Consonant and lexical tone interaction: Evidence from two Chinese dialects

Shi, M.

Chapter 5 Low-rising tone and onset consonant in Shuangfeng Xiang Chinese

5.1 The debate in Shuangfeng Xiang Chinese

In Chapter 3, the 'aspiration-induced tonal split' (ATS) phenomenon has been examined. ATS refers to the distinct lowering of *f0* of a lexical tone over syllables beginning with a voiceless aspirated obstruent, compared to that of syllables beginning with an unaspirated obstruent. The results of Lili Wu Chinese have shown that there has been a completed merger between lexical tones beginning with voiced and aspirated onsets. This merger, however, is conditioned by the MC tonal categories. In this chapter, another Chinese variety, namely Shuangfeng Xiang Chinese, will be the empirical test ground. Compared with Lili Wu Chinese, what is special for Shuangfeng Xiang Chinese?

As indicated by the description in Chapter 4, Shuangfeng Xiang Chinese features a three-way laryngeal contrast of obstruents. Like most Chinese dialects featuring the three-way-contrast system, there are co-occurrence constraints on onset and tone combinations in Shuangfeng Xiang Chinese (see Section 4.2 for more details). The two level tones (i.e., T1 and T5) are in complementary distribution: T1 (high–level) can only co-occur with voiceless onsets, while T5 (low–level) exclusively with voiced onsets. T3 (falling) and T4 (high–rising) both can co-occur with voiceless onsets, but not with voiced obstruent onsets. Thus far, there seems to be the [voiceless/H]-[voiced/L] co-occurrence pattern in Shuangfeng Xiang Chinese. However, the low-rising contour (i.e., T2) can co-occur with both voiceless and voiced onsets. This phenomenon has also been widely observed in other Old Xiang varieties (Yang, 1974). Given that the *f0* contours following all three onsets (i.e., voiceless unaspirated, voiceless aspirated, and voiced) show similar low-rising *f0* contours in the majority of Old Xiang varieties, a widely-discussed question is: Are there synchronic phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets which might condition/have conditioned the similar low-rising *fo* contours? In general, there exist two views.

The first view argues that the voicing contrast of onsets is the synchronic phonetic property featured by the onsets of syllables that bear the low-rising *fo* contour (hereafter the 'voicing contrast' view). This view seems to initially originate from the transcription of another Old Xiang variety, namely Wugang Xiang Chinese (武冈方言) (Chao, 1935 [Yang, 1974]).[38] In Wugang Xiang Chinese, all *fo* contours after different onsets are treated as an identical low-rising tone. This phenomenon is termed 'initial-associated tonal merger' ('*Yiniu tongdiao*' [异纽同调], Chao, 1959, ITM hereafter), which refers to different initial consonants being associated with one lexical tone. This treatment has become a convention for researchers who worked on other Old Xiang varieties since then. Although in Shuangfeng Xiang Chinese, some impressionistic studies have already suggested that voiced onsets tend to introduce a lower *fo* onset than voiceless onsets (Xiang, 1960; Yuan, 1960: 115; *Zihui*, 1989; Chen, 2004).

The second view, proposed by Zhu and Zou (2017), argues that the phonatory aspect (i.e., modal vs. breathy) of the laryngeal contrast conditions the co-occurrence of all onsets and the low-rising contours (hereafter the 'phonation contrast' view). They found that voiced onsets do not consistently differ from their voiceless counterparts in terms of phonetic voicing. Only 53% of the voiced onsets in their corpus have a

---

[38] This phonological system of Wugang Xiang Chinese was recorded by Yuen-Ren Chao in 1935 ([民国]二十四年十一月二日, lit. [Republic of China] the 24th year November 2nd) based on the investigation of two young consultants (18 years old and 19 years old). This transcription is included in the book titled *Report on a survey of the dialects of Hunan* under the name of Shih-Feng Yang published in 1974 (pp. 495–513).

negative-VOT realization. Furthermore, all low-rising-carrying syllables with a voiced onset in their experiment show a high magnitude of breathiness (based on the measurements of H1–H2) during the first interval of 30 ms of the vowel. They hence schematize a two-way phonatory contrast corresponding to the laryngeal contrast of voiceless vs. voiced onsets and posit two low-rising tones (i.e., modal – /24/ vs. breathy – /13/). They further claim that this analysis can be generalized to explain similar phenomena existing in other Xiang Chinese (e.g., Shuangfeng).

To sum up, the key controversy of the two views is centered on the synchronic phonetic properties shared by the onsets that condition the low-rising *f0* contours. The primary goal of this chapter is to test the veracity of the two views. Moreover, given that there is an 82-year interval between the two competing views when they were proposed (1935 vs. 2017), it is also possible that these two different views actually reflect sound changes over the last decades. Generational differences have been observed in a couple of studies focusing on the neighboring varieties of Shuangfeng Xiang Chinese. For example, Zhong and Chen (2012) show that in Xinhua (新化) Xiang Chinese, older speakers tend to produce more negative-VOT tokens than younger speakers. Accordingly, another particular interest is to examine whether the contradictory proposals in the existing literature may have resulted from generational differences. Exploring this question can not only advance our knowledge of the phonetic properties in Shuangfeng Xiang Chinese, but also sharpen our understanding of the typology of how languages structure the laryngeal contrasts and their interaction with lexical tones.

5.2 The current study

Generally speaking, previous studies attempting to solve the puzzle suffer from various inadequacies. One obvious problem is the scarcity of the empirical data, which results in the 'voicing contrast' view being a purely impressionistic assumption without solid phonetic evidence. The 'phonation contrast' view, although proposed with instrumental measurement data, still fails to take the phonatory characteristics of different onsets into consideration. We know that the magnitude of the differences between contrastive phonation types also varies across languages. For example, the modal phonation in Jalapa Mazatec (an Otomanguean language) presents approximately a 5 dB difference between H1 and H2 at 25 ms of vowels, while the breathy phonation shows a greater difference – more than 10 dB (Blankenship, 2002). In White Hmong (a Hmong-Mien language), however, the difference is more limited. Only a 3 dB difference was found at the vowel onset position in Esposito (2012). Given the various realizations of the phonatory contrast, H1–H2 values measured at one point of the vowel after a voiced onset alone cannot help reveal the full picture of the contrast between modal and non-modal phonation types. In a nutshell, none of these existing studies have managed to provide comprehensive and sound data to assess the debate rigorously. Multiple acoustic measures elicited from a large sample of speakers of different generations are urgently needed, hence the need for the current study.

Based on the literature, the following evidence is expected to support the competing views. If the 'voicing contrast' view is correct, given the acoustic realization of the three-way laryngeal contrast in Old Xiang Chinese (see Section 4.4 and references therein), the three onsets should differ significantly in VOT while the *f0* contours after different onsets are expected to be similar (i.e., ITM). On the contrary, if the 'phonation contrast' view is correct, given the breathy nature of the voiced onsets, the breathiness after voiced onsets is expected to be significantly greater than that after either voiceless unaspirated or aspirated onsets, while the phonatory state after the two voiceless counterparts should be similar. In terms of *f0*, given that the 'phonation contrast' view claims a contrast of /24/ and /13/ occurring with voiceless and voice onsets, respectively, the *f0* contour after voiced onsets (which can be characterized as breathy) is expected to be significantly lower than that after both voiceless unaspirated and aspirated onsets (which is not breathy), while *f0* contours after the two voiceless counterparts should be similar. The possible generational differences would predict that data from speakers of the older speakers support the 'voicing contrast' view, while that from the younger speakers support the 'phonation contrast' view.

5.3 Method

5.3.1 Stimuli

The stimulus list consists of a near-minimal set of 15 monosyllabic words carrying a low-rising *f0* contour (i.e., T2). All target words begin with i) a labial (/p p^h b/), ii) an alveolar (/t t^h d/), or iii) a velar (/k k^h g/) plosive. Within each place and manner of articulation, 5 example words are included. The three groups within alveolar are balanced in terms of the identity of the vowels (/i e a/) following the initial consonant. The three groups within both labial and velar, however, only contain one vowel (/e/) due to accidental gaps and phonotactic constraints in Shuangfeng Xiang Chinese (see Section 4.8). All stimuli were confirmed to be frequent and familiar words in Shuangfeng Xiang Chinese by an educated native speaker who also participated in the experiment. The full stimulus list is provided in Appendix II-1. Given the similarity of voiced plosive and affricate onsets in Shuangfeng Xiang Chinese (see Section 4.4), only plosives were selected in this study.

5.3.2 Participants

A total of 40 native speakers of Shuangfeng Xiang Chinese participated in the experiment and were recorded. However, only 37 participants were selected as qualified participants as the speech production of 3 participants turned out to be problematic due to an abundance of disfluency, which made speech segmentation difficult. Among the selected participants, there were 22 participants consisting of the old generation (9 males and 13 females born between 1950 and 1968, *M* = 58 years, *SD* = 6 years), and 15 participants consisting of the young generation (7 males and 8 females born between 1975 and 1990, *M* = 35 years, *SD* = 6 years). In addition to Shuangfeng Xiang Chinese, all participants were able to speak Standard Chinese but with different levels of proficiency. Younger participants generally had a higher level of proficiency than older participants. However, according to their self-reports, all considered Shuangfeng Xiang Chinese as the first and dominant language.

5.3.3 Procedure

The recordings were conducted in a quiet room in Shuangfeng county for all participants. The recording procedure was identical to that conducted for Lili Wu Chinese (see Section 3.3.3). Acoustic recordings and simultaneous EGG signals were coded in WAV format and digitized at a sampling rate of 44,100 Hz via the Field Phon (Feifeng) program. EGG data were recorded via Laryngograph microprocessor EGG-D800 connected to a laptop computer. The acoustic signal (Channel 1) and the simultaneous EGG signal (Channel 2) were recorded to separate channels.

In total, 1,110 tokens were collected (15 stimuli × 2 repetitions × 37 participants). The participants were asked to pronounce each word at their normal speaking rate. The whole recording of each participant usually lasted for 20 to 30 minutes. All were paid the equivalent of 10 euros in local currency for their participation.

5.3.4 Measurements

Similar to Lili Wu Chinese, the measurements of three sets of parameters were taken.

First, *f0* in Hz was automatically measured at twenty equidistant points over syllables starting from the first regular vocal pulse to the end of the syllable via VoiceSauce (Shue et al., 2011) using the Straight method (Kawahara et al., 1998). Furthermore, in order to eliminate the pitch range differences due to gender and to plot *f0* contours for visual inspection, the raw *f0* values at all points were normalized using the within-speaker *z*-score (Rose, 1987).

Second, the raw values of VOT of all voiced onsets were measured. Following Lisker and Abramson (1964) and Abramson and Whalen (2017), the point of voicing onset was visually determined by the first sign of a voice pulse in the waveform, which corresponds to the low-frequency voice bar in the spectrogram. The instant of release was found by fixing the point where the pattern showed an abrupt change in the waveform, which was associated with the sudden onset of energy in the spectrogram. The negative VOT was defined as the presence of voicing lead through much or all of the closure and was measured from the onset of voicing during plosive closure to closure release. It is worth noting that in a couple of tokens, voicing started but then ceased before the closure was released. Such tokens, therefore, had both a voiced and voiceless closure interval, which were also observed in some other languages (Abramson & Whalen, 2017; Coetzee et al., 2018). In this study, any token with closure voicing, whether full or partial, was therefore treated as having lead VOTs.

Third, contact quotient (CQ) values measured over the one-third, middle, and two-thirds point of the vowel were taken for the current study using EGGWorks (Tehrani, 2009). Identical to the method applied to Lili Wu Chinese, CQ was measured using the Hybrid method (see Section 3.4.2). Given a more direct reflection of articulatory measurements in observing laryngeal function (see Section 1.4.3), acoustic measurements (e.g., H1*–H2*) were no longer taken in this study.

5.3.5 Statistical analyses

The main purpose of this study is to examine the applicability of the two views debating on the phonetic properties shared by voiced and voiceless onsets, which condition the low-rising *f0* contours. Furthermore, another point worth exploring is that if any sound change is ongoning or have taken place in Shuangfeng Xiang Chinese. In order to achieve both goals, growth curve analysis (GCA), generalized linear mixed-effects models (GLMMs) and linear mixed-effects models (LMMs) were applied to compare normalized *f0* contours, VOT, and CQ, receptively. The procedure of modeling followed the general procedure illustrated in Section 3.3.5.

In this study, for statistical modeling of each dependent variable (i.e., *f0*, VOT, and CQ), the main effects of three independent variables, namely Consonant (aspirated vs. unaspirated vs. voiced), Generation (old vs. young), and their interaction (Consonant * Generation) were of particular interest. In addition, there were four control variables, namely Vowel (high vs. non-high), Gender (male vs. female), Place of Articulation (labial vs. alveolar vs. velar), and Repetition (first vs. second).

For *f0* contours, before evaluating all independent variables, to avoid overfitting, the best shape for capturing the changes of overall lexical contours was first determined. Given that T2 had a more convex contour shape (see Section 4.2), the model having a simple linear shape was then compared with the one having a curved shape. The procedure followed the way used in Section 3.3.5.

The distribution of the raw VOT values of all voiced onsets was demonstrated first. The percentage of tokens realized with negative VOT values was then calculated. In order to assess whether the productions of the old generation differ from the young generation in VOT, depending on the results of the VOT values, different types of LMMs were adopted. If all data had negative VOTs only, then LMMs using the *lmer* function would be applied. If there were positive VOTs, GLMMs with a logistic link function using the *glmer* function would be applied. In order to run GLMMs, following the procedure conducted in Coetzee et al (2018), positive VOT productions were coded as '1' and negative ones as '0', forming a new dependent variable labeled 'VOT-Index'. Models were then built on the basis of the general approach illustrated early.

As for the analysis of the CQ values, in order to assess whether different positions over the vowel differed from each other, Position (one-third vs. middle vs. two-thirds), Consonant, Generation, and their interactions were also taken as independent variables. If there was a significant three-way interaction, the dataset would be further divided into 6 subsets according to different generations (3 positions × 2 generations) and the position where the CQ measurements were made. Within each subset, LMMs were built following the general method. If Consonant significantly improved the model fit, multiple *post hoc* tests using Tukey's honestly significant difference (HSD) were further run to examine the CQ differences of vowels after different onsets.

The method of model comparisons and data visualization were identical to those employed in Section 3.3.5.

5.4 Results

The results consisted of three sections, corresponding to the three measurements (i.e., *f0*, VOT, and CQ). For the sake of simplicity, 'A' represented measured values with voiceless aspirated onsets, 'U' with voiceless unaspirated onsets, and 'D' with voiced onsets.

5.4.1 *f0* contour

The results of model comparisons showed that the second-order polynomial model significantly improved the model fit [χ^2 = 2551.5, p < .001]. The second-order polynomial, therefore, was applied to all subsets for further analyses. Table 5.1 shows the results of the effects of Consonant, Generation, and their interaction on *f0* in all time terms. As we can see from Table 5.1, except for Generation in the intercept term, all other factors showed significant effects in all three time terms.

Table 5.1 Results (χ^2) of model comparisons for the effect of Consonant, Generation, and Consonant * Generation on *f0*. Parameter-specific *p*-values (superscript) were indicated by Pr(>Chisq). n.s.: not significant.

Time term	Factor	χ^2
Intercept	Consonant	$2328.22^{<.001}$
	Generation	$.01^{n.s.}$
	Consonant * Generation	$2369.18^{<.001}$
Linear	Consonant	$973.1^{<.001}$
	Generation	$104.2^{<.001}$
	Consonant * Generation	$1147.07^{<.001}$
Quadratic	Consonant	$14.73^{<.001}$
	Generation	$99.28^{<.001}$
	Consonant * Generation	$190.66^{<.001}$

Given the significant interaction between Consonant and Generation in all time terms, this interaction was further decomposed into two subsets according to each of the two generations. Figure 5.1 demonstrates the *f0* of vowels following all plosives in normalized time (on the x-axis) and *f0* (on the y-axis represented by *z*-score values). It is plotted separately for each of the three onsets and generation of the speakers. Both generations show a prominent difference between *f0* contours after U and D. That is, *f0* contours after U are overall higher than those after D. However, *f0* contours after A vary across generations. As shown in Figure 5.1(a), for older speakers, *f0* contours after both A and D almost overlap. While for younger speakers, as shown in Figure 5.1(b), the *f0* contour after A tends to have a higher *f0* onset (above 0), which leads to a less acute rise to the syllable offset. However, the overall shape of the trajectory after A is still similar to that after D. It is also worth noting that all three *f0* contours of both generations converge to the same target towards the syllable offset, although it is slightly lower among younger speakers (around 0.5) than among older speakers (around 1).

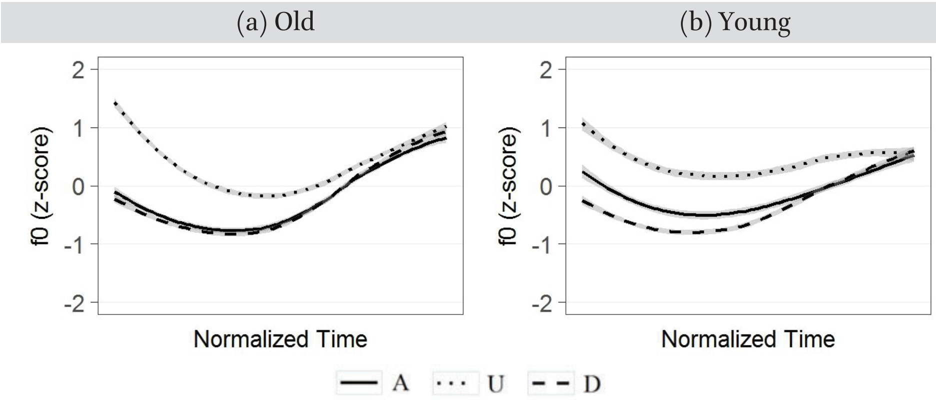

Figure 5.1 The *f0* realization of T2 for two generations. Mean normalized *f0* (gray areas indicate ± *SE*) after three laryngeal contrasts of plosives (i.e., A: voiceless aspirated, U: voiceless unaspirated, and D: voiced) based on 20 *f0* measurements taken at equally spaced intervals across the vowel. The x-axis represents the time-normalized duration of the vowel. The y-axis represents the *f0*-normalized value in the form of the within-speaker *z*-score. (a) and (b) panels show old and young generations' productions, respectively.

In order to further quantify the differences in the *f0* of different laryngeal contrasts, two sets of GCA models were run for each generation. Table 5.2 summarizes the results of both generations. The final models are attached in Appendix II-2.

Table 5.2 Results (β) of the effect of Consonant on *f0*. Reference: voiced onsets (D). n.s.: not significant.

Generation	Consonant	Intercept	Linear	Quadratic
Old	U	$.59^{<.001}$	$-2.05^{<.001}$	$.52^{<.05}$
	A	$.02^{n.s.}$	$-0.3^{n.s.}$	$-0.05^{n.s.}$
Young	U	$.7^{<.001}$	$-1.7^{<.05}$	$-0.3^{n.s.}$
	A	$.19^{n.s.}$	$-0.85^{<.05}$	$-0.01^{n.s.}$

As shown in Figure 5.1, for both generations, the *f0* contours after D consistently differ from that after U. For older speakers, the overall *f0* contours after U and D have a lower mean, steeper ramp, and sharpness of the centered peak. This was confirmed by significant effects in all three time terms, as shown in Table 5.2. A similar difference was also observed for younger speakers with significant results in both intercept and linear terms. Visual observations of the generational difference of the relationship between *f0* contours after A and D were confirmed by statistical results. As shown in Table 5.2, for the old generation, there was no significant difference in the *f0* contours of D and A. However, a marginally significant effect in the linear term was observed for the young generation [$\beta = -0.85$, $p < .05$], which likely resulted from the slightly higher *f0* onset after A. However, compared to the *f0* contour after U, the *f0* contour after D showed a more comparable pattern than that after A.

Generally speaking, all three curves showed a concave contour and converged to the same target towards the syllable offset. However, the trajectories of D presented an overall lower *f0* than those of U, which more or less confirmed the impressionistic description of the 'voicing contrast' view, which argues for a lower *f0* onset after voiced plosives (Xiang, 1960; Yuan, 1960: 115; *Zihui*, 1989; Chen, 2004). Nevertheless, there is one crucial discrepancy. For speakers of both generations, the *f0* contours after A were more similar to that after D. Note that for the younger speakers, there was a marginally significant difference in the linear term, suggesting a less steeper ramp of the *f0* contour after A. These findings thus challenge both views. The 'voicing contrast' view does not fully predict the findings, as this view predicts similar *f0* contours after all three onsets. The 'phonation contrast' view is not supported by the data either as this view predicts a tonal contrast associated with voiceless (i.e., unaspirated and aspirated) vs. voiced onsets.

5.4.2 VOT

Not all voiced onsets had a negative-VOT realization. Each token with voiced onsets was designated as '[negative VOT]' if there was any voicing during the closure phase of the plosive, and as '[positive VOT]' otherwise. As shown in the histograms of raw VOT values for voiced onsets in Figure

5.2, tokens with white bars represent voiced productions of plosives [negative VOT], while gray bars represent devoiced productions of plosives [positive VOT]. We can see that voiced onsets were not consistently realized with negative VOTs for both generations. On the one hand, tokens classified as [negative VOT] employed a region between –200 ms and 0 ms in the old generation, but a more limited region between –160 ms and 0 ms in the young generation. On the other hand, most tokens classified as [positive VOT] produced by the old generation showed a range from 0 ms to 40 ms, while a wider range from 0 ms to 80 ms was produced by the young generation. A calculation based on Figure 5.2 showed that in total 269 tokens were realized with [negative VOT]. The specific tokens produced by different generations, however, were different. The old generation turned out to have a higher rate (181/220 = 82%) of tokens with a negative-VOT realization than the young generation (88/150 = 59%). Moreover, it is clear that there was a shift to less VOT lead and more VOT lag for the young generation, which suggests that younger speakers tend to produce shorter negative VOTs than older speakers.

Given that not all voiced plosives had a negative-VOT realization, to further assess the contribution of Generation to the rate of positive-VOT productions, a series of GLMMs was run over the VOT values of these tokens with [negative] productions coded as '0' and [positive] productions as '1' (i.e., VOT-Index). As shown in Table 5.3, the only significant factor was found in Generation [$\chi^2 = 7.15$, $p < .01$]. All control variables and additional random structures did not show significant effects. Table 5.4 summarizes the results of generational differences. The rate of positive-VOT productions produced by the young generation was significantly higher than the rate produced by the old [$\beta = 1.76$, $p < .01$], confirming that younger speakers devoiced more than older speakers.

To sum up, irrespective of generations, voiced onsets with the low-rising tonal contour (i.e., T2) in Shuangfeng Xiang Chinese were partially realized with negative VOTs. This result generally confirmed the finding reported in Zhu and Zou (2017). However, a generational difference was also observed. The young-generation speakers significantly produced fewer negative-VOT tokens and shorter negative-VOTs than the old-generation speakers.

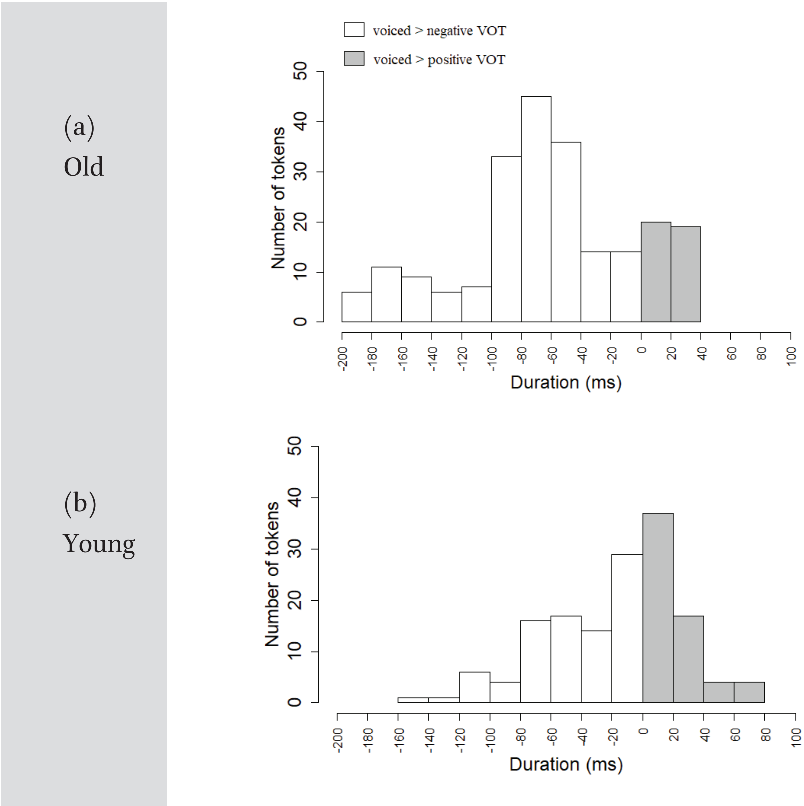

Figure 5.2 Histograms representing the distribution of VOT for productions of plosives for (a) old and (b) young generations. Each 20 ms bin includes all tokens that have a VOT value equal to or lower than the bin label (and higher than the label of the immediately lower bin). White bars represent voiced productions of plosives (negative VOT); gray bars represent devoiced productions of plosives (positive VOT).

Table 5.3 Results of model comparisons to the data of VOT-Index. n.s.: not significant.

Factor	*df*	logLik	χ^2	*p*
Generation	3	−176.24	7.15	< .01
Place	5	−176	.49	n.s.
Vowel	4	−176.2	.08	n.s.
Gender	4	−175.38	1.73	n.s.
Repetition	4	−175.41	1.67	n.s.
Intercept: Item	4	−176.24	0	n.s.
Slope: Speaker	Model failed to converge			

Table 5.4 Results of the effect of Generation on VOT-Index. Reference: Old generation.

Factor	Estimate (β)	*SE*	*z*	*p*
Intercept	−2.31	.47	−4.89	< .001
Generation: Young	1.76	.66	2.68	< .01

5.4.3 CQ

Table 5.5 shows the results of the effects of Consonant, Generation, Position, and their interactions on CQ. As we can see, both Consonant and Position significantly improved the model fit. Significant effects were observed in two two-way interactions [Consonant * Generation; Consonant * Position] and one three-way interaction [Consonant * Generation * Position]. Generation and its interaction with Position, however, did not show significances.

Table 5.5 Results of model comparisons for the effect of Consonant, Generation, Position, and their interactions on CQ. n.s.: not significant.

Factor	*df*	logLik	χ^2	*p*
Consonant	5	3491	73.69	< .001
Generation	6	3491.7	1.34	n.s.
Position	6	3574.5	167.09	< .001
Consonant * Generation	9	3582	14.97	< .01
Consonant * Position	8	3601.9	54.79	< .001
Generation * Position	10	3602.7	1.53	n.s.
Consonant * Generation * Position	14	3611.7	19.62	< .01

Given the interactions of Consonant with both Generation and Position, the entire dataset was further divided into 6 subsets according to each of Generation and Position (2 generations × 3 positions). Figure 5.3 presents the mean of CQ values of vowels following all three onsets, plotted separately for each of the three positions and generation of the speakers. An inspection of Figure 5.3 indicates that within each generation, U consistently introduces the overall highest CQ of vowels while A introduces the overall lowest CQ of vowels. D has intermediate CQ values, but there is again a generational difference in the specific patterns of CQ values. For older speakers, it is closer to its voiceless counterparts, but for younger speakers, it is closer to its aspirated counterparts. In addition, all differences wane after the midpoint (i.e., 50%) and almost vanish after the two-thirds position of vowels.

In order to further quantify the differences in CQ values after different laryngeal contrasts, a series of LMMs was run for each subset. As observed in Table 5.6, for both generations, Consonant consistently improved the model at P1 (one-third) and P2 (middle) of vowels, but they failed to improve the model fit at P3 (two-thirds). It indicated that none of the CQ differences triggered by laryngeal contrasts involved the whole vowel and were maintained only during the first half of the vowel.

(a) Old (b) Young

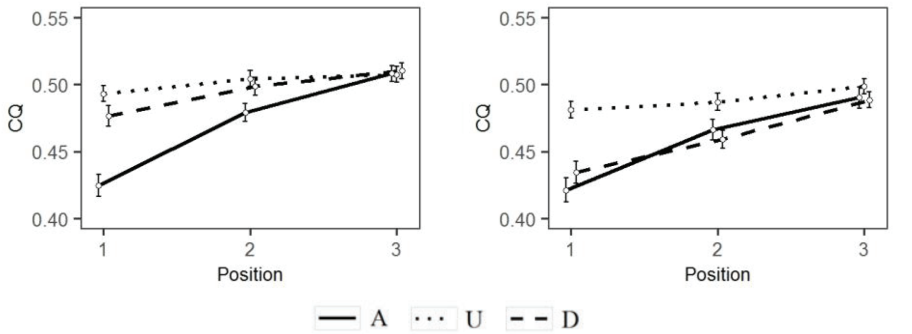

Figure 5.3 Mean CQ of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each generation. The x-axis represents the ratio of the vowel. The y-axis represents the CQ value. Error bars represent the standard error of the mean. (a) and (b) panels show old and young generations' productions, respectively.

Table 5.6 Results of the effect of Consonant on CQ at each position for each generation.

Generation	Position	*df*	logLik	χ^2	*p*
Old	P1	5	614.23	64.13	< .001
	P2	5	768.05	14.98	< .001
	P3	Failed to improve the model fit.			
Young	P1	5	516.55	52.32	< .001
	P2	5	558.81	13.87	< .001
	P3	Failed to improve the model fit.			

Given the significant effect of Consonant presented at P1 and P2 of vowels, in order to evaluate the CQ differences of vowels after different onsets, a series of *post hoc* tests using HSD was further conducted. Table 5.7 summarizes the results. For both generations, significant differences in CQ were observed between U and A. It is worth noting that the two generations exhibited a reversed relationship between CQ after D vs. that after the two voiceless onsets (i.e., U and A), which confirmed the visual inspection of Figure 5.3. For older speakers, CQ after D yielded more comparable patterns to that after U [old: D ≈ U], but showed stable differences from A [old: D ≠ A]. For younger speakers, however, CQ after D was more comparable to that after A [young: D ≈ A] but differed from that after U [young: D ≠ U]. The results further confirmed the change of the CQ pattern following D across generations. The final models to calculate the results of Table 5.7 are listed in Appendix II-3a. In addition, Appendix II-3b provides the results of pairwise comparisons. The results are consistent with the results of stepwise multilevel regression conducted in this section.

Table 5.7 HSD results (β) of the effect of Consonant on CQ at P1 (one-third) and P2 (middle) of vowels for each generation. Parameter-specific *p*-values (superscript) were indicated by Pr(>Chisq). n.s.: not significant.

Generation	Consonant	P1 (one-third)	P2 (middle)
Old	A vs. U	$.07^{<.001}$	$.03^{<.01}$
	D vs. A	$.05^{<.001}$	$.02^{n.s.}$
	D vs. U	$-0.02^{n.s.}$	$-0.01^{n.s.}$
Young	A vs. U	$.06^{<.05}$	$.02^{n.s.}$
	D vs. A	$.01^{n.s.}$	$-0.01^{n.s.}$
	D vs. U	$-0.05^{<.01}$	$-0.03^{<.05}$

In summary, there are three findings. First, the results showed that A introduced a higher magnitude of breathiness (indicated by lower CQ values) over the following vowel than both U and D. Second, the magnitude of the breathiness of vowels after D, however, varied depending on the speakers' age. Older speakers produced more comparable patterns of breathiness to that after U, while younger speakers produced more comparable patterns to that after A. This difference suggests that it is the younger speakers who started the trend of producing more breathiness in vowels after D. Last but not least, differences in phonation did not last over the whole vowel and vanished after the midpoint. Taken together, these findings cast serious doubts about both 'voicing contrast' and 'phonation contrast' views. The former neglects the phonatory differences of the laryngeal contrast, while the latter fails to predict greater breathiness after aspirated onsets. Furthermore, the phonatory contrast (i.e., breathy vs. modal) after voiced and unaspirated onsets was not consistently observed across generations.

5.5 Discussion

5.5.1 New light on the two views

The main purpose of this study is to tease apart two views on the phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets that condition similar low-rising *f0* contours. Generally speaking, the results suggest that neither the 'voicing contrast' view (e.g., Chao, 1935 [Yang, 1974]) nor the 'phonation contrast' view (Zhu & Zou, 2017) is fully correct.

The 'voicing contrast' view posits that the voicing contrast of onsets is the primary property that conditions the low-rising *f0* contours, and these *f0* contours are comparable ('initial-associated tonal merger', ITM). This view, however, is not supported by the results. First, voiced plosives in the low-rising tonal context were partially realized with negative VOTs. The young-generation speakers significantly produced fewer negative-VOT tokens and shorter negative-VOTs than the old-generation speakers. Moreover, there was a consistently lower *f0* contour after both voiceless aspirated and voiced onsets, compared to that after voiceless unaspirated onsets. This pattern was stable across generations. The prediction of comparable *f0* contours after different onsets (i.e., the one-tone prediction) did not bear out.

The 'phonation contrast' view proposes that the phonatory contrast of the onsets conditions the low-rising *f0* contours. It argues for a two-way phonatory contrast corresponding to the laryngeal contrast of voiceless vs. voiced onsets and posits a contrast of two low-rising tones (i.e., modal – /24/ vs. breathy – /13/; hence the two-tone prediction). This view is untenable as well. On the one hand, only younger speakers produced a higher magnitude of breathiness after voiced onsets, while older speakers did not. On the other hand, a higher magnitude of breathiness was consistently observed after aspirated onsets (the first half of the following vowel) across generations. However, the two generations produced comparable patterns of *f0*. In addition, the 'phonation contrast' view argues for two different rising tones: one after voiceless onsets and the other after voiced onsets. This prediction, however, did not bear out either. Instead, the results clearly suggest comparable *f0* contours after voiceless aspirated and voiced onsets, which are different from that after voiceless unaspirated onsets.

A subsequent issue is how to categorize the three low-rising contours within the tonal system of Shuangfeng Xiang Chinese. The answer to this question is couched in the potential correlations between *f0* contours and laryngeal properties (i.e., VOT and CQ).

Following the 'phonation contrast' view, a possibility is that there are two lexical tones based on the phonatory contrast after voiceless and voiced onsets (Zhu & Zou, 2017). However, we have observed that aspirated onsets consistently conditioned a higher magnitude of breathiness over the first half of the following vowel across generations. Furthermore, the phonatory state of voiced onsets is undergoing changes: only younger speakers produced a higher magnitude of breathiness after voiced onsets. Note that with regard to *f0* contours, there were no general differences. Conjointly, the results suggest that the low-rising *f0* contours produced by speakers from different generations are less likely to be associated with the phonatory contrast of onsets.

The alternative one-tone view (Chao, 1935 [Yang, 1974]) seems more likely for two reasons. First, all three *f0* curves showed a general rising contour and converged to the same target towards the syllable offset. In this way, the *f0* incongruence can be attributed to perturbation effects (i.e., C*F0*) triggered by different onsets. That unaspirated onsets co-occur with higher *f0* contours, while both aspirated and voiced onsets co-occur with comparable lower *f0* contours have been observed in Chinese dialects such as Lili Wu (Chapter 2) and Shanghai Wu Chinese (Chen, 2011). In addition, historically speaking, as suggested by studies such as Xiang (1960) and *Zihui* (1989), in Shuangfeng Xiang Chinese, low-rising-contour-carrying syllables with voiceless onsets (i.e., unaspirated and aspirated) consistently developed from the MC *Ru* tonal category, while syllables with voiced onsets originated from the MC *Ping* tonal category. The *f0* contours after both voiceless aspirated and voiced onsets, however, are comparable. This suggests that the phonological origin of tone-carrying syllables in MC has given in to the phonetic conditioning of the *f0* perturbation due to different onsets. Taken together, we conclude that the results of our study lend more support to the one-tone proposal.

5.5.2 The trading relationship between VOT and CQ: principal component analysis (PCA)

The generational differences observed in VOT and CQ seem to imply a sound change in Shuangfeng Xiang Chinese. Given the span of 82 years between the time when the two views were posited (i.e., 1935 vs. 2017), it is reasonable to argue that the two seemingly contradictory views represent the two stages of the realization of the voicing contrast in the low-rising tonal context, respectively. On the one hand, older speakers tend to produce more negative-VOT tokens and longer negative-VOTs but less breathiness (i.e., higher CQ values) of vowels following voiced onsets. On the other hand, younger speakers tend to produce fewer negative-VOT tokens and shorter negative-VOTs but more breathiness (i.e., lower CQ values) of vowels. Older speakers rely more on the negative VOT to signal the contrast of voiced onsets with the other counterparts, which aligns better with the 'voicing contrast' view. While younger speakers rely more on the breathiness cue, which is better predicted by the 'phonation contrast' view. Such differences between the two generations suggest a changing role of VOT and CQ to signal the voicing contrast of obstruent onsets in Shuangfeng Xiang Chinese.

In order to further investigate the relationship between VOT and CQ, a principal component analysis (PCA) was adopted. PCA is a descriptive tool that helps to reveal the internal structure of variables in a dataset (Jolliffe, 2002; Jolliffe & Cadima, 2016). The central idea of PCA is to reduce the complexity of a dataset consisting of a large number of interrelated variables while retaining as much as possible of trends and patterns present in the dataset. It is achieved by transforming the original dataset into a new dataset with fewer variables, namely, the principal components (PCs). PCs are uncorrelated and ordered so that the first few (usually two, PC1 and PC2) retain most of the variation present in original variables and act as summaries of variables. The PCs are defined as a linear combination of the data's original variables and are chosen to minimize the total distance between the data and their projection onto a certain PC by maximizing the total distance from the projected points to the origin (Lever et al., 2017). In detail, PC1 consists of the maximal sum of squared distances (SS) from the projected points of original variables onto an axis to the origin. One PC is uncorrelated with all previous PCs. Projection onto PC2 is geometrically orthogonal to projection onto PC1. SS from the projected points of original variables onto the orthogonal axis to PC1 is hence referred to as PC2. There is no physical meaning of both axes as they are combinations of original variables.

In this study, there were two original variables, namely VOT and CQ. A PCA and its corresponding biplot were obtained using the *gridExtra* (Auguie, 2015), *gdata* (Warnes et al., 2017), and *ggbiplot* (Vu, 2011) packages in R. Given that all differences of phonation vanished after the midpoint, the mean of CQ measured from P1 (one-third) and P2 (middle) was calculated as a new CQ variable. In total, the dataset consisted of two measurements of the voiced onsets on 37 speakers from two generations. Of the two variables, VOT presented the timing of the onset of voicing during plosive closure to closure release, while CQ was the measurement related to the magnitude of breathiness. Figure 5.4 gives the biplot for the correlation matrix PCA of the dataset and aims to account for the influences of the original variables. The variable markers (i.e., VOT and CQ) are displayed as arrows and the speaker markers as dots in black (Old) and gray (Young) indicating two different generations, respectively. The length of arrows represents how well the variable explains the distribution of the data: longer, stronger. The cut-point of a perpendicular from a point to an arrow shows the value of the variable the arrow represents. That is, along the arrow direction of VOT, speakers produce longer VOTs; along the arrow direction of CQ, speakers have greater CQ values, *vice versa*. The angle between variables implies the correlation. A small angle implies a positive correlation, a large angle suggests a negative correlation, and a 90° angle indicates no correlation between variables. The concentration ellipse in 68% probability (default) is plotted for each generation.

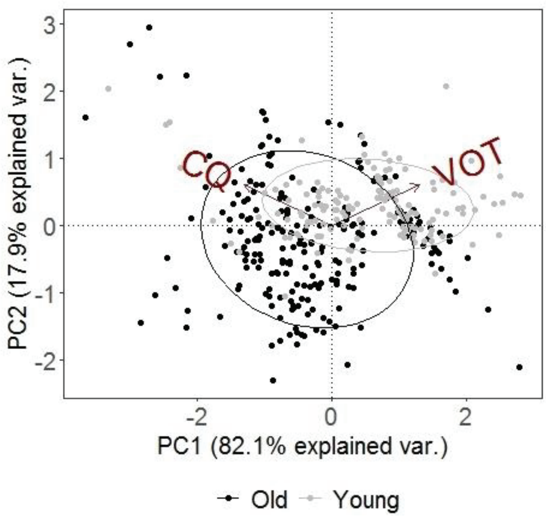

Figure 5.4 Biplot for the VOT-CQ data (correlation matrix PCA).

As shown in Figure 5.4, the first two PCs account for 82.1% (PC1) and 17.9% (PC2), respectively, so the two-dimensional biplot of the 37 speakers is a very good approximation to the original dataset.[39] It represents 100% of the total variation. The concentration ellipses of speakers of different generations are distributed separately. Dots representing the old generation (black) mainly concentrate near the left of the origin (0, 0), which means older speakers tended to have higher CQs but more negative VOTs. However, younger speakers tended to have lower CQs but more positive VOTs, as reflected by those dots (gray) near the right of the origin. In addition, VOT and CQ bear a negative correlation, as evidenced by a larger angle between the two arrows. Such a negative correlation was further confirmed by a Pearson's test [$r = -0.64$, $p < .001$], which suggests speakers who produced more positive VOTs tended to have lower CQs.

The results of PCA suggest that speakers of different generations utilize different properties to signal the voiced category in the low-rising tonal context. The trading relationship between VOT and CQ is prominent: the older speakers tend to employ more negative VOTs, while the younger speakers have changed to attach much more importance to the cue of lower CQs.

Trading relationships of acoustic cues are commonly believed to play a crucial role in the maintenance/change of language-specific phonological contrasts (e.g., Repp, 1982; Stevens et al., 1986; Toscano & McMurray, 2010; Stevens & Keyser, 2010; Clayards, 2018). Some studies in English have already shown, for onset voicing contrast, that VOT can trade against multiple cues, such as the onset frequency of F1 transition, the amplitude of the noise preceding the onset voicing, and onset *f0* (see Repp, 1982 and Kingston et al., 2008 for comprehensive reviews). However, in Shuangfeng Xiang Chinese, the phonatory state (i.e., breathiness) can also be traded against VOT. Younger speakers tend to employ more breathiness (i.e., lower CQ) of the following vowel to signal voicing

[39] The contribution of PCs to the total variation is calculated by variation for PCs. Please see Chapter 2 in Jolliffe (2002) for more details.

contrast. The increased breathiness can be viewed as an enhancement strategy (Garrett & Johnson, 2013: 79), where speakers enhance the magnitude of an existing feature in order to increase perceptual distance and maintain phonological equivalence. It is probable that the loss of phonetic voicing during the closure of voiced stop onsets threatens the voicing contrast, hence leading to enhanced breathiness over the following vowel among younger speakers in order to maintain the three-way laryngeal contrast in the low-rising tonal context.

The observed trading relationship between VOT and CQ in Shuangfeng Xiang Chinese is quite similar to an ongoing change reported in a Tai dialect spoken in Cao Bằng, Vietnam (Pittayaporn & Kirby, 2017). In this dialect, negative VOT is the primary acoustic feature distinguishing /b̤/ from /p/ of an older male speaker (75 years old). However, a younger female speaker (57 years old), in contrast to the older speaker, had a positive VOT of /b̤/ similar to voiceless unaspirated /p/, but distinguished it from /p/ via breathy voice. It suggests that speakers of different tonal languages can utilize a similar trading relationship between the negative VOT and the breathier phonation to maintain the voicing contrast. One reason that the two phonetic features can switch probably has to do with their similar compatibilities with the lowering effect on *f0*. Both negative VOT and breathier phonation have been widely reported to universally favor low tones across tonal languages (negative VOT – low tone: Hombert, 1978; breathiness – low tone: Gordon & Ladefoged, 2001), although breathier voices are not always inextricably associated to lower tones, such as in Yi (Kuang, 2013b) and Lili Wu Chinese (Chapter 3).

5.6 Conclusion

The primary goal of this chapter is to examine two views, namely the 'voicing contrast' view (e.g., Chao, 1935 [Yang, 1974]) and the 'phonation contrast' view (Zhu & Zou, 2017), concerning the phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets that condition the low-rising *f0* contours. To sum up, neither of them can fully account for the results observed in this study. In terms of VOT, voiced onsets were partially realized with negative VOTs. The old-generation speakers significantly produced more negative-VOT tokens and longer negative-VOTs than the young-generation speakers. In terms of CQ, on the one hand, a higher magnitude of breathiness was consistently observed after aspirated onsets (the first half of the following vowel) across generations. On the other hand, a higher magnitude of breathiness after voiced onsets (the first half of the following vowel) was only observed over tokens produced by younger speakers. In terms of *f0* contours, although all evidence indicated one underlying low-rising tone in Shuangfeng Xiang Chinese, different onsets did show *f0* perturbation effects over the initial part of the same low-rising tonal contour. Both voiceless aspirated and voiced onsets tended to co-occur with lower *f0* contours, different from voiceless unaspirated onsets (which co-occurred with an overall higher *f0* contour).

The results suggest that speakers from different generations can weigh the phonetic properties differently to signal the voiced category in the low-rising tonal context in Shuangfeng Xiang Chinese. The increased breathiness of vowels following voiced onsets has been adopted as an articulatory enhancement by younger speakers to maintain the three-way laryngeal contrast of obstruents.