Shifts in gestural timing as the basis for non-coronal fricative mergers in Southwestern Mandarin: acoustic evidence from a dialect island

Dongmei Rao[a,b], Jason A. Shaw[b,*]

*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 dialects of Chinese, particularly varieties of Southwest Mandarin. The sound changes that lead to merger in Southwest Mandarin dialects are bidirectional: in some, /f/ becomes /x/; in others /x/ becomes /f/. We conducted a study of phonetic variation in one such dialect, Zhongjiang Chinese, which has been reported to merge /x/ to /f/ in the environment of /w/. Our results confirm this basic pattern while revealing additional nuances, including a new environment for merger, /_oŋ/, and new phonetic details. In particular, /f/ exhibits a wide range of spectral variation, including tokens with a low frequency spectral peak, characteristic of a velar constriction. We interpret the pattern of spectral variation for /f/ in the Zhongjiang dialect evidence for a secondary velar articulation, /fˠ/. This result sheds new light on bidirectional sound change, as both directions of change, /fˠ/→/xʷ/ and /xʷ/→/fˠ/, can be understood in terms of the same mechanism, shifts in the relative timing of labial and dorsal gestural components of the fricatives.

Keywords

Velar fricatives, secondary articulation, labial-velar merger, sound change, gestural timing, dialect variation

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 dialects of Southwest Chinese. In a typological survey of 374 dialects spoken in Hunan, Hubei, Sichuan, and Yunnan provinces, He (2004) reports that 212 dialects have the /f/-/x/ merger. The merger can also be found in Southern Chinese dialects, e.g., Min dialect (Chen & Li 1991), Cantonese (Zhan 2002), Gan dialect (Sun 2007), Hakka dialect in west Guangdong (Li 1999) and the vernacular dialect of north Guangdong (Zhuang 2004).

Amongst these labial-velar mergers, there are dialects 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 Chinese labial-velar mergers has been largely documentational in nature. The patterns of change, dating back to medieval Chinese, have been recorded in detail by Chinese dialectologists (see references above). In this literature, dialects 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 dialect of Southwest Mandarin (Zhongjiang). 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, since Forrest et al. (1988), to characterize fricatives within and across languages. In particular, the mean energy of the spectrum, or Center of Gravity (CoG), is often reported (e.g., Gordon, Barthmaier, & Sands, 2002), which makes it a useful measurement for comparing across studies. 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 report 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. Our main analysis in this paper focuses on two spectral measurements: Center of Gravity (CoG) and Spectrum Standard Deviation (SD). To encourage additional analyses, the entire data set, including sound files and textgrids (see methods), has been submitted as a *Data in Brief* article. 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 have relatively straight-forward interpretations. 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. For /f/, the anterior constriction at the lips typically results in a diffuse (flat) spectrum, 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 Chinese dialects 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, /_oŋ/. These contexts are not random. They all involve a lip movement and tongue dorsum retraction, albeit with various 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 shifts in the relative timing of the constituent gestures of these fricatives lead to the observed sound changes. Tongue dorsum retraction, as required for /w/, /u/, /oŋ/, during the fricative /f/, gives rise to spectral properties that approach /x/. Similarly, lip rounding, as also required for /w/, /u/, /oŋ/, 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 Chinese Dialects. Section 3 describes the methods of four studies on the Zhongjiang dialect. 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.

2. Background on the /f/-/x/ merger in Southwest Mandarin dialects

2.1 Dialect types

Several surveys of Chinese dialects, 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 dialect variation in China. These studies covered hundreds of dialects across China using standardized methods and traced synchronic pronunciation patterns back to medieval Chinese. Before presenting our studies on synchronic phonetic variation within one dialect, we first situate this dialect within the broader Southwestern Chinese sprachbund, as characterized by the seminal dialect 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 dialects of Southwest Mandarin 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 dialect type. Examples of each dialect type are given in the first column.

The table shows that synchronic /f/ derives from a merger of three medieval Chinese categories: /f/, /f^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 /ɣ/. We describe these are proto-x, or *x. For reference, the first row of the table lists Standard Mandarin Chinese. In this dialect, there is no synchronic merger between /f/ and /x/. Other rows show different types of merger between /f/ and /x/ according to context. The dialects are divided into eight types. In discussing the dialects, 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 slashes, //. To facilitate comparison across language varieties, in addition to providing English glosses for words, we also provide Sino-graphs (Chinese characters).

In Type I dialects, /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 /xu/ 护 'protect' is pronounced as /fu/ in Zhongjiang, which is the same pronunciation as other words, such as 父 'father'. In Type II, /xu/ is also pronounced as /fu/, just as in Type I dialects. Additionally, in Type II, /x^w_/ is read as /f_/. Here, we use the underscore "__" to refer to indicate all following environments, except for /u/ and /oŋ/, which are listed in separate columns. Thus, in Type II, a word like /x^wæ̃/ (欢 'happiness') is pronounced as /fæ̃/, which becomes homophonous with 翻 'turn over'. /f/-/x/ remains distinct in the context of /oŋ/, e.g., /foŋ/ (缝 'sew') differs from /xoŋ/ (红 'red'). In Type III dialects, /fu/ is synchronically /xu/; /foŋ/ became /xoŋ/; in all other environments /f/ 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_/.

Notably, this set of dialect types and the mapping to medieval Chinese includes cases of bidirectional mergers. The labiodental fricative, *f, became /x^w/ in some dialects (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 Table 4, all *x^w became /f/. The other dialect 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 / types	*xu 护 *ɣu 户	*fu 付 *f^hu 赴 *vu 父	*foŋ 封 *f^hoŋ 蜂 *voŋ 缝	*xoŋ 烘 *ɣoŋ 红	*x^w_ 欢 *$ɣ^w$_ 幻	*f_ 反 *f^h_ 翻 *v_ 范	*x_ 汉 *ɣ_ 汗
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.

2.2 Dialect islands

The geographical distribution of the above eight dialects types across four Southwestern provinces of China is shown in the map in Figure 1. Type I occurs mostly in Sichuan province. Of the 374 documented dialects of Southwest provinces, there are 98 Type I dialects spoken in Sichuan, and 33 Type I dialects spoken in Yunnan. In addition to these areas (Sichuan and Yunnan), where the majority of Type I dialects are located, there are sporadic instances of Type I in Hubei Province and Hunan Province as well.

Type II is mostly in Hunan Province and occurs as well in some districts and counties in Hubei Province. In Sichuan, there are only six dialects of Type II. These are found in the districts and counties of Zhongjiang, Wusheng, Yongchuan, Lezhi, Suining, and Wuxi.

Type III dialects are less common. There are just 13 in Southwest China. Eight of them are found in Sichuan province, including in Jingyang and Luojiang counties, which border Zhongjiang. In addition, there are 5 Type III dialects scattered in Hunan Province and Hubei Province.

Type IV is found only in Tongcheng in Hubei and in Liling and Pingjiang in Hunan.
Type V is found in the middle of Huanan, such as Xiangxiang and Xinhua, and Baojing, which is located in southwest Hunan.

Type VI has been documented in Qiancheng in Hubei and Ningxiang in Hunan.

Type VII is mainly found at the junction of southwestern Hubei and northwestern Hunan, such as in Enshi in Hubei and Xinhua in Hunan. Two additional Type VII dialects are Ziyang and Zizhong, in the midwest region of Sichuan province.

Type VIII is only found in Linxiang, Hunan province.

Besides the 8 types of merger listed in the table, there are some other contexts which are related to the merger of /f/-/x/; for example, in Hejiang and Nanxi dialects (located in Sichuan), the tenseness of the high, back vowels influences merger of /f/ and /x/; /xu/ is read as /fu/, but /xʊ/ remains distinct, as /xʊ/. In addition, the voiced labiodental consonant /v/ is involved in the diachronic sound change in some dialects as well; for example, *xu is read as /fu/, while *ɣu is read as /vu/ (He 2004: 73).

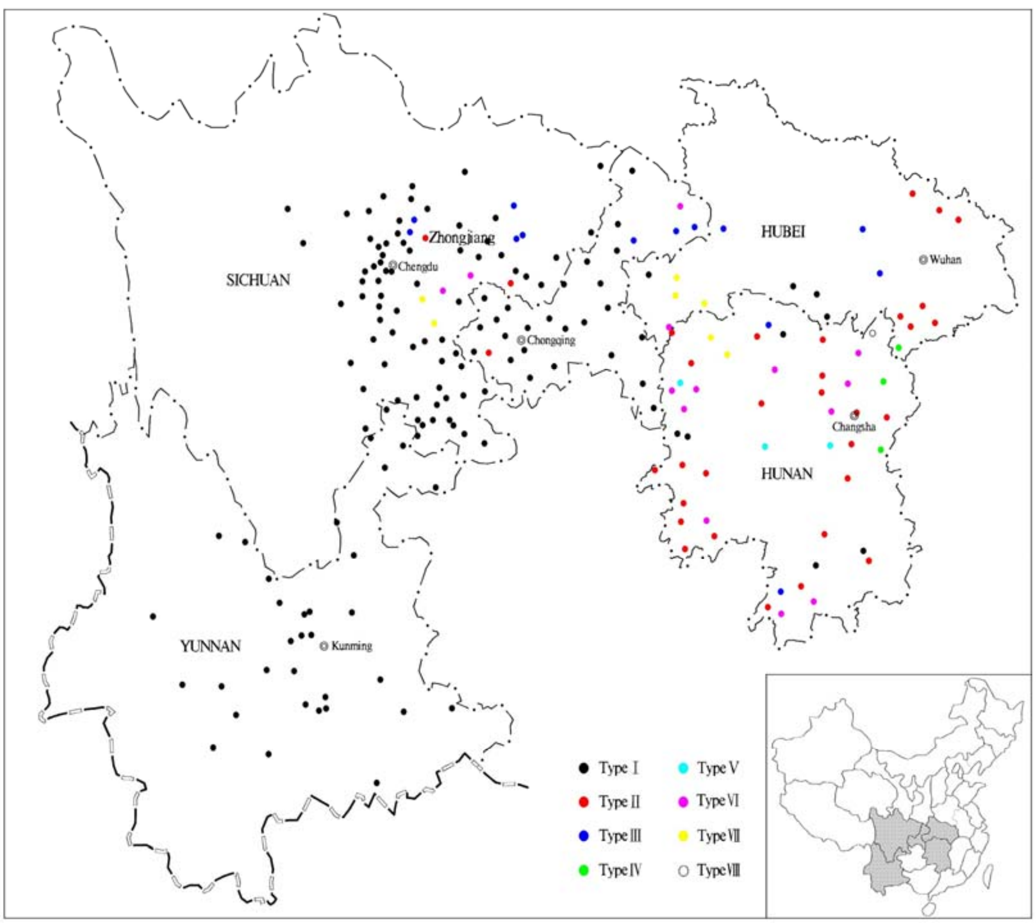


Figure 1. Geographical distribution of /f/-/x/ merger in Southwest four provinces

To summarize, as shown on Figure 1, there are 8 types of labial-velar merger. In the Southwest Mandarin of Yunnan, Sichuan and Hubei provinces, 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 dialect. 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. Human migration patterns have played a key role in dispersing dialect types, including dialect islands such as Zhongjiang. 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 the contemporary variation of dialects in Sichuan. Dialect islands constitute pockets of resistance to areal convergence in favor of dialect continuity over time.

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 dialect (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. In order to minimize the influence of Mandarin on our characterization of the Zhongjiang dialect, we focused on speakers over 55 years of age. 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

3.2 Materials

Materials consisted of a total 90 monosyllabic words beginning with /x/ or /f/. Given that the merger of /x/ and /f/ is related to different phonetic contexts, especially the labial-velar glide, /w/, we designed four sub-groups of stimuli to test different aspects of the Zhongjiang pattern. We describe these subsets as separate studies along with the specific objective of each.

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

As a Type II dialect, we expect *x and $*x^w$ in Zhongjiang to be contrastive: /x/-/f/. This is because $*x^w$ is expected to be /f/. In order to observe whether the change from $*x^w$ 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, /xa45/ ‘laughter’ (Sinograph and Pinyin: 哈, ha55) and /fa45/ ‘flower’(花, hua55[1]) are expected to be minimal pairs in Type II dialects. 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.

[1] Note that Standard Chinese Pinyin uses the same symbol ‘u’ to represent vowel /u/ and glide /w/, e.g., *hu2* /xu35/ ‘lake’ and *hua1* /x^wa55/ ‘flower’.

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-怀	hai35-huai35	/xai31/-/fai31/

Table 3. Contrastive pairs (*x_-*x^w_) in Zhongjiang Chinese

Study 2: merger of /f/ and /x^w/

In this group, we expect /f/ and /x^w/ to merge to /f/, resulting in non-contrastive pairs. In terms of phonetic measurements, we expect completely overlapping distributions for /f/ and /x^w/. 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 /x^w/ in Standard Mandarin ('hu' in Pinyin) produced as /f/ in Zhongjiang.

Sino-graph characters	Mandarin Pinyin	Type II expectation Zhongjiang IPA
发-华	fa55-hua35	/fa31/
肥-回 肺-会 飞-灰	fei35-hui35 fei51-hui51 fei55-hui55	/fei31/ /fei324/ /fei45/
翻-欢 烦-还 返-缓 饭-换	fan55-huan55 fan35-huan35 fan214-huan214 fan51-huan51	/fæ̃45/ /fæ̃31/ /fæ̃51/ /fæ̃324/
分-婚 坟-魂 奋-混	fen55-hun55 fen35-hun35 fen51-hun51	/fən45/ /fən31/ /fən324/
方-慌 房-黄	fang55-huang55 fang35-huang35	/faŋ45/ /faŋ31/

Table 4. Non-contrastive pairs (*f_-*x^w_) in Zhongjiang Chinese

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 dialect. 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 the Zhongjiang dialect.

Sino-graph characters	Mandarin Pinyin	Type II expectation Zhongjiang IPA
浮-猴	fu35-hou35	/fəɯ31/-/xəɯ31/
否-吼	fou214-hou214	/fəɯ51/-/xəɯ51/

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

Study 4: the /oŋ/ environment

From previous research, we expect the contrast between /foŋ/ and /xoŋ/ to be maintained in Zhongjiang dialect. 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 dialects. Table 6 included 3 pairs (6 items), which are expected to be contrastive in both standard Mandarin and Zhongjiang dialect: for example, /xoŋ31/ ‘red’(红，hong35) and /foŋ31/ ‘sew’(缝，fong35).

Sino-graph characters	Mandarin Pinyin	Type II expectation Zhongjiang IPA
风-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. They 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. Participants were asked to read each item when it appeared. The items were pseudo-randomized in one list, which was repeated 10 times over the course of the session. All tokens were recorded in mono channel at 44,100 Hz directly to a Thinkpad T440 laptop using an external Samson C03U microphone.

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) and spectrum Standard Deviation (SD). These measurements are 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/or spectrum SD. This resulted in the loss of 46 tokens or 0.5% of the data. Although we report average differences in spectral Center of Gravity across fricatives at each of the five timestamps, in the studies described above we focus on measurements taken only at the middle 20 msec interval of the fricatives, where the greatest average difference across fricatives was observed.

4. Results

For starters, we report the duration of the fricatives. Figure 2 shows a kernel density plot of contrastive pairs, i.e., tokens expected on the basis of past characterizations of Zhongjiang to be produced distinctly as /x/ and /f/. The figure collapses across the Study 1 words produced by all 10 speakers. The distributions overlap heavily, an observation which also extends to each of the 10 speakers individually and to the words in other studies. In short, /x/ and /f/ show vary similar durations in this corpus. The distribution for /x/ is slightly more peaky. /f/ tokens are slightly more probable at short durations. 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 as a fixed factor. A likelihood ratio test showed that the model including fricative did not significantly improve over the baseline model ($\chi^2 = 2.70$, $p > 0.1$) indicating that, when random effects are factored in, the difference in duration across fricatives is not significant.

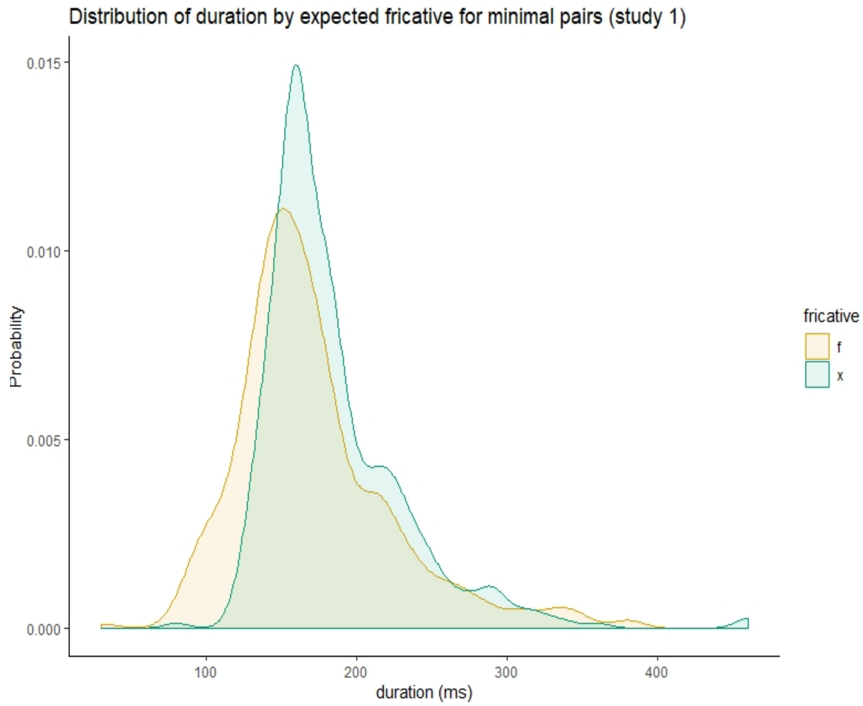


Figure 2. Distribution of segment duration by expected fricative: /f/-/x/

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 range 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 500-1,000 Hz are consistent with resonance in front of a velar constriction in the vocal tract.

/xe2/ (speaker 8) COG: 149 Hz SD: 289 Hz	/xəɯ3/ (speaker 7) COG: 200 Hz SD: 298 Hz	/xai3/ (speaker 1) COG: 370 Hz SD: 496 Hz

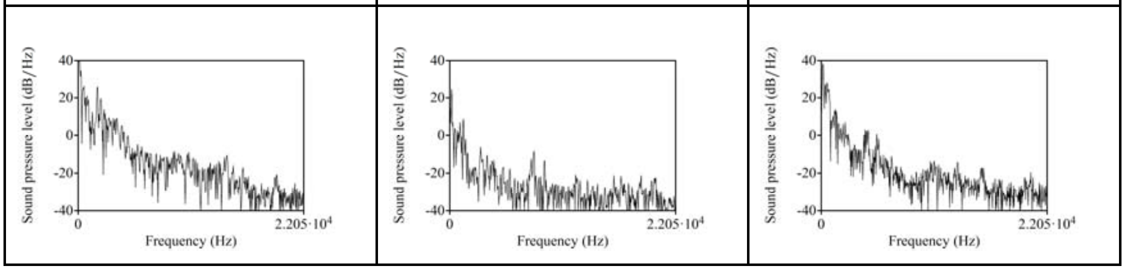


Figure 3. Spectra of /x/tokens

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 first two spectral moments: /f/ (Figure 4) has a substantially higher CoG and spectrum SD than /x/ (Figure 3).

/fe2/ (speaker8) COG: 4,813 Hz SD: 5,628 Hz	/fəɯ3/ (speaker7) COG: 3,729 Hz SD: 3,974 Hz	/fai2/ (speaker 1) COG: 5,173 Hz SD: 4,829 Hz

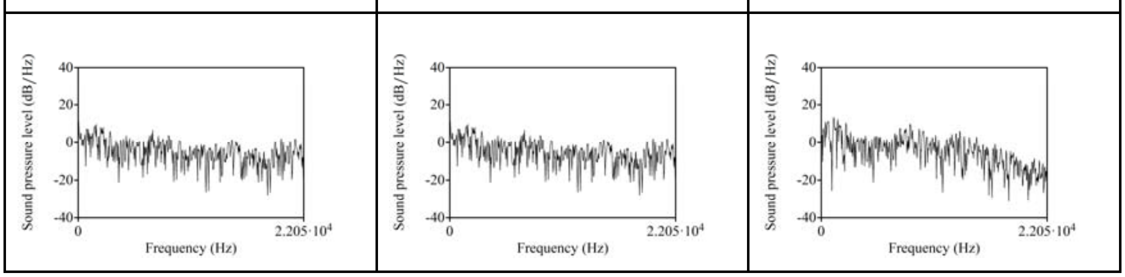


Figure 4. Spectra of /f/ tokens

The tokens in Figure 3 all sound unambiguously like /x/ to the authors and the tokens in Figure 4 sound unambiguously like /f/. However, there are numerous tokens in the corpus that straddle these perceptual boundaries. Figure 5 provides three 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 two tokens that both sound to the authors like /x/, even though they are historically distinct fricatives and synchronically distinct in other dialects of Chinese (including standard Mandarin). The third panel overlays spectra of fricatives that both sound like /f/ to the authors. 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 /x^w/.

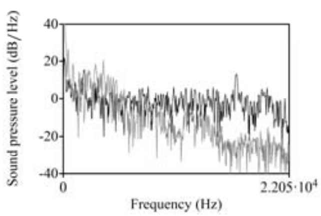
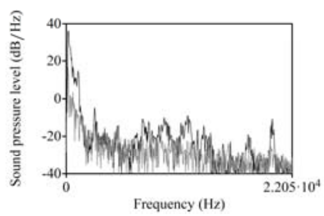
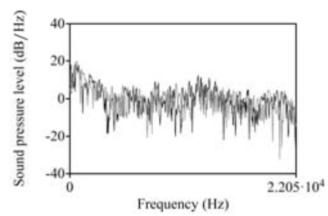
*x^w, /fe2/→[fe2] (speaker4) - black COG: 4,581 Hz SD: 6,185 Hz *x, /xe2/→[xe2] (speaker4) - grey COG: 134 Hz SD: 511 Hz	*x^w, /fən4/→[x^{w}ən4] (speaker7) - black COG: 567 Hz SD: 1,263 Hz *f, /fən4/→[x^{w}ən4] (speaker7) - grey COG: 265 Hz SD: 271 Hz	*f, /fei1/→[fei1] (speaker6) - black COG: 5,755 Hz SD: 6,137 Hz *x^w, /fei1/→[fei1] (speaker6) - grey COG: 3,888 Hz SD: 5,466 Hz
		

Figure 5. Spectral comparison of /f/ (black) and /x/ (grey) tokens

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.

The spectra reported above are based on the 20 msec window at the midpoint of the fricatives. We found that, on average, this is where the greatest difference between contrastive fricatives was found. Figure 6 illustrates the average difference in CoG between contrastive /x/-/f/ pairs from study 1 (Table 2). These are based on words that form minimal pairs in contemporary Zhongjiang Mandarin (Study 1 and Study 3 items). The figure shows that there is already a CoG difference in the first 20 msec of the /f/ and /x/ fricatives, i.e., at *t1*. This difference increases in the next 20ms window, *t2*, and peaks at the midpoint of the fricative, *t3*. The difference narrows substantially in the next 20 msec window, *t4*, and is lost altogether in the last 20 msec window, *t5*. Since we will be discussing cases of fricative merger, we focus on the midpoint of the fricatives, as this is the time window that shows the largest average difference between fricatives.

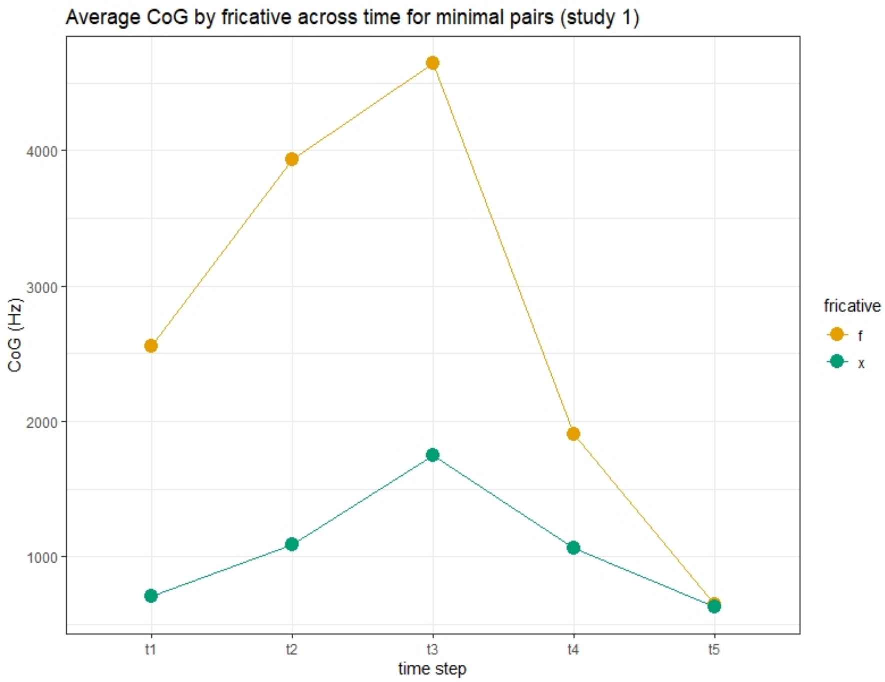


Figure 6. The average CoG between contrastive /x/-/f/ pairs

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 dialect (see Table I), as has been claimed in the literature. As a Type II dialect, *x^w has become /f/. The minimal pairs in this study contrast *x~*x^w, which are synchronically /x/~/f/. For example, /xa45/ 'laughter' (Sinograph and Pinyin: 哈, ha55) and /fa45/ 'flower'(花, hua55).

Figure 7 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.

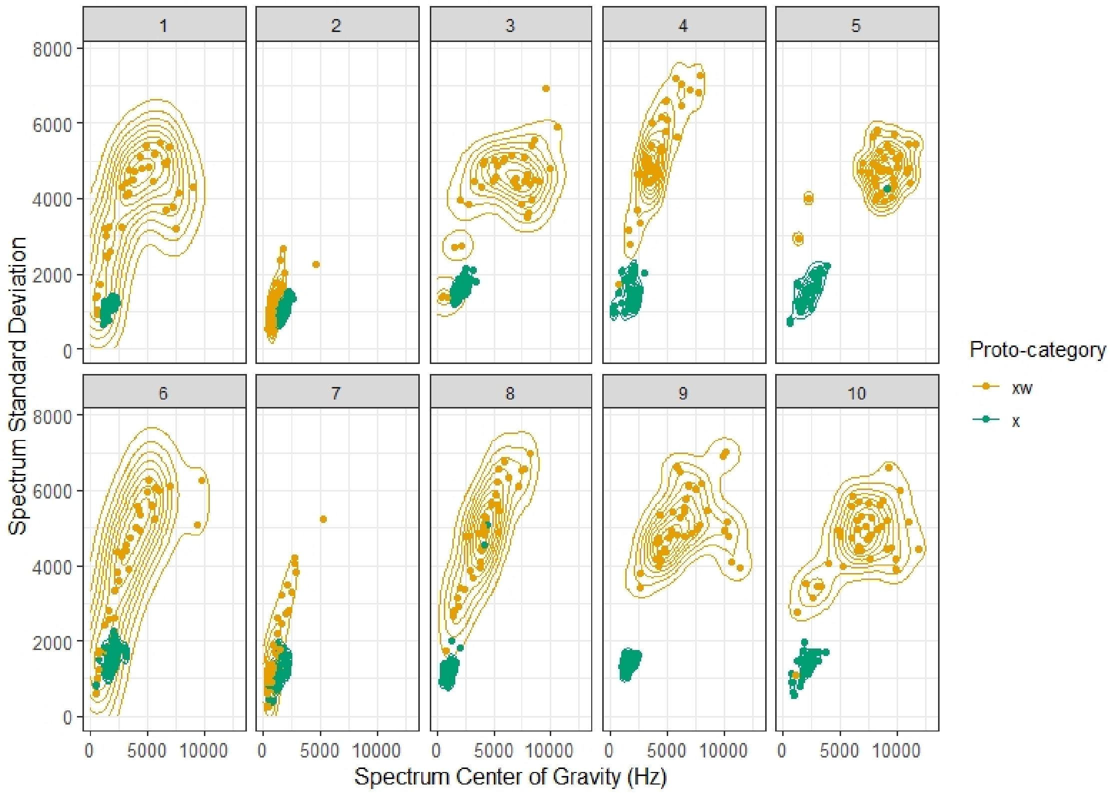


Figure 7.Center of Gravity and Spectrum Standard Deviation of contrastive pairs（*x_-*x^w_)

The results of study 1 indicate that, 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 dialect, 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 synchronically distinct in other Chinese dialects (including Standard Mandarin), but are expected to be homophonous in Zhongjiang. Figure 8 shows that, as expected, most speakers show a merger for these words. 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 7). 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 dialect.

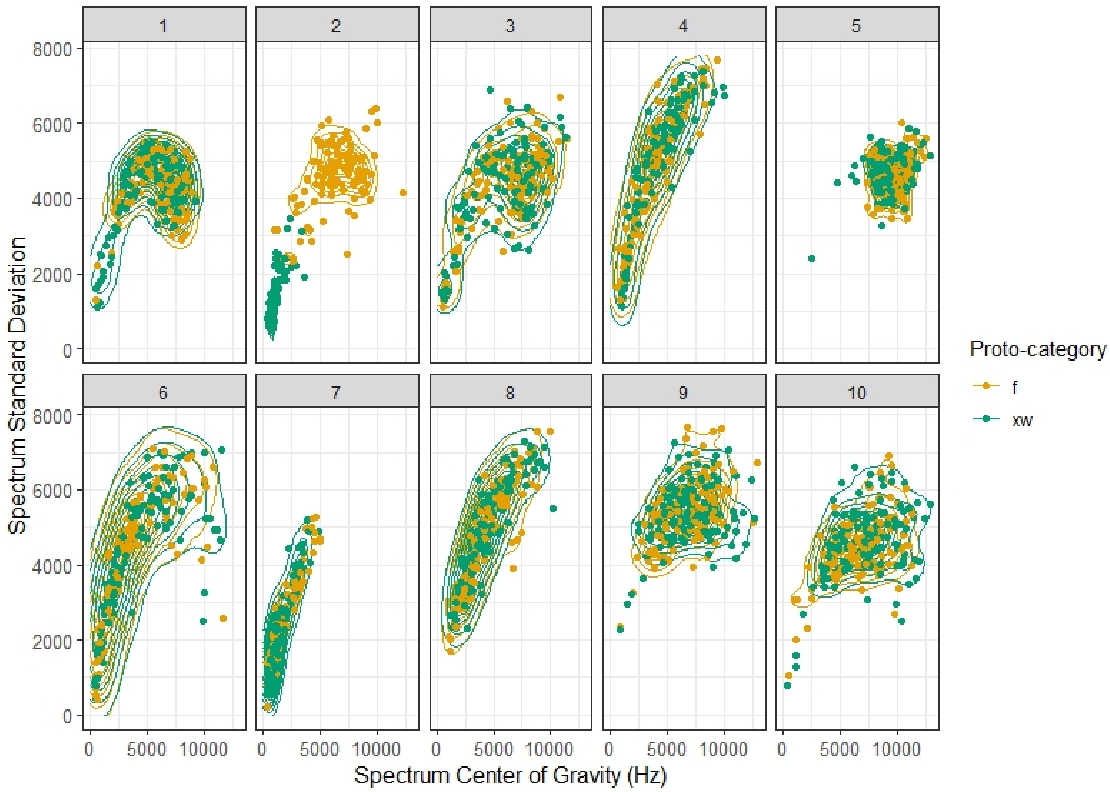


Figure 8. Center of Gravity and Spectrum Standard Deviation of non-contrastive pairs (*f_-*x^w_)

To summarize, the results of study 2 show that most speakers (9/10) produce *x^w as /f/, resulting in an increase of homophonic pairs in Zhongjiang relative to Standard Mandarin. The majority pattern verifies Zhongjiang a Type II dialect, 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 the Zhongjiang dialect and Standard Mandarin. The mergers (study 2) and contrasts (study 1) in the first two studies are specific to Zhongjiang (and other Type II dialects), 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 9 reports the results for Study 3. 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.

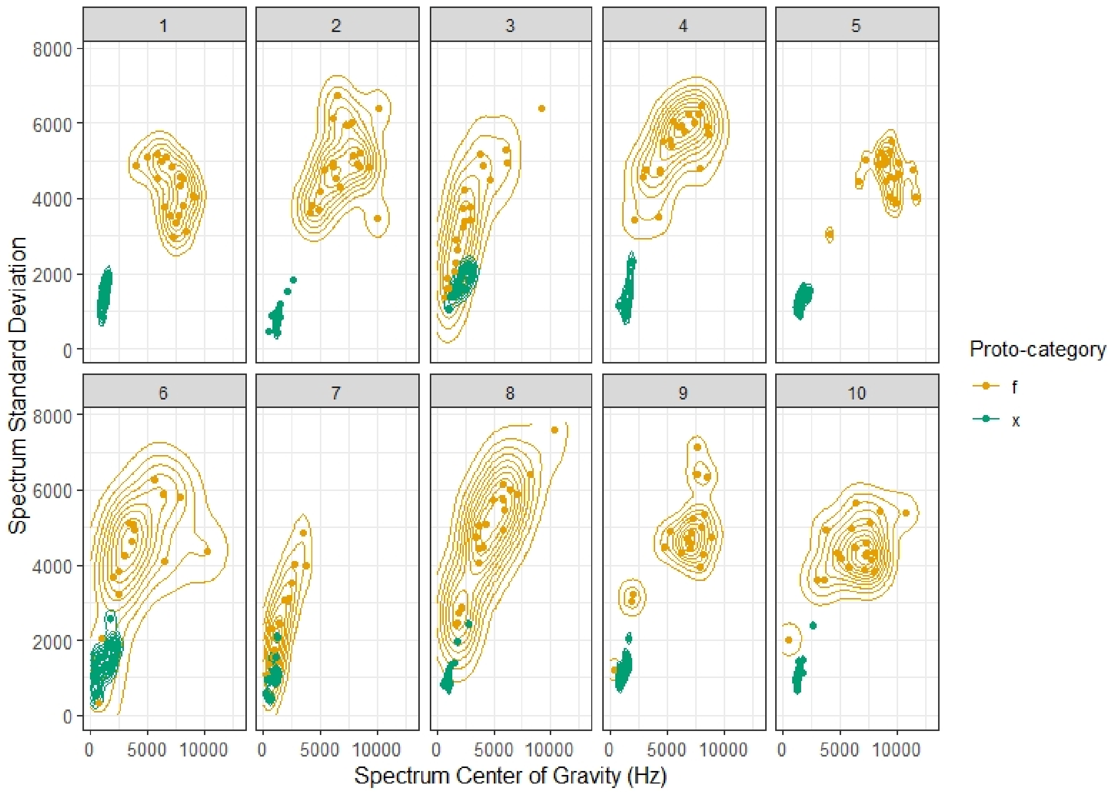


Figure 9. Center of Gravity and Spectrum Standard Deviation of contrastive pairs (*f_-*x_) with rime /əɯ/

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 dialects (see Table 1). As a Type II dialect, Zhongjiang is expected to maintain contrast in this environment. Figure 10 shows the results. Several speakers, including some that maintain contrast in other environments, show 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/.

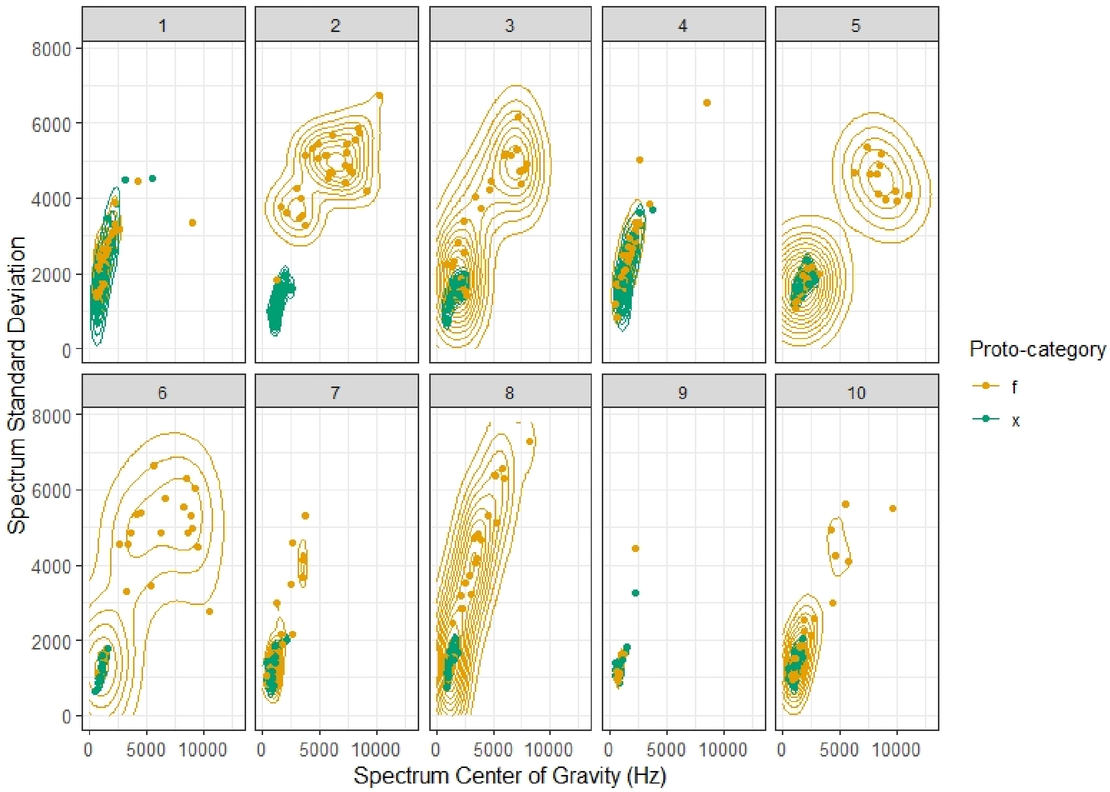


Figure 10. Center of Gravity and Spectrum Standard Deviation of non-contrastive pairs with rime /oŋ/

5 Discussion

Across the four studies reported above, we observed substantial individual differences in how /x/ and /f/ were produced across contexts. 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 *x^{w}. Three other subjects, S1, S6, S9, show only minor deviation from the dominant pattern.

Our reporting of the data focuses on two spectral measurements. 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. This is not necessarily a general finding but follows from the typical properties of velar and labiodental fricatives. We found no differences between /x/ and /f/ in duration. Ultimately, we opted to focus on a smaller number of phonetic measurements that have particularly clear articulatory interpretations; however, we have also submitted the entire data set, including additional measurements as well as sound files and text grids as a *Data in Brief* article.

5.1 Secondary velar articulation in Zhongjiang labiodental fricatives

A key characteristic of the dominant pattern is that /f/ shows a range of variation spanning from the CoG values for /x/ to much higher CoG values. Given the range of variation for /f/, it is possible that /f/ is variably velarized, i.e., as /f^{γ}/. The secondary velar constriction, a narrowing of the vocal tract in the region of the soft palate, would account for the low CoG observed for some tokens. The continuous variation from low CoG values, characteristic of /x/, to high CoG values, characteristic of /f/, may result from variation in the timing and magnitude of the two component gestures of /f^{γ}/, a raising of the lower lip for the /f/ component and a raising of the tongue body to the soft palate for the [// component. It is also notable in this context that /x^{w}/, which is unambiguously a complex segment, has the same duration as /f^{γ}/, which we propose, on the basis of the spectral variation, is also a complex segment. This account of the phonetic variation observed for *f may also be able to contribute to an account of the merger between *f and *x^{w}, which is also a characteristic of this group (Figure 8, Study 2).

If *f contains a secondary velar constriction, i.e., /f^{γ}/, then it is a complex segment, similar in composition to a labialized velar, /x^{w}/. Both /f^{γ}/ and /x^{w}/ 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 /x^{w}/; the /f/ component of /f^{γ}/. 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 /x^{w}/ and the /f/ component of /f^{γ}/. 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 /x^{w}/ and why they closely overlap with the range of values observed for /f^{γ}/.

To illustrate the similarity between a velarized labial fricative, /f^{γ}/, and a labialized velar fricative, /x^{w}/, we provide partial gestural scores for these sounds in Figure 11. A gesture is a discrete constriction event in the vocal tract, and the gestural score depicts how these events are organized in time (e.g., Browman & Goldstein, 1986). Gestures are defined by a set of dynamic parameters, including a target constriction location (CL) and constriction degree (CD). The gestural score in Figure 11 shows Tongue Dorsum and Lower Lip gestures for both /f^{γ}/ and /x^{w}/. The grey boxes indicate the activation duration, or temporal extent of the gestures, and they are labelled with CD and CL parameters. Notably, the gestures for /f^{γ}/ (left) and /x^{w}/ (right) are similar. Both involve the Tongue Dorsum and Lower Lip articulators. There are some small differences in the parameters for the gesture and the relative timing of the gestures. One parameter difference is the constriction location (CL) of the lower lip gesture. For /f^{γ}/, the lower lip CL is the upper teeth; for /x^{w}/, the CL is the upper lip. There is also a small difference in Lower Lip constriction degree (CD); it is “critical” for /f^{γ}/ and “narrow” /x^{w}/.

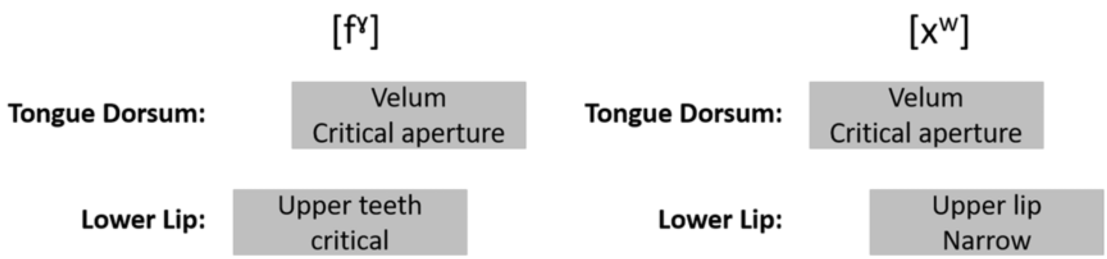


Figure 11. Partial gestural score for /$f^{ɣ}$/ (left) and /x^{w}/ (right). These complex segments are composed of similar gestures, both involving the lower lip and tongue dorsum articulators.

The gestural score in Figure 11 highlights the articulatory similarity of /$f^{ɣ}$/ and /x^{w}/. Maintaining acoustic distinctiveness between articulatorily similar sounds relies crucially on the precise timing and magnitude of the constituent gestures. Although we do not have articulatory data for these speakers, we speculate that the variability in the acoustics can derive naturally from variability in the relative timing and magnitude of constituent gestures. For /$f^{ɣ}$/, early formation of the tongue dorsum constriction at the velum could lower CoG and give rise to the percept of /x^{w}/. Likewise, for /x^{w}/, early formation of the lip constriction yielding a relatively flat spectrum (and high CoG) could give rise to the percept of /$f^{ɣ}$/. This change, *x^{w} to /$f^{ɣ}$/, is expected of Type II dialects, and is the dominant pattern in the Zhongjiang data reported above. However, there was variation within the Zhongjiang sample. A minority of speakers also showed the opposite pattern, /$f^{ɣ}$/ produced as /x^{w}/. This was exemplified most clearly by speaker S7. In Southwestern Mandarin dialects more broadly, we observe sound changes in both directions (Figure 1, map of dialects). Our data also revealed new nuances to the characterization of the Zhongjiang dialect, which are related to the proposal above. Study 4 revealed that some speakers that show the dominant pattern (consistent with Type II) whereby *x^{w} becomes /$f^{ɣ}$/ also show a shift from *f to /x/ in the environment of /oŋ/. Three speakers, S1, S4, and S9, showed this pattern. This indicates that there are also bidirectional mergers within speakers. Deriving the acoustic variability from temporal variation in articulation provides a natural account for the bidirectional nature of these mergers.

Given the aerodynamic requirements for these fricatives, normal degrees of spatial-temporal variation in articulation may have a disproportionate impact on the acoustics. That is, velar fricatives combined with 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 velar and labial components may be anti-stable in that a small amount of articulatory variation could give rise to relatively large acoustic consequences. Generating turbulence from a supralaryngeal constriction requires precise balance between the amount of air flow volume and the diameter of the constriction. In the case of /x^{w}/, 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. For /x^{w}/, 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 /x^w/, small reductions in velar constriction naturally give rise to acoustic conditions similar to /f^γ/.

Positing a secondary velar articulation for the labiodental fricative, /f^γ/, helps to explain the prevalence of alternations with /x^w/, both within specific speech communities, such as Zhongjiang, across the Mandarin-speaking world, and within individual speakers in different environments (e.g., S1, S4, S9). However, there may also be variation between /f/ and /f^γ/. 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, *x^w is produced as /f/. Even though this speaker is in the minority, by using /f/ instead of /f^γ/ there is a sense in which S5 could be considered the most canonical representative of a Type II dialect in our sample.

This brings our discussion to the highly relevant issue of dialect contact. The Zhongjiang dialect is island, a Type II dialect in a sea of Type I dialects. All speakers in our sample have at least some contact with both other Southwest Mandarin dialects, such as the Chengdu dialect (Type I), the language variety of the provincial (Sichuan) capital, a major urban area, and a prestige dialect 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 *x^w as /f/ is recognized as a local dialect feature that differs from both Chengdu dialect and Standard Mandarin. It is possible that the /f^γ/ variant that we have posited to explain the observed variation is a variant used by Zhongjiang speakers to approximate more standard varieties. From this standpoint as well, we can see S5 as the most prototypical (or possibly even stereotypical) example of a Type II dialect speaker.

The third speaker whose production patterns fell outside of the main pattern was S2. This speaker's productions are also interesting from the standpoint of dialect contact. This speaker had the range of variation for /f/ that we explained by positing a secondary velar articulation, /f^γ/. However, the distribution of phonetic categories across words did not correspond to our expectations for a Type II dialect. By and large, the distribution of variants corresponds more closely to Standard Mandarin. Study 1 showed that, for this speaker, *x^w maintained low CoG, just like *x. This is the pattern expected of Standard Mandarin. However, Study 2 showed that the *f words are produced as /f^γ/. We surmise that S2 has generalized the /f^γ/ realization, which developed in Zhongjiang from *x^w, such that, synchronically, /f^γ/ is used very generally for /f/, including even for *f.

To summarize, we verified aspects of the description of Zhongjiang as a Type II dialect. That is, *x^w is produced as a labiodental fricative. In addition, we exposed gradient phonetic variation amongst labiodentals, which we attributed to variation in timing between the labial gesture and a secondary velar gesture. A likely source of the secondary velar gesture in /f^{γ}/ is the velar component of *x^w. As sketched in Figure 11, the gestural components of /f^{γ}/ and /x^w/ are similar but differ in relative timing. In addition to these main trends, we identified a context, /oŋ/, in which *f is produced as /x^w/ by some speakers. This had not previously been documented for Zhongjiang and is not considered a characteristic of Type II dialects. Finally, we noted some cases of apparent dialect contact, in which Zhongjiang speakers showed characteristics of prestige dialects (Chengdu dialect, Standard Mandarin).

5.2 Labial-velar mergers more broadly

Although our study focuses on one particular dialect, the results have implications for labial-velar mergers more broadly. As mentioned in the introduction, labial-velar fricative mergers are common and have attracted the attention of many scholars. One hypothesis raised for labial-velar merger within South Chinese dialects points to labialization as a driver of the merger (Wan1998/2009, 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 /x^w/ is not a consonant cluster; rather, the velar fricative is labialized ('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 dialects. A variation on this account, proposed by Wan (1998/2009: 191), posits an intermediate step whereby /x^w/ becomes /ɸ/ on route to /f/. Whether the /ɸ/ stage is sufficiently stable or not, this approach can account for why the /w/ environment sometimes leads labial-velar mergers (see 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 /x^w/, which may contribute to the /f/ percept, as proposed by (Zhuang 2004).

Although /w/ is indeed a common environment for labial-velar merger in Chinese dialects, there are also cases in which sound change proceeds in the opposite direction. That is, *f changes to /x^w/. We observed this change as well in the Zhongjiang data, in the environment of /oŋ/. It is harder to see how Zhuang's account explains merger in the opposite direction, *f becoming /x^w/, because this requires the seemingly spontaneous emergence of a velar gesture, an observation also made by Sun (2007).

We have argued that the phonetic data for Zhongjiang is consistent with the presence of a secondary velar gesture for /f/, i.e., /f^{γ}/, across a range of vowel contexts. A change from /f^{γ}/ to /x^w/ is less mysterious than a change from /f/ to /x^w/, because there is a velar component to the labiodental fricative in /f^{γ}/. Given this secondary velar gesture, we can understand changes in both directions, /x^w/→/f^{γ}/ as well as /f^{γ}/→ /x^w/ as following from the same mechanism, namely, variation in the relative timing between the component gestures of complex segments.

Notably, our acoustic evidence for a secondary velar articulation in labiodentals is specific to Zhongjiang. The extent to which other Chinese dialects also have velarized labiodentals is an empirical question. We might speculate that velarized labiodentals exist in other dialects as well and that those dialects are more likely to exhibit changes from *f to $/x^w/$. In the absence of a velar gesture for the labiodental fricative, we'd expect *f to /x/ changes to proceed in environments in which the vowel provides a tongue dorsum retraction gesture, such as /u/ and /o/. These environments are indeed the typical conditioning environments for *f to /x/ sound changes, which may mean that a velarized labiodental, as we have observed in Zhongjiang, is not strictly necessary to condition this change.

Another possible consideration in predicting labio-velar mergers across speech communities is the specific quality of the conditioning vowels. For example, across dialects of Mandarin, there is variation in the quality of the /u/ vowel. In Zhongjiang, this vowel is not as rounded as it is in Standard Mandarin. When word-initial, not preceded by an onset consonant, Zhongjiang /u/ takes on a labiodental approximate quality, e.g., 五 /u3/ /ʋɯ/ 'five', or, variably, even a fricative quality, /vɯ/. In other environments as well, the lips are not as protruded for Zhongjiang /u/ as they are for Standard Mandarin. This small difference in labial specification for Zhongjiang may contribute to the ecology of a Type II dialect, favoring change of $*x^w$ to $/f^{\gamma}/$ or /f/. A detailed account of the variation in lip position for /u/ in Zhongjiang dialect is beyond the scope of this paper. We expect that there will be inter-speaker variation on this dimension as well. The change from $*x^w$ to $/f^{\gamma}/$ or /f/ may be particularly natural for speakers (or speech communities) that have a lower lip target for /u/ that approximates the upper teeth, as in /ʋɯ/, than for speakers that have higher degree of rounding and lip protrusion.

Our account of the labial-velar merger as well as the related proposal by Zhuang (2004) relies to some degree on the reinterpretation of temporal variation in articulatory gestures as a basis for sound change. There are other proposals that treat the labial-velar mergers as more standard instances of lenition. For example, He (2004: 123) notes that /f/ is sometimes produced weakly, or reduced, to a bilabial fricative, /ɸ/, in some dialects. 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 above, the progression of change passing through /ɸ/.

The types of patterns observed in Chinese dialects are representative of 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), which is consistent, we believe, with our gestural account of Zhongjiang. 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 dialects 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 Chinese dialects, synchronic phonotactic restrictions within Chinese 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. Chinese dialects 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 Southwest Mandarin dialects. Duanmu (2007) mentions that in Standard Mandarin there is an exception to the general absence of /f^w/, a single word /f^wo/ 'buddha'; but this exception is absent in Zhongjiang (and other dialects), 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 /f^j/ as outcomes of sound change in Chinese dialects.

The prevalent ban on /f^w/ across Chinese dialects 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 Southwest Mandarin dialects (Figure 1), /fu/ is a much more common outcome. Out of 212 documented dialects 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/. Anecdotally, as described above, we have observed that the labial component of /u/ in Zhongjiang is perhaps closer to that of /f/ than in Standard Mandarin. 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 labial velar fricative variation in Zhongjiang revealed a range of phonetic variation for /f/, which we argued derives from a secondary velar articulation: /f^{γ}/. The study comes in the context of areal variation amongst Southwest Mandarin dialects in labial velar fricative realization. In particular, these dialects exhibit bidirectional sound changes; both f* → /x^w/ and *x^w → /f/ changes are common. The identification of a secondary velar articulation points to a possible temporal basis for these sound changes that is also appropriately bidirectional, as shifts in the relative timing of labial and velar gestures can give rise to both outcomes.

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 Yale Phonology group, 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)*, Fuzhou: Fujian People Press.

Catford, J. C. 2001. *A practical introduction to phonetics* (2nd Edn). Oxford: Oxford University Press.

Chao, Y. R. et al,. 1948. *The report of Hubei dialects*. Taipei: The Institute of History and Philology.

Chen, Z. T., and Li, R. L. 1991. *The study of Min Chinese*, 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.

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

Cui, R. C. 1996. *The Xiang dialect in Sichuan*, 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 *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*. 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., and Tilsen, S. 2015. The articulatory tone-bearing unit: gestural coordination of lexical tone in Thai. *Proceedings of Meetings in Acoustics*. 22, 1–9.

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*. Ph.D dissertation. Chinese Academy of Social Sciences.

Li, R. L. 1999. *The report of Hakka Chinese in west Guangdong province*. Guangzhou: Jinan University Press.

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.

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.

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., 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. 676-680). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Sun, Y. Z. 2007. *The study of phonetics in Gan Chinese of Jiangxi*. Peking: Yuwen Press.

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

Wan, B. 1998/2009. *The study of the historical level of the initial consonants in Gan Chinese*. Peking: Commercial Press.

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

Xiang, M. B. 2005. *The phonology of Hakka Chinese*. Ph.D. dissertation. Peking University.

Xie, L. W. 2003. *The study of phonetic Hakka Chinese.* Peking: Chinese Social Science Press.

Yang, S. F. 1969. *The investigation report of Yunnan dialects.* Taipei: the Institute of History and Philology.

Yang, S. F. 1974. *The investigation report of Hunan dialects*. Taipei: the Institute of History and Philology.

Yang, S. F. 1984. *The investigation report of Sichuan dialects*. Taipei: the Institute of History and Philology.

Ye, X. F. 2008. The phenomenon of /f/-/x/ merger in Chinese Dialect. *East Linguistics(4)*, Shanghai: Shanghai Educational Press.

Zhan, B. H. 2002. *An introduction of Cantonese in Guangdong*. Guangzhou: Jinan University Press.

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*. 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.