# 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

## Consonant and lexical tone interaction

### Evidence from two Chinese dialects


This dissertation provides a comprehensive look at the consonant-tone interaction embedded in a description of the sound system of two under-documented Chinese dialects, namely Lili Wu Chinese and Shuangfeng Xiang Chinese. In the existing literature, consonant-tone interaction generally concerns a [voiceless/H]-[voiced/L] co-occurrence pattern. A high tone usually co-occurs with a voiceless consonant, while a low tone usually co-occurs with a voiced consonant. However, largely because of the high level of homogeneity in the languages sampled, and the lack of access to up-to-date statistical techniques, this [voiceless/H]-[voiced/L] pattern has veiled the full picture of consonant-tone interaction across the world's languages.

Based on a series of phonetic studies of phonological contrasts, there are two key findings that contribute to our understanding of the diversity in consonant-tone interaction. First, voiceless aspirated onsets can also co-occur with low tones. This finding is antagonistic to the [voiceless/H]-[voiced/L] pattern which posits that only contrastively voiced onsets can be in favor of low tones. Second, the realization of consonant-tone interaction is not only specific between languages but also within languages. Speakers of different generations of a given language can utilize phonetic cues differently to signal the same phonological contrasts.

This dissertation will be of interest to experimental phoneticians, laboratory phonologists, as well as to those interested in sound change. Understanding the interaction between consonant and tone also contributes to our knowledge of the sound systems of Chinese dialects and more broadly speaking, phonology-phonetics interface.



ISBN 978-94-6093-351-6

:LOT Netherlands Graduate School of Linguistics
Landelijke Onderzoekschool Taalwetenschap

:LOT 566



Menghui Shi

Consonant and lexical tone interaction


LUCL

Menghui Shi

# Consonant and lexical tone interaction

## Evidence from two Chinese dialects


:LOT



LUCL

# Consonant and lexical tone interaction

Evidence from two Chinese dialects

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461
e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: The tablature notation of <*Jiu Kuang*> (酒狂) [Wine Mad] recorded in <*Shenqi Mipu*> (神奇秘谱) [The Handbook of Spiritual and Marvelous Mysteries] (1425 CE), designed by Jinzhe Ouyang.

ISBN: 978-94-6093-351-6
NUR: 616

# Consonant and lexical tone interaction

## Evidence from two Chinese dialects

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 4 juni 2020
klokke 10.00 uur

door

Menghui SHI

geboren te Suzhou, China
in 1989

Promotores: Prof.dr. Yiya Chen
Prof.dr. Maarten Mous

Promotiecommissie: Prof.dr. Laura Downing
(University of Gothenburg)
Dr. Silke Hamann
(Universiteit van Amsterdam)
Dr. James Kirby
(University of Edinburgh)
Prof.dr. Claartje Levelt

For my parents, family, and those whom I love

献给我的父母、家人以及那些我爱的人

青春都一饷。
忍把浮名，
换了浅斟低唱。

The prime of one’s life is too short.
Better to barter empty fame
For the pleasures of good wine and sweet song.

– 柳永 (987?–1053?)
(Translated by Xianyi Yang 杨宪益 and Gladys Yang 戴乃迭)

# Contents

# List of Tables

# List of Figures

# Acknowledgments

First of all, I would like to acknowledge the financial support from the Chinese Scholarship Council (CSC), Leids Universiteits Fonds (LUF), and Leiden University Centre for Linguistics (LUCL). Without them, I would not have been able to come to Leiden and complete this dissertation.

On 5th September 2015, when I set foot on Leiden's soil, I had taken this journey into the ocean of my PhD life. I must admit without my supervisors – Yiya Chen and Maarten Mous, I would have drowned already and have become bubbles in the sea. Yiya, you accompanied me to grow up. There are currently 368 conversations containing more than 1,500 emails between us in my mailbox. Maarten, you are the beacon of hope – a place that keeps orienting my sailing direction. I am looking forward to our next encounter in the market on Saturdays.

I owe a debt of gratitude to all the participants and native speakers of my experiments, too numerous to list, who ended up teaching me the most about the languages I worked on. Particularly, I would like to thank my two principal consultants – Mr. Liangquan Cheng and Mrs. Suxin Xu for making Chapter 2 and Chapter 4 possible. I am also grateful to Mr. Haimin Li and Mr. Shijia Zhou for your coordination and arrangement for the fieldwork sessions conducted in Lili and Shuangfeng, respectively.

There are also a lot of senior scholars whom I owe a deep gratitude to: Prof. Huan Tao, for enlightening the entrance of linguistics for me; Prof. Zhongmin Chen, for bringing me the first breeze of international studies of linguistics; Prof. Rujie You, for introducing Leiden University to me; Prof. Zhongwei Shen, for inspiring me to take Lili Wu Chinese as a case language in my dissertation; Academician Dah-an Ho and Prof. Hsiu-Fang Yang, for encouraging me to go along my own path. Besides, I really appreciate those who agreed to evaluate my dissertation, including Prof. Laura Downing, Dr. Silke Hamann, Dr. James Kirby, and Prof. Claartje Levelt. Thank you all for spending time reading this work and providing excellent comments and suggestions.

This paragraph is for my paranymphs – Hang Cheng and Cesko Voeten. Even if it is time for us to grow apart, it doesn't change the fact that for these years we grew side by side.

This PhD project received great help from many other people. In particular, I am very grateful to Qian Li, for introducing growth curve analysis to me; to Min Liu, for instructing me model settings; to Yifei Bi, Ruiqing Shen, and Dan Yuan, for your comments on some chapters of this dissertation; to Feng Ling and Minghui Zhang, for your help on Praat scripts from time to time; to Jiayin Gao, Yan Gu, Mengru Han, Han Hu, Yaohua Jin, Xinyi Wen, Jiang Wu, Shengyi Wu, Qing Yang, Yang Yang, Tingting Zheng, and Jinlei Zhou, for your insightful discussion with me; to Jianjing Kuang, Jia Tian, and Yingyi Zhou, for your valuable remarks on the issue of phonation; to Feifan Wang, for helping me collect pilot recordings; to Qi Zhao, for helping me contact consultants; to Prof. Fang Hu, Prof. Yun Mai, Prof. Rujie Shi, Prof. Hongjun Wang, Prof. Ping Wang, Prof. Wei Zheng, Prof. Xiaonong Zhu, Ruoxi Cheng, Jian Ding, Chengyu Guo, He Huang, Yiming Sheng, Huaqiang Song, Lei Wang, Jingting Ye, Jingwei Zhang, Menghan Zhang, and Yifei Zheng, for sharing various literature and materials with me. The proofreading assistance from Kate Bellamy, Seamus Leith, Seumas Macdonald, Daan van de Velde, and Andrew Wigman are gratefully appreciated. I would also thank André Radtke for your audio editing guidance.

Further thanks go to my teachers in the Netherlands. You share your interesting thoughts with me on various aspects. They are: Aoju Chen, Lisa Cheng, Janet Grijzenhout, Carlos Gussenhoven, Willemijn Heeren, Maarten Kossmann, Constance Kutsch Lojenga, Jos Pacilly, Leticia Pablos Robles, Gijsbert Rutten, Niels Schiller, Rint Sybesma, and Jeroen Wiedenhof.

All colleagues at LUCL and friends in the Netherlands are a source of my strength during these years. Ami, Astrid, Gulnaz, Hanna, Lis, Renzi, Sarah, Xander, and Zhen, I will cherish the time I spent with you in our PhD Council. Ailin, Brother Sao, Chunhai, Feng, Hongling, Jing, Kejia, Libo, Meiyu, Shaoyu, Sister Xin, Wei, Weiqing, Xinxin, Yuan, Zexu, and Zhaole, I shall never forget those indelible moments with you. Others like Aliza, Adrian, Chams, Chen, David, Federico, George, Hana, Hongmei, Jie, Junru, Laura, Meike, Rasmus, Shuangshuang, Ting, Xiaowen, Xin, Yao, and Zhuoyi, each of you lent me continued encouragement and support.

At last, I own my thanks to Vijf Meibad and De Zijl – the two swimming pools in Leiden, which accompanied me through hundreds of evenings.

Leiden
February 2020

# Chapter 1 Introduction

## 1.1 General background

What is the relationship between the initial consonant and fundamental frequency (*f0*)? It has long been known that the initial consonant and *f0* interact irrespective of whether *f0* distinction is phonologized.

In non-tonal languages, very commonly, voiceless consonants co-occur with a higher *f0* onset in the following vowel, while voiced consonants introduce a lower *f0* onset (see reviews in Kingston & Diehl, 1994 and Dmitrieva et al., 2015). This binary pattern has been widely attested ([voiceless/H]-[voiced/L] hereafter) across the world's languages, although there is still a debate on possible sources of such a pattern in different languages (i.e., *f0* being lowered following voiced obstruents vs. *f0* being raised following voiceless obstruents, see Kirby & Ladd, 2016 and references therein). This interaction between the initial consonant and *f0* is known as 'obstruent intrinsic *f0*' (Hanson, 2009) and labeled as C*F*0 in Kingston (2007).

In tonal languages, the [voiceless/H]-[voiced/L] pattern between the initial consonant and onset *f0* has been further claimed to sow a seed in the phonologization of *f0*, which highly contributes to the development of lexical tonal contrasts (after Maspero, 1912). At a synchronic phonetic level, the appearance of high tone (usually having a higher *f0* onset) in a vowel adjacent to a voiceless consonant, and of low tone (usually having a lower *f0* onset) in a vowel adjacent to a voiced consonant is not fortuitous (e.g., Fromkin, 1972; Hombert, 1978; Hombert et al., 1979; Bradshaw, 1999; Yip, 2002; Tang, 2008). Although there is still a debate on the degree of relevance between C*F*0 and the process of phonologization of tonal contrasts (see Thurgood, 2002, 2007 for a direct relevance between voice quality distinctions on vowels and tone contrasts), the [voiceless/H]-

[voiced/L] pattern has been broadly observed in tonal languages spoken in Asia and Africa (see Tang, 2008: 25–6, Table 2 for more cases).

The [voiceless/H]-[voiced/L] pattern seems to be accepted as a well-established setting in the issue of the interaction of initial consonant and lexical tone (C-T interaction hereafter). However, such a pattern is challenged by data from two under-documented Chinese languages/dialects reported in this dissertation,[1] which will help us learn more about uncommon C-T interactions that are different from the well-known [voiceless/H]-[voiced/L] pattern. These uncommon interactions thus provide a more complete picture of how consonants and tones interact in different languages.

## 1.2 Property-based approach

Traditionally speaking, as summarized in Tang (2008), there are two approaches for exploring C-T interaction, namely the rule-based ('feature-based' in Tang's term) approach and the constraint-based approach. Both approaches echo well the two mainstream frameworks in the field of phonology over the past half-century (after Chomsky & Halle, 1968; see a recent review in Newell, 2018 and references therein).

The rule-based approach explores C-T interaction by assigning distinctive laryngeal features to different interactive patterns. For example, Stevens and Halle (1971) bridge a direct connection among the stiffness/slackness of the vocal folds, the high/low frequency of the tone, and the voiceless/voicing of the consonant. Due to the shared [+stiff vocal folds] feature, voiceless consonants can interact with a high tone; while voiced consonants can interact with a low tone due to the shared [+slack vocal folds] feature. This basic idea on accounting for C-T interaction is

[1] In this study, the term 'dialect' is consistently used for a language spoken in China and belonging to the Sinitic branch. Interested readers are referred to Norman (2003: 72) on the discussion of 'dialect' and 'language' in China.

further adopted by Yip (1995) and Bao (1999) via the introduction of the 'register' (h/l) dimension to accommodate new observations mainly from Chinese dialects. This proposal in Bradshaw (1999, 2000), termed the 'multiplanar approach,' is extremely simplified as a privative correlation between low tone and [voice] ([L/voice]).

The constraint-based approach mainly adopts the framework of Optimality Theory (OT) in accounting for C-T interaction. For example, Peng (1992) illustrates 32 logically possible combinations on C-T interaction constrained by groundings of different features (i.e., pairing four tonal features with two voice features). He further claims that only 16 of them can be observed in natural languages. In the dissertation of Lee (2008), this constraint-based approach is labeled as the 'Extended Tone Bearing Unit Theory' (xTBU theory), which relies on the type of nodes (i.e., the moraic vs. non-moraic root node) that associate tones and consonants. In addition, the constraint-based approach is also applied to a handful of case studies, such as on Yabem (Hansson, 2004), Thai (Morén & Zsiga, 2006; Perkins, 2013), and Nguni languages (Downing, 2018).

The illuminating classification of Tang provides us with an accessible lens to look into the existing literature of C-T interaction. Both approaches are particularly concerned with the abstract phonological representations of both consonants and tones (the 'abstract' approach hereafter, Schwarz et al., 2019) and tend to explore 'the logical properties of a specific model' for C-T interaction (Hyman, 2018: 5), which mainly relies on transcription at the level of phoneme. The abstract approach serves as an analytic perspective but should not be the only perspective. It needs to be augmented by substantial properties of actual phenomena.

To have a better understanding of the universality and diversity of C-T interaction observed in the world's languages, in this study, I will adopt a 'property-based' approach (Hyman, 2018). The goal of the property-based approach is in line with the core idea of laboratory phonology, which is to study the relationship between phonological feature representations and their phonetic realizations via experimentation with integrated methodologies (Cohn et al., 2018: 504). In doing so, we can have a better picture of the phonology and phonetics of C-T interaction.

It is worth noting that such an approach has helped to achieve great progress in our understanding of laryngeal features and their acoustic realizations across the world's languages. A particularly fruitful line of research is the relationship between voice onset time (VOT) and consonant voicing contrast. It has been shown that languages arbitrarily choose a 'modal' VOT value of their own to signal similar voicing contrasts (e.g., Keating, 1984; Kingston & Diehl, 1994; Cho & Ladefoged, 1999; Dmitrieva et al., 2015; Kirby & Ladd, 2016; Cho et al., 2019). For example, the phonological voiced onset is realized with a short-lag VOT in English (usually known as 'aspirating languages'), but with a lead VOT (prevoicing) in French (usually known as 'voicing languages'). Such an incongruence of the phonetic realization, consequently, brings about language variation and language specificity.

The abstract approach argues that the phonological representation of the laryngeal contrast can be an abstraction from the phonetic realization of the categories.[2] In contrast, the property-based perspective named 'laryngeal realism' (Honeybone, 2002) argues that differences in phonetic realization across different languages should be represented as underlyingly different phonological features (see a comprehensive review in Schwarz et al., 2019). The essential difference between these two approaches is couched in the mapping from phonological features to phonetic realizations. For the abstract approach, the mapping can be non-transparent, while for the laryngeal realism there needs to be a direct and transparent mapping between the two. The two approaches lead to different treatments of aspiration. Within the abstract approach, the same feature [voice] is sufficient for both aspirating and

[2] It is also referred to as the 'broad' interpretation in Hall (2001), or the 'traditional' approach in Honeybone (2002, 2005).

voicing languages, while the laryngeal realism suggests that for aspirating languages, the [spread glottis] feature instead of [voice] is needed (Beckman et al., 2013).

By examining both the phonetic realizations of segments and their phonological behaviors, an increasing body of research has shown the validity of the laryngeal realism approach in understanding laryngeal contrasts (e.g., Iverson & Salmons, 1995; Honeybone, 2002, 2005; Beckman et al., 2013 for languages featuring a two-way laryngeal contrast; and Harris, 1994: 135; Iverson & Salmons, 1995: 383; Schwarz et al., 2019 for languages featuring a relatively more complex contrast) (but see criticism in Schwartz & Arndt, 2018).

In the current study, I will follow the property-based approach for exploring C-T interaction and illustrate that it is important to adopt such an approach. A crucial reason for valuing phonetic properties in phonology is that it sometimes can help account for variation within phonological categories and interpret or foreshadow sound changes that historically happened or could potentially happen. In speech production, a phonological category is often signaled by a collection of co-varying phonetic cues (see recent reviews in Kuang & Cui, 2018; Yang & Sundara, 2019). Let's return to the topic of voicing contrast again. It has been argued that the phonetic nature of voicing contrast cannot be fully captured by VOT alone. Instead, different phonetic dimensions such as vowel formant transitions (e.g., Slis & Cohen, 1969) and C*F*0 (see a comprehensive review in Cho et al., 2019) are all at play to signal phonological voicing contrasts. For example, three languages spoken in Southeast Asia, two tonal and one non-tonal, all show similarities in VOT but differences in C*F*0. Based on these findings, Kirby (2018) concludes that the use of C*F*0 in conjunction with voicing contrast can be controlled by the speaker and contributes to the phonological system of a given language.

Moreover, C*F*0 has been argued to act as the source of tonogenetic sound change. It is not only assumed in research on the historical development of a non-tonal language to a tonal language (e.g., Vietnamese) but also has been observed at a synchronic level of modern languages. For example, in Seoul Korean, aspirated plosives have longer VOT values and higher *f0* onset than lenis plosives. Seoul Korean has also been reported to be undergoing a tonogenetic sound change with the loss of VOT distinction between aspirated and lenis plosives among younger speakers (see Kang, 2014 and references therein). The primary cue for the aspirated-lenis contrast is shifting from VOT distinction on the consonants to pitch difference on the vowels. A similar pathway has also been reported in Afrikaans (Coetzee et al., 2018) and Swiss German (Ladd & Schmid, 2018). Without the exploration of phonetic properties, such changes would have been impossible to uncover. In order to understand the cross-linguistic differences of C-T interaction, it is necessary to look into phonetic properties from multidimensional perspectives. These multidimensional phonetic properties may be phonologically redundant but phonetically informative in understanding universals and variations in C-T interaction across the world's languages.

In this current study, two Chinese varieties, namely Lili Wu and Shuangfeng Xiang, will be investigated. In doing so, I hope to achieve a better understanding of C-T interaction. More importantly, insights from experimental phonetic research can benefit our understanding of phonological typology, which, in turn, allows us to answer the question of 'how different languages systematize the phonetic substance available to all languages' (Hyman, 2009: 213).

## 1.3 Motivation

There are manifold studies on C-T interaction that have achieved fundamental success in the analysis of the [voiceless/H]-[voiced/L] pattern. However, two main biases can still be discerned in the existing literature as discussed below.

### 1.3.1 Typological bias

The first bias of the previous studies on C-T interaction is that of typological representation. A number of well-known studies on C-T interaction base their conclusion on comparisons of a range of languages. As we can see in the three dissertations related to this topic, all of them have tried to get as many languages as possible (i.e., 17 in Bradshaw, 1999; 61 in Tang, 2008; and more than 24 in Lee, 2008). However, these accomplishments suffer from the problem of language sampling. The ideal way to discover the typology of any phonological pattern is to examine all languages spoken in the world. This, however, is practically impossible. On the one hand, there are over 6,000 living languages in the world today (Ethnologue, 2019). On the other hand, only less than 10% of them have been decently described or documented (Evans & Levinson, 2009: 432). It is unrealistic for an individual researcher or study to compare such a large number of languages with only very limited and poor descriptions available. As one can imagine, language sampling is always needed (see Song, 2018, Chapter 5 for a detailed discussion).

Generally speaking, four geographical areas of the world contain many tonal languages: Africa, East and Southeast Asia, the Pacific, and the Americas (Yip, 2002). However, previous studies have mainly focused on C-T interaction in African languages (e.g., Bradshaw, 1999). The remaining areas are rarely mentioned and therefore have not been examined sufficiently. The overgeneralized claim of the privative correlation between low tone and [voice] argued by Bradshaw (1999), to a large extent, results from a highly geographical concentration of language sampling on African languages (also see previous criticisms in Tang, 2008 and Downing, 2009). Although in subsequent studies, Tang (2008) and Lee (2008) have tried to avoid such a bias by including more non-African data. Both, however, seem to help enlarge the inventory of languages that have similar patterns of C-T interaction. As suggested by Song (2018: 91), if studies contain a disproportionate number of languages with limited linguistic types, the outcome may not be complete. It is important to point out that the data of Chinese dialects have not been sufficiently exploited.

One must admit that descriptions and documentations abound in Chinese dialects with a surprisingly high degree of completeness and diversity. Almost every description contains a chart labeled as 'phonotactic constraints on onset, rhyme and tone' (声韵调配合关系), but the utilization of these charts is very limited. The most obvious reason is that it is due to the different conventions of the academic language as most descriptions are in Chinese rather than in English.

In addition to the language limitation, there are three other reasons. First, researchers adhering to the sinological tradition usually tend to analyze C-T interaction of Chinese dialects from a historical perspective with a wealth of terms inherited from Chinese philology (音韵学). Consequently, the existing literature remains obscure to general researchers with limited knowledge of the research conventions within sinology. Second, even for researchers who try to subsume Chinese data under their umbrella, they are still constrained by limitations of the accuracy of the material. For example, Ye (1983) describes a co-occurrence pattern between aspirated onsets and mid-tones in Songling Wu Chinese. Lee (2008: 41–2) hence cites it as one crucial example to argue for the universal incompatibility between [+spread glottis] and low tone. Aspirated onsets can co-occur with low tones in most Wujiang dialects including Songling (see Chapter 2 in this dissertation). Ye's description has been contested for a long while, at least starting from Zhang and Liu (1983). Such an accuracy bias is mainly due to the lack of accurate description of empirical data. Based on the poor quality of the material without a unified working frame and reliable quality,[3] it is a tremendous

[3] The symbols of sound very often are 'loosely and carelessly used without precise phonetic specification' (Catford, 1977: 203), let alone the idiosyncratic or even

challenge for researchers who aim to integrate a load of facts across the world's languages into one study. Third, there is a discrepancy of the ultimate goal between most of the Chinese dialectological studies and general studies. One crucial goal of Chinese dialectology is to identify the developmental trajectory from Middle Chinese (MC) to modern dialects, and further to evaluate the closeness of the genetic relationship between different dialects (e.g., Yuan, 1960: 14; Li & Xiang, 2009: 21). C-T interaction is always a key window to reconstruct the genetic relationship between dialects. This goal, however, keeps researchers from exploring synchronic universalities and underlying rules/hierarchies of C-T interaction across the world's languages.

Nevertheless, in previous studies, there are also some studies focusing on C-T interaction in Chinese dialects, but going beyond the scope of sinological perspective. These studies, generally speaking, examine C-T interaction via the instrumental method with empirical data (e.g., Howie, 1974, Chao, 1992, and Xu & Xu, 2003 for Beijing Mandarin; Zee, 1980 and Francis et al., 2006 for Hong Kong Cantonese; Lai et al., 2009 for Taiwan Min; Chen, 2011 for Shanghai Wu; Lai & Chen, 2011 for Taiwan Mandarin). However, in terms of the laryngeal contrast of obstruents, except for Chen (2011), the target languages of these studies show typological similarity. That is, all languages feature a two-way laryngeal distinction of obstruents (i.e., voiceless unaspirated vs. voiceless aspirated).

Although there exist various proposals on the classification of Chinese dialects, the seven groups, consisting of Mandarin, Wu, Xiang, Gan, Hakka, Min, and Yue, have been widely recognized (Yuan, 1960). Typologically speaking, the laryngeal contrast of obstruents among these varieties falls into either a two-way (i.e., voiceless unaspirated vs. voiceless aspirated) or a three-way (i.e., voiceless unaspirated vs. voiceless aspirated

---

controversial documentation of the descriptive conventions inherited among researchers focusing on different language areas.

vs. voiced) distinction. It has been widely argued that the two-way system developed from the three-way system inherited from MC via devoicing of the voiced category (e.g., Ting, 1982; Norman, 1988; Yang, 1989). According to a large-scale survey of 140 Chinese dialects covering all seven dialect groups (Ye, 2011: 12), it is true that the majority of Chinese dialects (98/140, 70%) maintain a two-way laryngeal contrast of obstruents. Among them, a vast majority (94/98, 96%) show a contrast of voiceless unaspirated vs. voiceless aspirated such as Standard Chinese and Cantonese. However, there are still 30% (42/140) of dialects, such as Shanghainese, predominately showing a three-way pattern with the contrast of voiceless unaspirated vs. voiceless aspirated vs. voiced (39/42, 93%). There has been very little research focusing on these dialects.

So why is it so important to investigate the Chinese dialects bearing a three-way laryngeal contrast of obstruents? A simple answer, in a nutshell, is that the C-T interaction in three-way-contrast dialects has been reported to show more complicated patterns compared to the common binary pattern (i.e., [voiceless/H]-[voiced/L]).

In two-way-contrast dialects, lexical tones co-occur with all initial consonants. For example, in Standard Chinese, the four lexical tones can combine with both /t/ and /$t^h$/. There is no doubt that lexical tones after onsets with different laryngeal configurations can be categorized as one identical toneme. For instance, /$ta^4$/ 'big' and /$t^ha^4$/ 'to tread' both can carry T4, the falling contour in Standard Chinese, although the acoustic study by Xu and Xu (2003) shows a lowering effect of aspirated onsets on *f0* of the following vowel. This lowering effect also exists in three other lexical tones but with different degrees. However, no matter how big the perturbation effects are, the two *f0* contours with the same lexical tone converge to the same pattern towards the syllable offset and are considered the *f0* realizations of the same lexical tone.[4]

[4] The lowering effect of aspirated onsets on *f0* onset is not consistently observed among dialects. The opposite direction, namely the rising effect, has also been reported in

In dialects with the three-way contrast, the situation is more complicated. Very commonly, distributional constrictions between initial consonants and lexical tones still stick to the [voiceless/H]-[voiced/L] pattern (Pan, 1982). For example, in Shanghai Wu Chinese (Chen & Gussenhoven, 2015), aspirated and unaspirated obstruent onsets allow for a two-way contrast with T1 (falling), T2 (high–rising), and T4 (short–high–level). When the onset is a voiced obstruent, only T3 (low–rising) and T5 (short–low–rising) are possible. T1, T2, and T4 all start within a high *f0* register, traditionally known as the *Yin-* (阴, high) register tones. Both T3 and T5 start within a low *f0* register, traditionally known as the *Yang-* (阳, low) register tones. Such interactions in Shanghainese can be summarized as a [voiceless/H]-[voiced/L] pattern as demonstrated in Figure 1.1(a).

However, the [voiceless/H]-[voiced/L] pattern is not always the case among the three-way-contrast dialects. Uncommon C-T interactions have also been observed. Typologically speaking, several pieces of the mosaic are still missing in the full picture of C-T interaction. Among these missing mosaics, two subtypes can be identified.

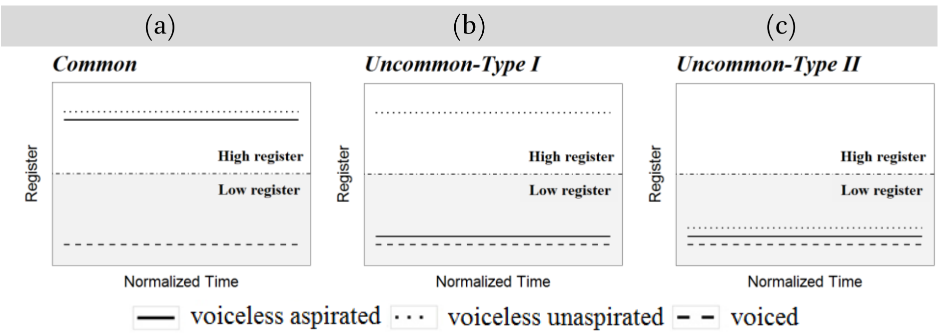



**Figure 1.1** Common and uncommon interactions within the three-way laryngeal contrast of initial consonant and lexical tone in Chinese dialects.

---

dialects such as Cantonese (Zee, 1980), Taiwan Min (Lai et al., 2009), and Taiwan Mandarin (Lai & Chen, 2011).

The first type, as shown in Figure 1.1(b), presents a significant *f*0-lowering effect in syllables with voiceless aspirated onsets, as if a split of the same tone into two as a function of voiceless unaspirated vs. aspirated onsets. Such a pattern suggests an invasion of voiceless aspirated onsets into the low register if assuming there exists a contrast of high and low tonal registers. This phenomenon, known as 'aspiration-induced tonal split' (Chao, 1928),[5] has been found mainly in the Wu and Xiang dialect groups, as well as some Kra-Dai and Hmong-Mien languages (Ho, 1989; F. Shi, 1998; Chen, 2005; Yue Xu, 2013). The second type, as shown in Figure 1.1(c), presents a further invasion of voiceless onsets into the low register. Not only the lexical tone of voiceless aspirated onsets but also that of voiceless unaspirated onsets get closer to the lexical tone of voiced onsets. This phenomenon is termed '*Yiniu tongdiao*' (异纽同调, Chao, 1959) among previous sinological literature and has been reported to appear in some Old Xiang Chinese and a few Wu Chinese (Chen, 2004).[6] '*Yiniu tongdiao*' can be translated as 'different initial consonants associated with one toneme' ('initial-associated tonal merger' hereafter). Among previous studies, both uncommon C-T interactions in Chinese dialects have not been figured in the literature on C-T interaction outside of the sinological research community, which highly motivated this study.

---

[5] In the existing literature, the same phenomenon has been referred to as '*Songqi fendiao*' [送气分调, lit. aspiration divides tones] (e.g., Ho, 1989), '*Qiliu fendiao*' [气流分调, lit. airflow divides tones] (Xu, 2006), or '*Ciqing fendiao*' [次清分调, lit. secondary voiceless divides tones] (Zhu & Xu, 2009). The English appellations include 'aspiration-conditioned tone-lowering' (Sagart, 1981), 'tone-split by aspiration' (Ho, 1989), 'aspirated tones' (Shen, 1994), 'tonal split based on the aspiration' (F. Shi, 1998), and 'tonal split following voiceless aspirated stop onsets' (Chen, 2011).

[6] Since voiceless and voiced onsets specifically refer to the voicing contrast of the consonantal onsets, '*Yiniu tongdiao*' is also called '*Qingzhuo tongdiao*' [清浊同调, lit. voiceless and voiced onsets associated with one toneme] in most literature.

### 1.3.2 Methodological bias

The second bias regards methodology. Initially, previous studies tended to analyze C-T interaction based on an impressionistic description only, i.e., without instrumental data. With the prevalence of various measurement instruments and the increased awareness of the limitations of impressionistic transcription, the situation has gradually improved, especially among researchers who take the laboratory phonology approach to study sound patterns (Cohn et al., 2018). Over the past forty years, an increasing number of studies have started to explore C-T interaction with experimental data. This trend does not only happen among Chinese languages (see Section 1.3.1 for case studies) but also among tonal languages across the world (e.g., Erickson, 1975, Gandour & Maddieson, 1976, Perkins, 2013 for Thai; Hombert, 1978 for Yoruba; Pearce, 2009 for Kera; Mathes, 2015 for Tsua; Gordon, 2015 for four indigenous languages of the Americas; Kirby, 2018 for Thai and Vietnamese). Particularly, it is noted that the work of Hombert (1978) marks an incipient and successful exploration of C-T interaction via the instrumental method. Some well-controlled experiments were conducted across different languages (i.e., American English, Yoruba, and French) to examine the perturbations caused by prevocalic consonants on *f0* of adjacent vowels. A notable innovation, in terms of measurements, is that *f0* values at different time points rather than an averaged or a peak of *f0* were measured. At least three general conclusions can be drawn. First, the *f0* values of vowels after voiceless plosives are higher than after voiced plosives. Second, the duration of the perturbations is shorter in a tonal language (i.e., Yoruba) than in a non-tonal language (i.e., American English). Third, the perturbation effect of voiceless aspirated plosives is not significantly different from that of voiceless unaspirated plosives. These findings are further utilized as evidence to explore the birth of tones in Hombert et al. (1979).

These studies, however, face two problems in methodology. One is related to the measurement of the *fo* contour. The other is of the measurement of the laryngeal contrast.

The problem of the measurement of the *fo* contour is two-fold. First, the majority of these studies tend to measure *fo* values at several equidistant time points over syllables. However, the method of *fo* measurement differs from study to study. For example, Zee (1980) and Perkins (2013: 55) make a measurement every 10 ms throughout the vowel, but Hombert (1978) every 20 ms. Lai et al. (2009) and Gordon (2015) measure *fo* at equidistant percentages of the vowel. Besides, some previous studies tend to focus on the measurement of the so-called onset *fo*. Again, its criterion varies. In Zee (1980), the onset *fo* is defined as the mean value of the first three time points (i.e., the mean value from 0 ms to 30 ms). However, the onset *fo* refers to the *fo* value of the first valid vocal cycle of the target syllable in Xu and Xu (2003), and the *fo* value at 10 ms in Perkins (2013), respectively. Although all results turn out to show a significant effect of different onsets on the so-called onset *fo*, their measurements are essentially different. Therefore, strictly speaking, these cross-linguistic studies are not directly comparable.

Furthermore, after the extraction of *fo* values at discrete time points, to compare the distinction of two *fo* contours, the traditional approach is to run separate analyses for individual time points like *t*-tests or ANOVAs. The conventional strategy of treating *p*-values smaller than 0.05 (or 0.01) functions as the statistical threshold for rejecting a null hypothesis. However, there is a trade-off between 'statistical power, which requires more data and therefore larger time windows, and temporal resolution, which requires smaller time windows but undermines statistical power' (Mirman, 2014: 5). In tonal languages, an *fo* contour is commonly believed to show distinctiveness based on lexical tone realization (Yip, 2002; Hyman, 2006). The *fo* contour is a gradual change that can be difficult to detect with discrete and discontinuous observations. To illustrate this point, let's imagine the results of 10 point-by-point *t*-tests of two *f*0 curves. In 5 of the 10 points, the difference is significant, but the significance regularly occurs at even points. How could we tell the similarity of both curves? Such a dilemma, to a large extent, is caused by decomposing gradual changes (i.e., *f*0 contours) into seemingly discrete, discontinuous differences. In the study of Tsua C-T interaction, Mathes and Chebanne (2018) have realized this flaw and tried to capture the dynamicity of the entire *f*0 contour via Smoothing Spline Analysis of Variance (SS ANOVA) (Gu, 2013). The basic idea of SS ANOVA is to compare curves along multiple reference points via an overlapping coefficient of confidence intervals. They determine the overlap of two curves when the coefficient of confidence intervals is higher than 95%. Indeed, SS ANOVA can help us to tell the differences between two curves at a general level. However, it cannot provide further information for understanding the source resulting in such differences. The differentiation of two *f*0 contours can be accounted for by multiple factors, such as height, direction or even concavity, which are extremely important to the tonal system of some Asian languages (e.g., the vast majority of Chinese dialects). Tracking *f*0 changes by one coefficient would cause a loss of optical resolution. In addition, traditional analyses (i.e., *t*-tests and ANOVAs) also fail to take random effects into consideration. Random effects come from either individual participants or individual items and correspond to the randomly sampled observational units over which we intend to generalize (Mirman, 2014: 21). It is not difficult to understand that in an organized system like language, individuals very often differ from the prototype. For example, in the study of Thai, Erickson (1975) reports that the perturbation effect of aspirated onsets seems to vary within speakers. Failure to include random effects inflates Type I error rates (false positive) (Barr et al., 2013). Traditional analyses, however, cannot remedy this.

The second problem in previous studies is related to the measurement of the laryngeal contrast of obstruents. Generally speaking, there are two aspects. One is about the neglect of phonatory differences in the realization of the laryngeal contrast of obstruents; the other is about the experimental bias in the measurement of phonatory differences.

First, the majority of studies only examine the laryngeal contrast in terms of VOT. On the basis of the time between the release of the consonant and onset of voicing, VOT of obstruents is divided into three types including short lag (voiceless unaspirated), long lag (voiceless aspirated), and lead (voiced) (Lisker & Abramson, 1964). Cross-linguistic data have widely shown that different types of VOT can serve as significant acoustic and perceptual cues to identify laryngeal differences of obstruents (e.g., Klatt, 1975; Lisker, 1975, 1978; Cho & Ladefoged, 1999; Abramson & Whalen, 2017). However, an increasing number of studies have also suggested that a language with more than a two-way laryngeal contrast would not consistently choose the contrast from the three possible VOTs (Ladefoged & Maddieson, 1996; Cho & Ladefoged, 1999). Phonatory contrasts involving the action of the larynx are also argued to play an important role in making voicing distinctions, especially in languages that employ more than a two-way laryngeal contrast (see a recent review in Cho et al., 2019) or tonal contrast (Gordon & Ladefoged, 2001; Kuang, 2013a; Gao, 2015; Kirby & Brunelle, 2017). However, previous studies on C-T interaction rarely pay attention to phonatory differences triggered by obstruents. The neglect of phonatory differences, to a certain extent, results from the bias of language sampling, where most languages bear a two-way contrast usually with less prominent phonatory contrasts.

Nevertheless, there are a handful of studies that try to involve the phonatory parameters in the study of a language with a three-way distinction of obstruents. There is still an experimental bias of the phonatory measurement among these studies. The experimental bias mainly consists of two aspects. First, among multiple acoustic measures of previous studies, spectral tilts are the often-tested set of parameters for differentiating phonation types. Spectral tilts generally refer to the ratio of the amplitude of the fundamental to that of higher frequency harmonics (Gordon & Ladefoged, 2001). The most frequently applied parameter of spectral tilts probably is of H1–H2, which indicates the amplitude difference between the first and second harmonics. H1–H2 is claimed to effectively differentiate the breathy phonation from the modal phonation in a number of languages (e.g., Ladefoged, 1983 for !Xóõ; Huffman, 1987 for Hmong; Ren, 1992 for Shanghainese; Gordon & Ladefoged, 2001 for Zapotec; Blankenship, 2002 for Mazatec; Khan, 2012 for Gujarati). However, in terms of speech production, acoustic analyses are more beneficial from the supplement of physiological experiments. Acoustic analyses alone, to a certain extent, cannot reveal the full picture of the state of the glottis. In order to have a more direct understanding of phonatory differences, articulatory measures should be carried out. Second, the comparison of phonatory differences often relies on values measured from one time point. For example, in the study by Perkins (2013) on Thai, H1–A1 (i.e., the amplitude difference between the first harmonic and the most prominent harmonic around the first formant) measured at 10% of the following vowel is used as an indication of the phonatory state. However, the phonatory aspects of the laryngeal contrast may change during the time course of the vowel (Gordon & Ladefoged, 2001; Blankenship, 2002). Therefore, examining values measured at one single time point is by no means a good way to look into the phonatory contrasts.

In conclusion, both inadequacies in the measurements of *f0* contour and laryngeal contrast prevent us from achieving a complete and accurate understanding of the phonetic properties of C-T interaction.

## 1.4 The current study

As summarized in Section 1.3, although there has been abundant research on C-T interaction, two main biases highly motivated this current study. This section will illustrate how this current research avoids both biases and benefits the study of C-T interaction accordingly.

### 1.4.1 Lili Wu and Shuangfeng Xiang: two Chinese dialects with uncommon C-T interactions

To a large extent, the typological bias is highly related to language sampling. As illustrated in Section 1.3.1, mainly due to a sampling of languages with a two-way laryngeal contrast, no sufficient attention has been given to uncommon C-T interactions that exist in Chinese dialects with a three-way contrast. To inlay these missing parts of the mosaic and to compare findings obtained in these Chinese dialects to patterns observed from languages across the world, the current study is urgently needed. I believe that it is beneficial to provide a study of the C-T interaction of Chinese dialects from the perspective of general linguistics. Chinese dialects, especially those showing typological distinctions from Standard Chinese, should be taken into consideration. Naturally, a subsequent question is which Chinese dialects could help us sharpen the understanding of the full picture of C-T interaction?

As reported in Ye (2011), among Chinese dialects, it is worth noting that more than 73% (31/42) of the three-way-contrast languages belong to the Wu (18) or Xiang (13) dialect groups. Considering this typological representation, in this current study, Lili Wu and Shuangfeng Xiang Chinese are the focus (see Figure 1.2).

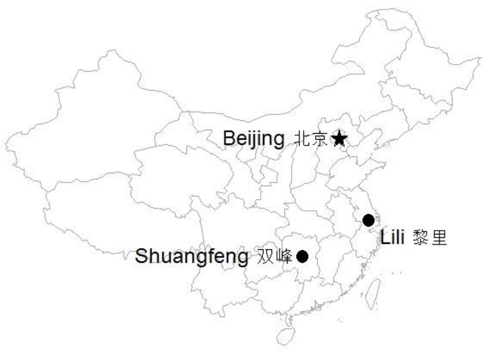



**Figure 1.2** Map showing the location of Beijing, Lili, and Shuangfeng. The star indicates the location of Beijing. The circles indicate the location of Lili and Shuangfeng.

Lili Wu Chinese (黎里方言) is commonly considered to belong to the Suhujia dialect cluster (苏沪嘉小片), which in turn is classified as a member of the Tai Lake subgroup (太湖片) of the Wu dialect group (吴语) (Wurm et al., 1987); while Shuangfeng Xiang Chinese (双峰方言) is commonly considered to belong to the Xiangshuang dialect cluster (湘双小片), which in turn is classified as a member of the Loushao subgroup (娄邵片) of the Xiang dialect group (湘语) (Bao & Chen, 2005). Both dialects bear a three-way laryngeal contrast of obstruents, conventionally labeled as voiceless unaspirated, voiceless aspirated, and voiced, respectively. The three-way laryngeal contrast of obstruents interacts with various lexical tones. However, as mentioned in Section 1.2.2, compared to the common [voiceless/H]-[voiced/L] pattern in the three-way dialects like Shanghainese, both Lili Wu and Shuangfeng Xiang Chinese show quite different patterns, which warrant the current study.

Lili Wu Chinese is well-known for the 'aspiration-induced tonal split' phenomenon. It has been argued that in Lili Wu Chinese, in comparison to syllables with voiceless unaspirated onsets, a clear $f_0$-lowering effect in syllables with voiceless aspirated onsets can be prominently observed. Probably due to the remarkable $f_0$-lowering effect after voiceless aspirated onsets, Lili Wu Chinese has therefore attracted much attention over the last six decades, which led to more than ten descriptive reports on Lili Wu Chinese (see Chapter 2 for more details). However, puzzles still exist. In the current study, two puzzles will be discussed. They are i) Is 'aspiration-induced tonal split' an ongoing change or a completed change? and ii) Is aspiration or breathiness synchronically related to 'aspiration-induced tonal split'?

Shuangfeng Xiang Chinese is famous for the 'initial-associated tonal merger' phenomenon. As claimed in the majority of previous studies, the low-rising tone (T2) is believed to co-occur with all three onsets; while the [voiceless/H]-[voiced/L] pattern still holds across the other lexical tones (references refer to Chapter 4). Among previous studies, there is a long-lasting debate, namely, are there synchronic phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets which might condition or have conditioned the similar lower-rising *f0* contours?

By examining these puzzles in both dialects, this current study aims to sharpen the understanding of the question of how different dialects realize the laryngeal contrasts and their interactions with lexical tones. Consequently, results of this study are expected to enrich the reservoir of knowledge on C-T interaction across the world's languages. In doing so, general linguists, especially phonological typologists, will be able to incorporate sound properties of various Chinese dialects into typological comparisons of C-T interaction and make a leap forward in our understanding of the phenomenon.

### 1.4.2 Multilevel regression models and the electroglottograph (EGG)

It is particularly important to base this study upon not only auditory impressions but also careful phonetic analyses. In order to overcome the methodological bias mentioned earlier, two techniques were applied to the current study. First, multilevel regression models were applied to this study. Second, an experimental device named the electroglottograph (also known as electroglottography) (EGG) was used for obtaining a direct measure of the laryngeal action.

Multilevel regression models are a family of methods specialized for the analysis of longitudinal and repeated-measures data (also called 'time course data'), which involve systematic relationships between observations at different time points (Mirman, 2014: 2, 22). Three types of multilevel regression models were adopted in this study, namely growth curve analysis (GCA), generalized linear mixed models (GLMMs), and linear mixed models (LMMs). A base model of multilevel regression has to contain two structures, namely the Level 1 structure and the random structure. The Level 1 structure is defined by $Y_{ij} = \beta_{0i} + \beta_{1i} \cdot \mathrm{Time}_j + \varepsilon_{ij}$, which describes the individual observations $Y_{ij}$ for individual *i* at Time *j* (Mirman, 2014: 22). The component $\varepsilon_{ij}$ is the residual error, which indicates the amount that actual observation $Y_{ij}$ differs from the predicted value. The most common random variable is either individual participants or individual items. The choice of the Level 1 structure determines the overall shape that describes the data (Mirman, 2014: 37).

Both GLMMs and LMMs contain a linear shape indicated by one-order polynomial term (i.e., *Time*). GLMMs are an extension of LMMs and are suitable for the analysis of non-normal data with different distributions, such as binary responses (Tuerlinckx et al., 2006). A model of GCA, however, can include multiple polynomial terms (i.e., more than one-order) for capturing non-linear time-varying changes such as *f0*. The advantage of GCA is that overall successive data can be taken into consideration (see the usage of this method for tonal contours in other Chinese dialects such as Li & Chen, 2016 for tonal realization in Tianjin Mandarin; and Zhang & Meng, 2016 for Shanghainese tones; see also the comparison between GCA and other statistical methods on tone analysis in Chen et al., 2017). The obvious advantage of GCA in the analysis of the *f0* contour is that it fits the discontinuous data into a successive curve and compares different curves via several core terms. In this way, the trade-off dilemma between statistical power and temporal resolution, to a general extent, can be overcome. In addition, as a member of the multilevel regression family, GCA can also efficiently quantify random effects. It is possible to take individual differences of *f0* contours into consideration as much of the structure of the data as possible.

In order to obtain a more direct measure of the laryngeal action, this study also employed EGG to capture the change in the vocal fold contact area. In recent decades, EGG has achieved some popularity for registering glottal activities during vocal fold vibration (see Baken & Orlikoff, 2000 for a comprehensive review). It has been extensively used in clinical treatment, such as with voice patients (e.g., Frokjaer-Jensen & Prytz, 1976) and for the hearing disorder (e.g., Wirz & Anthony, 1979). EGG relies on an electrical impedance technique for the direct examination of vocal fold contact during vibration through the measurement of electrical impedance changes (Childers & Krishnamurthy, 1985; Childers et al., 1986; Childers et al., 1990; Titze, 1990). There has been a productive line of research on the relationship between the EGG waveform and vocal fold vibration (i.e., physiological movement) (e.g., Fant et al., 1966; Rothenberg, 1982; Childers et al., 1990; Titze, 199; Baken, 1992; Baken & Orlikoff, 2000). In terms of linguistic research, EGG has been widely employed to measure articulatory properties – mainly phonation types. Compared to the acoustic measure of phonatory parameters such as spectral tilts (e.g., H1–H2), the measure of EGG provides more direct action of laryngeal function (i.e., the motion and contact between the vocal folds) without interfering with phonation. The change of the relative degree of contact between the two vocal folds during the glottal cycle is directly reflected by the change of the degree of electrical conductance collected by the EGG device. In addition, EGG is noninvasive, which is not achievable for any other practical means. More importantly, EGG can be performed at a relatively low cost and is portable equipment, making it useful for fieldwork. Due to these advantages (but see the problems and pitfalls in Colton & Conture, 1990; Childers et al., 1990), EGG has been widely employed in articulatory phonetics for assessing phonatory behavior across the world's languages (e.g., Vu-Ngoc et al., 2005 for Vietnamese; Mazaudon & Michaud, 2008 for Tamang; DiCanio, 2009 for Takhain Thong Chong; Esposito, 2012 for White Hmong; Khan, 2012 for Gujarati; Kuang, 2013a for Yi and Black Miao; Abramson et al., 2015 for Mon; Gao, 2015, Tian & Kuang, 2019 for Shanghainese).

Moreover, measurements over multiple time points were taken over the whole vowel. To my knowledge, different decisions have been made with regard to the number of measurements: 3-point (e.g., Garellek & Keating, 2011), 5-point (e.g., Dutta, 2007; Berkson, 2013; Gao, 2015), 9-point (e.g., Esposito, 2012; Esposito & Khan, 2012; Garellek, 2012), and 12-point (e.g., DiCanio, 2009; Kuang, 2013a) proposals. It is worth noting that there is no essential difference in methodology among these choices but that they just have different time windows. There is no obvious best proposal since it mainly depends on certain research purposes. For example, although Esposito (2012), Esposito and Khan (2012), and Garellek (2012) all took the same number of data points, they actually conducted analyses at different number of points: three points (1, 5, 9) in Esposito (2012), five points (1 to 5) in Esposito and Khan (2012), and all nine points in Garellek (2012). The goal of this study is not to track the subtle continuous changes in phonation over the time course of the vowel, but rather to examine whether certain phonatory state observed at the onset of the vowel can be still observed towards the later portions of the vowel. Following the method of Garellek and Keating (2011), Esposito (2012), and Tian and Kuang (2019), I took measurements over three time points (i.e., one-third, middle, and two-thirds).

By applying both techniques to this study, we are able to observe a more reliable and accurate picture of the uncommon C-T interactions in both Lili Wu and Shuangfeng Xiang Chinese. More importantly, with the improved methods, this study aims to provide an empirical base for further research on C-T interaction.

## 1.5 Outline

This dissertation covers two case studies of C-T interaction in Chinese dialects: Lili Wu Chinese and Shuangfeng Xiang Chinese. The C-T interaction in each dialect, generally speaking, is a property of the sound system of its own. I do not want to separate this property from the entire system. This illustrates how I think about C-T interaction in typology. To understand C-T interaction in both dialects, one has to know what kind of sound properties the dialect contains, how the phonological system functions, what its variation is, and what the role of the C-T interaction plays in this dialect. For this reason, I try to give a comprehensive phonetic description of the sound system of each dialect before presenting the case study. In doing so, certain background information, terms, or phenomena, which require an elaborate motivation and explanation, can be introduced to readers as early as possible. I believe that both descriptions are crucial for a deeper understanding of C-T interaction in Chinese dialects.

The current chapter (Chapter 1) introduced the research topic of this dissertation: C-T interaction in Chinese dialects. Particularly, this study followed the property-based approach and aimed to overcome the two biases (i.e., typological and methodological) observed in the existing literature. To this end, I focused on two Chinese dialects (i.e., Lili Wu and Shuangfeng Xiang) which have been reported to feature uncommon C-T interactions. Furthermore, two techniques (i.e., multilevel regression models and EGG) were applied to the data collection and analyses of both dialects. In the rest of this dissertation, Chapter 2 presents a comprehensive description of the sound system of Lili Wu Chinese. In Chapter 3, the C-T interaction of Lili Wu Chinese is investigated. A description of the sound system of Shuangfeng Xiang Chinese is provided in Chapter 4 and its C-T interaction is examined in Chapter 5. It is also worth noting that Chapter 2–5 are conceived as individual journal articles, and therefore may have some overlapping background information and citations. Last but not least, Chapter 6 concludes on findings of C-T interaction in both dialects. Furthermore, I discuss implications of the results concerning i) the typological value of C-T interaction found in Lili Wu and Shuangfeng Xiang Chinese and ii) the possible relationship between the Wu and Xiang dialect groups.

# Chapter 2 The sound system of Lili Wu Chinese[7]

## 2.1 Introduction

Lili Wu Chinese (黎里方言) is a Wu dialect (吴语, ISO 639-3; code: wuu) spoken by approximately 38,000 people who reside in the town of Lili (黎里镇), one of the ten major towns in the Wujiang district (吴江区). The Wujiang district belongs to the prefectural-level municipality of Suzhou city (苏州市) in Jiangsu province (江苏省), the People's Republic of China. It is located at the juncture area of the city of Shanghai (上海市), the city of Suzhou, and the city of Jiaxing (嘉兴市), as shown in Figure 2.1.

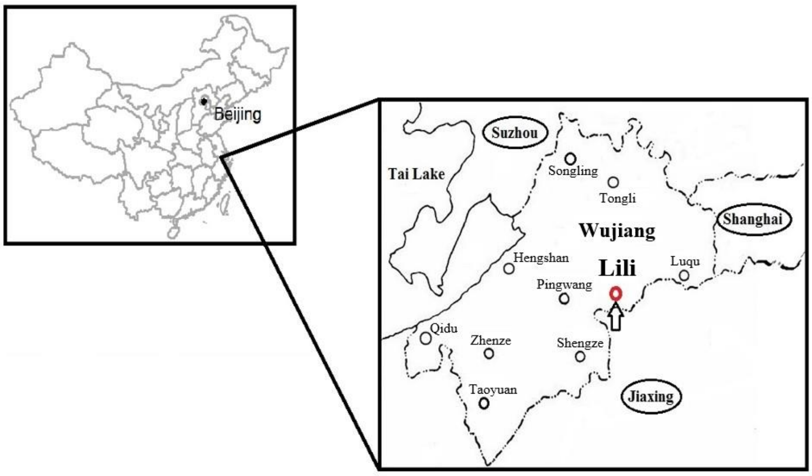



**Figure 2.1** Map of the Wujiang dialects (modified based on the map in Ye, 1983).

Lili Wu Chinese is commonly considered to belong to the Suhujia dialect cluster (苏沪嘉小片), which in turn is classified as a member of the Tai Lake subgroup (太湖片) of the Northern Wu dialect group (北部

[7] A version of this chapter has been accepted for publication by *Journal of the International Phonetic Association*. Shi, M., & Chen, Y. (forthcoming). Lili Wu Chinese.

吴语), a Sinitic branch within the Sino-Tibetan family (Wurm et al., 1987). The dialect is famous for the so-called 'aspiration-induced tonal split' phenomenon (ATS), which refers to the lowering of *f0* contours after voiceless aspirated obstruents in certain tonal contexts. Lili Wu Chinese has therefore attracted much attention over the last six decades, which led to a handful of descriptions not only on the dialect but also on its closely-related dialects in the Wujiang area, which appear to have similar aspiration-induced tonal splits (see Section 2.2 for further details on lexical tones and ATS). Perhaps because of this salient tonal-split feature in the dialect, much less attention has been paid to the segmental properties of Lili Wu Chinese in the existing literature.

The description is mainly accompanied by recordings of a sixty-eight-year-old male native speaker, who was born in 1948 and raised in Lili town. All acoustic data presented in this description were elicited from this consultant. The consultant spent most of his life living in Lili and speaking Lili Wu Chinese, except for the three years attending a college in a nearby city. According to his self-report, he can speak accented Standard Chinese and limited Shanghainese when the situation requires him to do so, but he speaks only Lili Wu Chinese at home.

## 2.2 Lexical tones and aspiration-induced tonal split (ATS)

### 2.2.1 Lexical tones

There are eight lexical tones in Lili Wu Chinese. Plotted in Figure 2.2 are the *f0* contours of the example morphemes listed in Table 2.1, labeled as T1 to T8, respectively. Generally speaking, lexical tones marked as odd numbers start within a higher *f0* range (above 160 Hz, high register hereafter), while those marked as even numbers start within a lower range (under 160 Hz, low register hereafter). T1 (black solid) has a level *f0* contour within a high register (high–level) while T2 (dark gray solid) is a low-register-rising tone (low–rising). T3 (black round dot) starts within the high register and falls (high–falling). T4 (dark gray round dot) is a low-register-level tone (low–level). T5 (black square dot) has a convex contour which starts at the high register, falls and ends with a slight rise (high–dipping). T6 (dark gray square dot) is realized with a similar *fo* contour to that of T5 but starts at the low register (low–dipping). Both T7 (black dash-dotted) and T8 (dark gray dash-dotted) are associated with syllables that have a much shorter duration than the other tone-bearing syllables. T7 starts within the high register and despite the slight falling contour, sounds like a high-register-level tone (short–high–level). T8 is a low-register-level tone (short–low–level).

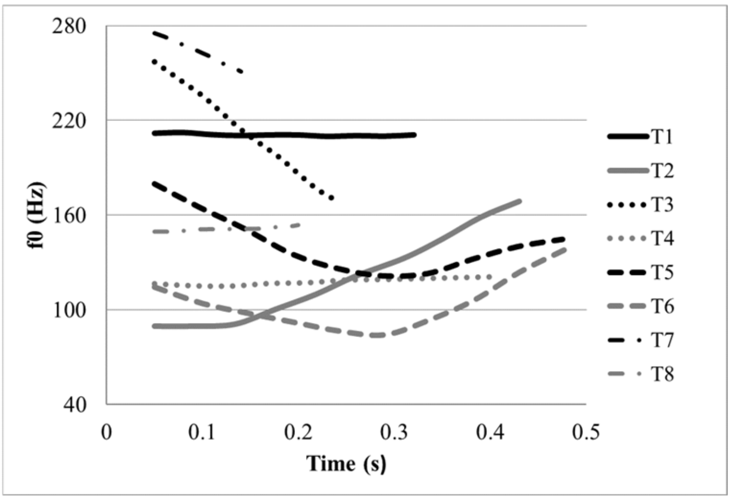



**Figure 2.2** *fo* contours of the lexical tones in Lili Wu Chinese.

**Table 2.1** Examples of the lexical tones in Lili Wu Chinese.

| Lexical tone | Tonal contour | Example | Orthography | Gloss |
|---|---|---|---|---|
| T1 | high–level | /tʊŋ¹/ | 东 | 'east' |
| T2 | low–rising | /dʊŋ²/ | 铜 | 'copper' |
| T3 | high–falling | /tʊŋ³/ | 懂 | 'to understand' |
| T4 | low–level | /dʊŋ⁴/ | 动 | 'action' |
| T5 | high–dipping | /tʊŋ⁵/ | 冻 | 'to freeze' |
| T6 | low–dipping | /dʊŋ⁶/ | 洞 | 'cave' |
| T7 | short–high–level | /tʊʔ⁷/ | 督 | 'supervision' |
| T8 | short–low–level | /dʊʔ⁸/ | 读 | 'to read' |

These eight lexical tones exhibit interesting co-occurrence patterns with both the onset and coda. Lili Wu Chinese features the three-way laryngeal contrast in obstruents, known as voiceless unaspirated, voiceless aspirated and voiced, respectively (see Section 2.3 for more details). Syllables with voiceless unaspirated onsets only allow high-register tones (T1, T3, T5, and T7); while voiced onsets co-occur with low-register tones (T2, T4, T6, and T8). T1 to T6 only co-occur with open syllables or syllables with a nasal coda (舒声) and are therefore also known as smooth/non-checked tones (Middle Chinese [MC] *Ping–Shang–Qu*; 平–上–去; level–rising–departing), while the T7 and T8 only co-occur with closed syllables with a glottal coda /ʔ/ (促声) and are known as abrupt/checked tones (MC *Ru*; 入; entering). MC is a sound system reconstructed based mainly on written records such as rhyme dictionaries (see a comprehensive introduction in Norman, 1988). In the development of the tonal system from MC to the modern Wu dialects, the four-way tonal contour contrast of MC (i.e., *Ping*, *Shang*, *Qu*, and *Ru*) split into a new dual-register, eight-tone system, conditioned by onset consonants, which is evident in modern Wu varieties (Pulleyblank, 1978; Ting, 1984; Norman, 1988).

In the vast majority of Northern Wu dialects such as Shanghainese (Chen & Gussenhoven, 2015), both voiceless unaspirated and aspirated onsets condition high-register tones, leaving voiced onsets to co-occur with low-register tones. What makes Lili Wu Chinese interesting is the effect of obstruent aspiration on lexical tonal realization, as exemplified by /tʰʊŋ¹/ 'unblocked', /tʰʊŋ⁴/ 'to unify', /tʰʊŋ⁶/ 'ache', and /tʰʊʔ⁸/ 'baldy'. Their *f0* contours are plotted in Figure 2.3 (labeled as T1-A, T3-A, T5-A, and T7-A where A indicates voiceless aspirated onsets), in comparison to the *f0* contours of the presumably same lexical tones realized after voiceless unaspirated onsets (indicated with U). Except for T1 (i.e., T1-U vs. T1-A), we see a clear *f0*-lowering effect in syllables with voiceless aspirated onsets. This lowering effect, as if a split of the same tone into two as a function of voiceless unaspirated vs. aspirated onsets, is known as ATS.

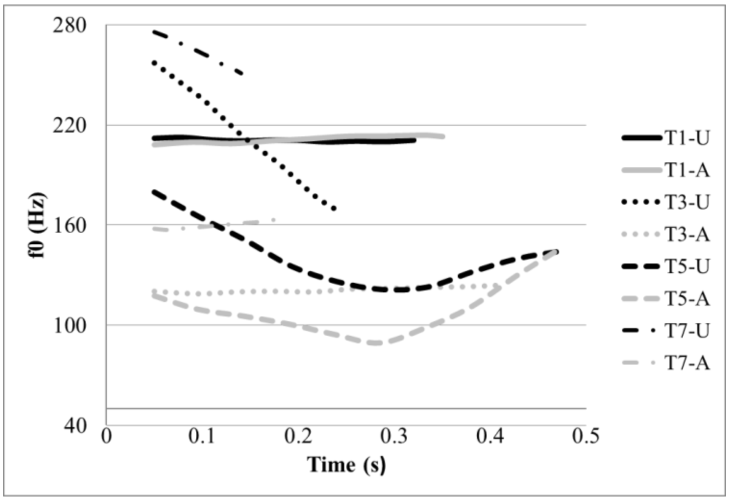



**Figure 2.3** *f0* contours of the lexical tones of the words with voiceless unaspirated onsets (U, black) and those with voiceless aspirated (A, light gray) onsets.

## 2.2.2 New analysis of ATS

Perhaps due to ATS, the tonal inventory in Lili Wu Chinese has been a point of debate in recent decades. To my knowledge, there are at least eleven descriptive works focusing on ATS in Lili Wu Chinese (Chao, 1928;

Ye, 1983; Zhang & Liu, 1983; Shi, 1992; Qian, 1992; Shen, 1994; Wang, 2008, 2010; Xu, 2009; Hirayama, 2010; Yanhong Xu, 2013).

Chao (1928) is the first to note this aspiration induced tonal-split phenomenon. Subsequent representative studies include Shi (1992), the first which reported acoustic data from two native speakers of Lili Wu Chinese; and Wang (2010), which is the first attempt at a comprehensive description of the phonological system of Lili Wu Chinese. However, researchers differ greatly in their treatment/interpretation of ATS. The main debate lies in the question of whether the *f0* contours of lexical tones after aspirated onsets have been merged with those after voiced onsets or emerged as distinct tonal categories independent of the existing eight tonal categories. As can be seen in Table 2.2, the only agreement in the existing literature on ATS is that it doesn't occur (marked as '–' in Table 2.2) in the MC *Ping* category. With respect to the other three MC tonal categories, researchers differ greatly in their treatment/interpretation of the tonal-split phenomenon. Chao (1928) reports an unmerged distribution (marked as '+' in Table 2.2) in the MC *Qu* and *Ru* categories. Represented by Ye (1983), a lot of researchers claim a three-way distinction of lexical tones in the MC *Shang*, *Qu*, and *Ru* categories. However, Shen (1994) argues for a merging of lexical tones following aspirated and voiced onsets (marked as '++' in Table 2.2) in the three MC categories. Wang (2008, 2010) reports a merging in *Shang* and *Qu*, and Yanhong Xu (2013) maintains a merging in *Qu* only.

How should we understand the discrepancies existing among these studies? Given that there is an over-80-year interval between the earliest and the latest studies, it is reasonable to assume that different proposals indeed reflect the actual changes of the lexical tonal system of Lili Wu Chinese. For example, if comparing the study by Chao (1928) to the study by Ye (1983), one can argue that it is not until the 1980s that ATS happened in the *Shang* category. However, it is worth pointing out that not all discrepancies can be explained by such diachronic interpretations. For instance, as we can see from the studies of Qian (1992) and Shen

(1994), both were finished in the early 1990s by recruiting the same consultants. However, Shen (1994) proposes an opposite treatment compared to Qian (1992). It is important to note that these previous studies typically explored ATS based on impressionistic descriptions (e.g., Chao, 1928), or with data from a very limited number of speakers (e.g., Shi, 1992 for one male and one female speaker; Shen, 1994 for two young speakers). The phonetics and phonology of the tonal system of Lili are still seriously under-studied, which motivated a large-scale experimental study with statistical analyses of the *f0* contours to know the general group-level patterns of *f0* contours of the lexical tones in Lili Wu Chinese.

**Table 2.2** Treatments of the lexical tones after voiceless aspirated onsets in Lili Wu Chinese. –: no tonal split; +: a three-way distinction; ++: a merger has happened.

| *Ping* | *Shang* | *Qu* | *Ru* | Study |
|---|---|---|---|---|
| – | – | + | + | Chao (1928) |
| – | + | + | + | Ye (1983); Zhang & Liu (1983); Shi (1992); Qian (1992); Xu (2009); Hirayama (2010) |
| – | ++ | ++ | ++ | Shen (1994)[8] |
| – | ++ | ++ | + | Wang (2008, 2010) |
| – | + | ++ | + | Yanhong Xu (2013) |

The results of a large-scale experimental study across three generations (see Chapter 3) show although *f0* contours after aspirated onsets show a trend of slightly higher *f0*, they can pattern more like those after voiced onsets, resulting in the merger of the *f0* contours of T3-A and T4 (MC *Shang*), T5-A and T6 (MC *Qu*), T7-A and T8 (MC *Ru*), respectively. However, there are comparable *f0* contours after voiceless aspirated (T1-A) and voiceless unaspirated (T1-U) onsets (MC *Ping*), both of which are

[8] According to the acoustic analyses in Shen (1994), *f0* contours after voiceless aspirated onsets in the MC *Ru* category should be transcribed as /33/. However, he finally argues that the tone after the aspirated onset has merged to that after its voiced counterpart and transcribes both as /23/ for the sake of systematicness.

realized within the high *f*0 register. It suggests that the lexical tonal system of Lili Wu Chinese includes two level tones (high–level T1 and low–level T4), one falling tone (high–falling T3), one rising (low–rising T2) and two dipping tones (high–dipping T5 and low–dipping T6). For short syllables with a glottal coda, two level tones are identified (short–high–level T7 and short–low–level T8). The numerical representations of the eight lexical tones and their co-occurrence patterns with onsets are provided in Table 2.3. Here, 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. The results confirm the treatment in Shen (1994).

**Table 2.3** Numerical representations of the lexical tones in Lili Wu Chinese. A single number refers to cases where the tone-carrying syllables have a short duration and only co-occur with the glottal coda /ʔ/.

| | | | | |
|---|---|---|---|---|
| Voiceless unaspirated (U) | T1 high–level 44 | T3 high–falling 53 | T5 high–dipping 423 | T7 short–high–level 5 |
| Voiced (D) | T2 low–rising 13 | T4 low–level 22 | T6 low–dipping 213 | T8 short–low–level 3 |
| Voiceless aspirated (A) | T1 high–level 44 | T4 low–level 22 | T6 low–dipping 213 | T8 short–low–level 3 |

T1 can co-occur with both voiceless onsets (i.e., unaspirated and aspirated). T3, T5, and T7, on the other hand, can only co-occur with voiceless unaspirated onsets. The three low-register tones (T4, T6, and T8) are licensed by both voiceless aspirated and voiced onsets, while T2 is only allowed after voiced onsets. It is important to note that the co-occurrence pattern (i.e., voiceless onsets co-occurring with high-register tones, while voiced onsets with low-register ones), which is commonly observed in most Northern Wu dialects, falls apart in Lili Wu Chinese where voiceless aspirated onsets can co-occur with low-register tones.

In addition, it is worth noting that in Lili Wu Chinese, sonorants (i.e., nasals and liquids) mainly co-occur with low-register tones and share the same tonal pattern with voiced plosives. A set of words initialed with nasals can also co-occur with high-register tones,[9] such as /mu³/ [məʊ³] 'bound morpheme for the literary address of mother'. With respect to fricatives, since there is only a two-way laryngeal distinction (i.e., voiceless vs. voiced), voiceless fricatives co-occur with high-register tones while their voiced counterparts with low-register ones.

## 2.3 Consonants

Lili Wu Chinese has 28 consonants. Corresponding keywords/bound morphemes are provided below the consonant chart.

[9] In Northern Wu Chinese, there is a group of shared words initialed with nasals (mainly including 妈/母/美/蛮/囡 'bound morpheme for mother/bound morpheme for the literary address of mother/beautiful/very/little darling') that can co-occur with high-register tones. Such a co-occurrence is argued to be relevant to the affective function of high tones (Zhu, 2004a).

| | Bilabial | | | Labio-dental | | Alveolar | | | Alveolo-palatal | | | Palatal | Velar | | | Glottal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plosive | pʰ | p | b | | | tʰ | t | d | | | | | kʰ | k | g | ʔ |
| Affricate | | | | | | tsʰ | ts | dz | tɕʰ | tɕ | dʑ | | | | | |
| Nasal | | | m | | | | | n | | | | | | | ŋ | |
| Fricative | | | | f | v | s | | z | ɕ | | | | | | | h |
| Approximant | | | w | | | | | | | | | j | | | | |
| Lateral approximant | | | | | | | | l | | | | | | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pʰ | pʰɛ$^{1}$ | 攀 | 'to climb' | tʰ | tʰɛ$^{1}$ | 瘫 | 'paralysis' | | | | | kʰ | kʰɛ$^{1}$ | 开 | 'to open' |
| p | pɛ$^{1}$ | 班 | 'class' | t | tɛ$^{1}$ | 单 | 'single' | | | | | k | kɛ$^{1}$ | 该 | 'ought to' |
| b | bɛ$^{2}$ | 赔 | 'to compensate' | d | dɛ$^{2}$ | 台 | 'platform' | | | | | g | gɛ$^{6}$ | 隑 | 'to lean against' |
| | | | | | | | | tsʰ | tsʰɛ$^{1}$ | 猜 | 'to guess' | tɕʰ | tɕʰi̟$^{1}$ | 欺 | 'to deceive' |
| | | | | | | | | ts | tsɛ$^{1}$ | 灾 | 'disaster' | tɕ | tɕi̟$^{1}$ | 鸡 | 'chicken' |
| | | | | | | | | dz | dzɛ$^{2}$ | 随 | 'at random' | dʑ | dʑi̟$^{2}$ | 奇 | 'strange' |
| m | mɛ$^{6}$ | 妹 | 'younger sister' | n | nɛ$^{2}$ | 难 | 'difficult' | | | | | ŋ | ŋɛ$^{2}$ | 颜 | 'color' |
| f | fɛ$^{1}$ | 翻 | 'to turn over' | s | sɛ$^{1}$ | 三 | 'three' | ɕ | ɕi̟$^{1}$ | 稀 | 'rare' | h | hɛ$^{1}$ | 虚 | 'turgescence' |
| v | vɛ$^{2}$ | 烦 | 'to bother' | z | zɛ$^{2}$ | 馋 | 'greedy' | | | | | ʔ | kʌʔ$^{7}$ | 割 | 'to cut' |
| w | wɛ$^{2}$ | 还 | 'to return' | l | lɛ$^{2}$ | 蓝 | 'blue' | | | | | j | jə̯o$^{1}$ | 优 | 'superior' |

Lili Wu Chinese features the three-way laryngeal contrast in obstruents, known as voiceless unaspirated, voiceless aspirated and voiced, respectively. This three-way contrast is a prominent feature of the Northern Wu dialects (Chao, 1967). The three-way laryngeal contrast, however, has different phonetic manifestations in initial vs. medial positions within a word. Generally speaking, in initial position, these obstruents feature a two-way distinction (i.e., short lag vs. long lag) in terms of voice onset time (VOT). As exemplified by the triplet (a) /tɑ$^{1}$/ 'knife', (b) /t$^{h}$ɑ$^{1}$/ 'billow', and (c) /dɑ$^{2}$/ 'peach' in Figure 2.4, both the voiceless unaspirated (/t/ as in /tɑ$^{1}$/, 2.4a) and voiced onsets (/d/ as in /dɑ$^{2}$/, 2.4c) have a short-lag VOT (15 ms and 19 ms), in contrast to the aspirated onset (/t$^{h}$/ as in /t$^{h}$ɑ$^{1}$/, 2.4b) which is realized with a long-lag VOT (89 ms).

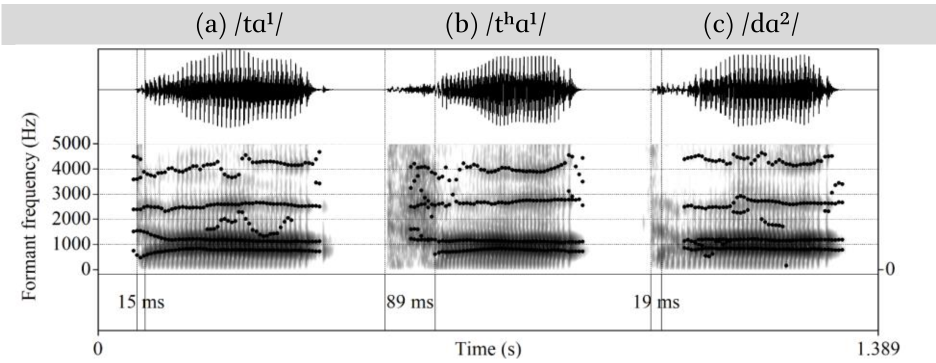



**Figure 2.4** Waveforms and spectrograms of (a) /tɑ$^{1}$/ 'knife', (b) /t$^{h}$ɑ$^{1}$/ 'billow', and (c) /dɑ$^{2}$/ 'peach'. Intervals indicate VOT values.

Table 2.4 shows the mean VOT values and their standard deviation of the three-way contrast of plosives in three places of articulation. The measurements were obtained from 90 monosyllabic morphemes consisting of 30 sets of triplets. As shown in Table 2.4, irrespective of the place of articulation, the voiceless aspirated plosives have longer VOT values than the other two counterparts. The voiceless unaspirated and voiced plosives have similar short VOT values.

**Table 2.4** VOT of unaspirated vs. aspirated vs. voiced plosives in different places of articulation in Lili Wu Chinese, based on 90 monosyllabic morphemes with plosive onsets.

| | Bilabial (10 triplets) | | Alveolar (10 triplets) | | Velar (10 triplets) | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Voiceless unaspirated | 10 ms | 6 ms | 12 ms | 6 ms | 16 ms | 7 ms |
| Voiceless aspirated | 94 ms | 16 ms | 89 ms | 18 ms | 102 ms | 19 ms |
| Voiced | 14 ms | 5 ms | 17 ms | 9 ms | 23 ms | 7 ms |

In medial position, as shown in Figure 2.5, the voiced onset /d/ (2.5c) is realized with noticeable voicing throughout the closure (−61 ms), leading to a three-way laryngeal distinction.

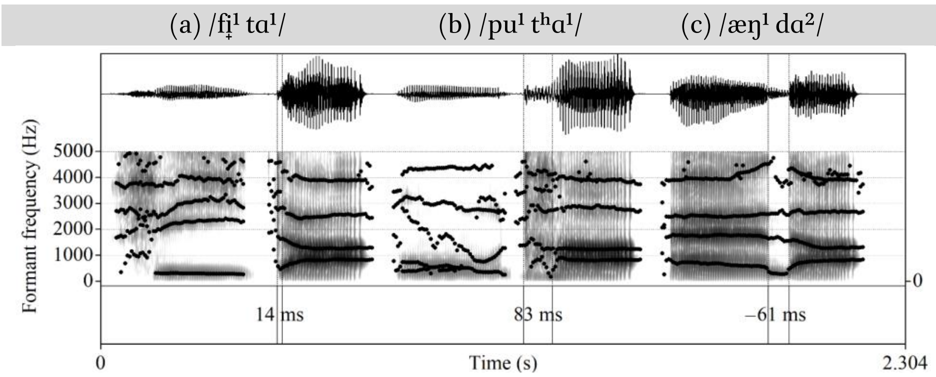



**Figure 2.5** Waveforms and spectrograms of (a) /fi̟$^1$ tɑ$^1$/ 'flying knife', (b) /pu$^1$ t$^h$ɑ$^1$/ 'great waves', and (c) /æŋ$^1$ dɑ$^2$/ 'cherry'. Intervals indicate VOT values.

In addition, in initial position, these obstruents also vary in their phonation from clearly modal (voiceless unaspirated), aspirated (voiceless aspirated), to breathy (voiced) (See Chapter 3 for more information). In Shanghainese, a Northern Wu dialect closely related to Lili Wu Chinese, there are other phonetic properties to signal the three-way laryngeal contrast in both initial and medial positions (e.g., Shen et al., 1987 on closure duration; Ren, 1992 on transillumination data; Gao, 2015 on motion-capture-system data; see also a review in Chen, 2011). Impressionistically speaking, Lili Wu Chinese behaves similarly to Shanghainese. Needless to say, more research is needed to examine if these properties also function in Lili Wu Chinese.

The alveolar consonants /tʰ t d/ tend to be denti-alveolar and /s z tsʰ ʦ ʣ/ apical alveolar, with /s z/ having a contact area slightly further forward than /tsʰ ʦ ʣ/. The alveolo-palatals /ʨʰ ʨ ʥ ɕ/ are produced via contacting the region between the alveolar ridge and the post-alveolar palatal, with the fricative /ɕ/ having a slightly more frontal contact area.

Fricatives have the voiceless vs. voiced two-way laryngeal contrast. Similar to the plosives and affricates, in initial position, their phonatory states vary from clearly modal in the voiceless ones to slightly breathy in the voiced ones. In medial position, the voiced category is realized with vigorous voicing, leading to a two-way contrast in terms of VOT. It is worth noting that the voicing contrast in fricatives is also signaled via their durational differences, similar to what have been reported for voicing contrast in English fricatives (e.g., Cole & Cooper, 1975), as shown in the following pairs: /f/ (/$fu^1$/ 'husband') vs. /v/ (/$vu^2$/ 'to support somebody with one's hand'); /s/ (/$s\varepsilon^1$/ 'three') vs. /z/ (/$z\varepsilon^2$/ 'greedy'). Figure 2.6 illustrates the acoustic realization of /f/ in /$fu^1$/ (2.6a) and /v/ in /$vu^2$/ (2.6b). Although neither is realized with regular vocal pulses (i.e., phonetically voiceless), the fricative duration of /f/ (131 ms; 29% of the total duration) is almost 2.4 times longer than that of /v/ (56 ms; 12% of the total duration).

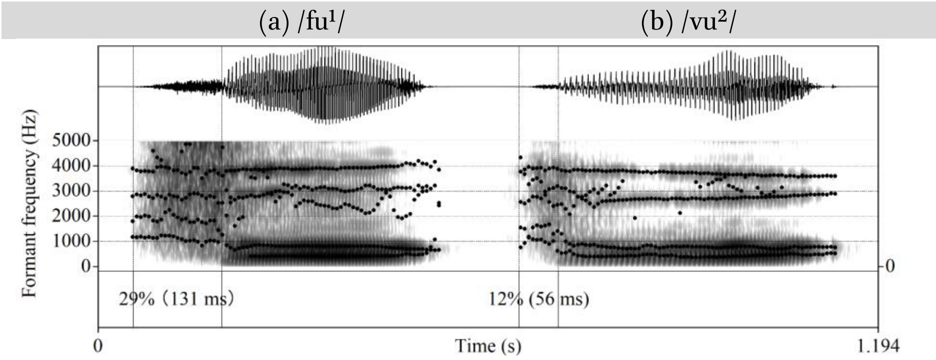



**Figure 2.6** Waveforms and spectrograms of (a) /fu[1]/ 'husband' and (b) /vu[2]/ 'to support somebody with one's hand'. Within a syllable, the percentage of the frication duration with absolute values (ms) between parentheses is indicated.

The durational differences of the voicing contrast in fricatives have not been well investigated in previous studies of Northern Wu Chinese. To further confirm these observations, ten minimal pairs for each minimal set of voicing contrast were elicited. All stimuli were lexemes of relatively high familiarity, as confirmed by the consultant. Both the absolute duration of the frication and the percentage of the frication duration over the whole syllable duration were calculated. The fricative duration was measured from the onset of clear frication noise to the first periodic cycle of the vowel. The results in Table 2.5 show that the percentage of the frication duration of voiceless onsets is significantly greater than that of their voiced counterparts, confirmed by the results of the independent samples *t*-tests (one-tailed) for each pair.

**Table 2.5** Average percentage of the frication duration and the independent samples *t*-test results (one-tailed) for each pair of voiceless vs. voiced. The Bonferroni method was adopted for *p*-value adjustments. Parentheses indicate absolute values of the average duration and the standard deviation (ms).

| Voiceless vs. voiced | Frication duration | *t*-test |
|---|---|---|
| f vs. v | 27% ($M$ = 124 ms, $SD$ = 7 ms) vs. 12% ($M$ = 59 ms, $SD$ = 4 ms) | $t(18) = 10.79$, $p < .001$ |
| s vs. z | 36% ($M$ = 149 ms, $SD$ = 8 ms) vs. 28% ($M$ = 117 ms, $SD$ = 6 ms) | $t(18) = 2.74$, $p < .01$ |

Lili Wu Chinese maintains the distinction of affricate /ʣ/ and fricative /z/, as evidenced in the minimal pair /dzε$^{2}$/ 'at random' and /zε$^{2}$/ 'talent'. In both phrase initial and medial positions, there is a clear frication duration difference between the fricative /z/ and the affricate /dz/, as shown in Figure 2.7 (initial position: /dz/ 21% vs. /z/ 30%) and Figure 2.8 (medial position: /dz/ 10% vs. /z/ 23%).

(a) /dzε$^{2}$/ (b) /zε$^{2}$/

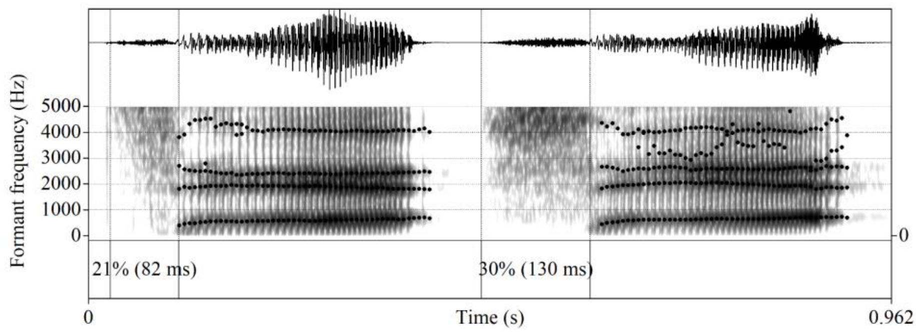



**Figure 2.7** Waveforms and spectrograms of (a) /dzε$^{2}$/ 'at random' and (b) /zε$^{2}$/ 'talent'. Within a syllable, the percentage of the frication duration with absolute values between parentheses is indicated.

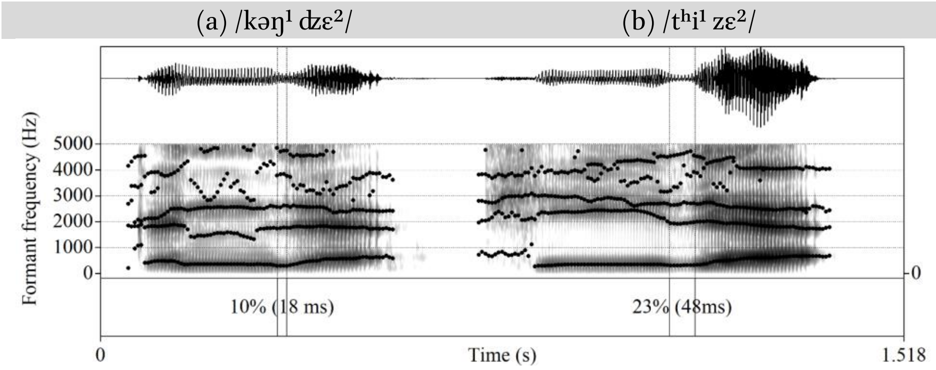



**Figure 2.8** Waveforms and spectrograms of (a) /kəŋ$^{1}$ ʥɛ$^{2}$/ 'to follow' and (b) /tʰi$^{1}$ zɛ$^{2}$/ 'genius'. Within a phrase, the percentage of the frication duration with absolute values (ms) between parentheses is indicated.

In addition, /dz/ and /z/ can sometimes be distinguished according to the degree of formality with which the lexical item is produced. Chao (1928) first observes the variation between /dz/ and /z/ in some morphemes. Very recently, on the basis of recordings of Shanghainese in 1853 by a British Protestant missionary – Joseph Edkins (1823–1905), Chen (2015) argues that the affricate phoneme /dz/ was borrowed from Hangzhou Chinese (i.e., the Hangzhou dialect spoken in the city center of Hangzhou (杭州市区)), and tends to be preserved in literary lexical items. In Lili Wu Chinese, for example, the morpheme /dzɛ$^{4}$/ 'crime' is pronounced as /dzɛ$^{4}$/ in the literary lexical item /dzɛ$^{4}$ ɪŋ$^{2}$/ 'crime', but as /zɛ$^{4}$/ in /zɛ$^{4}$ ku$^{5}$/ 'pitiful' – a more colloquial lexical item. However, /dz/ and /z/ are sometimes in free variation for the same lexical item, as exemplified in 'groceries' (/dzaʔ$^{8}$ hu$^{5}$/ vs. /zaʔ$^{8}$ hu$^{5}$/). This finding may imply that in Lili Wu Chinese, /dz/ and /z/ are undergoing a merger at the lexical level.

Last but not least, the glottal plosive /ʔ/ only appears in coda position as a phoneme and co-occurs with short syllables as in /paʔ$^{7}$/ 'hundred'. Phonetically speaking, the [ʔ] segment can also be observed at the beginning of onsetless syllables with the high-register tones (i.e., T1, T3, T5, and T7) (see Section 2.8 for further details on onsetless syllables).

## 2.4 Sonorants

/n l/ are typical laminal alveolar. The alveolar nasal /n/ is palatalized before high front segments (i.e., /i i̟ y j/), as in /ni$^{2}$/ [ɲi$^{2}$] 'year' and /njɛ$^{6}$/ [ɲjɛ$^{6}$] 'to read'. Labial and velar nasals can form syllable nuclei as in /ŋ̍$^{4}$/ 'five' and /m̩$^{4}$/ 'parcel of land'. These two syllabic nasals can be found in many Southern Chinese dialects (i.e., Wu, Min, Hakka, Xiang, and Yue) but are relatively rare in dialects belonging to the Mandarin family (Shen, 2006). In addition, /ŋ/ occurs as a nasal coda as well, but its acoustic realization varies according to the preceding vowel. After a front vowel, the nasal coda acquires the anterior feature and sounds like [n] (as in /zɪŋ$^{2}$/ [zɪn$^{2}$] 'to look for' and /tɕʏŋ$^{1}$/ [tɕʏn$^{1}$] 'army'), as in contrast to a non-front vowel (as in /dzəŋ$^{1}$/ 'deity' and /dʊŋ$^{2}$/ 'copper'). Following the treatment of Chen and Gussenhoven (2015) for Shanghainese, an underlying /ŋ/ in coda position is posited.

There are two glides /j/ and /w/ in Lili Wu Chinese. Glides are typically defined as vowel-like segments that function as consonants and belong to the approximant class (Ladefoged & Maddieson, 1996). In Lili Wu Chinese, /j/ and /w/ differ from the corresponding vowels (i.e., /i/ and /u/) in that both tend to be produced with a narrower constriction of the vocal tract indicated via lower F1 values. Following Maddieson and Emmorey (1985), I compared mean F1 of the beginning interval (i.e., 50 ms) of /j/ (/jə̠o$^{1}$/ 'surname, Ou') with /i/ (/i$^{1}$/ 'smoke') and /w/ (/wɛ$^{2}$/ 'to return') with /u/ (/u$^{2}$/ 'river'), respectively. The results showed that the F1 values of /j/ (265 Hz) and /w/ (314 Hz) were lower than the corresponding vowels (/i/: 271 Hz; /u/: 354 Hz). Existing descriptions of Lili Wu Chinese such as Wang (2010) have typically posited high vowels /i u/ instead of glides /j w/ in words like /jə̠o$^{1}$/ and /wɛ$^{2}$/ (/iəu$^{1}$/ and /uᴇ$^{2}$/ in Wang, 2010),[10] despite the consensus among sinologists that they are glides. The approximants /j

[10] Sinologists frequently use /ᴇ/ to describe the phoneme between /e/ and /ɛ/, which however, does not exist in the IPA.

w/ have been adopted to transcribe the sounds. Needless to say, large-scale and well-designed studies are needed to investigate further the differences in glides vs. vowels in Chinese dialects. Additionally, note that before rounded vowels /o/ and /ø/, /j/ is realized as [ɥ] as in /joʔ$^8$/ [ɥoʔ$^8$] 'bath' and /jø$^2$/ [ɥø$^2$] 'rounded'. Because of the complementary distribution, /ɥ/ is treated as a context-specific (i.e., before /o/ and /ø/) variation of /j/.

A controversial issue is whether it is necessary to posit /j/ after an alveolo-palatal affricate or fricative onset (i.e., /tɕ$^h$ tɕ dʑ ɕ/) in Wu Chinese (see a brief discussion in Chen & Gussenhoven, 2015). Historically, these alveolo-palatal onsets are commonly believed to develop from the velar or glottal onsets (i.e., /k$^h$ k g h/) due to the palatalization process triggered by the following high front segments (e.g., Wang, 1985: 394). Synchronically, there is no contrast between /tɕ$^h$ tɕ dʑ ɕ/ and /tɕ$^h$j tɕj dʑj ɕj/ in Lili Wu Chinese. More remarkably, the transition from the alveolo-palatal affricate to the following vowel is rather brief. Figure 2.9 illustrates the different transitional characteristics among /tɑ$^1$/ 'knife' (2.9a), where there is no glide, /tjɑ$^1$/ 'marten' (2.9b) and /tsjɑ$^1$/ 'scorched' (2.9c), where there is commonly recognized presence of /j/, and /tɕɑ$^1$/ 'to converge' (2.9d), where I propose absence of /j/. Adapted the method of Chitoran (2002), I marked the beginning of the transition at the start of the sonorant part (i.e., glide or vowel) and the end of the transition as the turning point from a falling F2 to an F2 steady-state, before it falls consistently less than 20 Hz. The F2 values were automatically measured in Praat with a window length of 5 ms. Note that one would have expected a much more stable realization of /j/ with a longer transition from /tɕ/ to /ɑ/ if one assumed the presence of a glide /j/ following /tɕ/. These observations motivated me not to posit an underlying /j/ after alveolo-palatal onsets (following the analysis of Chen & Gussenhoven, 2015 for Shanghainese). But it is worth stressing the importance of further experimental studies to investigate the phonological status and phonetic realization of /j/ after alveolo-palatals in Lili Wu Chinese as well as other Chinese dialects.

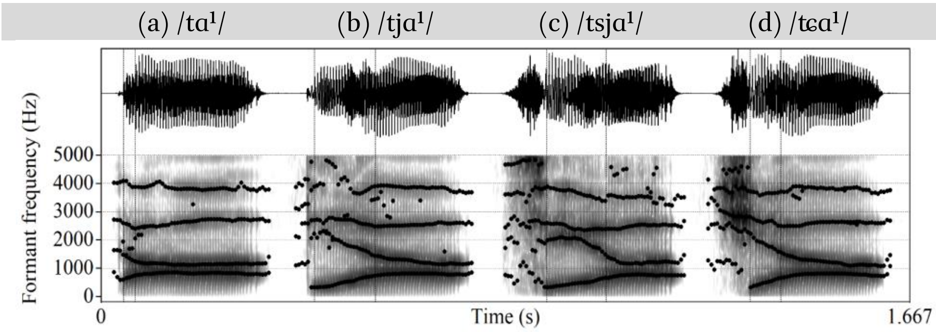



**Figure 2.9** Waveforms and spectrograms of (a) /tɑ$^1$/ 'knife', (b) /tjɑ$^1$/ 'marten', (c) /tsjɑ$^1$/ 'scorched', and (d) /tɕɑ$^1$/ 'to converge' respectively. Within each syllable, the transition from the end of the preceding consonant to the time that the F2 converges toward the value of /ɑ/ is indicated.

## 2.5 Vowels

Vowels in open syllables are plotted in Figure 2.10, and those in closed syllables in Figure 2.11. Except for /u/ (5), ten tokens of each vowel were used. The vowel ellipses were calculated based on the covariance of the tokens by following Pols et al. (1973). Formant values were automatically measured via VoiceSauce (Shue et al., 2011) using the Snack method. Formant values were converted from Hertz to Bark using the built-in formula[11] in Praat but were still indicated in corresponding Hertz values[12] in both plots for ease of comparison to the existing literature.

In open syllables, there are nine monophthongs (2.10a) and one diphthong (2.10b), as plotted in Figure 2.10. These nine monophthongs of Lili Wu Chinese (/i y ɨ̞ ɛ ø u o ɔ ɑ/) constitute a four-way distinction (i.e., close, close-mid, open-mid, and open) in height and a two-way distinction

---

[11] $\text{Bark} = 7 \cdot \ln(\text{Hertz} / 650 + \sqrt{(1 + (\text{Hertz} / 650)^2)})$.
[12] $\text{Hertz} = 650 \cdot \sinh(\text{Bark} / 7)$.

(i.e., front and back) in backness. /i y/ contrast in roundness. In addition, there is one diphthong occurring in open syllables, with /ɘ̯o/ gliding towards the back.

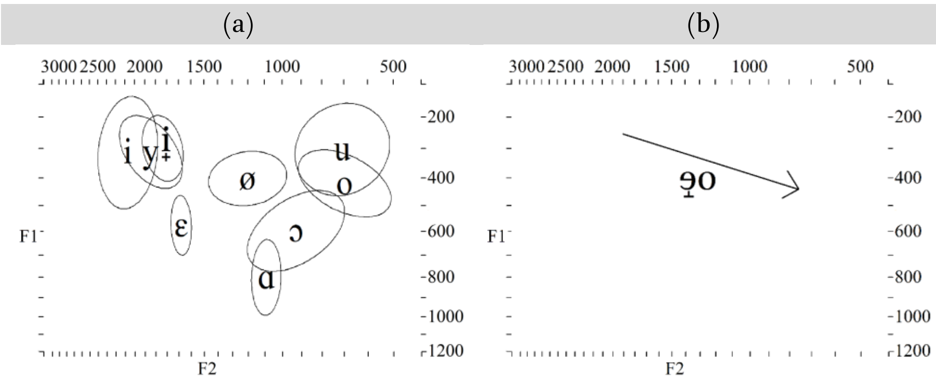



**Figure 2.10** Relative F1/F2 formant values of monophthongs and the diphthong in open syllables (a: for monophthongs; b: for the diphthong). The arrow demonstrates the trajectory of the gliding.

Monophthongs in closed syllables are plotted in Figure 2.11, where six (/ɪ ʏ æ ə ʊ ɑ/) occur in syllables closed by a nasal coda (2.11a) and four (/ɪ a ʊ ʌ/) in syllables closed by a glottal coda (2.11b). Compared to the vowels in open syllables, the number of vowels in closed syllables is largely reduced and so is their acoustic vowel space. Generally speaking, vowels in closed syllables are more central and lower than those in open syllables. Following Chen and Gussenhoven (2015), the same set of symbols (i.e., /ɪ/ and /ʊ/) for monophthongs followed by a nasal coda and those by a glottal coda was adopted, although their articulations do differ.

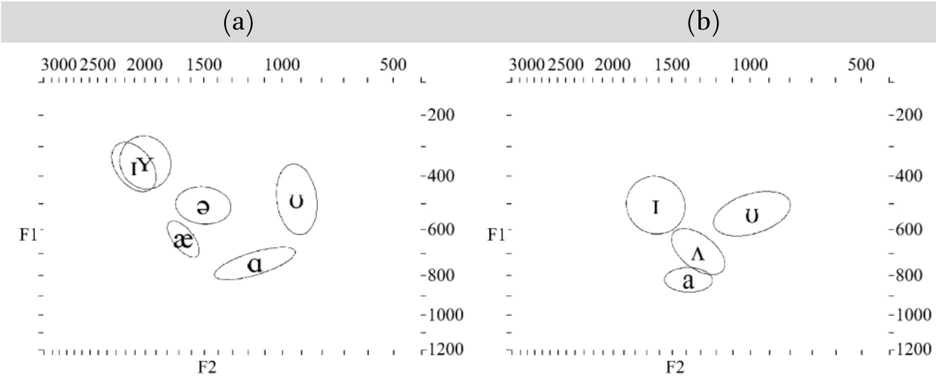



**Figure 2.11** Relative F1/F2 formant values of monophthongs in closed syllables (a: for vowels in syllables closed by a nasal; b: for vowels in syllables closed by a glottal coda).

Monophthongs and diphthong in open syllables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| i | ti$^{1}$ | 颠 | 'jolt' | u | u$^{1}$ | 乌 | 'crow' |
| y | ɕy$^{1}$ | 虚 | 'unreal' | ø | kø$^{1}$ | 肝 | 'liver' |
| i̟ | ti̟$^{1}$ | 低 | 'low' | o | ko$^{1}$ | 瓜 | 'melon' |
| ɛ | kɛ$^{1}$ | 该 | 'ought to' | ɔ | zɔ$^{2}$ | 柴 | 'firewood' |
| ɘ̞o | jɘ̞o$^{1}$ | 优 | 'superior' | ɑ | kɑ$^{1}$ | 高 | 'high' |

Monophthongs in closed syllables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ɪ | tɕɪŋ$^{1}$ | 金 | 'gold' | ʊ | kʊŋ$^{1}$ | 公 | 'public' |
| | tɕɪʔ$^{7}$ | 急 | 'hurry' | | kʊʔ$^{7}$ | 郭 | 'surname, Guo' |
| ʏ | tɕʏŋ$^{1}$ | 军 | 'army' | ə | kəŋ$^{1}$ | 跟 | 'to follow' |
| æ | kæŋ$^{1}$ | 庚 | 'age' | ʌ | kʌʔ$^{7}$ | 革 | 'to reform' |
| a | kaʔ$^{7}$ | 夹 | 'to clip' | ɑ | kɑŋ$^{1}$ | 刚 | 'just now' |

Lili Wu Chinese presents an interesting case of fricative vowel, as illustrated in Figure 2.12 which plots the spectrograms of the minimal pair /i/ in /ti$^{3}$/ 'dot' (2.12a) and /i̟/ in /ti̟$^{3}$/ 'bottom' (2.12b). The F2 of /i/ (2399 Hz) is higher than the F2 of /i̟/ (2009 Hz). Perceptually, a striking difference between /i/ and /i̟/ is the frication present in /i̟/ (Chao, 1928; Wang, 1987; Hu, 2007; Ling, 2007, 2011). Figure 2.13 exhibits narrow band spectrograms of /ti$^{3}$/ (2.13a) and /ti̟$^{3}$/ (2.13b). Harmonics can be clearly identified in /ti$^{3}$/ but are not in /tɨ$^{3}$/, especially in the frequency bands above 2 kHz. Furthermore, there is a substantial amount of aperiodic energy in the higher frequency region, particularly above 4 kHz in /tɨ$^{3}$/, which suggests the presence of strong fricative noise. This observation is further confirmed by the HNR (Harmonics-to-Noise Ratio) results, with /ɨ/ in /tɨ$^{3}$/ (8.1 dB) showing more noise than /i/ in /ti$^{3}$/ (9.8 dB).

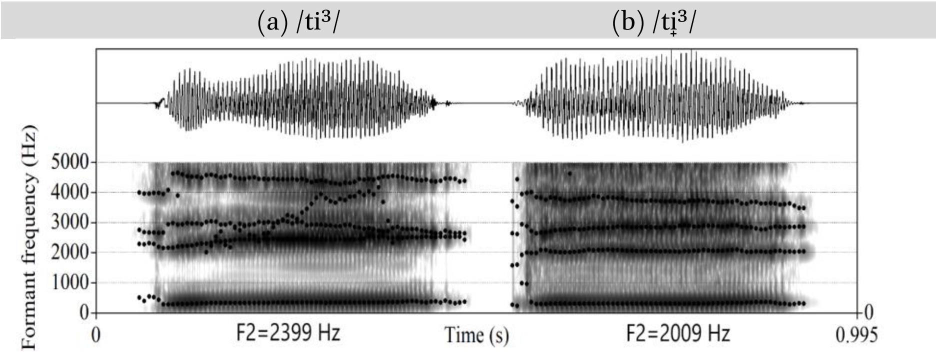



**Figure 2.12** Waveforms and spectrograms of (a) /ti$^{3}$/ 'dot' and (b) /tɨ$^{3}$/ 'bottom'. F2 values are indicated.

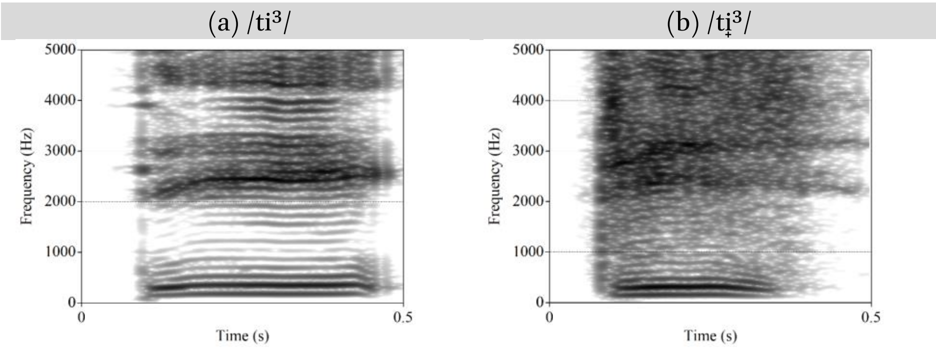



**Figure 2.13** Narrow band spectrograms of (a) /ti$^{3}$/ 'dot' and (b) /tɨ$^{3}$/ 'bottom'.

A similar contrast has been reported in Suzhou Wu Chinese. In order to illustrate the frication, Ling (2011) adopts the symbol /$i_z$/ for the phoneme and [ʒ̩̻] (i.e., the syllabic laminal postalveolar voiced fricative) for its phonetic realization. However, this treatment is problematic. First, a subscript /$_z$/ does not meet the convention of diacritic symbols in the IPA. Second, articulatory (i.e., palatographic, linguagraphic, and electromagnetic articulographic (EMA) studies) data of Suzhou Wu Chinese have shown that the constriction of /i̟/ is located at a more anterior position (Ling, 2007, 2011; Hu & Ling, 2019) than /i/. Consequently, the lengthening of the back resonating cavity lowers the F2 of /i̟/ as argued by Ling (2011) following Stevens (1989: 10). Third, it is also worth noting that the frication in Lili Wu Chinese, compared to that in Suzhou Wu Chinese, is not consistently audible for all /i̟/ words produced by the consultant and also not as strong as that in Suzhou Wu Chinese. For instance, there is little frication in /fi̟$^{1}$/ [fᵊɨ$^{1}$] 'to fly' (which tends to be diphthongal). Given the three reasons, the symbol ̟ has been adopted to highlight the more anterior constriction of /i̟/ and the less friction. Such an articulatory gesture is also accompanied by a remarkable raising of the lower jaw in words such as /i̟$^{1}$/ 'clothes', which however, is not observed in words such as /i$^{1}$/ 'smoke'. The contrast of high front vowels is an areal phenomenon in many Chinese dialects, especially in the Jianghuai Mandarin family (江淮官话) (R. Shi, 1998; Zhu, 2004b; Zhao, 2007). Similar contrasts have also been reported in some modern African languages, such as Len Mambila (Connell, 2007) and Ring languages (Faytak & Merrill, 2015).

Both /u/ (in /u$^{1}$/ 'crow') and /o/ (in /ko$^{1}$/ 'melon') are close/close-mid back monophthongs with compressed lip rounding. The lips for /o/ are more protruding but for /u/ they are less rounded and more compressed (similar to the /u o/ contrast in Shanghainese as discussed in Chen & Gussenhoven, 2015). After bilabial and labio-dental, /u/ is produced as [v̩] (i.e., the syllabic labiodental voiced fricative), as exemplified in /pu$^{1}$/ [pv̩$^{1}$] 'wave'. After alveolar, alveolo-palatal and velar consonants, /u/ is realized with diphthong quality (i.e., [əʊ]), as shown in /ku$^{1}$/ [kəʊ$^{1}$] 'song'. According to a Suzhounese syllabary named *A syllabary of the Suchow dialect*, recorded by A Committee of the Soochow Literary Association in 1892 for missionaries in acquiring Suzhounese, such a diphthongal realization of /u/ after alveolar, alveolo-palatal and velar consonants can be traced back to the beginning of the 20$^{th}$ Century.

The front vowel /ø/ tends to be produced with a lower F2 such as in /ø$^{1}$/ [ʔø̈$^{1}$] 'in safe' (1228 Hz) than in /jø$^{2}$/ [ɥø$^{2}$] 'rounded' (1425 Hz). Both, however, are produced with a lip rounding gesture. Following the treatment in Wang (2010), an underlying /ø/ is posited.

/ɘ̯o/ is a diphthong and only co-occurs with the glide /j/ (e.g., /vjɘ̯o$^{2}$/ 'to float' and /kjɘ̯o$^{1}$/ 'to tick off') or alveolo-palatals (e.g., /dʑɘ̯o$^{6}$/ 'used' and /ɕɘ̯o$^{1}$/ 'to rest'). It is transcribed as /əu/ in Wang (2010) but seems not to be supported by the formant trajectory produced by the consultant.

As for vowels in closed syllables, the contrast between /æ/ and /ɑ/ only exists before the nasal coda /ŋ/. As illustrated in /tsʰæŋ$^{6}$/ [tsʰæ̃$^{6}$] 'unimpeded' and /tsʰɑŋ$^{6}$/ [tsʰɑ̃$^{6}$] 'to sing', both vowels are consistently nasalized without recognizable velum closure at the end.

Vowels preceding a glottal stop coda show a much shorter duration. When high vowels (i.e., /ɪ/ and /ʊ/) occur before /ʔ/, a general displacement towards an open back position often results in a brief schwa after nuclei, such as /tɕɪʔ$^{7}$/ [tɕɪᵊʔ$^{7}$] 'hurry' and /kʊʔ$^{7}$/ [kʊᵊʔ$^{7}$] 'surname, Guo'.

## 2.6 Syllabic approximants

There are two syllabic approximants in Lili Wu Chinse, which are exemplified in /sɹ̩$^{1}$/ [sɹ̩̻$^{1}$] 'silk' and /sɹ̩̹$^{1}$/ [sʷɹ̩̹̻$^{1}$] 'book'. The syllabic approximant /ɹ̩/ [ɹ̩̻] in Lili Wu Chinese is similar to that in Standard Chinese.

With respect to /ɹ̩̹/, two features are to be further noted. First, the lip rounding gesture of the approximant contributes to the labialization of the preceding alveolar sibilant onset (i.e., /s/ [sʷ] before /ɹ̩̹/). Labialized alveolar sibilants are rare in the world's languages (but see Lao, a Kra-Dai language reported in Erickson, 2001). The rounding feature is believed to evolve from /u/ or /y/, the two rounded vowels reported to be present instead of /ɹ̩̹/ in other Wu dialects, such as /su$^{1}$/ in Danyang (丹阳) Wu and /ɕy$^1$/ in Songjiang (松江) Wu for ‘book’ (Qian, 1992: 88). In addition, /ɹ̹̍/ is articulated more laminally. Laminal consonants have been widely reported to exist in Australian languages (Butcher, 1990; Anderson & Maddieson, 1994). Such an articulatory gesture of /ɹ̹̍/ is reflected in Figure 2.14 as a lowered F4 (3375 Hz, compared to 4221 Hz of /ɹ̩/ in /sɹ̩$^1$/) and the proximity of F3 and F4. F4 lowering is generally said to be related to articulatory retraction (e.g., Fant, 1960; Stevens & Blumstein, 1975; Vaissière, 2011). The proximity of F3 and F4 is known as a consequence of weakly coupled resonators by forming a relatively larger frontal cavity (Stevens, 1989: 19). For instance, a significant convergence of F3 and F4 is observed in laminal alveolar and postalveolar fricatives in English, as well as in apico-laminal alveolars in French (Dart, 1991: 104). In short, /ɹ̹̍/ is produced with a more laminal articulation combined with a lip rounding gesture than its counterpart /ɹ̩/. Such differences were also noticed by the consultant who offered his native intuition voluntarily with us. Given the impressionistic nature of the description, needless to say, more instrumental studies (e.g., ultrasound) are needed for a precise description of their articulation and acoustic consequences.

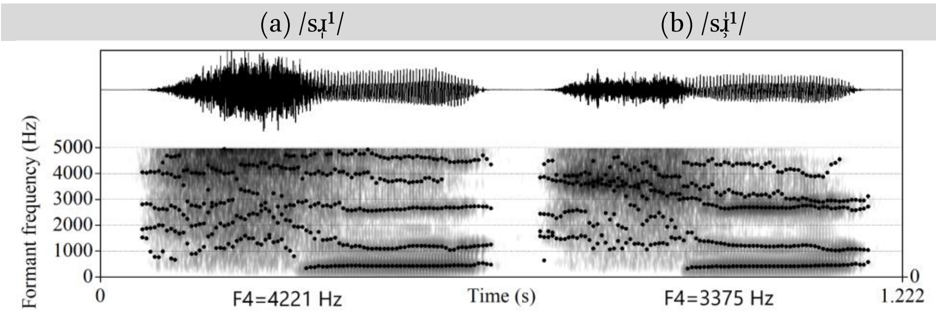



**Figure 2.14** Waveforms and spectrograms of (a) /sɹ̩$^1$/ ‘silk’ and (b) /sɹ̹̍$^1$/ ‘book’. F4 values are indicated.

It is worth noting that there exist different proposals to transcribe these sounds. For example, among sinologists (after Karlgren, 1915: 294), /ɹ̩/ and /ɹ̹̍/ have often been transcribed as /ɿ/ and /ʮ/, respectively, and are known as ‘apical vowels’. However, as criticized by Lee-Kim (2014: 264), both symbols 'are obsolete in the IPA and not informative for non-Chinese linguists' and are therefore avoided in this description. /ɹ̩/ is usually treated as [z̩] (e.g., Ladefoged & Maddieson, 1996: 314; Wiese, 1997: 239; Duanmu, 2000: 36 for Standard Chinese; Chen & Gussenhoven, 2015 for Shanghainese). Such a treatment, however, has been questioned with ultrasound imaging data (Lee-Kim, 2014; Faytak & Lin, 2015) and acoustic analyses (Howie, 1976: 10). Lee-Kim (2014) further argues that it is more appropriate to describe [z̩] as 'dental approximant [ɹ̪̍]' (see a similar description in Lee & Zee, 2003: 111).[13]

Last but not least, an increasing body of literature has shown that such syllabic approximants are known to affect diachronic changes of high vowels in different languages at different time points, across an overwhelmingly large number of Sino-Tibetan languages (e.g., Baron, 1974; R. Shi, 1998; Zhu, 2004b; Zhao, 2007; Hu & Ling, 2019).

## 2.7 Syllable structure

Generally speaking, eight syllable combinations can be identified in Lili Wu Chinese. The canonical syllable minimally consists of an obligatory nucleus (V) and a lexical tone as in /u$^{1}$/ 'crow' and /ø$^{2}$/ [ɦø̈$^{2}$] 'cold'. V stands for vowel or syllabic consonant (i.e., /ɹ̩ ɻ̩ m̩ ŋ̍/).[14] It may also contain up to three optional elements in the following linear structure: ($C_1$)(G)V($C_2$),[15] where $C_1$ can be any consonant in the consonant inventory except for /ʔ/, G is either /j/, as in /kjɘ̠o$^{1}$/ 'to tick off', or /w/, as in /kwɛ$^{1}$/ 'to close'; $C_2$ is either /ŋ/ or /ʔ/ as in /kʊŋ$^{1}$/ 'public' and /kaʔ$^{7}$/ 'to clip'. All combinations are demonstrated in Table 2.6.

[13] In Lee-Kim (2014: 264), the syllabic diacritic is not used for the sake for simplicity.
[14] In Lili Wu Chinese, /ɹ̩ ɻ̩/ are obligatory to contain an onset.
[15] Parentheses indicate optional constituents.

**Table 2.6** Syllabic combinations in Lili Wu Chinese.

| Combination | Example | | | | | |
|---|---|---|---|---|---|---|
| V | /u$^1$/ | 乌 | 'crow' | /ø$^2$/ | 寒 | 'cold' |
| GV | /wɔ$^4$/ | 坏 | 'broken' | /jø$^2$/ | 圆 | 'rounded' |
| $C_1V$ | /zɔ$^1$/ | 柴 | 'firewood' | /kɑ$^1$/ | 高 | 'high' |
| $VC_2$ | /ʊŋ$^1$/ | 翁 | 'surname, Weng' | /aʔ$^8$/ | 盒 | 'box' |
| $C_1GV$ | /kjə̯o$^1$/ | 勾 | 'to tick off' | /kwɛ$^1$/ | 关 | 'to close' |
| $C_1VC_2$ | /kʊŋ$^1$/ | 公 | 'public' | /kaʔ$^7$/ | 夹 | 'to clip' |
| $GVC_2$ | /jʊŋ$^2$/ | 熊 | 'bear' | /wʌʔ$^7$/ | 活 | 'alive' |
| $C_1GVC_2$ | /zjæŋ$^1$/ | 墙 | 'wall' | /kwʌʔ$^7$/ | 国 | 'country' |

As illustrated in Table 2.7, co-occurrence constraints on onset and rhyme combinations can be observed.

First, /i ɪŋ ɪʔ i̟/ behave similarly except that /i̟/ can appear after labio-dentals as in /fi̟$^1$/ 'to fly' and /vi̟$^2$/ 'fat'. /i/, on the other hand, is prohibited in this context (i.e., */fi/, */vi/). Second, /y ʏŋ/ are only allowed after alveolar sonorants and alveolo-palatals, or without an onset. Third, before /ø o ɔ ɑ æŋ/, labio-dentals are prohibited (*/fø fo fɔ fɑ fæŋ/) but /ɛ u/ are possible as in /fɛ$^1$/ 'to turn over' and /vu$^2$/ 'to support somebody with one's arm'. Fourth, the two syllabic approximants /ɹ̩ ɹ̩̹/ occur only after alveolar homorganic sibilant onsets /ʦ ʦʰ ʣ s z/.

The distribution of the two glides is summarized in Table 2.8. /j/ is allowed in the majority of cases (e.g., /pjɑ$^1$/ 'watch', /vjə̯o$^2$/ 'to float', /tjɑ$^1$/ 'marten', /kjə̯o$^1$/ 'to tick off', /hjə̯o$^3$/ 'to roar', and /jø$^2$/ 'rounded') except after alveolo-palatals. /w/, however, is more constrained and only allowed after velars (e.g., /kwɛ$^1$/ 'to close'), glottal fricative /h/ (e.g., /hwɛ$^1$/ 'dust'), or serves as a glide onset (e.g., /wɛ$^2$/ 'to be back').

**Table 2.7** Observed onset-rhyme combinations in Lili Wu Chinese. +: Yes; –: No.

| Onset | i ɪŋ ɪʔ | i̟ | y ɤŋ | ɛ u ɑŋ ʊŋ əŋ aʔ ʊʔ ʌʔ | ø o ɔ ɑ æŋ | ɹ̩ ɻ̩ | ɘ̠o |
|---|---|---|---|---|---|---|---|
| Bilabial (pʰ p b m) | + | + | – | + | + | – | – |
| Labio-dental (f v) | – | + | – | + | – | – | – |
| Alveolar plosives (tʰ t d) | + | + | – | + | + | – | – |
| Alveolar sonorants (n l) | + | + | +[16] | + | + | – | – |
| Alveolar sibilants (tsʰ ts dz s z) | + | + | – | + | + | + | – |
| Alveolo-palatal (tɕʰ tɕ dʑ ɕ) | + | + | + | +[17] | +[18] | – | + |
| Velar and Glottal (kʰ k g h ŋ) | – | – | – | + | + | – | – |
| j | – | – | – | +[19] | +[20] | – | + |
| w | – | – | – | + | +[21] | – | – |
| Zero onset | + | + | + | + | + | – | – |

[16] /nɤŋ/ and /lɤŋ/ cannot be observed.
[17] /tɕʰ tɕ dʑ ɕ/ + /ɛ u əŋ/ cannot be observed.
[18] /tɕʰ tɕ dʑ ɕ/ + /o/ cannot be observed.
[19] /jɛ/, /ju/, and /jəŋ/ cannot be observed.
[20] /jo/ cannot be observed.
[21] /wu/, /wø/ and /wɑ/ cannot be observed.

**Table 2.8** Observed onset-glide combinations in Lili Wu Chinese. +: Yes; –: No.

| Onset | j | w |
|---|---|---|
| Bilabial | + | – |
| Labio-dental | + | – |
| Alveolar | + | – |
| Alveolo-palatal | – | – |
| Velar and Glottal | + | + |
| Glottal | + | + |
| Glide onset | + | + |

## 2.8 Onsetless syllables

In onsetless syllables with high-register tones (i.e., T1, T3, T5, and T7), the phonetic segment [ʔ] can be observed at the onset of the tone-bearing syllable, as in /ø¹/ [ʔö¹] 'in safe' and /si̟¹ ø¹/ [si̟⁴⁴ ʔö⁴²] 'a city, Xi'an'. With respect to onsetless syllables with low-register tones (i.e., T2, T4, T6, and T8), phonetic realization of [ɦ] before a non-high vowel (e.g., /ø²/ [ɦö²] 'cold', /ɔ²/ [ɦɔ²] 'shoes', and /aʔ⁸/ [ɦaʔ⁸] 'box') is observed, in contrast to other cases when there is a high vowel or glide (e.g., /i²/ [ʝi²] 'salt', /jø²/ [ɥø²] 'rounded', /u²/ 'river', and /wʌʔ⁸/ 'alive'). [ɦ] disappears in non-initial position within a prosodic word, e.g., /tʰɑ⁴ ɔ²/ 'galoshes'. The general pattern is therefore similar to Shanghainese (Chen & Gussenhoven, 2015).

In Lili Wu Chinese, syllables with low-register tones show relatively stronger breathiness than those with high-register counterparts. As indicated by Figure 2.15, the Fast Fourier transform (FFT) spectrum of /ø¹/ [ʔö¹] 'in safe' (dark) and /ø²/ [ɦö²] 'cold' (light) shows the phonatory contrast in the vowel /ø/, taken within an interval of approximately 30 ms from the first regular vocal pulse of the vowel. As shown by the measurements on H1–H2 (i.e., amplitude difference between the first and second harmonics), there is a phonatory difference between the two vowels with /ø²/ (4.5 dB) showing more breathiness than /ø¹/ (2 dB). This contrast has also been observed in some other Northern Wu dialects (Cao & Maddieson, 1992).

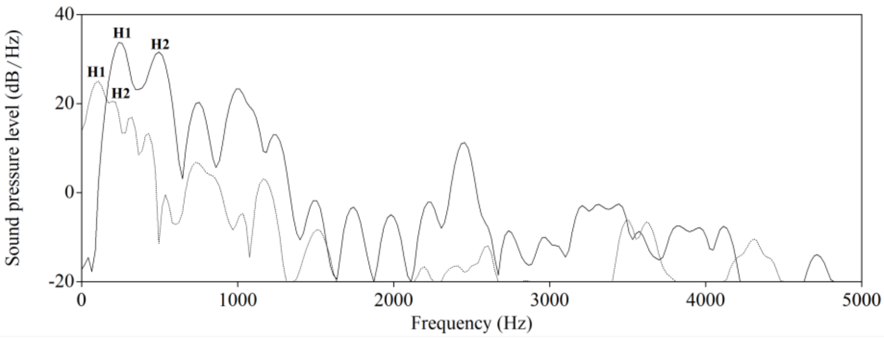



**Figure 2.15** FFT spectrum of /ø/ in /ø$^{1}$/ 'in safe' (dark) and /ø$^{2}$/ 'cold' (light), taken over an interval of approximately 30 ms from the first regular vocal pulse of the vowel. The first two harmonics (H1 and H2) of each syllable are indicated.

## 2.9 Tone sandhi

Lexical tones over monosyllabic morphemes undergo changes when they are combined into compounds or phrases. In this description, I offer some preliminary observations concerning tone sandhi variations in Lili Wu Chinese over disyllabic compounds (hereafter called the tone unit). Tonal realization is contingent upon the lexical tone of the second syllable ($\sigma_2$). Two general patterns are observed.

First, when $\sigma_2$ carries an abrupt tone (i.e., T7 and T8 over a glottal-coda syllable), regardless of the syllable structure of the first syllable ($\sigma_1$), only level *f0* contours surface, and the specific *f0* height is dependent on the lexical tone of $\sigma_1$. After a high tone, a low tone appears; while after a low tone, a high tone appears. Both patterns are illustrated in Figure 2.16, which shows T1 (high–level) + T7/T8 (2.16a, /tsʰəŋ$^{1}$ tsɪʔ$^{7}$/ 'the Spring Festival'; 2.16b, /tɕɪŋ$^{1}$ dʑɪʔ$^{8}$/ 'Peking Opera') and T6 (low–level) + T7/T8 (2.16c, /tʰɔ$^{6}$ kwʌʔ$^{7}$/ 'Thailand'; 2.16d, /tɕʰi̟$^{6}$ dɪʔ$^{8}$/ 'steam whistle').

Second, when $\sigma_2$ carries a non-abrupt tone (i.e., T1 to T6 over an open syllable or a syllable with a nasal coda), the lexical tonal contour of

$\sigma_1$ remains and affects the pitch realization of $\sigma_2$. The specific *fo* contour of $\sigma_2$ hinges upon the lexical tonal register of $\sigma_1$. When $\sigma_1$ is produced with a high-register tone (i.e., T1, T3, T5, and T7), $\sigma_2$ is typically realized with a falling *fo* contour, as shown in Figure 2.17 (e.g., 2.17a, /sɪŋ$^{1}$ zəŋ$^{4}$/ 'new kidney'; 2.17b, /kɛ$^{3}$ zɑ$^{4}$/ 'to remold'; 2.17c, /tɕɔ$^{5}$ zɑ$^{4}$/ 'introduction'; 2.17d, /kʊʔ$^{7}$ tʰu$^{4}$/ 'territory'). However, other patterns have also been observed. For example, in the combination of T7 + $\sigma_2$, when $\sigma_2$ bears T1 (e.g., /kʊʔ$^{7}$ kʰu$^{1}$/ 'orthopedics'), the monosyllabic citation form of T1 in /kʰu$^{1}$/ (high–level) is preserved, instead of a predictable falling contour like /tʰu$^{4}$/ in /kʊʔ$^{7}$ tʰu$^{4}$/ 'territory'.

When σ1 is pronounced with a low-register tone (i.e., T2, T4, T6, and T8), the sandhi pattern tends to be more complicated. The tonal contour of σ2 seems to also exert influence on the overall tonal realization. For example, Figure 2.18 shows the contrast of /pʰɔ$^{6}$ tɕʰi$^{4}$/ 'to dispatch' (2.18a) vs. /tɕʰi̟$^{6}$ pʰɑ$^{6}$/ 'bubble' (2.18b). Here, T4 in /tɕʰi$^{4}$/ completely loses its monosyllabic citation form (low–level) and is realized with a high-falling contour, similar to Shanghainese (Chen & Gussenhoven, 2015). However, the lexical tone of the preceding tone T6 in /pʰɑ$^{6}$/ (low–dipping) is only preserved to a certain extent. The same tone (i.e., T6) is realized with an audible pitch level difference: T6 in /pʰɔ$^{6}$ tɕʰi$^{4}$/ is overall lower than that in /tɕʰi̟$^{6}$ pʰɑ$^{6}$/. Further research is needed to investigate the extent to which listeners are sensitive to the differences.

Before closing off this section, it is worth pointing out that syllables with aspirated onsets show two different patterns of changes. They pattern either with syllables that have unaspirated onsets and carry T1, or with syllables that have voiced onsets and carry T4, T6, or T8. For example, the sandhi change of /tʰʊŋ$^{1}$ fʊŋ$^{1}$/ 'to ventilate' patterns with that of /tʊŋ$^{1}$ fʊŋ$^{1}$/ 'east wind'; while /tsʰʌ$^{8}$ djə̝o$^{2}$/ 'to stand out' patterns with /zʌ$^{8}$ djə̝o$^{2}$/ 'tongue'.

It is important to conclude here that even within the arguably simplest construction beyond a monosyllabic morpheme (i.e., disyllabic compounds), Lili Wu Chinese already exhibits different patterns of tonal realization from its neighboring Northern Wu dialects such as Shanghainese (Chen & Gussenhoven, 2015). It is not only subject to the influence of the preceding tone on tonal realization but also seems sensitive to tonal properties of the second syllable. In this illustration, I have just presented a preliminary glimpse into the pitch contours of disyllabic compounds in Lili Wu Chinese. Needless to say, more data and further research are needed.

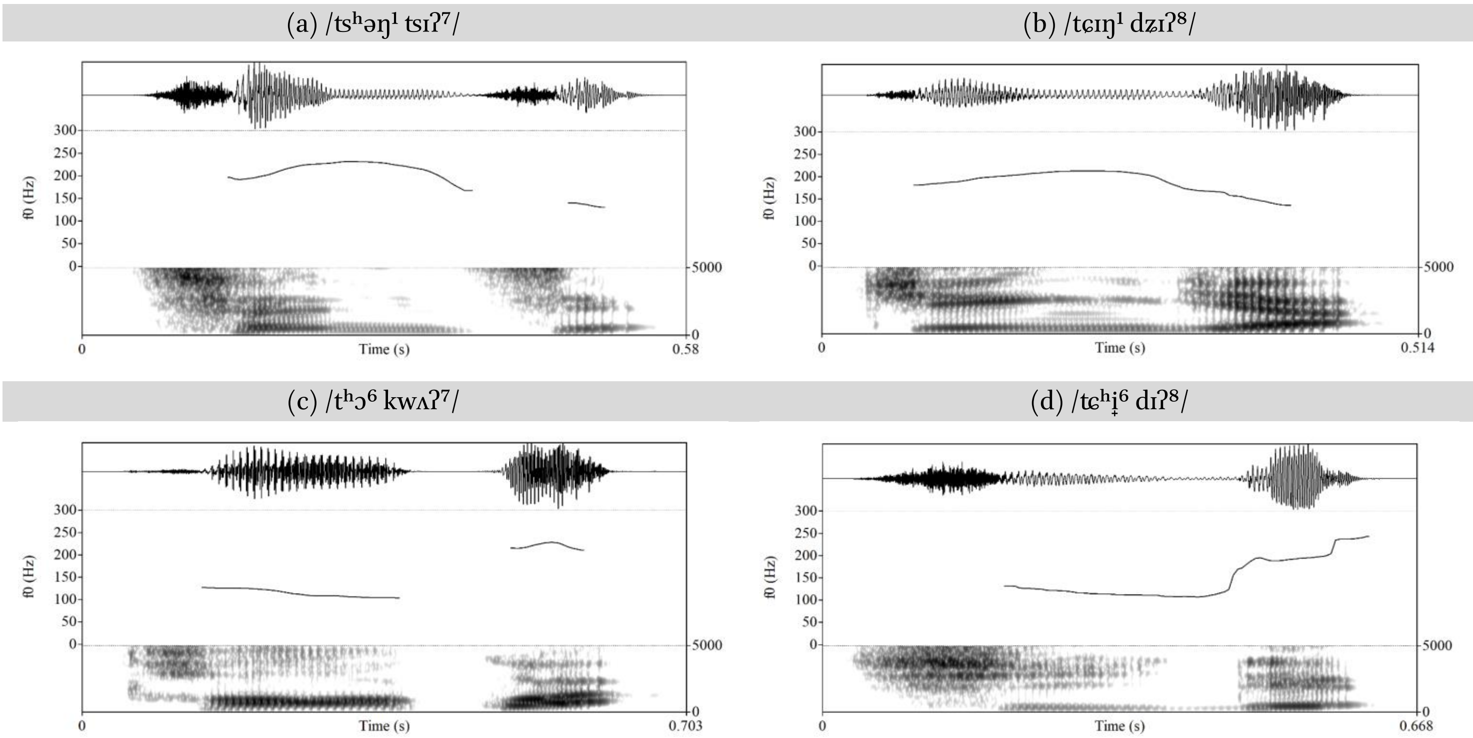



**Figure 2.16** Waveforms, *f0* tracks, and spectrograms of (a) /tsʰəŋ$^{1}$ tsɪʔ$^{7}$/ 'the Spring Festival', (b) /tɕɪŋ$^{1}$ dʑɪʔ$^{8}$/ 'Peking Opera', (c) /tʰɔ$^{6}$ kwʌʔ$^{7}$/ 'Thailand', and (d) /tɕʰɨ̥$^{6}$ dɪʔ$^{8}$/ 'steam whistle'.

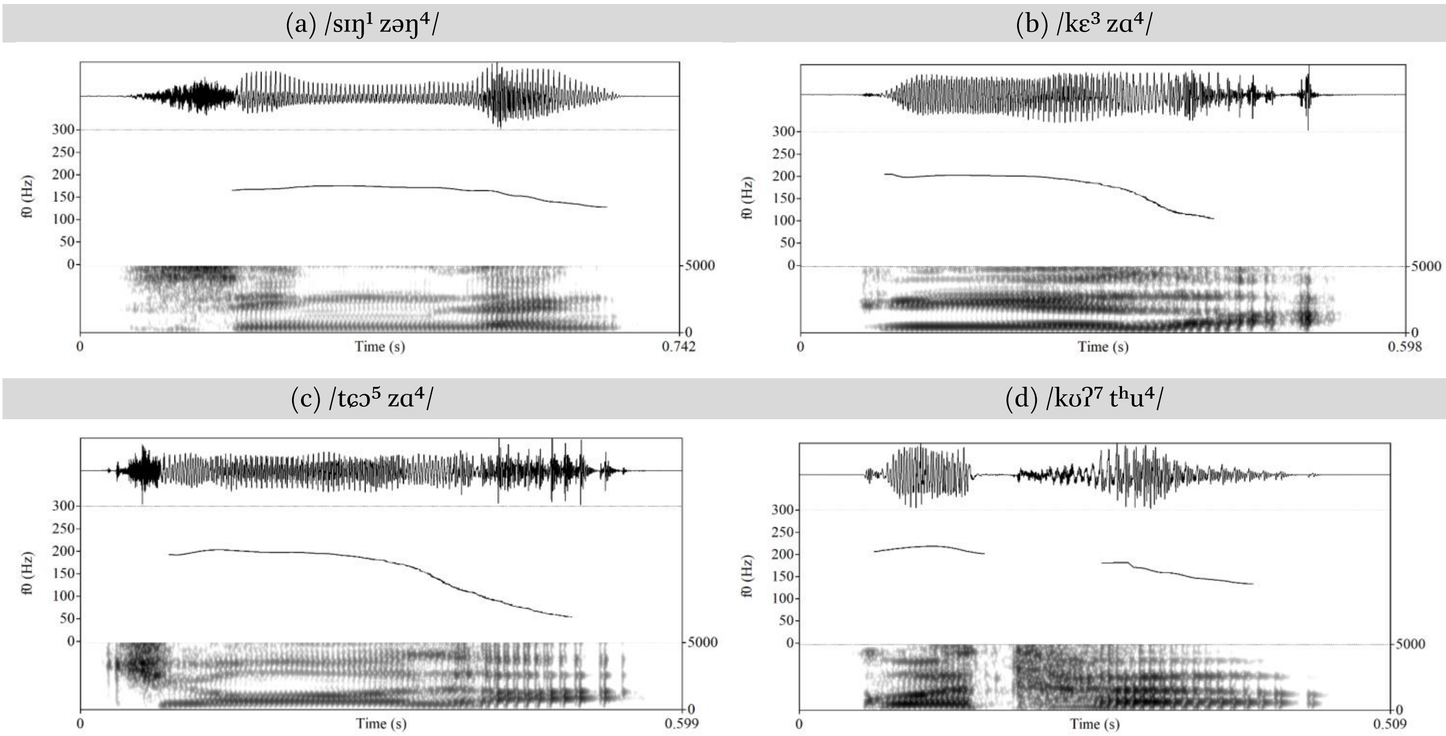
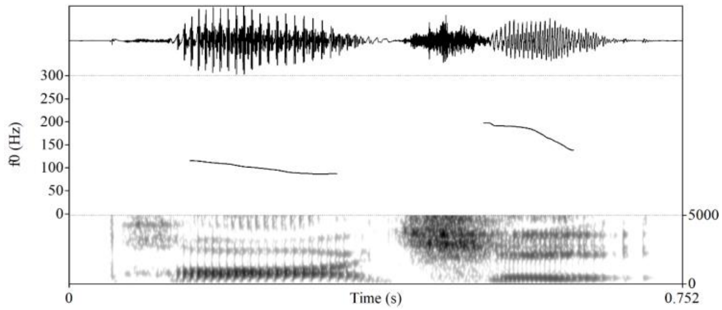



**Figure 2.17** Waveforms, *f0* tracks, and spectrograms of (a) /sɪŋ$^{1}$ zəŋ$^{4}$/ ‘new kidney’, (b) /kɛ$^{3}$ zɑ$^{4}$/ ‘to remold’, (c) /tɕɔ$^{5}$ zɑ$^{4}$/ ‘introduction’, and (d) /kʊʔ$^{7}$ tʰu$^{4}$/ ‘territory’.

(a) /p$^{h}$ɔ$^{6}$ tɕ$^{h}$i$^{4}$/




(b) /tɕ$^{h}$i̟$^{6}$ p$^{h}$ɑ$^{6}$/

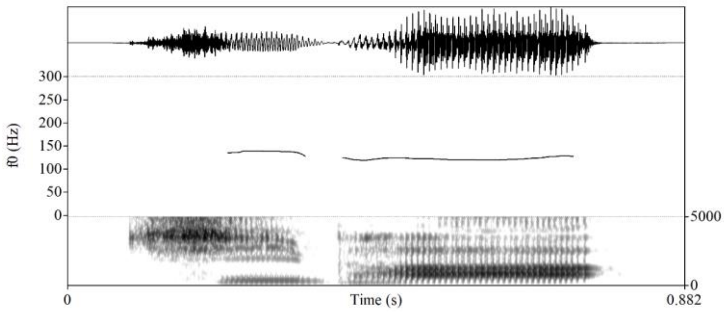



**Figure 2.18** Waveforms, *f0* tracks, and spectrograms of (a) /p$^{h}$ɔ$^{6}$ tɕ$^{h}$i$^{4}$/ 'to dispatch' and (b) /tɕ$^{h}$i̟$^{6}$ p$^{h}$ɑ$^{6}$/ 'bubble'.

# 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 *fo* contours of the other two types. In Stage III, the *fo* 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 *fo* 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.

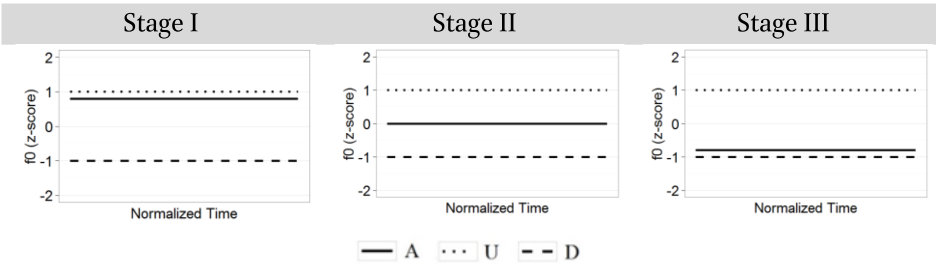



**Figure 3.1** Three expected stages of ATS based on the 'ongoing change' view. A: voiceless aspirated onsets; U: voiceless unaspirated onsets; D: voiced onsets.

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 i̽/ (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$^{213}$ lɨ̞$^{13}$ lɨ̞ 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: $y = \alpha + \beta \cdot \text{Time}_1 + \gamma \cdot \text{Time}_2 + \Delta$ would have three time terms: intercept ($\alpha$), linear ($\beta$), and quadratic ($\gamma$). 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) ($\chi^2$) 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: $\chi^2$ = 187.35, $p$ < .001; II: $\chi^2$ = 75.68, $p$ < .001; III: $\chi^2$ = 1638, $p$ < .001; IV: $\chi^2$ = 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 *f0*. 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 ($\chi^2$) 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.

| 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 *f0* 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 *f0* 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.

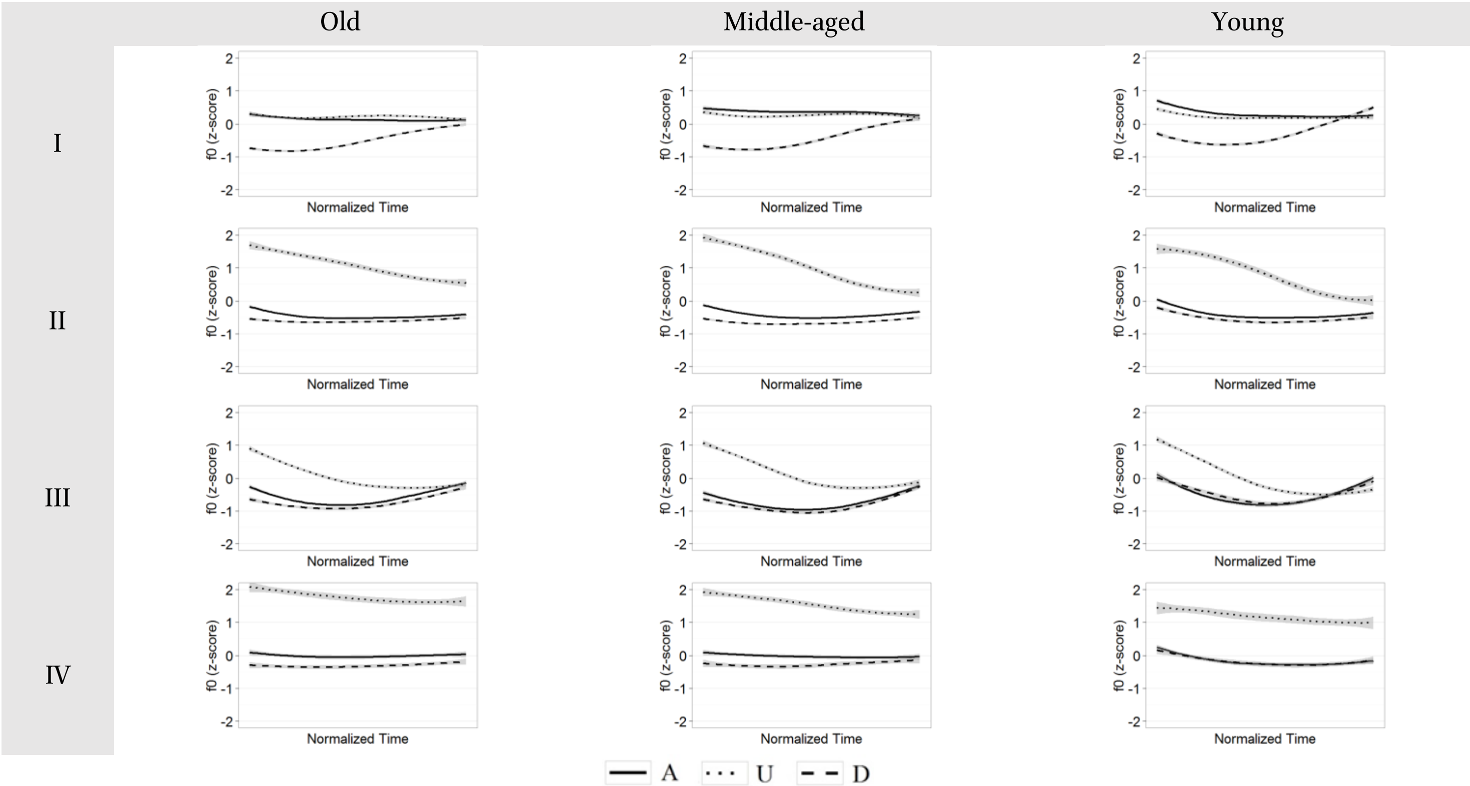



**Figure 3.2** The *f*0 realization of three generations. Mean normalized *f*0 (gray areas indicate ± *SE*) averaged across speakers and repetitions. The x-axis is normalized time and the y-axis is normalized *f*0.

**Table 3.4** Results ($\beta$) of the effect of Consonant on *f*o. 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$^{n.s.}$ | .14$^{n.s.}$ | −0.14$^{n.s.}$ | II-U | 1.51$^{<.001}$ | −1.45$^{<.001}$ | −0.25$^{<.05}$ |
| | I-D | −0.64$^{<.001}$ | 1.45$^{<.001}$ | .16$^{n.s.}$ | II-D | −0.15$^{n.s.}$ | .28$^{n.s.}$ | −0.19$^{n.s.}$ |
| Middle-aged | I-U | −0.1$^{n.s.}$ | .2$^{n.s.}$ | .01$^{n.s.}$ | 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$^{n.s.}$ | .28$^{n.s.}$ | −0.17$^{n.s.}$ |
| Young | I-U | −0.09$^{n.s.}$ | .24$^{n.s.}$ | −0.13$^{n.s.}$ | 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$^{n.s.}$ | .1$^{n.s.}$ | −0.07$^{n.s.}$ |
| Generation | III-A | Intercept | Linear | Quadratic | IV-A | Intercept | Linear | Quadratic |
| Old | III-U | .6$^{<.001}$ | −1.84$^{<.001}$ | −0.09$^{n.s.}$ | IV-U | 1.78$^{<.001}$ | −0.61$^{<.05}$ | .03$^{n.s.}$ |
| | III-D | −0.16$^{n.s.}$ | .19$^{n.s.}$ | −0.22$^{n.s.}$ | IV-D | −0.29$^{n.s.}$ | .18$^{n.s.}$ | −0.02$^{n.s.}$ |
| Middle-aged | III-U | .84$^{<.001}$ | −2.12$^{<.001}$ | −0.01$^{n.s.}$ | IV-U | 1.56$^{<.001}$ | −0.84$^{<.01}$ | −0.01$^{n.s.}$ |
| | III-D | −0.1$^{n.s.}$ | .15$^{n.s.}$ | −0.03$^{n.s.}$ | IV-D | −0.24$^{n.s.}$ | .39$^{n.s.}$ | .06$^{n.s.}$ |
| Young | III-U | .48$^{<.001}$ | −2.08$^{<.001}$ | −0.26$^{n.s.}$ | IV-U | 1.47$^{<.001}$ | −0.41$^{n.s.}$ | −0.32$^{<.01}$ |
| | III-D | .01$^{n.s.}$ | −0.17$^{n.s.}$ | −0.18$^{n.s.}$ | IV-D | −0.002$^{n.s.}$ | .16$^{n.s.}$ | −0.09$^{n.s.}$ |

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: $\beta = .12$, $p < .001$; middle-aged: $\beta = .11$, $p < .001$; young: $\beta = .06$, $p < .001$]. 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.

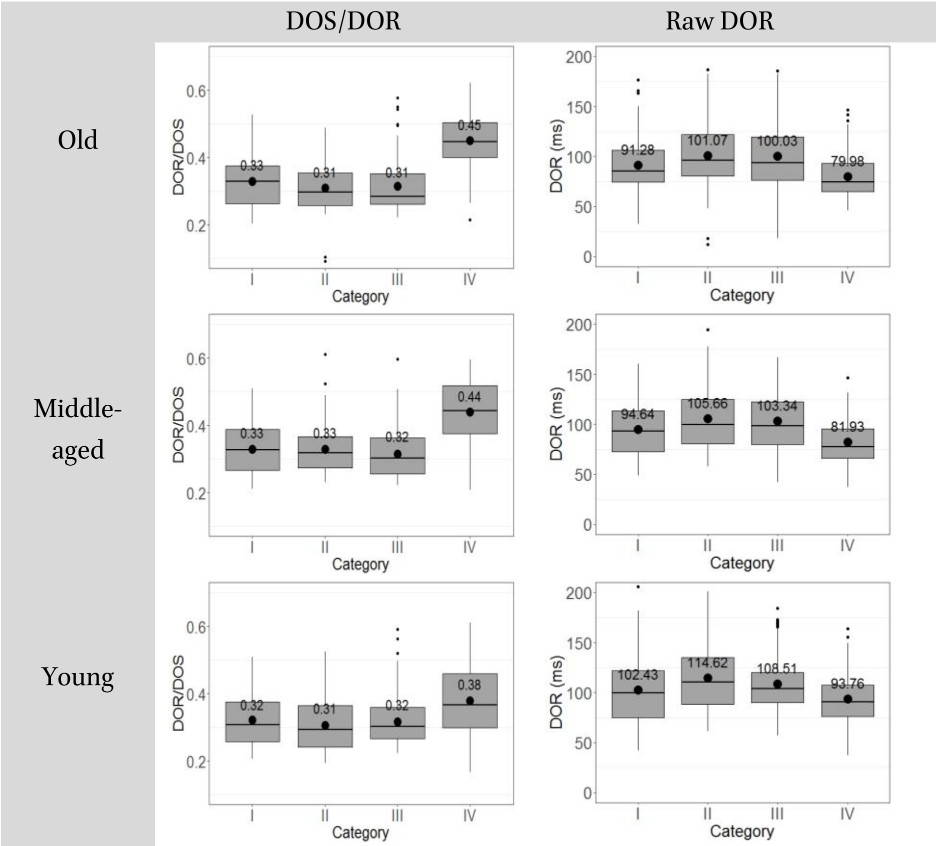



**Figure 3.3** Boxplots of the DOR/DOS ratio in target syllables of each MC tonal category for three generations. The bottom and top lines of a box represent the lower (or first) and upper (or third) quartiles of the data. The lines extending vertically from the box are known as the 'whiskers', which 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 ($\beta$) | $t$ | $p$ | Estimate ($\beta$) | $t$ | $p$ | Estimate ($\beta$) | $t$ | $p$ |
| 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 ($\chi^2$) 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 | $\chi^2$ | Two-way | $\chi^2$ | Three-way and four-way | $\chi^2$ |
|---|---|---|---|---|---|
| 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).

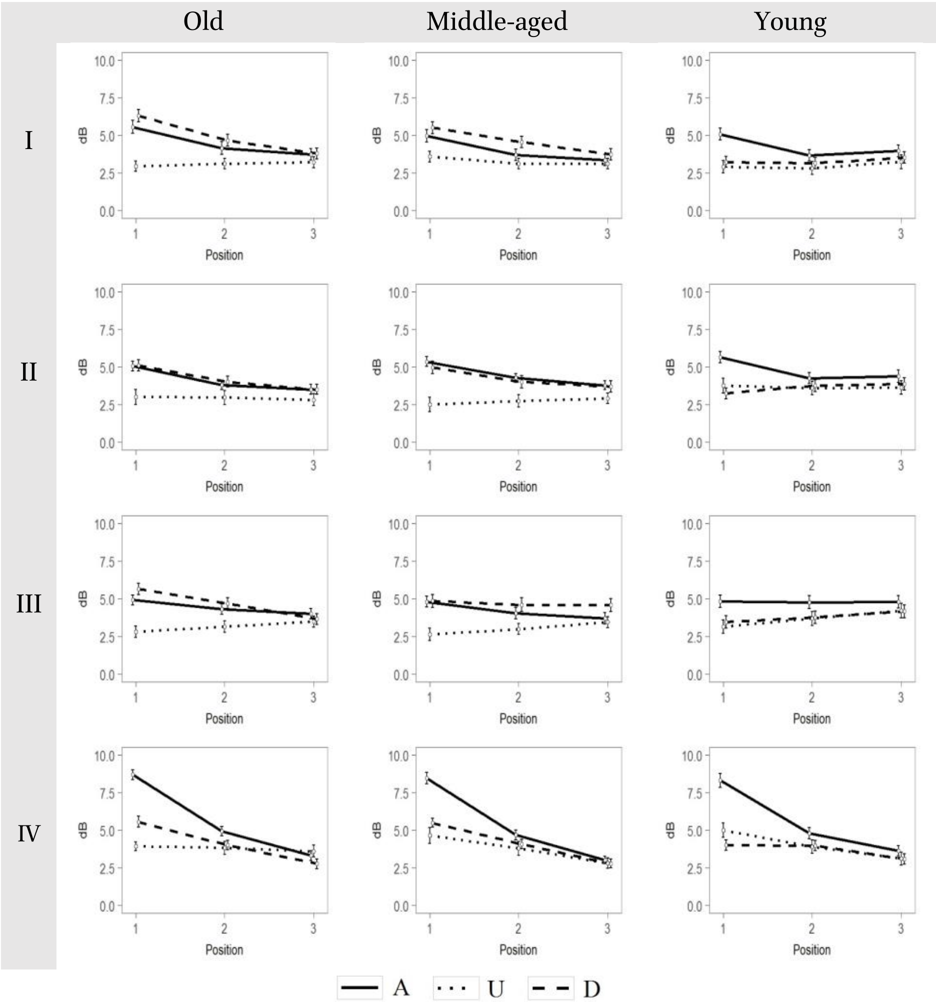



**Figure 3.4** Mean H1*–H2* of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each MC tonal category. Error bars represent the standard error of the mean.

**Table 3.7** Results ($\beta$) 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 | $.74^{n.s.}$ | $.51^{n.s.}$ | | II-D | $.04^{n.s.}$ | | |
| Middle-aged | I-U | $-1.37^{<.01}$ | $-0.59^{n.s.}$ | – | II-U | $-2.85^{<.05}$ | $-1.52^{n.s}$ | – |
| | I-D | $.55^{n.s.}$ | $.87^{n.s.}$ | | II-D | $-0.39^{n.s.}$ | $-0.25^{n.s.}$ | |
| 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 | $-2.12^{<.001}$ | $-1.16^{n.s.}$ | – | IV-U | $-4.75^{<.001}$ | $-1.12^{n.s.}$ | – |
| | III-D | $.72^{n.s.}$ | $.38^{n.s.}$ | | IV-D | $-3.12^{<.001}$ | $-0.94^{n.s.}$ | |
| Middle-aged | III-U | $-2.16^{<.05}$ | $-1.06^{n.s.}$ | – | IV-U | $-3.81^{<.001}$ | – | – |
| | III-D | $.05^{n.s.}$ | $.55^{n.s.}$ | | 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).

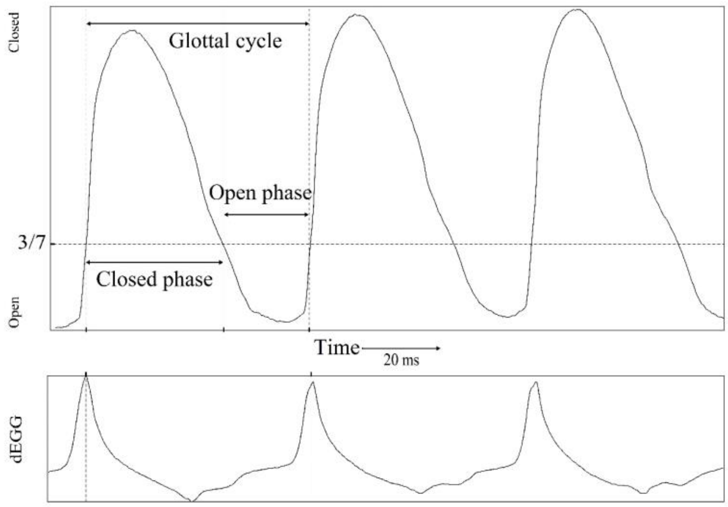



**Figure 3.5** EGG measurements (top) and the dEGG signal (bottom) exemplified by 20 ms from the beginning of /i̥/ in /ti̥¹/ 'low' produced by an old male speaker. The onset of each glottal cycle was determined by the steep rise (peak) of the corresponding dEGG signal. The threshold was set as 3/7 of 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 ($\beta$) 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 ($\chi^2$) 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 | $\chi^2$ | Two-way | $\chi^2$ | Three-way and four-way interactions | $\chi^2$ |
|---|---|---|---|---|---|
| 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 | $1.16^{n.s.}$ | | |

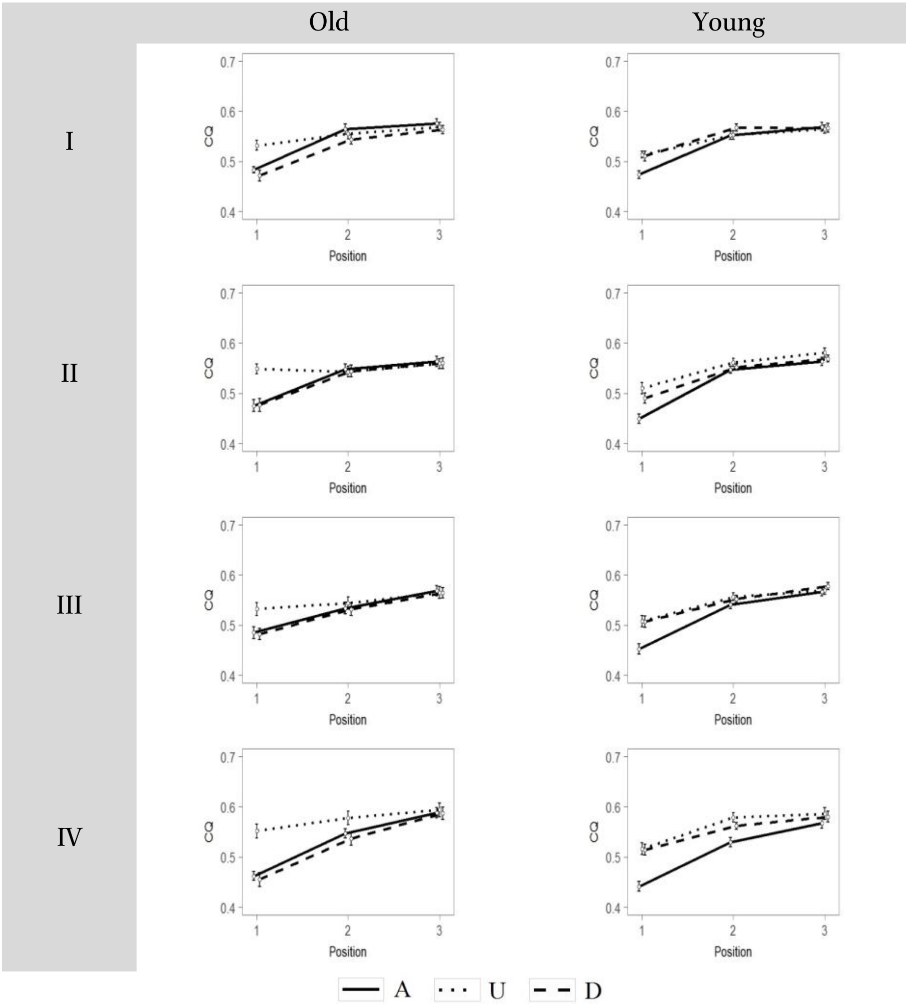



**Figure 3.6** Mean CQ of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each MC tonal category. The x-axis represents the ratio of the vowel. The y-axis represents the CQ value. Error bars represent 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 ($\beta$) 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 | $-0.01^{n.s.}$ | $-0.02^{n.s.}$ | $-0.01^{n.s.}$ | $-0.01^{n.s.}$ |

### 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.

| vs. | Voiceless aspirated | | Voiceless unaspirated | |
|---|---|---|---|---|
| | *Ping* | *Shang/Qu/Ru* | *Ping* | *Shang/Qu/Ru* |
| Voiced | VOT & *f0* | VOT | *f0* | |
| Voiceless aspirated | – | | VOT | VOT & *f0* |

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.

# Chapter 4 The sound system of Shuangfeng Xiang Chinese

## 4.1 Introduction

Shuangfeng Xiang Chinese (双峰方言) is a Xiang dialect (湘语, ISO 639-3; code: hsn) spoken in the town of Yongfeng (永丰镇). According to the census in 2000, there are approximately 88,000 residents in Yongfeng, which is the administrative town of Shuangfeng county (双峰县) (indicated by the solid circle in Figure 4.1). Shuangfeng county belongs to the prefectural-level municipality of Loudi city (娄底市) in Hunan province (湖南省), the People's Republic of China. It is located at the hinterland of Hunan province and is about 165 kilometers from to Changsha city (长沙市) (indicated by the hollow circle in Figure 4.1), the provincial capital of Hunan, about 1,500 kilometers to Beijing, the capital of China.

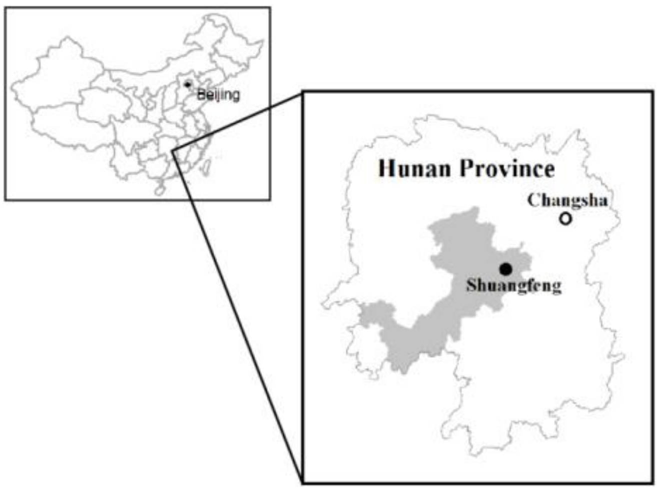



**Figure 4.1** Location of Shuangfeng county in Hunan province.

Shuangfeng Xiang Chinese is commonly considered part of the Xiangshuang cluster (湘双小片), which in turn is classified as a member of the Loushao subgroup (娄邵片) of the Xiang dialect group, a Sinitic branch (汉语支) within the Sino-Tibetan family (Bao & Chen, 2005). The Xiang dialect group is conventionally believed to be one of the seven major dialect groups in China (Yuan, 1960). As shown by the dark gray area in Figure 4.1, the Loushao subgroup dialect is also known as the Old Xiang Chinese (老湘语) in the majority of sinological studies, which is further differentiated from the Changsha dialect, known as one of the New Xiang Chinese (新湘语) (after Yuan, 1960: 102).[27] What distinguishes these two varieties is attributed mainly to their obstruent voicing contrasts. The Old Xiang group shows a three-way contrast (i.e., voiceless unaspirated vs. voiceless aspirated vs. voiced) while the New Xiang group has only a two-way obstruent contrast (i.e., voiceless unaspirated vs. voiceless aspirated) (Yuan, 1960; Yang, 1974; Tsuji, 1979; Bao & Yan, 1986; Wurm et al., 1987; Bao & Chen, 2005).

Shuangfeng Xiang Chinese has attracted much attention over the last six decades, which has resulted in a handful of descriptive works on not only Shuangfeng Xiang but also closely-related dialects in the Loushao subgroup which appear to have a similar three-way voicing contrast. Representative works include Yuan (1960), which provides the first introduction of the sound system of Shuangfeng Xiang Chinese in comparison to Changsha Xiang Chinese; Xiang (1960) on the first comprehensive description of Shuangfeng Xiang Chinese. *Hanyu fangyin zihui* (汉语方音字汇 [Lexicon of Chinese dialect pronunciations], *Zihui* hereafter) compiled by the Department of Chinese Language and Literature of Peking University (1989) provides a comprehensive reference work on the pronunciations of Chinese characters in Shuangfeng Xiang Chinese, with recordings of 2,961 monosyllabic words. Zhou (2005) and

[27] In some studies, the Old Xiang Chinese is also labeled as Southern Xiang Chinese and the New Xiang Chinese as Northern Xiang Chinese (e.g., Zhou & You, 1985).

Yuan (2005) are two MA theses, of which the first illustrates the historical sound change of the dialect, and of which the latter concentrates on dialectal variations across Shuangfeng county. Chen (2006) describes the typological features of the sound system of the Shuangfeng dialect in comparison with other Xiang dialects. Worth noting is Yang (1974), which reports the sound system of dialects spoken in 75 cities/counties of Hunan province. Unfortunately, Shuangfeng is not included, but the word list, together with its auditory recordings, provides a useful resource for comparative research on Shuangfeng Xiang within the Xiang dialect family.

It is important to note that the aforementioned studies tend to analyze the sound system of the Shuangfeng dialect based on auditory impression, and are prone to explore the synchronic system or variations from a historical perspective. Consequently, the existing literature remains obscure to researchers with limited knowledge of the research conventions within sinology.

This description is accompanied by recordings of a sixty-five-year-old female native speaker, who was born in 1953 and was raised in Yongfeng town. She spent most of her life living in Yongfeng and speaking Shuangfeng Xiang Chinese, except for five years when she was engaged in farming at a nearby village, where the local dialect shared a high level of intelligibility with Shuangfeng Xiang. According to her self-report, before her retirement, she sometimes spoke accented Standard Chinese with colleagues and customers who cannot understand Shuangfeng Xiang Chinese. In addition, she is also able to speak some Xiangtan Xiang Chinese (湘潭方言) (i.e., a Xiang variety spoken in the city center of Xiangtan) when the situation requires her to do so (e.g., in conversations with relatives who speak Xiangtan Chinese only).[28] Her primary language

[28] Xiangtan Xiang Chinese is classified as a dialect of the Changyi group (长益片) of the Xiang Chinese (Bao & Chen, 2005).

used at work and home is otherwise exclusively Shuangfeng Xiang Chinese.

## 4.2 Lexical tones

Figure 4.2 illustrates the five *f0* contours of the lexical tones in Shuangfeng Xiang Chinese, with each contour averaged over 20 words. The choice of these words was made by taking into consideration their segmental compositions; they constitute 20 near-minimal quintuplets. Example words are listed in Table 4.1. There are two level tones (i.e., T1 and T5), two rising tones (i.e., T2 and T4), and one falling tone (i.e., T3). Generally speaking, both T1 and T4 start in a higher *f0* range (above 180 Hz, high register hereafter), while both T2 and T5 start in a lower *f0* range (under 180 Hz, low register hereafter). Both T1 (black solid line) and T5 (gray long dashed line) are level tones, contrasting mainly in pitch height (high–level vs. low–level). T2 (gray solid line) and T4 (black long dashed line) are both rising, but differ in terms of not only their pitch registers but also their rising slopes. Specifically, T2 has a low-rising or low-dipping shape, whose pitch contour falls slightly within the lower pitch range initially, and then rises from the low to the mid-level (low–rising). Compared to T2, T4 starts from the mid-pitch range and stays mid till after the midpoint of the syllable and then rises sharply from the mid to the upper end of the speaker's pitch range (high–rising). T3 (black dotted line) is a falling tone, which falls from the upper end to the lower of the pitch range. Adopting the five-scale tonal descriptive system developed by Chao (1930), the five lexical tones can be broadly transcribed as /55/ (T1), /13/ (T2), /41/ (T3), /35/ (T4), and /22/ (T5).

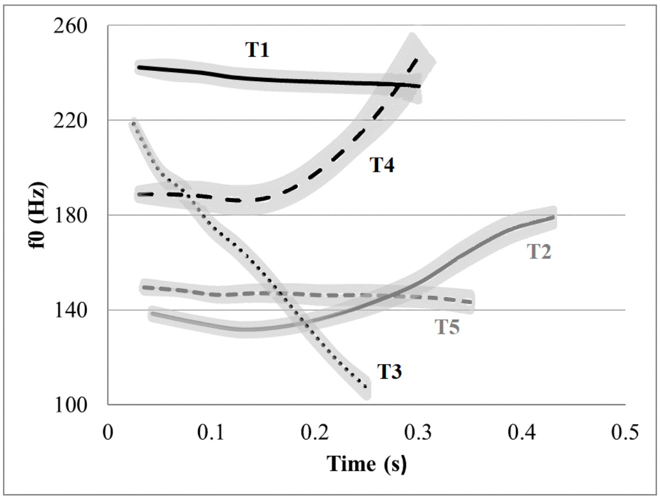



**Figure 4.2** *f0* contours of the lexical tones in Shuangfeng Xiang Chinese. For each lexical tone, mean *f0* (light gray areas indicate ± *SE*) of 10 equidistance points averaged over 20 words is plotted.

**Table 4.1** Examples of the lexical tones in Shuangfeng Xiang Chinese.

| Lexical tone | Tonal contour | Example | Orthography | Gloss |
|---|---|---|---|---|
| Tone 1 (T1) | high–level | /ti¹/ | 低 | 'low' |
| Tone 2 (T2) | low–rising | /di²/ | 啼 | 'to crow' |
| Tone 3 (T3) | falling | /ti³/ | 底 | 'bottom' |
| Tone 4 (T4) | high–rising | /ti⁴/ | 帝 | 'emperor' |
| Tone 5 (T5) | low–level | /di⁵/ | 地 | 'ground' |

There are some co-occurrence constraints on consonantal onsets and lexical tones, as shown in Table 4.2. In general, the five lexical tones can be divided into three subgroups, namely, i) T1 and T5, ii) T3 and T4, and iii) T2. The two level tones (i.e., T1 and T5) are in complementary distribution, where T1 can only co-occur with voiceless onsets (e.g., /pɘ¹/ 'bag'), while T5 occurs exclusively with voiced and sonorant onsets (e.g., /bɘ⁵/ 'to embrace'; /ljɘ⁵/ 'to predict'). T3 and T4 both can co-occur with voiceless and sonorant onsets, but not with voiced obstruent onsets. For example, T3 can appear in /pɘ³/ 'treasure', /pʰjɘ³/ 'to bleach', and /lɘ³/ 'old', while T4 in /pɘ⁴/ 'newspaper', /pʰɘ⁴/ 'bubble', and /mɘ⁴/ 'to emit'. T2 is a unique lexical tone in Shuangfeng Xiang Chinese, as there is no co-occurrence constraint as illustrated in /ti²/ 'target', /tʰi²/ 'to kick', /di²/ 'to mention', and /li²/ 'to leave'.

**Table 4.2** Co-occurrence constraints on onset-tone combinations in Shuangfeng Xiang Chinese. +: Yes; –: No.

| Onsets | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|
| Voiceless unaspirated | + | + | + | + | – |
| Voiceless aspirated | + | + | + | + | – |
| Voiced | – | + | – | – | + |
| Sonorant | – | + | + | + | + |

## 4.3 Consonants

Shuangfeng Xiang Chinese has 30 consonants. Corresponding key words/bound morphemes are provided below the consonant chart.

| | Bilabial | | | Labio-palatal | Alveolar | | | Postalveolar | | | Alveolo-palatal | | | Velar | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Plosive | pʰ | p | b | | tʰ | t | d | | | | | | | kʰ | k | g |
| Affricate | | | | | tsʰ | ts | dz | tʃʰ | tʃ | dʒ | tɕʰ | tɕ | dʑ | | | |
| Nasal | | | m | | | | n | | | | | | | | | ŋ |
| Fricative | | | | | s | | | ʃ | | | ɕ | | | x | | ɣ |
| Approximant | | | w | ɥ | | | | | | | | j | | | | |
| Lateral approximant | | | | | | | l | | | | | | | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pʰ | pʰə$^{1}$ | 抛 | 'to throw' | tʰ | tʰə$^{1}$ | 涛 | 'billow' | kʰ | kʰə$^{1}$ | 敲 | 'to knock' | | | | |
| p | pə$^{1}$ | 包 | 'bag' | t | tə$^{1}$ | 刀 | 'knife' | k | kə$^{1}$ | 高 | 'high' | | | | |
| b | bə$^{5}$ | 抱 | 'to embrace' | d | də$^{5}$ | 道 | 'path' | g | gen$^{2}$ | 乾 | 'heaven in eight trigrams' | | | | |
| tsʰ | tsʰə$^{1}$ | 操 | 'to hold' | tʃʰ | tʃʰɻ̩$^{1}$ | 痴 | 'infatuation' | tɕʰ | tɕʰə$^{3}$ | 悄 | 'quiet' | | | | |
| ts | tsə$^{1}$ | 糟 | 'terrible' | tʃ | tʃɻ̩$^{1}$ | 知 | 'to know' | tɕ | tɕə$^{1}$ | 焦 | 'scorched' | | | | |
| dz | dzə$^{2}$ | 曹 | 'surname, Cao' | dʒ | dʒɻ̩$^{2}$ | 池 | 'pond' | dʑ | dʑə$^{2}$ | 桥 | 'bridge' | | | | |
| m | mə$^{2}$ | 毛 | 'fur' | n | njə$^{5}$ | 尿 | 'urine' | ŋ | ŋə$^{2}$ | 熬 | 'to endure' | l | lə$^{2}$ | 劳 | 'labor' |
| s | sə$^{1}$ | 骚 | 'coquettish' | ʃ | ʃɻ̩$^{1}$ | 诗 | 'poem' | ɕ | ɕə$^{1}$ | 消 | 'to vanish' | x | xə$^{1}$ | 好 | 'good' |
| | | | | | | | | | | | | ɣ | ɣə$^{2}$ | 毫 | 'fine long hair' |
| w | wa$^{2}$ | 挖 | 'to dig' | ɥ | ɥa$^{2}$ | 曰 | 'to say, literary form' | j | ja$^{2}$ | 压 | 'to press' | | | | |

Generally speaking, Shuangfeng Xiang Chinese features a three-way contrast in plosives across three places of articulation (i.e., bilabial vs. alveolar vs. velar), known as voiceless unaspirated, voiceless aspirated, and voiced, respectively. Table 4.3 shows the mean VOT values and their standard deviation of the three-way contrast of plosives in three places of articulation. The measurements were obtained from 306 monosyllabic morphemes consisting of 102 sets of triplets. As shown in Table 4.3, irrespective of the place of articulation, the voiceless aspirated plosives have longer VOT values than the other two counterparts. The voiced plosives have negative VOT values. For further details, see below in the section on the three-way laryngeal contrast of obstruents (Section 4.4.1).

**Table 4.3** VOT of unaspirated vs. aspirated vs. voiced plosives in different places of articulation, based on 306 monosyllabic morphemes with plosive onsets.

| | Bilabial (35 triplets) | | Alveolar (64 triplets) | | Velar (3 triplets) | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Voiceless unaspirated | 14 ms | 7 ms | 15 ms | 5 ms | 23 ms | 8 ms |
| Voiceless aspirated | 85 ms | 17 ms | 105 ms | 21 ms | 107 ms | 14 ms |
| Voiced | –73 ms | 18 ms | –68 ms | 16 ms | –60 ms | 2 ms |

Identically to the plosives, there is also a three-way contrast in affricates across three places of articulation. /tsʰ ʦ ʣ/ are alveolars showing a more dental contact area than those in Standard Chinese. Alveolar-palatals /ʨʰ ʨ ʥ/ have a contact area of the alveolar ridge and the forward part of the palatal region; postalveolars consisting of /tʃʰ tʃ dʒ/. The postalveolar affricates in Shuangfeng Xiang Chinese are similar to those in both Beijing (Lee & Zee, 2003) and Tianjin Mandarin (Li et al., 2019). All are produced with the tongue tip raised against the postalveolar region and tend to be apical. More work on palatograms and linguograms can be informative to confirm the impressionistic observations.

Shuangfeng Xiang Chinese has five fricatives /s ɕ ʃ x ɣ/. /s ɕ ʃ/ have a slightly more anterior contact area than their corresponding affricates, respectively. Acoustically speaking, the center of gravity (COG) is argued to be an important indication of differentiating fricatives' places of articulation (Gordon et al., 2002). Therefore, COG values of ten minimal pairs of /s/ vs. /ɕ/ as well as ten additional morphemes which consist of the /ʃʅ/ syllable with various lexical tones were measured. As the stimuli contain different vowels, in order to exclude the coarticulation effect of the vowel, the mean of COG was measured within the first 30 ms interval from the beginning of each fricative. The results are plotted in Figure 4.3. The mean of /s/ is the highest (4999 Hz) while the mean of /ʃ/ the lowest (3198 Hz). A one-way ANOVA revealed a main effect for Place [$F(2, 27) = 56.42$, $p < .001$]. The results of *post hoc* comparisons (with Tukey HSD) suggested significant differences among the three places of articulation. It has been known that COG correlates well with the front cavity length: the lower COG, the larger front cavity (Ladefoged & Maddieson, 1996: 163). The results thus echo well with findings reported for /s ʃ/ in English (Jongman et al., 2000) and for /s ɕ ʃ/ in Standard Chinese (Svantesson, 1986; Li, 2008) and Swedish (Lindblad, 1980).[29]

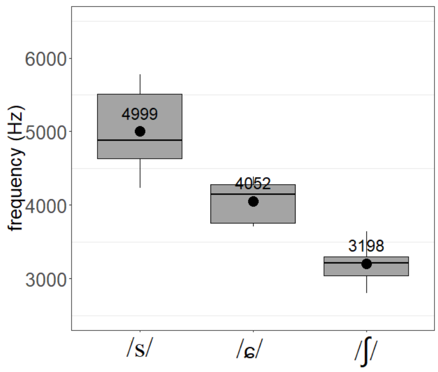



**Figure 4.3** Mean COG of /s/ vs. /ɕ/ vs. /ʃ/ measured within the first 30 ms interval from the beginning of each fricative.

[29] /ʃ/ is transcribed as /ʂ/ in both studies of Standard Chinese.

It is important to note that there is a two-way voicing contrast in velar fricatives but no voicing contrast in fricatives with other places of articulation. Such asymmetry has been proposed to be triggered by a parallel merger between voiced fricatives and voiced affricates, where the earlier non-velar /*z *ʑ *ʒ/ fricative onsets had merged with their affricate counterparts /dz dʑ ʤ/ (Xiang, 1960).

In a lot of Chinese dialects, a phonemic distinction between labiodental fricatives and velar/glottal fricatives can be recognized, such as /f/ vs. /x/ in Standard Chinese (Lee & Zee, 2003) or /v/ vs. /ɦ/ in Shanghai Wu Chinese (Chen & Gussenhoven, 2015).[30] However, in Shuangfeng Xiang Chinese, such a contrast in place of articulation has been lost. Words with the labiodental fricatives /f v/ have merged with their velar counterparts (i.e., /x ɣ/). For example, both 'wrong' and 'brightness' are pronounced as /xwi$^1$/ in Shuangfeng Xiang Chinese, while in Standard Chinese, they are /fei$^1$/ and /xwei$^1$/, respectively. In addition, /x ɣ/ are pronounced as glottal fricatives followed by a non-high vowel, as illustrated in /xɔn$^1$/ [χɔ̃$^1$] 'desolate' and /ɣan$^5$/ [ʁãn$^5$] 'slit'.

The realization of the voiced fricative /ɣ/ in Shuangfeng Xiang Chinese is contingent on the contexts. When /ɣ/ is pronounced in initial position, it is seldom realized as fully voiced. In most cases, the voicing is only partly observed, as shown in /ɣan$^5$/ 'slit', /ɣan$^5$/ (one morpheme of) 'phoenix', and /ɣjan$^5$/ 'lucky' in Figure 4.4, where the proportions of the voicing over the entire fricative vary from more to less to almost none. However, when /ɣ/ appears in non-initial position, voicing is consistently observed throughout the entire fricative sound. As shown in Figure 4.5, in initial position, /ɣ/ in the syllable /ɣjɔn$^2$/ 'frequent' fails to show voicing; while in non-initial position as in /t$^h$an$^1$ ɣjɔn$^2$/ 'usual', /ɣ/ is realized as a completely voiced sound.

[30] In some modern varieties of Wu Chinese, /ɦ/ is a phonetic segment only co-occurring with mid and open vowels and tends to disappear before high vowels (see Chen & Gussenhoven, 2015 and Section 2.8 for more information).

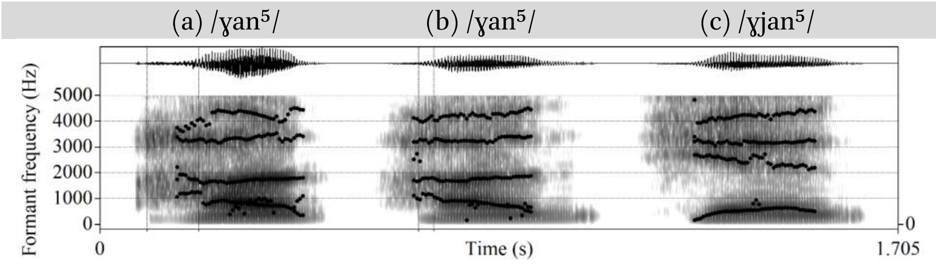



**Figure 4.4** Waveforms and spectrograms of (a) /ɣan$^5$/ 'slit', (b) /ɣan$^5$/ (one morpheme of) 'phoenix', and (c) /ɣjan$^5$/ 'lucky'. Intervals indicate the voicing over the entire fricative.

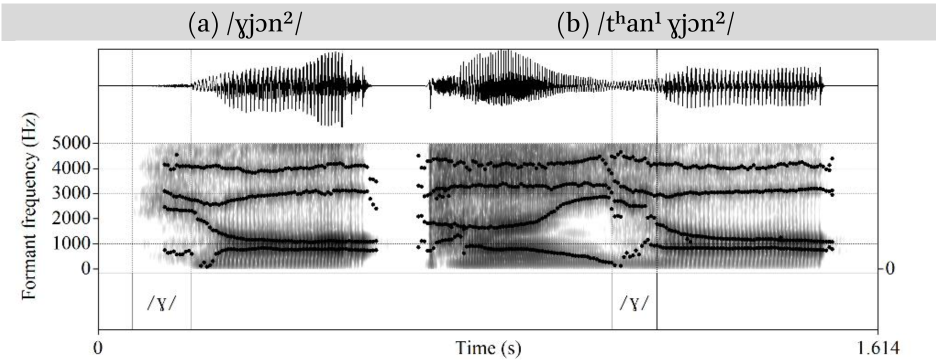



**Figure 4.5** Waveforms and spectrograms of (a) /ɣjɔn$^2$/ 'frequent' and (b) /t$^h$an$^1$ ɣjɔn$^2$/ 'usual'.

## 4.4 The three-way laryngeal contrast of obstruents

### 4.4.1 Phonetic realization

The three-way laryngeal contrast of obstruents in varieties of the Old Xiang group is commonly regarded as a crucial feature that makes the Xiang dialect group related to the Wu Chinese dialect group (e.g., Ting, 1982), the majority of which also have a three-way laryngeal contrast (after Chao, 1928). The phonetic manifestation of the three-way contrast in Shuangfeng Xiang Chinese, however, is very different from that observed in the Wu dialects. For example, in Shanghai Wu and Lili Wu Chinese, the three-way contrast in initial position cannot be distinguished via VOT. Instead, phonation has been argued to better distinguish from clearly modal (voiceless unaspirated), aspirated (voiceless aspirated), to breathy (voiced) (see Chen, 2011; Chen & Gussenhoven, 2015; Section 2.3 and references therein for further details). In Shuangfeng Xiang Chinese, voiced obstruents tend to be fully voiced, leading to a full-fledged distinction in terms of VOT, namely short-lag VOT (voiceless unaspirated), long-lag VOT (voiceless aspirated), and lead VOT (voiced). This is exemplified by the triplet /ti$^2$/ 'target', /t$^h$i$^2$/ 'to kick', and /di$^2$/ 'title' in Figure 4.6, respectively.

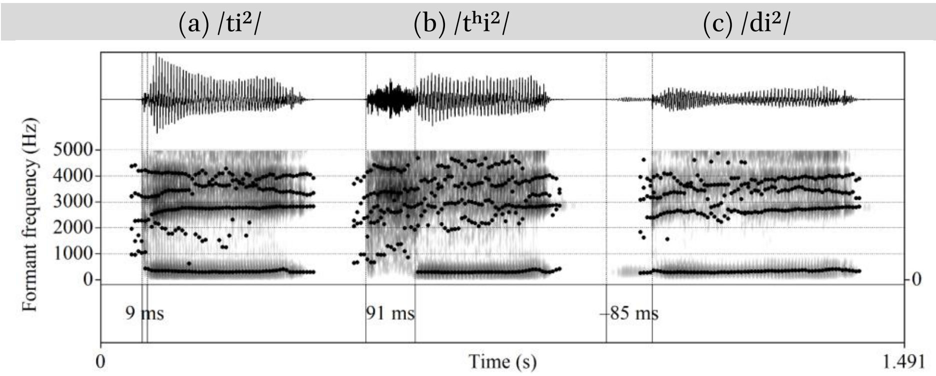



**Figure 4.6** Waveforms and spectrograms of (a) /ti$^2$/ 'target', (b) /t$^h$i$^2$/ 'to kick', and (c) /di$^2$/ 'title'. Intervals indicate VOT values.

Worth noting is that the acoustic realization of the voiced category also teems with variation. It has been observed that a voiced consonant may be produced with a relatively short VOT lead (/dų$^2$/ [d̤əu$^2$] 'pathway', –33 ms) or even a short VOT lag (/dɘ$^2$/ 'to escape', 13 ms). Even for the same word, voicing can be optional. Figure 4.7 shows two tokens of the syllable /dja$^2$/ 'jar' produced by the consultant. There is a lead VOT of –77 ms in the first instance, but only a short-lag VOT of 9 ms in the second.

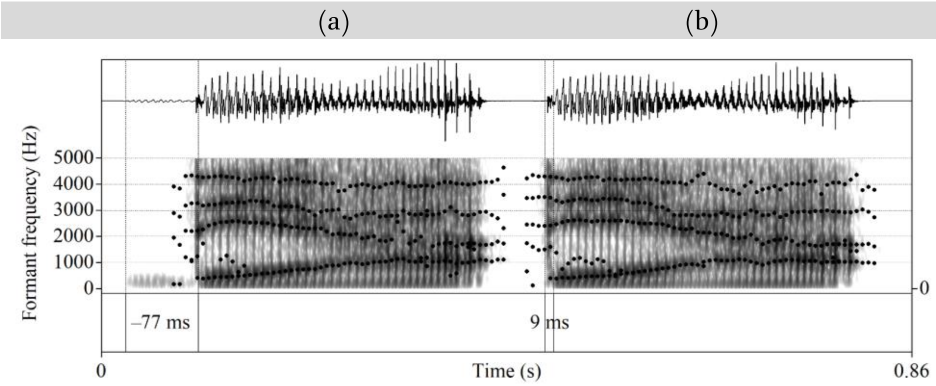



**Figure 4.7** Waveforms and spectrograms of two tokens of /dja²/ 'jar'. Intervals indicate VOT values. /d/ is realized with a lead VOT (–77 ms) in the first instance (a), but as a short-lag VOT (9 ms) in the second instance (b).

Voiced plosives in Shuangfeng Xiang may also be realized as implosives. As shown in Figure 4.8, /d/ is realized as a voiced plosive in /den²/ [d̪ẽĩ²] 'farmland', but an implosive [ɗ] in /djan⁵/ [ɗʝãn⁵] 'starch'. In the implosive [ɗ], the amplitude of vibrations gradually increases during the oral closure, in contrast with the voiced plosive [d], where one can identify a more typical biphasic and decreased amplitude during the closure (Lindau, 1984; Ladefoged & Maddieson, 1996). Such differences, to a large extent, can be attributed to the characteristic airflow during the articulatory closure in the production of obstruents. As explained by Ladefoged and Johnson (2010: 140), '[i]n the production of implosives, the downward moving larynx is not usually completely closed. The air in the lungs is still being pushed out, and some of it passes between the vocal cords, keeping them in motion.' Consequently, this particular airflow mechanism generates increasing amplitude visible in the waveforms. With regard to voiced plosives, the pulmonic airflow cannot continue through the glottis to maintain voicing. As more air flows to the mouth through the glottis, the pressure in the lungs returns to normal, and the vibration of vocal cords attenuates, resulting in a decrease of the amplitude over the closure time. Commonly, voiced implosives are regarded as allophones of voiced plosives in many languages, such as

Vietnamese and Khmer (Kirby, 2018), though implosives can also contrast with plosives in few languages as in Sindhi (Ladefoged & Johnson, 2010). In Shuangfeng Xiang Chinese, there is no phonemic contrast between voiced plosives and voiced implosives.[31]

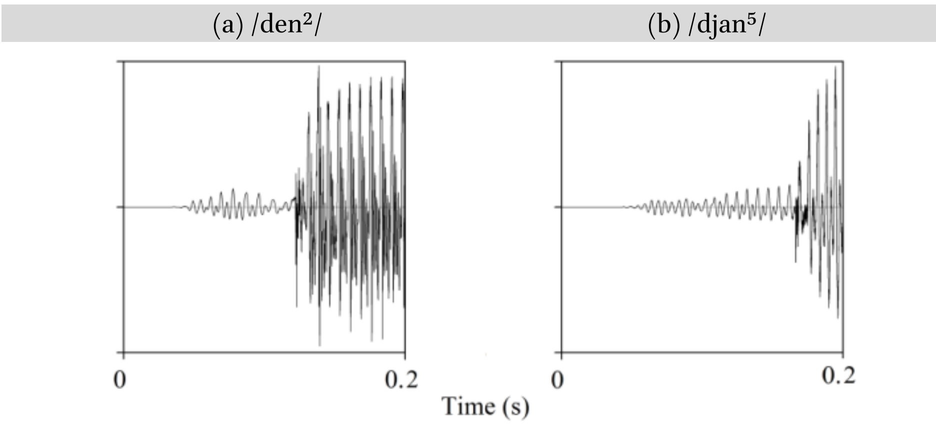



**Figure 4.8** Oscillograms of /d/ in (a) /den²/ 'farmland' and (b) /djan⁵/ 'starch'.

## 4.4.2 Historical development

It is important to note that the three-way contrast in modern Shuangfeng Xiang seems to be conditioned by the reconstructed MC tonal categories (e.g., Xiang, 1960; Zhou, 2005). Generally speaking, the contrast can be observed in syllables on the MC *Ping*, *Shang* and *Qu* tonal categories, but only a two-way obstruent contrast (i.e., voiceless unaspirated vs. voiceless aspirated) is maintained in syllables associated with the MC *Ru* tonal category. My corpus contains 899 high-frequency syllables selected from the Questionnaire of character for dialect surveys (*Fangyan diaocha zibiao*, 方言调查字表) with voiced obstruent onsets in MC. As indicated by

[31] As argued by Ladefoged and Maddieson (1996: 82), the relationship between voiced plosives and the so-called 'true' implosives is more gradient, rather than a clearly defined dichotomous distinction. This kind of relationship, to a certain extent, is reflected in diachronic changes, where the present-day implosives likely originated from earlier-time voiced plosives (Greenberg, 1970: 135; Cun, 2009; Ladefoged & Johnson, 2010: 143).

Table 4.4, none of the 149 syllables originating from the MC *Ru* tonal category is phonetically realized with voicing. With respect to the 750 syllables associated with the other three MC tonal categories (i.e., *Ping*, *Shang*, and *Qu*), the presence of the voice bar seems to be contingent upon the manner of articulation. Both plosives (48%) and affricates (50%) show higher percentages of the tokens with the voice bar than fricatives (30%). Compared to plosives, the higher devoicing-proclivity of voiced fricatives can be explained by the problematic mechanism of its articulation (Stevens, 1971; Ohala, 1983; Smith, 1997). The production of voiced fricatives is aerodynamically challenging, due to the opposing efforts required. On the one hand, a greater subglottal pressure with a weaker airflow should be kept for the maintenance of vibration of the vocal folds; on the other hand, a sufficient airflow above the glottis is required in order to generate frication noise. It is worth noting that typically partial voicing exists only for some voiced fricatives which do show the voice bar. In this description, I assume the voicing-present case (i.e., phonetically voiced) as long as the voice bar presents in the duration of the frication.

**Table 4.4** Percentage of the voiced category with the presence of the voice bar based on MC tonal categories.

| Voice bar | MC *Ru* (0/149) | | | MC *Ping*, *Shang*, and *Qu* (315/750) | | |
|---|---|---|---|---|---|---|
| | plosive | affricate | fricative | plosive | affricate | fricative |
| Presence | 0 | 0 | 0 | 48% | 50% | 30% |
| | (0/33) | (0/440) | (0/76) | (132/275) | (97/193) | (86/282) |

Interestingly, when taking a closer look at those items without a voice bar, two patterns can be observed. As shown in Table 4.5, in the MC *Ru* tonal category, 79% of the MC voiced plosives or affricates are produced as voiceless aspirated counterparts, while 21% become voiceless unaspirated counterparts, but their place of articulation remains the same. By contrast, plosives or affricates associated with the MC *Ping*, *Shang*, and *Qu* tonal categories are more likely produced as their voiceless unaspirated counterparts (unaspirated 92% vs. aspirated 21%) while maintaining their place of articulation.

**Table 4.5** Percentage of the voiced category without the presence of the voice bar based on the MC tonal categories.

| Voice bar | MC *Ru* (149/149) | | | MC *Ping*, *Shang*, and *Qu* (435/750) | | |
|---|---|---|---|---|---|---|
| | plosive & affricate | | fricative | plosive & affricate | | fricative |
| | aspirated | unaspirated | | aspirated | unaspirated | |
| Absence | 79% | 21% | 100% | 8% | 92% | 100% |
| | (58/73) | (15/73) | (76/76) | (18/239) | (221/239) | (196/196) |

Given these observations, it may conclude that in Shuangfeng Xiang Chinese, the three-way obstruent contrast is no longer a consistent feature but shows strong instability. First, voiced obstruents are realized with multiple variants. Furthermore, the three-way obstruent contrast seems to be largely conditioned by MC tonal categories. Onsets with a lexical tone belonging to the MC *Ru* tonal category, due to the ubiquitous loss of the voicing realization, can be regarded to show a two-way contrast only (i.e., voiceless unaspirated vs. voiceless aspirated). Onsets with a lexical tone belonging to non-*Ru* tonal categories (i.e., *Ping*, *Shang*, and *Qu*), to a large extent, still maintain a full-fledged three-way contrast (i.e. voiceless unaspirated vs. voiceless aspirated vs. voiced), although various realizations of the voiced category also exist. A similar phenomenon has also been reported for neighboring dialects of Shuangfeng Xiang Chinese, such as Xiangxiang (湘乡) and Loudi (娄底) (Chen, 2006: 27). Needless to say, further research, especially with data elicited from a large sample of speakers of different generations, are needed for a more precise description of their realizations and distributions.

## 4.5 Sonorants

All nasals (i.e., /m n ŋ/) can occur in onset position. The alveolar nasal /n/ is palatalized before high front segments (i.e., /i j ʏ/) such as in /ni$^2$/ [ɲi$^2$]

'Buddhist nun'. Sometimes, /ŋ/ tends to give rise to a strong nasalized quality over the following vowel due to progressive nasal assimilation, as shown in /ŋo²/ [ŋõ²] 'tooth'. /m n/ can also form syllable nuclei ( e.g., /m̩¹ mjaᴺ/ 'mother' and /n̩³/ 'you').

/n l/ contrast only before three high segments (i.e., /i ʏ j/), as in /ni²/ [ɲi²] 'muddy' vs. /li²/ 'to leave'; /nʏ³/ [ɲʏ³] 'female' vs. /lʏ³/ 'to travel'; and /njɔn²/ [ɲjɔ̃²] 'a form of address for an elderly married woman' vs. /ljɔn²/ [ljɔ̃²] 'grain'. Words distinguished by /n/ (e.g., /nau³/ 'brain') and /l/ (e.g., /lau³/ 'old') in Standard Chinese therefore become homophones /lɘ³/ in Shuangfeng Xiang Chinese. Due to this merger, the consultant failed to distinguish /n/ and /l/ in the non-/i j ʏ/ conditions. One observed exception is the loss of /n l/ contrast before /ja/. For example, 'difficulty' (/*n/) and 'blue' (/*l/) are homophones (i.e., /lja²/). This is due to the fact that these syllables (with /ja/ in modern Shuangfeng Xiang) have likely developed from the earlier /ã/ (Xiang, 1960) or /æ̃/ (*Zihui*, 1989), without the glide /j/.[32] Diachronically speaking, it is likely that the glide /j/ in /ja/ is an epenthetic segment that emerged after the merger of /n/ and /l/ before non-high-frontal segments, and therefore, was exempted from the condition which contributed to the merger of /n/ and /l/. The relative chronology of the observed exceptions can be delineated as the following: /*nã/ (e.g., 'difficulty') and /*lã/ (e.g., 'blue') first merged to /*lã/, and then became /lja/.

Typologically speaking, the lack of distinction between /n/ and /l/ seems to be rarely reported in the world's languages. However, such a merger has been commonly observed in Jianghuai and Southwest Mandarin (see Chen & Guo, forthcoming and references therein) as well as the Min and Xiang varieties (Xiang, 1960) (see Huang, 2007 for further

[32] In both studies, researchers also point out that /ã/ (/æ̃/) and /ja/ (/jɛ/) are in free variation for some speakers. My current consultant, however, tends to consistently produce all /ã/ as /ja/.

details on the different patterns of the /n l/ merger across Chinese dialects).

/n/ can serve as a coda which varies in its phonetic realization due to the articulatory height of the vowel. After /a/ and /ɜ/, /n/ is realized as a clear coda as in /tan$^1$/ [t̪ãn$^1$] 'east' and /dwɜn$^1$/ [d̪wɜ̃n$^2$] 'suckling pig'. By contrast, after an open-mid vowel /ɔ/, /n/ is not realized as a clear coda but triggers vowel nasalization (e.g., /tɔn$^1$/ [t̪ɔ̃$^1$] 'when'). Perhaps most prominently, when /n/ follows the closer vowel /e/ (e.g., /ten$^1$/ 'jolt'), the alveolar nasal seems to introduce no complete oral closure near the alveolar ridge and is realized as [t̪ẽĩ$^1$] where a nasalized [ĩ] element can be heard.

## 4.6 Approximants

Shuangfeng Xiang Chinese has three approximants in onset position, as indicated by the triplet /ja$^2$/ 'to press', /wa$^2$/ 'frog', and /ɥa$^2$/ 'to say, the form in classical Chinese'. All three approximants can serve as a glide, as in /tja$^1$/ 'to delay', /twa$^1$/ 'to hold something level with both hands', and /ɕɥe$^2$/ 'snow'. It is important to note that, sometimes, when /j/ occurs as a glide, compared to Standard Chinese, its quality appears to be characterized by a more rapid transition to the following vowel. Xiang (1960) argues that syllables containing the glide /j/ may have a closer relationship to the onset, such that /tja$^1$/ 'to delay' is realized as [tʲa$^1$].[33] However, I hasten to note that in the consultant, a very clear /j/ in /tja$^1$/ 'to delay' has been observed. After the alveolo-palatals (i.e. /tɕʰ tɕ dʑ ɕ/), only a brief transition from the alveolo-palatal onset to the following vowel (/ɕɘ$^1$/ 'to vanish') was observed, which suggests /j/-quality. However, given three reasons, I have adopted the analysis of Chen and Gussenhoven (2015) for Shanghainese and the analysis for Lili Wu Chinese (Section 2.4) and posited no underlying /j/ after alveolo-palatal

[33] In Xiang's proposal (1960), the glide /j/ is transcribed as /i/.

onsets. The first reason is the briefness of such transitive /j/ quality, which is much less than that in /tja$^{1}$/. The second is due to the predictability of such transitive /j/ quality after alveolo-palatal onsets. The third reason is the lack of contrasts like /ɕə/ vs. /ɕjə/ in Shuangfeng Xiang Chinese.

In Shuangfeng Xiang Chinese, there exist two syllabic approximants /ɹ̩ ɻ̩/ as in /tsɹ̩$^{3}$/ [t̪s̪ɹ̪̩$^{3}$] 'purple' and /tʃɻ̩$^{4}$/ 'to make'. The production of /ɹ̩ ɻ̩/ in Shuangfeng Xiang Chinese is similar to that in Standard Chinese. /ɹ̩/ only occurs after alveolar consonants /ts$^{h}$ ts dz s/ and /ɻ̩/ after postalveolars /tʃ$^{h}$ tʃ dʒ ʃ/. The tongue configuration of /ɹ̩/ is almost homorganic with the preceding consonant and therefore tends to be dental. For more information on both syllabic approximants, interested readers are referred to Lee-Kim (2014).

## 4.7 Vowels

Vowels in open syllables are plotted in Figure 4.9(a), and those in closed syllables with a nasal coda in Figure 4.9(b). The vowel ellipses were calculated based on 50 tokens of each vowel except for /ʮ/ (18) and /ʏ/ (30). The measurements of formant values are identical to Section 2.5. Example keywords/bound morphemes are provided below the plots. In open syllables, there are eight monophthongs (/i ʏ u ʮ e ɘ o a/) in Shuangfeng Xiang Chinese. These eight vowels constitute a four-way distinction (i.e., close, near-close, close-mid, and open) in height and a three-way distinction (i.e., front, central, and back) in backness. Compared to the vowels in open syllables, the number of vowels in closed syllables is reduced, and so is their acoustic vowel space. Phonemically, there are four vowels (/e ɜ ɔ a/) contrasting in height and backness.

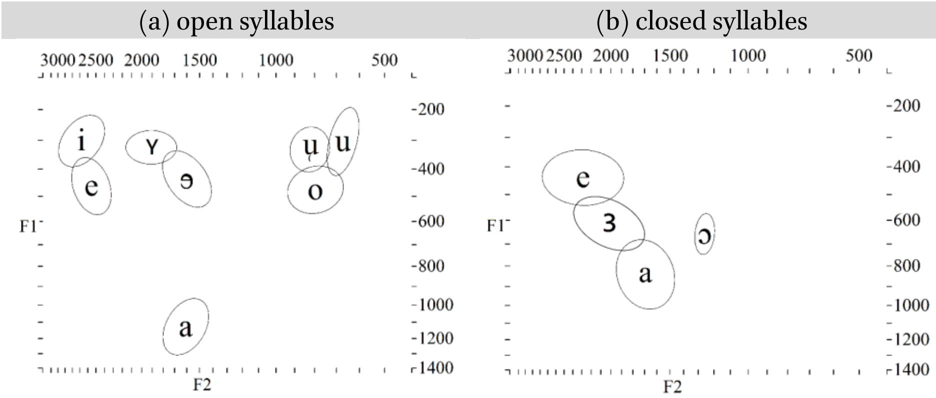



**Figure 4.9** Relative F1/F2 formant values of monophthongs produced in (a) open syllables and in (b) closed syllables.

Monophthongs in open syllables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| i | bi$^{2}$ | 皮 | 'skin' | u | bu$^{2}$ | 婆 | 'old woman' |
| ʏ | ʏ$^{2}$ | 鱼 | 'fish' | ṳ | bṳ$^{2}$ | 蒲 | 'calamus' |
| e | be$^{5}$ | 陪 | 'to accompany' | o | bo$^{2}$ | 爬 | 'to creep' |
| ɘ | bɘ$^{2}$ | 袍 | 'robe' | | | | |
| a | ba$^{2}$ | 牌 | 'card' | | | | |

Monophthongs in closed syllables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| e | ten$^{1}$ | 颠 | 'jolt' | ɜ | tjɜn$^{1}$ | 真 | 'real' |
| a | tan$^{1}$ | 东 | 'east' | ɔ | tɔn$^{1}$ | 当 | 'when' |

/ɘ/ is commonly transcribed as /ɤ/ (e.g., Xiang, 1960). In the consultant's productions, however, this vowel tends to be much more central. Hence, the IPA symbol /ɘ/ has been opted for.

/ʏ/ is consistently transcribed as /y/ in all previous studies. However, it is worth noting that there are remarkable differences between /ʏ/ in Shuangfeng Xiang Chinese and /y/ in Standard Chinese, as shown in Figure 4.10, where /y$^{2}$/ 'fish' in Standard Chinese (4.10a) and /ʏ$^{2}$/ 'fish' in Shuangfeng Xiang Chinese (4.10b) are plotted. Compared to /y$^{2}$/ in Standard Chinese, /ʏ$^{2}$/ in Shuangfeng Xiang Chinese has much lower formant values (F2: 2505 Hz vs. 1995 Hz; F3: 3088 Hz vs. 2708 Hz; F4: 4350 Hz vs. 3549 Hz), indicating an articulatory retraction. In this current study,

/ʏ/ instead of the widely-used /y/ is posited accordingly. Moreover, when /ʏ/ follows /tʃʰ tʃ dʒ/, due to coarticulation, the tongue body position of /ʏ/ as in /dʒʏ$^{2}$/ [dʒʏ̄$^{2}$] 'kitchen' is slightly further back than that in /ʏ$^{2}$/ 'fish', indicated by a lower F2 in /dʒʏ$^{2}$/ (1808 Hz, compared to 1995 Hz in /ʏ$^{2}$/).

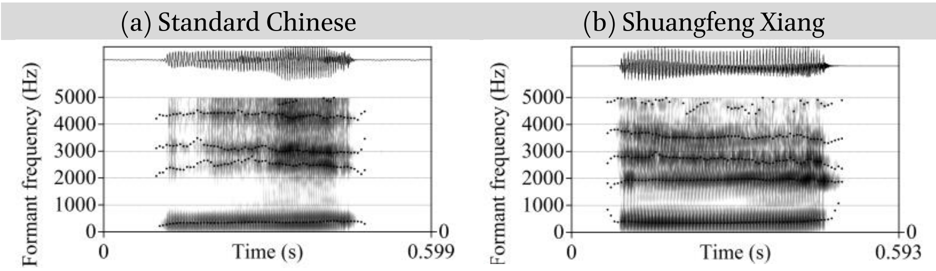



**Figure 4.10** Waveforms and spectrograms of 'fish' in (a) Standard Chinese and (b) Shuangfeng Xiang Chinese. The sound file of Standard Chinese was adopted from Lee and Zee (2003).

Shuangfeng Xiang Chinese has an interesting and rarely observed three-way contrast in high back vowels, as illustrated by the triplet of /bo$^{2}$/ 'to climb' vs. /bu$^{2}$/ 'old woman' vs. /bu̜$^{2}$/ 'calamus'. As shown in Figure 4.11, /o/ is produced with higher F1 and lower F2 values. The difference between /u/ and /u̜/ is mainly in F2 with /u/ having lower values than /u̜/. What is more salient is the lip gesture differences among these three back vowels as shown in Figure 4.12. The lip aperture is markedly smaller for the higher vowel /u/ than for the lower vowel /o/. However, both /o/ and /u/ differ from /u̜/ in having a more rounded and protruding lip constriction. /u̜/, in contrast, is produced with greater lip compression and less protuberance. The slightly higher F2 of /u̜/ probably results from the compressed lip gesture which shortens the length of the vocal tract accordingly. In some languages, vertical compression and horizontal protrusion are two lip-position parameters for vowel contrasts (Ladefoged & Maddieson, 1996: 295). For example, /y:/ is argued to differ from /ʉ:/ mainly with a more protruded lip position in Swedish (See Fant, 1983 and references therein). In Japanese, /u/ is usually phonetically transcribed as [ɯ] due to its unrounded and compressed lips (Okada, 1991). Compared to [ɯ] in Japanese, /u̜/ in Shuangfeng Xiang Chinese tends to be produced with similarly compressed lips. Given such a similarity, Zeng (2019) hence adopts /ɯ/ to transcribe this phoneme in Xiangxiang Xiang, an Old Xiang variety closely related to Shuangfeng Xiang. According to the results in Keating and Huffman (1984), the ratio of F1 and F2 for [ɯ] pronounced in word lists is around 0.33 (400 Hz: 1200 Hz, based on recordings of 7 young male speakers). However, the ratio of /u̜/ produced by the consultant is around 0.41 (350 Hz: 850 Hz), which suggests a different vowel quality to [ɯ] in Japanese. It is quite close to the ratio of /u/ in American English (0.42, based on average formant frequencies of 48 female speakers presented in Hillenbrand et al., 1995). Given these reasons, /u̜/ instead of /ɯ/ with the symbol ◌̜ has been adopted to highlight the less rounded and compressed lips of /u̜/.

In addition, /u̜/ is realized as a monophthong after bilabial onsets only, but with diphthong quality [əu] after other onsets or appearing without an onset, e.g., /du̜$^2$/ [dəu$^2$] 'pathway', /ku̜$^1$/ [kəu$^1$] 'alone', and /u̜$^1$/ [əu$^1$] 'to pollute'. Such a complementary realization of /u̜/ has also been found in some Northern Wu Chinese, such as Shanghainese (Chen & Gussenhoven, 2015) and Lili Wu Chinese (Section 2.5). In contrast, /u/ can appear as a monophthong not only after bilabial onsets, but also after other places of articulation, such as /tu$^1$/ 'many', /t͡su$^3$/ 'left', /ku$^1$/ 'song', and /ɣu$^2$/ 'river'.

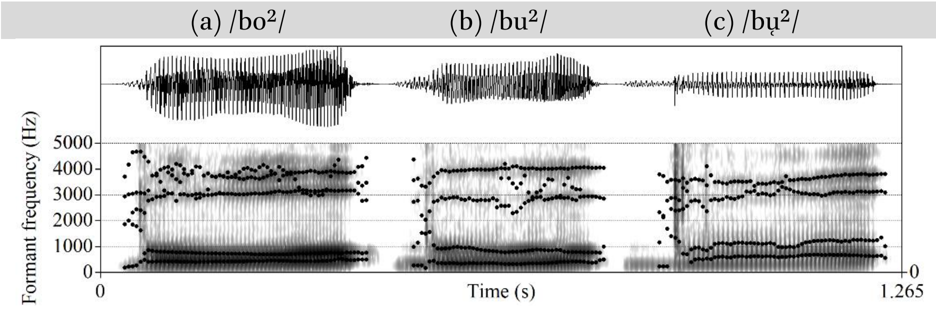



**Figure 4.11** Waveforms and spectrograms of (a) /bo$^{2}$/ 'to climb', (b) /bu$^{2}$/ 'old woman', and (c) /bu̜$^{2}$/ 'calamus'.

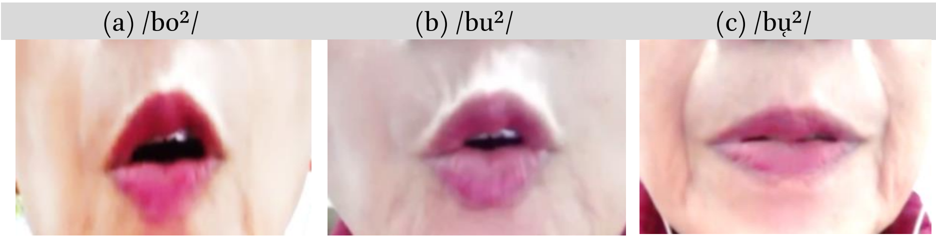



**Figure 4.12** Pictures of the maximal gesture of the lips in (a) /bo$^{2}$/ 'to climb', (b) /bu$^{2}$/ 'old woman', and (c) /bu̜$^{2}$/ 'calamus'.

There are four vowels that occur in closed syllables. Generally speaking, vowels in closed syllables are more central than those in open syllables. /ɔ/ is a back rounded open-mid vowel as in /tɔn$^{1}$/ [t̪ɔ̃$^{1}$] 'when'. /e/ is a close-mid unrounded front vowel as in /ten$^{1}$/ 'jolt' with a more salient diphthong quality, as discussed in Section 4.5 on sonorants. /a/ is a front low vowel as in /tan$^{1}$/ 'east'. /ɜ/ is a central open-mid vowel and is licensed after glides only. However, its actual realization is highly contingent upon the preceding glide due to coarticulation. As shown in Figure 4.13, it is pronounced as [ɛ] after the glide /j/, such as in /tjɜn$^{1}$/ [tjɛ̃n$^{1}$] 'real'; but [ɜ] after the more rounded glides (i.e., /w ɥ/), such as in /twɜn$^{1}$/ [twɜ̃n$^{1}$] 'army' and /ɕɥɜn$^{1}$/ [ɕɥɜ̃n$^{1}$] 'elder brother'. Each realization of /ɜ/ is plotted based on the measurements of ten real words produced by the consultant.

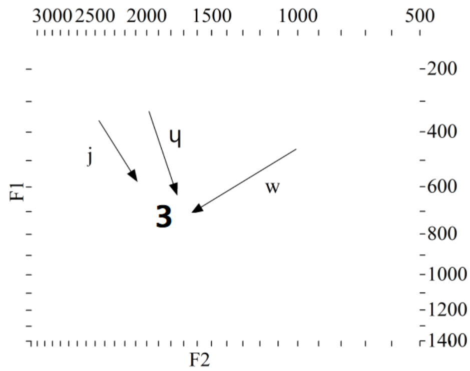



**Figure 4.13** Gliding trajectories of the phoneme /ɜ/ produced after glides /j/, /w/, and /ɥ/.

At last, /ja/ and /e/ occurring in open syllables are sometimes in free variation. For example, the morpheme 'north' can be pronounced either /pja$^{2}$/ or /pe$^{2}$/, and the morpheme 'to separate' can be either /kja$^{2}$/ or /ke$^{2}$/. As told by the consultant, the two pronunciations do not yield different lexical meanings.[34]

## 4.8 Syllable structure

Identical to a lot of Sinitic varieties, the canonical Shuangfeng syllable minimally consists of an obligatory nucleus (V) and a lexical tone as in /i$^{1}$/ 'clothes' and /u$^{1}$/ 'an old name of Japan'. V stands for vowel or syllabic consonant (i.e., /ɹ̩ ɻ̩ m̩ n̩/). It may contain up to three optional elements in

[34] It is worth pointing out that the exchangeable pronunciations of /ja/ and /e/ do not universally appear in Shuangfeng Xiang Chinese. /ja/ can be alternatively pronounced as /e/ only in syllables originating from the MC *Ru* tonal category diachronically, the majority of which became either /ei/ (e.g., /pei$^{3}$/ 'north') or /ɤ/ (e.g., /kɤ$^{2}$/ 'to separate') in Standard Chinese. The /e/ pronunciation of Shuangfeng Xiang probably emerged as an adaptation under the influence of Standard Chinese. Syllables with /ja/ but not developing from the MC *Ru* tonal category, which did not become /ei/ or /ɤ/ in Standard Chinese, do not bear the /e/ alternation. For example, /pja$^{1}$/ 'class' developed from the MC *Ping* tonal category. It is pronounced /pan$^{1}$/ in Standard Chinese and does not have the alternative pronunciation */pe$^{1}$/.

the following linear sequence: (C1)(G)V(C2), where C1 can be any consonant in the consonant inventory; G is either /j/, as in /ja²/ 'to press', /w/, as in /wi¹/ 'mighty', or /ɥ/, as in /ɥɐn²/ 'cloud'; C2 contains one element only, namely /n/, as in /an¹/ 'surname, Weng'. All combinations are demonstrated in Table 4.6.

**Table 4.6** Syllabic combinations in Shuangfeng Xiang Chinese.

| Combination | Example | | | | | |
|---|---|---|---|---|---|---|
| V | /i¹/ | 衣 | 'clothes' | /u¹/ | 倭 | 'an old name of Japan' |
| GV | /wi¹/ | 威 | 'mighty' | /ja²/ | 压 | 'to press' |
| C1V | /ti¹/ | 低 | 'low' | /ta²/ | 搭 | 'to put over' |
| VC2 | /en¹/ | 烟 | 'smoke' | /an¹/ | 翁 | 'surname, Weng' |
| C1VC2 | /ten¹/ | 颠 | 'jolt' | /tan¹/ | 东 | 'east' |
| GVC2 | /wen¹/ | 冤 | 'wronged' | /ɥɜn²/ | 云 | 'cloud' |
| C1GVC2 | /twen¹/ | 专 | 'specialized' | /tjɜn¹/ | 真 | 'real' |

There are some constraints of onset-rhyme combinations in Shuangfeng Xiang Chinese, as illustrated in Table 4.7.

For open syllables, bilabials /pʰ p b m/ are not licensed before /ʏ/. Alveolars /tʰ t d n l/ have the widest distribution, which can co-occur with the majority of vowels (except for /ɹ̩ ɻ̩/). The syllabic approximants /ɹ̩ ɻ̩/ only follow after /tsʰ ts dz s/ and /tʃʰ tʃ dʒ ʃ/, respectively. /tsʰ ts dz s/ and /tɕʰ tɕ dʑ ɕ/ do not occur in the environments where postalveolars /tʃʰ tʃ dʒ ʃ/ occur. The distribution of /kʰ k g ŋ x ɣ/ is quite similar to that of /tsʰ ts dz s/, except that /kʰ k g ŋ x ɣ/ are not allowed before /ɹ̩/.

Both /i ʏ/ are high front vowels. The distribution of /ʏ/, however, is more restricted. It can form an onsetless syllable, as in /ʏ²/ 'fish' or co-occur with alveolar sonorants /n l/ (e.g., /nʏ³/ 'female' and /lʏ³/ 'to travel'). Interestingly, /i/ is forbidden after /tʃʰ tʃ dʒ ʃ/, while /ʏ/ is allowed (/dʒʏ²/ 'kitchen'). The three back vowels /u ʯ o/ behave similarly. However, compared to /u/, /o/ cannot combine with /ɣ/ or form an onsetless syllable, while /ʯ/ is impossible to appear after alveolo-palatals or the glide /j/. Both glides /w ɥ/ cannot be followed by central and back vowels (i.e., /u ɥ̩ o ɘ/) in open syllables. /j/, however, can precede the majority of them (except /ɥ̩/), as in /ju²/ 'to swim', /jo³/ 'also', and /jɘ²/ 'to shake'.

For closed syllables, /en an ɔn/ can combine with alveolars, velars, and form onsetless syllables as in /en¹/ 'smoke', /an¹/ 'I', and /ɔn¹/ 'surname, Weng'. /en/ and /ɔn/ can follow after /w/ and /j/ as in /wen²/ 'round' and /jɔn²/ 'central'. /ɔn/ after bilabials (i.e., /pʰ p b m/), while /en/ after alveolo-palatals (i.e., /tɕʰ tɕ dʑ ɕ/) is forbidden. Worth noting is that the distribution of /ɜn/ is highly restricted. As discussed in Section 4.7, it can only co-occur with glides.

**Table 4.7** Observed onset-rhyme combinations in Shuangfeng Xiang Chinese. +: Yes; –: No.

| Onset | e a | i | ɤ | u | ų | o | ɘ | ɹ̩ | ɻ̩ | en | an | ɔn | ɜn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bilabial (pʰ p b m) | + | + | – | + | +[35] | + | + | – | – | + | + | – | – |
| Alveolar (tʰ t d n l) | + | + | +[36] | + | + | + | + | – | – | + | + | + | – |
| Alveolar sibilants (tsʰ ts dz s) | + | – | – | + | + | + | + | + | – | + | + | + | – |
| Postalveolar (tʃʰ tʃ dʒ ʃ) | – | – | + | – | – | – | – | – | + | – | – | – | – |
| Alverlo-palatal (tɕʰ tɕ dʑ ɕ) | – | + | – | + | – | + | + | – | – | – | + | + | – |
| Velar (kʰ k g ŋ x ɣ) | + | – | – | + | + | +[37] | + | – | – | + | + | + | – |
| j | + | + | – | + | – | + | + | – | – | – | – | + | + |
| w | + | + | – | – | – | – | – | – | – | + | – | – | + |
| ɥ | + | – | – | – | – | – | – | – | – | – | – | – | + |
| Zero onset | – | + | + | + | + | – | – | – | – | + | + | + | – |

[35] /m/ are not included.

[36] Alveolar plosives /tʰ t d/ are excluded. In some proposals (e.g., Xiang, 1960), /ɤ/ (/y/ in their transcriptions) can also co-occur with alveolar plosives, as in /dɤ²/ 'kitchen'. However, my consultant tends to produce these plosives as affricates.

[37] /ɣo/ cannot be observed.

## 4.9 Onset pitch perturbations

It is worth emphasizing the fact that the actual pitch contours of lexical tones in Shuangfeng Xiang Chinese can vary significantly due to the onset perturbation effects (i.e., C*F*0).

Take T2 as an example. It has already been noted in several impressionistic studies that voiceless unaspirated onsets introduce a higher *f*0 contour compared to voiced onsets (Xiang, 1960; Yuan, 1960; *Zihui*, 1989). Data in this study also suggest perturbation effects after voiceless aspirated and nasal onsets, as shown in Figure 4.14. Here, the four pitch contours of T2 are introduced by four types of onsets, namely voiceless unaspirated obstruents (solid line, 1-VU), voiceless aspirated obstruents (dotted line, 2-VA), voiced obstruents (dashed line, 3-D), and sonorants (dash-dotted line, 4-SN). Each contour was averaged over 50 samples, 10 of which form minimal quartets of onsets. Given that not all voiced onsets are realized with periodic voicing during the closure (i.e., lead VOT, see Section 4.4), the *f*0 values in Hz were measured at twenty equidistant points over the duration of the following vowel for all syllables using VoiceSauce in the Straight method automatically. For obstruents, vowel onset was defined as the onset of the first periodic pattern following the burst in the acoustic waveform. For sonorants (i.e., nasal and lateral), vowel onset was defined as the point with an abrupt change of F2 in the spectrogram. Vowel offset was defined as the cessation of periodicity in the waveform. As we can see in Figure 4.14, the trajectory after voiceless unaspirated onsets has an overall higher *f*0 contour than that after voiced onsets, which confirms the impressionistic observation. The contours after both voiceless aspirated and voiced onsets, however, show quite some similarity. The only difference seems to be that voiced onsets tend to trigger a slightly lower beginning than the voiceless aspirated counterparts. Very interestingly, the trajectory after sonorants is much closer to that after voiceless unaspirated counterparts, rather than after voiced counterparts in terms of the overall pitch height. All four *f0* contours, although starting with different pitch heights, converge towards a rising contour over the last one-third of the syllable.

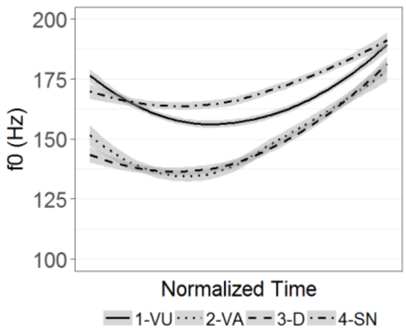



**Figure 4.14** Mean *f0* of T2 after different onsets. Each mean *f0* contour (light gray areas indicate ± *SE*) was averaged over 50 samples, 10 of which formed minimal quartets of onsets. VU: voiceless unaspirated; VA: voiceless aspirated; D: voiced; SN: sonorants.

A distinctive pitch contour characteristic of the lexical tone is consistently observed in other lexical tonal categories. Figure 4.15 shows onset perturbation effects on *f0* in T1 (4.15a), T3 (4.15b), T4 (4.15c), and T5 (4.15d), respectively. Each contour was extracted by averaging across different amounts of samples contrasting in onsets (i.e., 76 for T1; 43 for T3; 47 for T4; 61 for T5). Wherever possible, voiced, voiceless aspirated, voiceless unaspirated, and sonorant onsets have been included.

As observed in Figure 4.15(a–c), voiceless unaspirated onsets (solid line) consistently introduce a higher *f0* onset than its aspirated counterparts (dotted line). However, such an effect vanishes gradually and both *f0* contours tend to merge towards the end. Visual inspection of the contours of both T3 and T4 suggests that the perturbation effect of voiceless aspirated onsets is quite salient relative to that of voiceless unaspirated onsets, which ranges from 25 Hz in T3 to 50 Hz in T4.

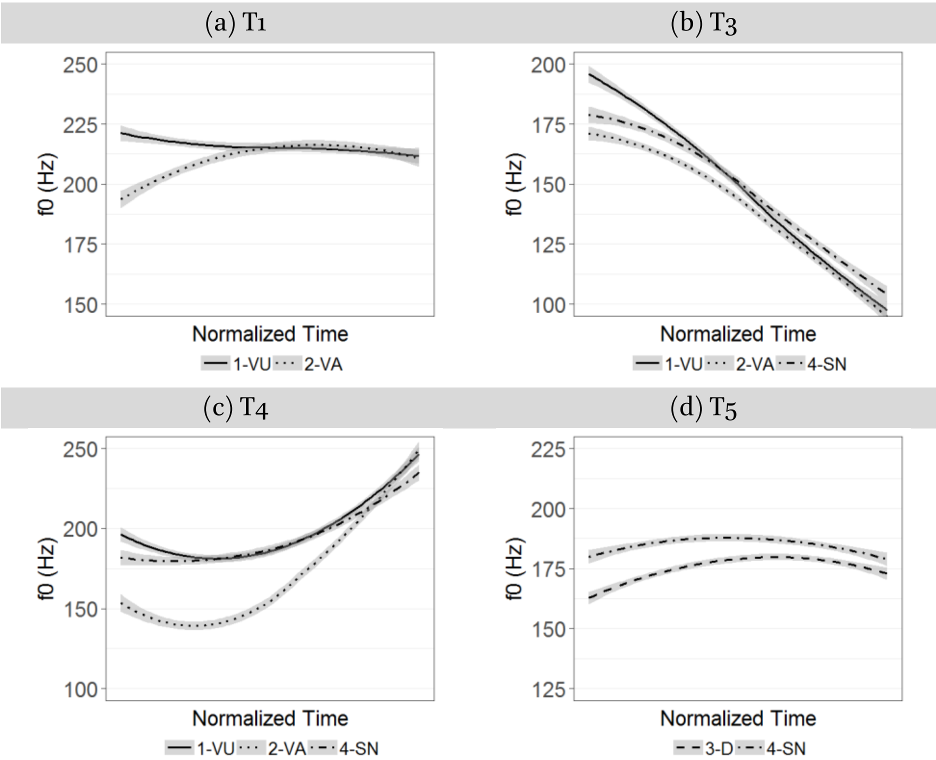



**Figure 4.15** Mean *f0* of (a) T1, (b) T3, (c) T4, and (d) T5 after different onsets. Light gray areas indicate ± *SE*. VU: voiceless unaspirated; VA: voiceless aspirated; D: voiced; SN: sonorants.

For the perturbation effect of sonorant onsets, as shown in Figure 4.15(b–c), the *f0* contours after sonorant onsets are similar to those after voiceless unaspirated counterparts for both T3 and T4. Furthermore, sonorant onsets tend to introduce a much higher *f0* contour than voiceless aspirated onsets. As mentioned earlier (Section 4.2), T5 does not co-occur with voiceless onsets. Figure 4.15(d) shows the *f0* realizations of T5 after voiced obstruent and sonorant onsets. Interestingly, the *f0* trajectory after sonorants shows a higher overall pitch than that after voiced onsets, similar to what has been observed in T2.

These observations suggest that in Shuangfeng Xiang Chinese, voiceless unaspirated onsets share more similar perturbation effects with sonorant counterparts, while voiced onsets resemble voiceless aspirated ones. Following Hanson's motivation of using nasals as a reference (Hanson, 2009), C*F*0 in Shuangfeng Xiang Chinese might be summarized that *f*0 is lowered relative to the sonorant baseline after voiced and voiceless aspirated onsets but is unaffected (or minimally affected) after voiceless unaspirated onsets. This trend, however, is opposite to the results reported in Kirby and Ladd (2016), which found that in 'true voicing' languages (i.e., French and Italian in their study), compared to *f*0 after a sonorant, voiceless onsets introduce a salient *f*0 raising but no distinguishable *f*0 change following the release of a voiced consonant. Needless to say, more research with a larger group of participants is needed to further verify the differences.

## 4.10 Tone sandhi

Lexical tones over monosyllabic morphemes may undergo changes when they come into contact with each other in connected speech, due to a range of factors (Chen, 2012). Given a disyllabic constitution, the most relevant contextual tonal variation is tone sandhi. Chinese dialects, especially those spoken in southern and eastern areas, such as Wu and Min dialects, are famous for their manifold sandhi variations. Interested readers are referred to the seminal work of Chen (2000), which gathers a constellation of descriptions on tone sandhi across Chinese dialects. (Tone sandhi is often difficult to be distinguished from tonal coarticulation; see Li et al., 2019 for further discussion with data from Tianjin Mandarin.)

A preliminary observation over all possible 25 disyllabic compounds (5 × 5) indicates that in Shuangfeng Xiang Chinese, no obvious sandhi change can be straightforwardly identified. Quite commonly, in a disyllabic compound, irrespective of the syllable position, both lexical tonal contours are maintained. For example, Figure 4.16 and 4.17 plot minimal pairs of /sja$^{1}$ pen$^{1}$/ 'three sides' (4.16a) vs. /sja$^{1}$ pen$^{4}$/

'three times' (4.16b), and /ɕan$^{1}$ tʃʅ$^{3}$/ 'a new sheet of paper' (4.17a) vs. /ɕan$^{4}$ tʃʅ$^{3}$/ 'letter paper' (4.17b), respectively. All tonal contours produced in the isolated form are held. This observation accords with the report on tonal variations by Xiang (1960). But tonal coarticulation does seem to exist, which requires quantitative data with more tokens produced by multiple speakers.

(a) /sja$^{1}$ pen$^{1}$/ (b) /sja$^{1}$ pen$^{4}$/

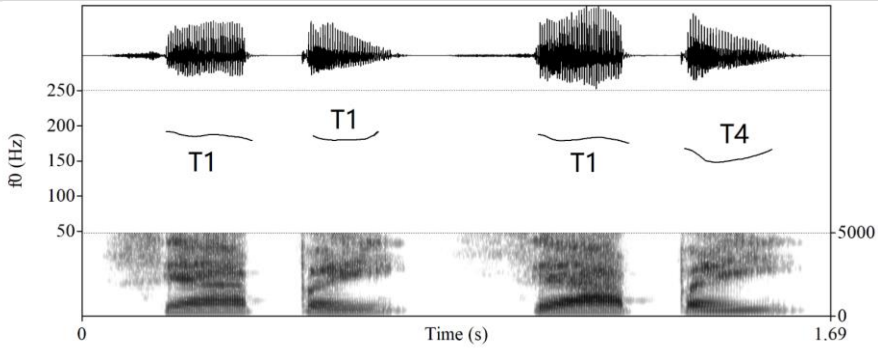



**Figure 4.16** Waveforms, *fo* tracks, and spectrograms of (a) /sja$^{1}$ pen$^{1}$/ 'three sides' and (b) /sja$^{1}$ pen$^{4}$/ 'three times'.

(a) / ɕan$^{1}$ tʃʅ$^{3}$/ (b) / ɕan$^{4}$ tʃʅ$^{3}$/

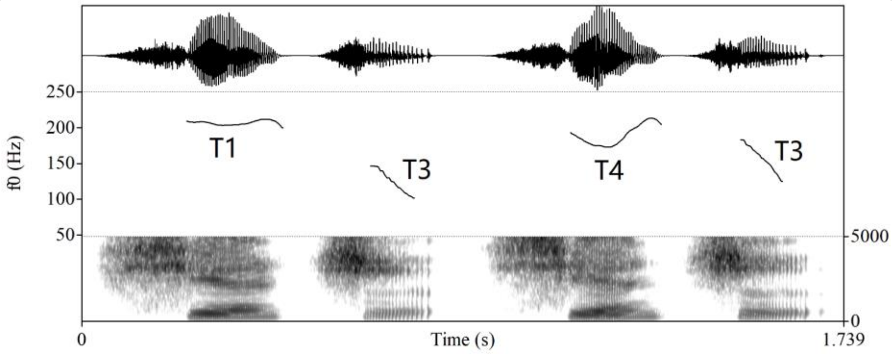



**Figure 4.17** Waveforms, *fo* tracks, and spectrograms of (a) /ɕan$^{1}$ tʃʅ$^{3}$/ 'a new sheet of paper' and (b) /ɕan$^{4}$ tʃʅ$^{3}$/ 'letter paper'.

Worth noting is that in Shuangfeng Xiang Chinese, there are some grammatical morphemes which are phonetically realized as neutral tone (Chao, 1968; Chen & Xu, 2006). Typical examples include the plural suffix

/li$^{N}$/ such as in /t$^{h}$o$^{1}$ li$^{N}$/ 'they' and the nominal suffix /tsɹ̩$^{N}$/ in /tja$^{1}$ tsɹ̩$^{N}$/ 'list' (also see Bao, 2015 for more examples). Here, the superscript $^{N}$ is used to mark syllables with a neutral tone.

Various tone sandhi changes have been reported in Xiangxiang Xiang Chinese. It is argued that the tonal contours of the second syllable are reduced to three level tones, depending on tonal identity of the syllable (Zeng, 2019). These changes, however, have not been observed in Shuangfeng Xiang Chinese. It is important to conclude here that even within the same subgroup of Xiang Chinese (i.e., Old Xiang Chinese), Shuangfeng Xiang Chinese already exhibits different sandhi patterns from its neighboring dialects such as Xiangxiang Xiang Chinese. It seems that tone sandhi patterns, at least within Old Xiang Chinese, can be highly dialect-specific. Needless to say, more data and further research are needed.

# 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 *f0* 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 *f0* 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 *f0* 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 *f0* 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 $f_0$ 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 $f_0$, given that the 'phonation contrast' view claims a contrast of /24/ and /13/ occurring with voiceless and voice onsets, respectively, the $f_0$ 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 [$\chi^2$ = 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 ($\chi^2$) 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.

| Time term | Factor | $\chi^2$ |
|---|---|---|
| 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 *f*0 of vowels following all plosives in normalized time (on the x-axis) and *f*0 (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 *f*0 contours after U and D. That is, *f*0 contours after U are overall higher than those after D. However, *f*0 contours after A vary across generations. As shown in Figure 5.1(a), for older speakers, *f*0 contours after both A and D almost overlap. While for younger speakers, as shown in Figure 5.1(b), the *f*0 contour after A tends to have a higher *f*0 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 *f*0 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).

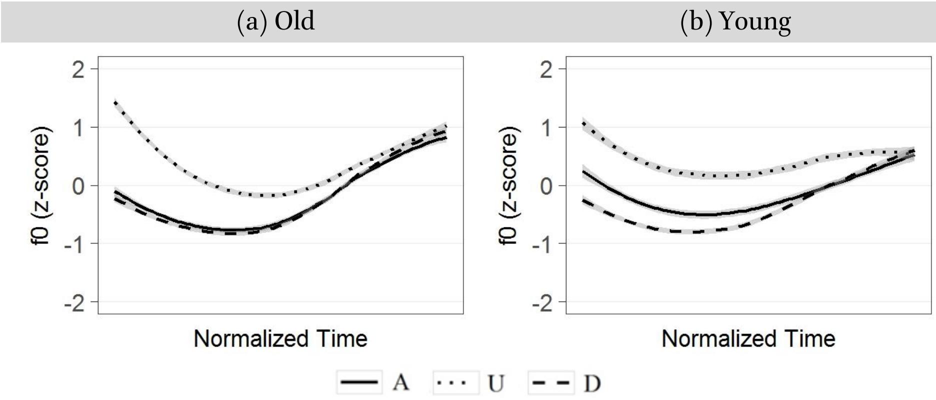



**Figure 5.1** The *f0* realization of T2 for two generations. Mean normalized *f0* (gray areas indicate ± *SE*) after three laryngeal contrasts of plosives (i.e., A: voiceless aspirated, U: voiceless unaspirated, and D: voiced) based on 20 *f0* measurements taken at equally spaced intervals across the 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 ($\beta$) 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 *f*0 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 *f*0 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 *f*0 onset after A. However, compared to the *f*0 contour after U, the *f*0 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 *f*0 than those of U, which more or less confirmed the impressionistic description of the 'voicing contrast' view, which argues for a lower *f*0 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 *f*0 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 *f*0 contour after A. These findings thus challenge both views. The 'voicing contrast' view does not fully predict the findings, as this view predicts similar *f*0 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.

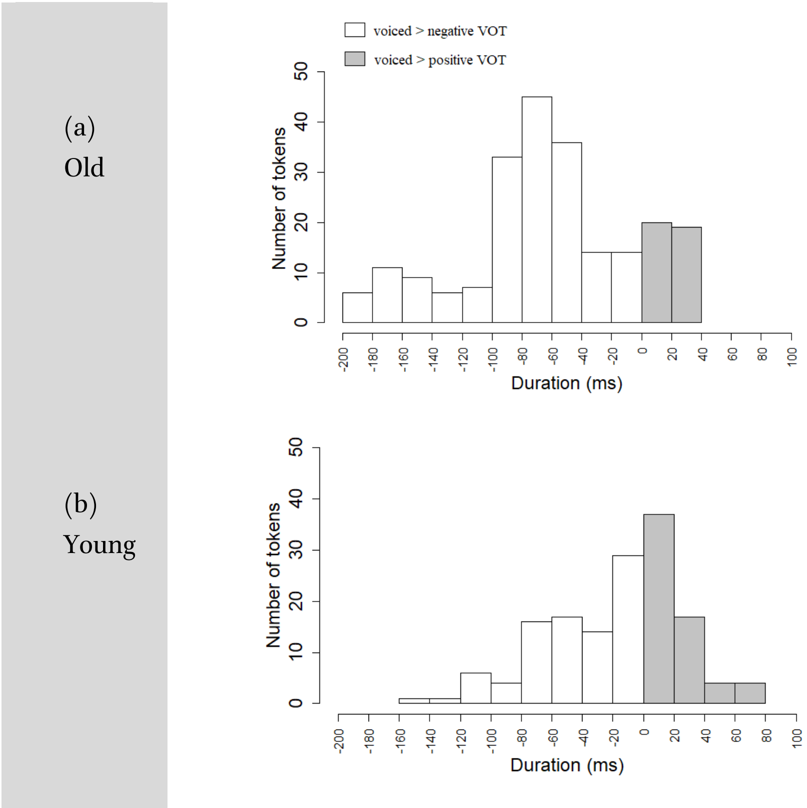



**Figure 5.2** Histograms representing the distribution of VOT for productions of plosives for (a) old and (b) young generations. Each 20 ms bin includes all tokens that have a VOT value equal to or lower than the bin label (and higher than the label of the immediately lower bin). White bars represent 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 | $\chi^2$ | *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 ($\beta$) | *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 | $\chi^2$ | *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.

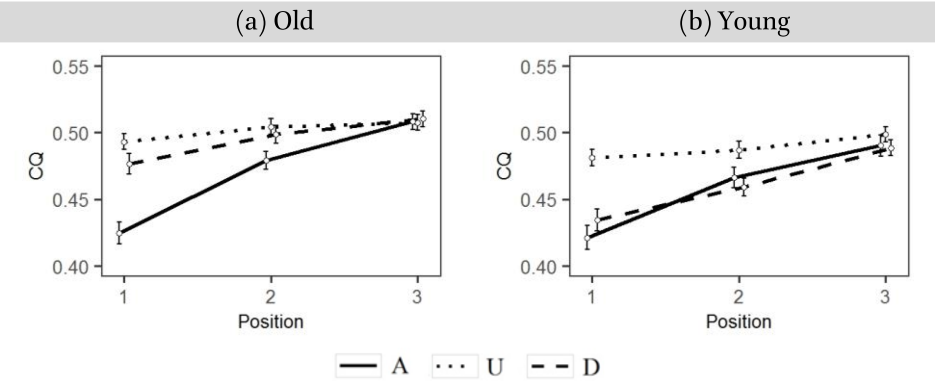



**Figure 5.3** Mean CQ of three positions (1: one-third; 2: middle; 3: two-thirds) from the onset of the first regular vocal pulse till the end of the rhyme in target syllables of each generation. The x-axis represents the ratio of the vowel. The y-axis represents the CQ value. Error bars represent the 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 | $\chi^2$ | *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 ($\beta$) 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 *f*0. 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 *f*0 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 *f*0 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 *f*0 contours, there were no general differences. Conjointly, the results suggest that the low-rising *f*0 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 *f*0 curves showed a general rising contour and converged to the same target towards the syllable offset. In this way, the *f*0 incongruence can be attributed to perturbation effects (i.e., C*F*0) triggered by different onsets. That unaspirated onsets co-occur with higher *f*0 contours, while both aspirated and voiced onsets co-occur with comparable lower *f*0 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.

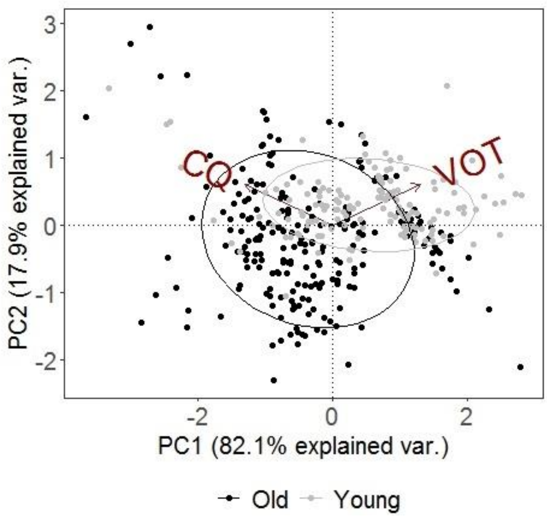



**Figure 5.4** Biplot for the VOT-CQ data (correlation matrix PCA).

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 *f*0. 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.

# Chapter 6 General discussion and conclusion

## 6.1 Main findings

The main goal of this dissertation was to investigate the interaction of consonant and lexical tone (C-T interaction hereafter) in two Chinese dialects, namely Lili Wu and Shuangfeng Xiang Chinese. To achieve this goal, the property-based approach was highlighted. As discussed in Chapter 1, two biases (i.e., methodological and typological) were identified in the existing literature, which motivated the current study.

Chapter 2 presented a comprehensive description of the sound system of Lili Wu Chinese. A number of methodological/analytical innovations and new perspectives with regard to not only lexical tones but also segmental features have been proposed. First, there are eight lexical tones in Lili Wu Chinese. A clear fundamental frequency (*f0*) lowering effect in syllables with voiceless aspirated onsets in certain tonal contexts (i.e., non-*Ping* Middle Chinese [MC] tonal categories) has been observed. This lowering effect, seemingly a split of the same tone into two as a function of voiceless unaspirated vs. aspirated onsets, is known as the 'aspiration-induced tonal split' (ATS) phenomenon. The co-occurrence pattern (i.e., voiceless onsets co-occurring with high-register tones, while voiced onsets with low-register ones) commonly observed in most Northern Wu dialects, falls apart in Lili Wu Chinese where voiceless aspirated onsets can co-occur with low-register tones. Second, voicing contrasts in fricatives such as /f/ vs. /v/ can be signaled via their durational differences. The percentage of the frication duration of voiceless onsets is significantly higher than that of their voiced counterparts. Third, the two high front vowels have been proposed to be better transcribed as /i/ (e.g., /ti$^3$/ 'dot') and /i̟/ (e.g., /ti̟$^3$/ 'bottom') with more anterior constriction for /i̟/. Fourth, there are two syllabic approximants in Lili Wu Chinese. /ɹ̩̹/, as in /sɹ̩̹$^{1}$/ ‘book’, is produced with a more laminal articulation combined with a lip rounding gesture in contrast to /ɹ̩/, as in /sɹ̩$^{1}$/ ‘silk’.

Chapter 3 focused on the issue of C-T interaction in Lili Wu Chinese. Controlled experiments were designed to examine two long-standing debates on ATS in previous literature. They are i) Is ATS an ongoing change or a completed change? and ii) Is aspiration or breathiness synchronically related to ATS? The present results suggest that ATS in Lili Wu Chinese is a completed sound change but that it is conditioned by certain tonal contexts (i.e., MC tonal categories). This pattern is quite consistent across generations. Regarding the second debate, the results suggest that neither aspiration nor breathiness is synchronically related to ATS. One ongoing sound change that is observed 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: it is therefore not a robust cue for the laryngeal contrast in Lili Wu Chinese.

Chapter 4 provided a comprehensive description of the sound system of Shuangfeng Xiang Chinese. There are four main findings. First, the voiced consonant has multiple laryngeal realizations: modal voiced, voiceless unaspirated, and implosive. Second, /n/ and /l/ contrast only before three high segments (i.e., /i ʏ j/) and are neutralized before the other segments. Third, Shuangfeng Xiang Chinese has an interesting and rarely observed three-way contrast in high back vowels, as illustrated by the triplet of /bo$^{2}$/ ‘to climb’ vs. /bu$^{2}$/ ‘old woman’ vs. /bu̜$^{2}$/ ‘calamus’. In addition to their formant differences, the three back vowels can be distinguished via strong visual cues, i.e., their distinct lip gestures. Both /o/ and /u/ differ from /u̜/ in having more rounding and protruding lip constriction, whereas /u̜/ is produced with greater lip compression and less lip protuberance than /o/ and /u/ are. Fourth, relative to the sonorant baseline, *fo* is lowered after voiced and voiceless aspirated onsets but is unaffected (or minimally affected) after voiceless unaspirated onsets. This pattern seems to be consistent across all tonal contexts.

Chapter 5 focused on C-T interaction in Shuangfeng Xiang Chinese. The question concerns the phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets that condition the low-rising *f0* contours. In the existing literature, *f0* contours after different onsets are usually treated as an identical low-rising tone (i.e., T2) which is associated with syllables with laryngeal contrast in voicing (e.g., Chao, 1935 [Yang, 1974]). This phenomenon is termed 'initial-associated tonal merger' (ITM). However, Zhu and Zou (2017) argue for a two-way phonatory distinction of the laryngeal contrast which co-occurs with two separate low-rising tones (i.e., modal – /24/ vs. breathy – /13/). The results suggest that neither voicing contrast nor phonation contrast could explain all findings. Furthermore, the phonetic properties that condition the low-rising *f0* contours have been undergoing changes. Specifically, there seems to be a trading relationship between voice onset time (VOT) and contact quotient (CQ) across generations. When a low-rising-contour-carrying syllable has a voiced onset, the old-generation speakers produce predominately negative VOT without significant breathiness (indicated by CQ) in the following vowel. The young-generation Shuangfeng Xiang speakers, however, produce fewer negative-VOT tokens as well as shorter negative-VOTs. But in contrast to the old-generation speakers, they enhance breathiness over the following vowel (over the first half).

## 6.2 Typological significance

Typologically speaking, what light can the results of Lili Wu and Shuangfeng Xiang Chinese shed on the topic of C-T interaction across the world's languages? In general, three points are of particular interest.

### 6.2.1 Tonal depression and [voice]

The most striking finding of this study is a comparable lowering effect on *fo* after aspirated and voiced onsets in both Lili Wu and Shuangfeng Xiang Chinese. A similar phenomenon of tonal lowering associated with a particular class of initial consonants has been reported in some studies of Nguni, Shona, and Khoisan languages spoken in Africa (see a recent review in Mathes & Chebanne, 2018). This lowering effect is called 'tonal depression' (e.g., Traill et al., 1987; Strazny, 2003) and is argued to be triggered by a series of consonants referred to as 'depressors' (Lanham, 1958: 66). The class of depressors may vary among African languages, but it usually includes both voiceless (i.e., unaspirated plosives and affricates) and voiced (i.e., clicks, fricatives, velar nasal, and nasal-voiced stop sequences) sounds (see a comprehensive review in Downing, 2018).

In her dissertation, Bradshaw (1999) makes a strong claim for a single-source approach for accounting for the depressor effects. Bradshaw's claim consists of at least two aspects. One is of the privative correlation between low tone and [voice] ([L/voice]) (pp. 43). The other, more phonetically, is that the central characteristics of the feature [voice] are associated with vocal fold vibration (pp. 163). However, with the extension of new data, both aspects face considerable challenges (also see previous criticisms in Tang, 2008; Downing, 2009, 2018).

As to the first aspect, cross-linguistic reports have shown that phonological [voiceless] consonants can also interact with low tones. The [voiceless] category mainly refers to voiceless aspirated/fricative consonants. For example, in Lili Wu Chinese, lexical tones with voiceless aspirated onsets have merged with tones with voiced onsets in certain tonal contexts. In another three-way-contrast language – Ikalanga (a Bantu language of the Shona group) – some voiceless aspirated obstruents also behave like depressor consonants, which can further block 'High Tone Spread' (Mathangwane, 1998). In Nambya, a Southern Bantu language closely related to Ikalanga, Downing and Gick (2001) show that a set referred to as 'depressor /f/', i.e., phonetically voiceless fricatives can also act as tonal depressors. Very recently, the acoustic data of Mathes and Chebanne (2018) show a quite comparable *f0* pattern following aspirated obstruents as opposed to the voiced counterparts in Tsua, a Khoisan language. All cases are further confirmed to have no historic source for voicing and grow out of historically voiceless segments. Interested readers are referred to Tang (2008: 26, Table 2) for more examples.

The second aspect, namely the correlation between vocal fold vibration and the phonological feature [voice], is more problematic. The studies of Lili Wu and Shuangfeng Xiang Chinese clearly suggest that the so-called voiced category of obstruents in both dialects fails to show stable voicing (indicted by lead VOT). In initial position, the voiced category is never realized with negative VOTs in Lili Wu Chinese; while in Shuangfeng Xiang Chinese, it is partially realized with negative VOTs. However, regardless of dialects, the so-called voiced category of obstruents is consistently correlated with a lowering of *f0*, contrasting with the *f0* contour after the voiceless unaspirated category. These observations are basically in accord with the study by Kingston and Diehl (1994), which has already shown that the [voice] distinction is not consistently realized with voicing lead through much or all of its closure (i.e., consequently lead VOT). However, *f0* is consistently depressed in vowels adjacent to [voice] onsets, regardless of their VOT values. In addition, Zulu and Xhosa Bantu have been reported to present a voicing-absent realization of the depressor onsets (Jessen & Roux, 2002; Chen & Downing, 2011).

In a nutshell, neither the phonological [voice] category nor phonetic voicing should be inextricably or exclusively correlated with the low tone/tonal depression. Cross-linguistic reports have already shown that although most tonal languages abide by the [voiceless/H]-[voiced/L] pattern, there are more language-specific patterns that need to be uncovered to further our understanding of the C-T interaction.

### 6.2.2 The effect of aspirated onsets on *fo*

What has been consistently observed in this study is the lowering of *fo* contour after aspirated onsets. The effect of aspirated onsets on *fo* seems to be highly language-specific (Carne, 2008; Chen, 2011) and even speaker-specific (e.g., see studies of Korean, Kagaya, 1974; and Thai, Erickson, 1975). In terms of typology, as argued by Chen (2011: 622), '[i]n some, often non-tonal languages, voiceless aspirated stops tend to show comparable *fo* as voiceless unaspirated stops, while in others, mainly tonal languages or languages with a full-fledged system of phonation and aspiration contrast, a voiceless aspirated stop tends to introduce lower *fo* than a voiceless unaspirated one.' It is worth noting that an aspiration-induced *fo* lowering effect also seems likely to exist in languages where obstruents have more than a binary laryngeal contrast. Asian languages such as Lili Wu Chinese, Shuangfeng Xiang Chinese, Kam (Sanjiang 三江 variety, Donohue & Wu, 2013), Vietnamese (Northern Vietnamese, Carne, 2008), and Tibetan (Qiuji and Tiebu, Sun, 2003) have all been reported to have a three-way laryngeal contrast of obstruents. It is also true for African languages such as Ikalanga (Mathangwane, 1998) and Tusa (Mathes, 2015: 34). Taken together, these studies suggest that, unlike the clearly binary effect on *fo* after voiceless and voiced onsets, perturbation effects after aspirated onsets seem to be more complicated.

A subsequent issue is why aspirated onsets can co-occur with lower *fo* contours. Breathiness should play a crucial role in such a co-occurrence. This is because in Lili Wu Chinese, irrespective of whether tonal split occurred or not, aspiration has a consistent perturbation effect on the following vowel, as evident in the strong degree of breathiness. Similar effects have been observed in Shuangfeng Xiang Chinese. This relationship between aspiration and breathiness has been reported in some other languages, such as Swedish (Gobl & Chasaide, 1988), English (Löfqvist & McGowan, 1992), and German (Chasaide & Gobl, 1993). The observed increased breathiness has been attributed to a delayed laryngeal adjustment after the release of an aspirated onset. Upon the release of an aspirated onset, the glottis may have a more abducted posture, which consequently causes the vocalis muscle's effect to be either weak or not as effective as after an unaspirated onset, as Chen (2011) argued for the mechanism of *f0* lowering in Shanghainese. In addition, aspiration-induced greater aperiodic noise may also be at play. It has been argued that aspiration is typically followed by considerably greater airflow (Stevens, 1971), which can result in a breathier transition between aspiration and vowel voicing (Sagart, 1981; Ren, 1992; Zhu & Xu, 2009).

In this way, breathiness transition, together with aperiodic noise of aspiration provides a possible pivot for linking onset aspiration to *f0* lowering. Aspirated onsets hence have the potential to behave like voiced onsets in introducing *f0* lowering. For example, similar lowering effects after aspirated and voiced onsets have been reported for Standard Thai (Gandour & Maddieson, 1976), Ikalanga (Mathangwane, 1998), and Tsua (Mathes & Chebanne, 2018). In Shanghainese, the perturbation effect of the aspirated category is similar to that of the voiced one in non-initial position (Chen, 2011). Moreover, lower *f0* and breathier phonation are found to correlate with both voiceless and voiced aspirated plosives in Nepali (Clements & Khatiwada, 2007; Khatiwada, 2008; Mazaudon, 2012) and Bengali (Mikuteit & Reetz, 2007). It is worth noting that in this current study, the breathier transition can be generally observed at the onset of the vowel after all aspirated onsets regardless of whether tonal split happened or not. This suggests that breathiness is not the direct trigger of lower *f0*. There should be no direct correlation between breathier voice and lower *f0*. As *f0* level is related to the degree of stiffness of the vocal cords while breathier voice is related to the glottal constriction and noise component, it is not difficult to imagine that the same glottal constriction and noise component can vary via different rates of vibration of the vocal folds (Ladefoged, 1973; Kuang, 2013b).

Last but not least, in comparison to voiceless unaspirated onsets, how should we understand those languages where voiceless aspirated onsets produce a comparable *f0* pattern (or even a raising effect as in Khmer, Central Thai, and Northern Vietnamese [Kirby, 2018])? The opposite effect on *f0* may result from different mechanisms for producing voiceless aspirated sounds. In terms of the state of the vocal folds, in those languages, speakers may employ active tension in order to inhibit vocal-fold vibration triggered by a more abducted glottis after both kinds of voiceless onsets (Hanson & Stevens, 2002), which then produces comparable *f0* patterns. The breathier transition after voiceless aspirated onsets is inhibited and hence limited. Needless to say, more research, especially aerodynamic experiments, is needed.

To summarize, in Lili Wu and Shuangfeng Xiang Chinese, not only voiced onsets but also aspirated onsets can introduce *f0* lowering.

### 6.2.3 ATS reported in other languages

Except for Wu Chinese, ATS has also been documented in some other languages. For example, in many Kra-Dai languages spoken in China, similar tonal-split phenomena associated with onset aspiration have been reported (Ho, 1989; F. Shi, 1998). Xu (2014) further argues that ATS serves as evidence for the retention of a Kra-Dai substratum of the proto-Wu. However, unlike Lili Wu Chinese, where *f0* contours after aspirated onsets have merged with those after voiced onsets, a few studies have shown that *f0* contours produced with aspirated onsets are allotones of those produced with unaspirated onsets in some Southern Kam (Dong 侗 in Chinese literature) varieties (Donohue & Wu, 2013; Long, 2018). What makes the story more complicated is the patterns reported for the Rongjiang (榕江) variety. The acoustic results of Zhu et al. (2016) show that there is shorter VOT lag (30–60 ms) but longer maintenance of breathiness (around 100 ms) over the following vowels for syllables featuring the so-called ATS. More importantly, ATS in Kam languages seemingly can happen to high-level tones (e.g., Rongjiang) and its historical condition is less clear-cut (F. Shi, 1998: 131). All these results are inconsistent with the findings observed in Lili Wu Chinese.

In addition, Gan Chinese (赣语) is another group where ATS has been reported in the existing literature. However, the historical condition of ATS in Gan Chinese is quite different from that observed in Wu Chinese. In Lili Wu Chinese, none of the aspirated onsets have a historic source for voicing and are known to grow out of historically voiceless segments. However, in dialects like Nanchang (南昌) Gan, ATS has been found to co-occur with aspirated onsets which developed from historically voiced obstruents (Xiong, 1979). A related question is whether the so-called ATS is the vestige of the low tones which co-occurred with voiced onsets in MC (Ho, 1989; F. Shi, 1998; Wang, 2010). A recent study of Zhajin (渣津) Gan Chinese (Zhou & Kirby, 2019) seems to indicate that the historical voiceless aspirated onsets first became voiced/breathy, and merged with MC voiced onsets. Synchronically speaking, VOT is no longer a robust cue to signal the laryngeal contrast of voiceless aspirated vs. voiced onsets. The merged category (covering MC voiceless aspirated and voiced onsets) is argued to further engage in the tone-developmental processes in Gan Chinese.

All evidence seems to point to a language-specific interpretation of the so-called ATS. To have a better understanding of ATS, two additional pieces of evidence are needed. One is the diachronic reconstructions on how the tonal inventories of proto-languages that existed in China evolved into those of modern Chinese dialects; and the other is more empirical data from under-documented languages, of which Kra-Dai languages especially.

## 6.3 The relationship between Wu and Xiang: from the perspective of C-T interaction

### 6.3.1 Unicity of Wu–Xiang?

The existence of a three-way laryngeal contrast in obstruents, labeled as voiceless unaspirated, voiceless aspirated, and voiced, is the most prominent feature shared by Wu (e.g., Lili) and Old Xiang Chinese (e.g., Shuangfeng) (Zhou & You, 1986). This shared phonological trait, therefore, has been adopted as an overarching criterion to assume that the Wu and Old Xiang dialect groups are much more closely related to each other in historical classification than either of them is with other dialect groups. This assumption seems to be first raised in Yuan (1960: 102) and is later more firmly established in Hashimoto (1978/1985). Given the shared three-way contrast in obstruents, Hashimoto (1985: 31) points out that the Wu and Xiang dialect groups once constituted a unified dialect phylum. This view was later called the 'Unicity of Wu–Xiang' (吴湘一体) in Chang (1999).[40] Given the shared phonological trait of the voiced category of obstruents, this phylum is further called the 'Corridor of voiced sounds' (浊音走廊) in Chen (2004). The assumed historical relatedness of the Wu and Old Xiang dialect groups, however, is not unproblematic. The challenges come from at least two aspects.

First, in terms of historical reconstruction, the three-way contrast of obstruents is *not* shared innovation but shared retention, which cannot be used as evidence to argue for the close relatedness of the Wu and Old Xiang dialect groups. It has long been established in historical linguistics that shared innovation is the only criterion for subgrouping (e.g., Trask, 1996: 182; Campbell, 2013: 197). Given the widespread consensus on the reconstruction of the three-way-contrast system in MC (e.g., Karlgren,

[40] Chang (1999) further includes some Jianghuai dialects, assuming a so-called 'Dialect Continuum of Wu–Chu–Jianghuai' (吴楚江淮方言连续体).

1915–1926), this single common phonological trait in the Wu and Old Xiang dialect groups is merely a trait inherited from the proto-language. Similar manipulation of subgrouping based on shared retentions seems to be a common fallacy in some studies of the classification of Chinese dialects, which has been sharply criticized by Sagart (2002, 2011). While it is reasonable to employ the *development* of the three-way-contrast system as a criterion to subgroup Chinese dialects (Ting, 1982; Li, 2005), the so-called close relatedness of the Wu and Old Xiang dialect groups postulated based on their shared retention of the three-way contrast of obstruents is unlikely to be tenable. Similar doubts have been raised in Guo (2015).

Nevertheless, no shared innovation relating to the development of the three-way-contrast system can be identified. Instead, the results of this current study show that the development of the three-way-contrast system is quite complicated. Two points are of particular note. First, although the aspirated onset in both dialects can co-occur with low tones, the condition that is responsible for this co-occurrence is different. In Lili Wu Chinese, the condition is more historical (i.e., MC tonal categories). Regardless of the tonal contours at present, this condition seems to take effect across all its neighboring dialects showing the same phenomenon (Wang, 2008; Yue Xu, 2013). However, the *f0*-lowering effect of the aspirated onset in Shuangfeng Xiang Chinese is more related to the synchronic tonal contexts, irrespective of its historical origin. The lowering effect of the aspirated onset is more prominent in the two rising tonal contexts (i.e., low–rising and high–rising) compared to the other contexts (i.e., high–level and falling). Such a context-dependent condition is in line with some previous findings. For example, in Standard Chinese (Xu & Xu, 2003), the aspirated onset introduces a greater perturbation effect for the rising tone than for high and falling tones. In Southern Standard Kam (Tang, 2008) and Northern Standard Vietnamese (Carne, 2008), both languages show an affinity for a prominent *f0*-lowering effect on the rising tone after an aspirated onset. This rising context can even extend beyond one syllable. For example, in Shanghainese (Chen, 2011), more salient perturbation effects on onset *f0* have been observed for syllables after low-rising tones than that after high-falling tones. However, the greater lowering effect of the aspirated onset in the rising context is not universal. Opposite cases can also be found. For example, in Cantonese (Francis et al., 2006), the perturbation effect after the aspirated onset presents a greater magnitude in the falling (/21/) tone than that in the level (/55/) and rising (/25/) tones. In a nutshell, the development of the aspirated category in the two dialects is not consistent.

More importantly, the development of the voiced category is also incongruent. This is mainly reflected by the realization of the voiced category in the two dialects. As shown in Table 6.1, the realization of the voiced category varies greatly between dialects. In Shuangfeng Xiang Chinese, negative VOT is partially realized to a higher degree among older speakers. In Lili Wu Chinese, none of the voiced onsets is realized with negative VOT. Furthermore, speakers of different generations within a dialect show differences. In Shuangfeng Xiang Chinese, speakers of the old generation produce less breathiness (indicated by higher CQ) than those of the young generation, while in Lili Wu Chinese, the relationship is reversed. Voiced onsets, however, are consistently realized with lower *f0* contours across dialects and generations. Given these characteristics of the phonologically voiced category in both dialects, it is questionable if the three-way laryngeal contrast of obstruents in the two dialects can be treated in a uniform way.

**Table 6.1** The realization of the voiced category in Shuangfeng Xiang and Lili Wu Chinese.

| Dialect | Generation | VOT | CQ | *f0* contour |
|---|---|---|---|---|
| Shuangfeng | Old | negative (82%) | higher | low |
| | Young | negative (59%) | lower | low |
| Lili | Old | positive (100%) | lower | low |
| | Young | positive (100%) | higher | low |

In summary, it is reasonable to argue that both Wu (e.g., Lili) and Old Xiang (e.g., Shuangfeng) dialect groups have inherited the phonological trait of the three-way laryngeal contrast of obstruents from MC. However, in terms of the development of the three-way-contrast system, no shared innovation can be identified. Instead, the two dialects behave quite differently and have developed their idiosyncratic features in the three-way laryngeal contrast of obstruents. Currently, there is no obvious evidence to support the 'Unicity of Wu–Xiang' assumption.

### 6.3.2 Today's Shuangfeng, tomorrow's Lili?

How can this study shed light on the relationship between the Wu and Old Xiang dialect groups? The most prominent finding is that the voiced category of obstruents is experiencing an ongoing change in both dialects (see Section 3.5.2 and Section 5.5.2). In general, an obvious trading relationship has been found between VOT and CQ. In Shuangfeng Xiang Chinese, the old-generation speakers produce predominately negative VOT without significant breathiness (indicated by CQ) in the following vowel. The young-generation Shuangfeng Xiang speakers, however, produce fewer negative-VOT tokens as well as shorter negative-VOTs. Furthermore, they enhance breathiness over the following vowel but only over the first half. In Lili Wu Chinese, a different cue-trading relationship has been observed. The old-generation speakers produce all 'voiced' obstruents with positive-VOTs but with a significantly higher level of breathiness over the first half of the following vowel. The young-generation speakers also produce positive-VOTs over the 'voiced' obstruents but with decreased breathiness over the first half of the following vowel. The four different relationships between VOT and CQ seem to imply a developmental trajectory of the voicing contrast varying from the old-generation Shuangfeng Xiang speakers to the young-generation Lili Wu speakers.

To sum up, VOT and CQ have started to show flexibility to signal the voicing contrast across dialects. The current situation in the two dialects, in terms of sound change, might give rise to two predictions. First, with the loss of phonetic voicing but the enhancement of breathiness among younger speakers of Shuangfeng Xiang Chinese, there seems to exist a chance that the current younger speakers of Shuangfeng Xiang Chinese today will speak like the current older speakers of Lili Wu Chinese in the future. Further longitudinal surveys are needed to test this possibility. Additionally, if no strategy of enhancement is taken by the younger speakers of Lili Wu Chinese, the loss of breathiness can potentially weaken the distinctness of voiced obstruents, leading to the neutralization of the voicing contrast. Such a tendency has already been observed in some of the young female speakers of Shanghainese (Gao, 2016) and in some speakers of Tamang dialects (Mazaudon, 2012).

The different trading relationships between laryngeal timing (in terms of VOT) and phonatory state (in terms of CQ) to signal voicing contrast in the two Chinese dialects highlight possible pathways for changes of cue weighting in the phonetic implementation of voicing contrasts in Asian tonal languages.

## 6.4 Conclusion

In the existing literature, C-T interaction generally concerns the [voiceless/H]-[voiced/L] co-occurrence pattern. However, largely because of the high level of homogeneity in the languages sampled, and the lack of access to portable articulatory and up-to-date statistical techniques, this [voiceless/H]-[voiced/L] pattern has veiled the full picture of C-T interaction across the world's languages.

There are two key findings in this dissertation that contribute to our understanding of the diversity in C-T interaction. First, voiceless aspirated onsets can also co-occur with low tones. This finding is antagonistic to the [voiceless/H]-[voiced/L] pattern which posits that only contrastively voiced onsets can be in favor of low tones. Second, the realization of C-T interaction is not only specific between languages but also within languages. Speakers of different generations of a given language can utilize phonetic cues differently to signal the same phonological contrasts.

Last but not least, I hope to highlight the importance of the property-based approach for exploring phonological contrasts. In doing so, I believe that it would be beneficial to incorporate more cross-linguistic data into the perspective of phonological typology.

# References

A Committee of the Soochow Literary Association. (1892). *A syllabary of the Soochow dialect*. Shanghai: American Presbyterian Mission Press.

Abramson, A. S., & Erickson, D. M. (1978). Diachronic tone splits and voicing shifts in Thai: Some perceptual data. *Haskins Laboratories: Status Report on Speech Research*, *53*(2), 85–96.

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86.

Abramson, A. S., Tiede, M. K., & Luangthongkum, T. (2015). Voice register in Mon: Acoustics and electroglottography. *Phonetica*, *72*(4), 237–256.

Anderson, V. B., & Maddieson, I. (1994). Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics*, *87*, 131–162.

Auguie, B. (2015). *gridExtra: Miscellaneous Functions for "GRID" Graphic*. Retrieved from https://cran.r-project.org/web/packages/gridExtra/index.html

Baken, R. J. (1992). Electroglottography. *Journal of Voice*, *6*(2), 98–110.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd ed.). San Diego: Thomson Learning.

Bao, H. (2015). Xiāng 湘 Dialects. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang, & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics*. Retrieved from https://referenceworks.brillonline.com/entries/encyclopedia-of-chinese-language-and-linguistics/xiang-dialects-COM_00000465

Bao, H. 鲍厚星, & Chen, H. 陈晖. (2005). 湘语的分区 (稿) [The classification of Xiang group]. *方言* [Dialects], (4), 261–270.

Bao, H. 鲍厚星, & Yan, S. 颜森. (1986). 湖南方言的分区 [The classification of the Hunan dialects]. *方言* [Dialects], (4), 273–276.

Bao, Z. (1999). *The structure of tone*. New York: Oxford University Press.

Baron, S. (1974). On the tip of many tongues: Apical vowels across Sino-Tibetan. Presented at the 7th International Conference on Sino-Tibetan Language and Linguistic Studies, Atlanta, USA.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*(2), 259–284.

Berkson, K. H. (2013). *Phonation types in Marathi: An acoustic investigation* (PhD Thesis). University of Kansas, Lawrence.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, *30*(2), 163–191.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. Retrieved from http://www.fon.hum.uva.nl/praat/

Bradshaw, M. M. (1999). *A crosslinguistic study of consonant-tone interaction* (PhD Thesis). Ohio State University, Columbus.

Bradshaw, M. M. (2000). Consonant-tone interaction in African languages. In V. Carstens, & F. Parkinson (Eds.), *Advances in African linguistics* (pp. 151–164). Asmara: Africa World Press.

Brown, J. M. (1975). The great tone split: Did it work in two opposite ways. In J. G. Harris, & J. R. Chamberlin (Eds.), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 33–48). Bangkok: Central Institute of the English Language.

Brunelle, M., & Kirby, J. (2016). Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass*, *10*(4), 191–207.

Butcher, A. (1990). "Place of articulation" in Australian languages. *Proceedings of the 3rd Australian International Conference on Speech Science and Technology*, 420–425. Melbourne, Australia.

Campbell, L. (2013). *Historical linguistics* (3rd edition). Edinburgh: Edinburgh University Press.

Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics*, *20*(1), 77–92.

Carne, M. J. (2008). Intrinsic consonantal F0 perturbation in 3-way VOT contrast and its implications for aspiration-conditioned tonal split: Evidence from Vietnamese. *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, 2294–2297. Brisbane, Australia.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Chang, K.-Y. 张光宇. (1999). 东南方言关系综论 [An overview of the relationship among Southeast dialects]. *方言* [Dialects], (1), 33–44.

Chao, H.-J. (1992). *Aspiration in Chinese* (PhD Thesis). University of Illinois, Urbana and Champaign.

Chao, Y.-R. 赵元任. (1928). *现代吴语的研究* [Studies of the modern Wu dialects]. Beijing: Tsing Hua College Research Institute.

Chao, Y.-R. (1930). ə sistim əv "toun-letəz" [A system of "tone-letters"]. *Le Maître Phonétique*, *8*(45), 24–27. [Reprinted in 1980 in *方言* [Dialects], (2), 81–83.]

Chao, Y.-R. 赵元任. (1959). 说清浊 [On the terms Ch'ing and Cho]. *中央研究院历史语言所研究集刊* [Bulletin of the Institute of History and Philology Academia Sinica], *30*(2), 493–497.

Chao, Y.-R. (1967). Contrastive aspects of the Wu dialects. *Language*, *43*(1), 92–101.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley & Los Angeles: University of California Press.

Chasaide, A. N., & Gobl, C. (1993). Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech*, *36*(2–3), 303–330.

Chen, H. 陈晖. (2004). 异纽同调和异纽异调—兼论湘语娄邵片与吴语在古全浊声母演变上的差异 [Initial-associated tonal merger and initial-associated tonal split—On the differences of the development of the voiced initial in Middle Chinese in dialects of Loushao Xiang and Wu Chinese]. *Proceedings of the 4th Postgraduate Research Forum on Linguistics*, 9–17. Hong Kong, China: University of Hong Kong.

Chen, H. 陈晖. (2006). *湘方言语音研究* [A study of phonology in the Xiang dialects]. Changsha: Hunan Normal University Press.

Chen, L. 陈立中. (2004). *湘语与吴语音韵比较研究* [A comparative study of phonology between Xiang and Wu Chinese]. Beijing: China Social Sciences Press.

Chen, L. 陈立中. (2005). 汉语方言声调送气分化现象初探 [An investigation on the Tone-Aspiration-Division in Chinese dialects]. *汉语学报* [Chinese Linguistics], *4*(4), 31–39.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Chen, S., Zhang, C., McCollum, A. G., & Wayland, R. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication*, *88*, 17–38.

Chen, Y. (2011). How does phonology guide phonetics in segment–*f0* interaction? *Journal of Phonetics*, *39*(4), 612–625.

Chen, Y. (2012). Tonal variation. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 103–114). Oxford: Oxford University Press.

Chen, Y., & Downing, L. J. (2011). All depressors are not alike: A comparison of Shanghai Chinese and Zulu. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 243–265). Dordrecht: Springer.

Chen, Y., & Guo, L. (forthcoming). Zhushan Chinese. *Journal of the International Phonetic Association*.

Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, *45*(3), 321–337.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – evidence from f0 patterns of neutral tone in Standard Chinese. *Phonetica*, *63*(1), 47–75.

Chen, Z. 陈忠敏. (2014). 江浙沪交界处方言送气分调及其相关问题 [Aspiration-induced tonal split in dialects of the juncture area of Jiang–Zhe–Hu and its related questions]. In X. Li 李小凡, & M. Xiang 项梦冰 (Eds.), *承泽堂方言论丛:王福堂教授八秩寿庆论文集* [Essays on dialects: A festschrift for Prof. Futang Wang] (pp. 123–138). Beijing: Language & Culture Press.

Chen, Z. 陈忠敏. (2015). 论 160 年前上海话声母 [dz]/[z] 变异—兼论北部吴语从邪澄崇船禅等母读音变异现象 [On the variation of the initial [dz]/[z] in the Shanghai dialect 160 years ago]. *方言* [Dialects], (4), 340–345.

Childers, D. G., & Krishnamurthy, A. K. (1985). A critical review of electroglottography. *Critical Reviews in Biomedical Engineering*, *12*(2), 131–161.

Childers, D. G., Hicks, D. M., Moore, G. P., & Alsaka, Y. A. (1986). A model for vocal fold vibratory motion, contact area, and the electroglottogram. *The Journal of the Acoustical Society of America*, *80*(5), 1309–1320.

Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L., & Lalwani, A. L. (1990). Electroglottography and vocal fold physiology. *Journal of Speech, Language, and Hearing Research*, *33*(2), 245–254.

Chitoran, I. (2002). A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association*, *32*(2), 203–222.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *72*, 52–65.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

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

Clements, G. N., & Khatiwada, R. (2007). Phonetic realization of contrastively aspirated affricates in Nepali. *Proceedings of the 16th International Congress of Phonetic Sciences*, 629–632. Saarbrücken, Germany.

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

Cohn, A. C., Fougeron, C., & Huffman, M. K. (2018). Laboratory phonology. In S. J. Hannahs, & A. Bosch (Eds.), *The Routledge handbook of phonological theory* (pp. 504–529). Abingdon & New York: Routledge.

Cole, R. A., & Cooper, W. E. (1975). Perception of voicing in English affricates and fricatives. *The Journal of the Acoustical Society of America*, *58*(6), 1280–1287.

Colton, R. H., & Conture, E. G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice*, *4*(1), 10–24.

Connell, B. (2007). Mambila fricative vowels and Bantu spirantisation. *Africana Linguistica*, *13*(13), 7–31.

Cun, X. (2009). *A phonetic study on implosives in China* (PhD Thesis). Hong Kong University of Science and Technology, Hong Kong.

Dart, S. N. (1991). Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics*, *79*, 1–155.

Davies, P., Lindsey, G., Fuller, H., & Fourcin, A. (1986). Variation of glottal open and closed phases for speakers of English. In R. Lawrence (Ed.), *Proceedings of the Institute of Acoustics*: *Vol. 8*(7) (pp. 539–546). Edinburgh, UK.

De Jong, K. (1995). On the status of redundant features: The case of backing and rounding in American English. In B. Connell, & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 68–86). Cambridge: Cambridge University Press.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, *39*(2), 162–188.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95.

Donohue, C., & Wu, M. (2013). Aspiration and F0 in Kam: Caught at the beginning? *Proceedings of the International Conference on Phonetics of the Languages in China*. Presented at the International Conference on Phonetics of the Languages in China, Hong Kong, China.

Downing, L. J. (2009). On pitch lowering not linked to voicing: Nguni and Shona group depressors. *Language Sciences*, *31*(2–3), 179–198.

Downing, L. J. (2018). Grounding Nguni depressor effects. In W. Kehrein, B. Köhnlein, P. Boersma, & M. Oostendorp (Eds.), *Segmental structure and tone* (pp. 109–146). https://doi.org/10.1515/9783110341263-005

Downing, L. J., & Gick, B. (2001). Voiceless tone depressors in Nambya and Botswana Kalang'a. *Berkeley Linguistics Society*, *27*, 65–80.

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dutta, I. (2007). *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt* (PhD Thesis). University of Illinois, Urbana and Champaign.

Erickson, B. (2001). On the origins of labialized consonants in Lao. *Papers from the 6th Annual Meeting of the Southeast Asian Linguistic Society*, 135–148. Tempe, USA: Arizona State University.

Erickson, D. (1975). Phonetic implications for a historical account of tonogenesis in Thai. In J. G. Harris, & J. R. Chamberlin (Eds.), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 110–111). Bangkok: Central Institute of the English Language.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. H. Hardcastle, & J. M. Beck (Eds.), *A Figure of Speech: A Festschrift for John Laver* (pp. 347–383). Mahwah: Lawrence Erlbaum Associates.

Esposito, C. M. (2010). Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association*, *40*(2), 181–198.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, *40*(3), 466–476.

Esposito, C. M., & Khan, S. ud D. (2012). Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association*, *42*(2), 123–143.

Ethnologue: Languages of the World. (2019). Retrieved January 15, 2019, from Ethnologue website: https://www.ethnologue.com/

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton & Co. N. V.

Fant, G. (1983). Feature analysis of Swedish vowels – a revisit. *Speech Transmission Laboratory – Quarterly Progress and Status Reports*, *24*(2–3), 1–19.

Fant, G., Ondrackova, J., Lindqvist, J., & Sonesson, B. (1966). Electrical glottography. *Speech Transmission Laboratory – Quarterly Progress and Status Reports*, *7*, 15–21.

Faytak, M., & Lin, S. (2015). Articulatory variability and fricative noise in apical vowels. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Faytak, M., & Merrill, J. (2015). Bantu spirantization is a reflex of vowel spirantization. *Proceedings of the 3rd Workshop on Sound Change*, 28–31. Berkeley, USA.

Fischer-Jørgensen, E. (1972). "p t k" et "b d g" français en position intervocalique accentuée [French "p t k" and "b d g" in intervocalic accented position]. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 143–200). The Hague: Mouton.

Francis, A. L., Ciocca, V., & Yu, J. M. C. (2003). Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America*, *113*(2), 1025–1032.

Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, *120*(5), 2884–2895.

Frokjaer-Jensen, B., & Prytz, S. (1976). Registration of voice quality. *Bruel and Kjaer Technical Review*, *3*, 3–17.

Fromkin, V. A. (1972). Tone features and tone rules. *Studies in African Linguistics*, *3*(1), 47–76.

Gandour, J., & Maddieson, I. (1976). Measuring larynx movement in Standard Thai using the cricothyrometer. *Phonetica*, *33*(4), 241–267.

Gao, J. (2015). *Interdependence between tones, segments, and phonation types in Shanghai Chinese: Acoustics, articulation, perception, and evolution* (PhD Thesis). Sorbonne Paris Cité, Paris.

Gao, J. (2016). Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology*, *1*, 166–186.

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, *40*(1), 152–161.

Garellek, M., & Keating, P. A. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185–205.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 51–97). Oxford: Oxford University Press.

Gobl, C., & Chasaide, A. N. (1988). The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study. *Speech Transmission Laboratory – Quarterly Progress and Status Reports*, *29*(2–3), 23–59.

Gordon, M. (2015). Consonant-Tone interactions: A phonetic study of four indigenous languages of the Americas. In H. Avelino, M. Coler, & L. Wetzels (Eds.), *The phonetics and phonology of laryngeal features in native American languages* (pp. 129–156). Leiden: Brill.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.

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

Greenberg, J. H. (1970). Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics*, *36*(2), 123–145.

Gu, C. (2013). *Smoothing Spline ANOVA Models*. New York: Springer.

Guo, C. 郭承禹. (2015). 论吴语、湘语中的"浊音走廊"说和元音高化现象 [On the corridor of voiced sounds and the vowel rising shift in the Wu and Xiang dialects]. *嘉兴学院学报* [Journal of Jiaxing University], *27*(1), 103–108.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report* (*MIT Research Laboratory of Electronics*), *101*, 198–213.

Hall, T. A. (2001). Introduction: Phonological representations and phonetic implementation of distinctive features. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 1–50). Berlin: Mouton de Gruyter.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America*, *101*(1), 466–481.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*(1), 425–441.

Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *The Journal of the Acoustical Society of America*, *112*(3), 1158–1182.

Hansson, G. Ó. (2004). Tone and voicing agreement in Yabem. *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 318–331. Davis, USA: Cascadilla Press Somerville, MA.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Harris, J. W. (1969). Sound change in Spanish and the theory of markedness. *Language*, *45*(3), 538–552.

Hashimoto, M. 桥本万太郎. (1985). *语言地理类型学* [Linguistic geographical typology] (Z. Yu 余志鸿, Trans.). Beijing: Peking University Press.

Haudricourt, A. G. (1954). De l'origine des tons en vietnamien [The origin of tones in Vietnamese] (M. Brunelle, Trans.). *Journal Asiatique*, *242*, 69–82. Retrieved from https://halshs.archives-ouvertes.fr/halshs-01678018/document

Haudricourt, A. G. (1972). Two-way and three-way splitting of tonal systems in some Far Eastern languages. In J. G. Harris, & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 58–86). Bangkok: Central Institute of the English Language.

Henrich, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America*, *115*(3), 1321–1332.

Herbst, C. (2004). *Evaluation of various methods to calculate the EGG contact quotient* (Diploma Thesis in Music Acoustics). Department of Speech, Music and Hearing, Stockholm.

Herbst, C., & Ternström, S. (2006). A comparison of different methods to measure the EGG contact quotient. *Logopedics Phoniatrics Vocology*, *31*(3), 126–138.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch and intensity of voice. *Folia Phoniatrica et Logopaedica*, *22*(1), 1–20.

Hirayama, H. 平山久雄. (2010). 次阴调形成的声学原因—以吴江方言为例 [On the phonetic cause of the evolution of the Ci-yin tones – the case of the Wujiang dialects]. *中国方言学报* [Journal of Chinese Dialectology], *2*, 17–23.

Ho, D.-A. 何大安. (1989). 送气分调及其相关问题 [The Tone-split by aspiration and other related problems]. *中央研究院历史语言所研究集刊* [Bulletin of the Institute of History and Philology Academia Sinica], *60*(4), 765–778.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 77–111). New York: Academic Press.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*(1), 37–58.

Honeybone, P. (2002). *Germanic obstruent lenition: Some mutual implications of theoretical and historical phonology* (PhD Thesis). University of Newcastle upon Tyne, Newcastle.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: The case of laryngeal specifications. In M. van Oostendorp, & J. van de Weijer (Eds.), *The internal organization of phonological segments* (pp. 319–354). Berlin: Mouton de Gruyter.

House, D., & Svantesson, J.-O. (1996). Tonal timing and vowel onset characteristics in Thai. *Pan-Asiatic Linguistics: Proceedings of the 4th International Symposium on Languages and Linguistics*, *1*, 104–113. Bangkok, Thailand: Mahidol University of Salaya.

Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice*, *9*(2), 163–172.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30*(3), 129–148.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Hu, F. 胡方. (2007). 论宁波方言和苏州方言前高元音的区别特征—兼谈高元音继续高化现象 [On the distinctive features for the high front vowels in Ningbo and Suzhou Wu dialects with reference to sound changes of high vowels]. *中国语文* [Studies of the Chinese Language], (5), 455–465.

Hu, F., & Ling, F. (2019). Fricative vowels as an intermediate stage of vowel apicalization. *Language and Linguistics*, *20*(1), 1–14.

Huang, Y. 黄燕. (2007). 古泥来母字在现代汉语方言中的分混情况 [The conditions of mixed and separate Ni Lai initial consonant in the contemporary dialects]. *宿州学院学报* [Journal of Suzhou University], *22*(5), 64–67.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America*, *81*(2), 495–504.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(2), 225–257.

Hyman, L. M. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences*, *31*(2–3), 213–238.

Hyman, L. M. (2018). What is phonological typology? In L. M. Hyman, & F. Plank (Eds.), *Phonological typology* (pp. 1–20). Berlin & Boston: Walter de Gruyter GmbH.

Institute of Linguistics, Chinese Academy of Social Sciences 中国社会科学院语言研究所 (Ed.). (1988). *方言调查字表* [Questionnaire of character for dialect surveys]. Beijing: The Commercial Press.

Institute of Linguistics, Department of Chinese Language and Literature of Peking University 北京大学中国语言文学系语言学教研室 (Ed.). (1989). *汉语方音字汇* [Lexicon of Chinese dialect pronunciations] (2nd edition). Beijing: Script Reform Press.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, *121*(4), 2283–2295.

Iverson, G., & Salmons, J. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*, 369–396.

Jacewicz, E., & Fox, R. A. (2015). Intrinsic fundamental frequency of vowels is moderated by regional dialect. *The Journal of the Acoustical Society of America*, *138*(4), EL405–EL410.

Jessen, M., & Roux, J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics*, *30*(1), 1–52.

Jolliffe, I. T. (2002). *Principal component analysis* (2ed edition). New York: Springer-Verlag.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

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

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*, 161–180.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76–90. https://doi.org/10.1016/j.wocn.2014.03.005

Karlgren, B. (1915–1926). *Études sur la phonologie chinoise* [Studies in Chinese phonology]. Archives d'études orientales 15(1–4). Leiden: Brill.

Kawahara, H., de Cheveigne, A., & Patterson, R. D. (1998). An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT-suite. *Proceedings of the 5th International Conference on Spoken Language Processing*, 659–663. Sydney, Australia.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319.

Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, *41*(4), 191–207.

Keating, P. A., Esposito, C. M., Garellek, M., & Kuang, J. (2010). Phonation contrasts across languages. *UCLA Working Papers in Phonetics*, *108*, 188–202.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168.

Khan, S. ud D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, *40*(6), 780–795.

Khatiwada, R. (2008). Some acoustic cues in the detection of the Nepalese aspiration. *The Journal of the Acoustical Society of America*, *123*(5), 3889–3889.

Kingston, J. (2007). Segmental influences on F0: Automatic or controlled? In *Tones and tunes* (Vol. 2, pp. 171–210). Berlin: Walter de Gruyter GmbH & Co. KG.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*(1), 28–54.

Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 228–246). Oxford: Oxford University Press.

Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, *71*, 326–354.

Kirby, J. P., & Brunelle, M. (2017). Southeast Asian tone in areal perspective. In R. Hickey (Ed.), *The Cambridge handbook of areal linguistics* (pp. 703–731). Cambridge: Cambridge University Press.

Kirby, J. P., & Hyslop, G. (2019). Acoustic analysis of onset voicing in Dzongkha obstruents. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3607–3611). Melbourne, Australia: Australasian Speech Science and Technology Association Inc.

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America*, *140*(4), 2400–2411.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*(4), 686–706.

Kuang, J. (2013a). *Phonation in tonal contrasts* (PhD Thesis). University of California, Los Angeles.

Kuang, J. (2013b). The tonal space of contrastive five level tones. *Phonetica*, *70*(1–2), 1–23.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, *71*, 194–214.

Kurpaska, M. (2010). *Chinese language(s): A look through the prism of the great dictionary of modern Chinese dialects*. Berlin & New York: Walter de Gruyter GmbH & Co. KG.

Labov, W. (1994). *Principles of linguistic change: Internal factors*. Malden & Oxford: Blackwell.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change: A multimedia reference tool*. Berlin: Walter de Gruyter GmbH & Co. KG.

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, *89*(1), 30–65.

Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics*, *71*, 229–248.

Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics*, *1*(1), 73–83.

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. Bless, & J. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351–360). San Diego: College Hill Press.

Ladefoged, P., & Johnson, K. (2010). *A course in phonetics* (6th edition). Boston: Thomson Wadsworth.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lai, Y., & Chen, M. (2011). Aspiration and tone. *The Journal of the Acoustical Society of America*, *129*(4), 2451–2451.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In D. Potter, & D. R. Storoshenko (Eds.), *Proceedings of the 2nd International Conference on East Asian Linguistics* (Vol. 2). Burnaby, Canada: Simon Fraser University.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*(8), 1137–1146.

Lanham, L. W. (1958). The tonemes of Xhosa. *African Studies*, *17*(2), 65–81.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In L. M. Hyman (Ed.), *Consonant types & tones* (Vol. 1, pp. 15–70). Los Angeles: The Linguistics Program University of Southern California.

Lee, S. J. (2008). *Consonant-tone interaction in optimality theory* (PhD Thesis). Rutgers University, New Brunswick. https://doi.org/10.7282/T32F7NS1

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

Lee-Kim, S.-I. (2014). Revisiting Mandarin "apical vowels": An articulatory and acoustic study. *Journal of the International Phonetic Association*, *44*(3), 261–282.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*(4), 419–425.

Lenth, R. (2020). emmeans: Estimaed Marginal Means, aka Least-Squares (Version 1.4.4). Retrieved from https://CRAN.R-project.org/package=emmeans

Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, *14*(7), 641–642. https://doi.org/10.1038/nmeth.4346

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

Li, F. K. (1977). *A handbook of comparative Tai*. Honolulu: University Press of Hawaii.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics*, *54*, 123–150.

Li, Q., Chen, Y., & Xiong, Z. (2019). Tianjin Mandarin. *Journal of the International Phonetic Association*, *49*(1), 109–128. https://doi.org/10.1017/S0025100317000287

Li, X. 李小凡, & Xiang, M. 项梦冰. (2009). *汉语方言学基础教程* [A basic textbook of Chinese dialectology]. Beijing: Peking University Press.

Li, X. 李小凡. (2005). 汉语方言分区方法再认识 [A review of the classification methodology of Chinese dialects]. *方言* [Dialects], (4), 356–363.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics*, *12*(2), 147–155.

Lindblad, P. (1980). Svenskans sje-och tje-ljud i ett allmänfonetisk perspektiv. *Travaux de l'Institut de Linguistique de Lund*, *16*.

Ling, F. (2007). The articulatory and acoustic study of fricative vowels in Suzhou Chinese. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 573–576). Saarbrücken, Germany: Saarland University.

Ling, F. 凌锋. (2011). 苏州话[i]元音的语音学分析 [A phonetic study of vowel [i] in Suzhou Chinese]. *语言学论丛* [Essays on Linguistics], *43*, 177–193.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America*, *57*(6), 1547–1551.

Lisker, L. (1978). In qualified defense of VOT. *Language and Speech*, *21*(4), 375–383.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422.

Löfqvist, A., & McGowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics*, *20*(1), 93–110.

Long, R. 龙润田. (2018). 侗语南部方言第二土语送气不分调成因 [Causes for non-splitting of tones of aspirated consonants in the second dialect of Southern Kam language]. *中央民族大学学报(哲学社会科学版)* [Journal of Minzu University of China (Philosophy and Social Sciences Edition)], *45*(2), 130–135.

Maddieson, I., & Emmorey, K. (1985). Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica*, *42*(4), 163–174.

Maspero, H. (1912). Etudes sur la phonétique historique de la langue annamite. Les initiales. *Bulletin de l'École Française d'Extrême-Orient*, *12*(1), 1–124.

Mathangwane, J. T. (1998). Aspirates: Their development and depression in Ikalanga. *Journal of African Languages and Linguistics*, *19*, 113–135.

Mathes, T. K. (2015). *Consonant-tone interaction in the Khoisan language Tsua* (PhD Thesis). New York University, New York.

Mathes, T. K., & Chebanne, A. M. (2018). An SS ANOVA (Smoothing Spline Analysis of Variance) study of high tone lowering in Tsua. *South African Journal of African Languages*, *38*(2), 137–148.

Mazaudon, M. (2012). Paths to tone in the Tamang branch of Tibeto–Birman (Nepal). In G. De Vogelaer, & G. Seiler (Eds.), *The dialect laboratory: Dialects as a testing ground for theories of language change* (pp. 139–177). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Mazaudon, M., & Michaud, A. (2008). Tonal contrasts and initial consonants: A case study of Tamang, a "missing link" in tonogenesis. *Phonetica*, *65*(4), 231–256.

Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech*, *50*(2), 247–277.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton: CRC Press.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory*, *24*(1), 113–178.

Newell, H. (2018). The syntax–phonology interface in Rule-based Phonology. In S. J. Hannahs, & A. Bosch (Eds.), *The Routledge handbook of phonological theory* (pp. 197–225). Abingdon & New York: Routledge.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Norman, J. (2003). The Chinese dialects: Phonology. In G. Thurgood, & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 72–83). New York: Routledge.

Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York & London: Academic Press.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189–216). New York: Springer–Verlag.

Okada, H. (1991). Japanese. *Journal of the International Phonetic Association*, *21*(2), 94–96.

Pan, W. 潘悟云. (1982). 关于汉语声调发展的几个问题—读王士元先生的 A note on tone development [Several problems in the development of Chinese tones]. *Journal of Chinese Linguistics*, *10*(2), 359–385.

Pan, W. 潘悟云, Li, L. 李龙, & Han, X. 韩夏. (2015). *斐风语言田野调查系统* [FieldPhone: A software for language fieldwork]. Retrieved from http://www.eastling.org/

Pearce, M. (2009). Kera tone and voicing interaction. *Lingua*, *119*(6), 846–864.

Peng, L. (1992). *A unified theory of tone-voice* (PhD Thesis). University of Arizona, Tucson.

Perkins, J. M. (2013). *Consonant-tone interaction in Thai* (PhD Thesis). Rutgers University, New Brunswick.

Pittayaporn, P., & Kirby, J. P. (2017). Laryngeal contrasts in the Tai dialect of Cao Bằng. *Journal of the International Phonetic Association*, *47*(1), 65–85.

Pols, L. C., Tromp, H. R., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America*, *53*(4), 1093–1101.

Pulleyblank, E. G. (1978). The nature of the Middle Chinese tones and their development to Early Mandarin. *Journal of Chinese Linguistics*, *6*(2), 173–203.

Qian, N. 钱乃荣. (1992). *当代吴语研究* [Studies of the contemporary Wu dialects]. Shanghai: Shanghai Education Press.

R Development Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from https://www.r-project.org/

Ren, N. (1992). *Phonation types and consonant distinctions: Shanghai Chinese* (PhD Thesis). University of Connecticut, Storrs.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110.

Rischel, J. (1986). Can "the great tone split" in Thai be phonetically explained? *Annual Reports of the Institute of Phonetics, University of Copenhagen*, *20*, 79–98.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, *6*(4), 343–352.

Rothenberg, M. (1982). Interpolating subglottal pressure from oral pressure. *Journal of Speech and Hearing Disorders*, *47*(2), 219–223.

Rothenberg, M., & Mahshie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech, Language, and Hearing Research*, *31*(3), 338–351.

Sagart, L. (1981). Aspiration-conditioned tone-lowering in Chinese dialects. Presented at the 14th International Conference on Sino-Tibetan Languages and Linguistics, Gainesville, USA.

Sagart, L. (2002). Gan, Hakka and the formation of Chinese dialects. In D.-A. Ho (Ed.), *Dialect variations in Chinese* (pp. 129–154). Taipei: Academia Sinica.

Sagart, L. (2011). Classifying Chinese dialects/Sinitic languages on shared innovations. Presented at the Séminaire Sino-Tibétain du CRLAO, Paris.

Schwartz, G., & Arndt, D. (2018). Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences*, *69*, 98–112.

Schwarz, M., Sonderegger, M., & Goad, H. (2019). Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics*, *73*, 113–127.

Shen, Z. (1994). The tones in the Wujiang dialect. *Journal of Chinese Linguistics*, *22*(2), 278–315.

Shen, Z. (2006). Syllabic nasals in the Chinese dialects. *Bulletin of Chinese Linguistics*, *1*(1), 81–108.

Shen, Z., Wooters, C., & Wang, W. S.-Y. (1987). Closure duration in the classification of stops: A statistical analysis. *Ohio State University Working Papers in Linguistics*, *35*, 197–209.

Shi, F. 石锋. (1992). 吴江方言声调格局的分析 [The analysis of the tonal space in the Wujiang dialects]. *方言* [Dialects], (3), 189–194.

Shi, F. 石锋. (1998). 送气声母对于声调的影响 [The influence of aspiration on tones]. *Journal of Chinese Linguistics*, *26*(1), 126–145.

Shi, F. 石锋. (2008). *语音格局: 语音学与音系学的交汇点* [Phonetic space: The intersection between phonetics and phonology]. Beijing: The Commercial Press.

Shi, R. 石汝杰. (1998). 汉语方言中高元音的强摩擦倾向 [The friction of the high vowel in Chinese dialects]. *语言研究* [Studies in Language and Linguistics], (1), 100–109.

Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1846–1849. Hong Kong, China: City University of Hong Kong.

Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, *40*(3), 477–490.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, *12*(2), 80–102.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, *25*(4), 471–500.

Song, J. J. (2018). *Linguistic typology*. Oxford: Oxford University Press.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *The Journal of the Acoustical Society of America*, *50*(4B), 1180–1192.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, *3*(4), 215–233.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, *65*(1), 81–106.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell, & D. H. Klatt (Eds.), *Symposium on invariance and variability of speech processes* (pp. 426–449). Hillsdale: Lawrence Erlbaum Associates.

Stevens, K. Noble, & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19.

Strazny, P. (2003). Depression in Zulu. In J. van de Weijer, V. J. van Heuven, & H. van der Hulst (Eds.), *The phonological spectrum: Segmental structure* (Vol. 1, pp. 223–240). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Sun, J. T.-S. (2003). Variegated tonal developments in Tibetan. In D. Bradley, R. J. LaPolla, B. Michialovsky, & G. Thurgood (Eds.), *Language variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff* (pp. 35–52). Canberra: Pacific Linguistics.

Svantesson, J.-O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics*, *14*(1), 53–70.

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, *75*(3), 983–992.

Tang, K. E. (2008). *The phonology and phonetics of consonant-tone interaction* (PhD Thesis). University of California, Los Angeles.

Tehrani, H. (2009). *EGGWorks: A program for automated analysis of EGG signals*. Retrieved from http://phonetics.linguistics.ucla.edu/facilities/physiology/egg.htm

Thavisak, A. (2004). F0 behavior of vowels influenced by aspirated and unaspirated initials in Southeast Asian languages: Implications for tonogenesis theories. *Proceedings of the 11th Annual Meeting of the Southeast Asian Linguistics Society*, 721–730. Nakornpatom, Thailand: Mahidol University.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, *19*(2), 333–363.

Thurgood, G. (2007). Tonogenesis revisited: Revising the model and the analysis. In J. G. Harris, S. Burusphat, & J. E. Harris (Eds.), *Studies in Tai and Southeast Asian linguistics* (pp. 263–291). Bangkok: Ek Phim Thai Co.

Tian, J., & Kuang, J. (2019). The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association*. https://doi.org/10.1017/S0025100319000148

Ting, P.-H. 丁邦新. (1982). 汉语方言区分的条件 [Phonological features for classification of the Chinese dialects]. *清华学报* [Tsing Hua Journal of Chinese Studies], *14*(1), 2.

Ting, P.-H. 丁邦新. (1984). 吴语声调之研究 [A study of tones in Wu Chinese]. *中央研究院历史语言所研究集刊* [Bulletin of the Institute of History and Philology Academia Sinica], *55*(4), 755–788.

Titze, I. R. (1990). Interpretation of the electroglottographic signal. *Journal of Voice*, *4*(1), 1–9.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.

Traill, A., Khumalo, J. S., & Fridjhon, P. (1987). Depressing facts about Zulu. *African Studies*, *46*(2), 255–274.

Trask, R. L. (1996). *Historical Linguistics*. London: Arnold.

Tsuji, N. 辻伸久. (1979). 湖南諸方言の分類と分布 [The classification and distribution of the Hunan dialects]. *中国語学* [Bulletin of the Chinese Linguistic Society of Japan], *1979*(226), 1–12.

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 225–255.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research*, *3*, 1–28.

Vaissière, J. (2011). On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. *Proceedings of the 17th International Congress of Phonetic Sciences*, 52–59. Hong Kong, China: City University of Hong Kong.

Vu, V. Q. (2011). *ggbiplot: A ggplot2 based biplot*. Retrieved from https://github.com/vqv/ggbiplot

Vu-Ngoc, T., d'Alessandro, C., & Michaud, A. (2005). Using open quotient for the characterisation of Vietnamese glottalised tones. *Proceedings of Eurospeech/Interspeech 2005*, 2885–2889. Lisbon, Portugal: International Speech Communication Association.

Wang, L. 王力. (1985). *汉语语音史* [The history of Chinese philology]. Beijing: China Social Sciences Press.

Wang, P. 汪平. (1987). 苏州音系再分析 [A reanalysis of phonology in the Suzhou dialect]. *语言研究* [Studies in Language and Linguistics], (1), 41–48.

Wang, P. 汪平. (2008). 吴江方言声调再讨论 [Tones in the Wujiang dialects: A revisit]. *中国语文* [Studies of the Chinese Language], (5), 455–461.

Wang, P. 汪平. (2010). *吴江市方言志* [A report on a survey of the Wujiang city dialects]. Shanghai: Shanghai Social Sciences Press.

Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., … Rogers, J. (2017). gdata: Various R programming tools for data manipulation (Version 2.18.0). Retrieved from https://CRAN.R-project.org/package=gdata

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. Lehmann, & Y. Malkiel (Eds.), *Directions for historical linguistics*. Austin: University of Texas Press.

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

Wiese, R. (1997). Underspecification and the description of Chinese vowels. In J. Wang, & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 219–249). Berlin & New York: Mouton de Gruyter.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, *1*(1), 7–18. https://doi.org/10.1093/jole/lzv003

Wirz, S. L., & Anthony, J. (1979). The use of the voiscope in improving the speech of profoundly deaf children. *British Journal of Disorders of Communication*, *14*(2), 137–151.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge: Cambridge University Press.

Wurm, S. A., Li, R., Baumann, T., & Lee, M. W. (1987). *Language atlas of China*. Hong Kong: Longman.

Xiang, X. 向熹. (1960). 湖南双峰县方言 [Hunan Shuangfeng dialect]. *语言学论丛* [Essays on Linguistics], *4*, 134–171.

Xiong, Z. 熊正辉. (1979). 南昌方言的声调及其演变 [Tones and its evolution in the Nanchang dialect]. *方言* [Dialects], (4), 275-283.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165–181.

Xu, Y. 徐越. (2006). 嘉善方言的气流分调 [The tone-split by aspiration of the Jiashan dialect]. *语言研究* [Studies in Language and Linguistics], *26*(3), 77–80.

Xu, Yue 徐越. (2013). 汉语方言中的气流分调现象 [Tone-split by aspiration in Chinese dialects]. *中国语文* [Studies of the Chinese Language], (3), 265–272.

Xu, Y. 徐越. (2014). 吴语次清弛化分调的特点 [Types of tonal split in Wu caused by slack-voicing of voiceless aspiration]. *语言研究* [Studies in Language and Linguistics], 34(2), 54–59.

Xu, Yanhong 徐燕红. (2013). *吴江方言声调实验研究* [An experimental study in Wujiang citation tones] (MA Thesis). Shanghai University, Shanghai.

Xu, Z. 许珍. (2009). *吴江方言声调研究* [A study of Wujiang citation tones] (MA Thesis). Shanghai Normal University, Shanghai.

Yang, H.-F. 杨秀芳. (1989). 论汉语方言中全浊声母的清化[The devoicing of obstruents among Chinese dialects]. *汉学研究* [Chinese Studies], *7*(2), 41–74.

Yang, M., & Sundara, M. (2019). Cue-shifting between acoustic cues: Evidence for directional asymmetry. *Journal of Phonetics*, *75*, 27–42.

Yang, S.-F. 杨时逢. (1974). *湖南方言调查报告* [Report on a survey of the dialects of Hunan]. Taipei: Institute of History and Philology, Academia Sinica.

Ye, X. 叶祥苓. (1983). 吴江方言声调再调查 [A reinvestigation of tones in the Wujiang dialects]. *方言* [Dialects], (1), 32–35.

Ye, X. 叶晓锋. (2011). *汉语方言语音的类型学研究* [A typological study of phonology in Chinese dialects] (PhD Thesis). Fudan University, Shanghai.

Yip, M. (1995). Tone in east Asian languages. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 476–494). Cambridge, Mass. & Oxford: Blackwell.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yokonishi, H., Imagawa, H., Sakakibara, K.-I., Yamauchi, A., Nito, T., Yamasoba, T., & Tayama, N. (2016). Relationship of various open quotients with acoustic property, phonation types, fundamental frequency, and intensity. *Journal of Voice*, *30*(2), 145–157.

Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, *24*(1), 187–214.

Yuan, J. 袁家骅. (1960). *汉语方言概要* [A synopsis of Chinese dialects]. Beijing: Script Reform Press.

Yuan, X. 袁先锋. (2005). *湖南省双峰县花门镇和永丰镇方言的语音比较研究* [The study of phonetic comparison between the dialects of Huamen town and Yongfeng town in Shuangfeng county, Hunan province] (MA Thesis). Hunan Normal University, Changsha.

Zee, E. (1980). The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics*, *49*, 90–97.

Zeng, T. (2019). The Xiangxiang dialect of Chinese. *Journal of the International Phonetic Association*, 1–24. https://doi.org/10.1017/S002510031800035X

Zhang, G. 张拱贵, & Liu, D. 刘丹青. (1983). 吴江方言声调初步调查 [A preliminary investigation of tones in the Wujiang dialects]. *南京师大学报(社会科学版)* [Journal of Nanjing Normal University (Social Science Edition)], (3), 36–40.

Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics*, *54*, 169–201.

Zhao, R. 赵日新. (2007). 汉语方言中的[i]>[ɿ] [Sound change of [i]>[ɿ] in Chinese dialects]. *中国语文* [Studies of the Chinese Language], (1), 46–54.

Zhengzhang, S., & Zheng, W. (2015). Wu dialect. In W. S.-Y. Wang, & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 189–199). New York: Oxford University Press.

Zhong, J. 钟江华, & Chen, L. 陈立中. (2012). 现代湘语和吴语浊音声母发音特征的比较. *湖北民族学院学报(哲学社会科学版)* [Journal of Hubei University for Nationalities (Philosophy and Social Sciences)], *30*(04), 126–129.

Zhou, H. 周洪涛. (2005). *永丰镇方言的语音研究* [The study on phonology of Yongfeng town] (MA Thesis). Jinan University, Guangzhou.

Zhou, Y., & Kirby, J. P. (2019). An acoustic study of Zhajin Gan tone. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1193–1197). Melbourne, Australia: Australasian Speech Science and Technology Association Inc.

Zhou, Z. 周振鹤, & You, R. 游汝杰. (1985). 湖南省方言区画及其历史背景 [The classification of the Hunan dialects and its historical background]. *方言* [Dialects], (4), 257–272.

Zhou, Z. 周振鹤, & You, R. 游汝杰. (1986). *方言与中国文化* [Dialects and Chinese culture]. Shanghai: Shanghai People's Publishing House.

Zhu, X. 朱晓农. (2004a). 亲密与高调:对小称调、女国音、美眉等语言现象的生物学解释 [Intimacy and high pitch: A biological explanation for the use of high pitch in diminutives, 'female Mandarin', 'Taiwan sisters', etc]. *当代语言学* [Contemporary Linguistics], (3), 193–222.

Zhu, X. 朱晓农. (2004b). 汉语元音的高顶出位 [Sound changes of high vowels in Chinese dialects]. *中国语文* [Studies of the Chinese Language], (5), 440–451.

Zhu, X. 朱晓农, & Xu, Y. 徐越. (2009). 弛化:探索吴江次清分调的原因 [Slack voice as the cause of tone split in Wujiang: A phonetic investigation into the pitch contour of the syllables with a voiceless aspirated onset in the Wujiang (Songling) Wu dialect]. *中国语文* [Studies of the Chinese Language], (4), 324–332.

Zhu, X. 朱晓农, Wei, M. 韦名应, & Wang, J. 王俊芳. (2016). 十五调和气调:侗语榕江县口寨方言案例 [Fifteen tones and breathy tones: A case of Kouzhai Dong in Rongjiang county]. *民族语文* [Minority Languages of China], (5), 12–24.

Zhu, X. 朱晓农, & Zou, X. 邹晓玲. (2017). 清浊同调还是气声分调—在音节学和类型学普适理论中安排湘语白仓话的声调事实 [One tone with voicing contrast or two tones with breathy contrast]. *南方语言学* [South Linguistics], *12*, 1–11.

# Appendix

## Appendix I

### I-1 Stimulus list of Lili Wu Chinese in Chapter 3

| MC tonal category | Unaspirated (U) | Aspirated (A) | Voiced (D) |
|---|---|---|---|
| *Ping* (I) | 低 tɿ̞ 'low' | 梯 tʰɿ̞ 'ladder' | 提 dɿ̞ 'to mention' |
| | 单 tɛ 'sole' | 瘫 tʰɛ 'paralysis' | 台 dɛ 'platform' |
| | 刀 tɑ 'knife' | 涛 tʰɑ 'billow' | 桃 dɑ 'peach' |
| *Shang* (II) | 底 tɿ̞ 'bottom' | 体 tʰɿ̞ 'body' | 弟 dɿ̞ 'younger brother' |
| | 胆 tɛ 'gallbladder' | 毯 tʰɛ 'mat' | 淡 dɛ 'light' |
| | 岛 tɑ 'island' | 讨 tʰɑ 'to ask for' | 稻 dɑ 'rice' |
| *Qu* (III) | 渧 tɿ̞ 'to drop' | 替 tʰɿ̞ 'to replace' | 地 dɿ̞ 'ground' |
| | 对 tɛ 'right' | 退 tʰɛ 'to retreat' | 代 dɛ 'dynasty' |
| | 到 tɑ 'to arrive' | 套 tʰɑ 'case' | 盗 dɑ 'robber' |
| *Ru* (IV) | 滴 tɪʔ 'drop' | 贴 tʰɪʔ 'to paste' | 敌 dɪʔ 'enemy' |
| | 得 tʌʔ 'to get' | 脱 tʰʌʔ 'to take off' | 特 dʌʔ 'specially' |
| | 搭 taʔ 'and' | 塔 tʰaʔ 'pagoda' | 达 daʔ 'to extend' |

## I-2 Final models to calculate the results presented in Table 3.4

| Generation-Category | Fixed structure | Random structure |
|---|---|---|
| Old-I&II; Middle-I; | (ot1 + ot2) * Consonant + Vowel + Repetition + Gender | (ot1 + ot2 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1+ ot2 \| Item) |
| Young-I | (ot1 + ot2) * Consonant + Vowel + Repetition + Gender | (ot1 + ot2 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) |
| Middle-II; Young-III | (ot1 + ot2) * Consonant + Vowel + Repetition | (ot1 + ot2 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1 \| Item) |
| Young-II; Old-III; Middle-IV; Young-IV | (ot1 + ot2) * Consonant + Vowel + Repetition | (ot1 + ot2 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1+ ot2 \| Item) |
| Middle-III | (ot1 + ot2) * Consonant + Vowel | (ot1 + ot2 \| Speaker) + (ot1 \| Speaker: Consonant) + (1 \| Item) |
| Old-IV | (ot1 + ot2) * Consonant + Vowel | (ot1+ ot2 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1 + ot2 \| Item) |

## I-3a Final models to calculate the results presented in Table 3.5

| Generation | Fixed structure | Random structure |
|---|---|---|
| Old & Middle-aged | Category + Vowel + Repetition | (1 + Category \| Speaker) + (1 \| Item) |
| Young | Category + Vowel | (1 + Category \| Speaker) + (1 \| Item) |

## I-3b Between-Category comparisons of the DOR/DOS ratio by Generation

A mixed model fitted to the DOR/DOS ratio with fixed factors Category, Generation, and their interaction was run. The random intercept for Speaker was also included. This was the maximal model which converged. Following the method used in Kirby (2018) and Kirby and Hyslop (2019), pairwise comparisons of the marginal means estimate were calculated using the *emmeans* function (Lenth, 2020). Fractional degrees of freedom computed using the Satterthwaite method, with *p*-values adjusted using the Tukey method for comparing a family of 4 estimates. n.s.: not significant.

| Contrast | Estimate | *SE* | *df* | *t*-ratio | *p* |
|---|---|---|---|---|---|
| Generation: Old | | | | | |
| I – II | .02 | .01 | 1379 | 2.24 | n.s. |
| I – III | .02 | .01 | 1379 | 1.69 | n.s. |
| I – IV | −0.12 | .01 | 1379 | −13.18 | < .001 |
| II – III | −0.01 | .01 | 1379 | −0.55 | n.s. |
| II – IV | −0.14 | .01 | 1379 | −15.44 | < .001 |
| III – IV | −0.14 | .01 | 1379 | −14.9 | < .001 |
| Generation: Middle-aged | | | | | |
| I – II | .01 | .01 | 1379 | .11 | n.s. |
| I – III | .01 | .01 | 1379 | 1.6 | n.s. |
| I – IV | −0.11 | .01 | 1379 | −12.15 | < .001 |
| II – III | .01 | .01 | 1379 | 1.5 | n.s. |
| II – IV | −0.11 | .01 | 1379 | −12.26 | < .001 |
| III – IV | −0.12 | .01 | 1379 | −13.75 | < .001 |
| Generation: Young | | | | | |
| I – II | .02 | .01 | 1379 | 1.68 | n.s. |
| I – III | .01 | .01 | 1379 | .62 | n.s. |
| I – IV | −0.06 | .01 | 1379 | −6.21 | < .001 |
| II – III | −0.01 | .01 | 1379 | −1.06 | n.s. |
| II – IV | −0.07 | .01 | 1379 | −7.89 | < .001 |
| III – IV | −0.06 | .01 | 1379 | −6.83 | < .001 |

## I-4a Final models to calculate the results presented in Table 3.7

| Generation-Category-Position | Fixed structure | Random structure |
|---|---|---|
| Old-I-P1 | Consonant + Vowel | (1 \| Speaker) + (1 \| Item) |
| Old-I-P2; Middle-I-P1; Middle-IV-P1 | Consonant + Vowel | (1 \| Speaker) |
| Middle-I-P2; Young-I-P1; Young-III-P1 | Consonant + Gender | (1 \| Speaker) |
| Old-II-P1; Middle-II-P1 | Consonant | (1 \| Speaker) + (1 \| Item) |
| Young-II-P1 | Consonant + Gender | (1 \| Speaker) + (1 \| Item) |
| Middle-II-P2 | Consonant + Vowel + Gender | (1 \| Speaker) + (1 \| Item) |
| Old-III-P1&P2; Old-IV-P1 | Consonant + Vowel | (1 + Consonant \| Speaker) |
| Middle-III-P1 | Consonant + Vowel | (1 + Consonant \| Speaker) + (1 \| Item) |
| Middle-III-P2 | Consonant + Vowel + Gender | (1 + Consonant \| Speaker) |
| Old-IV-P2 | Consonant + Vowel + Repetition | (1 + Consonant \| Speaker) |
| Young-IV-P1 | Consonant + Vowel + Gender | (1 \| Speaker) |
| Others | Consonant did not significantly improve the model fit. | |

## I-4b Between-Consonant comparisons of H1*–H2* by Generation, Category, and Position

A mixed model fitted to H1*–H2* with fixed factors Consonant, Generation, Category, and their full-fledged interactions was run. The random intercept for Speaker was also included. This was the maximal model which converged. Pairwise comparisons of the marginal means estimate were calculated with fractional degrees of freedom computed using the Satterthwaite method. The *p*-values were adjusted using the Tukey method for comparing a family of 3 estimates. n.s.: not significant.

| Contrast | Estimate | *SE* | *df* | *t*-ratio | *p* |
|---|---|---|---|---|---|
| Generation: Old, Category: I, Position: 1 | | | | | |
| A – U | 2.63 | .49 | 12891 | 5.34 | < .001 |
| A – D | −0.74 | .49 | 12891 | −1.51 | n.s. |
| U – D | −3.37 | .49 | 12891 | −6.85 | < .001 |
| Generation: Middle-aged, Category: I, Position: 1 | | | | | |
| A – U | 1.37 | .49 | 12891 | 2.79 | < .001 |
| A – D | −0.74 | .49 | 12891 | −1.11 | n.s. |
| U – D | −3.37 | .49 | 12891 | −3.9 | < .001 |
| Generation: Young, Category: I, Position: 1 | | | | | |
| A – U | 2.17 | .49 | 12891 | 4.41 | < .001 |
| A – D | 1.85 | .49 | 12891 | 3.76 | < .001 |
| U – D | −0.32 | .49 | 12891 | −0.65 | n.s. |
| Generation: Old, Category: II, Position: 1 | | | | | |
| A – U | 2.04 | .49 | 12891 | 4.13 | < .001 |
| A – D | −0.04 | .49 | 12891 | −0.07 | n.s. |
| U – D | −2.08 | .49 | 12891 | −4.21 | < .001 |
| Generation: Middle-aged, Category: II, Position: 1 | | | | | |
| A – U | 2.85 | .49 | 12891 | 5.79 | <.001 |
| A – D | .39 | .49 | 12891 | .79 | n.s. |
| U – D | −2.46 | .49 | 12891 | −5.01 | < .001 |
| Generation: Young, Category: II, Position: 1 | | | | | |
| A – U | 1.89 | .49 | 12891 | 3.84 | < .001 |
| A – D | 2.41 | .49 | 12891 | 4.89 | < .001 |
| U – D | .52 | .49 | 12891 | 1.05 | n.s. |
| Generation: Old, Category: III, Position: 1 | | | | | |
| A – U | 2.12 | .49 | 12891 | 4.31 | < .001 |
| A – D | −0.72 | .49 | 12891 | −1.47 | n.s. |
| U – D | −2.84 | .49 | 12891 | −5.77 | < .001 |

Continued.

| | | | | | |
|---|---|---|---|---|---|
| Generation: Middle-aged, Category: III, Position: 1 | | | | | |
| A – U | 2.16 | .49 | 12891 | 4.39 | < .001 |
| A – D | −0.05 | .49 | 12891 | −0.1 | n.s. |
| U – D | −2.21 | .49 | 12891 | −4.5 | < .001 |
| Generation: Young, Category: III, Position: 1 | | | | | |
| A – U | 1.67 | .49 | 12891 | 3.39 | < .01 |
| A – D | 1.36 | .49 | 12891 | 2.77 | < .05 |
| U – D | −0.3 | .49 | 12891 | −0.62 | n.s. |
| Generation: Old, Category: IV, Position: 1 | | | | | |
| A – U | 4.75 | .49 | 12891 | 9.65 | < .001 |
| A – D | 3.12 | .49 | 12891 | 6.34 | < .001 |
| U – D | −1.63 | .49 | 12891 | −3.31 | n.s. |
| Generation: Middle-aged, Category: IV, Position: 1 | | | | | |
| A – U | 3.81 | .49 | 12891 | 7.73 | < .001 |
| A – D | 2.97 | .49 | 12891 | 6.03 | < .001 |
| U – D | −0.84 | .49 | 12891 | −1.72 | n.s. |
| Generation: Young, Category: IV, Position: 1 | | | | | |
| A – U | 3.3 | .49 | 12891 | 6.7 | < .001 |
| A – D | 4.3 | .49 | 12891 | 8.74 | < .001 |
| U – D | 1 | .49 | 12891 | 2 | n.s. |
| Generation: Old, Category: I, Position: 2 | | | | | |
| A – U | 1.05 | .49 | 12891 | 2.14 | n.s. |
| A – D | −0.51 | .49 | 12891 | −1.04 | n.s. |
| U – D | −1.56 | .49 | 12891 | −3.18 | < .01 |
| Generation: Middle-aged, Category: I, Position: 2 | | | | | |
| A – U | .59 | .49 | 12891 | 1.2 | n.s. |
| A – D | −0.87 | .49 | 12891 | −1.77 | n.s. |
| U – D | −1.46 | .49 | 12891 | −2.97 | < .01 |
| Generation: Young, Category: I, Position: 2 | | | | | |
| A – U | .84 | .49 | 12891 | 1.71 | n.s. |
| A – D | .51 | .49 | 12891 | 1.04 | n.s. |
| U – D | −0.33 | .49 | 12891 | −0.68 | n.s. |
| Generation: Old, Category: II, Position: 2 | | | | | |
| A – U | .84 | .49 | 12891 | 1.7 | n.s. |
| A – D | −0.2 | .49 | 12891 | −0.41 | n.s. |
| U – D | −1.04 | .49 | 12891 | −2.1 | n.s. |
| Generation: Middle-aged, Category: II, Position: 2 | | | | | |
| A – U | 1.52 | .49 | 12891 | 3.08 | < .01 |
| A – D | .25 | .49 | 12891 | .51 | n.s. |
| U – D | −1.26 | .49 | 12891 | −2.57 | < .05 |
| Generation: Young, Category: II, Position: 2 | | | | | |

Continued.

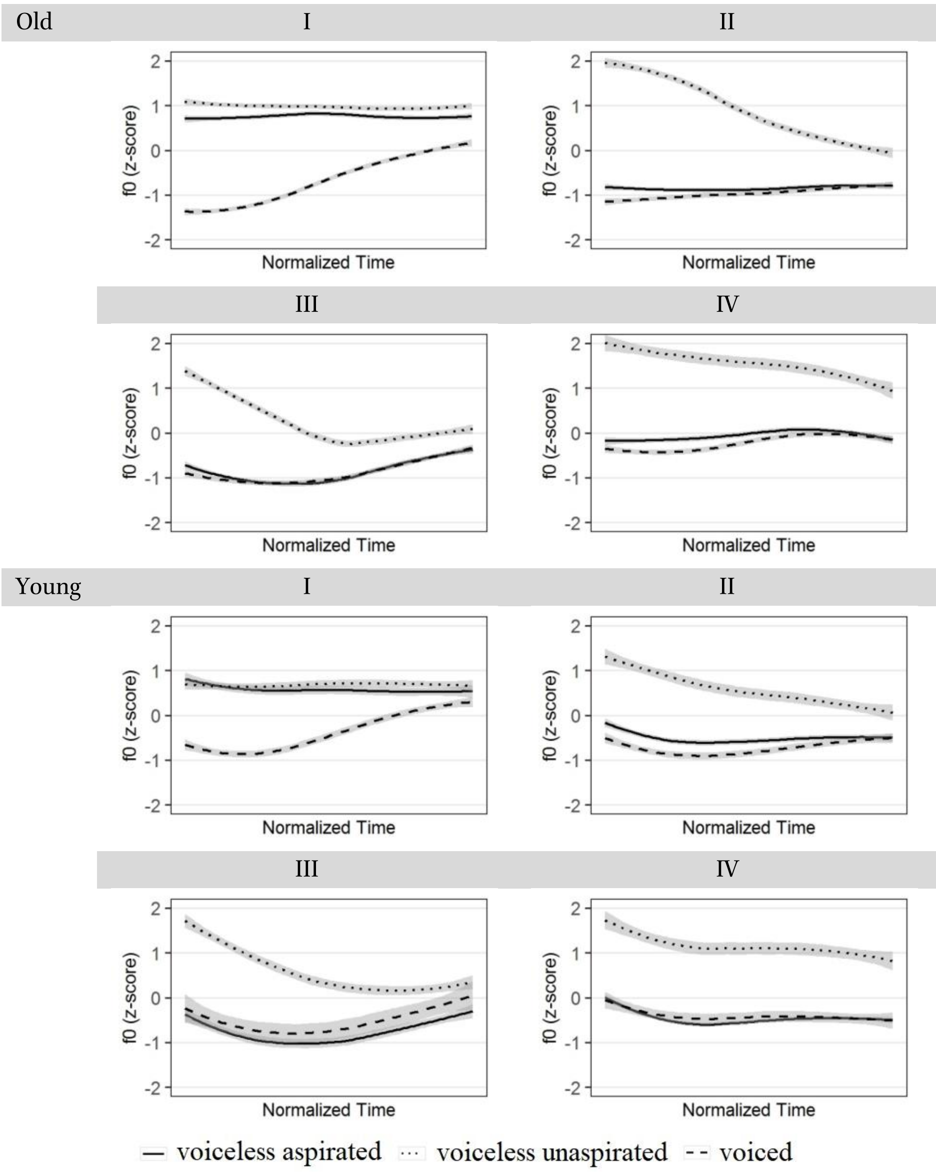
| | | | | | |
|---|---|---|---|---|---|
| A – U | .68 | .49 | 12891 | 1.37 | n.s. |
| A – D | .05 | .49 | 12891 | 1.02 | n.s. |
| U – D | −0.17 | .49 | 12891 | −0.35 | n.s. |
| Generation: Old, Category: III, Position: 2 | | | | | |
| A – U | 1.16 | .49 | 12891 | 2.35 | n.s. |
| A – D | −0.38 | .49 | 12891 | −0.77 | n.s. |
| U – D | −1.54 | .49 | 12891 | −3.12 | < .01 |
| Generation: Middle-aged, Category: III, Position: 2 | | | | | |
| A – U | 1.06 | .49 | 12891 | 2.15 | n.s. |
| A – D | −0.55 | .49 | 12891 | −1.12 | n.s. |
| U – D | −1.61 | .49 | 12891 | −3.27 | < .01 |
| Generation: Young, Category: III, Position: 2 | | | | | |
| A – U | 1.01 | .49 | 12891 | 2.21 | n.s. |
| A – D | 1.02 | .49 | 12891 | 2.07 | n.s. |
| U – D | −0.07 | .49 | 12891 | −0.15 | n.s. |
| Generation: Old, Category: IV, Position: 2 | | | | | |
| A – U | 1.12 | .49 | 12891 | 2.27 | n.s. |
| A – D | .94 | .49 | 12891 | 1.9 | n.s. |
| U – D | −0.18 | .49 | 12891 | −0.37 | n.s. |
| Generation: Middle-aged, Category: IV, Position: 2 | | | | | |
| A – U | .89 | .49 | 12891 | 1.81 | n.s. |
| A – D | .62 | .49 | 12891 | 1.25 | n.s. |
| U – D | −0.27 | .49 | 12891 | −0.55 | n.s. |
| Generation: Young, Category: IV, Position: 2 | | | | | |
| A – U | .89 | .49 | 12891 | 1.8 | n.s. |
| A – D | .83 | .49 | 12891 | 1.69 | n.s. |
| U – D | −0.05 | .49 | 12891 | −0.11 | n.s. |
| Generation: Old, Category: I, Position: 3 | | | | | |
| A – U | .53 | .49 | 12891 | 1.08 | n.s. |
| A – D | −0.06 | .49 | 12891 | −0.12 | n.s. |
| U – D | −0.59 | .49 | 12891 | −1.2 | n.s. |
| Generation: Middle-aged, Category: I, Position: 3 | | | | | |
| A – U | .22 | .49 | 12891 | .46 | n.s. |
| A – D | −0.39 | .49 | 12891 | −0.79 | n.s. |
| U – D | −0.61 | .49 | 12891 | −1.24 | n.s. |
| Generation: Young, Category: I, Position: 3 | | | | | |
| A – U | .71 | .49 | 12891 | 1.45 | n.s. |
| A – D | .43 | .49 | 12891 | .86 | n.s. |
| U – D | −0.29 | .49 | 12891 | −0.59 | n.s. |
| Generation: Old, Category: II, Position: 3 | | | | | |
| A – U | .69 | .49 | 12891 | 1.39 | n.s. |

Continued.

| | | | | | |
|---|---|---|---|---|---|
| A – D | .01 | .49 | 12891 | .01 | n.s. |
| U – D | −0.68 | .49 | 12891 | −1.38 | n.s. |
| Generation: Middle-aged, Category: II, Position: 3 | | | | | |
| A – U | .88 | .49 | 12891 | 1.79 | n.s. |
| A – D | .08 | .49 | 12891 | .16 | n.s. |
| U – D | −0.8 | .49 | 12891 | −1.63 | n.s. |
| Generation: Young, Category: II, Position: 3 | | | | | |
| A – U | .76 | .49 | 12891 | 1.54 | n.s. |
| A – D | .51 | .49 | 12891 | 1.03 | n.s. |
| U – D | −0.25 | .49 | 12891 | −0.5 | n.s. |
| Generation: Old, Category: III, Position: 3 | | | | | |
| A – U | .53 | .49 | 12891 | 1.08 | n.s. |
| A – D | .36 | .49 | 12891 | .74 | n.s. |
| U – D | −0.17 | .49 | 12891 | −0.34 | n.s. |
| Generation: Middle-aged, Category: III, Position: 3 | | | | | |
| A – U | .25 | .49 | 12891 | .5 | n.s. |
| A – D | −0.89 | .49 | 12891 | −1.8 | n.s. |
| U – D | −1.13 | .49 | 12891 | −2.3 | n.s. |
| Generation: Young, Category: III, Position: 3 | | | | | |
| A – U | .58 | .49 | 12891 | 1.19 | n.s. |
| A – D | .63 | .49 | 12891 | 1.28 | n.s. |
| U – D | .04 | .49 | 12891 | .09 | n.s. |
| Generation: Old, Category: IV, Position: 3 | | | | | |
| A – U | −0.29 | .49 | 12891 | −0.59 | n.s. |
| A – D | .56 | .49 | 12891 | 1.13 | n.s. |
| U – D | .85 | .49 | 12891 | 1.72 | n.s. |
| Generation: Middle-aged, Category: IV, Position: 3 | | | | | |
| A – U | .21 | .49 | 12891 | .43 | n.s. |
| A – D | .23 | .49 | 12891 | .46 | n.s. |
| U – D | .01 | .49 | 12891 | .03 | n.s. |
| Generation: Young, Category: IV, Position: 3 | | | | | |
| A – U | .52 | .49 | 12891 | 1.05 | n.s. |
| A – D | .54 | .49 | 12891 | 1.1 | n.s. |
| U – D | .02 | .49 | 12891 | .05 | n.s. |

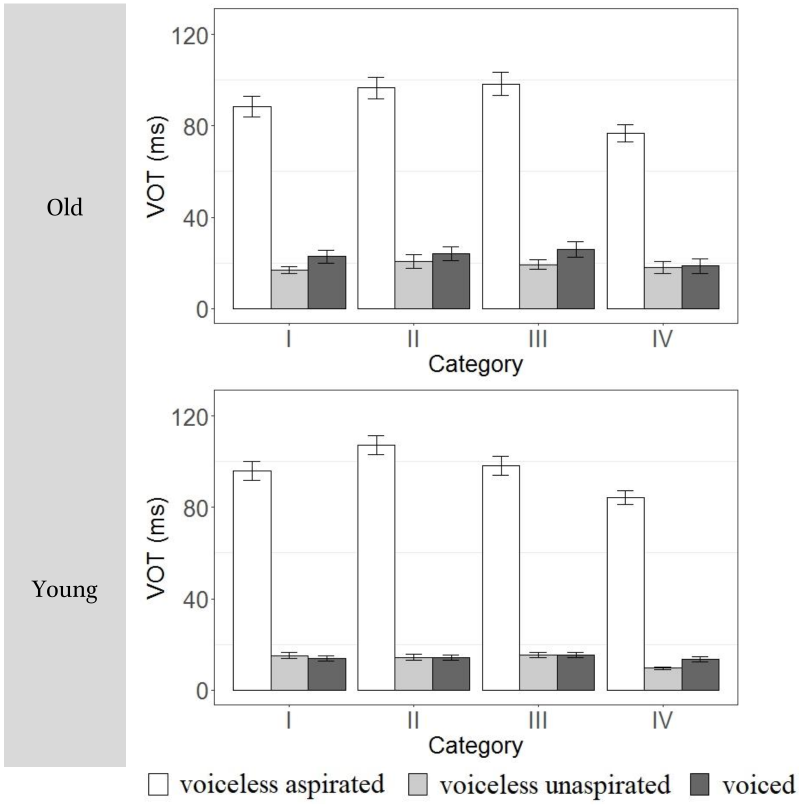
## I-5 Acoustic data of Section 3.4

### a. *f0* contour



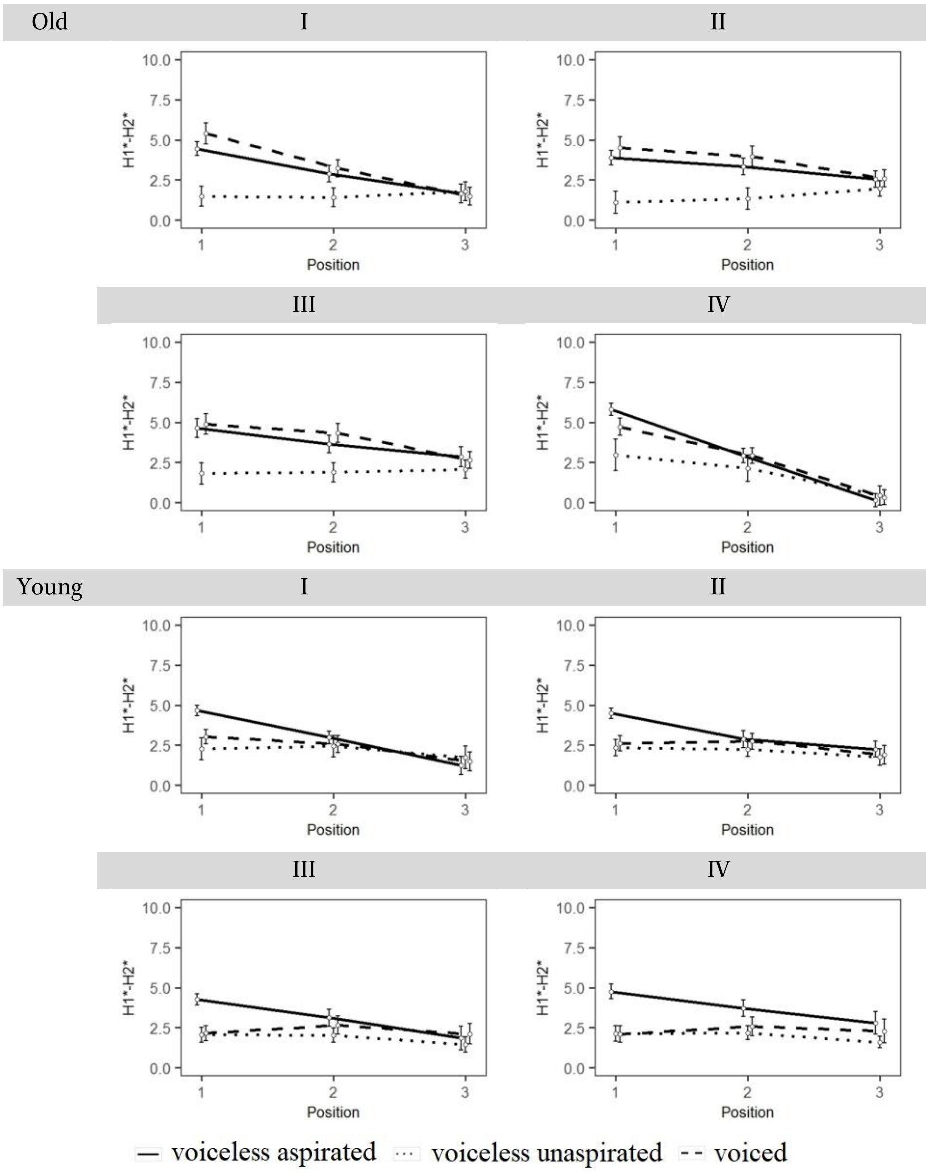

The general pattern of the *f0* contours is similar to the results reported in Section 3.3.6.1. That is, tonal split is absent in the MC *Ping* (I) tonal category. In the other three tonal contexts (i.e., MC *Shang* II, *Qu* III, and *Ru* IV), the *f*0 contours beginning with voiceless aspirated and voiced onsets merge. This pattern holds across generations.

b. VOT




In terms of VOT, the three-way laryngeal contrast in Lili Wu Chinese features a two-way distinction (i.e., short lag vs. long lag). The voiceless aspirated onsets have longer VOT values than the other two counterparts. The voiceless unaspirated and voiced onsets have similar short VOT values. None of the voiced onsets are realized with a lead VOT. This pattern holds across generations and is similar to the results reported in Section 2.3 as well as most Northern Wu dialects (Cao & Maddieson, 1992; Ren, 1992; Gao, 2015).

c. H1*-H2*




The general pattern of the H1*-H2* values is similar to the results reported in Section 3.3.6.3. First, a consistently higher H1*–H2* over one-third of vowels (Position 1) after aspirated onsets is observed. This pattern holds across generations. Second, all phonatory contrasts seem to vanish after the midpoint

(Position 2). This pattern again holds across generations. Third, across all MC tonal categories, the old-generation speakers show more comparable patterns of H1*–H2* of vowels after voiced and aspirated onsets, whereas the young-generation speakers show a minimized H1*–H2* difference, especially after voiced and unaspirated onsets.

## I-6a Final models to calculate the results presented in Table 3.9

| Generation-Category-Position | Fixed structure | Random structure |
|---|---|---|
| Old-I&III-P1 | Consonant + Vowel | (1 \| Speaker) |
| Young-I-P1 | Consonant + Vowel | (1 + Consonant \| Speaker) |
| Old-II-P1 | Consonant + Vowel + Gender | (1 \| Speaker) |
| Young-II-P1 | Consonant + Vowel | (1 + Consonant \| Speaker) + (1 \| Item) |
| Young-III-P1 | Consonant + Gender | (1 + Consonant \| Speaker) |
| Old&Young-IV-P1 | Consonant | (1 + Consonant \| Speaker) |
| Young-IV-P2 | Consonant + Vowel | (1 + Consonant \| Speaker) |
| Others | Consonant did not significantly improve the model fit. | |

## I-6b Between-Consonant comparisons of CQ by Generation, Category, and Position

A mixed model fitted to CQ with fixed factors Consonant, Generation, Category, and their full-fledged interactions was run. The random intercept for Speaker was also included. This was the maximal model which converged. Pairwise comparisons of the marginal means estimate were calculated with fractional degrees of freedom computed using the Satterthwaite method. The *p*-values were adjusted using the Tukey method for comparing a family of 3 estimates. n.s.: not significant.

| Contrast | Estimate | *SE* | *df* | *t*-ratio | *p* |
|---|---|---|---|---|---|
| Generation: Old, Category: I, Position: 1 | | | | | |
| A – U | −0.05 | .01 | 4090 | −4.16 | < .001 |
| A – D | .01 | .01 | 4090 | 1.02 | n.s. |
| U – D | .06 | .01 | 4090 | 5.23 | < .001 |
| Generation: Young, Category: I, Position: 1 | | | | | |
| A – U | −0.04 | .01 | 4090 | −3.47 | < .01 |
| A – D | −0.04 | .01 | 4090 | −3.24 | < .01 |
| U – D | .01 | .01 | 4090 | .24 | n.s. |
| Generation: Old, Category: II, Position: 1 | | | | | |
| A – U | −0.07 | .01 | 4090 | −6.11 | < .001 |
| A – D | .01 | .01 | 4090 | .05 | n.s. |
| U – D | .07 | .01 | 4090 | 6.18 | < .001 |
| Generation: Young, Category: II, Position: 1 | | | | | |
| A – U | −0.06 | .01 | 4090 | −5.26 | < .001 |
| A – D | −0.04 | .01 | 4090 | −3.53 | < .01 |
| U – D | .02 | .01 | 4090 | 1.71 | n.s. |
| Generation: Old, Category: III, Position: 1 | | | | | |
| A – U | −0.05 | .01 | 4090 | −4.07 | < .001 |
| A – D | .01 | .01 | 4090 | .29 | n.s. |
| U – D | .05 | .01 | 4090 | 4.37 | < .001 |
| Generation: Young, Category: III, Position: 1 | | | | | |
| A – U | −0.06 | .01 | 4090 | −4.85 | < .001 |
| A – D | −0.05 | .01 | 4090 | −4.51 | < .001 |
| U – D | .01 | .01 | 4090 | .36 | n.s. |
| Generation: Old, Category: IV, Position: 1 | | | | | |
| A – U | −0.09 | .01 | 4090 | −7.39 | < .001 |
| A – D | .01 | .01 | 4090 | .59 | n.s. |
| U – D | .09 | .01 | 4090 | 7.98 | < .001 |

Continued.

| | | | | | |
|---|---|---|---|---|---|
| Generation: Young, Category: IV, Position: 1 | | | | | |
| A – U | −0.08 | .01 | 4090 | −6.64 | < .001 |
| A – D | −0.07 | .01 | 4090 | −6.33 | < .001 |
| U – D | .01 | .01 | 4090 | .28 | n.s. |
| Generation: Old, Category: I, Position: 2 | | | | | |
| A – U | .01 | .01 | 4090 | .83 | n.s. |
| A – D | .02 | .01 | 4090 | 1.76 | n.s. |
| U – D | .01 | .01 | 4090 | .94 | n.s. |
| Generation: Young, Category: I, Position: 2 | | | | | |
| A – U | −0.01 | .01 | 4090 | −0.08 | n.s. |
| A – D | −0.02 | .01 | 4090 | −1.33 | n.s. |
| U – D | −0.01 | .01 | 4090 | −1.25 | n.s. |
| Generation: Old, Category: II, Position: 2 | | | | | |
| A – U | .01 | .01 | 4090 | .49 | n.s. |
| A – D | .01 | .01 | 4090 | .38 | n.s. |
| U – D | −0.01 | .01 | 4090 | −0.12 | n.s. |
| Generation: Young, Category: II, Position: 2 | | | | | |
| A – U | −0.01 | .01 | 4090 | −1.16 | n.s. |
| A – D | −0.01 | .01 | 4090 | −0.34 | n.s. |
| U – D | .01 | .01 | 4090 | .83 | n.s. |
| Generation: Old, Category: III, Position: 2 | | | | | |
| A – U | −0.01 | .01 | 4090 | −0.89 | n.s. |
| A – D | .01 | .01 | 4090 | .27 | n.s. |
| U – D | .01 | .01 | 4090 | 1.16 | n.s. |
| Generation: Young, Category: III, Position: 2 | | | | | |
| A – U | −0.02 | .01 | 4090 | −1.37 | n.s. |
| A – D | −0.01 | .01 | 4090 | −0.76 | n.s. |
| U – D | .01 | .01 | 4090 | .61 | n.s. |
| Generation: Old, Category: IV, Position: 2 | | | | | |
| A – U | −0.03 | .01 | 4090 | −2.37 | < .05 |
| A – D | .01 | .01 | 4090 | .85 | n.s. |
| U – D | .04 | .01 | 4090 | 3.21 | <.01 |
| Generation: Young, Category: IV, Position: 2 | | | | | |
| A – U | −0.05 | .01 | 4090 | −4.37 | < .001 |
| A – D | −0.03 | .01 | 4090 | −2.75 | < .05 |
| U – D | .02 | .01 | 4090 | 1.6 | n.s. |
| Generation: Old, Category: I, Position: 3 | | | | | |
| A – U | .01 | .01 | 4090 | .62 | n.s. |
| A – D | .01 | .01 | 4090 | 1.11 | n.s. |
| U – D | .01 | .01 | 4090 | .49 | n.s. |
| Generation: Young, Category: I, Position: 3 | | | | | |

Continued.

| | | | | | |
|---|---|---|---|---|---|
| A – U | .01 | .01 | 4090 | .39 | n.s. |
| A – D | .01 | .01 | 4090 | .19 | n.s. |
| U – D | −0.01 | .01 | 4090 | −.21 | n.s. |
| Generation: Old, Category: II, Position: 3 | | | | | |
| A – U | .01 | .01 | 4090 | .46 | n.s. |
| A – D | .01 | .01 | 4090 | .3 | n.s. |
| U – D | −0.01 | .01 | 4090 | −0.16 | n.s. |
| Generation: Young, Category: II, Position: 3 | | | | | |
| A – U | −0.02 | .01 | 4090 | −1.45 | n.s. |
| A – D | −0.01 | .01 | 4090 | −0.34 | n.s. |
| U – D | .01 | .01 | 4090 | 1.12 | n.s. |
| Generation: Old, Category: III, Position: 3 | | | | | |
| A – U | .01 | .01 | 4090 | .23 | n.s. |
| A – D | .01 | .01 | 4090 | .43 | n.s. |
| U – D | .01 | .01 | 4090 | .18 | n.s. |
| Generation: Young, Category: III, Position: 3 | | | | | |
| A – U | −0.01 | .01 | 4090 | −0.3 | n.s. |
| A – D | −0.01 | .01 | 4090 | −0.79 | n.s. |
| U – D | −0.01 | .01 | 4090 | −0.48 | n.s. |
| Generation: Old, Category: IV, Position: 3 | | | | | |
| A – U | −0.01 | .01 | 4090 | −0.21 | n.s. |
| A – D | .01 | .01 | 4090 | .08 | n.s. |
| U – D | .01 | .01 | 4090 | .29 | n.s. |
| Generation: Young, Category: IV, Position: 3 | | | | | |
| A – U | −0.02 | .01 | 4090 | −1.71 | n.s. |
| A – D | −0.01 | .01 | 4090 | −1.07 | n.s. |
| U – D | .01 | .01 | 4090 | .64 | n.s. |

# Appendix II

## II-1 Stimulus list of Shuangfeng Xiang Chinese in Chapter 5

| Unaspirated (U) | Aspirated (A) | Voiced (D) |
|---|---|---|
| 北 pe 'northern' | 撇 $p^h$e 'to put aside' | 陪 be 'to accompany' |
| 滴 ti 'drop' | 踢 $t^h$i 'to kick' | 题 di 'subject' |
| 得 te 'to gain' | 铁 $t^h$e 'iron' | 头 de 'head' |
| 答 ta 'to answer' | 塔 $t^h$a 'pagoda' | 桃 dɘ 'peach' |
| 结 ke 'knot' | 客 $k^h$e 'guest' | 狂 gɔn 'mad' |

## II-2 Final models to calculate the results presented in Table 5.2

| Generation | Fixed structure | Random structure |
|---|---|---|
| Old | (ot1 + ot2) * Consonant + Place + Vowel + Repetition | (1 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1 + ot2 \| Item) |
| Young | (ot1 + ot2) * Consonant + Place + Vowel | (1 \| Speaker) + (ot1 + ot2 \| Speaker: Consonant) + (ot1 + ot2 \| Item) |

## II-3a Final models to calculate the results presented in Table 5.7

| Generation-Position | Fixed structure | Random structure |
|---|---|---|
| Old-P1 | Consonant + Gender + Repetition | (1 + Consonant \| Speaker) |
| Old-P2 | Consonant + Vowel + Gender | (1 + Consonant \| Speaker) + (1 \| Item) |
| Young-P1 | Consonant + Vowel | (1 + Consonant \| Speaker) |
| Young-P2 | Consonant | (1 + Consonant \| Speaker) |

## II-3b Between-Consonant comparisons of CQ by Generation and Position

A mixed model fitted to CQ with fixed factors Consonant, Generation, Position, and the full-fledged interactions was run. The random intercept for Speaker was also included. This was the maximal model which converged. Pairwise comparisons of the marginal means estimate were calculated with fractional degrees of freedom computed using the Satterthwaite method. The *p*-values were adjusted using the Tukey method for comparing a family of 3 estimates. n.s.: not significant.

| Contrast | Estimate | *SE* | *df* | *t*-ratio | *p* |
|---|---|---|---|---|---|
| Generation: Old, Position: 1 | | | | | |
| A – U | −0.07 | .01 | 3293 | −8.96 | < .001 |
| A – D | −0.05 | .01 | 3293 | −6.81 | < .001 |
| U – D | .02 | .01 | 3293 | 2.16 | n.s. |
| Generation: Young, Position: 1 | | | | | |
| A – U | −0.06 | .01 | 3293 | −36.46 | < .001 |
| A – D | −0.01 | .01 | 3293 | −1.43 | n.s. |
| U – D | .05 | .01 | 3293 | 5.03 | < .001 |
| Generation: Old, Position: 2 | | | | | |
| A – U | −0.03 | .01 | 3293 | −3.32 | < .01 |
| A – D | −0.02 | .01 | 3293 | −2.57 | < .05 |
| U – D | .01 | .01 | 3293 | .75 | n.s. |
| Generation: Young, Position: 2 | | | | | |
| A – U | −0.02 | .01 | 3293 | −2.25 | n.s. |
| A – D | .01 | .01 | 3293 | .75 | n.s. |
| U – D | .03 | .01 | 3293 | 2.99 | < .01 |
| Generation: Old, Position: 3 | | | | | |
| A – U | .01 | .01 | 3293 | .09 | n.s. |
| A – D | −0.01 | .01 | 3293 | −0.27 | n.s. |
| U – D | −0.01 | .01 | 3293 | −0.35 | n.s. |
| Generation: Young, Position: 3 | | | | | |
| A – U | −0.01 | .01 | 3293 | −0.91 | n.s. |
| A – D | .01 | .01 | 3293 | .19 | n.s. |
| U – D | .01 | .01 | 3293 | 1.11 | n.s. |

# English Summary

This dissertation investigates the interaction of consonant and lexical tone (C-T interaction hereafter) in two Chinese dialects, namely Lili Wu and Shuangfeng Xiang Chinese. It has long been known that initial consonant and fundamental frequency (*f0*) interact irrespective of whether *f0* distinction is phonologized. In tonal languages, a high tone (having a higher *f0* onset) usually co-occurs with a voiceless consonant, and a low tone (having a lower *f0* onset) usually co-occurs with a voiced consonant ([voiceless/H]-[voiced/L] pattern hereafter). However, this [voiceless/H]-[voiced/L] pattern has been challenged by data from some Chinese dialects.

A comprehensive phonetic description of the sound system of each of the two dialects is given before presenting the two case studies. In doing so, certain background information, terms, or phenomena, which require an elaborate motivation and explanation, can be introduced to readers as early as possible.

**Chapter 1** introduces the research topic of this dissertation, namely C-T interaction in two Chinese dialects (i.e., Lili Wu and Shuangfeng Xiang). Particularly, this study follows a property-based approach and aims to overcome the two biases (i.e., typological and methodological) observed in the existing literature of C-T interaction.

**Chapter 2** presents a comprehensive description of the sound system of Lili Wu Chinese. A number of methodological/analytical innovations and new perspectives with regard to not only lexical tones but also segmental features are proposed. First, there are eight lexical tones in Lili Wu Chinese. The instrumental analyses show a clear *f0* lowering effect in syllables with voiceless aspirated onsets in certain tonal contexts (i.e., non-*Ping* Middle Chinese [MC] tonal categories). This lowering effect, seemingly a split of the same tone into two as a function of voiceless unaspirated vs. aspirated onsets, is known as the 'aspiration-induced tonal split' (ATS) phenomenon. The [voiceless/H]-[voiced/L] co-occurrence pattern commonly observed in most Northern Wu dialects, falls apart in Lili Wu Chinese where voiceless aspirated onsets can co-occur with low tones. Second, voicing contrasts in fricatives, such as /f/ vs. /v/, can be signaled via their durational differences. The percentage of the frication duration of voiceless onsets is significantly higher than that of their voiced counterparts.

Third, the two high front vowels are proposed to be better transcribed as /i/ (e.g., /ti$^{3}$/ 'dot') and /i̟/ (e.g., /ti̟$^{3}$/ 'bottom'), with more anterior constriction for /i̟/. Fourth, there are two syllabic approximants in Lili Wu Chinese. /ɹ̹̩/, as in /sɹ̹̩$^{1}$/ 'book', is produced with a more laminal articulation combined with a lip rounding gesture in contrast to /ɹ̩/, as in /sɹ̩$^{1}$/ 'silk'.

**Chapter 3** focuses on the issue of C-T interaction in Lili Wu Chinese. Controlled experiments were designed to examine two long-standing debates on ATS in previous literature. They are i) Is ATS an on-going change (e.g., Shi, 1992) or a completed change (Shen, 1994)? and ii) Is aspiration (e.g., Chao, 1928) or breathiness (e.g., Zhu & Xu, 2009) synchronically related to ATS? The present results suggest that ATS in Lili Wu Chinese is a completed sound change but conditioned by certain tonal contexts (i.e., MC tonal categories). This pattern is quite consistent across generations. Regarding the second debate, the results suggest neither aspiration nor breathiness is synchronically related to ATS. One ongoing sound change observed is that the breathiness of vowels after voiced onsets is disappearing among the younger generation of Lili Wu speakers. This is probably due to its superfluous role in cueing the three-way laryngeal contrast and it is therefore not a robust cue for the laryngeal contrast in Lili Wu Chinese.

**Chapter 4** provides a comprehensive description of the sound system of Shuangfeng Xiang Chinese. There are four main findings. First, the voiced consonant has multiple laryngeal realizations: modal voiced, voiceless unaspirated, and implosive. Second, /n/ and /l/ contrast only before three high segments (i.e., /i ʏ j/) and are neutralized before the other segments. Third, Shuangfeng Xiang Chinese has an interesting and rarely observed three-way contrast in high back vowels, as exemplified by the triplet of /bo$^{2}$/ 'to climb' vs. /bu$^{2}$/ 'old woman' vs. /bu̹$^{2}$/ 'calamus'. In addition to their formant differences, the three back vowels can be distinguished via strong visual cues, namely their distinct lip gestures. Both /o/ and /u/ differ from /u̹/ in having more rounding and protruding lip constriction, whereas /u̹/ is produced with greater lip compression and less lip protuberance than /o/ and /u/ are. Fourth, relative to the sonorant baseline, *fo* is lowered after voiced and voiceless aspirated onsets but is unaffected (or minimally affected) after voiceless unaspirated onsets. This pattern seems to be consistent across all tonal contexts.

**Chapter 5** focuses on C-T interaction in Shuangfeng Xiang Chinese. The question concerns the phonetic properties shared by voiced and voiceless (i.e., unaspirated and aspirated) onsets that condition the low-rising *f0* contours. In the existing literature, *f0* contours after different onsets are treated as an identical low-rising tone (i.e., T2) which is associated with syllables with laryngeal contrast in voicing (e.g., Chao, 1935 [Yang, 1974]). This phenomenon is termed 'initial-associated tonal merger' (ITM). However, Zhu and Zou (2017) argue for a two-way phonatory distinction of the laryngeal contrast which co-occurs with two separate low-rising tones (i.e., modal – /24/ vs. breathy – /13/). The results show that neither voicing contrast nor phonation contrast can explain all findings. Furthermore, the phonetic properties that condition the low-rising *f0* contours have been undergoing changes. Specifically, there seems to be a trading relationship between voice onset time (VOT) and contact quotient (CQ) across generations. When a low-rising-contour-carrying syllable has a voiced onset, the old-generation speakers produce predominately negative VOTs without significant breathiness (indicated by CQ) in the following vowel. The young-generation Shuangfeng Xiang speakers, however, produce fewer negative-VOT tokens as well as shorter negative VOTs. However, in contrast to the old-generation speakers, they enhance breathiness over the following vowel (over the first half).

**Chapter 6** first concludes the main findings reported in previous chapters. Furthermore, the typological significance of the results obtained from Lili Wu Chinese and Shuangfeng Xiang Chinese is discussed. In general, voiceless aspirated onsets can co-occur with low tones. Breathiness transition, together with aperiodic noise of aspiration, provides a possible pivot for linking onset aspiration to *f0* lowering. This finding is antagonistic to the [voiceless/H]-[voiced/L] pattern which posits that only contrastively voiced onsets can be in favor of low tones. It suggests that neither the phonological [voice] category nor phonetic voicing should be inextricably or exclusively correlated with the low tone/tonal depression. Additionally, a tentative comparison between the ATS observed in Lili Wu Chinese and that reported in other languages (Kra-Dai languages and Gan Chinese) is made. After the discussion of the typological significance, the relationship between the Wu and Old Xiang dialect groups is revisited from the perspective of C-T interaction. It is reasonable to argue that both Wu (e.g., Lili) and Old Xiang (e.g., Shuangfeng) dialect groups have inherited the phonological trait of the three-way laryngeal contrast of obstruents from MC. However, in terms of the development of the three-way-contrast system, no shared innovation can be identified. Currently, there is no obvious evidence to lend support to the 'Unicity of Wu–Xiang' assumption (Hashimoto, 1978). At last, the different trading relationships between laryngeal timing (in terms of VOT) and phonatory state (in terms of CQ) to signal voicing contrast in the two Chinese dialects are discussed. The results of this study highlight different possible pathways for changes of cue weighting in the phonetic implementation of laryngeal contrasts in Asian tonal languages.

# Nederlandse Samenvatting

Dit proefschrift onderzoekt de interactie tussen consonanten en lexicale toon (hierna: C-T-interactie) in twee Chinese dialecten, namelijk het Lili-Wu- en het Shuangfeng-Xiang-Chinees. Het is reeds lang bekend dat de beginconsonant en de grondfrequentie (*f0*) interacteren ongeacht de kwestie of *f0*-onderscheid gefonologiseerd is. In toontalen komt een hoge toon (met een hogere *f0*-aanzet) meestal voor met een stemloze consonant, en komt een lage toon (met een lagere *f0*-aanzet) meestal voor samen met een stemhebbende consonant (hierna: [stemloos/H]-[stemhebbend/L]-patroon). Dit [stemloos/H]-[stemhebbend/L]-patroon is echter in twijfel getrokken door data uit bepaalde Chinese dialecten.

Het proefschrift presenteert een omvattende fonetische beschrijving van het klanksysteem van ieder van de twee dialecten, waarna de twee case-studies gepresenteerd worden. Zodoende kunnen specifieke achtergrondgegevens, termen, of verschijnselen/fenomenen, die een uitgebreide motivering en toelichting behoeven, zo vroeg mogelijk aan lezers geïntroduceerd worden.

**Hoofdstuk 1** leidt het onderzoeksonderwerp van het proefschrift in, te weten C-T-interactie in twee Chinese dialecten (het Lili-Wu en het Shuangfeng-Xiang). In het bijzonder volgt deze studie een op eigenschappen gebaseerde benadering met als doel de twee vooringenomenheden (d.w.z. typologische en methodologische) te vermijden die in de bestaande literatuur over C-T-interactie te vinden zijn.

**Hoofdstuk 2** presenteert een omvattende beschrijving van het klanksysteem van het Lili-Wu-Chinees. Een aantal methodologische/analytische innovaties en nieuwe perspectieven worden voorgesteld, die niet alleen lexicale toon betreffen, maar ook segmentele kenmerken. Ten eerste: er zijn acht lexicale tonen in het Lili-Wu-Chinees. De instrumentele analyses laten een duidelijk *f0*-verlagend effect zien in syllaben met stemloze geaspireerde onsets in bepaalde tooncontexten (d.w.z. Middelchinese [hierna: MC] tooncategorieën anders dan *Ping*). Dit verlagingseffect – ogenschijnlijk een splitsing van één toon in tweeën als functie van stemloze ongeaspireerde tegenover geaspireerde onsets, staat bekend als het fenomeen 'door aspiratie veroorzaakte toonsplitsing' (DAVT). Het [stemloos/H]-[stemhebbend/L]-patroon, dat vaak gezien wordt in de meeste Noord-Wu-dialecten, komt tot afbraak in het Lili-Wu-Chinees, waar stemloze geaspireerde onsets samen met lage tonen kunnen voorkomen. Ten tweede: stemhebbendheidscontrasten in fricatieven, zoals /f/ t.o.v. /v/, kunnen aangegeven worden door middel van verschillen in duur. De percentuele fricatieduratie van stemloze onsets is significant hoger dan die van hun stemhebbende tegenhangers. Ten derde wordt betoogd dat de twee hoge voorvocalen beter getranscribeerd kunnen worden als /i/ (bijv. /ti$^3$/ 'dot') en /i̟/ (bijv. /ti̟$^3$/ 'bottom'), met meer anterieure constrictie voor /i̟/. Ten vierde: in het Lili-Wu-Chinees zijn er twee syllabische approximanten. De /ɹ̩̹/, in bijv. /sɹ̩̹$^1$/ 'boek', wordt geproduceerd met een meer laminale articulatie gecombineerd met een liprondingsbeweging, in contrast met de /ɹ̩/, in bijv. /sɹ̩$^1$/ 'zijde'.

**Hoofdstuk 3** richt zich op het punt van C-T-interactie in het Lili-Wu-Chinees. Ik heb gecontroleerde experimenten opgezet om twee langlopende debatten in eerdere literatuur over DAVT te bestuderen. Dit zijn: i) Is DAVT een lopende verandering (bijv. Shi, 1992) of een voltooide verandering (Shen, 1994)? en ii) Is aspiratie (bijv. Chao, 1928) of *breathiness* (bijv. Zhu & Xu, 2009) synchroon verwant aan DAVT? De huidige resultaten suggereren dat DAVT in het Lili-Wu-Chinees een voltooide klankverandering is, maar een die geconditioneerd werd door bepaalde tooncontexten (d.w.z. Middelchinese tooncategorieën). Dit patroon is vrij consistent over generaties. Wat het tweede debat betreft, suggereren de resultaten dat noch aspiratie, noch *breathiness* synchroon verwant is aan DAVT. Een lopende klankverandering die geobserveerd wordt, is dat de *breathiness* van vocalen na stemhebbende onsets aan het verdwijnen is onder de jongere generatie Lili-Wu-sprekers. Dit is waarschijnlijk te wijten aan de overbodige rol ervan in het aangeven van het drievoudige laryngale contrast, en het is daarom geen robuuste *cue* voor het laryngale contrast in het Lili-Wu-Chinees.

**Hoofdstuk 4** biedt een omvattende beschrijving van het klanksysteem van het Shuangfeng-Xiang-Chinees. Er zijn vier hoofdbevindingen. Allereerst: de stemhebbende consonant heeft meerdere laryngale realisaties: modaal stemhebbend, stemloos ongeaspireerd, en implosief. Ten tweede: /n/ en /l/ contrasteren alleen voor drie hoge segmenten (te weten /i ʏ j/) en worden geneutraliseerd voor de andere segmenten. Ten derde: het Shuangfeng-Xiang-Chinees heeft een interessant en zeldzaam drievoudig contrast in hoge achtervocalen, zoals verbeeld wordt in het triplet /bo$^2$/ 'klimmen' tegenover /bu$^2$/ 'oude vrouw' tegenover /bu̹$^2$/ 'kalmoes'. Naast hun formantverschillen kunnen de drie achtervocalen onderscheiden worden via sterke visuele cues, namelijk hun kenmerkende lipgebaren. Zowel /o/ als /u/ verschillen van /ṳ/ in dat zij meer ronding en uitstekende lipconstrictie hebben, terwijl /ṳ/ geproduceerd wordt met grotere lipcompressie en minder lipuitstulping dan /o/ en /u/. Ten vierde: relatief aan het sonorantijkpunt wordt *f0* verlaagd na stemhebbende en stemloze geaspireerde onsets, maar blijft het onbeïnvloed (of minimaal beïnvloed) na stemloze geaspireerde onsets. Dit patroon lijkt consistent over alle tooncontexten.

**Hoofdstuk 5** richt zich op C-T-interactie in het Shuangfeng-Xiang-Chinees. De vraag behelst de fonetische eigenschappen die gemeenschappelijk zijn aan stemhebbende en stemloze (d.w.z. ongeaspireerde en geaspireerde) onsets die de laag-stijgende *f0*-contouren conditioneren. In de vigerende literatuur worden *f0*-contouren na verschillende onsets behandeld als één en de zelfde laag-stijgende toon (d.w.z. T2), die geassocieerd wordt met syllaben met laryngaal contrast in stemhebbendheid. Dit fenomeen staat bekend als 'met begin geassocieerde toonsamenval' (MBGT). Zhu en Zou (2017) betogen echter een tweevoudig fonatorisch onderscheid van het laryngale contrast dat samen voorkomt met twee verschillende laag-stijgende tonen (d.w.z. modaal – /24/ tegenover *breathy* – /13/. De resultaten laten zien dat noch stemhebbendheid, noch fonatiecontrast alle bevindingen kunnen verklaren. Bovendien hebben zich veranderingen voltrokken in de fonetische eigenschappen die de laag-stijgende *f0*-contouren conditioneren. In het bijzonder lijkt er een *trading relationship* te zijn tussen *voice onset time* (VOT) en het contactquotiënt (CQ) over generaties. Als een syllabe die een laag-stijgende contour draagt een stemhebbende onset heeft, produceert de oudere generatie sprekers voornamelijk negatieve VOT's zonder noemenswaardige *breathiness* (aangegeven door het CQ) in de volgende vocaal. De jongere generatie Shuangfeng-Xiang-sprekers produceert daarentegen kortere negatieve VOT's en minder tokens met negatieve VOT's. In tegenstelling tot de oudere generatie sprekers versterken zij echter *breathiness* over de volgende vocaal (over de eerste helft).

**Hoofdstuk 6** besluit eerst de hoofbevindingen opgedaan in de voorgaande hoofdstukken. Daarnaast wordt de typologische significantie van de resultaten uit het Lili-Wu-Chinees en het Shuangfeng-Xiang-Chinees besproken. Over het algemeen kunnen stemloze geaspireerde onsets samen voorkomen met lage tonen. *Breathiness*-overgang, tezamen met aperiodische ruis van aspiratie, biedt een mogelijke spil om onsetaspiratie in verband te brengen met $f_0$-verlaging. Deze bevinding is antagonistisch aan het [stemloos/H]-[stemhebbend/L]-patroon dat stelt dat alleen contrastief stemhebbende onsets voorkeur kunnen hebben voor lage tonen. Dit suggereert dat noch de fonologische categorie '[stemhebbend]', noch fonetische stemgeving onlosmakelijk of exclusief verbonden zou moeten zijn aan de lage toon/tonale depressie. Daarnaast wordt een voorlopige vergelijking gemaakt tussen de DAVT geobserveerd in het Lili-Wu-Chinees en die gerapporteerd in andere talen (Kra-Dai-talen en het Gan-Chinees). Na de discussie van de typologische significantie wordt de relatie tussen de Wu- en de Oudxiang-dialectgroepen herbezien vanuit het perspectief van C-T-interactie. Redelijkerwijs wordt betoogd dat zowel de Wu- (waaronder het Lili) als de Oudxiang-dialectgroepen (waaronder het Shuangfeng) de fonologische trek van het drievoudige laryngale obstruentencontrast uit het Middelchinees hebben geërfd. Wat betreft de ontwikkeling van het drievoudigcontrastsysteem kan er echter geen gedeelde innovatie geïdentificeerd worden. Op dit moment is er geen overduidelijk bewijs voor de 'Eenheid van Wu–Xiang'-assumptie (Hashimoto, 1978). Tenslotte worden de verschillende *trading relationships* besproken tussen laryngale timing (in termen van VOT) en fonatorische status (in termen van CQ) om stemhebbendheidscontrast aan te geven in de twee Chinese dialecten. De resultaten van deze studie brengen verschillende mogelijke ontwikkelingsroutes naar voren voor veranderingen van *cue*-weging in de fonetische implementatie van laryngale contrasten in Aziatische toontalen.

# 中文撮要

本论文旨在调查黎里吴语及双峰湘语中出现的辅音与声调的交互现象（C-T 交互）。词首辅音（汉语中通常体现为声母）和基频（$f_0$）的交互现象在世界语言中普遍存在。就声调语言而言，高调通常与清声母共现，低调则通常与浊声母共现。这种共现类型一般被称为“清高浊低”模式。但本文所报道的汉语方言材料并不支持该种普遍观察到的模式。

在具体讨论黎里吴语及双峰湘语这两个方言的 C-T 交互前，本论文分别对每个方言的语音系统进行了较为细致的描写，以便读者能够尽早了解该方言的特定研究背景、相关术语及具体现象。

**第一章**引入本论文的研究话题，即，黎里吴语及双峰湘语中的 C-T 交互现象。尤其需要指出的是，本研究采用了以语音性质为基础的研究方法（property-based approach），旨在弥补前人研究中存在的语言采样及方法论上的不足。

**第二章**为对黎里吴语语音系统的描写。本章通过一些较新的研究方法，分别讨论了黎里吴语中的音段及超音段（主要是声调）特征。本章主要有四大发现。首先，黎里吴语中有 8 个单字调。基频数据显示，在特定调类中（中古非平声调类），清送气声母后的基频明显下降。该种下降导致清送气声母后的声调与清不送气声母后的声调产生分化，即所谓的“送气分调”（aspiration-induced tonal split，ATS）现象。“送气分调”现象打破了大多数吴语中“清高浊低”的共现模式。黎里吴语的数据明确显示，清送气声母可以与低调共现。第二，清浊擦音可以通过时长特征来区别。清擦音摩擦时长的占比显著高于浊擦音。第三，两个前高元音应被描写为/i/（如，/ti$^{3}$/“点”）和/i̟/（如，/ti̟$^{3}$/“底”）更为适宜。相较于/i/，/i̟/拥有更为靠前的收紧点。第四，黎里吴语中有两个自成音节近音/ɹ̩/（如，/sɹ̩$^{1}$/“丝”）和/ɹ̹̩/（如，/sɹ̹̩$^{1}$/“书”）。相较于/ɹ̩/，/ɹ̹̩/带有明显的舌叶及圆唇特征。

**第三章**探讨黎里吴语中的 C-T 交互问题。针对前人文献中存在的两大议题，本章进行了控制实验。该两大议题分别是：第一，“送气分调”是一个正在进行中的音变（如，石锋，1992），抑或是一个已经完成的音变（Shen，1994）？第二，共时上与“送气分调”联系更为紧密的是送气特征（如，赵元任，1928）还是气声特征（如，朱晓农、徐越，2009）？本章的实验结果显示，“送气分调”是一个已经完成的音变，

它的出现受到中古调类的制约（只出现在非平声调类中）。这种模式稳定存在于代际间。就第二个议题而言，实验结果显示，送气和气声在共时上与“送气分调”都没有直接联系。年轻人浊声母后的气声正在逐渐消失，是一个正在进行中的音变。这或许是由气声在区别三分声母中的冗余地位造成的。

**第四章**为对双峰湘语语音系统的描写。本章主要有四大发现。首先，浊声母有多种语音实现，主要包括常态浊音、清不送气音及内爆音。第二，/n/和/l/仅在前高元音/介音（/i ʏ j/）前构成对立，在其他成分前合并。第三，双峰湘语拥有三个对立的后高元音，即，/o/（如，/bo²/“爬”）、/u/（如，/bu²/“婆”）和/ṳ/（如，/bṳ²/“蒲”）的对立。除了共振峰的区别以外，三者还能够通过唇形区分。/o/和/u/拥有凸唇特征，圆唇性更强；而/ṳ/主要体现为展唇特征。第四，若以响音声母后的基频为参照，浊声母与清送气声母后的基频明显降低，而清不送气声母后的基频几乎没有受到影响（或较小）。该情况出现在所有调类中。

**第五章**关注双峰湘语中的 C-T 交互问题。本章主要探讨清浊声母（清不送气、清送气、浊）与低升调共现的具体语音性质。前人文献多认为双峰湘语中只有一个低升调（第 2 调）。该调能同时搭配清浊声母（如，赵元任，1935［杨时逢，1974］）。这种现象被称为“异纽同调”（initial-associated tonal merger，ITM）。但是，朱晓农和邹晓玲（2017）认为，第 2 调并不是一个调而应该是两个调，分调的依据是清浊声母的发声态区别（常态–/24/，气声–/13/）。本章的实验结果显示，前人的结论都不全面，低升调与清浊声母共现的语音条件并不固定。这主要体现在代际间浊音起始时间（voice onset time，VOT）与接触商（contact quotient，CQ）的不同交易关系（trading relationship）。在低升调与浊声母共现时，老年人产生的实现为负 VOT 的字更多，时长也更长，元音气声（通过 CQ 体现）则较弱；但年轻人产生的实现为负 VOT 的字较少，时长也更短，但元音前半部分的气声反而增强。

**第六章**首先总结了前五章的主要发现，接着讨论了这些发现的类型学意义。总体而言，清送气声母能够和低调共现。气声过渡段外加送气带来的不规则噪音为清送气声母后出现低调创造了有利的环境。该发现与“清高浊低”模式中只有浊声母才能与低调共现的规则并不一致。这意味着，所谓的浊音音位或实际语音上的带声与低调/声调降低都没有必然联系。此外，本章还尝试将黎里吴语中的“送气分调”现

象与其他语言中（侗台语和赣语）所发现的类似现象进行比较。在讨论完类型学显著性后，本章试图从 C-T 交互的角度出发，重新检讨吴湘关系。吴语（黎里）和老湘语（双峰）的确都继承了中古汉语塞音/塞擦音三分的音系特征，但就该特征的后续发展来看，并无法找到共享创新（shared innovation）。因此，目前并没有明确的证据能支持所谓的“吴湘一体”假说（桥本万太郎，1978）。最后，本章讨论了两个方言中喉部时序（从 VOT 的角度）与发声态（从 CQ 的角度）在标识浊声母时展现的不同的交易关系。这些结果显示，亚洲声调语言在实现喉部发音对立（laryngeal contrast）时，音征加权（cue weighting）可能存在不同的变化路径。

# Curriculum Vitae

Menghui Shi (史濛辉) was born in Suzhou in the People's Republic of China, on 14th May, 1989. After graduating from Suzhou No. 10 Middle School in 2008, he started his Bachelor's program in Chinese linguistics at Fudan University, Shanghai. In 2012, he continued his linguistic studies in the Master's program at Fudan University. In 2013, he spent one semester studying at Taiwan University, Taipei. For his Master's thesis, he focused on socio-phonetic research and conducted several investigations into language contact among Suzhou Wu, Shanghai Wu, and Standard Chinese, under the supervision of Prof.dr. Huan Tao. After obtaining his Master's degree, in September 2015, Menghui was granted the PhD scholarship (No. 201506100048) under the State Scholarship Fund offered by the China Scholarship Council (CSC) and started his PhD research at Leiden University Centre for Linguistics (LUCL), under the supervision of Prof.dr. Yiya Chen and Prof.dr. Maarten Mous. This dissertation is the main product of Menghui's PhD research.