroflex to tip-down bunched configurations for Mandarin syllabic and postvocalic rhotics. As mentioned earlier, this study utilizes the data from the same participants, with one participant excluded. Among the 18 participants in Chen's study, 8 used tip-up retroflex tongue shapes while the remaining 10 used bunched tongue shapes. Another ultrasound study by Xing (2021) reported a higher prevalence of retroflex tongue shapes, with 15 out of 18 Beijing speakers utilizing retroflex tongue shapes. Interestingly, Chen (2020) revealed that speakers consistently employed the same tongue shapes, either bunched or retroflex, in their production of syllabic and postvocalic rhotics. In addition, the tongue shape of Mandarin syllabic and postvocalic rhotics is not categorically influenced by vowel context. This is in contrast to American English where tongue shapes are affected by vowel context and syllable position. In terms of tongue movement dynamics, Mandarin syllabic and postvocalic rhotics involve 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).

Sub-dialectal variation has also been observed in the pronunciation of Mandarin rhotics. In Northeastern Mandarin, the postvocalic rhotic (r-suffix) is produced with a tip-down bunched gesture, whereas the syllabic rhotic sound (the rhotacized vowel [ɚ]) is produced with a tip-up tongue shape based on EMA data from three speakers (Jiang et al., 2019b). Huang et al. (2020) examined two female speakers with EMA and found that syllabic and postvocalic rhotics are exclusively produced with bunched tongue shapes in Southwestern Mandarin spoken in the western Hubei Province (Huang et al., 2020). Similarly, based on ultrasound data from 10 speakers, retroflex tongue shapes are not found in syllabic and postvocalic rhotics in Taiwan Mandarin (Huang et al., 2022).

1.5 This study

It has been well-established that L2 phonemes that do not exist in the native sound inventory pose great challenges for L2 learners, as exemplified by the learning of the English /ɹ/ by Japanese learners (Bohn & Flege, 1992; Boyce et al., 2016; Flege, 1992; Goto, 1971; Jun & Cowie, 1994; Munro et al., 1996). However, even for L2 sound categories that exist as phonemes in one's native language, it is still possible that the L2 sounds are incorrectly produced because the phonetic realizations of the shared phonemes might differ in the L1 and L2. A related issue is whether the phonetic similarities between L1 and L2 sounds would facilitate or hinder the acquisition process. This study aims to investigate the production of this type of sound by examining the production of the English /ɹ/ by Mandarin–English bilinguals.

As introduced in previous sections, Mandarin and English both have rhotic sounds, but there are also phonetic differences between them. The main similarities and differences between the Mandarin /ɹ/ and the English /ɹ/ are summarized in Table 1.

The first aim of this study is to examine how the English /ɹ/ is produced by Mandarin–English bilinguals. There are two logical possibilities in the production of this type of sound. First, bilinguals might fully copy the L1 sounds when producing the L2 sounds, transferring all phonetic features of L1 sounds into L2. Second, bilinguals can produce the L1 and L2 sounds differently,

Table 1. Articulatory and Acoustic Differences Between the English and Mandarin /ɹ/s.

	English	Mandarin
Articulation	1. Bunched and retroflex tongue shapes are found in all syllable positions 2. The tongue shapes are influenced by syllable position, vowel contexts and syllable structure	1. Both bunched and retroflex tongue shapes are used in the Mandarin non-prevocalic rhotics 2. Although there is no unanimous agreement on the presence of retroflex tongue shape in the production of Mandarin prevocalic rhotic, several studies suggest that retroflexion appears to be infrequent or is not consistently observed 3. The tongue shapes are not affected by vowel contexts
Acoustics	The English /ɹ/ is characterized by a low F3	1. The Mandarin /ɹ/ sound is also characterized by a low F3, but the F3 of the Mandarin /ɹ/ is higher than that of the English /ɹ/ 2. Frication noise was found in many tokens of prevocalic /ɹ/

with their L1 and L2 production approaching the phonetic targets in each language. SLM and the Revised Speech Learning Model (SLM-r) (Flege, 1995, 2003; Flege & Bohn, 2021) propose that if L2 learners can perceive the differences between L1 and L2 sounds, a new sound category would be established for the L2 sound. Therefore, if Mandarin–English bilinguals can perceive the differences between the Mandarin and English /ɹ/, they are predicted to produce /ɹ/ differently in English and Mandarin because a new category would be established for the English /ɹ/. The Perceptual Assimilation Model-L2 (PAM-L2) (Best & Tyler, 2007) is not directly designed to address questions in speech production, but its hypotheses can be extended to the production domain given that PAM-L2 assumes shared primitives in speech perception and production. According to PAM-L2, if only one L2 category is perceived as equivalent to the L1 category, learners will map the L2 sound onto the closest L1 category. Therefore, the English /ɹ/ will be directly mapped onto the Mandarin /ɹ/ at the phonological level. However, PAM-L2 also briefly mentioned that L2 learners might learn the different phonetic realizations for the same phonological category in each language. Bilinguals can have two different phonetic categories for L1 and L2 sounds under the common phonological category. PAM-L2, therefore, predicts that Mandarin–English bilinguals would map the English /ɹ/ onto the Mandarin /ɹ/ phonologically, and the language-specific phonetic details might still be learned. But it is not clearly predicted how and to what extent language-specific phonetic realizations can be learned.

The second aim is to investigate the influence of phonetic similarities on L2 sound acquisition. According to SLM and SLM-r, the greater the phonetic dissimilarity between the L2 category and its closest L1 category, the easier it is for L2 learners to discern the phonetic difference. Larger phonetic dissimilarity leads to the formation of a new phonetic category for the L2 sound, and hence a more native-like production. On the contrary, similarities between L1 and L2 sounds would hinder the establishment of a new L2 category because the phonetic differences are easily ignored by L2 learners. L2 learners might simply use the L1 category to substitute the L2 sound category, producing accented L2 sounds. The acquisition of the English /ɹ/ by Mandarin–English bilinguals provides a valuable opportunity to test this hypothesis. The positional allophones of the Mandarin /ɹ/ differ from those of the English /ɹ/ to varying degrees. The Mandarin prevocalic /ɹ/ is most dissimilar from the English /ɹ/ because it involves frication noise, and it has higher F3 and F2 than the

English /ɹ/. If Lee (1999) and Chen (2020) are correct about the tongue shapes of the Mandarin prevocalic rhotic, the Mandarin /ɹ/ and the English /ɹ/ further differ in that the Mandarin prevocalic rhotic does not show a binary variation between bunched and retroflex tongue shapes. The realizations of Mandarin syllabic and postvocalic /ɹ/s are more similar to those of the English /ɹ/ than those in the initial position, as they can be produced with both bunched and retroflex tongue shapes. The articulatory difference lies in the distribution of the two tongue shapes. The distribution of bunched and retroflex tongue shapes in English is conditioned by syllable position, vowel context, and syllable structure, whereas in Mandarin, it is conditioned only by syllable position. Besides articulatory differences, Mandarin syllabic and postvocalic /ɹ/s are also different from English syllabic /ɹ/ acoustically as they have a higher F3. Therefore, according to SLM, the English prevocalic /ɹ/ is easier to acquire than the syllabic and postvocalic /ɹ/s because Mandarin and English prevocalic /ɹ/ share fewer phonetic similarities than those of syllabic and postvocalic /ɹ/s.

The third aim is to investigate how the production of L2 sounds correlates with differences in the success of language learning. Many studies have shown that, with increasing L2 experience and proficiency, most bilinguals showed more native-like performance in speech production, such as more intelligible productions and more native-like accents (Best & Strange, 1992; Flege et al., 1996; Ingvalson et al., 2011; MacKain et al., 1981; Takagi & Mann, 1995). SLM and SLM-r, however, predict that high-proficiency bilinguals do not necessarily show more native-like production than low-proficiency bilinguals do when producing sounds that are similar in L1 and L2. SLM predicts that it would be easy for L2 learners to learn at the beginning stage because they can simply use the L1 categories to substitute the L2 sounds. But it might be challenging for more advanced learners when they want to achieve native-like performance by producing and perceiving subtle differences between the L1 and L2 sound categories.

To summarize, the /ɹ/ sound exists phonemically in both English and Mandarin, but it is realized differently in the two languages. This study aims to answer three research questions: (1) How is the English /ɹ/ produced by Mandarin–English bilinguals? Can Mandarin–English bilinguals produce language-specific phonetic realizations for Mandarin and English rhotics? (2) How does phonetic similarity between the English and Mandarin /ɹ/ affect L2 sound production? (3) How does the production of the English /ɹ/ change when Mandarin–English bilingual speakers' English proficiency improves?

2 Method

2.1 Participants

Seventeen Mandarin–English bilingual speakers (3 male and 14 female) and 16 American English native speakers (5 male and 11 female) participated in this study. Among the 17 Mandarin–English bilingual speakers, 5 were recorded in the United States and 12 were recorded in Hong Kong. They were all postgraduate students who could use English for academic purposes and daily communication. The bilinguals were all born and grew up in Northern China (5 from Beijing, 10 from Shandong Province, and 2 from Hebei Province). As mentioned in Section 1.4, r-suffixation is a common feature of Mandarin spoken in various regions of Northern China. Therefore, all participants in this study naturally used Mandarin r-suffixation in their daily communication, whether speaking Standard Mandarin or regional Mandarin dialects. Their average age was 23.3 years old (Range: 21–28, $SD = 1.99$). We also made sure that the participants learned and spoke with a rhotic English accent.

Table 2. Language Background of Mandarin–English Bilingual Speakers.

	Ultrasound system	Participant ID	Age	Sex	Total score in Standard test	Speaking score in Standard test	Age of acquisition (AOA)	Birthplaces
High-proficiency bilinguals	Siemens (United States)	H1	22	F	7	7	6	Shandong
		H2	23	M	7.5	6.5	10	Shandong
		H3	23	F	7.5	7	5	Shandong
		H4	23	M	7	6.5	6	Beijing
		H5	21	F	7.5	7.5	3	Beijing
	EchoB (Hong Kong)	H6	22	F	7	8	12	Shandong
		H7	23	F	7	6.5	8	Shandong
		H8	22	F	7.5	7.5	3	Hebei
		H9	28	F	8.5	8.5	3	Beijing
		H10	21	F	7.5	7	5	Shandong
		H11	22	F	7.5	7	4	Beijing
Low-proficiency bilinguals	EchoB (Hong Kong)	L1	26	F	6.5	6	4	Beijing
		L2	22	F	6.5	5.5	10	Hebei
		L3	22	F	6.5	6	7	Shandong
		L4	25	F	6.5	6	6	Shandong
		L5	24	M	6.5	5.5	6	Shandong
		L6	23	F	6.5	6	8	Shandong

The Mandarin–English bilinguals were divided into two groups—high-proficiency and low-proficiency—according to their scores in English language tests. As a proxy for participants' proficiency in oral English, the participants were also asked to report their speaking scores in the standard tests. The reported scores of standard tests included TOEFL (Test of English as a Foreign Language) and IELTS (International English Language Testing System) scores. For easy comparison, the TOEFL scores were converted into IELTS scores according to the official guideline from the Educational Testing Service (ETS, 2010; Papageorgiou et al., 2015). We use an IELTS score of 7 as a threshold. Speakers with a total of 7 or higher were grouped as the high-proficiency group. The details of the language proficiency and age of acquisition (AOA) are shown in Table 2. The average overall score is 7.41 ($SD=0.44$) out of 9 for the high-proficiency group, and 6.50 ($SD=0$) for the low-proficiency group.[2] The average speaking score of the high-proficiency group is 7.18 ($SD=0.64$) out of 9, and 5.83 ($SD=0.26$) for the low-proficiency group. The differences in IELTS scores between the two proficiency groups may not seem very large, but the actual proficiency levels are. According to IELTS score descriptors and IELTS speaking band descriptors,[3] the highest score of IELTS is 9 (native proficiency), and those with 6 cannot speak and write very well in English. The Mann–Whitney–Wilcoxon test showed that there are significant differences between the high- and low-proficiency groups in the total score ($w=66$, $p<.001$) and in the speaking score ($w=66$, $p<.001$). The average AOA of the high- and low-proficiency groups is 5.91 ($SD=2.98$) and 6.83 ($SD=2.04$) years old, respectively.

Although the bilingual participants in this study learned and spoke a rhotic English accent, one speaker in the low-proficiency group failed to produce rhotic sounds consistently. She only produced prevocalic /ɹ/ and postvocalic /ɹ/ after /a ɛ/, but failed to produce syllabic /ɹ/ and postvocalic /ɹ/ after /i u ɔ/. Her data were used in the study, with her production of English syllabic /ɹ/ marked as "no /ɹ/."

Sixteen monolingual English speakers were recorded reading English words using ultrasound imaging. As the articulation of the English /ɹ/ has been well-investigated in previous studies (Delattre & Freeman, 1968; Hagiwara, 1995; Mielke et al., 2010, 2016; Westbury et al., 1998), only the acoustic data were used in this study for a direct comparison of formant frequencies between native English production and bilingual production. Ten speakers were recorded with the Siemens system, and the other six were recorded with the EchoB system. The monolingual English speakers had an average age of 22.31 (Range: 19–28, *SD*=2.68), and spoke a rhotic accent of English. They all had very limited exposure to Mandarin or Cantonese. The ten speakers in the Siemens group were living in the United States and had no previous knowledge of either Mandarin or Cantonese. The six speakers in the EchoB group had very limited exposure to Mandarin Chinese or Cantonese as they were exchange students at the Chinese University of Hong Kong (CUHK) at the time of recording, but they did not learn Mandarin or Cantonese in any formal settings.

2.2 Stimuli

The English stimuli included 28 words containing prevocalic and postvocalic English /ɹ/ coarticulated with the /ɑ æ ɛ ɪ ɔ u ʌ/ vowels, and syllabic /ɹ/ (see Appendix A). The /ɹ/ sound was embedded in different syllable positions—20 prevocalic /ɹ/ in /#_V(C)/, /C_V(C)/ and /CC_V(C)/ words, 5 postvocalic /ɹ/ in /V_#/ words, 3 syllabic /ɹ/. The target words were produced in the carrier sentence "What a ___ again" when the word started with a consonant, and "Speak of ___ again" when the word started with a vowel.

Mandarin stimuli were included to compare the bilingual production of the English and Mandarin /ɹ/. Mandarin stimuli included 21 words containing prevocalic /ɹ/ coarticulated with the /ɭ̩ a ɤ u/ vowels, postvocalic /ɹ/ with the /i ɻ ɭ̩ y u a ɤ/[4] vowels, and syllabic /ɹ/ (see Appendix B). The Mandarin low vowel /a/ has three allophones /a/, /ɑ/, and /ɛ/ (Lin, 2007). The allophone /ɛ/ only occurs between /j y/ and the dental nasal /n/, such as in /jɛn/ "eye." The Mandarin prevocalic /ɹ/ cannot be combined with /a/ and /ɑ/ without a coda, so /ɹan/ and /ɹɑŋ/ were used. The Mandarin words were produced in the carrier phrase /tʂɤ kɤ ___ pa/ "This is ___." (/pa/ is a sentence-final particle in Mandarin).

The carrier phrases were designed to have as little coarticulatory effect as possible. In English, the target words were embedded between two schwas or between labiodental fricative [v] and a schwa. The [p] and [v] sounds do not have any lingual target, so they should have a lesser coarticulatory effect on the target word. The schwa in English has a relatively central tongue shape and has a smaller coarticulatory effect than other vowels. In Mandarin, the target words were embedded between the mid vowel /ɤ/ and the bilabial stop [p]. /tʂɤ51 kɤ51/ "this" in the carrier phrase is a function word, so the vowel in /kɤ51/ is reduced, and its phonetic realization is close to a schwa [ə]. The reduced vowel quality of /kɤ51/ is very different from content words such as [ɤ35] "goose" with a full vowel. Therefore, the carrier phrases have a very small coarticulatory influence on the target words. Both Mandarin and English speakers read the stimuli in their native languages. All stimuli were randomized and repeated eight times.

2.3 Procedure

The experiment included two sessions for Mandarin speakers—a Mandarin session and an English session, and only one English session for native English speakers. Before the experiment, the participants were briefed about the experiment procedure and ultrasound machine, making sure that speakers feel comfortable speaking with an ultrasound probe under their chin. They were also asked to read through the stimuli list to familiarize themselves with the words. During

the experiment, participants were seated in a sound-proof booth, facing a computer screen that displayed the prompts. At the beginning of each session, speakers were asked to swallow a sip of water. 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. Those two actions were used to capture the ultrasound image of the hard palate. All speakers repeated the two actions multiple times until the image of the hard palate was clearly captured. Participants were also instructed to say the peripheral vowels /a/, /i/ and /u/ out naturally to pinpoint their vowel space. In the experiment sessions, speakers read out the prompts shown on a computer screen. All Mandarin speakers started with the Mandarin session. Mandarin speakers had a 5-min break between the Mandarin and English sessions.

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 CUHK. Among the 17 Mandarin speakers, 5 speakers were recorded with the Siemens system in the Haskins Laboratories, and 12 were recorded with the EchoB system at CUHK. Among the 16 American English speakers, 10 speakers were recorded with the Siemens system, and the other 6 were recorded with the EchoB system. 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 the participants put their chins on the probe while talking. The probe was free to move with the jaw. The participants were asked to look at the screen in front of them where the stimuli were presented. 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. All ultrasound images with out-of-plane movements were excluded.

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 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 to get head movement information during recording. The head movement was represented by the movement of blue dots on the participants' heads and tracked by a tracking algorithm implemented by an in-house MATLAB procedure DotsTracking. The head movement was then corrected according to the blue dot positions using an optimization method in MATLAB (Chen et al., 2017). The details of the head movement correction procedure can be found in the Supplementary Materials.

For bunched gestures, the frame where the gesture reached the maximal constriction at the post-alveolar region was selected as the representative frame. For the retroflex tongue shape, part of the tongue front can be invisible in some ultrasound videos, and a bright white line shows up above the tongue surface. The white line is the reflection of the retroflexion, and this is the region where the tongue tip is expected (King & Ferragne, 2020; Mielke et al., 2016). So the frame with a bright white line above the tongue surface was selected as the representative frame for the retroflex tongue shape. There would be one or two frames containing the bright white line in the retroflex data. If there was more than one frame having a bright white line, the frame where the bright line was closest to the position of the post-alveolar region was selected. On the representative ultrasound frame, the tongue splines were drawn with an 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.

With the EchoB system, the articulatory and acoustic data were collected with the AAA software. The ultrasound probe was stabilized under the chin with a headset made by Articulate Instruments Ltd. to ensure that the relative position of the probe and the head was maintained (Articulate Instruments Ltd, 2008). The software recorded ultrasound videos and audio signals, 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 maximal 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. The splines were drawn on the lower boundary of the lighter line that represents the tongue–air interface in the ultrasound images. Each spline was exported as 124 equally spaced data points.

The major differences between the Siemens and EchoB systems are the frame rates and stabilization methods. The ultrasound videos collected from Siemens system have a frame rate of 36 frame/s, whereas videos from the EchoB system have a frame rate of 60 frame/s. A larger frame rate means that the ultrasound machine captures more ultrasound images in a second. When producing approximants, the tongue movements are relatively slow, and 30 frame/s frame rate has been shown to be sufficient to capture the tongue movements (Lawson et al., 2011; Mielke et al., 2016). In this study, the temporal resolution of both ultrasound systems was sufficient for the purpose of examining the tongue movements of Mandarin rhotic sounds. The higher frame rate in the EchoB system resulted in some consecutive images with the tongue in the same position so two adjacent frames might look very similar. 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 would not influence the reliability of the data. Also, the tongue shapes were first categorized as bunched or retroflex, and then compared with each other. The splines from the two ultrasound systems were never compared in one statistical model. Therefore, although two systems have been used in acquiring the ultrasound data, the data from the two systems are comparable for the purpose of this study, and caution has been taken to make sure that the data analysis is legitimate.

2.5 Ultrasound data analysis

The analysis of ultrasound data consists of two parts: tongue shape categorization and quantitative analyses of tongue splines. English and Mandarin tongue shapes were categorized as either bunched or retroflex. If the tongue tip was curling up, the tongue shape was categorized as retroflex; and if the tongue tip was pointing down, it was categorized as bunched, following the study by Mielke et al. (2010, 2016). Although it was sometimes difficult to tell the position of the tongue tip based on a single ultrasound frame, a sequence of tongue contour movements from the preceding segment to the following segment before and after the rhotic sound was examined. The first author and another trained phonetician experienced in ultrasound imaging did the categorization. The raters first did the categorization separately, and then discussed the different judgments with each other. If they had the same categorization, or agreed with each other after the discussion, the judgment of a particular token was marked as "same". The inter-rater agreement for all tokens after the discussion was 95.19%. If the raters disagreed with each other even after the discussion, the judgment was marked as "different", and that token was not analyzed.

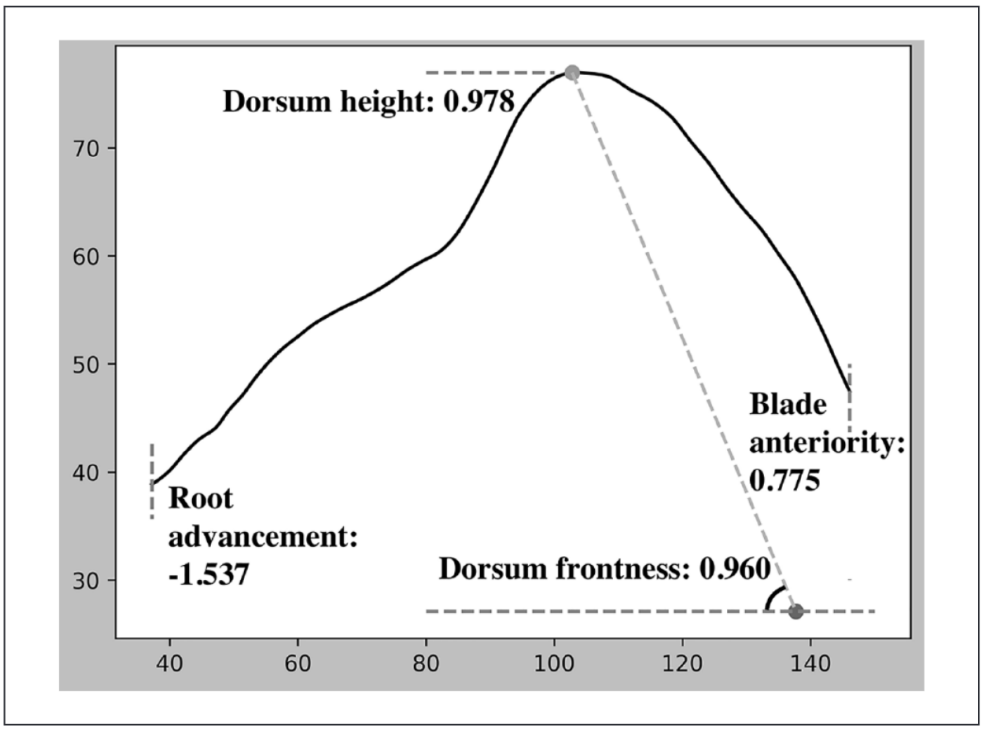


Figure 1. Illustration of the tongue spline measurements.

Smoothing spline analysis of variants (SSANOVAs) were used to quantify the tongue curves tracked from the ultrasound data using the "gss" packages in *R* (Gu, 2014). SSANOVA models the differences between two or more curves, and was widely used in comparing tongue splines in ultrasound studies (Ahn, 2018; Kochetov et al., 2014; Lee-Kim, 2014; Mielke, 2015). The result of SSANOVA is a plot containing average curves of each group of data and the 95% Bayesian confidence interval around the curves. If there are portions of the curves during which the confidence intervals do not overlap, it means that the two curves are significantly different there. Polar coordinates were used to model the tongue contours because it has been proposed that the tongue root position was better estimated with polar coordinates instead of Cartesian coordinates (Mielke, 2015). The data points exported from ultrasound videos were in Cartesian coordinates. They were first converted into polar coordinates to conduct SSANOVA analyses with the calculated origin for each speaker, and then converted back to Cartesian coordinates for plotting. The origin of the fan corresponds to the position of the probe during imaging.

To further quantify the tongue shapes, we measured tongue dorsum height, tongue dorsum frontness, tongue blade anteriority, and tongue root advancement following Hussain and Mielke (2021). Tongue dorsum height was measured by calculating the vertical distance from the origin to the highest point on the tongue dorsum. It was then normalized by dividing by each speaker's maximum dorsum height. For each speaker, a specific starting point (origin) was chosen, with its y-coordinate at 1% of the y range below the lowest point of all tongue trace and its x-coordinate at 1% of the x range to the right of the smallest maximum x value. To quantify tongue dorsum frontness, the angle in radians from the origin to the highest point on the tongue spline was measured. A larger value for dorsum frontness suggests a more fronted tongue dorsum. Tongue blade anteriority was measured as the x-value of the first point (most posterior point) on the tongue spline, whereas tongue root advancement was the x-value of the last point (most anterior point).

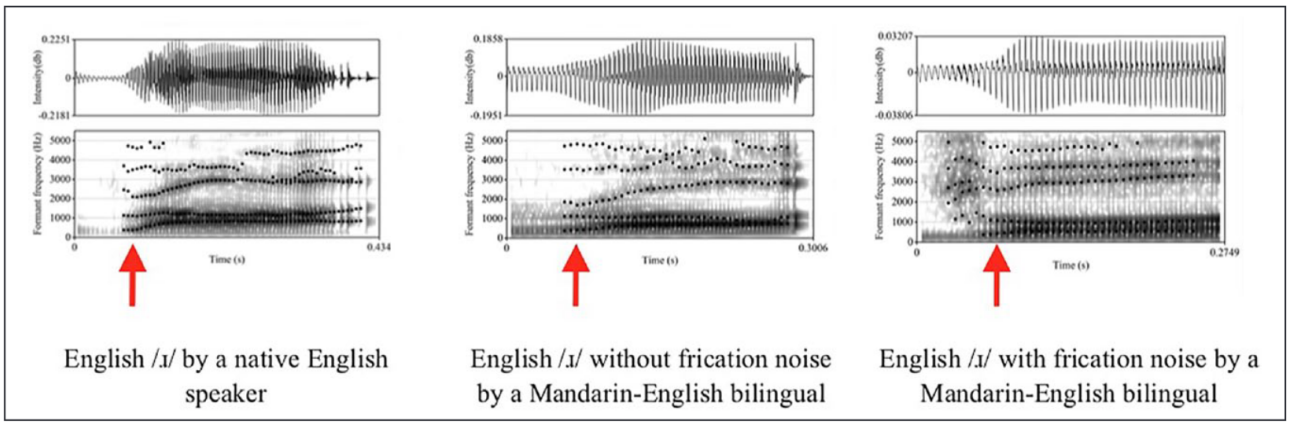


Figure 2. English word "raw" produced by an English native speaker and two bilingual speakers with and without frication noise. The red arrows point to the position where the formant values are analyzed.

Both of these measures were normalized using *z*-scores. The four tongue spline measurements are illustrated in Figure 1.

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; Rosenfelder et al., 2011) and then manually adjusted. The audio files from the EchoB system were exported from AAA and labeled manually. The rhotic sounds and the flanking vowels were labeled in PRAAT (Boersma, 2002). For prevocalic and postvocalic rhotics, we did not segment the rhotic sounds and the preceding/following vowels. The first three formants of the whole syllable (/ɹV/ or /Vɹ/) were tracked at 10 equidistant points for each syllable. The formants were measured using linear predictive coding (LPC) in PRAAT, and the maximum formant was set as 5,000 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. For the tokens where frication noise can be seen, we measured them at the point where the F3 reaches its minimum following the frication noise because the formants at the beginning of the syllable may not be reliably tracked due to the presence of frication noise. Figure 2 exemplifies the spectrograms of the English /ɹ/ produced by native English speakers and bilingual speakers with and without the frication noise, and the acoustic target of rhotic sound. The R algorithm ensured that the minimum F3 was extracted from the first half of an /ɹV/ syllable or the second half of a /Vɹ/ syllable. The raw formant data were plotted and visually inspected to make sure that no abnormal data were included. The first three formant values were then transformed into a Bark scale for further analysis.

The occurrence of frication noise was decided for each token and the frequency of occurrence was calculated by each participant for each stimulus, as each stimulus was repeated eight times. The percentage of occurrence was examined with linear mixed-effects models using the lmer() function from the "lme4" package (version 1.1-21) (Bates et al., 2015).

To quantify the frication noise more comprehensively, we measured the zero-crossing rate (ZCR) for all English and Mandarin syllables. ZCR is a measure commonly used to quantify the level of frication noise or turbulence in a speech signal. It indicates how frequently the amplitude of the signal crosses the zero point. It is calculated by dividing the count of zero-crossings by the window length. ZCR can provide information about the noisiness or frication in speech sounds, with a higher ZCR reflecting increased aperiodicity. As demonstrated by Shao and Ridouane (2023), apical vowels (also

known as fricative vowels) in Jixi-Hui Chinese exhibit a high ZCR at the syllable onset, corresponding to frication noise at the beginning of apical vowels. Following Shao and Ridouane's (2023) approach, we assessed points where the speech signal crossed zero in both upward and downward directions using a 40 ms sliding window, and examined 50 data points for every syllable. As will be elaborated in Section 3.2.1, frication noise was observed exclusively in word-initial /ɹ/ and not in /ɹ/ in consonant clusters, syllabic, or postvocalic contexts. Therefore, we computed the ZCR specifically for word-initial /ɹ/ in both English and Mandarin. The time-normalized ZCR was then modeled using Generalized Additive Mixed Models (GAMMs) through the "mgcv" package (Wood, 2023) and visualized using the "itsadug" package in R (van Rij et al., 2023).

For acoustic data, the primary measure for rhoticity was F2 and F3. A lower F3 and a smaller difference between F3 and F2 (F3–F2) indicate stronger rhoticity (McAllister Byun & Tiede, 2017). F3–F2 difference was proposed in some studies to better capture rhoticity because it partially corrected for the difference in the speaker's vocal tract, and thus was less influenced by individual differences in age, gender and height (McAllister Byun & Tiede, 2017). The first three formants and F3–F2 of the Mandarin and English /ɹ/s were examined with linear mixed-effects models using the lmer() function.

Linear mixed-effects models were conducted on F3, F2, and F3–F2 with Group (Native English, High-proficiency bilinguals, Low-proficiency bilinguals, Native Mandarin[5]) and Syllable position (Prevocalic, Postvocalic, Syllabic) as fixed effects, Participant and Item as random effects (both Participant and Item as a random intercept, and Participant as a random slope), and Syllable position as a random slope for each participant. The model with the best fit is presented with *p*-values calculated with the "lmerTest" package (Kuznetsova et al., 2017), and the post hoc comparisons were done with the "emmeans" package (Lenth, 2018).

3 Results

3.1 Articulatory characteristics of L2 English /ɹ/

3.1.1 Mandarin and English /ɹ/ tongue shapes by bilingual speakers. The tongue shapes of Mandarin and English rhotic sounds produced by the bilinguals were categorized as either bunched or retroflex shapes. Table 3 summarizes the tongue shapes used for the Mandarin and English /ɹ/s by each speaker. The categorization of English tongue shapes in each vowel context for all bilinguals can be found in the Supplementary Materials.

The bilinguals can be divided into three groups based on their tongue shapes. The first group of speakers categorically changed their tongue shape when switching from Mandarin to English (Speaker H6 postvocalic /ɹ/, H7 prevocalic /ɹ/, H11 syllabic /ɹ/). For example, Speaker H7 used a bunched tongue shape for Mandarin prevocalic /ɹ/, and changed to a retroflex tongue shape for English prevocalic /ɹ/. Recall that Mandarin prevocalic /ɹ/ is uniformly produced with bunched tongue shapes in our data, whereas English prevocalic /ɹ/ could be produced with either retroflex or bunched tongue shapes. The data showed that some Mandarin–English bilingual speakers adopted retroflex tongue shapes when producing L2 English /ɹ/ in prevocalic position. Also, Speaker H11 used a retroflex tongue shape for Mandarin syllabic /ɹ/, but changed to a bunched tongue shape for English syllabic /ɹ/. Figure 3 exemplifies the tongue shapes from speaker H11 who used a retroflex tongue shape when producing the Mandarin syllabic /ɹ/, and a bunched tongue shape in the English syllabic /ɹ/.

The second group of speakers used only bunched or retroflex tongue shapes in Mandarin, but used both tongue shapes in the same syllable position when they spoke English (Speaker H1, H2, H3, H6, H11, L4, L5). For example, speaker H1 used only bunched tongue shapes for Mandarin

Table 3. Tongue Shapes of Mandarin and English Rhotic Sounds Produced by Mandarin–English Bilinguals.

	Participants	Prevocalic		Syllabic		Postvocalic	
		Mandarin	English	Mandarin	English	Mandarin	English
High-proficiency group	**H7**	Bunched	**Retroflex**	**Retroflex**	**Retroflex**	**Retroflex**	**Retroflex**
	H2	Bunched	Mix	**Retroflex**	**Retroflex**	**Retroflex**	**Retroflex**
	H6	Bunched	Mix	**Retroflex**	**Retroflex**	**Retroflex**	Bunched
	H1	Bunched	Mix	**Retroflex**	Mix	**Retroflex**	Mix
	H3	Bunched	Bunched	**Retroflex**	Mix	**Retroflex**	Mix
	H11	Bunched	Mix	**Retroflex**	Bunched	**Retroflex**	Mix
	H4	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	H5	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	H8	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	H9	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	H10	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
Low-proficiency group	**L4**	Bunched	Mix	**Retroflex**	**Retroflex**	**Retroflex**	Mix
	L5	Bunched	Mix	Bunched	Bunched	Bunched	Mix
	L1	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	L3	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	L6	Bunched	Bunched	Bunched	Bunched	Bunched	Bunched
	L2	Bunched	Bunched	Bunched	No /ɹ/	Bunched	Bunched

Note. Participants who changed tongue shapes categorically in English and Mandarin are in bold. B: bunched tongue shape; R: retroflex tongue shape; M: mixing bunched and retroflex tongue shapes in different vowel contexts or repetitions.

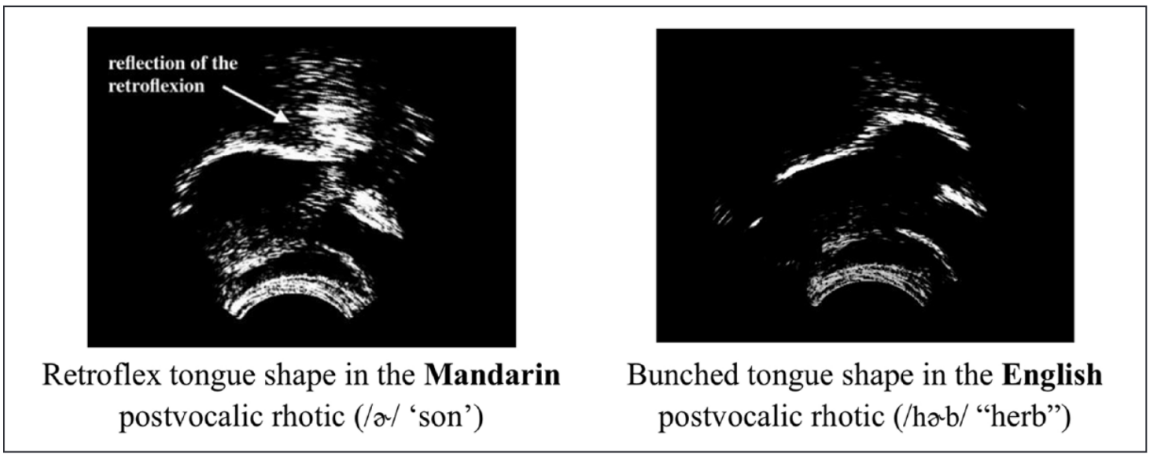


Figure 3. Tongue shapes of speaker H11 who categorically changed the tongue shapes when switching between Mandarin and English in syllabic position.

prevocalic /ɹ/, and only retroflex tongue shapes in syllabic and postvocalic positions. But when she spoke English, she mixed bunched and retroflex tongue shapes in all syllable positions. Figure 4 demonstrates the tongue shapes by Speaker H11 who used retroflex tongue shapes for Mandarin postvocalic /ɹ/, whereas mixed bunched and retroflex tongue shapes for English postvocalic /ɹ/.

The third group of speakers (Speaker H4, H5, H8, H9, H10, L1, L2, L3, L6) used bunched tongue shapes in both languages for all syllable positions. No categorical changes in tongue shapes were found when they switched between Mandarin and English. To examine if there are

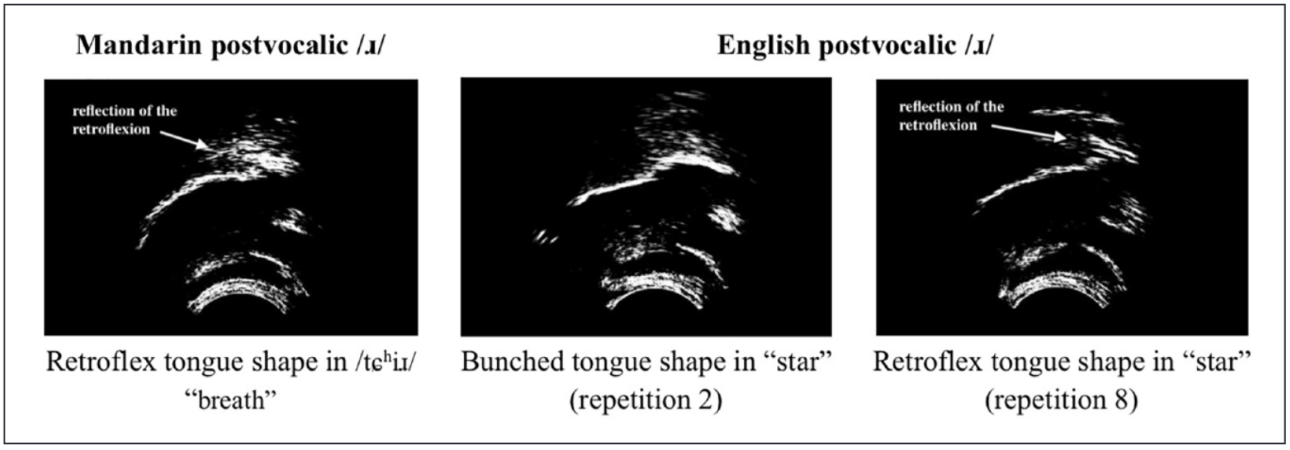


Figure 4. Tongue shapes of speaker H11 who used retroflex tongue shapes for Mandarin postvocalic /ɹ/ and mixed bunched and retroflex tongue shapes for English postvocalic /ɹ/.

within-category differences in tongue shapes between the English and Mandarin /ɹ/s for this group of speakers, their tongue shapes were compared using SSANOVAs. We examined the English and Mandarin rhotics in syllabic position (e.g.: English /hɚb/ vs. Mandarin /ɚ$_{51}$/ "two") rather than rhotics in the prevocalic and postvocalic position because prevocalic and postvocalic rhotics have coarticulatory influence from the flanking vowels. Figure 5 shows the tongue shape differences between English and Mandarin syllabic rhotics of eight speakers (speaker L2 failed to produce the syllabic /ɹ/, and produced the vowel /ə/ instead). There are significant differences in the tongue shapes of English and Mandarin syllabic /ɹ/ for these speakers. Based on the visual inspection of the SSANOVA images, there is a longer section of tongue splines with a significant difference for the high-proficiency group than the low-proficiency group, and the distance in tongue splines between the Mandarin and English /ɹ/ was larger for the high-proficiency group (especially H8, H9, H10). It suggests that the articulatory differences between the English and Mandarin rhotics are larger for the high-proficiency group.

Figure 6 shows the four articulatory measures of the tongue shape. The high-proficiency bilinguals produced the English /ɹ/ with a higher and more fronted tongue dorsum than the Mandarin /ɹ/. But the difference in tongue dorsum was smaller for the low-proficiency group. The pattern of the tongue blade and tongue root is less consistent across participants. Some participants produced the English /ɹ/ with a more fronted tongue blade (H4, H8, H9, H10, L1), whereas others had a more fronted tongue blade for the Mandarin /ɹ/ (H5, L3, L6). For H4, H8, H9, H10, and L6, a more advanced tongue root was observed in English /ɹ/, whereas others did not show differences between the two languages. Both groups of speakers can differentiate the tongue shapes of English and Mandarin rhotics, but the differences are larger for the high-proficiency bilinguals.

Taking the SSANOVAs and the articulatory measures together, we can see that although there are some individual differences, there are significant group differences between English and Mandarin syllabic /ɹ/s in at least some parts of the tongue. It suggests that bilinguals could differentiate the English and Mandarin /ɹ/s in articulation even if they use the same type of tongue shape, not simply fully copying the L1 /ɹ/ sound into L2. In addition, the difference in the tongue shape of Mandarin and English rhotics is larger for high-proficiency bilinguals.

In summary, we found both cross-category or within-category differences in the tongue shapes of Mandarin and English rhotics produced by the bilinguals. Six among the 11 participants in the high-proficiency group changed their tongue shapes categorically when switching between English

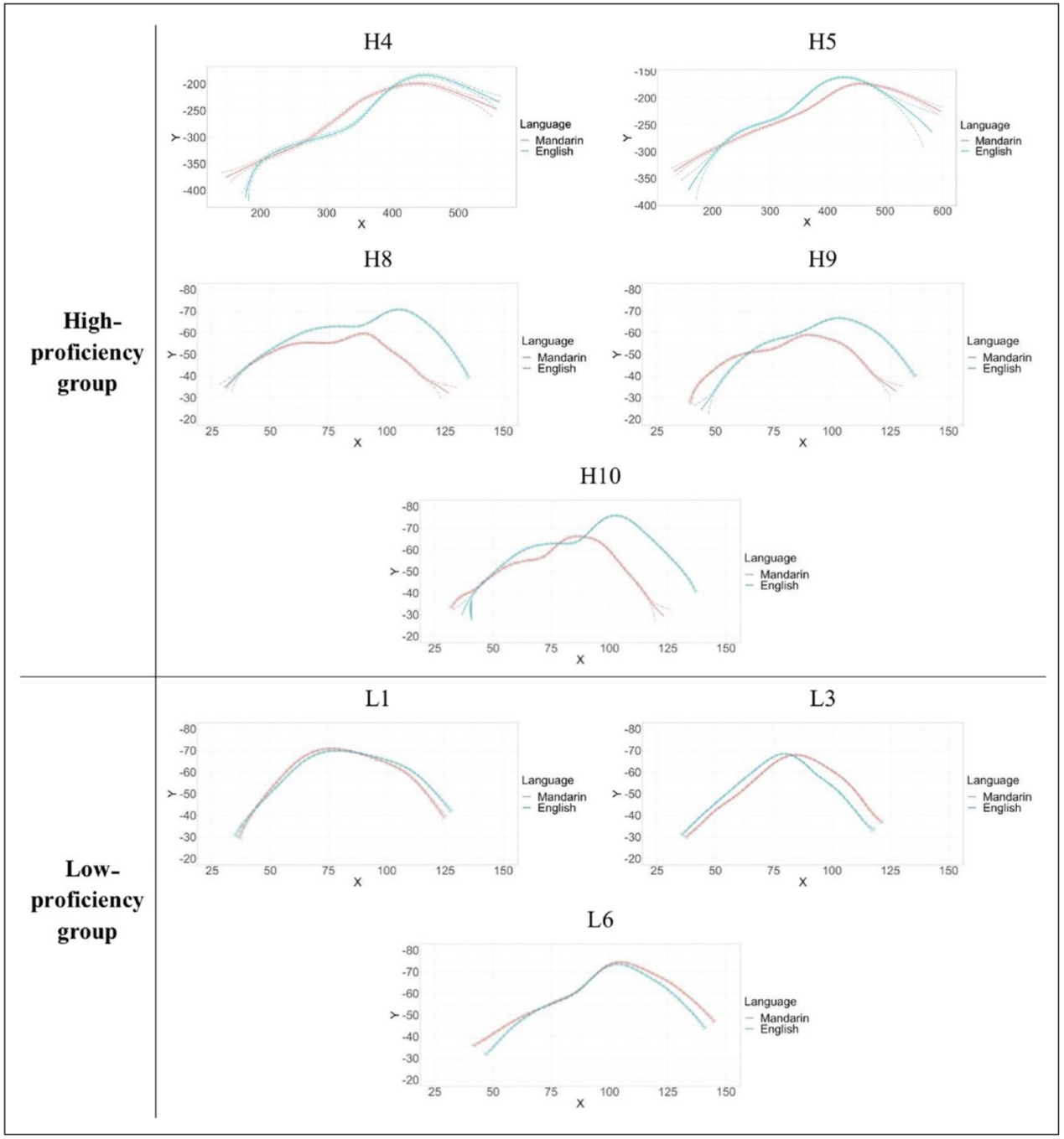


Figure 5. SSANOVAs comparing the tongue splines of Mandarin and English syllabic /ɹ/ by the eight bilingual speakers who did not change their tongue gesture categorically (Speakers H4, H5, H8, H9, H10, L1, L3, L6).

and Mandarin, whereas two among the six participants in the low-proficiency group did so. For speakers who did not change tongue shapes categorically, there are also significant differences in the tongue shapes.

3.1.2 Comparing native and L2 English /ɹ/. As shown in Table 3, both bunched and retroflex tongue shapes were observed in L2 English /ɹ/. In general, the retroflex tongue shape was less common than the bunched tongue shape. This is similar to the articulation pattern of native English

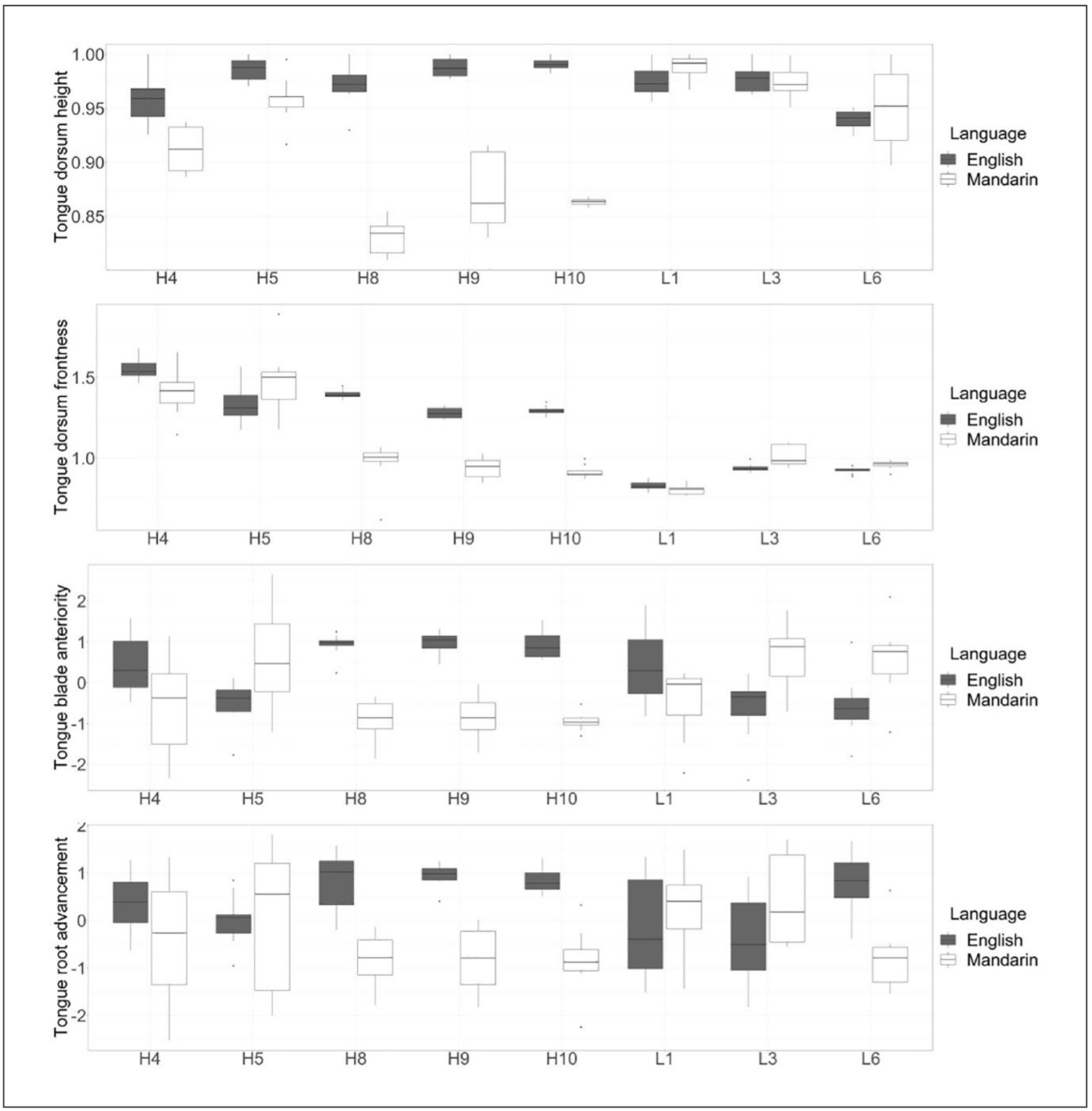
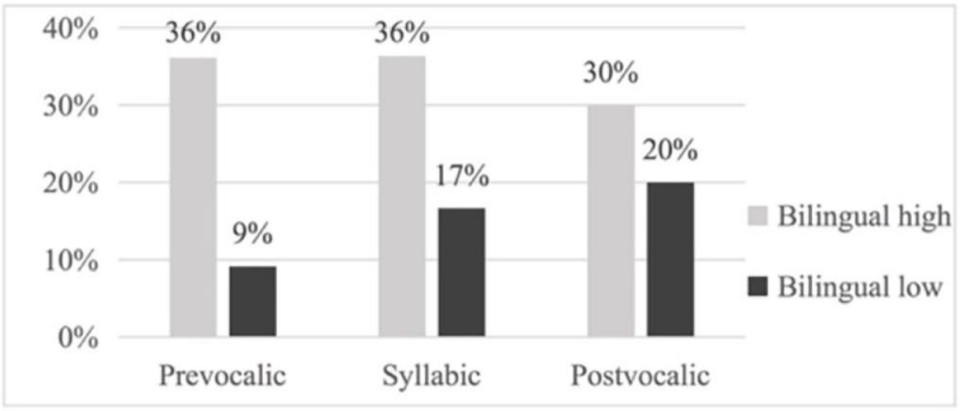


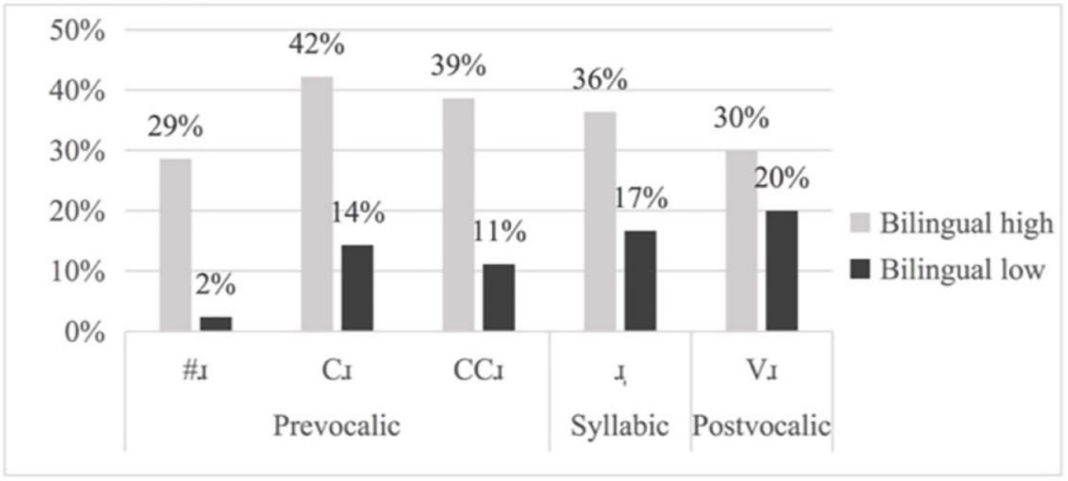
Figure 6. Tongue dorsum height, tongue dorsum frontness, tongue blade anteriority, and tongue root advancement of the eight speakers who used bunched tongue shapes for both Mandarin and English rhotics in syllabic position.

speakers. In English, retroflex tongue shapes were also less common compared with bunched tongue shapes (Mielke et al., 2010, 2016).

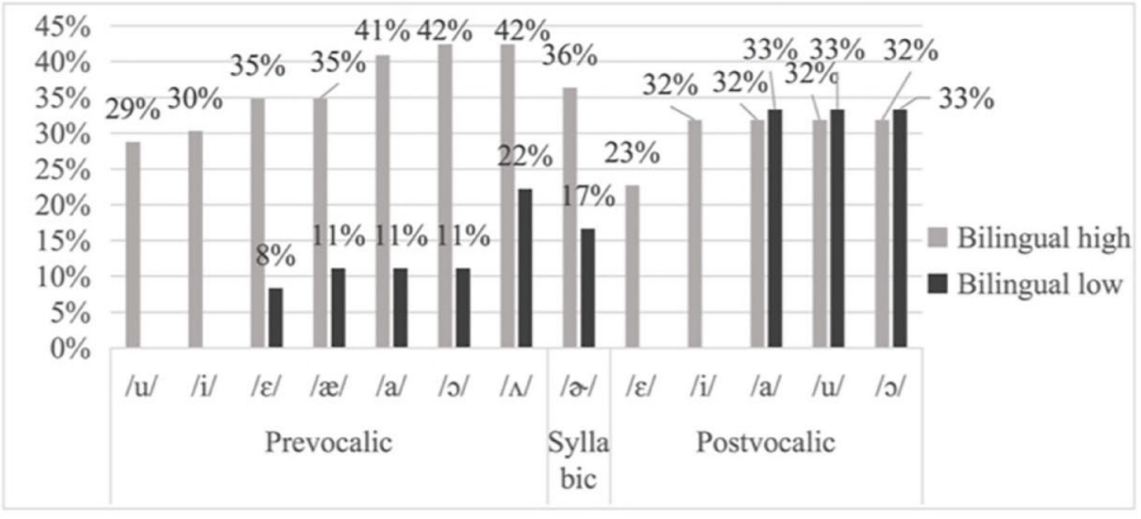
Figure 7 shows the percentage of retroflex tongue shapes (retroflexion rate) in different syllable positions, syllable structures, and vowel contexts. For the high-proficiency group, the retroflex tongue shape was most often used in syllabic position, with a similar ratio in prevocalic position, and was less often found in postvocalic position. In contrast, for low-proficiency speakers, the retroflex tongue shape occurred most often in postvocalic position, and then in syllabic position. It was seldom found in prevocalic position. In terms of syllable structure, high-proficiency bilinguals produced more retroflex tongue shapes in /Cɹ/ and /CCɹ/, and less often in syllable initial /ɹ/.



(1) Retroflexion rate in prevocalic, syllabic and postvocalic positions.



(2) Retroflexion rate in different syllable structures.



(3) Retroflexion rate in different vowel contexts.

Figure 7. Retroflexion rate in different syllable positions, syllable structures and vowel contexts by Mandarin–English bilinguals.

Table 4. Percentage of Prevocalic /ɹ/ Tokens with Frication Noise Among All Prevocalic /ɹ/ in English and Mandarin, and the Percentage Differences Between English and Mandarin for Each Speaker (Mandarin–L2 English).

	Speakers	L2 English (%)	Mandarin (%)	Mandarin–L2 English (%)
High-proficiency Group	H3	0	90	90
	H10	14	98	83
	H4	0	70	70
	H11	38	100	63
	H2	46	96	50
	H8	0	35	35
	H1	0	30	30
	H7	29	58	29
	H5	86	100	14
	H9	91	100	9
	H6	84	60	−24
Low-proficiency Group	L1	21	100	79
	L3	29	80	51
	L5	50	90	40
	L4	7	38	30
	L6	71	80	9
	L2	74	28	−46

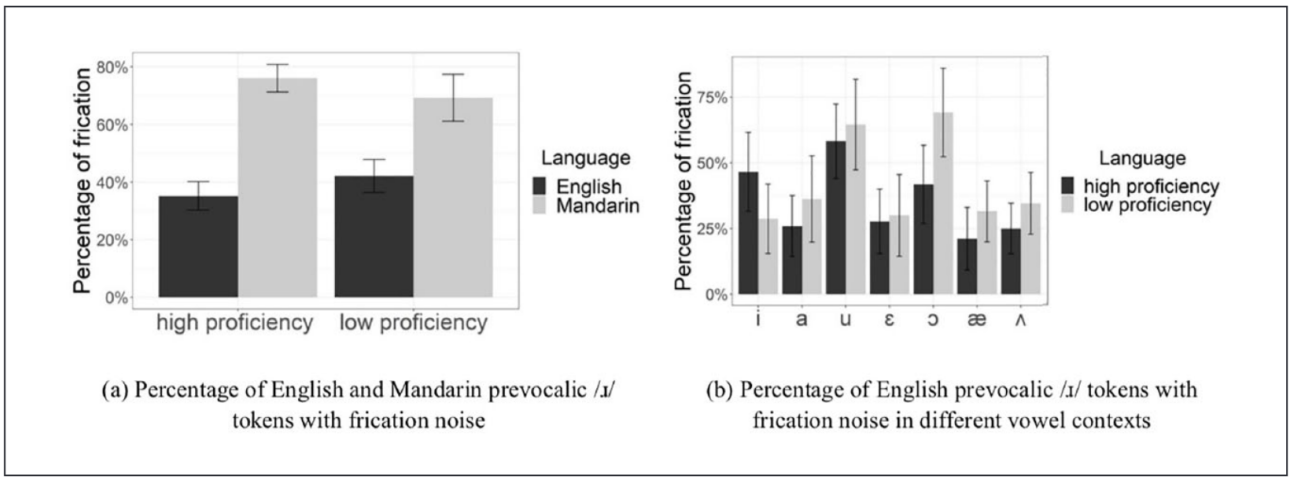
Similar patterns were found in low-proficiency speakers, with the highest percentage of retroflexion in /Cɹ/, followed by /CCɹ/, and least often in syllable initial /ɹ/.

Vowel contexts affect the percentage of retroflex tongue shapes. For high-proficiency speakers, retroflex tongue shapes were found most often before the back vowels /ʌ/ and /ɔ/, followed by the low vowel /a/, and then /æ ɛ/, and least often before /i/ and /u/ in prevocalic position. In postvocalic position, retroflex tongue shapes were less often found before the front vowel /ɛ/ compared with the other vowel contexts (/i a u ɔ/). For low-proficiency speakers, retroflex tongue shapes occurred most often in the /æ/ context, and then in /ʌ ɔ ɛ/ and /a/ for prevocalic /ɹ/. The retroflex tongue shape was never used in the /i u/ contexts. Low-proficiency speakers used more retroflex tongue shapes for postvocalic /ɹ/ in the /a u ɔ/ contexts, and never in the /i ɛ/ contexts.

3.2 Acoustic characteristics of L2 English /ɹ/

3.2.1 Frication noise in prevocalic /ɹ/. One important characteristic of Mandarin prevocalic /ɹ/ is that it often involves frication noise. This section examines whether Mandarin–English bilinguals would transfer the frication noise into their L2. Results showed that Mandarin–English bilinguals did produce frication noise in the English /ɹ/ sound (see Figure 2). The frication noise was only found in prevocalic /ɹ/, not in syllabic and postvocalic /ɹ/s. Also, the frication only occurred in word-initial /ɹ/, but not in consonant clusters. The percentage of frication noise in English and Mandarin prevocalic /ɹ/ produced by each speaker is summarized in Table 4.

Figure 8(a) shows the percentage of tokens where frication noise was found in English and Mandarin tokens for the two proficiency groups. Linear mixed-effected models were performed on the percentage of prevocalic /ɹ/ tokens with frication noise to examine the effects of Language (Mandarin vs. English) and Group (High- vs. Low-proficiency group). The best model included



(a) Percentage of English and Mandarin prevocalic /ɹ/ tokens with frication noise

(b) Percentage of English prevocalic /ɹ/ tokens with frication noise in different vowel contexts

Figure 8. Percentage of prevocalic /ɹ/ tokens with frication noise by Mandarin–English bilinguals. (a) Percentage of English and Mandarin prevocalic /ɹ/ tokens with frication noise. (b) Percentage of English prevocalic /ɹ/ tokens with frication noise in different vowel contexts.

Table 5. Best Linear Mixed-Effects Models on Frication Noise of L2 English /ɹ/ in Different Vowel Contexts.

	Estimate	*SE*	df	*t*-values	Pr(>|*t*|)
(Intercept)	0.296	0.097	32.949	3.042	0.005**
Vowel æ	−0.048	0.083	96.000	−0.572	0.569
Vowel i	0.107	0.083	96.000	1.281	0.203
Vowel u	0.309	0.083	96.000	3.704	<0.001***
Vowel ɛ	−0.010	0.083	96.000	−0.126	0.900
Vowel ɔ	0.219	0.083	96.000	2.626	0.010*
Vowel ʌ	−0.012	0.083	96.000	−0.143	0.886

Note. Formula: Percentage ~ Vowels + (1|Participant). Reference level: Vowels = ɑ.
*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$

Language as the fixed effect, and a random slope for Participant on Language. The model results suggested that there was a main effect of Language. There was significantly more frication noise in Mandarin prevocalic /ɹ/ than in English prevocalic /ɹ/ sound (Est. = 0.360, $SE = 0.089$, $t = 4.029$, $p = .001$). No significant differences were found in Group (High-proficiency group vs. Low-proficiency group), nor in the interaction between Language and Group.

Figure 8(b) summarizes the occurrence of frication noise found in different vowel contexts in L2 English /ɹ/ by the high- and low-proficiency groups. Linear mixed-effected models were performed on the percentage of prevocalic /ɹ/ tokens with frication noise to examine the effects of Group (High-proficiency group vs. Low-proficiency group) and Vowel (see Table 5). The results suggested that there was a main effect of Vowel. There was significantly more frication noise in the /u/ context than in the /a/ context [Est.] = 0.309, $SE = 0.083$, $t = 3.704$, $p < .001$), and significantly more frication noise in the /ɔ/ context than in the /a/ context (Est. = 0.219, $SE = 0.083$, $t = 2.626$, $p < .001$). No significant differences in Group and the interaction between Language and Group were found.

Individual differences in the production of frication noise in English were found. Most speakers produced frication noise in Mandarin prevocalic /ɹ/ and did not produce frication noise, or

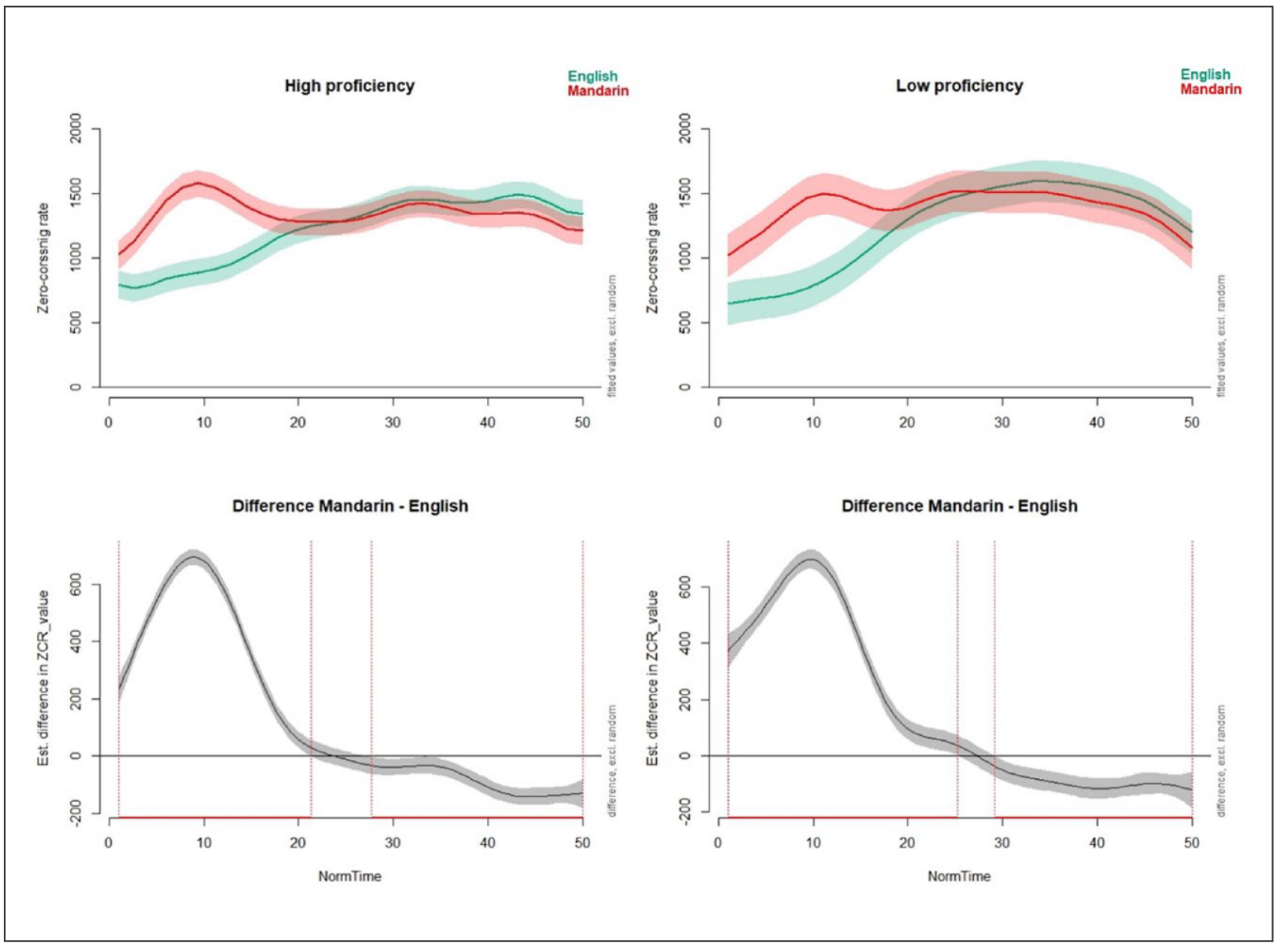


Figure 9. The Zero-crossing rate (ZCR) of English and Mandarin word-initial /ɹ/ modeled with Generalized Additive Mixed Models. The *x*-axis represents normalized time, and the *y*-axis represents the zero-crossing times per second. The shaded bands represent the pointwise 95% confidence interval. In the lower panel, differences between the two smooths comparing English and Mandarin are illustrated. Significance is indicated when the shaded pointwise 95% confidence interval does not intersect with the *x*-axis, marked by a red line.

reduced it to a lower percentage in English. For example, Speaker H3 produced frication noise in about 90% of Mandarin prevocalic /ɹ/, and did not produce any frication in English prevocalic /ɹ/ sound. Note that some speakers (Speaker H8, H1, H7, L4) produced only a little frication noise in their native Mandarin, and they produced frication in English to an even lower percentage. There is a large inter-speaker variation in the degree of reduction. Exceptions can be found in one speaker in the high-proficiency group (H6) and one speaker in the low-proficiency group (L2) who produced even more frication in English than in Mandarin. Speaker H6 increased frication noise from 60% to 84%, whereas speaker L2 increased frication noise from 28% to 74%.

The ZCR values of English and Mandarin syllables containing word-initial /ɹ/ and the following vowels were modeled with GAMMs (see Figure 9). The best model for both high- and low-proficiency groups included a non-linear pattern over normalized time for English and Mandarin syllables, a random intercept for each speaker, and a random smooth for each token. In Mandarin, a notably higher ZCR was observed in the first half of the syllable for both high- and low-proficiency groups (NormTime windows of significant differences were 1.000–21.292 and 27.727–50.000 for the high-proficiency group; NormTime windows of significant differences were 1.000–25.252 and 27.212–50.000 for the low-proficiency group). This heightened ZCR aligned with the frication

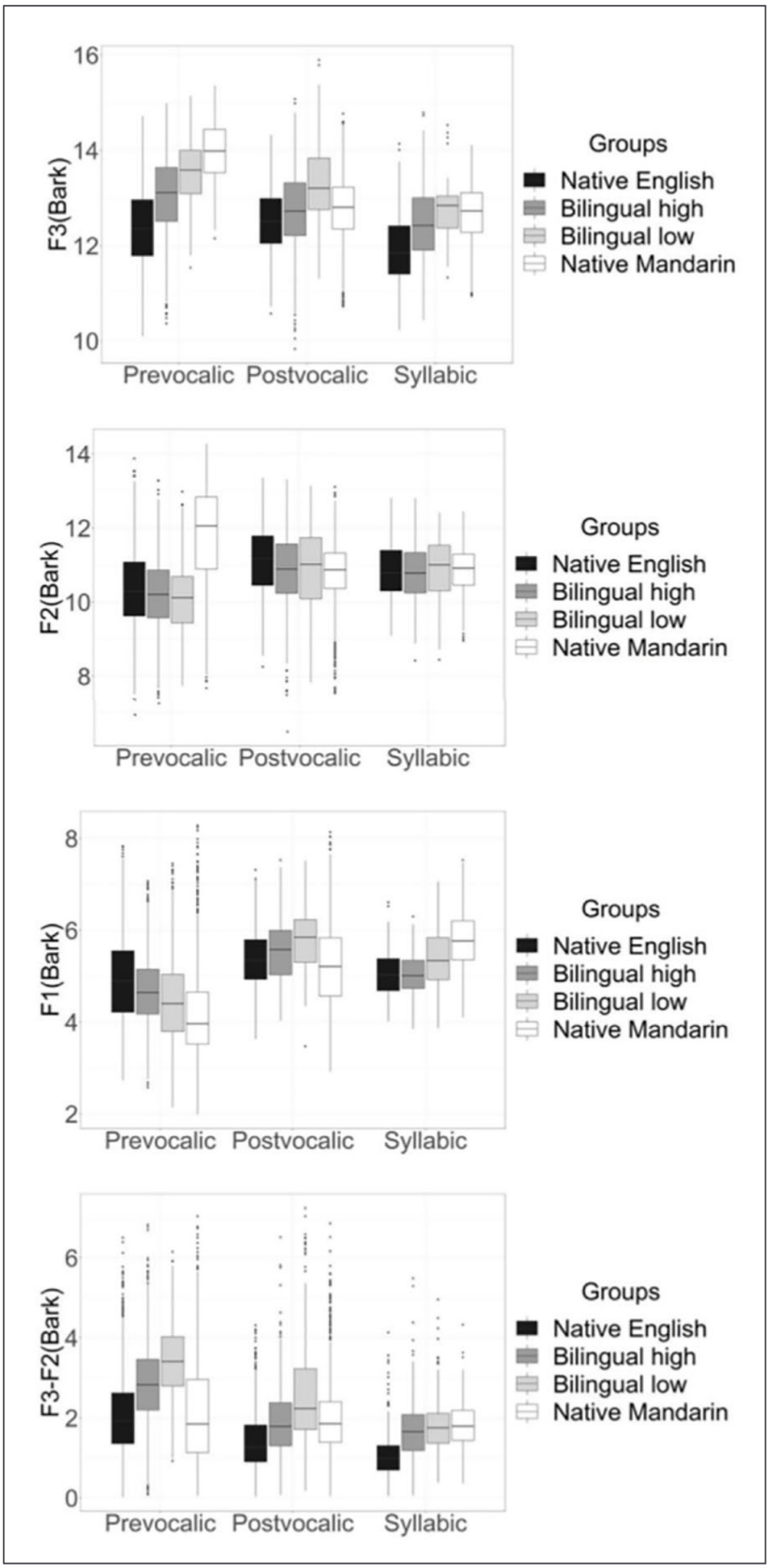


Figure 10. F3, F2, F1, and F3–F2 of English /ɹ/ produced by English native speakers, high-proficiency bilinguals, low-proficiency bilinguals, and Mandarin /ɹ/ produced by the two groups of bilinguals.

noise observed in the spectrograms, indicating significantly stronger frication noise at the beginning of the syllable in Mandarin compared with English for both proficiency groups.

3.2.2 Formant frequency. Figure 10 shows the formant frequencies of the Mandarin and English /ɹ/s in prevocalic, postvocalic, and syllabic positions. Linear mixed-effected models were performed on F3, F2, F1, and F3–F2 to examine the effects of Group (native English, L2 English by high-proficiency bilingual speakers, L2 English by high-proficiency bilingual speakers, native Mandarin) and Syllable position (prevocalic, postvocalic, and syllabic positions). The estimate, standard error, *t*-value, and *p*-value associated with the fixed factors can be found in Supplementary Materials.

The production of the English /ɹ/ by both groups of bilinguals deviated from that of native English speakers in F3 and F3–F2. The best model for F3 showed a main effect of Group, a main effect of Syllable position, and a significant two-way interaction between Group and Syllable position. To understand the nature of the interaction, post hoc analyses were performed on the formant values in each syllable position. The F3 of native English /ɹ/ was significantly lower than that of L2 English /ɹ/ by high-proficiency speakers in prevocalic position (Est.=–0.734, SE=0.169, t=–4.349, p=.001) and in syllabic position (Est.=0.508, SE=0.186, t=–2.724, p=.042). It was also significantly lower than that of low-proficiency speakers in all syllable positions (prevocalic: Est.=–0.910, SE=0.170, t=–5.357, $p<.001$; postvocalic: Est.=–0.518, SE=0.184, t=–2.812, p=.035; syllabic: Est.=0.518, SE=0.192, t=–2.692, p=.045).

The best model for F3–F2 showed that there were main effects of Language and Syllable position, and a significant interaction between Language and Syllable position. The F3–F2 of native English /ɹ/ was significantly lower than that of L2 English /ɹ/ by the high-proficiency group in prevocalic position (Est.=–0.770, SE=0.162, t=–4.739, $p<.001$). It was significantly lower that of L2 English /ɹ/ by the low-proficiency group in prevocalic position (Est.=–1.456, SE=0.167, t=–8.704, $p<.001$) and in postvocalic position (Est.=–0.916, SE=0.170, t=–5.389, $p<.001$).

To summarize, both groups of bilinguals deviated significantly from native speakers in prevocalic position. Only the low-proficiency group was significantly different from native speakers in postvocalic position. In prevocalic and postvocalic positions, L2 English /ɹ/ had a higher F3 and F3–F2, suggesting that L2 English /ɹ/ is less rhotic than native English /ɹ/. Neither group was significantly different from native speakers in syllabic position.

There are also significant differences between high- and low-proficiency bilinguals. L2 English /ɹ/ produced by high-proficiency speakers has a significantly lower F3 and F3–F2 than those of L2 English /ɹ/ by low-proficiency speakers in prevocalic position (F3: Est.=–0.176, SE=0.033, t=–5.285, $p<.001$; F3–F2: Est.=–0.687, SE=0.069, t=–9.935, $p<.001$) and in postvocalic position (F3: Est.=–0.306, SE=0.055, t=–5.610, $p<.001$; F3–F2: Est.=–0.576, SE=0.077, t=–7.464, $p<.001$). A significant difference was also found in F2 between high-proficiency bilinguals and low-proficiency bilinguals in prevocalic position (Est.=–0.345, SE=0.041, t=8.446, $p<.001$) and postvocalic position (Est.=0.297, SE=0.067, t=4.440, $p<.001$).

In addition, in prevocalic and postvocalic positions, the F3–F2 difference for native English and native Mandarin was quite similar, but there was a larger difference between L2 English /ɹ/ and native English /ɹ/. It seems that the bilinguals, especially low-proficiency bilinguals, hyper-articulated the English /ɹ/ in the prevocalic and postvocalic positions, exaggerating the difference between the Mandarin and English /ɹ/s.

The results of post hoc analyses comparing the formant values of the English /ɹ/ produced by native English speakers, high-proficiency bilinguals, and low-proficiency bilinguals in each syllable position are summarized in Table 6. In summary, the English /ɹ/ produced by

Table 6. Results of Post Hoc Analyses Comparing the Formant Values of the English /ɹ/ Produced by Native English Speakers, High-Proficiency Bilinguals, and Low-Proficiency Bilinguals in Prevocalic, Postvocalic, and Syllabic Positions.

		Native English versus high-proficiency group	Native English versus low-proficiency group	High-proficiency group versus low-proficiency group
F3	Prevocalic	Yes	Yes	Yes
	Postvocalic	NS	Yes	Yes
	Syllabic	Yes	Yes	NS
F2	Prevocalic	NS	NS	Yes
	Postvocalic	NS	NS	Yes
	Syllabic	NS	NS	NS
F3–F2	Prevocalic	Yes	Yes	Yes
	Postvocalic	NS	Yes	Yes
	Syllabic	NS	NS	NS

Note. "Yes" indicates a significant difference between the two groups, whereas "NS" indicates no significant difference.

the high-proficiency group was more similar to native English production than the English /ɹ/ produced by the low-proficiency group in F3 and F3–F2. Also, Mandarin–English bilinguals had more native-like production for syllabic and postvocalic /ɹ/s than for prevocalic /ɹ/.

In summary, Mandarin–English bilinguals showed categorical change between bunched and retroflex tongue shapes or within-category tongue shape differences for the Mandarin and English /ɹ/ sounds. They used various tongue shapes for the English /ɹ/ in different syllable positions, vowel conditions, and syllable structures, but the distribution pattern of bunched and retroflex tongue shapes was different from native English production. In terms of acoustics, the English /ɹ/ produced by the bilinguals had some frication noise in prevocalic position and a higher F3. A detailed summary of all acoustic and articulatory findings can be found in Appendix C.

4 Discussion

4.1 Production of L2 English /ɹ/

The data from this study showed that the production of the English /ɹ/ by Mandarin–English bilinguals deviated from native English production for both proficiency groups. The ultrasound data show that bilinguals can produce native-like bunched and retroflex gestures, but the distribution of tongue shapes is different from that of native speakers. In native English, retroflex tongue shapes are more common in prevocalic position and when next to low and/or back vowels. We found similar distribution patterns in the high-proficiency group but not in the low-proficiency group. Also, retroflex /ɹ/ is more often in syllable-initial position than in consonant clusters in native English production. Both groups of bilinguals, however, produced more retroflex /ɹ/ in consonant clusters than in syllable initial position. It should be noted that the /Cɹ/ clusters in this study included only labial consonants. Previous studies on the English /ɹ/ showed that there is not much tongue shape difference between syllable initial /ɹ/ and labial /Cɹ/ clusters (Mielke et al., 2010, 2016; Westbury et al., 1998). Therefore, we would expect similar retroflexion rate for syllable initial /ɹ/ and labial /Cɹ/ clusters in this study, but neither group of bilingual speakers showed such a pattern. Acoustically, the F3 and F3–F2 of L2 English /ɹ/ were significantly higher than those of native

English /ɹ/ production. Also, frication noise can be found in English prevocalic /ɹ/ because of the transfer from Mandarin. Our results suggest that the English /ɹ/ is a difficult sound for L2 learners to acquire. This difficulty arises from cross-linguistic phonetic differences and the complex articulatory characteristics of the English /ɹ/ sound.

Although the production of the English /ɹ/ by Mandarin–English bilinguals differs from native production, the bilinguals do not simply adopt the Mandarin rhotic category for the English /ɹ/. SLM and SLM-r (Flege, 1995, 2003; Flege & Bohn, 2021) propose that if L2 learners can perceive the differences between L1 and L2 sounds, a new sound category would be established for the L2 sound. Therefore, if Mandarin–English bilinguals can perceive the differences between the Mandarin and English /ɹ/s, they can produce the English and Mandarin /ɹ/s differently because a new category would be established for the English /ɹ/. According to PAM-L2, the different phonetic realizations for the same phonological category in each language can be learned by L2 learners because learners might establish two different phonetic categories for L1 and L2 sounds under the common phonological category. The results of this study support the predictions of the two models. We find that Mandarin–English bilinguals, regardless of their L2 proficiency, produced the English and Mandarin /ɹ/s differently. In terms of acoustics, the formant frequencies of the English and Mandarin /ɹ/s produced by the bilingual speakers were significantly different. Moreover, the frication noise in the prevocalic /ɹ/ was significantly less in the English /ɹ/ compared with the Mandarin /ɹ/. It suggests that the bilinguals realized that the English and Mandarin prevocalic /ɹ/s were different in terms of frication noise, but they failed to get rid of the influence from the L1 sound fully.

Compared with previous studies, which mostly analyzed acoustic evidence, this study further showed that bilingual speakers adopted language-specific phonetic details in articulation. The bilingual speakers either showed a categorical change between bunched or retroflex tongue shapes or within-category tongue shape differences for the English and Mandarin /ɹ/ sounds. The difference between English and Mandarin rhotics in the low-proficiency group was smaller. To be more specific, some speakers changed their tongue shapes categorically when they switched between their two languages. Some speakers did not change their lingual tongue shapes categorically, but significant difference was still found in the tongue shapes for the Mandarin and English /ɹ/s. It is expected that tongue shape variation would be hard to acquire for non-native speakers because gestural variation has minimal perceptual consequences. The results, however, showed that both groups of bilingual speakers managed to pick up the gestural variation in their L2.

In summary, the results showed that the English /ɹ/ is a challenging sound for Mandarin learners to acquire although the Mandarin rhotics are similar to the English rhotics to a certain extent. In addition, Mandarin–English bilinguals could produce language-specific phonetic realizations for the Mandarin and English rhotics. Bilingual speakers tried to use language-specific phonetic details to realize a phoneme existing in both their L1 and L2. The results supported the predictions of SLM that a new phonetic category can be established for L2 sounds. It also provided evidence for PAM-L2 that different phonetic realizations of L1 and L2 sounds can be learned for the same phonology category.

4.2 Effects of phonetic similarity

The second research question is how phonetic similarity between L1 and L2 sounds affects L2 production. According to SLM and SLM-r, when the L2 category is more phonetically different from its closest L1 category, L2 learners find it easier to distinguish the phonetic distinction. Larger phonetic dissimilarity results in the creation of a new phonetic category for the L2 sound, leading

to a more native-like production. SLM and SLM-r, therefore, predict that English prevocalic /ɹ/ is easier to learn than English syllabic and postvocalic /ɹ/s because there are larger phonetic differences between the Mandarin and English /ɹ/s in prevocalic position than in syllabic and postvocalic positions. The results, however, did not support this prediction. Among the three positional allophones, English prevocalic /ɹ/ produced by the bilinguals is the least native-like. It was different from native English production in three aspects—articulatory gestures, frication noise, and formant values. First, English native speakers were found to produce more retroflex tongue shapes in prevocalic position than in postvocalic position (Mielke et al., 2016). But for Mandarin–English bilinguals, fewer retroflex tongue shapes were observed. Second, frication noise could be found in many tokens of English prevocalic /ɹ/ for both groups of bilingual speakers. Third, the formant frequencies of prevocalic /ɹ/ deviated most from native speakers' production compared to the formant frequencies of syllabic and postvocalic /ɹ/s. When comparing F3–F2 of the English /ɹ/ produced by bilinguals and English native speakers, both groups of bilinguals deviated significantly from native speakers in prevocalic position. Only the low-proficiency group was significantly different from native speakers in postvocalic position. Neither group was significantly different from native speakers in syllabic position. Therefore, in terms of formant frequencies, the production of the English /ɹ/ was most similar to native production in syllabic position, and least similar in prevocalic position.

In summary, better production was observed for L2 English /ɹ/ in syllabic and postvocalic positions than in prevocalic position. Note that it does not mean that there was a larger improvement for English postvocalic /ɹ/ than for prevocalic /ɹ/. It shows that phonetic similarities between L1 and L2 sounds do not necessarily pose extra problems for L2 learners. This is in contrast with SLM's hypothesis about phonetic similarity and learnability of L2 sounds. Our data show that better learning outcomes can be found for the L2 allophone that is more similar to L1 sounds than the L2 allophone that is less similar. This is probably because, when learning more similar L2 sounds, the learning task is easier for learners, and less adjustment needs to be adopted to reach the L2 targets.

4.3 Proficiency differences in L2 speech production

The third research question is how the production of the English /ɹ/ changes when Mandarin–English bilingual speakers' English proficiency improves. According to SLM and SLM-r, high-proficiency bilinguals may not necessarily have more native-like production compared with low-proficiency bilinguals when producing similar sounds in their first and second languages. Initially, beginner L2 learners can easily substitute L2 sounds with their L1 categories. However, advanced learners face challenges when aiming for native-level performance by accurately producing and perceiving subtle phonetic differences between L1 and L2 sound categories. The results of this study, however, did not find a similar phenomenon in the production of English rhotics. We observed more native-like production of English rhotics in the high-proficiency group, which suggests improvement in the production of similar L2 sounds with increasing L2 proficiency. In terms of articulation, high-proficiency speakers alternated bunched and retroflex tongue shapes in a similar way as native speakers did. Recall that in English, retroflex /ɹ/ was more common in prevocalic position than in postvocalic position, and more common next to back and/or low vowels than high and/or front vowels (Mielke et al., 2010, 2016). We also found similar syllabic position effect and vowel effect for the English /ɹ/ sound produced by high-proficiency bilinguals. The production by low-proficiency bilinguals, however, was less native-like in two dimensions. Acoustically, the formant frequencies of the English /ɹ/ produced by high-proficiency speakers had a higher level of

rhoticity and were more similar to native production than were the low-proficiency group. However, high-proficiency speakers were not better than low-proficiency speakers in all dimensions. The performance of the two groups was similar in having frication noise in English prevocalic /ɹ/. But in general, better performance was observed in the high-proficiency group.

One limitation of this study is that we did not have the same number of speakers in the high- and low-proficiency groups due to practical reasons. We analyzed the subset data from six high-proficiency speakers and six low-proficiency speakers who were examined using the EchoB system, and similar patterns were observed. Therefore, we decided to report all the data from the 17 speakers for a more comprehensive investigation. Another limitation of this study is that the English proficiency of the low-proficiency group is not very low. This is because for students to be considered by the postgraduate programs in the two universities where we conducted the experiment, students need to meet the minimum language requirements. The difference in the production of English rhotics might be larger if the differences in English proficiency between the two groups were bigger. Future studies can consider using two groups with greater differences in English proficiency for comparison.

In summary, the results of this study show that, with increasing L2 proficiency, Mandarin–English bilinguals have more native-like performance in producing the fine phonetic details of L2 sounds, even in subtle articulation patterns which have minimal effects on acoustic patterns.

4.4 Using ultrasound imaging in L2 acquisition research

The acquisition of rhotic sounds, especially the acquisition of the English /ɹ/ sound has attracted much attention in the literature (Bohn & Flege, 1992; Boyce et al., 2016; Flege, 1992; Goto, 1971; Jun & Cowie, 1994; Munro et al., 1996). Most of these studies addressed this issue from a perceptual or acoustical approach (Bradlow, 1997; Bradlow et al., 1999; Goto, 1971; Sheldon & Strange, 1982). The articulation of the English /ɹ/ by second language learners has been largely ignored. As a crucial feature of the English /ɹ/, the variation in tongue shapes and multiple constrictions need to be examined with articulatory measures. This study exemplified that the use of articulatory measures allows us to have a more comprehensive view of the English /ɹ/ produced by bilingual speakers. Our findings show that, although the tongue shape variation can also be found in the Mandarin /ɹ/, the learners still have difficulties in having a native-like production. The distributional patterns of bunched and retroflex tongue shapes by bilinguals are different from those of native English speakers.

One limitation of this study is the absence of a consistent pattern observed in the tongue root. The data on the pharyngeal constriction is helpful because the production of the English /ɹ/ involves three supraglottal constrictions—a narrowing at the lips achieved by lip-rounding and protrusion, an oral constriction in the palatal region made by the tongue tip or tongue front, and a narrowing in the pharyngeal cavity made by the tongue root retracting toward the pharyngeal wall (Delattre & Freeman, 1968; Zhou et al., 2008). Previous studies suggest that the pharyngeal constriction is one of the most difficult gestures to acquire in the English /ɹ/ for L2 learners and English-speaking children (Boyce et al., 2011; Harper et al., 2016; Klein et al., 2013). Future studies could involve a larger sample size of speakers to examine the tongue root and the pharyngeal constriction in the production of L2 English /ɹ/. Furthermore, an intriguing aspect that warrants further investigation is the involvement of lip-rounding and protrusion in the production of L2 English /ɹ/. Future studies can make lip video recordings simultaneously with ultrasound imaging to gain a more comprehensive understanding of L2 English /ɹ/ production.

5 Conclusion

This study examined the articulatory and acoustic features of the English /ɹ/ produced by Mandarin–English bilinguals with two proficiency levels (one group with high English proficiency and one group with relatively lower proficiency). The English /ɹ/ produced by Mandarin–English bilinguals was different from native English /ɹ/ in both articulation and acoustics. The ultrasound data shows that bilinguals can produce native-like bunched and retroflex gestures, but the distribution of tongue shapes differs from that of native speakers. Acoustically, the F3 and F3–F2 of L2 English /ɹ/ were significantly higher than those of native English /ɹ/ production, and frication noise that is often found in Mandarin prevocalic /ɹ/ is transferred onto English prevocalic /ɹ/. Although the production by bilingual speakers is different from native production, they do not simply adopt the Mandarin rhotic category for the English /ɹ/. Both high- and low-proficiency bilinguals produced the English and Mandarin /ɹ/ differently. Bilinguals were able to produce different phonetic realizations for phonemes existing in both L1 and L2. In addition, as language proficiency grows, the production of the English /ɹ/ by Mandarin–English bilinguals becomes more native-like in both articulation and acoustics. Finally, the phonetic similarities between L1 and L2 sounds facilitated rather than hindered L2 sound acquisition. Mandarin–English bilinguals produced more native-like English /ɹ/ in syllabic and postvocalic positions—the English /ɹ/ allophones that were more similar to the Mandarin /ɹ/, and less native-like production for English prevocalic /ɹ/ which was more different from Mandarin prevocalic /ɹ/.

Acknowledgements

The Siemens ACUSON X300 system at Haskins Laboratories was available due to a generous loan agreement with Siemens Medical Solutions USA, Inc.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was supported by the graduate studentship and Global Scholarship Programme for Research Excellence scholarship from the Chinese University of Hong Kong to the first author. The experiment conducted at Haskins was supported by NIH grant DC-002717 to Haskins Laboratories.

ORCID iDs

Shuwen Chen https://orcid.org/0000-0002-0707-2619
D. H. Whalen https://orcid.org/0000-0003-3974-0084
Peggy Pik Ki Mok https://orcid.org/0000-0002-9284-6083

Supplemental material

Supplemental material for this article is available online.

Notes

1. This study represents a subsequent analysis of the data presented in Chen's doctoral dissertation (2020). It utilizes the data from the same participants, with the exclusion of one individual.
2. The participants in the low-proficiency group all have a score of 6.5 because this is the minimal entry requirement for the postgraduate programs at the two universities.
3. IELTS score descriptors: https://takeielts.britishcouncil.org/teach-ielts/test-information/ielts-scores-explained. IELTS speaking band descriptors: https://www.ielts.org/-/media/pdfs/ielts-speaking-band-descriptors.ashx

4. Mandarin has two apical vowels, [ɿ] and [ʅ]. These two symbols, [ɿ] and [ʅ], 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 apical vowel [ɿ] appears after dental affricates and fricatives /ts/, /tsʰ/ and /s/, whereas the apical vowel [ʅ] appears after post-alveolar consonants /tʂ/, /tʂʰ/, /ʂ/, and /ɹ/. The phonological status of the two sounds is also controversial, but it is not central to our study (see Lee-Kim, 2014 for a discussion on the phonological status of the two sounds).
5. Native Mandarin /ɹ/ was produced by the high-proficiency and low-proficiency bilinguals.

References

Ahn, S. (2018). The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics*, *71*, 451–467.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Articulate Instruments Ltd. (2008). *Ultrasound stabilisation headset users manual: Revision 1.4*.

Articulate Instruments Ltd. (2012). *Articulate assistant advanced user guide: Version 2.14*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 1–47). John Benjamins Publishing Company.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), 131–158.

Boyce, S. E., Combs, S., & Rivera-Campos, A. (2011). Acoustic and articulatory characteristics of clinically resistant /r/. *The Journal of the Acoustical Society of America*, *129*(4), 2625–2625.

Boyce, S. E., & Espy-Wilson, C. Y. (1997). Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America*, *101*, 3741–3753.

Boyce, S. E., Hamilton, S. M., & Rivera-Campos, A. (2016). Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements. *Clinical Linguistics & Phonetics*, *30*(3–5), 174–201. https://doi.org/10.3109/02699206.2015.1127999

Bradlow, A. R. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Yamada, R., & Tohkura, Y. (1995). The effect of training in /r/-/l/ perception on /r/-/l/ production by Japanese speakers. *Proceedings XIIIth International Congress of Phonetic Sciences*, *4*, 562–565.

Chao, Y. (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, S. (2020). *Production and perception of English rhotic sounds by Mandarin-English bilinguals* [PhD dissertation]. The Chinese University of Hong Kong.

Chen, S., & Mok, P. (2021). Articulatory and acoustic features of Mandarin /ɹ/: a preliminary study. In *Proceedings of the 12th International Symposium on Chinese Spoken Language Processing* (ISCSLP). IEEE.

Chen, W. R., Tiede, M., & Chen, S. (2017, October 4–6). *An optimization method for correction of ultrasound probe-related contours to head-centric coordinates* [Oral presentation]. Ultrafest VIII, Potsdam, Germany.

Chuang, Y., Wang, S., & Fon, J. (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*. The University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html

Chung, H., & Pollock, K. E. (2021). Acoustic characteristics of rhotic vowel productions of young children. *Folia Phoniatrica et Logopaedica 73*(2), 89–100.

Dalston, R. M. (1975). Acoustic characteristics of English /w, r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, *57*(2), 462–469.

Delattre, P. C., & Freeman, D. C. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics*, *6*(44), 29–68. https://doi.org/10.1515/ling.1968.6.44.29

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Educational Testing Service. (2010). *Linking TOEFL iBT scores to IELTS scores—A research report*. https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, *1*, 157–232.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller., & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–358). De Gruyter Mouton.

Flege, J. E., & Bohn, O. (2021). The revised Speech Learning Model (SLM-r). In W. Ratree (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), 1161–1173. https://doi.org/10.1121/1.414884

Foulkes, P., & Docherty, G. J. (2000). Another chapter in the story of /r/: "Labiodental" variants in British English. *Journal of Social Phonetics*, *4*(1), 30–59.

Fu, M. (1956). Phonemes and Pinyin symbols in the Beijing Speech. *Zhongguo Yuwen*, *5*, 3–12.[北京话的音位和拼音字母. 中国语文 5, 3–12.]

Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2007). A motor differentiation model for liquid substitutions in children's speech. *Proceedings of Meetings on Acoustics*, *1*(2008), 060003. https://doi.org/10.1121/1.2951481

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, *34*(1), 49–72.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Gu, C. (2014). Smoothing spline ANOVA models: R package gss. *Journal of Statistical Software*, *58*(5), 1–25.

Hagiwara, R. (1995). *Acoustic realizations of American /r/ as produced by women and men* [UCLA Working Papers in Phonetics No. 90]. https://escholarship.org/uc/item/8779b7gq

Harper, S., Goldstein, L. M., & Narayanan, S. S. (2016). L2 acquisition and production of the English rhotic pharyngeal gesture. In N. Morgan (Ed.), *Proceedings of Interspeech 2016* (pp. 208–212). University of Southern California.

Heyne, M., Wang, X., Derrick, D., Dorreen, K., & Watson, K. (2020). The articulation of /ɹ/ in New Zealand English. *Journal of the International Phonetic Association*, *50*(3), 366–388. https://doi:10.1017/S0025100318000324

Hu, F. (2020). *The vowel: A general introduction with reference to Chinese data [元音研究]*. Foreign Language Teaching and Research Press.

Huang, J., Hsieh, F., & Chang, Y. (2020). *Er-Suffixation in Southwestern Mandarin: An EMA and Ultrasound Study* (pp. 661–665). *Interspeech 2020*.

Huang, T., Chang, Y., & Li, H. (2022, 3–5 November). *Variable articulation for the Taiwanese Mandarin rhotic liquid /r/* [Paper presentation]. *UltrafestX*, *Manchester*.

Hussain, Q., & Mielke, J. (2021). An acoustic and articulatory study of rhotic and rhotic-nasal vowels of Kalasha. *Journal of Phonetics*, *87*, 101028.

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America*, *133*(6), 4232–4246. https://doi.org/10.1121/1.4802905

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*(4), 571–584. https://doi.org/10.1016/j.wocn.2011.03.003

Jiang, S., Chang, Y., & Hsieh, F. (2019a, May 24–25). *A Cross-dialectal Comparison of er-suffixation in Beijing Mandarin and Northeastern Mandarin: An Electromagnetic Articulography Study* [Conference session]. *HISPhonCog 2019: Hanyang International Symposium on Phonetics & Cognitive Sciences of Language*, *Hanyang University, Seoul, Korea*.

Jiang, S., Chang, Y., & Hsieh, F. (2019b). An EMA study of er- suffixation in Northeastern Mandarin monophthongs. In S. Calhoun, P. Escudero, M. Tabian, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3617–3621). Australasian Speech Science and Technology Association Inc.

Jun, S.-A., & Cowie, I. (1994). Interference for 'new' versus 'similar' vowels in Korean speakers of English. *Ohio State University Working Paper*, *43*, 117–130.

Karlgren, B. (1915–1926). *Studies on Chinese Phonology [Etudes sur la phonologie chinoise]*. K.W. Appelberg.

King, H., & Ferragne, E. (2020). Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics*, *80*, 100978. https://doi.org/10.1016/j.wocn.2020.100978

King, H., & Liu, A. (2017, October 4–6). An ultrasound and acoustic study of the rhotic suffix in Mandarin [Paper presentation]. Ultrafest VIII, University of Potsdam, Potsdam, Germany.

Klein, H. B., Byun, T. M., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech—Language Pathology*, *22*(3), 540–553. https://doi.org.easyaccess1.lib.cuhk.edu.hk/10.1044/1058-0360(2013/12-0137)

Knight, R., Dalcher, C. V., & Jones, M. J. (2007). *A real-time case study of rhotic acquisition in Southern British English* [Conference session]. *Proceedings of ICPHS XVI*, *Saarbrücken, Germany*.

Kochetov, A., Sreedevi, N., Kasim, M., & Manjula, R. (2014). Spatial and dynamic aspects of retroflex production: An ultrasound and EMA study of Kannada geminate stops. *Journal of Phonetics*, *46*, 168–184. https://doi.org/10.1016/j.wocn.2014.07.003

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Ladefoged, P., & Maddison, I. (1996). *The sounds of the world's languages*. Blackwell.

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, *15*(2), 256–268. https://doi.org/10.1111/j.1467-9841.2011.00464.x

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2013). Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics*, *41*(3), 198–210. https://doi.org/10.1016/j.wocn.2013.01.004

Lawson, E., Stuart-Smith, J., & Scobbie, J. M. (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, W.-S. (2005). A phonetic study of the "er-hua" rimes in Beijing Mandarin. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (pp. 1093–1096). International Speech Communication Association (ISCA).

Lee, W.-S. (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* (pp. 413–416). International Phonetic Association.

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112.

Lee-Kim, S. I. (2014). Revisiting Mandarin "apical vowels": An articulatory and acoustic study. *Journal of the International Phonetic Association*, *44*(3), 261–282.

Lenth, R. V. (2018). *emmeans: Estimated Marginal Means, aka LeastSquares Means* (R Package Version 1.1). https://CRAN.R-project.org/package=emmeans

Li, Y. (1996). Discussion on Mandarin er-hua and its phoneme. *Yuwen Yanjiu*, *59*, 21–26. [论普通话儿化韵及儿化音位. 语文研究 59, 21–26.]

Liao, R., & Shi, F. (1987). An experimental study on the sound quality of the r consonant of Mandarin Chinese. *Language Research*, *2*, 146–160. [汉语普通话r声母音质的实验研究. 语言研究2, 146–160]

Lin, X., & Wang, L. (2013). *A course in phonetics* (2nd ed.). [语音学教程(增订版)] Peking University Press.

Lin, Y. (1989). *Autosegmental treatment of segmental process in Chinese phonology*. University of Texas PhD dissertation.

Lin, B. (1992). Er-hua in Mandarin Chinese. *Yuyan Wenzi Yingyong* [普通话的儿化. 语言文字应用], *4*, 91–94.

Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge University Press.

Lindau, M. (1985). The story of /r/. In V. A. Fromkin (Ed.), *Phonetic linguistics. Essays in honor of Peter Ladefoged* (157–168). Academic Press.

Luo, S. (2020). Articulatory tongue shape analysis of Mandarin alveolar–retroflex contrast. *The Journal of the Acoustical Society of America*, *148*(4), 1961–1977.

Lyskawa, P. (2015). The ultrasound study of /ɹ/ in non-native speakers. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390. https://doi.org/10.1017/S0142716400009796

McAllister Byun, T., & Tiede, M. (2017). Perception-production relations in later development of American English rhotics. *PLOS ONE*, *12*(2), Article e0172022. https://doi.org/10.1371/journal.pone.0172022

Mcgowan, R. S., Nittrouer, S., & Manning, C. J. (2004). Development of [r] in young, Midwestern, American children. *The Journal of the Acoustical Society of America*, *115*(2), 871–884. https://doi.org/10.1121/1.1642624

Mielke, J. (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. https://doi.org/10.1121/1.4919346

Mielke, J., Baker, A., & Archangeli, D. (2010). Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Laboratory Phonology*, *10*, 699–730.

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, *92*(1), 101–140. https://doi.org/10.1353/lan.2016.0019

Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313–334.

Noiray, A., Ries, J., Tiede, M., Rubertus, E., Laporte, C., & Ménard, L. (2020). Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & optical linguo-labial articulation recording system. *Laboratory Phonology*, *11*(1), 14.

Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). *The association between TOEFL iBT test scores and the Common European Framework of Reference (CEFR) Levels*. Educational Testing Service. https://www.ets.org/Media/Research/pdf/RM-15-06.pdf

Preston, J. L., & Edwards, M. L. (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech & Hearing Services in Schools*, *38*(4), 297–308.

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite [Computer software]. https://fave.readthedocs.io/en/latest/index.html

Scobbie, J. M., Wrench, A. A., & Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In R. Sock., S. Fuchs., & Y. Laprie (Eds.), *Proceedings of 8th International Seminar on Speech Production* (pp. 373–376). INRIA.

Shao, B., & Ridouane, R. (2023). On the nature of apical vowel in Jixi-Hui Chinese: Acoustic and articulatory data. *Journal of the International Phonetic Association*, *53*(3), 977–1002.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

Shriberg, L. D., & Kent, R. D. (2003). *Clinical phonetics* (3rd ed.). Allyn & Bacon.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech, Language, and Hearing Research*, *37*(5), 1100–1126.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–798.

Smith, J. G. (2010). *Acoustic properties of English /l/ and /r/ produced by Mandarin Chinese speakers*. University of Toronto.

Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, *16*(4), 380–406. https://doi.org/10.1017/S0142716400066005

Tiede, M. K. (2018). *GetContours* (Version 1.3). https://github.com/mktiede/GetContours

Tiede, M. K., Boyce, S. E., Espy-Wilson, C. Y., & Gracco, V. L. (2010). Variability of North American English /r/ production in response to palatal perturbation. In B. Maassen & P. van Lieshout (Eds.), *Speech motor control in normal and disordered speech V* (pp. 53–67). Oxford University Press.

Twist, A., Baker, A., Mielke, J., & Archangeli, D. (2007). Are "Covert" /ɹ/ Allophones Really Indistinguishable? *University of Pennsylvania Working Papers in Linguistics*, *13*(2), Article 16.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2023). *Itsadug: Interpreting time series and autocorrelated data using GAMMs* (R Package Version, 2, 4). https://search.r-project.org/CRAN/refmans/itsadug/html/itsadug.html.

Wang, Z. (1993). *The geometry of segmental features in Beijing Mandarin* [PhD dissertation]. University of Delaware.

Wang, J. (2005). An integrated discussion of er-hua. *Yuyan Wezi Yingyong*, 3, 46–54. [儿化规范综论. 语言文字应用].

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, *24*, 203–225.

Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E. S., & Hailey, D. S. (2005). HOCUS, the Haskins optically-corrected ultrasound system. *Journal of Speech, Language, and Hearing Research*, *48*, 543–553.

Wood, S. (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

Wu, Z., & Lin, M. (1989). *Shiyan Yuyinxue Gaiyao*. [实验语音学概要] Higher Education Press.

Xing, K. (2021). *Phonetic and phonological perspectives on rhoticity in Mandarin* [PhD dissertation]. University of Manchester.

Yuan, J. (1960). *A Survey of the Chinese Dialects*. [汉语方言概要] Wenzi Gaige Press.

Zee, E., & Lee, W. S. (2001, September 3–7). An acoustical analysis of the vowels in Beijing Mandarin. In P. Dalsgaard, B. Lindberg, H. Benner, & Z. Tan (Eds.), *EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event* (pp. 643–646). International Speech Communication Association.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (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.

Appendix A

English stimuli

Words with Prevocalic /ɹ/.

	/i/	/ɑ/	/u/	/ɛ/	/ɔ/	/æ/	/ʌ/
/#_V(C)/	read /ɹid/	rod /ɹɑd/	rude /ɹud/	red /ɹɛd/	raw /ɹɔ/	rap /ɹæp/	rough /ɹʌf/
/C_V(C)/	breeze /bɹiz/	bra /bɹɑ/	prove /pɹuv/	press /pɹɛs/	broad /bɹɔd/	brag / bɹæg/	brush /bɹʌʃ/
/CC_V(C)/	spree /spɹi/	/	spruce /spɹus/	spread /spɹɛd/	sprawl /spɹɔːl/	Spratt /spɹæt/	sprung /spɹʌŋ/

Words with postvocalic /ɹ/.

beer	star	tour	bare	bore
/biɹ/	/stɑɹ/	/tuɹ/	/bɛɹ/	/bɔɹ/

Words with syllabic /ɹ/.

perp	Herb	purr
/pɚp/	/hɚb/	/pɚ/

Appendix B

Mandarin stimuli

Words with the Prevocalic Rhotic.

Vowel contexts	Words	Meanings	Chinese characters
ɿ	ɻɿ$_{51}$	sun	日
ɤ	ɻɤ$_{51}$	hot	热
u	ɻu$_{51}$	enter	入
a*	ɻan$_{35}$	but	然
ɑ	ɻɑŋ$_{51}$	“allow”	让

*/ɻa/ is phonotactically illegal in Mandarin.

Words with the Syllabic Rhotic (Rhotacized Vowel).

Words	Meanings	Chinese characters
ɚ$_{35}$	son	儿
ɚ$_{214}$	ear	耳
ɚ$_{51}$	two	二

Words with the Syllabic Rhotic (Rhotacized Vowel). Words with the Postvocalic Rhotic (r-suffix).

Vowel contexts	Words	Meanings	Chinese characters
i	tɕiɹ$_{55}$	chicken	鸡儿
	tɕ^{h}iɹ$_{51}$	“breath”	气儿
ʅ	sʅɹ$_{55}$	thread	丝儿
ɿ	tʂɿɹ$_{55}$	branch	枝儿
y	yɹ$_{55}$	fish	鱼儿
	yɹ$_{55}$	“small fish”	鱼儿
u	huɹ$_{35}$	soul	魂儿
	tʂuɹ$_{55}$	“pearl”	珠儿
a	paɹ$_{55}$	handle	把儿
	tʂ^{h}aɹ$_{55}$	“cross”	叉儿
ɤ	kɤɹ$_{55}$	song	歌儿
	xɤɹ$_{35}$	“small boxes”	盒儿
	tʂɤɹ$_{51}$	“here”	这儿

Appendix C

Summary of the Production Results by High- and Low-Proficiency Bilingual Speakers, and Comparison to Native English and Mandarin /ɹ/ Production.

		English /ɹ/	English /ɹ/ by high-proficiency speakers	English /ɹ/ by low-proficiency speakers	Mandarin /ɹ/
Articulation	Syllable position effect	Prevocalic > postvocalic (Mielke et al., 2010, 2016)	Prevocalic and syllabic > postvocalic	Postvocalic > syllabic > prevocalic	Prevocalic: bunched only Postvocalic/syllabic: bunched or retroflex
	Vowel effect	Back and/or low vowels (such as /a/) > high and/or front vowels (such as /i/) (Mielke et al., 2010, 2016)	Prevocalic: /ʌ ɔ/ > /a/ > /æ ɛ/ > /i/ > /u/ (low back vowels > low front vowels > high vowels) Postvocalic: /i a u ɔ/ > /ɛ/	Prevocalic: /æ/ > /ʌ ɔ ɛ/ > /a/ > /i u/ (low vowels > high vowels) Postvocalic: /a u ɔ/ > /i ɛ/	Same tongue shape for all vowel contexts
	Syllable structure effect	Syllable-initial > consonant clusters (but not much difference between /#ɹ/ and labial /Cɹ/ clusters) (Mielke et al., 2010, 2016)	Labial /Cɹ/ > /CCɹ/ > /#ɹ/	Labial /Cɹ/ > /CCɹ/ > /#ɹ/	NA (No consonant clusters in Mandarin)
Acoustics	Frication	No frication noise in the English /ɹ/	Less frication noise in the English /ɹ/ than in the Mandarin /ɹ/ (NS difference between the two proficiency groups	Less frication noise in the English /ɹ/ than in the Mandarin /ɹ/ (NS difference between the two proficiency groups)	Frication noise often found in Mandarin prevocalic /ɹ/ tokens
	Formants	A low F3, around 1,800 Hz	Significantly higher F3 than native production in prevocalic and syllabic position	Significantly higher F3 than native production in all syllable positions	A low F3, but higher than the F3 of English. Around 2,300 Hz in prevocalic position, and 1,900 Hz in syllabic and postvocalic positions

Note. The symbol ">" means more retroflex tokens (higher retroflexion rate) in all cells.