華語文教學研究 18.4: 1-28, 2021

How to Teach Mandarin Alveolo-palatal [ɕ]: Pedagogical Suggestions Based on Articulatory Comparisons of Mandarin [ɕ] and English [ʃ]*

Yung-Hsiang Shawn CHANG
Department of English
National Taipei University of Technology

Abstract

Mandarin alveolo-palatal [ɕ] is a difficult sound for learners of Chinese as a Second Language and is commonly replaced with [ʃ] by English-speaking learners. Various pedagogical instructions have been proposed in the literature: some focus on the difference in tongue height or lip configuration, while others recommend a focus on differences in both articulatory parameters. The divergent pedagogical practices may be due to researchers' comparing "base of articulation" characteristics of Mandarin [ɕ] and English [ʃ] that reflect little individual variation. This study took a different approach by comparing the articulations of [ɕ] and [ʃ] produced by bilingual Mandarin-English speakers. The articulations were examined in terms of tongue-palate contact (using the direct linguography and palatography methods), tongue posture (using ultrasound imaging), and lip configuration (using lip videos). The results showed that the two fricatives were produced with the same linguopalatal contact patterns and tongue postures; however, different degrees of lip protrusion were observed. Therefore, we suggest that Mandarin [ɕ] can be taught as English [ʃ] with less lip protrusion.

* This research was supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant No. 106-2410-H-027-008-MY2. I am grateful to two anonymous reviewers for their helpful and insightful comments.

Keywords: articulation, English, Mandarin, second language (L2) pedagogy, ultrasound

1. Introduction

The alveolo-palatal [ɕ] has been considered among the most difficult consonant sounds for learners of Mandarin Chinese as a Second Language (e.g., Chin 1972; Lin 2005; Chung 2015; Wang and Chen 2020). For English-speaking learners of Mandarin, researchers have found that [ɕ] is perceptually assimilated to [ʃ] (Hao 2012; Wang and Chen 2020). In production, Chung (2015) observed that Mandarin alveolo-palatal [tɕ, tɕʰ, ɕ] are often replaced by English post-alveolar [dʒ, tʃ, ʃ] respectively. Yang and Yu (2019) corroborated this observation in their acoustic study, where American English-speaking learners produced Mandarin alveolo-palatals characterized by a low spectral center of gravity, similar to their English post-alveolar counterparts.

Articulatorily, Ladefoged and Wu's (1984) X-ray and palatographic study of three Beijing Mandarin speakers showed that [ɕ] is an alveolo-palatal with either alveolar or post-alveolar constriction. The constriction extends posteriorly to the palatal region with the tongue body raised to various degrees, rendering [ɕ] a palatalized alveolar or post-alveolar. Lee's (1999) study, based on palatograms and linguograms taken from four Beijing Mandarin speakers, indicated [ɕ] was a post-alveolar articulated with the blade or the anterior of the tongue dorsum. She proposed that [ɕ] was produced with a raised tongue body given the large lateral contact areas on the tongue and in the palatal region. English [ʃ], on the other hand, has been described as a labialized palato-alveolar (or post-alveolar) produced with a domed tongue front (i.e., laminal constriction) based on Ladefoged's (1957) palatographic and X-ray data. Narayanan et al.'s (1995) MRI study observed that [ʃ] was articulated as a post-alveolar with a raised tongue blade. Two of the four speakers in their study also displayed a raised tongue body, and all four speakers rounded their lips. In another MRI study, Proctor et al. (2006) observed two articulatory patterns for English [ʃ]: a narrow constriction extending from the palatal region to the post-alveolar region formed with a raised tongue body, and an apical alveolar with a slightly retroflexed tongue posture. Notably, while three speakers in their study produced [ʃ] with lip rounding (smaller lip aperture) and protrusion, those lip gestures were not observed in the [ʃ] production of the other two speakers. Taken together, the differences between the articulations of Mandarin [ɕ] and English [ʃ] may lie in multiple parameters including tongue posture, lip configuration, and linguopalatal contact (i.e., which part of the tongue is used to make a contact with which part of the palate).

Based on the aforementioned articulatory descriptions of Mandarin [ɕ] and English [ʃ], different pedagogical suggestions have been made. Chin (1972) argued that the pronunciation of Mandarin [ɕ] should be taught in relation to English [ʃ] with a focus on the height of tongue front, which is higher for Mandarin [ɕ] than for English [ʃ]. Li (2003) and H. Lin (2005), on the other hand, suggested that Mandarin [ɕ] can be taught as English [ʃ] with spread lips; but neither mentioned tongue posture. In yet another suggestion, Y.-H. Lin (2007) stressed that the two fricatives differ in both the height of tongue front and labiality; therefore, learners could approximate [ɕ] by producing [ʃ] with different configurations of both articulators.

These divergent pedagogical suggestions are based on the assumption that the articulations of Mandarin [ɕ] and English [ʃ] are homogeneous, such that categorical distinctions between the two sounds can be readily made. For example, Mandarin [ɕ] has been described by Ladefoged and Wu (1984) as having distinct tongue postures in contrast to the [ʃ] described by Ladefoged and Maddieson (1996); specifically, [ɕ] was produced with a higher tongue body while [ʃ] was produced with a slightly raised tongue front. In light of this, Chin's (1972) pedagogical suggestion would be the most accurate. However, one of the Mandarin speakers described by Ladefoged and Wu (1984) had a tongue posture rather similar to the one Ladefoged and Maddieson (1996) reported for [ʃ]. In this regard, Li's (2003) and H. Lin's (2005) recommendation for teaching Mandarin [ɕ] would seem more appropriate. That being said, all the aforementioned pronunciation pedagogical suggestions are correct but incomplete in that they were based on the comparisons of "base of articulation" characteristics (Abercrombie 1967) that reflect little individual variation. A similar argument for the articulatory comparisons between French and English coronal consonants has also been made in Dart (1998).

With pedagogical development in mind, this study takes a different approach to investigate the articulatory parameters that distinguish Mandarin [ɕ] from English [ʃ]. We recruited bilingual Mandarin-English speakers (or specifically, proficient L1 Mandarin-L2 English speakers) to produce both fricatives. This approach was similarly employed in Chen and Mok (2019), who enlisted bilingual Mandarin-English speakers to examine cross-linguistic articulations of /r/. The advantage of this approach is that cross-linguistic articulatory differences can be better observed given controlled individual variation (e.g., palate morphology and tongue size). Developing pedagogical recommendations based on this analysis is expected to more accurately inform English-speaking learners of Mandarin as to what articulatory parameters to pay attention to when producing Mandarin [ɕ] with reference to English [ʃ]. A potential complication with this approach, however, is that the nativeness of [ʃ] productions may be called into question, thereby challenging the validity of such a comparison. Given that bilingual speakers generally produce phonological targets with acoustic realizations that differ from their monolingual counterparts' (e.g., Li et al. (2017) for sibilant place of articulation contrast in Spanish-English bilingual speakers; Yang (2021) for stop voice onset time contrast in Mandarin-English bilingual speakers), comparing the spectral measurements of our bilingual participants' [ʃ] with those of monolingual English speakers in the literature will lose sight of the unique mechanisms adopted by bilingual speakers to develop phonetic categories in both languages. Therefore, to ensure the nativeness of our bilingual speakers' [ʃ] production, we enlisted native English listeners to rate the nativeness of the [ʃ] productions (Section 2.9).

In light of the reviewed articulatory parameters that may distinguish Mandarin [ɕ] from English [ʃ], this study examines the articulations of the two fricatives in terms of tongue-palate contact (using the direct linguography and palatography methods), tongue posture (using ultrasound), and lip configuration (using lip videos).

2. Methodology

2.1 Participants

The participants included five Taiwan Mandarin speakers (3 male, 2 female), born and raised in Taiwan. All participants were 18 to 21 year-old undergraduate students at the author's institution. Based on their self-reported TOEIC scores (mean: 877; s.d.: 56), their English proficiency was considered between upper-intermediate and advanced, according to the CEFR (Common European Framework of Reference for Languages) scale. None of the participants reported any speech or hearing disorders. All received compensation for their participation.

2.2 Procedure

Production data was collected in two sessions: a linguography/palatography session and an ultrasound session. To avoid fatiguing the participants, the two sessions were scheduled on different days. Some participants completed the ultrasound task after the linguography/palatography task, while others were assigned the tasks in the opposite order.

2.3 Stimuli

The stimuli consisted of 38 words, six of which contained the palatal fricatives: Mandarin *xia*, *xi*, *xiou* ([ɕa, ɕi, ɕo][1] carrying Tone 4; 夏, 夕, 秀, respectively) and English *sha*, *she*, *show* ([ʃa, ʃi, ʃo]). The Mandarin words were written in traditional characters, and the English words in the Roman alphabets. The participants reviewed the entire word list before data collection to make sure they knew how to pronounce all the stimuli. They were told that the presentation of English words would precede that of Mandarin ones. During data collection, the stimuli were displayed on a screen. The instructions were in English and

[1] These three words can be alternatively transcribed as [ɕia], [ɕi] and [ɕio] (in light of Lee and Zee 2003) or [ɕja], [ɕi] and [ɕjo]) (in light of Lin 2007). However, it was argued in Ladefoged and Maddieson (1996:150) that '[F]rom a phonetic point of view, there is nothing other than a normal transition between the initial consonant and the following vowel'. The [ɕa, ɕi, ɕo] transcription was also used in phonetic studies like Chiu et al. (2020), Lee-Kim (2014) and Stevens et al. (2004). Therefore, this paper follows their practice and uses the [ɕa], [ɕi] and [ɕo] transcription.

Mandarin during the presentation of English and Mandarin words respectively to help speakers switch to the designated language mode.

For the linguography and palatography task, all stimuli were produced once. For the ultrasound experiment, all stimuli were repeated six times, randomized within each block. All the stimuli were produced in isolation.

2.4 Data Collection Apparatus

Linguographic and palatographic data were collected following the direct method introduced in Dart (1998) and Ladefoged (2003). For the linguograms, a mixture of olive oil and activated charcoal powder was applied to the palate; after the utterance of a stimulus, the tongue was photographed to observe where contact had been made. For the palatograms, the same mixture was applied to the tongue, and after articulation, a mirror was placed in the speaker’s mouth to photograph the reflected image of the palate. To facilitate subsequent interpretation of the palatograms in terms of where the sound had been articulated, an alginate impression was taken of each participant’s upper palate and then cast in plaster.

The ultrasound session took place in a sound-treated booth at the author’s institution. Tongue imaging was obtained using a Chison ECO1 portable ultrasound machine with a transvaginal probe (V6-A). Each participant wore an Articulate Instruments ultrasound transducer stabilizer, attached to which were a AKG head-mounted microphone and two lip cameras (YI 4K action cameras)—one for the frontal view and the other for the profile view of lip movement. The scanning frequency of the ultrasound was set at 8 MHz with a scan depth of 10.3 cm. The audio recording was conducted using a Marantz PMD500 digital recorder at 44.1 kHz. The audio was synchronized with ultrasound imaging into an ultrasound movie at 51 fps using an AverMedia capture card (CV710). The lip cameras recorded both the audio and lip imaging at 60 fps.

2.5 Acoustic Measurements

An acoustic analysis of Mandarin [ɕ] and English [ʃ] productions was included to ensure the speakers were producing acoustically distinct [ɕ] and [ʃ] tokens. We employed the spectral center of gravity (COG) of frication, a measure that is commonly used to characterize the acoustics of fricatives across languages (e.g., Lee et al.'s (2014) comparison of English and Mandarin fricatives and Li et al.'s (2007) comparison of English, Japanese and Mandarin fricatives). To this end, the audio recordings were automatically segmented using the PennPhonetics Lab Forced Aligner (Yuan and Liberman 2008). To obtain the COG values of the fricatives, a Praat script developed by DiCanio (2021) was used. In light of Shadle's (2012) advocacy of time-averaged spectra in the spectral analysis of fricatives, as opposed to ensemble-averaged spectra around fricative midpoint, this script generates spectral moments measures (including COG) from fricative spectra by averaging discrete Fourier transform (DFT) at 15-ms intervals across the duration of the fricative.

2.6 Articulatory Classification of the Linguographic and Palatographic Data

The contact locations on the tongue and palate for Mandarin [ɕ] and English [ʃ] were determined based on previous linguographic and palatographic studies on Mandarin (Lee 1999; Huang et al. 2016) and English coronal fricatives (Dart 1998). As illustrated in Figure 1, articulations of coronal fricatives can involve the apex (apical), blade (laminal), or the anterior part of the dorsum (antero-dorsal) of the tongue. The classification is based on the location of the narrowest channel on the tongue.

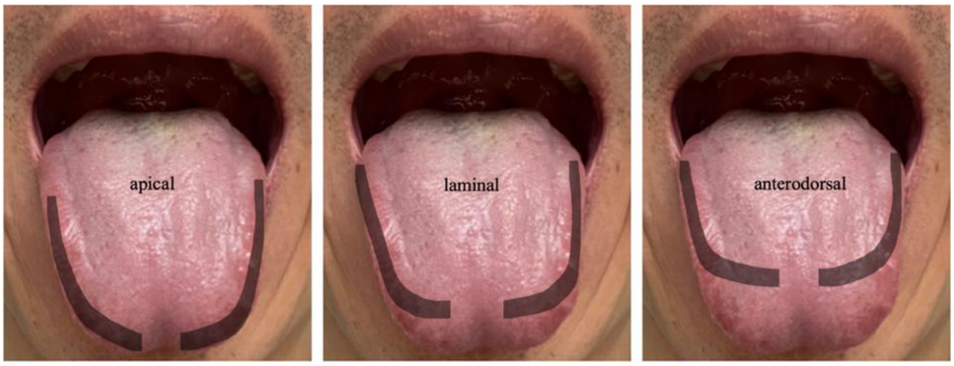


Figure 1: Linguographic Classifications.
Images are Reproduced from Figure 2 in Huang et al. (2016:750).

Figure 2 illustrates the way the palatograms are classified based on place of articulation, from dental to palatal. The place of articulation labels are based on the most anterior part of the contact area, regardless of the extent of posterior contact.

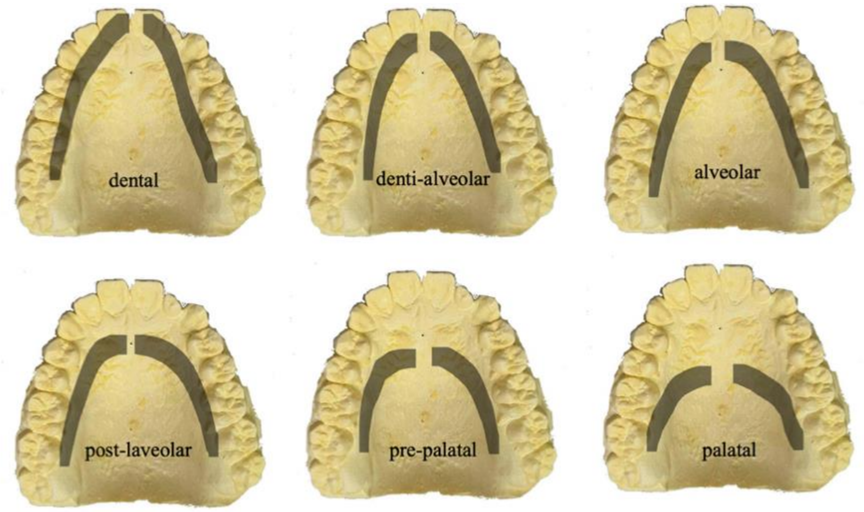


Figure 2: Palatographic Classifications.
Images are based on Figure 1 in Huang et al. (2016:749).

2.7 Ultrasound Tongue Data Processing

For tongue contour tracing, the ultrasound videos were imported into GetContours (Tiede 2015), a MATLAB-based platform, along with the accompanying TextGrid files. The ultrasound image corresponding to the midpoint of the Praat TextGrid-labeled fricative tokens was selected and the tongue contour was traced and optimized (Figure 3). Tiede (2020) provides more details on the tongue contouring procedure using GetContours.

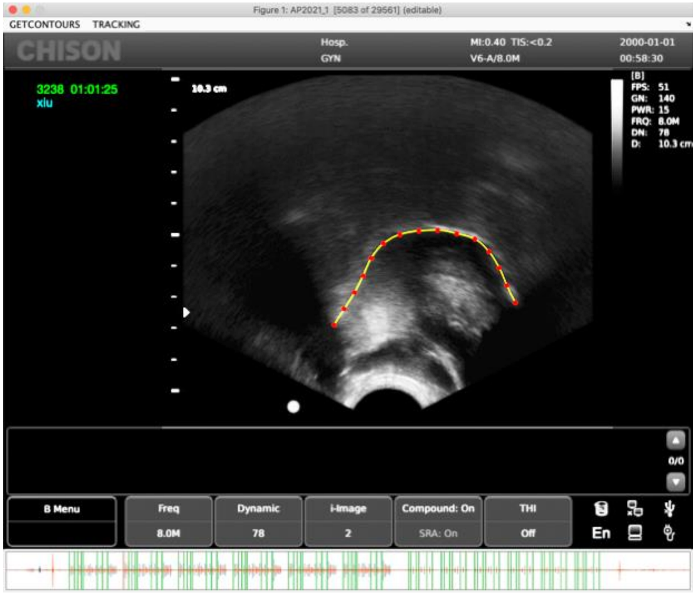


Figure 3: Display of GetContours

2.8 Lip Data Processing

Previous research on the lip gestures involved in labialized fricatives (e.g., Toda et al. 2003) has found that protrusion is more consistently employed than lip aperture. Protruding the lips lengthens the front cavity, considerably contributing to the lowering of the spectral frequencies of frication noise of post-alveolar fricatives. We therefore only analyzed the profile lip movement recordings, which captured lip protrusion. The ultrasound movie and lip video recording streams were synchronized using Chen’s (2021) Matlab script, which maximized the cross-correlation of the audio streams from the ultrasound movie and the lip video to compute their time lag. The TextGrid file used for the ultrasound movie was then time-shifted in Praat, giving us a TextGrid file that was time-aligned with the lip video. This process ensured that the ultrasound images and the lip images being analyzed corresponded to the same temporal point.

Along with the time-shifted TextGrid, the lip protrusion video was imported into ELAN (The Language Archive, Max Planck Institute for Psycholinguistics 2021), where we selected a neutral lip image (with the lips closed) and still lip images corresponding to the temporal midpoint of the labeled fricative. Lip protrusion was measured by calculating the difference between the participant's neutral lip position and lip displacement at the fricative midpoint, following the technique described in King and Ferragne (2020). Specifically, a reference line running parallel to the upper and lower panes of the neutral lip image was drawn to intersect the lip corner. Another line intersecting the upper and lower lip edges was also drawn. The value for the neutral lip position equaled the distance between the lip corner to the point where the two lines crossed. Using the same reference line, the procedure was performed again to obtain the value at the fricative midpoint. The lip protrusion measurement of a given fricative equaled the difference between the two values (see Figure 4). Before recording the lip movement, a ruler was positioned next to the participant's lips so the lip protrusion measurements could be scaled and converted to millimeters.

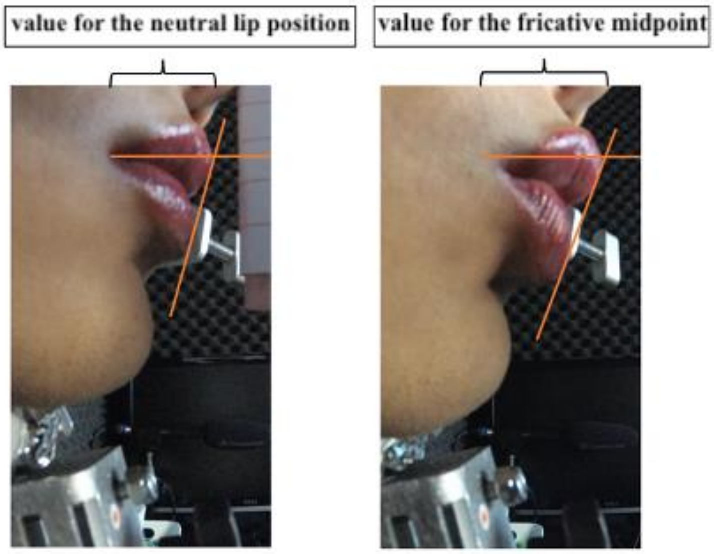


Figure 4: Lip Protrusion was Measured as the Difference between the Value for the Neutral Position and the Value for the Fricative Midpoint

2.9 Perceptual Judgments of English Tokens

To validate the direct comparisons of our five bilingual (L1 Mandarin-L2 English) speakers' Mandarin [ɕ] and English [ʃ] articulations, we first verified how natively they produced *sha*, *she*, and *show*. To this end, two native English listeners were asked to rate all the English [ʃ]-initial tokens. The tokens were randomized and presented using the E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA 2016). The two listeners were informed that they would hear English words starting with “sh”. They were instructed to rate the pronunciation of the consonant in each word on a 5-point Likert scale presented on a computer monitor, with 5 being the most native-like and 1 being the least. Three practice trials were given before the task began. After the listeners made their decision, two boxes appeared below the scale asking them whether they would like to listen and rate again or proceed with the next sound. The listeners were allowed to listen to each sound as many times as they deemed necessary. All responses were automatically logged in E-Prime.

3. Results

3.1 Acoustic Results

The COG values across vowel contexts for the group and for each participant are plotted in Figure 5 and Figure 6 respectively. From Figure 5, we can see that Mandarin [ɕ] had noticeably higher COG values than English [ʃ] in the [a] and [i] contexts, whereas the values for [ɕ] were lowest in the [o] context.

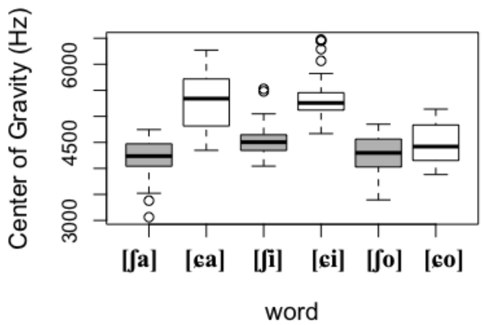


Figure 5: Spectral COG Values of Mandarin [ɕ] and English [ʃ] in the [a, i, o] Contexts from the Five Speakers. English Productions are Shaded for Clarity.

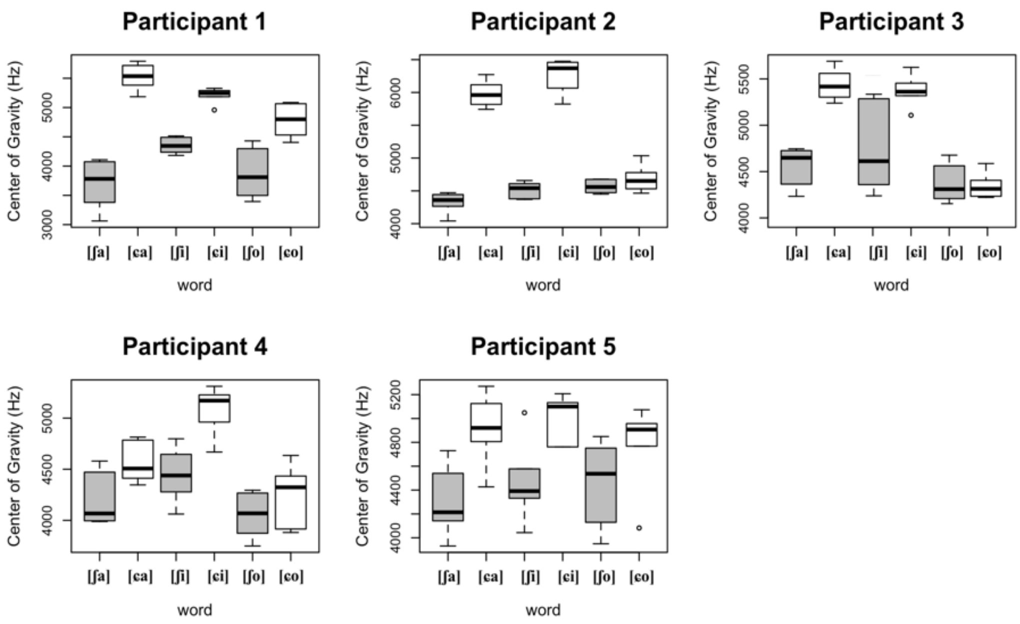


Figure 6: Spectral COG Values of Mandarin [ɕ] and English [ʃ] Across the [a, i, o] Contexts. English Productions are Shaded for Clarity.

The five participants' COG values were analyzed with a linear mixed-effects model, with LANGUAGE and VOWEL as fixed effects and PARTICIPANT intercept as the random effect. The analyses were conducted in R (R Development Core Team 2010) using the lme4 (Bates et al. 2014) and lmerTest packages (Kuznetsova et al. 2017). The best-fitting model (see Table 1) revealed the COG data were significantly influenced by LANGUAGE, VOWEL, and the LANGUAGE×VOWEL interaction. English [ʃ] in the [a] context served as the reference level. Specifically, it was shown that Mandarin [ɕ] had significantly higher COG than English [ʃ] (β = 1068.23, SE = 99.35, $p < .0001$) and that fricatives in the [i] context had significantly higher COG, in comparison to the [a] context (β = 327.19, SE = 99.35, $p = .0012$). The interaction further revealed significant lower COG for Mandarin [ɕ] in the [o] context (β = -844.6, SE = 140.5, $p < .0001$).

Table 1: Summary of the Best-fitting Linear Mixed-effects Model for Spectral COG

<table>
<tr><td colspan="5">Formula:
COG~LANGUAGE+VOWEL+LANGUAGE*VOWEL+(1|PARTICIPANT)</td></tr>
<tr><td>Effects</td><td colspan="4">Measures</td></tr>
<tr><td></td><td>Estimated Coefficient</td><td>Std. Error</td><td>t-value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>4214.64</td><td>119.00</td><td>35.418</td><td>< .0001</td></tr>
<tr><td>LANGUAGEMandarin</td><td>1068.23</td><td>99.35</td><td>10.753</td><td>< .0001</td></tr>
<tr><td>VOWELi</td><td>327.19</td><td>99.35</td><td>3.293</td><td>.0012</td></tr>
<tr><td>VOWELo</td><td>49.83</td><td>99.35</td><td>0.502</td><td>.6166</td></tr>
<tr><td>LANGUAGEMandarin: VOWELi</td><td>-222.87</td><td>140.50</td><td>-1.586</td><td>.1145</td></tr>
<tr><td>LANGUAGEMandarin: VOWELo</td><td>-844.60</td><td>140.50</td><td>-6.011</td><td>< .0001</td></tr>
</table>

The average spectral COG value of the five participants' English [ʃ] productions (calculated from all three vowel contexts) was 4340 Hz. This value appears comparable to the averaged spectral COG value of English [ʃ] and [ʒ][2] (4229 Hz) reported in Jongman et al. (2000). However, their COG measurement procedures (e.g., window size, window placement) differ from those in the current study; arguing that our bilingual speakers had similar English fricative realizations as the English speakers in Jongman et al. is therefore not warranted.

3.2 Perceptual Rating Results

The perceptual rating scores averaged from two native English listeners are listed for each participant in Table 2. Overall, the [ʃ] productions received a nativeness score of 4.94 out of 5.

[2] This value was averaged between [ʃ] and [ʒ] in Jongman et al. (2000); no separate values for [ʃ] and [ʒ] were reported in their study. Jongman et al. did note that English voiceless fricatives had slightly higher spectral COG than their voiced counterparts, although the difference was slight.

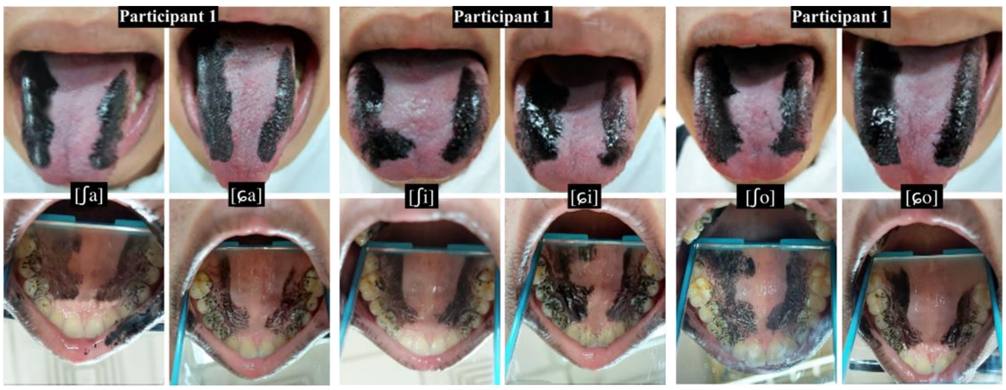
Table 2: Averaged Perceptual Rating Scores on the Participants' English [ʃ] Productions.

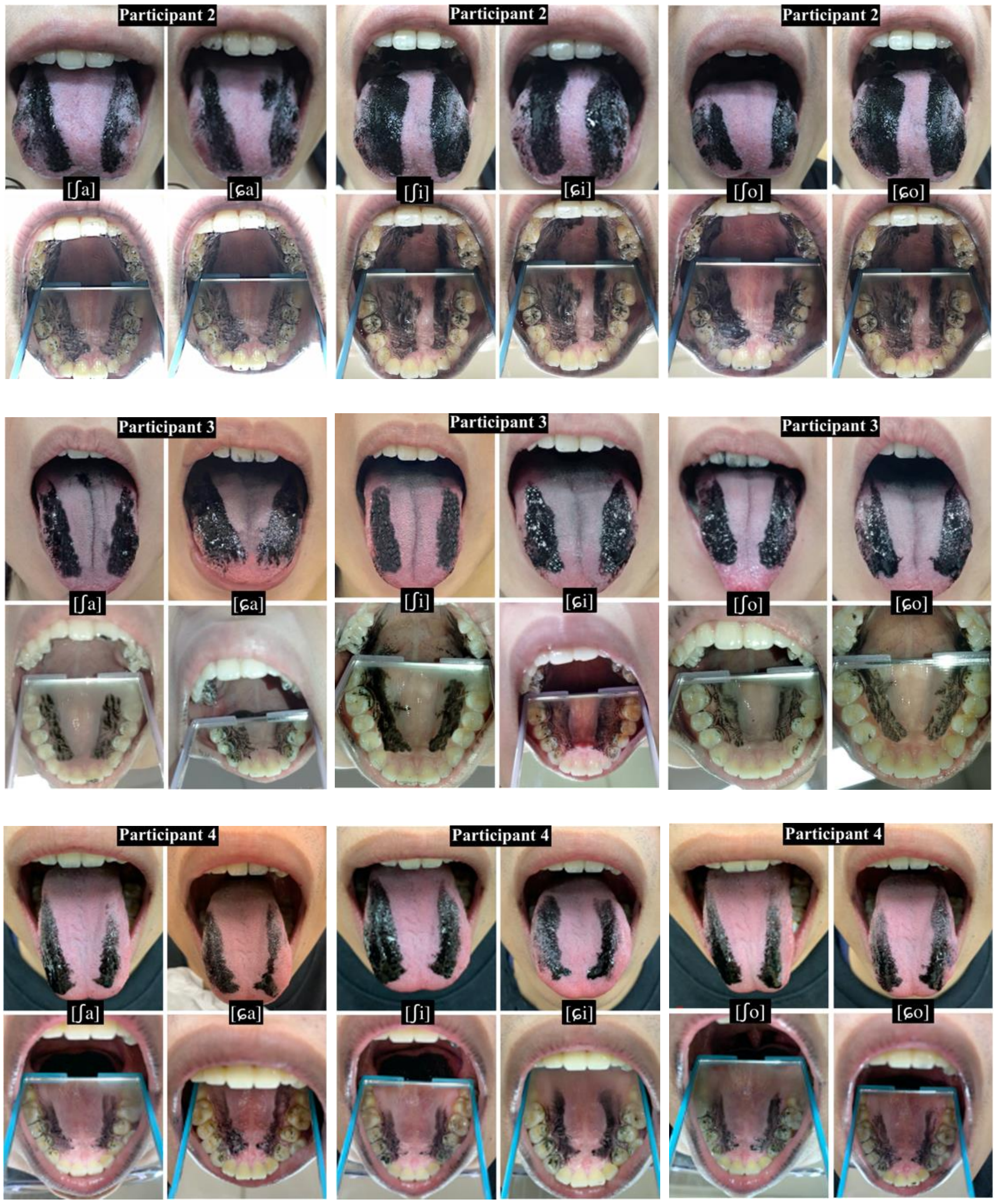
Participant	Vowel context		
	[a]	[i]	[o]
1	4.916	5	5
2	4.916	4.916	5
3	5	4.916	5
4	5	4.833	4.916
5	5	4.916	4.916

3.3 Articulatory Results

3.3.1 Linguopalatal Contact Patterns

The linguograms (Figure 7, top) showed that for [ɕ] and [ʃ] the narrowest part of the constriction channel was on the blade (for Participants 1-4) or anterodorsum (for Participant 5). The palatograms (Figure 7, bottom) revealed the maximal constriction of [ɕ] and [ʃ], across [a, i, o] vowel contexts, fell within the post-alveolar region. That is, the linguopalatal contact was similar for [ɕ] and [ʃ]: laminal post-alveolar or anterodorsal post-alveolar. Additionally, the linguographic and palatographic data showed that all the [ɕ] and [ʃ] productions featured relatively large lateral contact areas on the tongue and palate, respectively, which was similarly found in Huang et al. (2016) and Lee (1999). This pattern is indicative of palatalization, where the constriction was made with a raised tongue body that extends from the post-alveolar area to the palatal zone.




Participant 2
[ʃa]
[ɕa]
Participant 2
[ʃi]
[ɕi]
Participant 2
[ʃo]
[ɕo]
Participant 3
[ʃa]
[ɕa]
Participant 3
[ʃi]
[ɕi]
Participant 3
[ʃo]
[ɕo]
Participant 4
[ʃa]
[ɕa]
Participant 4
[ʃi]
[ɕi]
Participant 4
[ʃo]
[ɕo]

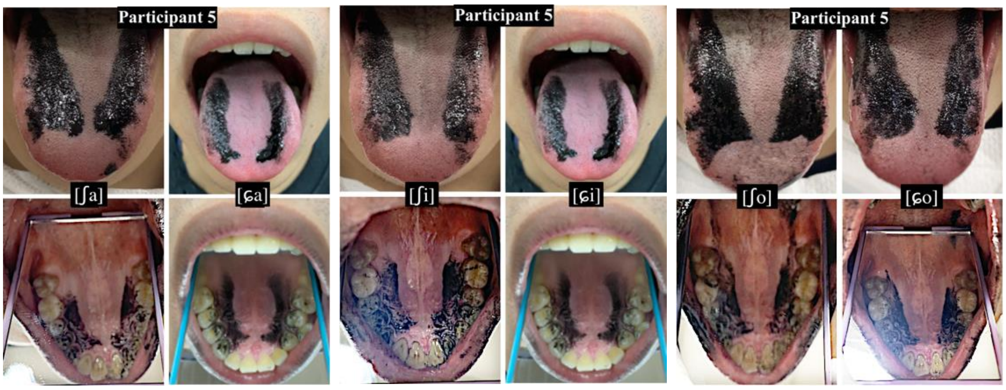


Figure 7: Linguograms (Top) and Palatograms (Bottom) of Mandarin [ɕ] and English [ʃ] Across [a, i, u] Vowel Contexts for the Five Speakers

3.3.2 Ultrasound Tongue Posture

Tongue postures traced in GetContours used the Cartesian coordinate system. They were first converted into polar coordinates, in light of Heyne and Derrick (2015), and then analyzed and visualized through smoothing-spline analysis of variance (SSANOVA) (Gu et al. 2004; Davidson 2006; Wang 2011), with 95% confidence intervals around the smoothed curves. This allows a visual representation of statistical significance. At any point the two curves are separate and their confidence intervals do not overlap, the differences between the curves are significant. However, when the curves are separate but their confidence intervals do overlap, the differences between the two curves are not significant. The *tongue_ssanova.r* function package developed by Mielke (2017) was used to perform the analysis.

In Figure 8 (with the tongue tip on the right side), we can see that the tongue shapes of Mandarin [ɕ] and English [ʃ], across all three vowel contexts, were nearly identical for all participants. The tongue postures all featured a high tongue body position, confirming our speculations about palatalization of both fricatives (Section 3.3.1).

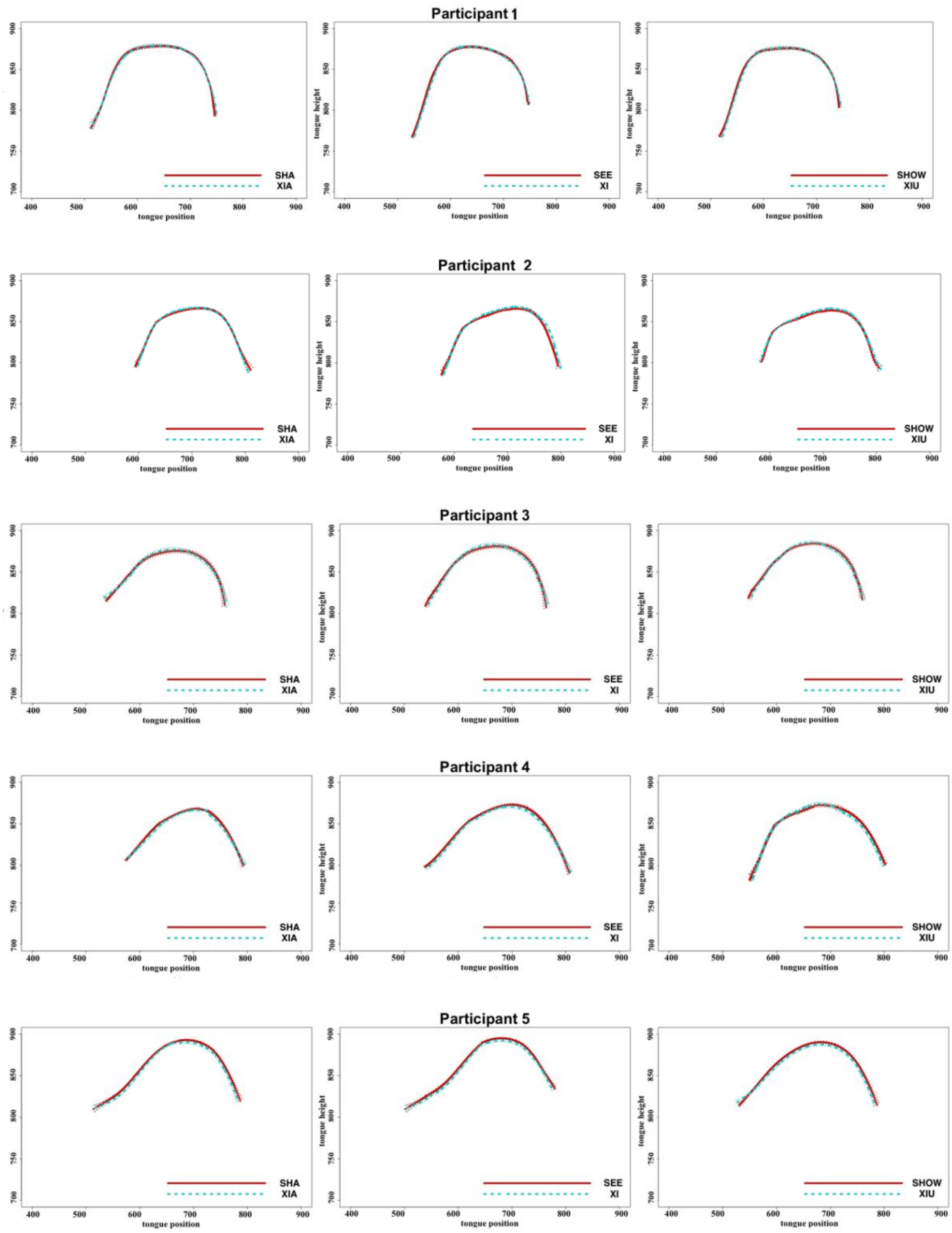


Figure 8: SS ANOVA Results of Mandarin [ɕ] and English [ʃ] in the [a, i, o] Contexts. The Tongue Tip is on the Right.

3.3.3 Lip Protrusion

The changes in lip protrusion from the neutral lip setting are plotted for the group and for each participant in Figures 10 and 11 respectively. It could be seen that the English [ʃ] tokens generally involved a greater degree of lip protrusion than their Mandarin counterparts. Some negative lip protrusion values were observed for Mandarin [ɕ] in the [a, i] vowel context, indicative of lip spreading.

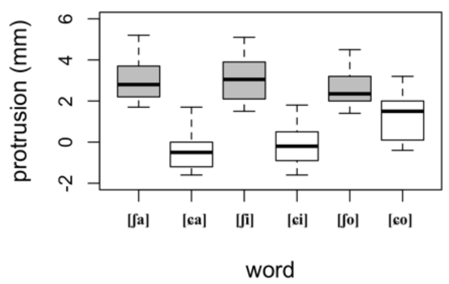


Figure 9: Lip Protrusion of Mandarin [ɕ] and English [ʃ] Across the [a, i, o] Contexts from the Five Speakers. English Productions are Shaded for Clarity.

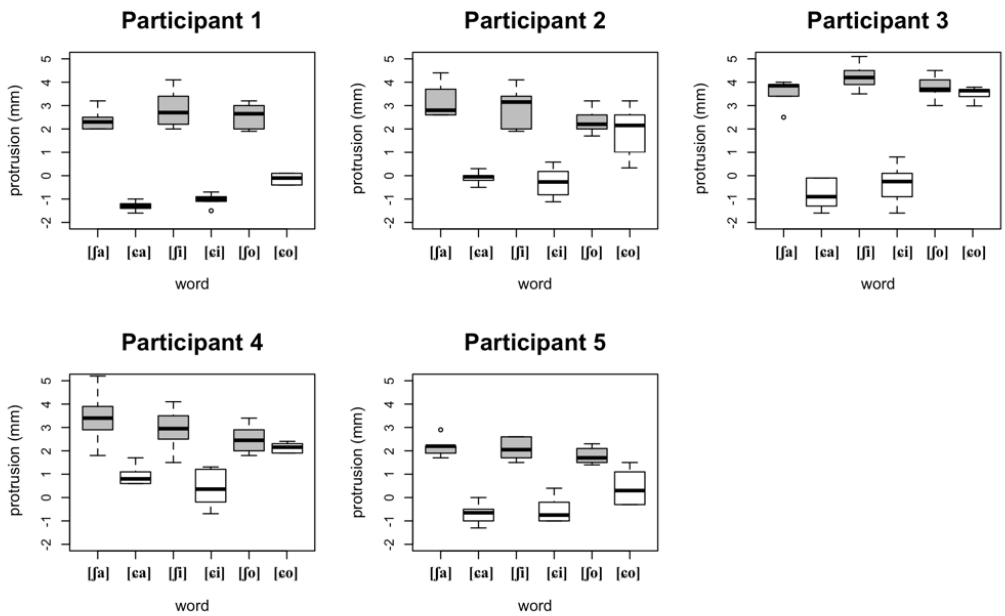


Figure 10: Lip Protrusion of Mandarin [ɕ] and English [ʃ] Across the [a, i, o] Contexts. English Productions are Shaded for Clarity.

Lip protrusion data were analyzed with a linear mixed-effects model, with LANGUAGE and VOWEL as fixed effects and PARTICIPANT intercept as the random effect. The best-fitting linear mixed-effects model (see Table 3) indicated that lip protrusion was significantly influenced by LANGUAGE, VOWEL and the LANGUAGE×VOWEL interaction. English [ʃ] in the [a] context served as the reference level. In particular, Mandarin [ɕ] was produced with significantly smaller degrees of lip protrusion than their English counterparts (β = -3.3367, SE = 0.1951, p < .0001), and marginally significantly more lip protrusion was involved in the [o] context (β = -0.36, SE = 0.1951, p = .0667). In addition, the interaction showed significantly more lip protrusion for Mandarin [ɕ] in the [o] context (β = 1.9467, SE = 0.2759, p < .0001).

Table 3: Summary of the Vest-fitting Linear Mixed-effects Model for Lip Protrusion

Formula: LIP~LANGUAGE+VOWEL+LANGUAGE*VOWEL+(1|PARTICIPANT)				
Effects	Measures			
	Estimated Coefficient	Std. Error	*t*-value	Pr(>|*t*|)
(Intercept)	2.9467	0.3006	9.801	< .0001
LANGUAGEMandarin	-3.3367	0.1951	-17.103	< .0001
VOWELi	0.0600	0.1951	0.308	.7588
VOWELo	-0.3600	0.1951	-1.845	.0667
LANGUAGEMandarin: VOWELi	0.1867	0.2759	0.677	0.4996
LANGUAGEMandarin: VOWELo	1.9467	0.2759	7.056***	< .0001

One reviewer pointed out that Mandarin [ɕ]'s having lower COG and being more protruded in the [o] context may contribute to the discussion of a long debate on the phonological status of [ɕ] (see Lu (2014) for a review)—whether [ɕ] should be considered an allophonic variant of [s] whose surface realization is conditioned by the following sound. To that end, however, [ɕ] followed by the rounded [y] or its homorganic glide [ɥ] must also be investigated. Per the reviewer's suggestion, we conducted a small follow-up experiment with Participant 1. Data was collected with the same apparatus described in Section 2.4. Participant 1 produced six repetitions of *xu*, *xue*, *xuan* ([ɕy, ɕɥe, ɕɥɛn] carrying Tone 4; 續, 穴, and 炫, respectively) in addition to *xia*, *xi*, and *xiu*. The COG data and lip protrusion data are plotted in Figures 11 and 12. It could be seen that when Mandarin [ɕ] was followed by [y] or [ɥ], the COG values were noticeably lower and the lips were more protruded than those of [ɕo]. Given that [y, ɥ, o] are all rounded sounds and should have similar coarticulatory effects on [ɕ], the most probable account may be that [ɕo] is in fact [ɕjo] (where an unrounded glide comes between the fricative and the rounded [o]) in relation to [ɕy, ɕɥe, ɕɥɛn] (where the fricative is immediately followed by a rounded sound). While this pattern must be interpreted in the context of a very small data set, it is promising for future studies with more participants.

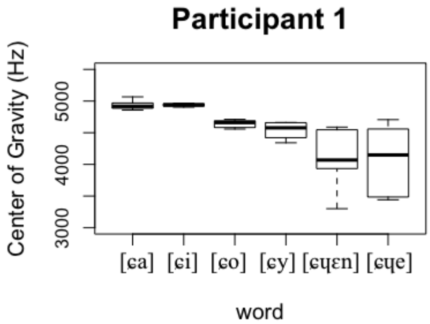


Figure 11: Spectral COG Values of Mandarin [ɕ] Across Various Vowel Contexts

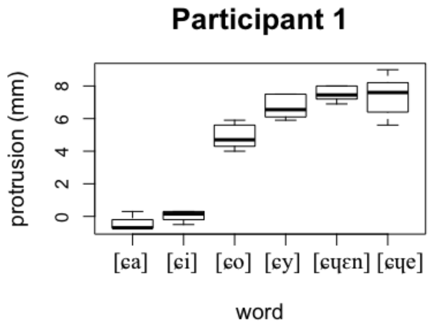


Figure 12: Lip Protrusion of Mandarin [ɕ] Across Various Vowel Contexts

4. General Discussion

The goal of this study was to determine the articulatory parameters critical to distinguishing Mandarin [ɕ] from English [ʃ]; in turn, these parameters could inform the optimal approach to teaching Mandarin [ɕ] with reference to [ʃ], a sound that is commonly substituted for [ɕ] by English-speaking learners of Mandarin. The literature diverges on the best pedagogical practice for teaching Mandarin [ɕ]. This may be due to researchers directly comparing the articulatory data of Mandarin [ɕ] and English [ʃ] from different sources, without considering the diversity of individual variation. Unlike acoustic data (e.g., vowel formant, amplitude), articulatory data (especially those that are relatively small in size, like the current study) are not easily normalized, making it difficult to directly compare individuals or languages. In this regard, this study took a different approach and compared [ɕ] and [ʃ] tokens produced by proficient L1 Mandarin-L2 English speakers. Assuming the participants' L2 English [ʃ] productions were native-like, then a direct comparison of their [ɕ] and [ʃ] articulations would be valid. As shown in Section 3.2, our bilingual participants' [ʃ] tokens were judged as very native-like (4.94 out of 5). As such, we could directly compare the articulatory data of the Mandarin [ɕ] and English [ʃ] productions.

Three articulatory parameters were investigated in this study: linguopalatal contact, tongue posture and lip protrusion. The linguographic and palatographic data showed that our participants produced the two fricatives with the same contact locations on the tongue and the palate: laminal post-alveolar or anterodorsal post-alveolar. The large lateral contact areas on the tongue suggested that the constriction was made with a raised tongue body. This was confirmed with the ultrasound data, which showed that all five speakers' [ɕ] and [ʃ] tokens were produced with the same tongue posture: a raised tongue body with the tongue front facing down. The Mandarin [ɕ] articulations observed in our study—a palatalized post-alveolar— generally aligned with Ladefoged and Wu's (1984) and Lee's (1999) articulatory descriptions of [ɕ] based on Beijing Mandarin speakers' data. The English [ʃ] articulations of our speakers had the same place of articulation (i.e., post-alveolar) as described in Ladefoged and

Maddieson (1996). However, while Ladefoged and Maddieson described [ʃ] as being articulated with a slightly domed tongue front posture, we observed a raised tongue body posture. The raising of the tongue body, nevertheless, matches one of the two articulatory patterns observed for English [ʃ] in Narayanan et al. (1995) and Proctor et al. (2006). As for lip protrusion, the speakers' lip video data indicated that their lips were more protruded for English [ʃ] than for Mandarin [ɕ] across [a, i, u] contexts. The presence of lip protrusion in English [ʃ] is generally in line with findings in Ladefoged (1957), Narayanan et al. (1995) and Proctor et al. (2006).

Based on our observation that the speakers in our study produced Mandarin [ɕ] and English [ʃ] with comparable tongue postures and linguopalatal contact but different degrees of lip protrusion, we suggest that English-speaking learners of Mandarin can be instructed to produce [ɕ] as [ʃ] with *less lip protrusion*. This pedagogical instruction applies to [ɕ] in both unrounded and rounded vowel contexts. Although Mandarin [ɕo] was articulated with more protruded lips (compared to [ɕa] and [ɕi]), the degree of protrusion was still significantly less than that in the production of its English counterpart, [ʃo].

Importantly, we do not claim that all bilingual Mandarin-English speakers produce the same articulations for Mandarin [ɕ] and English [ʃ]; in fact, one speaker in our pilot study (Chang 2019) made a clear tongue posture distinction between the two fricatives. Instead, this study argues that by varying the labial gesture, a Mandarin [ɕ] can be modified into a perceptually satisfactory English [ʃ] or vice versa. This is pedagogically viable, given that lip configurations can be easily observed, whereas learners have a harder time intuiting the tongue posture and tongue-palate contact that they cannot see. Future studies can test whether English-speaking learners of Mandarin could receive high nativeness rating scores on [ɕ] that is produced as [ʃ] with less lip protrusion.

References

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Chang, Yung-hsiang Shawn. 2019. A comparison of English and Mandarin palatal fricatives with articulatory data. Poster presented at *the Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019*, May 24, 2019. Seoul: Hanyang University.

Chin, Yin-Lien C. 1972. How to teach Mandarin retroflex and palatal sounds to native speakers of English. *Journal of the Chinese Language Teachers Association* 7.3: 77-81.

Chen, Shuwen, and Peggy Pik Ki Mok. 2019. Speech production of rhotics in highly proficient bilinguals: Acoustic and articulatory measures. In *Proceedings of ICPhS 2019*.

Chen, Weirong. 2021. *Align Wave*. MATLAB Central File Exchange. Accessed online, May 1, 2021. https://www.mathworks.com/matlabcentral/fileexchange/50473-align-wave

Chiu, Chenhao, Po-Chun Wei, Masaki Noguchi, and Noriko Yamane. 2020. Sibilant fricative merging in Taiwan Mandarin: An investigation of tongue postures using ultrasound imaging. *Language and Speech* 63.4: 877-897.

Chung, Raung-Fu. 2015. *Mandarin Phonetics and its Implications in Teaching* (2nd edition). Taipei: Cheng Chung Book Company.

Dart, Sarah N. 1998. Comparing French and English coronal consonant articulation. *Journal of phonetics* 26.1: 71-94.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120.1: 407-415.

DiCanio, Christian. 2021. *Time Averaging for Fricatives*. Praat script. Accessed online, May 1, 2021. http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives.praat

ELAN (Version 6.2) [Computer software]. 2021. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Accessed online, June 1, 2021. https://archive.mpi.nl/tla/elan

Gu, Jiayun, Tim Bressmann, Kevin Cannons, and Willy Wong. 2004. *The Ultrasonographic Contour Analyzer for Tongue Surfaces (Ultra-CATS).*

Toronto: University of Toronto.

Hao, Yen-Chen. 2012. *The Effect of L2 Experience on Second Language Acquisition of Mandarin Consonants, Vowels, and Tones*. Bloomington: Indiana University Ph. D. dissertation.

Heyne, Matthias, and Donald Derrick. 2015. Using a radial ultrasound probe's virtual origin to compute midsagittal smoothing splines in polar coordinates. *The Journal of the Acoustical Society of America* 138: EL509-EL514.

Huang, Ting, Yueh-Chin Chang, and Feng-fan Hsieh. 2016. Articulatory Characteristics of the Coronal Consonants in Malaysian Mandarin: With Special Reference to the Non-"Canonical" Sibilants. *Tsing Hua Journal of Chinese Studies* 46.4: 743-783.

Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108.3: 1252-1263.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune HB Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82.1: 1-26

King, Hannah, and Emmanuel Ferragne. 2020. Loose lips and tongue tips: The central role of the/r/-typical labial gesture in Anglo-English. *Journal of Phonetics* 80: 100978.

Ladefoged, Peter. 1957. Use of palatography. *Journal of Speech and Hearing Disorders* 22.5: 764-774.

Ladefoged, Peter. 2003. *Phonetic Data Dnalysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA: Wiley-Blackwell.

Ladefoged, Peter, and Zongji Wu. 1984. Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12.3: 267-278.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lee, Wai-Sum. 1999. An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin. In *Proceedings of the 14th International Congress of Phonetic Sciences*.

Lee, Wai-Sum, and Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the*

International Phonetic Association 33.1: 109-112.

Lee-Kim, Sang-Im. 2014. Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study. *Journal of the International Phonetic Association* 44.3: 261-282.

Lee, Chao-Yang, Yu Zhang, and Ximing Li. 2014. Acoustic characteristics of Mandarin voiceless fricatives. *Journal of Chinese Linguistics* 42: 150-171.

Li, Chris Wen-Chao. 2003. Phonetic detail in the teaching of Mandarin pronunciation. Presented in *the Chinese Language Teachers Association 2003 Spring Workshop*. May 2003. Stanford University.

Li, Fangfang, Jan Edwards, and Mary Beckman. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In *Proceedings of the XVIth International Congress of Phonetic Sciences*.

Li, Fangfang, Ferenc Bunta, and J. Bruce Tomblin. 2017. Alveolar and postalveolar voiceless fricative and affricate productions of Spanish–English bilingual children with cochlear implants. *Journal of Speech, Language, and Hearing Research* 60.9: 2427-2441.

Lin, Hua. 2005. Understanding problems in learning Mandarin consonants by monolingual speakers of English. *Journal of Canadian Teachers of Chinese as a Second Language* 15.1: 1-18.

Lin, Yen-Hwei. 2007. *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Lu, Yu-An. 2014. Mandarin fricatives redux: The psychological reality of phonological representations. *Journal of East Asian Linguistics* 23.1: 43-69.

Mielke, Jeff. 2017. *Tongue_ssanova. R function package*. Accessed online, July 1, 2021. https://phon.wordpress.ncsu.edu/lab-manual/ultrasound-and-video/working-with- data/

Narayanan, Shrikanth S., Abeer A. Alwan, and Katherine Haker. 1995. An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America* 98.3: 1325-1347.

Proctor, Michael, Christine Shadle, and Khalil Iskarous. 2006. An MRI study of vocalic context effects and lip rounding in the production of English sibilants. In *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*.

Psychology Software Tools, Inc. [E-Prime 3.0]. 2016. Accessed online, June 1, 2021. https://support.pstnet.com/

R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing* [computer program]. Accessed online, July 1, 2021. http://www.R-project.org/

Shadle, Christine. H. 2012. Acoustics and aerodynamics of fricatives. *Handbook of Laboratory Phonology*, eds. by Abigail C. Cohn, Cécile Fougeron, and Marie K. Huffman, 511-526. New York, NY: Oxford University Press.

Stevens, Kenneth N., Zhiqiang Li, Chao-Yang Lee, and Samuel Jay Keyser 2004. A note on Mandarin fricatives and enhancement. *From Traditional Phonology to Modern Speech Processing*, ed. by Gunnar Fant, 393-403. Beijing: Foreign Language Teaching and Research Press.

Tiede, Mark. 2015. *GetContours, Software Supporting Tongue Contour Extraction from Ultrasound Images*. Accessed online, June 1, 2021. https://github.com/mktiede/GetContours

Tiede, Mark. 2020. How to "GetContours" from ultrasound imaging. Presented at *Ultrasfest IX*. Indianapolis, IN.

Toda, Martine, Shinji Maeda, Andreas J. Carlen, and Lyes Meftahi. 2003. Lip protrusion/rounding dissociation in French and English consonants: /w/ vs. /S/ and /Z/. In *Proceedings of the 15th International Congress of Phonetic Sciences*.

Wang, Yuedong. 2011. *Smoothing Splines: Methods and Applications*. Chapman and Hall, Boca Raton, FL.

Wang, Xinchun, and Jidong Chen. 2020. The acquisition of Mandarin consonants by English learners: The relationship between perception and production. *Languages* 5.2: 20.

Yang, Chungshen, and Alan C. Yu. 2019. The acquisition of Mandarin affricates by American second language learners. *Taiwan Journal of Linguistics* 17.2: 91-122.

Yang, Jing. 2021. Comparison of VOTs in Mandarin–English bilingual children and corresponding monolingual children and adults. *Second Language Research* 37.1: 3-26.

Yuan, Jiahong, and Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. *Journal of Acoustical Society of America* 123.5: 3878.

[審查：2021.9.7 修改：2021.10.5 接受：2021.10.27]

張詠翔
Yung-Hsiang Shawn CHANG
10608 臺北市忠孝東路三段 1 號　國立臺北科技大學應用英文系
Department of English
National Taipei University of Technology
No.1, Zhongxiao E. Rd., Sec. 3, Taipei City 10608, Taiwan
shawnchang@mail.ntut.edu.tw

如何教中文齦顎擦音[ɕ]：以中文[ɕ]和英文[ʃ]構音比較來佐證

張詠翔
國立臺北科技大學應用英文系

摘要

對中文為外語之學習者來說，中文齦顎擦音 [ɕ] 的發音是相當困難的。這個輔音常常被英語為母語的中文學習者用英文的 [ʃ] 取代。有的文獻建議教學時只需著重中文 [ɕ] 和英文 [ʃ] 的舌位高低不同，也有文獻強調只需強調唇形差異，抑或有文獻強調教學時舌頭高度和唇形差異都需注意。教學建議歧異可能是因爲學者假設中文 [ɕ] 和英文 [ʃ] 都各只有單一構音型態、而沒有考慮個體變異性。為控制個體變異性，本研究採取直接比較中英雙語人士產製之 [ɕ] 和 [ʃ] 的構音差異。構音資料包括唇顎接觸（使用靜態舌面圖和假顎圖）、舌頭位置（使用超音波影像）、唇形（使用唇部攝影）。研究結果指出，說話者之 [ɕ] 和 [ʃ] 有相同的唇顎接觸型態和舌頭位置、但是不同程度的唇凸度。因此，本文建議在教英語母語人士 [ɕ] 的發音時，可以把中文 [ɕ] 視之為較少唇凸度的英文 [ʃ]。

關鍵詞：中文　英文　第二語言教學法　超音波　構音