Selection and implementation of socially motivated phonetic variants: The case of Mandarin sibilants

[1]Zhuo Chen and [2]Sang-Im Lee-Kim
[1]Hanyang University; [2]Yonsei University, Korea
[2]HIPCS (Hanyang Institute for Phonetics and Cognitive Sciences of Language), Hanyang University

chenzhuo0416@hanyang.ac.kr; sangimleekim@yonsei.ac.kr

Abstract

This study investigates how socially conditioned phonetic variation is integrated into the phonological system, focusing on the three-way place contrast among Mandarin sibilants (/s ʂ ɕ/). Previous research has shown that young female speakers tend to produce a more fronted /ɕ/, a variant associated with socially desirable speech styles. In a wordlist reading task, 40 young Mandarin speakers, balanced for gender, produced sibilants followed by /a/ or /u/. The results show that, in the /a/ context, female speakers exhibit higher spectral mean frequency (M1) values for both /ɕ/ and /s/ compared to male speakers. This pattern aligns with cross-linguistic findings linking higher-frequency fricatives to feminine speech styles, while simultaneously reinforcing a robust /s/-/ɕ/ contrast. In the /u/ context, however, increased M1 is observed only for /ɕ/, reducing its spectral separation from /s/. Under these conditions, the onset F2 cue becomes even more prominent, compensating for the diminished spectral contrast. Crucially, this enhancement of the F2 cue is more pronounced among female speakers, enabling the implementation of socially driven /ɕ/-fronting under contextual constraints. Taken together, these findings show that socially meaningful variation can be actively incorporated into the phonological system, balancing social expression with structural stability.

Key words: sibilant fronting, femininity, phonological contrast, cue weighting, Mandarin

1 Introduction

Language variation is shaped by multiple and sometimes competing forces. Speakers exploit phonetic variability to index identity, stance, and affiliation, yet linguistic systems must also preserve phonological contrasts to ensure lexical distinctions. This tension raises a central question of how speakers balance the social pressures shaping phonetic realization with the maintenance of phonological distinctiveness. Sibilants provide an especially revealing case, as their variability has been extensively documented cross-linguistically and shown to carry diverse social-indexical meanings (Stuart-Smith, Timmins, & Wrench, 2003; Zhang, 2005; Stuart-Smith, 2007; Stuart-Smith, Timmins, Pryce, & Gunter, 2013; Bouavichith, 2017; Zimman, 2017; Stuart-Smith & Alam, 2018; Stuart-Smith, 2020; Calder & King, 2022). Against this backdrop, Mandarin Chinese offers a compelling context for investigation, as its sibilant system exhibits complex contrasts maintained by multidimensional phonetic cues, enabling us to examine how such a system is reorganized in the face of socially conditioned variation.

This study focuses on fronted variants[1] of the Mandarin alveolopalatal sibilant /ɕ/ (hereafter, alveopalatal), which have become increasingly salient in recent years. Earlier impressionistic accounts described these variants as indexing femininity (Cao, 1986, 1987; Hu, 1988; Zhu, 2004). Yet experimental evidence on the realization of the Mandarin sibilant system and its social underpinnings remains limited, motivating the present investigation of its current state. Another central issue concerns how phonological distinctiveness is preserved when /ɕ/-fronting reduces the acoustic distance from /s/, thereby potentially weakening the robustness of the contrast. The empirical testing conducted here seeks to inform the broader inquiry of how languages accommodate socially driven phonetic change while maintaining phonological contrast.

1.1. Sibilant fronting: a cross-linguistic gendered speech pattern

Variability in speech production is often attributed to sociolinguistic factors, including gender, age, social class, occupation, and identity. Among these, gender has been extensively examined as a salient variable of cross-speaker variation in speech production. Early research suggests that gender-based differences in speech may stem from underlying physiological distinctions between males and females—such as females having shorter and thinner vocal folds, a shorter vocal tract, and a relatively smaller glottal source. Adult females, thus, generally exhibit higher formant frequencies and spectral noise frequencies than adult male (Peterson & Barney, 1952; Fant, 1966; Fitch & Giedd, 1999).

[1] In the present study, the term *fronting* is used in an acoustic sense to refer to higher-frequency realizations of sibilants. We agree with one of our reviewers that it is difficult to draw direct articulatory inferences from acoustic measures, especially for a parameter such as M1/COG, which may reflect multiple articulatory factors. However, the pattern observed in Mandarin sibilants—unlike in English—involves relatively salient shifts that have been described as crossing phonemic categories, from alveopalatal /ɕ/ toward alveolar /s/. This likely explains why previous work on Mandarin has described such patterns in terms of *fronting* (e.g., Li, 2017; Liao, 2023), a characterization that aligns with native speaker intuitions. Furthermore, while M1/COG can be influenced by both changes in cavity size or length and by variation in source strength (Jesus & Shadle, 2002), the latter appears less likely in the present case. Accordingly, we retain the term *fronting* as a descriptive label following previous literature, without implying a direct one-to-one correspondence with articulatory fronting.

However, mounting evidence from sociophonetic research suggests that gender-based patterns in speech are, at least in part, socially learned behaviors rather than purely anatomically determined (Johnson, 2006; Stuart-Smith, 2007; Simpson, 2009; Foulkes, Scobbie, & Watt, 2010; Fuchs & Toda, 2010; Zimman, 2017; Munson & Babel, 2019; Calder & King, 2022). For instance, Simpson (2009) found that female speakers tend to exhibit higher formant frequencies than males, yet the differences are not uniformly scaled across vowels. Point vowels such as /i/ and /a/ often show more extreme F1 and F2 values in female speech, resulting in a larger and more dispersed vowel space. Such evidence indicates that anatomical differences alone cannot fully account for the gender-based variation. Further support for the social learning account comes from research on children's speech. Both perception and production studies demonstrate that gender-related differences in speech emerge well before consistent sex-based anatomical dimorphism develops (Flipsen, Shriberg, Weismer, Karlsson, & McSweeny, 1999; Perry, Ohde, & Ashmead, 2001; Fox & Nissen, 2005; Li, Rendall, Vasey, Kinsman, Ward-Sutherland, & Diano, 2016; Barreda & Assmann, 2021; Fung, Schertz, & Johnson, 2021; Munson, Lackas, & Koeppe, 2022). Munson et al. (2022) found that young girls produced higher pitch than boys, and that listeners could reliably use pitch cues to identify the speaker's gender. This finding indicates that, even before puberty, female children may raise their pitch to signal gender identity. More broadly, frequency-based cues function as socially meaningful resources for gender expression acquired early in development.

In studies of sibilant variation, gender is typically discussed as a sociocultural identity rather than a biological attribute, encompassing patterns of variation that speakers use to construct or signal aspects of social meaning. Cross-linguistic research has consistently demonstrated that young female speakers are more likely than their male counterparts to employ fronted sibilant variants operationalized by higher spectral frequencies. In Korean, younger females often produce alveolarized alveopalatal affricates /tɕ/ before /a/ and alveolarized alveopalatal fricatives /ɕ/ before /i/ (Kong, Kang, & Seo, 2014; Kong & Kang, 2021). In English, female speakers tend to exhibit higher noise frequencies in /s/ production compared to male speakers (Stuart-Smith et al., 2003). Furthermore, gay men have been shown to produce /s/ with elevated noise frequencies, a pattern that parallels that of female speakers (Linville, 1998; Bouavichith, 2017; Munson, 2026).

Systematic variation in sibilant production is also found in male speech; however, its implementation shows qualitative differences from that of female speech. Cross-linguistically, male speakers are more likely to realize fronted but less sibilant-like fricatives, often approaching a /θ/-like quality. In Korean, male speakers frequently produce fronted, less sibilant-like fricatives, colloquially referred to as "short-tongue sounds," especially in non-/i/ vowel contexts (Kong & Kang, 2020; Holliday & Kong, 2025). In English, male speakers, particularly those in working-class communities, have been reported to produce /s/ with lower spectral peak and mean frequency values. Notably, working-class girls patterned with male speakers in their realization of /s/ (Stuart-Smith et al., 2003; Stuart-Smith, 2007). This practice has been interpreted as an indexical marker of working-class identity, enabling speakers to symbolically reject traditional notions of femininity and to differentiate themselves from middle-class female norms.

Mandarin is not an exception to these cross-linguistic patterns. Early impressionistic studies reported that young female speakers in Beijing often realized alveopalatal sibilants /ɕ tɕ $tɕ^h$/ as alveolar sibilants [s ts ts^h] (Cao, 1986, 1987; Hu, 1988; Zhu, 2004). More recent experimental research by Li (2017) confirmed that the productions of /ɕ/ by young female speakers, as well as /s/ and /ɕ/ by female children aged 4-5, exhibited higher spectral frequencies than those of males, indicating a more anterior articulation. Similarly, Liao (2023) found that gay men who adopted a self-positioning as "wife" in Chinese gay couples also produced fronted /s/ and /ɕ/, which was taken to index a refined and traditionally feminine "wife" persona. Moreover, /θ/-like variants were once frequently observed in Beijing Mandarin among male speakers and state professionals (Zhang, 2005). Taken together, these findings indicate that sibilant variation in Mandarin broadly aligns with cross-linguistic patterns of gender-based speech differences. This striking convergence suggests that gendered sibilant production is systematically conditioned rather than language-specific or incidental, thereby motivating the need for a theoretical framework that can account for these cross-linguistic patterns.

A particularly relevant theoretical proposal for interpreting cross-linguistic gender-related patterns is Ohala's Frequency Code (1983, 1984, 1996). According to this account, lower pitch is associated with larger body size, dominance, and authority, whereas higher pitch signals "smallness" and is linked to perceptual qualities such as submissiveness, friendliness, and uncertainty (Gussenhoven, 2016). From a sociolinguistic perspective, this framework highlights how speakers may exploit the sound-symbolic relationship between frequency and perceived size to construct and project aspects of identity. That is, female speakers often use higher pitch as a strategy to accentuate sex-based differences and thereby project a more feminine persona (Ohala, 1984, 1997). Importantly, this frequency-based symbolism also extends to fricative noise properties. A longitudinal study by Wong, Koeppe, Cychosz and Munson (2025) showed that by around age three, sex-related differences in the spectral centroid of English /s/ are already evident, with girls typically exhibiting higher centroids than boys, corresponding to more fronted articulations. This suggests that high-frequency noise functions as a resource for gender expression acquired early in development. Complementary perceptual studies in English further demonstrate that sibilants with higher-frequency spectral characteristics are consistently perceived as indexing femininity (Munson, Jefferson, & McDonald, 2006; Munson, 2007; Munson, 2011; Mack & Munson, 2012; Munson, Wruck, Benway, & Preston, 2024). In sum, these findings converge to suggest that female speakers often produce more anterior sibilants across languages, since fronting enhances high-frequency acoustic energy that functions as a socially meaningful resource for expressing female identity.

In the context of Mandarin, fronted /ɕ/ has likewise been linked to femininity. Several impressionistic phonetic studies (Cao, 1986, 1987; Hu, 1988; Zhu, 2004) suggest that this articulation is perceived as elegant, delicate, and distinctively feminine. Others have further highlighted broader cultural associations, reflecting admiration for regional cultural prestige. Specifically, Cao (1987) attributed /ɕ/-fronting to female speakers' emulation of fronted sibilant articulations in Suzhou and Shanghai dialects. These regions were historically associated with economic prosperity and cultural prestige and thus carried rich cultural connotations of refinement and aesthetic sophistication. In particular, the association of these dialects with a culturally celebrated ideal of feminine elegance enhanced the influence of their accents. Consequently, the accents of these regions, particularly the fronted realizations that are cross-linguistically associated with feminine qualities, were widely admired and imitated as a means of aligning one's speech with socially desirable styles.

What makes /ɕ/-fronting in Mandarin particularly interesting is that the language has a complex sibilant system, involving a three-way place contrast among alveolar /s/, retroflex /ʂ/, and alveopalatal /ɕ/ sibilants.[2] In this intricate system, fronting of /ɕ/ may have significant implications for maintaining robust phonological contrasts. In English, /s/-fronting does not risk reducing phonological contrasts with /ʃ/; if anything, it may increase phonetic distance between contrasting phonemes. In Mandarin, however, /ɕ/-fronting directly impacts phonological contrast, as it may encroach upon the acoustic space of /s/ and thus weaken the otherwise robust phonemic contrast. Therefore, Mandarin sibilants offer a valuable testbed for examining how speakers ensure optimal contrast among sibilants in response to /ɕ/-fronting. Building on this background, the present study offers foundational work toward a comprehensive understanding of the contemporary Mandarin sibilant system within its socially conditioned landscape of variation.

1.2. Multidimensional cues to Mandarin sibilants under CV coarticulation

Another source of sibilant variation that may weaken spectral separation is phonetically grounded: the coarticulatory influence of adjacent vowels. Coarticulatory vowel effects systematically shape the acoustic realization of sibilants, thereby creating a more complex phonetic landscape for understanding /ɕ/-fronting in Mandarin. The three-way sibilant contrast in Mandarin is sustained by two phonetic cues, which, as reviewed below, may be dynamically reorganized under coarticulatory influence.

The three sibilant categories in Mandarin differ in their place of articulation and tongue posture, which affects the size and shape of the front cavity and thus affects their acoustic characteristics. The retroflex /ʂ/ is typically produced with the tongue tip and blade raised near the post-alveolar region, yielding a long front cavity that is further extended by an extra sublingual cavity formed beneath the raised tongue front (Ladefoged & Wu, 1984; Hamann, 2003; Lee-Kim, 2011; Proctor, Lu, Zhu, Goldstein, & Narayanan, 2012; Lee-Kim, 2014). The alveolar /s/ is produced with the tongue tip approaching the alveolar ridge, yielding a much shorter front cavity. The alveopalatal /ɕ/ is articulated slightly anterior to the retroflex /ʂ/, involving a broad constriction formed by the tongue front against the hard palate. Unlike /ʂ/, it does not involve a sublingual cavity, and the resulting front cavity length is intermediate between those of /s/ and /ʂ/ (Ladefoged & Wu, 1984; Lee-Kim, 2014). These articulatory characteristics are well reflected in the acoustic domain, particularly in the first spectral moment (hereafter, M1)—the mean frequency (Hz) of the fricative spectrum (Forrest, Weismer, Milenkovic, & Dougall, 1988; Jongman, Wayland, & Wong, 2000). Because the spectral shape of a

---

[2] Past studies (Cao, 1986, 1987; Hu, 1987; Zhu, 2004) reported that all alveopalatal sibilants (/ɕ, tɕ, $tɕ^h$/) exhibit fronted variants in female speech. We assume this is likely to hold true, consistent with the general impression as well as with the expectation that fricatives and affricates within the sibilant series share broadly similar place-related acoustic properties. As in Li (2017), the present study concentrates on the sibilant fricatives, comparing the realization of /s/, /ʂ/, and /ɕ/.

fricative is determined by the size and shape of the front cavity (Fant, 1960; Stevens, 1999; Jongman et al., 2000), the three sibilant categories in Mandarin form a consistent M1-based hierarchy: the retroflex /ʂ/ exhibits the lowest M1, the alveopalatal /ɕ/ shows intermediate values, and the alveolar /s/ the highest (Chang & Shih, 2015; Chiu, Wei, Noguchi, & Yamane, 2020; Lee-Kim & Chou, 2022; Hauser, 2023; Lee-Kim & Chou, 2024; Lee-Kim & Tung, 2025). However, the effectiveness of M1 in differentiating /ɕ/ is limited, as its distribution may overlap to some extent with those of the other two sibilants.

Research on Mandarin has shown that F2 at vowel onset (hereafter, onset F2) serves as an important cue for distinguishing /ɕ/ from the non-palatal sibilants /s/ and /ʂ/ (Li, 2008; Chiu, 2009; Hauser, 2023). Onset F2, reflecting the consonant-to-vowel transition, captures underlying articulatory differences: /ɕ/ is characterized by a fronted tongue body posture, yielding a significantly higher onset F2 in the following vowel than is observed for /s/ and /ʂ/, which involve a relatively retracted tongue body (Ladefoged & Maddieson, 1996; Lee-Kim, 2014). The use of multiple acoustic cues is characteristic of languages with a large sibilant inventory. In Polish, which exhibits a parallel three-way sibilant system, onset F2 is significantly higher for /ɕ/ than for /s/ and /ʂ/, and serves as a perceptually salient cue for distinguishing /ɕ/ from the other sibilants (Nowak, 2006; Li, 2008; Chiu, 2009; Cristià, McGuire, Seidl, & Francis, 2011). For example, Nowak (2006) demonstrated that accurate identification of /ɕ/ was drastically reduced when the corresponding vowel transitions were absent, highlighting the indispensable role of onset F2 in perception. It is therefore likely that onset F2 is similarly crucial for the identification of alveopalatal sibilants in Mandarin.

These two phonetic cues to sibilant contrasts—M1 and onset F2—are both subject to contextual variation, with the following vowel further shaping them through consonant-vowel (CV) coarticulation. Importantly, the magnitude of this vowel-induced shift is not uniform across sibilant categories, as some are affected more than others. For example, Jongman et al. (2000) reported that in English, alveolar sibilants /s, z/ exhibited lowered spectral peaks in rounded vowel (/o, u/) contexts, whereas postalveolars /ʃ, ʒ/ remained relatively stable. A parallel asymmetry was observed for onset F2: /s, z/ showed greater variation across vowel contexts than /ʃ, ʒ/. Operationally, since alveolar sibilants typically have lower onset F2 than postalveolars, a greater reduction in onset F2 for alveolars in rounded vowel contexts would further enlarge the acoustic separation between the two sibilant classes. Comparable category-dependent vowel effects have also been observed in Mandarin. Hauser (2023) found that the M1 separation between /s/ and /ʂ/ was attenuated before /u/ relative to /a/, consistent with the possibility that the two sibilants differ in the extent to which they are affected by the rounded vowel context. Likewise, inspection of the reported F2 trajectories indicates that /ɕ/ shows a smaller decrease in onset F2 from /a/ to /u/ than the non-palatal sibilants /s/ and /ʂ/, remaining comparatively stable across vowels (Hauser, 2020). Taken together, these findings demonstrate that vowel context exerts systematic influences on both M1 and onset F2, with effects varying across sibilant categories and potentially further reshaping the acoustic separation among sibilants.

In sum, these vowel-induced shifts in M1 and onset F2 suggest that the acoustic realization of Mandarin sibilant contrasts is shaped by multiple interacting pressures. Mandarin sibilant variation therefore provides a particularly revealing window into how social and phonetic mechanisms jointly contribute to the maintenance and restructuring of sibilant contrasts. In rounded back vowel contexts, the spectral distinction between /s/ and /ɕ/ is likely to be reduced due to the coarticulatory pressure from the vowel, but onset F2 may instead compensate for this reduction. ***In such contexts, how might socially driven /ɕ/-fronting manifest itself, given that its implementation may further reduce spectral distance from /s/?*** The present study therefore extends the analysis beyond M1 to include onset F2, a dimension that has not been systematically examined in previous work. We first characterize how the acoustic space of the three Mandarin sibilants shifts in response to varying contextual pressures, thereby elucidating how multiple acoustic dimensions jointly sustain the /s/ and /ɕ/ contrast under additional social pressures.

1.3. Reorganization of the sibilant system in the face of variation

The preceding sections demonstrated that socially motivated and phonetically grounded sources of variation may jointly modulate the realization of Mandarin sibilants. For one, /ɕ/-fronting in Mandarin may affect phonological contrast, as it can encroach upon the acoustic space of /s/ and thereby weaken the otherwise robust phonemic distinction. For another, coarticulatory influences from surrounding vowels may also affect sibilant realization, particularly insofar as rounded back vowel contexts may attenuate the spectral contrast, while enhancing onset F2 separation between /ɕ/ and the non-palatal sibilants. In response to variation arising from both linguistic and extralinguistic sources, the Mandarin sibilant system may be dynamically reorganized through relative weighting of available cues, while sustaining an optimal level of phonological distinctiveness. The present section turns to how such adaptive responses may operate.

An immediately available response to /ɕ/-fronting, which may reduce contrast along M1, would be corresponding /s/-fronting. An increase in M1 resulting from /ɕ/-fronting can be offset by a concurrent increase in M1 due to /s/-fronting, thereby preserving sufficient spectral separation between the two categories. This idea has been characterized as a chain shift among sibilants by Li (2017). Specifically, Li (2017) examined the production of sibilants by both children and adults and found that, while fronting in adult female speech was limited to /ɕ/, children exhibited fronting patterns for both /s/ and /ɕ/. This result was taken to suggest an emerging chain shift in the speech of children, whereby /s/-fronting arises in response to earlier /ɕ/ fronting. However, the discrepancy between the adult and child groups in this study may be attributable to phonetic accommodation introduced by the experimental design: only the child participants engaged in a word-repetition game, raising the possibility that the reported fronting reflected mirroring of the auditory stimulus rather than spontaneous production.

A possible alternative interpretation is that both /s/ and /ɕ/ encode social meaning, such that fronted variants index femininity in the Mandarin speech community. In languages with relatively small sibilant inventories, /s/ is a common locus of social connotation (Stuart-Smith et al., 2003; Zimman, 2017; Munson, 2026), and female speakers tend to exhibit socially meaningful higher-frequency spectral properties in /s/ production compared to male speakers (Fuchs & Toda, 2010). In a system with a three-way sibilant contrast, as in Mandarin, /s/ as well as /ɕ/ may be recruited to achieve intended social goals in speech production.[3] Notably, this pattern can, as a consequence, enhance spectral separation between /s/ and /ɕ/, allowing speakers to fulfill both social-indexical and phonological optimization simultaneously. In contrast to the more extensively documented fronting of /ɕ/, evidence for /s/-fronting has been largely absent from earlier impressionistic phonetic descriptions of Mandarin. This may be attributed to the weaker perceptual consequences of increase in noise energy in the higher frequency range, which aligns with apparent differences in speakers' awareness of these phonetic variants within the speech community. Anecdotally, /ɕ/-fronting is widely discussed in public social media, whereas variation in /s/ tends to go largely unnoticed.

This study is not designed to determine whether sibilant variation in contemporary Mandarin reflects historical development involving a chain shift toward greater phonological distinctiveness, as proposed by Li (2017). This question cannot be addressed here, as the study does not include comparisons across age groups. Instead, our goal is to establish how the sibilant system is dynamically reorganized in response to social and phonetic pressures, which may or may not ultimately develop into a historical sound change. To this end, several methodological improvements are implemented to support a rigorous acoustic investigation. To obtain a robust empirical dataset, we expand the sample from 10 adults (five per gender) in Li (2017) to 40 adults (20 per gender). Speakers are recruited from Northern Mandarin varieties, including Beijing Mandarin, which is arguably a trend-setting variety in a metropolitan context. The speech materials are controlled for phonological context (i.e., rounded vs. unrounded vowel contexts) in which the sibilants occur, in order to account for known coarticulatory influence on their realization. This design enables us to examine how /ɕ/-fronting and /s/-fronting are implemented among present-day adults, potentially conditioned by gender, and thus to assess the extent of variation that may warrant restructuring of the system.

As discussed in the previous section, the back rounded vowel context poses an additional challenge to socially driven /ɕ/-fronting. In this environment, the effectiveness of spectral cues in frication noise for distinguishing /s/ and /ɕ/ may be reduced, leaving limited room for /ɕ/ to advance along the M1 dimension. Under such conditions, how might socially driven /ɕ/-fronting be realized, given that further fronting could reduce its spectral distance from /s/? Do female speakers continue to implement /ɕ/-fronting along M1, or do they suppress the use of phonetic variants for the sake of the phonological contrast? One possible strategy for coping with the reduced acoustic space along M1 is to recruit an additional dimension, namely onset F2. That is, compensatory adjustment of cue weighting may be employed, whereby speakers adjust the relative contribution of available acoustic cues so that a reduction in the effectiveness of one cue is offset by increased reliance on another. If so, do female speakers rely on F2 cues more than male speakers to implement /ɕ/-fronting as a resource for indexing gender identity?

[3] We thank the anonymous reviewer for insightful comments regarding the interpretation of /s/-fronting, particularly the possibility that it may itself be socially motivated, consistent with cross-linguistic trends in sibilant fronting, rather than functioning solely as a contrast-preserving response to /ɕ/-fronting.

Systematic adjustments of cue weighting have been well documented in the literature. For example, Shultz, Francis and Llanos (2012) found an inverse relationship between voice onset time (VOT) and onset F0 in English stop consonants, whereby speakers who relied more heavily on VOT showed reduced use of onset F0, and vice versa. A similar compensatory pattern has been documented in Mandarin sibilants. Hauser (2023) reported a trade-off between M1 and onset F2 in contrasts involving /ɕ/: speakers who produced greater M1 distinctiveness exhibited diminished onset F2 contrast, and vice versa. This pattern highlights inter-speaker variation in cue weighting strategies used to maintain phonological contrasts. Within individuals, cue reliance may also shift depending on contextual factors. In particular, vowel environment has been shown to modulate the relative contribution of different acoustic cues (Kuang & Cui, 2018; Penney, Cox, Miles, & Palethorpe, 2018). In Australian English, Penney et al. (2018) found that the relative weighting of vowel duration and glottalization as cues to coda stop voicing is conditioned by vowel context. In production, the duration cue was weaker in inherently short vowel contexts compared to inherently long vowels. In perception, glottalization exerted a stronger influence on voicing judgments when the preceding vowel was inherently short. Collectively, these findings underscore the importance of individual differences and contextual variability in cue weighting.

In the present study, differential cue weighting was assessed using Linear Discriminant Analysis (LDA; Fisher, 1936), a method that has been widely applied to estimating cue weights in phonetic contrasts (e.g., Shultz et al., 2012; Schertz, Cho, Lotto, & Warner, 2015; Hauser, 2023). LDA derives the linear combination of acoustic predictors—in this case, M1 and onset F2—that best discriminates the categories /s/ and /ɕ/. The coefficients from this analysis, interpreted as cue weights, indicate the relative contribution of each measure to the classification. This framework provides a means of testing whether females speakers adjust cue weights in rounded vowel contexts to facilitate the implementation of /ɕ/-fronting.

The discussion in this section gives rise to two predictions regarding how gender-related variation may be organized in the Mandarin sibilant system, schematized in Figure 1. The first prediction is that female speakers may show higher M1 values for both /s/ and /ɕ/ than male speakers, whereas /ʂ/ should show little gender difference (Figure 1A). This pattern is interpreted as reflecting selective fronting of sibilants in female speech. Because /ɕ/-fronting may reduce contrast along the M1 dimension, fronting of /s/ would help preserve sufficient spectral distinction between the two sibilants. The second prediction follows from the possibility of compensatory cue-weighting: if socially motivated /ɕ/-fronting along M1 reduce the spectral distance between /ɕ/ and /s/, female speakers may offset the resulting loss of contrast by relying more on onset F2, yielding greater F2 separation between /ɕ/ and the non-palatal sibilants than in male speech (Figure 1B). Such compensatory reliance on F2 is expected to be most evident in back rounded vowel contexts, where coarticulation weakens the effectiveness of M1 while increasing the separability afforded by onset F2.

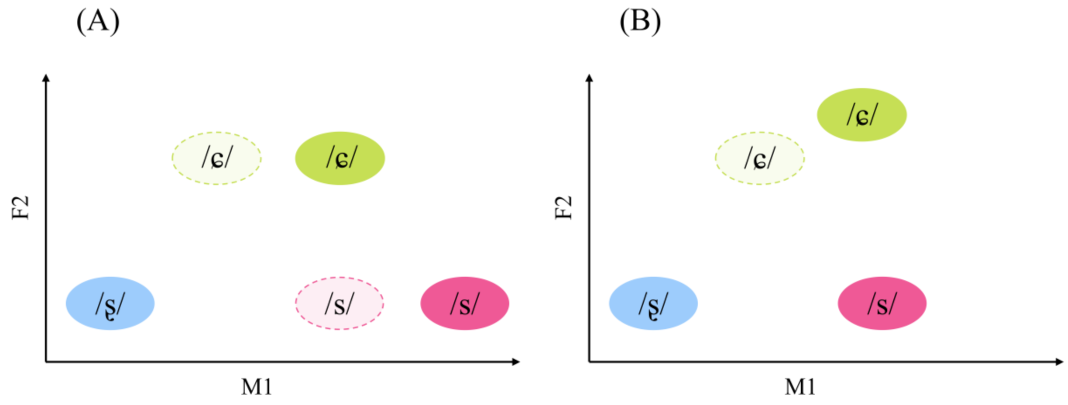

Figure 1. (Color online) Schematic illustration of two predicted patterns of gender-related variation in the Mandarin sibilant system in response to /ɕ/-fronting. The darker-colored circles represent female realizations, whereas the lighter-colored circles with dotted outlines represent the corresponding male realizations.

To evaluate these predictions, the following two research questions are considered:

(1) How is gendered sibilant production realized in Mandarin, a language with a rich sibilant inventory? Which sibilant categories are selected for the loci for the expression of social meaning, and how are these patterns manifested in the acoustic domain?

(2) How does socially motivated variation interact with phonetic pressures arising from coarticulatory contextual variation? How is the multidimensional acoustic space of Mandarin sibilants reorganized to accommodate both socially driven and contextually conditioned variation?

By addressing these questions, this study provides an account for understanding how the sibilant system is dynamically reorganized in response to competing demands to express socially meaningful phonetic variation while maintaining sufficient phonological distinctiveness.

2 Method

2.1. Participants

Forty speakers (20 female, 20 male; aged 18-30), who were college or graduate students at the time of recording, participated in the production experiments. Gender information was collected through a fill-in-the-blank item in the participant information questionnaire, which allowed speakers to self-report their gender. To avoid potential confounds from the /s/-/ʂ/ merger reported in many southern Chinese varieties (Zhang, 2006), all participants were recruited from northern cities of China. Among them, 12 speakers (4 female, 8 male) were from regions where Beijing Mandarin is spoken, and 11 speakers (5 female, 6 male) were from Hebei Province, a region geographically adjacent to Beijing. The language background questionnaire showed that the average self-rated proficiency score of Standard Mandarin was 4.45 (SD = 0.88), and the mean proportion of Standard Mandarin use in daily life was 74% (SD = 21.42). Five participants reported higher proficiency in their local subvarieties of Jilu Mandarin and Jin Chinese, whose sibilant realizations are comparable to those in Standard Mandarin. And two of them reported using their local Mandarin varieties more frequently than Standard Mandarin in daily life. All participants had parents from northern China except three with fathers from southern regions, who nonetheless reported native-level Standard Mandarin proficiency. None reported any hearing or speaking impairments.

2.2. Stimuli

The stimuli list included 30 disyllabic Mandarin target items and 120 fillers. Each target word began with one of three sibilants (/s, ʂ, ɕ/) followed by one of two vowel contexts (/a, u/)[4], yielding six sibilant-vowel combinations at word-initial position (e.g. /a/ context: 撒野 /**sa** jɛ/ 'act wild' ~ 沙发 /**ʂa** fa/ 'sofa' ~ 虾米 /**ɕa** mi/ 'small shrimp'). Five lexical items were selected for each sibilant-vowel combination, yielding 30 target items in total. To minimize potential effects of lexical tone on vowel quality, the word-initial syllable carried either T1 (X^{55}) or T2 (X^{35}). A complete list of stimuli is provided in Table 1.

The selection of /a/ and /u/ reflects phonological constraints on vowel distribution with Mandarin sibilants. Mandarin is generally assumed to have five underlying vowels, /i, y, u, ə, a/ (Duanmu, 2007), but their occurrence is subject to various restrictions: [i] and [y] appear only after /ɕ/, while [ə] appear only after /s/ and /ʂ/ (*[si], *[ʂi], ✓[ɕi]; *[sy], *[ʂy], ✓[ɕy]; ✓[sə], ✓[ʂə], *[ɕə]). Consequently, [a] and [u] provide the contexts in which all three sibilants can freely co-occur. In some analyses, alveopalatal-vowel sequences are transcribed as [ɕia] and [ɕiou] (Lee & Zee, 2003) or as [ɕja] and [ɕjou] (Lin, 2007). In this study, however, we follow previous work (Ladefoged & Maddieson, 1996; Proctor et al., 2012; Hauser, 2023) in adopting [ɕa] and [ɕu][5]. As the present analysis relies on acoustic measurements, these assumptions about the underlying phonemic representation are inconsequential.

Word frequencies were obtained from the BLCU Corpus Center (https://bcc.blcu.edu.cn/) and log-transformed. The data were then analyzed with a regression model including sibilant type as the predictor, implemented in the base function *lm()* in R v.4.3.3 (R Development Core Team, 2024). Using /s/ as the reference level, the analysis revealed no significant effects: neither /ɕ/ (p = .115) nor

[4] As shown in the stimulus set, two of the five words used in the /a/ condition for each sibilant contain the diphthongal rime [aʊ], a design choice that was difficult to avoid given constraints related to lexical tone, word frequency, and lexical diversity. Upon further examination, [aʊ] was found to exhibit a lower onset F2 than [a]. Including [aʊ] within the broader /a/ category therefore likely renders /a/ acoustically more similar to /u/, yielding a conservative estimate of vowel-context effects without changing the overall direction of the effects. Moreover, because [aʊ] tokens are systematically distributed across sibilant categories, this analytical decision is unlikely to affect the interpretation of gender- or sibilant-related patterns, the primary focus of the present study. The current grouping was therefore retained in order to preserve the original /a/-/u/ comparison, while acknowledging this limitation.

[5] The underlying form of the rhyme in /ɕu/ words is analyzed as /əu/, which surfaces as [ou] through rhyme harmony (Duanmu 2000; Lin 2007). Examination of the present data confirms that F1 shows a slight rise during the rhyme following /ɕ/, consistent with this diphthongal quality. However, since [o] and [u] are both back vowels with comparable backness, the distinction between [ou] and [u] is primarily reflected in F1 rather than F2. Given that the present study focuses on onset F2, this vowel difference is therefore unlikely to affect the relevant measurements, and /ɕu/ words are treated as equivalent to /u/ contexts for the purposes of the current analysis, consistent with previous studies (e.g., Hauser 2023). Accordingly, the labels /a/ and /u/ are used throughout the present study and are not intended to imply any particular phonological analysis of the individual stimuli.

/ʂ/ (p = .188) differed in word frequency from /s/. Table 1 summarizes the frequency information for the target items.

Table 1. Stimulus of the wordlist reading task.

	/a/ context		/u/ context	
	Stimulus	Freq	Stimulus	Freq
/s/	/**sa**^{55}jɛ^{214}/ 撒野 ‘act wild’	7.15	/**su**^{55}tʂou^{55}/ 苏州 ‘Suzhou city’	10.36
	/**sa**^{55}njaʊ^{51}/ 撒尿 ‘pee’	7.40	/**su**^{55}tsʰwei^{51}/ 酥脆 ‘crisp’	6.42
	/**sa**^{55}tɕaʊ^{55}/ 撒娇 ‘act cute’	8.93	/**su**^{35}tɕʰi^{51}/ 俗气 ‘vulgar’	7.59
	/**sa**ʊ^{55}ʐaʊ^{214}/ 骚扰 ‘harass’	9.07	/**su**^{35}y^{214}/ 俗语 ‘proverb’	7.77
	/**sa**ʊ^{55}tʰoʊ^{35}/ 搔头 ‘scratch head’	5.58	/**su**^{55}tan^{55}/ 苏丹 ‘Sudan’	8.80
/ʂ/	/**ʂa**^{55}fa^{55}/ 沙发 ‘sofa’	10.34	/**ʂu**^{55}tjɛn^{51}/ 书店 ‘bookstore’	9.67
	/**ʂa**^{55}tʂʰə^{55}/ 刹车 ‘brake’	8.52	/**ʂu**^{55}ny^{214}/ 淑女 ‘lady’	9.01
	/**ʂa**^{55}y^{35}/ 鲨鱼 ‘shark’	8.37	/**ʂu**^{35}tʂi^{55}/ 熟知 ‘know well’	8.29
	/**ʂa**ʊ^{35}tsi^{214}/ 勺子 ‘spoon’	7.70	/**ʂu**^{55}fu^{35}/ 舒服 ‘comfortable’	11.20
	/**ʂa**ʊ^{55}fan^{51}/ 烧饭 ‘cook’	7.41	/**ʂu**^{55}tsi^{214}/ 梳子 ‘comb’	7.53
/ɕ/	/**ɕa**^{55}mi^{214}/ 虾米 ‘small shrimp’	8.21	/**ɕu**^{55}ɕi^{55}/ 休息 ‘rest’	11.57
	/**ɕa**^{55}kwaŋ^{51}/ 瞎逛 ‘wander’	6.11	/**ɕu**^{55}tɕa^{51}/ 休假 ‘take a vacation’	8.99
	/**ɕa**^{35}kʰə^{51}/ 侠客 ‘swordsman’	7.31	/**ɕu**^{55}li^{214}/ 修理 ‘fix’	9.56
	/**ɕa**ʊ^{55}ɕi^{55}/ 消息 ‘message’	11.81	/**ɕu**^{55}kʰwei^{51}/ 羞愧 ‘ashamed’	8.56
	/**ɕa**ʊ^{55}jɛ^{51}/ 宵夜‘night snack’	9.47	/**ɕu**^{55}ny^{214}/ 修女 ‘nun’	8.31

2.3. Procedure

The experiment was conducted at Beijing Foreign Studies University in Beijing, China, using a randomized word list reading task. Recordings were collected using a Zoom H4n Pro Handy Recorder and a Shure head-worn dynamic microphone, with a sampling rate of 44.1 kHz and a bit depth of 16. The 150 items were randomly shuffled four times to create four distinct word lists, each containing the same items in a different order. Stimuli were presented in Chinese characters on PowerPoint slides, one word per slide, which remained visible for two seconds before automatically advancing. Participants were instructed to read the words in a natural speaking style.[6] After completing two lists, they took a short break before proceeding with the remaining two. The entire task lasted about 30 minutes, and participants received monetary compensation for their participation.

2.4. Acoustic analysis

[6] Representative sound files are available on OSF (https://osf.io/ugd7s/overview?view_only=70d16fcca8ea4779b343f12ff319a926).

The reading task resulted in four repetitions of 30 target words, yielding a total of 4,800 tokens from 40 participants. A total of 22 tokens were excluded due to speech errors, leaving 4,778 tokens for analysis. All tokens were initially aligned using the *Montreal Forced Aligner* (MFA; McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017) and subsequently manually adjusted in Praat (Boersma & Weenink, 2023) by the first author. The frication noise was marked between the onset and offset of the high-frequency noise. The onset of the following vowel was identified as the point at which the first and second formants became clearly visible in the spectrogram. This point typically corresponds to the onset of periodic voicing and the emergence of well-defined formant structure, indicating the beginning of vocalic articulation.

The frication noise was then saved individually and submitted to a multitaper spectral analysis (Blacklock & Shadle, 2003; Blacklock, 2004) in MATLAB. This spectral analysis estimates the spectrum by averaging multiple eigenspectra obtained from differently tapered versions of the signal, thereby reducing the variance of spectral estimates and yielding more stable spectral estimates (Reidy, 2013). It has been widely adopted in previous studies (e.g., Zygis, Pape, & Jesus, 2012; Koenig, Shadle, Preston, & Mooshammer, 2013; Shadle, Chen, & Whalen, 2016; Chodroff & Wilson, 2022; Lee-Kim & Chou, 2022, 2024). A 25-ms window was applied to three temporal intervals of the frication noise, and the mid-phase interval (Figure 2) was used for further analysis, as it was expected to show the most genuine acoustic properties of the sounds under investigation without much interference from surrounding segments. A high-pass filter at 1000 Hz was applied to the frication interval to filter out low-frequency noise to eliminate the effects of voicing from surrounding vowels. Spectral moments were calculated from the normalized linear power spectrum generated by the multitaper analysis.

Among the four spectral moments (Forrest et al., 1988) drawn from the multitaper spectral analysis, M1 was selected to characterize the noise properties of the fricatives. M1 represents the mean of the spectral energy distribution and has been widely adopted in fricative studies due to its strong correlation with place of articulation (Forrest et al., 1988; Jongman et al., 2000). In Mandarin Chinese, the three-way contrast among /s/, /ʂ/, and /ɕ/ involves differences in constriction location and tongue posture (Lee-Kim, 2014). Accordingly, M1 provides a useful proxy for capturing their place-related acoustic distinctions, especially for the /s/-/ʂ/contrast, with /s/ showing the highest values, /ʂ/ the lowest, and /ɕ/ falling in between (Chang & Shih, 2015; Li, 2017).

F2 values of the following vowel were obtained using custom Praat scripts. F2 trajectories were sampled at 5-ms intervals from 5 ms after vowel onset to 10 ms before offset, and onset F2 was measured at 20 ms after vowel onset (Figure 2). All automatically extracted values were manually corrected when necessary to ensure consistent and reliable data for subsequent statistical analyses.[7]

[7] Automatically extracted formant values were screened for potential outliers both by inspecting the raw data table for anomalous values and by visualizing the measurements using scatter plots. Tokens showing unusually large deviations within the same condition were revisited and examined in the spectrogram and formant tracks. When tracking errors were identified, the formant ceiling was adjusted and the formant values were corrected with reference to the formant bands visible in the spectrogram and the surrounding formant trajectories.

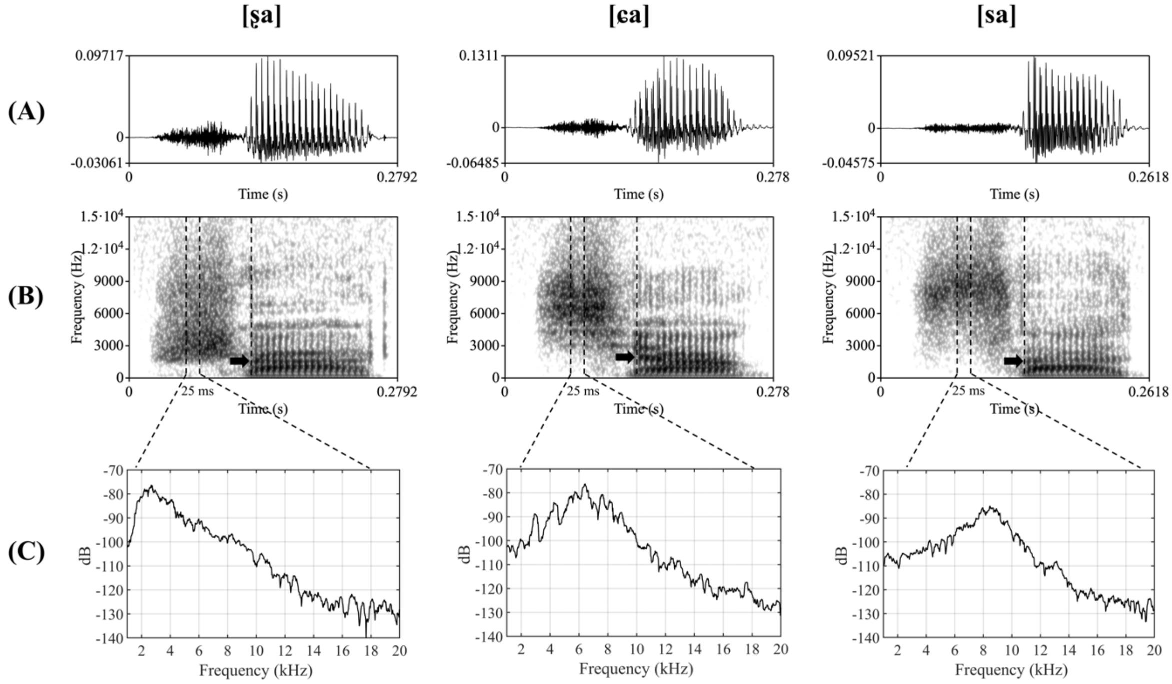

Figure 2. Examples of the spectral analysis. Waveform (A) and spectrogram (B) of [ʂa], [ɕa], and [sa], produced by a male speaker (M18). A 25-ms fricative segment was used for multitaper spectral analysis and the corresponding spectra (C) are shown below. Dotted lines on the vowel portion mark the time point for F2 extraction: 20 ms after vowel onset. The arrows in the spectrograms indicate the second formant (F2).

2.5. Assessing differences in cue weighting

In back rounded vowel contexts, where M1 is a less effective cue for distinguishing the /s/-/ɕ/ contrast, socially driven /ɕ/-fronting in female speech is expected to further reduce the M1 separation between the two categories. However, the /ɕ/-/ʂ/ contrast may remain largely unaffected; if anything, its M1 separation may even increase. Under these conditions, the onset F2 cue for /s/-/ɕ/ contrast may become even more prominent in female speakers, compensating for the diminished spectral contrast. Accordingly, the analysis of cue weighting focused on the /s/-/ɕ/ contrast. To investigate context-dependent cue-weighting strategies in the categorization of /s/ and /ɕ/ by individual speakers, Linear Discriminant Analysis (LDA; Fisher, 1936) was performed. Prior to the analysis, M1 and onset F2 were standardized using within-speaker z-scores to make the LDA coefficients comparable across the two acoustic dimensions (Shultz et al., 2012; Schertz et al., 2015). LDA was then performed separately for each speaker within each vowel context (/a/ and /u/) using the *lda()* function from the *MASS* package (Venables & Ripley, 2002) in R. The resulting discriminant coefficients were interpreted as cue weights, representing the relative contributions of M1 and onset F2 to the classification of the /s/-/ɕ/ contrast. A regression analysis was then conducted using the *lmer* function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R to quantify the shifts in F2 cue weights as a function of vowel context, gender, and their interaction.

3 Result

3.1. M1 realization in Mandarin sibilants

Figure 3 summarizes by-subject distributions of empirical M1 values of alveolars (pink), retroflexes (blue) and alveopalatals (green). The results are paneled by GENDER and VOWEL. Regression analysis was conducted to identify the sources of variation in sibilant production. The linear mixed-effects regression model was fitted using the *lmer* function in the *lme4* package (Bates et al., 2015) in R.[8] The dependent variable was M1 values (Hz). The fixed effects included GENDER (two levels: Female vs. Male), SIBILANT (three levels: Alveolar, Alveopalatal, Retroflex), and VOWEL (two levels: /a/ vs. /u/). The model used R's default treatment coding scheme, with female, alveopalatal, and /a/ vowel serving as the reference levels. In addition to the main effects, all possible interaction terms among the predictors were modeled to assess whether their effects on M1 values were further modulated by each other. The random effects included by-subject random intercepts and slopes for the SIBILANT and VOWEL interaction, as well as by-word random intercepts. Table 2 summarizes the results of the regression model.

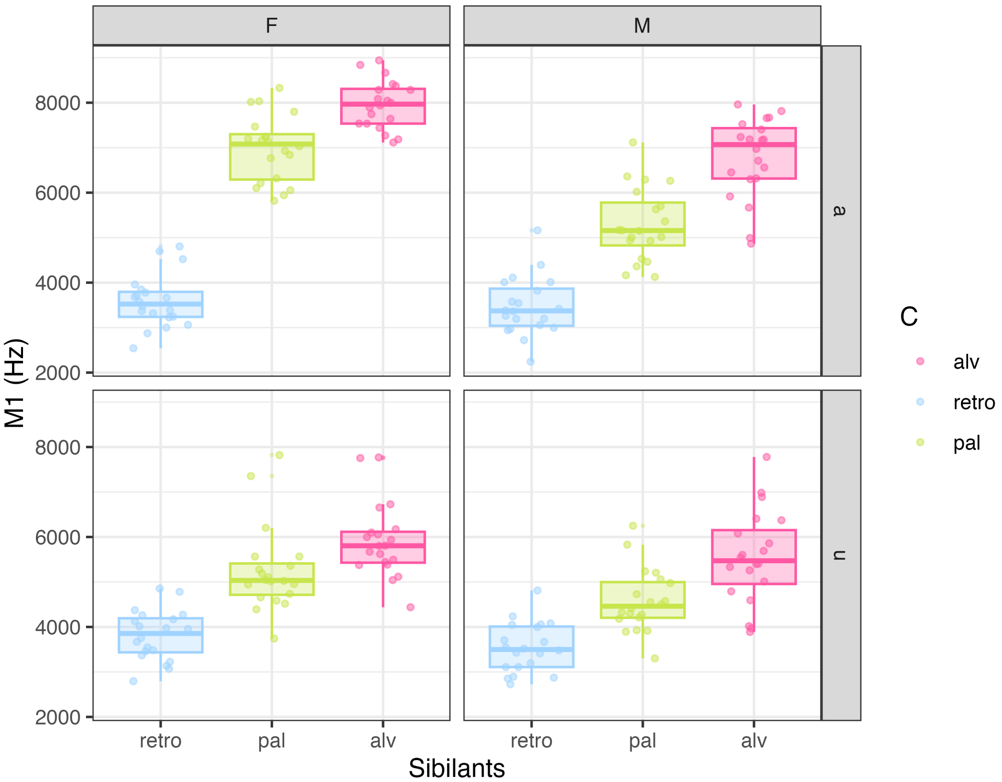

Figure 3. (color online) By-subject distributions of empirical M1 values (Hz) of alveolars (pink), retroflexes (blue) and alveopalatals (green). The results are divided by gender and vowel.

---

[8] All data files and R codes used in the analysis are available on OSF (https://osf.io/ugd7s/overview?view_only=70d16fcca8ea4779b343f12ff319a926).

Table 2. Summary of the mixed-effects regression Model. Formula: *M1* ~ GENDER*SIBILANT*VOWEL + (1 + SIBILANT*VOWEL|SUBJECT) + (1|WORD). Significant results are in bold.

Predictors	Coefficient	*SE*	*t* value	*p* value
Intercept	**6980.69**	**190.20**	**36.70**	**<0.0001**
SIBILANT$_{alv}$	**976.19**	**191.82**	**5.09**	**<0.0001**
SIBILANT$_{retro}$	**-3393.64**	**224.26**	**-15.13**	**<0.0001**
Vowel$_{/u/}$	**-1728.46**	**161.59**	**-10.70**	**<0.0001**
Gender$_{M}$	**-1692.81**	**244.74**	**-6.92**	**<0.0001**
Gender$_{M}$:Vowel$_{/u/}$	**1023.78**	**165.28**	**6.19**	**<0.0001**
Vowel$_{/u/}$:SIBILANT$_{alv}$	-310.00	217.97	-1.42	0.1608
Vowel$_{/u/}$:SIBILANT$_{retro}$	**1955.63**	**224.81**	**8.70**	**<0.0001**
Gender$_{M}$:SIBILANT$_{alv}$	**506.70**	**220.71**	**2.30**	**0.0273**
Gender$_{M}$:SIBILANT$_{retro}$	**1574.51**	**275.09**	**5.72**	**<0.0001**
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{alv}$	-212.59	212.70	-1.00	0.3239
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{retro}$	**-1182.46**	**226.43**	**-5.22**	**<0.0001**

Reference levels were set as: SIBILANT = alveopalatal, VOWEL = /a/, GENDER = female.

Starting with the M1 realization at the reference level (SIBILANT = alveopalatal, VOWEL = /a/, GENDER = female), both SIBILANT$_{alv}$ ($\beta = 976$, $p < .0001$) and SIBILANT$_{retro}$ ($\beta = –3{,}393$, $p < .0001$) exerted significant effects on M1, but in opposite directions. The /s/ productions (M = 7,957 Hz) exhibited markedly higher M1 values than those of /ɕ/ (M = 6,981 Hz), whereas /ʂ/ (M = 3,587 Hz) showed much lower M1 values than /ɕ/. This pattern reflects established articulatory differences: /s/ has a shorter front cavity than /ɕ/, while /ʂ/ involves the longest cavity due to the additional sublingual space. This overall pattern reveals a systematic M1 hierarchy across the three sibilant categories (M1: /s/ > /ɕ/ > /ʂ/), aligning with previous acoustic descriptions of Mandarin sibilants (Lee-Kim, 2011; Kallay & Holliday, 2012).

Building on these acoustic differences, we predicted that the M1 realization of /ɕ/ would be further shaped by both linguistic and extralinguistic factors. This question was addressed through the inclusion of the predictor VOWEL and GENDER. Consistent with expectations, VOWEL exerted a significant effect on the M1 values of /ɕ/ ($p < .0001$). This effect reflects the articulatory properties of /u/—tongue retraction and lip rounding—led to lower M1 values of /ɕ/ (M = 5,252 Hz) compared to /a/ (M = 6,981 Hz). This finding supports that vowel environment shapes the spectral characteristics of the sibilants through coarticulatory influences. Turning to extralinguistic influences, a significant effect of GENDER emerged in the realization of /ɕ/ ($p < .0001$). Female speakers produced /ɕ/ with significantly higher M1 values (M = 6,981 Hz) than their male counterparts (M = 5,288 Hz). Such gender-driven spectral differences are consistent with previous impressionistic reports in Mandarin that younger females often produce fronted /ɕ/ associated with feminine speech styles (Cao, 1986, 1987; Hu, 1988).

Furthermore, GENDER interacted with VOWEL ($p < .0001$), such that the magnitude of the gender difference in /ɕ/'s M1 varied across vowel contexts. The difference was smaller in the /u/ context (Δ = 669 Hz) than in the /a/ context (Δ = 1,693 Hz). As shown in Figure 3, the boxplots illustrate that female speakers consistently produced higher M1 values for /ɕ/ (green) than male speakers, although the separation between the two gender groups was substantially reduced in the /u/ panels (bottom row) relative to the /a/ panels (top row). This pattern suggests that the socially driven gender expression would be suppressed under the coarticulatory influence of /u/.

The patterns observed for /ɕ/ were not found consistently for other sibilants, as evidenced by several higher-order interactions involving the predictor SIBILANT. To examine the patterns of the /s/ and /ʂ/ categories, we fitted separate models using each as the reference levels. The results are summarized in Appendices 1 and 2. Focusing on the /s/ category, both VOWEL and GENDER showed significant effects on its M1 values ($p < .0001$). The results confirm that /s/, like /ɕ/, may show a degree of retraction in the /u/ context, while female speakers (M = 7,957 Hz) consistently produced higher M1 values than males (M = 6,771 Hz), consistent with spectral patterns typically associated with more anterior realizations. This pattern may reflect a tendency toward female-led fronting of /s/ in the /a/ context, consistent with the pattern observed for /ɕ/. In addition, a significant GENDER:VOWEL interaction also emerged ($p = .0018$), with the gender difference in /s/ was substantially reduced in the /u/ context, paralleling the reduced gender difference observed for /ɕ/. This finding reflects a common coarticulatory constraint imposed by /u/ on both categories.

Notably, /s/ and /ɕ/ share certain similarities: in both categories, female speakers exhibit higher M1 values than males, and the gender difference in M1 is attenuated in the /u/ context. However, as indicated by the significant GENDER:SIBILANT$_{alv}$ interaction ($\beta = –507$, $p = .0273$) in the main model, the gender difference in M1 values in the /a/ context was larger for /ɕ/ (M = 1,693 Hz) than for /s/ (M = 1,186 Hz), pointing to a stronger female-led fronting effect for /ɕ/. This is clearly illustrated in Figure 3: in the /a/ panels (top row), the boxplots reveal a smaller separation between female and male speakers for /s/ (pink) compared to /ɕ/ (green). Taken together, although both sibilants are shaped by the combined influence of VOWEL and GENDER, /ɕ/ shows a stronger tendency toward gender-related fronting than /s/. This finding underscores the particular sociolinguistic salience of /ɕ/ and suggests that it serves as a more prominent indexical marker of gendered speech style in Mandarin.

Next, the pattern for /ʂ/ diverges substantially from that of /ɕ/, as evidenced by the significance of all higher-order interactions involving SIBILANT$_{retro}$ in the main model. According to the results of the /ʂ/ model (see Appendix 2), the main effect of VOWEL on the M1 values of /ʂ/ was not significant ($p = .1284$). This suggests that the vowel context had little influence on the M1 realization of /ʂ/. One plausible explanation is that the retroflex configuration already involves a long front cavity, shaped jointly by the large sublingual cavity and slight lip rounding, which may collectively contribute to lower M1 values (Ladefoged & Maddieson, 1996; Hamann, 2003; Lee-Kim, 2014). As a result, although vowel context may induce some coarticulatory adjustments, these adjustments may not further alter the cavity configuration relevant to M1, which may account for the relatively stable M1 values of /ʂ/ across vowel contexts. Furthermore, the effect of GENDER, as well as its interaction with

VOWEL, were not significant (p = .5493, p = .2472), which shows the absence of gendered fronting for /ʂ/. Considering the significant gender differences observed for /s/ and /ɕ/, this result provides evidence against the anatomical account, which predicts uniform GENDER effects across all sibilants. Instead, it lends further support to a socially conditioned explanation, whereby speakers selectively modify the spectral characteristics of specific sibilants to construct socially meaningful speech styles.

The preceding analyses centered on the /a/ context, where gender effects reflecting social factors were clearly observed. Yet previous results also indicated that the /u/ vowel imposes a strong coarticulatory constraint, reducing the magnitude of the gender difference in M1. This raises the question of whether such gender-related differences in sibilant production, which have been interpreted as reflecting fronting, remain observable in /u/. To address this question, post-hoc comparisons were conducted on the fitted linear mixed-effects model. Pairwise contrasts between female and male speakers were extracted for each sibilant in the /u/ context using the *emmeans* package (Lenth, 2023) in R, with p-values adjusted by the Bonferroni method to account for multiple comparisons. The results of the post-hoc comparisons summarized in Appendix 3 revealed that in /u/, a robust gender-related difference was observed only for /ɕ/ (Δ = 669 Hz, p = .0109), whereas the estimated gender differences for /s/ (Δ = 375 Hz, p = .2037) or /ʂ/ (Δ = 277 Hz, p = .1139) did not reach statistical significance. Taken together, the pattern in the /u/ context reinforces the social account of gender-specific sibilant variation, whereby female speakers selectively enhance higher-frequency spectral energy in specific sibilants to construct socially meaningful speech styles. It further supports the interpretation that /ɕ/ functions as a stronger indexical resource than /s/, since the gender-related M1 difference is only consistently observed for /ɕ/ across both vowel contexts.

To summarize, gender-related variation was limited to a subset of Mandarin sibilants, which points to socially driven patterns as a source of variation in sibilant production. Crucially, this variation showed vowel-dependent asymmetry. In the /a/ context, both /s/ and /ɕ/ exhibited gender-related spectral differences, reflecting cross-linguistic patterns associated with feminine speech, while maintaining robust spectral contrasts between the two sibilants. In the /u/ context, however, this gender-related pattern was observed only for /ɕ/. The absence of a comparable gender effect for /s/ may signal a weakening of the M1-based distinction between the two sibilants. This raises the question of how female speakers maintain robust distinction between /s/ and /ɕ/ when M1 cue is weakened. To address this, we turn to onset F2 as a potential compensatory resource and examine how it is modulated by multiple factors.

3.2. Onset F2 realization in Mandarin sibilants

Data visualization of F2 is presented in two ways. Figure 4 presents GAM-smoothed (k = 7) F2 trajectories over normalized vowel duration for alveolars (pink), retroflexes (blue), and alveopalatals (green), based on the empirical data.[9] These trajectories illustrate the dynamic properties of the

[9] Figure 4 presents GAM smooths fitted to the raw F2 trajectories for descriptive purposes, intended to provide an intuitive illustration of trajectory patterns across sibilant and vowel context conditions. The confidence intervals reflect uncertainty in the smoothed trajectories but do not incorporate by-speaker or by-item random effects, and should therefore not be interpreted as inferential. Statistical

formant transition and reveal that /ɕ/ exhibits the steep falling F2 pattern, with the separation between /ɕ/ and the other two sibilants becoming more pronounced in the /u/ context than in /a/. Since the largest differences among the sibilant categories occur near the onset of the vowel, onset F2 was extracted at 20 ms into the vowel as a representative point along the transition. Figure 5 summarizes the empirical distributions of these onset F2 values together with the M1 distributions analyzed in the previous section, providing an integrated view of the two major acoustic cues. In general, the two-dimensional patterns show that /s/ and /ʂ/ remain well differentiated along the M1 dimension, while /ɕ/ is distinguished from the other two sibilants by its substantially higher onset F2. To statistically evaluate the onset F2 differences, a linear mixed-effects model, similar to that in Section 3.1, was fitted with onset F2 as the dependent variable. The fixed and random effects structure, as well as the coding scheme, were identical to those used in the M1 analysis. Table 3 summarizes the regression model fit.

The regression model revealed significant effects of all predictors on onset F2 values at the reference level (SIBILANT = alveopalatal, VOWEL = /a/, GENDER = female). Specifically, there were significant main effects of both $\mathrm{SIBILANT}_{alv}$ (p = .0006) and $\mathrm{SIBILANT}_{retro}$ (p = .0015), with /ɕ/ (M = 1,935 Hz) exhibiting a much higher onset F2 value than /s/ (M = 1,621 Hz) and /ʂ/ (M = 1,652 Hz). Notably, the high onset F2 of /ɕ/ is still retained despite its significant fronting, as verified in the M1 analysis. Although the frication noise shifts toward higher-frequency regions, approaching that of /s/ and reflecting a shortened front cavity, the fronted /ɕ/ variants in Mandarin nonetheless retain a robustly palatalized quality through the sustained vowel transitions. This pattern is evident in the female panels of Figure 4: the F2 transition following /ɕ/ remains substantially greater than those following /s/ and /ʂ/, closely paralleling the trajectory pattern observed for male speakers, who tend to produce a more canonical /ɕ/ realization. A comparable pattern has been reported in Polish female speech: the fronted variant of /ɕ/ exhibited higher-frequency frication noise than the standard /ɕ/, while its formant transitions into the following vowel remained largely stable (Czaplicki, Zygis, Pape, & Jesus, 2016). Consequently, the fronted /ɕ/ in Polish has been characterized as a shift from /ɕ/ to [sʲ], enhancing its social iconicity by redistributing spectral energy into higher-frequency regions while preserving palatalization cues. The fronted /ɕ/ in Mandarin shows a similar acoustic pattern, aligning it with cross-linguistic cases of socially meaningful fronting.

As observed in the M1 results, onset F2 values of /ɕ/ are further modulated jointly by phonetic and social factors. The significant main effect of VOWEL (p = .0038) shows lower onset F2 values of /ɕ/ in /u/ (M = 1,679 Hz) than in /a/ (M = 1,935 Hz), consistent with the well-known coarticulatory effects of the back rounded vowel. A significant main effect of GENDER was also observed (p < .0001), whereby female speakers (M = 1,935 Hz) produced higher onset F2 values than male speakers (M = 1,626 Hz), which may reflect more fronted tongue gestures associated with /ɕ/-fronting among female speakers, in line with the M1 results.

Beyond the main effects, a significant GENDER:VOWEL interaction (p = .0170) revealed that vowel context modulated the gender difference in onset F2, with gender difference being smaller in

inferences in this section are based on onset F2 values, the primary acoustic dimension of interest, as reported in Table 3.

the /u/ context (Δ = 218 Hz) than in the /a/ context (Δ = 309 Hz). This pattern aligns with the M1 results that the degree of /ɕ/-fronting in female speech was substantially attenuated in the presence of /u/. Importantly, the VOWEL:SIBILANT$_{alv}$ interaction (p = .0317) also reached statistical significance, indicating that /s/ (Δ = 511 Hz) underwent greater onset F2 lowering than /ɕ/ (Δ = 256 Hz) in the /u/ context. This result demonstrates their differential susceptibility to vowel-induced coarticulatory effects and supports the assumption that /ɕ/ exhibits stronger coarticulatory resistance to vowel contexts than /s/. Consequently, the onset F2 difference between /s/ and /ɕ/ was larger in /u/ (Δ = 568 Hz) than in /a/ (Δ = 314 Hz), due to the substantial further lowering of /s/. A similar effect emerged for /ʂ/ (VOWEL:SIBILANT$_{retro}$: p = .0059): its onset F2 decreased by 592 Hz from /a/ to /u/, resulting in a wider /ɕ/-/ʂ/ separation in /u/ (Δ = 618 Hz) than in /a/ (Δ = 282 Hz). This enhanced spectral separation may reflect increased reliance on onset F2 for maintaining the /s/-/ɕ/ and /ʂ/-/ɕ/ contrasts in /u/. As shown in Figure 4 and 5, /ɕ/ tokens are more clearly separated from /s/ and /ʂ/ in F2 space, especially at onset, in the /u/ context (bottom left) compared to the /a/ context (top left). Notably, this pattern is more pronounced in the female panels (left) than in the male panels (right), consistent with the significant GENDER:VOWEL:SIBILANT$_{alv}$ (p = .0490) and GENDER:VOWEL:SIBILANT$_{retro}$ (p = .0061) interactions. These results further suggest that female speakers may exhibit a greater increase in reliance on onset F2 cue in the /u/ context than male speakers.

Taken together, these findings underscore the pivotal roles of vowel context and gender in shaping the realization of onset F2 in the sibilant contrasts. The /u/ vowel, in particular, exerts strong coarticulatory pressure that lowers onset F2 values across sibilants, with /s/ showing greater susceptibility than /ɕ/. Consequently, onset F2 separation between the two categories is enhanced in /u/ relative to /a/. This vowel-dependent pattern is further modulated by gender, raising the possibility that female speakers may exhibit a greater increase in the weighting of onset F2 than male speakers in the back vowel context, thereby compensating for the weakening of the M1-based distinction. The upcoming analysis, therefore, aims to confirm this possibility by examining whether F2 cue weights differ across vowel contexts and between female and male speakers, particularly whether the effect of vowel context differs by gender.

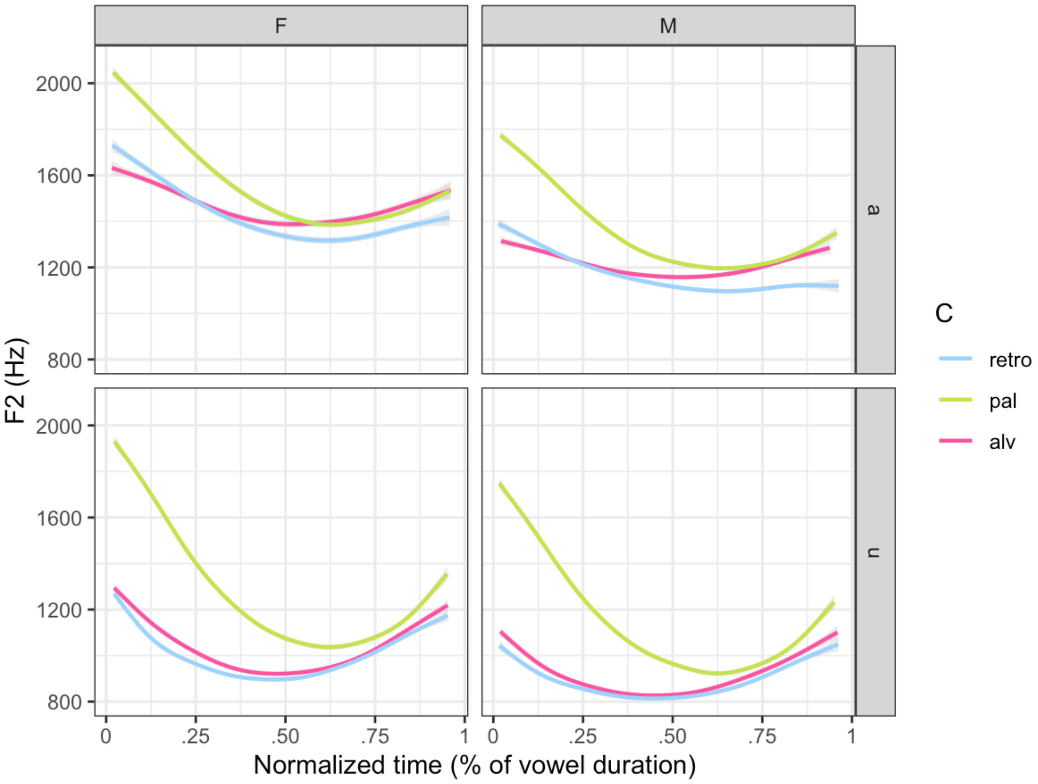

Figure 4. (Color online) GAM-smoothed F2 trajectories (Hz) over normalized vowel duration for vowels following alveolars (pink), retroflexes (blue), and alveopalatals (green), based on the empirical data. Light-gray shaded bands indicate standard errors from the GAM smooths. The results are divided by female (left) and male (right) speakers and by /a/ (top) and /u/ (bottom) vowels.

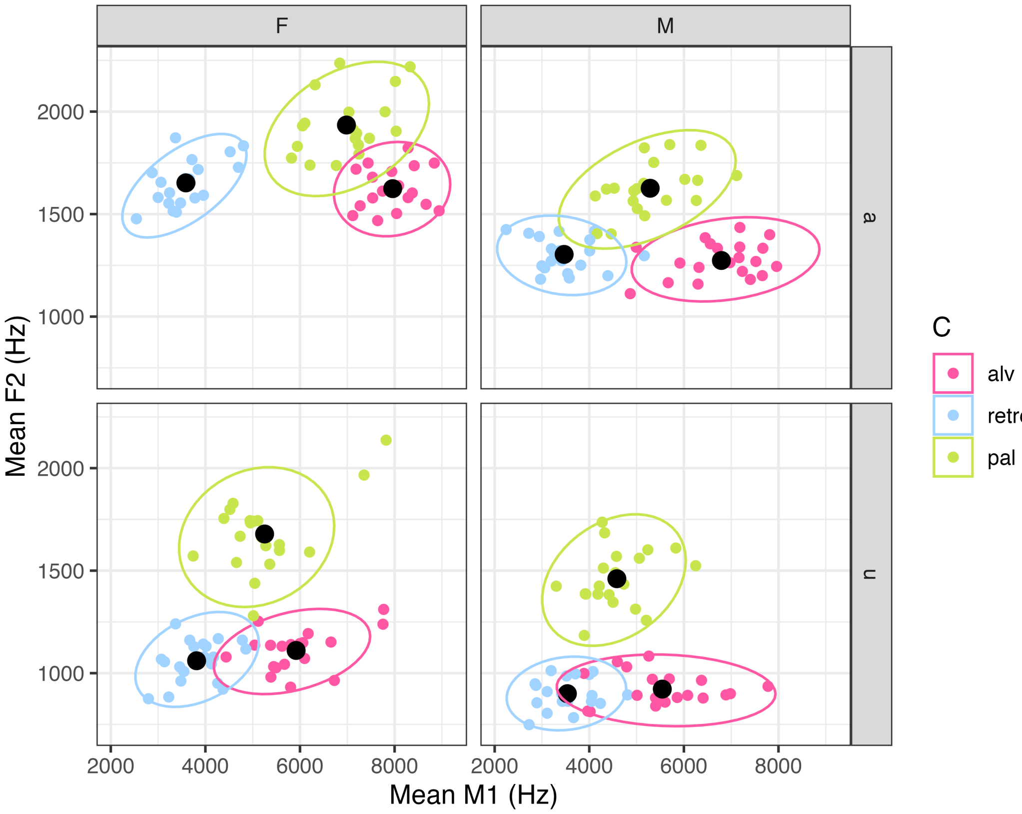

Figure 5. (Color online) By-subject distributions of empirical M1 values (Hz) and F2 values at vowel onset (Hz) of alveolars (pink), retroflexes (blue), and alveopalatals (green). Individual productions are shown as colored dots, with ellipses indicating the 95% confidence interval of a multivariate t-distribution. Black dots represent the across-speaker means for each sibilant category. The results are divided by female (left) and male (right) speakers and by /a/ (top) and /u/ (bottom) vowels.

Table 3. Summary of the mixed-effects regression Model. Formula: onset F2 ~ GENDER*SIBILANT*VOWEL + (1 + SIBILANT*VOWEL|SUBJECT) + (1|WORD). Significant results are in bold.

Predictors	Coefficient	*SE*	*t* value	*p* value
Intercept	**1934.78**	**62.58**	**30.92**	**<0.0001**
$\mathbf{SIBILANT_{alv}}$	**-313.60**	**80.21**	**-3.91**	**0.0006**
$\mathbf{SIBILANT_{retro}}$	**-282.48**	**79.79**	**-3.54**	**0.0015**
$\mathbf{VOWEL_{/u/}}$	**-256.01**	**81.48**	**-3.14**	**0.0038**
$\mathbf{GENDER_{M}}$	**-308.79**	**43.05**	**-7.17**	**<0.0001**
$\mathbf{VOWEL_{/u/}:SIBILANT_{alv}}$	**-254.39**	**112.09**	**-2.27**	**0.0317**
$\mathbf{VOWEL_{/u/}:SIBILANT_{retro}}$	**-335.78**	**111.90**	**-3.00**	**0.0059**
$\mathbf{GENDER_{M}:VOWEL_{/u/}}$	**90.66**	**36.32**	**2.50**	**0.0170**
$\mathrm{GENDER_{M}:SIBILANT_{alv}}$	-41.27	30.15	-1.37	0.1791
$\mathrm{GENDER_{M}:SIBILANT_{retro}}$	-40.11	27.85	-1.44	0.1580
$\mathbf{GENDER_{M}:VOWEL_{/u/}:SIBILANT_{alv}}$	**70.84**	**34.83**	**2.03**	**0.0490**
$\mathbf{GENDER_{M}:VOWEL_{/u/}:SIBILANT_{retro}}$	**97.63**	**33.60**	**2.91**	**0.0061**

Reference levels were set as: SIBILANT = alveopalatal, VOWEL = /a/, GENDER = female.

Figure 6 summarizes the by-subject cue weights of M1 and onset F2 for the /s/-/ɕ/ contrast, operationalized as speaker-level coefficients from the LDA analyses. The results are paneled by GENDER, while VOWEL is combined within each gender panel to facilitate comparison of cue weight distributions across vowel contexts. Each data point represents one speaker's cue-weighting strategy in a given vowel context; red and blue denote the /a/ and /u/ contexts, respectively. Solid lines show the fitted linear relationships between M1 and onset F2 weights for each vowel context within each gender; shaded areas indicate 95% confidence intervals. As shown in the figure, data points for the /u/ context cluster in the upper-left quadrant (high F2, low M1 weight), while those for the /a/ context are concentrated in the lower-right (high M1, low F2 weight). This pattern is particularly clear among female speakers, with a more pronounced separation between the /a/ and /u/ contexts, suggesting differences in cue weighting across vowel contexts. For example, Speaker F18 placed greater weight on M1 in the /a/ context (M1 weight = 3.74; onset F2 weight = 1.34), but shifted markedly toward onset F2 in the /u/ context (M1 weight = 0.29; onset F2 weight = 3.05). This context-sensitive shift, which mirrors the enhanced onset F2 separation between /s/ and /ɕ/ in /u/, shows that vowel environment influences how speakers distribute weight between M1 and onset F2 to ensure the robust /s/-/ɕ/ contrast. In addition, the relationship between M1 and onset F2 weights also appears to differ by gender: a negative association is visible among male speakers in both vowel contexts, whereas no comparable trend is observed among female speakers. In particular, in the /u/ context, female speakers maintain relatively high F2 weights regardless of their M1 weight, highlighting the importance of F2 for maintaining the /s/-/ɕ/ contrast when M1 separation is reduced.

To quantify the effects of vowel context, gender, and M1 cue weight on F2 cue weights, a linear mixed-effects regression analysis was conducted on the speaker-level LDA-derived cue weights for the /s/-/ɕ/ contrast in R. The dependent variable was onset F2 weight. Fixed effects included VOWEL (two levels: /a/ vs. /u/), GENDER (two levels: Female vs. Male), and M1 weight, as well as all interactions among these predictors. The model adopted R's default treatment coding scheme, with /a/ vowel and female serving as the reference levels. Random-effects structures including random slopes were initially considered; however, given the limited number of observations per speaker, such models were not identifiable. Accordingly, a simplified linear model with only random intercepts for speaker was adopted. Table 4 summarizes the results of the regression model.

The regression analysis revealed a significant effect of VOWEL on onset F2 weight (p = .0007), consistent with the enhanced spectral separation in onset F2 in the /u/ context. This finding suggests that female speakers increased their reliance on onset F2 in the /u/ context relative to /a/ context. A significant main effect of GENDER (p = .0302) further showed greater weighting of the F2 cue in female speakers than in males. Importantly, a significant GENDER:VOWEL interaction (p = .0171) indicates that the increase in F2 weight from /a/ to /u/ was larger for female speakers than for male speakers, paralleling the significant GENDER:VOWEL:SIBILANT interactions observed in the onset F2 analysis. Therefore, the observed increase in F2 weight for female speakers cannot be attributable solely to coarticulatory effects of the /u/ vowel, but also reflect a tendency for female speakers to compensate for the reduced spectral distinctiveness of the /s/-/ɕ/ contrast induced by /ɕ/-fronting.

Notably, although the interaction between M1 weight and GENDER did not reach statistical significant (p = .0529), the fitted regression lines in Figure 6 revealed different tendencies across gender groups. Male speakers show negative relationships between the two cue weights for (/a/: β = -0.32, p = .0363; /u/: β = -0.31, p = .0770; see Appendix 4), whereas little evidence of such a relationship was observed for female speakers in either vowel context (/a/: β = 0.07, p = .6026; /u/: β = 0.01, p = .9647). This pattern suggests a possible gender asymmetry in the relationship between the two cue weights: male speakers show a tendency toward a speaker-level trade-off between M1 and F2, comparable to the pattern reported by Hauser (2023); by contrast, in female speakers, F2 cue weighting appears to be more directly shaped by the contextual and social pressures associated with /ɕ/-fronting. This pattern is consistent with the possibility that socially motivated variation influences the organization of the female sibilant system, potentially resulting in a cue-weighting pattern distinct from that observed in male speakers.

Overall, these results point to systematic differences in cue weighting for the /s/-/ɕ/ contrast. In the back rounded vowel context, speakers increase their reliance on onset F2 relative to the /a/ vowel, offsetting the reduced effectiveness of the M1 cue and thereby maintaining robust phonological contrast. Crucially, this pattern is more pronounced among female speakers than males, indicating that, beyond a general phonetic tendency, the enhanced weighting of onset F2 also reflects an adjustment through which female speakers implement socially driven /ɕ/-fronting under contextual constraints.

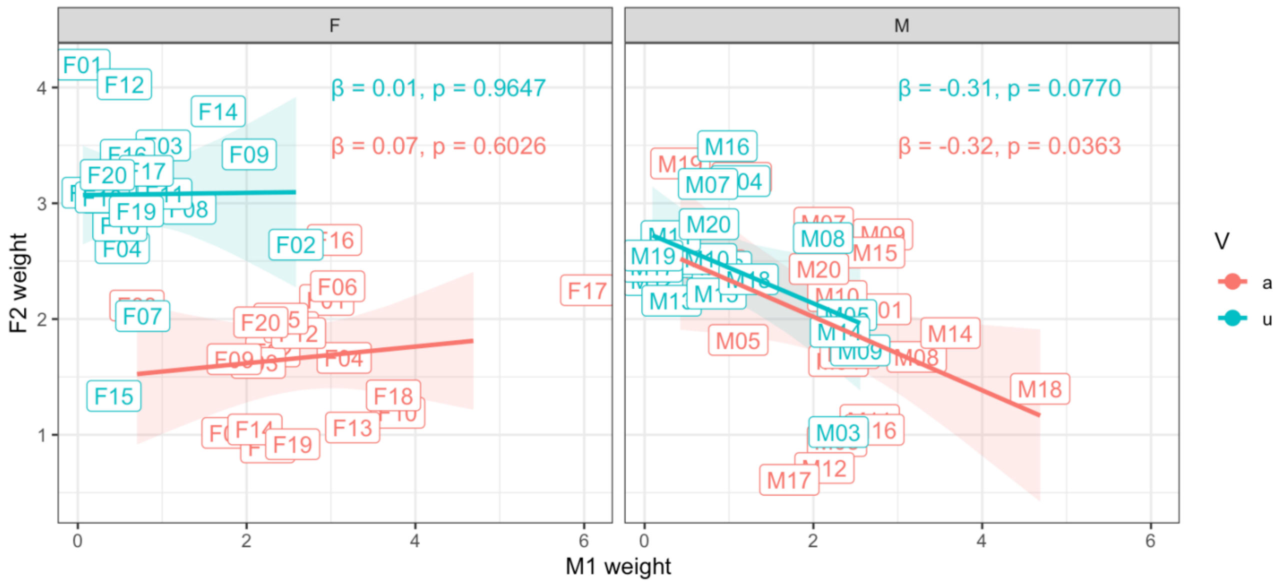

Figure 6. (Color online) By-subject weights of M1 and onset F2 for distinguishing /s/-/ɕ/ across gender, derived from LDA analyses. Each speaker is represented by two points corresponding to the /a/ (red) and /u/ (blue) vowel contexts, and labeled by gender (F/M) and speaker ID. Higher values indicate stronger reliance on the corresponding cue in categorizing the sibilants. Solid lines show fitted regression lines from the linear mixed-effects model (Table 4) for each vowel context within each gender; shaded areas indicate 95% confidence intervals.

Table 4. Summary of the mixed-effects regression model. Formula: F2 weight ~ VOWEL * GENDER * M1 weight + (1|SUBJECT). Significant results are in bold.

Predictors	Coefficient	*SE*	*t* value	*p* value
Intercept	**1.47**	**0.39**	**3.77**	**0.0003**
$\text{VOWEL}_{/u/}$	**1.60**	**0.45**	**3.54**	**0.0007**
GENDER_{M}	**1.18**	**0.53**	**2.21**	**0.0302**
M1 weight	0.07	0.13	0.54	0.5936
GENDER_{M}:$\text{VOWEL}_{/u/}$	**-1.50**	**0.61**	**-2.44**	**0.0171**
GENDER_{M}:M1 weight	-0.39	0.20	-1.97	0.0529
$\text{VOWEL}_{/u/}$:M1 weight	-0.06	0.26	-0.24	0.8123
GENDER_{M}:$\text{VOWEL}_{/u/}$:M1 weight	0.07	0.34	0.21	0.8342
Reference level was set as: VOWEL = /a/, GENDER = female.				

4. Discussion

4.1. Selection and implementation of socially motivated phonetic variants

This study provides a comprehensive investigation of the production of contemporary Mandarin sibilant fricatives based on a representative dataset. By examining two key acoustic dimensions—M1 and onset F2—the analysis provides empirical evidence for how socially motivated variation is implemented in the Mandarin sibilant system. Compared with males, female speakers exhibited fronted /ɕ/ variants in both /a/ and /u/ vowel contexts and fronted /s/ variants in the /a/ vowel context (Figure 3). In contrast, in the /u/ context, /s/-fronting was largely absent. Instead, the onset F2 separation between the two sibilants increased substantially (Figures 3 & 4). This increase is further supported by the LDA cue-weighting analysis, which shows that female speakers exhibited a greater increase in reliance on onset F2 in /u/ than in /a/ relative to male speakers (Figure 6).

Crucially, the results point to a gendered production pattern that is highly selective, rather than uniformly distributed across the Mandarin sibilant system. Only /s/ and /ɕ/ exhibit increased M1 values in female speakers, while /ʂ/ shows no comparable gender difference. This asymmetry suggests that, within the Mandarin sibilant system, social meaning is indexed specifically through sibilants characterized by higher-frequency noise. In particular, /s/ and /ɕ/, whose noise is concentrated in the higher-frequency region, participate in this pattern, whereas /ʂ/, whose frication noise is concentrated in the lower-frequency region, remains stable in female speech. Comparable patterns have been observed in vowel production in English, where gender-related increases in F2 are not uniformly distributed across categories, but are concentrated in front vowels, such as /i/, which shows a greater degree of F2 differentiation between female and male speakers than other vowels (e.g., Simpson, 2009). These findings are consistent with the view that gendered phonetic variation may be structured around a limited set of primary loci for expressing social meaning. The fronting of Mandarin sibilants may thus be interpreted as one such locus, reflecting cross-linguistic patterns commonly linked to feminine speech. According to Ohala's Frequency Code (1984), such fronting reflects the use of high-frequency cues associated with perceived smallness to evoke impressions such as friendliness and non-dominance. The present findings further support the view that fronted variants of /s/ and /ɕ/ function as socially meaningful resources associated with femininity.

Among the two sibilants that participate in indexing gender identity, the social meaning appears to be more strongly associated with /ɕ/. In the present study, /ɕ/ was shown to exhibit consistent fronting across vowel contexts, whereas /s/-fronting is absent in the /u/ context, suggesting a more limited role for /s/ in signaling social meaning. Within the /a/ context, where both sibilants show fronted realizations, /ɕ/ shows a greater degree of fronting than /s/. This aligns with impressionistic descriptions: while /ɕ/-fronting has been widely documented in previous studies, comparable evidence for /s/-fronting has been largely lacking. Together, these patterns suggest that /ɕ/ serves as a more robust index of social meaning than /s/ in the Mandarin sibilant system.[10]

This social salience of /ɕ/ may, at least in part, arise from its inherent phonetic properties. From a perceptual perspective, the fronting of /ɕ/ is likely more salient to listeners than the fronting of /s/,

[10] The present study was not designed to disentangle the relative contributions of sociophonetic and anatomical factors to the observed gender differences. Therefore, although the observed gender differences are consistent with socially conditioned fronting, they may also partly reflect anatomical differences, particularly for /s/, where the observed gender effect is relatively small and not significant in the /u/ vowel context. Notably, recent work shows that differences in peak frequency in English /s/ production may persist even after controlling for apparent vocal-tract length (e.g., Munson, 2026), suggesting a contribution of socially conditioned articulatory variation beyond physiological factors. This highlights the importance of accounting for anatomical differences when interpreting gender-related spectral differences. Although this issue was beyond the scope of the present study and was therefore not incorporated into our analysis, it raises an important avenue for future research.

offering a more efficient means of signaling socially meaningful speech variation. Due to the nonlinearity of auditory filters, frequency resolution is finer in the lower spectral range than in the higher range (Patterson, 1976; Glasberg & Moore, 1990). Consequently, a given spectral shift is perceptually more noticeable when it occurs in a moderate-high-frequency region such as that of /ɕ/, but becomes less distinct when it occurs in the higher-frequency range of /s/. In this sense, /ɕ/-fronting provides a perceptually efficient locus for signaling socially meaningful variation, whereas fronting /s/ would yield a comparatively weaker perceptual consequence. This difference may also help explain why public awareness within the speech community is relatively high for /ɕ/-fronting, whereas /s/-fronting has largely gone unnoticed.

Similar patterns have been noted in Polish, whose three-way sibilant system (/s, ʂ, ɕ/) is structurally comparable to Mandarin's and relies on similar acoustic cues (Ladefoged & Maddieson, 1996; Chiu, 2009). In this system as well, socially meaningful fronting predominantly involves /ɕ/, and has attracted considerable public and media attention. Czaplicki et al. (2016) show that young university-educated women produce fronted realizations of alveolopalatals with higher-frequency spectral peaks and higher center of gravity values relative to their standard counterparts. These patterns have been interpreted, in part, as indexing meanings such as smallness and affection, likely reflecting stronger peer-oriented social pressures within this age group. The recurrence of /ɕ/-targeting fronting in languages with larger sibilant inventories suggests that, when multiple high-frequency sibilant categories are available, /ɕ/ may offer a more perceptually responsive locus for expressing social variation. In systems lacking such an alveopalatal category, such as English, socially indexed fronting falls on /s/.

Beyond the linguistic-internal factors that may condition /ɕ/-fronting in Mandarin sibilant production, it is also worth considering the potential influence of their historical development through language contact. Fronted realizations of the alveolopalatal sibilant are widely attested in many Chinese varieties, such as Yue (e.g., Cantonese), Wu (e.g., Shanghainese and Suzhou), Min (e.g., Hokkien and Taiwanese), and certain Northern Mandarin subvarieties (Yuan, Li, & Zhou, 2001), and have been argued to reflect the retention of historically fronted pronunciations. For example, both 许 ('to allow') and 绪 ('beginning') are pronounced identically as [ɕy] in contemporary Beijing Mandarin, although they derive from different historical sources: the former from alveopalatal /ɕ/ and the latter from alveolar /s/. In varieties that preserve this distinction, the former remains [ɕy], whereas the latter surfaces as [sy]. Building on this, Cao (1987) proposed that contact with these dialects—resulting from population mobility and, in particular, frequent interaction between Beijing and Shanghai residents—may have exposed younger speakers to fronted alveopalatals and contributed to their imitation. This tendency may, in part, reflect the cultural prestige of these historically prosperous regions and their association with ideals of refinement and aesthetic sophistication, which underlie their perceived link to feminine elegance. In the contemporary context, such influences may be further amplified by increased population mobility and the wider circulation of regional speech through digital media. From this perspective, the fronted /ɕ/ observed in present-day Mandarin may be understood as the amplification or reactivation of a long-standing variant rather than the emergence of an entirely new form. This interpretation aligns with variationist and exemplar-based accounts of phonetic variation and change (Labov, 1994; Pierrehumbert, 2001; Foulkes & Docherty, 2006), which propose that linguistic innovations often emerge through the amplification and reorganization of fine-grained phonetic variants already present in the community's speech repertoire.

Taken together, gender-specific fronting of /ɕ/ and /s/ in Mandarin constitutes a socially motivated phonetic pattern, with /ɕ/ serving as a particularly strong locus for the expression of social meaning. The selective enhancement of /ɕ/-fronting appears to reflect the combined effects of perceptual salience and social prestige, whereby its relatively high perceptual efficiency enhances its salience, while its association with culturally prestigious regions—linked to feminine refinement—further reinforces its social meaning and facilitates its diffusion through increased language contact. Accordingly, socially conditioned variation is not uniformly distributed across the system, but is preferentially realized in phonetic domains that are both perceptually effective and socially interpretable.

4.2. Adaptation of gendered phonetic variants to contextual variation

Sibilant fronting in Mandarin is a complex pattern as contextually driven variation further shapes the realization of sibilants. By examining acoustic cues across two vowel contexts (/a/ and /u/), and conducting a subsequent cue-weighting analysis using LDA, the present study aimed to provide a comprehensive account of how socially meaningful variation is implemented across contexts. Compared to the /a/ context, the spectral energy is systematically shifted toward lower frequencies in the /u/ context (Figure 5), consistent with findings from other languages (Jongman et al., 2000; Hauser, 2023). Crucially, these /u/-induced coarticulatory effects are not uniform across sibilant categories. The M1 values of /s/ and /ɕ/ are lowered in the /u/ context relative to /a/, whereas the retroflex /ʂ/ shows little to no change. Onset F2 values are generally lower in /u/ than in /a/ across all three sibilants; however, the reduction is much smaller for /ɕ/ than for the other two categories. As a consequence, the M1 distance between /ɕ/ and /s/ is slightly reduced in the /u/ context, while the F2 distance between the two categories becomes larger in the /u/ context than in the /a/ context, pointing to a shift in the relative effectiveness of acoustic cues. Consistent with this pattern, the LDA results show that speakers increase their reliance on onset F2 in the /u/ context relative to /a/ (Figure 6).

The observed asymmetries in M1 and onset F2 reduction can be attributed to differences in inherent coarticulatory resistance across sibilant categories, arising from their degree of articulatory constraint. The more constrained an articulation, the stronger its resistance to vowel-driven coarticulation (Recasens, 1985; Farnetani & Recasens, 1993; Recasens & Espinosa, 2009; Iskarous, Fowler, & Whalen, 2010). For example, postalveolars produced with the tongue blade and dorsum are more constrained than alveolars articulated with the tip and blade, since the former require a larger contact area and involve the relatively sluggish tongue dorsum, whose fronting and raising inhibit coarticulatory activity in other tongue regions. Accordingly, postalveolars exhibit greater coarticulatory resistance than alveolars and thus are less susceptible to vowel-induced coarticulation. M1 and onset F2 are, therefore, adjusted by coarticulatory influence to different degrees depending on the sibilant category. In the case of Mandarin sibilants, /ɕ/ was shown to exhibit greater resistance to coarticulatory influence than /s/, consistent with its more constrained articulatory configuration. This asymmetry provides a principled account of the observed differences in how acoustic cues are modulated across sibilant categories, as well as the changes in M1- and F2-based separation between /s/ and /ɕ/ in the /u/ vowel context.

Under strong coarticulatory influence from vowels, sibilant fronting in female speech appears more constrained. Most notably, /s/-fronting is largely absent in the /u/ context (Recasens, 1985, 2012; Irfana & Sreedevi, 2016). The absence of concomitant /s/-fronting limits the available space along the M1 dimension, potentially placing additional pressure on the realization of socially driven /ɕ/-fronting. The findings of this study suggest that female speakers do not suppress socially meaningful variation in response to this contextual constraint. Instead, they maintain fronted /ɕ/ variants by further enhancing their reliance on another acoustic dimension, namely onset F2. This pattern is consistent with a process of compensatory adjustment of cue weighting, whereby a reduction in the effectiveness of one cue is offset by increased reliance on another. The enhancement of F2 cue in the /u/ context relative to /a/ reflects a general phonetic tendency shared across both gender groups. However, female speakers further strengthen this contextually enhanced F2 cue compared to their male counterparts. As shown in the LDA results (Figure 6), the increase in F2 weight from /a/ to /u/ is greater for female speakers than for males. This pattern suggests that, beyond phonetically driven coarticulatory effects, female speakers actively reinforce the F2 distinction in back rounded vowel environments, thereby compensating for the attenuation of M1-based contrasts induced by /ɕ/-fronting. In this way, female speakers maintain optimal phonological contrast while simultaneously expressing socially meaningful variation. The patterns observed across vowel contexts can be schematized as in Figure 7.

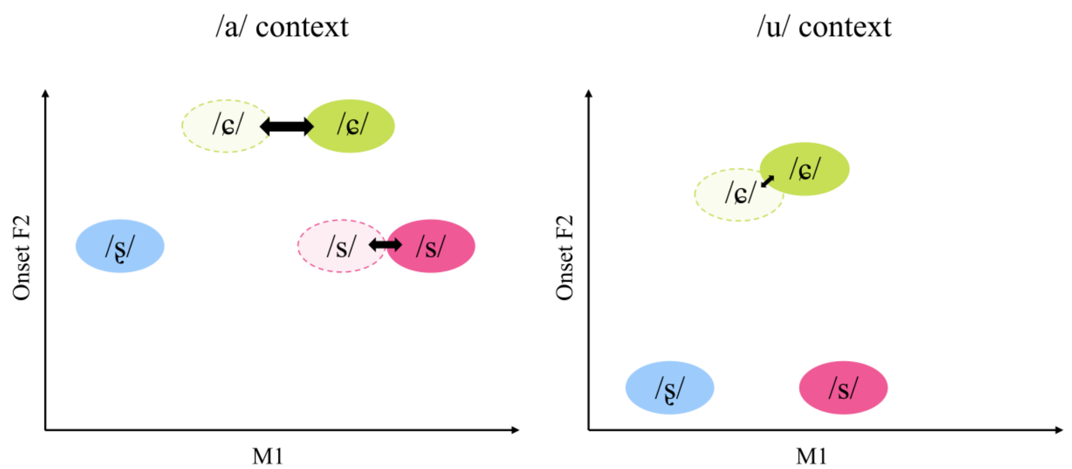

Figure 7. (Color online) Schematic illustration of context-dependent integration of sibilant fronting in Mandarin female speech. The darker-colored circles represent female realizations, whereas the lighter-colored circles with dotted outlines represent the corresponding male realizations. The double-headed arrows connecting the two indicate the magnitude of the

difference between male and female realizations for each category. In the /a/ context (left panel), both /ɕ/ and /s/ exhibit more fronted realizations in female speech, preserving spectral separation along the M1 dimension. In the /u/ context (right panel), where overall spectral energy is concentrated at lower frequencies than in the /a/ context, /ɕ/-fronting is reduced and /s/-fronting is largely absent, while the /s/-/ɕ/ contrast is sustained by the enhanced F2 cue.

Anticipating our ongoing work on how focus modulates the realization of place contrasts among the three sibilants, the preliminary results show that onset F2 for /ɕ/ is selectively enhanced under focus to a greater extent than for the other sibilants, whereas focus-induced increases in M1 appear to be relatively uniform across categories. This pattern further reinforces the role of onset F2 in distinguishing /ɕ/ from the other sibilants (Chen, Lee-Kim, & Cho, 2026). This convergence of evidence indicates that onset F2 serves as a feature-defining cue of /ɕ/ and can be actively exploited by female speakers for maintaining phonological contrast, allowing socially conditioned /ɕ/-fronting to be accommodated within the sibilant system.

Taken together, the findings show that, beyond a socially motivated pattern, Mandarin sibilant fronting is accommodated within the sibilant system through vowel-specific adjustments in cue efficiency. These adjustments reorganize the relative contribution of available acoustic dimensions in ways that are shaped by the acoustic properties of each vowel context, thereby maintaining a stable separation between the categories across contexts. More broadly, the results demonstrate that socially indexed phonetic detail can be integrated into phonological organization without compromising contrastive stability, reflecting the system's capacity to reconcile social meaning with communicative demands.

5. Conclusion

This study provides new empirical evidence for gendered sibilant production in Mandarin, and sheds light on the realization of socially motivated variation within a complex phonological system. Female speakers adopt fronted /ɕ/ variants in both /a/ and /u/ contexts, as well as fronted /s/ in /a/, potentially to align with socially meaningful speech styles. Among these, /ɕ/ appears to serve as a more prominent locus of variation, as evidenced by its more consistent and robust fronting, which may be related to its greater perceptual salience. In response to the reorganization of the sibilant acoustic space, female speakers employ flexible, context-specific adjustments, including increased reliance on a robust F2 cue in the /u/ context, to preserve phonological distinctiveness between sibilants.

Taken together, these findings offer a detailed account of how socially meaningful phonetic variation is realized alongside contextually conditioned variation within the Mandarin sibilant system. By examining a typologically richer system with multiple coronal sibilant contrasts that are sustained by multidimensional acoustic cues, the present study expands our understanding of sibilant systems beyond what has been reported for languages with simpler sibilant inventory. More broadly, the findings demonstrate that socially conditioned variation neither operates independently of the phonological system nor necessarily conflict with phonological structure. Rather, it may interact with the intrinsic phonetic organization, giving rise to adaptive cue adjustments that balance social expression with structural stability.

Appendix 1. Summary of the mixed-effects regression Model. Formula: *M1* ~ GENDER*SIBILANT*VOWEL + (1 + SIBILANT*VOWEL|SPEAK) + (1|WORD). Significant results are in bold.

Predictors	Coefficient	*SE*	*t* value	*p* value
Intercept	**7956.88**	**182.01**	**43.72**	**<0.0001**
SIBILANT$_{pal}$	**-976.19**	**191.82**	**-5.09**	**<0.0001**
SIBILANT$_{retro}$	**-4369.82**	**236.59**	**-18.47**	**<0.0001**
Vowel$_{/u/}$	**-2038.45**	**203.60**	**-10.01**	**<0.0001**
Gender$_{M}$	**-1186.10**	**232.02**	**-5.11**	**<0.0001**
Gender$_{M}$:Vowel$_{/u/}$	**811.18**	**240.89**	**3.37**	**0.0018**
Vowel$_{/u/}$:SIBILANT$_{pal}$	310.00	217.97	1.42	0.1608
Vowel$_{/u/}$:SIBILANT$_{retro}$	**2265.63**	**257.19**	**8.81**	**<0.0001**
Gender$_{M}$:SIBILANT$_{pal}$	**-506.70**	**220.71**	**-2.30**	**0.0273**
Gender$_{M}$:SIBILANT$_{retro}$	**1067.81**	**295.09**	**3.62**	**0.0009**
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{pal}$	212.59	212.69	1.00	0.3239
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{retro}$	**-969.87**	**287.26**	**-3.38**	**0.0017**

Reference levels were set as: SIBILANT = alveolar, VOWEL = /a/, GENDER = female.

Appendix 2. Summary of the mixed-effects regression Model. Formula: *M1* ~ GENDER*SIBILANT*VOWEL + (1 + SIBILANT*VOWEL|SPEAK) + (1|WORD). Significant results are in bold.

Predictors	Coefficient	*SE*	*t* value	*p* value
Intercept	**3587.05**	**159.35**	**22.51**	**<0.0001**
SIBILANT$_{pal}$	**3393.64**	**224.26**	**15.13**	**<0.0001**
SIBILANT$_{alv}$	**4369.82**	**236.59**	**18.47**	**<0.0001**
Vowel$_{/u/}$	227.17	146.86	1.55	0.1284
Gender$_{M}$	-118.30	195.79	-0.60	0.5493
Gender$_{M}$:Vowel$_{/u/}$	-158.69	135.03	-1.18	0.2472
Vowel$_{/u/}$:SIBILANT$_{pal}$	**-1955.63**	**224.81**	**-8.70**	**<0.0001**
Vowel$_{/u/}$:SIBILANT$_{alv}$	**-2265.63**	**257.19**	**-8.81**	**<0.0001**
Gender$_{M}$:SIBILANT$_{pal}$	**-1574.51**	**275.09**	**-5.72**	**<0.0001**
Gender$_{M}$:SIBILANT$_{alv}$	**-1067.81**	**295.09**	**-3.62**	**0.0009**
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{pal}$	**1182.46**	**226.43**	**5.22**	**<0.0001**
Gender$_{M}$:Vowel$_{/u/}$:SIBILANT$_{alv}$	**969.87**	**287.26**	**3.38**	**0.0017**

Reference levels were set as: SIBILANT = retroflex, VOWEL = /a/, GENDER = female.

Appendix 3. Post-hoc comparisons of M1 (Hz) between female and male speakers (F - M) in the /u/ vowel context, reported separately for each sibilant category. Significant results are highlighted in bold.

Sibilant	Contrast	Estimate	*SE*	*z* ratio	*p* value
alveolar	F - M	374.92	294.97	1.27	0.2037
retroflex	F - M	276.98	175.21	1.58	0.1139
alveopalatal	**F - M**	**669.03**	**262.79**	**2.55**	**0.0109**

Appendix 4. Simple-slope estimates of M1 weight predicting F2 weight for the /s/-/ɕ/ contrast, reported separately for female and male speakers in each vowel context. Significant results are highlighted in bold.

Gender	Vowel	Estimate (slope)	*SE*	*t* value	*p* value
Female	/a/	0.07	0.14	0.52	0.6026
Female	/u/	0.01	0.22	0.04	0.9647
Male	**/a/**	**-0.32**	**0.15**	**-2.13**	**0.0363**
Male	/u/	-0.31	0.17	-1.79	0.0770

Acknowledgement

We are grateful to the three anonymous reviewers for their constructive feedback, which greatly improved the manuscript. We thank Professor Manchun Dai for facilitating data collection in Beijing, and Yu He, Ziqian Du, and Dr. Richard Hatcher for their help with data processing. We also thank members of HIPCS for their valuable comments and discussions. The core findings of this study were presented at the Korean Society of Speech Sciences, the International Conference on Experimental Linguistics, and the 20th Conference on Laboratory Phonology. Finally, we remember with deep gratitude the late Dr. Christine Shadle, whose teaching and generous mentorship inspired the second author's interest in the spectral analysis of sibilants.

Funding

This work was supported by the research funds of Hanyang University (HY-202400000001642) and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2026S1A5A2A01003501).

Author Statement

Zhuo Chen: conceptualization, methodology, stimuli selection, data collection, labeling of the recorded data, formal analysis, writing of the original draft, and review and editing.

Sang-Im Lee-Kim: conceptualization, methodology, stimuli selection, data curation, formal analysis, funding acquisition, writing of the original draft, and review and editing.

Reference

Barreda, S., & Assmann, P. F. (2021). Perception of gender in children's voices. *The Journal of the Acoustical Society of America*, *150*(5), 3949–3963. http://doi.org/10.1121/10.0006785.

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

Blacklock, O. (2004). *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods.* University of Southampton. PhD dissertation.

Blacklock, O., & Shadle, C. (2003). Spectral moments and alternative methods of characterizing fricatives. *Journal of The Acoustical Society of America*, *113*(4), 2199–2199. http://doi.org/10.1121/1.4780184.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer*. In (Version 6.4) https://www.praat.org/.

Bouavichith, D. (2017). Cross-contextual consistency of /s/ length and spectral quality in gay men's speech. *The Journal of the Acoustical Society of America*, *142*(4), 2583–2583. http://doi.org/10.1121/1.5014451.

Calder, J., & King, S. (2022). Whose gendered voices matter?: Race and gender in the articulation of /s/ in Bakersfield, California. *Journal of Sociolinguistics*, *26*(5), 604–623. https://doi.org/10.1111/josl.12584.

Cao, Y. (1986). 北京话语音里的性别差异[Gender difference in the phonetics of Beijing Mandarin]. *汉语学习[Chinese Studies]*, *6*, 31－31.

Cao, Y. (1987). 北京话 tɕ 组声母的前化现象[The fronting phenomenon of the /tɕ/ class sounds in Beijing Mandarin]. *语言教学与研究[Language teaching and research]*, *3*, 84－91.

Chang, Y.-h. S., & Shih, C. (2015). Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin. *Journal of Phonetics*, *50*, 52–66. https://doi.org/10.1016/j.wocn.2015.02.001.

Chen, Z., Lee-Kim, S.-I., & Cho, T. (2026). Prosodic strengthening in Mandarin sibilants: Cue-selective contrast enhancement. Proceedings of HISPhonCog 2026, 113-114.

Chiu, C. (2009). Acoustic and auditory comparisons of Polish and Taiwanese Mandarin sibilants. *Canadian Acoustics*, *37*(3), 142–143.

Chiu, C., Wei, P.-C., Noguchi, M., & Yamane, N. (2020). Sibilant fricative merging in Taiwan Mandarin: An investigation of tongue postures using ultrasound imaging. *Language and Speech*, *63*(4), 877–897. http://doi.org/10.1177/0023830919896386.

Chodroff, E., & Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*, *98*(2), 250–289.

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, *39*(3), 388–402. https://doi.org/10.1016/j.wocn.2011.02.004.

Czaplicki, B., Zygis, M., Pape, D., & Jesus, L. (2016). Acoustic evidence of new sibilants in the pronunciation of young Polish women. *Poznan Studies in Contemporary Linguistics*, *52*(1), 1–42. http://doi.org/10.1515/psicl-2016-0004.

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

Fant, G. (1960). *Acoustic Theory Of Speech Production*. Mouton.

Fant, G. (1966). A note on vocal tract size factors and non-uniform f-pattern scalings. *STL-QPSR*, *4*, 22–30.

Farnetani, E., & Recasens, D. (1993). Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech*, *36 (Pt 2-3)*, 279–302. http://doi.org/10.1177/002383099303600309.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. http://doi.org/10.1121/1.427148.

Flipsen, P., Jr., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic characteristics of /s/ in adolescents. *J Speech Lang Hear Res*, *42*(3), 663–677. http://doi.org/10.1044/jslhr.4203.663.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *J Acoust Soc Am*, *84*(1), 115–123. http://doi.org/10.1121/1.396977.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409-438. https://doi.org/10.1016/j.wocn.2005.08.002.

Foulkes, P., Scobbie, J. M., & Watt, D. (2010). Sociophonetics. In *The Handbook of Phonetic Sciences* (pp. 703–754). https://doi.org/10.1002/9781444317251.ch19

Fox, R. A., & Nissen, S. L. (2005). Sex-related acoustic changes in voiceless English fricatives. *J Speech Lang Hear Res*, *48*(4), 753–765. http://doi.org/10.1044/1092-4388(2005/052).

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In (pp. 281-302). 10.1515/9783110226584.281

Fung, P., Schertz, J., & Johnson, E. K. (2021). The development of gendered speech in children: Insights from adult L1 and L2 perceptions. *JASA Express Letters*, *1*(1), 014407. http://doi.org/10.1121/10.0003322.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1), 103-138. https://doi.org/10.1016/0378-5955(90)90170-T.

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Top Cogn Sci*, *8*(2), 425–434. http://doi.org/10.1111/tops.12197.

Hamann, S. R. (2003). *The phonetics and phonology of retroflexes.* University of Utrecht. PhD dissertation.

Hauser, I. (2020). Coarticulation with alveopalatal sibilants in Mandarin and Polish: Phonetics or phonology? *Proceedings of the Annual Meetings on Phonology*, *8*. http://doi.org/10.3765/amp.v8i0.4688.

Hauser, I. (2023). Differential cue weighting in Mandarin sibilant production. *Language and Speech*, *66*(4), 1056–1090. http://doi.org/10.1177/00238309231152495.

Holliday, J. J., & Kong, E. J. (2025). The Role of Gender in the Social Evaluation of Korean Short Tongue Pronunciation. *Language and Speech*. http://doi.org/10.1177/00238309251388973.

Hu, M. (1988). 北京话"女国音"调查(1987) [A survey of the "female Mandarin" in Beijing (1987)]. *语文建设 [Language Planning]*, *1*, 28 – 33.

Iskarous, K., Fowler, C. A., & Whalen, D. H. (2010). Locus equations are an acoustic expression of articulator synergy. *The Journal of the Acoustical Society of America*, *128*(4), 2021–2032. http://doi.org/10.1121/1.3479538.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3 Pt 1), 1252–1263. http://doi.org/10.1121/1.1288413.

Kallay, J., & Holliday, J. (2012). Using spectral measures to differentiate Mandarin and Korean sibilant fricatives. Interspeech 2012, 118–121. http://doi.org/10.21437/Interspeech.2012-40.

Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: results from adolescents. *J Speech Lang Hear Res*, *56*(4), 1175–1189. http://doi.org/10.1044/1092-4388(2012/12-0038).

Kong, E. J., & Kang, J. (2020). Gendered spectral patterns of sibilant fricatives in Seoul Korean. *Studies in Phonetics, Phonology, and Morphology*, *26*(1), 1–19.

Kong, E. J., & Kang, J. (2021). Age and gender differences in the spectral characteristics of Korean sibilants. *Phonetics and Speech Sciences*, *13*(1), 37–44. http://doi.org/10.13064/KSSS.2021.13.1.037.

Kong, E. J., Kang, S., & Seo, M. (2014). Gender difference in the affricate productions of young Seoul Korean speakers. *The Journal of the Acoustical Society of America*, *136*(4), 329–335. http://doi.org/10.1121/1.4896420.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, *71*, 194–214. http://doi.org/10.1016/j.wocn.2018.09.002.

Labov, W. (1994). *Principles of Linguistic Change: Internal Factors* (Vol. 1). Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Wiley.

Ladefoged, P., & Wu, Z. L. (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *Journal of Phonetics*, *12*, 267–278.

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112. http://doi.org/10.1017/S0025100303001208.

Lee-Kim, S.-I. (2011). Spectral analysis of Mandarin Chinese sibilant fricatives. XVIIth ICPhS, 1178–1181.

Lee-Kim, S.-I. (2014). Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association*, *44*(3), 261–282. http://doi.org/10.1017/S0025100314000267.

Lee-Kim, S.-I., & Chou, Y.-C. (2022). Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Laboratory Phonology*, *13*(1). http://doi.org/10.16995/labphon.6446.

Lee-Kim, S.-I., & Chou, Y.-C. (2024). Unmerging the sibilant merger via phonetic imitation: Phonetic, phonological, and social factors. *Journal of Phonetics*, *103*, 101298. https://doi.org/10.1016/j.wocn.2024.101298.

Lee-Kim, S.-I., & Tung, H.-Y. (2025). Linguistic experience and social factors in speech perception: the case of merged speakers of Mandarin sibilants. *Laboratory Phonology*, *16*(1). http://doi.org/10.16995/labphon.15384.

Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. In (Version 1.8.5) R Foundation for Statistical Computing. https://CRAN.R-project.org/package=emmeans.

Li, F. (2008). *The Phonetic Development of Voiceless Sibilant Fricatives in English, Japanese and Mandarin Chinese.* The Ohio State University. PhD dissertation.

Li, F. (2017). The development of gender-specific patterns in the production of voiceless sibilant fricatives in Mandarin Chinese. *Linguistics*, *55*(5), 1021–1044. http://doi.org/10.1515/ling-2017-0019.

Li, F., Rendall, D., Vasey, P. L., Kinsman, M., Ward-Sutherland, A., & Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal of Phonetics*, *57*, 59–70. https://doi.org/10.1016/j.wocn.2016.05.004.

Liao, X. (2023). Performing 'husband and wife' relationships among Chinese gay couples: variations in pitch properties and sibilants in Mandarin. *University of Pennsylvania Working Papers in Linguistics*, *29*(1), 128–138.

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

Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatr Logop 1 February 1998*, *50*(1), 35-48. https://doi.org/10.1159/000021447.

Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198–212. https://doi.org/10.1016/j.wocn.2011.10.002.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text–speech alignment using Kaldi. Interspeech 2017, 498–502.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech*, *50*(Part 1), 125–142. http://doi.org/10.1177/00238309070500010601.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (L). *The Journal of the Acoustical Society of America*, *130*(5), 2631–2634. http://doi.org/10.1121/1.3641410.

Munson, B. (2026). *Getting the measures right: Sexuality and fricative variation revisited* [Poster presentation]. 20th Conference on Laboratory Phonology (LabPhon 20), Montréal, Canada.

Munson, B., & Babel, M. (2019). The phonetics of sex and gender. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (pp. 499–525).

Munson, B., Jefferson, S. V., & McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *J Acoust Soc Am*, *119*(4), 2427–2437. http://doi.org/10.1121/1.2173521.

Munson, B., Lackas, N., & Koeppe, K. (2022). Individual differences in the development of gendered speech in preschool children: Evidence from a longitudinal study. *Journal of Speech, Language, and Hearing Research*, *65*(4), 1311–1330. http://doi.org/10.1044/2021_JSLHR-21-00465.

Munson, B., Wruck, C., Benway, N. R., & Preston, J. L. (2024). Gender and age biases in the assessment of speech accuracy: A study of speech-language clinicians' ratings of /s/ accuracy. *International Journal of Language & Communication Disorders*, *59*(6), 2878–2895. https://doi.org/10.1111/1460-6984.13122.

Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, *34*(2), 139–152. https://doi.org/10.1016/j.wocn.2005.03.001.

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, *40*(1), 1–18. http://doi.org/10.1159/000261678.

Ohala, J. J. (1984). An ethological perspective on common cross-Language utilization of F_0 of voice. *Phonetica*, *41*(1), 1–16. http://doi.org/10.1159/000261706.

Ohala, J. J. (1996). Ethological theory and the expression of emotion in the voice. 4th International Conference on Spoken Language Processing (ICSLP 1996), 1812–1815.

Ohala, J. J. (1997). Sound Symbolism. Proceedings of the 4th Seoul International Conference on Linguistics (SICOL), 98–103.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, *59*(3), 640-654. https://doi.org/10.1121/1.380914.

Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, *66*, 161–184. https://doi.org/10.1016/j.wocn.2017.10.001.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988–2998. http://doi.org/10.1121/1.1370525.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. http://doi.org/10.1121/1.1906875.

Pierrehumbert, J. (2001). Exemplar Dynamics: Word Frequency, Lenition and Contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–157). John Benjamins. http://doi.org/10.1075/tsl.45.08pie

Proctor, M., Lu, L. H., Zhu, Y., Goldstein, L., & Narayanan, S. (2012). Articulation of mandarin sibilants: A multi-plane realtime MRI study. *Proc. Speech Sci. Tech.*, 113–116.

R Development Core Team. (2024). *R: A language and environment for statistical computing*. In (Version 4.3.3) R Foundation for Statistical Computing. https://www.R-project.org/.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, *28 (Pt 2)*, 97–114. http://doi.org/10.1177/002383098502800201.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America*, *125*(4), 2288–2298. http://doi.org/10.1121/1.3089222.

Reidy, P. F. (2013). *An introduction to random processes for the spectral analysis of speech data*. Ohio State University Working Papers in Linguistics, 60, 67-116.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183-204. https://doi.org/10.1016/j.wocn.2015.07.003.

Shadle, C., Chen, W.-R., & Whalen, D. (2016). Stability of the main resonance frequency of fricatives despite changes in the first spectral moment. *Journal of The Acoustical Society of America*, *140*, 3219–3220. http://doi.org/10.1121/1.4970153.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), 95–101. http://doi.org/10.1121/1.4736711.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x.

Stevens, K. N. (1999). *Acoustic Phonetics*. The MIT Press. https://doi.org/10.7551/mitpress/1072.001.0001.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 65–86). Mouton de Gruyter.

Stuart-Smith, J. (2020). Changing perspectives on /s/ and gender over time in Glasgow. *Linguistics Vanguard*, *6*(s1). http://doi.org/10.1515/lingvan-2018-0064.

Stuart-Smith, J., & Alam, F. (2018). *Sibilants and ethnicity in Glaswasian girls* [Poster presentation]. British Association for Academic Phoneticians (BAAP) Colloquium, University of Kent. http://doi.org/10.13140/RG.2.2.36032.15367.

Stuart-Smith, J., Timmins, C., Pryce, G., & Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language*, *89*(3), 501–536. http://doi.org/10.1353/lan.2013.0041.

Stuart-Smith, J., Timmins, C., & Wrench, A. (2003). Sex and gender differences in Glaswegian /s/. Proceedings of the 15th International Congress of Phonetic Sciences, 1851–1854.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S (4th ed.)*. Springer.

Wong, E., Koeppe, K., Cychosz, M., & Munson, B. (2025). Gendered speech development in early childhood: Evidence from a longitudinal study of vowel and consonant acoustics. *Journal of Child Language*, 1–28. http://doi.org/10.1017/S030500092500011X.

Yuan, J., Li, R., & Zhou, Z. (2001). *汉语方言概要[An Outline of Chinese Dialects]*. Beijing Language and Culture University Press[北京语言文化大学出版社].

Zhang, G. (2006). 漢語方言滋絲音的一些觀察[Some observations on sibilants in Chinese dialects]. *中國語文研究 [Studies in Chinese Linguistics]*, *21*(1), 87–102.

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in society*, *34*(3), 431–466.

Zhu, X. (2004). 亲密与高调——对小称调、女国音、美眉等语言现象的生物学解释[Intimacy and high pitch: A biological explanation for diminutive tone, female Mandarin, and meimei]. *当代语言学[Contemporary Linguistics]*, *6*(3), 193－222.

Zimman, L. (2017). Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics*, *55*(5), 993–1019. http://doi.org/10.1515/ling-2017-0018.

Zygis, M., Pape, D., & Jesus, L. (2012). (Non-)retroflex Slavic affricates and their motivation: Evidence from Czech and Polish. *Journal of the International Phonetic Association*, *42*(3), 281–329. http://doi.org/10.1017/S0025100312000205.