JASA ARTICLE

Variability in fricative contrast merger: Evidence from Quanzhou Southern Min and Mandarin

Caihong Weng,[1,2,a] Ioana Chitoran,[1,3,b] and Alexander Martin[2,c]
[1]*Laboratoire de linguistique formelle, CNRS, Université Paris Cité, 75013 Paris, France*
[2]*Center for Language and Cognition, University of Groningen, 9712 EK Groningen, The Netherlands*
[3]*Institut Universitaire de France, 1 Rue Descartes, 75231 Paris Cedex 05, France*

ABSTRACT:
This study investigates the merger of the Mandarin alveolar-retroflex contrast in two vowel contexts ([a] vs [u]) among bilingual speakers of Quanzhou Southern Min [first language (L1)] and Mandarin [second language (L2)], with a focus on the interaction between linguistic and social factors. Sixty-one bilingual speakers' productions were evaluated by L1 Mandarin listeners in a perceptual identification task, and *k*-means clustering analysis based on each fricative's perceptual accuracy was used to classify speakers as "distinctive" or "merged." Acoustic analysis of the spectral center of gravity (CoG) confirmed that L1 Standard Mandarin speakers maintained a larger CoG separation between /s/ and /ʂ/ than bilinguals, especially in the [u] context, reflecting individual variability within the bilingual group. For bilinguals, acoustic analysis further corroborated the perceptual classification: "Distinctive" bilinguals maintained clear CoG separation between /s/ and /ʂ/ categories, whereas merged bilinguals showed little to no difference in either vowel context. Furthermore, gender and age significantly influenced this contrast production. Notably, these effects were largely mediated by L2 exposure levels. Women and younger speakers tended to have greater exposure to Mandarin, which correlated with more distinctive productions. This result highlights the dynamic relationship between linguistic and social factors in shaping phonological contrast variability in bilingual speakers.
 https://doi.org/10.1121/10.0042467
(Received 22 April 2025; revised 16 January 2026; accepted 30 January 2026; published online 23 February 2026)
[Editor: Felicitas Kleber] Pages: 1659–1673

I. INTRODUCTION

A. Phonological contrast variation in bilingual contexts

In speech production, phonological contrast mergers refer to a type of change where two distinct phonemic categories in a language collapse, resulting in the loss of their phonetic differentiation, with only a single category remaining [see, e.g., Hoenigswald (1960) and Gordon (2013)]. Such mergers are especially prevalent in bilingual contexts, where speakers navigate the interaction between two different language systems, often facing challenges in maintaining distinct phonemic contrasts across their languages (Flege and Eefting, 1987). A key question in bilingual speech variation is, therefore, under what conditions are phonemic contrasts maintained or merged?

An increasing body of work shows that long-term bilingualism and sustained language contact can lead to phonological instability, manifesting in phonetic convergence, reduction of phonemic contrasts, and, in some cases, even phoneme mergers [e.g., Yao and Chang, (2016), Chang and Shih (2015), Muxika-Loitzate (2017), Weng and Lee-Kim (2023)]. Such contact-induced restructuring has been documented across a wide range of bilingual communities, including Catalan–Spanish (Simonet, 2014), Basque–Spanish (Amengual, 2024), Welsh–English (Mayr *et al.*, 2017), Shanghainese–Mandarin (Yao and Chang, 2016) and Southern Min–Taiwan Mandarin (Lee-Kim and Chou, 2022; Weng and Lee-Kim, 2023), where long-term coexistence of two languages results in hybrid realizations rather than simple transfer from one language to another. These findings suggest that bilingual speech often develops its own contact-conditioned phonetic patterns and that mergers may arise when competing phonological systems exert conflicting pressures on category maintenance.

Theoretical frameworks such as Flege's (1995) speech learning model (SLM) and its update SLM-r (Flege and Bohn, 2021), as well as Best's (1995) perceptual assimilation model (PAM) and its extension PAM-L2 (Best and Tyler, 2007), provide valuable insights into the role of acoustic similarity between phonemic categories across languages in shaping bilingual speech. Both models emphasize that cross-linguistic interaction is heavily mediated by the degree of acoustic overlap between first language (L1) and second language (L2) categories in a bilingual's two languages. In production, SLM-r predicts that when L2 categories are phonetically highly similar, they are more likely to be produced similarly, leading to a phonemic merger. For example, Flege (1995) documented the /i/ and /ɪ/ merger among Spanish–English bilinguals, where the reduced acoustic distinction led to overlapping phonemic categories. A similar pattern was reported by Simonet (2014) for the /o/

[a]Email: c.weng@rug.nl
[b]Email: ioana.chitoran@u-paris.fr
[c]Email: alexander.martin@rug.nl

and /o/ merger among Catalan–Spanish bilinguals. Conversely, when phonemic categories are sufficiently distinct, bilinguals may preserve or even exaggerate contrasts [see Hazan and Boulakia (1993) for differences in voice onset time (VOT) patterns in voicing contrasts among French–English bilinguals compared to monolinguals].

As perception-based models, PAM and PAM-L2 offer a useful theoretical background for understanding the conditions under which phonological mergers are likely to arise in bilingual contexts. Specifically, when two L2 sounds are assimilated to a single L1 category (a "single category assimilation"), they are unlikely to be distinguished in the mental phonological space. PAM also highlights that phonological context can increase or decrease the similarity between target categories due to coarticulation. This aligns with findings in production research showing that the acoustic distance between phonemes can vary, depending on segmental context, due to co-articulatory effects that may either amplify or diminish phonemic distinctions (Mann and Repp, 1980; Ohala, 1993). Early evidence for such context effects on fricatives comes from Fujisaki and Kunisaki (1978), who showed that Japanese listeners' identification of voiceless fricatives between /s/ and /s/ was systematically modulated by the surrounding vowel. Crucially, more recent work has emphasized that co-articulatory patterns themselves may be subject to restructuring over time, and that changes in production can be accompanied by corresponding changes in perceptual compensation during sound change [e.g., Beddor and Krakow (1999), Harrington *et al.* (2008), and Kleber *et al.*, 2012)]. This literature suggests that co-articulatory patterns may form part of a phonological system and that differences in such patterns can be relevant for perceptual categorization, providing motivation for the present study.

While linguistic and contextual factors shape bilingual phonology, individual extra-linguistic factors also play a crucial role in the likelihood of phonological mergers (Amengual, 2024; Casillas, 2024; Flege and Bohn, 2021). As we mentioned previously, according to SLM-r (Flege and Bohn, 2021), L2 exposure is a key determinant of category distinction: When L2 exposure level is limited, bilinguals may fail to establish robust categories in the L2, increasing the likelihood of a merger between similar phonemes. In contrast, speakers with extensive and consistent use of the L2 are more likely to maintain or enhance phonetic distinctions due to strengthened phonological representations. Empirical studies support these predictions. For instance, Mayr *et al.* (2019) showed that Galician–Spanish bilinguals with lower exposure to Galician were more likely to merge mid-vowel contrasts with their Spanish equivalents, highlighting the role of exposure in shaping bilinguals' phonological systems.

Independently of L2 exposure, social factors such as age and gender have also been linked to variation in phonemic contrast. In Labov's (2001) framework of sociolinguistic variation, age stratification is often interpreted as evidence of sound change in progress, with younger speakers more likely to show complete overlap [e.g., the "cot"–"caught" merger in American English (Labov (2001)]. Gender likewise shapes phonological variation, though such effects are not universal. In many Western sociolinguistic contexts, women have been observed to lead sound change or align more closely with prestige norms, while men may align more closely with local or non-prestige forms, a dynamic known as the "gender paradox" [see Labov (2001), p. 293]. Importantly, however, gendered linguistic behavior is highly sensitive to local social structure and cultural expectations, and does not necessarily generalize across societies. For example, Stanford (2008) documents a contact setting in Sui-speaking communities of southwest China in which immigrant women maintain clan-specific phonological features over long periods, illustrating that women do not universally lead convergence or change under contact.

Consistent with the role of sociocultural context, bilingual phonological research has also documented gender-related differences in contrast maintenance in some contact settings. For example, English–French bilingual women living and working in the United Kingdom have been found to produce longer VOTs and greater separation between voiced and voiceless stops than men (Pépiot and Arnold, 2021). Despite these insights, much of the existing research still focuses on group-level outcomes, sometimes overlooking the fine-grained variability found at the individual level. Although bilingualism and multilingualism have been widely studied, research has disproportionately focused on a limited set of well-studied language combinations, particularly involving English and major European languages, thus limiting our understanding of the generalizability of these findings across different linguistic situations (Kang and Guion, 2006). This imbalance has been noted more broadly in cognitive science, where over-reliance on English has been shown to constrain theoretical generalization and obscure cross-linguistic diversity (Blasi *et al.*, 2022). At the same time, a growing body of recent work has begun to address bilingual experience in a wider range of linguistic context, including contact settings involving East Asian languages [e.g., English–Korean community; Chang (2019, 2016)] as well as Sinitic language (e.g., Cantonese–Mandarin bilingualism; Mok *et al.* (2013)], highlighting the diversity of bilingual trajectories and outcomes.

Our study aims to explore the factors driving the merger of the /s/∼/ʂ/ contrast among Quanzhou Southern Min (QSM)–Mandarin speakers. Within this community, Standard Mandarin holds dominance in education and public life, while Southern Min remains primarily used in family settings and everyday informal interactions. Specifically, we seek to determine whether this merger represents a generalized phenomenon across the bilingual population or if significant individual variation exists in the extent and nature of the merger. Given that fricative contrasts are known to be socially conditioned across languages [e.g., Stuart-Smith (2020), Chang and Shih (2015), Lee-Kim and Chou (2022), and Lin (2018)], this situation offers an opportunity to examine the interaction of linguistic and social influences in shaping phonological outcomes. By linking

https://doi.org/10.1121/10.0042467

both linguistic and social factors, we aim to identify which speakers are more likely to exhibit this merger, who is more likely to maintain distinctions, and what conditions contribute to the observed variability.

B. Sibilant contrast merger in Quanzhou Mandarin

Quanzhou, a coastal city in south China's Fujian Province (see the geographical map in Fig. 1), provides a unique linguistic setting for investigating phonological contrast variation. Speakers in Quanzhou are bilingual, using QSM (泉州闽南话) and Mandarin in different aspects of daily life. Historically, Southern Min served as the primary language for local communication for over 1500 years and is deeply embedded in the region's culture and identity (Chappell, 2019). The promotion of Beijing Mandarin, officially known as Putonghua (普通话), began in 1956 when the Chinese government mandated its instruction in all schools within Han Chinese regions, including Quanzhou. As a result, local Quanzhou residents have been educated and immersed in Beijing Mandarin, gradually becoming Southern Min–Mandarin bilinguals (Chen *et al.*, 2015). However, in the earlier decades, the daily use of Mandarin remained limited, as Southern Min continued to dominate informal and community interactions. By the 1980s, Mandarin made significant progress in formal education, media, and most professional settings throughout mainland China. This shift increased exposure to Mandarin in these settings and, coupled with national language policies, significantly enhanced language contact. These changes have resulted in younger generations of Quanzhou speakers achieving greater proficiency in both languages, navigating a bilingual environment where Mandarin and QSM fulfill distinct social and functional roles.

FIG. 1. Geographical location of Quanzhou City in Fujian Province (CGTN, 2022).

QSM and Mandarin are two mutually unintelligible languages, each with distinct phonological systems, both segmental and tonal (Chappell, 2019; Duanmu, 2007). The extent of their phonological and lexical divergence reinforces the argument that QSM should not be treated merely as a dialect of Mandarin, but rather as a separate language (Tang and Van Heuven, 2009, 2015).

Standard Mandarin exhibits a distinct place contrast between alveolar /s/ and retroflex /ʂ/ sibilant fricatives, while QSM lacks a retroflex sibilant in its phonological inventory, relying solely on the alveolar sibilant /s/. Despite heavy exposure to Mandarin, this contrast is often merged in the Mandarin spoken by bilinguals of Southern Min. Bilingual speakers may map Mandarin retroflex /ʂ/ onto /s/ as the closest equivalent in QSM, leading to the observed merger. As detailed by Kubler (1985), this process illustrates how language contact with L1 Southern Min leads to a notable convergence of the retroflex sibilants towards an alveolar production in Mandarin. This merger is puzzling given the dominance of Standard Mandarin and the widespread use of Mandarin in formal domains.

However, the variation due to the influence of pure language contact with Southern Min is not widely accepted by later studies. Instead, the merger is more strongly influenced by vowel context and social factors (Chung, 2006; Chiu *et al.*, 2020; Chang and Shih, 2015; Chuang and Fon, 2010; Lee-Kim and Chou, 2022; Weng *et al.*, 2024; Lee-Kim and Tung, 2025).

Chang and Shih (2015) demonstrated a notable influence of vowel context on the center of gravity (CoG) between alveolar and retroflex fricatives in both Beijing Mandarin and bilingual speakers of Taiwan Mandarin–Southern Min. Both Beijing and Taiwan Mandarin speakers produced a smaller difference in CoG between target fricatives in the context of the vowel /u/, as coarticulation from the rounded vowel lowers the CoG in alveolar realizations, in turn decreasing the spectral distance between the two targets. Chiu *et al.* (2020) also confirmed in their study that this contrast merger is influenced by vowel context.

The exploration of variability in the merger of retroflex and alveolar sibilants extends, however, beyond purely linguistic dimensions. Some recent work has found that the merger is in fact not limited to one direction. While it is often assumed that, in a merger, bilingual speakers usually produce both fricatives similar to their L1 /s/ category, as described above, cases of hypercorrection have also been observed, where speakers produce both fricative categories using the retroflex variant, depending on a number of factors, including gender and age (Chung, 2006).

Indeed, a non-negligible amount of the production variability in the merger of this sibilant contrast appears to be captured by social factors (Lee-Kim and Tung, 2025; Chang and Shih, 2015; Chuang and Fon, 2010; Lee-Kim and Chou, 2022; Weng *et al.*, 2024). For example, Chuang and Fon (2010)

looked at the speech of Taiwan Mandarin–Southern Min bilinguals and found that whether sibilants were merged or not was highly dependent on gender, with men showing higher rates of sibilant merger than women. Lee-Kim and Chou (2022) also reported considerable individual variation in sibilant contrast mergers, with the mergers being more prevalent among male speakers. This merger was further constrained by age. Recent studies have shown that younger speakers born after the 1990s produce a consistent /s/-/ʂ/ distinction, actively adopting what is considered a more educated or cultivated Standard Mandarin. This distinction serves not only to align with the standard but also to differentiate their speech from the merged production of the older generation [see Beckman *et al.* (2014) in Songyuan Mandarin and Lin (2018) in Xiamen Mandarin]. The reasons why men and older speakers tend to merge this fricative contrast more frequently remain debatable. According to Chuang and Fon (2010), male speakers who exhibit this merger tend to use Southern Min more frequently. This aligns with the sociolinguistic perspective of Labov (2001) and Stuart-Smith (2020), which suggests that gender effects in language variation are linked to social roles, with women more likely to adopt standard linguistic forms than men due to socially determined normative gender roles [see also Gal (1978) on the interaction of language use and gender roles].

In the present study, we address this question through a case study of QSM–Mandarin bilinguals, focusing on variation in the production of the /s/∼/ʂ/ contrast. Using a combination of acoustic analyses and perceptual categorization by L1 Mandarin listeners, we examine whether bilingual speakers maintain a robust phonological contrast or produce merged realizations. Specifically, we investigate how contrast realization varies as a function of linguistic and extralinguistic factors within a setting of sustained bilingual language contact, situating our findings alongside recent work on production–perception relationships in ongoing mergers, which has focused on single language context, such as Norwegian (Evjen *et al.*, 2025).

C. Research questions

Grounded in L2 theoretical predictions from the SLM-r and the PAM-L2, we address the following research questions (RQ1–RQ3):

RQ1: Do bilingual QSM–Mandarin speakers show greater acoustic overlap between the Mandarin alveolar /s/ and retroflex /ʂ/ compared to L1 native Mandarin speakers?

We hypothesize that bilingual QSM–Mandarin speakers will demonstrate greater acoustic overlap between alveolar /s/ and retroflex /ʂ/ in Mandarin compared to L1 Mandarin speakers. This increased overlap is posited to reflect articulatory adjustments influenced by the phonological structure of QSM, which does not include retroflex sibilants. From the perspective of SLM-r, the absence of a corresponding retroflex category in the L1 inventory increases the likelihood that the Mandarin retroflex /ʂ/ will be perceived and produced as acoustically and articulatorily similar to the L1 alveolar /s/ category, resulting in a merged or reduced contrast in production.

RQ2: Is the degree of this fricative contrast merger modulated by vowel context?

We expect that vowel coarticulation, particularly in the context of the vowel [u], will affect the spectral distance between /s/ and /ʂ/, with differential effects between monolinguals and bilinguals due to their increased category overlap.

RQ3: How do social factors (e.g., L2 Mandarin exposure level, age, and gender) influence the degree of sibilant contrast among bilingual speakers?

We predict that younger speakers and women are likely to show greater phonological distinction, which may be attributed to their higher usage of and exposure to Mandarin. Therefore, rather than age or gender independently driving the distinction, we hypothesize that it is through Mandarin exposure that these variables play a key role (Chuang and Fon, 2010; Weng *et al.*, 2023). If exposure to and use of Mandarin is influential, we should expect speakers who differ in their frequency of Mandarin use to also differ in their sibilant contrast production.

These hypotheses are first tested in a production study, designed to examine bilingual speakers' speech patterns under certain controlled conditions, followed by acoustic analyses of their spectral properties and a perceptual judgment task carried out by L1 Standard Mandarin listeners.

The following section (Sec. II) outlines the methodological framework of the study. We will present first the perceptual component, since its categorization results serve as the basis for the *k*-means clustering analysis and guide the interpretation of the production data reported subsequently.

II. METHOD

This study consists of two main components designed to examine phonological contrast variation in bilingual speakers' production from both perceptual and acoustic perspectives. We start with the perceptual evaluation, in which a group of L1 Mandarin listeners categorized fricatives extracted from the production of QSM–Mandarin bilingual speakers. We then explain our production experiment, in which QSM–Mandarin bilingual speakers (the experimental group) and a separate group of L1 Mandarin speakers (the control group) completed a sentence reading task. Each component is described in detail in Secs. II A and II B below.

A. Perceptual evaluation by native Mandarin listeners

We conducted a perceptual coding study to assess how clearly this contrast was perceived in the productions of the QSM–Mandarin bilingual speakers. Participants were recruited in Paris, France, and tested at the Université Paris Cité. Ten L1 Mandarin listeners (five men, five women,

https://doi.org/10.1121/10.0042467

mean age 28 years) took part in this study. They all reported growing up in northern regions (Beijing, Tianjin, Hebei Province, and Xi'an) where local languages maintain a contrast between /s/∼/ʂ/. All of them were university students who had arrived in France less than 3 years prior to the study. Language background information was collected in a post-experiment questionnaire. Participants reported using Mandarin regularly in communication with family and friends, including via phone calls, messaging apps, and videos. Thus, even though we did not directly assess the French proficiency of this control group, according to their reports, Mandarin remains an important language in their daily lives.

These participants completed a two-alternative forced-choice identification task aimed at assessing the productions of the bilingual QSM speakers. We wanted to know whether the bilingual speakers were perceived as producing a reliable contrast between the fricatives or not.

Stimuli consisted of consonant-vowel (CV) syllables extracted from a sentence reading task performed by 60 bilingual QSM speakers (12 tokens × 60 speakers; see details in Sec. II B). For each trial, the L1 Mandarin listeners heard a token of one of the QSM speakers' productions (e.g., [sū]) and had to indicate if they thought it corresponded to /s/ or /ʂ/. Participants saw on a computer screen two simplified Chinese characters and corresponding Pinyin that indicated the response options (for example, "sū苏" or "shū书") and responded by clicking on one or the other. We used JSPSYCH (De Leeuw, 2015) to present the stimuli and collect response data.

To statistically evaluate L1 Mandarin listeners' perception of our bilingual QSM–Mandarin speakers' fricative productions, we fitted a logistical mixed-effects model predicting correct identification as a function of fricative (/s/ vs /ʂ/), vowel context ([a] vs [u]), and their interaction. The model included random intercepts for item, production speaker (QSM–Mandarin bilinguals), and listener (L1 Mandarin speakers). This analysis allowed us to examine the extent to which fricative type and vowel context influenced perceptual accuracy.

To further explore variation at the level of individual bilingual speakers, we conducted a *k*-means clustering analysis on the perceptual judgment data. The *k*-means algorithm groups speakers by minimizing within-cluster variance and maximizing between-cluster variance (Hartigan and Wong, 1979). Each speaker's production was represented by two accuracy scores: one for /s/ targets and one for /ʂ/ targets, reflecting how reliably their fricatives were identified by L1 Mandarin listeners. To determine the optimal number of clusters, we applied the elbow method separately for each vowel context. The results revealed a clear inflection point at $k = 3$ [see R MARKDOWN analysis in the Open Science Framework (OSF) repository at https://osf.io/b8ys4/?view_only=7c24ea57f1964005bc404fe9121a212b], indicating that three clusters provide the best balance between within-cluster homogeneity and model simplicity. Based on this result, we retained three distinct clusters for each vowel context in the subsequent analysis (see details in Table I).

TABLE I. Mean classification accuracy for each *k*-means cluster ($k = 3$) across vowel contexts.

		Perceived accuracy		
Cluster	Context	/s/	/ʂ/	Interpretation
1	[a]	0.91	0.09	Merged contrast
	[u]	0.90	0.11	
2	[a]	0.90	0.83	Distinctive contrast
	[u]	0.87	0.88	
3	[a]	0.25	0.52	Merged contrast
	[u]	0.20	0.56	

Based on the resulting clusters, we categorized speakers as producing either a distinctive or a merged /s/∼/ʂ/ contrast.

B. Sentence reading task

1. Stimuli

In the sentence reading task, all target words were embedded in a carrier sentence, 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 with the target fricative in the initial position (2 fricatives × 2 vowel contexts, [a] vs [u], × 3 examples) and realized with a high-level tone (tone 1) on the first syllable.[1] The target words were represented orthographically as two simplified Chinese characters (see Table II). For the purposes of lexical frequency balancing, the lexical frequency of each real word was controlled to be within the log frequency range of 3–5 according to the SUBTLEX-CH corpus (Cai and Brysbaert, 2010). The presentation of the stimuli and logging of participant responses were carried out with a script written in JSPSYCH (De Leeuw, 2015).

2. Participants

Two groups of speakers were recruited for the sentence reading task. First, as a control group, 15 L1 Mandarin speakers (7 men, 8 women) aged 18 to 30 years were recruited in Paris; none participated in the perceptual evaluation task, but their language background was comparable to those participants: All grew up in northern China, were university students in Paris, and had arrived in France within the 3 previous years. Mandarin remained an important language in daily life, and they consistently maintained a clear /s/∼/ʂ/ contrast.

Second, 61 bilingual speakers of QSM and Mandarin (29 men, 32 women) were recruited in Quanzhou, China, divided into two age ranges between 18 and 55 years (18–30

TABLE II. Mandarin /s/∼/ʂ/ target stimuli (real words), with each word presented as Chinese characters followed by their corresponding Pinyin in parentheses.

/s/	[a]	撒谎 (sāhuǎng)	撒娇 (sājiāo)	撒气 (sāqì)
	[u]	苏杭 (sūhuáng)	苏打 (sūdá)	苏醒 (sūxǐng)
/ʂ/	[a]	沙滩 (shātān)	沙子 (shāzi)	沙丘 (shāqiū)
	[u]	书本 (shūběn)	书包 (shūbāo)	书店 (shūdiàn)

years, 27 participants; 31–55 years, 34 participants). All participants self-reported native-level fluency in QSM and Mandarin; they had all spent their childhood in Quanzhou and were living there at the time of recording. Mandarin was used as a metalanguage in experimental materials (including on-screen instructions), but all verbal communication with the experimenter (the first author) was conducted exclusively in QSM before, during, and after the experimental sessions. This decision was motivated by the fact that the first author is a bilingual speaker of QSM and Mandarin and self-reports exhibiting a /s/∼/ʂ/ merger in her Mandarin speech. Using QSM, which does not have this contrast, as the language of communication ensured that participants were not auditorily exposed to the potentially merged Mandarin sibilant realizations during the experiment.

3. Procedure

The recordings were made in a sound-attenuated booth in Paris (for L1 Mandarin speakers) and in a quiet room in Quanzhou (for QSM bilinguals) at a sampling rate of 44.1 kHz using the same equipment: a Neumann TLM102 microphone (Neumann, Berlin, Germany) and a USBPre 2 audio mixer by Sound Devices (Reedsburg, WI). To ensure high-quality recordings in Quanzhou, we placed a noise-canceling filter (Alctron VB 860, Alctron Electronics Company, Ningbo, China) 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 using a Benetech GM1356 digital sound decibel noise level meter tester (Benetech, Shenzhen, China). Consequently, the recordings in Quanzhou achieved booth-quality standards (see Fig. 2 for the comparison of sound recording quality in the two environments).

At the end of the task, bilingual participants completed a post-task questionnaire about their language background, age, and how frequently they used Mandarin and QSM during their childhood and currently in interactions with family, friends, and colleagues. The sentence reading session took 15–20 min to complete. All participants read and signed consent forms prior to data collection.

4. Data processing

One bilingual participant of QSM–Mandarin who was orally fluent reported difficulty reading certain Chinese characters during the production task; we therefore excluded this participant's data from the analysis. All production target stimuli were segmented and annotated automatically using the Montreal Forced Aligner (McAuliffe *et al.*, 2017). The first author then verified the output manually and corrected the segmentation boundaries in PRAAT (Boersma and Weenink, 2023). The onset of the fricative was marked at the positive zero crossing nearest to the presence of high-frequency energy (i.e., energy exceeding 1000 Hz). The offset of the fricative was identified as the positive zero crossing closest to the onset of periodic energy from the subsequent vowel.

5. Acoustic and statistical analyses

One widely used approach to characterize the acoustic properties of sibilant fricatives is through the analysis of four spectral moments, which quantify the distribution of spectral energy [for a review, see Forrest *et al.* (1988)]. A wealth of studies have focused on distinguishing sibilant fricatives in various languages by analyzing spectral moments in both L1 and L2 speech. The findings generally indicate that distinctions in the place of articulation of fricatives can be captured through the analysis of these four key spectral moments [see Jongman *et al.* (2000), Koenig *et al.* (2013), Maniwa *et al.* (2009), Shadle (2023), Spinu and Lilley (2016), and Stevens and Harrington (2016) for English, Reidy (2016) for Japanese, Nirgianaki (2014) for Greek, Wikse Barrow *et al.* (2022) for Swedish, Kochetov (2017) for Russian, Beristain (2022) and Egurtzegi *et al.* (2024) for Basque, and Chang and Shih (2015) and Li *et al.* (2014) for Mandarin).

In terms of sibilants, the CoG has cross-linguistically been shown to clearly differentiate fricatives by place of articulation and has remained a prominent measure (Egurtzegi *et al.*, 2024; Chang and Shih, 2015; Chiu *et al.*, 2020; Jongman *et al.*, 2000; Koenig *et al.*, 2013), often without considering the additional spectral moments. However, as recent studies have shown, spectral moments are not without

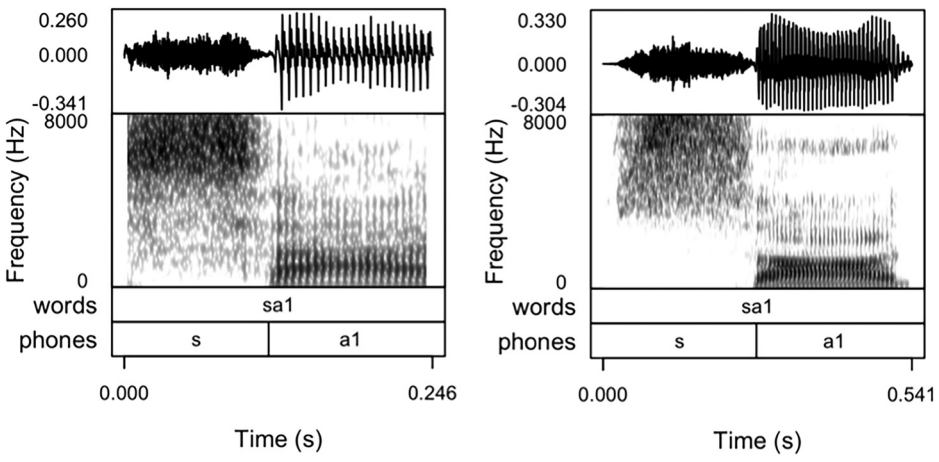

FIG. 2. Example of [sa] recording in a sound-attenuated booth with a male L1 Mandarin speaker (left) vs in a quiet room with a male bilingual QSM speaker (right).

https://doi.org/10.1121/10.0042467

limitations and may not fully capture the spectral complexity of fricatives. Shadle (2023) points out that many studies computing spectral moments rely on single discrete Fourier transform (DFT), which can produce an inefficient and inconsistent spectral estimate: The error is proportional to the spectral magnitude and tends to be greatest at the peaks. These methodological limitations are often overlooked. One proposed improvement by Shadle (2012, 2023), which we adopt in our study, is to compute spectral measures from time-averaged spectra, obtained by averaging magnitude-squared spectra across multiple short windows during the central portion of the fricative. This approach yields more stable and representative estimates of the energy distribution across the fricative segment, ensuring that the spectral moments measurements better reflect underlying articulatory and aerodynamic properties.

In the present study, the acoustic analysis of the fricative productions was conducted using a PRAAT script adapted from DiCanio (2021) that extracts the intensity and the first four spectral moments from fricative spectra. We focused on the center 80% of the fricative, excluding the initial and final 10% to avoid potential boundary effects. The DFTs were averaged using time-averaging (Shadle, 2012, 2023). The relationship between the window size and the number of DFTs was guided by the principle that the duration of these DFTs (window size) multiplied by the number of DFTs (window number) should equal a value no greater than 1.6 times the duration of the sound, as per DiCanio (2021). We therefore selected six windows of 15 ms each at equally spaced intervals across the center 80% of the fricative duration. This window size was chosen to ensure that the total windowed duration of 90 ms remained within the limits set by the fricative's overall duration. This configuration was effective for analyzing fricatives with durations as short as 56 ms, while maintaining a maximum of 50% overlap between windows, preventing over-weighting of any specific part of the fricative.

L1 Mandarin speakers produced fricatives with an average duration of 153 ms (±38 ms). As for bilingual QSM speakers, the average fricative length was 164 ms (±44 ms). Accordingly, all fricative productions were included in the analysis based on their duration.

We then utilized the Random Forest classification method (Breiman, 2001) to identify the most important acoustic parameters that differentiate the alveolar /s/ and retroflex /ʂ/ fricatives in our L1 Mandarin speakers. The model included four spectral moments and intensity and showed high classification performance (96.3% out-of-bag accuracy). CoG emerged as the strongest predictor, showing the highest values for both *Mean Decrease in Accuracy* and *Mean Decrease Gini*, confirming its key role in the fricative contrast. We thus focus specifically on CoG in our analyses.

To assess how fricative production patterns (/s/ vs /ʂ/) vary as a function of vowel ([a] vs [u]) and speaker group (L1 Mandarin control group vs bilingual QSM–Mandarin target group) as well as their interaction, we analyzed CoG values using a mixed-effects model implemented in R using the LME4 package (Bates *et al.*, 2015).

The full model included all predictors using deviation coding, and random intercepts were specified for participants and items to account for individual and item-level variability. We compared this full model to reduced models using likelihood ratio tests. The full model was specified as follows:

$$\texttt{cog} \sim \texttt{fricative} \times \texttt{vowel} \times \texttt{group} + (1 \mid \texttt{participant}) + (1 \mid \texttt{item})$$

To assess the contribution of each predictor, we conducted a series of nested model comparisons using likelihood ratio tests, systematically removing one factor at a time (target, vowel, or group, as well as their interactions) from the full model.

III. RESULTS

A. Native speakers' perceptual evaluation of QSM fricative productions

We begin by presenting the results of the perceptual evaluation, in which L1 Mandarin listeners identified the fricatives produced by QSM–Mandarin bilingual speakers. This provides a functional measure of /s/∼/ʂ/ contrast preservation: If the contrast is perceptually ambiguous for a given speaker, it suggests a potential merger.

Overall, the L1 Mandarin listeners' identification accuracy of /s/ targets for all QSM speakers is higher than for /ʂ/ targets in both [a] and [u] contexts (see Fig. 3).

Specifically, in the context of [a], the mean identified accuracy for /s/ targets was 81.5% (±24.9%), compared to 34.4% (±36.5%) for /ʂ/ targets. Similarly, in the context of [u], the mean identified accuracy was 77.4% (±29.4%) for /s/, compared to 32.6% (±36.4%) for /ʂ/, showing that most of QSM speakers' /s/ productions were identified as Mandarin /s/ in both [a] and [u] contexts. Many QSM speakers' /ʂ/- target productions were also identified as Mandarin /s/ in both vowel contexts ([ʂa]: 65.6% vs [ʂu] 67.4%).

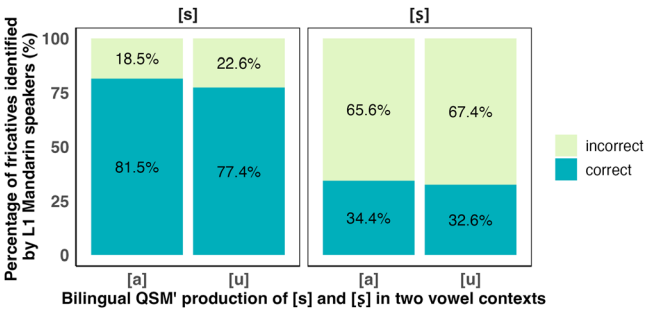

FIG. 3. Percentages of /s/- and /ʂ/- target productions across two vowel contexts by QSM–Mandarin bilinguals that were identified by L1 Mandarin listeners as either "correct" (matching the target) or "incorrect" (mismatching the target).

The results of the logistic mixed-effects model (see Sec. II A for the model details) confirmed these trends. A significant main effect of fricative [$\beta = -2.87$, standard error (SE) = 0.34, $z = -8.53$, $p < 0.001$] indicated that /ʂ/ targets were substantially less likely to be identified correctly than /s/ targets. In terms of vowel context, the vowel [u] had a significant effect on identification accuracy for both fricative targets ($\beta = 0.89$, SE = 0.41, $z = 2.18$, $p < 0.05$). The interaction between fricative and vowel was also significant ($\beta = -2.37$, SE = 0.57, $z = -4.13$, $p < 0.001$), suggesting that /ʂ/ targets produced in the [u] context were especially difficult for listeners to categorize correctly. These findings align with our descriptive results and support the interpretation that QSM–Mandarin bilinguals tend to produce /ʂ/ in a way that is perceptually confusable with /s/, particularly in the [u] context, where identification accuracy dropped significantly. This pattern is also consistent with a merger of the contrast between the two fricatives, modulated by vowel context (Chang and Shih, 2015; Chiu *et al.*, 2020).[2]

Figure 4 summarizes how the L1 Mandarin listeners identified the sibilant contrast 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. Based on the elbow decision described in Sec. II A, three clusters were retained for each context. Participants from cluster 1 (blue numbers, dashed outline) have high accuracy for /s/ but low accuracy for /ʂ/. Participants from cluster 2 (pink numbers, solid outline) show high accuracy for both /s/ and /ʂ/. Participants from cluster 3 (yellow numbers, dotted outline) show reduced accuracy for /s/ and/or /ʂ/. Context-specific mean perceived accuracies for each cluster are provided in Table I.

We classified speakers as either "merged" or "distinctive" based on the outcomes of the clustering analysis. Speakers in cluster 2 (pink) are classified as "distinctive," since they produced both /s/- and /ʂ/- targets with high accuracy. We considered that they produced a perceptibly reliable contrast between the two sibilants. Thus, bilingual QSM participants such as speaker 12, who demonstrated high perceived accuracy rates for both /s/ and /ʂ/, are classified as "distinctive" speakers. In contrast, many participants fall into the "merged" clusters [clusters 1 (blue) and 3 (yellow)] due to their significantly lower perceived accuracy rate in at least one of the fricative targets (shown inside the boxes in the top left corner and the bottom of the figures).

Consequently, in the [a] context, we identified 16 QSM speakers who produced a reliably "distinctive" contrast, as opposed to 44 speakers who were categorized as "merged." In the [u] context, we identified 11 QSM speakers who contrasted the fricatives, against 49 who merged them. Among those classified as "merged," variability in contrast accuracy persists. For example, speaker 33 was classified as "merged" in both contexts 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 clear case of hypercorrection.

B. Production results

1. L1 Standard Mandarin vs bilinguals' production

The perceptual results showed that L1 Mandarin listeners were more accurate in identifying /s/ targets than /ʂ/ targets, and that /ʂ/ targets produced in the [u] context were especially difficult to categorize correctly. To examine how these perceptual patterns relate to speakers' productions, we next analyzed the spectral CoG of fricative tokens. We first compared the productions of bilingual speakers to those of an L1 Standard Mandarin control group, using a linear mixed-effects model with fixed effects of fricative, vowel context, speaker group (L1 Standard Mandarin vs bilingual), and their interactions. Random intercepts were specified for participants and items to account for individual and item-level variability. This analysis establishes whether bilingual speakers overall exhibit production patterns comparable to those of L1 Mandarin speakers, setting aside for the moment

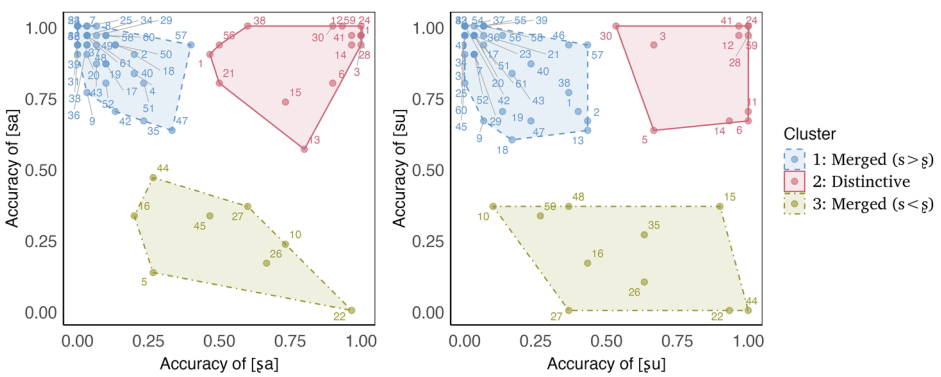

FIG. 4. L1 Mandarin listeners' identification accuracy of QSM bilinguals' /s/- and /ʂ/- target productions in the context of [a] (left) and [u] (right).

https://doi.org/10.1121/10.0042467

the variation that occurs within the bilingual group. The full models were compared to reduced models excluding one of the fixed effects or their interaction using likelihood ratio tests.

Fricative contrast CoG values for L1 Standard Mandarin speakers and QSM–Mandarin bilinguals are shown in Fig. 5.

Model comparisons first indicated that fricative made a significant contribution to model fit: $\beta = -2497.4$, SE = 235.2, $\chi^2(1) = 83.1$, $p < 0.001$. This result reveals that /s/ targets were produced overall with higher mean CoG values than /ʂ/. This pattern is expected given articulatory and acoustic theory. The alveolar sibilant /s/, produced with a more anterior constriction near the alveolar ridge, results in a shorter front cavity and higher-frequency noise, yielding higher CoG values (Jongman *et al.*, 2000; Stevens, 1998). In contrast, the retroflex /ʂ/, articulated further back in the vocal tract with tongue curling, creates a long front cavity that amplifies lower frequencies, resulting in lower CoG values (Li *et al.*, 2014; Duanmu, 2007).

The factor vowel also significantly affected model fit [$\beta = -633.4$, SE = 245.2, $\chi^2(1) = 6.7$, $p < 0.01$], indicating higher CoG values in the [a] context than before [u] across the two groups. This acoustic value reduction before [u] may explain the perceptual finding that /ʂ/ tokens were especially difficult to identify in the [u] context. Consistent with previous studies (Yu, 2019; Fujisaki and Kunisaki, 1978; Mann and Repp, 1980) of contextual coarticulation, the [u] vowel, through lip rounding, lowers CoG and thereby reduces the acoustic distinctiveness of /s/∼/ʂ/.

Moreover, the full model provided a significantly better fit to the data compared to a model excluding the fricative and group interaction [$\beta = 3794.7$, SE = 460.5, $\chi^2(1) = 51.8$, $p < 0.001$], revealing that the CoG contrast between /s/ and /ʂ/ was significantly smaller in the QSM bilinguals than in the L1 Mandarin group, reflecting a reduction in acoustic distinctiveness (at least when considering only CoG) for bilingual speakers. However, this result can be interpreted in different ways.

One possibility is that bilingual QSM speakers can make a distinction between /s/ and /ʂ/, but the acoustic difference they produce is uniformly smaller across the group. Alternatively, the variability within the bilingual QSM group may reflect individual differences in how clearly speakers distinguish this contrast. As observed in Fig. 5, the broader spread in the bilingual data compared to the L1 Mandarin group suggests that some bilingual QSM individuals maintain a robust distinction between /s/ and /ʂ/, while others appear to merge the contrast.

This inter-speaker variability is also reflected in the bimodal shape of the [a]-context distributions in Fig. 5, where two visible peaks likely correspond to the "distinctive" and "merged" subgroups identified in the perceptual analysis. Indeed, this is the pattern revealed by the perceptual analysis reported above. This likely contributes to the overall reduction in effect size at the group level.

We therefore sought to examine whether the perceptual distinction between "distinctive" and "merged" speakers is reflected in their CoG values and to better understand the variability within the bilingual group.

2. Distinctive vs merged bilinguals' productions

We are now focusing on our target QSM bilingual group. Fricative contrast CoG values for "distinctive" and "merged" QSM speakers are shown in Fig. 6.

We fit a similar mixed-effects model, this time focusing only on the bilingual QSM–Mandarin speakers. Instead of the factor group (L1 Mandarin vs bilingual), the model included profile ("distinctive" vs "merged," based on perceptual classification) together with fricative, vowel, and their interactions, and by-item as well as by-participant random intercepts. As before, the full model was compared to simpler models excluding one of the fixed effects or an interaction using likelihood ratio tests.

Model comparisons confirmed that including the factor vowel significantly improved model fit. CoG values were higher before [a] than before [u] [$\beta = -736.4$, SE = 93.1, $\chi^2(1) = 60.5$, $p < 0.001$]. A main effect of the factor Fricative was also observed [$\beta = -1351.2$, SE = 83.3, $\chi^2(1) = 210.5$, $p < 0.001$], with /ʂ/ target exhibiting lower CoG values than /s/ target.

The fricative and profile interaction significantly affected model fit [$\beta = 2705.9$, SE = 166.6, $\chi^2(1) = 217.7$,

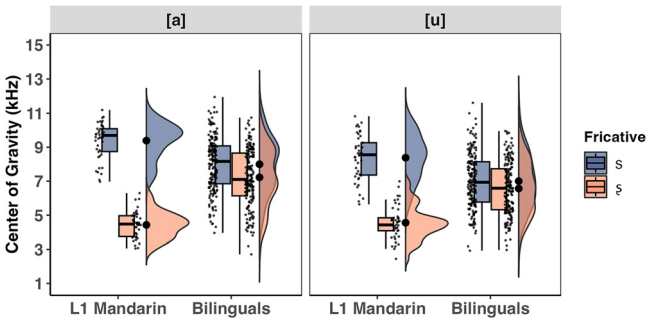

FIG. 5. Comparison of CoG for /s/ and /ʂ/ across two vowel contexts between L1 Standard Mandarin speakers and QSM–Mandarin bilingual speakers. Black dots represent mean CoG values for each condition.

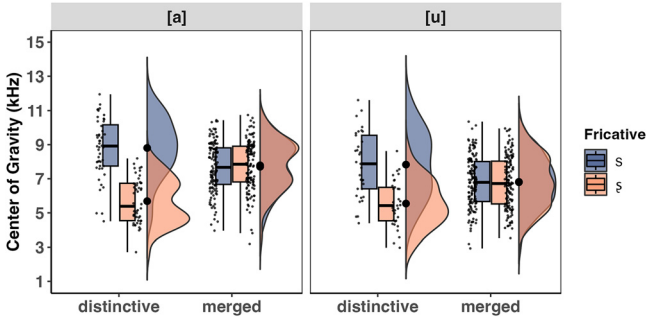

FIG. 6. CoG distributions (in kilohertz) for /s/ and /ʂ/ across two vowel contexts between distinctive and merged bilingual QSM–Mandarin speakers. Each panel combines a boxplot with a half-violin plot. Black dots represent mean CoG values for each condition.

$p < 0.001$]. This indicates that "distinctive" speakers maintained a larger acoustic separation between /s/ and /ʂ/ than did merged speakers, who showed little or no contrast. The three-way fricative, vowel, and profile interaction was also significant [$\beta = -882$, SE = 333.2, $\chi^2(1) = 7.0$, $p < 0.01$], reflecting that the difference between the fricative targets was larger before [a] than before [u], but only for the distinctive group and not for the merged group. As visible in Fig. 6, slight bimodality persists in the [a] context distributions of both groups. For "distinctive" speakers, this reflects fine-grained individual differences in the strength of the /s/∼/ʂ/ contrast, with some producing a more robust separation than others. For merged speakers, the bimodality arises because the merger direction varies across individuals: Some speakers neutralize the contrast toward an alveolar-like realization, whereas others merge toward a retroflex-like realization. These patterns illustrate that even within each profile, bilinguals exhibit gradient and multidirectional variation rather than fully homogeneous production patterns.

Taken together, these results demonstrate that the perceptual classification of the bilingual speakers is corroborated by their production patterns: "Distinctive" speakers maintain a clear CoG separation between /s/ and /ʂ/, whereas "merged" speakers do not. Moreover, the weakening of the contrast before [u] among "distinctive" speakers mirrors the perceptual finding that /ʂ/ targets in this context were especially difficult for Mandarin listeners to categorize, leading to a marked drop in accuracy relative to the [a] context. Thus, the contrast is perceptually confusable because it is acoustically weakened before [u].

Finally, the close alignment between the acoustic results based on CoG values and perceptual judgments reinforces the role of CoG as a strong and reliable spectral measure for distinguishing the /s/∼/ʂ/ contrast in this bilingual population.

3. Extra-linguistic effects

Having examined linguistic effects, we now turn to the role of extra-linguistic factors, specifically, the influence of Mandarin exposure level, age and gender.

Given that our QSM participants all self-identified as bilingual, we focus on their L2 usage frequency. For assessing the extent of Mandarin exposure and use, we relied on their responses to our post-test language use questionnaire. This questionnaire was developed as part of our earlier work [following Weng *et al.* (2023) and Weng *et al.* (2024)] and involves 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 Mandarin use on a five response levels corresponding from 0% to 100% Mandarin (with the complementary proportion attributed to QSM) 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.

Specifically, we analyzed the mean spectral CoG difference between /s/ and /ʂ/ per participant [$\Delta M = M$(/s/ target) – M(/ʂ/target), where larger ΔM values indicate a greater spectral separation between the two fricatives]. ΔM was modeled as a function of individual Mandarin exposure scores, vowel, gender, age group, and speaker profile (distinctive vs merged). We also included the interaction between the Mandarin exposure score and each speaker's profile, as well as the interactions of the Mandarin exposure score with age group and with gender.

Our analysis revealed that speakers with a higher Mandarin exposure score tended to produce a larger contrast between the target fricatives ($\beta = 102.45$, SE = 38.19, $t = 2.68$, $p < 0.01$). The interaction between a speaker's classification and their Mandarin exposure score also shows a significant effect ($\beta = -199.2$, SE = 72.6, $t = -2.7$, $p < 0.01$), indicating that Mandarin exposure scores have a stronger influence on CoG values for distinctive compared to merged speakers (see Fig. 7).

Concerning the other social factors (age group and gender), age group also affected CoG difference ($\beta = -854.39$, SE = 186.14, $t = -4.59$, $p < 0.001$). Specifically, older speakers exhibited smaller CoG differences, suggesting that their production of the /s/∼/ʂ/ contrast was less distinct compared to younger speakers. This reduction in CoG difference with age may reflect a tendency for older speakers to produce more similar fricatives. Gender also had a significant effect on CoG differences ($\beta = 905.57$, SE = 188.99, $t = 24.79$, $p < 0.01$), with women showing larger differences than men. Meanwhile, interaction effects revealed that the impact of Mandarin exposure score on CoG difference varies significantly by age ($\beta = -168.75$, SE = 76.52, $t = -2.21$, $p < 0.05$). Mainly, younger QSM bilingual speakers exhibit a stronger positive effect of Mandarin exposure on CoG difference. In other words, increased Mandarin exposure leads to a stronger acoustic distinction between /s/∼/ʂ/ for younger speakers [$\mu = -0.1$, standard deviation (SD) = 2.2], while older QSM bilingual speakers show less enhancement with more Mandarin exposure ($\mu = -0.7$, SD = 2.3). Interactions with gender were also significant ($\beta = 194.64$, SE = 76.26, $t = 2.55$, $p < 0.05$), suggesting

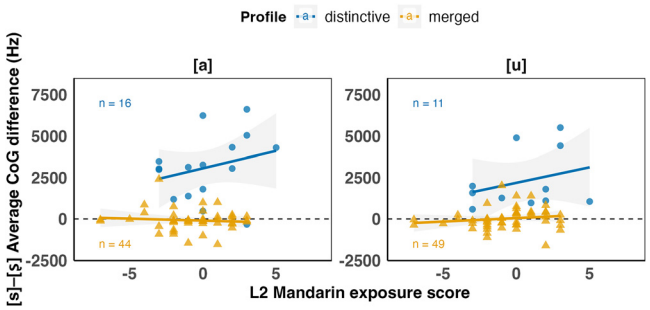

FIG. 7. Relationship between L2 Mandarin exposure and /s/∼/ʂ/ average CoG difference (hertz) for distinctive (blue) and merged (orange) bilingual speakers in [a] (left) and [u] (right) contexts. Lines show linear fits with 95% confidence intervals; sample sizes are indicated in boxes within the figure panels.

https://doi.org/10.1121/10.0042467

that the impact of Mandarin exposure on /s/∼/ʂ/ CoG difference is different for men ($\mu = -1.1$, SD = 2.6) and women ($\mu = 0.1$, SD = 2.2).

IV. DISCUSSION

In this study, we tested the production of the Mandarin sibilant fricative contrast /s/∼/ʂ/ by bilingual speakers of QSM (L1) and Mandarin (L2) in two different vowel contexts (before [a] and before [u]). Our first research question concerned whether bilingual QSM–Mandarin speakers show greater acoustic overlap between /s/ and /ʂ/ than L1 Standard Mandarin speakers. Perceptual identification by L1 Mandarin listeners revealed that many bilingual speakers produced fricatives that were difficult to distinguish, particularly /ʂ/, a pattern consistent with PAM-L2 predictions (Best and Tyler, 2007) for single category assimilation when two L2 sounds are assimilated to a single L1 category. Acoustic analysis of CoG values further confirmed that, compared to L1 Mandarin speakers, our bilinguals demonstrated significantly reduced spectral differences between the two sibilants. Within the SLM-r framework, the absence of a retroflex category in QSM does not in itself predict merger; rather, reduced contrast is expected when the Mandarin /ʂ/ is equivalently classified with an existing L1 sibilant, which can limit new category formation and lead to overlapping acoustic realizations (Flege and Bohn, 2021, p. 23).

The second research question addressed whether the fricative contrast is modulated by vowel context. As predicted, the perceptual evaluation showed that QSM–Mandarin bilinguals tended to produce /ʂ/ in a way that was perceptually confusable with /s/, particularly in the [u] context, where identification accuracy dropped significantly. The acoustic analysis revealed a parallel pattern: Both bilinguals and L1 Standard Mandarin speakers showed a clear co-articulatory effect, with the spectral values of both /s/ and /ʂ/ lowered in the [u] context relative to [a].

Through a clustering analysis based on the perceptual judgments of L1 Mandarin speakers, we categorized our bilingual speakers into two subgroups: "distinctive" and "merged" speakers. Both subgroups showed a coarticulation effect such that CoG values of both fricatives were lower before the vowel [u], but only the "distinctive" group showed a reliable difference between /s/ and /ʂ/ in both vowel contexts, in addition to an interaction such that the distinctive contrast was larger before [a] than before [u]. The coarticulatory pattern for the "distinctive" group therefore closely matches the coarticulatory pattern of L1 Mandarin speakers. This alignment goes beyond what would be expected from well-attested vowel-driven effects on sibilant acoustics alone (Yu, 2019; Fujisaki and Kunisaki, 1978; Mann and Repp, 1980; Ohala, 1993), and instead reflects a structured similarity in co-articulatory behavior. This result connects to broader work on sound change and near-mergers, which has shown that shifts in co-articulatory structure are closely linked to perceptual compensation and category stability (Egurtzegi *et al.*, 2024; Harrington *et al.*, 2008; Kleber *et al.*, 2012). A natural next step would be to test whether these bilinguals also show Mandarin-like perceptual compensation for these co-articulatory effects. Such group-level divergence within a bilingual community would parallel observations in monolingual sound change and provide valuable insight into how production and perception evolve jointly under language contact.

Our third question concerned the role of sociolinguistic variables in predicting contrast strength. We examined whether level of L2 exposure, age, and gender influenced the production of the target fricatives in our bilingual group. Our results revealed that a higher level of L2 exposure tended to coincide with a larger acoustic contrast and that this effect was stronger for "distinctive" speakers than for "merged" speakers. Younger speakers, as well as women, were generally more likely to have higher levels of L2 exposure (Chuang and Fon, 2010; Labov, 1990) and demonstrated greater differentiation in their production of the target fricatives. Importantly, these age- and gender-related patterns should be interpreted within the broader language contact context of the community. Rather than reflecting individual bilingualism alone, the reduced contrast observed among older merged speakers may reflect acquisition of a local variety of Mandarin shaped by long-term contact with QSM and earlier stages of L2 learning, in which the /s/∼/ʂ/ distinction was already weakened. In contrast, younger speakers are growing up in a sociolinguistic environment where greater institutional and social emphasis is placed on Standard Mandarin, leading to increased exposure to standard contrast realization. From this perspective, the observed effects of age and gender appear to be mediated by differences in language exposure and by shifting community norms, reflecting an interaction between individual bilingual experience and the longer-term outcomes of language contact and transmission.

It should be noted that we took a rather coarse measure of L2 exposure, confounding different aspects of language use. Specifically, our compound score was based on five-level self-reports of the proportional use of Mandarin with family, friends, and colleagues, as well as during childhood. While this gives us a good overview of QSM compared to Mandarin use for our participants, a more detailed questionnaire about language use [e.g., the Bilingual Language Profile, as described by Birdsong *et al.* (2012)] could provide more nuanced insights into the relationship between participants' use of QSM and Mandarin and their contrast merger or preservation.

A key observation in this study was the large amount of contrast merger within our bilingual QSM–Mandarin speakers' productions. The majority of our participants did not reliably produce a /s/∼/ʂ/ distinction in either vowel context. Our analysis also identified different patterns among the "merged" speakers' productions, suggesting a variety of profiles. Some speakers neutralized the contrast by producing both /s/ and /ʂ/ as [s] (the most common pattern), while others hypercorrected the contrast, producing both /s/ and /ʂ/ as [ʂ]. These patterns raise the question of whether such

production differences reflect underlying differences in perceptual sensitivity. A growing body of work on contrast mergers and individual differences has explored how perception and production are connected, showing that speakers who perceive distinctions more robustly tend to maintain them more reliably in production (Lee-Kim and Tung, 2025; Yu, 2019; Evjen *et al.*, 2025; Martin *et al.*, 2022; Schertz *et al.*, 2015; Zhang *et al.*, 2022). Although these studies examine different contrasts and speech communities, they demonstrate that production–perception relationships are central to understanding how phonological mergers develop. This body of work suggests that the variability observed in our bilingual speakers' productions may arise, at least in part, from parallel variability in perception. At the same time, studies of mergers and near-mergers have shown that perception and production do not always align perfectly: Some speakers can perceptually distinguish contrasts they do not reliably produce or fail to perceive distinctions they occasionally maintain (Beddor and Krakow, 1999; Bowie, 2000; Kleber *et al.*, 2012). Determining whether bilingual speakers who merge the contrast in production also merge it in perception, or whether they nonetheless perceive distinctively produced contrasts, would therefore be a crucial next step to better understand this bilingual contact situation.

Interestingly, the merger observed in Quanzhou parallels a similar pattern observed in Taiwan Mandarin, where the loss of /s/∼/ʂ/ contrast was initially neutralized under the influence of Taiwan Southern Min (Kubler, 1985). Over time, this change has evolved from a purely phonetic merger to a marker of social identity in Taiwan, reflecting a divergence from Standard Beijing Mandarin norms (Chung, 2006). In Quanzhou, however, the sociolinguistic dynamics appear to be different. Unlike in Taiwan, where local identity is strongly tied to linguistic practices in Mandarin (Lee-Kim and Tung, 2025; Chung, 2006), the Quanzhou region is characterized by strong societal bilingualism: Mandarin dominates public and institutional domains, such as education, media, and professional domains, but QSM is widely spoken in family and community settings. Standard Mandarin explicitly prescribes a clear distinction between /s/ and /ʂ/, making the persistence of this merger and its variability in Quanzhou particularly intriguing.

Despite this widespread exposure to Standard Mandarin norms, most speakers in our study exhibited a tendency to merge the contrast. This suggests that individual exposure alone cannot fully account for the observed phonological variation. Instead, we propose that social identity may play a crucial role in shaping these outcomes, in line with theories of social meaning in sound change (Eckert, 2018; Labov *et al.*, 2006), where speakers subconsciously align with emerging local linguistic norms that go beyond mere language contact. This would suggest that a hybrid identity is emerging in Quanzhou, shaped by both societal-level language norms and individual experiences. Some Quanzhou speakers may distinguish the contrast more clearly to align with the prestige of Mandarin, while others may preserve the merger to affirm their local roots. Clearly much more work is needed to explore the sociolinguistic aspects of this contrast merger, and in particular how they might interact with phonological perception.

In this study, we used CoG as the primary acoustic measure to quantify the spectral differences between the Mandarin sibilant fricatives /s/ and /ʂ/. CoG has been widely used in fricative research due to its simplicity, interpretability, and sensitivity to differences in place of articulation (Li *et al.*, 2014; Forrest *et al.*, 1988; Jongman *et al.*, 2000). While CoG, and spectral moments more generally, remain a common and interpretable choice in sibilant fricative research, they have known limitations, including sensitivity to spectral noise (Reidy, 2016; Shadle, 2023). To address this, we followed the recommendation of Shadle (2012, 2023) to compute CoG from time-averaged magnitude-squared spectra, improving the stability and consistency of our estimates. Despite these precautions, we are aware that CoG captures only the average energy distribution and may overlook important spectral features such as peak location, bandwidth, or spectral shape asymmetries (Shadle, 2023). Recent studies have proposed alternative or complementary measures such as spectral peak tracking (Jannedy and Weirich, 2017) and cepstral coefficients (Reidy, 2016; Spinu and Lilley, 2016; Thomson and Haley, 2014). Future work could incorporate additional acoustic features, such as spectral peaks, band energy ratios, or multi-taper-based estimates, which may provide richer descriptions of fricative acoustics.

While acoustic parameters can reliably distinguish different fricative contrasts cross-linguistically, most studies of sibilant fricatives have treated their spectra as static, by extracting spectral moments from a single window at either the midpoint of the fricative or by averaging measures taken at several time points, with no specific analysis of how the moments vary over time. However, research has shown that spectral characteristics can vary considerably over the time frame of a fricative, indicating that dynamic acoustic parameters can be more effective for capturing contextual differences for sibilants (Iskarous *et al.*, 2011; Koenig *et al.*, 2013; Reidy, 2016). In evaluating cases of fricative mergers, the dynamic characteristics of the fricative productions should thus also be taken into account in future work.

V. CONCLUSION

To conclude, we have explored the variation in the production of the Mandarin phonemic contrast /s/∼/ʂ/ in the speech of bilingual L1 QSM and L2 Mandarin speakers. Our findings reveal significant inter-speaker variability, underscoring the importance of analyzing individual-level data to capture the full complexity of bilingual speech. We noted two important subgroups within our bilingual sample: merged speakers who did not maintain the target contrast in their Mandarin and distinctive speakers who not only maintained the contrast, but actually showed an L1 Mandarin-like coarticulatory pattern. Additionally, we have shown that this phonemic contrast is influenced by both linguistic factors,

such as vowel context, and extra-linguistic factors, including gender, age, and L2 exposure level, highlighting the complex interplay between linguistic and social factors in shaping phonological variation in bilingual speech production.

Future research should further investigate the connection between speech perception and production, through the lens of the interaction between social identity and phonological outcomes in bilingual communities. By adopting this perspective, future studies could provide a deeper understanding of speech variation in bilingual populations and the sociolinguistic dynamics underlying such variation.

ACKNOWLEDGMENTS

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 and from the IdEx Université Paris Cité (ANR-18-IDEX-0001). We thank the editor and the anonymous reviewers for feedback that has greatly improved the quality of the paper and the audience at ISSP 2024 for valuable feedback. We also thank Martijn Wieling for assistance with the statistical analysis.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Ethical approval for this study was obtained at the Université Paris Cité (IRB No. 00012022-95). Informed consent was obtained from all participants.

DATA AVAILABILITY

Scripts and anonymized data are openly available at the Open Science Framework repository: https://osf.io/b8ys4/?view_only=7c24ea57f1964005bc404fe9121a212b. This OSF repository also contains the PRAAT and R scripts, the dynamic spectral analysis procedures, and additional methodological details referred to in the main text.

[1][i] was not included as neither [si] nor [ʂi] is phonotactically well-formed in Mandarin

[2]To assess whether L1 Mandarin listeners' identification responses were driven by acoustic properties of the stimuli, we fitted an additional mixed-effects model predicting response choice from CoG. This analysis confirmed that listeners' categorizations were driven by CoG differences, further reinforcing the role of CoG as a primary acoustic cue in sibilant contrast. Owing to space constraints, this analysis is not reported in the main text, but the full model details are provided in the OSF repository at https://osf.io/b8ys4/?view_only=7c24ea57f1964005bc404fe9121a212b.

Amengual, M. (**2024**). "Phonetics of early bilingualism," Annu. Rev. Linguist. **10**(1), 191–210.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.
Beckman, M. E., Li, F., Kong, E. J., and Edwards, J. (**2014**). "Aligning the timelines of phonological acquisition and change," Lab. Phonol. **5**(1), 151–194.
Beddor, P. S., and Krakow, R. A. (**1999**). "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," J. Acoust. Soc. Am. **106**(5), 2868–2887.
Beristain, A. (**2022**). "Spectral properties of anterior sibilant fricatives in Northern Peninsular Spanish and sibilant-merging and non-merging varieties of Basque," J. Int. Phonetic Assoc. **52**(3), 421–452.
Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, London, UK), pp. 171–204.
Best, C. T., and Tyler, M. (**2007**). "Nonnative and second-language speech perception," in *Language Experience in Second Language Speech Learning*, edited by O.-S. Bohn and M. J. Munro (John Benjamins Publishing Company, Amsterdam, The Netherlands), pp. 13–34.
Birdsong, D., Gertken, L. M., and Amengual, M. (**2012**). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism* (COERLL, University of Texas at Austin, Austin, TX).
Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., and Majid, A. (**2022**). "Over-reliance on English hinders cognitive science," Trends Cogn. Sci. **26**(12), 1153–1170.
Boersma, P., and Weenink, D. (**2023**). "Praat: Doing phonetics by computer (version 6.3.09) [computer program]," http://www.praat.org (Last viewed 15 December 2025).
Bowie, D. (**2000**). "Perception and production in processes of merger," Deseret Lang. Linguist. Soc. Symp. **26**, 18.
Breiman, L. (**2001**). "Random forests," Mach. Learn. **45**, 5–32.
Cai, Q., and Brysbaert, M. (**2010**). "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles," PLoS One **5**(6), e10729.
Casillas, J. V. (**2024**). "The speech production of bilingual adults," in *The Cambridge Handbook of Bilingual Phonetics and Phonology*, edited by M. Amengual (Cambridge University Press, Cambridge, UK), pp. 407–425.
CGTN (**2022**). "Map of the geographical location of Quanzhou City, Fujian Province," https://news.cgtn.com/news/2022-08-11/Quanzhou-a-world-heritage-city-with-a-thriving-economy–1cnJb2eiCBy/index.html (Last viewed 21 April 2025).
Chang, C. B. (**2016**). "Bilingual perceptual benefits of experience with a heritage language," Bilingualism **19**(4), 791–809.
Chang, C. B. (**2019**). "Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research," J. Phonetics **74**, 96–113.
Chang, Y.-H. S., and Shih, C. (**2015**). "Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin," J. Phonetics **50**, 52–66.
Chappell, H. (**2019**). "Southern Min," in *The Mainland Southeast Asia Linguistic Area*, edited by A. Vittrant and J. Watkins (De Gruyter Mouton, Berlin, Germany), pp. 176–233.
Chen, Y., Xu, Y., and Guion-Anderson, S. (**2015**). "Prosodic realization of focus in bilingual production of southern min and mandarin," Phonetica **71**(4), 249–270.
Chiu, C., Wei, P.-C., Noguchi, M., and Yamane, N. (**2020**). "Sibilant fricative merging in Taiwan Mandarin: An investigation of tongue postures using ultrasound imaging," Lang. Speech **63**(4), 877–897.
Chuang, Y.-Y., and Fon, J. (**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*, Chicago, IL (ISCA, Stockholm, Sweden).
Chung, K. S. (**2006**). "Hypercorrection in Taiwan Mandarin," J. Asian Pac. Commun. **16**(2), 197–214.
De Leeuw, J. R. (**2015**). "jsPsych: A JavaScript library for creating behavioral experiments in a Web browser," Behav. Res. **47**, 1–12.
DiCanio, C. (**2021**). *Spectral Moments of Fricative Spectra Script in Praat (Version 4.0)* (Haskins Laboratories & SUNY Buffalo, Buffalo, NY).
Duanmu, S. (**2007**). *The Phonology of Standard Chinese* (Oxford University Press, Oxford, UK).
Eckert, P. (**2018**). *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics* (Cambridge University Press, Cambridge, UK).
Egurtzegi, A., Krajewska, D., Carignan, C., and Urrestarazu-Porta, I. (**2024**). "An acoustic exploration of sibilant contrasts and sibilant merger in Mixean Basque," J. Int. Phonetic Assoc. **54**(2), 677–706.
Evjen, M., Roettger, T. B., and Stausland, S. (**2025**). "Investigating potential merged and distinct speakers in an ongoing merger—Evidence from production and perception," Lab. Phonol. **16**(1), 1–42.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, London, UK), 233–277.
Flege, J. E., and Bohn, O.-S. (**2021**). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by R. Wayland (Cambridge University Press, Cambridge, UK), pp. 3–83.
Flege, J. E., and Eefting, W. (**1987**). "Cross-language switching in stop consonant perception and production by Dutch speakers of English," Speech Commun. **6**(3), 185–202.
Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**(1), 115–123.
Fujisaki, H., and Kunisaki, O. (**1978**). "Analysis, recognition, and perception of voiceless fricative consonants in Japanese," IEEE Trans. Acoust. Speech Signal Process. **26**(1), 21–27.
Gal, S. (**1978**). "Peasant men can't get wives: Language change and sex roles in a bilingual community," Lang. Soc. **7**(1), 1–16.
Gordon, M. J. (**2013**). "Investigating chain shifts and mergers," in *The Handbook of Language Variation and Change*, edited by J. K. Chambers and N. Schilling (Wiley, Hoboken, NJ), pp. 203–219.
Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**(5), 2825–2835.
Hartigan, J. A., and Wong, M. A. (**1979**). "Algorithm as 136: A *k*-means clustering algorithm," J. R. Stat. Soc. C **28**(1), 100–108.
Hazan, V. L., and Boulakia, G. (**1993**). "Perception and production of a voicing contrast by French-English bilinguals," Lang. Speech **36**(1), 17–38.
Hoenigswald, H. M. (**1960**). *Language Change and Linguistic Reconstruction* (University of Chicago Press, Chicago, IL).
Iskarous, K., Shadle, C. H., and Proctor, M. I. (**2011**). "Articulatory–acoustic kinematics: The production of American English /s/," J. Acoust. Soc. Am. **129**(2), 944–954.
Jannedy, S., and Weirich, M. (**2017**). "Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives," J. Acoust. Soc. Am. **142**(1), 395–405.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.
Kang, K.-H., and Guion, S. G. (**2006**). "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals," J. Acoust. Soc. Am. **119**(3), 1672–1683.
Kleber, F., Harrington, J., and Reubold, U. (**2012**). "The relationship between the perception and production of coarticulation during a sound change in progress," Lang. Speech **55**(3), 383–405.
Kochetov, A. (**2017**). "Acoustics of Russian voiceless sibilant fricatives," J. Int. Phonetic Assoc. **47**(3), 321–348.
Koenig, L. L., Shadle, C. H., Preston, J. L., and Mooshammer, C. R. (**2013**). "Toward improved spectral measures of /s/: Results from adolescents," J. Speech Lang. Hear. Res. **56**(4):1175–1189.
Kubler, C. C. (**1985**). "The influence of Southern Min on the Mandarin of Taiwan," Anthropol. Linguist. **27**(2), 156–176.
Labov, W. (**1990**). "The intersection of sex and social class in the course of linguistic change," Lang. Var. Change **2**(2), 205–254.
Labov, W. (**2001**). *Principles of Linguistic Change: Social Factors* (Blackwell, Oxford, UK).
Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, Berlin, Germany).
Lee-Kim, S.-I., and Chou, Y.-C. (**2022**). "Unmerging the sibilant merger among speakers of Taiwan Mandarin," Lab. Phonol. **13**(1), 1–36.
Lee-Kim, S.-I., and Tung, H.-Y. (**2025**). "Linguistic experience and social factors in speech perception: The case of merged speakers of Mandarin sibilants," Lab. Phonol. **16**(1), 1–31.
Li, Z., Zhang, Y., and Li, X. (**2014**). "Acoustic characteristics of voiceless fricatives in Mandarin Chinese/普通话清擦音的声学特征," J. Chin. Linguist. **42**, 150–171.
Lin, Y. (**2018**). "The role of educational factors in the development of lexical splits," Asia Pac. Lang. Var. **4**(1), 36–72.
Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**(6), 3962–3973.
Mann, V. A., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," Percept. Psychophys. **28**(3), 213–228.
Martin, A., van Heugten, M., Kager, R., and Peperkamp, S. (**2022**). "Marginal contrast in loanword phonology: Production and perception," Lab. Phonol. **13**(1), 1–36.
Mayr, R., López-Bueno, L., Fernández, M. V., and Lourido, G. T. (**2019**). "The role of early experience and continued language use in bilingual speech production: A study of Galician and Spanish mid vowels by Galician-Spanish bilinguals," J. Phonetics **72**, 1–16.
Mayr, R., Morris, J., Mennen, I., and Williams, D. (**2017**). "Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English," Int. J. Bilingualism **21**(3), 245–267.
McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (**2017**). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Interspeech 2017*, Stockholm, Sweden (ISCA, Stockholm, Sweden), pp. 498–502.
Mok, P. P., Zuo, D., and Wong, P. W. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**(3), 341–370.
Muxika-Loitzate, O. (**2017**). "Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano," Languages **2**(4), 25.
Nirgianaki, E. (**2014**). "Acoustic characteristics of Greek fricatives," J. Acoust. Soc. Am. **135**(5), 2964–2976.
Ohala, J. J. (**1993**). "Coarticulation and phonology," Lang. Speech **36**(2-3), 155–170.
Pépiot, E., and Arnold, A. (**2021**). "Cross-gender differences in English/French bilingual speakers: A multiparametric study," Percept. Mot. Skills **128**(1), 153–177.
Reidy, P. F. (**2016**). "Spectral dynamics of sibilant fricatives are contrastive and language specific," J. Acoust. Soc. Am. **140**(4), 2518–2529.
Schertz, J., Cho, T., Lotto, A., and Warner, N. (**2015**). "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," J. Phonetics **52**, 183–204.
Shadle, C. H. (**2012**). "On the acoustics and aerodynamics of fricatives," in *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. Huffman (Oxford University Press, Oxford, UK), pp. 511–526.
Shadle, C. H. (**2023**). "Alternatives to moments for characterizing fricatives: Reconsidering Forrest *et al.* (1988)," J. Acoust. Soc. Am. **153**(2), 1412–1426.
Simonet, M. (**2014**). "Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals," J. Phonetics **43**, 26–37.
Spinu, L., and Lilley, J. (**2016**). "A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives," J. Phonetics **57**, 40–58.
Stanford, J. N. (**2008**). "A sociotonetic analysis of sui dialect contact," Lang. Var. Change **20**(3), 409–450.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Stevens, M., and Harrington, J. (**2016**). "The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English," J. Phonetics **58**, 118–134.
Stuart-Smith, J. (**2020**). "Changing perspectives on /s/ and gender over time in Glasgow," Linguist. Vanguard **6**(s1), 20180064.
Tang, C., and Van Heuven, V. J. (**2009**). "Mutual intelligibility of Chinese dialects experimentally tested," Lingua **119**(5), 709–732.
Tang, C., and Van Heuven, V. J. (**2015**). "Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures," Linguistics **53**(2), 285–312.
Thomson, D., and Haley, C. (**2014**). "Spacing and shape of random peaks in non-parametric spectrum estimates," Proc. R. Soc. A **470**(2167), 20140101.
Weng, C., Chitoran, I., and Martin, A. (**2023**). "Bilingual phonological contrast perception: The influence of Quanzhou Southern Min on Mandarin non-sibilant fricative discrimination," JASA Express Lett. **3**(7), 075202.
Weng, C., Chitoran, I., and Martin, A. (**2024**). "Sibilant contrast production by bilingual speakers of Quanzhou Southern Min and Mandarin," in *Proceedings of ISSP 2024—13th International Seminar on Speech Production*, Autrans, France (ISCA, Stockholm, Sweden), pp. 258–261.
Weng, W.-C., and Lee-Kim, S.-I. (**2023**). "Loss of unreleased final stops among Mandarin-Min bilinguals: Structural convergence of languages in contact," J. Phonetics **101**, 101279.

Wikse Barrow, C., Włodarczak, M., Thörn, L., and Heldner, M. (**2022**). "Static and dynamic spectral characteristics of Swedish voiceless fricatives," J. Acoust. Soc. Am. **152**(5), 2588–2600.
Yao, Y., and Chang, C. B. (**2016**). "On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese," Language **92**(2), 433–467.
Yu, A. C. L. (**2019**). "On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation," Lab. Phonol. **10**(1), 2.
Zhang, J., Graham, L., Barlaz, M., and Hualde, J. I. (**2022**). "Within-speaker perception and production of two marginal contrasts in Illinois English," Front. Commun. **7**, 844862.