UNIVERSITY OF CALIFORNIA
Los Angeles

Phonation in Tonal Contrasts

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Jianjing Kuang

2013

ABSTRACT OF THE DISSERTATION

Phonation in Tonal Contrasts

by

Jianjing Kuang

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2013

Professor Patricia Keating, Chair

Phonation is used in many tonal languages, but how it should be incorporated into tonal systems is not well understood. The purpose of this dissertation thus is to examine the role of phonation in tonal contrasts, and to investigate how phonation and pitch interact in the tonal space. This dissertation presents close studies of tonal contrasts from three language families that provide different answers to these questions.

From the case of Yi languages, where the tense/lax phonation ("register") contrast is orthogonal to the tone contrast, we show that phonation production can be independent from tonal production. Phonation in these languages is not only phonologically contrastive, but phonetically completely independent. Tone is purely pitch, and phonation is purely voice quality.

From the case of Mandarin, where non-modal phonation (specifically, creaky voice) is known to be an allophonic cue to the low-dipping Tone 3, we show that the presence of creak in Mandarin

is purely driven by pitch range, and can occur with any of the low-pitch targets. Moreover, we show that in a corpus of pitch sweeps that rose or fell over large pitch ranges, Mandarin speakers' voice quality co-vary with pitch in a wedge-shaped function.

Finally, we investigate the case of Black Miao, a language with five-level-tone contrasts, the case of maximum use of pitch contrast attested among languages. Due to limitations on pitch production and perception, it should be very hard to make these contrasts purely by pitch. Both production and perception experiments show that non-modal phonations provide very important cues in both tonal production and perception, but in different ways. On the one hand, the three mid-range tones (T22, T33, T44) have very similar pitch cues, but T33 is quite distinct from T22 and T44 by its breathy voice quality. On the other hand, the extra-high (T55) and extra-low (T11) tones show the same kind of pitch-dependent voice quality that was found in Mandarin: the extra-low pitches tend to be creaky while the extra-high pitches tend to be tense.

Taking these cases together, we establish: 1) there are two different uses of phonation types, based on their relationship with pitch: pitch-dependent and pitch-independent, which can either enhance the perceptual distinctiveness of extreme pitch targets, or serve as a contrastive dimension for tones with similar pitch values; 2) tonal contrast is multidimensional, and the dimensionality of tonal contrasts in part depends on the size of the tonal inventory. We propose a new tone space model, and principles for good dispersion of tonal contrasts. This dissertation thus extends our knowledge about tonal contrasts, and provides a better understanding of the interaction between phonation and pitch in tone languages.

The dissertation of Jianjing Kuang is approved.

Bruce Hayes

Sun-Ah Jun

Jody Kreiman

Megha Sundara

Patricia Keating, Committee Chair

University of California, Los Angeles

2013

To my family and teachers

Table of Contents

List of Figures

List of Tables

Acknowledgements

This work could not have been done without the support of many people. I am especially grateful to my advisor, Pat Keating. She has been a great mentor for both my professional and personal life. None of my achievements thus far would have been possible without her generous support and guidance. I could never thank her enough for her tireless effort; she has commented thoroughly on every single version of every draft, and on every aspect, whether it be the bigger theoretical picture, the development of discussions, or even the smallest typos and errors. She is my role model in every area of life, and she has set a perfect example for me as a scholar, teacher, and mentor.

My deepest gratitude also goes to my wonderful committee: Bruce Hayes, Sun-Ah Jun, Jody Kreiman and Megha Sundara. I thank them for sharing with me their expertise and insight from different perspectives, for helping me clarify my thinking with all their inspiring questions, and for their insightful comments and constructive criticism at every stage of the dissertation. Meetings with my committee always had the magic power of bringing my work to a whole new level. I thank them for all their encouragement and support during the important moments of my career. I am also grateful to all my other teachers at UCLA, especially Kie Zuraw, Pam Munro and Robert Daland; they all have inspired and helped me in many ways.

Many thanks to all my dearest friends in the p-lab: Marc Garellek, Jamie White, Jason Bishop, Grace Kuo, Kristine Yu, Chad Vicenik, Roy Becker, Michael Lefkowitz, Yun Kim, Dustin Bowers, Adam Chong, Yu Tanaka and Nancy Ward. Other members graduated years before me, but they have also been great collaborators and friends: Christina Esposito, Sameer Khan, Jie Zhang and Kuniko Neilson. I especially thank the company of Marc Garellek, Jamie White, Jason Bishop and Grace Kuo, during the stressful season of dissertating and job search. We emotionally supported each other and celebrated each other's achievements. All of these moments are truly precious memories.

I also would like to thank my advisors at Beijing University: Professors Baoya Chen and Hongjun Wang. They introduced me to linguistics, and have supported me in every possible way since. It would not have been possible for me to come to UCLA and pursue my dream without their encouragement and help.

I owe special thanks to Professors Jiangping Kong and Feng Wang at Beijing University for supporting my fieldwork. I thank Feng Wang for taking me to my first field trip to the Jiangcheng Yi village in 2006, and for sharing with me his 2000-word corpus resulting from many years of hard work. I thank Professor Kong for generously sharing his Hani data with me, and lending me his EGG for my field trips in China. These field trips also could not happen without the help of Yan Lu, Defu Shi, Lindy Mark and Haichao Yang. I was so fortunate to have met many great consultants of Yi, Bo, Hani and Miao, who patiently shared their linguistic knowledge with me and kindly participated in the tiresome experiments.

I would like to thank Abeer Alwan and her students Yen-Liang Shue and Gang Chen for providing invaluable technical support on speech processing, especially on voice quality analysis and pitch detection. Thanks also go to Henry Tehrani for his support with the recordings and EGG signal processing. Without their help, the extensive voice analysis that was undertaken in this dissertation would not have been possible.

I am also grateful to my dedicated undergraduate research assistants: Spencer Lin, Kristen Toda, Ayu Hasegawa, and Jackie Pasche, for their time and effort in processing and analyzing the EGG and acoustic recordings.

I thank Hui Zhou, Ming Xue, Chen Chen, Min Li & Jie Yu, Yang Cao, Guoqiong Song and many other friends who have always been by my side these past five years, for sharing all my pains and joys, for comforting me when I felt down, and for pulling me out of my room for good food.

I thank my parents Yongxing Kuang and Yunfei Liu, my departed grandmother Yufei Fan, and my "brother" Weihan Zheng, for their boundless love and support over the years. I love you all.

Finally, the experiments and fieldwork in this dissertation were supported by NSFgrant BCS-0720304 to Patricia Keating, Abeer Alwan, and Jody Kreiman, summer research rewards from the UCLA Linguistics Department, and by a fieldwork fellowship from the International Institute. Chapter 2 is based on Kuang and Keating (2013); Chapter 4 has been submitted to Phonetica, and I am grateful for the thoughful comments from anonymous reviewers.

Vita

2008 -- 2011: M.A. Linguistics
University of California, Los Angeles

2005 – 2008: M.A. Linguistics
Chinese Language and Literature Department, Peking University
Beijing, China

2001 – 2005: B.A. Linguistics
Chinese Language and Literature Department, Peking University
Beijing, China

Chapter 1 Introduction

1. Non-modal phonation in tonal contrasts

The purpose of this study is to examine the contribution of phonation to tonal contrasts. Traditionally, phonation is not a part of the phonological representation of tones, as tone is defined as pitch contrasts (for overview, see Yip, 2002; Clements *et al.*, 2010; Hyman, 2010). Tonal contrast studies have mainly focused on pitch levels and movement (see Zsiga, 2012 for review), and fewer studies have addressed the role of phonation in tonal contrasts.

However, tone and phonation can co-occur in many languages. It has been reported that languages can contrast different phonations independently of tone. For example, Mpi (Silverman, 1997) combines two phonations with three level and three contour tones. Jalapa Mazatec combines three phonations with three level tones (plus contours) (Garellek and Keating, 2011). In some other languages, phonation can co-vary with certain tonal categories. For example, Mandarin Tone 3 is usually produced with a creaky voice (Belotel-Grenié and Grenié, (1994); Vietnamese has six tones, some of which are accompanied by creaky or breathy voice (Brunelle, 2009). The second kind of case is trickier, because the relative importance of phonation and pitch for these tonal contrasts is not clear.

To settle this question, perception studies have been done in various languages. For example, it has been shown that the presence of creaky voice is optional for Mandarin speakers, and pitch

cues are sufficient for tonal identification, but this non-modal phonation can facilitate the reaction time in identifying tone 3 (Belotel-Grenié and Grenié, 1994; 1997; 2004; Yang, 2011) . A similar effect is also found for Cantonese, where the presence of creak can bias the identification towards the lowest tone (Yu and Lam, 2011). Therefore, for Mandarin and Cantonese, phonation is only an allophonic cue to tonal contrasts. However, sometimes a phonation cue is more important than pitch for a tonal contrast. For example, largnealization is the primary cue for identifying one of the rising tones in Vietnamese (Brunelle, 2009); native listeners attend only to the phonation cues when they are asked to distinguish the three falling tones in Green Mong (Andruski and Ratliff, 2000); phonation and duration are the contrastive cues for three out of the six tones in Sgaw Karen (Brunelle and Finkeldey, 2011) . Therefore, for these languages, phonation is a phonemic cue in tonal contrasts. The matter becomes even more complicated when phonation is more important than pitch for some tones, but less important than pitch for other tones. For example, breathy phonation is the primary cue in distinguishing between two high-falling tones for White Hmong listeners, but creaky phonation plays little role in identifying the low falling tone (Garellek *et al.*, 2013).

The cases listed above have raised important questions for tonal studies:

1. What is tone? Is phonation a part of tonal contrasts?

2. If phonation should be included in phonetic tonal spaces, how does it interact with pitch?

3. Why is it sometimes allophonic, but sometimes contrastive? When and why do languages have either or both of these non-modal phonations?

The cases reviewed above have suggested a yes to the first question. In fact, some theoretical studies have attempted to incorporate phonation as "tonal registers" (to be reviewed in 2.1 of this chapter), but because the rest of the questions remain unanswered, the attempts have not been fruitful. This dissertation will answer these questions from two sides: on the mechanism side, we will gather a better understanding of the interaction between pitch and phonation; on the structural side, we will look into the role of phonation in the contrastiveness of tonal categories. Specifically, we will provide close studies of three cases from different language families to investigate the different roles of phonation in tonal contrasts. The rest of the introduction will provide background for related issues for this topic.

2. Dimensionality of tonal contrasts

2.1. The phonological proposals

Although phonological theories have proposed different representations of tones, most of them have defined tone with one single phonetic dimension -- pitch. With this assumption, tonal contrasts are essentially a "pitch scalar system" (Hyman, 2010; c.f. Chao's number). Therefore, the question of how many contrasting pitch levels are needed in a tonal feature system becomes a central issue in tone theory. Chao (1948) proposed that languages have at most five level tones, and transcribed them as 1, 2, 3, 4 and 5 (low to high). This system has been adopted in the IPA. However, typological studies (e.g. Maddieson, 1978) found that most languages only have two or three levels. There has not been a physiological explanation for this typological limit, but phonologists have tried to capture this limit in their feature systems.

To find the right number of features has been challenging. A single binary feature such as [+/-high] (and underspecification) is not sufficient to describe tonal systems with more than three levels. A system without underspecification but using both [+/-high] and [+/- low] can also deal with a three-level system, but not four, since [+high, +low] is logically excluded. However, three-binary-feature systems, such as Wang (1967) and Woo (1969), have the problem of over-generating levels, because three binary features can in principle be combined in eight ways. So redundancy rules have to be induced to trim the number of levels. Moreover, too many features complicate description for the languages with fewer levels. As Clements (1983: p. 146-147) addressed, "a language with two tone levels can be described in ten different ways" using three binary features.

Autosegmental phonology enriched the understanding of tonal structures and inspired the multi-tier tonal representations. In this framework, a "register model" (Yip, 1980) or a "two-feature model" (Clements, 1983) were proposed. The common core among many proposals (Hyman, 1986; Pulleyblank, 1986; Duanmu, 1990; Snider, 1990; Bao, 1999) is shown in (1), with the focus on the representations for multiple levels. In general, these models posit a tonal register feature, e.g. [+/-upper] (according to Yip), which divides the entire pitch range into two primary sub-ranges. Within each register, the pitch range is further sub-divided into two levels, e.g. [+/-high]. Taken together this hierarchical structure predicts a tonal system with up to four contrastive levels.

(1)	+Upper	+high	=	extra high	55
		- high	=	high	44
	- Upper	+high	=	mid	33
		- high	=	low	11

Four levels are sufficient for most tonal languages. This very well solves the paradox between sufficiency and simplicity of features. Moreover, the [+/-upper] registers are functionally similar (but not directly correspondent) to the *yin* and *yang* registers (Haudricourt, 1972; Bao, 1999), which were historically conditioned by onset voicing contrasts. Therefore, this model insightfully reveals the connection between tones and laryngeal properties.

The typological challenge for this model is tonal systems with a fifth level tone. Though very rare, several Asian languages and Central American languages with five-level tones have been reported (a complete list see Edmondson and Gregerson, 1992). The two-feature system suffers from the problems of ambiguity and insufficiency: It is very ambiguous how to characterize five levels into two registers, and two by two features would not be sufficient for specifying five levels.

A possible solution is to allow a mid level in each register, such as [+upper, -high, -low] (Edmondson and Gregerson, 1992). But this predicts six-level-tone systems, which are not typologically attested.The two most central mid tones [+upper, -high, +low] and [-upper, +high, -low], have to be defined as phonetically identical so as to collapse the number of levels into five.

(2) +Upper +high, -low = 55
- high, -low = 44
- high, +low = 33

- Upper +high, -low = 33
- high, -low = 22
- high, +low = 11

A prediction of this model is that languages differ in how many levels belong to each register, since 33 could go either way. Even so, the problem of assigning tonal features does not completely go away. In the Yip (1980) and Bao (1999) models, registers are assumed to be correspondent to underlying consonant voicing contrasts; so correspondent tonal pairs should form a natural class, such as [+high], when distinguished by registers. However, for the five-level-tone languages, even referring to the historical source of voicing contrasts, not all contrastive levels can be clearly grouped into natural classes. The mid tones are especially ambiguous. For example, 33 and 44 can be either [+upper] or [-upper], varying across languages (Edmondson and Gregerson, 1992).

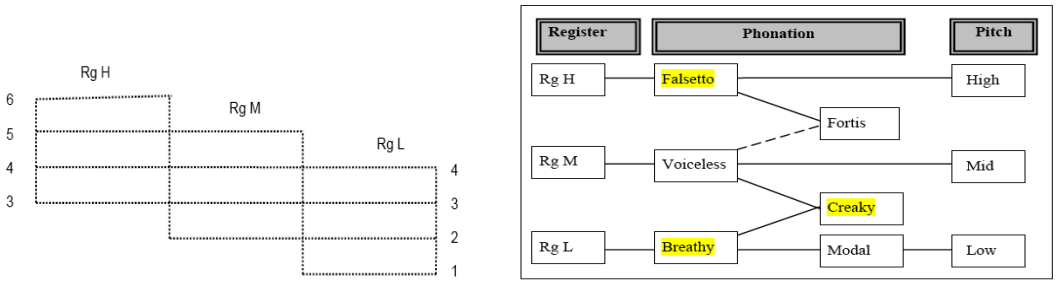
An alternative solution is to propose a third register. It is possible for five levels to be distinctive in three registers. Duanmu (1990) proposed a model with two laryngeal features ([+/-stiff] and [+/-slack]) and two pitch features ([+/-above] and [+/- below]), as shown in (3).

(3)

+stiff	[+above, -below]
- slack	[-above, -below]
	[-above, +below]
-stiff	[+above,-below]
-slack	[-above, -below]
	[-above, +below]
-stiff	[+above, -below]
+slack	[-above, -below]
	[-above, +below]

The laryngeal features, i.e. [+/-stiff] and [+/- slack], which were originally used to specify pitch levels (Halle and Stevens, 1971), were adopted in this model to indicate registers. This model appears to have too many levels (up to nine), but Duanmu suggests that these tonal registers are not purely tonal, but might be accompanied by phonation. For example, [+slack] register could be associated with breathy phonation, and [+stiff] is related to "clear" voice. The model insightfully recognizes that tonal split is not necessarily two-way, and registers could be conditioned by phonation. Since this model has no limitations on how pitch will interact with phonation, there appears to be no universal prediction of how tonal features should be assigned to the pitch levels. Because of its redundancy, Duanmu (2002) dropped this model and came back to a two-register-two-pitch system. Most recently, Zhu (2012, most recent) provides a clearer proposal of how three registers could interact with pitch levels. Zhu's model has three registers: H, M and L; and each register has four pitch levels. More importantly, these three registers are conditioned by six discrete phonation types.

(4) Zhu's multi-register and four-level model (reproduced from Zhu 2012)



As shown in (4), this model predicts up to six pitch levels, since overlapping pitch values across registers are allowed. As in Duanmu (1990), multiple levels could contrast in various ways, either multiple pitch levels within one register, or fewer contrastive levels with multiple registers. This model explicitly suggests that phonation is the phonetic correlate of tonal registers, and also insightfully suggests that certain phonation types are associated with certain pitch ranges. However, this model also risks overgenerating, as it allows four contrasting levels in each register; and the physiological mechanisms of including these seven phonation types are not clear.

Overall, models (3) and (4) make a significant improvement in tonal representation. The original pitch-based tonal registers (Yip, 1980) are not synchronic natural classes in Cantonese, as there is no phonological process based on the register features (Clements *et al.*, 2010). Moreover, such registers also do not provide perceivable cues for native speakers (Mok and Wong, 2010). So there is no mental reality of this kind of tonal registers. The phonation-based registers in (3) and (4) are certainly more likely to be synchronic natural classes, as tone sandhi is usually constrained by this kind of register (Bao, 1999). However, despite the idea that phonation is

related to tonal registers, it has been unclear 1) what kind of phonation types could be involved (different proposals involve different numbers and types of phonation), 2) how phonation and pitch interact with each other, and 3) how many contrastive pitch levels can occur in each register (three in Duanmu, 1990; four in Zhu, 2012). Moreover, allophonic non-modal phonation as in Mandarin is usually not taken into account. With extensive experimental studies on different languages, this dissertation will provide some insight on these questions.

2.2. Perceptual cues in tonal contrasts

As already mentioned, perception experiments have been employed to decide which cues are most important to certain tonal contrasts. Most tonal perception studies have focused on pitch cues, including mean F0, direction of change (e.g. rising or falling) and slope of change (or F0 difference within a syllable). For example, for Cantonese listeners, both the direction of F0 change and the slope of change are important dimensions in tonal perception (Khouw and Ciocca, 2007). Native listeners are sensitive to the subtle movements of pitch within a syllable. For example, Zsiga and Nitisaroj (2007) showed that Thai listeners use peak alignment as the main perceptual cue in tonal distinctions. Listeners with different language experience have different strategies in grouping pitch contours in a perceptual space. For example, Gandour and Harshman (1978) played different pitch contours to speakers of Yoruba (level tone only), English (no tone), and Thai (both level and contour tones), and Multidimensional-Scaling analysis identified five perceptual dimensions for tone: mean F0, contour shape (level, rising and falling), duration, endpoints, and slope (level and contour). They found that listeners with different language experience use the same five dimensions, but weight their importance differently: English

speakers mostly only attend to mean F0, but speakers of tonal languages also attend to slope of F0; Thai speakers weight contour shape more than Yoruba speakers. The authors thus concluded that the direction and slope of pitch change should be included in the universal feature set. This claim directly opposes Autosegmental Theory (Goldsmith, 1979), which proposed that contour tones are decomposed into multiple levels.

Most recently, phonation cues have also been taken into consideration. As mentioned at the beginning, studies have shown that phonation plays an important role in the perception of certain tonal contrasts. For example, allophonic creak can facilitate the tonal identification of the lowest tones in Mandarin (Yang, 2011, most recently) and Cantonese (Yu and Lam, 2011); phonation is the primary cue for identifying certain tonal categories in Vietnamese (Brunelle, 2009), Green Mong (Andruski and Ratliff, 2000), Sgaw Karen (Brunelle and Finkeldey, 2011) and White Hmong (Garellek *et al.*, 2013).

When all the dimensions that can contribute to tonal contrasts are considered, it becomes non-trivial to model the tonal space in which dispersion can be understood. Previous studies of tonal dispersion have instead used very simple spaces. In comparing the production effort for tones in normal vs. noisy environments, Zhao and Jurafsky (2007; 2009) modeled tonal dispersion as variability of the overall pitch range. Adopting this method, Alexander (2010) also only used a one-dimensional cue, i.e. mean F0, to define tonal spaces. As she noted, this method was not adequate to capture the perceptual separability of tonal contrasts. Tonal space models that allow contour cues significantly improve the separability of the tonal space. For example, in a study

comparing Cantonese tone productions by normal-hearing adults, normal-hearing children and cochlear-implanted children, Barry and Blamey (2004) defined the tonal space by F0 onset x F0 offset. This method enables the authors to capture the dynamic factors, such as direction and slope, of the tones. Yu (2011) demonstrated that mean F0 + pitch change, rather than mean F0 alone, can better model the distribution of tonal inventories. Taken together, these studies suggest that a tonal space should incorporate multi-dimensional cues to reflect the actual perceptibility of tonal contrasts, contours as well as levels. However, no tonal space models so far have incorporated cues like duration and phonation. There is no proposal yet about how to incorporate allophonic and phonemic phonation in the tonal space at the same time, when they both exist in a language. This dissertation will consider all possible dimensions in tonal contrasts, at least including pitch, duration and phonation, and examine how they contribute to the dispersion of a tonal space.

3. Linguistically meaningful non-modal phonations

3.1 Phonation types

Due to the continuous and large variability of phonation (Gerratt and Kreiman, 2001; Redi and Shattuck-Hufnagel, 2001), definitions of phonation types vary by different research purposes and disciplines (see Gerratt and Kreiman, 2001 for overview). Non-modal phonation is relative to "modal" phonation, which refers to the most usual and default voice quality in speech, typically involving full closure and full opening of the glottis in each cycle of vibration. This section gives a brief review of linguistically meaningful non-modal phonations.

"Creaky" and "breathy" phonations have been the most widely used terms in linguistics, and many languages use such a contrast to distinguish word or utterance meanings. A typical breathy voice is defined as having the vocal folds fairly abducted (relative to modal and creaky voice) and with little longitudinal tension (Laver, 1980; Gordon and Ladefoged, 2001; Gobl and Ní Chasaide, 2012). By contrast, a typical creaky voice is defined as having the vocal folds tightly adducted and vibrating irregularly (Gordon and Ladefoged, 2001). However, these terms are very ambiguous; as many speech studies (Gerratt and Kreiman, 2001; Gordon and Ladefoged, 2001; Keating *et al.*, 2012) have pointed out, creakiness/breathiness vary continuously along a continuum of glottal closure. It has been well-established that Open Quotient (OQ, or its reflex in the EGG signal, Contact Quotient (CQ)) is the main difference between creakier sounds versus breathier sounds. Acoustically, this property is usually reflected in the difference between the first two harmonics (H1-H2) of the spectrum. Creakier sounds have much smaller OQ (greater CQ) and smaller H1-H2 values than breathier sounds.

Moreover, creaky voice is not only ambiguous in its relations to modal and breathy voice, but also ambiguous in its actual physiological mechanisms. Closer investigations of phonation variability (Gerratt and Kreiman, 2001; Redi and Shattuck-Hufnagel, 2001) have revealed that linguistically creaky voice could be classified into more refined subcategories. We will not get into the details of different criteria. Here we will only highlight two different types of largyngealization based on their affiliation with pitch: tense voice and vocal fry.

Stiff/tense voice is when the vocal folds have a high longitudinal tension (Gobl and Ní Chasaide, 2012), which naturally happens when pitch is close to the highest (without changing into falsetto). In the singing literature, this kind of phonation is also called pressed voice (Sundberg, 1987). One can also produce it with a non-high pitch when lifting heavy objects. When pitch is even higher, people tend to turn to falsetto phonation. When producing falsetto, there is a long narrow open leakage between the vocal folds (Kong, 2007), and this causes a high value of OQ. The sinusoid-like pulse shape is another basic feature of falsetto. In the singing context, falsetto is the phonation of the highest pitch range (Sundberg, 1987; Titze, 1988).

In contrast to tense voice, classic vocal fry is usually associated with a very low pitch and it is related to adductive tension and compression of the vocal folds (Gerratt and Kreiman, 2001; Kong, 2007). During vocal fry, strong damping occurs in the glottal pulses, which often leads to the failure of pitch tracking in acoustics, but it is very salient in perception (Gerratt and Kreiman, 2001).

All in all, despite the great variability of phonation, the relative glottal opening has been regarded as the primary property of phonation contrasts. There is physiological evidence that the control of glottal opening is relatively independent from the control of pitch (Laver, 1980; Gobl and Ní Chasaide, 2012). Therefore, we expect that phonation can be a contrastive dimension in tonal contrasts, totally independent from pitch contrasts. Languages with phonation and pitch crossed are good test cases to validate this hypothesis.

3.2 Phonation can vary along the pitch scale

Linguistic studies on phonation contrasts have mainly focused on the property of the relative glottal opening, assuming that phonation production is relatively independent from pitch production. However, phonation can also vary along the pitch scale. This kind of phonation continuum has not been explicitly addressed in any linguistic literature, but it has been a major issue in studies of singing.

It has been proposed that pitch range is divided into three "registers" (Hollien and Michel, 1968; Hollien, 1974; Titze, 1988; Roubeau *et al.*, 2009), basically three pitch sub-ranges, and each register is related to a certain type of phonation. The default is called modal register, defined as the pitch range that is used for normal speech and singing. Hollien and Michel (1968) found that the modal register for female singers is about 144 – 495 Hz, and for male singers is about 78 – 275 Hz. Titze (1994: p. 263) found that both female and male singers tend to have a major involuntary transition at around 300 Hz. The lowest pitch of falsetto register for male singers is 200 Hz (Sundberg, 1994: p. 51). However, the comfortable pitch range found for speech is much smaller (Baken and Orlikoff, 2000: p. 174, summary from various studies): about 90 (+/-10) Hz – 165 (+/-10) Hz (median = 142 Hz) for male English speakers, and 160 (+/-10) Hz – 250 (+/-10) Hz (median =201 Hz) for female English speakers (averaging over spontaneous speech and read speech).

When pitch goes higher than these limits, singers will switch into the so-called loft register, where falsetto is used, in order to reduce the vocal fold tension (Titze, 1994). On the other hand,

when pitch goes to the low end, it turns into the so-called pulse register, where vocal fry is used. Perception experiments (Hollien and Wendahl, 1968; Keidar *et al.*, 1987) showed that any pitch below 70 Hz were treated as vocal fry by trained musicians; and both falsetto and vocal fry were highly correlated with F0 in perception. These results suggested that falsetto and vocal fry are the high and low ends of the frequency scale, and that they are driven by pitch production.

In this dissertation, we will look into this kind of phonation variation more closely. The hypothesis is that, if allophonic non-modal phonation, such as creaky voice in Mandarin, is driven by extreme pitch targets, we would expect that it can be turned off when pitch range is changed. We also hypothesize that this kind of non-modal phonation plays a different role from the non-modal phonation that is independent from pitch range.

4. Complex phonation effects on F0 and the development of non-modal phonation

Glottal states affect F0. It has been well-documented that consonants and non-modal phonation on vowels can affect F0, both synchronically and diachronically. Breathy phonation appears to be associated with pitch lowering in the majority of languages (Hombert *et al.*, 1979; Gordon and Ladefoged, 2001; Kong, 2001). The cases for so-called creaky phonation are mixed, however, and may be due to the variations of laryngealization. According to the literature, creaky voice is associated with pitch lowering in many languages, e.g., Mam and Northern Iroquoian languages (see Gordon and Ladefoged, 2001 for overview), San Lucas Quiaviní Zapotec (Chávez Peón, 2010), Santa Ana del Valle Zapotec (Esposito, 2010), Yucatec Maya (Frazier, 2009), Coatzospan Mixtec (Gerfen and Baker, 2005). However, in many Tibeto-Burman

languages such as Northern Yi and Bai, creaky voice is usually associated with a pitch raising effect. According to Kong (2001), creaky voice in these languages is usually some form of tense voice.

The issue of phonation effects on pitch has received special attention because it has been proposed that phonation is responsible for the genesis of contrastive tones. A substantial amount of language evidence supports the hypothesis that tones emerged from the laryngeal articulations of onset or coda consonants (Haudricourt, 1954; Hombert *et al.*, 1979; Kingston, 2005). The original stage is the F0 perturbation caused by the glottal settings of the consonants (Halle and Stevens, 1971; Hombert *et al.*, 1979; Thurgood, 2002; Tang, 2008). However, it is not clear how these glottal features spread from consonants to entire vowels and lead to contrastive pitches, as glottal features of consonants are not sufficient to account for tonal changes (Tang, 2008; Blankenship, 2002). Thongkum (1990) and Thurgood (2002) proposed that consonants do not lead to tonogenesis directly, but through coarticulated non-modal phonation types. There must be a stage of non-modal phonation governing whole vowels. DiCanio (2012) also observed that the presence of non-modal phonation could enhance the F0 perturbation effects. This might suggest a beginning of tonogenesis, that is, F0 perturbation is facilitated by the adoption of a coarticulated non-modal phonation. These non-modal phonations are then responsible for tonal splitting. This "missing link" is found in the dialects of these languages that nowadays still keep a phonation contrast in the tonal system, e.g. Tamang (Mazaudon and Michaud, 2006; 2009). Phonation register contrasts that are commonly found in Mon-Khmer and Tibeto-Burman languages could in principle induce tonal splitting, e.g. Suai (Abramson *et al.*, 2004) and Khmu'

(Abramson *et al.*, 2007). The current hypothesis in the field, then, is that creaky phonation and breathy phonation could be the intermediate stage of tonogenesis and could facilitate pitch contrast (Thurgood, 2002; Abramson and Luangthongkum, 2009).

5. Interaction between phonation and pitch: constraints from production and perception

Phonological structures tend to prefer some contrasts to others due to human limitations on speech perception and production. Tone production has been found to be limited in many ways; for example, there is a robust negative correlation between duration and pitch height (Faytak and Yu, 2011): the lower the tone, the longer the duration. And yet, F0 change in falling tones is much faster than in rising tones (Hombert, 1977). F0 change speed also has a maximum limit that will cause undershoot tonal targets in fast speech (Xu and Sun, 2002).

Although few studies have explicitly discussed interactions between pitch and coarticulated phonations, this is also constrained. It appears that phonation contrasts are constrained by pitch range, though it is not clear how. For example, the Southern Yi phonation contrast does not occur with high tone (Kuang, 2011); in Jalapa Mazatec (Garellek and Keating, 2011), the phonation contrast in high tone is less frequent; in Santa Ana del Valle Zapotec (Esposito, 2010), when the falling tones are spoken at a higher pitch, as in focus position, the phonation contrast between breathy and creaky may be neutralized. Relatedly, the correlation between phonation and F0 is also limited to certain ranges of F0. For example, Iseli *et al.* (2007) found that H1-H2 and F0 had a positive correlation when F0 is below 175 Hz. We notice that this number is close to the upper limit of males' modal pitch register for speech (Baken and Orlikoff, 2000). In

addition, a categorical perception experiment on the phonation contrast in Southern Yi (Kuang, 2011) suggested that native listeners performed slightly but significantly better in the mid tone than in the low tone. Taken together, these findings suggest that phonation contrasts are best in the middle of the pitch range.

On the other side, pitch contrasts are also constrained by phonation types. For example, in languages with multiple tones, such as Hmong and Zapotec languages, most F0 contrasts occur with modal voice; fewer tones occur with non-modal phonations. On the perception side, Silverman (2003) showed that pitch discrimination by English speakers is less good during breathy phonation than during modal phonation. He speculated that the Just Noticeable Difference (JND) increase might be due to the reduction of the harmonic-to-noise ratio during breathy phonation relative to modal phonation.

In sum, there are natural constraints on pitch ranges which are good for phonation contrasts, and phonation types which are good for pitch contrasts. That is why phonation contrasts and tonal contrasts happen in certain ways. A good tone model should incorporate these constraints.

6. Summary of research goals

This dissertation will examine the role of phonation in tonal contrasts, and develop a tonal model that integrates the interaction between phonation and pitch. In this introduction, we identified two different relationships between pitch and phonation: phonation can be independent from

pitch or can co-vary with pitch. Therefore, there are two possible types of non-modal phonation: pitch-independent and pitch-dependent. In the following chapters, we will examine whether they have the same functions in tonal contrasts. To preview, the body of this dissertation will cover three language cases: tense vs. lax phonation contrasts in Yi languages, as a case of tonal contrasts plus phonemic non-modal phonation; creaky voice in Mandarin, as an example of tonal contrast with allophonic non-modal phonation; finally, Black Miao's five-level-tone system, as a case where phonation helps with crowded tonal contrasts. In these language-case studies, general mechanisms of the interaction between the two will be explored as well.

Chapter 2 Yi languages: A case of phonation independent from tone

1. Introduction

Phonation contrasts and tonal (pitch) contrasts can be independent and crossed, resulting in more contrastive syllables in a language. Two-way phonation contrasts are often found in Tibeto-Burman languages. For example, Mpi has six tones and all of them can occur both modally and laryngealized, resulting in a total of twelve contrasts (Silverman, 1997). In a rare case, Jalapa Mazatec, an Otomanguean language, has a three-way phonation contrast (breathy, modal, laryngealized) and three tonal contrasts (high, mid, low), fully crossed (Silverman, 1997). Nonetheless, as acoustic phonation-related measures all have interactions between phonation and tone for Mazatec (Garellek and Keating, 2011), it is still not clear whether phonation contrasts and tonal contrasts can be completely independent in articulation.

The type of phonation contrast that will be the focus of this chapter is the two-way tense vs. lax phonation contrast found in some tonal Tibeto-Burman languages. For example, in Southern Yi (a Tibeto-Burman language spoken in China), the syllable /be^{21}/ with a lax phonation means "mountain", whereas /b$\underline{\text{e}}^{21}$/ with a tense phonation means "foot". (Here, tense phonation is indicated by an underscore, following Ma (2003), and the superscript numbers indicate the tone). Both kinds of phonation occur with mid and low tones[1]. For example, Southern Yi has not only /b$\underline{\text{e}}^{21}$/ ("foot") vs. /be^{21}/ ("mountain"), but also /b$\underline{\text{e}}^{33}$/ ("shoot") vs. /be^{33}/ ("fight"). This chapter

[1] Most such languages, including Southern Yi, also have a high 55 tone, but it does not have phonation

investigates the production of the tense vs. lax phonation contrasts in three of these languages: Southern Yi, Bo and Luchun Hani. Extensive acoustic and electroglottographic (EGG) analyses will be presented to show that the two kinds of contrasts are indeed independent and engage separate production mechanisms.

2. Background

2.1 Languages

Southern Yi, Bo, and Hani are Yi (also called Loloish) languages in the Tibeto-Burman family of the Sino-Tibetan phylum. The name "Yi" refers to both the whole Yi (Loloish) branch of languages and the Yi language, because it has the most population in this language family branch. Sometimes Yi, Burmese and Zaiwa are collectively called Burmese-Lolo. The Yi language branch includes perhaps fifty languages, for example Yi, Hani, Lisu, Lahu, Naxi and Bo, among many others. Yi languages are geographically distributed in Yunnan, Sichuan and Guizhou provinces of China, and are spoken by more than six million people (Ethonologue, 2012).

The inventories of Yi languages share the following common typological properties (Ma, 2003). First, voicing is the most important distinctive feature for consonant inventories. All the obstruents and laterals have a voicing contrast; Northern Yi even contrasts voicing in nasals. Second, syllable structure in Yi languages is very simple: no onset clusters and no codas, and thus syllables are typically CV. Third, all Yi languages are tonal languages, typically with 3

tones, namely, High, Mid, and Low (noted as 55, 33, and 21). Tones do not contrast by contours. Fourth, vowel inventories mainly consist of monophthongs, with diphthongs very rare. Fifth, vowels contrast by registers: tense vs. lax contrasts are the hallmark feature of Yi languages. (Some languages in this family, such as Nu, even have a third "register", nasalization, and therefore have four-way register contrasts in vowels.) Table 2-1 gives the tones and registers in Southern Yi, Bo and Hani.

Table 2-1 Tones and Registers in Yi languages: Tense vs. lax contrast in the mid and low tones

	Low	Mid	High
Lax	21L	33L	55L
Tense	21T	33T	

2.2 Production of phonation types

Linguistic voice quality has been defined along a continuum of the (average) aperture between the arytenoid cartilages (Ladefoged, 1971; Gordon and Ladefoged, 2001) . Four types of non-modal phonations have been widely referred to: breathy, creaky, lax and tense. Breathy phonation is said to involve minimal adductive tension, weak medial compression and low longitudinal tension (Laver, 1980). There is an increase in the aperture between the vocal folds, such that the posterior portion to the midline of one vocal fold never comes in contact with the other fold (Laver, 1991; Gobl and Ní Chasaide, 2012). The constant glottal leakage often leads to audible frication noise. Lax phonation has a similar articulatory configuration as breathy phonation, but is less extreme. Breathy phonation has a greater vocal fold aperture and higher

amplitude of noise components (Pennington, 2005). A "slack" or "lax" configuration of the vocal folds often leads to lower pitched and softer sounds (Gobl and Ní Chasaide, 2012).

Both creaky and tense phonations are characterized by increased adductive tension and medial compression (Gobl and Ní Chasaide, 2012), which leads to decreased aperture between the vocal folds. Because of the high adductive tension, only the ligamental part of the vocal folds is vibrating. Tense phonation also involves a higher degree of tension in the entire vocal tract as compared to the neutral setting. The increased muscular tension associated with tense phonation is likely to affect the respiratory system as well as the supralaryngeal tract (Gobl and Ní Chasaide, 2012). The major difference between tense phonation and creaky phonation is whether the vocal folds have periodic vibration (Childers *et al.*, 1990); tense phonation is thus closer to modal phonation.

2.3 Previous studies of Tibeto-Burman phonation types

A few previous studies have examined a small set of acoustic correlates associated with tense vs. lax contrasts in Tibeto-Burman languages. Maddieson and Ladefoged (1985) and Maddieson and Hess (1986) investigated tense vs. lax contrasts in four Tibeto-Burman languages (Hani, Eastern Yi, Jingpo, and Wa) and found that despite the cross-language variation in other acoustic correlates (e.g. F0, F1, duration, and VOT), the amplitude difference between the second harmonic and the fundamental frequency (H1-H2) is consistently higher for the lax phonation across languages. Moreover, the lax phonation was also found to have a higher rate of airflow and pressure. Kong (2001) found that H1-H2 is a successful measure to distinguish tense vs. lax

phonation contrast in several Tibeto-Burman languages (e.g. Northern Yi, Zaiwa, Jingpo), though H1-A1 and H1-A2 are better measures for Northern Yi. A laryngoscopic study (Esling *et al.*, 2000) showed that a harsh quality of the tense phonation of Northern Yi can be partially attributed to retraction of the tongue root. Similarly, a recent acoustic study of Southern Yi (Xinping village; Shi and Zhou, 2005) also showed that H1-H2 is an important acoustic correlate of phonation contrasts, and they also found that F1 is consistently higher for the vowel with tense phonation. Kuang (2011) examined both electroglottographic and acoustic properties of tense vs. lax contrast in Southern Yi (the same corpus used here, in part from Xinping village), and found that H1-H2 and H1-A1 are the best acoustic correlates of the phonation contrast, and furthermore that these acoustic measures both correlate with the contact quotient of vocal fold vibration. Moreover, using a cepstral measure that will be described below, she found that the tense phonation is more periodic than the lax phonation.

In the present study we will compare different electroglottographic measures of phonation to better understand both the production of tense vs. lax phonations in Tibeto-Burman languages. We will also compare these physiological measures to a range of acoustic measures of the audio signals.

3. Methods

3.1 Recordings

All the data in this study were obtained from recordings made during a trip to Yunnan province of China in the summer of 2009. Before the recordings were made, a wordlist of two thousand words[2] was elicited and archived in Excel as a small lexical database. These words covered things and events in everyday life, and had been used in fieldwork for many other Yi languages. The phonological system was then sorted out from this word pool and items were grouped into phonemes. Then this word pool was elicited again to check if the items had been correctly transcribed. This procedure needed to be repeated several times until the consultants agreed with all the homophones and minimal contrasts. The phonation register difference was easy to identify in the minimal pairs. Finally, a word list of monosyllable minimal pairs with all possible combinations of tone × phonation × vowels (about 40 pairs per language) was made for the purpose of this phonation contrast study. The number of tokens actually produced varies among speakers, because speakers were instructed to skip any pairs they did not know.

All the speakers were recruited from Southern Yi villages (Xinping and Jiangcheng), Bo villages (Shizong and Xingfucun) and a Hani village (Luchun). Twelve speakers per language were recorded for Bo and Yi, and eight speakers were recorded for Hani, with gender balanced. For all 32 speakers, simultaneous electroglottograph (EGG) and audio recordings were made. The signals were recorded directly to a computer via its sound card, in stereo, using Audacity, at the

[2] The corpus, built by Professor Baoya Chen and Feng Wang in Peking University, was a collection of high frequency words across various Tibeto-Burman languages, for the purpose of historical comparison among related languages.

sampling rate of 22050 Hz per channel. The audio signal was recorded through a Shure SM10A microphone as the first channel. EGG data were obtained by a two-channel electroglottograph (Model EG2, Glottal Enterprises, with a 40 Hz high-pass cutoff frequency) and recorded as the second stereo channel. Each word was repeated twice.

The analysis presented here includes recordings from all speakers. Since the phonation contrast does not occur with high tone, the data matrix of phonation by tone is not balanced. In order to be able to examine the interaction between tone and phonation, we thus exclude high tone tokens from the current analysis.

3.2 Measurements

3.2.1 Acoustic Measures

Comprehensive acoustic measures potentially reflecting different phonation properties, as described above, were made using VoiceSauce (Shue *et al.*, 2011). The frequencies of harmonics are estimated from the fundamental frequency, which is in turn estimated by the STRAIGHT algorithm (Kawahara *et al.*, 1999), and the location of formants is estimated by the Snack Sound Toolkit (Sjölander, 2004). Phonation-related acoustic measures include: The difference between the first harmonic (H1) and the second (H2), H1*-H2* (with formant corrections by Iseli *et al.*, 2007), controversially reflecting open quotient of the vocal folds (Holmberg *et al.*, 1995; Gobl and Ní Chasaide, 2012); for questions and issues see (Gerratt and Kreiman, 2001; Kreiman *et al.*, 2007), which has been found to successfully distinguish contrastive phonations across languages

(Gordon and Ladefoged, 2001; Keating *et al.*, 2011; 2012); amplitude of H1 relative to the amplitudes of the harmonics nearest to F1, F2, and F3 (H1*-A1*, H1*-A2*, H1*-A3*), indicating the strength of higher frequencies in the spectrum, which might be related to closing velocity of the vocal folds (Stevens, 1977), and have been found reliably distinguish between breathy vs. non-breathy phonation types (Blankenship, 2002; DiCanio, 2009; Esposito, 2012); individual harmonic amplitudes, H1*, H2* and H4*, which have been found to be important spectral landmarks of voice perception (Kreiman *et al.*, 2007); H2*-H4*, which has been found significantly correlated to gender (Kuang, 2011; Bishop and Keating, 2012); Cepstral peak prominence (CPP) (Hillenbrand *et al.*, 1994), reflecting the harmonics-to-noise ratio and periodicity, which has been found to be an indicator of contrastive breathy phonation (Blankenship, 2002; Garellek and Keating, 2011).

This set of measures has become fairly standard in recent literature (e.g. Green Mong: Andruski and Ratliff, 2000; Mazatec: Blankenship, 2002; Suai/Kuai: Abramson *et al.*, 2004; Javanese: Thurgood, 2004; Ju|'hoansi: Miller, 2007; Takhian Thong Chong: DiCanio, 2009; Santa Ana Valle Zapotec: Esposito, 2010; Garellek and Keating, 2011; Southern Yi: Kuang, 2011; White Hmong: Esposito, 2012; Gujarati: Khan, 2012).

3.2.2 EGG parameters

Electroglottography (EGG) is a popular method which measures variations in the vocal fold contact area during phonation. A small, high frequency current is passed between two electrodes that are placed on each side of the larynx. Variation in the electrical impedance across the larynx

is produced by the opening and closing of the vocal folds. The EGG signal is related to the contact area of the vocal folds: the larger the contacted area, the larger the measured admittance. Since the signal is not calibrated, it reflects relative rather than absolute contact. EGG is non-invasive, and does not interfere with speakers' natural production (compared to invasive methods); it can thus be used to study complex speech events, and is convenient for use outside the laboratory. In recent years it has been widely used in studies of linguistic phonation, and plays an important role in documenting uses of non-modal phonations in various under-described languages (e.g. Maa: Guion *et al.*, 2004; Tamang: Mazaudon and Michaud, 2006; Takhain Thong Chong: DiCanio, 2009; Mazaudon and Michaud, 2009; Santa Ana Del Valle Zapotec: Esposito, 2010; White Hmong: Esposito, 2012; and Gujarati: Khan, 2012).

The EGG signals corresponding to the audio tokens were processed by EggWorks (Tehrani, 2012) to obtain the traditional landmark-based parameter measures. Figure 2-1 illustrates the EGG parameters. Contact Quotient (CQ), which is defined as the proportion of the vocal fold contact during each single vibratory cycle (Rothenberg and Mashie, 1988) is estimated with four methods: 1) EGG threshold (Rothenberg and Mashie, 1988): the contact event is defined as the time point when the signal strength exceeds a threshold of peak-to-peak amplitude (CQ method in Figure 2-1); 2) dEGG (Henrich *et al.*, 2004): the contact and opening events are defined on peaks in the derivative of the EGG signal (CQ_PM method in Figure 2-1) Hybrid (Howard, 1995): use the dEGG contacting peak for detecting the glottal contact event, and an EGG-based 3/7 threshold for detecting the glottal opening event (CQ_H method in Figure 2-1); 4) Tehrani hybrid method (c.f. documentation of EggWorks; same threshold was used in (Davis *et al.*,

1986): the contact event is defined by dEGG contacting peak, and the opening event is defined by the y-value of the dEGG contacting peak (CQ_HT method in Figure 2-1). Two measures are made from the dEGG signal: Peak Increase in Contact (PIC in Figure 2-1) (Keating *et al.*, 2011), defined as the amplitude of the positive peak on the DEGG wave, corresponding to the highest rate of increase of vocal fold contact (Michaud, 2004); Peak Decrease in Contact (PDC in Figure 2-1), defined as the amplitude of the negative peak of dEGG. Finally, closing duration and opening duration are measured at a 10% threshold (Marasek, 1997), and Speed Quotient, which reflects the skewness of the pulses (Esling, 1984; Holmberg *et al.*, 1988; Dromey *et al.*, 1992; Marasek, 1996), is computed as the ratio between closing and opening duration.

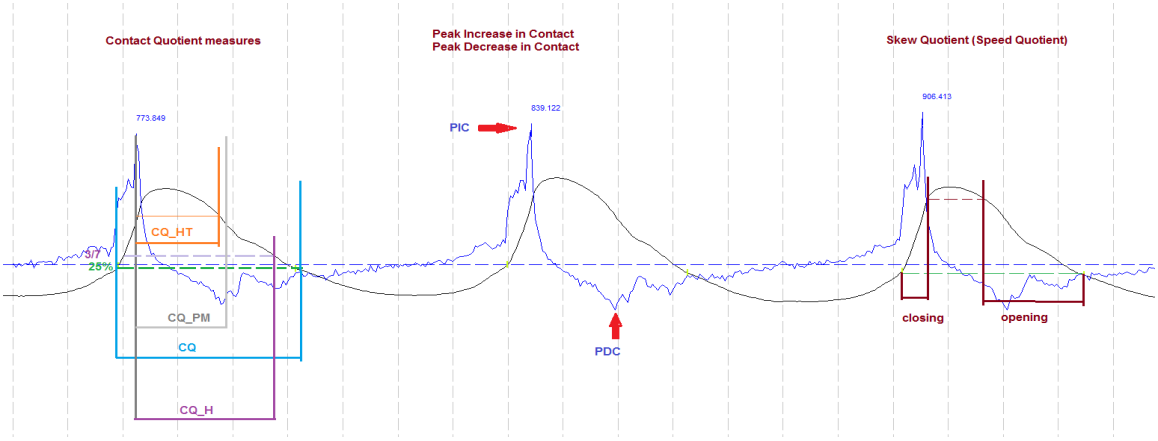


Figure 2-1 Demonstration of EGG measures from EggWorks. The black line is the EGG pulses, and the blue line is the dEGG signal. The first pulse demonstrates four methods of estimating CQ, the middle pulse demonstrates PIC and PDC from dEGG, and the third pulse demonstrates measures related to skewness of the pulse: closing duration and opening duration (SQ=T-closing/T-opening). The thresholds shown are schematic only.

4 Results

4.1. Linear Mixed-Effect models

Since preliminary analysis showed that these three languages are quite similar, we collapse them together for analysis here. A series of mixed-effect models (one per measure) were employed to evaluate the main effects and interactions of phonation and tone on acoustic voice measures, using the *lme4* package in R, with phonation and tone as the main effects, and speaker as the random effect. In order to model our data, we created three kinds of models with predetermined random effects structures: models with by-speaker random intercept but no random slopes; models with by-speaker random slopes for tone and phonation, but no random intercepts; models with both random slops and random intercepts. In order to select the best model, the *anova* function is used to compare the performance of different models. Goodness of fit is decided by the log likelihood (loglik in the models) and Chi-square tests. As representative examples, (1) (2) and (3) below present the model comparisons for CQ, H1*-H2* and F0, the most important measures for phonation and tone.

(1) Model comparison for F0
Models:
Model 1: strF0_mean ~ Tone * Phonation + (1 | Speaker)
Model 2: strF0_mean ~ Tone * Phonation + (1 + Tone + Phonation | Speaker)
Model 3: strF0_mean ~ Tone * Phonation + (Tone + Phonation | Speaker)

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
f0.1	6	37478	37515	-18733			
f0.2	11	37339	37408	-18659	191.6	5	<2e-16***
f0.3	11	37339	37408	-18659	0	0	1

(2) Model comparison for H1*-H2*
Models:
Model 1: H1H2c_mean ~ Tone * Phonation + (1 | Speaker)
Model 2: H1H2c_mean ~ Tone * Phonation + (1 + Tone + Phonation | Speaker)
Model 3: H1H2c_mean ~ Tone * Phonation + (Tone + Phonation | Speaker)

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
h1h2.1	6	21139	21176	-10563			
h1h2.2	11	21086	21155	-10532	62.611	5	3.50E-12***
h1h2.3	11	21086	21155	-10532	0	0	1

(3) Model comparison for CQ
Models:
Model 1: CQ_H_mean ~ Tone * Phonation + (1 | Speaker)
Model 2: CQ_H_mean ~ Tone * Phonation + (1 + Tone + Phonation | Speaker)
Model 3: CQ_H_mean ~ Tone * Phonation + (Tone + Phonation | Speaker)

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
cq.1	6	-8824.5	-8786.9	4418.2			
cq.2	11	-8889.9	-8821	4455.9	75.423	5	7.59E-15***
cq.3	11	-8889.9	-8821	4455.9	0	0	1

As can be seen here, larger log likelihood ratios and large Chi-squared values indicate that the models with by-speaker random slopes for phonation and tone (model 2 and model 3) are significantly better than the model without random slopes (model 1) (p-values < 0.001); and there is no difference between models with both random slopes and intercept (model 2) and models with random slope only (model 3). In other words, the by-speaker random effects of phonation and tone are really on slopes. In sum, model 3 is the best model for our data.

The current version of the lme4 package in the R statistical software does not provide p-values for t- and F-tests. A popular way to obtain p-values is to use R's *pvals.fnc*, which is based on the Markov chain Monte Carlo (MCMC) method (Baayen, 2010). However, this function fails to estimate the degree of freedom when there is a random slope, and so it cannot be used in our

study. Therefore, we must resort to an alternative method, two-tailed *t*-tests with the degrees of freedom at the upper bound (observations minus fixed effect). It has been demonstrated that this upper bound works reasonably well for large data sets with over 100 observations as the *t*-distribution approximates the normal distribution. A simple way of assessing significance at the 5% significance level is to check whether the absolute value of the *t*-statistic exceeds 2. Therefore, we report statistical significance by exact student *t*-value and its *p*-value based on the upper bound degree of freedom (Bates *et al.*, 2008; Baayen, 2010).

4.2. Distinctive main effects of phonation and tone

The main effects of tone (low, mid) and phonation (tense, lax) in the three languages are summarized in Table 2-2. Original tables of outputs for each measure are in appendix. Only significant effects at a $p < .05$ level are reported in the tables, and direction is noted.

Table 2-2. Main effect of phonation and tone. Only significant effects (p <.05) are reported and direction is noted, t-values are in parentheses.

	Tone	Phonation
H1*		Tense<Lax (8.32)
H1*-H2*		Tense<Lax (10.69)
H1*-A1*	Mid<Low (4.71)	Tense<Lax (8.77)
H1*-A2*	Mid<Low (4.15)	Tense<Lax (8.16)
H1*-A3*	Mid<Low (2.47)	Tense<Lax (6.78)
CPP	Mid>Low (8.6)	Tense>Lax (3.19)
H2*		
H4*	Mid<Low (4.04)	
H2*-H4*	Mid<Low (2.28)	
F0	Mid>Low (10.5)	
CQ		Tense>Lax (7.34)
PIC		Tense<Lax (4.06)
Closing Dur		Tense<Lax (3.15)
SQ		

Table 2-2 shows that phonation and tone have some distinctive effects on pitch and voice measures, and some shared effects. This distinction is the most salient between EGG measures and F0 (bottom of the table): CQ, PIC and closing duration only have phonation effects but no tone effect; by contrast, F0 has only a tone effect but no phonation effect. As expected, the tense

phonation consistently has a greater CQ than the lax phonation[3], suggesting a smaller open quotient in the vocal folds. In addition, the pulse shapes of lax and tense phonation are generally quite similar, as there are no significant phonation effects on SQ, but the closing duration is significantly shorter for the tense phonation. Finally, the consistently smaller PIC values for the tense phonation arguably suggest the tense phonation has a slower glottal closure. In general, the glottal articulations involved in the tense vs. lax contrasts are not extreme, and they are controlled independently from pitch so that speakers are able to keep pitch constant when producing different phonation types.

The independence between phonation and tone can also be seen in some of the spectral measures. There is a general agreement across the three languages on the phonation effect. Phonation has a significant main effect on H1* and the strength of H1* relative to the higher-frequency harmonics (H1*-H2* and H1*-An*). In general, tense phonation has a less prominent H1*, and a shallower spectral slope (H1*-A3*). As indicated by a higher CPP, tense phonation is also either more periodic, or has more harmonic energy. In contrast to the H1* related measures, H2*-H4* and H4* only show a main effect of tonal categories. This result for all three languages is consistent with the previous study on Southern Yi only (Kuang, 2011). It is interesting that the two spectral ranges (H1-H2 vs. H2-H4) have distinctive functions in a language with both phonation and tonal contrasts: H1-H2 for phonation, and H2-H4 for tone. Mixed results are found for the spectral measures that cover larger ranges of frequency (e.g. H1*-A1*, H1*-A3*). This is probably because these measures cover the ranges that are important for both tone (H2-

[3] Different methods have the same pattern.

H4) and phonation (H1-H2). Overall, then, phonation and tone have quite distinctive effects, indicating that tone and phonation are produced with different physiological configurations.

4.3 Interactions between phonation and tone

Since the tense vs. lax contrast is crossed with tonal categories in these three Yi languages, the interaction between phonation and tone was also considered in the mixed-effect models.

Neither the EGG measures (CQ, PIC and Closing Duration) nor F0 show significant interactions between phonation and tone (ref. appendix for *t*-values), further supporting the hypothesis that phonation and tone are very independent from each other. Figure 2-2 and Figure 2-3 demonstrate the distinctive articulatory properties between phonation and pitch: CQ has only the main effect of phonation, and no interaction between phonation and tone; F0 has only the main effect of tone, and no interaction between phonation and tone either.

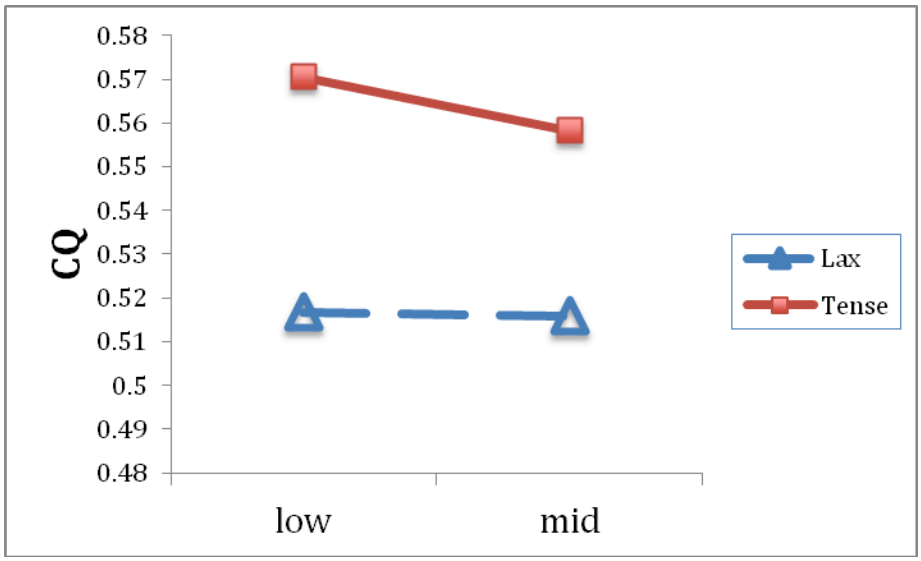


Figure 2-2 phonation and tone effects on CQ: The tense phonation has a greater CQ; the two lines are almost parallel, indicating that there is no significant interaction. Tense=solid line; Lax=dashed line

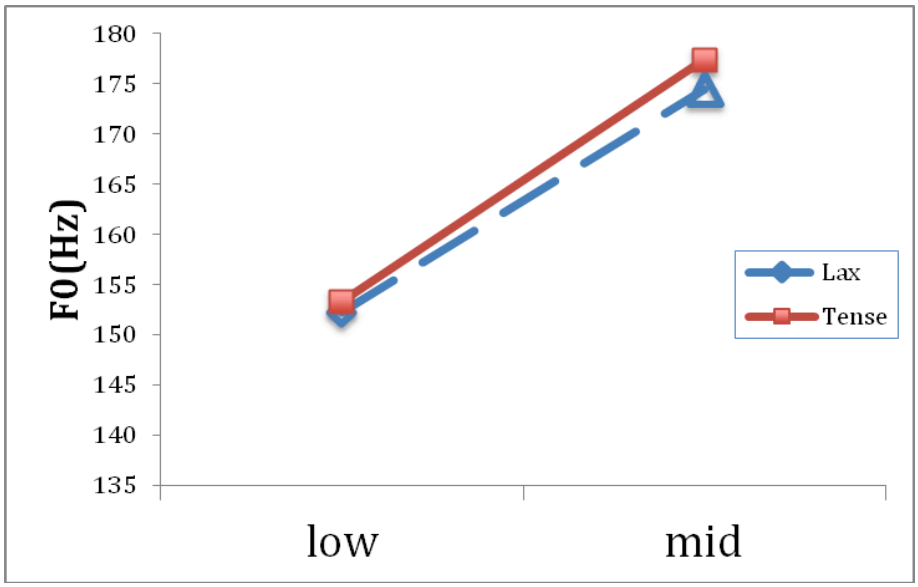


Figure 2-3 Phonation and tone effects on F0: mid tones have higher F0; the two lines are overlapping, indicating that there is no significant interaction. Tense=solid line; Lax=dashed line.

Although distinct from each other in articulation, phonation and tone can slightly interact with each other in spectral measures. For example, consistent significant tone by phonation interactions were found for H1*-H2* and H1*-A1* in all three languages. In general, the tense vs. lax contrast is better distinguished in low tone than in mid tone (Figure 2-4). The H1*-H2* difference between tense and lax phonation in low tone is consistently around 3 dB, whereas it is only 1 dB in mid tone (Hani has a better distinction, with a 2 dB difference for mid tone). Similarly, the phonation contrast in H1*-A1* is about 3.5-4 dB in low tone, 1 dB in mid tone. This might explain why the tense vs. lax contrast cannot occur with high tone, because the phonation contrast gradually becomes less and less distinguished as pitch increases. A similar effect is also observed in Mazatec (Garellek and Keating, 2011). Overall, phonation contrasts are independent from tonal contrasts in these languages, but with the phonation distinctiveness gradually reduced for the higher tones.

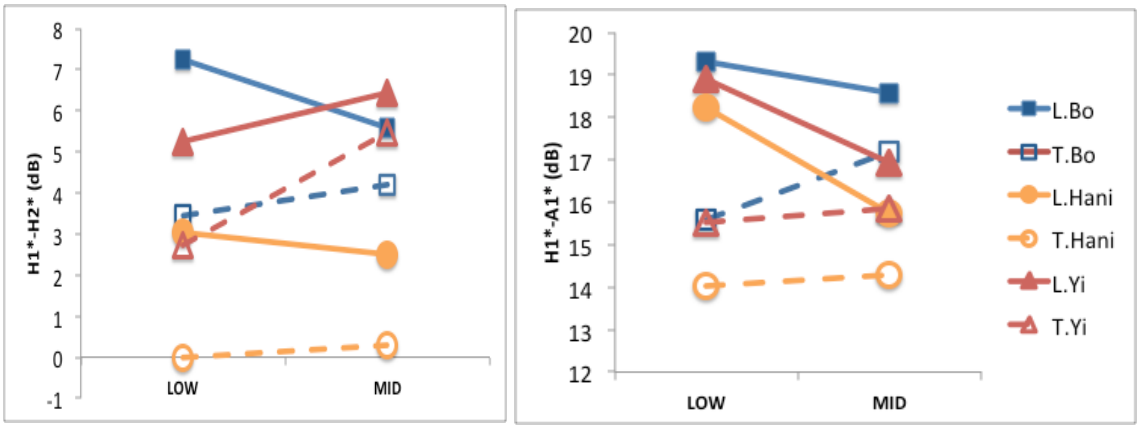


Figure 2-4 The interaction between phonation and tone for H1*-H2* (left panel) and H1*-A1* (right panel). Head pattern and color indicates languages (square=Bo, round=Hani, triangle=Yi), line pattern indicates phonation (solid=Lax, dashed=Tense).

4.4 The distinctive mechanisms of phonations vs. pitch

4.4.1 Understanding the mechanism of the phonation contrast

Since multiple cues are involved in the phonation contrast, we would like to know which phonetic correlates make the greatest contributions. Therefore, a series of forward stepwise logistic regressions were employed to evaluate the relative importance of different measurements for the phonation contrast.[4] Acoustic and EGG measures introduced in the previous sections were included as the predictors, and phonation (tense vs. lax) was the dependent variable. The relative importance of each measure was estimated by *p*-values based on Wald Chi-Square test (three languages are combined). A predictor with more importance should have a smaller *p*-value. In order to visualize the relative importance among measures in a more intuitive way, we plot the $-\log_{10}(p\text{-value})$ so that more significant measures have higher bars than less significant ones, as shown Figure 2-5. The models for acoustic measures suggest that H1*-H2* and H1*-A1*, i.e. the lower frequency range of the spectra (relative to F2, F3), are the best acoustic correlates to distinguish phonation contrasts across the three languages; but all H1* related measures, both H1* itself and its relative strength (H1*-Hn*, H1*-An*) measures, make significant contributions to the phonation contrast. CPP also makes a significant contribution, but H2*, H4*, and H2*-H4* make no or very little contribution to the phonation contrast. As for the models for EGG measures, CQ, the ratio of the contact phase to the entire glottal cycle, is the single best measure.

[4] The backward stepwise method does not work for this kind of dataset, where measures are highly correlated with each other, because the dropping process will kill the variables which are most correlated with the best contributing variable.

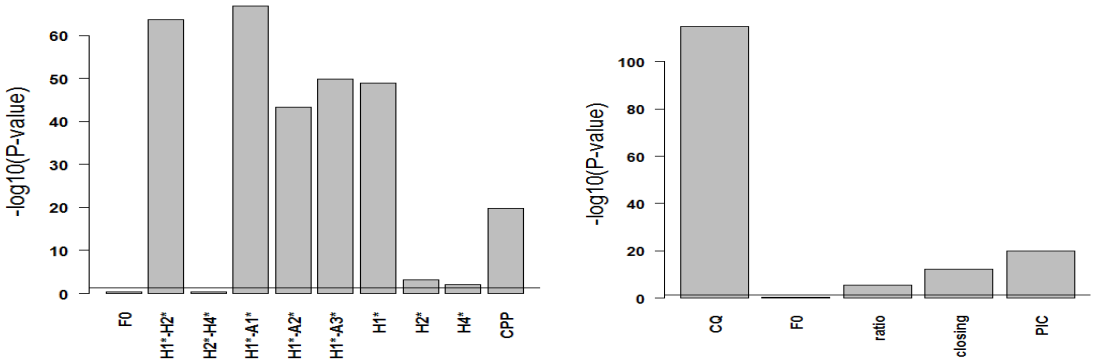


Figure 2-5 Contributions of measures to the phonation contrast: stepwise logistic regressions for acoustic measures (left) and EGG measures (right). *p*-values are estimated from Wald Chi-square test, and converted into positive integers by $-\log_{10}(p\text{-value})$. Higher $-\log_{10}(p\text{-value})$ indicates higher significance of contribution. The horizontal line is $p=0.05$. Scales are different.

A series of Spearman correlations were done to understand the relationships between EGG parameters and acoustic correlates. The results show that all H1* related measures are significantly correlated with CQ, and H1* (r=-0.59, $p<0.01$), H1*-H2* (r=-0.59, $p<0.01$) and H1*-A1* (r=-0.69, $p<0.01$) are best correlated with CQ. Sundberg (1987) has suggested that greater opening results in greater peak glottal flow, which in turn gives a higher H1. Also as expected, an inverse correlation with CQ indicates that smaller CQ values can lead to a more prominent H1*. Therefore, these measures together suggest that the tense vs. lax phonations primarily differ in the relative degree of glottal opening.

4.4.2 The distinctive mechanism of tonal contrasts

To compare with the results in Figure 2-5, another series of logistic regressions were employed to evaluate the relative importance of different measures for tonal contrasts. The same acoustic and EGG measures were included as the predictors, and tone (mid and low) was the dependent variable. Again, the relative importance of each measure was estimated by *p*-values based on the Wald Chi-Square test. We again visualize the strength of significance by converting these original *p*-values into positive integers by $-\log_{10}(p\text{-value})$, as shown in Figure 2-6.

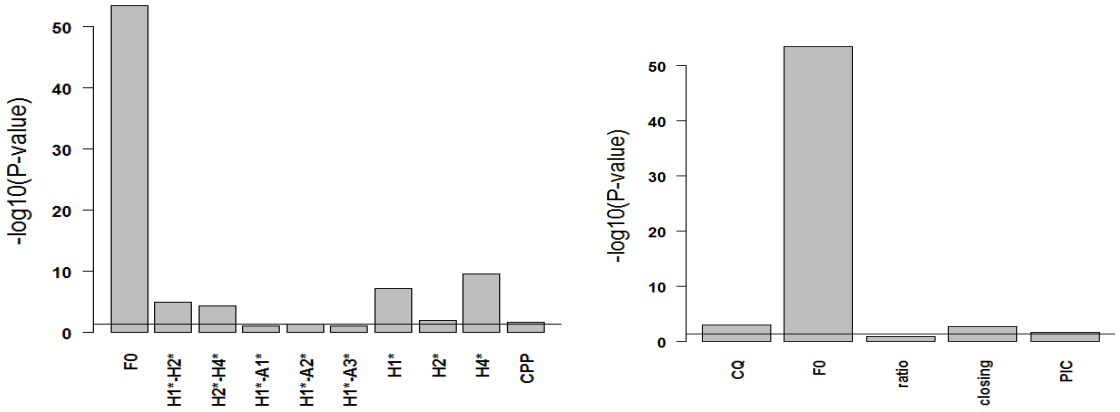


Figure 2-6 Contribution of measures to tonal contrasts: stepwise logistic regressions for acoustic measures (left) and EGG measures (right). P-values are estimated from Wald Chi-square test, and converted into positive integers by $-\log_{10}(p\text{-value})$. Higher $-\log10(p\text{-value})$ indicates higher significance of contribution. The horizontal line is p=0.05.

As can be seen here, Figure 2-6 has a very distinct pattern from Figure 2-5. F0 is the dominant phonetic correlate for tonal contrasts; the measures that are related to the degree of glottal opening, e.g. CQ and H1*-related measures, make very little contribution to the tonal contrasts. Interestingly, H2*-H4* only makes a significant contribution to tonal contrasts but not to phonation contrasts. This result is consistent with the results in 4.2. Although H1*-A1*, H1*-

A2* and H1*-A3* showed significant main effects of tone (section 4.2), they do not make significant contributions to tonal contrasts in the logistic regression model, suggesting that they are not reliable cues for tonal contrasts. Again, the results demonstrate that tonal contrasts and phonation contrasts are subject to different mechanisms.

A series of Spearman correlations were done between F0 and phonation-related voice measures, in order to understand the relationships between pitch and phonation. Results are shown in Table 2-3.

Table 2-3 Correlation coefficients (r values) between F0 and phonation-related measures

	H1*	**H1*-H2***	**H1*-A1***	**H1*-A2***	**H1*-A3***	**CQ**
F0	0.26	0.14	-0.15	-0.14	-0.1	0.03

It can be seen that none of these phonation-related measures have good correlations with F0. Noticeably, CQ, the most important EGG correlate for the phonation contrast, has no correlation with F0. Acoustically, H1*and H1*-An* have very weak correlations with F0 as well. This means that the control of glottal opening is generally independent from pitch control.

5. Discussion

This chapter has investigated the phonation and tonal contrasts of 32 speakers of three Yi languages, and provides evidence of the independence of pitch and phonation in these contrasts. The EGG data clearly indicated that the tense-lax register contrasts in these languages involve

phonation (laryngeal) differences, while the low-mid tone contrasts do not: EGG measures have strong main effects of phonation, but no main effect of tone; by contrast, F0 has a strong main effect of tone, but no effect of phonation; neither EGG measures nor F0 show significant interactions between phonation and tone. These results together demonstrated that phonation and tone have very independent articulations from each other. Phonation can be kept constant while changing pitch, and pitch can be kept constant while changing phonation.

Moreover, according to the regression analysis, the most important articulatory measure of the phonation contrasts is Contact Quotient, which means that the different degrees of glottal opening is the essential mechanism of the tense vs. lax contrast. Finally, the correlation analysis showed that CQ is uncorrelated with F0 in these languages, which means that the control of glottal opening is independent from the control of pitch. This result is supported by previous physiological studies on phonation and pitch production (Laver, 1980; Gobl and Ní Chasaide, 2012): glottal opening is controlled by the interarytenoid and lateral cricoarytenoid muscles, but pitch is controlled by the cricothyroid muscle.

Although not as strong as the EGG data, acoustic data also demonstrate the independence between pitch and phonation. In this chapter, we identified six successful acoustic measures for this type of phonation contrast: H1*, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3* and CPP, possibly reflecting the open quotient (H1*, H1*-H2*, H1*-A1*), the abruptness (H1*-A3*, H1*-A2*), and the periodicity (CPP) of the vibration. Among these reliable measures, H1*-H2* and H1*-A1*, the acoustic indicators of the degree of glottal opening, turn out to be the most

important acoustic correlates. This is consistent with the result of the EGG data. Results from linear mixed-effect models as well as logistic regression models suggest that H1*-related spectral measures are not reliable cues for tonal contrasts, since they either have no main effect of tone or make no significant contributions in predicting tonal contrasts. In the previous study on Mazatec (Garellek and Keating, 2011), no main effect of tone is found for any H1* related measures as well. Therefore, these languages have in common that phonation contrasts are indeed independent from tonal (pitch) contrasts. The EGG data from this chapter provides strong evidence for the distinctive mechanisms of pitch and phonation productions.

In contrast to the H1* related measures, our results suggest that H2*-H4* is significantly affected, and only affected, by the tonal contrasts. It has a significant main effect of tone, but no main effect or interaction of phonation. It is also a significant contributor to the tonal contrasts. The physiological mechanism behind this measure is still unclear, but it seems to be related to tonal contrasts (Kuang, 2011) and gender difference (Bishop and Keating, 2012). We speculate that, although physiologically distinctive, pitch and phonation can affect different ranges of the spectrum, so that spectral measures can show more complicated interaction with tones. For example, in this chapter, we showed that the tense vs. lax contrast is acoustically more distinguished in low tone than in mid tone.

In sum, this chapter has demonstrated that phonation can be very independent from pitch, so that phonation contrasts and tonal contrasts can occur orthogonally in languages. In the next chapter,

we will look into a different relationship between pitch and phonation, where phonation is part of the pitch scale.

Chapter 3 Mandarin: A case study of phonation dependent on tone

1. Introduction

As we have seen from the previous chapter, for Yi languages non-modal phonation is a primary cue for distinguishing lexical meanings, and phonation cues are independent from pitch cues. Articulatorily, as shown in Kuang and Keating (2012), F0 in these languages can be kept constant while changing vocal fold vibration patterns. This chapter will look into a different type of case, where non-modal phonation is an allophonic or secondary cue for tonal contrasts. In this kind of case, non-modal phonation occurs only optionally, with certain tonal categories. Mandarin is such a language, perhaps the most well-known one. As has been well-documented, this language has four lexical tones, represented with Chao's numbers as 55 (high level), 35 (rising), 214 (low-dipping), and 51(falling)[5]. Among the four, Tone 3 (214) is usually, though not always, realized with a creaky voice (Hockett, 1947; Chao, 1956; Davison, 1991; Belotel-Grenié and Grenié, 1994, to cite a few).

Creak in Mandarin Tone 3 shows various larygealization properties in acoustic waveforms: aperiodic, or period doubled, or low-frequency pulse-like vibratory patterns (Gerratt and Kreiman, 2001). Figure 3-1 shows examples of creaky Mandarin Tone 3; in contrast, Figure 3-2 shows an example of non-creaky Tone 3.

[5] 55 is really 44, as it is not extra high as defined in the IPA; 35 is really 24, as it is not extra high either; 214 is really 213, as it ends lower than Tone 1 and Tone 2.

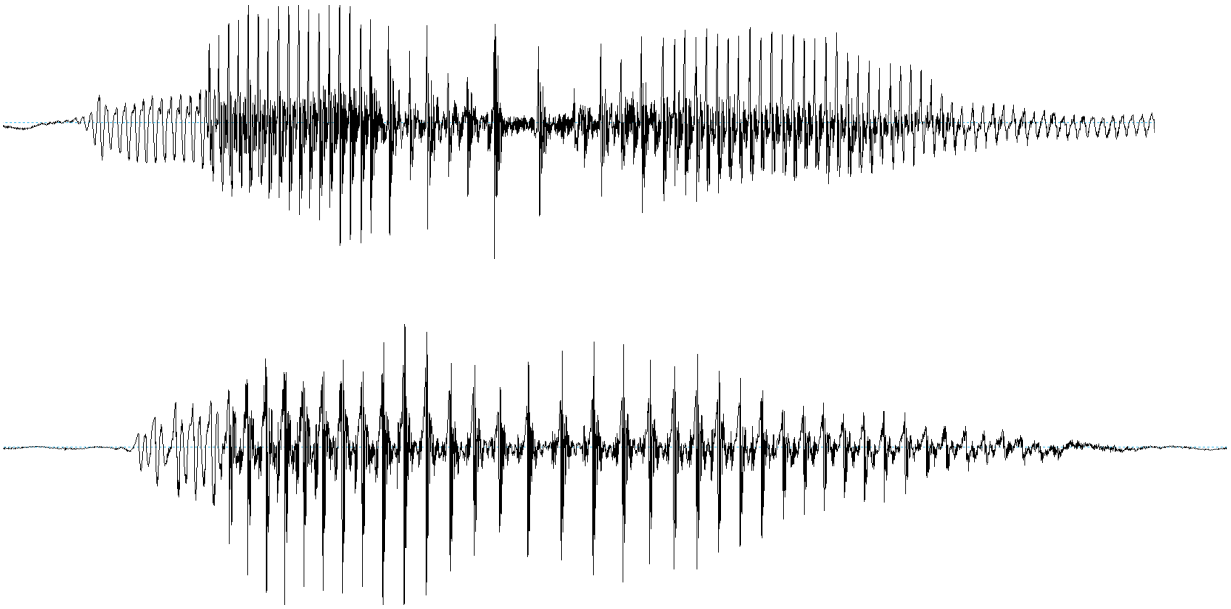


Figure 3-1 Examples of typical creaky Tone 3. Upper: example of aperiodic vibration; lower: example of low-frequency pulse-like vibration.

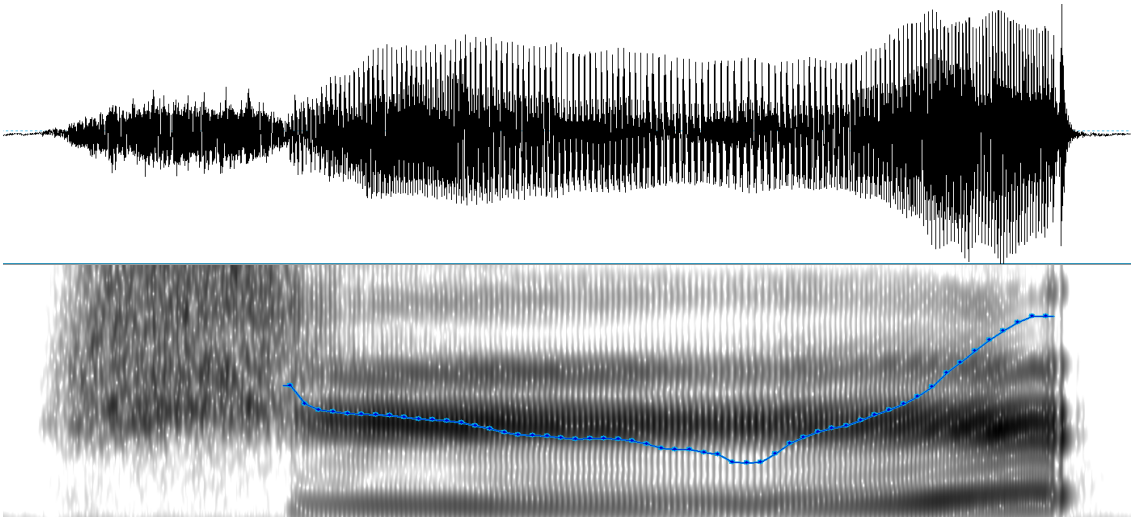


Figure 3-2 An example of non-creaky Tone 3[6]

[6] The last pulses of this syllable is a glottal stop

Because the presence of creak is variable within and across speakers, phonological studies and a majority of perception studies usually do not take it into account. However, some Mandarin perception studies have showed that the presence of creaky voice can facilitate Tone 3 identification. For example, in a tonal identification experiment with both creaky and non-creaky natural Tone 3 stimuli, Belotel-Grenié and Grenié (1997) found that listeners recognize Tone 3 faster in the creaky condition. A more recent study with resynthesized stimuli (Yang, 2011) suggested that creaky voice is especially helpful to distinguish Tone 3 (214) from Tone 2 (35), which are the most phonetically similar tonal pair in Mandarin. Relatedly, a similar effect is also found in Cantonese (Yu and Lam, 2011): the presence of allophonic creak can bias the recognition of the lowest tone (Tone 21) from the tone with similar pitch values (Tone 22). Therefore, non-modal phonation is an important phonetic cue of these lowest tones in these two languages.

With its optionality in production and salience in perception, some questions are raised about Mandarin creak: what mechanism leads to its presence in the Mandarin tonal contrasts? What functions does it play in the tonal system?

To answer these questions, we need to understand whether the presence of creak is naturally driven by phonetic conditions or is exclusively conditioned by certain tonal categories (e.g. Tone 3). In the former case, we would expect that creak always co-occurs with very low pitch targets in any tone, as creak naturally happens when the pitch is very low (e.g. Titze, 1990). That is, both Tone 3 (214) and Tone 4 (51) should have creaky voice present at the low target portion of

the tones, i.e. the /1/ part in Chao's representation. Indeed, Belotel-Grenié and Grenié (1994; 1997) found that not only Tone 3 but also Tone 4, and even some Tone 2, can involve creak. It is quite possible that native speakers regard creak as the lowest pitch (Keating and Esposito, 2007). However, it is also possible that creak in Mandarin is independent from pitch and tied to certain tonal categories (Davison, 1991). In that case, we would expect that Tone 3 should show creak even when the tone's pitch is not especially low, e.g. when spoken in a high pitch range.

This chapter will provide some evidence to evaluate these two hypotheses. First, in order to see whether creak is strictly tied to Tone 3 or instead occurs with all low pitch targets, we will replicate Belotel-Grenié and Grenié (2004)'s study, counting the presence of creak in a small corpus. To provide further evidence, in the second study we test whether manipulating pitch range can change the voice quality in tonal production. We will also try to understand the relation(s) between pitch and voice quality across the pitch range of Mandarin.

2. Experiment 1: The presence of vocal fry – T3-specific, or for all low targets?

This small corpus study is to test the question: whether creak is Tone 3 specific or occurs with low targets in both Tone 3 and Tone 4. Mandarin minimal tone sets with the syllable /ma/ are retrieved from a database of EGG and audio recordings from 12 (6M/6F) native Beijing Mandarin speakers (http://www.phonetics.ucla.edu/voiceproject/voice.html). Speakers were college students in Beijing, aged between 18 and 25. All the test words are monosyllabic and were read in isolation in the speakers' most comfortable pitch range, and each token was

repeated five times. In order to investigate the correlation between creaky voice and tones with low targets, we further subset Tone 3 and Tone 4 from the dataset, a total of 120 tokens.

Creak is measured by coding its presence, i.e. the occurrence of creak in a given signal will be coded as 1, otherwise 0. Frequency of occurrence of creak in T3 and T4 thus can be counted. Pitch values before and after vocal fry portions are measured so that we can know the pitch range when vocal fry can occur. The presence of creak is defined if the token had audible creak, and if there were: alternating cycles of amplitude and/or frequency or irregular glottal pulses in the waveform, or missing values or discontinuities in the f0 track determined by Praat's autocorrelation algorithm. The results are shown in Table 3-1, along with the results from previous studies.

Table 3-1 Frequency of presence of creak in current and previous studies

	Style	**Speakers**	**Tone 1**	**Tone 2**	**Tone 3**	**Tone 4**
Current study	Laboratory	6F/6M	--	--	60/ 60	39/60
Belotel-Grenié & Grenié 1997	Laboratory	4M/3F	0/53	8/44	40/51	18/56
Belotel-Grenié & Grenié 2004	Broadcast	1F	0/55	1/40	17/64	4/121

In general, we found that not just all Tone 3 tokens but also most Tone 4 tokens in this small corpus present creak, which means that the presence of creak is not Tone 3 specific, but is sensitive to all low targets. This result can be compared with Belotel-Grenié and Grenié (1997; 2004)'s results. As shown in Table 3-1, in their study, not only Tone 3 but also a considerable amount of Tone 4, even small proportion of Tone 2, were produced with creak, though less than

in the present study. It is worth mentioning that Belotel-Grenié and Grenié (2004) also counted creaky voice in neutral tones, and it turned out that a great proportion of the neutral tones (6/17) were produced with creak. In general, then, creak happens very often in Mandarin speech, and is not just limited to Tone 3. Across studies, the actual amount of creak varies depending on both speakers and speech style.

Furthermore, in our corpus creak happens at similar pitch values for both Tone 3 and Tone 4, below around 180 Hz for female speakers and 95 Hz for male speakers. This is shown in Figure 3-3, which plots the highest F0 values at which creak was observed (averaged across speakers). This further suggests that creaking naturally takes place when F0 falls below certain values.

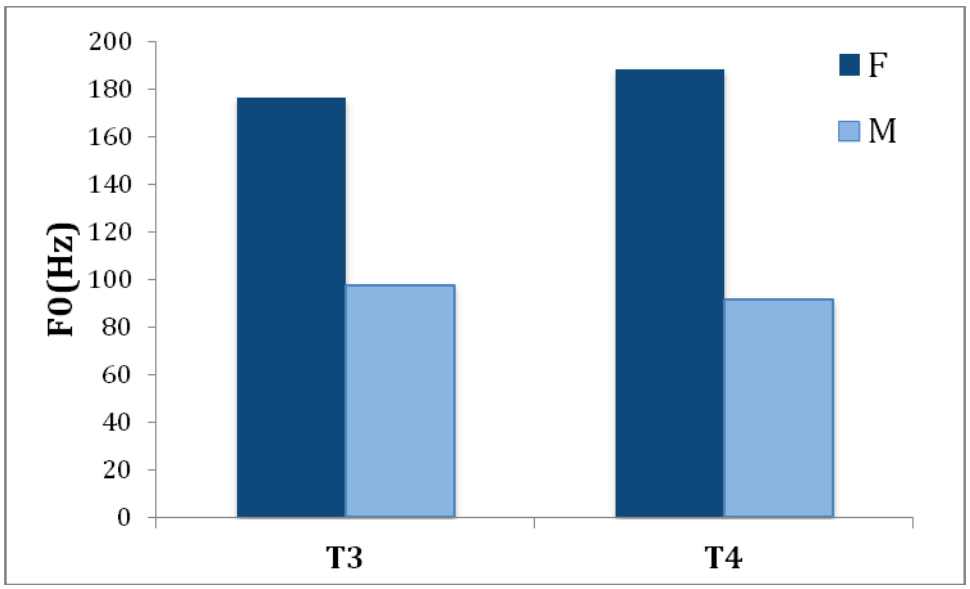


Figure 3-3 Pitch values around creak for Tone 3 and Tone 4

In sum, this small corpus study replicates previous studies and supports the hypothesis that only F0 values matter for the presence of vocal fry in Mandarin. In other words, vocal fry is the sign

of lowest pitch; whenever speakers reach the bottom of their pitch ranges, they tend to creak, no matter with what tonal categories. Of course, as a result of the pitch differences among the tones, the frequency and magnitude of creak are not evenly distributed among them. Since the incidence (and probably the extent as well) of creak is greater in Tone 3, it is apparently more salient in Tone 3 than in any other tone.

The hypothesis that only F0 values matter for the presence of creaky voice in Mandarin will be further validated in a corpus that consists of three production conditions: exclamatory speech (with pitch range raised), normal speech (with default pitch range) and low speech (with pitch range lowered). The hypothesis predicts that if creak is correlated to F0, then when a speaker's pitch range is raised, there will be less frequent creak, or less-creaky voice.

3. Experiment 2: Whether pitch range can affect the voice quality of tonal production

3.1. Understanding pitch ranges in normal, low and exclamation conditions

The corpus recorded for Keating and Kuo (2012)'s pitch range study was retrieved from http://www.phonetics.ucla.edu/voiceproject/voice.html. For our purpose, tonal productions in isolation from 22 (11M/11F) Mandarin speakers will be reanalyzed here. In this corpus, speakers were instructed to produce the Mandarin four lexical tones with the syllable 'shi' (orthographic transcription of " 师 时 使 是") in three different conditions: (1) Normal pitch: words are produced in speakers' most comfortable pitch ranges; (2) Exclamation: words are produced as if there is an exclamation mark after the word (e.g. shi!); (3) Low pitch: words are produced in a lower pitch range (some of the speakers also produced these in a quieter manner). These

recordings were made at a 44.1 kHz sampling rate and a 32bit quantization rate. F0 values were measured using the STRAIGHT algorithm (Kawahara *et al.*, 1999) in VoiceSauce (Shue *et al.*, 2011). F0 values during creak voice have been manually checked and corrected. Figure 3-4 and Figure 3-5 present the effects of three conditions on pitch ranges.

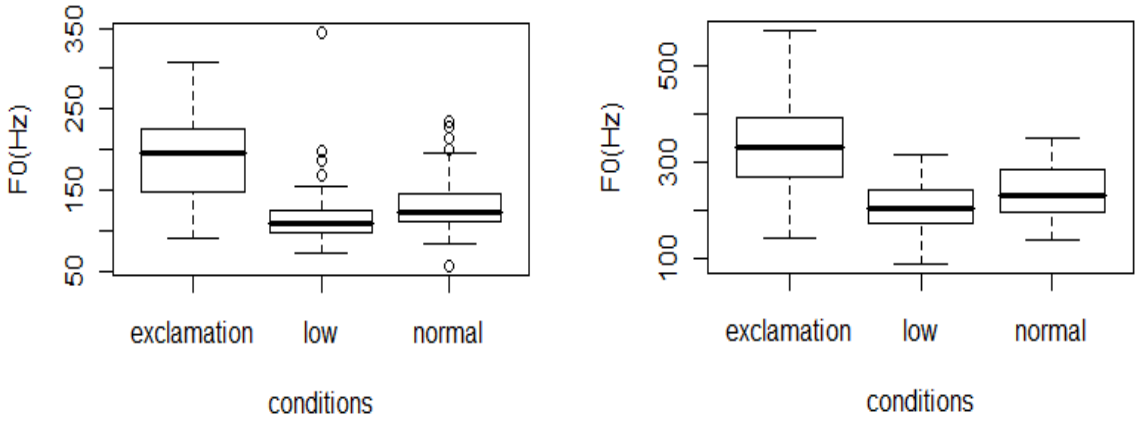


Figure 3-4 Overall pitch ranges in three production conditions, left panel = male, right panel= female. x-axis represents the three production conditons; y-axis represents the F0 values in Hz.

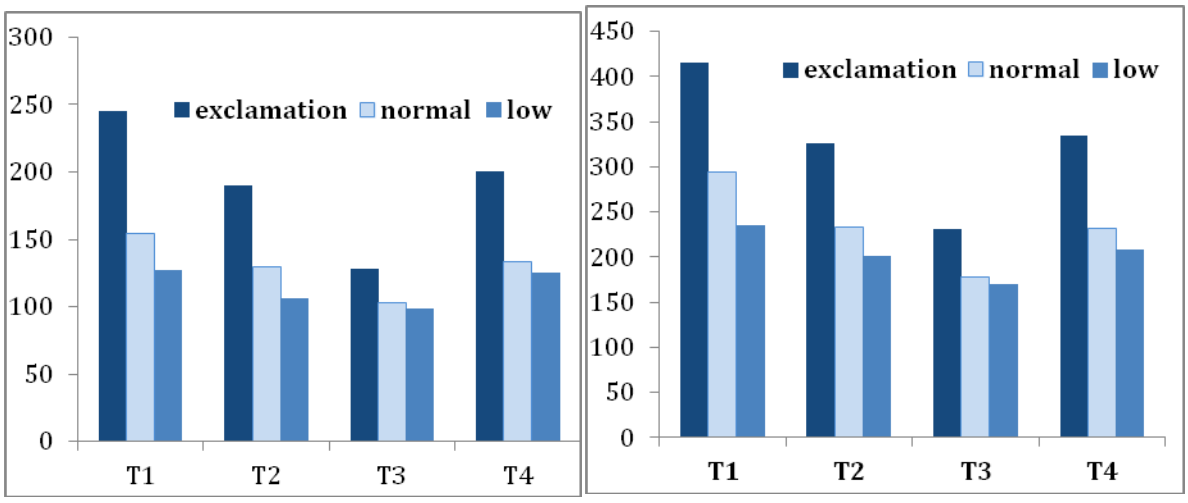


Figure 3-5 Mean pitch values in three production conditions. Left panel = male; right panel = female. y-axis represents the F0 values in Hz; x- axis represents the four lexical tones.

As can be seen here, the pitch ranges exhibit the expected pattern, which is consistent across four tonal categories: compared to the normal condition, tones have a higher pitch range in the exclamation condition and a lower pitch range in the low condition. This means that the speakers understood the task and produced the target words properly. Moreover, it is noted that the overall pitch range of the exclamation condition is not just raised but also significantly larger than that of the other two conditions. Particularly, the lower limit of the pitch range does not change significantly, and the enlarged pitch range is mainly effected by the increase in the upper limit. In contrast to the exclamation condition, the overall pitch range of the low condition is slightly compressed and lower relative to the normal condition.

With this understanding of the three pitch-range conditions, we now can examine their effects on creak in Mandarin tonal production. Unlike the preliminary study, we will not count the presence of non-modal phonation; instead, we will measure the voice quality during the tonal productions, and examine whether voice quality can be affected by different pitch-range conditions. If voice quality is associated with pitch range, as concluded from the preliminary study, we would expect that Tone 3 is produced with a creakier phonation in a lower pitch range, but with a breathier phonation in a higher pitch range. Furthermore, if the conclusion is true, we should not only expect a change in voice quality for the low pitch targets, but should also see a significant change for the high pitch targets, depending on the pitch range.

3.2. Method

The amplitude difference between the second harmonic and the fundamental frequency, H1*-H2* (corrected for formants and bandwidth using the algorithm by Iseli *et al.*, 2007), is selected to represent the property of voice quality. Keating and Shue (2009) found it most relevant to F0 changes in Mandarin, and this measure also has been found to be the most successful acoustic indicator of phonation across languages (c.f. the previous chapter). For each vowel, 12 time intervals are extracted, and both F0 and H1*-H2* are automatically measured by VoiceSauce (Shue *et al.*, 2011). Both overall mean H1*-H2* and values from certain portions of the vowels will be used here.

3.3. Results: whether pitch range affects tonal production

3.3.1 Overall effects on each tonal category

A Linear mixed-effect model is employed to decide whether pitch range has effects on overall voice quality (i.e. mean H1*-H2* of the entire vowels) in producing four tonal categories. Production conditions (exclamation, normal and low), tonal categories (Tone 1, 2, 3 and 4) and gender and all their interactions are set as the fixed effects, and speaker as the random intercept. Overall effects are estimated by the *anova* function. Highly significant main effects are found for both production conditions ($F\ [2,765]=14.9$, $p<0.01$) and tonal categories ($F\ [3,764]=9.8$, $p<0.01$), but there is no gender effect ($F\ [1,766]=0.58$, $p>0.5$). Therefore, the results indicate that overall, different pitch ranges can affect voice quality in tonal production.

Moreover, significant interactions between conditions and tonal categories are also found ($F[6,756]=7.0$, $p<0.01$). Therefore, effects of production conditions are further examined within each tonal category, presented in Table 3-2. Only significant effects are reported here (*p*-value (MCMC) < 0.05), and "normal condition" is set to be the reference level in the mixed-effect models.

Table 3-2 Main effects of production conditions on each tonal category (only significant *p*-values estimated by MCMC methods are reported here)

	Tone1	**Tone2**	**Tone3**	**Tone4**
Exclamation	**0.017**		**0.05**	**0.02**
Low	**<0.0001**		**0.03**	

As indicated in Table 3-2, the salient effects of production conditions are seen in Tone 1 and Tone 3, the highest and the lowest tones; Tone 2 seems to be not affected by the change of pitch range; Tone 4 is affected by the exclamation production, but not by the low-pitched condition. The mixed finding seems to suggest that tonal categories are more likely to be affected by the change of pitch range when there are sustained extreme pitch targets, e.g. Tone 1 (high) and Tone 3 (low). The overall weak effects on Tone 2 and Tone 4 are possibly due to complex pitch targets and more dynamic changes of pitch, which can cancel out each other's effects over time. Figure 3-6 shows the mean H1*-H2* values in each condition for each tonal categories; recall that a high H1*-H2* indicates a more breathy/less creaky voice quality.

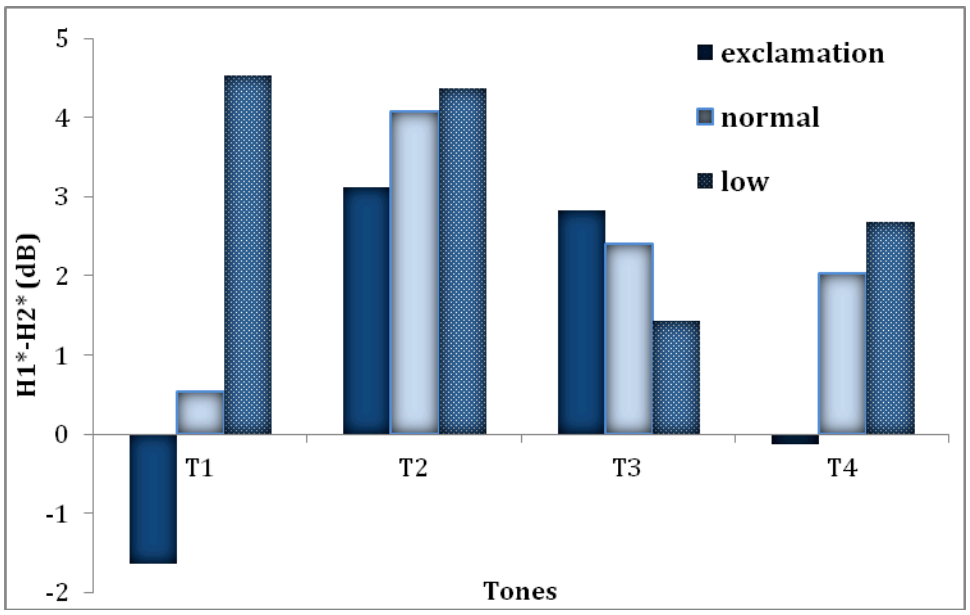


Figure 3-6 Mean H1*-H2* values in each production condition, broken down into tonal categories

Two directions of pitch-range effects are shown in Figure 3-6. The voice quality in Tone 1 and Tone 4 becomes much tenser in the exclamation condition, indicated by a decreased H1*-H2* value compared to normal condition; on the other hand, when pitch range is lowered, their voice quality become much breathier, indicated by increased H1*-H2* values. The difference is very salient for Tone 1, but less salient for Tone 4, and weak in Tone 2. Interestingly, in contrast to the tones involving high pitch targets, pitch range has an opposite impact on Tone 3, the lowest tone. The voice quality of Tone 3 becomes less creaky (greater H1*-H2*) when pitch range is raised, and creakier (smaller H1*-H2*) when pitch range is lowered. This result is consistent with the conclusion from the preliminary study that creak in Tone 3 is associated with a certain pitch range, and thus sensitive to the change of pitch range.

3.3.2 High pitch targets vs. Low pitch targets

To more clearly test the effects of pitch range on pitch targets and to break down the complex targets in contour tones, the dataset is rearranged based on pitch targets. The subset comprising high targets includes mean F0 values for Tone 1 (we assume that there is only one pitch target for this tone), and the first third F0 values for Tone 4. And the subset comprising low targets includes the second third F0 values of Tone 3 and the last third F0 values of Tone 4. Separate mixed-effect models are run for each subset, with production condition, tonal categories and gender (and their interactions) as the fixed effects, and speaker as the random intercept.

Overall, production conditions have a strong effect on the voice quality of the high targets ($F[2,399]=17.883$, $p<0.01$); the gender effect is also very salient, as there are both a significant main effect of gender ($F[1,400]=6.3$, $p<0.01$) and a significant interaction between production conditions and gender ($F[2,395]=14.3$, $p<0.01$). But there is no effect of tonal categories ($F[1,400]=0.04$, $p>0.1$) and no significant interaction between condition and tonal categories ($F[2,395]=1.25$, $p>0.1$), which means that tonal categories do not matter. The mean H1*-H2* values of high targets (whole vowel for Tone 1, first third for Tone 4) in three production conditions are presented in Figure 3-7, conditioned by tonal categories (left) and by gender (right).

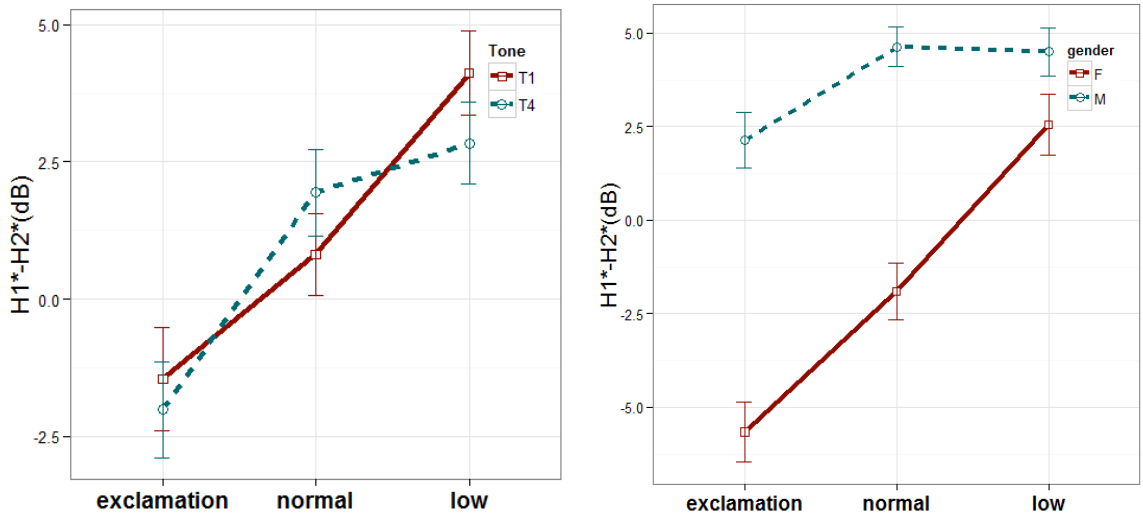


Figure 3-7 Mean H1*-H2* values for high target intervals, by tonal categories (left) and by gender (right)

We can see that the voice quality of high pitch targets shows a consistent pattern in Tone 1 and Tone 4: consistent with the findings in the previous section, the voice quality of high pitch targets is breathier when the overall pitch range is lower, and is tenser when the overall pitch range is higher. In addition, female speakers have a more dramatic change in the exclamation condition than male speakers.

As for the low targets (middle third of Tone 3, last third of Tone 4), overall significant effects on their voice quality are found for production conditions ($F[2,357]=3.68$, $p<0.05$), gender ($F[1,358]=7.4$, $p<0.01$) and tonal categories ($F[1,358]=14$, $p<0.01$). There are also significant interactions between production conditions and gender ($F[2,353]=3.6$, $p<0.05$). Crucially, there is no significant interaction between condition and tonal categories ($F[2, 353]=1.1$, $p>0.1$), which

means that conditions have similar effects on low targets in both Tone 3 and Tone 4. To understand the tonal differences, two more mixed-effect models are run for these target intervals in Tone 3 and Tone 4 separately, and significance is estimated by MCMC p-values. Compared to the normal condition, voice quality of the low target in Tone 3 is significantly different in both exclamation ($p < 0.05$) and low ($p < 0.01$) conditions; however, voice quality of the low target in Tone 4 is only significantly different in the exclamation condition ($p < 0.05$) but not in the low condition ($p > 0.1$). The mean H1*-H2* values of low targets in three production conditions are presented in Figure 3-8, by tonal categories (left) and by gender (right).

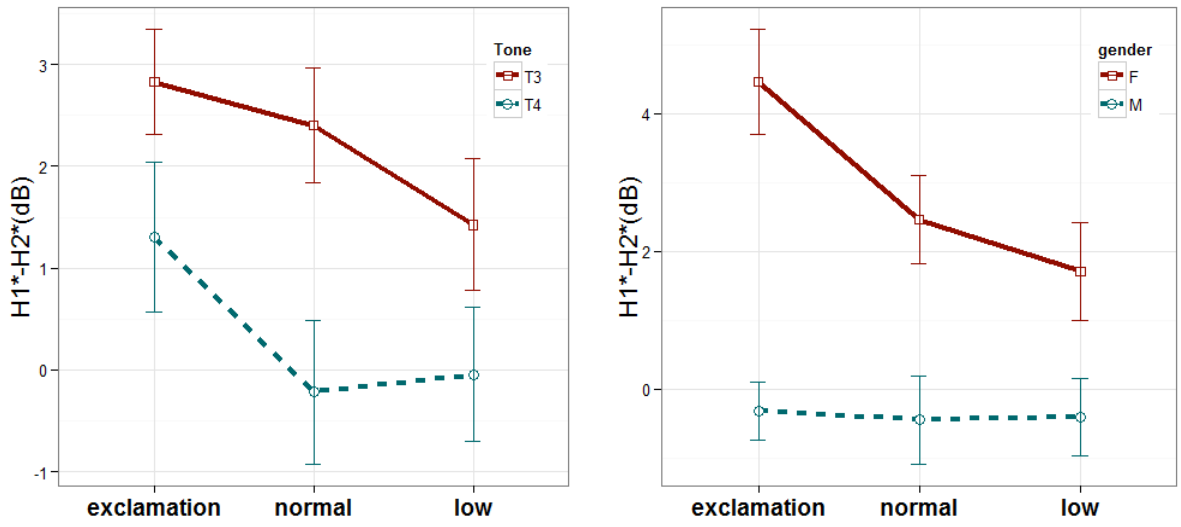


Figure 3-8 Mean H1*-H2* values for low target intervals, by tonal categories (left) and by gender (right)

As shown in Figure 3-8, low targets in Tone 3 and Tone 4 have a consistent pattern, except that low speech does not significantly change the voice quality of Tone 4. This could be because

these two conditions do not have different pitch ranges. To test this possibility, a mixed-effect model is run for F0 values of Tone 4 low targets, with condition as the fixed effect and speaker as random intercept. The result shows that significantly different F0 values are only between normal and exclamation conditions, but not between normal and low conditions. Overall, the results shown in Figure 3-8 are consistent with previous results: low targets are creakier in the low condition and breathier in the exclamation condition. And again, female speakers have more salient changes in voice quality than male speakers.

To validate these statistic results, we randomly pick several speakers from the corpus to see whether production conditions would affect the presence of creak. Indeed, we found that creaky speakers can turn off creak in their exclamation speech, while non-creaky speakers can turn on creak in their low speech. Figure 3-9 is an example of within-speaker variations.

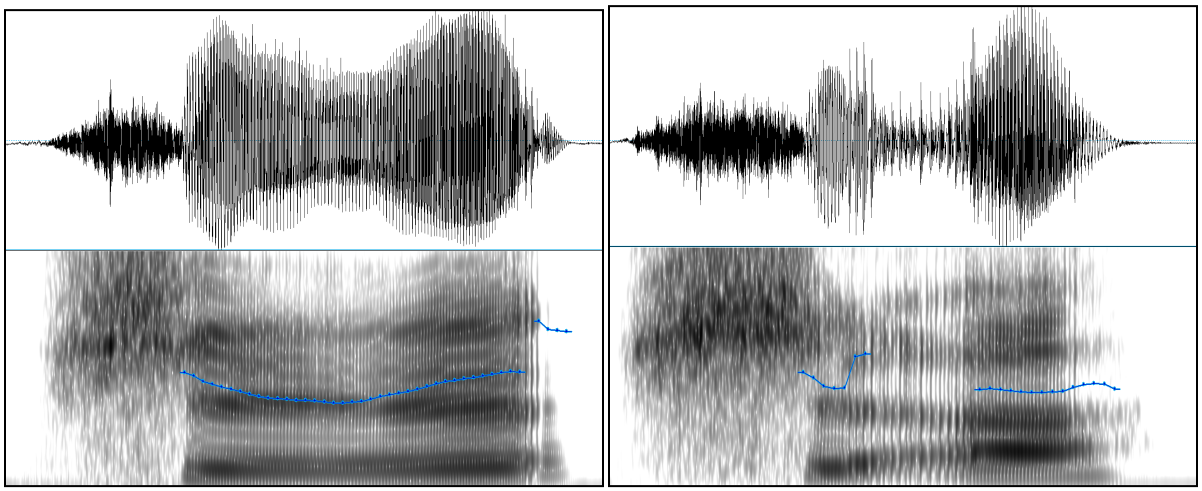


Figure 3-9 An example of within-speaker variations of Tone 3: a non-creaky speaker (speaker number=F35September10) produces creak in the low condition (right). Left=normal speech; right = low speech.

To sum up this section, Figure 3-7 and Figure 3-8 form a mirror image, and they together demonstrate that different pitch ranges affect the voice quality in both low targets and high targets of Mandarin tones: low targets become breathier when pitch range is raised, but creakier when pitch range is lowered; by contrast, high targets become tenser when pitch range is raised, but breathier when pitch range is lowered. These results support the claim based on the first corpus study, that non-modal phonation in Mandarin is very sensitive to pitch range, and creak is naturally driven by a very low pitch. To integrate these effects of pitch ranges, in the following section, further analyses will be provided to investigate how voice quality co-varies more generally with the pitch scale.

4. Experiment 3: Variations of voice quality along the pitch scale

4.1 Dynamic voice quality changes during tonal productions

In the previous sections, we analyzed tonal production in a static way. But a close investigation reveals that the voice quality of tones changes dynamically over time, with pitch, especially for Tone 2 and Tone 4. In order to show the dynamic change of pitch and correspondent voice quality over time, Figure 3-10 and Figure 3-11 plot the F0 tracks as well as the H1*-H2* tracks over 12 time-intervals for both female and male speakers. Tonal categories are identified by their colors and patterns. These figures provide more detailed information about the effects of pitch ranges on voice quality.

The upper panels of Figure 3-10 and Figure 3-11 show the F0 contours for the Mandarin four tones in three conditions, and the lower panels of these figures gives the correspondent H1*-H2*

contours. It can be seen that Tone 3 in the lower pitch range appears to have a longer duration of creaky voice, although the other two conditions also can have creak. For the females, Tone 3 in the exclamation condition can be very breathy. More importantly, these two figures show that voice quality and pitch do not co-vary linearly. For example, as shown in Figure 3-10 and Figure 3-11, the H1*-H2* track for Tone 4 is a complex contour, and the timing of the peak varies depending on different pitch-range conditions. Similar effect can be found for Tone 2 as well. There is no simple linear explanation for these different contours. Therefore, to understand better the correlation between voice quality and pitch, we analyze a third corpus, comprising unprompted pitch sweeps over maximum F0 ranges.

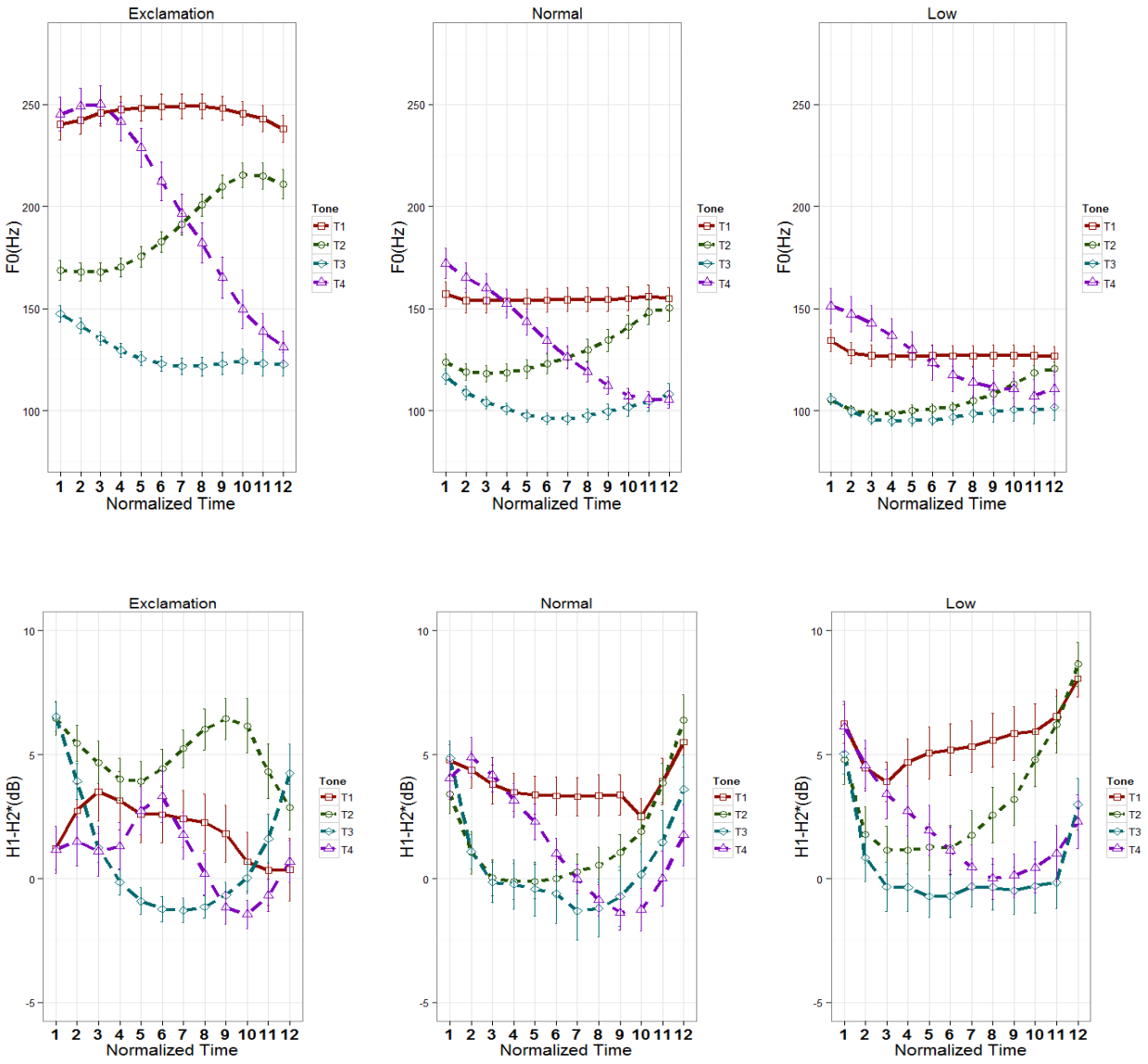


Figure 3-10 Dynamic pitch and phonation changes in tonal production, in three production conditions (male speakers).

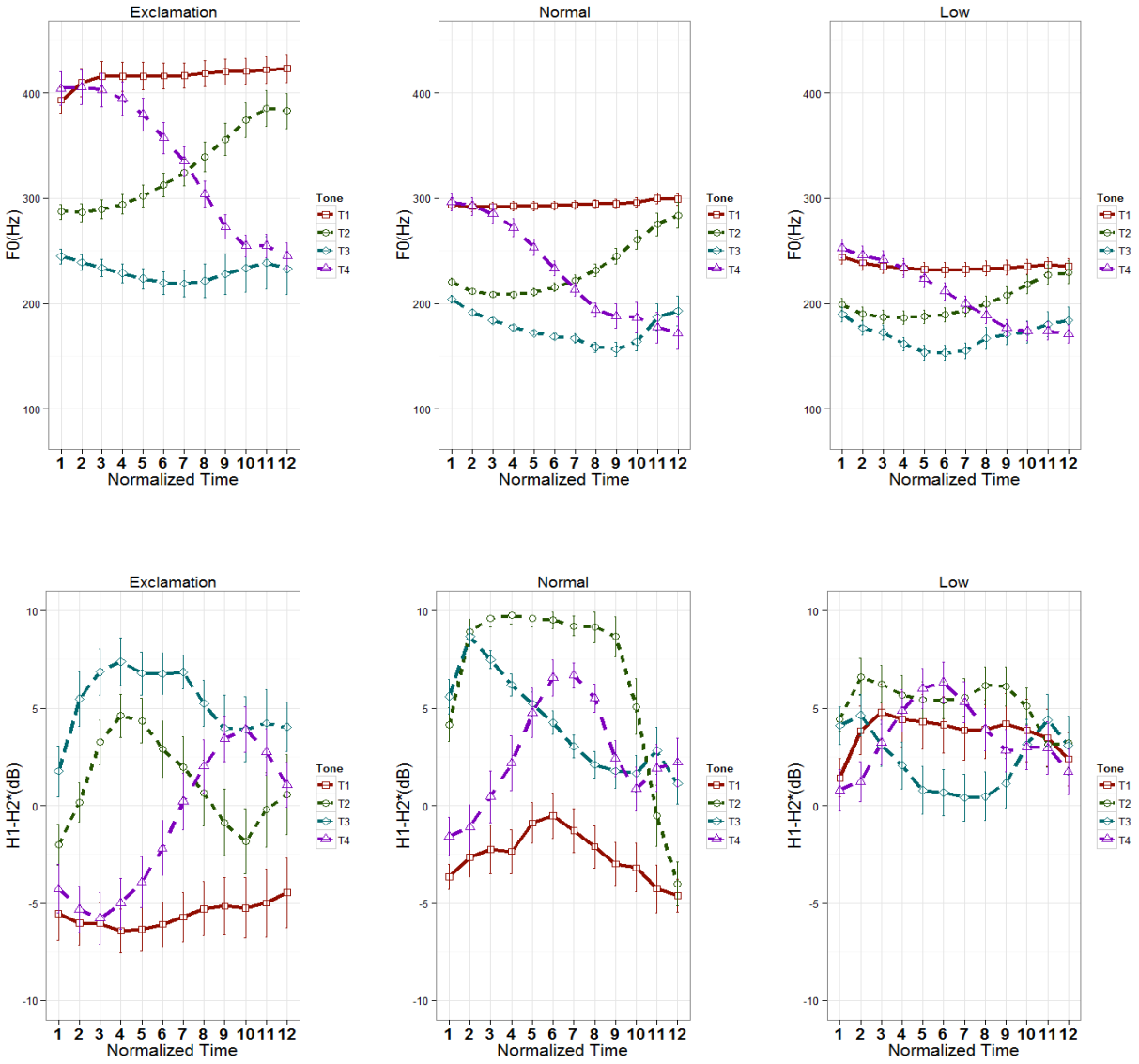


Figure 3-11 Dynamic F0 and voice quality changes during tonal production, in three production conditions (Female speakers).

4.2. Unprompted pitch glides: Correlations between voice quality and pitch

4.2.1 Method

The corpus of unprompted sweeps from 21 (10F/11M) Mandarin speakers was originally recorded for Keating and Kuo (2012). These recordings were also used in Keating and Shue (2009)'s preliminary analysis of within-speaker correlations between voice quality and pitch, the results from which will be compared with the results from the new analysis. For this corpus, speakers were instructed to start at a comfortable, normal pitch, and then sweep in three different ways: 1) gradually rise until they feel their voice break; 2) gradually lower with a breathy voice to avoid creak; 3) gradually lower and fall into creak. As a demonstration, Figure 3-12 shows mean F0 tracks of the three types of sweeps produced by the female speakers.

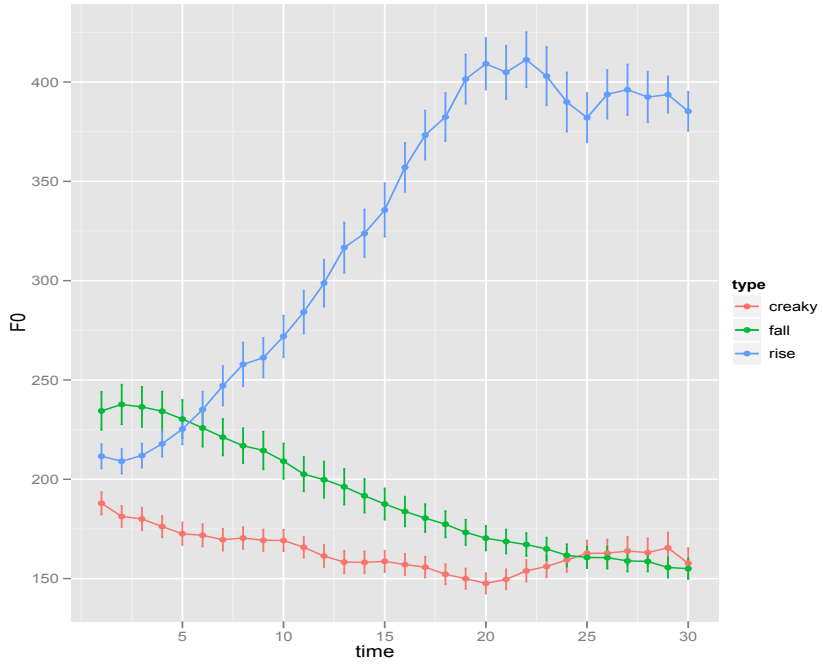


Figure 3-12 Demonstration of unprompted pitch sweeps and change in voice quality (10 Mandarin female speakers). Blue=rise, green=falling with a breathy voice, red=falling into creak

F0 values in these sweeps with 30 time-intervals were measured by the STRAIGHT algorithm in VoiceSauce, and tokens with missing values are excluded. Because aperiodic creak happens, there are many missing values at the ends of the creaky sweeps, so the minimum F0 values shown in Figure 4-12 might not be perfectly accurate. F0 values during creak were measured manually in Keating and Kuo (2012), and the minimum F0 is about 136-150 Hz for female speakers, and 77-85 Hz for male speakers. For our purpose, the automatic measures are accurate enough to show the pitch ranges.

As the previous study (Keating and Shue, 2009) showed that H1*-H2* is the best correlated voice measure (r^2=0.5) with F0 for Mandarin speakers, we will continue using this measure to indicate voice quality. More importantly, Keating and Shue (2009) showed that the correlation between F0 and H1*-H2* is non-linear. Therefore, we will not use linear regression to estimate the correlation between pitch and voice quality. Instead, in a more exploratory way, we will use *loess* function (Local Polynomial Regression fitting) to identify the non-linear co-varying relationship between F0 and H1*-H2*[7]. Data points are pooled into different sweep types (e.g. rise, fall and creak) and gender; for each time interval, a mean H1*-H2* value is calculated over all speakers; and finally average H1*-H2* values with 30 time-intervals are plotted and fitted with a local polynomial regression line (loess smoother).

[7] The disadvantage of this method is that it does not provide a regression function that is easily represented by a mathematical formula, or a reliable estimate of of r-square.

4.2.2. Results

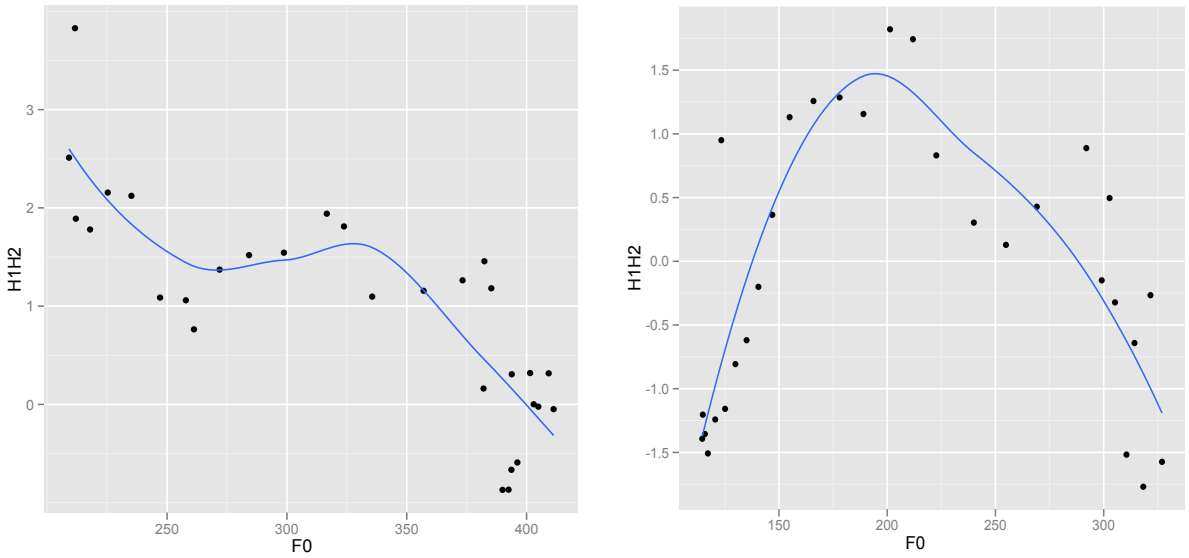


Figure 3-13 Relationship between F0 and H1*-H2* for rise sweeps. Left=female; right=male. Because the x-axis = F0, time runs from left to right.

Figure 3-13 shows the estimated relationships between F0 and H1*-H2* for the rise sweeps. Both female and male speakers show a non-linear correlation between F0 and H1*-H2*, but the patterns are quite different for female speakers and male speakers. For male speakers, the relationship between F0 and H1*-H2* is wedge-shaped: the two measures have a positive correlation until reaching around 180 Hz, and then the correlation becomes negative (though with a blip around 290 Hz not captured by the fitted function). For female speakers, although F0 and H1*-H2* generally have a negative correlation, there is also a turning point at around 270 Hz, followed by a small plateau.

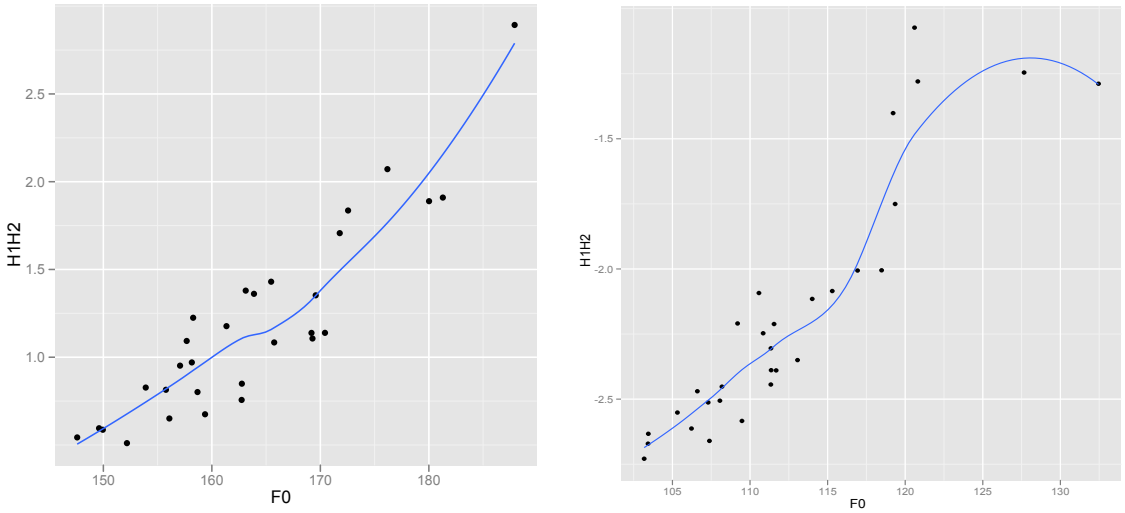


Figure 3-14 Relationship between F0 and H1*-H2* in creaky falling sweeps. Left=female, and right=male. Because x-axis=F0, time here runs from right to left.

Figure 3-14 depicts the relationship between F0 and H1*-H2* in the creaky falling sweeps. Overall, the two measures have a very good positive correlation (r^2=0.75 with a linear regression), which means that the lower, the creakier. But this pitch range is fairly small, 150 Hz – 200 Hz for females, and 100 Hz – 130 Hz for males. In addition, more strictly speaking, their relationship is not completely linear as well. The slopes of the fitted lines become steeper as F0 increases.

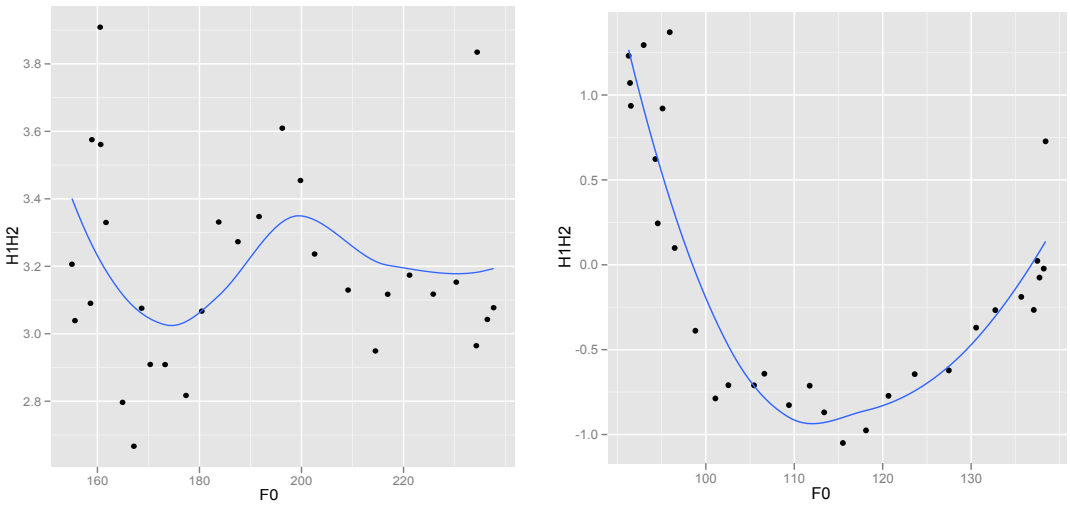


Figure 3-15 Relationship between F0 and H1*-H2* for breathy falling sweeps. Left=female, and right=male. Because x-axis=F0, time here runs from right to left.

Figure 3-15 shows the relationship between F0 and H1*-H2* in the breathy falling sweeps. The patterns here are quite different from Figure 3-14. For females, as F0 decreases (from right to left), voice quality gradually becomes breathier, and reaches a peak around 200 Hz; then from 200 Hz to 170 Hz, voice quality gradually becomes creakier, and reaches a valley around 170 Hz; after this point, voice quality gradually becomes breathier. In contrast, the male speakers show a U-shape. From 150 Hz to 110 Hz, voice quality and pitch have a negative correlation (the lower the creakier), but at around 110 Hz, voice quality turns to the breathy direction. These turning points are very interesting, as they have similar values to the ones previously found where creak or vocal fry is about to occur (shown in Figure 3-3). Vibration status must have changed at this moment to avoid falling into creak, which is the default way of reaching the lowest pitch (Figure 3-14).

4.3 Summary of Experiment 3

To summarize the complex relationships shown in Figure 3-13, 3-14 and 3-15:

(1) Female speakers: their natural breathiest voice quality (peak H1*-H2* in Figure 3-15 left) happens in the lower half of the entire pitch range (about 200 Hz), and voice quality goes creakier or tenser as F0 goes below or above this point. Below 200 Hz, there is a positive correlation between voice quality and pitch; below 170 Hz, it appears that creak will occur by default unless otherwise changes are made in glottal settings. Above 200 Hz, there is a negative correlation between voice quality and pitch; at around 270 Hz, there is another voice break to interrupt the locally linear relationship between voice quality and F0 (Figure 3-13).

(2) Male speakers: their natural breathiest voice quality (peak H1*-H2* in Figure 3-13 right) happens in the lower half of the entire pitch range (about 180 Hz), and voice quality goes creakier or tenser as F0 goes below or above this point. Below 180 Hz, there is a positive correlation between voice quality and pitch; below 110 Hz, creak is about to occur by default unless otherwise changes in glottal settings. Above 180 Hz, there is a negative correlation between voice quality and pitch.

Therefore, the overall relationship between F0 and H1*-H2* is like a "wedge" (Keating and Shue, 2009), as demonstrated with female productions in Figure 3-16. Male speakers have a similar pattern, but with different F0 values for the turning points[8].

[8] The Keating and Kuo (2012) dataset also includes English speakers. So we also cross-validated this wedge pattern with English speakers. We found very similar patterns for English speakers as well. Therefore, we speculate that this non-linear co-varying effect is a general physiological property of human beings, and different groups of people only differ in its size and actual turning points.

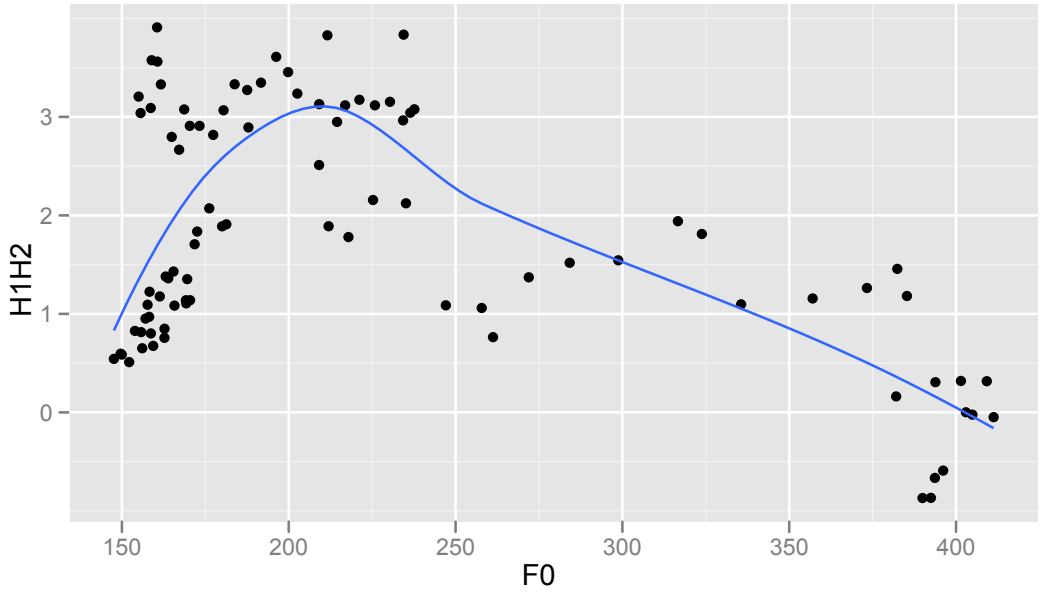


Figure 3-16 The overall relationship between H1*-H2* and F0 (showing female speakers). Data points from Figure 13 – 15 are put together. The blue line is not perfectly fitted as sweep types are mixed, and the dipping at 250 Hz is not well captured by the fitting function.

These patterns can be cross-validated with the results from previous sections. It can be easily understood that low targets become breathier when pitch range is raised (Figure 3-7), and high targets becomes breathier when pitch range is lowered (Figure 3-8). The dynamic changes shown in Figure 3-10 and Figure 3-11 can also be understood now. One can check the peaks and valleys of H1*-H2* for Tone 2 and Tone 4, and find that they follow the pattern shown in Figure 3-16. For example, peaks in female productions also show up around 200 Hz, the most comfortable range for females (c.f. Figure 3-12). All in all, in this section, we show that voice quality can closely co-vary with F0, but not along a linear scale. Non-modal phonation is likely to happen when the pitch range is lower or higher than certain values.

4. Discussion

4.1 Non-modal phonation can be part of the pitch scale

This chapter looks into Mandarin, a case where non-modal phonation is an allophonic cue in tonal contrasts. Three corpus studies are presented here. The first study demonstrates that the presence of creak in Mandarin is not exclusively limited to Tone 3, but can possibly apply to all low targets in Mandarin tones, e.g. Tone 3 and Tone 4, and even Tone 2, when it happens to go low; moreover, creak occurs at a similar pitch value for both tones.

In the second study, we validate and extend this claim in two ways: First, we demonstrate that manipulating pitch ranges can affect the voice quality in producing Mandarin tones. For example, low targets in Tone 3 are less creaky when pitch range is raised, but can be creakier when pitch range is lowered. This further confirms that vocal fry in Tone 3 is driven by very low pitch targets. Moreover, this conclusion can be extended to high targets as well. For example, we have shown that Tone 1 is breathier when pitch range is lowered, and is tenser when pitch range is raised.

Finally, an integrated study on co-varying relationship between voice quality and pitch is also presented in this chapter. Voice quality co-varies with pitch in a wedge-shaped way, with breathiest voice quality in the mid range, and creakier and tenser voice quality as pitch moves lower or higher. Moreover, non-modal phonation such as creak is likely to happen when F0 exceeds certain points. For example, we found that 170 Hz (female) and 110 Hz (male) are the critical points for changing voice quality, corresponding to Figure 4-3. In sum, these corpus

studies show that extreme pitch targets in Mandarin tonal contrasts can introduce non-modal phonations; or in other words, non-modal phonations can be part of the pitch scale.

The presence of non-modal phonation in Mandarin makes a contrast with the case of Yi languages. As discussed in the previous chapter, in Yi languages phonation production and pitch production are independent and no correlations are found between voice quality and pitch.

4.2 Contributions of creak: Make some contrasts easier to recognize

Although non-modal phonation in Mandarin is a redundant cue, and by hypothesis it has a purely physiological basis, it still appears to be perceptually salient. As reviewed in the introduction, this kind of non-modal phonation is especially useful for distinguishing the extreme tones from non-extreme tones, e.g. Tone 3 vs. Tone 2 in Mandarin, and Tone 4 (21) and Tone 6 (22) in Cantonese. Therefore, it is a very useful enhancement cue.

However, is worth noting that, as an optional cue, it functions differently from phonemic non-modal phonation. For example, Garellek *et al.* (2013) found an interesting distinction between two kinds of phonation types in White Hmong: creaky phonation is not very useful in distinguishing the low level tone (22) from the low falling (21 with creak at the end); whereas breathy phonation is the primary cue for distinguishing mid/high falling (breathy voice) from high falling (modal voice). In fact, even in cases such as Mandarin, creak does not always help as well. For example, Gårding et al. (Gårding *et al.*, 1986) used resynthesized stimuli in a two-choice identification experiment between Tone 3 and Tone 4, and creak did not play any role in

distinguishing Tone 3 from Tone 4. Yang (2011) also had the same result for Tone 3 and Tone 4, but recall that this study found that creak is important for distinguishing Tone 3 from Tone 2. Therefore, it is quite possible that listeners do not necessarily pay attention to creak unless pitch cues are not sufficient in certain contrast contexts.

4.3 Non-modal phonation for high pitch targets

Although Mandarin is well-known for non-modal phonation in its lowest tone, we show in this chapter that non-modal phonation can also be natural in the highest tone when the pitch range is raised. But in normal speech, Tone 1 usually does not have very high pitch values. Mandarin has no need to produce a super-high tone, as the low tone is already distinctive enough via its contour as well as via vocal fry, and also there is no mid-level tone in this language.

The missing case here then, is non-modal phonation in a super-high tone. Falsetto is naturally used for singing high pitch, but it has been rarely reported in languages. Rose (1997) observed such a case. The Pakphanang dialect of Southern Thai (PPhN for short) has a super-high tone, which is often produced with audible falsetto. The highest F0 of this tone is up to 285 Hz (Rose, 1997: Table 2) for a female speaker. Peng and Zhu (2010) also reported a super-high tone with falsetto in some Chinese dialects. The highest pitch for a male speaker is around 250 Hz. Despite these observations, quantified acoustic measures other than F0 for these tones have not been attempted. To produce one's highest pitch, one might use only a tense/pressed voice, as most untrained singers do in singing, rather than falsetto. It is not yet clear whether the voice quality associated with super-high tones in languages has the same phonation property as in singing.

Nonetheless, these cases are still of interest because of their super-high F0s, which are beyond the range of normal speech in most languages. Moreover, these languages all have multiple level tones. It is reasonable to assume that the presence of falsetto is related to super-high pitch targets, just as vocal fry is related to super-low pitch targets. It is quite possible that this voice quality provides the cue to the highest pitch, when the tonal contrasts in the mid-range are very crowded. For example, PPhN has seven tones, among which two are mid levels tones, and one a low falling tone. In this case, falsetto (or possibly tense voice) would be very helpful to produce a distinctive high tone.

In sum, this chapter demonstrates a case where phonation and pitch can be closely correlated in tonal productions; this kind of non-modal phonation is quite different from the one we have discussed in Yi languages. In next chapter, we will show that how these two kinds of non-modal phonation types can work together to define a complex tonal space.

Chapter 4 Black Miao: A case study of a mixed system

1. Introduction

How many contrasting pitch levels can tonal languages have? Linguists (Chao, 1948; Maddieson, 1978) have observed that while people are able to produce many phonetically different levels of pitch in speech, no known language makes a phonological contrast of more than five pitch levels. Five-level phonetic transcriptions of tones, e.g. Chao's numbers and the IPA, have been found to be sufficient for known tonal languages. In fact, typologically the number of contrastive levels is even more restricted. According to Maddieson (1978)'s cross-linguistic survey, five- and four-level tonal languages are extremely rare, compared to languages with fewer contrastive levels. A two-level contrast is the most frequently attested type among tonal languages.

Why is the number of possible contrastive levels so limited? Why do languages generally prefer fewer contrastive tonal levels? Dispersion Theory (Lindblom, 1986; 1990) and similar views (Martinet, 1952; 1955; Lindblom and Maddieson, 1988; IPA, 1999; Flemming, 2004) can shed light on these questions. The basic idea of Dispersion Theory is that the structure of inventories is subject to two goals: maximize auditory contrasts but minimize articulatory effort. So an optimal inventory space is the counterbalance between the constraints of speakers and listeners. In other words, the form of the tonal space should follow from its function of maximizing phoneme contrasts while requiring minimal articulatory effort.

Considering the limitations of pitch production and perception, we can understand why multi-level tone systems are not preferable – to maintain multiple level-pitch contrasts is extremely hard for people. On the one hand, the pitch range used in normal speech is fairly small, said to be usually no more than 100 Hz (Baken and Orlikoff, 2000; Keating and Kuo, 2012, for isolated words and passages). Here we can validate this pitch range estimate by doing a quick survey in the cross-linguistic corpus from the UCLA voice quality project. Multi-speaker recordings and a large range of voice measures (Keating *et al.*, 2012) are available online (http://www.phonetics.ucla.edu/voiceproject/voice.html). F0 has been calculated using the STRAIGHT algorithm (Kawahara *et al.*, 1999). The mean F0 value of each token produced by each male speaker from nine languages was retrieved. The boxplot in Figure 4-1 displays the overall distributions of these F0 values. As can be seen here, across languages, the overall pitch ranges (the range between the two whiskers) for male speakers are fairly similar, mostly around a 100 Hz range, lying between 80 Hz and 180 Hz; medians are all around 140 Hz. This means that the physiological limits of pitch-range production are quite universal across languages. A 100 Hz range is about the pitch range for their modal (i.e. with the least articulatory effort) vocal register (Hollien and Michel, 1968; Hollien, 1974; Titze, 1988; Baken and Orlikoff, 2000). Although some tonal languages tend to have slightly larger ranges (e.g. Bo and Mazatec, compared to English), the differences are small, suggesting that the pitch range is rather physiologically restricted, at least at default physiological settings. (Of course, pitch ranges can be enlarged by extra vocal effort or stronger glottal airflow, for example for exclamatory utterances.)

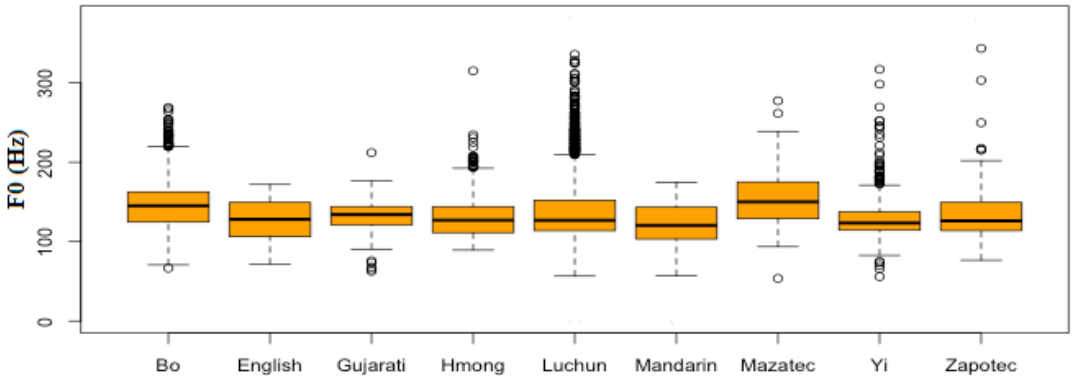


Figure 4-1 Speech pitch range of male speakers across languages. The measure "strF0_mean" for all tokens in the corpus is plotted here. For each language, the plot indicates: the median (the horizontal line in the box), the highest 25% of the datapoints (the upper whisker), the lowest 25% of the datapoints (the lower whisker), and 50% of the datapoints (within the box between the upper and lower quartiles); outlier datapoints are shown as circles.

On the other hand, the just-noticeable-difference (JND) for lexical tone appears to be not less than 9 Hz (indicated by Silverman, 2003: Figure 17.3), and languages usually require a much larger difference than the JND to maintain a phonological contrast; a 20-30 Hz (around 2 semitones in the speech pitch range) difference is just marginally enough. Furthermore, Harris and Umeda (Harris and Umeda, 1987) found that JNDs are much greater in natural sentences. This is so even though the pitch JND for pure tones is fairly small, about 3 Hz for frequencies below 500 Hz (Kollmeier *et al.*, 2008), meaning that speech pitch discrimination is much harder than non-speech pitch discrimination.

For example, Cantonese, a Yue dialect of Chinese, has four level tones (i.e. 11, 22, 33, 55 in Chao's representation). A perception experiment (Mok and Wong, 2010) shows that tones 22 and 33 are the most confusable in perception, though they still have a pitch difference of 30 Hz. Their language survey shows that young speakers frequently merge these two tones. Therefore, three levels appear to be the maximum contrasts that listeners are able to perceive within a pitch range of 100 Hz. Four or five levels of pitch contrasts violate both dispersion principles: the distinction is very hard to produce and extremely hard to perceive.

How then is a five-level-tone system possible, since several such languages have been reported (Edmondson and Gregerson, 1992)? According to Dispersion Theory, to maintain the maximum contrasts, one possible consequence of increasing the size of inventories is that the overall acoustic space of the inventories would be enlarged. This is true for vowel spaces. Cross-linguistic studies on the dispersion of vowels (Lindblom, 1986; Becker-Kristal, 2010) among many others) demonstrate that the size of acoustic space is positively correlated with the size of vowel inventories. That is, languages with more vowels occupy a larger acoustic space than languages with fewer vowels.

Very few studies have addressed tonal dispersion, but Maddieson (1978: p. 339), observes a similar enlarging effect on tonal spaces: languages with more tonal levels tend to employ a relatively larger pitch range than languages with fewer tonal levels, as reproduced below (Table 4-1). For example, Yoruba, a three-level language, has an overall pitch range of 79 Hz, while Toura, a four-level language, has an overall pitch range of 90 Hz.

Table 4-1 Pitch intervals between tones in different languages (Maddieson, 1978), combining various sources, averaged across gender.

	Two levels		Three Levels			Four levels
	Siswati	Kiowa	Yoruba	Thai	Taiwanese	Toura
						50
			52	28	32	30
	18	22	27	16	18	10
Lowest tone	0	0	0	0	0	0

However, a recent quantitative cross-linguistic investigation (Alexander, 2010) claims that tonal space size is rather fixed across level-tone languages, and that size of inventory has little effect on size of pitch range. For example, the pitch range of Cantonese (four-level) is not significantly different from that of Yoruba (three-level) and Igbo (two-level) at mid-point and offset positions of the tones. However, as these pitch ranges were calculated across level and contour tones, they actually reflect the overall pitch ranges of the languages instead of the dispersion of just the level tones. So her results do not counter the findings of Maddieson (1978). Moreover, finding fixed overall pitch ranges is consistent with our survey in Figure 5-1. Therefore, it can be inferred from Alexander's study that the ability of speakers to expand the pitch space is rather limited.

The second way to optimize the dispersion of large inventories is to add contrastive dimensions. For example, Lindblom and Maddieson (1988) found that as the size of consonant inventories increases, more and more articulatory dimensions are utilized. Therefore, for tonal contrasts, in addition to linearly enlarging the pitch range, the second possibility is to rely on cues other than pitch. One of these cues that can contribute to tonal contrasts is duration. Duration is a distinctive

cue for vowel length contrasts, e.g. in Cantonese, Thai; and it is also an enhancement cue for the tone 35 vs. tone 213 contrast of Mandarin (Tseng, 1981; Blicher *et al.*, 1990). Pitch contours are the other most common cues for tonal contrasts. In addition to phonemic contour tones in languages (e.g. Chinese dialects, Vietnamese dialects, Thai dialects), optional falling is also commonly found for low tones (Zhu, 2012). Finally, another, less addressed, cue is phonation, which is commonly found in many tonal languages and has been found to be a salient cue in perception. Sometimes phonation functions as an allophonic cue, e.g. creaky voice on the low tone of Mandarin (Belotel-Grenié and Grenié, 1994; Yang, 2011) and Cantonese (Yu and Lam, 2011); other times phonation functions as a phonemic dimension in addition to pitch: Green Mong (Andruski, 2006), White Hmong (Esposito, 2012; Garellek *et al.*, 2012), Southern Yi (Kuang, 2011), and Northern Vietnamese (Brunelle, 2009), to cite just a few.

When all the dimensions that can contribute to tonal contrasts are considered, it becomes non-trivial to model the tonal space in which dispersion can be understood. Previous studies of tonal dispersion have instead used very simple spaces. In comparing the production effort for tones in normal vs. noisy environments, Zhao and Jurafsky (2007; 2009) modeled tonal dispersion as variability of the overall pitch range. Adopting this method, Alexander (2010) also only used a one-dimensional cue, i.e. mean F0, to define tonal spaces. As she noted, this method was not adequate to capture the perceptual separability of tonal contrasts. Tonal space models that allow contour cues significantly improve the separability of the tonal space. For example, in a study comparing Cantonese tone productions by normal-hearing adults, normal-hearing children and cochlear-implanted children, Barry and Blamey (2004) defined the tonal space by F0 onset x F0

offset. This method enables the authors to capture the dynamic factors, such as direction and slope, of the tones. Yu (2011) demonstrated that mean F0 + pitch change, rather than mean F0 alone, can better model the distribution of tonal inventories. Taken together, these studies suggest that a tonal space should incorporate multi-dimensional cues to reflect the actual perceptibility of tonal contrasts, contours as well as levels. However, no tonal space models so far have incorporated cues like duration and phonation. Since previous tonal studies have been limited to pitch, we will test two competing hypothesized models: pitch (+ duration) cues only vs. pitch + phonation (+ duration) cues.

In sum, the question asked in this chapter is, given normal hearing and speaking ability, how can native speakers produce and hear multiple contrasting level tones? We will try to address this question by exploring the tonal production and perception of a language with five level tones, the most contrasting levels to our knowledge (Edmondson and Gregerson, 1992). I will argue that these tonal contrasts are much more than pitch contrasts. When pitch contrasts get crowded, other cues must be involved to enhance the contrasts (e.g. pitch contour, phonation cues), resulting in an expanded tonal space.

2. Black Miao

The five-level-tone language that will be discussed in this paper is a Black Miao dialect, called Qingjiang Miao (Ch'ing Chiang Miao), belonging to the Hmong-Mien or Miao-Yao family. This dialect is spoken at Shidong Kou (Shih-Tung-K'ou), Taijiang (T'ai-Kung) county of Guizhou (Kweichow) province. Figure 4-2 is a map showing the location of this language. This particular

Black Miao dialect was first documented by Fang-Kuei Li in the 1940s, and since then has been the most famous five-level-tone language in tonal studies (e.g. Yip, 2002; among many others). I went to the same village to conduct the experiments reported here.

Figure 4-2 Map showing the location of Taijiang county of Guizhou province, reproduced from the geological study of Guo et al. (2005). The triangle indicates the location of Taijiang.

According to Li's transcription, there are eight tones (I-VIII) in this dialect; five of them are level tones, two rising and one falling (using Chao's tonal representation), as shown in Table 4-2. Tone VIII (11), tone IV (22), tone VI (33), tone I (44), and tone III (55) are the five levels.

Table 4-2 Black Miao tonal system.

I	II	III	IV	V	VI	VII	VIII
44	51	55	22	45	33	13	11

This chapter is organized as follows: We start by examining whether the five level tones are well dispersed in a pitch-based tonal space. A follow-up perception experiment is then presented to reveal the dispersion of tonal categories in listeners' perceptual space. Finally, the dispersion of the five level tones is further examined in a tonal space incorporating phonation cues. Tonal spaces with different dimensionalities will be compared.

3. A pitch-based tonal space

3.1 Production recordings

A wordlist of minimal monosyllabic sets for the eight tones was created based on Li's transcriptions, which were partially reported in Kwan (1966) and Chang (1947). A list of 23 minimal sets of words was compiled from these sources. These words were first elicited from a fifty-year-old male speaker, who had the best knowledge about Black Miao; ten complete minimal sets were confirmed with him. These words were then checked and rehearsed with every participant in the production experiment, and some additional words not in complete sets were identified. Each speaker confirmed from 100 to 120 monosyllabic words.

The wordlist was then elicited in minimal sets by the experimenter with each speaker; speakers were instructed to skip the items that they could not recognize and to note any items that in their judgment had identical pronunciations. To avoid tone sandhi in continuous speech, the test monosyllabic words were spoken in isolation. Some monosyllables are morphemes that do not normally occur by themselves, but speakers can say them if instructed to. Elicitation was carried out in Southwestern Mandarin, as all the speakers were able to understand and speak this local Mandarin dialect of Guizhou Province. Simultaneous electroglottographic (EGG) and audio signals were recorded in a quiet room, directly to a computer via its sound card, in stereo using Audacity. The sampling rate per channel was 22050 Hz. The audio signal, from a Shure SM10A head-mounted microphone, was the first channel of the recordings. The EGG signal, from a two-channel Glottal Enterprises Electroglottograph, model EG2, was the second channel. Each token was produced as many times as needed until two good repetitions were obtained.

A total of 14 native speakers of Black Miao were recorded. Nine male speakers are native speakers of this particular dialect; five females were also recorded, but they had married into this village and were not native speakers of this dialect. For the purpose of consistency, we therefore only report the data of the male speakers here, but the data of the female speakers are available upon request. One of the male speakers failed to produce the tonal distinctions in most tokens, and thus is also excluded from the current analysis. Therefore, results from eight native male speakers will be presented here.

3.2 Measurements

Pitch values of nine time intervals were automatically obtained by VoiceSauce (Shue *et al.*, 2011), using the STRAIGHT algorithm (Kawahara *et al.*, 1999). For all tokens, pitch was measured across the complete pitch-carrying portion of the rime, as segmented in Praat textgrid files. In cases where the pitch tracking failed, the circumstances were noted (such as glottalization or breathiness), and these tokens were manually double checked in Praat (Boersma and Weenink, 2012). In general, STRAIGHT was successful in extracting correct values for most tokens, even for many vocal fry tokens. Mean F0 values were calculated over nine time sub-intervals, and three contour-related pitch measures were made: onset (first ninth), offset (last ninth) and ΔF0 (the range of pitch change within a syllable). In addition, rime duration was obtained from the Praat textgrids.

3.3 Results – pitch analysis

A series of pairwise mixed-effect models were used to decide which measures significantly distinguish one tone from another, with mean F0, ΔF0, F0 onset, F0 offset and duration as the dependent variables. Duration does not contribute to any tonal contrasts, so time-normalized F0 is appropriate. Mean F0 is significantly different between every pair of eight tones. ΔF0 can distinguish contour tones (i.e. T51, T13, T45) from level tones (i.e. T22, T33, T44, T55), and ΔF0 of T11 is also slightly different from T22. F0 onset and offset mostly help distinguish contours with different directions. Therefore, the only pitch cue needed for the level tones is the average pitch value. Figure 4-3 shows the average pitch trajectories of the five level tones for

eight male speakers. F0 is presented in Hertz in order to show the actual physical pitch space of these tonal categories.

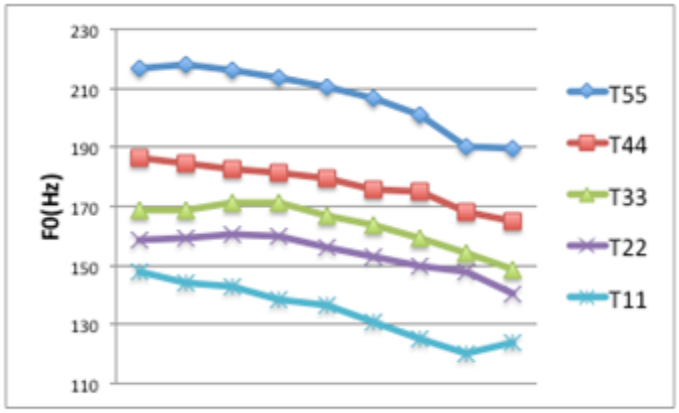


Figure 4-3 Pitch trajectories of five level tones for eight male speakers (time normalized).

Since mean F0 reaches significance for all level tones, Li's transcription based on a five-level pitch scale is correct. However, as shown in Figure 4-3, the five-level-tone space is very crowded, especially for the mid-range tones. Although the pitch difference between Tone 22 and 33 reaches statistical significance, their difference is less than 10 Hz, just about the JND. Likewise, the difference between 33 and 44 is only 20 Hz, which is also a very small difference. These small pitch differences make us wonder whether these tones are able to contrast in the tonal space. To see how these level tones are distributed in a physical tonal space, the distribution of the tonal categories is visualized by Multi-Dimensional Scaling (MDS) in a low dimensional space. Typically, distances among contrasting categories (e.g. vowels) are calculated as Euclidean distances in a low-dimensional physical space (e.g. a 2-dimensional vowel-formant space). In contrast, with MDS the physical space can be based on a large number of phonetic

measures. Each measure represents one dimension, and the values of the measure are the coordinates of the dimension. Here, each token is represented by five measures: mean F0, ΔF0, F0 onset, F0 offset, and duration. The MDS function is able to calculate the distances among tokens in a high-dimensional space (here, five-dimensional), and project the distances into a lower-dimensional and interpretable space.

The MDS solution was obtained by Kruskal's Non-metric Multidimensional Scaling algorithm (*isoMDS* function in R), and physical distances among the five level tones were calculated with the five measures mean F0, ΔF0, F0 onset, F0 offset, and duration (all scaled) as the coordinates. A 2-D solution accounting for 84% of the variance (with stress value of 0.02) is presented as Figure 4-4.

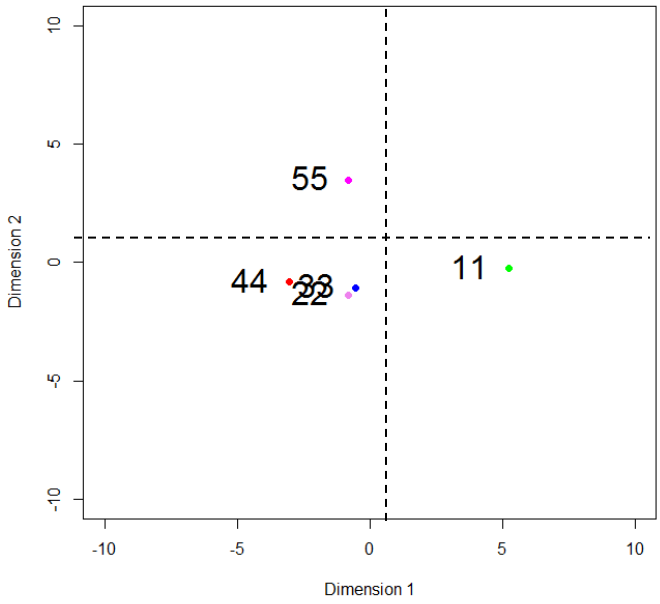


Figure 4-4 Tonal space derived by MDS from pitch and duration measures, level tones only. This is a physical space showing acoustic differences. The dashed lines are added for visual convenience and are not part of the MDS solution.

The intuitive interpretation of the plot is that the more distant the tokens are in the space, arguably the more contrastive the tonal categories are. Although we include all pitch and duration cues in the model, consistent with the results from mixed-effect models, mean F0 is the primary cue that is responsible for the dispersion of the tonal categories. As indicated in Figure 4-4, the tonal space seems to be structured by two pitch-height features (i.e. high and low), and tones cluster into three pitch ranges. Dimension 1 distinguishes low tone (T11) from non-low tones, while Dimension 2 separates high tone (T55) from non-high tones; mid tones (T22, T33 and T44) cluster together in a non-high and non-low range. T22 and T33 are not distinctive in

the tonal space at all, and they are only marginally contrastive with T44. This is not surprising given the pitch differences in Figure 4-3: recall that the difference between T22 and T33 is less than 10 Hz, and T44 and T33 have only a 20 Hz difference. If this tonal space is accurate, native listeners should not be able to hear those contrasts reliably. In general, this space suggests that mean F0s can distinguish only three levels of tonal contrasts.

4. Perceptual space of tonal contrasts

The goal of the perception experiment is to determine whether native listeners are able to hear all the tonal contrasts in Black Miao, and to examine how these tones are distributed in a perceptual space.

4.1 Methods

4.1.1 Stimuli

The stimuli were a minimal set of eight real monosyllabic words with syllable /pa/: /pa44/ “send”, /pa51/ “drop”, /pa55/ “(water) full”, /pa22/ “net”, /pa45/ “pig”, /pa33/ “fail”, /pa13/ “father”, and /pa11/ “drive away (duck)”. These words were chosen because they are frequently used in Black Miao people’s daily life, and were found to be the set that was most easily recognized and produced by native speakers during the elicitation. This choice could partially overcome the possible influence of relative lexical frequency of the eight words (in the absence of an exhaustive source for an accurate estimation of lexical frequency for this language). Potential lexical frequency bias was further eliminated by a familiarity phase in the experiment (see next section).

A male native speaker produced all of these words in isolation; one token of each word was used in this experiment. This male speaker had a good education background and used to be a Black Miao language teacher. All participants in our study were personally familiar with him, which made his voice comfortable to them. He also recorded the experimental instructions in Black Miao. In the instructions, these monosyllabic targets were explained in Black Miao and used in appropriate contexts so that the subjects would unambiguously understand these words. For example, they would hear "/pa51/, as in 'I dropped my money'" (in Black Miao). The time-normalized F0 tracks of the stimuli are shown in Figure 4-5. Similar to Figure 4-3, the speaker's pitch range for the mid-level tones is only around 30 Hz, and T33 is barely distinctive from either T22 or T44. T11 is also very close to T22. Therefore, these particular tokens produced by the speaker are representative of the community's productions as seen in Figure 4-3.

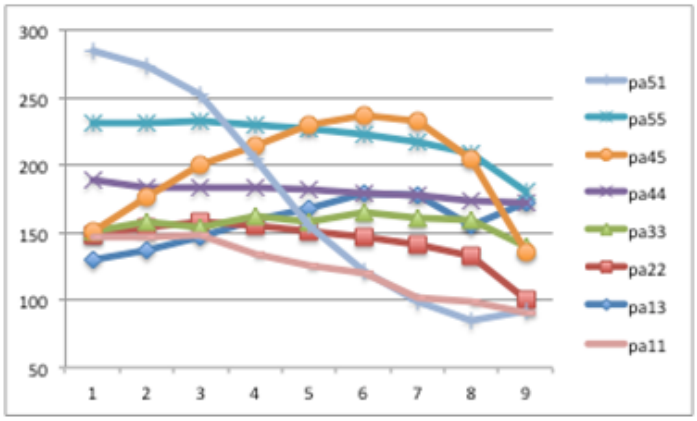


Figure 4-5 F0 values of the stimuli.

4.1.2 Procedures

The experiment was run by a Matlab script on a laptop in a quiet room, and audio stimuli were played through SONY MDR-NC60 noise-attenuating headphones. The experiment had two phases. The first phase was a familiarity phase, during which subjects were asked to listen to an audio introduction in which they were presented with the 8 test words and told that they would hear two of them in each trial. This was to create the same expectation for all the words and thus overcome any prior bias about the test words. The instructions could be heard as many times as needed until a listener fully understood and memorized the words that would be presented in the following test. When they were ready, they were asked to produce the eight words by themselves first, repeating each word twice. This was to make sure these words were fully accessible for them.

The second phase was an AX discrimination task. In each trial, two audio stimuli were presented, and two possible responses, “different” and “same” in Chinese, were displayed on the screen. Subjects were asked to judge whether the sounds they had just heard were same or different words. The stimuli were all possible pairs among the eight stimuli in a random order. Thus in the “same” trials a single stimulus was played twice. The task was repeated three times.

4.1.3 Subjects

A total of 18 subjects, eight males and ten females, with self-report of no hearing disabilities, were recruited from Shidong village. Listeners ranged in age from 25 to 55 with an average of 34 years. Most of them had been to local elementary school, so they were able to understand

Chinese and were comfortable with reading Chinese characters on the computer screen. There were four females, not native speakers of this particular Black Miao dialect, who were excluded from the current analysis, leaving 14 subjects.

4.2 Results and discussion

Table 4-3 is the summary dissimilarity matrix of all eight tones from the discrimination task. Dissimilarity is calculated from the percentage of "different" responses to the tone pairs across all listeners. Responses for a stimulus pair in the two possible orders (e.g. T33 vs. T55, and T55 vs. T33) are averaged. In Table 5-3, the dissimilarity for the "different" pairs is close to 1, while for the "same" pairs it is close to 0. Thus in general, listeners showed high accuracy rates among all tonal pairs, including the "same" pairs, indicating that listeners are able to hear the tonal distinctions in the "different" pairs without any strong bias to answer "different" to all pairs. Among the level tones, the most confusable pair is T22 vs. T44, with 70% of the responses correct for that pair. Surprisingly, T33 is not confusable with either T44 (98% correct) or T22 (95% correct), even though Figure 5-3 shows that they barely differ in pitch. Similarly, the other two pairs of adjacent tones, i.e. T11 vs. T22 and T44 vs. T55, are also nearly perfectly discriminated (93% and 92% respectively).

Table 4-3 Dissimilarity matrix for all listeners. (1.00= perfectly discriminated; 0.00=not at all; therefore, we expect 0 for the same pairs, and 1 for the different pairs)

	T11	T13	T22	T33	T44	T45	T51	T55
T11	0.05							
T13	0.94	0						
T22	0.93	0.88	0.03					
T33	0.97	0.78	0.95	0.05				
T44	0.98	1	0.7	0.98	0.03			
T45	0.94	1	1	1	1	0		
T51	0.94	1	1	1	1	1	0	
T55	0.95	1	1	1	0.92	0.88	0.88	0

Another Multidimensional-Scaling (MDS) solution was founded to map the confusability of the five level tones, now in a "perceptual space", with distances calculated from the perceptual dissimilarity (Shepard, 1972; Kreiman *et al.*, 1990). The more distinctively the listeners perceive them, the more distant the tokens are in the space. A 2-D solution which accounts for 61.6% of the variance (with stress value of 0.009) is presented in Figure 4-6.

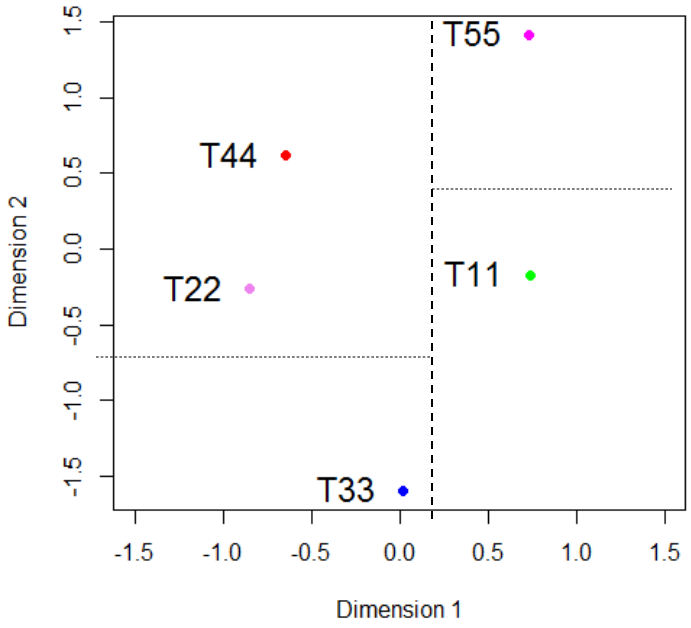


Figure 4-6 Perceptual space of Black Miao five level tones derived by MDS from discrimination responses. The dashed lines are added for visual convenience and are not part of the MDS solution.

As seen in Figure 4-6, the five level tones are well distinguished in native listeners' minds. Dimension 1 divides the tonal space into a mid-range (T22, T33 and T44) vs. extremes (T11 and T55). Dimension 2 mostly follows the low-to-high pitch scale, as seen for T22 vs. T44 and T11 vs. T55, with the exception of T33. It seems that the mid-range space is further divided into two: Tone 33 occupies its own space, which is very distinct from the space occupied by T22 and T44.

Ideally, this perceptibility of tonal categories should be reflected in the corresponding production space. However, comparing Figure 4-6 to Figure 4-4, we can see that the two spaces are very different. Only a very weak correlation ($r=0.17$, $P<0.05$) is found between the production distance-matrix and perceptual distance-matrix. Therefore, the pitch-and-duration-based model fails to reflect the actual distinctiveness of the well-separated five-level-tone contrasts.

In sum, the results from the perception experiment suggest that the tonal categories are well dispersed in a perceptual space; however, the relative perceptibility of the five level tones is different. Tones with extreme pitch values are well distinguished from the mid-range tones; the mid-range space is further divided into tone T33 vs. tone T22 and T44. Even though pitch-wise T33 is very close to both T22 and T44 (Figure 4-6), T33 is not confusable with them; and T22 and T44, the tonal pair with the larger pitch difference, are actually the most confusable. Since there is no significant contribution of pitch contours or duration, it is very mysterious why and how native listeners are able to hear these contrasts. Thus a more sophisticated tonal space model, which incorporates phonation cues, will be tested.

5. A pitch-phonation tonal space

Failing to reflect the perceptual distinctiveness for native listeners indicates that a tonal space modeled only on pitch (+ duration) cues is not sufficient to account for the Black Miao five-level-tone contrasts; therefore, in this section, we will evaluate the competing hypothesis: both phonation and pitch contribute to tonal dispersion.

5.1 Measurements

In additional to pitch measures, comprehensive acoustic measures reflecting different phonation properties were also included: H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, H1*, H2*, H4*, and CPP. The idea here is to include all possible acoustic information without bias from any specific preconception of phonation and tone. Three EGG measures were extracted: Contact Quotient (CQ), Peak Increase in Contact (PIC) and Speed Quotient (SQ). CQ was calculated using the "hybrid" method (Howard *et al.*, 1990): using the positive peak of dEGG to define closing events, and a 3/7 threshold to define opening events.

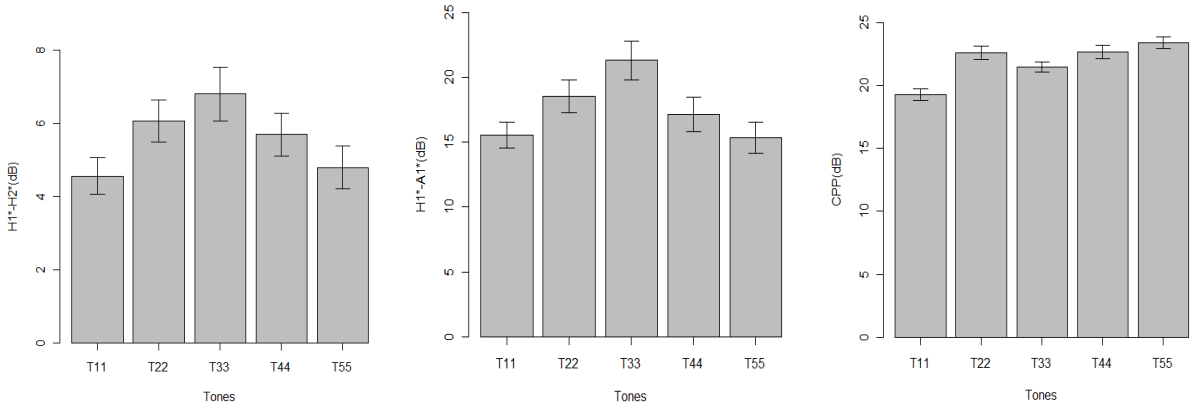
5.2 Phonation cues in five level tones

5.2.1 Acoustic phonation cues

A classification regression tree (*rpart* function in R) is first run to determine which measures are the most important to the tonal contrasts, with all pitch, duration, and voice measures as the predictors. The results show that the most important acoustic cues for classifying the five level tones are (mean) F0 ($p<0.001$), H1*-H2* ($p<0.001$), H1*-A1* ($p=0.002$) and CPP ($p<0.001$). These cues together can correctly classify 70% of the data. Two spectral measures (H1*-H2* and H1*-A1*) that are related to open quotient of the vocal folds, and one measure (CPP) that reflects noise ratio and periodicity, are included in this set of significant cues. A series of pairwise mixed-effect models are used to decide the tonal effects on the three voice measures, with tonal categories as the fixed factor and speaker as the random factor. For H1*-H2* and H1*-A1*, significance is found for all tonal pairs, except for T22 and T44; as for CPP, there is no significant difference among T22, T44 and T55, but these three tones are all significantly

different from T11 and T33. Figure 4-7 (a-c) shows the values of these three measures for the five tonal categories.

The patterns of H1*-H2* and H1*-A1* are very consistent. They both show that T33 is breathier than any other tones, as it has significantly greater H1*-H2* and H1*-A1* (Figure 5-7a, 5-7b). On the other hand, T11 and T55 are more constricted/laryngealized than any other tones, as they have significantly smaller H1*-H2* and H1*-A1*. T22 and T44 have a similar voice quality, which is in between the breathier T33 and the more laryngealized tones (i.e. 11 and 55). These two figures suggest a three-way phonation distinction among the five level tones. Figure (7c) suggests some different information about these different phonations. CPP groups tones T11 and T33 together vs. the others. As discussed before, CPP reflects the periodicity and harmonic-to-noise ratio of phonation, so this means that T11 and T33 are less periodic than the other tones. It is likely that lower CPP for T33 is caused by higher breathy noise in the spectrum, and lower CPP for T11 is caused by irregularity of vibration. Therefore, although T55 and T11 are both laryngealized, T55 is much more periodic than T11, suggesting that they are actually not the same type of phonation, with T11 being more creaky.



(a) H1*-H2* **(b) H1*-A1*** **(c) CPP**

Figure 4-7 Acoustic measures related to phonation contrast, by tone.

5.2.2 Physiological mechanisms

We now try to understand the physiological mechanisms involved in the tonal contrasts, especially the interaction between pitch and phonation. A Principal Component Analysis (PCA) biplot is employed to classify the five level tones, and evaluate the roles of phonation and pitch in tonal contrasts. CQ (Contact Quotient), SQ (Speed Quotient), PIC (Peak Increase in Contact) and mean F0 are the variables of the model. The inner interactions among these variables are also visualized by the biplot.

Figure 4-8 presents the first two principal components, which together account for 97.4% of the variance in the data. The biplot indicates the interactions among the variables. There are several kinds of information that can be read from this plot.

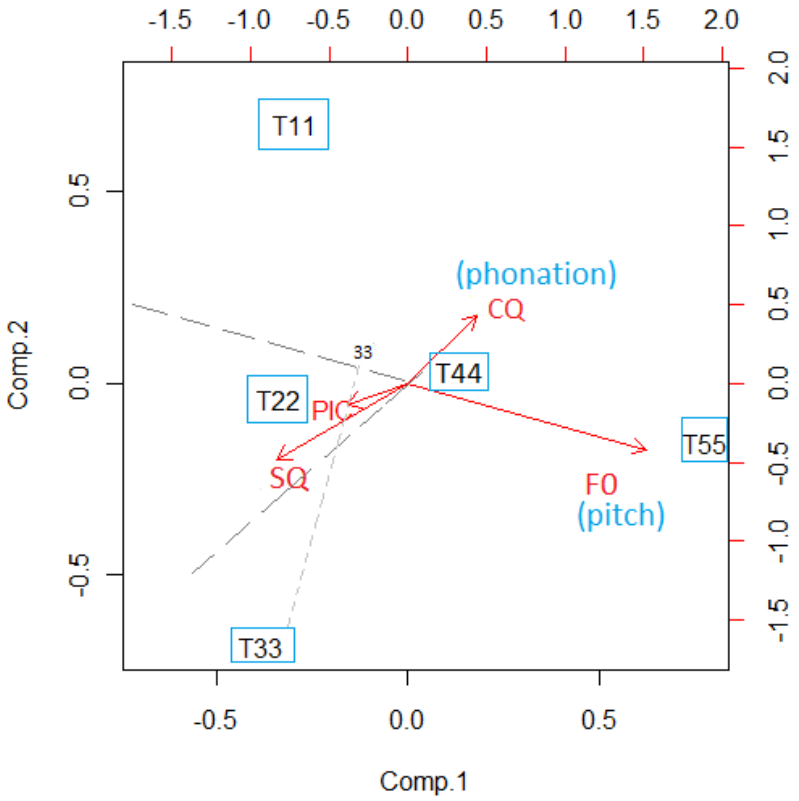


Figure 4-8 PCA biplot from factor analysis of the interaction between pitch and laryngeal parameters. Length of line=strength of this parameter, arrow=direction, angle between the lines=correlation. Tonal categories' positions are determined by the interaction of the parameters. E.g. T33 pitch-wise is located between 22 and 44 (ref. the projection of T33 on the pitch dimension), and phonation-wise is the breathiest among the tonal categories.

First, in this kind of plot, the length of the lines approximates the variances of the variables (direction is indicated by arrow). The longer the line, the higher the variance (i.e. the more important is the cue). Reading this figure, F0 has the highest variance among the variables in the biplot, as it has the longest arrow line; while PIC has the lowest, as it has the shortest arrow line. Intuitively, this means that F0 difference is the most important mechanism for tonal contrasts,

which is not surprising. Among the EGG parameters, SQ, the skewness of the glottal pulse, and CQ, indicating the open quotient of the vocal folds, are the major phonation mechanisms.

Second, the plot also shows the roles of phonation and pitch in these tonal contrasts. Consistent with the results of the initial analysis of the acoustic phonation cues, T11 and T55 have the greatest CQ values, and smallest SQ and PIC, suggesting that these two tones are laryngealized such that the glottal pulses have small open quotients, skewed shape, and slow contact speed. By contrast, T33 has the smallest CQ, and greatest SQ and PIC, suggesting that this tone has a breathy phonation, the glottal pulses having greater open quotient, more symmetric shape and faster contact speed. T22 and T44 have the most similar voice quality, which is intermediate.

Third, the angle between the lines approximates the correlation between the variables they represent. The closer the angle is to 90, or 270 degrees (i.e. lines are perpendicular to each other), the smaller the correlation; whereas an angle of 0 or 180 degrees (i.e. lines are overlapping or in the exactly opposite direction) reflects a strong correlation of 1 or -1, respectively. Therefore, the biplot shows strong correlations among CQ, PIC and SQ, which almost fall on a single line. Meanwhile, F0 has only a weak positive correlation with the phonation parameters, and forms a largely independent dimension by itself. The weak correlation can be explained by the well-established mechanism that the longitudinal tension of the vocal folds increases as pitch increases, which leads to increased CQ values. Indeed, T22, T44 and T55 follow this pattern. T55 is more laryngealized than the tones with lower pitches, suggesting that T55 is produced

with a tense (or stiff) phonation. However, the locations of T11 and T33 cannot be explained by this mechanism. The high CQ of T11 is likely to be due to vocal fry, caused by the compression in the vocal folds, which naturally occurs with low pitches. Thus different phonatory mechanisms can explain the different acoustic properties of T11 and T55. Finally, the distinctive breathy phonation of T33 cannot be explained by the pitch production mechanism, but instead must have to be learned by native speakers, independent of the pitch contrasts.

5.2.3 Pitch-phonation tonal space

With this understanding of phonation cues for the five tones, we now turn to the main question: how do these phonation cues contribute to the dispersion of the tonal space? Incorporating acoustic phonation cues as well as pitch (+ duration) cues, we generate a new MDS tonal space (Figure 4-9), which accounts for 66.2% of the variance in this larger dataset (with stress value < 0.00001). We can see significant improvements from Figure 4-4: First of all, T33 now is well distinguished from T22 and T44, occupying its own quadrant; second, the scale of the space is much bigger than Figure 4-4, which indicates a better dispersion in general. The enhancement of distinctiveness is very important given that tonal contrasts are realized in a very limited pitch range. The new production space now matches better with the perceptual space. A strong correlation (r=0.87) of the perception distance-matrix and production distance-matrix is found. This result indicates that non-modal phonations in Black Miao are very important in production, and by inference, also in perception.

As indicated in Figure 4-9, similar to Figure 4-6, the tonal space is divided into a mid range (T22, T33 and T44) and an extreme range (T11 and T55). The extreme pitch ranges are also related to laryngealization (Figure 4-8). The mid-range space is further divided into two parts, breathy tone T33 vs. modal tones T22 and T44. Tones in the different quadrants benefit from both phonation and pitch cues, whereas tones in the same quadrants are primarily distinguished by pitch cues. The laryngealization in T11 and T55 enhances the difference between tones with extreme pitch values and tones within the mid-range. On the other hand, breathy phonation creates an additional quadrant for the distinction between T33 and the other mid tones.

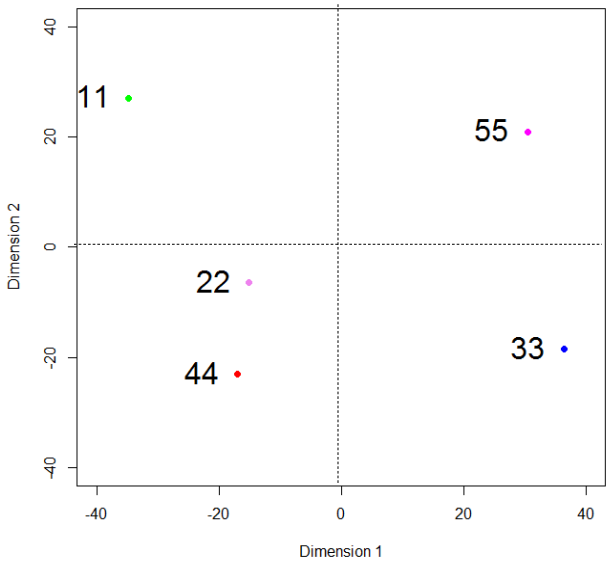


Figure 4-9 MDS tonal space with pitch, duration, and phonation measures, level tones only. Note the scale of Figure 5-9 is much larger than that of Figure 5-4.

4. Discussion – tonal space model

In this study, we conducted both production and perception experiments with Black Miao, to explore how native speakers produce and perceive the contrasting five level tones. We confirmed that pitch is not the only cue in tonal contrasts for this language, and non-modal phonations appear to be very important cues in both tonal production and perception. T55 and T11 can benefit from both pitch cues and phonation cues so that they have very good separability from the mid tones. For the mid-range tones that have very similar pitch cues, T33 is distinctive from T22 and T44 primarily by the phonation cue. T22 vs. T44, the tonal contrast with only a pitch difference, is the hardest to produce and perceive distinctively.

The interaction between pitch and phonation leads to the well-dispersed tonal spaces shown in Figure 4-6 and Figure 4-9. The two tonal spaces share a similar dispersion pattern: Pitch range is first divided into three ranges, i.e. high, mid and low, where the high and low ranges are also cued by laryngealization in the vocal folds. The mid-range space is further divided into breathy vs. modal parts, so that tone 33, the least dispersed tone in a pitch-based tonal space (Figure 4-4), is well distinguished in the new pitch-phonation tonal space. Figure 4-10 generalizes the organization of the tonal spaces from Figure 4-6 and Figure 4-9.

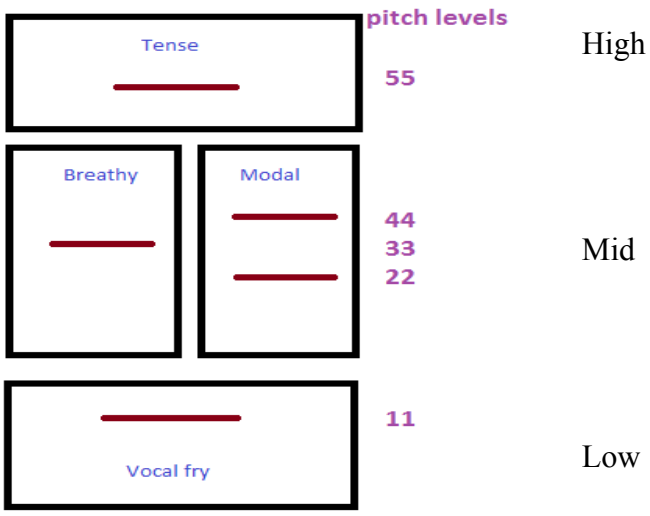


Figure 4-10 Phonation registers of the five contrasting levels: a model of Black Miao tones

In this schema, the five level tones are divided into different quadrants based on different phonations, as in Figure 4-9. The tones with the highest pitch and the lowest pitch form their own quadrants, and the tones with mid-range pitches can be further divided into two quadrants: T33 in the breathy quadrant, and T22 and T44 in the modal quadrant. With these dimensionalities, the burden of pure pitch contrasts reduces to T22 vs. T44 only.

Comparing Figure 4-9 with Figure 4-4, the non-modal phonations contribute to the improvement of tonal distinctiveness in two ways: On one hand, the phonation cues enhance the contrasts for T11 and T55, so that the general distinctiveness of the tone space is enlarged; on the other hand, the breathy phonation creates an independent dimension for T33, so that T33 is very distinct from the other mid-range tones, i.e. T22 and T44.

These two functions reflect the different relationships between pitch and non-modal phonations. The first kind of non-modal phonations are parts of the pitch scale, such as vocal fry, falsetto and tense voice. Vocal fry is coarticulated with the lowest pitch range, and falsetto or tense voice is usually associated with the highest pitch range. Referring to Figure 4-3, the mean F0 of the highest tone is around 220 Hz, which is a remarkably high pitch for male speakers, much higher than the average 175 Hz upper limit of the male speech range across languages (Baken and Orlikoff, 2000). If not doing anything to reduce the longitudinal tension in the vocal folds, then these high pitches must be produced with tense voice (Kong, 2007). This tension results in a greater CQ in EGG signals. Likewise, when pitch goes to the lowest end, e.g. below 75 Hz for males, speakers have to produce these pitches with creaky voice (e.g. vocal fry), which also leads to a greater CQ. This is what we saw in Chapter 3 with Mandarin.

Unlike these pitch-driven non-modal phonations, the second type of non-modal phonation, such as breathy, is relatively independent from pitch. This type of non-modal phonation can create an independent dimension for tonal contrasts, so that tones with similar pitches (T33 vs. T22 and T44) but in different phonation registers are rarely confused. This is similar to what we saw in Chapter 2 with Yi languages.

Why do tonal languages need these two kinds of non-modal phonations? We can account for it from the view of optimizing the dispersion of tonal inventories. When level tone inventories are large, pitch cues are no longer sufficient, requiring too much perceptual and articulatory effort to maintain the crowded contrasts. As discussed earlier, there are two possible ways to optimize the

tonal spaces with large size of inventories: expand the pitch space for tonal contrasts or add an additional contrastive dimension. Indeed, the pitch-driven phonations can help to produce extreme F0 targets, either super high or low, and thus enhance the perceptual pitch distinctiveness for the highest and lowest tones. On the other hand, pitch-independent phonations create an independent dimension for tonal contrasts so that tones with similar pitches can be distinguished from each other. In sum, the well-distinguished five-level-tones of Black Miao can be attributed to both kinds of non-modal phonations. Lindblom and Maddieson (1988) proposed that small inventories can be distinguished on just the "basic" dimensions, while larger inventories have to expand to more complicated dimensions. This principle was proposed on the basis of a study of consonant inventories, and the present study shows that the same principle also holds for tones. Pitch cues are sufficient to distinguish small tonal inventories, but larger tonal inventories require more complicated dimensions.

5. Conclusions

This study investigates the dispersion of multi-level tonal contrasts by exploring the cues used in producing and perceiving five level tones of Black Miao. Both production and perception experiments show that non-modal phonations are very important cues for these tonal contrasts. A model is proposed in which two different kinds of non-modal phonations that either enhance pitch contrasts or provide an additional contrastive cue divide tonal levels into several registers so as to optimize the distinctiveness of the tonal space. Thus Black Miao, with its large inventory of level tones, makes use of both kinds of phonation demonstrated in the two previous chapters: pitch-dependent and pitch-independent.

Chapter 5 General discussion and conclusion

1. Summary of the three case studies

Zsiga (2012: p. 207) in her survey of research on contrastive tone, notes that "the interaction of voice quality and tone (...) is an especially active research area" and asks (Zsiga, 2012: p. 198) "should the definition of 'tone' be revised to include laryngeal contrasts other than pitch?". This dissertation has presented close studies of tonal contrasts from three language families that provide different answers to this question.

First, Chapter Two investigated phonemic phonation contrasts in three Yi languages, Southern Yi, Hani, and Bo. In these languages, the tense/lax phonation ("register") contrast is orthogonal to the tone contrast: low and mid tones (though not high tones) freely combine with tense and lax phonations. Acoustic and electroglottographic studies of the three languages, conducted in the field, showed that phonation production is independent from tonal production. Statistical comparison of various acoustic and EGG measures from different tone/phonation combinations showed that tone and phonation contrasts use different phonetic dimensions. Crucially, the phonation contrast has no effect on F0, the most important phonetic correlate of tone; and the tone contrast has no effect on Contact Quotient (CQ), the most important phonetic correlate of phonation for these languages. In addition, these two key dimensions, CQ and F0, are not correlated. That is, in these languages, tone and phonation are not only phonologically

contrastive, but phonetically completely independent. Tone is purely pitch, and phonation is purely voice quality.

In contrast to these Yi languages, Chapter Three looked into Mandarin, a case where non-modal phonation (specifically, creaky voice) is known to be an allophonic cue to the low-dipping Tone 3. We showed that the presence of creak in Mandarin is not exclusively limited to Tone 3, but can occur with any of the low targets in the Mandarin tones, e.g. Tone 4, and even Tone 2. Specific group threshold F0s were identified for male and female speakers: below 170 Hz for females and 110 Hz for males, speakers are very likely to use creaky voice, regardless of the tone. Moreover, manipulating speakers' overall pitch ranges does not much affect these thresholds, and therefore speakers' use of creaky voice varies as overall pitch range is varied: Tone 3 is less creaky when the overall pitch range is raised, but more creaky when the overall pitch range is lowered. Finally, we showed that in a corpus of pitch sweeps that rose or fell over large pitch ranges, Mandarin speakers' voice quality co-varied with pitch in a wedge-shaped function. Speakers produced their breathiest voice quality in the mid-pitch range, and creakier and tenser voice as pitch moved lower or higher. In sum, voice quality in general, and use of creaky voice in particular, seems to be quite systematically tied to F0 in Mandarin. Voice quality varies across the different tones because they differ in their F0s. While the present study does not go so far as to provide a function predicting all aspects of voice quality from F0 in Mandarin, it seems likely that voice quality is indeed highly predictable from F0 in this language. Mandarin, then, appears very different from the Yi languages. Because of its apparently tight coupling

between pitch and phonation, tone in Mandarin involves laryngeal differences in addition to pitch.

Finally, Chapter Four investigated the case of Black Miao, a language with five-level-tone contrasts. Black Miao has been widely cited as evidence of the maximum use of pitch contrast attested among languages, and therefore is a very important case for understanding the nature of tonal contrasts. Based on considerations of comfortable pitch production and reliable perception, it was hypothesized that these five tones could not contrast purely in pitch: if five tones are squeezed into the pitch range of modal voice (by definition the range of pitches that is comfortable to produce), then the tones' F0 differences must fall too close to the JND for tones for them to be reliably distinctive. Some other phonetic dimensions must be involved: tonal contours, or durational differences, or voice quality. Both production and perception experiments were conducted in the field to explore how native speakers produce and perceive the contrasting five level tones. We confirmed that the tones are indeed similarly level in pitch, that they do not differ in duration, and that pitch is not the only cue in tonal contrasts for this language. Non-modal phonations appear to be very important cues in both tonal production and perception, but in different ways. On the one hand, the three mid-range tones (T22, T33, T44) have very similar pitch cues, but T33 is quite distinct from T22 and T44 by its breathy voice quality. On the other hand, the extra-high (T55) and extra-low (T11) tones show the same kind of pitch-dependent voice quality that was found in Mandarin: the extra-low pitches tend to be creaky while the extra-high pitches tend to be tense. These tones are sufficiently distinct from the other tones in pitch, but the voice quality differences enhance their perceptual distinctiveness. In contrast, T22

and T44, the only tonal contrast based purely on pitch, with no reliable voice quality differences, are the most confusable of the tones.

This system, then, is a mixed one, combining both of the uses of non-modal phonation seen in the previous chapters. As in Yi, non-modal phonation can be pitch-independent and thus phonemic: here, the mid-tone contrast that depends on breathy vs. modal voice. But as in Mandarin, non-modal phonation can also be pitch-dependent: in both languages, the enhancement of extremes of pitch by voice quality.

Based on the 2-D tonal spaces derived for Black Miao, shown in Figure 4-10, a new tonal model was proposed for this mixed-system language. In this schema, the five level tones are divided into different quadrants based on different phonations. The tones with the highest pitch and the lowest pitch form their own quadrants, and the tones with mid-range pitches can be further divided into two quadrants: T33 in the breathy quadrant, and T22 and T44 in the modal quadrant. With these dimensionalities, the burden of pure pitch contrasts reduces to T22 vs. T44 only. This model sheds light on many important issues concerning tonal contrasts, but needs to be extended beyond Black Miao to tone systems more generally The following sections present several proposals that arise from our results.

2. There are two uses of non-modal phonation in tone systems

2.1 Pitch-independent non-modal phonation

As reviewed in Chapter One, it is well-known that languages can contrast different phonations independently of tone. For example, Mpi (Silverman, 1997) combines two phonations with three

level and three contour tones. Jalapa Mazatec combines three phonations with three level tones (plus contours), and Garellek and Keating (2011) showed that F0 does not differ between phonation categories. Esposito (2012) showed that White Hmong (at least, the female speakers) contrast two high-falling tones by breathy vs. modal voice, and Garellek *et al*. (2013) showed that Hmong listeners rely on this phonation difference and ignore F0 differences. This dissertation has presented two further cases in some detail. In three Yi languages, the low-mid tone contrast and tense-lax phonation contrast were shown to be completely independent on various production measures. In Black Miao, the mid tone (T33) was shown to be produced with a breathy voice quality that made it distinct from two other tones with very similar pitches (T22 and T44). Thus the very traditional view of phonation types is further supported by the present study.

Contrastive phonation categories have been presented by Ladefoged (Ladefoged *et al.*, 1978; Gordon and Ladefoged, 2001) in terms of a continuum of glottal aperture, as shown in Figure 5-1. In this kind of model, lax voice lies between breathy and modal, while tense voice lies between modal and creaky[9]. The control of glottal opening, by the interarytenoid and lateral cricoarytenoid muscles, is relatively independent from the control of pitch, by the cricothyroid muscles (Laver, 1980; Gobl and Ní Chasaide, 2012). This allows tones and phonations to combine more or less freely in a language, at least within a comfortable pitch range.

[9] The categories are not meant in absolute terms, instead are relative.

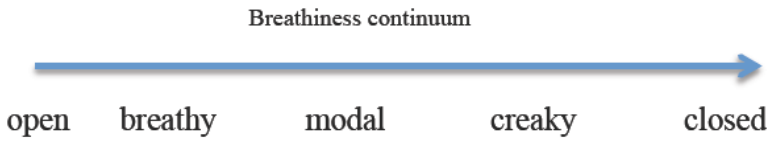


Figure 5-1 Continuum of glottal constriction (after Ladefoged 1971)

2.2. Pitch-dependent non-modal phonation

At the same time, we have shown that phonation can equally well vary with pitch in a seemingly closely coupled way. As shown in the study of pitch sweeps (Chapter 3), voice quality can co-vary with pitch, albeit in a non-linear manner, with the mid-range more breathy and extreme ranges more creaky/tenser. When pitch moves below or above a comfortable (modal) F0 range, non-modal phonation will readily occur. This kind of variability of phonation types has been well-documented in the singing literature (e.g. Titze, 1990; Sundberg, 1994), as indicated in Figure 5-2. As indicated in the figure, very low pitches will be produced with vocal fry, while very high pitches will be produced with tense voice, or even in falsetto.

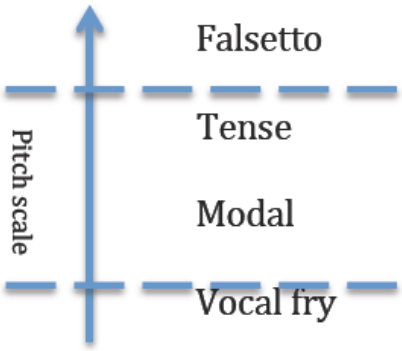


Figure 5-2 Variation of phonation types along the pitch scale

Voice quality varies along this scale because pitch control involves complicated glottal configurations, at least including height of larynx, vocal fold tension, and subglottal pressure (Ohala, 1973). Raising pitch above a comfortable range is mostly controlled by the criocothyroid muscle (CT), which can basically stretch the vocal folds and increase the longitudinal tension, and thus affect the adduction of the vocal folds (Sundberg, 1994). This less relaxed phonation in high pitch is usually called tense voice. In order to reduce vocal fold tension, trained singers are likely to switch to falsetto at this point. Unlike for pitch raising, the CT muscle is not really involved in pitch lowering. Lowering pitch below a comfortable range is mainly realized by the sternohyoid (SH) muscle (Erickson, 1993), which can lower the larynx; but low tones are usually also produced with a compressed larynx (Moisik *et al.*, 2010), which naturally leads to creak.

As shown in our Mandarin and Black Miao case studies, pitch-dependent non-modal phonation of this kind is a useful enhancement cue for extreme pitch values. The presence of creak is a salient cue of "the lowest pitch", and the presence of tense is salient cue of "the highest pitch", which makes the extra low tones (e.g. T213, T11) in Mandarin and Black Miao and the extra high tone (i.e. T55) in Black Miao perceptually more distinctive from mid-range tones (e.g. T35 in Mandarin, T22, T44 in Black Miao).

Note that according to this scale, and our results, there are two kinds of creaky voice (or laryngealization): vocal fry and tense voice. While both can be produced with a constricted glottis[10], they are very different in terms of their natural pitch affiliations. Tense voice is

[10] Utterance-final creak can be produced with a somewhat spread glottis (Slifka, 2006)

correlated with high pitch, while vocal fry is associated with low pitch. This may partially explain why mixed findings, both raising pitch and lowering pitch, have been reported for creaky phonation. As suggested in the Black Miao case, although tense voice and vocal fry both have a greater CQ and smaller H1-H2, they are different in periodicity or regularity of vibration. This confusion among different types of "creaky voice" has happened because of a lack of understanding of the second type of non-modal phonation – pitch-dependent phonation – associated with extremes in pitch ranges.

Several threshold F0 values were identified for male and female Mandarin speakers, based on the data from unprompted pitch sweeps (c.f. Figure 3-16): the naturally breathiest voice quality (peak H1*-H2*) happens in the most comfortable pitch range (at about 200 Hz for females, and 150 Hz for males), and voice quality goes creakier or tenser (both with smaller H1*-H2* values) as F0 goes below or above this point, showing a wedged-shape relationship between F0 and voice quality. The best positive correlation between voice quality and pitch happens in the lower pitch range, 110 – 180 Hz for males, 170 – 200 Hz for females; and the best negative correlation happens in the higher pitch range, 180 – 250 Hz for males, 200 – 270 Hz for females [11]. Finally, 110 Hz for males and 170 Hz for females are the threshold F0 values for the natural occurrence of creaky voice; and there is a clear voice break at 250 Hz for female speakers. These threshold F0 values for voice quality suggest that glottal configurations might require significant changes when passing through these pitch values. These "turning points" can be compared with previous studies on the modal register. Most of the studies on pitch-driven phonation have been done by

[11] The turning points are less reliable for the higher range because fewer data points are available, and the pattern for female speakers is clearer than the pattern for male speakers.

singing scientists. Very few studies have explicitly discussed the thresholds of vocal registers. The comfortable pitch ranges found for speech are much smaller (Baken and Orlikoff, 2000 p. 174, summary from various studies): from 90 (+/-10) Hz to 165 (+/-10) Hz (median = 142 Hz) for male English speakers, and from 160 (+/-10) Hz to 250 (+/-10) Hz (median =201 Hz) for female English speakers (averaging over spontaneous speech and read speech). It is possible that with a much weaker subglottal pressure and less activation for the CT muscle, the comfortable range for speech cannot be as large as the singing range. The thresholds found for female speakers presented here are similar to those in Baken and Orlikoff (2000), but the values for male Mandarin speakers here are higher. In sum, the acoustic evidence suggests that voice quality may change when passing certain pitch thresholds, but these values need further validation by physiological methods in the future.

3. Tonal contrasts are multi-dimensional

Although phonological theories have proposed different representations of tones, most of them have defined tone with one single phonetic dimension -- pitch. With this assumption, tonal contrasts are essentially a "pitch scalar system" (Hyman, 2010; c.f. Chao's number). Because of this monodimensionality, binary-feature representations are always problematic and ambiguous in many cases, especially for mid tones, as pitch levels are relative and continuous (c.f. Mid tone problem discussed in Hyman, 2010). Compared to tones, vowels and consonants have fewer difficulties and confusions with features. Clements (2010) reasons that this is because vowels and consonants are defined at least along two dimensions, e.g. frontness and height, or place of articulation and manner, respectively. A feature system is not quite applicable when there is only

one phonetic correlate. Frustrated by the impossibility of unambiguously defining pitch levels, Clements and Hyman suggested that tonal analysis should abandon tone features completely, and rely only on direct F0 tracks.

However, if tonal theories or tonal models abandon all theoretical explanations of possible contrasts, then we cannot explain why there are typological limitations on possible pitch contrasts, which is one of the purposes of a feature system (Yip, 2002). As we know, no known languages have more than five-level-tone contrasts; and as we have shown in this study, in fact five-level-tone contrasts do not purely rely on pitch contrasts. Clements' and Hyman's discussions of the relativity of pitch contrasts are certainly true, but it does not fully capture the nature of tonal contrasts. Based on the cases in this study, we have argued that tonal contrasts are not really monodimensional, and just like vowels and consonants, many cues can be involved in tonal contrasts, and make tonal contrasts much easier.

As shown in previous chapters, five level tones contrast in Black Miao in a space defined by both phonation and pitch, so the burden of pitch contrasts is not heavy at all; the low target of Mandarin Tone 3 is enhanced by the creaky voice, so the difference between Tone 2 and Tone 3 is more salient; multiple-level pitch contrasts are certainly possible, as long as they are divided in more than one dimensions/registers, such as in Yi languages (and many other Tibeto-Burman languages). Previous studies also documented some redundant cues in tonal contrasts, although they were not addressed as a necessary part of those contrasts. For example, in addition to pitch contour and phonation difference, Tone 3 is also slightly longer than Tone 2 (Tseng, 1981),

which makes these two tones even more distinctive from each other. It should be highlighted that even in Mandarin, where pitch contrasts should be sufficient, tonal contrasts are not purely pitch contrasts.

To push further, we shall argue that even pitch cues are actually not monodimensional, and that they include at least contour shapes as well as pitch levels. Autosegmental representations have treated contours as multiple pitch levels. Contour tones in African languages are quite decomposable (Goldsmith, 1979), but tend to function as single units in Asian languages (Abramson, 1978; Gandour, 1978; Xu, 2004; Roengpitya, 2007). Change of F0 within a syllable is an independent cue from pitch levels. Previous tonal dispersion studies (Gandour and Harshman, 1978; Alexander, 2010; Yu, 2011) have shown that a tonal space (for Thai, Mandarin or Cantonese) that is defined by both change of F0 and pitch levels is better than one that is defined by pitch levels only.

A real monodimensional pitch contrast would be a case with only pitch-level contrasts, but such cases are extremely limited. According to the Black Miao case, as shown in Figure 4-10, a five-pitch-level contrast is certainly not possible, and in fact the pure pitch-level contrast in that language only comprises two levels. We expect that even three pitch levels should require additional cues. Indeed, as shown in the Yi case, where phonologically there are three level tones, the lowest tones are actually falling, transcribed as either 21 or even 31. To further validate this hypothesis, we screened through several other level-tone languages in the UCLA Phonetic Archives (http://archive.phonetics.ucla.edu/archive.htm): Ewe (two levels), Igbo (two

levels), and Yoruba (three levels). Among the three, only Ewe has two perfectly-level contrastive tones; for Yoruba, even for Igbo, the low tones are falling and creaky, just as in Cantonese, Mandarin, Yi and Black Miao. Of course, the examples available in the archive are very few, and one can conduct more sophisticated studies on these and other languages in the future. This is a strong prediction that will require extensive testing in the future.

In sum, tone is not just pitch, and the nature of tonal contrasts is really multidimensional. Adopting the view that Lindblom and Maddieson (1988) espoused for consonant contrasts, we can say that tonal contrasts have different dimensionality depending on the size of the inventory. When the tonal inventory is very small, e.g. two levels, using the basic dimension of pitch levels would be enough. However, due to the physiological limitations on pitch perception and production, if the tonal inventory is larger, e.g. more than two levels, additional dimensions such as contours, duration, and phonation should be involved as well.

4. Sorting out cases from previous studies

This new model can explain cross-linguistically different uses of non-modal phonation in various tonal languages, and shed light on the controversy about phonation effects on pitch.

4.1 Phonation effects on pitch

It has been widely documented that phonation contrasts lead to pitch differences in some languages. For example, breathy phonation is reported to be associated with pitch lowering (c.f. Gordon and Ladefoged, 2001), as breathy phonation usually originates from voiced onsets. However, it needs to be stressed that this lowering effect is not automatic and thus not always

found, as pitch control is independent from phonation control within a comfortable pitch range. For example, F0 does not co-vary with phonation contrasts at all in Yi languages; and the breathy phonation effect on F0 is also very weak for White Hmong (only for males, Esposito, 2012). By contrast, for Gujarati, a non-tonal language, F0 is significantly different in different phonations (Khan, 2012). Esposito and Khan (2012) reason that this is because White Hmong (and Yi) are tonal languages, but Gujarati is non-tonal. As a non-tonal language, Gujarati is free in its F0 variation, so it can lengthen the duration of the glottal opening phase to produce breathy voice; but tonal languages, such as White Hmong and Yi, need to keep phonation contrasts more independent from F0. Therefore, whether non-modal phonation affects F0 depends on a language's strategies for producing a given phonation type, and perhaps on whether the language is tonal or non-tonal.

4.2 Non-modal phonation can be commonly found in two situations

First, we have proposed that non-modal phonation tends to be involved when a language has extra high or low tones. We found a long list of such low tones, here just to list a few: Cantonese (Yu and Lam, 2011), Mandarin (Chapter 3), White Hmong (Esposito, 2012), Black Miao (Chapter 4), Yoruba (from the UCLA Phonetic Archive), and Trique (DiCanio, 2008). No matter how many tones in a language, as long as one tone is extra low (11/21), it tends to have creak or vocal fry. For languages with large pitch spaces, such as Black Miao, the highest tone is also cued by a non-modal phonation. Although very rare, falsetto is reported to co-occur with an extra high tone, e.g. Chinese Yueyang dialect (Peng and Zhu, 2010), Gaoba Kam (Zhu, 2012); Pakphanang Thai (Rose, 1997). All in all, as been discussed, it seems likely that non-modal

phonations in these languages are not the goal, but the side-effects of producing extreme pitch targets.

The second situation for non-modal phonation being involved is when there are minimal pairs or triples which share the same pitch level/contour. The non-modal phonation serves as a contrastive dimension, which is independent from pitch, so that tones with similar pitch values are not confusable. Most of this kind of phonation contrast (at least for Tibeto-Burman, Mon-Khmer, Hmong-Mien families) has a historical origin of onset voicing contrasts or coda types, and the tones with similar pitch values are of the same tonal categories in the proto-languages (Thurgood, 2007). Therefore, phonation contrasts can happen across all tones in a system, e.g. Mpi (Silverman, 1997; Blankenship, 2002), Chinese Wu dialects (Cao and Maddieson, 1992), Eastern Yi (Maddieson and Ladefoged, 1985), Jalapa Mazatec (Silverman, 1997; Garellek and Keating, 2011). Some of the register contrasts have been neutralized, so tone and phonation are partially crossed, e.g. Southern Yi, Bo and Hani (Chapter 3). This kind of phonation and tone crossed system is demonstrated with Southern Yi (Figure 5-3).

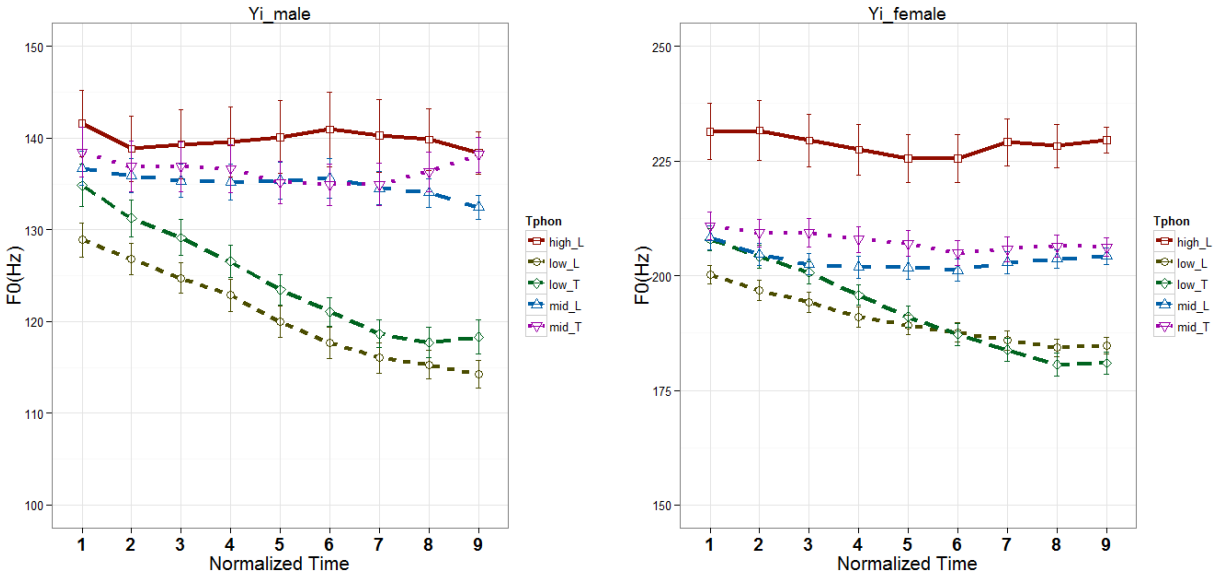


Figure 5-3 Phonation and tone crossed system: Southern Yi.

Pitch-independent phonation sometimes only occurs with one or two tones in languages, e.g. White Hmong high-falling (Esposito, 2012, shown in Figure 5-4), Green Mong mid-falling and low-falling (Andruski and Ratliff, 2000), Black Miao mid-level (Chapter 4), and Vietnamese mid-rising (Brunelle, 2009). Although the development of these non-modal phonations can be traced back to either onset or coda laryngeal properties, there is also a structural reason for the preservation of non-modal phonations: keep the tones with similar pitch levels/contours distinct from each other. As shown in Chapter 4, the breathy phonation is the distinctive cue for Black Miao T33, so that it is not confusable with the other two mid tones. This distinctive function also applies to contour tones. Figure 5-4 shows the tonal system of White Hmong, in which the two high falling tones have similar shape and pitch values, and they are mainly distinguished by

phonation types (Garellek *et al.*, 2013)[12]. Likewise, the presence of syllable-medial laryngealization is the primary cue for the distinction between the two rising tones in Vietnamese (Brunelle, 2009). In addition, languages sometime also involve pitch-dependent phonations as well, e.g. White Hmong and Black Miao, so that phonation and pitch have different weights in different tones (Garellek *et al.*, 2013).

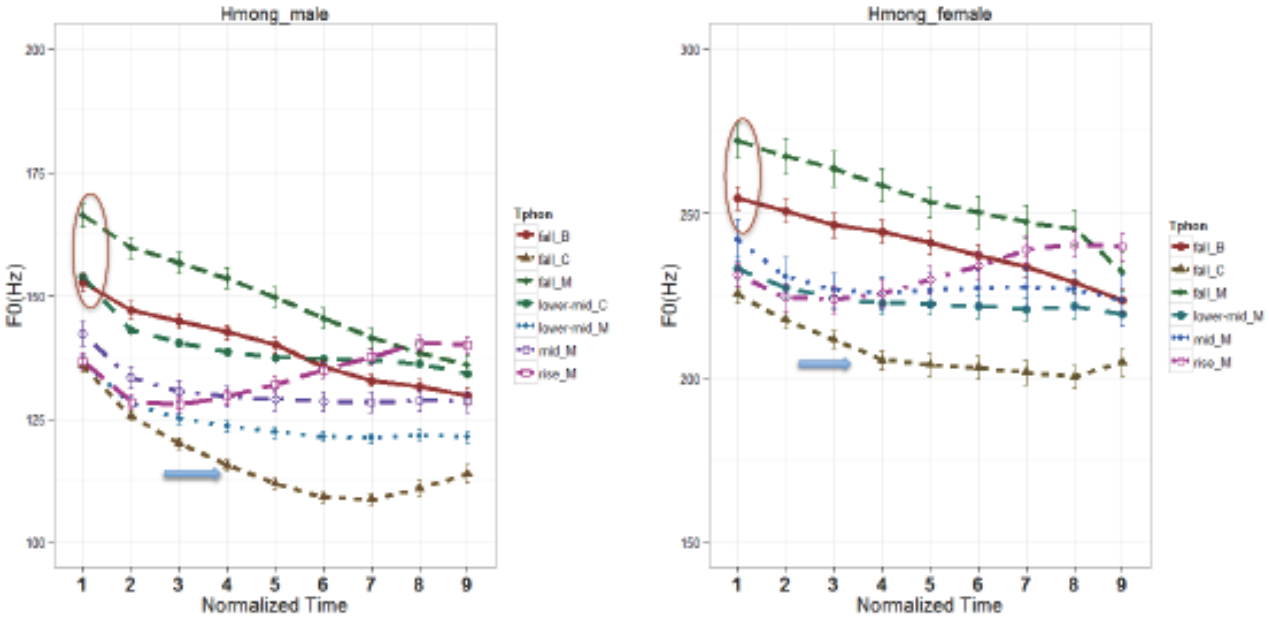


Figure 5-4 The tone by phonation system of White Hmong: mixed system. Breathy vs. modal is the distinctive cue for the two high-falling tones (circled tones), but creaky phonation is allophonic for the low falling tone (pointed by a blue arrow) (c.f. Garellek et al. 2013). Data from http://www.phonetics.ucla.edu/voiceproject/voice.html

In sum, the distinction between these two types of non-modal phonation can help us understand why and when non-modal phonation can happen in tonal languages; and to what extent the

[12] Similar to Black Miao, the two modal mid tones in this languages may also be undergoing merger.

variability of phonation types is language specific (e.g. degree of breathiness for phonation contrasts), or universal (e.g. lowest tones are naturally creaky).

5. An integrative proposal about preferred tonal spaces

One of the major goals of this dissertation is to reexamine the nature of tonal contrasts. We have shown that tonal contrasts are multidimensional, and the space is defined at least by both phonation and pitch. Because tonal contrasts are not limited to pitch contrasts, the question of maximum possible number of contrastive pitch levels, which used to be one of the central issues for tonal studies, is less important; instead, the question of how to make good tonal contrasts is more of interest. Based on the interactions between phonation and pitch that have been observed in this study, we propose the following principles for a preferred tone space:

(1) **Pitch range preference**: Tonal contrasts prefer to occur within the most comfortable pitch-range for speech, i.e. about 100 – 180 Hz for males, and 170 – 250 Hz for females;

(2) **Pitch contrast limitation**: A pure pitch difference within the comfortable range is only sufficient for contrasts of two levels or parallel contours, and additional contrasts prefer to have additional cues, e.g. duration, phonation;

(3) **Non-modal phonations as a distinctive dimension**: In fact, pitch-independent non-modal phonations are preferred when tonal systems have similar contours/level tones, and they can make tones with similar pitch values distinctive;

(4) **Non-modal phonations as an enhancement cue**: Extra low tones are naturally falling and creaky, and extra high tones are naturally rising and tense (even falsetto);

(5) **The good pitch range for phonation contrasts**: Pitch-independent phonation contrasts are more likely to occur with the comfortable speech range (ranges listed in (1)) than with the extra low or extra high ranges; and two-way phonation contrasts are more preferred than three-way contrasts.

(6) **The good phonation for pitch contrasts**: Tonal contrasts generally prefer modal phonations to non-modal phonations.

These principles together make some interesting predictions about possible contrasts and preferred tonal spaces, and they are generalized into the tonal register model in Figure 5-5. This expandable tone space has two dimensions: The horizontal dimension is the breathiness continuum, and the vertical dimension is the pitch scale. Non-modal phonation can happen along both of these two dimensions. The hierarchy of preference is also noted: The boxes (registers) and tones with thicker lines are preferred spaces and contrasts, and dashed registers and tones are less preferred contrasts.

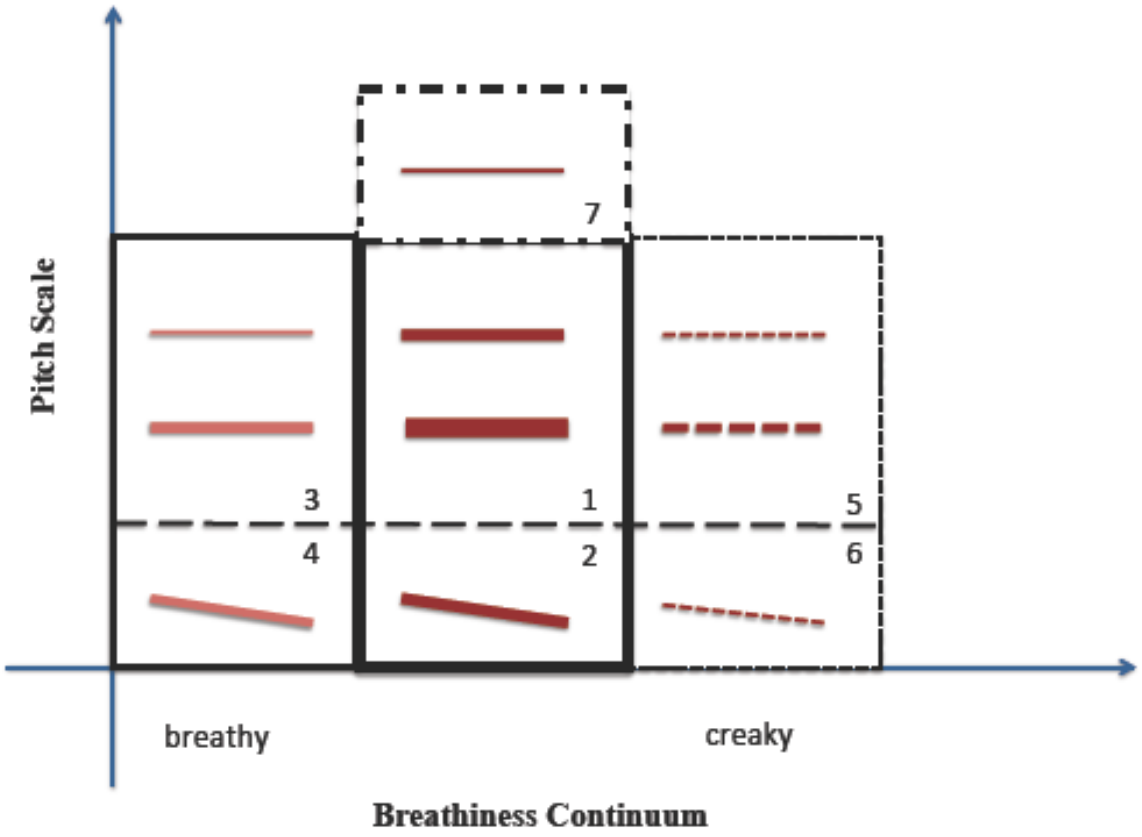


Figure 5-5 A model of the expandable tone space

As shown in Figure 5-5, the most basic tonal space is the #1 register, where pitch contrasts happen within the comfortable pitch range and with a modal voice, but it only allows two pitch levels/contours. This kind of system can be seen in Ewe (data from the UCLA Phonetic Archive). When there is a third level, it is likely to fall into the #2 register, where the lowest tone is naturally produced with vocal fry and slightly falling pitch. The dashed line between #1 and #2 indicates that the boundary is not absolute, and the non-modal phonation is allophonic. #1 and #2 are the most preferred registers for tonal contrasts. A majority of tonal languages utilize this kind of tone space, e.g. Mandarin. It is noted that the high tones in these languages are still produced

within the comfortable pitch range, not with a super high pitch. When the tonal space keeps expanding along the pitch scale as pitch contrasts increase, less commonly, languages can utilize the extra high register (#7), which is beyond the default pitch range for speech, and possibly produced with a tense or even falsetto voice; such cases include Black Miao, Yueyang dialect, and probably Cantonese. We also speculate that when pitch is very high, phonation contrasts are unlikely to happen.

Alternatively, the tonal space can expand along the horizontal dimension. It is possible to introduce a pitch-independent phonation into the language, so that the tones in register #3 and #4 are distinctive from the tones with similar pitch values in register #1 and #2. The solid line between #1 and #3 indicates that the non-modal phonation is phonemic. Phonation contrasts can fully cross with all pitch contrasts, as in Eastern Yi and Mpi, or partially cross with pitch contrasts, as in Southern Yi (c.f. Figure 5-3). A two-way contrast along the breathiness continuum is preferable, and only very rarely, languages would use a third phonation (registers #5 and #6) to contrast along the breathiness continuum, e.g. Jalapa Mazatec. Since this language is the only known case with a fully crossed three by three system, we shall take a close look at it. Figure 5-6 is the tone by phonation system of Jalapa Mazatec.

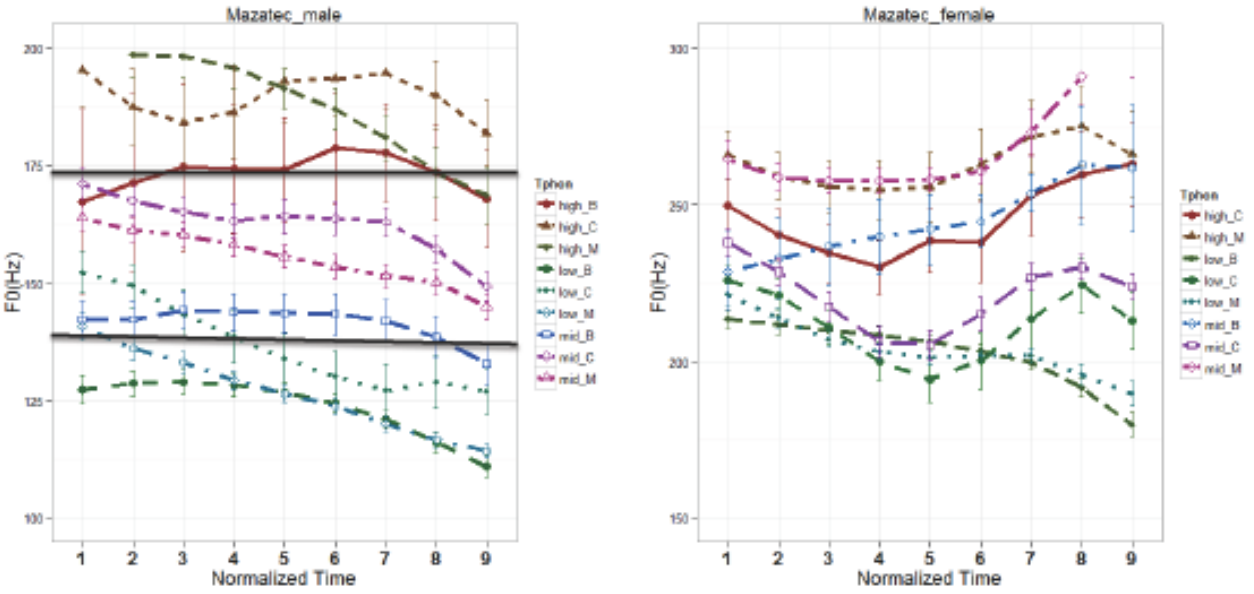


Figure 5-6 The tonal by phonation system of Jalapa Mazatec. The lines indicate the regions for the three tones. Data from http://www.phonetics.ucla.edu/voiceproject/voice.html.

As shown here, the three by three system is only found for the males' production. For females, high_modal and mid_modal are not contrastive, and the tones are generally not level. More importantly, for tones with the same contours/levels, there are only two-way phonation contrasts. Therefore, the system of female speakers is also a system with phonation and pitch crossed, but only with a two-way contrast, similar to Southern Yi. For the productions by male speakers, phonation contrasts, especially in mid and low tones, are indeed three-way, suggesting that a three-way phonation contrast in tonal languages is possible. For the high tones, it seems likely that the falling contour with high-modal also helps with the contrasts.

In sum, the dimensionality of tonal contrasts depends on the size of the tonal inventory; more registers are unfolded as tonal contrasts increase. It should be noted that, 1) although we have focused on phonation, in principle, other cues such as duration can also be a contrastive dimension for tonal contrasts; 2) although we only illustrate with level tones in Figure 5-5, the principles listed in this section are applicable to contours as well.

6. Comments on tonal registers

The idea of tonal registers is not new in phonological studies, but the organization of the model in Figure 5-5 is radically different from previous proposals about tonal registers presented in Chapter 1 and reviewed here.

Recall that tonal registers were originally proposed by Yip (1980), defined as sub-ranges of the pitch scale. For example, Cantonese four level tones can be represented by two binary features H vs. L and +/-upper: 11/21=[-upper, L], 33=[+upper, L], 22=[-upper, H], and 55=[+upper, H]. 11 and 22 belong to the lower register, and 33 and 55 belong to the upper register. This kind of register is motivated by the well-known fact that proto-Chinese tones split into two sets, based on onset voicing, resulting in two tonal registers, i.e. Yin and Yang in traditional terms (Haudricourt, 1972). However, Clements (2010) has questioned whether such registers are synchronic natural classes in Cantonese, as there is no phonological process based on the register features. Moreover, such registers do not provide perceivable cues for native speakers, as 22 vs. 33, with opposite values on the register feature, as well as the tone feature, are still the most confusable tonal pair (Mok and Wong, 2010).

Some proposals based on Chinese languages (Duanmu, 1990; Bao, 1999; Zhu, 2012) have then suggested that phonation is the phonetic correlate of Yip's tonal registers, given the fact that most of the languages with tonal registers actually involve phonation contrasts. Phonation-based registers are very common in numerous Wu dialects. Such registers are certainly more likely to be synchronic natural classes, as tone sandhi is usually constrained by this kind of register (Bao, 1999). However, despite the idea that phonation is related to tonal registers, it has been unclear 1) what kind of phonation types could be involved (different proposals involve different numbers and types of phonation), 2) how phonation and pitch interact with each other, and 3) how many contrastive pitch levels can occur in each register (three in Duanmu, 1990; four in Zhu, 2012). Moreover, allophonic non-modal phonation as in Mandarin is usually not taken into account.

Our new proposal provides a better understanding of these three questions:

1) Since phonation contrasts are continuous, it is nearly impossible to exclusively list all attested concrete phonation types in the world. In this dissertation, we instead propose two dimensions that phonation can vary along. Phonation can vary either independently from pitch or dependently along the pitch scale, and these two different non-modal phonation types have different functions in tonal contrasts. The distinction between these two types of non-modal phonation can help us understand why and when non-modal phonation can happen in tonal languages; and to what extent the variability of phonation types is language-specific (e.g. degree of breathiness for phonation contrasts), or universal (e.g. lowest tones are naturally creaky).

2) The previous section has provided an extensive answer this question. We would like to add some additional comments to one of the controversies: which register should have a higher pitch range, and whether two registers can overlap in pitch ranges (c.f. Duanmu, 1990; Zhu, 2012). Since the pitch difference between two contrastive registers is a secondary cue and not perceptually salient in languages (e.g. Green Mong, White Hmong and Black Miao), the contrastive registers are essentially parallel to each other. However, the pitch ranges for pitch-driven registers are fixed, and they can be either the highest or the lowest. This is a major difference from previous proposals, e.g. Zhu (2012).

3) Several previous proposals have at least three levels in each register, but we found this is overgenerating given the apparent limitation on pure-pitch contrasts. Based on the study on Black Miao, we speculate that pure pitch contrast is only good for two levels; and that pitch contrasts are more likely to happen with in modal register than non-modal registers. For example, among the Black Miao mid tones, there is only one breathy tone (T33), but two modal tones (T22 and T44).

6. Conclusions and future work

This dissertation has extended our knowledge about tonal contrasts, and provided a better understanding of the interaction between phonation and pitch. We demonstrate that pitch contrasts are limited in languages, and that tonal contrasts are more than pitch contrasts; therefore, future tonal studies should consider all possible cues that contribute to tonal contrasts. We have proposed a new tone space model, and the principles for good dispersion of tonal

contrasts. More extensive phonetic work, with more languages from more language families, should be done to validate the predictions by the new model. We will need to bear in mind that tonal processing in African and American languages might not be the same as in Asian tonal languages, and perhaps information from higher levels such as prosody should be taken into account as well.

This study also sheds light on the modeling of tonal spaces. Previous studies on tonal spaces mainly focused on pitch contrasts, and phonation cues were not taken into account. We have shown that distinctiveness should be calculated from all useful cues, and MDS tonal spaces are a good method to integrate multidimensional cues for the dispersion among the tonal categories. Future dispersion work, not limited to tones, should all consider using a multivariate tool to integrate all available cues.

In this study, we have not proposed a learning model for tonal contrasts with phonation cues, but this study has provided a possible way to weight phonation cues in tonal contrasts: contrastive non-modal phonation should be weighted higher than pitch differences, and pitch-driven non-modal phonation should be weighted lower than pitch differences. When weighting allophonic non-modal phonation, we should also consider the distinctiveness of pitch cues between potentially confusable pairs.

Furthermore, in studying the co-varying relationship between phonation and pitch in Chapter 3, we found several pitch thresholds for voice quality and we speculate that glottal configurations

might change at these thresholds. This proposal should be validated by direct physiological methods, such as EGG and high-speed laryngeal imaging.

Finally, the different functions of non-modal phonation should be further validated with psychoacoustic studies. Based on this study, we speculate that pitch-driven non-modal phonations can affect the perception of pitch location, as they are cues for extreme pitch values. This kind of cue is apparently not effective in the mid pitch range (Bishop and Keating, 2012), but can be helpful in judging the highest/lowest pitches. This is potentially a useful cue for listeners' tone normalization. For example, if we present pitch continuua with different phonation types (e.g. modal, vocal fry), we expect listeners will be biased towards low tones for the pitch continuum with vocal fry.

All in all, we believe this study has made a significant contribution to the better understanding of tones, and opens a door for various future work on tonal processing.

Appendix

Phonation and tone effects and interaction on different voice measures

```
Formula: H1*-H2* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                  Estimate Std. Error t value
(Intercept)         5.0013     0.5229   9.564
Tonemid            -0.3263     0.3155  -1.034
PhonationT         -3.0897     0.2890 -10.690
Tonemid:PhonationT  1.4507     0.2219   6.538
```

```
Formula: H1* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                  Estimate Std. Error t value
(Intercept)        15.0224     0.9466  15.869
Tonemid             0.6663     0.4314   1.544
PhonationT         -3.5216     0.4231  -8.323
Tonemid:PhonationT  1.2244     0.3032   4.039
```

```
Formula: H1*-A1* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                  Estimate Std. Error t value
(Intercept)        19.1717     0.6797  28.207
Tonemid            -2.1628     0.4586  -4.717
PhonationT         -5.0049     0.5706  -8.771
Tonemid:PhonationT  2.7770     0.3957   7.017
```

```
Formula: H1*-A2* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                  Estimate Std. Error t value
(Intercept)        18.4715     0.7685  24.037
Tonemid            -2.2024     0.5307  -4.150
PhonationT         -4.9961     0.6127  -8.155
Tonemid:PhonationT  2.3564     0.5105   4.616
```

```
Formula: H1*-A3* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         10.7502     0.8014  13.415
Tonemid             -1.4170     0.5720  -2.477
PhonationT          -5.8359     0.8610  -6.778
Tonemid:PhonationT   2.7848     0.5750   4.843


Formula: H2*-H4* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)          2.8917     0.4882   5.923
Tonemid             -0.8685     0.3808  -2.281
PhonationT          -0.2374     0.2526  -0.940
Tonemid:PhonationT   0.2522     0.3351   0.753


Formula: H2* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         10.0059     0.8857  11.297
Tonemid              0.9852     0.6070   1.623
PhonationT          -0.4258     0.3694  -1.153
Tonemid:PhonationT  -0.2164     0.3251  -0.666


Formula: H4* ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)          7.1665     0.7223   9.921
Tonemid              1.8288     0.4525   4.042
PhonationT          -0.2594     0.3080  -0.842
Tonemid:PhonationT  -0.4011     0.3443  -1.165


Formula: CPP ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         22.2309     0.4743   46.87
Tonemid              2.2628     0.2606    8.68
PhonationT           0.8187     0.2563    3.19
Tonemid:PhonationT  -0.3221     0.1447   -2.23
```

```
Formula: F0 ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         153.007      8.714  17.559
Tonemid              21.339      2.022  10.556
PhonationT            1.428      1.501   0.952
Tonemid:PhonationT    1.432      1.323   1.082


Formula: PIC ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)          976.91      53.75  18.174
Tonemid               43.68      35.63   1.226
PhonationT          -123.98      30.56  -4.057
Tonemid:PhonationT    23.37      20.61   1.134


Formula: CQ ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                    Estimate Std. Error t value
(Intercept)         0.515334   0.011441   45.04
Tonemid             0.005814   0.004981    1.17
PhonationT          0.054313   0.007399    7.34
Tonemid:PhonationT -0.010392   0.004303   -1.41


Formula: SQ ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         0.46778    0.09434   4.959
Tonemid             0.10879    0.08465   1.285
PhonationT         -0.12560    0.08646  -1.453
Tonemid:PhonationT -0.06850    0.10266  -0.667


Formula: Closing Duration ~ Tone * Phonation + (Tone + Phonation | speaker)

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         1.44830    0.13287  10.900
Tonemid            -0.14908    0.08562  -1.741
PhonationT         -0.25892    0.08228  -3.147
Tonemid:PhonationT  0.01326    0.09981   0.133
```

Bibliography

Abramson, A. S. (**1978**). "Static and dynamic acoustic cues in distinctive tones," Language and Speech **21**, 319-325.

Abramson, A. S., and Luangthongkum, T. (**2009**). "A fuzzy boundary between tone languages and voice-register languages," in *Frontiers in Phonetics and Speech Science*, edited by G. Fant, H. Fujisaki, and J. Shen (The Commercial Press, Beijing), pp. 149-155.

Abramson, A. S., Luangthongkum, T., and Nye, P. W. (**2004**). "Voice register in Suai (Kuai): An analysis of perceptual and acoustic data," Phonetica **61**, 147-171.

Abramson, A. S., Nye, P. W., and Luangthongkum, T. (**2007**). "Voice register in Khmu': Experiments in production and perception," Phonetica **64**, 80-104.

Alexander, J. A. (**2010**). "The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems," (Ph.D thesis, Northwestern University, Evanston).

Andruski, J. E. (**2006**). "Tone clarity in mixed pitch/phonation-type tones," Journal of Phonetics **34**, 388-404.

Andruski, J. E., and Ratliff, M. (**2000**). "Phonation types in production of phonological tone: the case of Green Mong," Journal of the International Phonetic Association **30**, 37-61.

Baayen, R. H. (**2010**). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge，UK).

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice* (Singular Publishing Group, San Diego).

Bao, Z. (**1999**). *The Structure of Tone* (Oxford University Press, New York).

Barry, J. G., and Blamey, P. J. (**2004**). "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese," Journal of the Acoustical Society of America **116**, 1739-1748.

Bates, D., Maechler, M., and Dai, B. (**2008**). "lme4: Linear mixed-effects models using S4 classes."

Becker-Kristal, R. (**2010**). "Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus," (Ph.D thesis, University of California Los Angeles).

Belotel-Grenié, A., and Grenié, M. (**1994**). "Phonation types analysis in standard Chinese," in *Proceedings of Spoken Language Processing* pp. 343-346.

Belotel-Grenié, A., and Grenié, M. (**1997**). "Types de phonation et tons en chinois standard," Cahiers de linguistique - Asie orientale **26**, 249-279.

Belotel-Grenié, A., and Grenié, M. (**2004**). "The creaky voice phonation and the organisation of chinese discourse," in *International Symposium on Tonal Aspects of Languages: With Emphasis of Tone Languages*, pp. 5-8.

Bishop, J., and Keating, P. (**2012**). "Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex," Journal of the Acoustical Society of America **132**, 1100-1112.

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

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement," Journal of Phonetics **18**, 37-49.

Boersma, P., and Weenink, D. (**2012**). "Praat: http://www.fon.hum.uva.nl/praat/"

Brunelle, M. (**2009**). "Tone perception in Northern and Southern Vietnamese," Journal of Phonetics **37**, 79-96.

Brunelle, M., and Finkeldey, J. (**2011**). "Tone perception in Sgaw Karen," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 372-375.

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

Chang, K. (**1947**). "Tones of the Miao and Yao languages (Chinese). Shiyusuo Jikan, " Bulletin of the Institute of History and Philology **16**, 93-110.

Chao, Y. R. (**1948**). *Mandarin Primer* (Harvard University Press, Cambridge, MA).

Chao, Y. R. (**1956**). "Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese," in *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*, edited by M. Halle, H. Lunt, H. McLean, and C. V. Schooneveld (Mouton), pp. 52-59.

Chávez Peón, M. E. (**2010**). "The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec, " (Ph.D. thesis, The University of British Columbia, Vancouver).

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

Clements, G. N. (**1983**). "The hierarchical representation of tone features," Current Approaches to African Linguistics **1**, 145-176.

Clements, G. N., Michaud, A., and Patin, C. (**2010**). "Do we need tone features," in *Tones and Features: Phonetic and Phonological Perspectives*, edited by J. A. Goldsmith, E. Hume, and L. Wetzels (Wlater de Gruyter, Berlin), pp. 3-24.

Davies, P., Lindsey, G., Fuller, H., and Fourcin, A. (**1986**). "Variation in glottal open and closed phases for speakers of English," in *Proceedings of the Institute of Acoustics*, pp. 539-546.

Davison, D. S. (**1991**). "An acoustic study of so-called creaky voice in Tianjin Mandarin," UCLA Working Papers in Phonetics **78**, 50-57.

DiCanio, C. T. (**2008**). "The phonetics and phonology of San Martín Itunyoso Trique," (Ph.D thesis, University of California Berkeley), pp. 162-188.

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

DiCanio, C. T. (**2012**). "Coarticulation between tone and glottal consonants in Itunyoso Trique," Journal of Phonetics **40**, 162-176.

Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (**1992**). "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," Journal of Voice **6**, 44-54.

Duanmu, S. (**1990**). "A formal study of syllable, tone, stress, and domain in Chinese languages," (Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA).

Duanmu, S. (**2002**). *The Phonology of Standard Chinese* (Oxford University Press, New York).

Edmondson, J. A., and Gregerson, K. J. (**1992**). "On five-level tone systems. ," Language in Context: Essays for Robert E. Longacre, 555-576.

Erickson, D. (**1993**). "Laryngeal muscle activity in connection with Thai tones," Research Institute of Logopedics and Phoniatrics Annual Bulletin **27**, 135-149.

Esling, J. H. (**1984**). "Laryngographic study of phonation type and laryngeal configuration," Journal of the International Phonetic Association **14**, 56-73.

Esling, J. H., Edmondson, J. A., Harris, J. G., Li, S., and Ziwo, L. (**2000**). "The aryepiglottic folds and voice quality in Yi and Bai languages: Laryngoscopic case studies," Minzu Yuwen **6**, 47-53.

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

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

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

Ethonologue (**2012**). "Ethonologue: http://www.ethnologue.com/ "

Faytak, M., and Yu, A. C. L. (**2011**). "A typological study of the interaction between level tones and duration. ," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 659-663.

Flemming, E. (**2004**). "Contrast and perceptual distinctiveness," in *The Phonetic Bases of Markedness*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge, UK), pp. 232-276.

Frazier, M. (**2009**). "The production and perception of pitch and glottalization in Yucatec Maya," (Ph.D. thesis, University of North Carolina, Chapel Hill).

Gandour, J. T. (**1978**). "The perception of tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 41-76.

Gandour, J. T., and Harshman., R. A. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Language and Speech **21**, 1-33.

Gårding, E., Kratochvil, P., Svantesson, J.-O., and Zhang, J. (**1986**). "Tone 4 and Tone 3 discrimination in modern standard Chinese," Language and Speech **29**, 281-293.

Garellek, M., Esposito, C. M., Keating, P., and Kreiman, J. (**2012**). "Perception of spectral slopes and tone identification in White Hmong," UCLA Working Papers in Phonetics **110**, 24-45.

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

Garellek, M., Keating, P., Esposito, C. M., and Kreiman, J. (**2013**). "Voice quality and tone identification in White Hmong," Journal of the Acoustical Society of America **133**, 1078-1089.

Gerfen, C., and Baker, K. (**2005**). "The production and perception of laryngealized vowels in Coatzospan Mixtec," Journal of Phonetics **33**, 311-334.

Gerratt, B. R., and Kreiman, J. (**2001**). "Toward a taxonomy of nonmodal phonation," Journal of Phonetics **29**, 365-381.

Gobl, C., and Ní Chasaide, A. (**2012**). "Voice source variation and its communicative functions," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Wiley-Blackwell, Oxford), pp. 378-423.

Goldsmith, J. A. (**1979**). *Autosegmental phonology* (Garland Publishing House, New York).

Gordon, M., and Ladefoged, P. (**2001**). "Phonation types: a cross-linguistic overview," Journal of Phonetics **29**, 383-406.

Guion, S. G., Post, M. W., and Payne, D. L. (**2004**). "Phonetic correlates of tongue root vowel contrasts in Maa," Journal of Phonetics **32**, 517-542.

Guo, Q. J., Strauss, H., Liu, C. Q., Zhao, Y. L., Pi, D. H., Fu, P. Q., Zhu, L. J., and Yang, R. D. (**2005**). "Carbon and Oxygen Isotopic Composition of Lower to Middle Cambrian Sediments at Taijiang, Guizhou Province, China," Geological Magazine **142**, 723-733

Halle, M., and Stevens, K. N. (**1971**). "A note on laryngeal features," Quarterly Progress Report, MIT Research Laboratory of Electronics **101**, 198-212.

Harris, W. J., and Umeda, N. (**1987**). "Difference limens for fundamental frequency contours in sentences," Journal of the Acoustical Society of America **81**, 1139-1145.

Haudricourt, A. G. (**1954**). "De l'origine des tons en Viênamien," Journal Asiatique **242**, 69-82.

Haudricourt, A. G. (**1972**). "Two-way and three-way splitting of tonal systems in some far eastern languages (translated by Christopher Court)," in *A Conference on Tai Phonetics and Phonology*, edited by J. G. Harris, and R. B. Noss (Mahidol University), pp. 58-86.

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

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy voice quality," Journal of Speech and Hearing Research **37**, 769-778.

Hockett, C. F. (**1947**). "Peiping Phonology," Journal of the American Oriental Society **67**, 253-267.

Hollien, H. (**1974**). "On vocal registers," Journal of Phonetics **2**, 125-143.

Hollien, H., and Michel, J. F. (**1968**). "Vocal fry as a phonational register," Journal of Speech and Hearing Research **11**, 600-604.

Hollien, H., and Wendahl, R. W. (**1968**). "Perceptual study of vocal fry," Journal of the Acoustical Society of America **43**, 506-509.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," Journal of the Acoustical Society of America **84**, 511-529.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," Journal of Speech and Hearing Research **38**, 1212-1223.

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

Hombert, J. M. (**1977**). "Difficulty of producing different F0 in speech," UCLA Working Papers in Phonetics **36**, 12-19.

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

Howard, D. M., Lindsey, G. A., and Allen, B. (**1990**). "Toward the quantification of vocal efficiency," Journal of Voice **4**, 205-212.

Hyman, L. M. (**1986**). "The representation of multiple tone heights," in *The Phonological Representation of Suprasegmentals: Studies on African Languages Offered to John M. Stewart on His 60th Birthday*, edited by K. Bogers, H. Van der Hulst, and M. Mous (Foris Publications, Dordrecht), pp. 109-152.

Hyman, L. M. (**2010**). "Does tone have features," in *Tones and Features: Phonetic and Phonological Perspectives*, edited by J. A. Goldsmith, E. Hume, and L. Wetzels (Wlater de Gruyter, Berlin), pp. 50-80.

IPA (**1999**). *Handbook of the International Phonetic Association* (Cambridge University Press).

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

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction," Speech Communication **27**, 187-207.

Keating, P., and Esposito, C. (**2007**). "Linguistic voice quality," UCLA Working Papers in Phonetics **105**, 85-91.

Keating, P., Esposito, C., Garellek, M., Khan, S., and Kuang, J. (**2011**). "Phonation contrasts across languages," in *Proceedings of the International Conference of Phonetic Sciences*, pp. 1046-1049.

Keating, P., Kuang, J., Esposito, C., Garellek, M., and Khan, S. (**2012**). "Multi-dimensional phonetic space for phonation contrasts," Poster presented in *LabPhon 13* (Stuttgart, Germany).

Keating, P., and Kuo, G. (**2012**). "Comparison of speaking fundamental frequency in English and Mandarin," Journal of the Acoustical Society of America **132**, 1050-1060.

Keating, P., and Shue, Y.-L. (**2009**). "Voice quality variation with fundamental frequency in English and Mandarin," Journal of the Acoustical Society of America **126**, 2221.

Keidar, A., Hurtig, R. R., and Titze, I. R. (**1987**). "The perceptual nature of vocal register change," Journal of Voice **1**, 223-233.

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

Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," Journal of Phonetics **35**, 104-117.

Kingston, J. (**2005**). "The phonetics of Athabaskan tonogenesis," in *Athabaskan Prosody* (John Benjamins), pp. 137-184.

Kollmeier, B., Brand, T., and Meyer, B. (**2008**). "Perception of speech and sound," in *Springer Handbook of Speech Processing*, edited by J. Benesty, M. Sondhi, and Y. Huang (Springer, Berlin), pp. 61-82.

Kong, J. (**2001**). *On Language Phonation* (The Central University of Nationalities Press, Beijing, China).

Kong, J. (**2007**). *Laryngeal Dynamics and Physiological Model* (Publishing House of Peking University, Beijing).

Kreiman, J., Gerratt, B., and Antonanzas-Barroso, N. (**2007**). "Measures of the glottal source spectrum," Journal of Speech, Language, and Hearing Research **50**, 595-610.

Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). "Listener experience and perception of voice quality," Journal of Speech and Hearing Research **33**, 103-115.

Kuang, J. (**2011**). "Production and perception of the phonation contrast in Yi," (M.A. thesis, University of California Los Angeles).

Kuang, J., and Keating, P. (**2012**). "Glottal articulations of phonation contrasts and their acoustic and perceptual consequences," UCLA Working Papers in Phonetics **111**, 123-161.

Kwan, J. C. (**1966**). "A phonology of a Black Miao dialect," (M.A. thesis, University of Washington, Seattle).

Ladefoged, P. (**1971**). *Preliminaries to Linguistic Phonetics* (University of Chicago Press, Chicago).

Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L. (**1978**). "Generating vocal tract shapes from formant frequencies," Journal of the Acoustical Society of America **64**, 1027-1035.

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge University Press).

Laver, J. (**1991**). "Description of voice quality in general phonetic theory," in *The Gift of Speech: Papers in the Analysis of Speech and Voice*, edited by J. Laver (Edinburgh University Press, Edinburgh, Scotland), pp. 184-208.

Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala, and J. J. Jaeger (Academic Press, Orlando), pp. 3-42.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. Hardcastle, and A. Marchal (Kluwer, Dordrecht), pp. 403-439.

Lindblom, B., and Maddieson, I. (**1988**). "Phonetic universals in consonant systems," in *Language, Speech and Mind, Studies in Honor of Victoria A. Fromkin*, edited by L. M. Hyman, and C. N. Li (Routledge, London), pp. 62-78.

Ma, X. (**2003**). *An introduction to Sino-Tibetan Languages* (Minzu Publishing House, Beijing).

Maddieson, I. (**1978**). "Universals of tone," in *Universals of Human Language*, edited by J. H. Greenberg, C. Ferguson, and E. A. Moravcsik (Stanford University Press, Palo Alto), pp. 335-365.

Maddieson, I., and Hess, S. (**1986**). ""Tense" and "Lax" revisited: more on phonation type and pitch in minority languages in China," UCLA Working Papers in Phonetics **63**, 103-109.

Maddieson, I., and Ladefoged, P. (**1985**). ""Tense" and "lax" in four minority languages of China," Journal of Phonetics **13**, 433-454.

Marasek, K. (**1996**). "Glottal correlates of the word stress and the tense/lax opposition in German," in *Proceedings of the International Congress of Phonetic Sciences* (University of Stuttgart, Stuttgart), pp. 1573-1576.

Marasek, K. (**1997**). "Electroglottographic description of voice quality," in *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS)* (University of Stuttgart).

Martinet, A. (**1952**). "Function, structure, and sound change," Word **8**, 1-32.

Martinet, A. (**1955**). *Economie des Changements Phonétiques* (Francke, Berne).

Mazaudon, M., and Michaud, A. (**2006**). "Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitch-plus-voice-quality tones (Vietnamese data)," in *Proceedings of Speech Prosody* pp. 823-826.

Mazaudon, M., and Michaud, A. (**2009**). "Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link'in tonogenesis," Phonetica **65**, 231-256.

Michaud, A. (**2004**). "A measurement from electroglottography: DECPA, and its application in prosody," in *Proceedings of Speech Prosody 2004*, pp. 633-636.

Miller, A. L. (**2007**). "Guttural vowels and guttural co-articulation in Ju|'hoansi," Journal of Phonetics **35**, 56-84.

Moisik, S. R., Lin, H., and Esling, J. H. (**2010**). "An investigation of laryngeal behavior during Mandarin tone production using simultaneous laryngoscopy and laryngeal ultrasound," in *Proceedings of the 9th Phonetics Conference of China*, pp. 1548-1558.

Mok, P., and Wong, P. (**2010**). "Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables," in *Proceedings of Speech Prosody*, pp. 1-4.

Ohala, J. J. (**1973**). "The physiology of tone," Southern California Occasional Papers in Linguistics **1**, 1-14.

Peng, J. G., and Zhu, X. N. (**2010**). "Falsetto in Yueyang dialect, " Journal of Contemporary Linguistics **1**, 24-32.

Pennington, M. (**2005**). "The phonetics and phonology of glottal manner features," (Ph.D thesis, Indiana University, Bloomington).

Pulleyblank, D. (**1986**). *Tone in Lexical Phonology* (Springer).

Redi, L., and Shattuck-Hufnagel, S. (**2001**). "Variation in the realization of glottalization in normal speakers," Journal of Phonetics **29**, 407-429.

Roengpitya, R. (**2007**). "The variations, quantification, and generalizations of standard Thai tones," in *Experimental Approaches to Phonology* edited by M.-J. Solé, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford), pp. 270-302.

Rose, P. (**1997**). "A seven-tone dialect in Southern Thai with super-High: Pakphanang tonal acoustics and physiological inferences," in *Southeast Asian Linguistic Studies in Honour of Vichin Panupong*, edited by A. S. Abramson (Chulalongkorn University Press, Bangkok, Thailand), pp. 191-208.

Rothenberg, M., and Mashie, J. J. (**1988**). "Monitoring vocal fold abduction through vocal fold contact area," Journal of Speech and Hearing Research **31**, 338-351.

Roubeau, B., Henrich, N., and Castellengo, M. (**2009**). "Laryngeal vibratory mechanisms: The notion of vocal register revisited," Journal of Voice **23**, 425-438.

Shepard, R. N. (**1972**). "Psychological representation of speech sounds," in *Human Communication: A Unified View*, edited by E. E. David, and P. B. Denes (McGraw-Hill, New York), pp. 67-113.

Shi, F., and Zhou, D. (**2005**). "An acoustic study of tense and lax vowels in Southern Yi," Yuyan Yanjiu **25**, 19-23.

Shue, Y.-L., Keating, P. A., Vicenik, C., and Yu, K. (**2011**). "VoiceSauce: A program for voice analysis," in *Proceedings of the International Congress of Phonetic Sciences*, pp. 1846-1849.

Silverman, D. (**1997**). *Phasing and Recoverability* (Routledge, London).

Silverman, D. (**2003**). "Pitch discrimination between breathy vs. modal phonation," in *Laboratory Phonology 6*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 293-304.

Sjölander, K. (**2004**). *The Snack Sound Toolkit* (KTH Stockholm, Sweden).

Slifka, J. (**2006**). "Some physiological correlates to regular and irregular phonation at the end of an utterance," Journal of Voice **20**, 171-186.

Snider, K. L. (**1990**). "Tonal upstep in Krachi: evidence for a register tier," Language **66**, 453-474.

Stevens, K. N. (**1977**). "Physics of laryngeal behavior and larynx modes," Phonetica **34**, 264-279.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (University of Chicago Press, Chicago).

Sundberg, U. (**1994**). "Tonal and temporal aspects of child directed speech," Lund University Department of Linguistics and Phonetics Working Papers **43**, 128-131.

Tang, K. E. (**2008**). "The phonology and phonetics of consonant-tone interaction," (Ph.D. thesis, University of California Los Angeles).

Tehrani, H. (**2012**). "EggWorks: http://www.linguistics.ucla.edu/faciliti/facilities/physiology/EGG.htm"

Thongkum, T. L. (**1990**). "The interaction between pitch and phonation type in Mon: phonetic implications for a theory of tonogenesis," Mon-Khmer Studies **16**, 11-24.

Thurgood, E. (**2004**). "Phonation types in Javanese," Oceanic Linguistics **43**, 277-295.

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

Thurgood, G. (**2007**). "Tonogenesis revisited: Revising the model and the analysis," in *Studies in Tai and Southeast Asian Linguistics* edited by J. Harris, S. Burusphat, and J. Harris, pp. 241-262.

Titze, I. R. (**1988**). "A framework for the study of vocal registers," Journal of Voice **2**, 183-194.

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

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs).

Tseng, C. Y. (**1981**). "An acoustic phonetic study on tones in Mandarin Chinese," (Ph.D. thesis, Brown University, Providence).

Wang, W. (**1967**). "Phonological features of tone," International Journal of American Linguistics **33**, 93-105.

Woo, N. (**1969**). "Prosody and phonology," (Ph.D. thesis, Massachusetts Institute of Technology).

Xu, Y. (**2004**). "Understanding tone from the perspective of production and perception," Language and Linguistics **5**, 757-797.

Xu, Y., and Sun, X. (**2002**). "Maximum speed of pitch change and how it may relate to speech," Journal of the Acoustical Society of America **111**, 1399-1413.

Yang, R. X. (**2011**). "The Phonation factor in the categorical perception of Mandarin tones," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 2204-2207.

Yip, M. J. W. (**1980**). "The tonal phonology of Chinese," (Ph.D. thesis, Massachusetts Institute of Technology).

Yip, M. J. W. (**2002**). *Tone* (Cambridge University Press, Cambridge, UK).

Yu, K. M. (**2011**). "The learnability of tones from the speech signal," (Ph.D. thesis, University of California Los Angeles).

Yu, K. M., and Lam, H. W. (**2011**). "The role of creaky voice in Cantonese tonal perception," in *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 2240-2243.

Zhao, Y., and Jurafksy, D. (**2007**). "The effect of lexical frequency on tone production," in *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 477-479.

Zhao, Y., and Jurafksy, D. (**2009**). "The effect of lexical frequency and Lombard reflex on tone hyper-articulation," Journal of Phonetics **37**, 231-247.

Zhu, X. N. (**2012**). "Multi registers and four levels: A new tonal model," Journal of Chinese Linguistics **40**, 1-17.

Zsiga, E., and Nitisaroj, R. (**2007**). "Tone features, tone perception, and peak alignment in Thai," Language and Speech **50**, 343-383.

Zsiga, E. C. (**2012**). "Contrastive tone and its implementation," in *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. K. Huffman (Oxford University Press, Oxford), pp. 196-207.