Sibilant contrast production by bilingual speakers of Quanzhou Southern Min and Mandarin

Caihong Weng[1], Ioana Chitoran[1], Alexander Martin[2]

[1]*Université Paris Cité*

[2]*University of Groningen*

caihong.weng@etu.u-paris.fr, ioana.chitoran@u-paris.fr, alexander.martin@rug.nl

Abstract

The merger of the Mandarin [s]~[ʂ] contrast, known as "deretroflexion", frequently occurs in Mandarin spoken by bilingual Southern Min speakers, whose L1 lacks the retroflex category. This study explores the production of the Mandarin alveolar-retroflex contrast by bilingual speakers of Quanzhou Southern Min (L1) and Mandarin (L2) in two different vowel contexts ([a] vs. [u]). Our bilingual speakers' contrast production was evaluated using a perceptual identification task by L1 Mandarin speakers, showing only a small subset of our sample who maintained the [s]~[ʂ] contrast. We found significant Center of Gravity (CoG) differences between the two target fricatives for "distinctive" speakers, with this difference being larger in the context of [a] than [u]. For all speakers, the acoustic difference between the target fricatives increased with increased exposure to and use of Mandarin.

Keywords: sibilant fricatives, contrast merger, Mandarin, Quanzhou Southern Min

1. Introduction

A merger of the Mandarin sibilant fricative contrast [s]~[ʂ] has been observed in Mandarin spoken by bilingual L1 Southern Min speakers, a phenomenon commonly characterized as "deretroflexion". This process, detailed in Kubler (1985), underscores how language contact with L1 Southern Min, which lacks the retroflex phone, led to a notable convergence of the retroflex sibilants towards an alveolar pronunciation in Mandarin. This reflects a broader pattern of L2 phonological adaptation in response to the phonological inventories of the languages in contact.

Other linguistic factors, such as vowel context, have also been noted to influence this contrast merger, but some conflicting results have emerged. On the one hand, Chang and Shih (2015) demonstrated a notable influence of vowel context on the spectral differentiation between alveolar and retroflex fricatives in both Beijing Mandarin and bilingual speakers of Mandarin and Taiwan Southern Min. In comparison to the [a] vowel context, it was observed that, in the [u] context, speakers from both regions exhibited a reduced spectral contrast. On the other hand, Chiu et al. (2020) applied ultrasound imaging techniques to the variablility of sibilant contrast production, and found that the tongue postures for [s] and [ʂ] showed more "context-dependent overlap" in the context of [a].

The exploration of variability in the merger of retroflex and alveolar sibilants extends, however, beyond purely linguistic dimensions. Recent research suggests that production variability in the merger of this sibilant contrast can additionally be captured by considering social factors, such as age, gender, and language exposure level (Chang and Shih 2015; Chuang and Fon 2010; Lee-Kim and Chou 2022).

The present paper explores variation in the production of the Mandarin [s]~[ʂ] contrast among a sample of bilingual speakers of Quanzhou Southern Min (泉州闽南话, henceforth QSM) [L1] and Mandarin [L2] and thus examines different linguistic and social factors at both the group and individual levels.

2. Method

61 bilingual speakers of QSM and Mandarin (29 men, 32 women) were recruited in Quanzhou, China, divided into three age ranges between 18 and 55 (18–30: 27 participants, 31–40: 18 participants, 41–55: 16 participants). These participants all have self-reported native-level fluency in Quanzhou Southern Min and Mandarin. They had all spent their childhood in Quanzhou and were living there at the time of the study. Mandarin was used as a metalanguage in experimental materials (including on-screen instructions), but all communication with the experimenter before, during, and after experimental sessions was conducted in Quanzhou Southern Min.

Each participant took part in a sentence reading task with target words embedded into carrier sentences, e.g., "请阅读单词X八遍", "Please read the word X eight times". Targets were all real Mandarin words of the form CVCV (2 fricatives × 2 vowel contexts ([a] vs. [u]) × 3 examples) and realized with a high level tone (tone 1) on the first syllable ([i] was not included as neither [si] nor [ʂi] are phonotactically well-formed in Mandarin). Target words were all represented orthographically as two Simplified Chinese characters. The lexical frequency of each real word was controlled to be within the log frequency range of 3 to 5 according to the SUBTLEX-CH corpus (Cai and Brysbaert 2010). Recordings were made in a quiet room at a sampling rate of 44.1 kHz using a Neumann TLM102 microphone, and a USBPre 2 audio mixer by Sound Devices. To guarantee the quality of the recording, we placed Alctron's VB 860 noise-canceling filter around the recording setup and installed soundproofing foam on both the window and the door of the room. We also ensured that the noise levels were maintained below -48 dB with the help of a Benetech GM1356 Digital Sound Decibel Noise Level Meter Tester. Before conducting the experiment, ethical approval was obtained at the Université Paris Cité (IRB Number: 00012022–95).

3. Results

Data from one participant were excluded because he had difficulty reading the Chinese characters during the task. Data ana-

lyzed are from the remaining 60 participants. Spectral moments were extracted at the mid-point of each fricative using the Praat script of DiCanio (2013). We focus here on Center of Gravity (CoG). We first compared speakers' [s] and [ʂ] productions in the two vowel contexts, shown in fig. 1. We employed mixed-effects models to investigate the effects of Contrast and Vowel, as well as their interaction, on CoG values while accounting for individual variability with by-participant random intercepts. We compared this full model to reduced models using likelihood ratio tests and found that the full model was a significantly better fit to the data than models which excluded the factors Fricative ($\beta = -343.2$, $SE = 48.02$, $\chi^2(1) = 49.40$, $p < 0.001$) and Vowel ($\beta = -548.28$, $SE = 48.02$, $\chi^2(1) = 119.47$, $p < 0.001$). The full model did not significantly differ from a reduced model which excluded the interaction between the two factors ($\chi^2(1) = 3.05$, $p > 0.05$). While the statistical model underscores significant contrast at the group level, along with a general coarticulatory effect (lower CoG values in the context of [u]), the graphical representation shows considerable overlap among the data points. We therefore sought to measure how individual participants produced the target fricatives, in order to categorize individuals as producing a reliable contrast or merging the two fricatives.

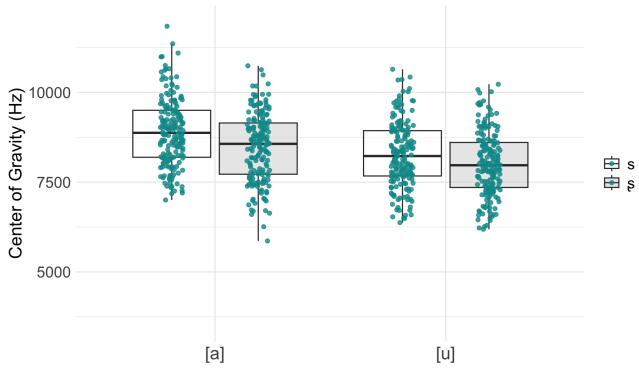

Figure 1: Comparison of CoG value for [s] and [ʂ] fricatives across vowel contexts within bilingual QSM Speakers

To investigate individual-level variability, we conducted a perceptual coding study involving ten native Mandarin speakers (5 men and 5 women, with a mean age of 28 years). These participants were recruited to perform a two-alternative forced-choice identification task. The aim was to assess the judgments of native Mandarin speakers regarding the productions of our bilingual QSM speakers. Stimuli consisted of CV syllables extracted from the sentence reading task performed by the 60 QSM speakers (12 tokens × 60 speakers). For each trial, the L1 Mandarin listeners heard a token of one of the QSM speakers' productions and had to indicate if they thought it corresponded to [s] or [ʂ]. Participants saw two Simplified Chinese characters and corresponding pinyin which indicated the response options, for example "sū苏" or "shū书".

We computed the L1 Mandarin speakers' identification overall accuracy for each QSM speaker in both [a] and [u] contexts. Figure 2 and fig. 3 summarize how the native listeners identified the sibilants produced by individual QSM speakers. The x-axis represents identification accuracy of individuals' [ʂ]-targets and the y-axis represents identification accuracy of individuals' [s]-targets. Following Chang and Shih (2015), we consider reliable productions to be above the threshold of 60% identification accuracy. Individuals had to produce both [s]- and [ʂ]-targets above this accuracy threshold in order to be considered to make a reliable sibilant contrast; they are shown inside the box in the top right corner of the figures. Consequently, QSM participants such as speaker 12, who demonstrated accuracy rates above 60% for both [s] and [ʂ] are classified as "distinctive" speakers. In contrast, many participants fall into the "merged" category due to their significantly lower accuracy rate (below 60%) for both fricatives. Among those classified as "merged", variability in contrast accuracy persists. For example, speaker 7 was classified as "merged" due to the significantly lower identification accuracy rate of their [ʂ]-targets (close to zero), despite an [s]-target accuracy nearly reaching 100%. Such speakers are producing fricatives that are perceived by L1 Mandarin speakers as [s] across the board (yielding high accuracy for [s]-targets and near-zero accuracy for [ʂ]-targets). These speakers are clustered in the top-left of the figures. On the other hand, speaker 22's [ʂ]-target production (in the bottom-right of the figures) achieves close to 100% accuracy but this speaker's [s]-targets were identified with near-zero accuracy. This speaker is producing fricatives that are perceived by L1 Mandarin speakers as [ʂ] across the board, a likely case of hypercorrection. Other speakers fall somewhere between these two extremes, producing some fricatives that are accurately perceived by L1 Mandarin speakers, but not above the 60% threshold.

We identified 9 QSM speakers who produced a reliably "distinctive" contrast in both vowel contexts, in contrast to 48 speakers who were categorized as "merged" in both vowel contexts. There were three additional participants who demonstrated the ability to distinguish the target contrast in one vowel context but not the other ([a]: speakers 15 and 30; [u]: speaker 5). For the sake of brevity, we focus in the rest of the paper on the 57 distinctive and merged participants.

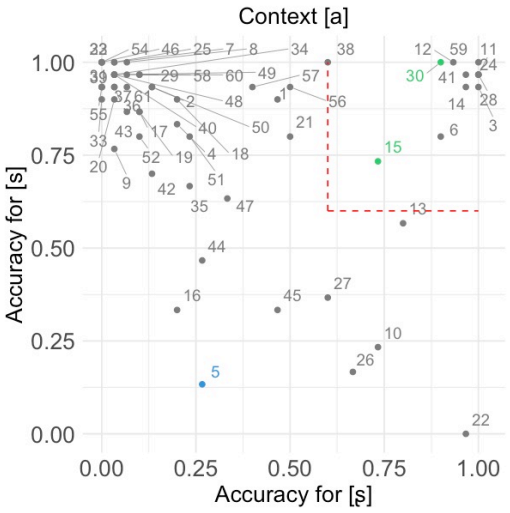

Figure 2: L1 Mandarin speakers' identification accuracy of QSM bilinguals' [s]- and [ʂ]-target productions in the context of [a].

3.1. Linguistic effects

CoG values for "distinctive" and "merged" speakers are shown in fig. 4. For both "distinctive" and "merged" speakers, we

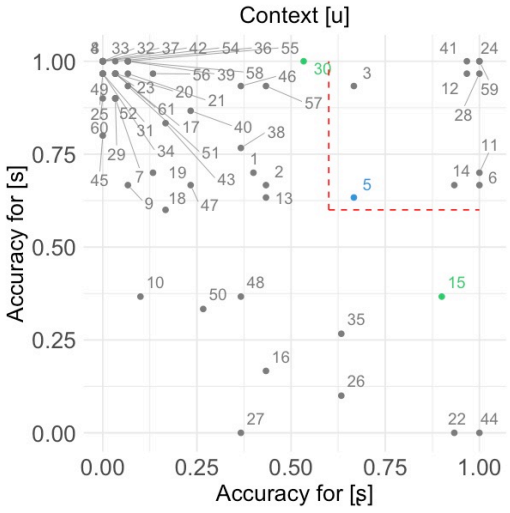

Figure 3: L1 Mandarin speakers' identification accuracy of QSM bilinguals' [s]- and [ʂ]-target productions in the context of [u].

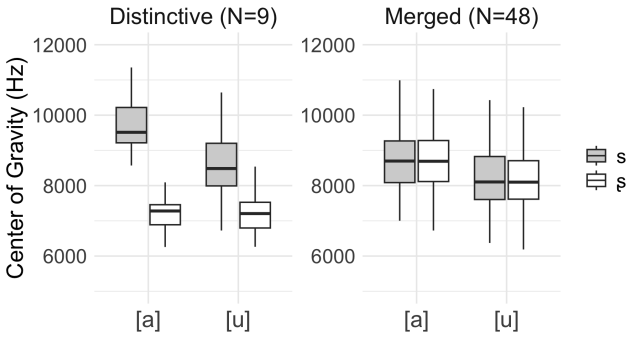

Figure 4: Comparison of CoG value for [s] and [ʂ] fricatives across vowel contexts between "distinctive" and "merged" bilingual QSM speakers

employed mixed-effects models to investigate the effects of Fricative and Vowel (both included using deviation coding), as well as their interaction (Fricative × Vowel), as fixed factors on CoG values, while accounting for individual variability of QSM speakers with by-participant random intercepts. We compared this full model to simpler models excluding one of the fixed effects or their interaction using likelihood ratio tests. For "distinctive" speakers, the full model was a significantly better fit to the data than models which excluded the factors Fricative ($\beta = -1978.2$, $SE = 127.9$, $\chi^2(1) = 123.8$, $p < 0.001$), Vowel ($\beta = -531.4$, $SE = 127.9$, $\chi^2(1) = 16.3$, $p < 0.001$), and their interaction ($\beta = 1082.4$, $SE = 255.8$, $\chi^2(1) = 16.9$, $p < 0.001$). This finding confirms that for these speakers who were perceived as producing different [s]- and [ʂ]-targets, their CoG values significantly differed according to the target fricative. Additionally, alongside the previously reported general coarticulatory effect (lower CoG values before [u]), the "distinctive" speakers exhibited a greater difference in the CoG values between alveolar and retroflex fricatives in the context of [a] than in the context of [u]. This suggests that "distinctive" speakers are able to maintain a greater spectral contrast between [s] and [ʂ] when followed by [a]. This observation aligns with the research presented by Chang and Shih (2015) (cf. Chiu et al. 2020), which noted that speakers displayed a larger spectral contrast distance in the [a] context compared to the [u] context, the rounded vowel tending to reduce the CoG in the realization of alveolar and retroflex fricatives.

For "merged" speakers, the analysis revealed that only the factor Vowel significantly affected model fit ($\beta = -544.8$, $SE = 40.2$, $\chi^2(1) = 158.3$, $p < 0.001$). This significant effect underscores again that the CoG values for [s] and [ʂ] are affected by the vocalic context, with both showing higher CoG values in the context of [a] compared to [u]. However, the factor Fricative does not exert a significant effect on the model fit for "merged" speakers ($\chi^2(1) < 1$). Similarly, the interaction between Vowel and Fricative does not contribute significantly to the model fit ($\chi^2(1) < 1$). These results suggest that, for "merged" speakers, the contrast between [s] and [ʂ] is not reliably maintained in production and that the fricatives that are being produced are all similarly affected by vowel context.

3.2. Extra-linguistic effects

As mentioned in the introduction, extra-linguistic (social) factors might also influence the variation we observed. We examined whether exposure to and use of Mandarin, age group, and gender influenced productions of the target fricatives. Given that our QSM participants all self-identified as highly bilingual, we focus on their L2 usage frequency. For assessing the extent of Mandarin exposure and use, we based on their responses to our post-test language use questionnaire. We followed Weng, Chitoran, and Martin (2023), which involved assigning an overall Mandarin exposure and use score to each participant. This score, which ranged from -8 to 8, was based on self-reported frequency of use of Mandarin and QSM on a five-point scale from "always QSM, never Mandarin" (-2), to "half QSM/half Mandarin" (0) to "always Mandarin, never QSM" (2) across four contextual domains: language used in childhood, within family settings, with friends, and among colleagues. A higher score indicates greater and more consistent exposure to and use of Mandarin relative to QSM. We observed variation in participants' responses ($M = -0.46$, $SD = 2.45$; recall that a score of 0 represents balanced Mandarin/QSM usage).

Because CoG values were found to significantly differ according to vowel context for both "distinctive" and "merged" speakers, we looked at data from each vowel context separately (see fig. 5). For each vowel context, we created a linear regression model to predict the average CoG differences of the participants in that context (each participant's average [s]-target CoG − their average [ʂ]-target CoG). The predictors included individual Mandarin exposure scores, gender, age group, and speaker classification (distinctive vs. merged), as well as the interaction between each speaker's classification and Mandarin exposure level score. Our analysis revealed that, for both vowel contexts, speakers with a higher Mandarin exposure score tended to produce a larger contrast difference between the target fricatives ([a] context: $\beta = 176.0$, $SE = 68.2$, $t = 2.5$, $p < 0.05$; [u] context: $\beta = 216.0$, $SE = 51.5$, $p < 0.001$). Moreover, both models indicated a significant negative effect of being classified as a merged speaker ([a] context: $\beta = -2435.5$, $SE = 193.6$, $t = -12.5$, $p < 0.001$; [u] context: $\beta = -1301.9$, $SE = 146.4$, $t = -8.8$, $p < 0.001$), again reflecting that distinctive speakers maintained a larger

CoG difference between the target fricatives, showing a clear acoustic contrast. It appears that despite observing an increase in contrast CoG difference between distinctive and merged speakers as Mandarin exposure score rise, the interaction between a speaker's classification and their Mandarin exposure score does not show a significant effect on CoG differences for either [a] or [u] context ([a] context: $\beta = -157.7$, $SE = 82.3$, $t = -1.9$, $p = 0.06$; [u] context: $\beta = -109.3$, $SE = 62.3$, $t = -1.7$, $p = 0.08$). The limited sample size of nine data points for the distinctive group in each vowel context may be a contributing factor to this outcome. Such a small dataset can limit the statistical power of the study, potentially obscuring real effects that might emerge with a larger number of observations. Consequently, while increased Mandarin exposure seems to be associated with the production of larger CoG differences, the current evidence does not conclusively support a differential impact based on speaker classification.

Concerning the other social factors (age group, gender), for the [a] context, speakers in the middle age range (41–55) showed a significant difference with the youngest group ($\beta = -434.46$, $p < 0.01$; all others $p > 0.05$). In the context of [u], significant effects were observed for gender and age, with women [compared to men] ($\beta = -283.98$, $p < 0.05$) and younger speakers [compared to the middle and older groups] producing more distinct contrasts (young vs. middle: $\beta = -471.88$, $p < 0.001$; young vs. older: $\beta = -423.93$, $p < 0.01$).

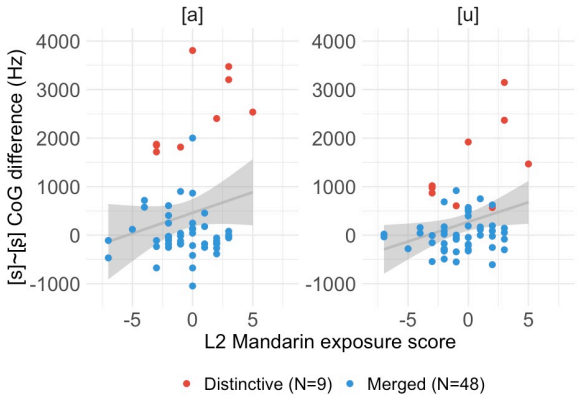

Figure 5: Participants' mean CoG difference as predicted by L2 Mandarin exposure level in each vowel context. More positive scores represent higher exposure to and use of Mandarin compared to QSM; more negative scores represent higher exposure to and use of QSM compared to Mandarin.

4. Discussion

In this study, we tested the production of a Mandarin sibilant fricative contrast by bilingual speakers of Quanzhou Southern Min (L1) and Mandarin (L2) in two different vowel contexts. Our results indicate that both the following vowel and a speaker's Mandarin exposure level are significant predictors of how this contrast is produced. Through the perception judgments of L1 Mandarin speakers, we categorized our bilingual speakers into two distinct groups: "distinctive" and "merged". Both groups showed a coarticulatory effect such that CoG values of each fricative were lower before the vowel [u]. Meanwhile, the exposure to and use of Mandarin appeared to relate to how strong of a contrast a speaker was likely to produce (more exposure to Mandarin was correlated with a larger CoG difference between the target fricatives). However, a significant interaction effect was not observed for either group. We speculate that this may be due to the disparity in sample sizes, with only 9 "distinctive" speakers compared to 48 who were categorized as "merged". This imbalance could potentially alter the interpretation of interaction effects. Further research with more "distinctive" speakers is needed to make these findings clearer and see if the trend we noticed (with a stronger effect for distinctive as compared to "merged" speakers) holds true.

Our analysis also identified patterns of hypercorrection and hypocorrection among the "merged" speakers' productions, suggesting a variety of profiles. This raises the question: what makes a speaker likely to distinguish or merge the contrast in the first place? Future work might benefit from including a measure of acuity alongside the factors explored here. Additionally, it is yet to be determined if speakers who merge contrasts in production also do so in their perception, highlighting a potential area for future research to explore the relationship between production and perception in bilingual populations.

5. Acknowledgements

Funding for this study was provided through a doctoral contract with Université Paris Cité, and through partial funding from the ANR-10-LABX-0083-LabEx EFL. The authors declare no conflicts of interest.

6. References

Cai, Qing and Marc Brysbaert (2010). "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles". In: *PloS one* 5.6, e10729.

Chang, Yung-Hsiang Shawn and Chilin Shih (2015). "Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin". In: *Journal of Phonetics* 50, pp. 52–66.

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". In: *Language and speech* 63.4, pp. 877–897.

Chuang, Yu-Ying and Janice Fon (2010). "The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech". In: *Speech Prosody 2010-Fifth International Conference*.

DiCanio, Christian (2013). *Spectral moments of fricative spectra script in Praat*. Haskins Laboratories & SUNY Buffalo.

Kubler, Cornelius C (1985). "The influence of Southern Min on the Mandarin of Taiwan". In: *Anthropological Linguistics* 27.2, pp. 156–176.

Lee-Kim, Sang-Im and Yun-Chieh Chou (2022). "Unmerging the sibilant merger among speakers of Taiwan Mandarin". In: *Laboratory Phonology* 13.1, pp. 1–36.

Weng, Caihong, Ioana Chitoran, and Alexander Martin (2023). "Bilingual phonological contrast perception: The influence of Quanzhou Southern Min on Mandarin non-sibilant fricative discrimination". In: *JASA Express Letters* 3.7.