UNIVERSITY OF CALIFORNIA

Los Angeles

Production and Perception of the Phonation Contrast in Yi

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Linguistics

by

Jianjing Kuang

2011

The thesis of Jianjing Kuang is approved.

Sun-Ah Jun

Jody Kreiman

Patricia A. Keating, Committee Chair

University of California, Los Angeles

2011

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT OF THE THESIS

Production and Perception of the Phonation Contrast in Yi

by

Jianjing Kuang

Master of Art in Linguistics

University of California, Los Angeles, 2011

Professor Patricia A. Keating, Chair

The purpose of this thesis is to provide a comprehensive understanding of how contrastive phonation is produced and perceived by native speakers of a language with both tonal and phonation contrasts, Yi. In the production experiment, we measure a wide scope of relevant physiological and acoustic parameters, which show substantial physiological-acoustic coupling: a Contact Quotient (CQ) distinction is the essential property of the phonation contrast, while H1*-H2* and H1*-A1*, which are significantly correlated with CQ, are the best acoustic measures for the phonation contrast. The bandwidth of the first formant (B1) and the Cepstral peak prominence (CPP) are effective acoustic cues too. In addition to the well-established contributions of the vocal folds, this study gives insight into the role of supraglottal settings for the phonation contrast. A consistent F1 difference for the phonation contrast in Yi indicates a shape change in the vocal tract, supporting the multidimensional phonation model proposed by Edmondson & Esling (2006). Such a supraglottal effect is similar to that in ATR (Advanced Tongue Root) languages, related to a change in pharyngeal size, though in Yi the direction is RTR (Retracted Tongue Root). With more detailed discussion about variations in the tense vs. lax contrast across speakers as well as across languages, the relativity of phonation contrasts is highlighted.

The perception experiment looks at the effectiveness of voice quality, F0 and F1 for the perception of phonation categories by Yi listeners. We draw a perception map from a perceptual confusion matrix and then compare that with a production map. It is found that listeners heavily rely on F1 for the perception of the tense vs. lax contrast, though it is a secondary cue in production. However, the perception map generally agrees with the production map, though the perception map varies across listeners. The variation found in the production experiment suggests a possible sound change in the phonation contrast. The variation in the perception map further indicates the future direction of the sound change: Vowel quality might eventually take over as the distinctive feature of the tense vs. lax contrast.

1 Introduction

Phonation types are used in many languages allophonically, as prosodic cues or enhancement cues of other distinctive features. Very few languages around the world use phonation itself as a phonemic dimension, and thus phonation phonemes are less understood than phonological contrasts in consonants, vowels and tones. Yi is such a language, with a tense vs. lax contrast based on phonation differences, combined with orthogonal phonemic tones. The contrast in Yi is one type of so-called register contrast, but not all register contrasts involve phonation. To better understand phonation-based register contrasts in languages like Yi, this thesis will conduct experimental studies from both production and perception points of view.

The tense vs. lax contrast involves multidimensional production mechanisms, including both glottal configurations and supraglottal settings. However, the relationships of glottal configurations to their acoustic outcomes have not been well established, and little is known about the effect of supraglottal settings on voice quality. Therefore, from both acoustic and physiological perspectives, an extensive production experiment is conducted to comprehend the production-acoustics coupling in the Yi phonation contrast.

In addition to investigating the mechanism of tense vs. lax contrast production, a perception experiment is designed to obtain a perceptual map of the Yi phonation contrast and to reveal the mental reality of the phonation contrast for native speakers.

1.1 About Yi

The Yi (彝) language (own name in standard dialect: Nuosu), also known as Loloish, belongs to the Yi branch of the Tibeto-Burman family of the Sino-Tibetan phylum. The name "Yi" refers to both the Yi language and the whole Yi branch of languages (Loloish), because it has the most population in this language family branch. Sometimes Yi, Burmese and Zaiwa are collectively called Burmese-Lolo. Yi people are geographically distributed in Yunnan, Sichuan and Guizhou provinces of China. Yunnan province has the largest Yi population, and Yi people are distributed in most counties of that province. The Yi language is grouped into four main dialects by geographical distribution. The Northern dialect is mainly distributed in Sichuan province; Xide dialect is a representative. The Southern dialect is primarily in southern Yunnan and is represented by Xinping dialect. The Eastern dialect is in the eastern part of Yunnan and Guizhou provinces; Luquan is a representative. The Western dialect is mainly in western Yunnan and is exemplified by Weishan.

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, no diphthongs, no codas; and thus syllables are typically CV. Third, all Yi languages are tonal languages, typically with 3 tones, namely, High, Mid, and Low． 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.

Yi dialects mainly vary in two features. First, voiced consonants: In some dialects, the eastern dialect particularly, voiced obstruents are realized with prenasalization. Secondly, tense vs. lax contrasts are distributed asymmetrically across dialects. They are realized with middle and high tones in the northern dialect, conversely with middle and low tones in the southern dialect, and fully across all tones in the eastern dialect. Most dialects have a tense vs. lax contrast for all vowels, but some dialects only keep the contrast in high vowels. This distribution is the reverse of Burmese languages (e.g. Jingpo).

The dialect of the two villages discussed in this thesis is a Southern dialect. The villages are in Yunnan province, on the border of China, Vietnam and Laos. The tense vs. lax contrast in this dialect can be realized in all vowels but only with middle and low tones. Hereafter, the notation for a tense vowel will be an underscore, following Ma's convention. Many researchers use the creaky diacritic for tense register, but we will demonstrate that this is not accurate, since tense voice does not necessarily mean creaky (ref. section 2.6).

There are 29 consonants, 7 vowels, 3 tones and 2 phonations in this dialect, listed in Table 1:

Table 1. Inventory of Southern Yi.

1A. Consonants: voicing contrast across manners and places except nasals

p		t		k
p^h		t^h		k^h
b		d		g
	ts		tɕ	
	ts^h		$tɕ^h$	
	dz		dʑ	
f	s		ɕ	x(h)
v	z		ʑ	ɣ
m		n		ŋ
		l		
		ɬ		

1B. Distribution of tense vs. lax contrast in vowels and tones:

Vowels and Registers: Tense vs. lax contrast in all vowels

LAX	i	ɛ	a	ɿ	ə	o	u
TENSE	i̱	ɛ̱	a̱	ɿ̱	ə̱	o̱	u̱

Tones and Registers: Tense vs. lax contrast in the mid and low tones

	LOW	MID	HIGH
LAX	21L	33L	55L
TENSE	21T	33T	

1.2. Tense vs. lax phonation contrast in Yi

1.2.1 Tense vs. lax contrast in Yi and related languages is phonation based

A tense vs. lax contrast is the most widely shared phonological feature among Tibeto-Burman, especially Loloish branch, languages. According to previous fieldwork studies (among them, Ma 1948 is the earliest, see Ma 2003) for comprehensive summary of all of his works), the list includes Yi, Jingpo, Zaiwa, Langsu, Nusu, Lisu, Hani, Lahu, Bai and more than ten other languages. Traditionally, linguists referred to a language as having a tense vs. lax contrast when the vowel inventory in the language can be separated into two registers by some kind of tenseness gesture.

The tense and lax feature more generally has a long history and refers to various linguistic phenomena. Tense vs. Lax in vowels in Germanic languages is related to the muscular tension in the tongue. Phonetically, tense vowels have higher tongue position and longer duration than lax vowels. They are more peripheral in the vowel space, suggesting a more extreme articulatory gesture. African languages with ATR harmony are also said to have distinctive tense vs. lax vowels. Tenseness in these languages is related to pharyngeal cavity size (Ladefoged 1964). Vowels with advanced tongue root ([+ATR]), usually with pharynx expanded, are referred to as "tense". Tense vowels in these languages are more front and higher in the vowel space. Vowel harmony between tense and lax registers ("yang" and "yin" in traditional literature) is also found in Mongolian, an Asian language. Bao (1992) measured F1, F2, F1-F2 and the pharyngeal cavity size by calculating the vocal tract channel area for Mongolian. Results suggested that tense ("yang") vowels in this language are more back and lower, which is the reverse of African and Germanic languages.

The terms tense and lax in Tibeto-Burman languages and other southeastern Asian languages refer to a third type of contrast, which is related to phonation status but may be accompanied by some non-phonatory features. Even within this type, the mechanism of tense vs. lax contrasts still varies across languages.

Table 2. Phonetic realization of tense vs. lax registers in four languages (Maddieson & Ladefoged 1985).

	Hani	**Eastern Yi**	**Jingpo**	**Wa**
flow/pressure ratio	lax greater	lax greater	lax greater	lax greater
ratio of F0 to second harmonic (H1-H2)	lax greater	lax greater	lax greater	lax greater
height of F1	lax lower	no difference	no difference	no difference
vowel duration	lax longer	lax longer in falling tone	no difference	lax slightly longer
overall F0	lax slightly lower	lax slightly lower	lax higher[1]	no difference
F0 onset	lax sometimes rising	no consistent difference	lax rising with high tone	no difference
voice onset time (VOT)	no difference	lax somewhat shorter	lax longer	lax longer
Other consonantal properties	Tense: final glottalization	lax: voiced stops less prevoicing		lax: nasals longer

In a landmark study of tense and lax contrasts, Maddieson & Ladefoged (1985) found that the four minority languages of China that they investigated (Yi, Jingpo, Hani and Wa) have different phonetic properties, as shown in Table 2.

Although the consistent difference in the ratio of F0 to H2 and in airflow/pressure confirms that all four languages have a phonation contrast, these contrasts can roughly be

[1] These results for Wa and Jingpo are based on the later report in Maddieson & Hess 1987. The original report shows no difference for Jingpo and lax is slightly higher in Wa.

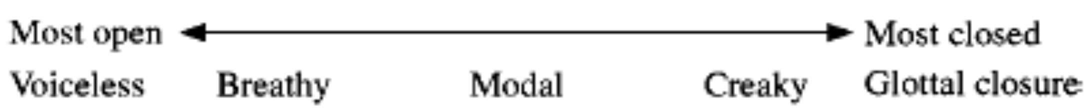
classified into 2 groups. Jingpo and Wa form one group, with longer VOT and higher pitch in lax syllables; the other group, Hani and Yi, has longer VOT and higher pitch in tense syllables. These findings can be related to historical comparisons. The tense/lax contrasts in Wa (Diffloth 1980) and Jingpo (Dai 1979) were initially derived from onset voicing contrasts. The lax syllables initially had voiced onsets. By contrast, the tense/lax contrasts in Hani and Yi were derived from former checked syllables (Dai 1979, Bradley 1979, Wheatley 1982). The tense syllables come from checked codas.

To summarize, tense vs. lax contrasts in southeastern Asian languages are phonation based and originated from at least two different coarticulation processes: vowels can be affected either by the preceding consonants or by the following consonants, and generate different non-modal phonation types. As a result of the original voiced onsets, the lax phonation in Wa and Jingpo is breathy or slack, contrasting with modal as the tense phonation (Maddieson & Ladefoged 1985, Bao1990). In contrast, due to the original checked codas, the tense phonation in Hani and Yi is creaky or stiff, contrasting with modal as the lax phonation (Maddieson & Ladefoged 1985).

1.2.2 Complication in tense vs. lax contrasts and our specific goals

Further investigation found that tense vs. lax vowels in Wa and Jingpo do not have exactly the same mechanism (Maddieson & Hess 1986). For example, phonation has no effect on Wa's pitch, but has a significant effect on Jingpo's tonal pitch. In fact, there is a more essential difference between the two languages. Wa belongs to the Mon-Khmer language family, without tonal contrasts. All the voiced consonants, including all sonorants, have an aspiration contrast, which is the hallmark property of this language (Zhou & Yan 1984). The tense vs. lax contrast can only occur after non-aspirated consonants. Lax vowels cannot occur after aspirated consonants. These facts are the opposite of Jingpo, in which aspirated consonants can only be followed by lax vowels. It might be the case that the phonation contrast in Wa is three-way: aspirated, modal and breathy (or slack) (Zhu 2009). Therefore, the terms tense and lax can ambiguously refer to different phonation types. The ambiguity can be seen as reflecting the relativity of glottal aperture distinctions, as proposed by Ladefoged (1971), Ladefoged & Maddieson (1996), and Gordon & Ladefoged (2001). On this view, some languages have a contrast in a more breathy part of the overall voice quality range while the others have a contrast in a more creaky part of the range (figure 1).

Figure 1. Continuum of phonation types (after Ladefoged 1971).



Maddieson & Hess's work (1986, 1987) provoked further important questions about tense vs. lax contrasts. The Luquan Yi dialect they investigated first is an eastern dialect, which exemplifies a phonation contrast, in that H1-H2 is highly significantly different between tense and lax. Reflecting the checked coda origin, tense syllables are higher in pitch and shorter in duration, and there is no audible vowel quality difference. In contrast, the Liangshan dialect investigated in their subsequent study is a northern dialect.

Contradictory to expectations, they found that H1-H2 in this dialect is not significantly different between tense and lax, while the tense syllables can have a salient “harsh” voice quality. In addition, the language has a clear vowel quality difference between tense vs. lax pairs. These facts raise two issues.

First, is the Liangshan Yi contrast a phonation contrast, but H1-H2 is not the right measure for it? Perhaps the harsh voice is crucial, but H1-H2 does not distinguish this voice quality from others. H1-H2 might also be unreliable because phonation types in Yi mainly contrast in high vowels: the F1 of high vowels is very close to the H2 frequency, which boosts the amplitude of H2. Another possible factor is the tone contrast of Yi: Blankenship (1997) found that H1-H2 does not always distinguish phonation contrasts on mid and low tones. An answer to this puzzle was suggested by Kong (2001), who found evidence for a phonation contrast in Liangshan Yi from spectral tilt measures (i.e. H1-A1, H1-A2). These measures are less sensitive to low values of F1, and Blankenship had found that H1-A2 was a better measure of phonation contrasts on non-high tones.

The second issue is whether the vowel quality difference is a crucial part of the tense/lax contrast, perhaps even more important than the voice quality difference. Esling *et al.* 2000 has claimed that this is the case for (dialect) Yi. Furthermore, they showed that harsh voice can be produced by supraglottal constrictions alone, with tense syllables having a reduced resonating space due to 1) extreme narrowing of the supraglottic tube; 2) tongue retraction over the larynx; 3) larynx raising itself. These gestures could affect F1 frequency as well as voice quality, and thus both the harsh voice quality and the tense vowel qualities could arise from this single supraglottal mechanism. Along the lines of Laver (1981), the auditory voice quality would result from both glottal and supraglottal settings. As in African ATR languages, the pharyngeal cavity change would be key, and this mechanism would provide a path for a sound change from voice quality to vowel quality. This scenario fits well with Edmondson & Esling's (2006:187-88) expanded model in which six major valve mechanisms control articulatory postures in the lower vocal tract[2]. This model not only covers the uni-dimensional glottal stricture continuum as proposed by Ladefoged (1971) (as valve one of the six valve mechanisms), but also emphasizes that supraglottal settings also play important roles in phonation production.

Thus the tense/lax contrasts in Yi dialects seem to vary from pure phonation, to mixed voice/ vowel quality, to pure vowel quality contrasts. Here we examine a new, southern, dialect. All in all, in order to get a better understanding of the articulatory-acoustic coupling in tense vs. lax contrasts, a comprehensive experimental study with both acoustic and physiological analysis is necessary. This is the first and main goal of this thesis.

In addition to investigating the mechanism of tense vs. lax contrast production, we want to know how the tense vs. lax contrast is perceived by native speakers. Perception studies of phonation contrasts have been very few. A recent study (Esposito 2010) found that

[2] These mechanisms are: (1) adduction or abduction of the glottal vocal folds,(2) ventricular incursion, (3) engagement of the aryepiglottic sphinctering mechanism of the laryngeal constrictor, (4) retraction of the tongue and epiglottis in laryngeal constriction, and lingual closure against the posterior pharyngeal wall, (5) raising of the larynx in laryngeal constriction or, conversely, lowering of the larynx, usually associated with nonconstricted contexts and (6) narrowing of the lateral walls of the pharynx, usually associated with laryngeal constriction.

listeners from a phonation contrast language are more sensitive to phonation differences. But it has not been settled which cues/dimensions can possibly contribute to phonation-based tense vs. lax contrasts. Across Yi dialects, tone, vowel quality and phonation have been found to be related to the Yi tense vs. lax contrast, and the register contrast is not evenly distributed across all phonological categories. Therefore, obtaining a perception map of this tense vs. lax contrast under different phonological conditions by an identification perception experiment is important to reveal the mental reality of the phonation contrast for native speakers. This experiment will provide more knowledge about the role of tone and supraglottal settings in the tense vs. lax contrast.

2 Production experiment

This chapter will investigate the physiological and acoustic properties of the phonation contrast in Yi, based on extensive analysis of the speech of 12 native speakers of southern Yi. At the beginning of the chapter, we will briefly review the methods of measuring phonation production, focusing on the use of electroglottography and on the acoustic theories of the production of phonation contrast; the second section of this chapter will introduce the experimental design and statistical models of data analysis. The results section will report both acoustic and physiological measures, evaluating the previous interpretations of three important issues: 1) the relationship between physiology and acoustics in phonation contrasts; 2) the interaction between tone and phonation in a language that has contrasts on both dimensions; 3) the involvement of supraglottal settings. Based on all the data, we will use statistical methods to model the production of the phonation contrast in this language considering all three aspects. With solid statistical models, at the end of the chapter we come back to the research questions: how do phonation types work as a phonemic dimension? Why are there so many varieties across languages? To explore native speakers' phonological knowledge of phonation contrasts, closer investigation of individual varieties will be discussed. We shall argue that there is no fixed articulatory target in phonation contrast production. As a phonemic dimension, phonation types are more relative and more based on auditory effects.

2.1 Measurements of phonation production

2.1.1 Acoustic measures of phonation

In this section, we will review the acoustical parameters involved in phonation production. Up until now, the mechanisms that have been well understood are mainly about the larynx, but much less about supraglottal settings. Summarizing previous literature about the glottis (Stevens 1977, Holmberg 1995, Fant 1976, 1979 (a, b), 1986, Klatt & Klatt 1990, Hanson1997, Gobl & Ni Chasaide 1988), the most basic aspects of glottal activity that can affect phonation production include: 1) ligamental and cartilaginous glottal abduction, 2) open quotient; 3) abruptness and speed of closure; and 4) vocal fold tension. These glottal configurations can be reflected in various acoustic measures, especially in spectral shapes, such as H1-H2, H1-A3, H1-A1, etc., though the precise relations between physiologic and acoustic aspects have only begun to be studied and much remains unclear.

These acoustic parameters have been evaluated from the viewpoint of listeners. For example, Klatt & Klatt (1990) investigated non-contrastive perceptual voice quality based on auditory spectral analysis of synthesized signals. They found that the degree of perceived “breathiness” is affected by several acoustic cues: 1) spectral slope; 2) aspiration noise at higher frequency harmonics; 3) formant bandwidth.

The acoustic measures used in this thesis are discussed below in light of both production and perception studies, in order to capture the whole picture of phonation variation.

H1-H2: The relationship between the amplitude of the fundamental and its second harmonic has been the most widely used measure of phonation contrasts across languages. H1-H2 is correlated with the ratio of the open phase to the entire glottal cycle (the open quotient, OQ hereafter) (Holmberg *et al.* 1995). The following physiological interpretation has been widely accepted: The greater the amplitude of glottal vibration, the greater the H1, and therefore the greater H1-H2. Also, all else being equal, greater vibratory amplitude will mean larger OQ, and when OQ increases, the glottal waveform can more closely approximate a sinusoid of frequency F0, and therefore the amplitude of the first harmonic increases relative to the amplitudes of the higher harmonics. However, the relationship between OQ and H1-H2 is more complicated than that. Hanson (1997) found that H1-H2 is not strongly correlated with other spectral tilt measures, which might suggest that H1-H2 has a distinctive property from other spectral measures. Kreiman *et al.* (2007) also found H1-H2 is statistically independent from other spectral tilt measures. More crucially, the correlation between OQ and H1-H2 can be varied significantly by different voice models and measuring methods. (Henrich *et al.* 2001 and Shue *et al.*

2009a) Therefore, as a keynote of phonation study, the physiological basis of H1-H2 is still open to question.

Perception studies have found that language experience significantly influences perceptual sensitivity to H1-H2 (Kreiman *et al.* 2009, 2010). Interestingly, in addition to the listeners from a phonation contrast language (Gujarati), the listeners from a tonal language without phonation contrast (Mandarin) are also very sensitive to H1-H2. Kreiman *et al.* (2009) propose that it might be due to tonal language speakers paying close attention to F0 (and thus to H1). But in their 2010 paper, they found that Thai listeners (also a tonal language) were more like English, not like Mandarin. So the case in Mandarin might be actually due to allophonic voice quality in this language. All in all, although H1-H2 has been a popular measure for phonation studies, the actual physiological property it reflects and the auditory effects it is responsible for are not settled yet.

H1-A1 (B1): The relationship between the amplitudes of the fundamental and the harmonic nearest the first formant. This measure has been an alternative successful measure to distinguish the phonation types in many languages (Ladefoged 1983, Kirk 1984, Gobl & Ni Chasaide 1992, Bao 1992, Kong 2001). This measure is related to the bandwidth of the first formant (B1). B1 in turn reflects subglottal coupling (Hanson 1996) as well as vocal tract wall impedance (Fant 1976). If a speaker has a posterior glottal opening, then spectral tilt, aspiration noise, and first formant bandwidth will all be increased. Hanson *et al.* (2001) hypnotized that this measure in particular reflects breathiness due to open arytenoids. However, similar to H1-H2, this is still open to question. (Henrich *et al.* 2001 and Shue *et al.* 2009b)

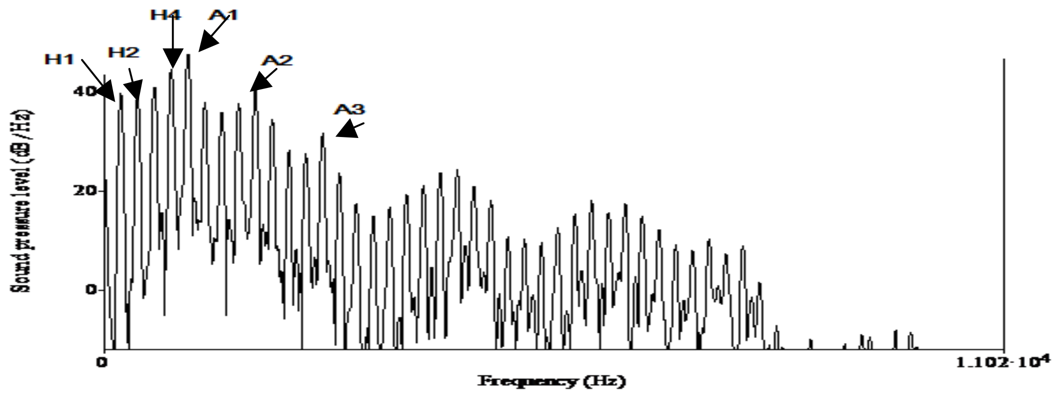
H1-A2: The relationship between the amplitudes of the fundamental and the harmonic nearest the second formant. Blankenship (1997) found that in Mpi, H1-H2 is a more reliable indicator of phonation type for high tone than for either mid or low tone, whereas H1-A2 was more useful for differentiating phonation contrasts in mid and low tone vowels than in high tone vowels.

H1-A3: The relationship between the amplitudes of the fundamental and the harmonic nearest the third formant. Hanson (1997) and Klatt & Klatt (1990) use this measure to represent the overall spectral slope. Stevens (1977) suggested that spectral slope correlates with the abruptness of vocal fold closure. Abruptness affects the high frequency energy in the source. The stronger the glottal closure, the higher the high frequency energy, and thus the shallower the slope. A strong closure usually means a less symmetrical pulse since the opening is usually more gradual, while a smoother closure usually means a more sinusoidal signal.

H2-H4: The relationship between amplitudes of the second harmonic and fourth harmonic. This is a new spectral measure. Kreiman *et al.* (2007) found that H2-H4 is one of the four principle factors accounting for spectral variability.

Figure 2 indicates the locations of H1, H2, H4, A1, A2 and A3 in a harmonic spectrum

Figure 2. Measuring H1, H2, H4, A1, A2 and A3.



So far, none of the measures directly reflect noise in the spectrum, which is also a very important property of voice quality, especially for breathy voice (Klatt & Klatt 1990, Hillenbrand *et al.* 1994, Kreiman *et al.*2007, Esposito 2010). Although breathiness is a relative property, increased spectral noise, particularly at higher frequencies, can characterize breathy phonation. Creaky phonation is also characteristically associated with aperiodic glottal pulses, which introduce noise into the spectrum.

The measure of aperiodicity in this thesis is cepstral peak prominence (**CPP**). According to the review by Blankenship (1997), a cepstrum is an inverse spectrum generated by taking the FFT of the log magnitude values of a power spectrum. The spectrum of a highly periodic signal shows well-defined harmonics; its cepstrum has a prominent peak at a location corresponding to the duration of the F0 cycle. Less periodic signals such as those often produced in breathy or creaky phonation have a spectrum with less defined harmonics, resulting in a cepstrum with a low peak. Hillenbrand *et al.* (1994) proposed the CPP measure and claimed that it is mostly responsible for American English listeners' ratings of perceived breathiness. Esposito (2009) found CPP is important for distinguishing breathy phonation from creaky in White Hmong, while Esposito 2010 showed that English listeners use is as a perceptual cue for distinguishing modal from breathy

Table 3 is the summary of a comparison of 8 measures for modal vs. breathy phonations across languages by Esposito (2006), and the checks indicate which measures were significantly different in which languages:

Table 3. Measures across languages (cited from Esposito 2006) (uncorrected spectral measures, low vowels).

Languages and/or dialects	Measures							
	CPP	H1-H2	((H1+H2)/2) -A1	H1-A1	H1-A2	H1-A3	H2-H4	A2-A3
Chong	✓	✓	--	✓	✓	✓	✓	✓
Fuzhou	✓	✓	✓	✓	✓	--	✓	✓
Green Hmong	✓	✓	--	--	--	✓	--	--
White Hmong	✓	✓	--	--	--	✓	--	--
Mon	✓	--	✓	--	✓	✓	✓	--
SADV Zapotec	✓	✓	--	✓	✓	✓	--	--
SLQ Zapotec	✓	✓	--	✓	✓	✓	✓	--
Tlacolula Zapotec	✓	✓	✓	✓	✓	✓	--	--
Tamang	✓	--	--	--	✓	✓	--	✓
!Xóõ	✓	✓	--	✓	--	✓	--	--

Other measures:

Energy: Gordon & Ladefoged (2001) conclude that breathy phonation is associated with a decrease in overall acoustic intensity in many languages. Creakiness also triggers a reduction in intensity.

F0: Generally, non-modal phonation is associated with pitch lowering effects (Gordon & Ladefoged 2001). But the relationship is not consistent. Creaky phonation has been found to be responsible for F0 increase in many languages (Hombert *et al*. 1979, Maddieson & Ladefoged 1985, Kong 2001). The effect may be related to raising or lowering the larynx in different phonation types, or to differences in vocal fold tension. Non-modal phonation is sometimes used in tonal languages as an enhancing cue of a tonal contrast. For example, creakiness is a very important cue for Mandarin Tone 3 (Davison 1991, Belotel-Grenié & Grenié 2004, Yu 2010). In languages with both tonal and phonation contrast, the interaction is still not well understood. Within the Zapotecan languages, San Lucas Quiaviní Zapotec shows a correlation between tone and phonation, whereas in San Juan Guelavía Zapotec tone and phonation are completely independent (Esposito 2005).

F1: The frequency of the first formant. This measure can directly indicate a change in vocal tract shape. Many studies have shown significantly different F1 values between phonation types. Swerdlin *et al.* (2010) found that the voice source affects vocal tract resonances. Specifically, creaky voice had a small effect on F1 (about 45 Hz,), while voiceless aspiration a large effect (about 225 Hz). Kirk (1993) found that in Mazatec F1 is higher during creaky phonation, presumably due to a raising of the larynx and concomitant shortening of the vocal tract during creaky voice. However, this is not supported by a recent larger study of Kirk's recordings (Garellek & Keating 2010). Maddieson & Ladefoged (1985) also found that F1 is higher for tense vowels in Hani. Thongkum (1987) reports that breathiness is associated with a lowering of F1 in Chong. Samely (1991) also found that breathy vowels have lower first and second formant values than modal vowels in Kelang. Maddiesson & Hess (1986) and Kong (2001) found a salient F1 difference in the tense vs. lax contrast in northern Yi.

It is important to remember that the reliability of spectral measures can be affected by the vocal tract transfer function (Ni Chasaide & Gobl 1997): The comparison of H1 and H2 levels may be a valid measure when F1 is high and F0 low, but when F1 is low or F0 is high (or both), the levels of H1 or H2 may be boosted depending on their proximity to the F1 peak. Therefore, Hanson (1997) first proposed corrected harmonic amplitude measures to remove the amount of amplitude boost by the first formant from lower harmonics. This method improves the accuracy of spectral measures and makes possible the comparison across vowels and speakers. Iseli *et al*. (2007) extended the harmonic correction to use F2 and F3 and the formant bandwidths. This thesis will present the corrected version of spectral measures, which will be marked with asterisks. All the acoustic measures were made using the program VoiceSauce (Shue *et al*. 2009a).

2.1.2 EGG measurements of phonation

Electroglottography (EGG) is an ideal method to measure variations in the vocal fold contact area during phonation thanks to its easiness and non-invasiveness. 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.

The parameter, which can reflect the duration of the vocal fold contact during each single vibratory cycle, is often known as the "contact quotient" (CQ) (Rothenberg 1988). (In some studies, CQ is referred to as "closed quotient" (Nair 1999).) The contact quotient is defined by comparing the duration of the contact phase to the period of the vibratory cycle. During the last 15 years, several methods of calculating the CQ have been developed. See Henrich *et al*. (2004) and Herbst & Ternström (2006) for detailed reviews.

The essential difference among the methods is the way to define the glottal opening and closing instants. Here is a brief summary based on Henrich *et al.* (2004):

1. EGG threshold: Developed by Rothenberg (1988). The contact event is defined as the time point when the signal strength exceeds a certain threshold level, which is usually indicated as a percentage of the peak-to-peak amplitude (CQ method in Figure 3) Levels between 20% ~50% have been used in studies, depending on the target phonation. A higher threshold is better for creakier voice and a lower threshold is better for breathier voice. Thresholds at 20% and 25% are found to be best correlated with the Contact Quotient obtained by videokymographic imaging (Herbst & Ternström (2006)).

2. DEGG: Proposed by Henrich *et al.* (2004). This method detects the contact and opening events relying on peaks in the derivative of the EGG signal (CQ_PM method in

Figuire 3). The DEGG algorithm correlates with Open Quotient measurements derived from the inverse-filtered glottal flow. The contact event is defined by the strong positive peak and the opening event is defined by the weak negative peak. However, comparison study with Photoglottography (PGG) signals (Baer *et al.*1983) has pointed out that there is no identical correspondent glottal opening time between PGG and EGG waves, since glottal opening is more gradual. And the mismatch is especially worse for female voices. Given that, the main problem of the DEGG method is the accuracy of the opening time.

3. DEGG + threshold. The hybrid combination of the above two methods (CQ_H method in Figure 3). Howard (1995) used the DEGG contacting peak for detecting the glottal contact event, and an EGG-based 3/7 threshold for detecting the glottal opening event. The threshold was set at a level of 25% by Orlikoff (1991). (A later version of this method by Tehrani (ref. documentation of EggWorks) proposes a new threshold for the opening event, which is the y-value of the DEGG contacting peak, CQ_HT method in Figure 3)

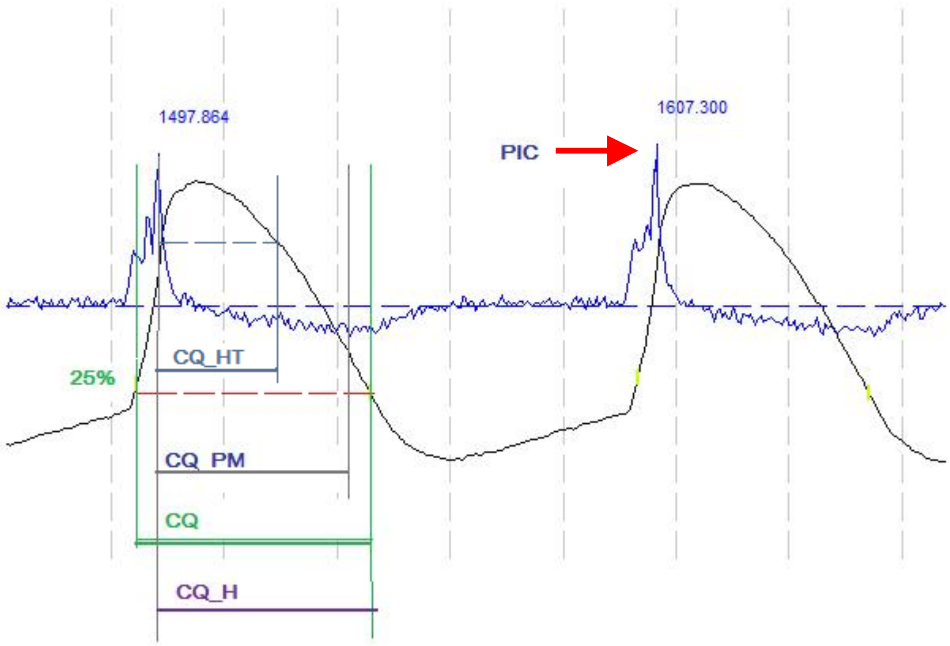
The EGG analysis in our study is done by EggWorks, a free program developed by Henry Tehrani in UCLA Phonetics Lab. The outputs of the program include all the different methods of measuring: CQ by the threshold method, CQ_PM by the DEGG method, CQ_H by the Howard's hybrid method (but using Orlikoff's 25% threshold), and CQ_HT by the Tehrani's hybrid method. Different CQ measures are displayed in Figure 3.

Recently, a new measure has been employed by Michaud (2004), related to earlier measures of average rate of change in increasing contact (see Baken & Orlikoff 2000 for review). Derivative-EGG Closure Peak Amplitude (DECPA) is the amplitude of the positive peak on the DEGG wave, corresponding to the highest speed in increase of vocal fold contact, which is thought to be reached at the glottis-closure-instant.

"Peak increase in contact" (PIC) is a more transparent name for this measure (Keating *et al.* 2010). Michaud (2004) found that prosodic accent is correlated with the maximal PIC reached. F0 and phonation should both affect PIC: a relatively high PIC value is expected to be a significant cue for creaky voice; extra high F0 is expected to show the lowest PIC. However, a phonation study of Hmong (Esposito *et al.* 2009) found the opposite. They found that PIC can distinguish breathy voice from non-breathy voice in the earlier half of the syllables and distinguish creaky from non-creaky in the last 4/9 proportion, but breathy voice has higher PIC whereas creaky voice has lower PIC. The physiological basis as well as acoustic correlates of the new measure still needs better understanding.

Four CQ measures plus PIC measure are visually presented in Figure 3.

Figure 3. EGG measures in EggWorks. Both EGG and DEGG signals are displayed here. CQ is the 25% threshold method; CQ_H and CQ_HT both hybrid method: using DEGG to find the closing phase and threshold to define open phase, CQ_H uses a 25% threshold while CQ_HT uses same y-value of closing phase. CQ_PM is the DEGG method, both closing phase and opening phase are defined by DEGG signal. PIC is the y-value of the positive peak of DEGG, indicated by an arrow.



2.2 Materials

2.2.1 Speakers

All the data in this study were obtained during a trip to Yunnan province of China in summer 2009. I visited the villages of Xinping and Jiangcheng, and made recordings from six native speakers (three males and three females) per village. Xinping is close to Kunming, which is located in the northeast of Yunnan; Jiangcheng is close to the border with Vietnam and Laos, a very southern area of Yunnan. Speakers from Xinping are all around 40 ~50 years of age, using Yi as the primary language in everyday communication. Speakers from Jiangcheng vary more, from 18 to 60 years of age. Yi is also the primary language in the village. They all can also speak Southwestern Mandarin to some extent, as it is the common language among groups of people in Yunnan.

2.2.2 Recording Material

The fieldtrip consisted of three stages: word collection, production recording and perception experiment. The word list used in this thesis was made in the first stage. To start, 2000 words were elicited from speakers and archived in Excel as a small lexical database. These words covered things and events in everyday life, and had been used in many fieldworks for other Yi dialects. 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 was made for the purpose of this phonation contrast study (see Appendix 1). In Yi, phonation contrasts do not occur with high tone. While it would be an interesting question to determine the phonetic nature of the neutralized phonation that occurs with high tones, in this study high tone words were excluded. In that way unbalanced data structures were avoided, making the statistical analysis easier.

2.2.3 Procedures:

For all 12 speakers, both 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 sampling rate of 22050 Hz per channel. The audio signal was the first channel. EGG data were obtained by a two-channel electroglottograph (Model EG2, Glottal Enterprises) and recorded as the second stereo channel.

Before making the recording, the speakers were asked to go over the word list, checking the contrasts in the minimal pairs. Non-contrastive pairs were excluded from the data analysis. The speakers were wearing the EGG device and a microphone at the same time when they read the word list. Each word was repeated twice.

2.3 Statistical method

Table 4 is the summary of measures analyzed in the thesis.

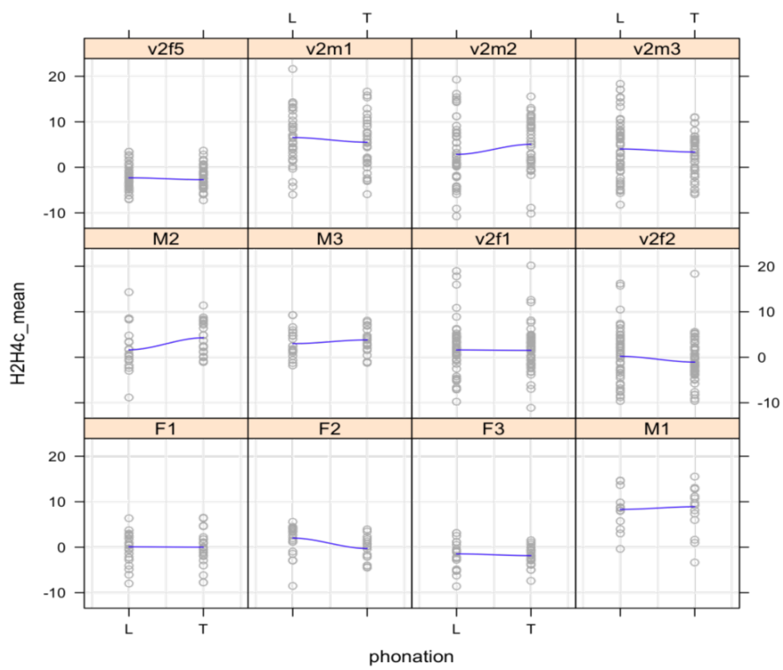
Table 4. Summary of measures analyzed in this thesis.

Spectral tilts	H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*
Amplitudes of individual harmonics	H1*, H2*, H4*, A1*, A2*, A3*
Formants and bandwidths	F1, F2, F3, B1, B2
Pitch, periodicity, intensity	F0, CPP, energy
EGG measures	CQ, PIC

For each measure, VoiceSauce extracts the overall mean value for a segment as well as the average value for each of nine time intervals. All statistical analysis is based on the overall means; the values of the nine intervals are used only for plotting the contours of measures along their time course.

Since we only have 12 speakers, which is a small number, the individual differences will affect the general result significantly. Different kinds of linear mixed effect model (West, 2006)- random intercept models and random coefficient models-were tested. The pitch range of tones and the range of voice quality differ across speakers, not just in intercept but also in slope (the comparison of means in two categories). (See Figure 4.)

Figure 4. Example panel plot for phonation contrast in individuals.



Hence, a random coefficients model was employed to model our dataset in this study. In this random coefficients model, both gender and phonation category have been specified as fixed effects, and speaker has a random effect on both intercept and slope. ANOVA (Chambers *et al.* 1992) was used to compare the goodness-of-fit of different mixed models; better models can explain more variance and also better fit the data (i.e. Lower AIC, BIC, higher Loglik).

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. In this section, we report statistical significance by exact student t value and its p-value based on the upper bound degree of freedom. (Bates & Maechler 2010, Baayen 2010)

Finally, a forward stepwise logistic regression model was utilized to evaluate the independent contributions of different measurements to tone or phonation differences. (the backward stepwise method does not work for the data, since it kills the variables which are most correlated with the best contributing variable.) The quantity – log10 (p-value) was used as an indicator of this contribution.

In the logistic regression model (Hosmer 2000), I did not include the random effects in the syntax of the models. The reason is that Mixed-effect Regression requires a larger dataset than regular regression. If there are not enough data points, the models will not be stable. We found that Mixed-effect models fail to converge for the data, failing to give stable predictions. Since here the regression models only serve as additional evidence, which can be compared with mixed-effect models, I decided to use regular logistic regression models.

Goodness-of-fit of logistic regression is evaluated by AUC (Hosmer & Lemeshow 2000, pp. 162). AUC is the area under the receiver-operating characteristic curve, formed by taking the predicted values from the regression model as a diagnostic test for the event in the data. The minimum value is 0.5; the maximum is 1.0. The rule of thumb is that an AUC value between 0.7 to 0.8 shows acceptable discrimination, values of 0.8 to 0.9 indicate excellent discrimination, and values >=0.9 to show outstanding discrimination.

2.4 Results

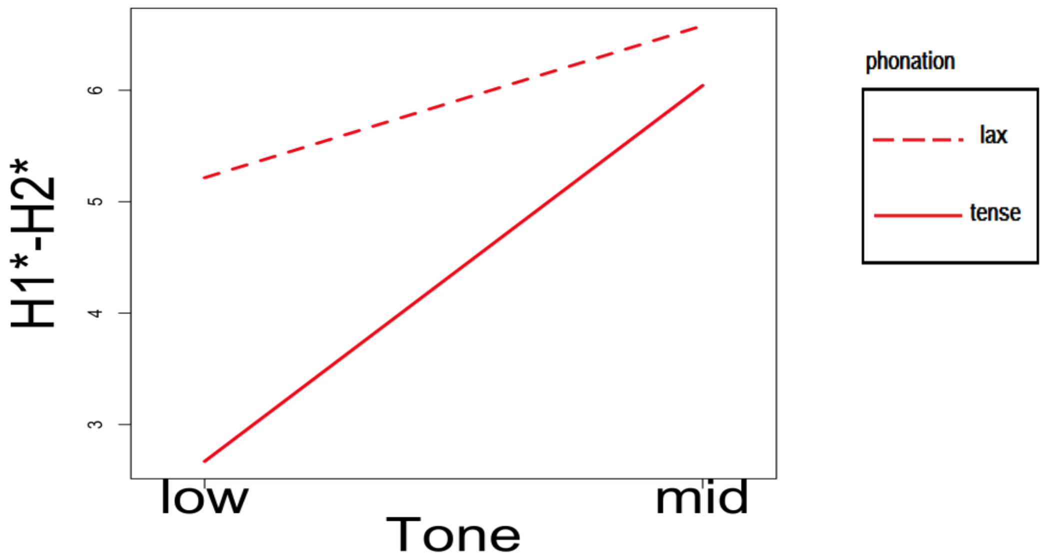
Preliminary analyses showed that there is no main effect of village, so data from the two villages are combined in all subsequent analyses.

2.4.1 Spectral tilt measures:

(1) H1*-H2*

Main effects of both tone ($t= 2.06$, $p=0.04$) and phonation ($t= -2.95$, $p =0.003$) are found. H1*-H2* is overall higher in the lax phonation and the mid tone. The interaction between tone and phonation shown in Figure 5 is also significant ($t= -2.64$, $p=0.008$). A pairwise *post hoc* test for H1*-H2* shows that each of the four tone × phonation combinations is distinctive from the others ($p<0.01$). There is no main effect of or interactions with gender. Figure 5 plots the 2-way interaction of phonation by tone for H1*-H2* (the dashed line indicates the lax phonation). It appears that the lower tone has a larger phonation contrast than the higher tone.

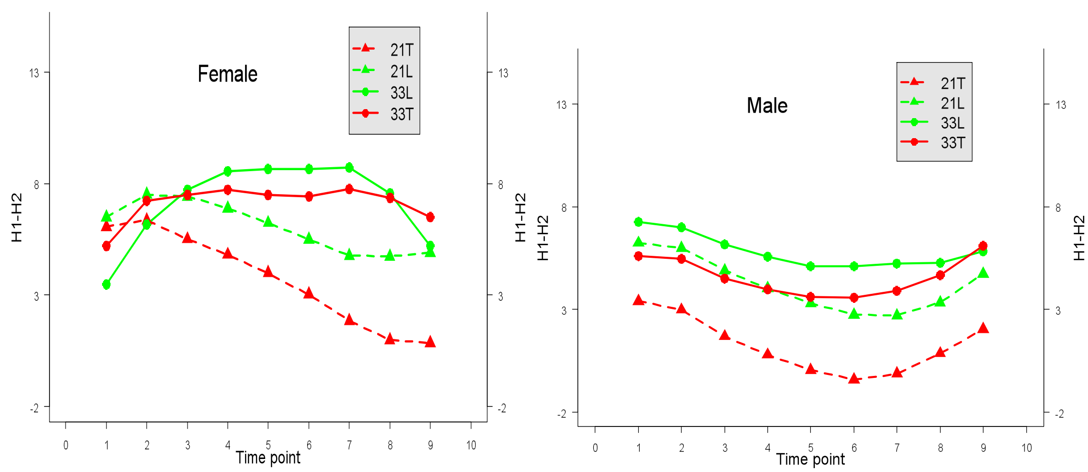
Figure 5. 2-way Interaction plot for H1*-H2*. Line type shows phonation.



As expected, larger H1*-H2* values of the lax phonation indicate that the lax phonation is breathier than the tense phonation, possibly reflecting a longer open quotient or more gradual closure of vocal folds (Holmberg *et al*. 1995). Moreover, the relationship between the first two harmonics is sensitive to tones. As indicated in Figure 5, the higher tone has overall higher H1*-H2*. This suggests that the higher tone is breathier than the lower tone.

Although the statistics show no main effect of gender on the overall mean values, the temporal plots suggest a slight difference in the contour shape. As shown in Figure 6, male speakers have a dipping point at the sixth time point. Female speakers have an earlier turning point at the second time point. Female speakers also appear to have a stronger tone effect than male speakers.

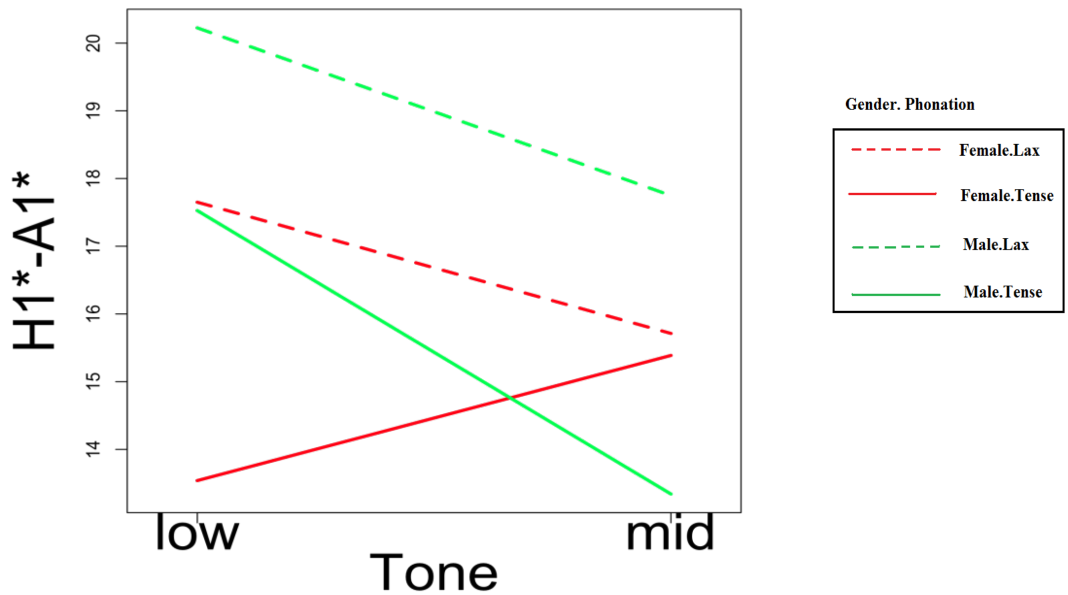
Figure 6. Temporal change of H1*-H2*. Color indicates gender and Line type shows phonation.



(2) H1*-A1*

Only phonation has a main effect on H1*-A1* ($t= -7.32$, $p<0.001$). The lax phonation has overall higher H1*-A1*. There are no main effects of gender or tone. The interaction between phonation and tone is significant ($t= -2.192$, $p =0.03$), and so is the three-way interaction of phonation by tone by gender ($t= 2.17$, $p =0.03$). Pairwise *post hoc* tests show that different phonation types are well distinguished by H1*-A1*; tones are only significantly different by H1*-A1* when the phonation is lax, but not when the phonation is tense. Figure 7 shows the complicated three-way interaction of phonation by gender by tone for H1*-A1*. Females have a more distinctive phonation contrast in the low tone, while for males, a more distinctive phonation contrast appears in the mid tone.

Figure 7. 3-way interaction plot for H1*-A1*. Color shows gender and line type shows phonation.



Larger H1*-A1* as well as H1*-H2* values in the lax phonation demonstrate that the breathier lax phonation has a more dominant H1 in the power spectrum. But reverse to the trend for H1*-H2*, H1*-A1* of lax phonation is overall lower in the higher tone. The other remarkable difference from H1*-H2* is that H1*-A1* has different tone and phonation interaction between genders. Tone affects tense phonation for both males and females, but in opposite directions – females have a less distinctive phonation contrast in the mid tone while males have a less distinctive phonation contrast in the low tone. The different effects cancel each other out, which leads to no main effect of tone on tense phonation.

These differences support Hanson (1997)'s claim that H1*-H2* and H1*-A1* are independent from each other. She hypothesized that H1*-A1* reflects the bandwidth of the first formant (B1), which might in turn reflects posterior opening of the vocal folds. The above results can be cross-referenced with the B1 data below (see section 2.4.5).

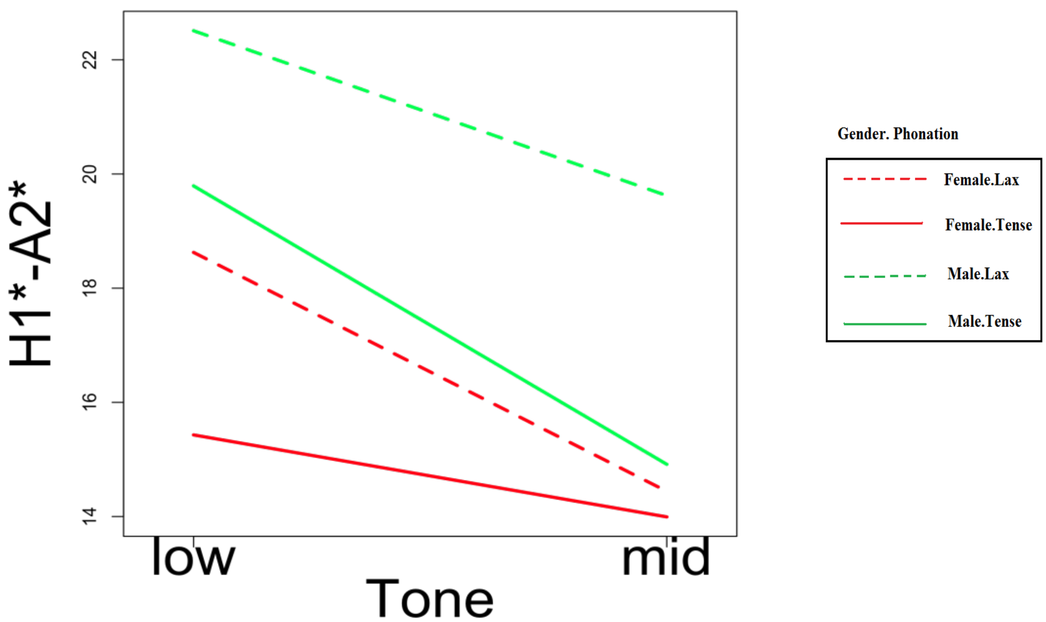
Since the measurement accuracy of H1*-A1* heavily relies on formant tracking, in order to verify the correctness of the results presented here, we randomly pulled out a small number of sound samples and manually checked the formant and bandwidth measures in Praat. The values from Praat confirm the above results (see Appendix 2).

(3) H1*-A2*

Phonation ($t= 23.72$, $p<0.001$), tone ($t= 14.18$, $p<0.001$) and gender ($t= 8.58$, $p<0.001$) all have main effects on H1*-A2*. Consistent with the previously discussed spectral tilt measures, the lax phonation has higher H1*-A2* values than the tense phonation. Distinctively from the other measures, however, the main effect of gender is significant for H1*-A2*. Female speakers have overall lower H1*-A2* than male speakers. This measure is also sensitive to tone, with the higher tone having higher H1*-A2* values.

The interaction between tone and phonation is significant ($t= 2.13$, $p =0.03$), as is the three-way interaction of tone by phonation by gender ($t=-2.19$, $p=0.03$). Pairwise *post hoc* tests show that H1*-A2* can differentiate all phonation by tone combinations. Figure 8 shows the interaction of phonation by tone by gender for H1*-A2*. Female speakers have a more distinctive phonation contrast in the low tone, while male speakers have a more distinctive phonation contrast in the mid tone.

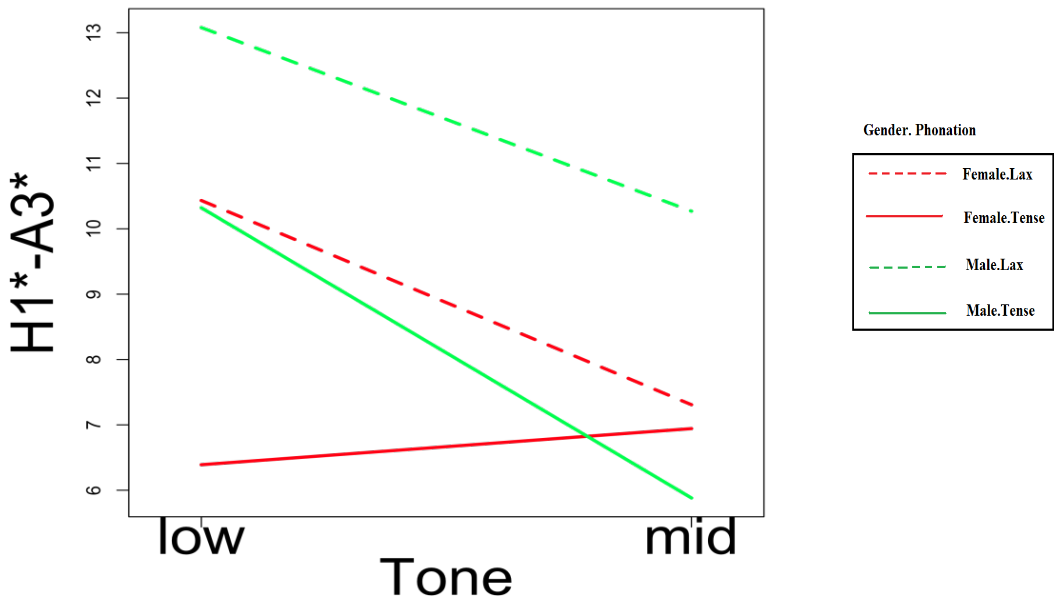
Figure 8. 3-way interaction plot for H1*-A2*. Color shows gender and line type shows phonation.



(4) H1*-A3*

Main effects of phonation (t= 19.28, p <0.001) and tone (t=7.43, p <0.001) are significant. As expected, lax phonation has higher H1*- A3* than tense phonation. The higher tone generally has lower H1*-A3*. The interaction between phonation and tone is also significant (t=-2.40, p =0.02), as is the three-way interaction of phonation by gender by tone (t=2.03, p=0.04). Pairwise *post hoc* tests show that H1*-A3* cannot distinguish low vs. mid tones with tense phonation. This is similar to H1*-A1*, with the tonal effect only reaching significance in the lax phonation. The three-way interaction (Figure 9) is caused by the opposite performance of the two genders in the tense phonation: females have more distinctive phonation contrast in the low tone, while males do so in the mid tone.

Figure 9. 3-way interaction plot for H1*-A3*. Color shows gender, line type shows phonation.

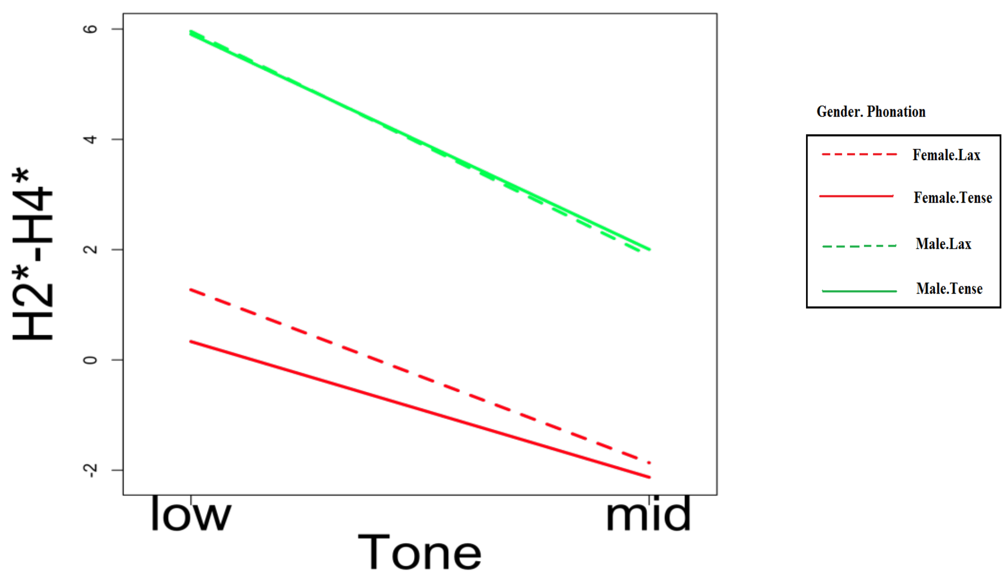


H1*-A3* is believed to be correlated with the abruptness of vocal fold closure (Stevens 1977). The stronger the closure, the greater the energy in the high frequency range. Therefore, higher H1*-A3* may suggest a more gradual closure.

(5) H2*-H4*

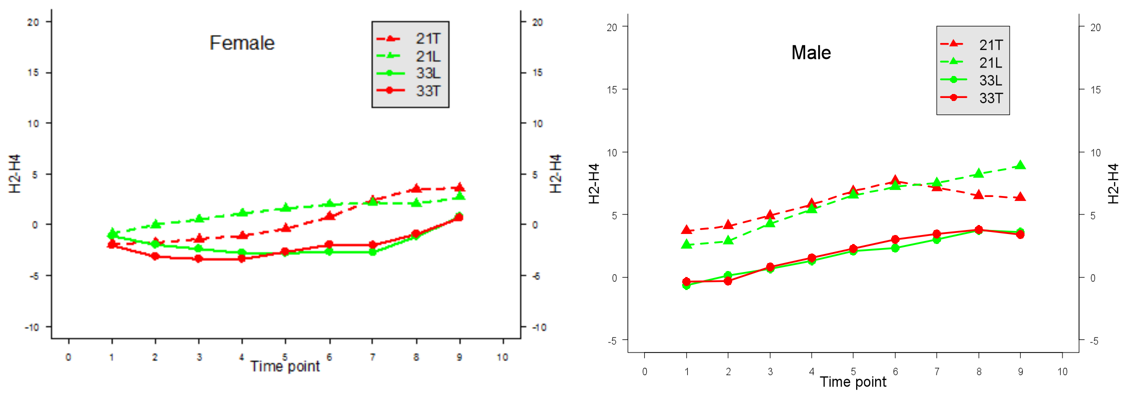
This new measure has a very distinctive performance from the other spectral tilt measures. The main effect of tone is highly significant (t= -3.29, p =0.001), with H2*-H4* lower in the mid tone. The main effect of gender is also strong for H2*-H4* (t=3.6, p<0.001), with males having overall higher values. However, there is neither a main effect of phonation nor an interaction between tone and phonation. So this spectral tilt measure apparently has nothing to do with the phonation contrast, but only with tone and gender. The interaction plot (Figure 10) shows the relationships among the three factors; it can be seen that the tense and lax phonations are not distinct. Further analysis shows that a significant negative correlation is found ($r=-0.5$, $p<0.001$) between F0 and H2*-H4* across all speakers.

Figure 10. 3-way Interaction plot for H2*-H4*. Color shows gender and line type shows phonation.



Temporal plots (cross-reference to Figure 11 and Figure 13) can further illustrate the correlation between H2*-H4* and F0. H2*-H4* is generally a rising contour, while F0 falls in the low tone.

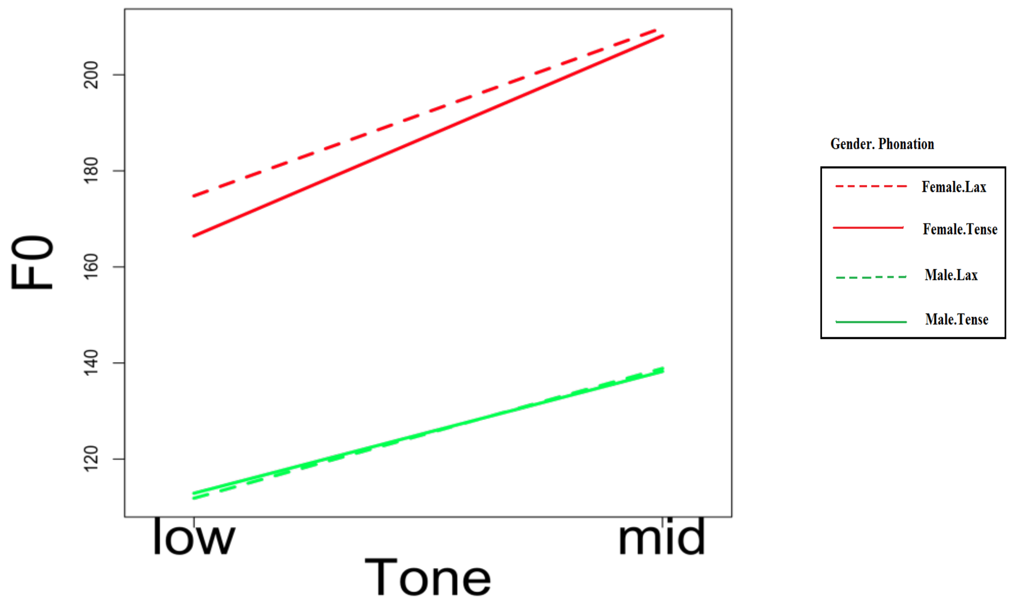
Figure 11. Temporal change of H2*-H4*. Color shows gender and line type shows phonation.



2.4.2 Pitch (F0)

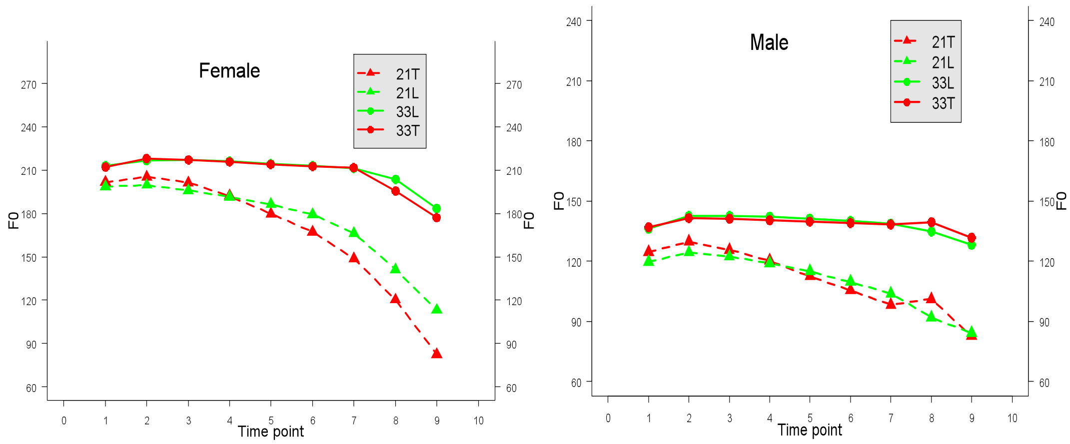
As expected, strong main effects of tone ($t= 5.57$, $p <0.001$) and gender ($t= 4.63$, $p<0.001$) are found for F0, but there is no main effect of phonation, and no interaction between phonation and tone. Figure 12 shows the interaction of phonation by gender by tone for F0.

Figure 12. 3-way interaction plot for F0. Color shows gender and line type shows phonation.



The contours of the two tones are plotted along time in Figure 13. It can be seen that the low tone is a falling tone. The onsets of the mid and low tones are very close, and they are mainly distinct in their offsets.

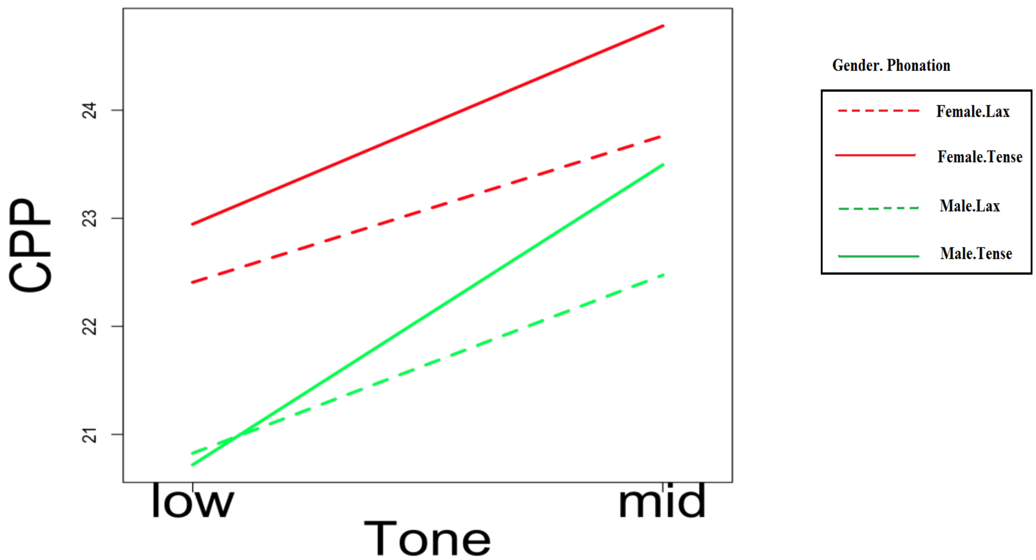
Figure 13. F0 contours of mid and low tones. Color shows gender and line type shows phonation.



2.4.3 Periodicity (CPP)

Phonation (t= 2.57, p =0.01), tone (t= 2.14, p=0.03), and gender (t= 3.7, p<0.001) all have significant main effects. CPP values are larger in the tense phonation, the mid tone and female speakers. But there is no interaction between factors. The interaction plot (Figure14) illustrates all three main effects.

Figure 14. 3-way interaction plot for CPP. Color shows gender and line type shows phonation.



CPP reflects the harmonic to noise ratio in the spectrum. It is expected that, if the lax register is breathier, then it would have a larger ratio of aspiration noise and thus have smaller CPP values. The pattern shown in Figure 14 supports this interpretation.

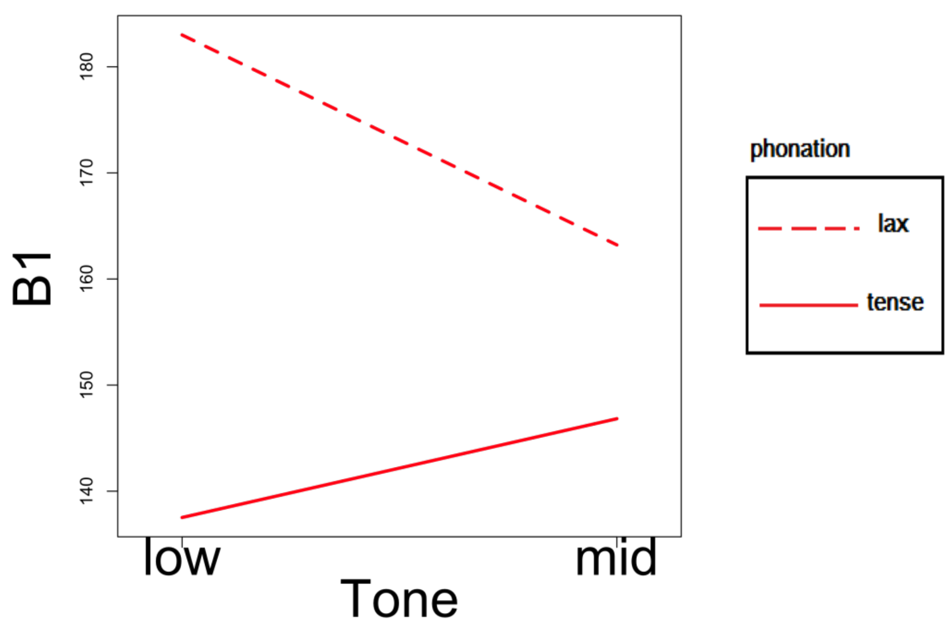
2.4.4 Energy

Only phonation has a weak main effect on Energy (t=2.2, p=0.03). As expected, tense phonation has stronger intensity than lax phonation. There is no interaction among the factors.

2.4.5 Bandwidth of the first formant (B1)

The main effect of phonation is highly significant for B1 ($t=9.8$, $p<0.001$), with the lax phonation having overall larger B1 values. Consistent with the result for H1*-A1* presented earlier, the two-way interaction between phonation and tone ($t=3.54$, $p<0.001$) is significant. Pairwise *post hoc* analysis shows that the phonation effect solely exists in the low tone, and B1 cannot distinguish tense mid from lax mid. This pattern is also illustrated in the interaction plot (Figure 15).

Figure 15. 2-way interaction plot for B1. Line type shows phonation.



Since B1 is thought to reflect the posterior opening in the glottis, it is expected that the breathier voice will have larger B1 values than the creakier voice, which is supported by our data. The loss of this contrast in the mid tone might be attributed to more tension in the vocal tract. See Appendix 2 for B1 values measured in Praat.

2.4.6 EGG measurements:

(1) Contact Quotient

EggWorks provides us with CQ measurements from four methods; therefore, before presenting the data, a brief measure comparison is necessary. To do this, small amounts of data are randomly pulled out, and a mixed-effect model is run to examine the main effect of phonation for the different CQ values. The statistical parameters of AIC, BIC, log likelihood and variance are considered as the criteria of goodness-of-fit of models. A good model would show small AIC and BIC values but a large value of log likelihood.

Table 5. Model comparison of four methods.

Methods	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
CQ_H	10	-2018.69	-1971.86	1019.35	NA	NA	NA
CQ_PM	10	-1716.87	-1670.03	868.43	0	0	1
CQ	10	-1930.45	-1883.61	975.22	213.58	0	<2e-16 *
CQ_HT	10	-1741.91	-1695.07	880.95	0	0	1

From Table 5, we can tell that all the models are almost equally good, except for the traditional CQ measure. Given that CQ_H shows smallest AIC and BIC but biggest log likelihood value, CQ_H is proven to be the best model.

We also compare the CQ values extracted by the four methods in Figure 16:

Figure 16. CQ value comparison across four measures.

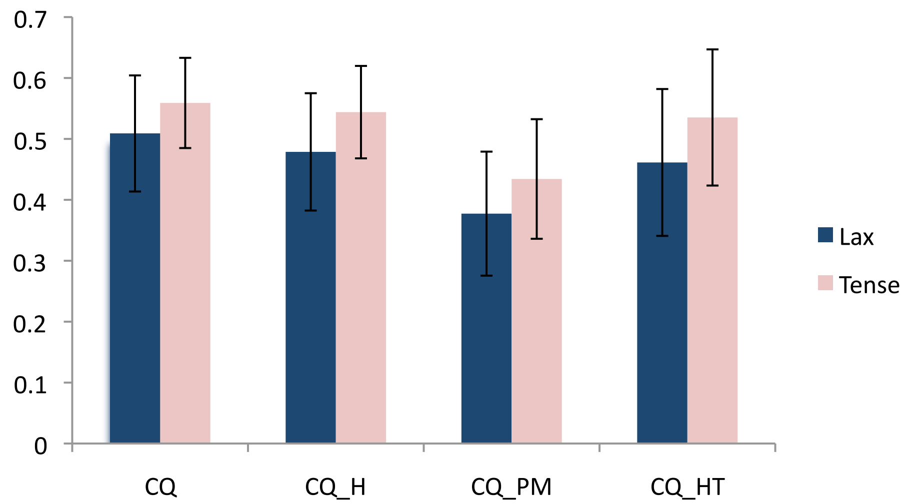
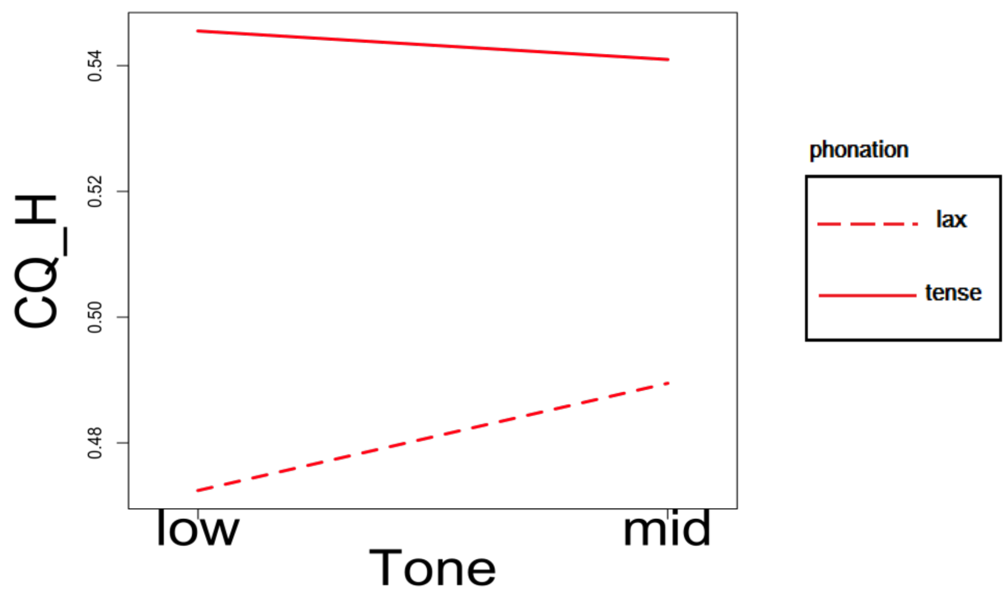


Figure 16 exhibits the consistent pattern that tense phonation has overall higher CQ than lax phonation, although the actual values slightly differ across methods. This means all the measures are successful; indeed they are almost equally good, at capturing the phonation contrast. Although model comparison slightly suggests that the CQ_H method best differentiates the two phonations, there is no significant difference in explaining the variance of the data among the methods (CQ and CQ_H have relatively smaller deviations). Therefore, it is safe to report the results from any one of the methods. This study will present CQ_H values as the measure of Contact Quotient. (Hereafter all the "CQ" results refer to CQ_H.)

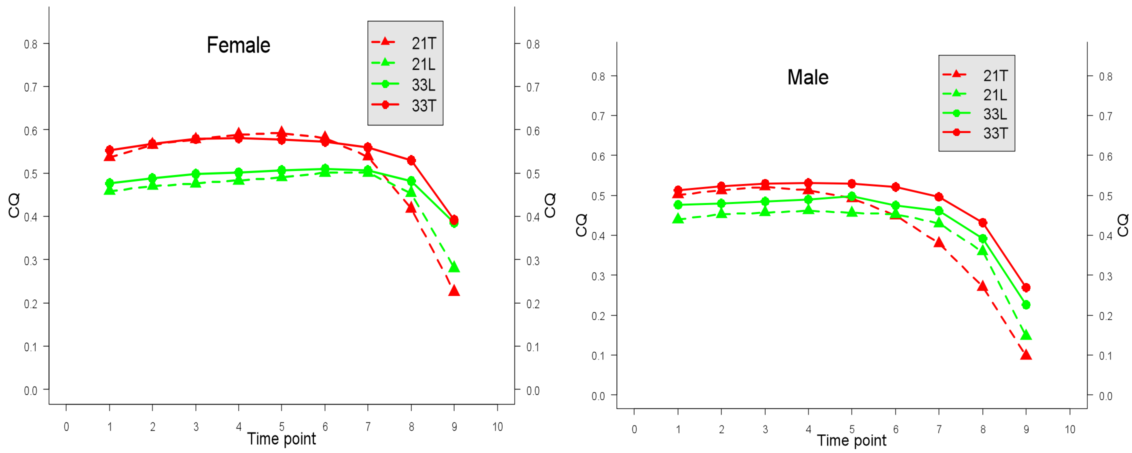
A strong main effect of phonation is found for CQ ($t=18$, $p<0.001$). Contact quotient is much higher in the tense phonation. But there are no main effects of tone or gender, and no interaction between tone and phonation. Figure 17 shows the interaction of phonation by tone for CQ, where only the phonation effect is significant.

Figure 17. 2-way interaction plot for CQ. Line type shows phonation.



Temporal contours (Figure 18) show an overall falling trend (that is, all vowels become breathier), with phonations better separated at the beginning portion of the syllable. Close scrutiny reveals that the contrast is kept until the 7th of the 9 intervals.

Figure 18. Temporal change of CQ. Color shows gender and line type shows phonation.



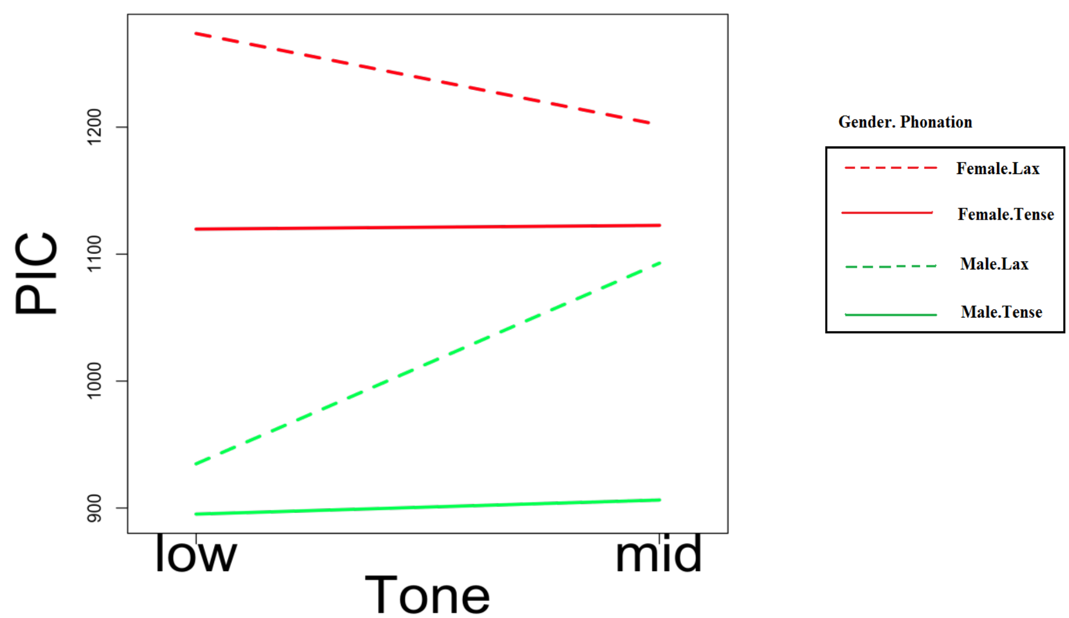
(2) Peak Increase in Contact (PIC)

A significant main effect of phonation is found (t=5.8, p<0.001). Lax phonation has larger PIC values than tense phonation. The direction is the same as in Hmong (Esposito 2010), which also has a phonation contrast; but the opposite of Michaud (2004), a study of prosodic prominence (if prominence is considered to be tense phonation).

There are no main effects of tone or gender, but a significant interaction is found between gender and tone (t=2.63, p=0.008). The three-way interaction of phonation by gender by tone is also significant (t=-2.24, p=0.02). Pairwise *post hoc* analysis shows that PIC is well distinguished between the different phonation types, but there is no difference between 21T and 33T. The tonal effect only reaches significance for lax phonation. As indicated in Figure 19, similar to H1*-A1* and H1*-A3*, the relationship among the factors is essentially a three-way interaction of phonation by tone by gender for PIC.

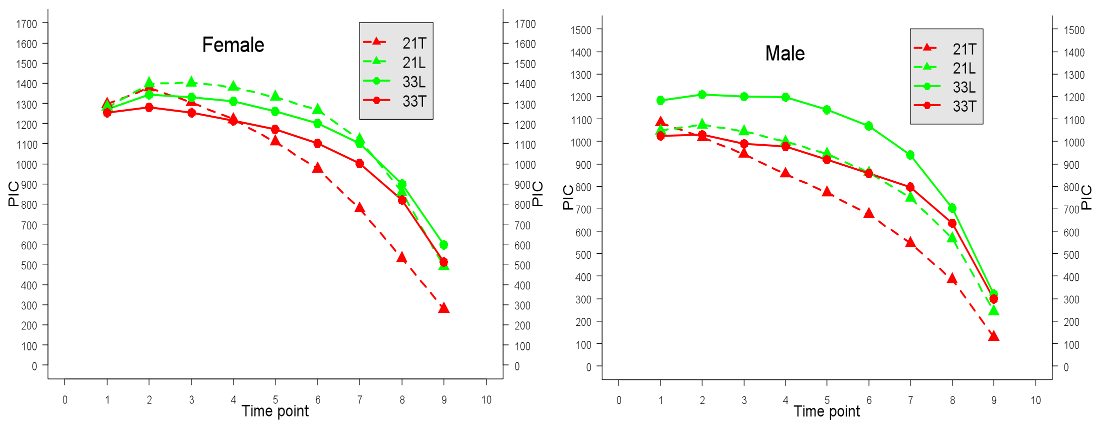
Only lax phonation is sensitive to tonal change, and the influence is opposite for male and female. PIC becomes lower in higher tone for female speakers, but goes higher in higher tone for male speakers. The direction for males is the opposite of Michaud (2004)'s prominence comparison. The different results might be due to different ranges of F0 of the sound samples. Michaud (2004) discussed more extreme voice qualities, such as low tone fry and super high-F0 voice. Results from Yi and Hmong are more concerned with the normal range of voice.

Figure 19. 3-way interaction plot for PIC. Color shows gender and line type shows phonation.



Temporal contours (Figure 20) indicate that the PIC keeps dropping from the beginning. PIC is more distinctive in the later portion of the syllable. Tense phonation is differentiated from lax phonation mainly by an overall lower peak increase in contact and the steeper slope of the drop in these values.

Figure 20. Temporal change of PIC. Color shows gender and line type shows phonation.



Comparison of physiological mechanisms in phonation contrast:

To better understand the articulatory movements involved in the phonation contrast, sample EGG signals are presented here (male, mid tone). CQ and PIC values are also displayed in the plots. Please note that the y-axis scaling, which is automatically generated by the program, is smaller in Figure 22 than in Figure 21.

Figure 21. EGG signal of lax phonation (black) and its derivative (blue), with superposed calculated values of CQ (black numbers at the top) and PIC (blue numbers near the peaks in the derivative).

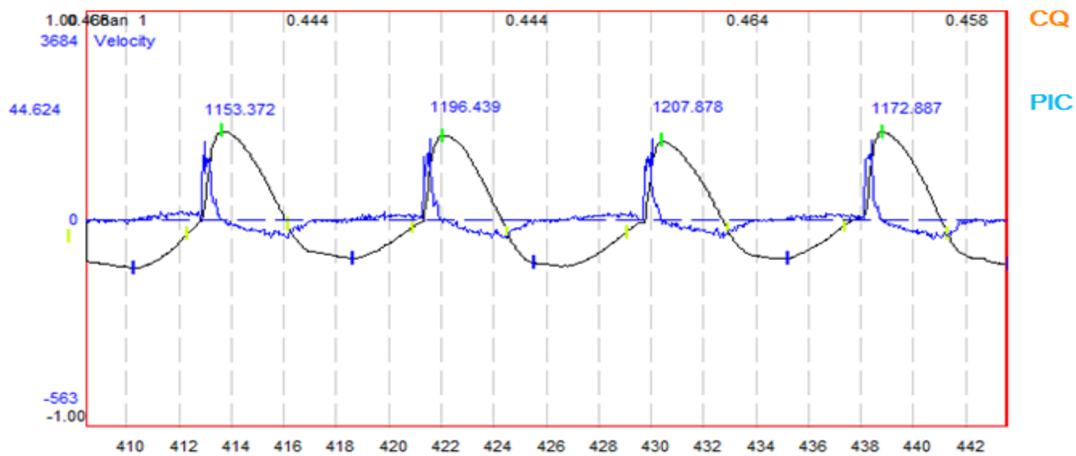
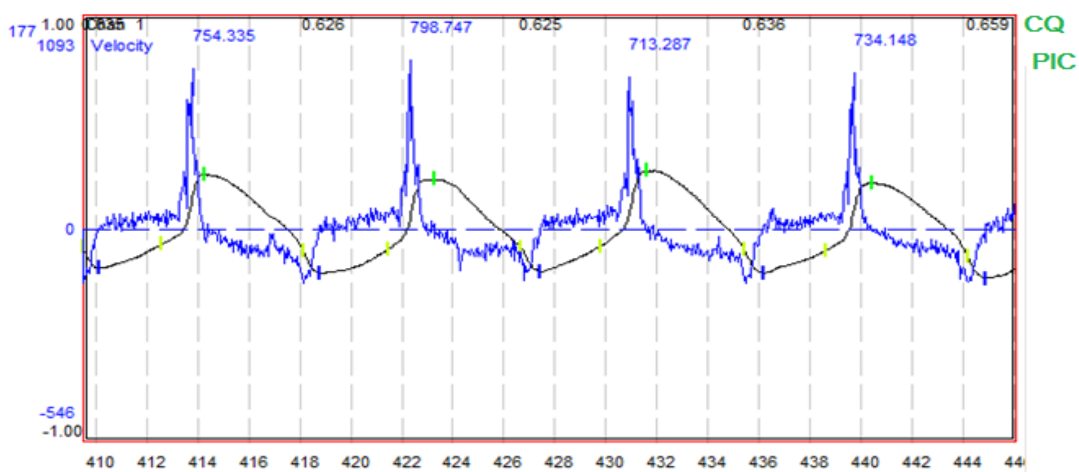


Figure 22. EGG signal of tense phonation (black) and its derivative (blue), with superposed calculated values of CQ_H (black numbers at the top) and PIC (blue numbers near the peaks in the derivative).



As illustrated in the Figure 22 and Figure 21, the tense and lax phonations are different in the following ways:

First, the tense phonation has a larger contact quotient than the lax phonation, as seen the black numbers on the top (around 0.44 for the lax and 0.63 for the tense), which is the principle difference in the phonation contrast. Second, the tense phonation has stronger abruptness than lax phonation, indicated by sharper derivative closing peaks and notable opening peaks in Figure 22. The vibration of lax phonation is more gradual (Figure 21). Third, tense phonation has smaller vibration amplitude and overall lower PIC, as seen from the blue numbers near the derivative peaks (about 1180 for the lax and 750 for the tense).

2.5 Discussion and further analyses

2.5.1 The physiological and acoustic properties of the phonation contrast – focusing on glottal settings

The findings of section 2.4 are summarized in the following tables:

(1) The main effects of tone, phonation and gender on each measurement, in Table 6

(2) Pairwise *post hoc* analysis for significant interactions, in Table 7

Table 6. Main effects of tone, phonation and gender.

	Tone	Phonation	Gender
H1*-H2*	Mid tone higher	Tense lower	
H2*-H4*	Mid tone lower		Female lower
H1*-A1*		Tense lower	
H1*-A2*	Mid tone lower	Tense lower	Female lower
H1*-A3*	Mid tone lower	Tense lower	
B1		Tense lower	
CPP	Mid tone higher	Tense higher	Female higher
Energy		Tense higher	
F0	Mid tone higher		Female higher
CQ		Tense higher	
PIC		Tense lower	

Table 7. Pairwise *post hoc* analysis (check means significant).

	21T vs. 21L	33T vs. 33L	21T vs. 33T	33L vs. 21L
H1*-H2*	✓	✓	✓	✓
CPP		✓	✓	✓
H1*-A1*	✓	✓		✓
H1*-A2*	✓	✓	✓	✓
H1*-A3*	✓	✓		✓
B1	✓			✓
CQ	✓	✓		
PIC	✓	✓		✓

A key result here is that the distinctiveness of the EGG measures confirms that the tense vs. lax contrast in southern Yi is essentially a phonation contrast. There are also very consistent acoustic differences that indicate a breathier voice vs. a creakier voice contrast, particularly in spectral tilt measures. All the spectral tilts relative to H1* are lower for tense phonation. Correlation analysis shows that all of these measures are correlated to

CQ to some extent: H1*-H2*(r=-0.51, p<0.01), H1*-A1* (r=0.49, p<0.05), H1*-A2*(r=-0.32, p<0.05) and H1*-A3*(r=0.27, p<0.05). Among those, H1*-H2* and H1*-A1* are the most highly correlated[3]. The degree of correlation is relative to gender: H1*-A1* has a stronger correlation with CQ in female speakers (r= -0.59 vs. 0.28) and H1*-H2* has a stronger correlation with CQ in male speakers (r= -0.6 vs. 0.34). The other phonation production hypotheses are also supported by our data. As the breathier voice, lax phonation has relatively higher B1, lower CPP (less prominent harmonics) and slightly weaker energy.

The most novel finding of this section is about H2*-H4*. While all the other spectral tilts are related to the phonation contrast, this measure has no phonation effect at all; instead, it is strongly related to tones. A highly significant though modest correlation is found between F0 and H2*-H4* (r=-0.5, p<0.01). The correlation is even stronger when using tonal categories to stratify the observations. This suggests that for tonal languages, the relationship between the second and fourth harmonics might be a potential cue for tonal perception.

Further analysis shows that the prominence of H2* is highly correlated with F0 (r=0.6, p<0.01) but not correlated with CQ (r=0.01). It is H1* that correlates with CQ (r=0.48, p<0.01). That is to say, the opening quotient (indicated by CQ in this paper) is best

[3] A multiple regression was used to evaluate composite relationship of acoustic measurements with CQ. R-squared is 0.3 (corresponding r is 0.57), which means the composite effect from the whole set of acoustic measurements does not explain CQ much more than H1*-H2* does, convincing us that H1*-H2* is the best acoustic measure correlated with CQ to our knowledge.

acoustically reflected by the prominence of H1. This fact gives us better knowledge about H1-H2. That acoustic parameter simultaneously carries information about pitch or tone (through H2) and phonation (through H1), so it is especially important for a language contrasting both tone and phonation. This may also explain why H1-H2 is more independent from the other spectral tilt measures (Hanson1997, Kreiman *et al*. 2007).

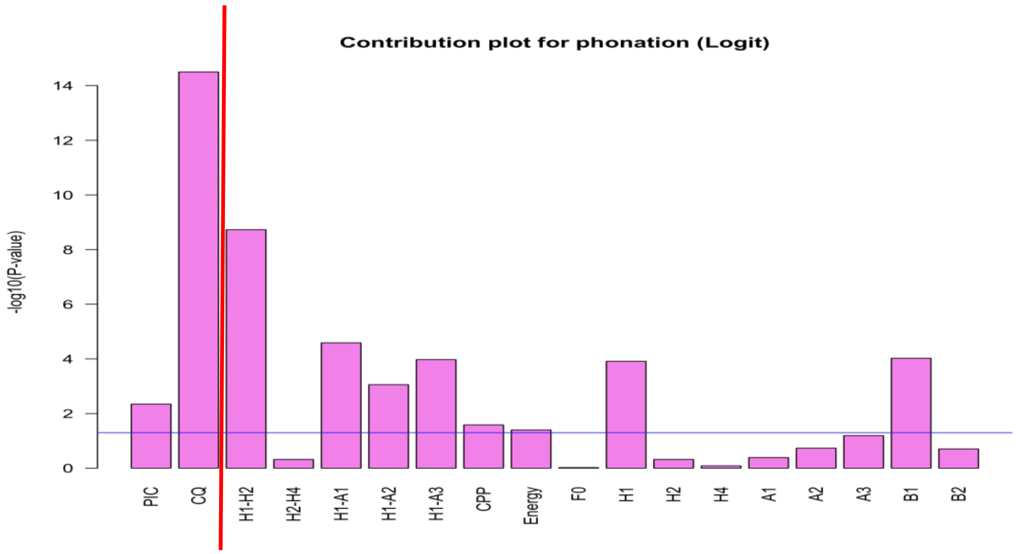
We also employ correlation analysis to explore the relationship between PIC and acoustic parameters. There is no strong correlation found with any measure. The relatively best correlated measure is A3* ($r=0.27$, $p<0.05$) and H1*-A3* is also slightly correlated with PIC. This suggests that PIC can affect the energy of the high frequency spectrum to some degree (the quicker the change in contact, the more high frequency energy), which has been thought to be related to abruptness and speed of vocal fold closing (Stevens 1977). But the result here is in the unexpected direction. As discussed in the previous section, the tension of the vocal folds affects PIC in that tense phonation has a lower PIC (though a quicker, more sharply defined peak), yet nonetheless more high-frequency energy

There are complicated interactions between the phonation types and tonal categories. On the one hand, the phonation contrast is overall more distinctive in the low tone, across measurements. But when there is a three-way interaction (H1*-A1*, H1*-A2*, H1*-A3*, PIC), then females and males have opposite behavior – females have more distinctive phonation in the low tone, but males have more distinctive phonation in the mid tone. On the other hand, as indicated in Table 6, several measures (PIC, H1*-A1*, H1*-A3*) are only sensitive to tonal categories when they are in the lax phonation. In the tense phonation, when the vocal folds are constricted, the F0 difference has less impact on the status of the vocal folds. This phenomenon is also observed by Michaud (mentioned in his 2004 paper). However, it remains unclear about the correlation direction between PIC and F0. Higher tones possibly have higher PIC, as suggested by the average PIC value of our data and supported by the data from Hmong (Esposito 2010, Keating *et al*. 2010), but it is also quite possible that females have a different pattern from males (Figure 20).

2.5.2 Modeling the production of the phonation contrast

We ran a logistic regression to look at the contributions of the various measurements to predict the two phonation types. Results are shown in Figure 23. The Area under the curve (AUC) of this model is 0.77, which shows our model is moderately accurate.

Figure 23. Contributions of measures to phonation contrast production (EGG on the left, acoustic on the right).

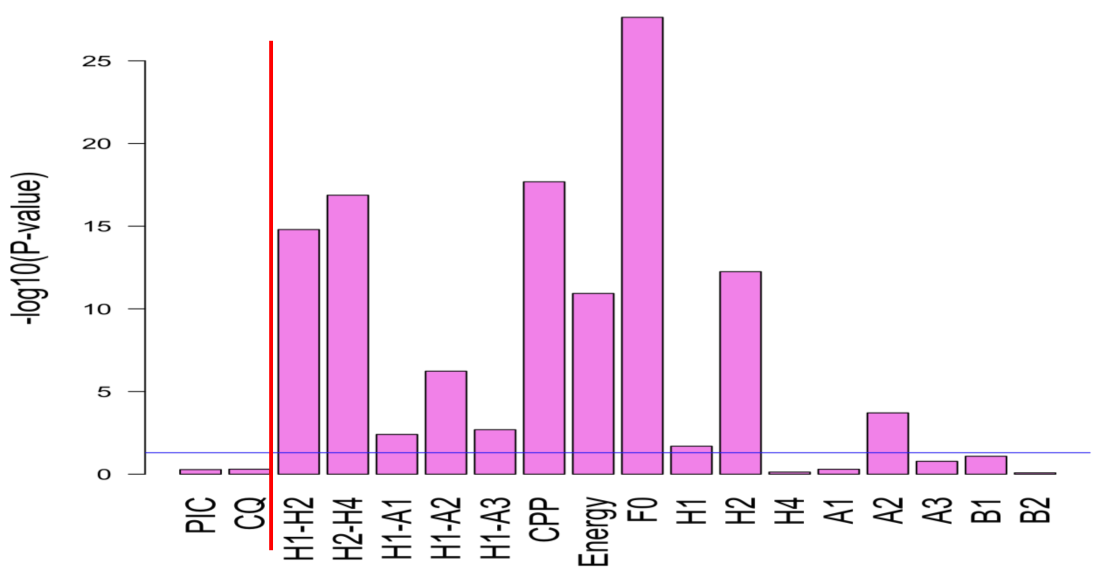


The horizontal line in the plot marks the significance threshold, $p<0.05$. Compared to the peak increase in contact, the contact quotient is the primary physiological difference in the phonation contrast. H1*-H2*, the measure best correlated with CQ, contributes the most among the acoustic measures. The other spectral tilts, H1*-A1*, H1*-A2*, and H1*-A3*, also make salient contributions. Among the amplitude measures of individual harmonics, only H1* has a significant contribution. This illustrates the essential acoustic characteristic of breathiness -- the prominence of H1 in the spectrum, which shows up in all the tilt measures relative to H1*. Other than the harmonic measures, the bandwidth of the first formant, which is related to degree of subglottal coupling and the tension of the vocal tract (here, we believe, the vocal folds themselves here) is also a very important property of phonation contrast production. Additionally, the noise ratio measure CPP makes a significant but not strong contribution to the phonation contrast. The energy difference is very subtle though just significant.

2.5.3 Modeling the production of the tonal contrast

We ran another logistic regression to look at the contributions of the same measurements to predicting the two tones. The Area under the curve (AUC) of this model is 0.83, which shows our model is moderately accurate.

Figure 24. Contributions of measures to tonal contrast production.



The plot in Figure 24 is quite distinct from the one for phonation above. As expected, F0 contributes most to the tonal contrast. Interestingly, CPP is the second most important cue for the tonal contrast. The mid tone has been found to have a larger CPP than the low tone, i.e. more prominent harmonics and/or less noise. This may be due to the fact that the low tone has dynamic F0, which means that F0 is smeared over a time window. Or, the low tone could be breathier, as happens in some languages (but the other measures argue against this). Moreover, H2* and its spectral tilts H1*-H2*, H2*-H4* make quite outstanding contributions too. Recall that H2*-H4* and H2* make no contribution to the phonation contrast. B1, another salient acoustic cue to the phonation contrast, is not significant here either. Physiologically, CQ and PIC do not contribute at all to the tonal contrast.

2.5.4 Interaction between phonation and tone

Figure 23 and Figure 24 well illustrate the distinctive properties of tone and phonation. It is convincing that they are generated by different gestures of the vocal folds. Some measures simply contribute to the tonal contrast, e.g. F0 and H2*-H4*, whereas other measures merely contribute to the phonation contrast, e.g. CQ, PIC, and B1.

Interaction between phonation and tone is mostly found in the spectral tilt measures H1*-H2*, H1*-A1*, H1*-A2*, and H1*-A3*. H1*-H2* is the most vital one among those. As discussed in the previous sections, the first harmonic better reflects the property of phonation whereas the second harmonic better reflects the property of tone. This suggests that spectral tilts contain rich information from multiple dimensions, so listeners probably listen to H1 for phonation and H2 for tone, but the two combined give an auditory impression of voice quality variation that applies to both phonations and tones. Hereafter, we shall use "phonation contrast" to refer to the physiological contrast, but use "voice quality contrast" to refer to the auditory effect.

The interaction of phonation and tone shows up in various ways. First of all, the mid tone has overall higher spectral tilt than the low tone. That means higher tones might sound breathier than lower tones, although CQ and PIC are not distinguished for mid tense vs. low tense. Moreover, the low tone has a more distinctive tense vs. lax contrast than the mid tone for many acoustic measures, although CQ has no significant interaction between phonation and tone. In general, the low tone can better keep the phonation contrast than the mid tone. That could be the reason why the phonation contrast is neutralized with the high tone in Yi. Compared to the reported H1*-H2* JND values across languages (Kreiman *et al.* 2009, 2010), average differences of H1*-H2* in Yi are very small: 2.55dB for the low tone and 0.54dB for the mid tone, which are even smaller than the observed best case JND (Gujarati JND =2.60dB). This raises the issue of the distinction between statistical significance and perceptibility. It is possible that the native speakers may not be able to hear the difference in the mid tone. However, as described below, our perception experiment finds that native listeners are able to hear the contrast in mid tone even better than in low tone (ref. section 3.2 and 3.3 for details). It is possible that Yi speakers are very sensitive to subtle acoustic difference related to voice quality. This higher perceptibility in mid tone may also be attributed to other acoustic cues. Another regression analysis was run to see the contributions of acoustic measures to the phonation contrast in the two different tones. We found that the best acoustic contributor for the mid tone contrast is H1*-A1* (p=4.01E-04). In addition to the spectral tilts, the noise measure CPP is highly significant (p=3.32E-03) while it's not significant in low tone (p=0.23). This suggests that pitch range can affect phonation contrast in such a way that the dominant contributors can be changed. Therefore, although phonation and tone involve distinctive articulations of vocal folds, they can closely interact with each other in the acoustic space.

In addition, as discussed in the earlier section, tense and lax phonations can have different relationships with tone. Although the CQ and PIC of tense phonation do not vary with tone at all, the PIC of lax phonation is sensitive to tonal categories (Figure 19). Therefore, it seems that tonal production is not the same in the different phonation types.

2.5.5 Effect of gender

In general, gender has no main effect on most measures indicating the phonation contrast, but complicatedly interacts with phonation and/or tone for some measures, i.e. PIC, H1*-A1*, B1, and H1*-A3*. Those measures share the properties that the tonal effect appears only in lax phonation, and that males and females can possibly have effects in opposite directions (e.g. PIC). This complication has been partially explained in the previous section, but is not ready for any definitive conclusion.

By contrast, gender has a strong main effect on most measurements related to tone, namely F0, H2*-H4*, and CPP. For F0, female speakers have much higher values than male speakers. For H2-H4, which shows a strong negative correlation with F0, female speakers have much lower values than male speakers. For CPP, the measure reflecting periodicity and harmonic prominence, female speakers have higher values than male speakers, which means the voice quality of female speakers could sound tenser or generally clearer. This pattern is the opposite of English speakers (Hillenbrand *et al*. 1994).

2.5.6 Phonation and the vowel space – the supraglottal settings

(1) Effect of phonation on formant frequencies

The purpose of this section is to examine the proposal that supraglottal settings are involved in the tense vs. lax contrast in Yi. Formant frequencies of vowels are measured automatically by VoiceSauce. Since F1s of low vowels are distant from F0, the program occasionally tracks F0 as F1. To avoid mistakes, results are double checked manually in Praat.

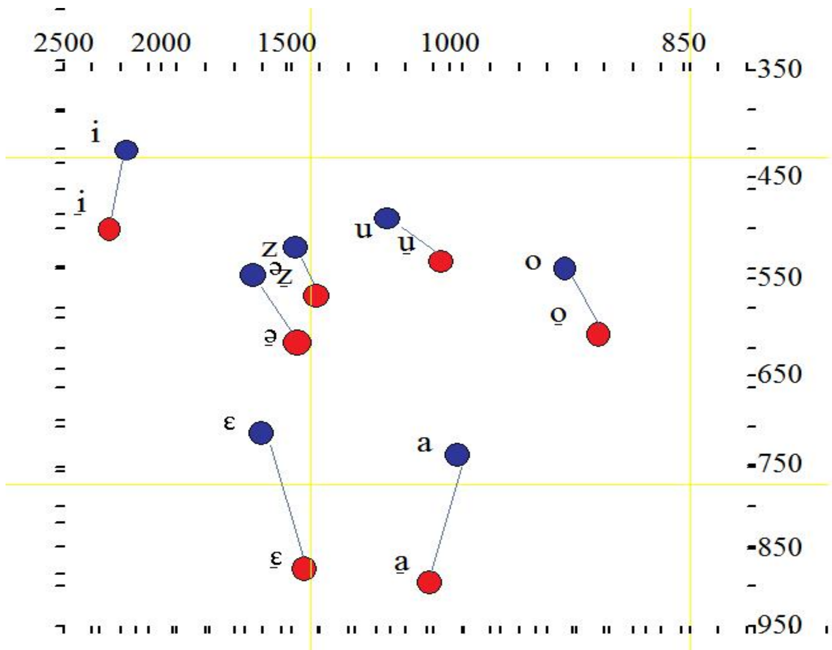
Formant frequencies can only be compared within the same vowel quality. A mixed effect model is used for examining the effects of phonation on vowel formant frequencies. The analysis is done in pairwise comparisons between the minimal phonation contrast vowel pairs. Speaker is set as the random effect in order to normalize the different scales among speakers.

Table 8. Summary of effect of phonation on formant frequencies of Yi vowel pairs, checks indicating significant difference ($p<0.05$) between vowels in a pair.

	ɛ vs.ɛ̱	ə vs. ə̱	i vs i̱	u vs u̱	a vs. a̱	ɿ vs, ɿ̱	o vs. o̱
F1	✓	✓	✓	✓	✓	✓	✓
F2	✓	✓		✓	✓		✓
F3				✓		✓	

Table 8 summarizes the effect of phonation among vowel pairs. As shown in the table, phonation has a significant effect on F1 for all seven vowel pairs. There is no consistent effect on F2 and F3, though most pairs show a significant difference in F2. Inspection of the direction of the effect shows that vowels with tense phonation have higher F1 than vowels with lax phonation, which means tense vowels are lower than lax vowels. The vowel space is plotted in Praat (Figure 25):

Figure 25. The vowel space of Yi, tense (red) vs. lax (blue) vowels.



It can be seen that the tense phonation has lower tongue positions. Mid front vowels and back vowels are also differentiated in F2—tense vowels are slightly more back than lax vowels.

(2) Discussion: Supraglottal settings in phonation contrast

Formant frequencies reflect the shape of the vocal tract. The consistently lower F1 values reveal that lax vowels have a relatively bigger resonator than their tense counterparts. The trend that breathy vowels usually have lower F1 than the corresponding creakier vowels has been observed in several languages (e.g. Kirk *et al.*1984, Maddieson & Ladefoged

1985, Samely 1991). It is believed to be related to lowering vs. raising of the larynx, which leads to a change in the pharyngeal space. Edmondson *et al.* (2001) claim that tenseness in northern Yi is produced by retracting the tongue root and raising the larynx, which is responsible for the consistently higher F1 value for tense vowels. Since a similar vowel space pattern is found in our data, it is possible that a similar mechanism occurs in the tense vs. lax contrast in southern Yi. It is also possible that higher F1 is directly due to tongue lowering. In any case, it seems clear that some supraglottal change is involved in the phonation contrast, in addition to the glottal change.

This leads to the following question: what kind of contrasts can be considered phonation contrasts? If the tongue root is considered as an articulator of phonation, a larger scope of languages should be included in phonation studies. Among those, ATR languages intrigue us most. In these languages, tongue root position is a phonemic feature, with vowel inventories separated into two registers: [+ATR] and [-ATR]. [+ATR] vowels are consistently higher and more front than [-ATR] vowels. This contrast of tongue root position is comparable with Yi and other Tibeto-Burman languages. Could ATR languages also involve a phonation component? Guion *et al*. (2004) made EGG and acoustic measurements for one of the ATR languages -- Maa. It is strikingly found that the answer is yes. Their results are summarized in Table 9:

Table 9. Significant effects for the ATR pairs in Maa (reproduced from Guion *et al.* 2004).

Measure	Result
Duration	ns.
F2	ns.
F1	[–ATR]>[+ATR] for i/ɪ, e/ɛ, o/ɔ, u/ʊ
Normalized A1–A2	[+ATR]>[–ATR] for (e/ɛ, u/ʊ)
Closure quotient[a]	[–ATR]>[+ATR] for all vowels

As expected, F1 in Maa is generally higher for the [-ATR] vowels. This contrasts with the F1 difference in Yi, where F1 is higher for the tense vowels. In Maa, the [-ATR] vowels sound less breathy than the [+ATR] vowels, and Guion *et al*. found that CQ (determined by the threshold method) for [-ATR] vowels is consistently higher than for [+ATR] vowels. Therefore, Maa's ATR contrast involves a phonation difference too. We put the two languages together in Table 10, comparing the phonation measurements and phonation types. Because of this linkage between glottal and tongue root behavior, we can agree with Edmondson & Esling (2001, 2006) that the tongue root can be a phonation articulator.

Table 10. Tongue root movement and phonation types in two languages.

Language	breathier	creakier	F1	CQ
Yi	lax(-RTR)	tense (+RTR)	T>L	T > L
Maa	[+ATR]	[-ATR]	[-ATR] >[+ATR]	[-ATR] > [+ATR]

From this table, the relationship between continuous gestures and voice quality is suggested to be:

(1) Lowered larynx and advanced tongue root can produce a breathy voice; raised larynx and retracted tongue root can produce a creaky voice.

(2) Creakier sounds (tense and [-ATR]) have more constricted vocal folds than breathier sounds (lax and [-ATR]).

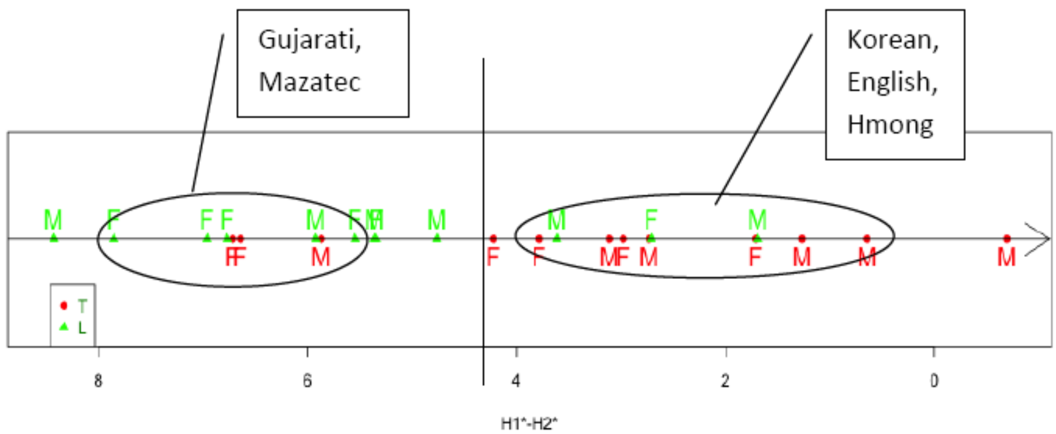
Essentially, advancement/retraction of the tongue root leads to a shape change in the vocal tract, which can change the amount of energy in the higher part of the spectrum as well as the first formant bandwidth. Both of these contribute to an auditory “brighter” voice quality. (Ladefoged & Maddieson 1996: 301-302).

2.6 Further discussion about production variation

2.6.1 The variety of phonation types in tense vs. lax contrasts

“Tense” and “lax” are phonological labels for the Yi phonation contrast. We try to locate these phonological categories on a phonetic voice quality continuum. Since H1*-H2* is the most widely successful phonation measure, it well serves the purpose of cross-linguistic comparison. Here, the Yi values for tense and lax are compared to the modal voice ranges of other languages (data from Garellek 2010, Esposito *et al*. 2009, and Khan 2010). In this way we can see if the Yi contrast is more like a modal vs. creaky contrast, a modal vs. breathy contrast, or a creaky vs. breathy contrast.

Figure 26. Variation in the Yi phonation contrast along H1*-H2* across speakers, compared to modal phonation ranges in 5 other languages. Colored F and M represent the tense (red) and lax (green) means of individual Yi speakers. Ellipses enclose ranges for modal phonations in 5 other languages. Breathier values are on the left and creakier values on the right. X-axis is the reversed H1*-H2* values, in order to match the direction of phonation continuum proposed by Ladefoged (1971) (See Figure1).



Unfortunately, it is impossible to define the phonation types of Yi from Figure 26. Compared to Gujarati and Mazatec, the phonation contrast in Yi is more like modal vs. creaky; however, the contrast is more like breathy vs. modal when compared to Korean, English and Hmong. As can be seen from this figure, voice quality ranges (for both modal and non-modal phonations) vary a lot across languages (Keating *et al.* 2010) and individual speakers within a language. We shall argue therefore that it is not critical to tag the contrastive phonation types with any absolute label, i.e. breathy, modal or creaky. Essentially, the tense vs. lax contrast is a relative feature in Yi; the contrast itself is much more important than the actual types involved in the contrast. Any two distinctive phonation types can serve as the tense vs. lax contrast. That explains why the tense vs. lax contrast has a great variety of phonetic properties among languages, as reviewed in section 1.2. The argument can be further supported by the fact that variation in tense and lax can even be found among individual speakers within the given language.

Figure 27. Variations of CQ and F0 across speakers. Colored F and M represent the tense (red) and lax (green) means of individual Yi speakers. Numbers identify individual speakers. X-axis is mean value of contact quotient. The breathier values are on the left and the creakier values on the right.

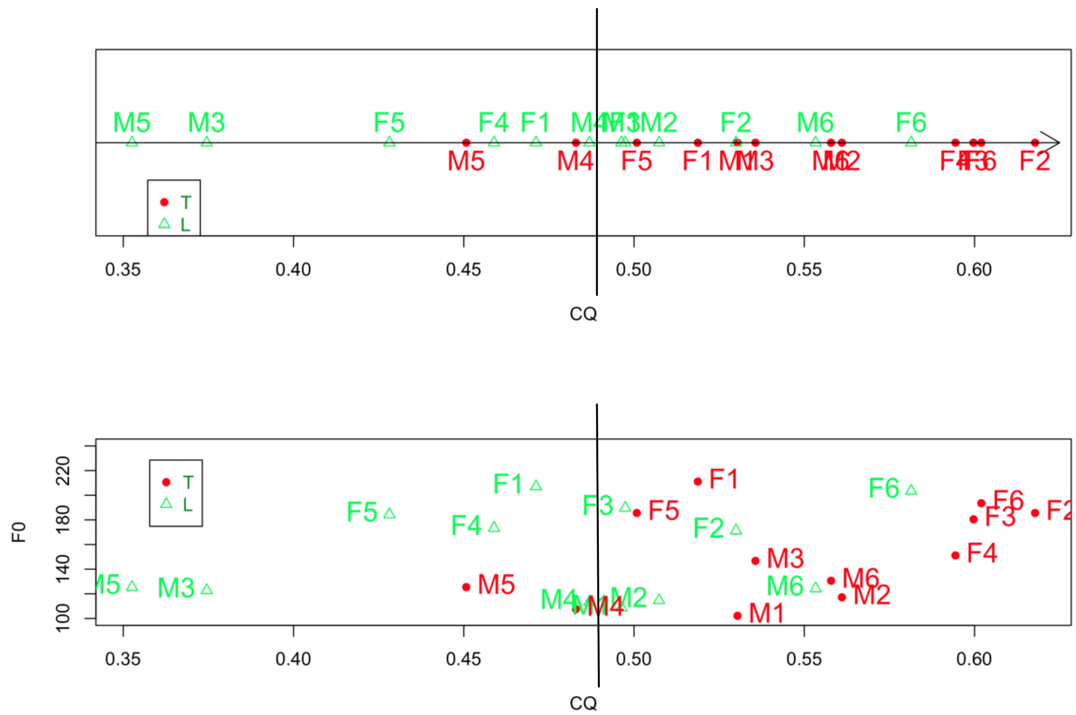


Figure 27 shows the cross-speaker variation of the Contact Quotient from the EGG signal. The bottom figure also displays the relationship between phonation and F0, as phonation has an effect on F0 in eastern Yi dialect (Maddieson & Ladefoged 1985), and for some of our speakers (See Table 11). Although there is much overlap between the two categories, the cutoff line between the CQ values of the two phonation types is around 0.48. Some people are generally breathier than others (e.g. M5), while some speakers are relatively more in the creaky range (e.g. F2). This suggests that the actual voice qualities people use vary among individuals, as long as the voice qualities are distinct from each other within each individual.

2.6.2 The variety of phonetic cues used by individual speakers

It is noticeable that in Figure 27 some speakers have a more distinctive phonation contrast while other speakers' phonation contrast tends to be minimal, especially speaker M4. If all speakers keep the phonological contrast in their production, what parameters do they use? The following table (Table 11) is a summary of the cues used by the individual speakers, that is, measures which are significantly different between the two phonations:

Table 11. Contrasting cues of individual speakers. Check means P-value<0.05; Numbers indicate marginal significant p values.

Speaker	CQ	PIC	H1*-H2*	H2*-H4*	H1*-A1*	H1*-A2*	H1*-A3*	CPP	F0	F1
F1	✓	✓	✓				✓			✓
F2	✓		✓		✓	✓	✓	✓		0.07
F3	✓		✓						0.08	✓
M1	✓		✓							✓
M2	✓	✓	✓	✓						✓
M3	✓	✓	✓		✓	✓	✓	✓		
F4	✓	✓	✓		✓	✓	✓		✓	✓
F5	✓	✓			✓					0.07
F6	✓				✓			✓	✓	
M4					✓	✓				✓
M5	✓		✓							✓
M6	✓	✓			✓		✓			✓

CQ distinguishes the phonation contrast for all the speakers except M4. This indicates that the contrast may be fundamentally one of differences in glottal contact. PIC does not work for every speaker; several speakers use CQ but not PIC. However, despite this clear physiological basis of the contrast, no single acoustic measure so consistently distinguishes the categories. H1*-H2* and H1*-A1* seem to be alternative cues for all the speakers, in that at least one of these distinguishes the categories, but only 3 speakers use both. F1 is also a strong parameter for the tense vs. lax contrast, especially for the speakers who make a weaker EGG contrast (i.e. PIC is not significant, or CQ is less contrastive as in Figure 27). M4 is the extreme case of this kind, who does not have an EGG contrast at all. However, this speaker still has significant acoustic cues (i.e. H1*-A1*, H1*-A2*, F1) to keep the tense vs. lax contrast.

As the higher part of the spectrum can be affected by vocal tract shape, and in light of Edmondson *et al.* (2001)'s study about northern Yi (whose speakers can possibly make a contrast without glottal activity), we speculate that the tense vs. lax contrast of this speaker may solely rely on supraglottal settings, whose existence is partially evident from the F1 difference.

Phonation categorygenerally has no main effect on F0, but several female speakers (F3, F4, and F6) have consistently higher F0 in the tense phonation (Figure 27). The effect is similar to Maddieson & Ladefoged (1985)'s observation from eastern Yi. Therefore, variations of tense vs. lax contrast seen across dialects can exist in one given dialect.

In addition, M4 and F6, the two speakers who have the weakest phonation contrast, turn out to be the youngest speakers (18 years of age). Therefore, it is possible that the phonation contrast in southern Yi is undergoing sound change, and may even eventually vanish. But for the language, the phonological tense and lax contrast would nonetheless be maintained, as it can be realized by other cues, such as tongue height.

2.7 Summary of the production experiment

This chapter investigated the phonation contrast in southern Yi. EGG data confirms that Yi has a phonation contrast, with CQ the most basic property of its production. Acoustically, various measures contribute to the phonation contrast. Among those, H1*-H2* and H1*-A1*, which are strongly correlated with CQ, are the best acoustic measures for the phonation contrast. Other effective correlates include B1 and CPP. F0 has no main effect on phonation but interacts with phonation contrast in the acoustic space. The Yi tonal contrast has distinct physiological mechanisms from the phonation contrast, but it can interact with phonation in the acoustic space. Spectral tilt measures not only reflect the phonation contrast, but also carry information about the tonal contrast. This is particularly clear for H1*-H2*; however, we showed that this is because H1* is correlated with CQ (and thus H1*-H2* differs between phonations) while H2* is correlated with F0 (and thus H1*-H2* differs between tones). As a result, the tones differ to some extent in phonation type as well as in F0. There is no gender effect on phonation in general, but an interaction of tone by gender is found for lax phonation in several measures: PIC, H1*-A1*, H1*-A3* and B1. All of these reveal the complicated physiological-acoustic coupling in phonation production.

Moreover, in addition to the glottal settings, supraglottal settings are also very important for the tense vs. lax contrast. A consistent F1 difference in the phonation contrast in Yi indicates a shape change in the vocal tract, supporting the multidimensional phonation model proposed by Edmondson & Esling (2006). This is like an RTR mechanism, which can be compared with ATR languages, and it was found that while tongue root is an independent articulator in these languages, there is a relationship between continuous gesture and voice quality (Table 9). Furthermore, as B1 is strongly correlated with F1, a larger pharyngeal size (which raises F1) can contribute to an auditory breathier voice quality via its effect on B1. Therefore, for the ATR and RTR languages, tongue root movement is involved in phonation production and responsible for the separation of vowel registers.

In the final discussion of variation, we showed the essential phonological knowledge of tense vs. lax contrast: the contrast itself is more important than the actual types, and thus various ways of making the contrast are possible. All that native speakers need to know is that the tense voice should be creakier and brighter than the lax one. With this understanding, the complicated and diverse variations across Yi dialects can be understood under the same frame.

3 Perception experiment

The purpose of this experiment is to explore the mental reality of the tense vs. lax contrast of native speakers. We already learned from the inventory of Yi (Table 1) that the tense vs. lax contrast is not evenly distributed across all phonological conditions. This implies that some contrast pairs might be easier to perceive (or produce) than others. So it is very intriguing to know how contrastive pairs are organized in native speakers' minds, i.e. what their "perception map" for the contrast looks like.

Such a perception map can further serve to evaluate the role of voice quality, tone and vocal tract shape in the tense vs. lax contrast from the view of perception. The production experiment has found that the tense vs. lax contrast is realized by multidimensional cues, with both vocal folds and supraglottal settings involved. Tones do not have a main effect on tense vs. lax contrast production in the southern dialect. We nonetheless found that tone and phonation can to some extent interact with each other in the acoustic space. So it is worthwhile to evaluate this interaction in the perceptual space.

Lastly, it was also shown that individual speakers have great variation in production, which points to the possibility of sound change in progress. Is there perception variation too?

The experiment is designed around the basic idea of cue weighting. It is a common observation that an auditory category has multiple distinctive acoustic cues (e.g. Repp 1982 for review), but usually one cue is more important to listeners than the others. For example, as found in Hillenbrand *et al*. (2000), native English listeners rely much more on formant frequencies than vowel duration in categorizing /i/ and /I/. Cue weighting judgments always are language specific, dialect specific and even speaker specific (Holt & Lotto 2006), based on the mental reality or the perception map of native listeners. The perception map varies due to the different linguistic experience and phonological knowledge of listeners. Esposito (2010) found that listeners from a language with a phonation contrast are more sensitive to the different phonation types; and that listeners from different languages relied on different cues, though only one of the listeners' languages had a contrast. So variation in categorical perception of phonation contrasts across speakers can give a better understanding of native speakers' knowledge of phonation contrasts.

A standard and ideal approach to test for cue weighting is synthesizing stimulus continua that vary the values of the cues being tested, and using an identification task to obtain categorizations of stimuli. If these categorizations are used to estimate similarities among the stimuli, Multidimensional Scaling can be employed to illustrate which cues can best separate the stimuli into categories (Iverson & Kuhl 1996). However, this approach is not practical in a fieldwork study. An identification task with synthesized signals is relatively hard for listeners, since the differences between stimuli are very subtle. It requires a very quiet experimental environment as well as highly attentive listeners. In addition, synthesized voice quality is often unnatural, which makes the task even more challenging. We tried this approach in the field and found that near half of the subjects failed in the task, i.e. their answers were by chance on items that should have been easily categorized.

We then tried an alternative approach, using natural stimuli. To make the task even easier and more acceptable for the listeners, an AXB instead of AX identification task is used in this experiment. The challenge of this approach is how to control the variation in the stimuli. Instead of using synthesis, following Jiang *et al.* (2007) and Esposito (2010) among many other such studies, we use statistical methods to model the cues of the stimuli *post hoc* rather than manipulate them *a priori*.

Based on our data, two statistical methods are applied for two main purposes: Model the dissimilarity and then determine the contributors to the dissimilarity.

First, Multidimensional Scaling (MDS) is applied to the stimulus set to provide a visual representation of the dissimilarity distances between testing pairs under different phonological conditions. By knowing which testing pairs are acoustically more confusable than the others, we can explain the confusion bias in the perception space.

Second, logistic regression (along with t-tests) is applied the same way as in the production chapter, to determine which signal properties significantly distinguish the test stimuli from the two phonation categories. Since the test stimuli are a subset of the entire corpus, we expect the results to be the same as in the production study.

3.1 Experimental design

3.1.1 Subjects

Ten listeners from Xinping village participated in this experiment, five males and five females (i.e. m1~m5, f1 ~f5). Only one of the female listeners also participated in the production experiment. The ages of the listeners ranged from 18 to 50. Southern Yi is their native language and also the primary language in daily life. The subjects who participated in the perception experiment generally have more education background than those in the production experiment.

3.1.2 Stimuli

The task is aimed to examine the tense vs. lax contrast in different phonetic environments and is supposed to answer the question: in what conditions are tense vs. lax syllables more likely to be confused?

As shown at the beginning of this thesis, the phonation contrast in Yi occurs across all vowels and with mid and low tones. We thus produced four groups of stimuli with all combinations of tone and vowel height:

1) High vowel with mid tone (bu33/bu̲33)
2) High vowel with low tone (bu21/bu̲21)
3) Low vowel with mid tone (bɛ33/bɛ̲33)
4) Low vowel with low tone (bɛ21/bɛ̲21) (for convenience in plotting, this vowel is hereafter transcribed as [e])

A dataset containing all testing types, namely "e21 vs. e̲21","e33 vs. e̲33","u21 vs. u̲21","u33 vs. u̲33", was retrieved from the production corpus. Unpaired t-tests on all measures demonstrated that this subset did not differ reliably from the whole dataset (all p-value >0.05) and thus can represent it.

3.1.4 Procedures - AXB identification task

The audio stimuli were retrieved from the recordings of the six speakers from Xinping village, labeled as F1, F2, F3, M1, M2 and M3, and put into an AXB task. The pronunciations of M1 and F1 were chosen as the standards for the AXB identification task, as they seem to maintain a good contrast between tense vs. lax phonations. The minimal phonation contrast pairs that were produced by these two speakers served as the A and B, and the Xs were the pronunciations of all six speakers. For example:

F1_bu33(A), F2_bu33(X), F1_bu̲33(B)	M1_bu33(A), F2_bu33(X), M1_bu̲33(B)
…	…
F1_bu33(A), M3_bu33(X), F1_bu̲33(B)	M1_bu33(A), M3_bu33(X), M1_bu̲33(B)

Therefore, the listeners heard 20 stimuli in each group (half compared to F1 and half compared to M1), thus 80 stimuli in total. This stimulus set was presented three times to each listener.

The task was run by a Praat script on a computer. Audio stimuli were played through SONY MDR-NC60 headphone. On the screen, the listeners could see three buttons, labeled as (A), (X) and (B) (The buttons of A and B were in yellow and clickable) The listeners heard three stimuli in sequence separated by 0.5 second, and had to decide whether the second (X) is more similar to the first (A) or to the third (B). Listeners had to make a response for every trial by clicking either A or B. They were able to replay the audio as often as necessary before responding, and they also could “regret” and go back to re-listen to the previous sequence. There was an introduction and a practice session before the formal test. For those who had difficulties operating the computer, I asked the subjects simply to point on the screen, and I assisted them in clicking the mouse. The duration of the experiment by design was under 40 min. Listeners could pause if they felt tired.

3.2 Results

Listener m2 failed to perceive the differences in the stimuli at all, and thus is excluded from the data analysis. Thus the data from 9 listeners are presented here.

Figure 28. Four-fold displays for four conditions. The dark shading indicates the correct types whereas the light shading represents the wrong types.

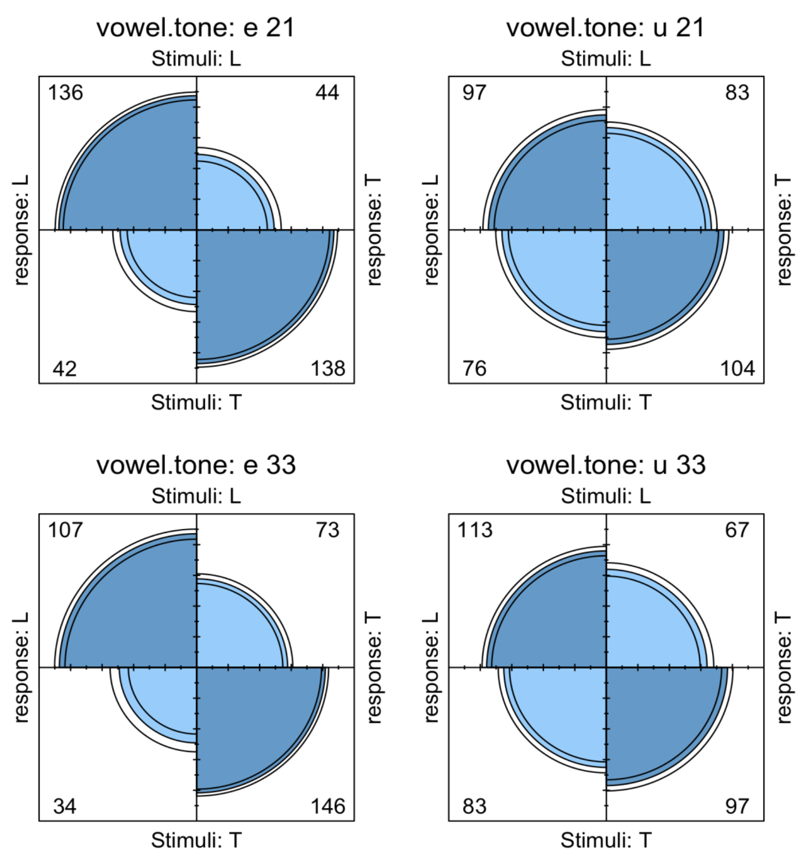


Figure 28 is a set of four-fold displays (Friendly, M. 1994). A four-fold display shows the frequencies in a 2 x 2 table in a way that depicts the correctness ratio and the distribution of responses. In this display the frequency of responses in each cell is shown by a quarter circle, so each quarter circle represents one of four types of answers, relative to the X stimuli (i.e. stimuli: response= L:L, L:T, T:T and T:L). The radius is proportional to the square root of the count, so the area indicates the proportion. An association between the stimulus and response is shown by the tendency of diagonally opposite cells, with wrong answer types (i.e. stimulus: response= LT, TL) in one direction, and correct answer types (i.e. stimulus: response= LL, TT) in the other direction. We use color and shading to distinguish the directions: the dark shading indicates the correct types whereas the light shading represents the wrong types. Confidence rings for the observed data provide a visual test of the null hypothesis of no association.

For example, the top left panel shows that for 360 stimuli under the condition of low tone + low vowel, 136 responses to Lax stimuli are L while 44 are T, and 138 responses to Tense stimuli are T, while 42 are L. So there are about 76% correct answers in total under this condition. Comparing the answer rates across panels, it can be concluded that low vowels generally have higher correctness rates than high vowels. Interestingly, the correctness rate for the mid tone is slightly better than the low tone, although the difference does not reach significance.

3.3 Discussion

(1) The perception map of the tense vs. lax contrast

The unbalanced accuracy rate of low vowel pairs and high vowel pairs suggests a bias in the perceptual distances of tense vs. lax vowels. We visually present the perceptual map for the tense vs. lax contrast under four phonological conditions, following the approach of Johnson (2003). Suppose that “A” refers to lax and “B” refers to tense, with capitals used for stimuli and lower case for response, and that 6 of 10 “bu33” sound like “bu̲33”; then if P stands for proportion, then we can define

(a)	(A)“bu33”	(B)“bu̲33”
[a][bu33]	PAa	PBa
[b][bu̲33]	0.6(PAb)	PBb

Similarly, PBa is the proportion of how many tense “bu̲33” are heard as lax syllables.

With this kind of P value matrix, we can calculate the similarity of different vowels, given in Table 12, by the following equation:

$S_{ij} = (P_{ij} + P_{ji})/(P_{ii} + P_{jj})$

Table 12. Similarity matrix (all possible pairs of 8 stimuli).

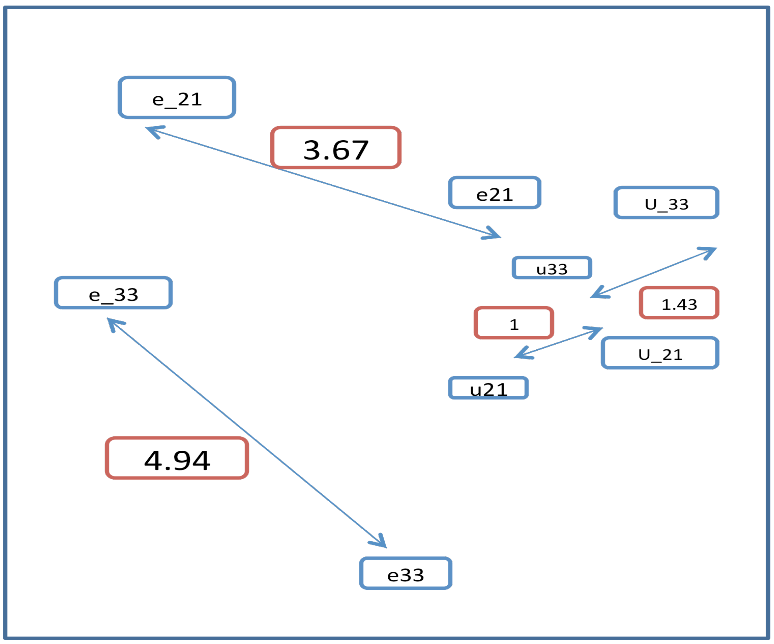
	be̲31	be31	be̲33	be33	bu̲31	bu31	bu̲33	bu33
be̲31	0	1.159						
be31	1.159	0						
be̲33			0	0.861				
be33			0.861	0				
bu̲31					0	0.234		
bu31					0.234	0		
bu̲33							0	0.336
bu33							0.336	0

The negative of the natural log of the similarity is used to calculate the perceptual distance (dissimilarity):

$d_{ij} = - \ln(S_{ij})$

Figure 29 is the resulting perceptual map of the tense vs. lax contrast.

Figure 29. Perception map of tense vs. lax contrast. Numbers in red boxes are distances between vowels in a minimal pair.



In this map, the tense vs. lax contrast is perceptually more distinguishable in low vowels than in high vowels. That accounts for why low vowels have much better correctness rate than high vowels.

Another important observation is that the mid tone pairs have slightly larger perceptual distances than low tone pairs, though this is not significant. The fact is particularly interesting since our production experiment found that spectral tilts (e.g. H1*-H2*) distinction of tense and lax is generally less salient in mid tone (ref. section 2.5.4 for discussion). This means that the perceptibility of the tense vs. lax contrast in mid tone is no worse than that in the low tone. To reach the same level of perceptibility, there must be some cue compensation or substitution; for example, CQ can be reflected in acoustic cues other than H1*-H2* in mid tone. Since the tonal effect is not significant, we will not go further on this issue.

(2) *Post hoc* analysis of stimuli

The reason why low vowel pairs have larger perceptual distances might be because the production of low vowel pairs is more distinct. To test which stimulus pairs are more alike than the others, we need to know the overall acoustic similarity among the stimuli. The distribution of the stimuli in acoustic space can then be calculated and plotted by MDS.

MDS has often been successfully applied to perceptual data in linguistic studies, but here it is applied to production data. MDS can visually present observation points in a lower dimension space. The algorithm works as following (Kruskal 1978: 27-28): 1) Using distance functions (e.g. Euclidean distance) to compute distances (matrix D) among categories in an original high k-dimensional space (here 12 acoustic measures are the coordinates of a 12-dimension space); 2) Find a low p-dimension space (p can be any number between 1 to k-1, here 1~11) to best visually present the distances among categories. To do so, 2a) Compute the distances (e.g. Manhattan distance) among all pairs of points, to form their dissimilarity matrix (d) in this low p-dimension. 2b) Compare this matrix (d) with the input data matrix (D) by evaluating the stress function. The smaller the stress value, the greater the correspondence between the two. Adjust coordinates of each point in the direction that best minimizes the stress until the stress won't get any lower. (For our case, a 2-dimension space is adequate to present the data.)

Simply speaking, MDS sums up effects from all the individual measurements and reveals the overall acoustic distance between every tested pair. In this way, we can get a “production map” of the stimuli, which can clearly illustrate the general acoustic dissimilarity of the testing pairs.

Two popular distance functions can be used to calculate the dissimilarity based on physical measurements: Euclidean distance and Manhattan distance. We employ the Manhattan distance in this study since it discounts the influence of different scales among different measurements. Here is the formula for the distance between p and q over i dimensions:

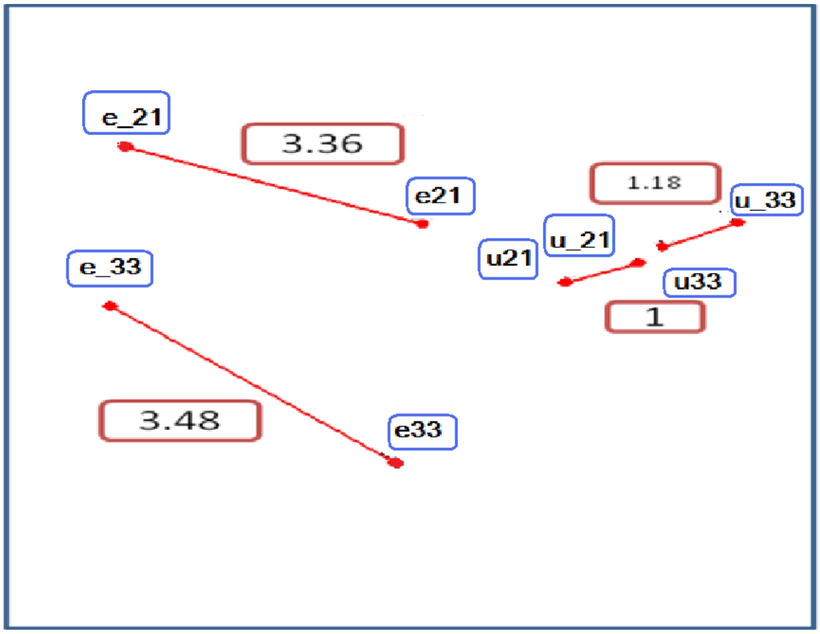
$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|,$$

The stress function, which measures (inversely) the degree of correspondence between the distances among points implied by the MDS map and the matrix input by the user, is as follows:

$$\sqrt{\frac{\sum \sum (f(x_{ij}) - d_{ij})^2}{scale}} \tag{4-1}$$

The MDS presented here was performed in R using the isoMDS function (Venablesn & Ripley (2002)). Figure 30 is the resulting production map of the tense vs. lax contrast.

Figure 30. Production map for tense vs. lax contrast. Numbers in red boxes are distances between vowels in a minimal pair.



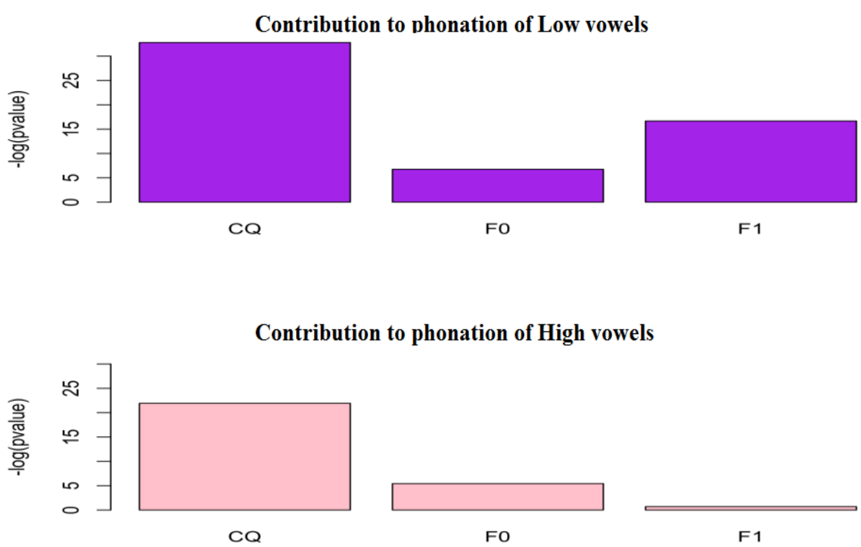
As shown in the above plot, the low vowel pairs have larger dissimilarity than the high vowel pairs, and thus it should be easier to hear the difference. That is presumably why low vowel pairs have a relative higher perceptual accuracy rate. Comparing the perception map (Figure 29) with the production map (Figure 30), it can be seen that listeners are able to perceive the variation in pronunciation, perfectly matching the production map.

What cues contribute to this unbalanced dissimilarity in the tense vs. lax contrast?

A logistic regression model was run for the stimuli's production data to evaluate the contributions of voice quality, tone and vowel quality. To simplify the statistical model, only three measures were analyzed, each taken to be the best representative of a phonological dimension: CQ as the phonation measure, F0 as the tone measure and F1 as the vowel quality measure. Since the various acoustic measures, which contribute to the phonation contrast, essentially reflect CQ, and CQ can account for much larger variance than any of the acoustic measures (e.g. H1*-H2*), we use CQ instead of any acoustic measure to indicate phonation.

As indicated in Figure 31, vowel quality (i.e. F1 here) makes a significant contribution to the low vowel contrastive pairs but little contribution to the high vowel pairs. For both vowel heights, voice quality is always the most important cue for tense vs. lax, consistent with our findings in the production experiment.

Figure 31. Contribution to the tense vs. lax contrast for different vowel qualities.



Since the regression model can only reveal the relative weights of contributions (the heights of the columns in Figure 31), a paired t-test is employed to show the degree of absolute difference between counterparts in the minimal pair. It turns out that the tense vs.

lax contrast has a much more distinctive CQ in the high vowel pairs ($t(54)= -5.78$, $p=3.894e-07$), compared to that in the low vowel pairs ($t(54)=2.23$, $p=0.03$). Therefore, phonation is the only cue to the contrast in the high vowels, and this phonation difference is more extreme. For the low vowels, the phonation distinction is much smaller, but the tense vs. lax contrast has F1 as an additional minor contributor. The accuracy rate of identification in the different conditions can suggest the cue dimensions listeners rely on. If listeners only rely on the phonation cue, we expect the low vowel stimuli should have a lower accuracy rate than the high vowels, since the CQ difference is smaller in low vowels. Alternatively, if listeners rely on vowel quality, then a higher accuracy rate should be found for low vowels. The actual perception data supports the second hypothesis.

A second logistic regression model was run for the perception data to predict the perceptual phonation types, with the vowel, tone, gender and speaker of the stimuli as the predictors. This model suggests that the vowel factor makes a highly significant contribution to the responses ($p<0.001$) but tone does not ($p=0.49$). This result further confirms that vowel quality plays a crucial role for perception of the tense vs. lax phonation contrast.

(3) Perceptual variation and indication of sound change

From the above discussion, the conclusion can be drawn that vowel quality (F1) is a very robust cue for tense vs. lax categorical perception, though voice quality is still the distinctive feature for the tense vs. lax contrast.

Can we predict the future sound change route? As synchronic variation is the clue to diachronic sound change, it is very interesting to closely look at individual listeners' performance, given in Table 13:

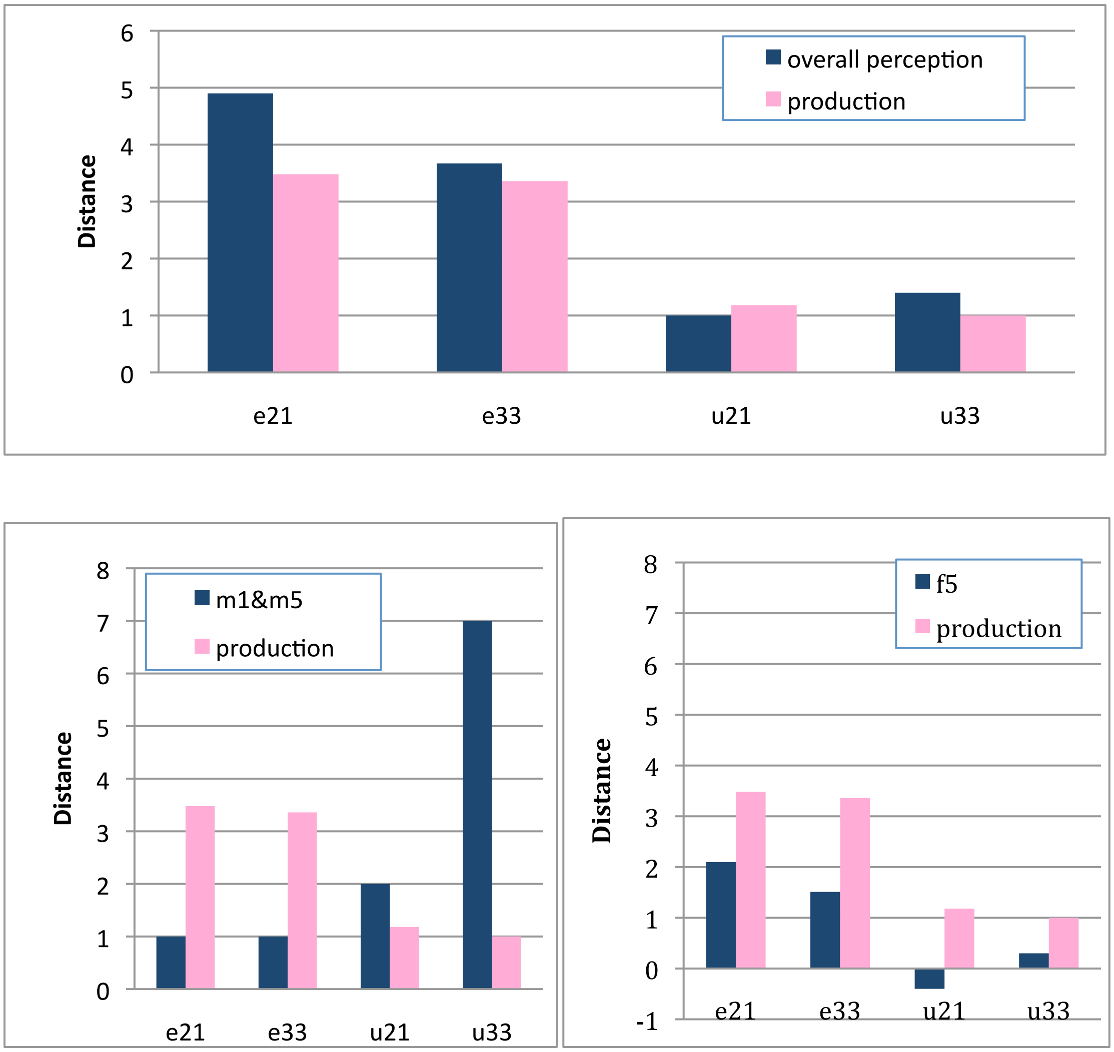
Table 13. Perceptual variation of individual listeners, false alarm and hit rates by vowel.

	f1	f2	f3	f4	f5	m1	m3	m4	m5
e-false	0.15	0.09	0.04	0.06	0.16	0.25	0.08	0.13	0.25
e-hit	0.35	0.41	0.46	0.44	0.34	0.25	0.43	0.37	0.25
u-false	0.22	0.16	0.11	0.2	0.26	0.23	0.28	0.26	0.22
u-hit	0.28	0.34	0.39	0.3	0.24	0.28	0.23	0.24	0.28
overall-correct	0.63	0.74	0.84	0.74	0.59	0.53	0.65	0.61	0.53

All the listeners are able to hear the difference between tense and lax (the overall correct ratio is bigger than 0.5). Most listeners hear the difference much better in the low vowel pair, when voice quality has a smaller difference but vowel quality has a significant contribution; however they perform less well on the high vowel pairs, when the vowel quality difference is absent though the voice quality difference is much bigger. This suggests that these listeners are more sensitive to vowel quality differences than to voice quality. With the absence of vowel quality differences, f5 fails to perceive the different phonation categories for [u]. However, we also notice that some listeners, e.g. f3, are good at hearing the difference for both low vowel and high vowel pairs, suggesting that they are sensitive to both voice quality and vowel quality differences. More interestingly, m1 and m5 perform differently from the other listeners. These two listeners hear the contrast better in the high vowel pair but totally fail to distinguish tense vs. lax categories in the low vowel pair, even when vowel quality difference (F1) is a very strong cue for the low vowel pair. That means these listeners do not pay attention to the F1 distinction. They use only voice quality as the cue for the tense vs. lax contrast. Given that CQ is more saliently different in the high vowel pair than in the low vowel pair, the listeners could hear a difference better in high vowels than in low vowels. Therefore, we can speculate that vowel quality seems to be growing as the alternative distinctive feature of tense vs. lax contrast for native listeners.

The different perception maps are illustrated in the following distance plots (Figure 32). The scaling of the distances is standardized and compared with the production map:

Figure 32. Variation in perception maps: Blue bars represent the perception distances (all listeners, combination of m1 and m5, and f5 respectively for top graph, left bottom and right bottom) and production distance (for all stimuli) is in pink.



3.3 Summary

This chapter further illustrates the mental reality of the southern Yi tense vs. lax contrast. The vowel quality difference examined (F1) contributes a lot to the dispersion distance in the production map, as the voice quality cue examined (CQ) is relatively weaker in low vowels than in high vowels. This vowel quality cue is also robust for categorical perception. It is shown that listeners apparently rely heavily on this cue, as they show accordingly high accuracy in low vowel pairs.

Generally, the perception map reflects the production map well. More interestingly, it is found that the perception map also varies across listeners. Listeners prefer different cues in categorization: some use both vowel quality and voice quality cues; some prefer vowel quality to voice quality (e.g. listener f5); and some only use voice quality (e.g. listener m1). These variations suggest the orientation of possible sound change. We speculate that the phonation contrast might be gradually replaced by a vowel quality distinction. Such a sound change will be eventually finished when the listeners no longer pay attention to the voice quality difference but solely rely on the vowel quality difference. At the current stage, females are more sensitive to the vowel quality cue than the males in general.

4 General summary

In southern Yi, the tense vs. lax contrast is essentially a phonation contrast, but it is complicatedly accompanied by multiple phonetic properties. In order to reveal how phonation types serve as the phonemic dimension in the language, we conducted a production experiment to better understand the relationship between physiological mechanisms and acoustic properties in the phonation contrast, and also a perception experiment to explore the mental reality of the phonation contrast in native speakers.

Comprehensive analysis of the production data provides substantial findings of physiological-acoustic coupling. Compared with Peak Increase in Contact (PIC), the Contact Quotient (CQ) is the more essential property of the phonation contrast. Acoustically, H1*-H2* and H1*-A1*, which are strongly correlated with CQ, are the best acoustic measures for the phonation contrast. Other effective cues include the bandwidth of the first formant (B1) and Cepstral Peak Prominence (CPP). F0 does not contribute to the phonation contrast in this language.

However, the data reveal a complicated interaction between tone and phonation. The tonal contrast has distinct physiological mechanisms from the phonation contrast, but it can interact with phonation in acoustic space. Spectral tilts are not only responsible for the phonation contrast but also carry information about the tonal contrast. H1*-H2* is particularly important for both phonation and tonal contrasts, as H1* is correlated with CQ while H2* is correlated with F0. In addition, it suggests that tonal production differs in phonation type. There is no gender effect on phonation in general, but an interaction of tone by gender is found for lax phonation in PIC, H1*-A1*, H1*-A3* and B1. All of these reveal the complicated physiological-acoustic coupling in phonation production.

Another main contribution of this thesis is that other than the glottal settings, supraglottal settings are also very important for the phonation contrast. Consistent F1 differences in the phonation contrast in Yi indicate a shape change in the vocal tract, supporting the multidimensional phonation model proposed by Edmondson & Esling (2006). The RTR mechanism was compared with ATR languages, and a consistent pattern is found. As B1 is strongly correlated with F1, larger pharyngeal size can account for an auditorily breathier voice quality. Cross-linguistically, tongue root movement is involved in phonation production and is responsible for the separation of vowel registers.

Therefore, we argued that tense and lax are relative and auditory features. The contrast itself is more important than the actual types involved in the contrast, and that is why various ways of making the contrast are possible.

Further cross-language and within-language study of the phonation types involved in this tense vs. lax contrast reveals that the phonation contrast is relative and the contrastive types vary among individual speaker and differ across languages. The variation in production suggests some possible sound change of the phonation contrast. A perception experiment was conducted to obtain more direct evidence to support the conclusions reached from production. It was found that listeners heavily rely on vowel quality for the perception of the Yi tense vs. lax contrast, though it is only a secondary cue in production. The perception map of the tense vs. lax contrast varies across listeners, which further indicates the future direction of sound change: Vowel quality might eventually upgrade as the distinctive feature of the tense vs. lax contrast.

APPENDIX

Appendix 1

WORD LISTS

Minimal pairs of tense vs. lax

	minimal pairs	English	Chinese	Gloss	English	Chinese	Gloss
1	u/u̱	frighten	吓唬	gu21	bent	弯	gu̱21
2		chopstick	筷子	dzu21	waist	腰	dzu̱21
3		muscle	筋	dzu33	fear	怕	dzu̱33
4		stick	棍子	du21	get fire	着火	du̱21
5		hoe	锄头	kʰu21	year	年	kʰu̱21
6		sky	天	mu33	blow	吹	mu̱33
7		orange	桔子	lu21	enough	足够	lu̱21
8		soft	软	nu33	fur, hair	毛	nu̱33
9		silver	银子	tʰu21	dull	钝	tʰu̱21
10		hoe	锄头	tsʰu21	six	六	tsʰu̱21
11		praise	称赞	pu33	curl up	蜷缩	pu̱33
12	ə/ə̱	count	数v	ɣə21	needle	针	ɣə̱21
13		dew	露水	tsə21	wrinkled	皱	tsə̱21
14		kitchen	厨房	tə21	hole	洞	lu33 tə̱21
15		herd	放牧	ɬə21	dry in the sun	晒	ɬə̱21
16		quantifier for book	一本[书]	pə21	dregs	渣	pə̱21
17		pillow	枕头	gə21	foot	脚	gə̱21
18		mountain	山	bə21	trotter	[猪]蹄	bə̱21
19		toward	去	lə21	drill	钻子	lə̱21
20		sun set	[太阳]落	də21	wear(shoes)	穿[鞋]	də̱21
21		have	[碗里]有	nə33	cut with scissors	剪v	nə̱33
22		dig	刨	dzə33	lie down	躺	dzə̱33
23	ie/ie̱	paper	纸	jie33	chicken	鸡	jie̱33
24		sister	姐姐	tɕie21	saw	锯子	tɕie21
25		iron	铁	ɕie21	breath	气	ɕie̱21
26		tile	瓦	phie21	leaf	叶子	phie̱21
27	o/o̱	brother	哥	ko33	complain	告状	ko̱33
28		tea leaf	茶叶	lo21	tiger	老虎	lo̱21
29		get sick	病	no21	how much	多少	no̱21
30		sound nice	好听	ʂo21	seek	寻找	ʂo̱21
31	i/i̱	heavy	重	li33	street	街	li̱33
32		slow, late	慢	phi21	throw up	呕吐	phi_21

33		water	水	ʑi21	sleep	睡觉	ʑi̱21
34		dog	狗	tɕʰi33	wink	眨眼	tɕʰi̱33
35	ɯ/ɯ̱	stove	灶	tɯ21/tə21	thoughts	一斗[米]or 想法	tɯ̱21/tə̱21
36		nod	低/点头	tɯ33/tu33	quantifier(one handful of) rice	一把[米]	tɯ̱33/tṳ33
37	ɿ/ ɿ̱	cough	咳嗽	tsɿ21	astringent	涩	tsɿ̱21
38		weeds	杂草	sɿ33	sweep	扫	sɿ̱33
39		seven/seed	七/种子	sɿ21	die	死	sɿ̱21
40		the day before yesterday	前天	sɿ33	screw	拧	sɿ̱33/s ɿ̱21
41		borrow	借	tsʰɿ33	grinding	磨[面]	tsʰɿ̱33

Phonation contrast on vowels

number	Vowels	English	Chinese	Gloss	English	Chinese	Gloss
42	i	give	给	bi21	heavy	重	li33
43	i̱	rupture	裂开	bi̱21	to market	赶街	li̱33
44	ɿ	twig	树枝	sɿ21	dog	狗	tsʰɿ33
45	ɿ̱	new/thirsty	新/渴	sɿ̱21	wink	眨	tsʰɿ̱33
46	ʅ	seven	七	ʂʅ121	wine	酒	dʐʅ 121
47	ʅ̱	screw	拧	ʂʅ33	sew	缝	dʐʅ̱=121
48	u	worm	虫	bu33	flow out	泻	fu33
49	u̱	full	饱	bu̱33	maggot	蛆	fu̱33
50	ɛ	cut	割	dɛ33	pick up	拿	kɛ33
51	ɛ̱	plain	淡（咸淡）	dɛ̱33	scold	骂	kɛ̱33
52	ə	mountain	山	bə21	sun set	（太阳）落	də21
53	ə̱	hoof	蹄	bə̱21	to fruit	结（果子）	də̱21
54	ɚ	bottle	瓶	bɚ21	pillar	柱	zɚ̱21
55	a	light	轻	la21	open	开	kʰa21
56	a̱	scoop up	捞	la21	carry on shoulder	扛	kʰa̱21
57	o	sell	卖	ɣo21	put	放	to33

58	o̠	obtain	得到	ɣo̠21	chop	剁（肉）	to̠33
59	ie	put into	装（进）	tie33	come	来	lie21
60	ie̠	hold	抱	tie̠33	hand	手	lie̠21

Phonation contrast on tones:

33L		33T		21L		21T		55L	
例词	音标	例词	音标	例词	音标	例词	音标	例词	音标
铜	dzɿ33	砍	dzɿ̠33	太阳	dzɿ21	断代	dzɿ̠21	皮	dzɿ55
借	tʂʰʅ33	热	tʂʰʅ̠33	臭	tʂʰʅ21	削	tʂʰʅ̠21	甜	tʂʰʅ55
矮	ti33	浸泡	ti̠33	顶、抵	ti21	砸	ti̠21	推	ti55
打闹	bɛ33	射	bɛ̠33	掉	bɛ21	纠缠	bɛ̠21	壶	bɛ55
粘	na33	把（量）	na̠33	你	na21	唠叨	na̠21	水果名	na55
吃	tʂo33	撞	tʂo̠33	中（打中）	tʂo21	罩	tʂo̠21	筛	tʂo55
倒	bɚ33	围起	bɚ33	山	bɚ21	蹄	bɚ=21	堆	bɚ55
虫	bu33	饱	bu̠33	背（东西）	bu21	霉	bu̠21	啼	bu55
说	ɣɯ33	喊	ɣɯ̠33	切（菜）	ɣɯ21	揉	ɣɯ̠21	舅舅	ɣɯ55
短/钝	də33	踩	də̠33	落（太阳落）	də21	结（果子）	də̠21	酒曲	də55
割（割肉）	dɛ33	淡（盐淡）	dɛ̠33	给	bi21	裂开	bi̠21		
朵	pu33	反（反面）	pu̠33	开门	kʰu21	扛	kʰu̠21		
放	to33	剁（剁肉）	to̠33	轻	lo21	捞	lo̠21		
天	mu33	吹（喇叭）	mu̠33	平/慢	pʰi21	吐（痰）	pʰi̠21		
软	nu33	羽毛	nu̠33	树枝	sɿ21	新/渴	sɿ̠21		
泻	fu33	蛆	fu̠33	斗（一斗）	tu21	点火/燃烧	tu̠21		
筋	dʑu33	害怕	dʑu̠33	箍儿	vɛ21	搓（搓绳）	vɛ̠21		
锄草	tʂʰu33	秋	tʂʰu̠33	干净	xu21	元（一元）	xu̠21		
				切（切菜）	ɣə21	针	ɣə̠21		
				二	nji21	压/按	nji̠21		

Appendix 2

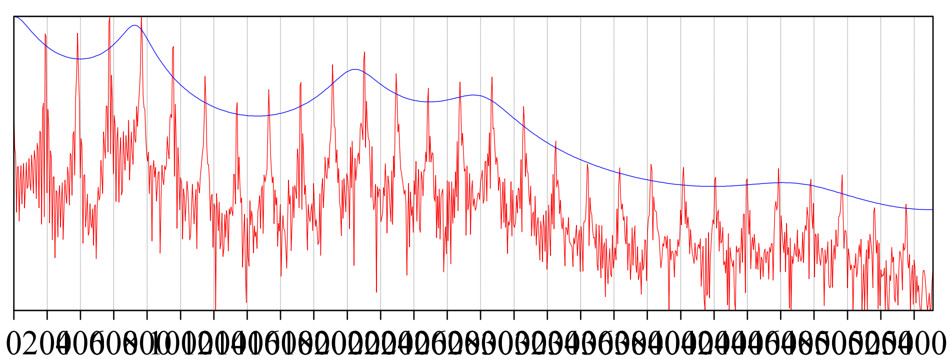
Manual check of formants and bandwidth:

The following items are pulled out: be21 be33 be21 be33 bu21 bu33 bu21 bu33, for 2 men and 2 women.

Formants and bandwidth were checked in Praat. The following plots come from one female and one male. Generally, the lax phonation has larger B1 values than the tense phonation. However, this distinction is only kept in the low tone. For the mid tone, there is no consistent trend.

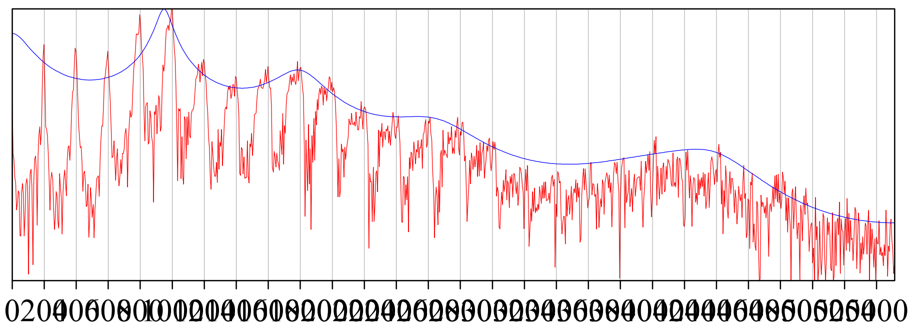
FEMALE:

be21



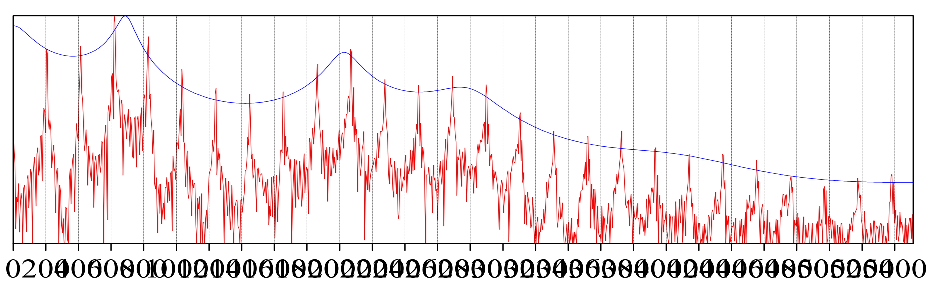
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	749	2001	2612	3385
Bandwidths:	178	259	357	2529

be21



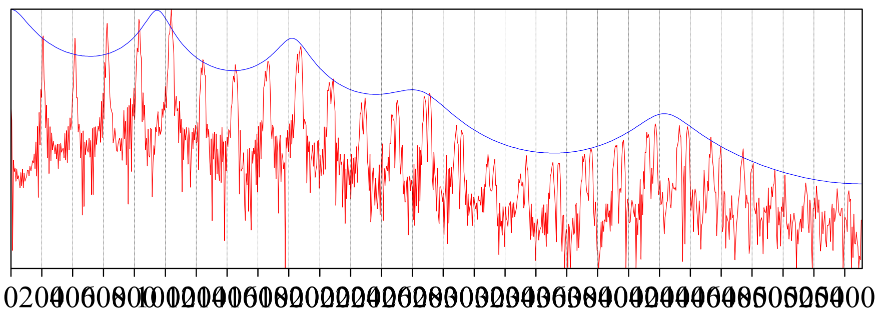
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	939	1786	2636	3892
Bandwidths:	94	270	590	618

be33



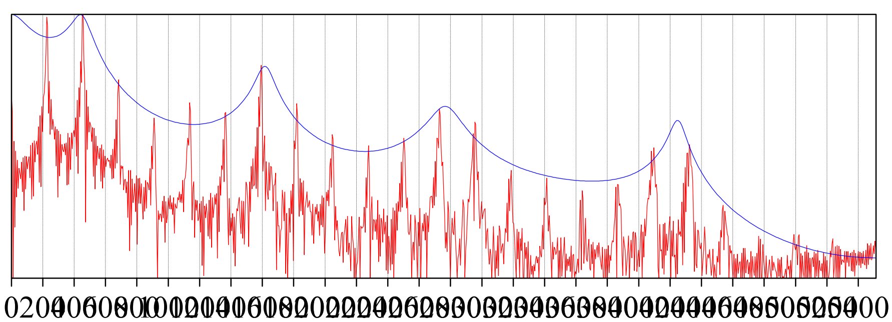
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	691	2030	2527	3106
Bandwidths:	107	181	3762	401

be̲33



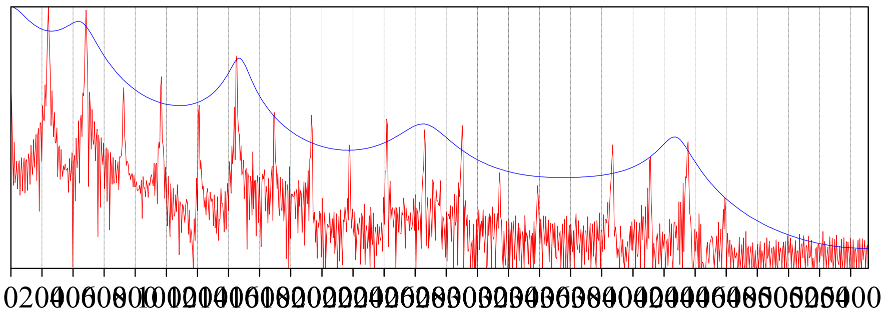
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	957	1821	2639	4215
Bandwidths:	139	184	336	485

bu33



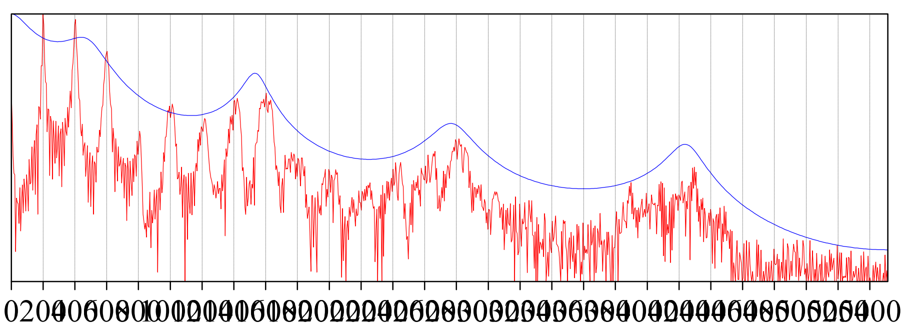
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	450	1611	2790	3426
Bandwidths:	100	100	173	1608

bu̲33



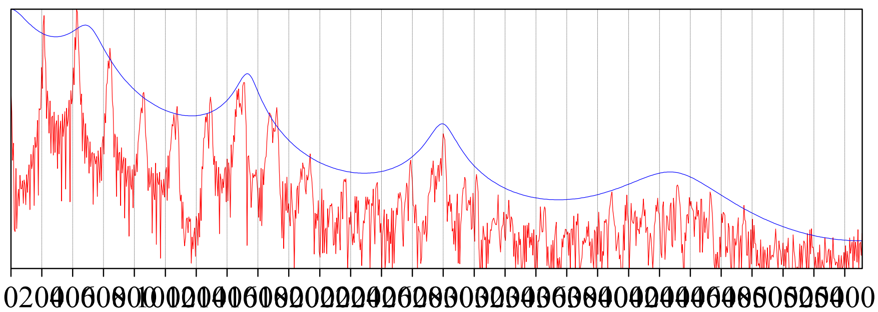
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	465	1475	2639	3681
Bandwidths:	163	111	270	964

bu21



Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	495	1543	2646	3736
Bandwidths:	259	493	106	1897

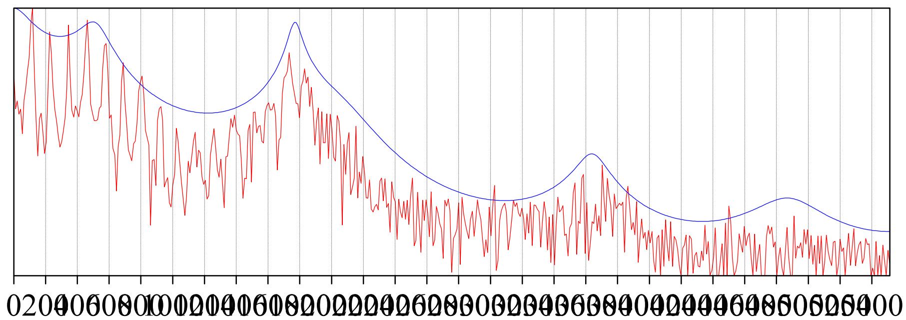
bu21



Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	500	1531	2785	4153
Bandwidths:	178	105	170	1633

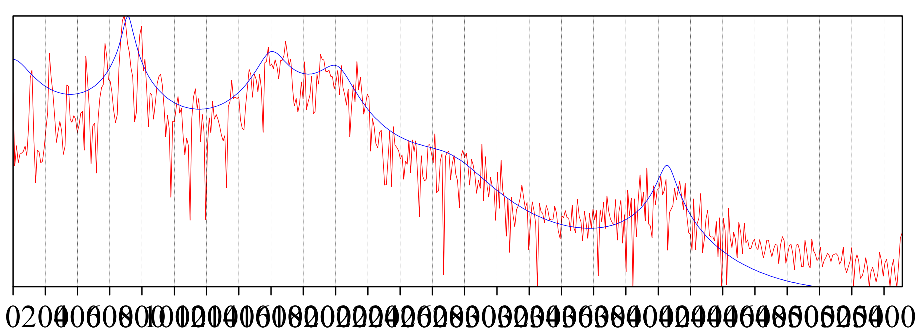
MALES:

be21



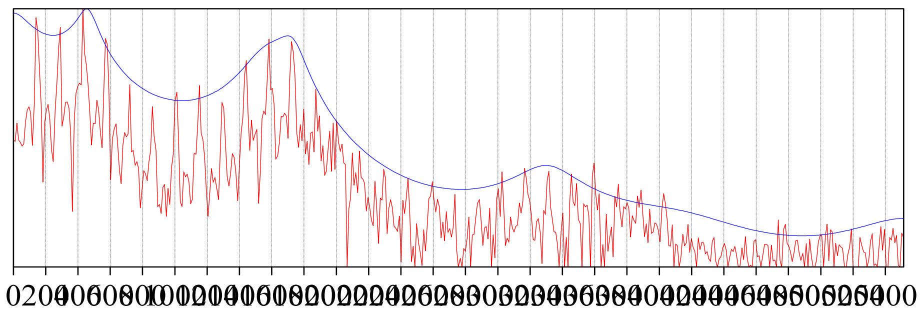
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	519	1765	1985	3635
Bandwidths:	144	121	741	208

be21



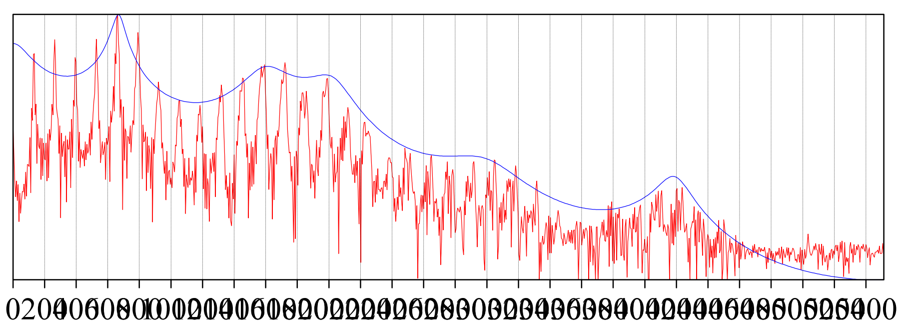
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	719	1619	2026	2787
Bandwidths:	69	177	181	598

be33



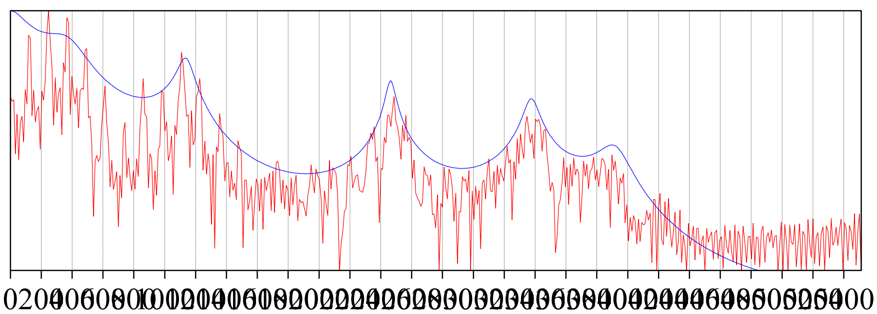
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	465	1562	1741	3339
Bandwidths:	126	312	140	508

be̲33



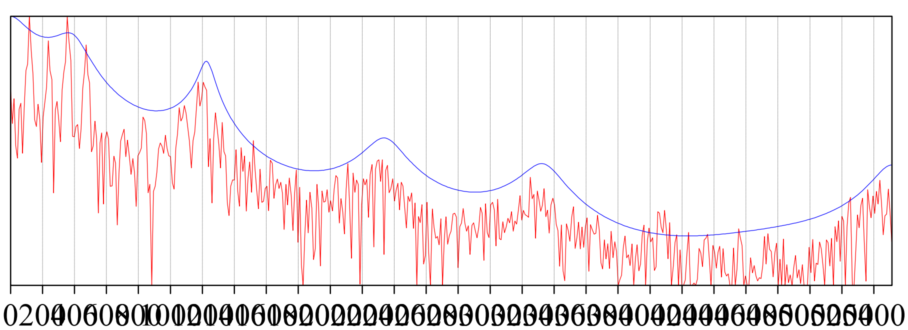
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	679	1581	1992	2996
Bandwidths:	73	229	252	369

bu21



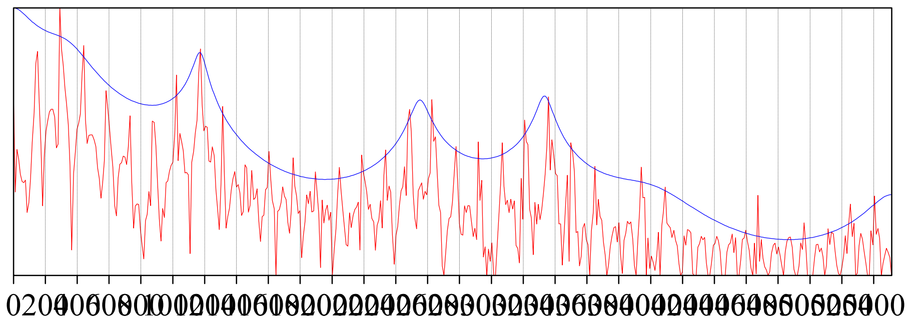
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	380	1151	2467	3374
Bandwidths:	249	97	60	91

bu̲21



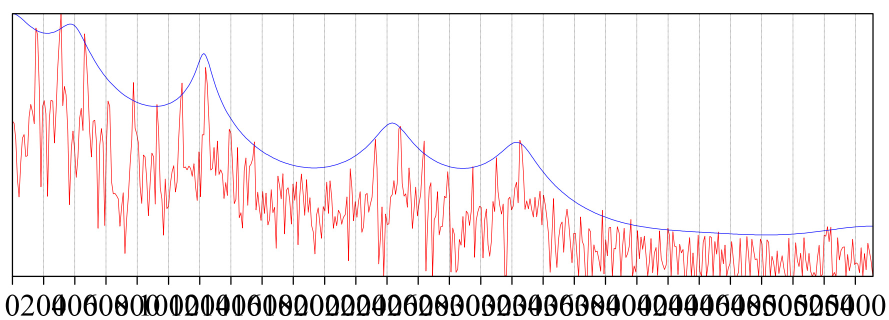
Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	388	1221	2298	3348
Bandwidths:	189	111	233	233

bu33



Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	353	1172	2554	3331
Bandwidths:	311	78	94	96

bu̲33



Formant Analysis:	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
Frequencies:	391	1225	2462	3255
Bandwidths:	156	88	118	145

BIBLIOGRAPHY

Baayen, R.H. (**2010**). languageR: Data sets and functions with "*Analyzing Linguistic Data: A practical introduction to statistics*". R package version 1.0. http://CRAN.R-project.org/package=languageR

Baer, T., Lofqvist A., McGarr, N.S. (**1983**). "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques", J. Acoust. Soc. Am. **73**(4), 1304-1308.

Baken, R. J. and Orlikoff, R.F. (**2000**). *Clinical measurement of speech and voice* (2nd. ed). San Diego, CA: Singular Publishing Group

Bates, D. and Maechler, M. (**2010**). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-37. http://CRAN.R project.org/package=lme4

Belotel-Grenié, A. and Grenié, M. (**2004**). "The creaky voice phonation and the organisation of Chinese discourse", In TAL-2004, 5-8.

Blankenship, B. (**1997**). "The time course of breathiness and laryngealization in vowels," Ph.D. dissertation, UCLA.

Bradley, D. (**1979**). "Proto-Loloish," in *Scandinavian institute of Asian studies Monograph* series 39. Curzon Press, London and Malm.

Chambers, J. M., Freeny, A. and Heiberger, R. M. (**1992**). "Analysis of variance; designed experiments," Chapter 5 of *Statistical Models in S,* edited by J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Childers, D. G., Moore, G. P., Naik, J. M., Larar, J. N., and Krishnamurthy, A. K. (**1983a**). "Assessment of laryngeal function by simultaneous, synchronized measurement of speech, electroglottography and ultra-high speed film," in *Transcripts of the eleventh symposium : Care of the professional voice,* edited by L. V. , New York : Voice Foundation, pp. 234–44.

Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore, G. P. (**1983b**). "Electroglottography, speech and ultra-high speed cinematography," in *Vocal fold physiology and biophysics of voice*, edited by I. Titze and R. Scherer , Denver, CO : Denver Center for the Performing Arts, pp. 202–20.

Childers, D. G. and Larar, J. N. (**1984**). "Electroglottography for laryngeal function assessment and speech analysis," IEEE Trans. Biomedical Engineering BME-**31**, 807–17.

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

Childers, D. G., Hicks, D. M., Moore, G. P., and Alsaka, Y. A. (**1986**). "A model for vocal fold vibratory motion, contact area, and the electroglottogram," J. Acous. Soc. Am. **80**, 1309–1320.

Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L. and Lalwani, A. L. (**1990**). "Electroglottography and vocal fold physiology," J. Speech Hear. Res. **33**, 245–254.

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

Diffloth, G. (**1980**). "The Wa Languages," Linguistics of the Tibeto-Burman area. **5.2**, pp. 1-182.

Edmondson J. A., Esling, J.,Harris J. G., Li, S. and Ziwo, L. (**2001**). "The aryepiglottic folds and voice quality in the Yi and Bai languages: laryngoscipic case studies," Mon-Khmer Studies, **31**, 83-100.

Edmondson, J. A. and Esling, John H. (**2006**). "The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies," Phonology, **23**(2), 157-191.

Edmondson, J. A., Padayodi, C. M., Hassan, Z. M. and Esling, J. H. (**2007**). "The laryngeal articulator: Source and resonator," In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, vol. 3 (pp. 2065-2068). Saarbrücken: Universität des Saarlandes.

EggWorks: http://www.linguistics.ucla.edu/faciliti/facilities/physiology/egg.htm

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 (in Chinese).

Esposito, C. M. (**2005**). "An Acoustic and Electroglottographic Study of Phonation in Santa Ana del Valle Zapotec," Poster presented at the 79th meeting of the Linguistic Society of America, San Francisco, CA.

Esposito, C. M. (**2006**). “The effects of Linguistic Experience on the Perception of Phonation”, Ph.D. dissertation, UCLA.

Esposito, C. M., Ptacek, J. and Yang, S. (**2009**). “An acoustic and electroglottographic study of White Hmong phonation,” Acoustical Society of America, San Antonio, TX.

Esposito, C. M. (**2010**). “The effects of linguistic experience on the perception of phonation,” J. Phonetics. **38**(2), 306-308.

Fant, G., Nord, L. and Branderud, P. (**1976**). “A note on the vocal tract wall impedance,” STL-QPSR 4/1976, 13-20.

Fant, G. (**1979a**). “Glottal source and excitation analysis,” STL-QPSR 1/1979, 85-107.

Fant, G. (**1979b**). “Vocal source analysis, a progress report,” STL-QPSR 3-4/1979, 31-53.

Fant, G. (**1982**). “The voice source - acoustic modelling,” STL-QPSR 4/1982, 28-48.

Fant, G. (**1986**). “Glottal flow, models and interaction,” From the 1985 symposium in Gotland, Sweden. Journal of Phonetics **4** (3/4) Theme issue, Voice Acoustics and Dysphonia, 393-399.

Fant, G. and Lin, Q. (**1987**). “Glottal source – vocal tract acoustic interaction,” STL-QPSR 1/1987, 13-27.

Fant, G. (**1997**). “The voice source in connected speech,” Speech Communication. **22**, 125-139.

Friendly, M. (**1994**). “A fourfold display for 2 by 2 by k tables,” Technical Report. **217**, York University, Psychology Department.

Gobl, C. and Ní Chasaide, A. (**1988**). “The effect of adjacent voiced/voiceless consonants on the vowel voice source: a cross-language study,” Speech Transmission Laboratory Quarterly Progress and Status Report **2-3**, 23-59.

Gobl, C. and Ní Chasaide, A. (**1992**). “Acoustic characteristics of voice quality,” Speech Communication**. 11**, 481-490.

Gordon, M & Ladefoged, P. (**2001**). “Phonation types: a cross- linguistic overview,” J. Phonetics, **29**, 383–406.

Garellek, M. and Keating, P. (**2010**). “The acoustic consequences of phonation and tone interactions in Jalapa Mazatec,” (Submitted).

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

Hanson, H. (**1997**). “Glottal characteristics of female speakers: Acoustic correlates,” J. Acous. Soc. Am. **101**, 466-481.

Hanson, H. (**1999**). “Glottal characteristics of male speakers: Acoustic correlates and comparison with female data,” J. Acous. Soc. Am. **106** (2) 1064-1077.

Hanson, H. M., Stevens, K. N., Kuo, H. J., Chen, M. Y. and Slifka, J. (**2001**). “Towards models of phonation,” J. Phonetics. **29**, 451-480.

Herbst, C. and Ternstrom, S. (**2006**). “A comparison of different methods to measure the EGG contact quotient,” Logopedics Phoniatrics Vocology **31**,126-138.

Henrich, N., d'Alessandro, C., and Doval, B. (**2001**). "Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data," Proc. of Eurospeech, pp. 47-50.

Henrich, N., d'Alessandro, C., Castellengo, M. and Doval, B. (**2004**). “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” J. Acous. Soc. Am. **115** (3), 1321-1332.

Hillenbrand, J.M., Cleveland, R.A. and Erickson, R.L. (**1994**). “Acoustic correlates of breathy vocal quality,” J. Speech Hear. Res. **37**, 769-778.

Hillenbrand, J.M., Clark, M.J., and Houde, R.A. (**2000**). “Some effects of duration on vowel recognition,” J. Acous. Soc. Am. **108**, 3013-3022.

Holmberg, E.B., Hillman, R. E., Perkell, J. S. , Guiod, P.C. and Goldman, S.L. (**1995**). “Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice,” J. Speech Hear. Res. **38**, 1212-1223.

Holt, L. L. and Lotto, A. J. (**2006**). “Cue weighting in auditory categorization: Implications for first and second language acquisition,” J. Acous. Soc. Am. **119**, 3059-3071.

Hombert, Jean-Marie, Ohala, J. J. and Ewan W. G.(**1979**). “Phonetic explanations for the development of tones,” Language **55**. 37–58.

Hosmer, D. W. and Lemeshow, S. (**2000**). *Applied Logistic Regression*, 2nd ed. New York; Chichester, Wiley.

Howard, D. M., Lindsey, G. A. and Allen, B. (**1990**). "Toward the quantification of vocal efficiency. J. Voice **4**(3), 205-12.

Howard, D. M. (**1995**). "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," J. Voice **9**(2): 163-72.

Huffman, M. K. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495–504.

Iseli, M., Shue, Y.-L. and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," J. Acoust. Soc. Am.**121**, 2283-2295.

Iverson, P. and Kuhl, P. K. (**1996**). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," J. Acoust. Soc. Am. **99**, 1130–1140.

Jiang, J., Alwan, A., Keating P., Auer, E. and Bernstein, L. (**2007**) "Similarity structure in visual speech perception and optical phonetic signals," Perception and Psychophysics **69** (7), 1070-1083.

Johnson, K. (**2003)**. *Acoustic and auditory phonetics*. Oxford: Blackwell Publishing.

Keating, P., Esposito, C., Garellek, M., Khan, S. and Kuang, J. (**2010**) "Phonation contrasts across languages," poster presented at LabPhon12 in Albuquerque NM.

Khan, S. D. (**2010**). "Breathy phonation in Gujarati: an acoustic and electroglottographic study". Poster presented at the 159th meeting of the Acoustical Society of America, Baltimore, 23 April.

Kirk, P., Ladefoged, P. and Ladefoged, J. (**1984**). "Using a spectrograph for measures of phonation types in natural language", UCLA Working Papers in Phonetics **59**, 102-113.

Kirk, P. L., Ladefoged, J. and Ladefoged, P. (**1993**). "Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec," In *American Indian Linguistics and ethnography in honor of Lawrence C. Thompson*. A. Mattina and T. Montler, editors., Missoula, MT: Universtiy of Montana Press.

Klatt, D. H. and Klatt,L. C. (**1990**). “Analysis, synthesis and perception of voice quality variations among female and male talkers,” J. Acoust. Soc. Am. **87**, 820–857.

Kreiman, J., Gerrat, B. G. and Antoñanzas-Barroso, N. (**2007**). “Measures of the glottal source spectrum,” J. Speech Lang. Hear. Res. **50**:595-610.

Krishnan, A., and Gandour, J. T. (**2009**). “The role of the auditory brainstem in processing linguistically-relevant pitch patterns,” Brain and Language, **110**,135–148.

Kreiman, J., Gerrat, B. G. and Khan, S.(**2010**) “Effects of native language on perception of voice quality,” J. Phonetics , doi: 10.1016/j.wocn.2010.08.004.

Kruskal, J.B. and Wish, M. (**1978**). *Multidimensional Scaling*. Sage Publications.

Ladefoged P. (**1964**). *A phonetic study of west African languages*. Cambridge: Cambridge University.

Ladefoged, P. (**1971**). *Preliminaries to linguistic phonetics*. Chicago: The University of Chicago Press.

Ladefoged, P. (**1983**). “The linguistic use of different phonation types,” *Vocal fold physiology. Contemporary research and clinical issues*, edited by Diane M. Bless and James H. Abbs, 351-360. San Diego: College-Hill Press.

Ladefoged, P., Maddieson, I. and Jackson, M. (**1988**). “Investigating phonation types in different languages,” In: O. Fujimura, Editor, *Vocal fold physiology: voice production, mechanisms and functions*, pp. 297-317. New York: Raven Press.

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

Laver, J. (**1981**). *Phonetic Description of Voice Quality*. Cambridge: University of Cambridge Press.

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

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**,103-9.

Michaud, A. (**2004**). “A measurement from electroglottography: EDCPA, and its application in prosody," ISCA speech prosody 2004.

Nair, G. (**1999**). Voice Tradition and Technology: A State-of-the-Art Studio.

Ní Chasaide, A. and Gobl, C. (**1997**). "Voice Source Variation," *Handbook of Phonetic Sciences*, ed. by William J. Hardcastle and John Laver, 427-461.

Orlikoff, R. F. (**1991**). "Assesment of the Dynamics of Vocal Fold Contact From the Electroglottogram: Data from Normal Male Subjects," J. Speech Hear. Res. **34**, 1066-1072.

R Development Core Team (**2010**). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rothenberg, M. and Mahshie, J. J. (**1988**). "Monitoring vocal fold abduction through vocal fold contact area," J. Speech Hear. Res. **31**, 338-51.

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

Samely, U. (**1991**). *Kedang, (Eastern Indonesia), some aspects of its grammar*. Hamburg: Helmut Buske Verlag.

Shue, YL, Keating, P. and Vicenik, C. (**2009a**) "VoiceSauce: A program for voice analysis," poster presented at the Fall 2009 meeting of the Acoustical Society of America in San Antonio.

Shue, Y. L., Kreiman, J. and Alwan, A. (**2009b**) "A novel codebook search technique for estimating the open quotient," Proceedings of Interspeech, Brighton, UK, pp. 2895–2898, August 2009.

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

Stevens, K. N. (**1988**). "Modes of vocal fold vibration based on a two-section model," In O.Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms, and function* (pp. 357–371). NewYork: Raven Press.

Stevens, K. N. and Hanson, H. (**1995**). "Classification of glottal vibration from acoustic measurements," In Fujimura, O. and Hirano, M. (Eds.), *Voice quality control* (pp. 335–342). San Diego: Singular Publishing Group Inc.

Swerdlin, Y., Smith, J. and Wolfe, J. (**2010**) “The effect of whisper and creak vocal mechanisms on vocal tract resonances,” J. Acoust. Soc. America. **127**, 2590-2598.

Tiede. (**1993**). “An MRI-based study of pharyngeal volume contrasts in Akan,” Haskins Laboratories Status Report on Speech Research SR-**113**,107-130

Thongkum, T. (**1987**). “Phonation types in Mon-Khmer languages,” UCLA Working Papers in Phonetics **67**.

Venables, W. N. and Ripley, B. D. (**2002**) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

West, B., Welch, K. B. and Galecki, A. T. (**2006**). *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: Chapman & Hall/CRC.

Wheatley, J. K. (**1982**). “Comments on the Hani dialects of Loloish,” Linguistics of the Tibeto-Burman Area **7.1**, 1-38.

Yu, K.M., Lam, H. M. and Li, S. Y. (**2010**). “An acoustic and electroglottographic study of Cantonese tone,” Poster, Acoustical Society of America, Baltimore, MD.

Ziwo, L., Li, S., Edmondson, J. A., Esling, J. H. and Harris, J. G. (**2001**). “The aryepiglottic folds and voice quality in the Yi and Bai languages: laryngoscopic case studies,” Mon-Khmer Studies **31**,83-93.

鲍怀翘，周植志（Bao, Huaqiao and Zhou, Zhizhi）(**1990**).佤语浊送气声学特征分析,《*民族语文*》1990（2）62-70.

鲍怀翘, 吕士楠（Bao, Huaqiao and Lu, Shinan）(**1992**).蒙古语察哈尔元音的声学分析,《*民族语文*》1992(1) 61-68.

戴庆夏（Dai, Qingxia）(**1979**). 我国藏缅语族松紧元音来源初探,《*民族语文*》1979（1）31-39.

戴庆夏（Dai, qingxia）(**1993**).关于纳西语的松紧元音问题——兼论彝缅语语音历史演变的研究方法,《*民族语文*》1993(1) 27-36.

戴庆夏, 杨春燕（Dai, Qingxia and Yang, Cunyan）(**1994**). 景颇语两个语音特点的统计分析,《*民族语文*》1994(5) 24-43.

盖兴之(Gai, Xingzhi) (**1994**). 藏缅语的松紧元音,《*民族语文*》1994(5) 49-53.

江荻 (Jiang, Di) (**2001**). 藏缅语言元音的上移和下移演化,《*民族语文*》2001(5) 14-27.

孔江平 (Kong, Jiangping) (**1996**). 哈尼语发生类型声学研究及音质概念的讨论,《 *民族语文*》1996(1) 40-46.

孔江平 (Kong, Jiangping) (**2001**). 《*论语言的发声*》(On Language Phonation), 中央民族大学出版社.

马学良 (Ma, Xueliang) (**2003**). 《*汉藏语概论*》(An introduction to Sino-Tibetan Languages), 民族出版社.

石锋，周德才(Shi, Feng and Zhou, Decai) (**2005**). 南部彝语松紧元音的声学表现,《*语言研究*》 2005(1) 60-65.

周德才(Zhou, Decai) (**2005**). 彝语方言松紧元音比较研究,《*云南民族大学学报*》22 (5)152-155.

周植志，颜其香 (Zhou, zhizhi and Yan, Qixiang) (**1984**). 《*佤语简志》*, 民族出版社.

朱晓农, 周学文(Zhu ,Xiaonong and Zhou, Xuewen) (**2008**) 嘎裂化：哈尼语紧元音,《*民族语文*》2008(4) 9-18.