Differential cue weighting in Mandarin sibilant production

Ivy Hauser
University of Texas at Arlington

Abstract

Standardized varieties of Mandarin traditionally exhibit a three-way place contrast between retroflex, alveopalatal, and alveolar sibilants. Phonetic realization of the sibilants varies across speakers, including a merger between the retroflex and alveolar in some varieties, such as Taiwan Mandarin. This study examines data from Mandarin speakers with origins in mainland China and finds individual variation in use of spectral center of gravity (COG) and the second formant of the following vowel (F2). Production cue weights of COG and F2 are inversely correlated across speakers, demonstrating a trade-off relationship in cue use. Production cue weight of COG is also positively correlated across contrasts, such that the speakers that exhibit lower alveolar-retroflex COG contrast due to merger simultaneously lose COG contrast with the alveopalatal. This suggests that the merger does not only involve overlap of the alveolar-retroflex categories, but rather restructures the acoustic space for all sibilants.

Keywords: Mandarin, sibilants, merger, cue weighting, individual differences

Introduction

Standardized Mandarin is often described as exhibiting a three-way place contrast among alveolar, retroflex, and alveopalatal sibilant fricatives (Duanmu, 2007) while the alveolar and the retroflex sibilant are often merged in Taiwan Mandarin (e.g. Chang & Shih, 2015; Kubler, 1985). The merger is sometimes described as also occurring in other southern/eastern varieties, though most of the work on the merger has focused exclusively on Taiwan Mandarin (cf. Zhu, 2012, on Shanghai Mandarin). This paper examines acoustics of Mandarin sibilants to investigate differential cue weighting patterns in individual speakers' productions. The data include acoustic analyses of sibilants produced by Mandarin speakers with origins in mainland China who are currently living in the United States.

This material is partially based upon work supported by the National Science Foundation under Grants No. 1451512 and 823869. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. I would also like to acknowledge the contributions of several research assistants: Zachary Sun, Allison Chen, Borui Zhang, and Emily Graham.

Despite none of the speakers being from regions typically associated with the merger, interspeaker variation in degree of merger and cues to sibilant contrast are observed. A trade-off relationship between use of spectral center of gravity (COG) and second formant of the following vowel (F2) emerges within sibilant contrasts—speakers that use COG more also use F2 less (and vice versa). The data also indicate that the /s ʂ/ merger does not simply involve overlap of those categories, but rather restructures the acoustic space for sibilants, shifting remaining contrasts away from use of COG and towards use of F2.

Mandarin sibilants

Though [s] and [ʂ] are typically used in phonetic transcription of Mandarin, the sibilants have been described as having a variety of different places of articulation including alveoar, dental, and denti-alveolar for [s], and retroflex, laminal post-alveolar, and apical post-alveolar for [ʂ]. Chang and Shih (2015) provides a review of claims about exact places of articulation. In this paper, I use the terms alveolar and retroflex (following e.g. Chang & Shih, 2015; Duanmu, 2007; Ladefoged & Wu, 1984). There is less controversy about the place of articulation of alveopalatal [ɕ] (though sometimes the place is also termed "alveolopalatal").

There is an allophonic restriction on sibilants requiring [ɕ] before high front vowels (e.g. Duanmu, 2007; Lin, 2014), but there is disagreement in the literature as to how [ɕ] patterns outside of the high front vowel contexts. The alveopalatal sibilant is represented as 'x' in Pinyin script, a romanized quasi-phonemic orthographic system, and the literature diverges regarding how orthographic 'xia' and 'xiu' sequences are represented at phonemic and surface levels. Some analyses posit that these are pronounced as represented one-to-one in the orthography, [ɕia] [ɕiu] (e.g. Lee & Zee, 2003). Under this assumption, [ɕ] only appears before high front vowels, and is in complementary distribution with the other sibilants and the velar fricative. Other analyses posit that 'xia' and 'xiu' orthographic sequences are pronounced as [ɕa] and [ɕu/ɕəu] respectively (e.g. Duanmu, 2007; Lee-Kim, 2011; F. Li, 2008; W.-C. Li, 1999; Lin, 2014). Under this assumption, there is a surface contrast between all three sibilants before [a] and [əu/u]. Analysts who assume these surface pronunciations have taken different approaches to phonological analysis. Duanmu (2007) analyzes [ɕ] as being derived from underlying /si/ sequences, such that /sia/ → [ɕa], while others analyze all three sibilants as belonging to separate underlying phonemes (Lee-Kim, 2011; W.-C. Li, 1999).

In this paper, I follow the assumption of a three-way place contrast among sibilants in the standardized Beijing variety in /a/ and /u/ vowel contexts, and treat 'xia' and 'xiu' orthographic sequences as [ɕa] and [ɕu]. This assumption and the phonemic analysis of the Mandarin sibilant system, however, is not crucial for any of the results presented here. The key results of individual variability and the relationships between cue weights hold regardless of the phonological representation of alveopalatal sequences.

Regarding acoustics, Beijing Mandarin production (and other varieties with the three-way place contrast), has been described as exhibiting a three-way distinction in spectral center of gravity (COG; also sometimes referred to as "spectral mean"), with [ʂ] having the lowest COG and [s] having the highest COG (Kallay & Holliday, 2012; Lee, 1999; Lee-Kim, 2011). COG has been shown to be influenced by coarticulation with following vowels, such that COG is lower when followed by a rounded vowel (Jeng, 2006; Y. Li, 2009). Though

Hu (2008) found that individual differences, not vowel context effects, were the major source of variability in both articulation and acoustics.

While some speakers exhibit a three-way COG contrast, other studies have reported a two-way center of gravity contrast between the alveolar sibilant and the other two sibilants with contrast in the second formant at the onset of the following vowel distinguishing the alveopalatal from the retroflex (e.g. Stevens, Li, Lee, & Keyser, 2004). These results demonstrate variation in phonetic realization of the sibilant contrasts even among speakers who distinguish all three sibilants. Specifically, these speakers can differ in the relative weight of COG in distinguishing the alveopalatal /ɕ/ from retroflex /ʂ/. Some speakers collapse the COG distinction between /ɕ ʂ/ and primarily rely on F2, while others maintain COG distinction between all three sibilants including /ɕ ʂ/. Both patterns are observed in the data presented here.

There is typically less of a distinction between /s ʂ/ in Taiwan Mandarin in both acoustics (e.g. Chen, 1999; Chung, 2006; Jeng, 2006) and articulation (Chiu, Wei, Noguchi, & Yamane, 2020). It has been proposed that this is due to contact with Southern Min, which lacks retroflex consonants (Chuang & Fon, 2010; Kubler, 1985). Although the merger is typically associated with Taiwan Mandarin, /s ʂ/ overlap has been observed in Beijing Mandarin and other northern varieties as well (e.g. Chung, 2006). For example, Chang (2011) analyzes corpus data and finds that Taiwan speakers have higher /ʂ/ COG relative to Beijing speakers, but both groups display overlapping /s ʂ/ categories to at least some degree.

The merger is typically described as a loss of retroflexion or a substitution of retroflex sibilants with the alveolar sibilant (e.g. Chang & Shih, 2015; Chung, 2006). Within Taiwan Mandarin speakers, there is interspeaker variability in phonetic realization of the merger. Chiu et al. (2020) documents varying degrees of merger, from complete merger with total category overlap to no merger with no category overlap. The use of retroflexion also has socio-indexical value; it is associated with higher education levels and distinguishes standardized Mandarin pronunciation from "dialect-accented" Mandarin (Chang, Shih, & Allen, 2013). Hypercorrection of retroflexion, using [ʂ] for underlying /s/, has also been documented in Taiwan speakers who exhibit the merger (Chung, 2006).

Vowel context, sociolinguistic factors, formality of task, and prosodic context have all been shown to enhance the alveolar-retroflex contrast (Chang & Shih, 2012, 2015; Chuang & Fon, 2010; Chung, 2006; Jeng, 2006; Y. Li, 2009). Though vowel context effects are well-documented, existing literature provides conflicting results on which contexts facilitate contrast enhancement versus merger. Chung (2006) describes retroflexion (and therefore /s ʂ/ distinction) as being more common before back rounded vowels /o u/, and offers an articulatory explanation that the back articulation facilitates anticipatory tongue retraction. Similarly, Chiu et al. (2020) finds that some speakers have a pattern of contextually dependent merger only in the /a/ context, producing more distinction before back rounded /o/. By contrast, Y. Li (2009), Jeng (2006), and Chang and Shih (2012) all find that the sibilants are less distinct in the back rounded /u/ context. It is possible that these conflicting results were obtained because of differences in type of experimental task. The present study includes two vowel contexts: back rounded /u/ and low central /a/ all for an additional test case of vowel context effects.

In perception, multiple studies have found the primary perceptual cue for the

alveolar-retroflex contrast to be COG or the position of the lowest spectral prominence (F. Li, 2008; Wu & Lin, 1989), with dialectal variation in discrimination boundary according to the status of the /s ʂ/ merger. Chang (2013) compares perception between Taiwan and Beijing Mandarin listeners and finds that Beijing listeners categorize ambiguous stimuli as /s/ at lower COG frequencies relative to Taiwan listeners. In other words, Taiwan listeners permit more /s/-like tokens to be categorized as /ʂ/, mirroring the pattern of merger in production. While COG is the primary cue for the /s ʂ/ contrast, it is not sufficient to distinguish the alveopalatal from the other two sibilants, regardless of the status of the alveolar-retroflex merger. Instead, the primary cue distinguishing /ɕ/ from /s ʂ/ is the second formant frequency at the onset of the following vowel (F. Li, 2008). This mirrors production results which show that some speakers do not produce a COG distinction between /ɕ ʂ/ (Stevens et al., 2004).

In sum, previous work on Mandarin sibilants demonstrates variation in how the sibilant contrasts are realized phonetically among individual speakers and regional dialects. While F2 is consistently the primary perceptual cue for contrasts involving /ɕ/, there is variability in degree of COG distinction produced for these contrasts. There is also a merger between the alveolar and retroflex sibilants in some varieties, which can vary across speakers and contexts. The /s ʂ/ merger is well-documented in Taiwan Mandarin, though there is less work on /s ʂ/ contrast in other varieties. Given these findings, we expect to see individual differences in sibilant realization, even among individuals not from regions typically associated with the merger. The results here present data from Mandarin speakers with origins in mainland China and demonstrate variability in degree of /s ʂ/ merger and differential use of COG versus F2 to distinguish sibilant contrasts.

Cue weighting in production and individual differences

There is a growing body of work revealing systematicity in individual talker differences. For example, several studies have documented covariation of a single phonetic value across multiple segments (e.g. Chodroff & Wilson, 2017; Tanner, Sonderegger, & Stuart-Smith, 2020) and covariation of phonetic values across multiple cues (e.g. Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Clayards, 2018; Tanner et al., 2020). Less is known about related individual differences in "contrast signaling" (Clayards, 2018), or systematicity of cue weighting across individuals. Examining weights of multiple cues can shed light on the nature of individual differences, in particular, whether they are attributable to speech style and degree of clear speech or cue trading and contrast maintenance. In this paper, I investigate the relationship between degree of contrast across multiple cues in Mandarin sibilant production. I take "degree of contrast" to mean the relative strength of a particular phonetic dimension in predicting phonological category, or production cue weight. In the data here, cue weights differ across speakers such that not all speakers use the same dimension as the primary cue to the sibilant contrasts.

Cue weight in production is typically measured using a classification algorithm (e.g., discriminant analysis, logistic regression) which assigns relative weights to each predictor dimension (see Schertz & Clare, 2020, for a review). Differences in cue weighting patterns for the same contrast have been observed in production between native speakers of the same language (Shultz, Francis, & Llanos, 2012), native and non-native speakers (Schertz, Cho, Lotto, & Warner, 2015), non-native speakers with different levels of L2 exposure (Kong

& Yoon, 2013), and speakers of a language undergoing sound change (Bang et al., 2018; Coetzee, Beddor, Shedden, Styler, & Wissing, 2018; Kuang & Cui, 2018). However, previous work on the relationship between primary and secondary cue weights in production provides conflicting results with limited case studies.

There are three possible relationships between two cues to a contrast: no correlation, positive correlation, and inverse correlation (Schertz & Clare, 2020). These relationships can be described at token and talker levels (following discussion in Clayards, 2018). Cue relationships at the token level compare phonetic values of two cues across tokens. Cue relationship at the talker level are the focus of this paper, and relevant correlations are between individual talker cue weights for COG and F2.

A positive correlation indicates that speakers who use one cue more distinctively also use the other cue more distinctively. At the token level, this may be the case if the two cues are produced by the same articulatory mechanism (instrinsic linkage as in Wang & Fillmore, 1961). At the talker level, this could be the case if speech style modulates the relationship between primary and secondary cue weights. Speakers who adopt a clearer speech style may be expected to enhance contrast on all dimensions. This results in a positive correlation between cue weights across speakers and has been observed in production of English stops. Clayards (2018) found a positive correlation between voice onset time (VOT) and fundamental frequency (F0) cue weights across speakers; those who produced VOT more distinctively also produced F0 more distinctively. However, this was not statistically significant. For the present study, if individual differences are rooted in speech style, we expect a positive correlation between COG and F2 cue weight across speakers—speakers who use COG more distinctively should also use F2 more distinctively.

The relationship between cues could also be modulated by cue trading (e.g. Repp, 1982), which would result in an inverse correlation between cues. Trading refers to the pattern where speakers exaggerate the value of one cue when producing an ambiguous value of the other cue. The term "cue trading" is traditionally used to describe this pattern at the token level. In a trading relationship at the talker level, an inverse correlation between cue weights indicates that speakers who produce one cue more distinctively produce the other cue less distinctively. To distinguish between cue trading relationships at the token and talker levels, I will refer to the talker-level trading relationship between cue weights as a "trade-off" relationship rather than "cue trading". The talker-level trade-off relationship between cue weights is examined in this paper.

A trade-off relationship between cue weights has been found in previous work on stop production. For example, Bang et al. (2018) examines F0 and VOT cue weighting for Korean stops and finds that speakers that have higher production cue weight for F0 have lower cue weight for VOT. They argue that this indicates the emergence of F0 as the primary cue in a quasi-tonogenetic sound change. The same relationship has also been found in English stops (Clayards, 2008; Shultz et al., 2012), where it is thought to be related to individual speaker differences rather than community-level sound change. A trade-off relationship between primary and secondary cue weights across speakers indicates a pressure for contrast maintenance—the contrast between categories is maintained by all speakers using different relative contributions of cues. For the present study, if individual differences are rooted in cue trade-off, we expect an inverse correlation between COG and F2 across speakers such that speakers who use COG more distinctively use F2 less distinctively.

The relationship between cue weights across speakers bears on the nature of individual differences, specifically whether differences in contrast signaling are modulated by speech style or contrast maintenance. Ultimately, previous work is unclear on the relationship between primary and secondary cue weights across speakers and has been limited to studies of VOT and F0 of stops in few languages. More case studies are needed to understand the nature of individual differences in production cue weighting. The present study provides a new test case using sibilant fricatives in Mandarin.

Experimental design

Participants

All speakers were between the ages of 18-24 and recruited from the student population at the University of Massachusetts Amherst through linguistics courses and email advertisements to the "Taiwanese and Chinese Students' Association". All recruitment materials (emails, sign-up info, etc.) were distributed in Mandarin orthography. Participants were compensated with course credit or $15 USD per hour for their participation. 20 speakers (13F) were recorded and two speakers were excluded because they did not complete the task.

All speakers acquired Mandarin natively in China, relocated to the United States for their undergraduate studies, and continue to use Mandarin on a daily basis. None of the speakers reported early L1 experience with any other languages. Eleven speakers reported origins in southern/eastern areas which are geographically close to Taiwan and have been associated with the /s ʂ/ merger in previous literature (Shanghai, Jiangsu, Fujian). Nine speakers reported origins in northern areas which are typically associated with /s ʂ/ distinction (Beijing).

Stimuli

The stimuli were words and rare words which were expected to behave as non-words. Because the Mandarin writing system is logosyllabic, use of non-words presents problems for participant reading. Instead of attempting to design new and orthographically natural characters, we used rare words with existing characters as "non-words". Each stimulus was presented with the simplified Mandarin orthographic character and the Pinyin romanization. With the Pinyin presented alongside the logosyllabic characters, the participants were able to pronounce the intended stimulus even if they were unfamiliar with the word or Mandarin character. No participants reported trouble reading either orthographic system when asked upon study completion. The stimuli were read in the carrier phrase "wǒ bǎ X dú yī biàn" ('I read X once').

To ensure the rare words were actually unknown to participants and could be analyzed as non-words for purposes of lexical frequency balancing, a word frequency judgment task was also constructed for use after recording. This task was a paper survey which took about 1-2 minutes to complete. The survey asked the question 'How common are each of these words? Circle your answer' and participants answered for all stimuli. Possible answers included 'common', 'moderately common', 'rare', and 'I don't know this word'. The results supported analyzing the rare words as non-words—rare words were all unknown to all participants.

Table 1

Example stimuli

C	vowel	stimulus (IPA)	stimulus (pinyin)	tone	frequency
s	a	sá	sa	4	low
s	a	sā	sa	1	high
s	a	sá	sa	4	rare
ʂ	a	ʂā	sha	1	low
ʂ	a	ʂǎ	sha	3	high
ɕ	a	ɕā	xia	1	low
ɕ	a	ɕà	xia	2	high
ɕ	a	ɕá	xia	4	rare
s	u	sú	su	4	rare
s	u	sū	su	1	high
s	u	sǔ	su	3	low
ʂ	u	ʂū	shu	1	low
ʂ	u	ʂǔ	shu	3	high
ʂ	u	ʂù	shu	2	rare
ɕ	u	ɕǔ	xiu	3	low

The stimuli were crossed according to the following factors: sibilant (3 levels: s ʂ ɕ) × vowel context (2 levels: a u) × word status (3 levels: high frequency/low frequency/non-word) × number of syllables (2 levels) × tone (4 levels). Due to limitations of the lexicon, some of the tones are not fully crossed with all other factors. However, in post-hoc analyses there were no significant effects of tone on F2 or COG values (see results section for further discussion). There were a total of 137 distinct sibilant stimuli. Additional stimuli with word-initial affricates and stops were included as fillers. Word-initial non-sibilant fricatives were not included in the task. Example stimuli are provided in Table 1.

Recording

The participants were recorded in a sound-attenuated booth using Audacity software (Audacity Team, 1999-2021). Recordings were collected using an M-Audio Fast Track Pro Mobile Audio Interface and a Shure SM10A head-worn microphone with a sampling rate of 44.1 kHz and a bit depth of 16. The participants were presented with stimuli in simplified Mandarin characters and Pinyin orthography on a laptop computer inside the booth. They were asked to produce the phrases as naturally as possible. Experimenters were trained to give feedback to encourage natural production, which included things like suggesting the participant speak as if they were talking to a friend rather than giving a presentation. The stimuli were recorded in four separate blocks, each with a different random order, totaling four repetitions of each stimulus for analysis.

Data processing and analysis

The recordings from each speaker were first scanned by the author and trained research assistants for speech errors.[1] The recordings were then force aligned using the Montreal Forced Aligner (MFA; McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017) with a pretrained Mandarin model.[2] This created Praat (Boersma, 2001) TextGrids marking the boundaries between segments. All sibilant and vowel boundaries of the MFA TextGrids were then hand edited to ensure accuracy of extracted measurements.

A Praat script based on DiCanio (2013) was used to extract spectral center of gravity (COG) of the fricatives and formant values of the following vowels. COG was computed from a time-averaged spectrum over the middle 80% of the fricative interval. The formants were estimated using the Burg method and extracted at 10 ms intervals throughout the duration of the vowel. The formant measurements were also all hand checked and manually adjusted whey they contained formant tracking errors. All results presented below use F2 values at 20 ms into the vowel, following Nowak (2006), who analyzes a similar sibilant system in Polish.

All analysis and visualization was done in R (R Core Team, 2013).

Results

This section provides results on individual differences in contrast implementation, cue weighting patterns at the group level, cue relationships within contrasts across speakers, and effects of the /s ʂ/ merger on cue weights across contrasts.

Individual differences in contrast implementation

In accordance with previous literature, we found individual differences in use of COG versus F2 to distinguish the sibilants. Most speakers in this sample exhibit three distinct sibilant categories; only two speakers appear to fully merge /s ʂ/ in both vowel contexts. However, many speakers exhibit some degree of overlap between /s ʂ/. All speakers consistently exhibit higher F2 values for the alveopalatal sibilant. There is also intraspeaker variation between vowel contexts, with most speakers exhibiting more COG contrast before /a/ than before /u/. This was true for speakers that had distinct or partially merged /s ʂ/ categories.

For example, Figure 1 provides representative data from two speakers who exhibit three non-overlapping sibilant categories in both vowel contexts. In all graphs in Figures 1-3, COG values are given on the x-axis and F2 values at 20ms into the following vowel are given on the y-axis. The /a/ context is in the left panel and the /u/ context is in the right panel. The sibilants colors are denoted in the key with their Pinyin orthographic symbols: 's' for /s/, 'sh' for /ʂ/, and 'x' for /ɕ/. These and all other graphs in this paper were created using the ggpplot2 package (Wickham, 2009) in R (R Core Team, 2013).

Speaker 16 in the top panel of Figure 1 is the speaker with the greatest COG distinction between /s ʂ/ in both vowel contexts. While the contrast between /ʂ ɕ s/ is

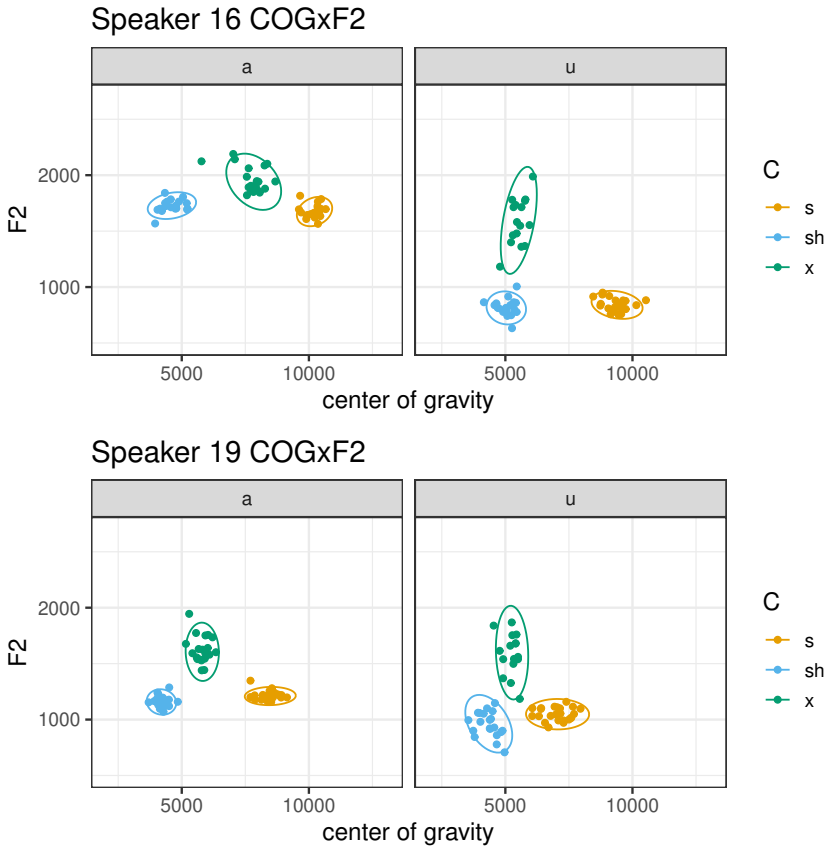
[1] Tokens were excluded if they were speech errors (wrong word produced as determined by a native speaker), or included non-speech vocalization during production (coughing, etc.). Participant 13 produced 5 such tokens, and all other participants produced 1-3.

[2] Available at https://montreal-forced-aligner.readthedocs.io/en/latest/user_guide/models/acoustic.html.

primarily a COG contrast in the /a/ context, the /ʂ ɕ/ contrast is primarily an F2 contrast in the /u/ context, such that there is no 3-way COG contrast in this context. Though not as drastic, the same pattern is demonstrated by several other speakers, including speaker 19, given in the bottom panel of Figure 1.

Figure 1

More COG contrast before /a/ for non-mergers. F2 and COG in Hz.

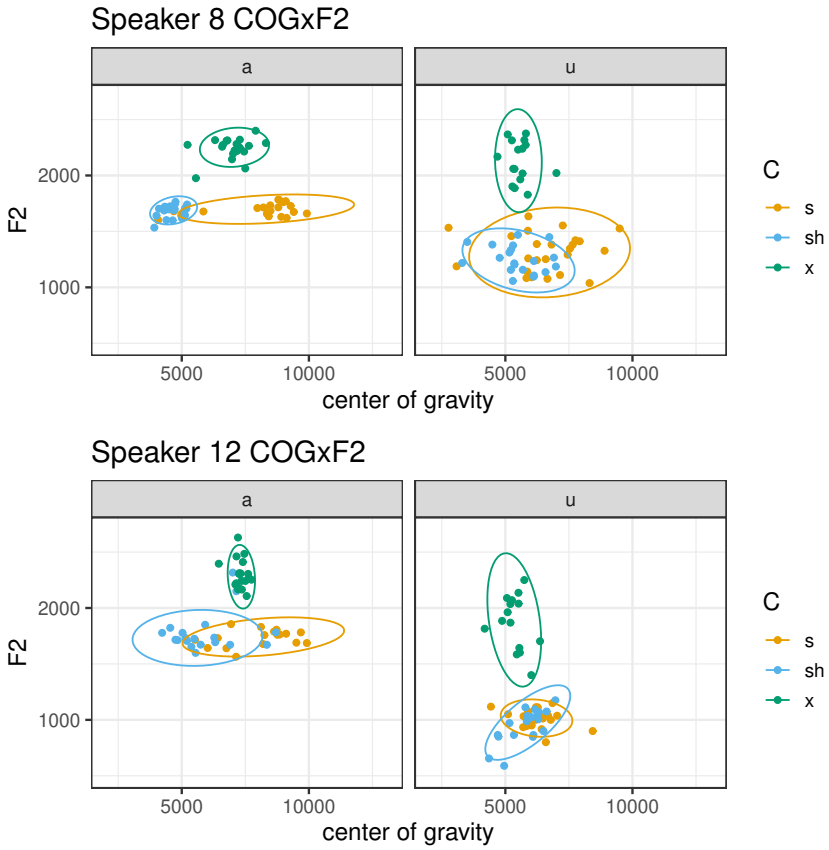


For speakers that appear to have a partial merger (some /s ʂ/ overlap in at least one of the vowel contexts), the categories are also more distinct before /a/ than before /u/. Data from two representative speakers with this pattern are given in Figure 2. There were two speakers that appear to have a complete or near-complete merger, with /s ʂ/ categories that are almost entirely overlapping in both vowel contexts. They are given in Figure 3.

In these data, interspeaker variation appears to be primarily attributable to individual differences rather than regional trends. According to previous work on the /s ʂ/ merger, we expect northern speakers (n=8) to exhibit distinction and southern speakers (n=11) to tend towards merger. In this data, the two speakers which did appear to exhibit a complete /s ʂ/ merger were both from southern provinces, as expected. However, the two speakers which displayed the largest distinction between /s ʂ/ were also from southern provinces. Thus, major region does not appear to correlate with degree of merger in these data. As the analyses in this paper deal with individual variation, the absence of discrete regional

Figure 2

More COG contrast before /a/ for partial mergers. F2 and COG in Hz.



patterns is not problematic as degree of merger is quantified with cue weight.

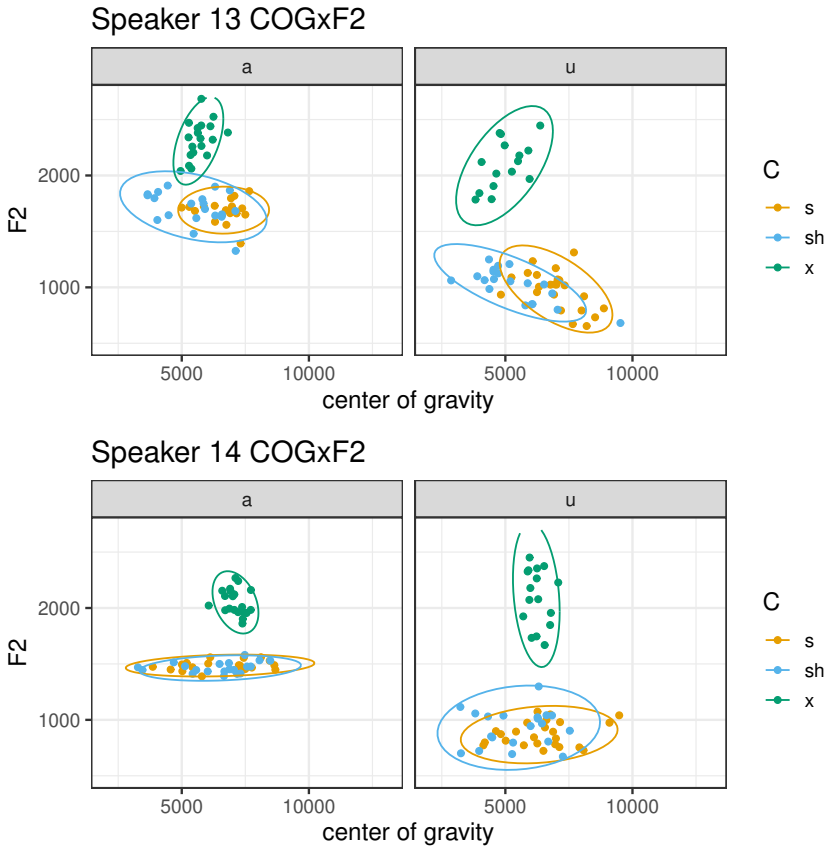
Group-level cue weighting patterns

I use Linear Discriminant Analysis to quantify cue weight in production (LDA; Duda, Hart, & Stork, 2012; Fisher, 1936; Fukunaga, 1990). LDA is a classification method that relates continuous predictor variables to category labels. The purpose of LDA is to find the linear function[3] that best discriminates a set of categories (here, the sibilant categories) given a set of predicting features (here, the acoustic measures of COG and F2). There is precedent in phonetic literature for using LDA to quantify cue weight in production (see Schertz & Clare, 2020, for a review). Following this, I use the coefficients of linear discriminants as the measure of cue weight from LDA. The coefficients are regression weights used to calculate the probability of category membership (James, Witten, Hastie, & Tibshirani, 2013). They indicate the contribution of each predictor variable to the discriminant func-

[3]LDA therefore only determines *linear* separability. Similar discrimination algorithms can be applied which utilize different functions for separability (e.g., quadratic discriminant analysis). The methodology for using these would be similar, though they are not directly examined here.

Figure 3

Minimal COG distinction for full mergers. F2 and COG in Hz.



tion; higher values indicate more contribution to the discriminant function. These weights can be interpreted as indexing the strength of each individual predictor.

An analysis of lexical properties was first conducted to determine the best space for performing the LDAs. Separate linear mixed effects models were fit for COG and F2 values using the lme4 package (Bates, Sarkar, Bates, & Matrix, 2007) in R (R Core Team, 2013). In both models, sibilant, vowel context, the sibilant × vowel interaction, experimental block, lexical frequency, and tone were predictors with random intercepts for speaker and item. ANOVAs of both models showed significant effects of sibilant, vowel context, the sibilant-vowel interaction, and the random effect of speaker. This is expected as COG and F2 values have been shown to differ across sibilants, vowel contexts, and speakers. Neither model showed significant effects for tone, block, lexical frequency, or the random effect of item on COG or F2, thus data were collapsed over these factors when performing LDAs.

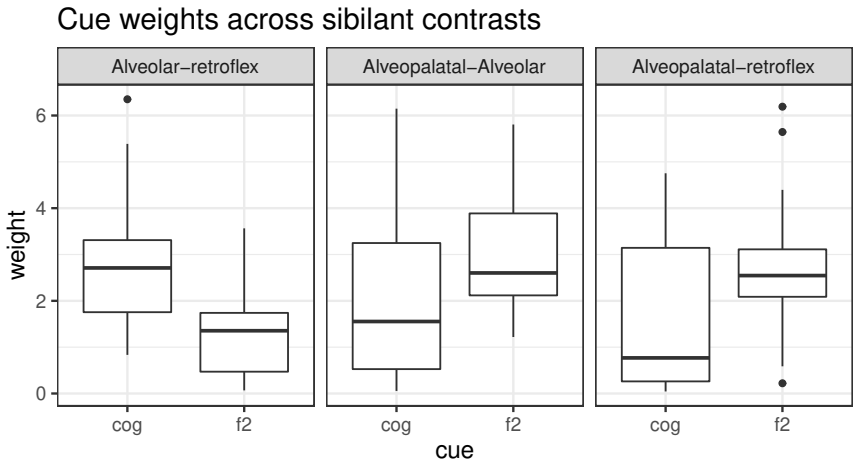
A set of LDAs was performed for each speaker within vowel contexts between each pair of sibilants using COG and F2 as the relevant predictors. This was done using the lda function from the MASS R package (Ripley et al., 2013). Prior to performing the LDAs, the acoustic measurements for COG and F2 were standardized using within-speaker z-scores so the LDA coefficients would be comparable across dimensions (as in Schertz et al., 2015).

Some coefficients are presented with reverse polarity such that larger values always indicate higher cue weight for that dimension.[4]

The COG and F2 coefficients for each contrast are plotted in Figure 4. For each pair of contrasting sounds, the two cues of COG and F2 are indicated on the x-axis, with their respective weights on the y-axis. For the contrasts involving alveopalatal /ɕ/, the average F2 cue weight is higher than the average COG cue weight. This is expected based on previous work noting F2 as the primary cue distinguishing the alveopalatal from the other sibilants for many speakers. The average F2 cue weight for these contrasts is also higher than the average F2 cue weight for the /s ʂ/ contrast, where COG cue weight is higher. This also is in line with previous work—the primary cue distinguishing /s ʂ/ is COG, but the primary cue distinguishing /s ʂ/ from /ɕ/ is F2, with interspeaker variation in relative contribution of COG to alveopalatal contrast.

Figure 4

Cue weights for each contrast collapsed over speakers and vowel contexts



Cue relationships within contrasts across individuals

The main question of interest within contrasts is whether individual differences in cue weighting are suggestive of a trade-off relationship or degree of clear/careful speech. If F2 and COG cue weights are in a trade-off relationship, we expect an inverse correlation such that speakers who use one cue more contrastively use the other less contrastively. If F2 and COG cue weights are instead related by speech style, we expect a positive correlation such that speakers who use one cue more contrastively also use the other cue more contrastively.

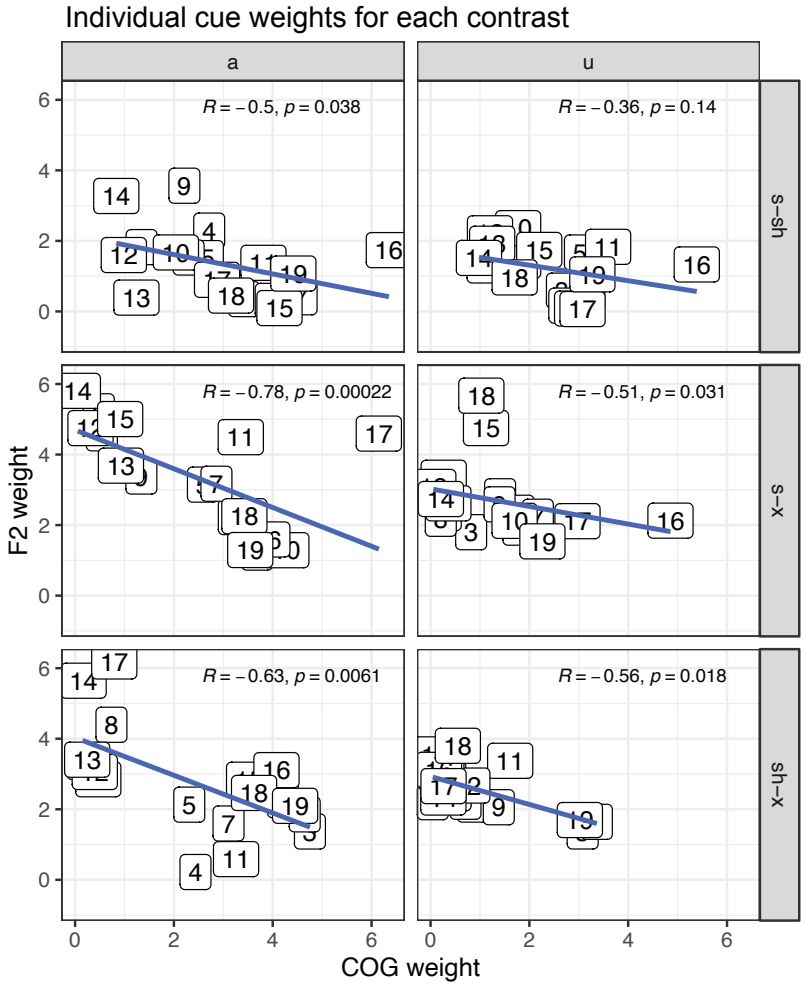
Graphs showing each individual speaker's cue weights are given in Figure 5, which is partitioned by sibilant contrast and vowel context. Sibilant contrast is given in the rows, and within each row, the /a/ vowel context is in the left panels and the /u/ vowel context is in the right panels. Within each panel, COG cue weights are on the x-axis and F2 cue

[4]The polarity of the coefficients will depend on the coding of factors, and reversing polarity for negative coefficients ensures that higher values always indicate more weight when presenting results (Schertz et al., 2015; Shultz et al., 2012).

weights are on the y-axis. Each point represents the cue weights of an individual speaker. The points are labeled with the speakers' participant numbers and can be cross-referenced with the graphs in Figures 1-3. The linear best-fit regression line is shown in each panel along with a Spearman correlation test.

Figure 5

Cue weights for each contrast across speakers



The individual weights for each contrast exist on a continuum, including the alveolar-retroflex /s ʂ/ contrast, which is involved in a merger. There is no clear delineation between speakers who exhibit the merger and speakers who contrast /s ʂ/ (at least with respect to the cue weight values here). Within each contrast and vowel context (each individual panel of Figure 5), there is a negative correlation. This means that speakers who exhibit higher cue weight for COG also exhibit lower cue weight for F2 for each contrast in each vowel contrast. To estimate the strength of the correlations, these data were submitted to a linear

mixed effects regression where the dependent variable is F2 cue weight. The predictors are vowel context, contrast, and COG cue weight with random intercepts for speaker. Results for fixed effects are given in Table 2.

Table 2

Fixed effect table for linear mixed effects regression. Dependent variable: F2 cue weight, Predictors: vowel context, contrast, COG cue weight, random intercepts for speaker. Reference level is alveolar-retroflex contrast in the /a/ context.

Effect	Estimate (se)	t	p
Vowel - /u/	−0.74 (0.19)	−3.81	0.001^{***}
Contrast - /s ɕ/	1.70 (0.50)	3.74	$< 0.001^{***}$
Contrast - /ʂ ɕ/	1.35 (0.45)	3.02	0.003^{**}
COG cue weight	−0.29 (0.13)	−2.22	0.03^{*}
Contrast - /s ɕ/ × COG cue weight	−0.11 (0.16)	−0.65	0.51
Contrast - /ʂ ɕ/ × COG cue weight	−0.17 (0.16)	−1.048	0.30

The negative estimate and significant effect of Vowel indicate that F2 cue weight is significantly lower in the /u/ vowel context relative to the /a/ context. The positive estimates and significant effects of Contrast indicate that F2 cue weight is significantly higher for the contrasts that involve /ɕ/ relative to the alveolar-retroflex contrast (the reference level). This is expected as F2 has been noted as the primary cue for those contrasts. This effect can also be observed in the group weights given in Figure 4, which shows that F2 weights are, on average, higher for contrasts involving /ɕ/. The negative estimate of COG cue weight indicates that there is an inverse relationship between COG cue weight and F2 cue weight for the intercept contrast /s ʂ/. The interactions with the other contrasts show negative estimates, indicating that the inverse relationship between COG cue weight and F2 cue weight for these contrasts is stronger relative to the intercept /s ʂ/ contrast. However, these interactions are not significant, so there is no significant difference in the relationship between COG cue weight and F2 cue weight relative to /s ʂ/. This means that COG cue weight is significantly negatively correlated with F2 cue weight for all contrasts.

Effects of /s ʂ/ merger on cue weight across contrasts

The previous section analyzes the relationships between cue weights within contrasts and results suggest a trade-off relationship between use of F2 and use of COG. In this section, I analyze how speakers use COG across contrasts to examine the effect of the /s ʂ/ merger on the other sibilant contrasts in the system.

Because of the way the merger tends to be realized, the loss of /s ʂ/ COG contrast simultaneously results in loss of COG contrast with /ɕ/. Figure 3 provides examples of speakers with a (near) complete merger of /s ʂ/. These are the speakers who have some of the lowest COG cue weights for distinguishing /s ʂ/, and their categories are almost entirely overlapping in both vowel contexts. For these and other speakers that tend towards merger in this dataset, the merger seems to be realized primarily as an increase in within-category

COG variability for /s/ and /ʂ/, and not as a shift in the phonetic values of one category towards the other. This means that any COG distinctions between /s ʂ/ and /ɕ/ also collapse, as /ɕ/ COG values are typically between those of /ʂ/ and /s/. For speakers with the merger, the merged /s ʂ/ category occupies all the COG space in an individual's system, leaving no space for COG contrast with /ɕ/.

If this pattern holds across speakers, we would expect a positive correlation between COG cue weights of different contrasts. Speakers that have low COG cue weights due to the /s ʂ/ merger should also exhibit low COG cue weights for the /s ɕ/ and /ʂ ɕ/ contrasts. Speakers that maintain distinction between /s ʂ/ (and therefore have higher COG cue weights for /s ʂ/) should also be more likely to distinguish /ɕ/ using COG, as their phonetic space permits such a distinction.

This positive trend can be observed in the graphs in Figure 6. COG weight for the /s ʂ/ contrast is given on the x-axis, with COG weight for the /s ɕ/ contrast on the y-axis in the top panel and COG weight for the /ʂ ɕ/ contrast in the bottom panel. As in Figure 5, the graphs are partitioned by vowel context and each point represents the weights of an individual speaker. Within each panel, there is a positive correlation across speakers, indicating that speakers who merge /s ʂ/ (and have lower /s ʂ/ COG cue weight values on the x-axis) also have lower COG cue weight for the /s ɕ/ and /ʂ ɕ/ contrasts.

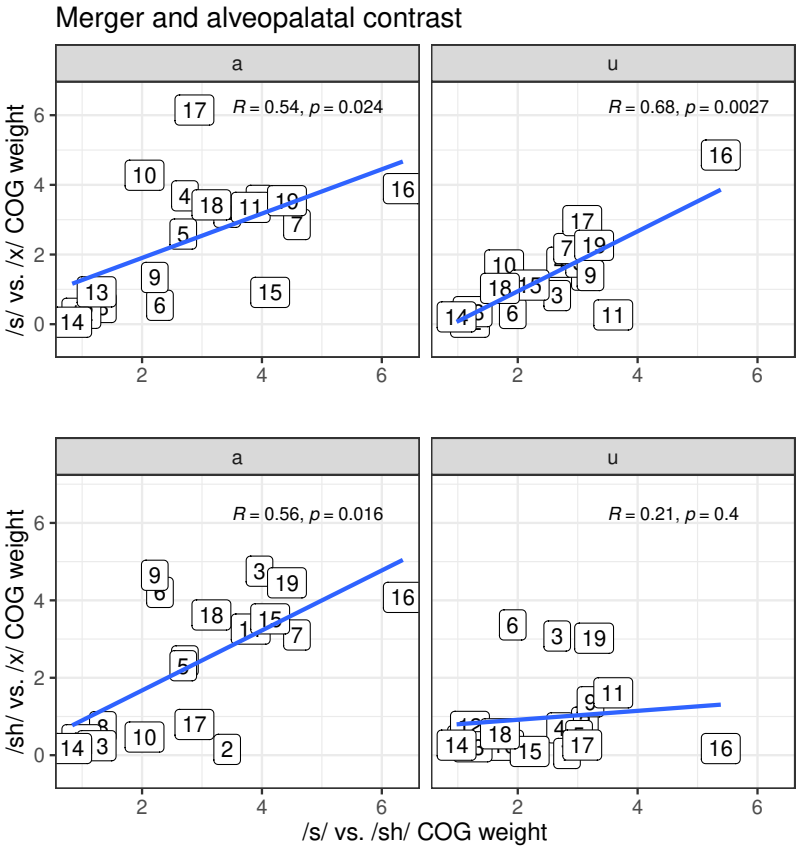
To estimate the strength of the correlations shown in Figure 6, a linear mixed effects model was performed with /s ʂ/ COG cue weight as the dependent variable. Predictors are vowel context and the interactions between vowel context and COG cue weight for the /s ɕ/ and /ʂ ɕ/ contrasts, respectively, with random intercepts for speaker. Results are provided in Table 3. The effect of vowel context compares the /u/ context to the reference level /a/ context and shows no significant difference in COG cue weight for /s ʂ/ between the two contexts. The positive estimates and significant effects of both COG weights in each vowel context indicate a significant positive relationship between COG cue weights for the /s ʂ/ contrast and COG cue weights for the other contrasts in the system (/s ɕ/ and /ʂ ɕ/). This means that speakers who merge the alveolar and retroflex sibilants also exhibit low COG cue weight for the other contrasts, and speakers who maintain COG distinction between the alveolar and the retroflex also maintain COG distinction between these categories and alveopalatal /ɕ/. These results indicate that the /s ʂ/ merger does not only involve overlap of those two categories, but represents an overall decrease in COG contrast throughout the sibilant system.

Discussion

This paper has examined the relationship between cue weights in the Mandarin sibilant system at the talker level, both between and across sibilant contrasts. Within contrasts, the results reveal an inverse relationship between COG cue weight and F2 cue weight across speakers. This is consistent with a cue trade-off account of individual differences in contrast signaling. Speakers that produce one cue less distinctively tend to produce the other cue more distinctively. Previous work on weights of multiple cues in production has yielded conflicting results on whether cue trade-off or speech style modulates the relationship between cues. The results here strengthen existing support for a trade-off relationship between cue weights (e.g. Bang et al., 2018; Shultz et al., 2012, on stops in Korean and English).

Figure 6

COG cue weights across contrasts



Though correlations between cue weights like the ones here have been observed in production, these relationships do not seem to be predictive of relative cue weights in perception. Multiple studies have examined the relationship between cue weights in production and cue weights in perception for the same individuals and have typically found no significant relationship (e.g. Kim & Clayards, 2019; Schertz et al., 2015; Shultz et al., 2012), though some have observed positive trends, potentially indicating a weak relationship between cue weight in production and cue weight in perception. Based on this, we would not expect the differences in COG and F2 cue weight observed here to necessarily be predictive of how the speakers would use COG versus F2 in a perception task. An area for further work would be to determine the relationship between production cue weight and individual differences in perception of Mandarin sibilants.

The cue trade-off relationship within contrasts also has implications for the phonetic realization of all sibilants when /s ʂ/ are merged, as a decrease in COG contrast increases reliance on F2. Between sibilant contrasts, there is a positive relationship of COG cue weights across speakers, such that speakers that have low COG cue weights for the /s ʂ/ contrast due to merger also have low COG weights for the /s ɕ/ and /ʂ ɕ/ contrasts. This is because a complete merger effectively collapses all COG contrast in the system, shifting

Table 3

Fixed effect table for linear mixed effects regression. Dependent variable: COG cue weight for the /s ʂ/ contrast, Predictors: Vowel context, COG cue weight for the /s ɕ/ contrast × vowel context, COG cue weight for the /ʂ ɕ/ contrast × vowel context, random intercepts for speaker.

Effect	Estimate (se)	t	p
Vowel–/u/	0.26 (0.45)	0.58	0.46
Vowel–/a/ × /s ɕ/ COG weight	0.36 (0.11)	3.29	0.002**
Vowel–/u/ × /s ɕ/ COG weight	0.67 (0.16)	4.13	< 0.001***
Vowel–/a/ × /ʂ ɕ/ COG weight	0.45 (0.11)	4.12	< 0.001***
Vowel–/u/ × /ʂ ɕ/ COG weight	0.37 (0.18)	2.11	0.04*

the contrasts involving /ɕ/ to rely less on COG to maintain distinction. Therefore, the /s ʂ/ merger does not only involve overlap of those categories, but rather restructures the acoustic space for sibilants.

While the /s ʂ/ merger is associated with Taiwan Mandarin and other southern varieties, these results demonstrate varying degrees of merger from speakers with origins in northern and southern mainland China. The two speakers with a near complete merger and the two speakers with the highest COG cue weight were all from southern provinces. This could reflect the large amount of interspeaker variation in these varieties (in line with Chiu et al., 2020, who found a range of /s ʂ/ realizations among Taiwan speakers, the variety most associated with the merger). The high COG cue weights from these speakers could also be the result of hyperarticulation and/or hypercorrection of the contrast (in line with Chung, 2006). It is possible that some speakers who would merge in casual contexts produced distinction in the laboratory context. Further work specifically manipulating formality of task to elicit hypo- and hyperarticulated sibilants from the same participants could be done to disentangle these explanations.

Conclusion

In this paper, I have presented the results of a production study comparing COG and F2 cue weight of sibilants across Mandarin speakers with origins in mainland China who are currently living in the United States. Despite none of the speakers being from Taiwan, multiple speakers exhibit substantial overlap of these categories in production. Interspeaker variation in degree of merger was also observed, and contrast between /s ʂ/ seems to exist on a continuum such that speakers cannot be discretely divided into groups of mergers and non-mergers. These differences also appeared to be related to individual patterns and did not significantly correlate with region of origin.

These individual differences were quantified using cue weights of the COG and F2 dimensions. The observed relationship between COG cue weight and F2 cue weight is one of trading. For all sibilant contrasts, the speakers that displayed the least distinction in COG also displayed the most distinction in F2, and vice versa. Because of the increase in

COG variability associated with the /s ʂ/ merger, speakers who merge /s ʂ/ simultaneously collapse COG contrast between /s ʂ/ and alveopalatal /ɕ/. Therefore, the merger does not only involve overlap of the retroflex and alveolar categories. It restructures the acoustic space for sibilants such that COG contrast is generally reduced in the sibilant system, as merging speakers shift towards relying fully on F2 to distinguish the remaining contrasts.

References

Audacity Team. (1999-2021). Audacity(r): Free Audio Editor and Recorder [Computer software manual]. http://audacity.sourceforge.net/.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144. doi: 10.1016/j.wocn.2017.09.005

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, *2*(1), 74.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5:9/10*, 341-345.

Chang, Y.-H. (2011). A Corpus Study of Retroflex Realizations in Beijing and Taiwan Mandarin. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 440–443).

Chang, Y.-H. (2013). *Variability in cross-dialectal production and perception of contrasting phonemes: the case of the alveolar-retroflex contrast in Beijing and Taiwan Mandarin* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Chang, Y.-H., & Shih, C. (2012). Using map tasks to investigate the effect of contrastive focus on the Mandarin alveolar-retroflex contrast. In *Speech prosody 2012.*

Chang, Y.-H., & 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.

Chang, Y.-H., Shih, C., & Allen, J. B. (2013). Dialectal variation in the perception of phonological contrasts. In *Proceedings of the International Conference on Phonetics of the Languages in China* (pp. 115–118).

Chen, P. (1999). *Modern chinese: History and sociolinguistics.* ERIC.

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.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47. doi: 10.1016/j.wocn.2017.01.001

Chuang, Y.-Y., & Fon, J. (2010). The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech. In *Proceedings of speech prosody 5.*

Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication*, *16*(2), 197–214.

Clayards, M. (2008). *The ideal listener: making optimal use of acoustic-phonetic cues for word recognition* (Unpublished doctoral dissertation). University of Rochester.

Clayards, M. (2018). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica*, *75*(1), 1–23.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. doi: 10.1016/j.wocn.2017.09.009

DiCanio, C. (2013). *Time averaging for fricatives 2.0.* Praat script published online.

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

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification.* John Wiley & Sons.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition.* Elsevier.

Hu, F. (2008). The three sibilants in Standard Chinese. In *Proceedings of the 8th international seminar on speech production* (pp. 105–108).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jeng, J.-Y. (2006). The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies*, *40*(1), 27–48.

Kallay, J., & Holliday, J. (2012). Using spectral measures to differentiate Mandarin and Korean sibilant fricatives. In *Proceedings of the annual conference of the international speech communication association.*

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, *31*, 1–18. doi: 10.1080/23273798.2019.1582787

Kong, E. J., & Yoon, I. H. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Phonetics and Speech Sciences*, *5*(4), 81–90. doi: 10.13064/ksss.2013.5.4.081

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. doi: 10.1016/j.wocn.2018.09.002

Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, *27*(2), 156–176.

Ladefoged, P., & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics*, *12*(3), 267–278.

Lee, W.-S. (1999). An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 413416).

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112.

Lee-Kim, S.-I. (2011). Spectral analysis of Mandarin Chinese sibilant fricatives. In *Proceedings of the 17th international congress of phonetic sciences* (pp. 1178–1181).

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese* (Unpublished doctoral dissertation). Ohio State University, Columbus.

Li, W.-C. (1999). *A diachronically-motivated segmental phonology of Mandarin Chinese*

(Vol. 37). Peter Lang Publishing.

Li, Y. (2009). *Effects of lexical frequency and neighborhood density on incomplete neutralization in Taiwan Mandarin* (Unpublished doctoral dissertation). CCU.

Lin, Y.-H. (2014). Segmental phonology. In *The handbook of Chinese linguistics* (pp. 400–422). John Wiley & Sons.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal forced aligner [computer program] version 0.9.0, retrieved from http://montrealcorpustools.github.io/montreal-forced-aligner/.*

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.

R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *CRAN Repository*. Retrieved from `https://cran.r-project.org/web/packages/MASS/`

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. doi: 10.1016/j.wocn.2015.07.003

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2). doi: 10.1002/wcs.1521

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), EL95–EL101. doi: 10.1121/1.4736711

Stevens, K. N., Li, Z., Lee, C.-Y., & Keyser, S. J. (2004). A note on Mandarin fricatives and enhancement. In *From traditional phonology to modern speech processing* (pp. 393–403). Foreign Language Teaching and Research Press.

Tanner, J., Sonderegger, M., & Stuart-Smith, J. (2020). Structured speaker variabiltiy in Japanese stops: Relationships within versus across cues to stop voicing. *Journal of the Acoustical Society of America*, *148*(793), 1. doi: 10.1121/10.0001734

Wang, W. S., & Fillmore, C. J. (1961). Intrinsic cues and consonant perception. *Journal of Speech and Hearing Research*, *4*(2), 130–136.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer New York.

Wu, Z., & Lin, M. (1989). *Overview of experimental phonetics*. Higher Education Press, Beijing.

Zhu, L. (2012). *Retroflex and non-retroflex merger in Shanghai accented Mandarin* (Unpublished master's thesis). University of Washington, Seattle, WA.