UCLA Working Papers in Phonetics, No.111 (2012), pp. 123-161

Glottal articulations of phonation contrasts and their acoustic and perceptual consequences[1]

Jianjing Kuang and Patricia Keating
(kuangjianjing@ucla.edu, keating@humnet.ucla.edu)

Abstract

This study explores the properties of one type of phonation contrast - the tense vs. lax phonation contrasts of Yi (Loloish) languages – in terms of their glottal articulations, acoustic correlates, and perceptual salience. To determine the glottal articulations involved in the phonation contrasts, we adopted Functional Data Analysis to analyze entire EGG glottal pulse shapes. This method is found to capture differences in the abruptness of contact, a key property that is not captured by traditional EGG parameter measures. The primary contact patterns for phonation contrasts are very consistent across languages, speakers, and genders, and there is only one underlying articulatory pattern for these tense vs. lax contrasts. Overall, we found consistent correlations among the different kinds of measures, which means that speakers and listeners are able to establish a stable link between articulation, acoustic signals and perception.

1. Introduction

Voice qualities are used in many languages allophonically, as prosodic cues (e.g. stress and focus in German: Mooshammer, 2010) or as enhancement cues of other distinctive features (e.g. creak in Mandarin low dipping tone: Belotel-Grenié and Grenié 1994). But relatively few languages around the world use phonation by itself as a phonemic dimension -- that is, where voice quality distinguishes lexical meanings just as consonants, vowels, and tones do. Various types of phonation contrasts have been attested across languages. Some languages contrast two phonation types, e.g. breathy and modal in Gujarati (Fischer-Jørgensen, 1967; Khan, 2012), while other languages contrast three phonation types, e.g. breathy, modal, and creaky in White Hmong (Ratliff, 1992; Esposito, 2012)[2] and in Mazatec (Kirk et al., 1993; Garellek and Keating, 2011), or even four (e.g. Ju|'hoansi: Miller 2007). The type of phonation contrast that will be the focus

---

[1] This work was supported by NSF grant BCS-0720304 to Prof. Patricia Keating. Thanks to Yan-Liang Shue for VoiceSauce and Henry Tehrani for EggWorks. Fieldwork was largely helped by Prof. Jiangping Kong and Feng Wang of Peking University.

[2] Huffman (1987) and Garellek et al. (2012) have treated the creaky phonation as non-contrastive.

[3] Most such languages, including Southern Yi, also have a high 55 tone, but it occurs only with lax phonation

[2] Huffman (1987) and Garellek et al. (2012) have treated the creaky phonation as non-contrastive.

[4] Note that the kind of tense vs. lax contrast discussed here, a phonation contrast, is different from the so-

of this chapter is the two-way tense vs. lax phonation contrast found in some Tibeto-Burman languages. For example, in Southern Yi (a Tibeto-Burman language spoken in China), the syllable /be^{21}/ with a lax phonation means “mountain”, whereas /be̱^{21}/ with a tense phonation means “foot”. (Here, tense phonation is indicated by an underscore, and the superscript numbers indicate the tone). These Tibeto-Burman languages are also tonal, and both kinds of phonation occur with mid and low tones[3]. For example, Southern Yi has not only /be̱^{21}/ (“foot”) vs. /be^{21}/ (“mountain”), but also /be̱^{33}/ (“shoot”) vs. /be^{33}/ (“fight”). That is, tone and phonation type are independently contrastive in these languages. This study investigates the production of the tense vs. lax phonation contrasts in three of these languages: Southern Yi, Bo and Luchun Hani.

Tense vs. lax phonations[4] are generally considered to be more like modal phonation than are creaky and breathy phonations. In terms of Ladefoged’s (1971) continuum model, which conceptualizes phonations in terms of glottal strictures that fall in between the most open (voiceless) and the most closed (glottal stop), tense phonation is a phonation in the creaky-modal range whereas lax phonation is a phonation in the breathy-modal range. A slightly more complex version of this idea emerges from a recent cross-linguistic study of 24 phonation categories from ten languages (Keating *et al.*, 2012) – e.g. the two categories of Gujarati and the three of Hmong. In this study, many acoustic measures of phonation were rendered in a 2-dimensional physical space by multi-dimensional scaling, and the 24 language categories organized in the space, as seen in Figure 1. The tense and lax categories of the three Tibeto-Burman languages in the present study are included in Figure 1: Southern Yi’s phonations are labeled YiT and YiL, Luchun Hani’s are labeled LuT and LuL, and Bo’s are labeled BoT and BoL.[5] The five broad cross-language phonation types of the Ladefoged continuum model are also labeled: breathy, lax, modal, tense, and creaky. It can be seen that in this acoustic space, the 5 types form a V-shaped pattern. Dimension 1 in the space goes from least modal (breathy, creaky) to most modal, while Dimension 2 is like a rather crowded glottal stricture continuum.[6] On the two dimensions combined, the differences among the phonation types are more apparent than on either dimension alone. Nonetheless, it can be seen that the tense and lax phonations

---

[3] Most such languages, including Southern Yi, also have a high 55 tone, but it occurs only with lax phonation.

[4] Note that the kind of tense vs. lax contrast discussed here, a phonation contrast, is different from the so-called tense vs. lax vowel contrast in Germanic languages, which is mainly related to the muscular tension in the tongue (Ladefoged, 1964), and at least in German does not involve consistent glottal and pharyngeal articulations (Marasek 1996,1997). 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.

[5] The other languages are Zapotec (3 categories), Black Miao (4 categories), Mandarin (2 categories – creaky Tone 3 vs. other tones), and English (1 category). The Mazatec breathy and Zapotec creaky categories are outliers, different from others of those types.

[6] Dimension one is related to spectral measures in the mid-frequency range, e.g. H1*-A1*, H1*-A2*, and dimension two is related to spectral measures in the low-frequency range, e.g. H1*-H2*.

cluster near each other, suggesting that they share acoustic phonetic properties; and it can be seen that they both are similar to the modal phonations (much nearer to modal than to the breathy or the creaky cases), supporting the traditional idea that they are not extremes of non-modal phonation. These similarities lead us to explore more thoroughly the basis of these contrasts.

The main goal of the present study is thus to better understand Tibeto-Burman tense vs. lax phonation contrasts. We will extend the acoustic methods used by Kuang (2011) to analyze Southern Yi to two additional languages. We will also analyze in detail electroglottographic (EGG) recordings from the three languages, in order to understand better how these contrasts are produced. In order to understand the glottal events in tense and lax phonations, we need in turn to better understand the information provided by EGG signals: how do EGG pulse landmarks and EGG pulse shapes relate to glottal events, to acoustic measures, and to perception of phonation? In this connection, we will follow Mooshammer (2010) in applying Functional Data Analysis to EGG pulse shapes, and we will relate the components of the shapes obtained in this way to other production measures, as well as to perception. Thus the secondary goal of the present study is to contribute to the understanding of EGG measures of phonation type.

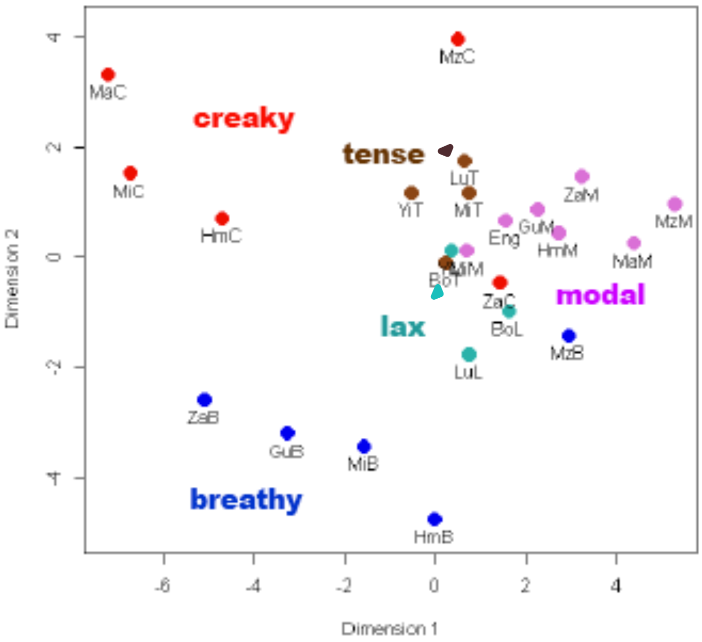

FIG. 1. Acoustic space for 24 phonation categories from ten languages (from Keating *et al*. 2012). The first 2 letters of each label indicate the language: Bo, Gujarati, Hmong, Luchon Hani, Mandarin, Miao, Mazatec, Yi; the third letter, plus the color, indicates the phonation (breathy/blue, creaky/red, lax/turquoise, modal/lavender, tense/brown); Eng indicates English. The circled phonation types are the tense vs. lax. They are similar to each other, and near the modal range.

2. Background

2.1 Languages

Southern Yi, Bo, and Hani are Yi (also called Loloish) languages in the Tibeto-Burman family of the Sino-Tibetan phylum. The name "Yi" refers to both the whole Yi (Loloish) branch of languages and the Yi language, because it has the most population in this language family branch. The Yi language branch includes perhaps fifty languages, for example, Yi, Hani, Lisu, Lahu, Naxi and Bo, among many others. Yi languages are geographically distributed in Yunnan, Sichuan and Guizhou provinces of China, and are spoken by more than six million people (Ethnologue, 2012). Yi languages typically have a CV syllable structure, a seven-vowel system and a three-tone system (low, mid and high, noted as 21, 33 and 55), and two phonation registers: tense vs. lax.

2.2 Production of phonation types

Linguistic voice quality has been defined along the continuum of the aperture between the arytenoid cartilages (Ladefoged, 1971; Gordon and Ladefoged, 2001). Four types of non-modal phonations have been widely referred to: Breathy, creaky, lax and tense (ref. Figure 1). Breathy phonation is said to involve minimal adductive tension, weak medial compression and low longitudinal tension (Laver, 1980; Ní Chasaide and Gobl, 1997). There is an increase in the aperture between the vocal folds, such that the posterior portion to the midline of one vocal fold never comes in contact with the other fold (Laver, 1991). The constant glottal leakage often leads to audible frication noise. Lax phonation has a similar glottal configuration as breathy phonation, but is less extreme. Breathy phonation has a greater vocal fold aperture and higher amplitude of noise components (Pennington, 2005). A "slack" or "lax" configuration of the vocal folds often leads to lower pitched and softer sounds (Ní Chasaide and Gobl, 1997)

Both creaky and tense phonations are characterized by increased adductive tension and medial compression (Ní Chasaide and Gobl, 1997), which leads to decreased aperture between the vocal folds. Because of the high adductive tension, only the ligamental part of the vocal folds is vibrating. Tense phonation involves a higher degree of tension in the entire vocal tract as compared to the neutral setting. The increased muscular tension associated with tense phonation is likely to affect the respiratory system as well as the supralaryngeal tract (Ní Chasaide and Gobl, 1997). The major difference between tense phonation and creaky phonation is whether the vocal folds have periodic vibration (Childers *et al.*, 1990); tense phonation is also closer to modal phonation (ref. Figure 1).

2.3 Electroglottographic measures of phonation types

Electroglottography (EGG) is a popular method which measures variations in the vocal fold contact area during phonation. A small, high frequency current is passed between two electrodes that are placed on each side of the larynx. Variation in the electrical impedance across the larynx is produced by the opening and closing of the vocal folds. The EGG signal is related to the contact area of the vocal folds: the larger the contacted area, the larger the measured admittance. Since the signal is not calibrated, it reflects relative rather than absolute contact. EGG is non-invasive, and does not interfere with speakers' natural production (compared to invasive methods); it can thus be used to study complex speech events, and is convenient for use outside the laboratory. In recent years it has been widely used in studies of linguistic phonation, and plays an important role in documenting uses of non-modal phonations in various under-described languages (e.g. Maa (Guion *et al*., 2004), Santa Ana Del Valle Zapotec (Esposito, 2010), Tamang (Mazaudon and Michaud, 2006, 2008), Takhain Thong Chong (DiCanio, 2009), Gujarati (Khan, 2012), and White Hmong (Esposito, 2012)). Therefore, it is worth understanding more precisely the meaning of the signals acquired by EGG.

EGG signals are most commonly analyzed in terms of a parameter which reflects the relative amount of vocal fold contact during each single vibratory cycle; this parameter is often known as the "Contact Quotient", or CQ[7] (Rothenberg *et al*., 1988; Baken and Orlikoff, 2000). The Contact Quotient is defined as the ratio of the duration of the contact phase to the period of the vibratory cycle. Although Hanson (2012) claims that the EGG signal "has not been found to correlate with voice quality", previous studies have in fact shown that CQ varies with linguistic voice quality. Tense, creaky or other laryngealized phonations usually show a greater CQ, reflecting a more closed glottis; breathy phonations usually show a smaller CQ, as the glottis usually has a wide resting aperture. For example, Mooshammer (2010) found that CQ reliably varies with lexical stress and vocal effort in German, though an earlier study by Marasek (1996) did not find such an effect.[8] CQ shows stronger and consistent patterns for contrastive linguistic phonations: for example, CQ distinguishes breathy from creaky and modal phonations in White Hmong (Esposito, 2012), breathy vs. modal in Gujarati (Khan, 2012), three-way contrasts in Santa Ana del Valle Zapotec (Esposito, 2010), four-way phonation contrasts in Takhian Thong Chong (DiCanio, 2009), and even the allophonic creaky vs. modal phonations in Mandarin (Keating *et al.*, 2012).

Nonetheless, CQ is not unproblematic. It depends on estimates of the closing and opening moments in the EGG cycle, yet it is hard to unambiguously define these landmarks. Many efforts have been made to compare EGG signals with more direct physiological signals, for example, subglottal pressure (Kitzing *et al.*, 1982) or various imaging techniques (Anastaplo and Karnell, 1988; Karnell, 1989; Baer et al., 1983a,b; Childers *et al.*, 1990; Berke *et al.*, 1987; Herbst and Ternström, 2006; Henrich *et al.*, 2004; among

---

[7] In some studies, CQ is referred to as the "closed quotient" (e.g. Nair 1999), but this term is possibly misleading since EGG cannot necessarily detect full closure vs. opening.

[8] This discrepancy might be in part due to their different methods of defining CQ, an issue addressed below.

many others). While the derivative of the EGG signal usually shows a very strong positive peak, which is taken to be a clear and strong indicator of vocal fold closing, the negative peak is not so clear, making the moment of vocal fold opening much less certain. Therefore various thresholds have been employed to define opening (and sometimes also closing) events; however, studies have disagreed on what threshold to use. It appears that different thresholds might be sensitive to different types of phonation (Herbst and Ternström, 2006; Marasek, 1997). For example, a higher threshold is better for creakier voice and a lower threshold is better for breathier voice (Herbst and Ternström, 2006). Therefore currently several different methods of defining CQ are in use, reviewed in section 4.1.1.

Another potentially important parameter of vocal fold vibration is the Speed Quotient (SQ; Holmberg *et al.*, 1988; Dromey *et al.*, 1992), defined as the ratio between closing duration and opening duration, and thus a measure of the symmetry of the glottal pulses. As creaky phonations are usually shorter in closing and longer in opening, the EGG pulse shows a steeper slope in the closing phase, and thus the pulse shape is skewed. On the other hand, breathy phonations usually have similar closing and opening duration, so the pulse shows a more symmetrical shape. Note that the SQ derived from EGG signals is not directly comparable to the SQ derived from glottal flow, since flow signals reflect degree of opening while EGG signals reflect degree of contact. Indeed, although SQ of glottal flow has been found to consistently increase in higher vocal effort, SQ from EGG signals cannot replicate this result (Dromey *et al.*, 1992). Nonetheless, Esling (1984) found that skewness of EGG signals is a useful indicator of phonation types. But more recent studies (e.g. Mooshammer (2010) on German stress, Keating *et al.* (2012) on phonation contrasts across languages) have not found that EGG SQ reliably varies with phonation in the way that CQ does. Therefore, while SQ is less subject to measurement uncertainties (though still affected by the selection of a threshold), it seems to be less reflective of differences in phonation.

In addition to SQ and CQ, a new measure has been employed by Michaud (2004), related to earlier measures of average rate of change in increasing contact (see Baken and 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 rate of increase of vocal fold contact. “Peak increase in contact” (PIC) is a more transparent name for this measure (Keating *et al.,* 2011). This peak increase is thought to be reached at the instant of glottal closure, and it is often assumed that faster closure results in more high-frequency energy in the voice spectrum. This measure has the advantage of not depending on landmarks in the EGG pulse. Michaud found that in Naxi (another Yi language), syllables under emphasis have higher values of PIC, and sometimes (especially for Low and Mid tones) that was the only acoustic correlate of focus; but for creaky tones, syllables in focus have lower values of PIC. He proposed that the interpretation of PIC values thus depends on the "laryngeal mechanism" used to produce a given pitch range or phonation type. Later studies of contrastive phonation types in other languages have shown that breathy, creaky, and/or modal phonation categories differ in PIC, with breathy voice having the highest values (e.g. Kuang, 2011b; Esposito, 2012; Keating *et al.*, 2012).

Researchers (Titze 1984, 1989, 1990; Childers *et al.*, 1986; Larson *et al.*, 1994; among many others) have found that changes in certain basic geometric properties of EGG pulses reflect specifiable changes in vocal fold adjustment and behavior, especially the contact patterns (see Baken and Orlikoff 2000: 422 for detailed review). Modeling studies (e.g. Marasek, 1997) have tried to describe the contact patterns of the vocal folds in terms of the entire EGG pulse shape. However, since Marasek used annotated landmarks to represent the pulse shape, this approach is problematic in cases where the landmarks are not detectable. Moreover, Marasek found that there is a great deal of individual and gender variability in EGG pulses, which make it harder to find consistent properties of voice qualities. Both of these shortcomings make the modeling of EGG glottal pulses very tricky.

In a different approach to overcoming the difficulties in defining landmarks, and to quantify the variability of contact patterns indicated by EGG glottal pulses, more advanced statistical approaches have been explored. Makhtari *et al*. (2003) and Kreiman *et al*. (2007) applied principle component analysis (PCA) to glottal pulses derived from audio pulses by inverse filtering. For example, Kreiman *et al*. (2007) identified five pulse shape factors that are important for voice quality: steepness of the opening (Asymmetry), open quotient, shape of return to zero, shape and duration of closing and opening. These factors are highly correlated with LF voice source parameters. More recently, Mooshammer (2010) adopted Functional Data Analysis (FDA) to study EGG pulse shapes as a function of lexical stress in German; she found that EGG pulses from stressed syllables have a shorter open phase and a steeper rise in the pulse.

The FDA statistical technique, which was developed by Ramsay (1982, also Ramsay and Silverman 1997, 2002) and has been a powerful approach in analyzing the patterns of articulatory movements (e.g. lips: Ramsay *et al*., 1996; tongue: Lee *et al*., 2006), can consider entire and continuous EGG pulses from different speech conditions by comparing the overall pulse shapes, without the need to define landmarks in the signal. FDA involves the definition of traditional principle component analysis in functional analytic terms (i.e. a functional version of PCA), and "expresses the modes of variation of trajectories in a form similar to the trajectories themselves" (Ramsay *et al*., 1997). According to Ramsay, this approach has several advantages: 1) it takes account of the underlying continuity of the physiological system generating the behavior; 2) it displays temporal and spatial dependencies of articulators; 3) it can quantitatively factor the separable components of complex multidimensional time series data. We will adopt this method to describe the contact patterns in tense vs. lax phonation contrasts, as it can factor out non-phonation variability in the glottal pulses (e.g. gender, speaker), and reveal the essential gestural patterns of phonation contrasts without the problem of defining landmarks.

In sum, EGG has the benefits of being non-invasive and not interfering with speech, and thus practical for linguistic fieldwork, and measures derived from landmarks or other key points in the signal or its derivative have been shown to be useful in characterizing phonation types. However, the EGG signal is complex, and its relation to phonation is not as well understood as we would like, especially in the face of large inter-speaker and inter-gender variability. Therefore we will apply FDA to EGG pulse shapes.

2.4 Previous studies of Tibeto-Burman phonation types

A few previous studies have examined a small set of acoustic correlates associated with tense vs. lax contrasts in Tibeto-Burman languages. Maddieson and Ladefoged (1985) and Maddieson and Hess (1986) investigated tense vs. lax contrasts in four Tibeto-Burman languages (Hani, Eastern Yi, Jingpo, and Wa) and found that despite the variation in other acoustic correlates (e.g. F0, F1, duration, and VOT), the amplitude difference between the second harmonic and the fundamental frequency (H1-H2) is consistently higher for the lax phonation across languages. Meanwhile, the lax phonation was also found to have a higher rate of airflow and pressure. Kong (2001) found that H1-H2 is a successful measure to distinguish tense vs. lax phonation contrast in several Tibeto-Burman languages (e.g. Northern Yi, Zaiwa, Jingpo), though H1-A1 and H1-A2 are better measures for Northern Yi. A laryngoscopic study (Esling *et al.*, 2000) showed that a harsh quality of the tense phonation of Northern Yi can be partially attributed to retraction of the tongue root. Similarly, a recent acoustic study of Southern Yi (Xinping village, Shi and Zhou, 2005) also showed that H1-H2 is an important acoustic correlate of phonation contrasts, and they found that F1 is consistently higher for the vowel with tense phonation. Kuang (2011) examined both electroglottographic and acoustic properties of tense vs. lax contrast in Southern Yi (the same corpus used here, in part from Xinping village), and found that H1-H2 and H1-A1 are the best acoustic correlates of the phonation contrast, and furthermore that these acoustic measures both correlate with the contact quotient of vocal fold vibration. Moreover, using a cepstral measure that will be described below, she found that the tense phonation is more periodic than the lax phonation.

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

3. Analysis 1: Acoustic correlates of phonation

3.1 Methods

3.1.1 Recordings

All the data in this study were obtained from recordings made during a trip to Yunnan province of China in the summer of 2009. Before the recordings were made, a wordlist of two thousand words[9] was elicited, and minimal phonation-contrast pairs of monosyllabic words (around 40 pairs for each language) were selected. The number of tokens produced varies among speakers, because speakers were instructed to skip any pairs they did not know. All the speakers were recruited from Southern Yi villages (Xinping and Jiangcheng), Bo villages (Shizong and Xingfucun) and a Hani village (Luchun). Twelve speakers per language were recorded for Bo and Yi, and eight speakers were recorded for Hani, with gender balanced. For all 32 speakers, simultaneous electroglottograph (EGG) and audio recordings were made. The signals were recorded directly to a computer via its sound card, in stereo, using Audacity, at the sampling rate of 22050 Hz per channel. The audio signal was recorded through a Shure SM10A microphone as the first channel. EGG data were obtained by a two-channel electroglottograph (Model EG2, Glottal Enterprises, with a 40 Hz high-pass cutoff frequency) and recorded as the second stereo channel. Each word was repeated twice.

The analysis presented here includes recordings from all speakers. Since the phonation contrast does not occur with high tone, the data matrix of phonation by tone is not balanced. In order to be able to examine the interaction between tone and phonation, we thus exclude high tone tokens from the current analysis. In section 3.2, we will present the analysis for the acoustic signals, and in section 4.2, we will present the analysis for the corresponding EGG signals.

3.1.2 Measurements

Comprehensive acoustic measures potentially reflecting different phonation properties, as described above, were made using VoiceSauce (Shue *et al.*, 2011). The frequencies of harmonics are estimated from the fundamental frequency, which is in turn estimated by the STRAIGHT algorithm (Kawahara *et al.*, 1999), and the location of formants is estimated by the Snack Sound Toolkit (Sjölander, 2004). Phonation-related acoustic measures include: H1*-H2* (corrected for formant frequency and bandwidth using the

---

[9] The corpus was a collection of high frequency words across various Tibeto-Burman languages, built to do historical comparison among related languages.

algorithm by Iseli *et al*., 2007); H1*-A1*, H1*-A2*, H1*-A3*; individual harmonic amplitudes, H1*, H2* and H4*; H2*-H4*; Cepstral peak prominence (CPP). This set of measures has become fairly standard in recent literature (e.g. Gujarati: Khan, 2012; Mazatec: Blankenship, 2002; Garellek and Keating, 2011; Green Mong: Andruski and Ratliff, 2000; Takhian Thong Chong: DiCanio, 2009; Santa Ana del Valle Zapotec: Esposito, 2010; White Hmong: Esposito, 2012; Suai/Kuai: Abramson *et al*., 2004; Javanese: Thurgood, 2004, Ju|'hoansi: Miller, 2007; Southern Yi: Kuang, 2011).

3.2 Results

For each language, a series of mixed-effect models were employed to evaluate the main effects of phonation and tone on acoustic voice measures, using the *lme4* package in R. Both tone and phonation were specified as fixed effects, and speaker had a random effect on the intercept. The current version of the lme4 package in the R statistical software does not provide *p*-values for *t*- and *F*-tests. Thus *p*-values are estimated by Markov chain Monte Carlo (MCMC) method, implemented as the *pvals.fnc* function in R (Baayen, 2008). The main effects of tone (low, mid) and phonation (tense, lax) in the three languages are summarized in Table 1. Only significant effects at a $p < .05$ level are reported in the tables, and direction is noted. Pair-wise post-hoc tests were performed among all phonation x tone contrastive pairs (e.g. 21T vs. 21L, 21T vs. 33T). Unless a given measure successfully distinguished all pairs, the specific distinguished pairs are noted.

Table 1. Main effect of phonation in Yi, Bo and Hani. Only significant effects ($p < .05$) are reported and direction is noted.

	Yi	**Bo**	**Hani**
H1*	T < L	T < L	T < L
H2*			21T < 21L
H4*			
A1*			21T > 21L
A2*			
A3*			T > L
H1*-H2*	T < L	T < L	T < L
H2*-H4*			21T < 21L
H1*-A1*	T < L	T < L	T < L
H1*-A2*	T < L	T < L	T < L
H1*-A3*	T < L	T < L	T < L
CPP	T > L	T > L	T > L
F0		T >L	

Table 1 shows a general agreement across the three languages on the phonation effect. The reliable acoustic measures that can consistently distinguish tense vs. lax contrasts in three languages are: H1*, H1*-H2*, H1*-A1*, H1*-A2*, H1*- A3* and CPP (shaded measures in Table 1). Phonation also has a significant main effect on the strength of H1* relative to the higher-frequency harmonics (H1*-H2* and H1*-An*). In general, tense phonation has a less prominent H1*, and a shallower spectral slope (H1*-A3*). As indicated by a higher CPP, tense phonation is also either more periodic, or has a higher harmonic-to-noise ratio.

Since the tense vs. lax contrast is crossed with tonal categories, the interaction between phonation and tone was also considered. Consistent significant tone by phonation interactions were found for H1*-H2* and H1*-A1* among all three languages. In general, the tense vs. lax contrast is better distinguished in low tone than in high tone (Figure 2). The H1*-H2* difference between tense and lax phonation in low tone is consistently around 3 dB, whereas it is only 1 dB in mid tone (Hani has a better distinction, with 2 dB difference for mid tone). Similarly, the phonation contrast on H1*-A1* is about 3.5-4 dB in low tone, 1 dB in mid tone. The tone by phonation interaction for H1*, H1*-A2* and H1*-A3* varies among languages. For example, a significant tone by phonation interaction for H1*-A2* is only found in Yi, and for H1*-A3* only in Hani. Tone has a main effect on CPP, such that mid tones generally have higher CPP than low tones, but tone and phonation have no significant interaction on this measure.

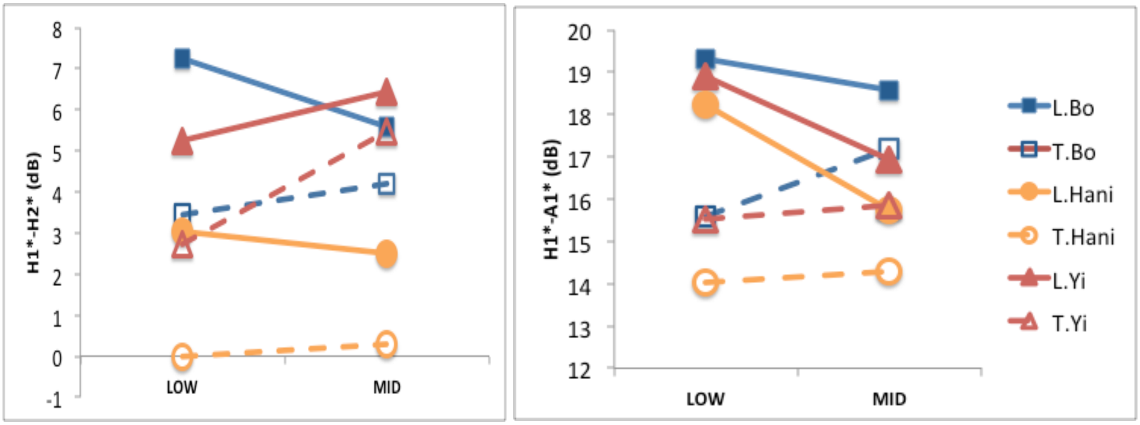

FIG. 2. The interaction between phonation and tone for H1*-H2* (left panel) and H1*-A1* (right panel). Head pattern and color indicates languages (square=Bo, round=Hani, triangle=Yi), line pattern indicates phonation (solid=Lax, dashed=Tense).

Finally, we would like to know which acoustic correlates make the greatest contributions to the phonation contrasts. Therefore, for each language, a series of forward stepwise logistic regressions were employed to evaluate the relative importance of different measurements for the phonation contrast.[10] Measures that reliably distinguished the phonation contrast across three languages were included as the predictors, and phonation (tense vs. lax) was the dependent variable. The relative importance of each measure was estimated by *p*-values based on Wald Chi-Square test (results for individual languages are in Appendix 1). A predictor with more importance should have a smaller *p*-value. The results suggest that H1*, H1*-H2* and H1*-A1*, i.e. the lower frequency range of the spectra (relative to F2, F3), are the best acoustic correlates to distinguish phonation contrasts across the three languages; this result is consistent with the cross-language solution seen in Figure 1.

In this section, we identified six successful acoustic measures for the tense vs. lax type of phonation contrasts: H1*, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, and CPP. Among these reliable measures, H1*-H2* and H1*-A1* turn out to be the most important acoustic correlates. They also both show a significant and similar interaction of tone and phonation: in general, acoustic correlates of phonation are more distinctive in low tone than in mid tone. Various articulatory interpretations of these measures have been proposed in the literature. First, H1-H2 is thought to be related to the ratio of the open phase to the entire glottal cycle (the Open Quotient, or OQ, the converse of the EGG CQ), because when OQ increases, the glottal closing may be less abrupt, so that the glottal area function may approximate a sinusoid of frequency F0. H1 will then dominate the spectrum, and H1-H2 will be large. Second, it has been proposed by Hanson (1997, Hanson *et al*. 2001) that A1 (related to the bandwidth of F1) decreases with DC glottal flow through any inter-arytenoid gap, and thus a larger H1-A1 indicates a breathier phonation. However, this proposal likewise has had limited support (Henrich *et al*. 2001, Shue *et al*. 2009). Third, Stevens (1997:467) hypothesized that a “more abrupt termination” of a glottal pulse, with a “sharp change in slope when it reaches a minimum flow” should result in more energy in high-frequency harmonics, and thus in larger values for higher formant amplitudes (A2, A3) and lower values of H1-A2 and H1-A3. Finally, CPP (or any harmonic-to-noise ratio) will be influenced by two different aspects of glottal articulation. CPP is large when harmonics are strong relative to spectral noise. Thus CPP is smaller in the presence of glottal noise, as in breathy voice. It is likewise smaller, even in the absence of a noise source, when harmonics are less clearly defined, e.g. when vibration is not quite periodic, as in creaky voice. In light of these hypotheses and expectations, we turn now to analysis of the articulatory EGG patterns for the tense vs. lax contrasts, and their relation to the acoustic measures.

---

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

4. Analysis 2: EGG correlates of phonation

4.1 Measurements

4.1.1 EGG parameters

The EGG signals corresponding to the audio tokens analyzed in section 3.2 (12 speakers from Bo and Yi, 8 speakers from Hani) were processed by EggWorks (Tehrani, 2012) to obtain the traditional landmark-based parameter measures. Figure 3 illustrates the EGG parameters. Contact Quotient (CQ) is estimated with four methods: 1) EGG threshold (Rothenberg *et al*., 1988): the contact event is defined as the time point when the signal strength exceeds a threshold of peak-to-peak amplitude (CQ method in Figure 3); 2) dEGG (Henrich et al. 2004): the contact and opening events are defined on peaks in the derivative of the EGG signal (CQ_PM method in Figure 3) Hybrid (Howard 1995): use the dEGG contacting peak for detecting the glottal contact event, and an EGG-based 3/7 threshold for detecting the glottal opening event (CQ_H method in Figure 3); 4) Tehrani hybrid method (ref. documentation of EggWorks): the contact event is defined by dEGG contacting peak, and the opening event is defined by the y-value of the dEGG contacting peak (CQ_HT method in Figure 3). Two measures are made from the dEGG signal: Peak Increase in Contact (PIC in Figure 3), defined as the amplitude of the positive peak of dEGG; Peak Decrease in Contact (PDC in Figure 3), defined as the amplitude of the negative peak of dEGG. EggWorks also provides the time of PIC and PDC. Finally, closing duration and opening duration are measured at a 10% threshold (Marasek, 1997), and Speed Quotient is computed as the ratio between closing and opening duration.

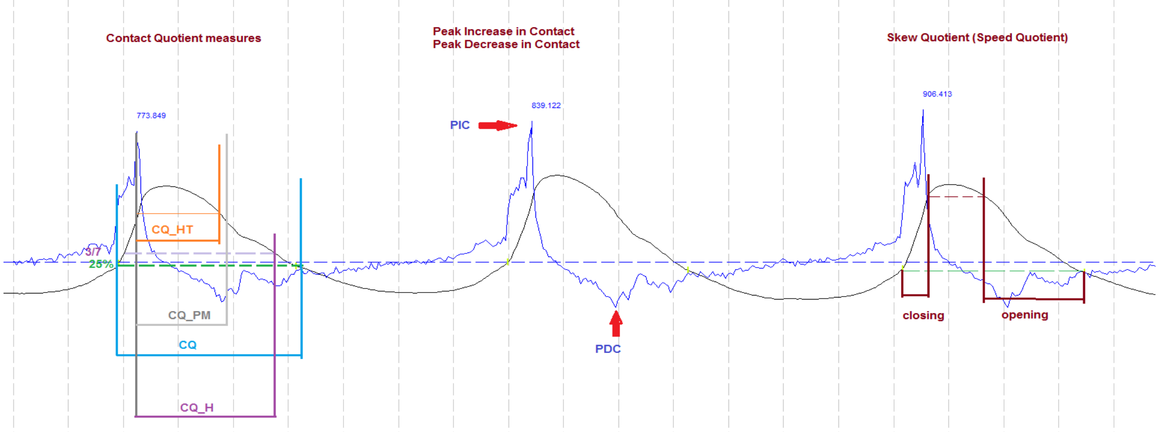

FIG. 3. Demonstration of EGG measures from EggWorks. The black line is the EGG pulses, and the blue line is the dEGG signal. The first pulse demonstrates four methods of estimating CQ, the middle pulse demonstrates PIC and PDC from dEGG, and the third pulse demonstrates measures related to skewness of the pulse: closing duration and opening duration (SQ=T-closing/T-opening). The thresholds shown are schematic only.

4.1.2 EGG pulse shape

Functional Data analysis of EGG pulses follows the procedure by Mooshammer (2010). EGG signals of 6 speakers (3 males and 3 females) per language were selected for shape analysis; for each speaker, a subset of EGG files of the syllable [be] or [be̠] (each with two repetitions) was chosen from the entire recording corpus. So this small dataset controls for vowel and consonant, but varies in phonation type and tone. For each EGG waveform file, two periods during the steady portion of the vowel were extracted using the program PCQuirerX. A 25% threshold was used to define the beginning and the end of a duty circle. In order to cancel the effect of the length of the pulses, the selected pulses were time-normalized to a uniform length of 1000 samples. This was done by a Matlab script using linear interpolation. The data were then amplitude normalized to an amplitude of 1 for the contact peaks, and 0 for the opening peaks. This effect of this amplitude normalization is illustrated in Figure 4.

The entire shapes of the EGG pulses were analyzed by functional principal component analysis (FPCA) using the FDA package in R (Ramsay and Silverman, 1997). Before analysis, the two pulses were pre-processed by a Fourier function, as recommended for periodic waves (for choices of different basis functions, see Ramsay and Silverman 2002). Following Mooshammer (2010), the number of Fourier coefficients was set to 200, and the smoothing parameter lambda was set to 10E−12. After pre-processing, a FPCA was applied (Ramsay and Silverman 1997, 2002) to identify the main factors of variability in the EGG pulses.

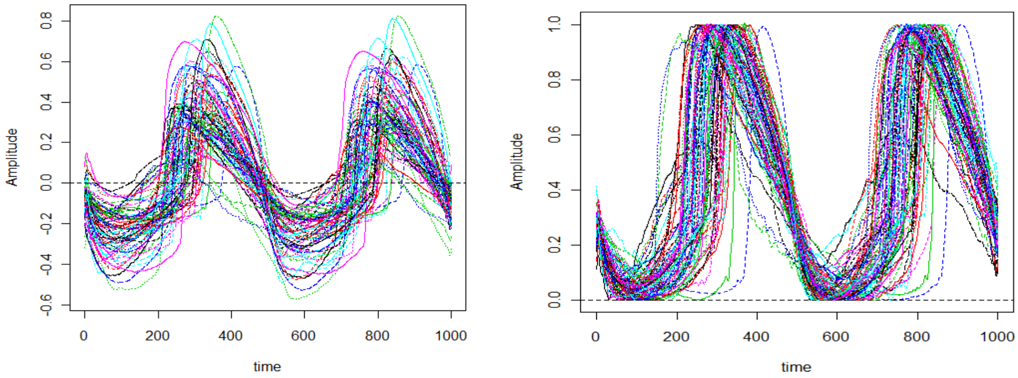

FIG. 4. Lx pulses before amplitude normalization (left panel) and after amplitude normalization (right panel), time is normalized.

4.2 Results – EGG parameters

For each language, a series of mixed-effect models were employed to evaluate the main effects of phonation and tone on EGG parameter measures, with tone and phonation as fixed effects, and speaker as the random effect. The main effects of tone (low, mid) and phonation (tense, lax) in the three languages are summarized in Table 2. Only significant effects (at a $p < .05$ level) are reported in the tables, and direction is noted. Pair-wise post-hoc tests were performed among all phonation x tone contrastive pairs (e.g. 21T vs. 21L, 21T vs. 33T). Unless the given measure successfully distinguishes all pairs, the specific distinguished pairs are noted. PDC values are presented as absolute values, as we care only about the amplitude of the peak.

Table 2. Main effects of phonation in Yi, Bo and Hani on 8 EGG parameter measures.

	Yi	**Bo**	**Hani**
CQ	T > L	T > L	T > L
SQ	T < L	33T<33L	T < L
Closing	T < L	T < L	T < L
Opening		T > L	T < L
PIC	T < L	T < L	T < L
|PDC|	T < L	T < L	T < L
PIC_time	T < L	T < L	T < L
PDC_time	T < L	T < L	T < L

All the EGG parameter measures (except for opening duration) successfully distinguish the lax phonation from the tense phonation for all three languages, and the pattern is very consistent. As expected, the tense phonation consistently has a greater CQ than the lax phonation[11], suggesting a smaller open quotient in the vocal folds. Figure 5 shows the CQ_H[12] values for the three languages. The CQ difference between the two phonations, though significant, is rather small (0.05 ratio), and the mean range for the phonation contrasts lies between 0.4 and 0.5, a very "modal" range. Therefore, the tense vs. lax contrast does not seem to require an extreme gestural change in the vocal folds.

The parameters that are related to the relative duration of closing and opening provide further information on the pulse shape. SQ is the ratio between closing duration and opening duration, so the smaller its value the more skewed is the EGG pulse. Referring to Table 2, the tense phonation generally has a smaller SQ value, suggesting more left-skewed pulses. The skewness is mostly due to the variation of closing duration, as it is significantly shorter for the tense phonation. Finally, the measures from dEGG provide some information about contacting speed. The moment of PIC has been used to define the closing event, but the physiological mechanism of PIC is less clear. It is arguably related to the speed of glottal closure. Taking this view, the consistently smaller PIC

---

[11] Different methods have the same pattern.

[12] For discussion of the choice of this method, see footnote 13 in section 4.4.

values for the tense phonation suggest the tense phonation has a slower glottal closure. Other seldom discussed aspects of dEGG signals turn out to be significant as well. The amplitude of PDC of the tense phonation is generally smaller than the lax phonation. So the tense phonation has generally smaller amplitudes for both opening peak and closing peak, which might suggest a general slower vibration speed in the vocal folds. The time measures of the peaks (PIC_time and PDC_time) suggest that both closing and opening peaks come earlier for the tense phonation.

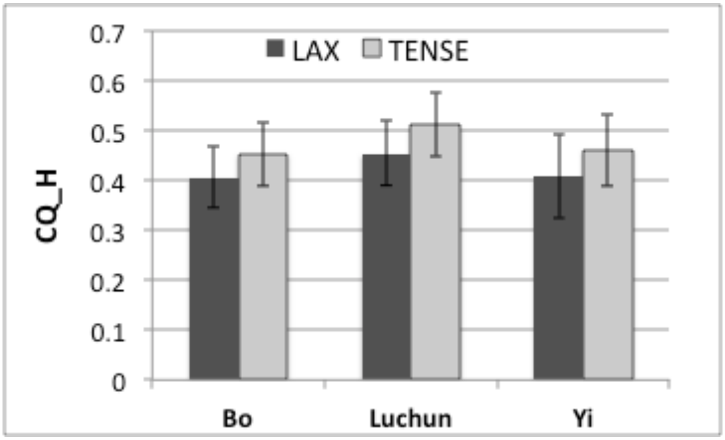

FIG. 5. Means and standard deviations of Hybrid CQ measure (CQ_H) in Bo, Luchun and Yi. The tense phonation consistently has a greater CQ. The range of CQ is similar across languages, which is between a ratio of 0.4 – 0.5, a very "modal" range.

Similar to the acoustic analysis (section 3.2), the interaction between phonation and tone is tested in mixed-effect models. However, unlike the consistent interaction found for acoustic correlates, no consistent interaction between tone and phonation is found for EGG parameter measures. Cross-linguistically, the tone that co-occurs with the lax phonation has a weak effect on CQ, with the mid-lax tone (33L) having a slightly greater CQ than the low-lax tone (21L). However, this interaction between phonation and tone reaches statistical significance only in Hani. Therefore, the physiological mechanism of tense vs. lax phonation is generally independent from that of tonal production.

Finally, for each language, a series of forward stepwise logistic regressions were employed to evaluate the relative importance of different measurements for phonation contrasts. Measures that reliably distinguish phonation contrast across three languages are included as the predictors, and phonation (tense vs. lax) is the dependent variable. The relative importance of each measure is estimated by *p*-values based on Wald Chi-Square test (results for individual languages are in Appendix 2). The results suggest that CQ, closing duration, PIC and PDC are the parameters that contribute most to the phonation contrasts.

In sum, the parameter analysis of EGG is generally successful, with CQ and closing duration the most important properties of the EGG waveforms, and the amplitudes of the positive and negative peaks of dEGG also important parameters. That is, most of the EGG parameters tested proved to differ between the phonation types. Other parameters could doubtless be defined over the EGG and dEGG signals, and it is likely that at least some of those would also reflect the phonation contrast.

4.3 Results--FDA analysis of EGG pulses

We first did separate FDA analyses for the individual languages, and the patterns were very similar. Therefore we collapsed the languages together and did a single FDA analysis for tense vs. lax contrasts across all three languages. The factor scores for the first four principle components (PC1 to PC4) are presented in Figure 6.

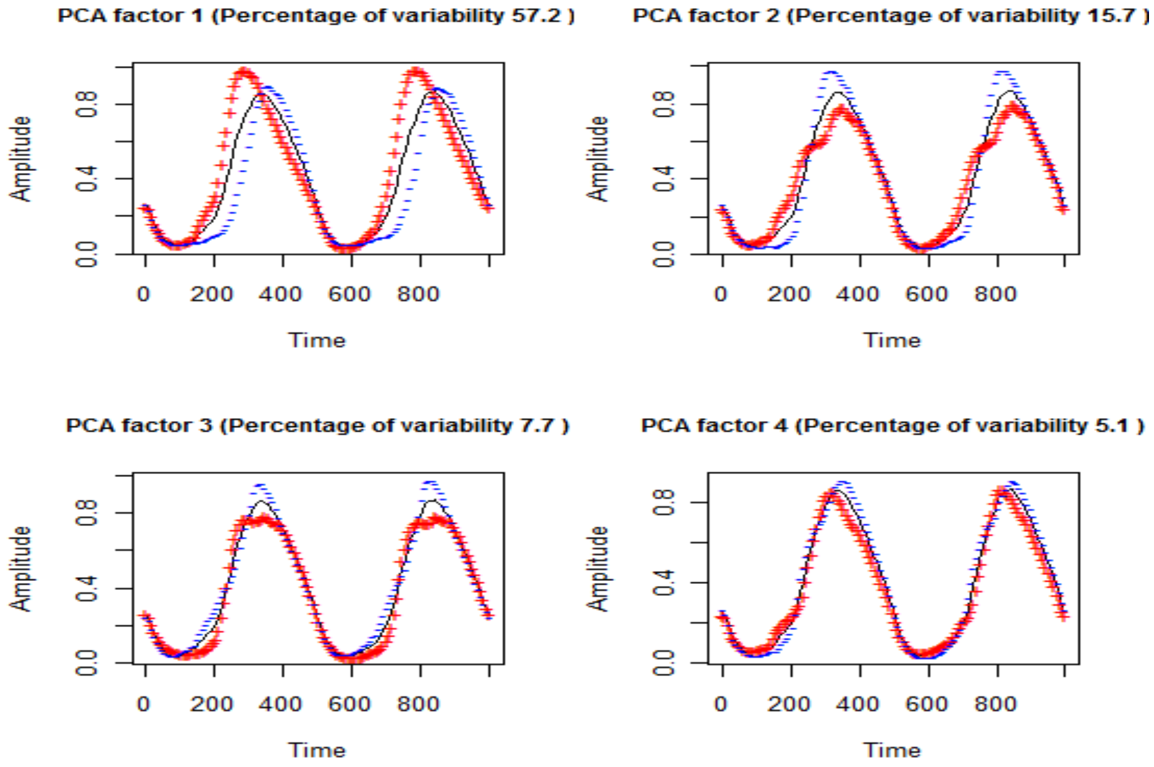

FIG. 6. The shapes of the EGG signals. Negative factor values are indicated by minus signs (blue lines), positive factor values by plus signs (red lines), and the mean curve by a solid line, for the first four factors.

The four panels in Figure 6 show the first four principle components of FPCA, the negative factor values indicated by minus signs, the positive factor values by plus signs, and the mean curve by a solid line. PC1 and PC2 account for the most variability of the pulse shapes, 57.2% and 15.7% respectively; PC3 and PC4 together account for another

12.8% of the variance. Taken together, the first four factors account for 85.7% of the variability of the pulse shapes.

To help understand the meaning of the positive/negative deviations from average pulses seen in Figure 6, Figure 7 provides a graphical representation of the relevance of the experimental factors on each of the four principal components of the shapes, following a method used by Mooshammer (2010). For each gender by phonation by tone condition (there are 8 of these, e.g. female/lax/low), the mean score on the component is plotted; the higher these scores are, the greater the approximation to a positive-deviation pulse shape in Figure 6. We then look at these plots for any patterns within each component.

PC1 shows a strong phonation effect: positive values in Figure 7 are the tense phonation and negative values are the lax phonation. Therefore, the plus line for PC1 in Figure 6 can be taken to mostly represent the tense phonation pulses, and the minus line the lax phonation pulses. In contrast, PC2 and PC3 mostly show a strong gender effect in Figure 7, as females overall have lower values than the males; the females' 33T is the exception in PC3. Therefore, the plus lines for PC2 and PC3 in Figure 6 can be taken to mostly represent the male pulses, and the minus line the female pulses. Finally, PC4 in Figure 7 singles out the 33T. Therefore, the plus line for PC4 in Figure 6 can be taken to mostly represent the 33T pulses, and the minus line all the other pulses. In sum, this visualization suggests that a phonation effect is mostly seen in PC1; a gender effect is mostly seen in PC2 and PC3; and an effect of 33T is seen in PC3 and PC4.

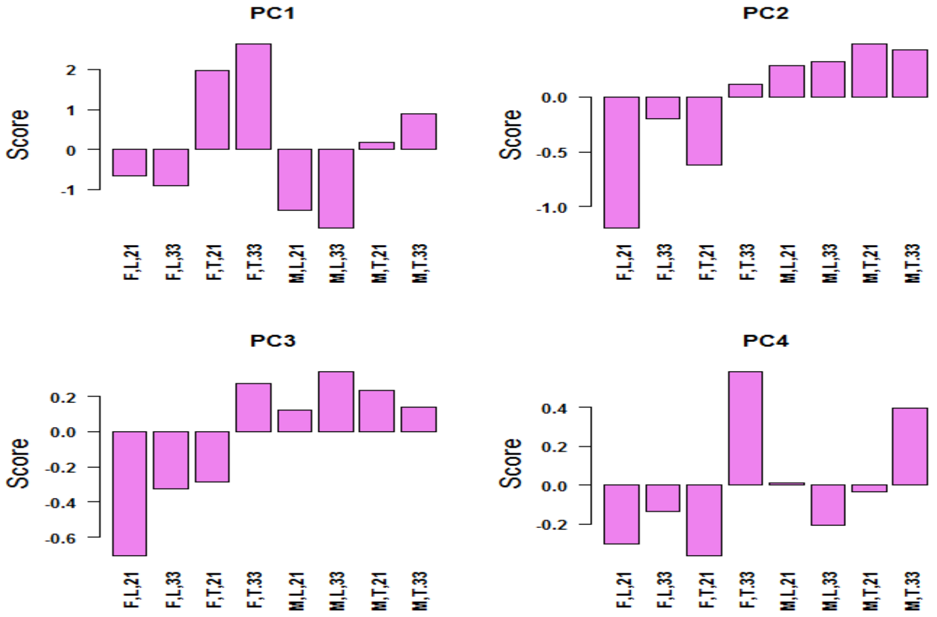

FIG. 7. Mean scores of factors in each principle component for each gender/phonation/tone condition. F=female, M=male, L=lax phonation, T=tense phonation, 21=low tone, 33=mid tone.

A more quantitative approach to understanding the principal components and Figure 6 involves the correlations between the component scores and the experimental factors (language, gender, phonation and tone). These are given in Table 3, and essentially confirm our qualitative observations. The dominant factors for each principle component are in bold. PC1 is by far the most strongly related to phonation; PC2 and PC3 more weakly related to gender; and PC4 to tone (and phonation). Language has no significant effect on any of the principle components, confirming that the three languages share very similar gestural patterns to make their tense vs. lax contrasts.

PC1 accounts for the most variance in pulse shapes, and is mostly related to phonation contrasts. Therefore, phonation is the most important factor in determining the EGG pulse shapes. Referring to Figure 6, it can be seen that the contacting part of the pulse is the main difference between the lax and tense phonations. Generally speaking, the tense phonation (the plus line in PC1) has a greater extent of contact (more complete contact), larger contact quotient and shorter contacting duration (steeper positive slope); whereas the lax phonation (the minus line in PC1) has a smaller extent of contact (less complete contact), smaller contact quotient and longer contacting duration. Moreover, compared to the lax phonation, the tense phonation has clearer "knees" in contacting and de-contacting, indicating a more abrupt change in contact (Rothenberg *et al.*, 1988; Childers *et al.*, 1990; Baken and Orlikoff, 2000). As a result of the more abrupt contact, the slope of the closing phase is steeper for the tense phonation. By contrast, the lax phonation has more gradual and less complete contact, and greater opening. The gesture must be very smooth, as the pulse is very symmetric and sine-like.

As also seen in Figure 6, extent of contact is also involved in the gender effects seen in PC2 and PC3. PC2 also involves the width of the pulse.

Table 3 Correlation between Principle Component and experimental factors (gender, phonation and tone). Bold shows the factor most strongly correlated with each PC.

	Language	**Phonation**	**Tone**	**Gender**
PC1	-0.10	**0.46**	0.05	-0.19
PC2	0.06	-0.03	-0.08	**-0.19**
PC3	0.05	0.06	0.08	**0.23**
PC4	0.06	-0.15	**-0.18**	-0.12

4.4 Correlations between acoustic measures and EGG measures

What are the consequences of the articulatory patterns just discussed in the previous section? Spearman correlation coefficients calculated between factor scores (from the

FDA of pulse shapes) and acoustic measures or EGG parameters (from signal landmarks) are shown in Table 4. As can be seen in the table, most acoustic measures and EGG measures that were found to contribute to phonation contrasts are most strongly correlated to the first component of EGG pulse shapes (PC1), which also mostly reflects the phonation contrasts. Consistent with the results in section 3.2 above, the acoustic correlates that are significantly correlated with PC1 are H1*, H1*-A1*, H1*-H2*, H1*-A2*, H1*-A3*, and CPP. Recall that PC1 shows that the tense vs. lax phonations mainly differ in abruptness of contact (and de-contact), the extent of contact, and a shorter/steeper closing phase. Therefore, these several acoustic measures at least in part reflect these variations, but no single acoustic measure does so more than the others.

Table 4. Correlation coefficients between factor scores from FDA of EGG pulse shapes, and the acoustic measures (top) and EGG parameters (bottom). Only significant coefficients (*p* < .05) are reported here, and highly significant ones (*p* < 0.001) are highlighted in bold.

	PC1	**PC2**	**PC3**	**PC4**
H1*	**-0.525**			
H2*				
H4*				
A1*				
A2*			0.175	
A3*		0.128		-0.142
H1*-H2*	**-0.486**		**-0.234**	
H2*-H4*		**-0.263**		
H1*-A1*	**-0.586**		-0.104	
H1*-A2*	**-0.492**		-0.190	
H1*-A3*	**-0.464**			
CPP	0.306			0.120
CQ	0.689		**-0.416**	
CQ_H	**0.696**			-0.224
CQ_PM	0.581	-0.208		
CQ_HT	0.420	**-0.495**	**0.393**	**-0.271**
PIC	0.251	-0.140	0.174	
PIC_time			-0.289	0.140
PDC	0.410			0.166
PDC_time			-0.290	0.139
Closing	-0.282		**-0.369**	
Opening		-0.223		**-0.252**
SQ	**-0.472**		-0.310	-0.191

Likewise, most EGG parameters, which reflect different aspects of the pulse shape, are significantly correlated with PC1, suggesting that PC1 is able to capture a complete picture of the EGG glottal pulse. The holistic pulse shape provides us with a more precise understanding of the EGG pulses at least in two ways. First, it is able to reveal the articulatory variations that are concealed by the relative CQ values. According to PC1, there are multiple aspects of the vibratory pattern that will increase CQ, e.g. increase the extent of contact, shorten the closing phase, or increase the width of the pulse. Different methods of CQ also reflect different kinds of change in the glottal pulses.[13] So CQ is an ambiguous measure for describing vibratory patterns of phonation types, as two phonation types may have similar CQ values but involve quite different vibrations. Second, the FDA components are able to capture the abruptness of the glottal contacting (the "knees" in PC1), which cannot be captured by any current EGG parameters. Abruptness of contact is a very important property of vocal fold vibration, as it has a significant impact on glottal flow and acoustics. In sum, the results in Table 4 indicate that one underlying gesture (PC1) can lead to a series of consequences in acoustic correlates and EGG parameters, and thus that PC1 appears to be the single best measure of the EGG pulse.

Another series of Spearman correlations were done, this time between EGG parameters and acoustic measures, in order to understand the relationships between individual EGG parameters and acoustic correlates. The significant correlations are shown in Table 5. As can be seen here, there are no one-to-one correlations, in part because the measurements in each domain (either acoustic or EGG) are likely to be correlated with each other as well. For example, correlation analysis shows that CQ is significantly correlated with SQ (r=0.65, $p < .001$), and with closing duration (r=0.79, $p < .001$). So it is not clear whether the significant correlation between H1*-H2* and SQ is because SQ can affect this spectral measure, or because SQ is correlated with CQ, and the latter is the main influence of H1*-H2*. To tease apart such nested correlations, in addition to the Spearman correlation test between acoustic measures and EGG parameters, a multiple regression for each acoustic measure is applied to decide which EGG parameter has a major or independent influence on this given acoustic measure, with all significantly correlated EGG parameters as the predictors. The significant EGG parameters from the regression models are highlighted in bold in Table 5.

---

[13] Notes about different CQ measures: the hybrid method (CQ_H) best reflects the properties of the EGG pulses. The threshold method (CQ) is another good method for this kind of phonation contrast. However, looking at PC2 through PC4, it seems that the different CQ measures reflect different properties of the EGG pulse shapes. For example, the maximum contact variability is best captured by CQ_HT, as shown with PC2 and PC3. CQ (threshold method) is sensitive to an earlier closing phase, as shown with PC3, and CQ_H is good at capturing the opening contact variations, as shown with PC4.

Table 5. Correlation coefficients between EGG parameters and acoustic measures. Only significant coefficients ($p < .05$) are reported here. Multiple regressions were run for each acoustic measure, with all EGG parameters as the predictors. Significant EGG parameters from the multiple regression models are highlighted in bold (read these by rows).

	CQ	**PIC**	**|PDC|**	**Closing**	**Opening**	**SQ**
H1*	**-0.59**	0.24	0.49	**0.47**	0.25	0.62
H2*				0.30	**0.39**	0.28
H4*					**0.22**	
A1*	0.22	0.20			**0.28**	-0.19
A2*	0.27			-0.34	0.23	-0.25
A3*	0.32	0.39	0.23	-0.26	0.25	-0.19
H1*-H2*	**-0.59**		0.48	**0.28**		**0.52**
H2*-H4*	-0.22			0.38	0.19	0.32
H1*-A1*	**-0.69**		0.32	0.50		**0.64**
H1*-A2*	**-0.53**		0.26	**0.58**		0.60
H1*-A3*	**-0.48**	-0.17		0.44		0.49
CPP	**0.30**				-0.18	-0.12

Indicated by the bolded cells in Table 5 (read row by row), H1* is mostly affected by CQ and closing duration (i.e. the time spent in the increasing-contacting phase). As indicated by the positive correlation, a longer closing duration can lead to a greater value of H1*; this makes sense if longer closing duration corresponds to a greater amplitude of opening (which however cannot be seen by EGG), since Sundberg (1987) has suggested that greater opening results in greater peak glottal flow, which in turn gives a higher H1. Also as expected, an inverse correlation with CQ indicates that smaller CQ values can lead to a more prominent H1*. H2*, H4* and A1* are all affected by the opening duration of the vocal folds. According to Ní Chasaide and Gobl (1997), a longer opening duration (RG in the glottal flow pulse) can boost the range of H2*, as a characteristic of "creaky" voice. H1*-H2*, one of the best acoustic correlates of the tense-lax phonation contrast (see section 3.2), is affected by CQ, closing duration, and SQ; similarly, H1*-A2*, the other best acoustic correlate of the phonation contrast, is affected by CQ and SQ. CQ is also the main contributor to H1*-A2*, H1*-A3* and CPP. It has been previously proposed that H1*-A2* and H1*-A3* are related to closing duration or SQ (Marasek 1997), or abruptness of contact (Stevens 1977), or speed of contact (Esposito, 2012). These two acoustic measures indeed are significantly correlated with SQ and closing duration, but the main cause is still CQ. So H1*-A2* and H1*-A3* may not be reliable reflexes of skewness of glottal pulses.

Table 5 also sheds light on the understanding of the dEGG measures. The two dEGG peaks are relatively independent from CQ and SQ, as no significant correlation is found between PIC and CQ_H[14] (r=-0.01, p >.05), or PIC and SQ (r=0.03, p >.05); and only a weak correlation is found between PDC and CQ_H (r=0.27, p < .05). It is noted that neither of these two dEGG peaks is found to be significant predictors of spectral measures in the multiple regression analysis (in Table 5, none of these cells are bolded). This result is consistent with Michaud (2004) and Esposito (2012)'s claim that PIC is not a measure of spectral slope. However, as seen in the correlation analyses (as opposed to multiple regression) of our current dataset, PIC is somewhat correlated to the energy in higher-frequencies, as A3* is the best correlated acoustic measure with PIC. If higher frequencies are boosted by a more abrupt contacting (Ní Chasaide and Gobl, 1997), then PIC somewhat reflects this property.

In sum, this section includes two correlation analyses to understand the relationship between EGG measures and acoustic correlates. Both Table 4 and Table 5 show that there are no one-to-one correspondences between these two domains, and the one-to-many correlation pattern in Table 4 is more consistent and reliable. Essentially, the complex consequences in either EGG parameters or acoustic correlates are caused by one underlying articulatory pattern (PC1).

5. Relation of production to perception

Finally, we would like to know whether the variations of articulatory patterns in the vocal folds can be perceived by native listeners, and which gestures are most important in the perception of phonation contrasts. In this section, we will analyze the EGG glottal pulses of the stimuli in a previous categorical perception study of Yi (Kuang 2011), and examine the correlations between perception responses and gestural variations.

5.1 Method

A set of [be] and [bu] syllables contrasting on both phonation (tense vs. lax) and tone (21 vs. 33) were retrieved from Southern Yi two-channel recordings, the audio signals of which were used in Kuang (2011) to investigate the perceptibility of the tense vs. lax contrast in Southern Yi. The stimulus set included productions from six speakers (3 females, and 3 males), and each speaker produced the following eight words as isolation forms:

A. High vowel with mid tone, lax vs. tense (bu33/bu̠33)
B. High vowel with low tone, lax vs. tense (bu21/bu̠21)
C. Low vowel with tone, lax vs. tense (be33/be̠33)
D. Low vowel with tone, lax vs. tense (be21/be̠21)

---

[14] A significant correlation is found between PIC and CQ (threshold method), r=-0.3, p < .05.

In Kuang (2011), these 48 test stimuli were used in an AXB identification task. In an AXB task, A and B serve as standards for two different categories, and the listener must decide which standard the X is more like. Here the A and B standards were pronunciations of a tense-lax minimal pair by speaker M1 or speaker F1, chosen because these speakers seemed to maintain a good contrast between tense and lax phonations. The Xs were the 48 test tokens, produced by all six speakers. Half of the trials used speaker M1 as the standard, and half used speaker F1. When the test token was by speaker M1 or F1, the other speaker was used for the standard. Therefore, the 48 tokens yielded 80 stimuli in total. These 80 stimuli were presented three times to each listener (in a random order). Nine Yi listeners participated in the experiment, giving 2160 (80 stimuli x 3 repetitions x 9 listeners) responses for tense vs. lax identification. The task was run by a Praat script on a computer. Stimuli were played through SONY MDR-NC60 headphones. On the screen, the listeners could see three buttons, labeled as (A), (X) and (B). The buttons for A and B were in yellow and clickable, and X was in grey. The listeners heard three stimuli in sequence separated by 0.5 second, and had to decide whether the second word (X) is more similar to the first word (A) or to the third word (B). Listeners had to make a response for every trial by clicking either A or B.

Kuang's results showed that the listeners were largely able to perceive the phonation contrasts in across the vowels and tones in this stimulus set. Counting accuracy by vowel, the average accuracy rate for low vowels is 73.5% ($\chi^2(1, 540)=77.5$, $p < .001$) compared to 57% ($\chi^2(1, 540)=5.4$, $p < .05$) for high vowels, while counting accuracy by tone, the average accuracy rates for the two tone conditions are nearly identical: 66% ($\chi^2(1, 540)=36.1$, $p < .001$) for the low tone and 64.5% ($\chi^2(1, 540)=29$, $p < .001$) for the mid tone. Such less-than-perfect perception performance is well-suited to statistical testing of listeners' attention to different potential cues in the stimuli.

In sum, for these 48 tokens, there are existing perceptual data, i.e. how identifiable each token was in Kuang's (2011) experiment. In addition, because we have the audio signals, we can carry out the same acoustic analyses as in our analysis 1 above, and because we have the accompanying EGG signals, we can carry out the same EGG analyses as in our analysis 2 and 3 above. Thus we are able to directly calculate the correlations between aspects of the acoustics, the EGG pulses, and perception.

5.2 Results

5.2.1 Replications of acoustic and EGG analyses

Figure 8 shows the first four PCA components of the EGG pulse shapes of the 48 test tokens. The first two components (accounting for 74.4% of the variance) are strikingly similar to the first two components in Figure 6, confirming that the EGG pulse shapes for tense vs. lax phonation have a very consistent pattern. Just as for the full dataset in section 4.3 above, results from repeated measures ANOVAs on the effects of the factors in Table 6 on the principle component scores also show that PC1 is dominated by phonation, and PC2 is related to the interaction between phonation and gender. Females' lax phonation generally has smaller principle component scores. PC3 has both phonation and gender effects, and PC4 is mostly related to tone. Vowel quality does not affect glottal pulse shapes.

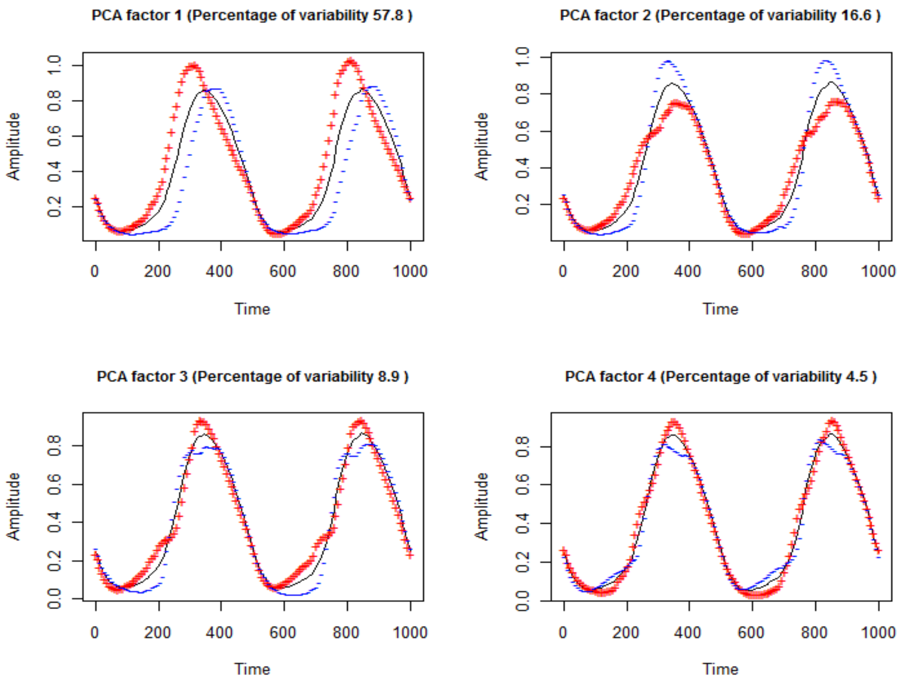

FIG. 8. First four principle components of glottal pulse shapes. Positive factors are indicated by (red) plus signs, and negative factors are indicated by (blue) minus signs.

Table 6. Results of repeated measures ANOVA and pairwise t-tests for PC1–PC4 and the factors of phonation, tone, gender and vowel. Significant F values are highlighted in bold and their *p*-values are given.

	Df.	**Phonation**	**Tone**	**Gender**	**Vowel**
PC1	1,32	**33.87** (*p*<.0001)	0.02	4.48	0.06
PC2	1,32	0.05	0.99	**4.5** (*p*<.05)	0.07
PC3	1,32	**6.6** (*p*<0.05)	0.07	**3.7** (p<.05)	0.9
PC4	1,32	0.8	**3.4** (*p*<0.1)	1.5	0.2

Also as before, mean scores of the factors in each principle component are calculated in order to understand the meanings of plus and minus lines in Figure 8. PC1 is mostly related to phonation, and its minus line can be interpreted as representing the lax phonation, its plus line the tense phonation. PC2 is mostly related to speaker gender, and its minus line can be interpreted as representing the female speakers, its plus line the male speakers. Finally, PC4 is mostly related to tone, and its minus line can be interpreted as representing the low tone, its plus line the mid tone.

Table 7. Spearman correlation coefficients between PC scores and perception accuracy, acoustic correlates and EGG parameters. Only significant coefficients (*p* <.05) are shown here, and highly significant coefficients (*p* < .001) are highlighted in bold.

	PC1 (57.8%) Phonation	**PC2 (16.6%) Gender**	**PC3 (8.9%) Phonation+Gender**	**PC4 (4.5%) Tone**
H1*	**-0.54**		0.18	
H2*				
H4*				
A1*			-0.20	-0.20
A2*				
A3*	0.16	0.21		0.14
H1*-H2*	**-0.53**	**-0.24**		
H2*-H4*		**0.37**		
H1*-A1*	**-0.53**		**0.29**	
H1*-A2*	**-0.39**		0.23	
H1*-A3	**-0.34**			
CPP	**0.44**	0.27	**0.31**	**-0.47**
CQ	**0.75**	0.29	0.19	-0.16
CQ_H	**0.81**		**-0.31**	
CQ_PM	**0.75**			
CQ_HT	0.49		**-0.45**	
PIC				
PDC	0.22			**0.32**
Closing	**-0.39**		**0.42**	
Opening		**-0.32**	-0.30	**0.30**
SQ	**-0.34**		**0.38**	-0.27
Perception Accuracy	**0.49**		0.13	

Correlation coefficients between principle component scores and the various acoustic EGG parameters, for the 48 stimuli as a group, are given in Table 7. As can be seen here, the two major EGG pulse shape components PC1 and PC2 have different acoustic consequences. Further confirming the earlier analysis with the full dataset, here PC1, the phonation-related component of EGG pulse shape, is mainly reflected in H1* and its relative prominence to the harmonics in higher frequency range (H1*-H2*, H1*-An*). Slightly different from Table 4, the noise/periodicity measure CPP is also an important acoustic correlate, as it has a better correlation with PC1 (compared to Table 4). By contrast, PC2, the gender-related component, is mainly reflected in H2* related acoustic correlates, e.g. H2*-H4* and H1*-H2*. This finding is consistent with Kuang (2011) and Bishop and Keating (2012), which also found that H2*-H4* has a significant gender effect. Kuang (2011) also found H2*-related measures are significantly correlated with tone and pitch. Also consistent with the earlier analysis, phonation-related pulse shape components (PC1 and PC3) are best correlated with the CQ_H measure[15]. SQ and closing duration are also very important EGG parameters that can capture the phonation related pulse shapes. In general, the patterns shown in Table 7 are consistent with the results in Table 4.

5.2.2 Production-perception relations

Correlation coefficients between principle component scores and mean perception accuracy for all listeners combined, for the 48 stimuli as a group, are also given in Table 7, at the bottom. It can be seen that only PC1 is related to listeners' performance. Specifically, as listeners had to attend to the phonation contrast in order to classify the stimuli, the identification accuracy rates are only significantly correlated to PC1, while PC2, a shape component related to gender, is successfully factored out by listeners. Although PC3 and PC4 are not major components here, we still can notice the sensitivity of detecting phonation-related gestures/acoustic correlates, as there is a weak correlation between PC3 (related in part to phonation) and perception accuracy rates.

Crucially, we are interested in which kind of measure (acoustic, EGG parameter, or EGG pulse shape) is the best predictor of perceptual identification accuracy. Three multiple regressions were done to predict perception accuracy scores from (1) acoustic correlates (H1*, H1*-H2*, H1*-An*, CPP), (2) EGG parameters (CQ_H, SQ and duration of closing phase), and (3) principle components of EGG pulse shapes (PC1, PC2, PC3, PC4). Among the acoustic measures, H1* is the best predictor, though other related spectral measures (e.g. H1*-H2*, H1*-An*) have significant contributions as well. The correlation coefficient between perception accuracy and H1* is 0.28 ($p < .01$). Among the EGG parameters, CQ_H is the best predictor, and SQ and closing duration make

---

[15] CQ_H has the strongest correlation with PC1 (r=0.81, $p< .001$), and a significant correlation with PC3 (r=-0.31, $p< .001$). By contrast, CQ and CQ_PM are less sensitive to the pulse shape difference in PC3; CQ_HT is sensitive to PC3, but much less sensitive to the major pulse shape of PC1. A correlation analysis between CQ measures and perception accuracy rates shows that CQ_H also has the best correlation with perception accuracy rate (r= 0.26, $p< .05$).

significant contributions as well. The correlation between CQ_H and perception accuracy is 0.35 ($p < .001$). All of these relations are weaker than the relation already with PC1 of the EGG pulse shape – its correlation with perception accuracy is 0.49 ($p < .001$). Therefore, it can be concluded that the overall shape of the EGG pulse is a better predictor of the perception of the phonation contrast than partial reflexes (e.g. CQ, H1*) of it. This is not surprising, since the whole pulse shape contains more information than any one parameter. In fact, as the regression analyses revealed, more than one acoustic correlate/EGG parameter contribute to the perception of phonation contrasts; these complex discrete measures are the consequences of a series of gestures reflected in PC1.

In sum, in this section we first presented the FDA analysis of EGG pulse shapes of the Southern Yi test stimuli, which were found to be very similar to the pulse shapes of the larger set of cross-language stimuli reported above. Thus the pulse shapes of the tense vs. lax phonations are very consistent, regardless of language, vowel, and speaker sample. More importantly, we found that listeners' perception is very sensitive to the aspects of EGG pulse shape seen in the first principle component of pulse shape (PC1). This single glottal pulse shape factor is a much more powerful predictor of perceptibility of phonation contrasts than any single measure/parameter from acoustic signals and EGG parameters. Listeners apparently are able to unpack the acoustic signals and factor out the variability related to other, non-phonation, influences on pulse shape.

6. Discussion

The goal of this study has been to explain the properties of a type of phonation contrast: tense vs. lax phonation contrasts of three Yi (Loloish) languages. We explored the following questions: 1) the important acoustic correlates of the phonation contrast from among a large set of voice measures; 2) the important EGG parameters of the phonation contrast; 3) the glottal events involved in the phonation contrast; 4) the perceptual salience of these phonation contrast-related parameters and measures.

In the first part of our analysis, we identified six successful measures for this tense vs. lax type of phonation contrast: H1*, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3* and CPP, possibly reflecting the open quotient (H1*, H1*-H2*, H1*-A1*), the abruptness (H1*-A3*, H1*-A2*), and the periodicity (CPP) of the vibration. Among these reliable measures, H1*-H2* and H1*-A1* turn out to be the most important acoustic correlates. They also both have a significant and similar interaction of tone and phonation. In general, acoustic phonation correlates are more distinctive in low tone than in mid tone.

Consistent cross-linguistic contrast-patterns were also found for EGG parameters. CQ and closing duration are the most important properties of the EGG waveforms, and the amplitudes of positive and negative peaks of dEGG are also important parameters. These measurements give some hints about the glottal gestures that might be involved and the possible glottal pulse shapes.

Therefore, following Mooshammer (2010), we adopted FDA (or FPCA), a method previously found to be useful in illustrating articulatory movements (e.g. Lee *et al*. 2006), to analyze the variability of entire glottal pulse shapes of the tense vs. lax phonation contrasts. Now we are able to quantitatively generalize the glottal vibration pattern of this type of phonation contrasts across speakers and languages. As shown in Figure 6, despite the great deal of variability among speakers, there is essentially only *one* underlying articulatory pattern involved in these tense vs. lax contrasts: As indicated by PC1, these two less extreme non-modal phonations, tense and lax, mainly differ in extent of contact, abruptness of contact, and duration of closing phase. This articulatory pattern is consistent across all speakers from all three languages. Abruptness of contact (known to be an important aspect of the glottal flow waveform) is a property not captured by any EGG parameters, but captured here by FDA. Our understanding of the articulatory processes of the tense vs. lax contrast is thus much improved by FDA.

The results from FDA also improve our understanding of the relationship between articulatory patterns and their acoustic and parameter consequences. The principle component that is responsible for phonation contrasts (PC1 in Figure 6) is correlated with a series of EGG parameters and acoustic correlates, which have been demonstrated to be successful measures. Specifically, PC1 is correlated to the relative prominence of H1* in the spectrum, and to CQ, SQ and closing duration of EGG waveforms. The correlations between PC1 and EGG parameters and acoustic correlates are very consistent among speakers and languages as well. This suggests that various EGG parameters and acoustic measures are the consequences of one articulation pattern (PC1). In other words, the relationship between laryngeal articulation and its consequences is very stable, but it is a one-to-many kind of relationship.

Perception is a very important complement to articulatory processes: listeners should be able to correctly decode the information packed in acoustic signals. In this study, we have presented a possible way to evaluate the perceptual salience of articulatory patterns. We found that listeners are very sensitive to phonation contrast-related articulatory factors. They are apparently able to factor out the variability due to speaker differences, including gender. Again, we showed that a single glottal pulse shape factor is a much more powerful predictor of perceptibility of phonation contrasts than any single measure/parameter from acoustic or EGG signals, suggesting that the perception of certain phonological contrast relies on multidimensional cues that are related to underlying articulatory patterns.

In sum, this paper discusses three aspects of the articulation of the tense vs. lax contrast: glottal vibration patterns, acoustic correlates, and perceptual salience. Though the Tibeto-Burman type of phonation contrast, tense vs. lax, may be less distinctive than a creaky vs. breathy contrast, we found that in the three Loloish languages studied here these phonations are well distinguished in production, and reasonably well perceived. Overall, we found consistent correlations among the three aspects, which means that speakers and listeners are able to establish a stable link between articulation, acoustic signals and perception. The methodology adopted in this study improved our understanding of EGG waveforms, and shed light on future studies on cross-linguistic phonation types.

Appendix 1

Stepwise Logistic regression models of acoustic measures for individual languages:

Bo:

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		1641.2	1643.2			
H1*-H2*	1	1550.1	1554.1	91.185	**2.20E-16**	***
H1*-A1*	1	1613.8	1617.8	27.44	**1.62E-07**	***
H1*	1	1589	1593	52.294	**4.78E-13**	***
H1*-A2*	1	1619.2	1623.2	22.01	2.71E-06	***
H1*-A3*	1	1623.3	1627.3	17.943	2.28E-05	***
CPP	1	1639.8	1643.8	1.405	0.2358	

Yi

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		1463.8	1465.8			
H1*-H2*	1	1400	1404	63.868	**1.33E-15**	***
H1*-A1*	1	1422	1426	41.877	**9.72E-11**	***
H1*	1	1432.7	1436.7	31.146	**2.39E-08**	***
H1*-A2*	1	1454.8	1458.8	8.989	0.002716	**
H1*-A3*	1	1445.3	1449.3	18.574	1.63E-05	***
CPP	1	1461.3	1465.3	2.505	0.113488	

Hani

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		2255.4	2257.4			
H1*-H2*	1	2064.7	2068.7	190.75	**2.20E-16**	***
H1*-A1*	1	1934.1	1938.1	321.36	**2.09E-09**	***
H1*	1	1971.5	1975.5	283.88	**4.72E-08**	***
H1*-A2*	1	2025.4	2029.4	230.05	0.708252	
H1*-A3*	1	1998.2	2002.2	257.22	0.000103	***
CPP	1	2196.9	2200.9	58.52	0.005985	**

Appendix 2

Stepwise Logistic regression models of EGG parameter measures for individual languages: CQ, closing duration, PIC and PDC

Bo

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		1641.2	1643.2			
CQ	1	1482.8	1486.8	158.476	**2.20E-16**	***
SQ	1	1640.5	1644.5	0.789	0.37451	
Opening	1	1640.6	1644.6	0.63	0.42718	
Closing	1	1612.4	1616.4	28.81	**7.99E-08**	***
PIC	1	1618.2	1622.2	23.071	**1.56E-06**	***
PDC	1	1543.1	1547.1	98.146	**2.20E-16**	***
PIC_time	1	1637.6	1641.6	3.642	0.05633	.
PDC_time	1	1637.7	1641.7	3.533	0.06017	.

Yi

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		1463.8	1465.8			
CQ	1	1337.6	1341.6	126.199	**2.20E-16**	***
SQ	1	1462.8	1466.8	1.023	0.3118631	
Opening	1	1462.9	1466.9	0.943	0.3315606	
Closing	1	1463	1467	0.787	0.3749866	
PIC	1	1450.4	1454.4	13.455	**0.0002444**	***
PDC	1	1402.4	1406.4	61.407	**4.64E-15**	***
PIC_time	1	1458.8	1462.8	5.072	0.0243112	*
PDC_time	1	1458.9	1462.9	4.888	0.0270397	*

Hani

	Df	Deviance	AIC	LRT	*p*-value(Chi)	
<none>		2255.4	2257.4			
CQ	1	1954.7	1958.7	300.712	**2.20E-16**	***
SQ	1	2082.7	2086.7	172.702	**2.20E-16**	***
Opening	1	2238.4	2242.4	17.062	3.62E-05	***
Closing	1	2163.1	2167.1	92.366	**2.20E-16**	***
PIC	1	2188.2	2192.2	67.26	**2.38E-16**	***
PDC	1	2100.7	2104.7	154.694	**2.20E-16**	***
PIC_time	1	2224.8	2228.8	30.64	3.11E-08	***
PDC_time	1	2225.7	2229.7	29.741	4.94E-08	***

References

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

Anastaplo, S., and Karnell, M. P. (**1988**). “Synchronized videostroboscopic and electroglottographic examination of glottal opening, ” J. Acoust. Soc. Am. **83**, 1883–1890.

Andruski, J. E. and Ratliff, M. (**2000**). “Phonation types in production of phonological tone: the case of Green Mong, ” J. Int. Phonetic Assoc. **30**, 37–61.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics* (Cambridge University Press, Cambridge), pp. 1–390.

Baer, T., Löfqvist, A., and McGarr, N. S. (**1983a**). “Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques, ” J. Acoust. Soc. Am. **73**, 1304–1308.

Baer, T., Titze, I. R., and Yoshioka, H. (**1983b**). “Multiple simultaneous measures of vocal fold activity, ” in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by A. Bless (College Hill, San Diego), pp. 229–237.

Baken, R.J., and Orlikoff, R.F. (**2000**). *Clinical measurement of speech and voice* (Singular Publishing Group).

Belotel-Grenié, A., and Grenié, M. (**1994**). “Phonation types analysis in Standard Chinese,” Proc. Spoken Language Processing **94**, 343–346.

Berke, G. S., Moore, D. M., Hantke, D. R., Hanson, D. G., Gerratt, B. R., and Burstein, F. (**1987**). “Laryngeal modeling: Theoretical, *in vitro, in vivo*,” Laryngoscope **97**, 871–881.

Bishop, J., and Keating, P. (**2012**). “Perception of pitch location within a speaker’s range: Fundamental frequency, voice quality and speaker sex,” J. Acoust. Soc. Am **132**, 1100–1112.

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

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. Acoust. 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. Sp. Hear. Res. **33**, 245–254.

DiCanio, C. (**2009**). "The phonetics of register in Takhian Thong Chong," J. Int. Phonetic Assoc. **39**, 162–188.

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

Esling, J. H. (**1984**). "Laryngographic study of phonation type and laryngeal configuration," J. Int. Phonetic Assoc. **14**, 56–73.

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

Esposito, C. M. (**2010**). "Variation in contrastive phonation in Santa Ana del Valle Zapotec," J. Int. Phonetic Assoc. **40**, 181–198.

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

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels," Indian Linguistics **28**, 71–139.

Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," J. Int. Phonetic Assoc. **41**, 185–205.

Garellek, M., Esposito, C. M., Keating, P. and Kreiman, J. (To appear). "Voice quality and tone identification in White Hmong," J. Acoust. Soc. Am.

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

Guion, S., Post, M., and Payne, D. (**2004**). "Phonetic correlates of tongue root vowel contrasts in Maa," J. Phonetics **32**, 517–542.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hanson, H. M. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**, 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.

Hanson, H.M. (**2012**). "Methodologies used to investigate laryngeal functions and aerodynamic properties of speech", in *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. Huffman (Oxford University Press, Oxford), pp. 496–511.

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. Eurospeech **7**, 47–50.

Henrich, N., Doval, B., and Castellengo, M. (**2004**). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," J. Acoust. Soc. Am. **115**, 1321–1332.

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

Hillenbrand, J.M., Cleveland, R.A., and Erickson, R.L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Sp. Hear. Res. **37**, 769–778.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.

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. Sp. Hear. Res. **38**, 1212–1223.

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

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

Huffman, M.K. (**1987**). "Measures of phonation contrasts 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.

Karnell, M. P. (**1989**). "Synchronized videostroboscopy and electroglottography," J. Voice **3**, 68–75.

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

Keating, P., Esposito, C., Garellek, M., Khan, S., and Kuang, J. (**2011**). "Phonation contrasts across languages," Proc. ICPhS **17**, 1046–1049.

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

Khan, Sameer ud Dowla. (**2012**). "The phonetics of contrastive phonation in Gujarati," J. Phonetics **40**, 780–795.

Kirk, P., 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 Laurence C. Thompson*, edited by A. Mattina and T. Montler (Universtiy of Montana Press, Missoula, MT), pp. 435–450.

Kitzing, P. (**1982**). "Photo- and electrophysiological recording of the laryngeal vibratory pattern during different registers," Folia Phoniatr. **34**, 234–241.

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

Kreiman, J., Gerratt, B., and Antoñanzas-Barroso, N. (**2007**). "Measures of the glottal source spectrum," J. Sp. Lang. Hear. Res. **50**, 595–610.

Kreiman, J., Gerrat, B. G., and Khan, S. (**2010**). "Effects of native language on perception of voice quality," J. Phonetics **38**, 588–593.

Kuang, J. (**2011**). "Production and Perception of the Phonation Contrast in Yi," MA thesis, University of California, Los Angeles.

Ladefoged, P. (**1964**). *A phonetic study of West African languages* (Cambridge University Press, Cambridge).

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

Larson, K. K., Ramig, L. O., and Scherer, R. C. (**1994**). "Acoustic and glottographic voice analysis during drug-related fluctuations in Parkinson's disease," J. Med. Speech Lang. Pathol. **2**, 227–239.

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

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

Lee, S., Byrd, D., and Krivokapic, J. (**2006**). “Functional data analysis of prosodic effects on articulatory timing,” J. Acoust. Soc. Am. **119**, 1666–1671.

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**,103–109.

Marasek, K. (**1996**). “Glottal correlates of the word stress and the tense/lax opposition in German,” Proc. CSLP (University of Stuttgart, Stuttgart) **96**, 1573–1576.

Marasek, K. (**1997**). “Electroglottographic description of voice quality.” AIMS Working Papers Stuttgart, Vol. 3.

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

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

Michaud, A. (**2004**). “A Measurement from Electroglottography: DECPA, and its application in prosody,” Proc. Speech Prosody 2004, 633–636.

Miller, A. L. (**2007**). “Guttural vowels and guttural co-articulation in Ju|’hoansi,” J. Phonetics **35**, 56–84.

Mokhtari, P., Pfitzinger, H. R., and Ishi, C. T. (**2003**). “Principal components of glottal waveforms: towards parameterisation and manipulation of laryngeal voice quality,” Proc. VOQUAL, 133–138.

Mooshammer, C. (**2010**). “Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German,” J. Acoust. Soc. Am. **127**, 1047–1058.

Nair, G. (**1999**). *Voice tradition and technology: A state-of-the-art studio* (Singular Publishing Group, San Diego, CA).

Ní Chasaide, A., and Gobl., C. (**1997**). “Voice source variation,” in *The Handbook of PhoneticSsciences*, edited by W.J. Hardcastle and J. Laver (Blackwell Publishing Ltd, Oxford), pp. 427–461.

Pennington, M. (**2005**). “The Phonetics and Phonology of Glottal Manner Features,” PhD thesis, Indiana University.

Ramsay, J. O. (**1982**). "When the data are functions," Psychometrika **47**, 379–396.

Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (**1996**). "Functional data analyses of lip motion," J. Acoust. Soc. Am. **99**, 3718.

Ramsay, J. O., and Silverman, B. W. (**1997**). *Functional Data Analysis* (Springer, New York).

Ramsay, J. O., and Silverman, B. W. (**2002**). *Applied Functional Data Analysis: Methods and Case Studies* (Springer, New York).

Ratliff, M. (**1992**). *Meaningful Tone: A study of tonal morphology in compounds, form classes and expressive phrases in White Hmong*, Monograph Series on Southeast Asia (Northern Illinois University, Center for Southeast AsianStudies, DeKalb, IL), pp. 1–279.

Rothenberg, M., and Mahshie, J. J. (**1988**). "Monitoring vocal fold abduction through vocal fold contact area," J. Sp. Hear. Res. **31**, 338–351.

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

Shue, Y-L., Kreiman, J., and Alwan, A. (**2009**). "A novel codebook search technique for estimating the open quotient," Proc. Interspeech 2009, pp. 2895–2898.

Shue, Y-L., Keating, P. A., Vicenik, C., and Yu, K. (**2011**). "VoiceSauce: A program for voice analysis," Proc. ICPhS **17**, 1846–1849.

Sjölander, K. (**2004**). *The Snack sound toolkit* (KTH Stockholm, Sweden).

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

Stevens, K. N. (**1997**). "Articulatory-Acoustic-Auditory Relationships," in *The Handbook of Phonetic Sciences*, edited by W.J. Hardcastle and J. Laver (Blackwell Publishing Ltd, Oxford), pp. 462–506.

Stevens, K. N. (**2000**). *Acoustic Phonetics* (MIT press, Cambridge, MA).

Tehrani, H. (**2012**). http://www.linguistics.ucla.edu/faciliti/sales/software.htm, accessed by Dec. 5th 2012.

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

Titze, I. R. (**1984**). "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am. **75**, 570–580.

Titze, I. R. (**1989**). "A four-parameter model of the glottis and vocal fold contact area," Speech communication **8**, 191–201.

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