Consonant and lexical tone interaction: Evidence from two Chinese dialects

Shi, M.

Author: Shi, M
Title: Consonant and lexical tone interaction: Evidence from two Chinese dialects
Issue Date: 2020-06-04

Chapter 3 Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese[22]

3.1 Introduction

In this chapter, the uncommon consonant-tone interaction (C-T interaction) in Lili Wu Chinese, namely the 'aspiration-induced tonal split' (ATS) phenomenon, will be examined in detail. Particularly, two long-standing debates lingering in previous literature will be discussed. They are i) Is ATS an ongoing change or a completed change? and ii) Is aspiration or breathiness synchronically related to ATS?

Mergers and splits are generally believed to be the two main processes of phonemic change in the development of language (Labov, 1994; Campbell, 2013). In previous studies, mergers have drawn a great deal of attention to a variety of linguistic properties (e.g., Labov, et al., 2006 for vowels in North American English; Harris, 1969 for consonants in Latin American Spanish; and Alan, 2007 for tones in Cantonese). The opposite process, namely splits, defined as 'the division of a preexisting phoneme to create a new phonemic distinction' (Labov, 1994: 331), is commonly argued as a conditioned, complicated and unusual event. Nevertheless, tonal split has been observed in many Asian languages with an already established tonal system (Haudricourt, 1972; Brown, 1975; Brunelle & Kirby, 2016). A great deal of work has been done on two aspects. One is diachronic reconstructions on how the tonal inventories of proto-languages evolved into those of modern languages (e.g., Haudricourt, 1972; Li, 1977); the other is of phonetic explanations for

[22] A version of this chapter has been accepted for publication by *The Journal of the Acoustical Society of America*. Shi, M., Chen, Y., & Mous, M. (forthcoming). Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese.

mechanisms of tonal split using acoustic or perceptual experiments (e.g., Abramson & Erickson, 1978; Rischel, 1986; House & Svantesson, 1996; Thurgood, 2007). However, very few studies have access to a living language where the implementation of tonal split can be systematically and synchronically observed without indirect assumptions or inferences.

Fortunately, the language of focus in this chapter, Lili Wu Chinese provides us with just such an opportunity to fill in the hiatus for our understanding of tonal split. The language has been reported to show fundamental frequency (*f0*) lowering after voiceless aspirated onsets relative to voiceless unaspirated onsets, and this lowering effect has been argued to be phonologized and to have resulted in the splitting of an existing lexical tone and the forming of new tonal categories (after Chao, 1928).

3.1.1 Wu Chinese and tonal split with aspiration onsets in Lili Wu Chinese

Wu Chinese is commonly classified as one of the seven major dialect groups within the Sinitic branch of the Sino-Tibetan language family (Yuan, 1960).[23] The most prominent feature of Wu Chinese is the existence of a three-way laryngeal contrast in obstruents, known as voiceless unaspirated, voiceless aspirated, and voiced, respectively (Chao, 1967). The three-way laryngeal contrast has different manifestations in initial position as opposed to medial position (see Chen, 2011 and references therein). In initial position, these obstruents vary in their phonation from clearly modal (voiceless unaspirated), to aspirated (voiceless aspirated), to breathy (voiced). In medial position, however,

[23] Wu Chinese is spoken by approximately 70 million people who reside in the area of southern Jiangsu (江苏南部), Shanghai (上海) and Zhejiang (浙江) provinces, as well as some part of Anhui (安徽), Jiangxi (江西), and Fujian (福建) provinces in the People's Republic of China (Zhengzhang & Zheng, 2015).

voiced obstruents are fully voiced, leading to a three-way laryngeal distinction in voice onset time (VOT).

Another distinct feature of modern Wu dialects is that the majority of them are known to still preserve, to a large extent, an eight-way tonal system that developed from Middle Chinese (MC) (see Section 2.2 for more details). The four-way tonal contour contrast of MC, traditionally labeled as *Ping* (level), *Shang* (rising), *Qu* (departing), and *Ru* (entering), split into a new dual-register, eight-tone system, conditioned by onset consonants. In the majority of modern Wu varieties, generally speaking, syllables with voiceless onsets are argued to be produced within a higher *f0* range, referred to as the high register, while voiced onsets condition a lower *f0* range, known as the low register.

Let us take Lili Wu Chinese as an example. As illustrated in Section 2.2.1, there are eight lexical tones in Lili Wu Chinese. As shown in Table 3.1, the eight lexical tones of Lili Wu Chinese exhibit a remarkable co-occurrence pattern between consonantal onsets and lexical tones. Syllables with voiceless onsets only license tones that start in the high register while those with voiced onsets allow tones that are in the low register. In Lili Wu Chinese, as shown in Table 3.2, syllables beginning with voiceless aspirated obstruents in three of the MC tonal categories (i.e., *Shang*, *Qu* and *Ru*), are reported to introduce distinctively lower tones than syllables beginning with unaspirated obstruents (see a comprehensive review in Wang, 2008). This unusual tonal phenomenon is termed ‘aspiration-induced tonal split’ (ATS) by sinologists. ATS is not considered to be a widespread phenomenon that occurred across Chinese languages/dialects. Existing literature rather indicates that ATS has been reported for varieties spoken in only 39 cities/counties of China (Yue Xu, 2013), mainly including dialects of the Wu, Gan, and Xiang groups as well as some languages belonging to the Kra-Dai and Hmong-Mien language families (Ho, 1989; F. Shi, 1998; Chen, 2005). Note that in Lili Wu Chinese, this phenomenon is absent in words within the MC tonal category *Ping*, where syllables with voiceless aspirated obstruents still bear a high tone.

Table 3.1 Co-occurrence constraints on onset-tone combinations in Lili Wu Chinese. The tonal transcription system developed by Chao (1930) was adopted. In this system, 5 indicates the highest end of a speaker's pitch range into levels and 1 the lowest. A single number refers to cases where the tone-carrying syllables have a short duration and only co-occur with the glottal coda /ʔ/.

MC / Onset	*Ping*	*Shang*	*Qu*	*Ru*
Voiceless unaspirated (high register)	T1 high–level 44	T3 high–falling 53	T5 high–dipping 423	T7 short–high–level 5
Voiced (low register)	T2 low–rising 13	T4 low–level 22	T6 low–dipping 213	T8 short–low–level 3

Table 3.2 ATS in Lili Wu Chinese.

MC / Onset	*Ping*	*Shang*	*Qu*	*Ru*
Voiceless unaspirated	high–level 44	high–falling 53	high–dipping 423	short–high–level 5
Voiceless aspirated	still high	lower	lower	lower

3.1.2 Two debates on ATS

Most studies on ATS in Chinese dialects are impressionistic descriptions. To my knowledge, Lili is thus far the only dialect that has been investigated in a number of studies. According to Wang (2008), the consensus on the condition of the occurrence of the ATS phenomenon is that ATS is only present in words within non-*Ping* tonal categories (i.e., *Shang*, *Qu*, and *Ru*). However, researchers differ greatly in their analyses of the ATS phenomenon, which has resulted in various debates. Among them, two debates have long been a focus. The first debate regards the progress of lexical tones beginning with aspirated onsets – Is tonal split an ongoing change or a completed one? The second debate concerns the phonetic properties of ATS, namely, is aspiration or breathiness synchronically related to ATS? The two debates focus on different aspects of ATS in Lili Wu Chinese. From a broader perspective, both, however, can be regarded as a facet of what Weinreich et al. (1968) call the 'constraints' in their foundational work on language change. A possible approach for investigating constraints of sound change is to study asymmetries between sound changes and phonetic patterns. As argued in Garrett and Johnson (2013), asymmetries in sound changes usually reflect asymmetries in sound patterns. In Lili Wu Chinese, given the asymmetry of ATS conditioned by different MC tonal categories, it motivated me to pay more attention to the incongruent patterns of lexical tones between tonal categories (Debate one) and its possible phonetic biases (Debate two). Exploring both debates can not only further sharpen our understanding of the lexical tonal system of Lili Wu Chinese but it is also pivotal for answering general issues on constraints of sound change, especially of the tonal-split phenomenon.

3.1.2.1 Debate one: ongoing or completed?

Sound change in progress can be synchronically observed (Labov, 1994). However, the estimate of the stage of completion at which one particular linguistic change finds itself at any given time is always debatable. The first point of contention concerns the independence of lexical tones beginning with voiceless aspirated obstruents, namely whether tonal split is an ongoing change or a completed one. Two opposing views have been proposed.

On the basis of two speakers (one male and one female without report of their exact ages), Shi (1992) argues that tonal split of syllables beginning with aspirated onsets is an ongoing change (hereafter referred to as the 'ongoing change' view). Specifically, Shi (1992) claims that lexical tones beginning with aspirated onsets are independent of lexical tones beginning with unaspirated onsets and are merging toward lexical tones beginning with voiced onsets. Shi (2008: 227) further posits a stepwise lowering of lexical tones beginning with aspirated onsets across three MC tonal categories (i.e., *Shang*, *Qu,* and *Ru*) which Shi assumes should be observed from speakers of different generations within the speech community. This assumption is then confirmed by Zhu and Xu (2009), based on acoustic data collected from speakers of two generations (three old speakers whose ages are 60, 66, and 61 years; and three young speakers whose ages are 35, around 30, and 28 years) in Songling Wu, a variety spoken in the administrative-level town of the Wujiang area with a similar phenomenon of tonal split to Lili Wu Chinese. They claim that two old speakers show a three-way contrast of lexical tones conditioned by initial onsets in the MC *Ru* tonal category, while the merging of lexical tones beginning with aspirated and voiced onsets is happening in the speech of one young speaker.

In contrast to the 'ongoing change' view, Shen (1994) maintains that there is no so-called ongoing change, but rather a completed merger between lexical tones beginning with voiced and aspirated onsets in Lili Wu Chinese (hereafter referred to as the 'completed change' view). Based on acoustic data obtained from two young speakers (high school students without exact ages) of Lili Wu Chinese collected in 1985, Shen (1994) claims that a completed merger is observable in three MC tonal categories, namely *Shang*, *Qu*, and *Ru*.

3.1.2.2 Debate two: aspiration or breathiness?

The phonetic properties of speech sounds undergoing change are also a widely discussed issue. In Lili Wu Chinese as well as other dialects featuring ATS, onset aspiration seems to be the most prominent feature to have actuated the change. However, researchers differ in their analyses of how onset aspiration is responsible for ATS, which forms the second debate.

A traditional view is that tonal split in Lili Wu Chinese is directly related to onset aspiration (e.g., Ye, 1983; Wang, 2008) (hereafter the 'aspiration' view). Chao (1928) first delineates the phonological condition (i.e., voiceless aspirated onset) for ATS but without mentioning its phonetic substances and mechanisms. This view, however, has been widely adopted by a lot of impressionistic reports after Chao. They tend to attribute ATS directly to the distinction between the voiceless unaspirated and the voiceless aspirated onsets. For example, Wang (2008) explicitly argues that ATS results from onset aspiration via inhibition of vocal fold vibration. Generally speaking, the 'aspiration' view is in line with the idea that the development of tones may result from different articulatory reinterpretations of segmentally-induced perturbations in intrinsic *f0* (Haudricourt, 1954; Hombert et al., 1979).

The alternative view argues for a phonation-based account, which emphasizes an important role of voice quality during the process of tonal split (see Thurgood, 2002, 2007; Chen et al., 2017 for reviews of such work). It claims that breathiness due to onset aspiration is the primary trigger of ATS (Sagart, 1981; Ho, 1989; F. Shi, 1998; Zhu & Xu, 2009; Hirayama, 2010; Chen, 2014) (hereafter the 'breathiness' view). This view seems to have been initiated by Sagart (1981), who assumes a correlation between the breathy phonation and the low tonal onset. Subsequent studies attempt to provide more elaborate interpretations. For example, Zhu and Xu (2009) found that the magnitude of breathiness at the 30–40 ms interval of vowels after aspirated onsets is higher than that after unaspirated onsets. Hirayama (2010) further argues that a higher magnitude of breathiness can be observed throughout the entire vowel (or most part of the vowel). Recently, Chen (2014) attempts to attribute the correlation between breathier phonation types and lower tones to the intrinsic aerodynamic property of an aspirated stop release suggested by Ohala (1978). The 'breathiness' view is in line with the observation that non-modal phonation types and *f0* contours correlate with one another (e.g., Laver, 1994; Gordon & Ladefoged, 2001; see a comprehensive review in

Esling & Harris, 2005). Increasing evidence also shows that breathier phonation types are commonly associated with *f0* lowering cross-linguistically (e.g., Gordon & Ladefoged, 2001; see also a detailed review in Kuang, 2013b).[24]

3.2 The current study

The overarching goal for this current study is to shed light on the two aforementioned debates concerning tonal split in Lili Wu Chinese. Generally speaking, previous studies on both debates suffer from manifold inadequacies. For the first debate, the 'ongoing change' view fails to rule out the speaker-specific possibility of a three-way contrast of lexical tones as a function of the tone-bearing syllable onset, due to the small sample size recruited in each generation (e.g., Zhu & Xu, 2009). The 'completed change' view, however, draws its conclusion on data from speakers with a limited age range (Shen, 1994).

With respect to the second debate, adherents to the 'aspiration' view base their analyses of tonal categories on impressionistic observations only (after Chao, 1928). Furthermore, it is known that the *f0*-lowering effect of aspirated onsets varies across languages and speakers of

[24] In the existing literature, three non-modal phonation types, namely, slack/lax voice, breathy voice, and whispered voice have been argued to be produced with a larger glottal aperture and less glottal constriction. All are regarded to be 'breathier', hence are further classified into the so-called 'breathier voice' (see a comprehensive review in Tian & Kuang, 2019). However, these different types of 'breathier voice' differ in the size of glottal aperture and the rate of flow of air. For example, breathy voice is argued to have a greater glottal aperture and a higher rate of flow of air than slack voice has (Ladefoged & Maddieson, 1996: 57–66). Whispered voice, however, is produced with a substantial amount of aperiodic noise (Catford, 1977). Tian and Kuang (2019) argue that the non-modal phonation in Shanghai Wu Chinese would be better categorized as 'whispered voice'. In this study, I have no intention of exploring the differences among them, hence do not distinguish them strictly due to their similar distribution of energy in the fundamental and higher frequencies (Ladefoged & Maddieson, 1996: 317).

the same language (Thavisak, 2004; Chen, 2011). The 'breathiness' view also lacks empirical evidence. The study by Zhu and Xu (2009) includes the results of breathiness but lacks control of the lexical properties of the stimuli as well as details on the method of measurements. It is therefore not clear how the connection between breathiness and ATS has been established. Moreover, although breathier voices are commonly associated with lower *f0*, such an interaction is not inevitable. For example, Hirano et al. (1970) demonstrate that vowels within very high pitches tend to be breathy due to the relaxation of muscles. Kuang (2013a) reports that vowels with the tone /33/ are significantly breathier than vowels with two lower tones /11/ and /22/ in Black Miao, a Hmong-Mien language.

In brief, none of the existing studies provide comprehensive and empirically sound data for assessing the two debates. Consequently, the tonal system of Lili Wu Chinese as well as the acoustic properties of lexical tones and tone-bearing syllables, require further investigation. To this end, multiple acoustic measures are needed to gain insights into the above issues. The current study was therefore designed to elicit data from a large sample of speakers of different generations. Studying linguistic variables across age groups in apparent time is commonly considered to be the first and most straightforward approach to studying a linguistic change across decades in real time (Labov, 1994: 45–6; also see Labov et al., 2013 for a comprehensive review). Based on the literature, the following evidence is expected to be observed to support the competing views of each debate.

As for the first debate, namely the ongoing or completed change of tonal split with aspirated onsets, it is predicted that if ATS has indeed been an ongoing change within the speech community, different stages of this change should be reflected by generational data. Figure 3.1 shows the scenario assumed by studies holding the 'ongoing change' view. In Stage I, identical to most Wu varieties, lexical tones beginning with unaspirated and aspirated onsets in Lili Wu have the same *f0* contours. In Stage II, lexical tones beginning with aspirated onsets bifurcate from those beginning with unaspirated counterparts and become independent as new tonal categories, distinct from both *f0* contours of the other two types. In Stage III, the *f0* lowering trend continues and finally leads to the merging of contours beginning with aspirated onsets with those beginning with voiced onsets. If the 'ongoing change' view is true, the three stages of tonal categorization and the stepwise lowering of *f0* contours beginning with aspirated onsets are expected across different age groups (from old to middle-aged to young). On the contrary, according to the 'completed change' view, a merger of lexical tones beginning with aspirated and voiced onsets is expected for all three generations.

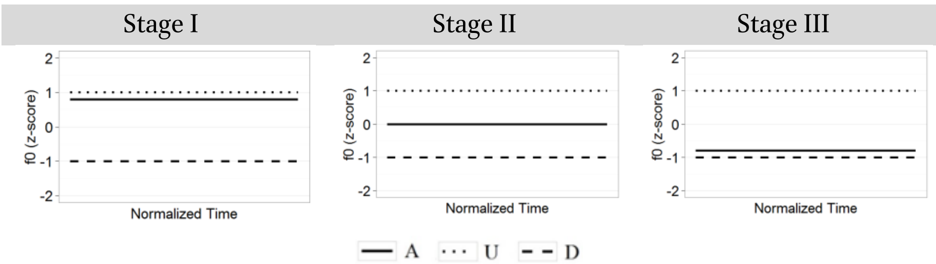

Figure 3.1 Three expected stages of ATS based on the 'ongoing change' view. A: voiceless aspirated onsets; U: voiceless unaspirated onsets; D: voiced onsets.

For the second debate, the phonetic properties of tonal split, the 'aspiration' view predicts that similar patterns of onset aspiration will lead to similar patterns of tonal split. That is to say, if ATS does not occur in the MC *Ping* tonal category, onset aspiration in the *Ping* category is expected to show a significant difference from that in the other three MC tonal categories (i.e., *Shang*, *Qu*, and *Ru*) where ATS is observed. As for the 'breathiness' view, it predicts that similar patterns of breathiness will lead to similar patterns of tonal split. That is to say, if ATS does not occur in the MC *Ping* tonal category, breathiness in the *Ping* category is expected to show a significant difference from that in the other three MC tonal categories (i.e., *Shang*, *Qu*, and *Ru*) where ATS is observed. Moreover, Hirayama (2010) argues a higher magnitude of breathiness throughout the entire vowel (or most part of the vowel). That is to say, as long as ATS happens, a consistent higher magnitude of breathiness is expected over at least the majority of the vowel interval and is not just localized at the onset of the adjacent vowel.

3.3 Acoustic experiment

3.3.1 Stimuli

The stimulus list consists of a minimal set of 36 real monosyllabic words (3 consonant onsets × 3 vowels × 4 MC tonal categories) with three laryngeal-alveolar contrasts /t tʰ d/ (voiceless unaspirated, voiceless aspirated, and voiced) combined with three vowels /ɑ ɛ i̟/ (low, middle, and high) (but /a ʌ ɪ/ for the *Ru*-category words). Each syllable is further associated with four tonal categories, covering all potential CV (CVʔ in *Ru*) combinations between consonantal onsets and lexical tones. All words are selected from the Questionnaire of character for dialect surveys (*Fangyan diaocha zibiao*, 方言调查字表) compiled by the Institute of Linguistics of the Chinese Academy of Social Sciences (1988). The characters in the questionnaire are listed according to their MC pronunciations, which guaranteed the reflex of the tonal system from modern Lili Wu Chinese to MC (see a detailed introduction of this questionnaire in Kurpaska, 2010, Chapter 7). The full stimulus list is provided in Appendix I-1. The place of articulation of the plosive onset is restricted to alveolar so as to minimize variation due to the intrinsic effect of different places of articulation on VOT (Cho & Ladefoged, 1999; Lai et al., 2009). All stimuli were confirmed to be frequent and familiar words in Lili Wu Chinese by an educated native speaker who has spent most of his life living in Lili and also took part in the experiment.

3.3.2 Participants

A total of 68 native speakers were recruited in the experiment. However, only 60 participants were selected as qualified participants as the speech production of 8 participants turned out to be problematic either due to too much disfluency or equipment failure. So, only 60 participants' data were included for further acoustic analysis. Among the selected participants (*M* = 47 years; *SD* = 17 years), there were 20 participants for each age group (old: 12 males and 8 females born between 1933 and 1956, *M* = 67 years, *SD* = 6 years; middle-aged: 11 males and 9 females born between 1961 and 1976, *M* = 48 years, *SD* = 6 years; young: 8 males and 12 females born between 1978 and 1994, *M* = 27 years; *SD* = 6 years). In addition to Lili Wu Chinese, all participants were speakers of Standard Chinese but with different levels of proficiency. Younger participants generally achieved a higher level of proficiency than older participants. However, according to their self-reports, all considered Lili Wu Chinese their first and dominant language.

3.3.3 Procedure

The recordings were conducted for all participants in a quiet room in Lili town. The utterances were recorded on an external sound card (Cakewalk UA-1G) with a Sennheiser PC 151 Headset condenser microphone. The signal was digitized at a 22,050 Hz sampling rate. Stimuli were presented twice as differently randomized lists via the Field Phon (Feifeng) program (Pan et al., 2015). The participants first heard a pre-recorded question in Lili Wu Chinese which was read by a native male speaker and then answered the question verbally with the target words on the screen. The pre-recorded question was 'What is it called in Lili Wu Chinese?'[25] In this

[25] IPA transcription: /kɛ^{44} kʌʔ joŋ^{213} lḭ^{13} lḭ tʰu^{22} u naʔ^{3} haŋ u^{213}/. Tones are marked for each prosodic unit on the basis of the tone of the initial syllable due to tone sandhi.

way, the discourse context (Lea, 1973) was controlled. This made sure that all target words were uttered as information elicited by a *wh*-question in the same controlled discourse context (Chen, 2011). In total, 4,320 tokens were collected (36 target words × 2 repetitions × 60 speakers). The participants were asked to pronounce each word at their normal speaking rate. To make sure that the task was correctly understood, all participants undertook five practice trials (with no target words included) to become familiar with the procedure before the real recording, but none knew the purpose of the experiment. All were paid the equivalent of 10 euros in local currency for their participation.

3.3.4 Measurements

Segments were identified manually with Praat (Boersma & Weenink, 2018) based on the periodicity in the acoustic waveform, supplemented by spectrographic analyses (Lehiste & Peterson, 1961; Turk et al., 2006). In order to explore the two debates on tonal split, three sets of acoustic measurements were extracted.

For the exploration of whether ATS is an ongoing change, *f0* in Hz was measured at twenty equidistant points over syllables with a long duration (i.e., *Ping*, *Shang*, and *Qu*), but ten points over syllables with a short duration (i.e., *Ru*) starting from the first regular vocal pulse to the end of the syllable using a custom-written script (Chen, 2011). Furthermore, in order to eliminate the pitch range difference due to individual variation 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). The plotted tonal contours were then averaged across speakers in each group on the basis of the mean *z*-score.

With respect to the debate on the phonetic properties of tonal split, two points are worth noting. First, since the 'aspiration' view argues that onset aspiration is the obvious connection of tonal split, the absolute duration of onset release (DOR) was measured. DOR was defined as the interval between the beginning of the release burst of the aspirated onset and the onset of voicing for the following vowel (identified as the onset of the low-frequency voicing energy on the spectrogram) (Francis et al., 2003). It has been known that speaking rate can affect the production of VOT of stop consonants, especially for long-lag VOT (Kessinger & Blumstein, 1997). For better control of the speaking rate, the raw DOR values were further divided by the duration of the tone-carrying syllable (DOS). Both the raw DOR as well as the DOR/DOS ratio served as the dependent variable for statistical analysis, respectively. Second, since the 'breathiness' view maintains that the breathiness of vowels is the more direct property for tonal split, the corrected H1*–H2* values were taken (see Hanson, 1997; Iseli et al., 2007 for formant corrections denoted with an asterisk). The acoustic cue H1*–H2*, which is the amplitude difference between the first and second harmonics, has been widely adopted as an indication of the phonatory state across languages with higher H1*–H2* values signaling breathier phonation (e.g., Gordon & Ladefoged, 2001; Blankenship, 2002; Keating et al., 2010; Kuang, 2013a). For the breathy contrast here, the main interests are whether breathiness is different in the *Ping* vs. non-*Ping* tonal categories, and whether the difference is maintained throughout the whole vowel. To this end, three points were taken (i.e., the one-third, middle, and two-thirds of the vowel). This was automatically obtained by using VoiceSauce (Shue et al., 2011) with a 25 ms window size.

3.3.5 Statistical analyses

For statistical modeling of each dependent variable (i.e., *f0*, VOT, and H1*–H2*), multilevel regression models were used (see Section 1.3.2 for more details). In general, all base models included the Level 1 structure and the by-speaker random effect on the Level 1 structure without any other independent variables in question. The effect of adding each independent variable (depending on different research purposes, see below for more details) was evaluated by using model comparisons. If a significant effect of interaction between independent variables is observed, the overall data would be further divided into several subsets according to different research questions. Within each subset, a base model was first established containing only the Level 1 structure and the by-speaker random structure on all time terms. The independent variables in question were then added. In addition, there were three control variables, namely Vowel (high vs. middle vs. low), Gender (male vs. female), and Repetition (first vs. second). All of them were known to exert an effect on the three dependent variables (e.g., Jacewicz & Fox, 2015 for vowel on *f0*; Lam & Watson, 2010 for repetition on *f0*; Simpson, 2012 for gender on *f0*; Fischer-Jørgensen, 1972 for vowel on VOT; Swartz, 1992 for gender on VOT; Esposito, 2010 for gender on H1–H2), however, none of these variables were related to the research questions, hence were not discussed. In order to improve model fit, control variables were further tested in a stepwise fashion via model comparisons to determine which of them should be considered in final models. At last, Speaker by Consonant and Item were also tested as random structures.

Given that each of the three measurements was tested for achieving separate goals, it is worth addressing that except for the general analysis, particular analyses were also applied to different dependent variables (i.e., *f0*, VOT, and H1*–H2*).

Growth curve analysis (GCA) was employed for examining the changes in *f0* contours. The basic idea of GCA is to build orthogonal polynomials including multiple polynomial terms for capturing the change of real data. For example, a second-order polynomial model: $y = \alpha + \beta \cdot Time_1 + \gamma \cdot Time_2 + \Delta$ would have three time terms: intercept (α), linear (β), and quadratic (γ). These terms index the overall mean of the curve, the direction of curve change, such as rising vs. falling, and the steepness of curve rising or falling, respectively (Li & Chen, 2016). If two curves differ from each other, it is expected to show a statistical difference in at least one aspect of the three terms.

In order to choose the best polynomial order and avoid overfitting the data, before following the general analysis, the best shape for capturing the changes of overall *f0* contours was first determined. Both practical and statistical reasons were taken into consideration (Mirman, 2014: 46–7). According to the tonal system of Lili Wu Chinese shown in Section 2.2, the most complex *f0* contour has a convex contour shape. Therefore, the model having a simple linear shape was compared with the one having a curved shape. Following the method suggested in Winter and Wieling (2016), two base models with different time terms (i.e., Level 1 structure, linear shape: ot1 vs. curved shape: ot1 + ot2) and individual participants (i.e., Speaker) varying in the random intercept were built for model comparisons within each MC tonal category. Independent variables were then entered. Given the first debate on the differences of ATS across generations, I was interested in the main effect of three independent variables, namely Consonant (aspirated vs. unaspirated vs. voiced), Generation (old vs. middle-aged vs. young), and their interaction Consonant * Generation. If there was an interaction between the two variables, data were further divided to explore the *f0* as a function of different onsets. Separate subsets according to each of the generations were built; Consonant and additional control variables were then introduced stepwise via model comparisons within each subset.

With regard to the analysis of DOR-related (i.e., raw DOR and ROR/DOS ratio) and H1*–H2* data, linear mixed-effects models (LMMs) were built. For the analysis of the DOR-related data of aspirated onsets, as fixed effects, Category, Generation, and their interaction were entered into the models in a stepwise fashion. The intercept for Speaker was also taken as a random structure. If the interaction between Category and Generation showed a significant effect, separate LMMs were further built based on each of the generations. Each base model was first built only with the random intercept of Speaker. Category and additional control factors (i.e., Vowel, Repetition, and Gender) were then introduced in a stepwise manner for model comparisons. The intercept for Item as well as the slope for Speaker by Consonant were also tested as random effects via model comparisons.

The analysis of H1*–H2* was similar to the DOR-related data. Consonant, Generation, Position (one-third vs. middle vs. two-thirds), Category, and their interactions were first entered as fixed effects into the models in a stepwise fashion. All models kept the random intercept of Speaker consistent. If there was an interaction of the four factors, data were further divided to explore the H1*–H2* of vowels as a function of different onsets. To this end, separate LMMs were carried out according to each of MC tonal categories, generations, and time positions. Each base model contained the random intercept of Speaker only. Consonant and additional control factors were then introduced stepwise via model comparisons. The operation of other random effects was developed in the same manner applied to the analysis of the DOR-related data.

In comparisons of all model data, an improvement of model fit was obtained through likelihood ratio tests indicted by Pr(>Chisq) (χ^2) for each model with the effect in question against the model without the effect in question. Under any circumstances where the model failed to converge, the newly added structure was then dropped. All data were analyzed with the *lme4* package (Bates et al., 2014) and plotted with the *ggplot2* package (Wickham, 2009) in *R* (R Development Core Team, 2019).

3.3.6 Results

3.3.6.1 *f0* contour

In the following, the four MC tonal categories (i.e., *Ping*, *Shang*, *Qu*, and *Ru*) were labeled as I to IV, respectively. 'A', 'U', and 'D' represented measured values with voiceless aspirated, voiceless unaspirated, and voiced onsets, respectively.

The results showed that the second-order polynomial model significantly improved the model fit for all four MC tonal categories [I: χ^2 = 187.35, p < .001; II: χ^2 = 75.68, p < .001; III: χ^2 = 1638, p < .001; IV: χ^2 = 20.9, p

< .001]. The second-order polynomial was therefore applied to all data for further analyses.

Table 3.3 shows the results of model comparisons for the effect of Consonant, Generation, and their interaction on *f*0. Except for Consonant in the quadratic term in IV, both Consonant and its interaction with Generation had significant effects on all three time terms across all MC tonal categories. However, the Generation factor did not show a significant main effect except for the quadratic term in I, the intercept term in III, and the linear term in IV. Given the across-the-board significance of the interaction between Consonant and Generation across categories, the data were further decomposed according to each of the three generations.

Table 3.3 Results (χ^2) of model comparisons for the effect of Consonant, Generation, and Consonant * Generation on *f*0. Parameter-specific *p*-values (superscript) were indicated by Pr(>Chisq). n.s.: not significant.

Factor	Time term	I	II	III	IV
Consonant	Intercept	$6578.4^{< .001}$	$14814.28^{< .001}$	$5017.56^{< .001}$	$11663.88^{< .001}$
	Linear	$2163.19^{< .001}$	$2252.9^{< .001}$	$2680.76^{< .001}$	$161.99^{< .001}$
	Quadratic	$190.63^{< .001}$	$50.44^{< .001}$	$11.86^{< .01}$	$2.37^{n.s.}$
Generation	Intercept	$3.46^{n.s.}$	$1.01^{n.s.}$	$16.03^{< .001}$	$1.14^{n.s.}$
	Linear	$.55^{n.s.}$	$3.81^{n.s.}$	$5.81^{n.s.}$	$7.61^{< .05}$
	Quadratic	$17.81^{< .001}$	$1.45^{n.s.}$	$5.34^{n.s.}$	$1.57^{n.s.}$
Consonant * Generation	Intercept	$6871.14^{< .001}$	$15115.7^{< .001}$	$5428.58^{< .001}$	$11911.77^{< .001}$
	Linear	$2200.92^{< .001}$	$2374.6^{< .001}$	$2761.41^{< .001}$	$187.16^{< .001}$
	Quadratic	$241.64^{< .001}$	$63.2^{< .001}$	$26.02^{< .001}$	$10.97^{n.s.}$

Figure 3.2 displays the *f*0 contours of each MC tonal category (I–IV) within the factor Generation. Table 3.4 shows corresponding results tested by GCA, where A was set as the reference group. The final models to calculate the results of Table 3.4 are attached in Appendix I-2 for interested readers.

As shown in Figure 3.2-I, across generations, both aspirated (I-A) and unaspirated (I-U) onsets introduce comparable *f*0 contours within the high register, but voiced onsets (I-D) introduce low-rising contours. Within the factor Generation, compared to I-A, no time terms of I-U showed any significant effect. Irrespective of generation, both intercept and linear terms of I-D consistently showed significant differences from those of I-A, as supported by the significant differences of I-D shown in Table 3.4. Moreover, compared to I-A, the *f0* contour with I-D presents a slightly concave trajectory produced by the middle-aged- and young-generation speakers, as indicated by the significant quadratic term of I-D in both groups.

As shown in Figure 3.2-II, aspirated (II-A) and voiced (II-D) onsets lead to more comparable *f0* contours, which are realized within the low register, while unaspirated onsets (II-U) introduce falling contours within the high register. The results from Table 3.4 indicated that there were significant differences between *f0* contours with II-A and II-U in all three time terms. However, no significant difference was observed in any time term between II-A and II-D. Such a pattern held across generations.

Figure 3.2-III suggests that contours beginning with aspirated (III-A), voiced (III-D) and unaspirated (III-U) onsets are consistently realized with concave trajectories. There is no prominent difference between *f0* contours beginning with III-A and III-D. However, both are produced with a lower *f0* than III-U is. This difference was reflected by the significant results in both intercept and linear terms. Again, this pattern held across generations.

Finally, as shown in Figure 3.2-IV, similar low-level contours beginning with aspirated (IV-A) and voiced (IV-D) onsets are again observed within the factor Generation. This was reflected by the lack of significant results in all time terms presented in Table 3.4. Both, however, are significantly different from the *f0* contours beginning with unaspirated onsets (IV-U), which basically show high-level trajectories. Worth noticing is that the contour for IV-U produced by young-generation speakers, as compared to that for I-A, showed a less concave trajectory as indicated by the significant quadratic term.

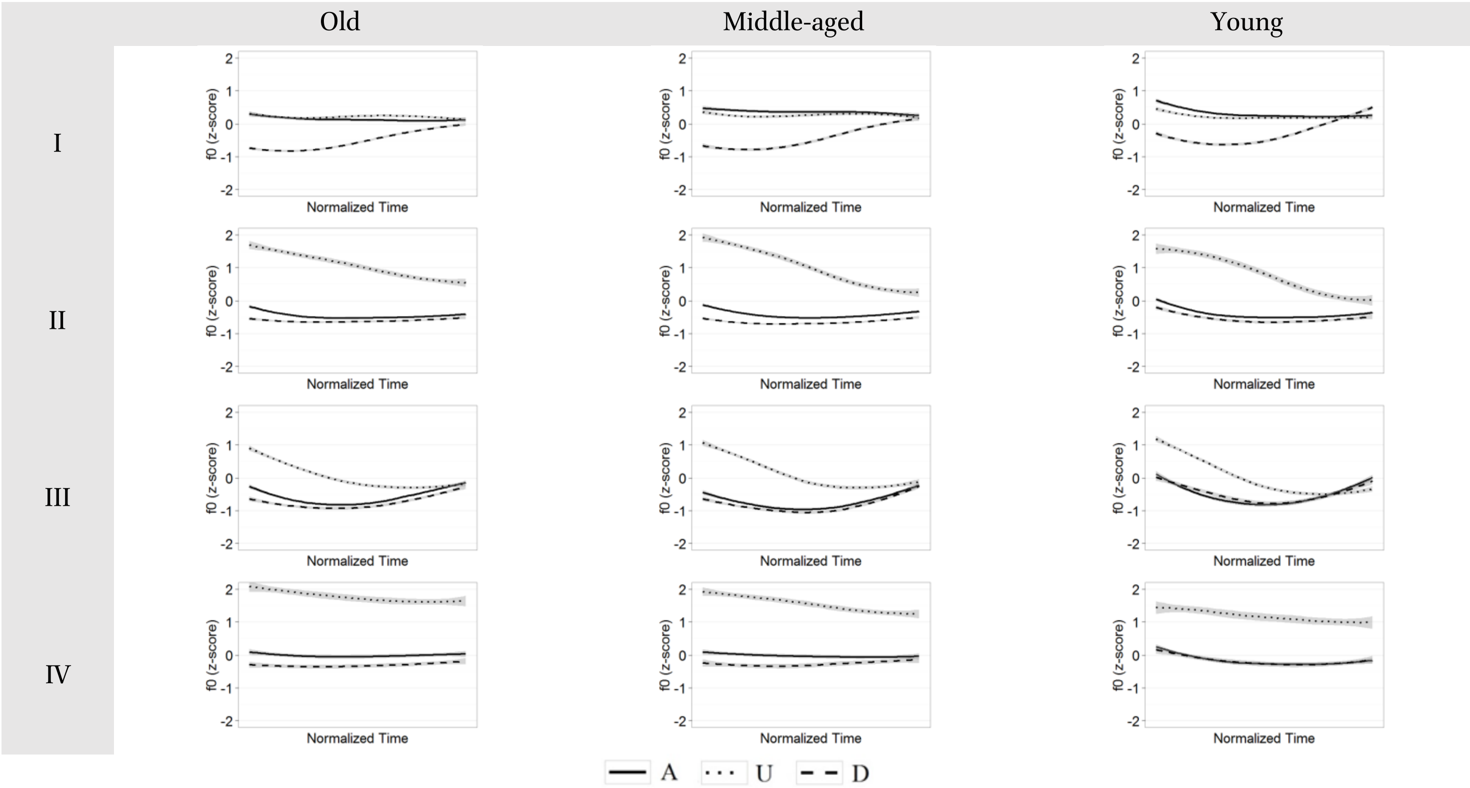

Figure 3.2 The *f*0 realization of three generations. Mean normalized *f*0 (gray areas indicate ± *SE*) averaged across speakers and repetitions. The x-axis is normalized time and the y-axis is normalized *f*0.

Table 3.4 Results (β) of the effect of Consonant on *f0*. Reference: voiceless aspirated onset (A). Parameter-specific *p*-values (superscript) were estimated using the normal approximation. n.s.: not significant.

Generation	I-A	Intercept	Linear	Quadratic	II-A	Intercept	Linear	Quadratic
Old	I-U	$.07^{\text{n.s.}}$	$.14^{\text{n.s.}}$	$-0.14^{\text{n.s.}}$	II-U	$1.51^{<.001}$	$-1.45^{<.001}$	$-0.25^{<.05}$
	I-D	$-0.64^{<.001}$	$1.45^{<.001}$	$.16^{\text{n.s.}}$	II-D	$-0.15^{\text{n.s.}}$	$.28^{\text{n.s.}}$	$-0.19^{\text{n.s.}}$
Middle-aged	I-U	$-0.1^{\text{n.s.}}$	$.2^{\text{n.s.}}$	$.01^{\text{n.s.}}$	II-U	$1.43^{<.001}$	$-2.38^{<.001}$	$-0.3^{<.01}$
	I-D	$-0.78^{<.001}$	$1.64^{<.001}$	$.42^{<.001}$	II-D	$-0.23^{\text{n.s.}}$	$.28^{\text{n.s.}}$	$-0.17^{\text{n.s.}}$
Young	I-U	$-0.09^{\text{n.s.}}$	$.24^{\text{n.s.}}$	$-0.13^{\text{n.s.}}$	II-U	$1.15^{<.001}$	$-1.98^{<.001}$	$-0.48^{<.05}$
	I-D	$-0.54^{<.001}$	$1.59^{<.001}$	$.54^{<.001}$	II-D	$-0.12^{\text{n.s.}}$	$.1^{\text{n.s.}}$	$-0.07^{\text{n.s.}}$
Generation	III-A	Intercept	Linear	Quadratic	IV-A	Intercept	Linear	Quadratic
Old	III-U	$.6^{<.001}$	$-1.84^{<.001}$	$-0.09^{\text{n.s.}}$	IV-U	$1.78^{<.001}$	$-0.61^{<.05}$	$.03^{\text{n.s.}}$
	III-D	$-0.16^{\text{n.s.}}$	$.19^{\text{n.s.}}$	$-0.22^{\text{n.s.}}$	IV-D	$-0.29^{\text{n.s.}}$	$.18^{\text{n.s.}}$	$-0.02^{\text{n.s.}}$
Middle-aged	III-U	$.84^{<.001}$	$-2.12^{<.001}$	$-0.01^{\text{n.s.}}$	IV-U	$1.56^{<.001}$	$-0.84^{<.01}$	$-0.01^{\text{n.s.}}$
	III-D	$-0.1^{\text{n.s.}}$	$.15^{\text{n.s.}}$	$-0.03^{\text{n.s.}}$	IV-D	$-0.24^{\text{n.s.}}$	$.39^{\text{n.s.}}$	$.06^{\text{n.s.}}$
Young	III-U	$.48^{<.001}$	$-2.08^{<.001}$	$-0.26^{\text{n.s.}}$	IV-U	$1.47^{<.001}$	$-0.41^{\text{n.s.}}$	$-0.32^{<.01}$
	III-D	$.01^{\text{n.s.}}$	$-0.17^{\text{n.s.}}$	$-0.18^{\text{n.s.}}$	IV-D	$-0.002^{\text{n.s.}}$	$.16^{\text{n.s.}}$	$-0.09^{\text{n.s.}}$

These findings confirm descriptions in the existing literature that there is no tonal split in the MC *Ping* (I) tonal category.[26] This implies that ATS is not an across-the-board phenomenon in Lili Wu Chinese, but rather that its appearance is conditioned by certain tonal contexts (i.e., MC tonal categories). More importantly, in those tonal contexts where ATS occurred (i.e., MC *Shang*, *Qu*, and *Ru*), the *f0* contours beginning with voiceless aspirated and voiced onsets completely merged. Both were significantly lower than the *f0* contours beginning with unaspirated onsets. Such a pattern of ATS was stable across all three generations.

3.3.6.2 Raw DOR and DOR/DOS ratio

For the raw DOR, there was a significant main effect of Category [$\chi^2 = 145.98$, $p < .001$]. However, both the main effect of Generation [$\chi^2 = 6.19$, $p > .05$] and its interaction with Category [$\chi^2 = 2.74$, $p > .05$] did not show a significant effect. The insignificant interaction impeded us from dividing the data. For the DOR/DOS ratio, the results showed both a significant main effect of Category [$\chi^2 = 477.76$, $p < .001$] and a significant interaction of Category and Generation [$\chi^2 = 50.14$, $p < .001$], but there was no significant main effect of Generation [$\chi^2 = 3.88$, $p > .05$]. A subset of data was then generated for each generation. Separate models were run for each subset to examine the difference of the DOR/DOS ratio between the MC *Ping* and the other three MC tonal categories. Figure 3.3 depicts the DOR/DOS ratio of MC tonal categories (I–IV) of each generation. Although there was no statistical significance for the factor Generation, a trend of difference is observable as plotted in Figure 3.3. Table 3.5 shows corresponding results tested by LMMs, where I was set as the reference group. The final models to calculate the results of Table 3.5 are attached in Appendix I-3a for interested readers. In addition, Appendix I-3b provides

[26] Except for Chao (1928), which reports the absence of ATS in the MC *Shang* tonal category. A further question to be discussed is whether this is due to change over the MC *Shang* category not having started yet at that time (see Section 2.2).

the results via *post hoc* pairwise comparisons (i.e., estimated marginal means). The results of pairwise comparisons are consistent with the results of stepwise multilevel regression conducted in this section.

As shown in Figure 3.3, the mean of the DOR/DOS ratio of IV remains highest across all generations [old: 0.45; middle-aged: 0.44; young: 0.38]. Correspondingly, the mean of the raw DOR of IV is the shortest [old: 79.98 ms; middle-aged: 81.93 ms; young: 93.76 ms]. This visual inspection was further supported by a consistently significant difference between I and IV across generations [old: β = .12, p < .001; middle-aged: β = .11, p < .001; young: β = .06, p < .001]. This result was attributed to the short tone-carrying syllable of IV, which reduced the duration of vowels and increased the ratio accordingly. The expected difference between I and the other three categories, however, was not observed. As shown in Table 3.5, within each generation, I did not show any significant difference from the other two counterparts, namely II and III. This pattern held across generations.

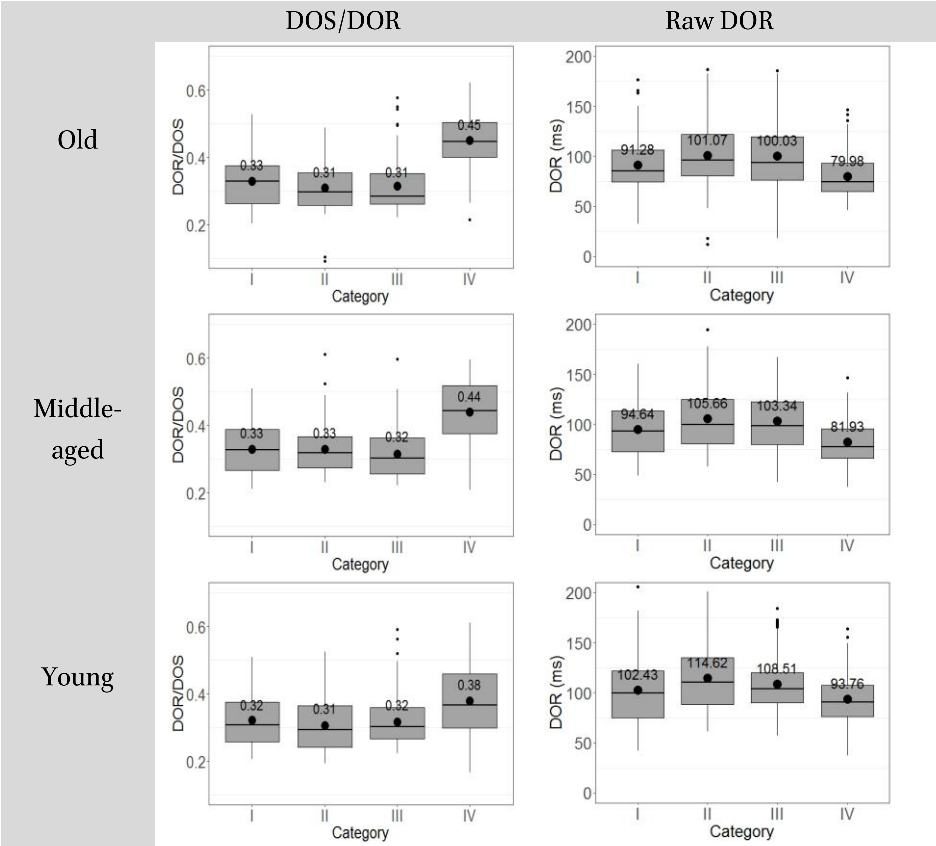

Figure 3.3 Boxplots of the DOR/DOS ratio in target syllables of each MC tonal category for three generations. The bottom and top lines of a box represent the lower (or first) and upper (or third) quartiles of the data. The lines extending vertically from the box are known as the 'whiskers', which are used to indicate variability outside the lower and upper quartiles. The two ends of the 'whiskers' represent the lower and upper extremes, respectively. The solid point in the box represents the mean and the line within the box the median. Individual dots outside the box are outliers.

Table 3.5 Results of models fit to the DOR/DOS ratio of each generation. Reference: I. n.s.: not significant.

	Old			Middle-aged			Young		
	Estimate (β)	t	p	Estimate (β)	t	p	Estimate (β)	t	p
Intercept	.32	23.55	< .001	.33	23.8	< .001	.31	23.06	< .001
II	−0.02	−1.44	n.s.	−0.01	−0.07	n.s.	−0.02	−1.38	n.s.
III	−0.02	−1.21	n.s.	−0.01	−1.08	n.s.	−0.01	−0.53	n.s.
VI	.12	8.13	< .001	.11	6.42	< .001	.06	3.3	< .001

Table 3.6 Results (χ^2) of model comparisons for the effect of Consonant, Generation, Category, Position, and their interactions on H1*–H2*. Parameter-specific p-values (superscript) were indicated by Pr(>Chisq). n.s.: not significant.

One-way	χ^2	Two-way	χ^2	Three-way and four-way	χ^2
Consonant	$250.88^{<.001}$	Consonant * Generation	$56.65^{<.001}$	Consonant * Generation * Category	$88.23^{<.001}$
Generation	$.05^{n.s.}$	Generation * Category	$17.4^{<.01}$	Generation * Category * Position	$184.16^{<.001}$
Category	$47.43^{<.001}$	Category * Position	$153.91^{<.001}$	Category * Position * Consonant	$255.04^{<.001}$
Position	$191.95^{<.001}$	Position * Consonant	$104.72^{<.001}$	Position * Consonant * Generation	$177.1^{<.001}$
		Consonant * Category	$59.01^{<.001}$	Consonant * Generation * Category * Position	$314.73^{<.001}$
		Generation * Position	$24.45^{<.001}$		

3.3.6.3 H1*–H2*

As summarized in Table 3.6, the results indicated that, except for Generation, all other factors (i.e., Consonant, Category, and Position) showed a significant effect on H1*–H2*. Moreover, four factors significantly interacted in all orders (i.e., two-way, three-way, and four-way). There were multiple scenarios to further quantify the interactions. Given the purpose of comparing H1*–H2* values of vowels beginning with different onsets, the dataset was divided into twelve subsets according to each of the generations, MC tonal categories and time positions where H1*–H2* values were measured.

To help visualize the interactions, Figure 3.4 plots the mean H1*–H2* measured over the three positions for all twelve subsets. A series of LMMs was run over each subset. Table 3.7 summarizes the results with A as the reference category. The final models to calculate the results are attached in Appendix I-4a for interested readers. In addition, Appendix I-4b provides the results of pairwise comparisons. The results are consistent with the results of stepwise multilevel regression conducted in this section.

Figure 3.4 shows that in I, II, and III of both old and middle-aged groups, there is little H1*–H2* difference between vowels after A and D. Both, however, show higher H1*–H2* values than vowels after U. This observation was supported by the results of Table 3.7. On the one hand, H1*–H2* of vowels after A (old and middle-aged: I-A, II-A, and III-A) was consistently different from that after U (old and middle-aged: I-U, II-U, and III-U). On the other hand, it did not differ significantly from that after D (old and middle-aged: I-D, II-D, and III-D). However, the young-generation speakers showed a very different pattern. As shown in Figure 3.4-I/II/III, it is quite clear that the H1*–H2* of vowels after A is much higher than that after U and D. Very different from the pattern of old- and middle-aged-generation speakers, H1*–H2* of vowels after D of young-generation speakers tends to be lower. It leads to an approximation of

H1*–H2* of vowels after U and D. As seen in Table 3.7, significant differences existed between A (young: I-A, II-A, and III-A) and its two counterparts (young: I-U, II-U, III-U, I-D, II-D, and III-D).

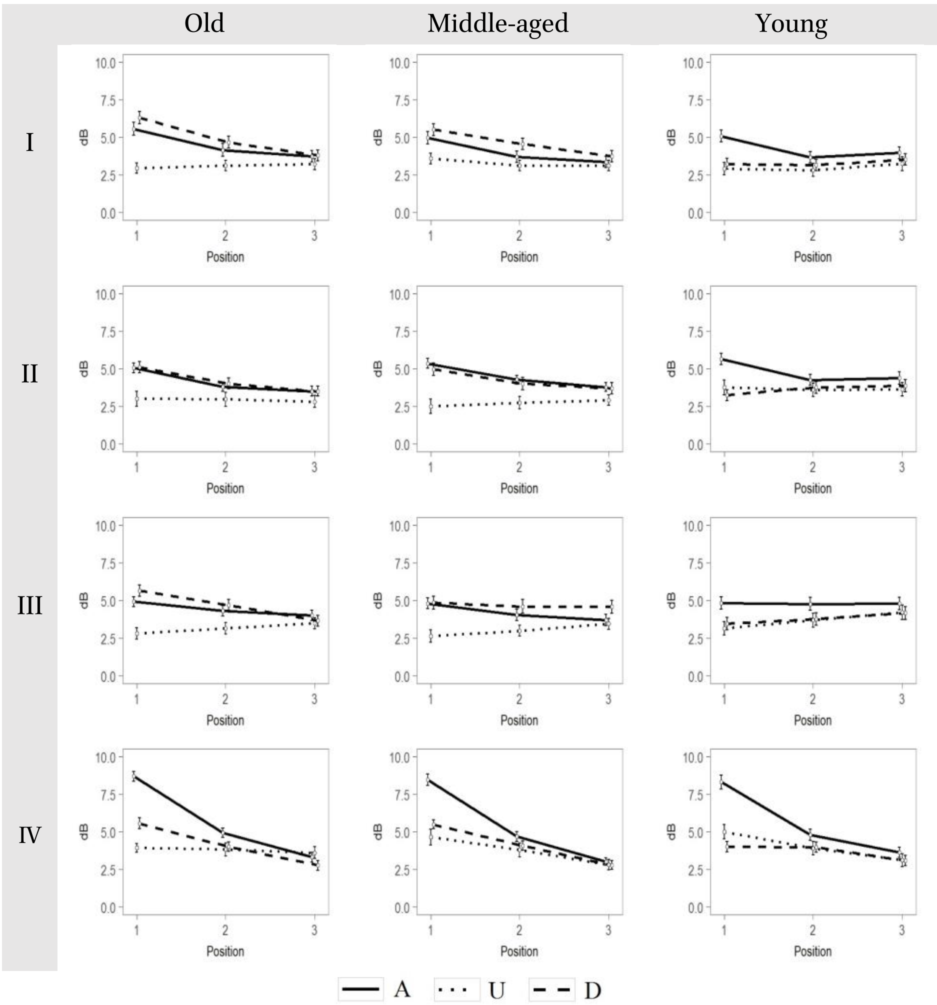

Figure 3.4 Mean H1*–H2* of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each MC tonal category. Error bars represent the standard error of the mean.

Table 3.7 Results (β) of model comparisons for the effect of Consonant on H1*–H2*. Reference: voiceless aspirated onset (A). P1 to P3 represent the three time positions where the H1*–H2* measurements were made. Parameter-specific *p*-values were superscripted. –: Consonant factor did not improve the model fit. n.s.: not significant.

Generation	I-A	P1	P2	P3	II-A	P1	P2	P3
Old	I-U	$-2.63^{< .001}$	$-1.05^{< .05}$	–	II-U	$-2.04^{< .05}$	–	–
	I-D	$.74^{\text{n.s.}}$	$.51^{\text{n.s.}}$		II-D	$.04^{\text{n.s.}}$		
Middle-aged	I-U	$-1.37^{< .01}$	$-0.59^{\text{n.s.}}$	–	II-U	$-2.85^{< .05}$	$-1.52^{\text{n.s}}$	–
	I-D	$.55^{\text{n.s.}}$	$.87^{\text{n.s.}}$		II-D	$-0.39^{\text{n.s.}}$	$-0.25^{\text{n.s.}}$	
Young	I-U	$-2.17^{< .001}$	–	–	II-U	$-1.89^{< .05}$	–	–
	I-D	$-1.85^{< .001}$			II-D	$-2.41^{< .01}$		
Generation	III-A	P1	P2	P3	IV-A	P1	P2	P3
Old	III-U	$-2.12^{< .001}$	$-1.16^{\text{n.s.}}$	–	IV-U	$-4.75^{< .001}$	$-1.12^{\text{n.s.}}$	–
	III-D	$.72^{\text{n.s.}}$	$.38^{\text{n.s.}}$		IV-D	$-3.12^{< .001}$	$-0.94^{\text{n.s.}}$	
Middle-aged	III-U	$-2.16^{< .05}$	$-1.06^{\text{n.s.}}$	–	IV-U	$-3.81^{< .001}$	–	–
	III-D	$.05^{\text{n.s.}}$	$.55^{\text{n.s.}}$		IV-D	$-2.97^{< .001}$		
Young	III-U	$-1.67^{< .01}$	–	–	IV-U	$-3.3^{< .001}$	–	–
	III-D	$-1.36^{< .01}$			IV-D	$-4.3^{< .001}$		

The situation of IV was different from all other tonal categories (i.e., I, II, and III). As shown in Figure 3.4-IV, across generations, H1*–H2* of vowels after A is always higher than that for its two counterparts (i.e., U and D). Supported by the results in Table 3.7, a significant effect was consistently found between IV-A and IV-U as well as IV-A and IV-D across all generations. Moreover, as observed from Figure 3.4-IV, the difference of H1*–H2* between vowels after U and D is also obvious within each generation. H1*–H2* of vowels after U is lower than that after D in speakers of old and middle-aged generations, but higher than that after D in the young-generation speakers.

When focusing on the middle (P2) and two-thirds positions (P3), as shown in Figure 3.4, all differences presented at P1 tend to be diminished across generations. This pattern was also confirmed by the results of Table 3.7. In the majority of cases, Consonant did not significantly improve the model fit (indicated by '–'), which suggested that there was no significant difference of $H1^*–H2^*$ of vowels after the three onsets. In six cases of P2 (old: I, III, and IV; middle-aged: I, II, and III), Consonant did help to improve the model fit. However, five of them did not show significant results between A and its two counterparts. The only significant result was found in I, where $H1^*–H2^*$ after A was higher than that after U.

In summary, there are three findings. First, regardless of whether ATS happened or not, a consistently higher $H1^*–H2^*$ over one-third of vowels after aspirated onsets was observed. This pattern held across generations. Second, the consistently higher $H1^*–H2^*$ of vowels only presented at the beginning of adjacent vowels (i.e., one-third position), but vanished after the midpoint regardless of whether ATS happened or not. This pattern again held across generations. Third, an interesting finding is that across all MC tonal categories, the two older groups (old and middle-aged) showed more comparable patterns of $H1^*–H2^*$ of vowels after voiced and aspirated onsets, whereas the young group showed a minimized $H1^*–H2^*$ difference, especially after voiced and unaspirated onsets. It suggests that the phonatory state of vowels after voiced onsets is experiencing an ongoing change across generations.

3.3.7 Summary I

There are four general findings in this section. First, ATS is not an across-the-board phenomenon in Lili Wu Chinese, but, rather, its appearance is conditioned by certain tonal contexts (i.e., MC tonal categories). More importantly, in those tonal contexts where ATS occurred (i.e., II *Shang*, III *Qu*, and IV *Ru*), the $f0$ contours beginning with voiceless aspirated and voiced onsets completely merged. Both were significantly lower than the $f0$ contours beginning with unaspirated onsets. Such a pattern of ATS was stable across all three generations. Second, in terms of aspiration, I (i.e., *Ping*) failed to show differences. Instead, it was IV, which was consistently distinguished from the other three categories. This pattern held across generations. Third, in terms of breathiness, H1*–H2* of vowels after aspirated onsets did not show differences across tonal categories. Regardless of whether ATS happened or not, a higher H1*–H2* in vowels after voiceless aspirated onsets was consistently observed, which, however, was only present at the beginning of the following vowel (i.e., at the one-third position). This pattern again held across generations. Last but not least, a serendipitous finding is that the reduced breathiness of vowels following voiced onsets seemed to be an ongoing sound change.

In order to have more accurate measurements on the phonatory state of vowels after different onsets, an articulatory experiment was conducted.

3.4 Articulatory experiment

3.4.1 Design and procedure

Given that in the acoustic experiment, the phonatory state of vowels between the old- and middle-aged-generation speakers was not significantly different, it motivated me to focus mainly on the difference between the old- and young-generation speakers. Therefore, participants from two generations (old vs. young) were recruited in the articulatory experiment. A total of 26 native participants of Lili Wu Chinese were recorded. All speakers were born and spent most of their lifetime in Lili. According to the clarity of articulation and the lower noise of the EGG signal, 20 of them were selected for further analysis with 10 participants for each generation (old: 5 females and 5 males speakers born between 1939 and 1957, M = 67 years, SD = 6 years; young: 5 females and 5 males born between 1981 and 1994, M = 29 years, SD = 4 years). The stimulus list was identical to the one used in Section 3.2.1. The first repetition of one old male participant (S10) was discarded due to high levels of noise during the recording process. All were paid the equivalent of 10 euros in local currency for their participation.

Acoustic recordings and the simultaneous EGG signal 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 the laptop computer. The acoustic signal (Channel 1) and the simultaneous EGG signal (Channel 2) were recorded to separate channels. Two gold-plated electrodes of the laryngograph were held to either side of the participant's thyroid cartilage and held stable using a velcro neckband around the participant's neck. Each electrode consists of an inner disk surrounded by an outer guard-ring. A test of the signal was conducted until it was confirmed that the location of the electrodes was adequate. In addition, throughout the recording session, the electrodes were relocated when the signal did not present to be reliable on the system. The electrodes were cleaned of sweat and dust by alcohol wipes before each recording. The recording procedure was identical to Section 3.3.3.

Noting that since the articulatory experiment and the acoustic experiment were conducted separately, the participants of the two experiments were not consistent. Fifteen participants in the articulatory experiment (old: 4 females and 3 males; young: 4 females and 4 males) also participated in the acoustic study. One may have an interest in the acoustic data collected from the participants in this experiment. Generally speaking, the three acoustic measurements (i.e., *f*0, VOT, and H1*–H2*) pattern with the results presented in Section 2.1, Section 2.3, and Section 3.3. Figures are attached in Appendix I-5 for interested readers. In this section, I focused on the articulatory data only.

3.4.2 Measurements and analyses

For the articulatory measurement, contact quotient (CQ) was taken for the current study using EGGWorks (Tehrani, 2009). CQ is defined as the proportion of vocal fold contact area during every single vibratory cycle (Rothenberg & Mahshie, 1988). The degree of CQ is reflected in the degree of electrical conductance collected by the EGG device. Generally speaking, higher conductance corresponds to greater vocal fold contact, while lower conductance corresponds to lesser contact. There are multiple methods to identify CQ in the consecutive EGG signal. In this study, CQ was measured using the Hybrid method (Davies et al., 1986; Howard, 1995; Henrich et al., 2004) based on not only the EGG signal but also on the dEGG signal (i.e., the first derivative of the EGG signal). The dEGG signal is argued to yield more reliable indicators of glottal closing instants than the raw EGG signal (Childers et al., 1986; Henrich et al., 2004). As illustrated in Figure 3.5, the Hybrid method calculates closed quotient with the positive peak (the maximum value) of the dEGG signal (bottom) identifying the onset of the closed phrase and a fixed threshold, in this case, 3/7 of the difference between the minimum and maximum values over a glottal period marking its offset (top). Because of the reliable detection of the glottal closing instants, the Hybrid method to calculate CQ has been argued to give more accurate results (see more discussion on different methods for measuring CQ in Henrich et al., 2004; Herbst, 2004; Herbst & Ternström, 2006; Yokonishi et al., 2016). It has been widely adopted in studies of tonal languages (e.g., Keating et al., 2010 for Mazatec; Esposito, 2012 for White Hmong; Kuang, 2013a for Southern Yi and Black Miao).

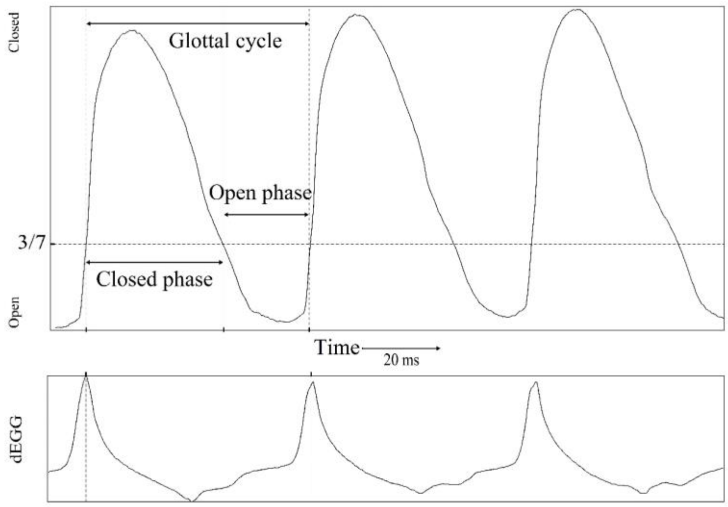

Figure 3.5 EGG measurements (top) and the dEGG signal (bottom) exemplified by 20 ms from the beginning of /i̟/ in /ti̟^{1}/ 'low' produced by an old male speaker. The onset of each glottal cycle was determined by the steep rise (peak) of the corresponding dEGG signal. The threshold was set as 3/7 of the difference between the minimum and maximum values over a glottal period, separating each cycle into closed and open phases.

As for the analysis of the CQ values, consistent with the measurements of H1*–H2*, CQ values were measured over the one-third, middle, and two-thirds points of the vowel. Following the general method illustrated in Section 3.2.5, LMMs were built. All settings were identical to the analyses of H1*–H2*.

3.4.3 Results

As shown in Table 3.8, the results indicated that Generation, Category, Generation * Category, and Generation * Position did not show significant effects on CQ. However, the four factors significantly interacted in various orders. Following the manipulation of H1*–H2*, the dataset was divided into eight subsets according to each of the generations, MC tonal categories and time points where CQ values were measured.

To help visualize the four-way interaction, Figure 3.6 is plotted with the mean of CQ measured over the three positions for all eight subsets. A series of LMMs was run over each subset. Table 3.9 summarizes the results with A as the reference category. The final models to calculate the results are attached in Appendix I-6a for interested readers. In addition, Appendix I-6b provides the results of pairwise comparisons. The results are consistent with the results of stepwise multilevel regression conducted in this section.

Table 3.9 Results (β) of the effect of Consonant on CQ. Reference: voiceless aspirated onset (A). P1 to P3 represent the three time positions where the H1*–H2* measurements were made. Parameter-specific p-values were superscripted. –: Consonant factor did not improve the model fit. n.s.: not significant.

Generation	I-A	P1	P2	P3	II-A	P1	P2	P3
Old	I-U	$.05^{<.001}$	–	–	II-U	$.07^{<.001}$	–	–
	I-D	$-0.01^{n.s.}$			II-D	$-0.01^{n.s.}$		
Young	I-U	$.04^{<.001}$	–	–	II-U	$.06^{<.01}$	–	–
	I-D	$.04^{<.05}$			II-D	$.04^{<.001}$		
Generation	**III-A**	**P1**	**P2**	**P3**	**IV-A**	**P1**	**P2**	**P3**
Old	III-U	$.05^{<.001}$	–	–	IV-U	$.09^{<.01}$	–	–
	III-D	$-0.01^{n.s.}$			IV-D	$-0.01^{n.s.}$		
Young	III-U	$.05^{<.01}$	–	–	IV-U	$.08^{<.01}$	$.05^{<.001}$	–
	III-D	$-0.05^{<.001}$			IV-D	$.07^{<.001}$	$.03^{<.05}$	

Table 3.8 Results (χ^2) of model comparisons for the effect of Consonant, Generation, Category, Position, and their interactions on CQ. Parameter-specific *p*-values (superscript) were indicated by Pr(>Chisq). n.s.: not significant.

One-way	χ^2	Two-way	χ^2	Three-way and four-way interactions	χ^2
Consonant	$87.97^{<.001}$	Consonant * Generation	$42.84^{<.001}$	Consonant * Generation * Category	$29.6^{<.05}$
Generation	$.01^{n.s.}$	Generation * Category	$.72^{n.s.}$	Generation * Category * Position	$25.62^{<.001}$
Category	$4.77^{n.s.}$	Category * Position	$13.73^{<.01}$	Category * Position * Consonant	$39.35^{<.001}$
Position	$909.5^{<.001}$	Position * Consonant	$93.64^{<.001}$	Position * Consonant * Generation	$25.18^{<.01}$
		Consonant * Category	$23.03^{<.001}$	Consonant * Generation * Category * Position	$204.05^{<.001}$
		Generation * Position	$1.16^{n.s.}$		

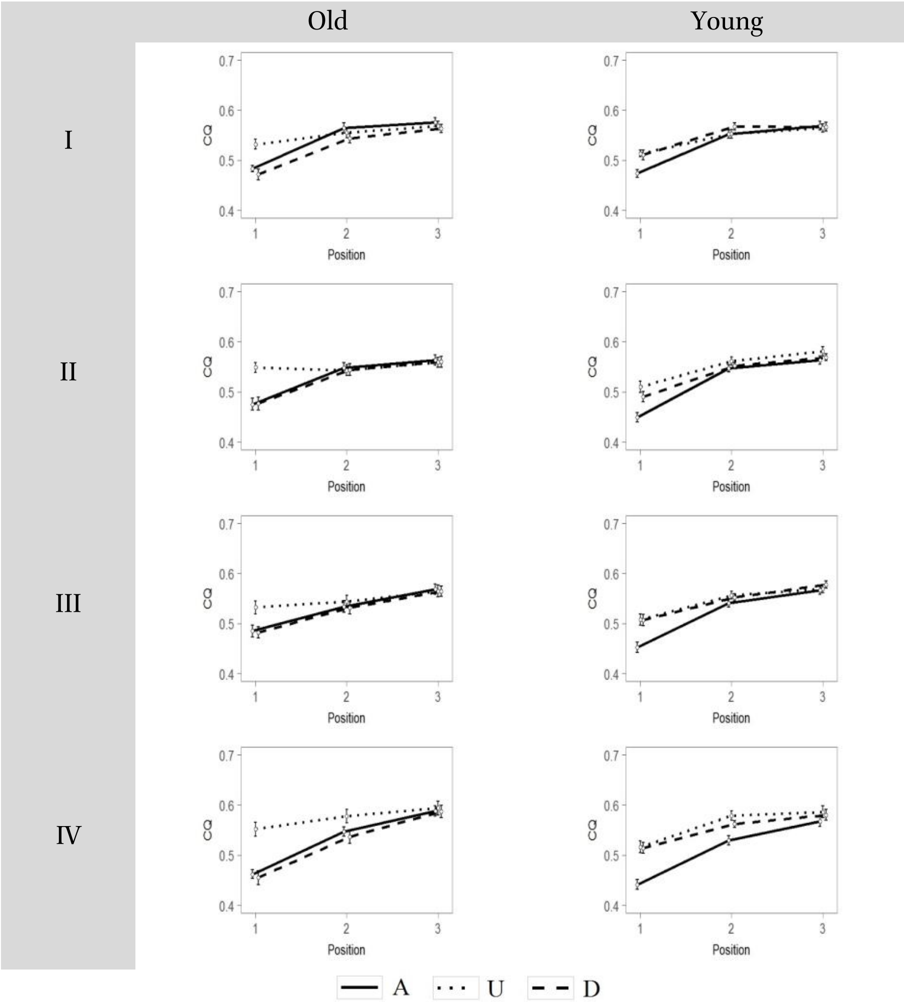

Figure 3.6 Mean CQ of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each MC tonal category. The x-axis represents the ratio of the vowel. The y-axis represents the CQ value. Error bars represent the standard error of the mean.

As seen in Figure 3.6, there are two general findings. First, CQ after A is consistently lower than that after U. However, CQ after D seems to be shifting across generations. It patterns quite similar to CQ after A in speakers of the old generation, but to CQ after U in speakers of the young generation. Second, the differences in CQ mainly exist at P1 (one-third) and generally vanish after P2 (middle). Both observations were supported by the results demonstrated in Table 3.9. First, regardless of tonal categories, CQ after A was significantly different from that after U in the old generation only, while it was significantly different from CQ after both U and D (reflected by the significant results at P1). Second, except for VI of the young generation, Consonant failed to improve the model fit at both P2 and P3 (indicated by '–').

Given that A was set as the reference category, for speakers of the young generation, the relationship between CQ after U and D is still not clear. If CQ after D indeed patterns with that after U, no significant effects are expected. In order to examine the assumption, multiple *post hoc* tests using Tukey's honestly significant difference (HSD) were further run to evaluate the differences between onsets at P1. As shown in Table 3.10, the results indicated that CQ after D maintained more comparable patterns to that after U [D ≈ U] and differed from that after A [D ≠ A]. It confirmed the visual inspection of Figure 3.6, which suggested a change of the CQ pattern following D in speakers of the young generation.

Table 3.10 HSD results (β) of the effect of Consonant on CQ at the one-third position (P1) of vowels for the young generation. Parameter-specific p-values were superscripted. n.s.: not significant.

Generation	Consonant	I	II	III	IV
Young	A vs. U	$.04^{<.001}$	$.06^{<.01}$	$.06^{<.05}$	$.08^{<.05}$
	D vs. A	$.04^{<.05}$	$.04^{<.01}$	$.05^{<.001}$	$.07^{<.001}$
	D vs. U	$-0.01^{\text{n.s.}}$	$-0.02^{\text{n.s.}}$	$-0.01^{\text{n.s.}}$	$-0.01^{\text{n.s.}}$

3.4.4 Summary II

The results of CQ conformed with those obtained via H1*–H2* (Section 3.3.6.3). Such a correlation of CQ and H1*–H2* thus echoes well with findings reported in Esposito (2012). Regardless of whether ATS happened (i.e., II *Shang*, III *Qu*, and IV *Ru*) or not (i.e., I *Ping*), a consistently lower CQ of vowels after voiceless aspirated onsets was observed but was only present at the beginning of adjacent vowels (i.e., one-third position). This pattern held across generations. The generational differences of the phonatory state of vowels after voiced onsets were further confirmed via the articulatory parameter (i.e., CQ). It is safe to conclude that the breathiness of vowels after voiced onsets is becoming weaker among speakers of the young generation.

3.5 Discussion

3.5.1 New light on the two existing debates

The primary goal of this study is to examine two long-standing debates. First, is 'aspiration-induced tonal split' (ATS) an ongoing change or a completed change? Second, is aspiration or breathiness synchronically related to ATS?

With respect to the first debate, a stepwise lowering of lexical tones beginning with aspirated onsets as described in Figure 3.1 has not been observed across generations. The 'ongoing change' view, therefore, is challenged. The results of growth curve analysis (GCA) instead tend to favor the 'completed change' view. A two-way categorization of *f0* contours conditioned by Middle Chinese (MC) tonal categories was consistently observed across generations. For the MC *Ping* (I) tonal category, both voiceless-onset types (i.e., aspirated and unaspirated) introduced similar high-level *f0* contours, while voiced onsets introduced low-rising contours. For the remaining three MC tonal categories (i.e., *Shang* II, *Qu* III, and *Ru* IV), *f0* contours of aspirated onsets exactly patterned with contours of voiced onsets. Both, however, differed from *f0* contours of unaspirated onsets. This pattern held across generations.

As regards the second debate, the results lend no support to either claim. Specifically, the results of the DOR/DOS ratio for aspirated onsets were similar in the MC *Ping*, *Shang*, and *Qu* tonal categories. All, however, were significantly lower than that of the *Ru* category across generations. This raises doubts about the 'aspiration' view, which argues that the aspirated onset is synchronically responsible for tonal split (e.g., Wang, 2008). If it was true, the DOR/DOS ratio for *Ping* should be significantly different from that for the other three MC tonal categories (i.e., *Shang*, *Qu*, and *Ru*). On the other hand, the results of H1*–H2* and CQ showed a consistently higher magnitude of breathiness in vowels after aspirated onsets (although only presented at the beginning of adjacent vowels), regardless of whether tonal split (i.e., *Shang*, *Qu*, and *Ru*) happened or not (i.e., *Ping*). Again, this pattern remained constant across generations. This finding is inconsistent with the expected results of the 'breathiness' view, which predicts a significant difference of breathiness in *Ping* vs. non-*Ping* and greater breathiness throughout the vowel (e.g., Hirayama, 2010) for the MC tonal categories showing ATS. Taken together, we may conclude that synchronically speaking, there is no consistent link between either aspiration or breathiness and tonal split in Lili Wu Chinese. The relationship between them is less transparent and more complex.

3.5.2 An ongoing change: the phonatory state of voiced onsets

One serendipitous finding is the reduced breathiness of vowels following voiced onsets as an ongoing sound change. Consequently, two questions are of particular note. First, why did this change happen? Second, what is the effect of this change on the phonological system? I will approach both questions from the perspectives of cue redundancy and robustness for signaling phonological contrasts.

In the Northern Wu dialects, there are a variety of acoustic cues for signaling the three-way laryngeal contrast, demonstrating a robustly encoded phonological contrast. For example, in Shanghainese, breathiness has been argued to act as a secondary cue for enhancement on vowels after voiced onsets, while the *f0* contour of a lexical tone is taken to be the primary cue for the contrast between syllables beginning with voiced and voiceless unaspirated onsets (Gao, 2015; Chen & Gussenhoven, 2015). In Dzongkha, a Tibetic language spoken in Bhutan, breathiness has also been regarded as a concomitant property to help enhance the perceptibility of voicing (Kirby & Hyslop, 2019). The status of breathiness in Lili Wu Chinese seems comparable. As demonstrated in Table 3.11, breathiness on vowels after voiced onsets has a superfluous role in cueing the three-way laryngeal contrast. First, for the contrast between voiced and voiceless aspirated onsets, VOT combined with *f0* suffices as a robust cue in the MC *Ping* category and VOT suffices as a robust cue in the MC *Shang*, *Qu*, and *Ru* categories. Second, with regard to the contrast between voiced and voiceless unaspirated onsets, lexical tonal contrast serves as a robust cue. Finally, for the contrast between voiceless aspirated and unaspirated onsets, VOT serves as the prominent cue in the MC *Ping* category and both VOT and lexical tonal contrast serve as primary cues for the MC *Shang*, *Qu*, and *Ru* categories. Given the superfluous role of breathiness in signaling any laryngeal contrasts of Lili Wu Chinese, it is not difficult to understand the reduced degree of breathiness in vowels following voiced onsets produced by younger speakers.

Table 3.11 Acoustic cues used for signaling the three-way laryngeal contrast in Lili Wu Chinese.

<table>
<tr><th>vs.</th><th colspan="2">Voiceless aspirated</th><th colspan="2">Voiceless unaspirated</th></tr>
<tr><th></th><th>Ping</th><th>Shang/Qu/Ru</th><th>Ping</th><th>Shang/Qu/Ru</th></tr>
<tr><td>Voiced</td><td>VOT & f0</td><td>VOT</td><td colspan="2">f0</td></tr>
<tr><td>Voiceless aspirated</td><td colspan="2">–</td><td>VOT</td><td>VOT & f0</td></tr>
</table>

With respect to the second question, generally speaking, the decrease of breathiness of vowels with voiced onsets can potentially threaten the three-way laryngeal contrast. The contrast between voiced and voiceless unaspirated onsets is in jeopardy of losing its cue robustness. The weakening of cue robustness, to a large extent, introduces more bias, which then can reduce the precision of the contrast (Kirby, 2013). On the other hand, a contrast is more likely to survive when more cues signal it (Stevens & Keyser, 1989; de Jong, 1995; Wright, 2004). Predictably, if no strategy of enhancement is taken by younger speakers, the loss of the redundant cue can eventually lead to the three-way laryngeal contrast becoming less robust. Very likely, with the weakening of cues for distinguishing the voiced vs. voiceless unaspirated contrast, neutralization of the three-way contrast may be triggered.

3.6 Conclusion

This chapter has provided a substantial amount of experimental data collected from Lili Wu Chinese to examine two debates of 'aspiration-induced tonal split' (ATS) in previous literature. In conclusion, the results suggest that ATS in Lili Wu Chinese is a completed sound change but conditioned by certain tonal contexts (i.e., Middle Chinese tonal categories). Synchronically speaking, a consistent link between either aspiration or breathiness and tonal split is not tenable in Lili Wu Chinese. This fact is reminiscent of the statement in Chen (2011: 622): '(…) [S]peakers may use different strategies to produce aspirated stops in different languages which lead to different perturbation effects'. Based on the findings observed in Lili Wu Chinese, different strategies can also be adopted by speakers even within the same language.

One ongoing sound change observed serendipitously is that the breathiness of vowels after voiced onsets is disappearing among the young generation of Lili speakers. This is probably due to its superfluous role in cueing the three-way laryngeal contrast which makes it a less robust cue for the laryngeal contrast in Lili Wu Chinese. It may expect that this ongoing language change can lead to the loss of this cue for the three-way laryngeal contrast in the future.