Speech production of rhotics in highly proficient bilinguals: Acoustic and articulatory measures

Shuwen Chen, Peggy Pik Ki Mok

The Chinese University of Hong Kong
chenshuwen@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

ABSTRACT

One of the biggest unresolved questions in bilingual speech production is how L2 phonological categories are constructed given the well-established L1 categories and how phonological categories are realized phonetically. The current study aimed to investigate how bilinguals' two languages interact when the same phoneme in two languages was realized differently in phonetics. We examined the acoustic and articulatory features of English and Mandarin /ɹ/ produced by highly proficient Mandarin-English bilinguals. Ten Mandarin-English bilingual speakers and ten American English monolingual speakers were recorded with ultrasound imaging technique. Results showed that highly proficient Mandarin-English bilinguals produced English /ɹ/ without L1 influences. They used language-specific phonetic details for English and Mandarin /ɹ/. It implies that language-specific phonetic realizations can be learned for the same phonological category by highly proficient bilingual speakers.

Keywords: Speech production; rhotics; English; Mandarin; high proficiency late bilinguals.

1. INTRODUCTION

Bilingual speakers are speakers who regularly use two languages, regardless of their proficiency, age of acquisition, or sequence of language learning [1]. Bilinguals' two languages interact at all facets of languages, including syntactic structure, semantic encoding, phonological category formation, and phonetic realization of individual sounds [2][3][4]. A big challenge in bilingual speech production is to uncover the mechanisms of phonological encoding of L2 sounds and to reveal the source of accented L2 speech. Models on bilingual speech production has attributed accented L2 production to inaccurate perception and wrong phonological categorization [5][6]. The current study aims to investigate the interaction of bilinguals' two phonological systems (Mandarin and English) and their phonetic realization in terms of acoustics and articulation of their rhotic sounds.

2. LITERATURE REVIEW

Previous models on bilingual speech production have proposed testable hypotheses about inaccurate L2 speech production. The Speech Learning Model (SLM) [5] attributes accented production to mismatches in the L2 phonological category formation. Late L2 speakers merge sounds similar in their two languages into the same phonological category, and thus produce them identically. The Perceptual Assimilation Model (PAM) has been extended to address problems in L2 phonological encoding and phonetic realization, which is called PAM-L2 [6]. PAM-L2 proposes three types of category formation: one L2 sound to one L1 category, two L2 sounds to one L1 category, and no L1-L2 category assimilation. Though PAM-L2 does not directly address cross-linguistic interaction in speech production, we can extend the hypotheses to the production domain since PAM assumes shared primitives in speech perception and production.

This study examines the acoustic and articulatory features of English rhotics produced by Mandarin-English bilinguals. Rhotics are prominent in perception and complex in production, and contribute notably to foreign accents. Studies have shown that English /ɹ/ is a difficult sound to acquire for English-speaking children and L2 speakers [7][8][9]. The reason that English /ɹ/ is hard to acquire is two-fold. First, the production of /ɹ/ requires more than one constriction in the vocal tract. It involves constrictions at lips, tongue blade, and pharyngeal cavity [10]. Second, various lingual gestures can be adopted, from tongue tip-up to tip-down, to produce similar sounds that can be identified as /ɹ/ [11][12]. In Mandarin Chinese, there is a rhotic sound that is perceptually similar to English /ɹ/, and has similar phonological distribution. Mandarin /ɹ/ is represented as "r" in the official romanization system *pinyin* and taught in school. Mandarin-English bilinguals, therefore, categorize Mandarin and English /ɹ/ sounds as the same phoneme. This provides a unique opportunity for us to test if bilingual speakers would produce identical sounds for the same phoneme in two languages or language-specific phonetic realization for each language.

By comparing the articulatory and acoustic features of English and Mandarin rhotics, the current study helps to shed light on the mechanisms of bilingual speech production and the interaction of bilinguals' two languages at the phonological and phonetic levels. To compare English and Mandarin rhotics, we also systematically described the acoustic and articulatory features of Mandarin rhotics, setting a baseline for cross-linguistic comparisons.

3. METHOD

English /ɹ/ and Mandarin /ɹ/ produced by Mandarin-English speakers were compared with English /ɹ/ produced by American English native speakers. Ten American English speakers and ten Mandarin-English bilingual speakers were recorded. The bilingual speakers grew up in Northern China (Beijing, Hebei and Shandong) and learnt English at around eight years old at school. At the time of recording, they have lived in Connecticut, USA for around 2.23 years. They have an average TOFEL speaking score of 23.6 out of 30. Their Mandarin speech were presented to two Mandarin monolingual speakers who were born and raised in Beijing, and were judged as not accented and not different from monolingual speakers. The English native speakers all lived in the east coast of the United States, and spoke a rhotic accent of English.

3.1 Stimuli

English stimuli included words containing prevocalic, postvocalic and intervocalic English /ɹ/ coarticulated with /ɑ æ ɛ ɪ ɔ u ʌ/ vowels, and syllabic /ɹ/. The /ɹ/ sound was embedded in different syllable positions – 47 prevocalic /ɹ/ in [#_V(C)], [C_V(C)] and [CC_V(C)] words, 15 postvocalic/ɹ/ in [V_#] and [V_C] words, 10 syllabic /ɹ/ in [C_C], [C_#] and [#_C] words, and 3 intervocalic words. The target words were produced with the carrier sentence "what a ___ again" when the word started with a consonant, and "Speak of ___ again" when the word started with a vowel. The carrier sentences were designed to have minimal lingual influence on the target words. The tongue was in a neutral position when producing a schwa (in *a*) and the production of a labiodental fricative (in *of*) involves no lingual movements. Mandarin stimuli included words containing prevocalic /ɹ/ coarticulated with /ʅ a ɤ u/ vowels, postvocalic /ɹ/ with /i ʅ ʅ y u a ɤ/ vowels, and syllabic /ɚ/[1] (see Table 1 for examples). The Mandarin words were produced with the carrier phrase / tʂɤ kɤ ___ ba/ "This is ___." (/ba/ is a sentence final particle in Mandarin). In order to control the length of the experiment, the English stimuli for Mandarin-speaking participants were a subset of that for the English participants, including 28 prevocalic, postvocalic, intervocalic and syllabic /ɹ/. Both English and Mandarin stimuli were repeated for ten times in a random order.

Table 1: Sample words from the Mandarin stimuli used in the experiment.

Syllable Positions	Words	Gloss	Mandarin
Prevocalic	/ɹɤ/	"hot"	热
Postvocalic	/tʂɤɹ/	"here"	这儿
Syllabic	/ɚ/	"son"	儿

3.2 Procedure

The articulatory data were collected with a Siemens ACUSON X300 ultrasound system at a frame rate of 36Hz. The participants were seated on a chair with their jaw rested on the ultrasound probe. The stimuli were presented on a laptop screen in front of the participants. The participants were instructed to read out the words on the screen. During recording, the participants' heads were unconstrained. The tongue surface contours were corrected to head-based coordinates using fifteen blue dots on participants' head and on the ultrasound probe. Participants' lips were painted blue to enhance contrast for tracking their movement. At the beginning of each session, the participants were asked to swallow some water. This was to capture the palate position of each speaker. The participants were also asked to say the peripheral vowels /a/, /i/ and /u/ in order to pinpoint the vowel space.

3.3 Statistical analysis

The recordings were segmented using the FAVE-align forced aligner and manually checked [13]. The formants were measured using LPC from PRAAT and logarithmically transformed [14]. The formant values were normalized by calculating z scores for each speaker. Since the phoneme /ɹ/ is an approximant that is heavily coarticulated with flanking vowels, we analyzed formant trajectories of /ɹ/ together with its adjacent vowels rather than a single time point to get a better picture about the acoustic characteristics of /ɹ/.

The tongue splines were drawn with an interactive MATLAB procedure 'GetContours' [15]. The splines were drawn on the lower boundary of the lighter line which represents the tongue-air interface in the ultrasound images. One hundred equally spaced data points on each spline were exported for analysis.

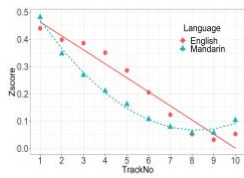
Smoothing spline ANOVA (SS-ANOVA) [16] and Growth curve analysis (GCA) [17] were used to compare tongue curves tracked from ultrasound data and formant trajectories using the “gss” and “lme4” packages in R [18]. In the GCA analysis, the formant trajectories were modeled with second-order orthogonal polynomials and fixed effects of Language (English/ Mandarin) or Group (bilingual/monolingual English) on all time terms. The models also included random effects of participants on all terms. SSANOVA and GCA yielded consistent results on all comparisons. Only GCA results were below reported due to page limit.

4. RESULTS

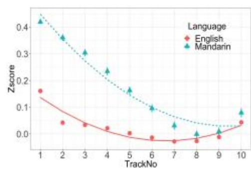
4.1 Acoustic and articulatory characteristics of Mandarin /ɹ/

Both acoustic and articulatory data showed that Mandarin prevocalic /ɹ/ was different from postvocalic and syllabic /ɹ/ in both acoustics and articulation. Similar to English /ɹ/, Mandarin postvocalic and syllabic /ɹ/ was characterized by a low F3 target, but the formant trajectories between English and Mandarin /ɹ/ were different. GCA on postvocalic and syllabic /ɹ/ showed that Language had a significant effect on F3 trajectories on linear term for postvocalic and syllabic positions (χ^2 (1) = 14.764, p = .000; χ^2 (1) = 8.973, p = .002). The modeled differences in F3 between English and Mandarin /ɹ/ are illustrated in Figure 1. Ultrasound data showed that Mandarin postvocalic and syllabic /ɹ/ was produced with either a retroflex or bunched tongue shape, similar to the tongue shape variations of English /ɹ/ (Figure 2).

Figure 1: Modeled F3 trajectories of Mandarin and English rhotics produced by native speakers.



(1) Postvocalic position



(2) Syllabic position

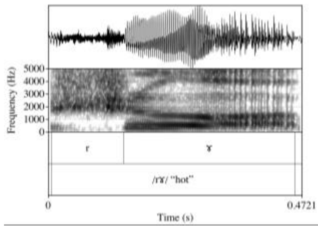
Figure 2: Tongue shapes of Mandarin postvocalic and syllabic /ɹ/ by Mandarin native speakers.

(1) A bunched tongue shape of Mandarin postvocalic /ɹ/ by speaker W1

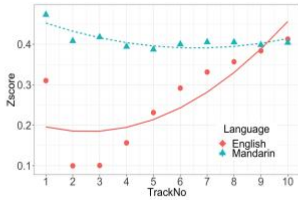
(2) A retroflex tongue shape of Mandarin postvocalic /ɹ/ by speaker W1

For Mandarin prevocalic /ɹ/, a clear frication was consistently observed at the beginning of each syllable, resembling the features of fricatives (Panel 1 in Figure 3). F3 lowing was observed in the transition from /ɹ/ to the following vowels, as shown in the second panel in Figure 3. The lingual gestures of Mandarin prevocalic /ɹ/ did not show retroflexion, and did not resemble English bunched /ɹ/ either. It was articulated in the same way as the post-alveolar affricates (/tʂ/ and /tʂʰ/), which were called “retroflex consonants” in Mandarin (Panel 3 of Figure 3).

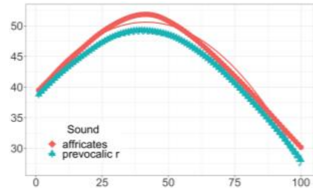
Figure 3: Formant trajectories and spectrogram of Mandarin prevocalic /ɹ /.



(1) Waveform and spectrogram of Mandarin prevocalic /ɹ/ in word /ɹɤ4/ “hot” by speaker W1



(2) Modeled F3 trajectories of Mandarin and English prevocalic rhotics and the following vowels



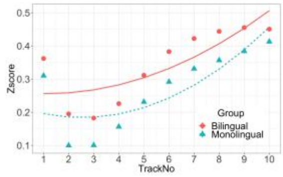
(3) Modeled tongue contours of Mandarin prevocalic /ɹ/ compared with Mandarin post-alveolar fricative /tʂ/ and /tʂʰ/ produced by speaker W1

In summary, the phonetic realization of Mandarin /ɹ/ was conditioned by syllable positions. Mandarin postvocalic and syllabic /ɹ/ was characterized by a low F3, while prevocalic /ɹ/ was realized with an initial frication followed by a lowering of F3 when transitioning to the following vowels.

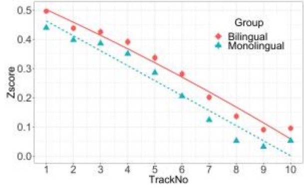
4.2 Production of English /ɹ/ by monolingual and bilingual speakers

GCA comparing bilingual and monolingual speakers’ production of English syllabic /ɹ/ showed that Group (bilingual/monolingual English) did not have an effect on the model fit on any time terms (Figure 4). For prevocalic and postvocalic conditions, the effect of Group only had a significant effect on the intercept (χ^2(1) = 15.956 , p = .000 ; χ^2 (1) = 4.052, p = .044). It indicated that bilinguals’ production of English /ɹ/ was not significantly different from native speakers’ production in F3 in all syllable positions. The pattern was consistent in consonant and vowel initial words.

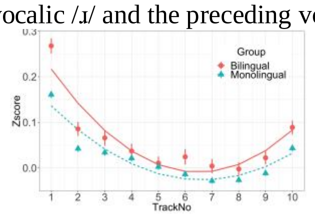
Figure 4: Modeled F3 trajectories of English /ɹ/ in prevocalic, postvocalic and syllabic positions produced by monolingual and bilingual speakers.



(1) Modeled F3 trajectories of English prevocalic /ɹ/ and the following vowels



(2) Modeled F3 trajectories of English postvocalic /ɹ/ and the preceding vowels

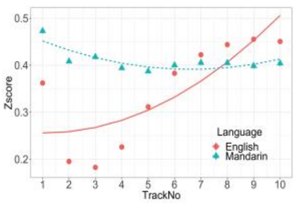


(3) Modeled F3 trajectories of English syllabic /ɹ/

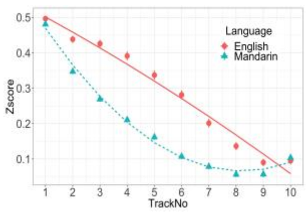
4.3 Production L1 and L2 /ɹ/ by bilingual speakers

GCA on bilingual acoustic data showed that Language (English/Mandarin) had a significant effect on the model fit on quadratic terms for prevocalic and postvocalic positions ($\chi^2(1) = 4.052$, $p = .044$; $\chi^2(1) = 250.142$, $p = .000$) and had a significant effect on linear terms for syllabic positions ($\chi^2(1) = 203.882$, $p = .000$) (Figure 5). It indicated that Mandarin-English bilinguals produced Mandarin and English /ɹ/ differently in all syllable positions.

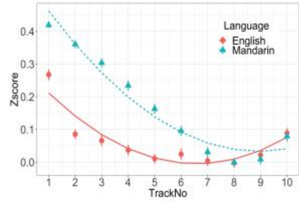
Figure 5: Modeled F3 trajectories of English and Mandarin rhotics by bilingual speakers.



(1) Modeled F3 trajectories of English and Mandarin prevocalic /ɹ/ with following vowels



(2) Modeled F3 trajectories of English and Mandarin postvocalic /ɹ/ and the preceding vowels



(3) Modeled F3 trajectories of English and Mandarin syllabic /ɹ/

5. DISCUSSION

There are three main results. First, Mandarin /ɹ/ has two different phonetic realizations. Mandarin postvocalic and syllabic /ɹ/ was characterized by a low F3, while prevocalic /ɹ/ was featured by an initial frication before F3 lowering. The result supports previous proposals that Mandarin phoneme /ɹ/ resembled fricatives phonetically when in prevocalic position [19]. This feature posts challenges for late bilinguals when learning English /ɹ/ because it is uniformly realized in all syllable positions in English, but has allophonic variations in Mandarin.

Second, Mandarin-English bilinguals successfully learned English /ɹ/ without interferences from the perceptually-similar Mandarin /ɹ/. Though Mandarin prevocalic /ɹ/ is distinct phonetically to its English equivalence, Mandarin-English bilinguals produced the correct phonetic realizations of English /ɹ/. SLM proposed that a new phonetic category would establish when a L2 sound was phonetically different from the closest L1 sound [5]. Our results showed that this hypothesis was not totally true. Phonetic differences did not lead to a brand new phonological category, but a new positional allophone with a distinct phonetic realization. Our results is consistent with the proposal in PAM-L2 that bilingual speakers could equivalent a L2 category to a L1 category if they shared the same phonological distribution, but could still be aware of the phonetic difference between the two sounds [6]. Similar phenomenon has been observed in French-English bilinguals as well [6].

Third, highly proficient Mandarin-English bilinguals produced distinct phonetic realizations for /ɹ/ in each language. This suggests that, at least for speakers who are highly proficient in their L2, late bilinguals could still learn language-specific phonetic details for their two languages.

In summary, the current study showed that highly proficient Mandarin-English bilinguals produced English /ɹ/ without L1 influences, and produced language-specific phonetic details for English and Mandarin /ɹ/.

Acknowledgements: The study was supported by NIH grant DC-002717 to Haskins Laboratories. The authors want to thank the comments from the two reviewers. The authors also want to thank Dr. Mark Tiede and Dr. Wei-rong Chen for his technical support during the ultrasound data collection and analysis, and Dr. Christine Shadle for helpful comments on the acoustic results.

7. REFERENCES

[1] Grosjean, F. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lang.*, 36(1), 3–15.

[2] Yip, V., Matthews, S. 2007. *The bilingual child.* Ernst Klett Sprachen.

[3] Heredia, R., Brown, J. 2004. Bilingual memory. *Handbook of Bilingualism*, 225–249.

[4] Runnqvist, E. , Fitzpatrick, I., Strijkers, K., Costa, A. 2012. An appraisal of the bilingual language production system: quantitatively or qualitatively different from monolinguals? *Handbook of Bilingualism and Multilingualism*, 244–265.

[5] Flege, J., Schirru, C., MacKay, I. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40(4), 467–491.

[6] Best C., Tyler M. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing Company, 1–47.

[7] Bradlow, A., Pisoni, D., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101(4), 2299–2310.

[8] Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *J. Phon.* 32(2), 233–250.

[9] Adler-Bock, M., Bernhardt, B., Gick, B., Bacsfalvi, P. 2007. The Use of Ultrasound in Remediation of North American English /r/ in 2 Adolescents. *Am. J. Speech Lang. Pathol.*, 16(2), 128–139.

[10] 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 *J. Acoust. Soc. Am*, 123(6), 4466-81.

[11] Delattre, P., Freeman, D. C. 1968. A dialect study of american R'S by x-ray motion picture. *Linguistics*, 6(44), 29–68.

[12] Mielke, J., Baker, A., Archangeli, D. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, 92(1), 101–140.

[13] Rosenfelder I., Fruehwald, J., Evanini, K., Yuan, J. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite.

[14] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer. http://www.praat.org/.

[15] Tiede, M., Whalen, D. W. 2016. GetContours: an interactive tongue surface extraction tool. *Ultrafest VII*.

[16] Gu, C. 2014. Smoothing Spline ANOVA Models: R Package gss. *J. Stat. Softw*. 58(5), 1-25. http://www.jstatsoft.org/v58/i05/.

[17] Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Tayler & Francis.

[18] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw*. 67(1), 1-48.

[19] Duanmu, S. 2007. *The phonology of standard Chinese*. New York : Oxford University Press.

---

[1] [ɿ ʅ] refer to two front high vowels in Mandarin, which are usually called apical vowels. The apical vowel [ɿ] appears only after dental fricative [s] and affricates [ts, tsh], and [ʅ] appears only after post-alveolar consonants [ʂ, tʂ, tʂh, ɹ].