The role of gestural timing in non-coronal fricative mergers in Southwestern Mandarin: acoustic evidence from a dialect island

Dongmei Raoa,b, Jason A. Shawb,*

*Corresponding author: jason.shaw@yale.edu, Tel: +1 47 5441 2015

a School of Literature and Journalism, Xihua University, Chengdu, 610075 Sichuan, China
b Department of Linguistics, Yale University, New Haven, 06511 Connecticut, USA

Abstract
Merger between a voiceless labiodental fricative, [f], and a voiceless velar fricative, [x], is common across languages, including many varieties of Chinese, particularly those spoken in Southwestern China. The sound changes that lead to merger in Southwestern Mandarin varieties are bidirectional: in some, [f] becomes [x]; in others [x] becomes [f]. We conducted a study of phonetic variation in one such variety, Zhongjiang (中江) Chinese, which has been reported to merge [x] to [f] in the environment of [w] and [u]. Our results confirm this basic pattern while revealing additional nuances, including a new environment, [_oŋ], which conditions merger in the opposite direction, [x] becomes [f], and new phonetic details. In particular, [x] exhibits a particularly low spectral Center of Gravity (CoG) and [f] exhibits a wide range of spectral variation, including tokens with low CoG, characteristic of a velar constriction. We interpret these patterns in the context of areal variation, proposing a pathway to change that relates spectral variation attributable to gestural overlap to diachronic observations of labio-velar merger.

Keywords
Velar fricatives, secondary articulation, labial-velar merger, sound change, gestural timing, regional variation, pathways of change
1. Introduction

Merger between a voiceless labiodental fricative, [f], and a voiceless velar fricative, [x], henceforth labial-velar merger, is a common sound change, attested in Germanic, Romance, Celtic, Slavic, and Uralic (Hickey 1984) as well as many varieties of Southwest Chinese. In a typological survey of 374 varieties spoken in Hunan, Hubei, Sichuan, and Yunnan provinces, He (2004) reports that 212 have the [f]-[x] merger. The merger can also be found in Southern Chinese, e.g., Min (Chen & Li 1991), Cantonese (Zhan 2002), Gan (Sun 2007), Hakka in west Guangdong (Li 1999) and the vernacular of north Guangdong (Zhuang 2004).

Amongst these labial-velar mergers, there are languages in which [f] has become [x] and others in which [x] has become [f]. That is, the sound change is bidirectional. The focus of research to date on labial-velar mergers in China has been largely documentational in nature and have not incorporated facts about phonetic variation. The patterns of change, dating back to medieval Chinese, have been recorded in detail by Chinese dialectologists (see references above). In this literature, language varieties tend to be characterized categorically, e.g., words are described as being produced with either [f] or [x]. Less is known about patterns of synchronic phonetic variation and how they could relate to the observed sound changes. The aim of this paper is to establish this relation, considering patterns of synchronic phonetic variation alongside documented patterns of sound change.

To this end, we provide an examination of phonetic variation in one language of Southwest China (Zhongjiang 中江). Numerous phonetic measurements have been used to characterize variation in fricatives (e.g., Jongman, Wayland, & Wong, 2000; McMurray & Jongman, 2011). Spectral moments have been used commonly, particularly since Forrest et al. (1988), to characterize fricatives within and across languages. The mean energy of the spectrum, or Center of Gravity (CoG), is often used even in the absence of other spectral moments (e.g., Gordon, Barthmaier, & Sands, 2002). In their study of English fricatives, Shadle & Mair (1996) also included two additional spectral measures, dynamic amplitude and the slope of a line fit from the maximum frequency to 16.97 kHz. Dynamic amplitude picked up some consistent differences between English fricatives while the slope of the line captured variation in speaker effort. In a study of eight English fricatives, Jongman et al. (2000) investigated the acoustic separability of fricatives, and reported that the variance of the spectrum was a particularly robust acoustic cue to place of articulation. Similarly, Shadle & Mair (1996) found that the related measure of spectrum standard deviation to be useful in differentiating English fricatives. Similar methods have been used to analyze Mandarin Chinese fricatives. For example, Svantesson (1986)’s 2D ‘fricative space’ consists of measures of Center of Gravity (CoG) and a measure of spectral dispersion that is closely related to spectrum standard deviation. Other studies of Mandarin have followed this method, finding that all five fricatives can be clearly differentiated using these two dimensions (Ran 2008) or some normalized version of these measures (Ran & Shi 2012). Following this work, our main analysis in this paper focuses on spectrum Center of Gravity (CoG), which we also plot alongside Spectrum Standard Deviation (SD). To encourage additional analyses, the entire data set, including sound files and textgrids (see methods), has been submitted as a
Although spectral moments are rather coarse descriptions of the spectrum, for the specific case of labial-velar fricatives, the interpretation of CoG and Spectrum Standard Deviation are relatively straight-forward. The posterior constriction for velar fricatives, [x], usually ensures a low CoG, due to resonance of the long cavity in front of the constriction, and low SD, due to the relatively sharp spectral peaks (Stevens, 1998: 370-372). For [f], the anterior constriction at the lips typically results in a diffuse (flat) spectrum (Stevens, 1998: 389-398), indexed by high spectrum SD, and high CoG, due to resonance of either a very short cavity in front of the constriction or no detectable front cavity resonance at all. Phonetic variation in these measures, conditioned in part by coarticulatory influences, provides some clues to understanding the diachronic patterns across the languages of China more generally.

One important characteristic of the labial velar merger is that, like many sound changes, it tends to proceed in specific phonological environments. There are three environments in particular that favor merger. Mergers are more likely when the fricative precedes: (1) a labiovelar glide, [w]; (2) a high back rounded vowel, [u]; and (3) the VC sequence consisting of a mid-round vowel and velar nasal coda, [on]. These contexts are not random. They all involve a lip movement and tongue dorsum retraction, albeit with possibly different modes of coordination. To preview our results, we find that the range of acoustic variation in the production of [f] and [x] across these and other environments suggests that there may be a temporal basis for the labial-velar merger. Specifically, we argue that the relative timing of the constituent gestures of these fricatives and surrounding gestures contribute crucially to the observed sound changes. Tongue dorsum retraction, as required for [w], [u], [on], during the fricative [f], gives rise to spectral properties that approach [x]. Similarly, lip rounding, as also required for [w], [u], [on], during the fricative [x], brings the acoustics closer to [f]. Overlap in time between the component gestures of the fricative and other gestures in the local environment has the effect of neutralizing acoustic differences between [f] and [x].

The remainder of the paper is structured as follows. Section 2 provides background on the labial-velar merger in Southwest China. Section 3 describes the methods of four studies on the language variety of Zhongjiang. Study 1 provides a descriptive baseline, reporting measurements of [x] and [f] fricatives in minimal pair environments. Against this baseline, studies 2-4 report the same phonetic measurements in environments related to conditional mergers. Section 4 reports the results of our phonetic analysis. Section 5 discusses synchronic and diachronic issues related to the merger in light of the phonetic results. Section 6 briefly concludes.

---

1 Glides and corresponding vowels in Mandarin have typically been analyzed as single sounds; relatedly a consonant followed by a glide is analyzed as a single sound (Chao 1934:42; Duanmu 2007:25). The reason is that it is assumed that there is only one slot in the onset, which C and G must share. Glides do not contrast with corresponding high vowels, [i,u,y], and the two sets can be treated as variants of each other. Duanmu (2007:23) uses the vowel symbol when the sound is the syllable nucleus or the second part of a diphthong, such as [u] in [mu] ‘wood’ and [mau] ‘cat’, and uses the glide symbol when the sound precedes the nuclear vowel, such as [w] in [sawan] ‘sour’.

---
2. Background on the \([f]-[x]\) merger in Southwest China

Several surveys of Chinese language varieties, undertaken in the 1940’s and published in large Chinese volumes in the decades that followed (e.g., Chao, 1948; Yang, 1969, 1974, 1984), provide a comprehensive starting point for the study of language variation in China. These studies covered hundreds of language varieties across China using standardized methods and traced synchronic pronunciation patterns back to medieval Chinese. Before presenting our studies on synchronic phonetic variation within one variety, we first situate this language within the broader Southwestern Chinese sprachbund, as characterized by the seminal surveys. The data described here is based on *The Report of Sichuan Dialects* (Yang 1984), *The Report of Hunan Dialects* (Yang 1974), *The Report of Yunnan Dialects* (Yang 1969), and the *The Report of Hubei Dialects* (Chao 1948). Of 374 documented languages of Southwest China in these studies, 212 are differentiated from Standard Mandarin in patterns of phonological variation between a labiodental fricative, \([f]\), and a velar fricative \([x]\) (He 2004). The patterns of variation come in eight types. These are classified in Table 1 with reference to medieval Chinese. The top row lists relevant medieval Chinese proto-phonemes. There are often multiple sources for synchronic phonemes; however, for convenience, we’ve provided a single label for each proto-category (described below) in the second row of the table. Each row below shows how that proto-category is realized in a synchronically attested merger situation. Examples of each merger situation are given in the first column.

The table shows that synchronic \([f]\) derives from a merger of three medieval Chinese categories: \([f]\), \([\text{f}^\text{h}]\), and \([v]\). We henceforth refer to these medieval sources of modern \([f]\) as proto-\(f\), or *\(f\). Synchronic \([x]\) derives from two categories in medieval Chinese: \([x]\) and \([\text{ɣ}]\). We describe these as proto-\(x\), or *\(x\). For reference, the first row of the table lists Standard Mandarin Chinese. In this variety, there is no synchronic merger between \([f]\) and \([x]\). Other rows show different types of merger between \([f]\) and \([x]\) according to context. The attested mergers are divided into eight types. In discussing the mergers, we use both Standard Mandarin Chinese and medieval Chinese as points of reference. In the discussion that follows, and throughout the paper, we refer to the medieval Chinese categories with “*“ and we place all IPA symbols in square brackets, \([\text{]}\). To facilitate comparison across Chinese language varieties, in addition to providing English glosses for words, we also provide sino-graphs (Chinese characters).

In Type I mergers, \([f]\) and \([x]\) remain distinct except in the context of the vowel \([u]\). In this environment, \([x]\) became \([f]\). This merger created new homophones. For example, the Standard Mandarin pronunciation of \([\text{xu}]\) ‘protect’ is pronounced as \([\text{fu}]\) in Zhongjiang, which is the same pronunciation as other words, such as 父 ‘father’. In Type II, \([\text{xu}]\) is also pronounced as \([\text{fu}]\), just as in Type I mergers. Additionally, in Type II, \([\text{x} \_\_]\) is read as \([\text{f} \_\_]\). Here, we use the underscore “_” to indicate all following environments, except for \([u]\) and \([oŋ]\), which are listed in separate columns. Thus, in Type II, a word like \([\text{x}^\text{w} \text{æ̃}]\) (欢 ‘happiness’) is pronounced as \([\text{f}^\text{w} \text{æ̃}]\), which becomes homophonous with 翻 ‘turn over’. \([f]-[x]\) remains distinct in the context of \([oŋ]\), e.g., \([\text{foŋ}]\) (缝 ‘sew’) differs from \([\text{xoŋ}]\) (红 ‘red’). In Type III mergers, \([\text{fu}]\) is synchronically \([\text{xu}]\); \([\text{fon}]\) (缝 ‘sew’) differs from \([\text{xon}]\) (红 ‘red’).
is synchronically \([x^w]\), e.g., \([fæ̃]\) (范 ‘law’) is homophonous with \([x^wæ̃]\) (幻 ‘fantasy’). In Type IV, \([xu]\) became \([fu]\); \([xoŋ]\) became \([foŋ]\); and, \([x^w\_]\) became \([f\_]\). In Type V, \([xu]\) became \([fu]\) but \([f\_]\) became \([x^w\_]\). In Type VI, \([foŋ]\) became \([xoŋ]\), while other contexts maintain the labial-velar fricative distinction. In Type VII, \([xu]\) became \([fu]\), while \([foŋ]\) became \([xoŋ]\) and \([f\_]\) became \([x^w\_]\). In Type VIII, \([fu]\) became \([xu]\) and \([foŋ]\) became \([xoŋ]\), while \([x^w\_]\) became \([f\_]\).

This set of merger types and the mapping to medieval Chinese includes cases of bidirectional mergers. The labiodental fricative, \(*f\) became \([x^w]\) in some languages (e.g., Type III, Type V and Type VII); in others, \(*x^w\) became \([f]\) (Type II, Type IV and Type VIII). Type III and Type IV are polar opposites: in Type III, all \(*f\) became \([x^w]\); in Type 4, all \(*x^w\) became \([f]\). The other merger types show mergers in more restricted environments, e.g, Type VI shows a merger in just one environment: \(*foŋ\) became \([xoŋ]\) but \([f\_]\)–\([x^w\_]\) remain distinct otherwise. Of note is that it is always a round back vowel \([u]\), glide \([w]\), or \([oŋ]\) that conditions the merger or exceptions to a general pattern.

| Medieval Chinese merger types | *xu  | *fu  | *foŋ | *xoŋ | *x^w\_ | *f\_ | *x\_ |
|------------------------------|------|------|------|------|--------|------|------|
| Our label for the proto-category | *xu  | *fu  | *foŋ | *xoŋ | *x^w\_ | *f\_ | *x\_ |
| Standard Mandarin | xu   | fu   | foŋ   | xoŋ | x^w\_ | f\_ | x\_ |
| Type I | fu   | fu   | foŋ   | xoŋ | x^w\_ | f\_ | x\_ |
| Type II (Zhongjiang) | fu   | fu   | foŋ   | xoŋ | f\_ | f\_ | x\_ |
| Type III | xu   | xu   | xoŋ   | xoŋ | x^w\_ | x^w\_ | x\_ |
| Type IV | fu   | fu   | foŋ   | foŋ | f\_ | f\_ | x\_ |
| Type V | fu   | fu   | foŋ   | xoŋ | x^w\_ | x^w\_ | x\_ |
| Type VI | xu   | fu   | xoŋ   | xoŋ | x^w\_ | f\_ | x\_ |
| Type VII | fu   | fu   | xoŋ   | xoŋ | x^w\_ | x^w\_ | x\_ |
| Type VIII | xu   | xu   | xoŋ   | xoŋ | f\_ | f\_ | x\_ |

Table 1. \([f\_]\)–\([x]\) merger types in Sichuan, Hunan, Hubei and Yunnan provinces. The underscore, “\(\_\)”, indicates all following environments, except for \([u]\) and \([oŋ]\), which are listed in separate columns.
The geographical distribution of the above eight merger types across four Southwestern provinces of China is shown in the map in Figure 1.

In the Mandarin of Yunnan, Sichuan and Hubei provinces, as well as the area around Chongqing, the majority pattern is Type I, while multiple types of merger exist in Hunan and areas of Hubei that are adjacent to Hunan. In addition, there are a few different types of labial-velar merger in hilly areas in the middle Sichuan, which are also distributed in Hunan and Hubei. Zhongjiang is one such language. It is a “dialect island” in the sense that it shows a Type II pattern in an otherwise predominantly Type I area.

“Dialect islands” such as Zhongjiang show the influence of historical migration on contemporary variation. Hunan migration patterns have played a key role in dispersing language varieties, including the Zhongjiang language. Due to periods of war, famine and plague, the population in Sichuan province was dramatically reduced in Ming (1368 to 1644) and Qing (1644 to 1912) dynasties. Communities originally from Hunan, Hubei, and Guangdong provinces were encouraged to move to Sichuan in order to recover production and open up more land for farming. Vast waves of migration, termed “Huguang Fill Sichuan”, spanned 150 years (Cao 1997) and contributed to contemporary variation in Sichuan. “Dialect islands” constitute pockets of resistance to areal convergence in favor of language continuity over time. We address the role that this language contact situation may have on
It is evident that the phonetic variation in the discussion.

3. Methods

To investigate [x]-[f] variation, we conducted a phonetic analysis of a corpus of ~9,000 tokens; 10 speakers (5 female) were recorded producing 10 repetitions each of 90 monosyllabic words beginning with [x] or [f].

3.1 Speakers

Zhongjiang is a city in Sichuan province. The center of the city is a densely populated urban area. The outskirts of the city are more rural. The variety of Mandarin spoken in the more rural outskirts of the city is considered to be different from the urban variety (Cui 1996). For this study, speakers were recruited from the urban areas of Zhongjiang City. Each speaker was born and raised in the urban area of Zhongjiang and spent no more than half a year living outside of Zhongjiang. Younger speakers in Zhongjiang tend to be more influenced by Standard Mandarin, a general trend across Southwest China (e.g., Stanford & Evans, 2012). Across China, older speakers educated in local varieties tend to be more conservative with respect to adopting features of Standard Mandarin (Zhang 1995). In order to minimize the influence of Mandarin on our characterization of the Zhongjiang language, we focused on speakers over 55 years of age who had all of their front teeth. The gender, age, and occupation of our speakers are given in Table 2.

| speaker | gender | age  | occupation               |
|---------|--------|------|--------------------------|
| 1       | M      | 61   | factory worker           |
| 2       | F      | 55   | hospital nurse           |
| 3       | M      | 62   | elementary school teacher|
| 4       | F      | 56   | elementary school teacher|
| 5       | M      | 60   | factory worker           |
| 6       | F      | 60   | factory worker           |
| 7       | F      | 68   | elementary school teacher|
| 8       | M      | 65   | elementary school teacher|
| 9       | M      | 61   | city hall office worker  |
| 10      | F      | 59   | businessman              |

Table 2. Gender and age by speaker ID

Materials consisted of a total 90 monosyllabic words that can be represented orthographically as single Simplified Chinese characters. In Standard Mandarin, the syllables start with either...
[f] (N = 37) or [x] (N = 53). Of the 53 syllables that start with [x] in Standard Mandarin, 33
are in the environment of [w]. It is these 33 syllables that are expected, on the basis of the
characterization of Zhongjiang as exhibiting a Type II merger, to be produced as [f]. The
complete list of words, including the Simplified character used for presentation, Standard
Mandarin IPA, pinyin (standard Romanized orthography), and the expected IPA for a Type II
language variety, is provided in the appendix in the same pseudo-randomized order as it was
presented to participants (see Methods).

The list of 90 items results from balancing a number of design constraints. We attempted to
elicit each fricative in a range of different environments but also to balance the rime context,
including the tone, across fricatives and to focus on relatively high frequency words. This last
concern is because low frequency words could potentially elicit more literary, archaic or
Standard Mandarin pronunciations, which might not accurately reflect the contemporary
speech patterns of the regional Zhongjiang variety. There are systematic phonotactic
constraints prohibiting certain fricative vowel combinations, e.g. [f] cannot be followed by [w]
or [au], as well as some accidental gaps which limit the number of perfect minimal pairs.
Focusing on well-known words also limited the number of minimal pairs. Of the 90 mono-
syllabic words, 44 words enter into a minimal pair (a total 22 minimal pairs). We made sure
that there are at least some minimal pairs for each of the vowel environments relevant to
characterizing conditional mergers (see Table 1), which we describe in further detail below.

We deal with the lack of balance in the corpus in two ways. First, we use mixed effects
modelling to assess the effect of fricative while accounting for contextual variation with a
combination of random and fixed effects. Second, we run follow-up analyses on subsets of
the corpus consisting of minimal pairs controlled to address specific aspects of the
Zhongjiang pattern, including context. We describe these minimal pair subsets below as
separate studies along with the specific objective of each.

Study 1: minimal pairs from *x and *xw
As a Type II variety, we expect *x and *xw in Zhongjiang to be contrastive: [x]-[f]. This is
because *xw is expected to be [f]. In order to observe whether the change from *xw to [f] is as
expected, we chose 4 minimal pairs (8 items), in which we expect the contrast between [x]
and [f] to be maintained. For example, [xa\textsuperscript{45}] ‘laughter’ (Simplified Chinese characters and
Pinyin: 哈, ha\textsuperscript{55}) and [fa\textsuperscript{45}] ‘flower’(花, hua\textsuperscript{55}) are expected to be minimal pairs in Type II
languages. This contrasts with Standard Mandarin Chinese, in which ‘哈’ and ‘花’ have the
same onset consonant: [x]. If our speakers produce these words as expected, study one will
provide minimal pairs offering a baseline for phonetic differences between [x] and [f] outside
the conditioning environments for merger.
| Sino-graph characters | Mandarin Pinyin | Type II expectation Zhongjiang IPA |
|----------------------|----------------|----------------------------------|
| 哈-花 | ha₅₅-hua₅₅ | [xa₄₅]-[fa₄₅] |
| 害-坏 | hai₅¹-huai₅¹ | [xai₃²₄]-[fai₃²₄] |
| 黑-或 | hei₅₅-huo₅¹ | [xe₃¹]-[fe₃¹] |
| 还(还有)-怀 | hai₃⁵-huai₃⁵ | [xai₃¹]-[fai₃¹] |

Table 3. Contrastive pairs (*x_-*xʷ_) in Zhongjiang Chinese

**Study 2:** merger of [f] and [xʷ]

In this group, we expect [f] and [xʷ] to merge to [f], resulting in non-contrastive pairs. In terms of phonetic measurements, we expect completely overlapping distributions for [f] and [xʷ]. We chose 13 pairs (26 items), which are contrastive in standard Mandarin but are expected to have merged in Zhongjiang. We refer to these as ‘non-contrastive’ pairs; an example is [fei₃¹] ‘fat’ (肥, fei₃⁵) and [fei₃¹] ‘return’ (回, hui₃⁵). The merger in Zhongjiang results from [xʷ] in Standard Mandarin (‘hu’ in Pinyin) produced as [f] in Zhongjiang.
Table 4. Non-contrastive pairs (*f_-*xw_) in Zhongjiang Chinese

| Sino-graph characters | Mandarin Pinyin | Type II expectation | Zhongjiang IPA |
|-----------------------|-----------------|---------------------|----------------|
| 发(发财)-华          | fa55-hua35      |                     | [fa31]         |
| 肥-回                 | fei35-hui35     |                     | [fei31]        |
| 肺-会                 | fei51-hui51     |                     | [fei324]       |
| 飞-灰                 | fei55-hui55     |                     | [fei45]        |
| 翻-欢                 | fan55-huan55    |                     | [fa45]         |
| 烦-还                 | fan35-huan35    |                     | [fa31]         |
| 返-缓                 | fan214-huan214  |                     | [fa31]         |
| 饭-换                 | fan51-huan51    |                     | [fa324]        |
| 分-婚                 | fen55-hun55     |                     | [fan45]        |
| 坟-魂                 | fen35-hun35     |                     | [fan31]        |
| 奋-混                 | fen51-hun51     |                     | [fan324]       |
| 方-慌                 | fang55-huang55  |                     | [fan45]        |
| 房-黄                 | fang35-huang35  |                     | [fan31]        |

Table 5. Contrastive pairs (*f_-*x_) with vowel [əɯ] in Zhongjiang Chinese

| Sino-graph characters | Mandarin Pinyin | Type II expectation | Zhongjiang IPA |
|-----------------------|-----------------|---------------------|----------------|
| 浮-猴                 | fu35-hou35      |                     | [fu31]-[xou31] |
| 否-吼                 | fou214-hou214   |                     | [fu51]-[xou51] |

Study 3: contrast continuity, [f]-[x]

The items in this study provide an important control comparison. They consist of words that were contrastive in medieval Chinese and are still contrastive in both Standard Mandarin and the Zhongjiang language. For example, [fəɯ51] ‘deny’ (否, fou214) and [xəɯ51] ‘yell’ (吼, hou214). There are 2 such pairs (4 items) in our wordlist, which are contrastive both in standard Mandarin and Zhongjiang.

Study 4: the [oŋ] environment

From previous research, we expect the contrast between [fəŋ] and [xəŋ] to be maintained in Zhongjiang. However, based on our perceptions in the field, we have noticed some variations
between [f] and [x] in the context of final [oŋ] and this is an environment that is known to
condition variation across Chinese languages. Table 6 included 3 pairs (6 items), which are
expected to be contrastive in both standard Mandarin and Zhongjiang: for example, [xoŋ³¹]
‘red’(红, hong³⁵) and [foŋ³¹]‘sew’(缝, fong³⁵).

| Sino-graph | Mandarin | Type II expectation |
|------------|----------|---------------------|
| characters | Pinyin   | Zhongjiang IPA       |
| 風-轰     | fong⁵⁵-hong⁵⁵ | [xoŋ⁴⁵]-[foŋ⁴⁵]     |
| 冯-洪     | fong³⁵-hong³⁵ | [xoŋ³¹]-[foŋ³¹]     |
| 红-缝     | fong³⁵-hong³⁵ | [xoŋ³¹]-[foŋ³¹]     |

Table 6. Contrastive pairs (*f_*-*x_*) with final [oŋ] in Zhongjiang Chinese

3.3 Procedure

Participants were recorded in a sound-attenuated room at Wucheng hotel in Zhongjiang City. The data reported in this paper were part of a longer recording session, including spontaneous speech and other elicitation materials. The list of monosyllables reported here was recorded immediately after the spontaneous speech portion of the session, in which participants were asked to talk about their life in Zhongjiang or to introduce some aspect of Zhongjiang life: food, popular local attractions, etc... Before recording the monosyllables, all participants were given the complete list on paper to look over. The order of the items in the list, also the order of items listed in the appendix, was pseudo-random. Each participant confirmed that they knew all of the words on the list. After this familiarization stage, the target items were displayed one at a time on a computer screen, in the same pseudo-randomized order as they appeared in the paper list. Participants were asked to read each item when it appeared on the screen. The items were presented continuously until 10 repetitions of the list were elicited.

The hotel where we did the recording has carpet on the floors and wallpaper on the walls. We chose a comparatively small room without a window to avoid noise from outside, and hanged curtains on the walls to minimize echo. Before we started the recording, we closed the air conditioner, electric fan, fluorescent lamp and phone, and used Audacity software to test that the noise was under -48 dB, and the volume of the speakers was in the range of -18 to -6 dB.

We used BYLY software to record the sound, which can automatically save each monosyllable recording as a single sound file, and display the waveform automatically and synchronously, which is helpful to test the quality of each token. All the tokens were saved as .wav format. All tokens were recorded in mono channel at 44,100 Hz directly to a Thinkpad T440 laptop using an external Samson C03U microphone, which was stabilized with a microphone stand and set to unidirectionality mode.
Segment boundaries were determined by forced alignment, using the Montreal Forced Aligner. We created two sets of segment boundaries, one based on the pre-trained Standard Mandarin aligner and another based on a Zhongjiang-specific aligner, trained on our recordings. To evaluate the forced alignment, we hand-segmented 100 tokens from one speaker and assessed correlations between the hand-segmented and force-aligned tokens both for segment duration and also for the dependent measures of interest for the study (see below). The Zhongjiang-specific aligner showed higher correlations with the hand-measured set than the Standard Mandarin aligner, so we proceeded by using the Zhongjiang-specific aligner throughout. A total of 9 tokens (0.1%) were excluded due to alignment failure.

Spectral measurements were extracted using Praat (Boersma & Weenick, 2016), with reference to the segment boundaries from forced alignment. We extracted measurements at five different timestamps in the target fricatives, the first 20 ms of the fricative, the second 20 ms of the fricative, the middle 20 ms of the fricative, the penultimate 20 ms time window and the final 20 ms of the fricative. Our main analysis focuses on spectrum Centre of Gravity (CoG). This measurement is known to be sensitive to the frequency range of the analysis (e.g., Shadle & Mair, 1996). Since our recordings are studio-quality, we opted to use the maximal frequency range at our disposal, basing our analyses on the Nyquist frequency: 22,500 Hz. We discarded extreme outliers, defined as tokens that were greater than three standard deviations from the mean CoG and spectrum SD calculated across the entire data set. This resulted in the loss of 46 tokens or 0.5% of the data.

4. Results

4.1 Fricative Duration

We start by reporting the duration of the fricatives. Figure 2 shows a kernel density plot of fricative duration across the entire corpus based on the proto-category of the fricative. The figure collapses across speakers and items. The distributions for all three proto-category fricatives, *[f, *xʷ, *x, overlap heavily, an observation which also extends to each of the 10 speakers individually. In short, [x] and [f] show very similar durations in this corpus. The peaks of distributions show subtle differences: *f is slightly shorter than *xʷ which is slightly shorter than *x. To investigate whether this difference is statistically significant, we fit two nested linear mixed effects models to the duration data using the lme4 package (Bates et al 2019) in R (version 3.9.2). The baseline model contained only random effects: a random intercept for speaker and a random intercept for item. To this baseline, we added fricative (proto-category) as a fixed factor. A likelihood ratio test showed that the model including fricative did not significantly improve over the baseline model ($\chi^2 = 1.19$, $p = 0.55$) indicating that the difference in duration across fricatives is not significant. This includes both the difference between *f and *xʷ, hypothesized to have merged in this variety, and between *f and *x, which are expected to maintain contrast.
4.2 Fricative Spectra

Before discussing the measurements of CoG and spectrum SD for each of the four studies, we first exemplify the range of spectral variation observed in the data and how that variation corresponds to our measurements. The fricative spectra shown in the figures below were extracted using the middle 20 milliseconds of each fricative. They show the average power of the fricative during this time window across the frequency range from 0 to 22,500 Hz. The power is expressed in dB relative to the reference value of 0.00002 Pa.

We begin with the velar fricative [x]. Figure 3 shows three examples from the corpus. The distribution of energy in these tokens has a long right tail with a peak at low frequency. Since most of the energy is concentrated in the lower frequencies, these tokens are characterized by a low CoG and a low standard deviation. This is expected for fricatives with a posterior constriction in the vocal tract. Aperiodic energy generated at the posterior constriction will resonate in the portion of the oral cavity in front of the turbulent energy source. The longer the cavity in front of the constriction, the lower resonant frequency. Peaks in the spectra for [x] in the range of 600-1,000 Hz are consistent with resonance in front of a velar constriction in the vocal tract.
Figure 3. Spectra of [x] tokens

Figure 4 shows spectra for tokens of [f] in the corpus with typical CoG values. Compared to [x], these tokens do not have the same degree of low frequency energy. The decrease in energy with higher frequencies is more gradual. Energy remains closer to the reference level 0dB at higher frequencies in the [f] spectrum than in the [x] spectrum. These two characteristics are reflected in the spectral moments: [f] (Figure 4) has a substantially higher CoG and spectrum SD than [x] (Figure 3).

Figure 4. Spectra of [f] tokens

The tokens in Figure 3 all sound unambiguously like [x] to the authors and the tokens in Figure 4 sound unambiguously like [f]. Figure 5 provides direct comparisons of [x] and [f] tokens. The first panel overlays the spectra of tokens of [x] and [f] that sound distinct. The spectrum for [x], shown in grey, is more peaky than for [f], shown in black, and the [f] spectrum maintains energy at higher frequencies. The second panel overlays spectra of fricatives that both sound like [f] to the authors; one of these derives from *f and the other from *xw. These tokens both have a relatively flat energy profile (diffuse spectrum) particularly at frequencies above 5,000 Hz. This spectral profile yields a high CoG and high spectrum SD for both tokens, even though one of them is historically derived from [xw].
| Token | Speaker | COG (Hz) | SD (Hz) |
|-------|---------|----------|---------|
| *x₁, [xe³¹] | (speaker4) - black line | 1,986 | 1,623 |
| *x₂ | (speaker6) - black line | 5,755 | 6,137 |
| *x₃, [fe³¹] | (speaker4) - grey line | 4,581 | 6,185 |
| *x₄ | (speaker6) - grey line | 3,888 | 5,466 |

Our brief exemplification of the spectra above serves in part to motivate our choice of using CoG and Spectrum SD to phonetically characterize the Zhongjiang fricatives. These two spectral measures offer only a very sparse characterization of the spectrum, but we have found that by and large the differences observed in these numbers correspond with our subjective impressions of the auditory classifications of the sounds.

4.3 Variation in CoG across Fricatives

In order to assess how non-coronal fricatives are realized in Zhongjiang, we fit a series of linear mixed effects models to CoG extracted at five different times in the fricatives. Models were fit to each measurement of CoG at each time point separately, using the lme4 and lmerTest packages (Bates et al., 2015) in R (version 3.6.2). Our baseline model contained only random effects, a random intercept for speaker and a random intercept for item. This was the maximal random effects structure that allowed all models to converge. To this baseline model, we added consonant duration as a fixed effect, which significantly improved over the baseline for all models, as indicated by a significant likelihood ratio test and lower Akaike Information Criterion (AIC). The addition of rime as a fixed factor also led to significant improvements at all timepoints as did the addition of fricative. Our final model added the interaction between fricative and rime as an additional fixed factor, which also led to significant improvements at all timepoints. Rime and fricative were contrast coded. For rime, there were 32 levels, with [əɯ³¹] serving as the reference level. For fricative, there were three levels, *f₁, *x₁, *x₃, which are based on the proto categories for the fricatives (and also consistent with Standard Mandarin), and [f₁] was the reference level. This coding schema
allows us to interpret the model intercept as the CoG estimate for [f] and the \( \beta \) coefficients for
[x] and [x\text{"}] as deviations from [f]. We therefore expect, based upon the spectra in 4.2 as well
as past research that [x] should have a negative coefficient, since the CoG for [x] is lower
than the CoG for [f]. At issue for the characterization of Zhongjiang as a Type II merger is
whether the coefficient for [x\text{"}] will be significantly different from [f].

A summary of the nested linear mixed effects models fit to CoG from five different
timestamps is summarized in (1):

(1) Summary of linear mixed effects models

\[
\begin{align*}
m0: & \quad \text{COG} \sim (1 | \text{speaker}) + (1 | \text{item}) \\
m1: & \quad \text{COG} \sim \text{C\_dur} + (1 | \text{speaker}) + (1 | \text{item}) \\
m2: & \quad \text{COG} \sim \text{C\_dur} + \text{rime} + (1 | \text{speaker}) + (1 | \text{item}) \\
m3: & \quad \text{COG} \sim \text{C\_dur} + \text{fricative} + \text{rime} + (1 | \text{speaker}) + (1 | \text{item}) \\
m4: & \quad \text{COG} \sim \text{C\_dur} + \text{fricative\_rime} + (1 | \text{speaker}) + (1 | \text{item})
\end{align*}
\]

Table 7 summarizes the modelling results. Each column shows a different timestamp: t1 is
the first 20ms of the fricative, t2 is the second 20ms, midpoint is the middle 20ms, t4 is the
penultimate 20ms window and t5 is the last 20ms of the fricative. The first three rows report
summaries from m3, as defined in (1): the intercept, which serves as an estimate for [f], the
\( \beta \) coefficient for [x], and the \( \beta \) coefficient for [x\text{"}]. The estimates from m3 are not complicated
by the interaction with rime, a topic which we take up later in the paper, including through an
analysis of minimal pair subsets in Section 4.4. The final three rows of Table 7 show the \( \chi^2 \)
statistic from model comparison via anova for the effects of rime (anova comparison of m1
and m2), fricative (anova comparison of m2 and m3), and the interaction between them
(anova comparison of m3 and m4).

As mentioned above, the addition of the fixed factors, consonant duration (m1), rime (m2),
fricative (m3), and rime\_fricative (m4) each led to significant improvements in model fit. The
direction of the consonant duration effect depended on the timestamp of the measurement--
negative at t1 and t2 (t1: \( \beta = -6.8428^{**} \); t2: \( \beta = -4.3686^{**} \)), positive at the midpoint (\( \beta = 5.9953^{**} \)), and positive but much weaker at later timestamps (t4: \( \beta = 0.6951 \); t5: \( \beta = 0.4993^{*} \)). The effect of consonant duration on the midpoint might be related to voicing, as
voicing may be more likely to bleed into the fricative midpoint at shorter durations. In
contrast, the negative effect at early timestamps could be related to the sensitivity of the
forced aligner to fricative intensity, i.e., longer fricatives may be picked up earlier in time
before constrictions have formed.

However, the magnitude of the improvement varied depending on the timestamp at which
CoG was extracted. For rime, the strongest effect was found on the later timestamps, t4 and
t5--this is indicated by the larger \( \chi^2 \) at t4 (198.71^{***}) and t5 (184.46^{***}) than at earlier
timestamps (t1: 141.07***; t2: 149.63***; midpoint: 152.12***). For fricative, on the other hand, the strongest effects were found at earlier timestamps, indicated by larger $\chi^2$ at t1 (116.17***), t2 (128.32***), and the midpoint (127.81***) than at later timestamps (t4: 77.74***; t5: 26.66***). This indicates that the influence of rime, including both the vowel and tone, on fricative CoG increases over the course of the fricative while the influence of fricative type, including the contrast between *f, *x, *x w, decreases towards the end of the fricative.

Inspection of the $\beta$ coefficients for fricative in (m3) provides a more nuanced perspective on how fricative type influences differences in CoG over time. Across the models of CoG measurements at five timepoints, the intercept is always significant at $p < .01$ as is the $\beta$ coefficient for *x. In contrast, the $\beta$ coefficient for *x w is only significant at the midpoint and at t4. The magnitude of the differences between *f and *x w (402 Hz at the midpoint) is much smaller than the difference between *f and *x (-2863 Hz at the midpoint). This pattern is in the direction of a Type II merger, but it suggests that the merger may be incomplete. We delve into this issue in greater depth when we evaluate the minimal pair subsets in Section 4.4.

| Model | 1st 20ms | 2nd 20ms | midpoint | pre-final 20ms | final 20ms |
|-------|----------|----------|----------|----------------|------------|
| m3: Intercept *f | 3312.06*** | 4185.02*** | 3640.16*** | 2499.17*** | 913.07*** |
| m3: $\beta$ *x | -1569.71*** | -2400.99*** | -2863.72*** | -1489.30*** | -515.80** |
| m3: $\beta$ *x w | -72.69 (n.s.) | -135.84 (n.s.) | -402.50** | -376.28** | -92.83 (n.s.) |
| m2: $\chi^2$ - rime | 141.07*** | 149.63*** | 152.12*** | 198.71*** | 184.46*** |
| m3: $\chi^2$ - fricative | 116.17*** | 128.32*** | 127.81*** | 77.74*** | 26.66*** |
| m4: $\chi^2$ - rime*fricative | 63.64*** | 86.50*** | 112.18*** | 135.44*** | 198.30*** |

Table 7. Summary of linear mixed effects modelling of CoG at different timestamps. Models were fit to the entire corpus, 90 monosyllables produced 10 times each by 10 participants (see methods). The first three rows summarize the intercept and $\beta$ coefficients for fricative in the best fitting model: m3 in (1). Asterisks indicate statistical significance: *** for $p < .001$; ** for $p < .01$; * for $p < .05$. The intercept shows the estimate for *f, which is significant at all timestamps. The estimate for *x (second row) is always negative and statistically significant at either *** or ** levels. *x w (third row) is not significant at any timepoint, indicating merger with *f. The bottom two rows report chi-squared statistic from model comparisons evaluated the fixed effect of rime (fourth row) and fricative (fifth row). These fixed factors lead to statistically significant improvement at all timepoints; however, the degree of improvement varies across time, with fricative (fifth row) having a stronger effect earlier in time and rime (fourth row) having a strong effect later in time.

The absolute value of the estimate for [x] is greatest at the fricative midpoint ($\beta = -2863.72$) and decreases at the end of the fricative to -1489 Hz at t4 and then -515.80 Hz at t5. This decrease in CoG difference between [f] and [x] at the end of the fricative reflects an
increasing influence of rime on CoG towards the end of the fricative, as also revealed by model comparisons discussed above. The significant interaction between fricative and rime (m4) at all time points indicates that the effect of rime on CoG is not the same for each fricative. Interpreting this interaction within the entire corpus is complicated by the unbalanced nature of the corpus, due to phonotactics, accidental gaps, and the restriction of target items to high frequency words. We therefore revisit the analysis of fricative CoG within subsets of the corpus controlled for rime type.

To summarize the results so far, we investigated fricatives corresponding to three proto-categories, *f, *x w, *x. The proto-categories correspond to three distinct fricatives in some other Chinese language varieties, including Standard Mandarin. In terms of duration, the three proto-categories are indistinguishable in Zhongjiang. Our analysis of CoG showed that *x is spectrally distinct from *f at all time points, the beginning, middle and end of the fricative, while differences between *f and *x w were smaller and significant only at the midpoint and t4. These results are largely consistent with past reports, based on impressionistic listening, that Zhongjiang exhibits a Type II merger. Our analysis is based on 8,945 tokens, repetitions of 90 monosyllabic items, which vary in the composition of their rimes. The analysis above also revealed that the CoG of fricatives is significantly impacted by the identity of the following rime. The effect of rime on CoG is significant throughout the fricative, i.e., at early time points as well as later timepoints, but it is strongest at the end of the fricatives (later time windows) with corpus imbalance factored into the analysis through a combination of fixed and random effects. In particular, we found that rime, one of the factors that was not perfectly controlled for in the materials, had a significant effect in the model, across time points (although it was strongest towards the end of the fricative). To provide a complementary perspective, we now turn our attention to subsets of the larger corpus that constitute minimal pairs, i.e., are controlled for rime context.

4.4 Minimal pair comparisons

We now turn to a spectral analysis for each subset of the data, comprising four studies.

Study 1: minimal pairs from *x and *x w

The purpose of study 1 was to establish phonetic differences between [x] and [f]. For this purpose, we selected four sets of words that we expect to form minimal pairs in contemporary Zhongjiang. A key assumption underlying our selection of these words as minimal pairs is that Zhongjiang is a Type II language (see Table 1), as has been claimed in the literature. As a Type II variety, *x w has become [f]. The minimal pairs in this study contrast *x~*x w, which are synchronically expected to be [x]~[f]. For example, [xf45] ‘laughter’ (Simplified Chinese characters and Pinyin: 哈, ha45) and [fa45] ‘flower’ (花, hua45).

Figure 7 shows the average CoG over time for each fricative proto-category. Differences in CoG between *x and *x w are already present at the earliest time point, t1. The CoG for both fricatives increases over the course of the fricative, with a maximum CoG at the midpoint of the fricative. This is also the timepoint with the greatest difference between categories. The
difference in CoG decreases dramatically toward the end of the fricative (t4) and is completely neutralized by the end of the fricative (t5).

![Study 1 Items](image)

*Figure 7. Average CoG for *x and *xʷ proto-categories at five timepoints.*

Since we will be discussing cases of fricative merger, we focus our subsequent discussion on inter-speaker variation on the midpoint of the fricatives, as this is the time window that shows the largest average difference between fricatives.

![proto_category](image)

Figure 8 shows the CoG and SD measurements for Study 1 words, minimal pairs contrasting [x]–[f]. As expected, the [x] words consistently show low CoG and low SD. Many tokens of [f] are also as expected, showing high CoG and high SD. However, there is variation in the [f] category. Some tokens of [f] are closer to the [x] category and some overlap with [x] substantially. One speaker, S2, does not show a clear contrast between [x] and [f], producing [f] tokens as [x]. Another speaker, S7, shows substantial overlap between [x] and [f]. Speakers S1, S3, S4, S6, S8, show a smaller number of [f] tokens that overlap phonetically with [x]. The remaining speakers, S5, S9, S10, show clearer separation between the fricatives in this minimal pair context.
To evaluate the statistical significance of the trends in Figure 8, we fit linear mixed effects models of the same structure shown in (1), m0, m1, m2, m3, m4, to the study1 words. The results for study1 were similar to the results for the larger corpus, except that the effect of rime was not significant in the study one words--m2 did not show significant improvement over m1 ($\chi^2 = 0.996$) and m4 did not show significant improvement over m3 ($\chi^2 = 2.08, p = 0.14911$). This indicates that the subset of rimes included in study1 words do not impact CoG at the midpoint. As in the larger model, m1, the model adding duration, showed significant improvement over m0 ($\chi^2 = 4.36^*$); and m3, adding fricative proto-category, showed significant improvement over m2 ($\chi^2 = 27.45^*$). The estimate for *x, based on m3 was 1063 Hz. This is higher than the estimate of 776 Hz for the main model, which can be obtained by adding the effect of fricative (-2863.72) to the intercept (3640.16) reported in Table 7. The model estimate for *x w in study1 words is 3929 Hz, which is similar, though somewhat higher, to the larger model estimate of 3238 Hz for *x w. The study 1 words indicate a significant differences between *x and *x w; however, as Figure 8 shows, even in minimal pair contexts, the majority of the speakers in this sample show some phonetic overlap between [x] and [f].

**Study 2:** merger of [f] and [x w]

Study 2 features 13 pairs of words that we expected, on the basis of the characterization of Zhongjiang as a Type II variety, to be fully merged. These are pairs of words that historically derive from a contrast between *f and *x w. In Zhongjiang, *x w is reported to have changed to [f]. This change produced homophones from minimal pairs. For example, 肥 ‘fat’ (Pinyin fei35) and 回 ‘return’ (hui35) were phonetically distinct in medieval Chinese and are
synchronously distinct in other Chinese languages (including Standard Mandarin), but are expected to be homophonous in Zhongjiang.

Figure 9 plots the average CoG at each time point of analysis. The CoG is similar across proto-categories at all time points but maximally distinct at the midpoint. At the first time window, *xʷ, is already similar to the values reported in study1, which are slightly lower than *f.

Figure 9 plots the token-by-token variation by speaker. As expected, most speakers show a merger for these words—the CoG and SD values tend to overlap. Moreover, the range of phonetic values for these words corresponds on a speaker-by-speaker basis to the range of values observed for [f] in minimal pairs (Study 1, Figure 8). The one exception to this pattern is S2. S2 produces a contrast between these pairs that is in the direction of what would be expected for Standard Mandarin; however, S2’s production of the *f→[f] category shows a range of phonetic variation for [f] that is common amongst Zhongjiang speakers. That is, there is a range of values extending from high CoG and high SD down to low CoG and low SD. In other words, S2 maintains a pattern of contrast across words that is similar to Standard Mandarin using a [f] that is phonetically like the Zhongjiang [f].
Figure 10. Center of Gravity and Spectrum Standard Deviation of non-contrastive pairs (*f*-*xw*).

To assess the statistical significance of the differences in Figure 10, we fit the same set of linear mixed effects models, shown in (1), to the Study2 words. The addition of duration ($\chi^2 = 30.69***$) and rime ($\chi^2 = 27.81***$) led to significant improvements. Beyond that, the addition of fricative proto-category also led to significant model improvement ($\chi^2 = 24.48***$). This indicates that the merger between *f* and *xw* is not complete. We also tested interaction between rime and fricative, which was not significant ($\chi^2 = 19.53$, $p > 0.05$). The estimate for *xw* was 3815Hz, which is similar to the larger study (see Table 7). The estimate for *f* was 4351Hz, an increase of 536Hz. This significant effect is driven entirely by speaker 2. If we re-run the models without speaker 2, then the effect of fricative proto-category is not significant ($\chi^2 = 2.24$, $p > 0.05$) and the estimates for *f* and *xw* are nearly identical: 3537Hz for *f* and 3572 for *xw*.

To summarize, the results of study 2 show that most speakers (9/10) produce *xw* as [f], resulting in an increase of homophonic pairs in Zhongjiang relative to Standard Mandarin. The majority pattern verifies Zhongjiang as a Type II language, as claimed in past work.

**Study 3:** contrast continuity, [f]-[x]

Study 3 provides an important control case for interpreting the results of study 1 and study 2. The items in study 3 consist of pairs that are predicted to be distinct in both Zhongjiang and Standard Mandarin. The mergers (study 2) and contrasts (study 1) in the first two studies are specific to Zhongjiang (and other Type II languages), but the words in Study 3 represent contrasts more broadly. This is because there is no labial glide involved in the contrast. Both
pairs of words in Study 3 were contrastive in medieval Chinese and remain contrastive synchronically. This comparison is a useful control case because it allows us to investigate whether the range of variation found for [f] persists even in contexts that have been stable over time.

Figure 11 plots the average CoG at each time point of analysis. The results closely resemble those for the study1 items. The difference in CoG is present even in the early windows. CoG rises over time, but it rises faster for [f] than for [x], with the maximum difference occurring at the midpoint, before falling in later windows.

Figure 12 shows the range of variation by subjects. All speakers maintain a distinction between [f] and [x] for these words, although the categories overlap slightly for some speakers. All speakers show a range of phonetic variation for [f] that is comparable to studies 1 and 2.

We conclude that the variation found for [f] in studies 1 and 2 cannot be attributed only to a historical or synchronic connection to the labial glide. Study 3 shows that even [f] in the environment of [əɯ] shows a wide range of variation.
Figure 12. Center of Gravity and Spectrum Standard Deviation of contrastive pairs (*f*-*x*) with rime [ɔu̯]

**Study 4: the [oŋ] environment**

The final study investigated the environment of [oŋ], as this environment has been known to condition mergers and splits in other Chinese language varieties (see Table 1). As a Type II language, Zhongjiang is expected to maintain contrast in this environment. Figure 13 plots the average CoG over time. Compared to other contexts, the average CoG for *f* is lower in the [oŋ] environment, leading to reduced contrast with *x.*
Figure 13. Average CoG for *f and *x proto-categories at five timepoints.

Figure 14 plots the variation in CoG at the midpoint by speaker. Several speakers, including some that maintain contrast in other environments, show mergers or partial mergers in the context of [oŋ]. Speakers, S1, S4, S9, who all maintain contrast in the study 1 words, show a merger between [x] and [f] in the environment of [oŋ]. Notably, the direction of the merger is such that CoG for *f is lowered. This contrasts with the merger that we’ve observed elsewhere, in which *x in the environment of [w] increases in CoG.
5 Discussion

To summarize the main results, we found that CoG for *xw and *f are largely merged in the Zhongjiang variety of Chinese, with *xw generally produced as [f]. This result is supported by the mixed effects model of the entire corpus, as well as study 2 words, subsets for which the rime context is closely controlled. We found no differences between [x] and [f] in duration. At the beginning of the fricatives, t1, the first 20 ms window, and t2, the second 20 ms of the fricative, there was no significant difference between *xw and *f in the larger corpus. Later in the fricative, including the middle 20 ms, there was a small but significant difference, with *xw having a lower CoG than *f ($\beta = -402.5$ Hz, $p < 0.01$). This decrease in CoG leaves *xw still much closer to *f than to *x, suggesting perhaps incomplete neutralization of *xw and *f. When we controlled for rime context--the minimal pairs in study 2--we found that 9 out of 10 speakers showed a complete merger, i.e, no significant differences between *xw and *f. This result is consistent with Zhongjiang as a Type II variety, although there was some inter-speaker variation. The most common pattern was shared by four subjects: S3, S4, S8, and S10. This group includes two male participants (S3 & S8) and two female participants (S4 & S10). For this group, *x in the environment of [w] is produced with the range of variation characteristic of synchronic [f]. Essentially, this group shows a synchronic merger between *f and *xw, the expected Type II merger. Three other subjects, S1, S6, S9, show only minor deviation from the dominant pattern. We return to the issue of inter-speaker variation in greater detail below.
We also found that the environment [on] conditions lower CoG in *f, such that the CoG for *f moves close to [x]. This direction of change, *f → [x] is the opposite direction of the *xw → [f] merger, observed in other rime environments. This aspect of the Zhongjiang pattern has not previously been documented and is inconsistent with prior expectations for a Type II variety. Finally, we found that [f], including both *f and *xw, shows a wide range of CoG and spectrum SD variation.

Before continuing with our discussion of how synchronic phonetic variation for [f] and [x], may be related to patterns of merger in Zhongjiang and other varieties, we first consider how the measurements of CoG we report for Zhongjiang compare with other measurements of CoG for [f] and [x] reported in the literature.

5.1 Comparison of Zhongjiang fricative CoG with other language varieties

Table 8 provides descriptive statistics for CoG, comparing the current study (top row) with others in the literature. The table includes the mean CoG, the standard deviation (SD) where available, and the number of tokens (N) that entered into each analysis. For the current study, we also provide the CoG estimates from our mixed models, which take into account control variables and random effects. As far as we know, there are no other quantitative phonetic analyses of Zhongjiang non-coronal fricatives. For reference, the table includes a seminal study on Standard Mandarin fricatives (Svantesson 1986) as well as a more recent update (Ran 2008), a recent study on another variety of Mandarin, Jianghuai Mandarin, with a larger number of tokens than the other reference points (Wu 2020), and a study on Cairene Arabic (Norlin 1983), which also has a contrast between labiodental and velar fricatives. Most of the studies present CoG values in Hz, but some present results in critical band units (CBU). Svantesson (1986) reports CoG in both Hz and CBU.

Across studies, it is always the case that [f] has a higher CoG than [x], although the magnitude of this difference varies. Our study reports the largest difference between [x] and [f]. Average CoG for [f] across studies ranges between 4,224 Hz and 5,303 Hz. This is a similar range to that reported by Gordon et al. (2002) in their typological survey of fricatives. On average, the CoG for [f] in Zhongjiang is relatively similar to other language varieties with [f]. If anything, the average CoG for Zhongjiang is slightly higher than other reported values. We note, however, that there is substantial variability in the CoG measurements for Zhongjiang [f]. This is apparent both in the high SD (3,095 Hz) as well as the figures of Section 4.4. Variability is harder to compare across studies because of the different numbers of tokens--the standard deviation of CoG is possibly undersampled in the smaller studies. The number of tokens analyzed here is at least an order of magnitude greater than other studies comparing the CoG of [x] and [f]. We note as well that the model estimate for [f], 3,640 Hz, which factors control variables into the analysis, is substantially lower than the average CoG, 5,303 Hz. The large standard deviation for *f CoG of 3,095 Hz reflects the observation in Figures 8, 10, 12, 14, which all show a wide range of CoG variation for *f, including values that encroach on *x.
The average CoG for [x] in our data is quite a bit lower than other language varieties. Estimates from other studies in Table 8 ranged from 2,979 Hz (Norlin 1983) to 4,744 Hz (Wu 2020). The average CoG for [x] in our study was 1,782 Hz. It is possible that the lower CoG for [x] is a reflex of the variability in CoG found for [f]. That is, the spectral properties of fricatives may be subject to auditory dispersion in a manner similar to vowels (Liljencrants and Lindblom 1972). Spectral variation for [f] resulting in lower CoG in some environments may provide pressure to further lower the CoG for [x]. Articulatorily, pressure to lower CoG could result in more posterior constrictions in the vocal tract, which increase the length of the cavity in front of the constriction, thereby lowering resonance frequency. Increasing the amplitude of a low resonance serves to lower CoG. Indeed, The range of CoG values for [x] in our data overlaps values for [x] and for [χ], [h], [h], as reported for Arabic in Norlin (1983).

| CoG estimates of non-coronal fricatives | [x]            | [f]            |
|----------------------------------------|----------------|----------------|
| Current study: Zhongjiang Mandarin     | 1,782 Hz (β = 776) | 5,303 Hz (β = 3,640) |
|                                        | 1,474           | 3,095          |
|                                        | 2,994           | 5,951          |
| Norlin (1983): Cairene Arabic          | 2,979 Hz        | 5,034 Hz       |
|                                        | 659             | 1,331          |
|                                        | 6               |                 |
| Wu (2020): Jianghuai Mandarin         | 4,744 Hz        | 5,280 Hz       |
|                                        | NA              | 1,050          |
|                                        | 598             | 600            |
|                                        |                 |                 |
| Svantesson (1986): Standard Mandarin   | 3,266 Hz 16.28 CBU | 4,224 Hz 18.29 CBU |
|                                        | 1,262           | 1,050          |
|                                        | 8               | 8              |
| Ran (2008): Standard Mandarin         | 15.99 CBU       | 21.01 CBU      |
|                                        | 3.63            | 1.45           |
|                                        | 25              | 25             |

Table 8. Summary of CoG measurements from this paper (first row) and others in the literature.

Our reporting of the data focuses primarily on spectral CoG, a measure which enables broad comparison to past results, as summarized above. We also explored other ways to summarize fricative spectra including additional spectral moments, such as skew and kurtosis, formant values, and duration. In exploring these additional phonetic measures, we found that many of them were correlated. For example, skew and kurtosis, the third and fourth spectral moments, were closely correlated with the mean energy (CoG) and variance of the spectrum, the first and second spectral moments. From the plots in Figures 8, 10, 12, 14, it is also possible to observe a positive correlation between spectrum CoG and spectrum SD. This is not a general finding for fricatives but follows from the typical properties of velar and labiodental fricatives (see also Section 1). To facilitate additional exploration of the data, including cross-linguistic comparison, we have submitted our entire data set, including additional measurements as well as sound files and text grids as a Data in Brief article.
In the remainder of the paper, we discuss how some of the observations above might be related. In particular, we explore sources of variability for [f] and for [x] and how these might contribute to the patterns of non-coronal fricative merger that we also observe in Zhongjiang and other varieties. Section 5.2 elaborates on sources of variability. Section 5.3 hypothesizes the pathway to sound change in Zhongjiang. Section 5.4 considers labial-coronal mergers more broadly.

5.2 *x → *f merger in Zhongjiang: a pathway of change

As mentioned in the introduction, labial-velar fricative mergers are common and have attracted the attention of many scholars. In this section, we consider accounts of labial-velar mergers in light of the new phonetic facts uncovered in our study. In particular, we seek to link patterns of phonetic variation to patterns of merger, on the assumption that the phonetic variation exemplified by Zhongjiang is characteristic of Type II merger varieties.

One hypothesis raised for labial-velar merger within South Chinese languages points to labialization as a driver of the merger (Wan 1998/2009, Xie 2003, Zhuang 2004/2016, Xiang 2005, Sun 2007, Ye 2008). Zhuang (2004)’s proposal is that production of [(w) overlaps in time with [x] such that the lower lip gesture for [(w] comes close to the upper teeth, yielding the percept of a labial-dental fricative [f]. Zhuang (2017) further clarifies the proposal, arguing specifically that a sequence of [x] followed by [(w] is not a consonant cluster; rather, the velar fricative is labialized (described as ‘labialized aspect of the velar’ (软腭音的形容性唇化成分)), resulting in a complex segment, at least in South Chinese language varieties (see also footnote 1). A variation on this account, proposed by Wan (1998/2009: 191), posits an intermediate step whereby [xw], a complex segment, becomes [(f] on the way to becoming [f]. Whether the [(f] stage is sufficiently stable or not, this broad approach, treating labialization as a driver of merger, can account for why the *x → [f] change only happens in labial environments: [u] (Type I, table 1), [w] (Type VIII, table 1), or both [w] and [u] (Type II, table 1). The account is also consistent with claims about the temporal basis of complex segments (Shaw et al., 2019). Shaw et al. (2019) propose that complex segments, e.g., segments involving multiple gestures, including secondary articulations, differ from segment sequences in how the gestures are coordinated in time. Specifically, complex segments are proposed to have gestures coordinated according to gestural onsets. On this hypothesis, the labial component of [w] would begin early in the segment [xw], which may contribute to the [f] percept, as proposed by (Zhuang 2004).

This hypothesis is consistent with some aspects of our data. For example, we found no difference between *xw and *f at early time points in the fricative, t1 and t2. However, this lack of difference could also be due to the fairly late stage of the Zhongjiang merger. That is, synchronically, there may be no real difference between *xw and *f, with both being realized as [f]. However, this account does not link the merger to the phonetic observations. Specifically, we observed substantial variation in CoG for [f], regardless of whether the words come from *f or *xw. Additionally, we observed that the CoG for *x is substantially
lower than other varieties, another fact which is not integrated into this account of the merger.

In the remainder of this sub-section, we sketch a pathway to change that links the pattern of phonetic variation that we have observed to the Type II merger.

Our proposed pathway to change decomposes the diachronic process into five steps, which are summarized in Figure 15. The dimensions of each box represent a fricative space defined by CoG (x-axis) and spectrum dispersion (y-axis), as per Figures 8, 10, 12, and 14 of our results. In the first step (top, left panel) we assume that \( *f, *x, *x^w \), the three relevant proto-categories, are all produced with similar ranges of variation. We also assume that the CoG for \( *x^w \) is slightly higher than for \( *x \), owing to coarticulation with the labial glide. This part of the account is consistent with Zhuang (2004), as summarized above. The next step is characterized by increased variability for \( *f \), which causes this category to encroach on \( *x \). Since \( *x^w \) has higher CoG than \( *x \), increased variability from \( *f \) first threatens \( *x^w \). In the Type II scenario, \( *x^w \) mergers with \( *f \). This is shown as a partial merger in the top right panel. In the next proposed step, \( *x \) differentiates from \( *x^w/\ f \) by lowering CoG (and spectrum dispersion). We surmise that this lowering helps to stabilize a new system in which \( *x^w \) and \( *f \) are completely merged.

---

**Figure 15.** Proposed pathway to change for \( *x^w \rightarrow [f] \) merger in Zhongjiang. Each panel represents a diachronic stage showing three proto-fricative categories in phonetic space, defined by spectrum CoG, x-axis, and spectrum SD, y-axis.

Our proposal, summarized in Figure 15, builds on past work, while integrating the findings from this study, namely, the variability of the synchronic \([f]\) category and the low CoG of \([x]\). The account remains agnostic about the source of variability for \([f]\) which contributes to the merger, an issue we take up in the next subsection.
5.3 Sources of variability for [f] and [x]

Our account of the Type II merger adopts the hypothesis that *x* was phonetically more similar to *f* than to *x* (in other contexts), owing to labialization by the [w] component. We therefore assume that coarticulation impacts the phonetics of the velar fricative in ways that contribute to merger and, specifically to the *w* environment leading the change. Another important aspect of the proposal is the variability of *f*. In this subsection, we discuss some sources of variability for *f*.

5.3.1 Fricative weakening

For starters *f* may be intrinsically variable. Acoustically, since the constriction is in the anterior portion of the vocal tract, there is little or no cavity in front of the constriction to provide stable resonance. Additionally, the sound source may depend on the dental status of the speaker, i.e., whether they are missing teeth, have gaps between them, etc. Failing to generate turbulence at the teeth or lips would have substantial impact on the spectrum, presumably lowering the amplitude of low frequency signal components. There are proposals that treat labial-velar mergers as relatively standard instances of lenition, i.e., weakening of *f*. For example, He (2004: 123) notes that [f] is sometimes produced weakly, or reduced, to a bilabial fricative, [ɸ], in some Chinese language varieties. Others have assumed a diachronic progression from [f] to [h] to [∅], initiated by weakening of [f] to [ɸ] (Pellegrini 1980: 69; Jungemann 1955: 142). Variation along this progression could be perceived as [x], contributing to labial-velar merger in some cases. These approaches have in common with Wan (1998/2009:191), described below, the progression of change passing through [ɸ].

5.3.2 Vowel coarticulation as a source of variability

Additionally, the spectrum for *f* may be sensitive to variation in tongue body position, conditioned by coarticulation with the following vowel. Just as the labial component of [w] may influence the phonetics of [x], raising CoG, the context following [f] may influence its phonetic realization. Specifically, vowels with a posterior constriction, could potentially lower the CoG of [f]. Our analysis of the full corpus (Table 7) showed a significant effect of rime, even at the earliest onset of the fricative, t1, although the effect was strongest towards the end of the fricative. This result suggests that the range of variation for [f] is indeed conditioned, at least in part, by the following rime. An extreme case of rime-conditioned CoG lowering comes in the [ɔŋ] environment. In this environment, the CoG of [f] lowered to the point that it overlapped [x] substantially, with some speakers showing near complete merger (Figure 14). However, when we controlled for rime, zeroing in on a single non-posterior rime environment, [ɔɯ], in Study3, we still observed a wide range of CoG variation for *f* (Figure 12). In particular, speakers 4, 6, 8, produced a similar range of CoG variation for *f* in Study 3 (Figure 12) as they did in Study 1 and 2, which featured a wide range of rime environments. This indicates that, although rime is indeed a significant contributor to CoG, rime cannot explain the entire range of variation. Specifically, we still observe low CoG tokens of [f], even in rime environments that do not involve a posterior constriction.
5.3.3 Timing variability of secondary articulation as a source of variability

Another possibility is that [f] itself is variably velarized, i.e., as [fˠ]. A secondary velar constriction, i.e., a narrowing of the vocal tract in the region of the soft palate, would account for the low CoG observed for some tokens, even in vowel environments that do not involve a posterior constriction. The continuous variation ranging from low CoG values, characteristic of [x], to high CoG values, characteristic of [f], may result from variation in the timing and magnitude of two component gestures of [fˠ], a raising of the lower lip for the [f] component and a raising of the tongue body to the soft palate for the [ˠ] component. A secondary velar articulation could explain why we variably observe low CoG values for *f, even in Study 3.

If *f contains a secondary velar constriction, i.e., [fˠ], then it is a complex segment, similar in composition to a labialized velar, [xʷ]. Both [fˠ] and [xʷ] contain two constrictions. One constriction is formed by the tongue dorsum at the soft palate. This is the constriction that gives rise to turbulent airflow, the [x] component [xʷ]; the [ˠ] component of [fˠ]. The resonance of turbulent energy in the relatively long cavity in front of the constriction contributed to low CoG. The second constriction involves the lips, the [w] component of [xʷ] and the [f] component of [fˠ]. Notably, the labial constriction location is anterior to the velar constriction. Variation in the timing and magnitude of these two constriction gestures could explain the range of CoG values for [xʷ] and why they closely overlap with the range of values observed for [fˠ]. This account is very similar to what we expect from vowel-based anticipatory coarticulation, except that [fˠ] predicts variably lowered CoG even in the environment of vowels that do not have a posterior constriction, as in Study 3. Although Chinese language varieties are typically analyzed as having secondary articulations (see, e.g., Chao 1934; Duanmu 2007; see also footnote 1 and discussion in section 5.4), composed of an obstruent and glide component, the specific proposal of a secondary velar articulation for labiodentals is novel, as far as we are aware.

5.3.4 Sociophonetic sources of variation

The Zhongjiang variety is a “dialect island”, a Type II language in a sea of Type I languages (see Figure 1). All speakers in our sample have at least some contact with both other language varieties in Southwest China, such as the Chengdu variety (Type I), the language variety of the provincial (Sichuan) capital, a major urban area, and a prestige variety from the perspective of Zhongjiang speakers. Our participants also have some exposure to standard Mandarin, at least through mainstream Chinese media. Within Zhongjiang City, the production of *xʷ as [f] is recognized as a local language feature that differs from both Chengdu Chinese and Standard Mandarin. It is possible that the [f] variant that we have proposed as a possible explanation of observed variation is a variant used by Zhongjiang speakers to approximate more standard varieties.

One of our Zhongjiang speakers, S5, shows only a narrow range of acoustic variation for *f. Across studies, S5 shows two fairly distinct fricative clusters within the CoG-SD acoustic space. In study 1, which focuses on minimal pairs, S5 maintains clear separation between [f] and [x]. In study 2, which looks at mergers to [f], the majority of S5’s tokens have high CoG
and SD. This contrasts with the dominant pattern, which shows variation overlapping with canonical [x]. Interestingly, S5 shows the typical Type II merger, *xwx is produced as [f]. Even though this speaker is in the minority, by showing less variability for [f], possibly using [f] instead of [β], there is a sense in which S5 could be considered the most canonical representative of a Type II language in our sample.

Another speaker whose production patterns fell outside of the main pattern was S2. This speaker’s productions are also interesting from the standpoint of the sociophonetics of language contact. This speaker had the range of variation for *f in Study 3 that motivated us to posit a secondary velar articulation, [fˠ]. However, the distribution of phonetic categories across words did not correspond to our expectations for a Type II merger. By and large, the distribution of variants for S2 corresponds more closely to Standard Mandarin. Study 1 showed that, for this speaker, *xw maintained low CoG, just like *x. This is the pattern expected of Standard Mandarin. However, Study 2 showed that the *f words are produced with some lowered CoG values, possibly due to a [fˠ] variant or fricative weakening. Possibly, S2 has generalized the [fˠ] realization, which developed in Zhongjiang from *xw, such that, synchronically, [fˠ] is used very generally for [f], including even for *f.

The speakers that deviate from the dominant phonetic patterns within our Zhongjiang sample (S02, S05), do so in ways that can be linked transparently to language contact; these speakers showed characteristics of prestige varieties (Chengdu dialect, Standard Mandarin).

To summarize, with the data that we have, we are not able to tease apart the possible sources of CoG variability for [f] enumerated above and these various sources of CoG are not mutually exclusive. Fricative weakening may coincide with rime coarticulation, making it difficult to ascertain the specific articulatory contributions to spectral variation. If the phonetic influence of back vowels on [f] generalizes to other contexts (non-front vowels), such as those in Study 3, it would support our proposal for a secondary velar articulation, but this account also does not preclude velar weakening as an independent source of variation. Finally, these sources of variation may be socially evaluated by speakers in ways that drive change towards perceived social prestige.

5.4 Labial-velar mergers more broadly

Although our study focuses on one particular language variety dialect, the results have implications for labial-velar mergers more broadly.

We have shown that [f] in Zhongjiang shows substantial spectral variability and argued that this contributed to the labial-velar merger. Although many of the factors that we mentioned as possibly contributing to [f] variability are relevant for other language varieties as well, not all language varieties show the same degree of [f] variability. This was illustrated in Table 8, which shows that, in contrast to Zhongjiang, Standard Mandarin as reported by Svantesson (1986) and Ran (2008) report greater CoG variability for [x] than for [f].

The aerodynamic requirements for velar and labiodental fricatives to maintain perceptual
distinctiveness may make them particularly susceptible to merger across languages. Generating turbulence from a supralaryngeal constriction requires precise balance between the amount of air flow volume and the diameter of the constriction (e.g., Shadle, 1991). In the case of [xʷ], small variation in constriction degree could make the difference between generating turbulent energy at the tongue dorsum or not. Even in the absence of turbulent energy at the tongue dorsum, turbulent energy could still be generated at a more anterior location, the lips or teeth. For [xʷ], the articulatory conditions for a small reduction in tongue dorsum height to yield qualitatively different acoustic outcomes are met. Lenition of velars, quantified as a reduction in tongue dorsum constriction degree, is particularly common for a number of reasons (for a recent discussion see, e.g., Shaw et al., 2020). In the case of [xʷ], small reductions in velar constriction naturally give rise to acoustic conditions similar to [f] or [fˠ]. That is, in these cases, normal degrees of spatial-temporal variation in articulation may have a disproportionate impact on the acoustics. Velar fricatives combined with a labial constriction may be anti-stable from the standpoint of articulatory-acoustic correspondence. Non-linearity in the mapping between articulation and acoustics has been proposed as a basis for stable phonetic categories, i.e., the quantal regions of Stevens (1989). Weak fricatives with overlapping velar and labial components may be anti-stable in that a small amount of articulatory variation could give rise to relatively large acoustic consequences. Such anti-stability may contribute to bidirectional sound change, regardless of whether the gestures are components of a complex segment or overlapping segments.

Although [ʷ] is indeed a common environment for labial-velar merger in Chinese languages, there are also cases in which sound change proceeds in the opposite direction. That is, *f changes to [xʷ], as in the Type III/V/VII varieties in Table 1. Notably, this direction of change does not follow from labialization of a velar fricative, as in Zhuang (2004, 2016, 2017), because this requires the seemingly spontaneous emergence of a velar gesture, an observation also made by Sun (2007). If there is a secondary velar component to *f in some varieties, a possibility which we raise in section 5.3.3, the change is less mysterious. In the presence of a secondary velar articulation, we can understand changes in both directions, [xʷ]→[fˠ] as well as [fˠ]→[xʷ] as following from the same mechanism, namely, variation in the relative timing between the component gestures of complex segments. However, it is also possible that the general anti-stability of these non-coronal fricatives in conjunction with other sources of variability may be sufficient to trigger bi-directional merger. More specifically, listeners could conceivably interpret variation in the labiodental fricative as deriving from [xʷ], even in the absence of a secondary velar articulation on the labiodental.

The types of patterns observed in Chinese languages are representative of broader cross-linguistic patterns as well. For example, round vowels are typical environments for labial-velar merger in the Spanish of Chicano children and adults (Greenlee 1992). In earlier stages of Spanish, the [w] environment supported resistance to lenition of [f] (to [h]) (Lass & Anderson 1975). In these cases, the labial feature of the [w] is viewed as reinforcing or “strengthening” the labial component of the [f] (Hickey 1984). In Sentani, as described by Cowan (1965), patterns of assimilation conditioned by [w] reveal temporal spreading of both labial and dorsal components; [w] conditions assimilation of an adjacent alveolar nasal, [n],
to a velar nasal, [ŋ], and conditions assimilation of an adjacent glottal fricative to a labial fricative (Ohala and Lorentz 1977). Thus, [h] before [w] is optionally realized as [f] or [ɸ]. These patterns suggest that the types of labial-velar mergers observed across Chinese languages derive from a general mechanism of coordinating labial and dorsal gestures in time, whether as a part of a complex segment or as a sequence of segments.

Although labial-velar mergers are not unique to the languages of China, synchronic phonotactic restrictions within Chinese languages may also contribute to the outcomes. We have discussed at length the role of the labial-velar glide, [w], in conditioning variation and change in labial-velar mergers. Many Chinese varieties exhibit phonotactic restrictions on how [w] can combine with onset consonants. Duanmu (2007) discusses the situation in Standard Mandarin. There are 18 onset consonants and three glides. Free combination of the onset consonants and glides would yield 54 possibilities, of which 29 are found. One of the missing combinations is [f w]. This combination is missing as well in Zhongjiang and many other varieties in Southwest China. Duanmu (2007) mentions that in Standard Mandarin there is an exception to the general absence of [f w], a single word [f w o] ‘buddha’; but this exception is absent in Zhongjiang (and other varieties), where ‘buddha’ is pronounced [fu]. Duanmu (2007) argues convincingly that CG sequences in Standard Mandarin are single sounds, i.e., complex segments, as opposed to segment sequences. On this account, the attested CG gaps, including [f w], follow in part from the constraint that an articulator can only be specified once per segment. Conflicting labial specifications for [f] and [w] rule out [f w]. The phonotactic constraint against [f w] may encourage [x w] (and maybe even [fˠ]) as outcomes of sound change in Chinese language varieties.

The prevalent ban on [f w] across Chinese languages juxtaposes with [fu] as a frequent outcome of sound change, including cases of *xu → [fu] (Table 1). There are also cases in which *fu changes to [xu], but across the survey of merger types (Figure 1), [fu] is a much more common outcome. Out of 212 documented varieties with a labial-velar merger, 184 (~87%) have resulted in [fu]. The prevalence of [fu] (c.f., [xu]) as the outcome labial-velar sound changes has encouraged some speculation about phonological/articulatory bases. For example, Li (1995) proposed that both [f] and [u] having labial features might contribute to the merger of [xu]→[fu]. If Li’s proposal is correct, then it suggests an interesting dichotomy; consecutive labial specifications are preferred across CV sequences but dispreferred across CG.

This dichotomy between CV and CG, with respect to consecutive labial specifications may have a structural basis, as proposed in Duanmu (2007), as well as a temporal basis that is specific to tone languages. In CV sequences, lexical tone languages are known to differ from non-tone languages in that there is an increased temporal lag between the consonant and the vowel (Gao 2009; Hu 2016; Karlin & Tilsen 2015; Karlin 2018; Zhang et al. 2019; Geissler et al. 2020). In contrast, complex segments are heavily overlapped in time (e.g., Catford 2001: 103; Shaw et al. 2019). Possibly, CV sequences in Mandarin allow different specifications of identical articulators for the C and V gestures because these gestures are temporally separated in time. Given the differences in temporal organization (between CV
and CG in tone languages), the ban on multiple labial specification would follow from a
temporal version of the Obligatory Contour Principle (OCP), such as that proposed in Gafos
(2002). This temporal basis for the differences between CV and CG as domains for multiple
labial gestures parallels the structural basis for the distinction proposed by Duanmu (2007). If
the temporal difference between CV and CG is of direct relevance to the phonotactics, we
might expect different patterns in non-tonal languages, which tend to show greater overlap
between consonant and vowel gestures in CV sequences.

5 Conclusion

A phonetic study of non-coronal fricatives in Zhongjiang Chinese, a “dialect island” showing
contextually dependent labial-velar fricative mergers, revealed: (1) that [f] exhibits a wide
range of spectral variation; (2) that the spectral CoG of [f] is lowered in the environment of
back vowels, particularly [on]; and (3) that [x] shows particularly low spectral CoG.

We interpret these results in the context of the Zhongjiang labial-velar merger, one of eight
common labial-velar merger types in Southwest Chinese language varieties. We propose a
pathway to change triggered by spectral variability in [f], which encroaches on [x],
particularly when labialized, e.g., [xʷ], leading to merger of [xʷ] and [f]. We interpret the
low CoG of [x] as resulting from adaptive dispersion, which, following merger of [xʷ] and [f],
affectively facilitates contrast between [x] and [f] in other environments. When [f] shows
spectral variability, as in Zhongjiang, gestural overlap between the labial and velar
components of [xʷ] cause this environment to be the first to merge with [f].

Various factors could contribute to the observed pattern of spectral variability for [f], which
we argue instigates labial velar merger in Zhongjiang. These include fricative weakening, due
to factors intrinsic to [f], vowel coarticulation, sociophonetic factors related to language
contact and the possibility of a secondary velar articulation, i.e., [fˠ], which could lower
spectral CoG even in the context of non-back vowels, a pattern also observed in our data.

Acknowledgements

This research was supported by “Young Scholars” program of Xihua University (Chengdu,
China). The authors thank Rikker Dockum for help with the forced alignment, and thank the
three anonymous reviewers, Jeff Mielke, the Associate Editor, and the Yale Phonologroup,
particularly Mark Tiede, Natalie Weber, and Christopher Geissler for their comments on this
paper.

References

Boersma, P., and Weenick, D. 2016. “Praat (Version 6.0.23)”. University of Amsterdam:
Institute of Phonetic Science.
Cao, S. J. 1997. The history of immigration in China (5). 中国移民史. Fuzhou: Fujian People
Press.
Catford, J. C. 2001. A practical introduction to phonetics (2nd Edn). Oxford: Oxford
University Press.

Chao, Y. R. 1934. The non-uniqueness of phonemic solutions of phonetic systems, Bulletin of the Institute of History and Philology, Academia Sinica, 4.4: 363-97, Reprinted in Marin Joos (ed.), 1957. Readings in Linguistic I. Chicago: University of Chicago Press.

Chao, Y. R. et al. 1948. The report of Hubei dialects. 湖北方言调查报告. Taipei: The Institute of History and Philology.

Chen, Z. T., and Li, R. L. 1991. The study of Min Chinese, 闽方言研究. Peking: Yuwen Press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... and Fox, J. 2019. lme4 package: linear mixed effects models using 'Eigen' and S4. R package version, 1.1-21.

Cowe, H. K. J. 1965. Grammar of the Sentani language. The Hague: Martinus Nijhoff.

Cui, R. C. 1996. The Xiang dialect in Sichuan, 四川境内的湘方言, Taipei: The Institute of History and Philology.

Duanmu, S. 2007. The phonology of Standard Chinese (2nd edition). Oxford: Oxford University Press.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. 1988. Statistical analysis of word-initial voiceless obstruents: preliminary data, The Journal of the Acoustical Society of America. 84, 115–124.

Gafos, A. 2002. A grammar of gestural coordination. Natural Language and Linguistic Theory. 20, 269-337.

Gao, M. 2009. Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. Chinese Journal of Phonetics. 2, 43–50.

Geissler, C., Shaw, J., Hu, F., and Tiede, M. 2020. Consistent C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. Presented at the 94th Annual Meeting of the Linguistic Society of America.

Goldstein, L., Browman, C. P. 1986. Representation of voicing contrasts using articulatory gestures. Journal of Phonetics. 14, 339-342.

Gordon, M., Barthmaier, P., and Sands, K. 2002. A cross-linguistic acoustic study of voiceless fricatives. Journal of the International Phonetic Association. 32(2), 141-174.

Greenlee, M. 1992. Perception and production of voiceless Spanish fricatives by Chicano children and adults. Language and Speech. 35(1, 2), 173-187.

He, D. A. 2004. Regulations and directions: phonological structure in change. 规律与方向:变化中的音韵结构, Peking: Peking University Press.

Hickey. R. 1984. On the nature of labial velar shift. Journal of Phonetics. 1984(12), 345-354.

Hu, F. 2016. Tones are not abstract autosegmentals. Proceedings of Speech Prosody, 302-306.

Jongman, A., Wayland, R., and Wong, S. 2000. Acoustic characteristics of English fricatives. The Journal of the Acoustical Society of America. 108(3), 1252-1263.

Karlin, R. 2014. The articulatory TBU: Gestural coordination of lexical tone in Thai. Cornell Working Papers in Phonetics and Phonology. 124-149.

Karlin, R. P. 2018. Towards an Articulatory Model of Tone: A Cross-Linguistic Investigation. Ithaca, NY: Cornell University.
Lass, R., and Anderson, J. 1975. *Old English Phonology*. Cambridge: Cambridge University Press.

Li, L. 1995. *The Study of comparison between consonant initials and tones in Southwest Mandarin*. 西南官话中的声母与调类比较研究. Ph. D dissertation. Chinese Academy of Social Sciences.

Li, R. L. 1999. *The report of Hakka Chinese in west Guangdong province*. 粤西客家话调查报告. Guangzhou: Jinan University Press.

Liljencrants, Johan, and Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839-62.

McMurray, B., and Jongman, A. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118(2), 219.

Norlin, K. 1983. Acoustic analysis of fricatives in Cairo Arabic. Working papers, *Phonetics Laboratory*, Lund, 25.

Ohala, J., and Lorentz, J. 1977. The story of /w/: an exercise in the phonetic explanation for sound patterns, *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, 577-599.

Pellegrini, G. B. 1980. Substrata. In *Trends in Romance linguistics and philology. Volume 1: Romance comparative and historical linguistics* (Posner and Green, Eds). The Hague: Mouton.

Ran, Q. B. 2008. *The phenomena and features of consonants: experiments study based on plosives, fricatives and affricates of Standard Mandarin*. 辅音现象与辅音特性：基于普通话的汉语阻塞辅音实验研究. Tianjin: Nankai University Press.

Rao, D., Shaw, J.A., Dockum, R. (submitted). Data on acoustic phonetic properties of non-coronal fricatives in monosyllabic words of Zhongjiang Chinese. *Data in Brief*.

Rao, D., Dockum, R., Shaw, J.A. 2020. Acoustic recordings of Zhongjiang Chinese non-coronal fricatives. *Mendeley Data*, V1, doi: 10.17632/mptnd5p8db.1

Roman J., Fant, C. G. M., and Halle, M. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge: Acoustics Laboratory, Massachusetts Institute of Technology.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of phonetics*, 19(3-4), 409-424.

Shadle, C. H., and Mair, S. J. 1996. Quantifying spectral characteristics of fricatives. Proc. Int. Conf. Spoken Lang. Proc. (ICSLP), 1521–1524.

Shaw, J. A., and Chen, W. R. 2019. Spatially conditioned speech timing: evidence and implications. *Frontiers in Psychology* (10) 2726.

Shaw, J. A., Carignan, C., Agostini, T., Mailhammer, R., Harvey, M., and Derrick, D. (to appear). Phonological contrast and phonetic variation: the case of velars in Iwaidja. *Language*, 1-38.

Shaw, J. A., Durvasula, K., and Kochetov, A. 2019. The temporal basis of complex segments. In Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019 (pp. 38-42.
Sun, Y. Z. 2007. *The study of phonetics in Gan Chinese of Jiangxi*. 江西赣方言语音研究. Peking: Yuwen Press.

Stanford, J. N., & Evans, J. P. (2012). The influence of Mandarin Chinese on minority languages in rural southwest China: a sociolinguistic study of tones in contact. *International journal of the sociology of language*, 2012(215), 79-100.

Stevens, K. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

Stevens, K. 1998. *Acoustic Phonetics*. First MIT Press paperback edition, Cambridge: The MIT press.

Svantesson, Jan-Olof. 1986. Acoustic analysis of Chinese Fricatives and Affricatives. *Journal of Chinese Linguistics*, 14(1), 53-70.

Wan, B. 1998/2009. *The study of the historical level of the initial consonants in Gan Chinese*. 赣方言声母的历史层次研究. Peking: Commercial Press.

Wu, B. 2020. A study of acoustic features and variations of the fricatives in Jianghuai Mandarin. *Bulletin of Linguistics Studies*, 25, 329-343.

Wu, Y. W. 1995. *Cooccurrence restrictions in Mandarin Chinese*. Ph.D. dissertation (Chapter 4). University of Toronto.

Xiang, M. B. 2005. *The phonology of Hakka Chinese*. 客家话音系. Peking University.

Xie, L. W. 2003. *The study of phonetic Hakka Chinese*. 客家话语音研究. Peking: Chinese Social Science Press.

Yang, S. F. 1969. *The investigation report of Yunnan dialects*. 云南方言调查报告. Taipei: the Institute of History and Philology.

Yang, S. F. 1974. *The investigation report of Hunan dialects*. 湖南方言调查报告. Taipei: the Institute of History and Philology.

Yang, S. F. 1984. *The investigation report of Sichuan dialects*. 四川方言调查报告. Taipei: the Institute of History and Philology.

Ye, X. F. 2008. The phenomenon of [f]-/x/ merger in Chinese Dialect. 汉语方言中的[f]-/x/混读现象. *East Linguistics*, (4), Shanghai: Shanghai Educational Press.

Zhan, B. H. 2002. *An introduction of Cantonese in Guangdong*. 广东粤方言概要. Guangzhou: Jinan University Press.

Zhang, S. Z. 1995. The influence of Mandarin Chinese on the phonetics of Chinese dialects. 试论普通话对方言语音的影响.*Applied Linguistics*, (16), 94-98.

Zhang, M., Geissler, C., and Shaw, J. A. 2019. “Gestural representations of tone in Mandarin: evidence from timing alterations,” in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren, (Canberra, ACT: Australasian Speech Science and Technology Association Inc.), 1803–1807.

Zhuang, C. S. 2004. *A study of phonology of the vernacular dialect of north Guangdong*. 粤北土话音韵研究. Peking: Chinese Social Science Press.
Zhuang, C. S. 2016. The labialized velar initial consonants in Hakka and other Southeast Chinese Dialects. *Fangyan (Dialect)*, 38 (2), 158-168.

Zhuang, C. S. 2017. Further investigation of the motivation for labialization of Xiao-group characters with closing medial in Southern Chinese Dialects. *Chinese Language*, 378, 332-338.

## Appendix

| Syllable | Simplified Chinese Characters | Pinyin (standard romanized orthography) | Standard Mandarin IPA¹ | Type II expectation Zhongjiang IPA |
|----------|-------------------------------|----------------------------------------|------------------------|-----------------------------------|
| syllable01 | 法 | fa²¹⁴ | [fa²¹⁴] | [fa³¹] |
| syllable02 | 发 (发财) | fa⁵⁵ | [fa⁵⁵] | [fa³¹] |
| syllable03 | 付 | fu⁵¹ | [fu⁵¹] | [fu²²⁴] |
| syllable04 | 喝 | he⁵⁵ | [xe⁵⁵] | [xe⁴⁵] |
| syllable05 | 画 | hua⁵¹ | [xʷa⁵¹] | [fa²²⁴] |
| syllable06 | 肥 | fei³⁵ | [fei³⁵] | [fei³¹] |
| syllable07 | 黑 | hei⁵⁵ | [xei⁵⁵] | [xe³¹] |
| syllable08 | 后 | hou⁵¹ | [xu⁵¹] | [xʊə²²⁴] |
| syllable09 | 放 | fang⁵¹ | [fan⁵¹] | [fan³²⁴] |
| syllable10 | 黄 | huang³⁵ | [xʷaŋ³⁵] | [fan³¹] |
| syllable11 | 红 | hong³⁵ | [xʊŋ³⁵] | [xʊŋ³¹] |
| syllable12 | 烦 | fan³⁵ | [fan³⁵] | [fei³¹] |
| syllable13 | 混 | hun⁵¹ | [xʷən⁴¹] | [fan³²⁴] |
| syllable14 | 分 | fen⁵⁵ | [fən³⁵] | [fən⁴⁵] |

¹ The IPA used for Standard Mandarin follows Duanmu (2007).
| Syllable | Character | Pinyin | Traditional | Simplified |
|----------|-----------|--------|-------------|------------|
| 15       | 飞        | fei55  | [fei55]     | [fei45]    |
| 16       | 怀        | hua135 | [xwa135]    | [fai131]   |
| 17       | 范        | fan51  | [fan51]     | [fæ324]    |
| 18       | 或        | huo51  | [xwo51]     | [fè31]     |
| 19       | 还（还有）| hai35  | [xai35]     | [xai11]    |
| 20       | 华        | hua35  | [xwa35]     | [fà31]     |
| 21       | 罚        | fa35   | [fa35]      | [fà31]     |
| 22       | 还（还钱）| huan35 | [xwan35]    | [fè31]     |
| 23       | 否        | fou214 | [fsu214]    | [fou51]    |
| 24       | 方        | fang55 | [fan55]     | [fan45]    |
| 25       | 何        | he35   | [x35]       | [xo31]     |
| 26       | 慌        | huang55| [xañ55]     | [fan45]    |
| 27       | 洪        | hong35 | [xuñ35]     | [xun31]    |
| 28       | 回        | huei35 | [xuei35]    | [fèi31]    |
| 29       | 服        | fu35   | [fu35]      | [fù31]     |
| 30       | 核        | he35   | [x35]       | [xe31]     |
| 31       | 厚        | hou51  | [xou51]     | [xou324]   |
| 32       | 害        | hai51  | [xai51]     | [xai324]   |
| 33       | 划        | hua135/hua35 | [xwa135]/[xwa35] | [fà324]/[fa31] |
| 34       | 伐        | fa35   | [fa35]      | [fà31]     |
| 35       | 合        | he35   | [x35]       | [xo31]     |
| 36       | 房        | fang35 | [fan31]     | [fan31]    |
| 37       | 粪        | fen51  | [fèn51]     | [fèn324]   |
| 38       | 灰        | huei35 | [xuei35]    | [fei45]    |
| Syllable | Character | Pinyin | Tones | Mandarin | Application |
|----------|-----------|--------|-------|----------|-------------|
| syllable39 | 冯 | feng³⁵ | [fəŋ³⁵] | [fəŋ³¹] |  |
| syllable40 | 号 | hao⁵¹ | [hau⁵¹] | [xau²⁴] |  |
| syllable41 | 返 | fan²¹⁴ | [fan²¹⁴] | [fə̆⁵¹] |  |
| syllable42 | 叩 | hou²¹⁴ | [xəu²¹⁴] | [xəu⁵¹] |  |
| syllable43 | 复 | fu⁵¹ | [fu⁵¹] | [fu³¹] |  |
| syllable44 | 盒 | he³⁵ | [xə³⁵] | [xo³¹] |  |
| syllable45 | 昏 | hun⁵⁵ | [xə̆n⁵⁵] | [fən⁴⁵] |  |
| syllable46 | 费 | fei⁵¹ | [fei⁵¹] | [fei3²⁴] |  |
| syllable47 | 疲 | fa³⁵ | [fa³⁵] | [fa³¹] |  |
| syllable48 | 浮 | fu³⁵ | [fu³⁵] | [fəu³¹]/[fu³¹] |  |
| syllable49 | 皇 | huang³⁵ | [xuaŋ³⁵] | [fən³¹] |  |
| syllable50 | 宏 | hong³⁵ | [xəŋ³⁵] | [xəŋ³¹] |  |
| syllable51 | 话 | hua⁵¹ | [xə⁵¹] | [fa³²⁴] |  |
| syllable52 | 翻 | fan⁵⁵ | [fan⁵⁵] | [fə̆⁵⁵] |  |
| syllable53 | 恢 | hui⁵⁵ | [xəei⁵⁵] | [fei⁵⁵] |  |
| syllable54 | 婚 | hun⁵⁵ | [xə̆n⁵⁵] | [fən⁵⁵] |  |
| syllable55 | 废 | fei⁵¹ | [fei⁵¹] | [fei3²⁴] |  |
| syllable56 | 府 | fu²¹⁴ | [fu²¹⁴] | [fu⁵¹] |  |
| syllable57 | 候 | hou⁵¹ | [xəu⁵¹] | [xəu³²⁴] |  |
| syllable58 | 花 | hua⁵⁵ | [xə̆a⁵⁵] | [fa⁴⁵] |  |
| syllable59 | 缝 (缝纫) | feng³⁵ | [fəŋ³¹] | [fəŋ³¹] |  |
| syllable60 | 欢 | huan⁵⁵ | [xə̆n⁵⁵] | [fə̆⁴⁵] |  |
| syllable61 | 坟 | fen³⁵ | [fəŋ³⁵] | [fəŋ³¹] |  |
| syllable62 | 鸿 | hong³⁵ | [xuaŋ³⁵] | [xəŋ³¹] |  |
| Syllable | Character | Pinyin | Tones |
|----------|-----------|--------|-------|
| 63       | 芳        | fāng⁵⁵ | [fəŋ⁵⁵] | [fəŋ⁵⁵] |
| 64       | 父        | fù¹¹   | [fu¹¹] | [fu³²⁴] |
| 65       | 晃        | huáng⁵¹ | [xwəŋ⁵¹] | [fəŋ⁵¹] |
| 66       | 封        | fēng⁵⁵ | [fuŋ⁵⁵] | [fəŋ⁴⁵] |
| 67       | 豪        | hào³⁵  | [xau³⁵] | [xau³¹] |
| 68       | 荷        | hé³⁵   | [xɤ³⁵] | [xʊ³¹] |
| 69       | 肺        | fèi⁵¹  | [fei⁵¹] | [fə²³⁴] |
| 70       | 饭        | fàn⁵¹  | [fan⁵¹] | [fəⁿ³²⁴] |
| 71       | 猴        | hóu³⁵  | [xəu³⁵] | [xəu³¹] |
| 72       | 蜂        | fēng⁵⁵ | [fuŋ⁵⁵] | [fəŋ⁴⁵] |
| 73       | 续        | huàn²¹⁴ | [xwən²¹⁴] | [fəⁿ⁵¹] |
| 74       | 耗        | hào¹¹  | [xau¹¹] | [xau³²⁴] |
| 75       | 防        | fáng³⁵ | [fan³⁵] | [fəⁿ³¹] |
| 76       | 轰        | hōng⁵⁵ | [xʊŋ⁵⁵] | [xʊŋ³¹] |
| 77       | 奋        | fèn³¹  | [fən³¹] | [fən³²⁴] |
| 78       | 魂        | hún³⁵  | [xwən³⁵] | [fəⁿ³¹] |
| 79       | 粉        | fěn²¹⁴ | [fn²¹⁴] | [fəⁿ³¹] |
| 80       | 谎        | huáng²¹⁴ | [xwəŋ²¹⁴] | [fn⁵¹] |
| 81       | 辜        | hūn⁵⁵  | [xwəŋ⁵⁵] | [fəⁿ⁴⁵] |
| 82       | 汇        | huì¹¹  | [xwi⁵¹] | [fei³²⁴] |
| 83       | 环        | huàn³⁵  | [xwən³⁵] | [fəⁿ³¹] |
| 84       | 风        | fēng⁵⁵ | [fuŋ⁵⁵] | [fəŋ⁴⁵] |
| 85       | 哈        | hā³⁵   | [xa³⁵] | [xa⁴⁵] |
| 86       | 海        | hǎi²¹⁴ | [xai²¹⁴] | [xai³¹] |
| syllable  | 中文 | Pinyin | 注音 | 注音 |
|-----------|------|--------|------|------|
| syllable87 | 好 | hao²¹⁴ | [xau²¹⁴] | [xau⁵¹] |
| syllable88 | 坏 | huai⁵¹ | [x³'ai⁵¹] | [fai³²] |
| syllable89 | 会 | hui⁵¹ | [x³'ei⁵¹] | [fei³²] |
| syllable90 | 换 | huan⁵¹ | [x³'an⁵¹] | [fie³²] |