Title
Vocal fold vibratory patterns in tense versus lax phonation contrasts

Permalink
https://escholarship.org/uc/item/3fq7x9mb

Journal
The Journal of the Acoustical Society of America, 136(5)

ISSN
0001-4966

Authors
Kuang, Jianjing
Keating, Patricia

Publication Date
2014-11-01

DOI
10.1121/1.4896462

Peer reviewed

Vocal fold vibratory patterns in tense versus lax phonation contrasts

Jianjing Kuang

Department of Linguistics

University of Pennsylvania

619 Williams Hall

Philadelphia, PA 19104

Patricia Keating

Department of Linguistics

University of California, Los Angeles

Los Angeles, CA 90095-1543

Suggested running title: Tense-lax phonation contrasts

Correspondence address:

Jianjing Kuang
Department of Linguistics
University of Pennsylvania
619 Williams Hall
Philadelphia, PA 19104

Telephone: (215)746-3136
Email: kuangj@sas.upenn.edu

This study explores the vocal fold contact patterns of one type of phonation contrast – the tense vs. lax phonation contrasts of three Yi (Loloish) languages. These contrasts are interesting because neither phonation category is very different from modal voice, and because both phonations are largely independent of the languages' tonal contrasts. Electroglottographic (EGG) recordings were made in the field, and traditional EGG measures were derived. These showed many small but significant differences between the phonations, with tense phonation having greater contact quotients and briefer but slower changes in contact. Functional Data Analysis was then applied to entire EGG pulse shapes, following Mooshammer (2010). The resulting first Principal Component was found to be mostly strongly related to the phonation contrasts, and correlated with almost all the traditional EGG measures. Unlike the traditional measures, however, this component also seems to capture differences in abruptness of contact. Furthermore, previously-collected perceptual responses from native speakers of one of the languages correlated better with this component than with any other EGG measure or any acoustic measure. The differences between these tense and lax phonations are not large, but apparently they are consistent enough, and perceptually robust enough, to support this linguistic contrast.

PACS numbers: 43.70.Kv, 43.70.Mn

I. INTRODUCTION

Voice qualities are used in many languages allophonically, either as prosodic cues (e.g. stress and focus in German: Mooshammer, 2010) or as enhancement cues of other distinctive features (e.g. creak in the Mandarin low dipping tone: Belotel-Grenié and Grenié, 1994). But relatively few languages use phonation by itself as a phonemic dimension -- that is, with two, three, or even four voice qualities distinguishing lexical meanings just as consonants, vowels, and tones can do. The focus of this paper is the two-way tense vs. lax phonation contrast found in many Tibeto-Burman languages. These languages are generally also tonal, and both phonation types occur with both mid and low tones[1]; that is, tone and phonation type are largely independently contrastive. For example, in Southern Yi (spoken in China), the low-tone syllable /be^{21}/ with a lax phonation means “mountain”, whereas /be̲^{21}/ with a tense phonation means “foot”, while mid-tone lax /be^{33}/ means “fight” and tense /be̲^{33}/ means “shoot”. (Here, tense phonation is indicated by an underscore, and the superscript numbers indicate the starting and ending pitches of the tone, 1 to 5 indicating lowest to highest. Such superscript numbers inside brackets always refer to tones, never footnotes.) Despite the clear linguistic role of the phonation difference, these phonations are little-studied in terms of either their production – laryngeal muscle activity, glottal configuration, or vocal fold vibration patterns – or their perception.

The main goal of the present study is to better understand the articulation of these tense vs. lax phonation contrasts. Electroglottographic (EGG) recordings from three Tibeto-Burman languages – Southern Yi, Bo, and Luchun Hani – will be analyzed in detail to understand better how these contrasts are produced. In order to understand the glottal events in tense and lax phonations, it is necessary in turn to better understand the information provided by EGG signals: how do EGG pulse landmarks and EGG pulse shapes relate to each other and to the perception of phonation? In this connection, this study will follow Mooshammer (2010) in applying Functional Data Analysis to EGG pulse shapes, and will further relate the components of the shapes obtained in this way to other EGG measures, as well as to perception. Thus our secondary goal is to contribute to the understanding of EGG measures of phonation type.

II. BACKGROUND

A. 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 largest population in this language family branch. Its approximately fifty languages are geographically distributed in Yunnan, Sichuan and Guizhou provinces of China, and spoken by more than six million people (Ethnologue, 2012). Yi languages typically have a CV syllable structure, a seven-vowel system, a three-tone system (low, mid and high, noted as 21, 33 and 55), and two phonation registers: tense vs. lax.

B. Production of phonation types

Ladefoged (1971) proposed that phonations can be seen as lying along a single-dimension continuum of glottal aperture, bounded by the voiceless extremes of, on one end of the continuum, the most open glottis to, on the other end of the continuum, the most closed glottis (glottal stop). In this scheme, voicing varies from a relatively open glottis (breathy voice), through a normal glottal position (modal voice), to a relatively closed glottis (creaky voice). Intermediate positions along this continuum are commonly referred to as “tense” and “lax” voice or phonation[2]. Tense phonation is then in the creaky-modal range whereas lax phonation is in the breathy-modal range. (Other linguistic voice qualities, such as harsh or strident voice, would involve additional, supraglottal, mechanisms.)

These four types of non-modal phonations (breathy, lax, tense, creaky) are defined relative to modal phonation. The summary descriptions that follow are largely from Gobl and Ni Chasaide (2012), in turn based in part on Laver (1980). In modal phonation, the vocal folds vibrate along their entire lengths, often with full glottal closure, and their vibration is periodic. In breathy phonation, muscle tensions (e.g. crico-thyroid, vocalis and/or inter-arytenoid) may be low, the overall glottal aperture is increased (either because the peak area is larger, or because there is a constant leakage), the closure of the glottis is incomplete or of short duration, the vibratory pulse (glottal airflow or area) is more symmetrical in shape (Sundberg and Gauffin, 1979; Chen et al., 2013), and glottal airflow is high, leading to a strong first harmonic (both in absolute terms, and relative to higher harmonics) and audible frication noise (Sundberg and Gaufin, 1979; Klatt and Klatt, 1990; Gordon and Ladefoged, 2001; Pennington, 2005). Lax phonation is supposed to have a similar glottal configuration, but less extreme, being only “slightly breathy” (Ladefoged and Maddieson, 1996).

In contrast, creaky voice involves small movements by thick and short vocal folds, with long closed phases and low glottal flow (Gerratt and Kreiman, 2001). Due to increased muscle tension, the glottal aperture is reduced relative to modal, and only the membranous parts of the folds vibrate, with complete and abrupt closures (Gobl and Ni Chasaide, 2012, who use the term "ligamental" instead of "membranous"). Vibration is either low-frequency (such that the individual pulses may be perceptually resolved – this type especially is often called vocal fry, or pulse register), or irregular (due to greater tension and possible incursion of the ventricular folds, e.g. Childers *et al.*, 1990)[3]. The vibratory pulse shape is more skewed, increasing the energy in the higher frequencies. Glottal airflow is lower, and the first harmonic weaker, than in modal voice. Tense phonation is supposed to be similar, but again less extreme. Laryngealization (the term used by e.g. Klatt and Klatt, 1990) and pressed voice (the term used by e.g. Sundberg and Gauffin, 1979) are more general terms for phonation with a constricted glottis due to greater vocal fold adduction; these terms are used especially when the vibration frequency is neither low nor irregular (so, not perceptually creaky). Tense phonation is thus an instance of laryngealization or pressed voice.

These descriptions would seem to suggest that tense and lax phonations are produced quite similarly to creaky and breathy phonations, respectively (Gobl and Ni Chasaide, 2012). However, tense and lax phonations are sometimes both considered to be more like modal phonation than either are like creaky or breathy phonation (Gordon and Ladefoged, 2001), though since they contrast in languages, they cannot both be simply modal. Some insight into the relations among these phonations emerges from a recent cross-linguistic acoustic study of 24 phonation categories from ten languages (Keating *et al.*, 2012). Breathy, lax, modal, tense, and/or creaky phonations were obtained from multiple speakers of each language, and 14 acoustic measures of phonation were extracted. The tense and lax categories of the three Tibeto-Burman languages in the present study were included in the earlier study. The acoustic measurements were then rendered in a 2-dimensional physical space by multi-dimensional scaling. (See Keating *et al.*, 2012 for details.)

The 24 categories organized in the resulting space are shown in Figure 1. In this figure, 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. (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.) The five broad cross-language phonation types of the Ladefoged-style continuum model are also labeled, with the lax and tense tokens circled. It can be seen that in this acoustic space, the five 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 Ladefoged's continuum, but with substantial overlap among the categories. 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*; these measures are described below. 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 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.

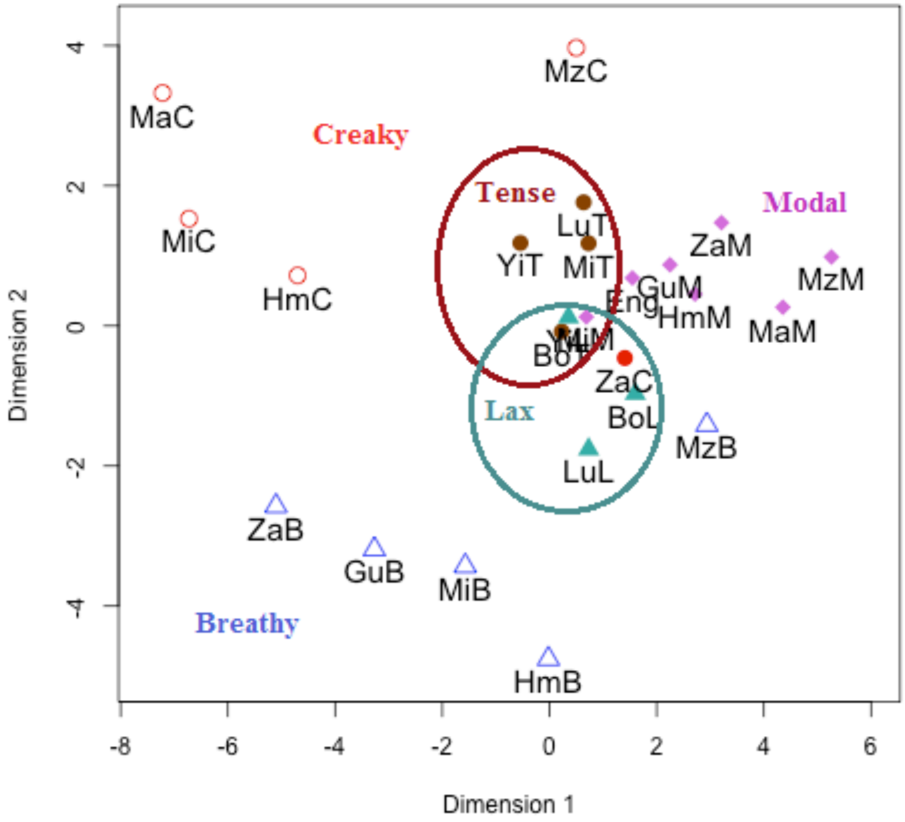


FIG. 1. Acoustic space for 24 phonation categories from ten languages (from Keating *et al*., 2012). The first 2 letters of each label indicate the language: Bo, Gujarati, Hmong, Luchun Hani, Mandarin, Miao, Mazatec, Yi, Zapotec; the third letter, plus the plotting symbol and the color (online only), indicates the phonation (breathy/ open triangles, creaky/ open circles, lax/ filled triangles, modal/ filled diamonds, tense/ filled circles); Eng indicates English. The circled phonation types are the tense vs. lax. They are similar to each other, and near the modal range.

C. Previous studies of Tibeto-Burman phonation types

The acoustic correlates associated with tense vs. lax contrasts in Tibeto-Burman languages have received some attention in the literature. Several studies have identified the amplitude difference between the first and second harmonics (H1-H2) as differentiating the two phonations, with the lax phonation having higher values, i.e.

relatively more energy in the fundamental (Maddieson and Ladefoged, 1985 for Jingpo, Hani, Eastern Yi, Wa; Maddieson and Hess, 1986 for Jingpo, Liangshan Yi, Wa; Shi and Zhou, 2005 for Southern Yi from Xinping village; Kong, 2001 for Northern Yi, Zaiwa, Jingpo; Kuang, 2011 for Southern Yi (the same corpus used here, in part from Xinping village)). Other acoustic correlates that show that lax phonation has relatively more energy in the fundamental, and tense phonation has more overall harmonic energy are the harmonics-to-noise-ratio measure Cepstral Peak Prominence (Kuang, 2011) and the amplitude difference between the fundamental and the harmonic nearest F1 (H1-A1: Kong, 2001; Kuang, 2011) or nearest F2 (H1-A2: Kong, 2001). Maddieson and colleagues (Maddieson and Ladefoged, 1985; Maddieson and Hess, 1986) have found differences (for some though not necessarily all of these languages) in fundamental frequency, vowel duration, and Voice Onset Time (VOT) of the preceding consonant (where VOT is the duration from the consonant release to voicing onset). The first formant frequency F1 also differs (Maddieson and Ladefoged, 1985; Shi and Zhou, 2005).

A full acoustic analysis of the corpus analyzed in the present paper was given by Kuang and Keating (2012). Six successful acoustic correlates of the contrasts were identified: H1*[4], H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, and Cepstral Peak Prominence (CPP). In regression analyses (Kuang, 2013:39), CPP was relatively less important than H1*-related measures; H1*-H2* and H1*-A1* were the most important, and were more distinctive in low tone than in mid tone. In sum, it has been well-established that Tibeto-Burman tense-lax contrasts involve multiple acoustic correlates.

On the other hand, there have been few studies of physiological correlates of these contrasts. Maddieson and Ladefoged (1985) found that lax phonation has greater oral flow, indicating a more open glottis offering less resistance. A laryngoscopic study of Northern Yi by Esling *et al*. (2000) showed that the tongue root is retracted for the tense phonation, contributing to a harsher quality in that language. Finally, Kuang (2011) found that in the present Southern Yi corpus, the electroglottographic Contact Quotient was higher for the tense phonation. This latter finding is the only direct evidence to date that the contrasts are clearly glottal in nature. The present study pursues this approach.

D. Electroglottographic measures of phonation types

Electroglottography (EGG) measures changes in the vocal fold contact area during phonation (Fabre, 1957; Fourcin, 1971; see Baken and Orlikoff, 2000). A small, high frequency current is passed between two electrodes placed on each side of the larynx. Variation in the electrical impedance across the larynx is produced by the opening and closing of the glottis: the larger the contacted area, the larger the measured admittance. Since the signal is neither calibrated nor likely to be linear, and since most EGG recordings do not preserve the DC component of the signal, it reflects relative rather than absolute contact. EGG is non-invasive and does not interfere with natural speech production; 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 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, 2009), Takhain Thong Chong (DiCanio, 2009), Gujarati (Khan, 2012), and White Hmong (Esposito, 2012)).

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[5] (Rothenberg *et al.*, 1988; Baken and Orlikoff, 2000). CQ 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", she may have been referring to *individual speaker* voice quality; certainly, previous studies have in fact shown that CQ varies with *linguistic* voice quality. Mooshammer (2010) found that CQ reliably varies with lexical stress and vocal effort in German (stressed syllables have a shorter open phase).[6] 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 wider average aperture. Such CQ differences among contrastive phonations have been demonstrated for several languages by Mazaudon and Michaud (2009), DiCanio (2009), Esposito (2010, 2012), Khan (2012), and for allophonic phonations in Mandarin by Keating *et al.* (2012).

Nonetheless, CQ is not unproblematic. It depends on estimates of the glottal closing and opening moments from the EGG cycle, yet these moments are not necessarily seen in the EGG signal. Many efforts have been made to compare EGG signals with more direct physiological signals (e.g. Kitzing, 1982; Anastaplo and Karnell, 1988; Karnell, 1989; Baer et al., 1983a,b; Childers *et al.*, 1990; Berke *et al.*, 1987; Holmberg *et al.*, 1995; Herbst and Ternström, 2006; Henrich *et al.*, 2004; among many others), and these have suggested that the positive and negative peaks in the derivative of the EGG signal indicate glottal closing (contacting along the entire length of the vocal folds) and opening respectively. Unfortunately, while the positive peak is usually very strong and clear, the negative peak often is not, making the moment of glottal opening at best uncertain, and CQ potentially unreliable.

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 contacting duration and decontacting duration, and thus a measure of the symmetry of the glottal pulses[7]. As creaky phonation pulses are usually shorter in contacting and longer in decontacting, their pulse shape is skewed, while breathy phonation pulses, with similar contacting and decontacting durations, show a more symmetrical shape. 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.

A new measure employed by Michaud (2004) is 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 in the dEGG, 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). As noted above, 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 (just as a greater negative peak in the derivative of the glottal flow boosts high-frequency energy (Gobl and NiChaisade, 2012)). This measure has the advantage of not depending on landmarks in the EGG pulse, though it has the disadvantage of being potentially very sensitive to signal artifacts. Michaud used the measure to study focus prosody, but later studies of contrastive phonation types have shown that breathy, creaky, and/or modal phonation categories differ in PIC, with breathy voice having the highest values (e.g. Kuang, 2011; Esposito, 2012; Keating *et al.*, 2012). These differences have been found even though a measure like PIC can be influenced by many factors affecting the EGG signal and its derivative, suggesting that comparing relative differences across linguistic categories within a language is fairly robust to signal noise. The direction of the difference is perhaps surprising, since breathy voice is generally thought to involve the least abrupt vocal fold closing, yet shows the highest values of PIC. The relation of the PIC measure to abruptness of contact will be considered as part of the present study.

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 sex variability in EGG pulses, which make it harder to find consistent properties of voice qualities. More advanced statistical approaches, which do not require landmarks to be defined, have also been explored. Makhtari *et al*. (2003) and Kreiman *et al*. (2007) applied principal component analysis (PCA) to glottal pulses derived from audio pulses by inverse filtering, and Chen *et al.* (2013) applied to PCA to glottal area pulses. Recently, Mooshammer (2010) adopted Functional Data Analysis (FDA) to study EGG pulse shapes as a function of lexical stress in German. This 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 compare the overall shapes of entire and continuous EGG pulses from different speech conditions. In FDA, traditional multivariate analyses such as Principal Component Analysis are expressed in functional analytic terms, and thus FDA "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. This study will adopt this method to describe the EGG contact patterns in tense vs. lax phonation contrasts, as it can factor out non-phonation variability in the glottal pulses (e.g. sex, speaker), and reveal the essential gestural patterns of phonation contrasts without the problem of defining landmarks.

III. METHOD

A. Recordings

All the data in this study were obtained from recordings made during a trip to Yunnan province of China in summer 2009. All the speakers were recruited from Southern Yi villages (Xinping and Jiangcheng), Bo villages (Shizong and Xingfucun) and a Hani village (Luchun). Nine speakers (5M, 4F) were recorded for Bo, twelve speakers (6F, 6M) for Yi, and nine speakers (5M, 4F) for Hani. The speakers all self-reported to have good vocal health, but differed in age and other factors; all such speaker differences will be subsumed statistically by random effects for speaker. Before the recordings were made, one or two speakers of each language served as consultants from whom a wordlist of two thousand words[8] was elicited for each language. From these lexicons, minimal phonation-contrast pairs of monosyllabic words with a variety of consonants and vowels (around 40 pairs for each language) were selected. (Since in these languages the phonation contrast does not occur with high tone, no high tone tokens were elicited.) The minimal pairs for each language were then checked with each of the speakers. Individual speakers were instructed to produce only the pairs they knew; as a result, the number of tokens produced varies across speakers. This restriction to minimal pairs is important in guaranteeing that all comparisons are within-pairs: each item finds its own control in the other item in its pair.

For all 30 speakers, simultaneous EGG and audio recordings were made. Each word was repeated twice. The signals were recorded directly to a computer via its sound card, in stereo, using Audacity, at a 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) and recorded as the second stereo channel. A total of 3867 tokens were analyzed. All of these recordings, audio and EGG, are freely available for download at http://www.phonetics.ucla.edu/voiceproject/voice.html.

Details of the EGG signal depend on how the signal was acquired, though published studies rarely give any relevant information about this. This Glottal Enterprises model offers a choice of high-pass cutoff frequencies. Rothenberg generally recommends 20 Hz (indeed, newer models have this frequency as the sole, hard-wired, option), which is thought to be the cutoff used in other EGG brands, but has sometimes recommended higher cutoffs. 40 Hz was used here to minimize vertical shifts in the signal. Comparisons in our laboratory indicated that there is little difference in the signal shapes for 20 vs. 40 Hz cutoffs[9] (Rothenberg, 2002), with the distortion greater in the decontacting phase of the pulse, and thus affecting pulse symmetry. In any event, the signal was captured by the laptop's soundcard, which is AC-coupled and so does not preserve any low-frequency components. Thus the signals here reflect only contact changes, not absolute contact levels; and the signal is systematically distorted relative to signals acquired down to DC (Baken and Orlikoff, 2000: 424; Rothenberg, 2002).

Furthermore, a constant low-frequency hum at 50 Hz is present in these EGG recordings, which will also influence signal shape. As EGG recordings of women very often have lower signal to noise ratio (SNR) than those of men, due to sex differences in the shape of the thyroid cartilage, this low-frequency hum is more noticeable for some women's signals. The overall SNRs for individual speakers' recordings are shown in the Appendix 1; sub-sample analyses, not presented here, indicate that the data patterns are very similar across SNRs.

B. Measurements

1. EGG parameters

The EGG signals were processed by EggWorks (Tehrani, 2012) to obtain the traditional landmark-based parameter measures. Figure 2 illustrates the EGG parameters. Contact Quotient (CQ) is estimated by the “hybrid” method (Howard, 1995), which uses the dEGG contacting peak for detecting the glottal contact event, and a 3/7 threshold on the EGG signal for detecting the glottal opening event (CQ_H method in Figure 2). This method has previously been shown to be the best for such comparisons (Kuang, 2011). Two measures are made from the dEGG signal: Peak Increase in Contact (PIC in Figure 2), defined as the amplitude of the positive peak of dEGG; Peak Decrease in Contact (PDC in Figure 2), defined as the amplitude of the negative peak of dEGG. EggWorks also provides the times of PIC and PDC. Finally, contacting duration and decontacting duration are measured at a 10% threshold (Marasek, 1997), and Speed Quotient is computed as the ratio between these durations. All of these measures, especially the threshold-based measures and the measures from the decontacting phase, will be influenced by signal noise and filtering, and the absolute values of the measures are not meaningful. Only relative differences will be discussed, keeping in mind that variability in the measurements due to noise should make it harder to find consistent differences between phonations.

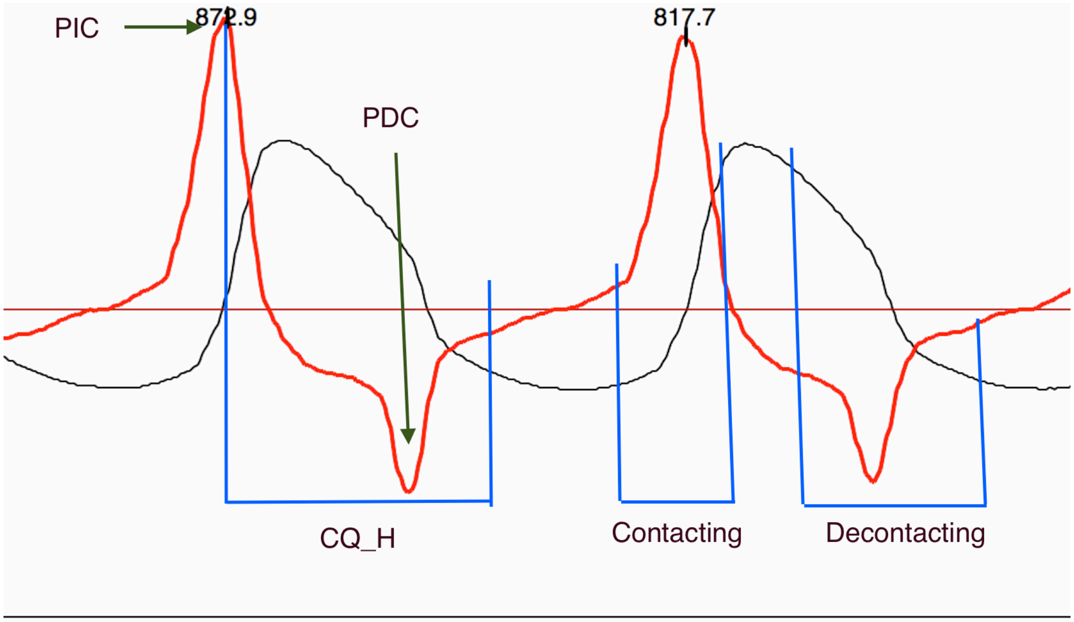


FIG. 2. (Color online) Illustration of EGG measures from EggWorks. The darker line is the EGG pulses, and the lighter line is the dEGG signal. The left pulse illustrates the hybrid method of estimating CQ (Howard, 1990), and PIC and PDC from dEGG, and the right pulse demonstrates measures related to skewness of the pulse: contacting duration and decontacting duration (SQ=T-cont/T-decont).

2. EGG pulse shape components

All aspects of the Functional Data analysis of EGG pulses presented here follows the procedure by Mooshammer (2010). A subset of the tokens in the corpus was selected for this analysis: the consonant-vowel combination [be] was chosen, with all values of phonation type (lax vs. tense) and tone (low vs. mid). Six speakers (3 males and 3 females) were selected, based on their familiarity with the greatest number of the set of minimal pairs, as an index of their knowledge of the language (a criterion that favors older speakers over younger ones). 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 cycle, and each selection was checked by hand. In order to cancel the effect of the length of the pulses due to f0 differences, the selected pulses were time-normalized to a uniform length of 1000 samples by a Matlab script using linear interpolation, just as in Mooshammer (2010). The pulse pairs were then amplitude normalized by setting both decontacting peaks in the EGG signal to 0, and then setting the amplitude of the highest EGG peak in each pair to 1. This amplitude normalization is illustrated in Figure 3.

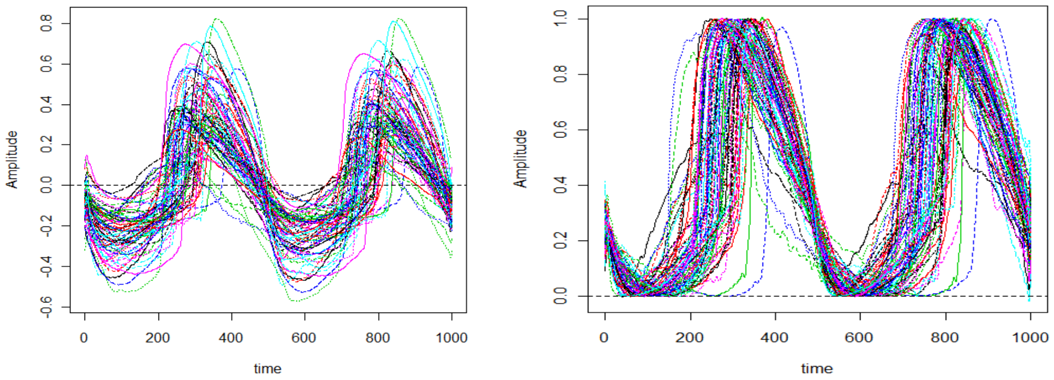


FIG. 3. (Color online) EGG pulses before amplitude normalization (left panel) and after amplitude normalization (right panel); in both panels time is normalized into 1000 intervals.

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 data smoothing is done with the third order of derivative whose square is integrated and weighted by lambda as 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.

Because different statistical methods are used for the different analyses, these methods will be described separately where relevant in the following section.

IV. RESULTS

A. EGG parameters

For each language, a series of mixed-effect models (R package *lme4* (Bates *et al.* 2013)) 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. P-values were obtained using R's *pvals.fnc*, which is based on the Markov chain Monte Carlo (MCMC) method (Baayen, 2010). The main effects of phonation (tense, lax) in the three languages are summarized in Table I, and detailed mean and standard deviation values are in Appendix 2. Only significant effects (at *p*-MCMC < .05 level) are reported in the table, 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. Figure 4 shows sample EGG and dEGG pulses, from a Yi speaker.

(TABLE I about here)

There was no main effect of tone on any measure[10] (other than f0, as expected). In contrast, all the non-f0 EGG parameter measures (except for decontacting duration and SQ, two measures likely to be affected by signal filtering) successfully distinguish the lax phonation from the tense phonation for all three languages, and the pattern is very consistent. Thus the physiological mechanism of tense vs. lax phonation is generally independent from that of tonal production. For each language, a series of forward stepwise logistic regressions was then employed to evaluate the relative importance of the different parameter measures for phonation contrasts. The measures that reliably distinguish the phonation contrast across the 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 tests (results for individual languages are in Appendix 3). The results suggest that CQ, contacting duration, PIC and PDC are the parameters that contribute most to the phonation contrasts in these EGG signals.

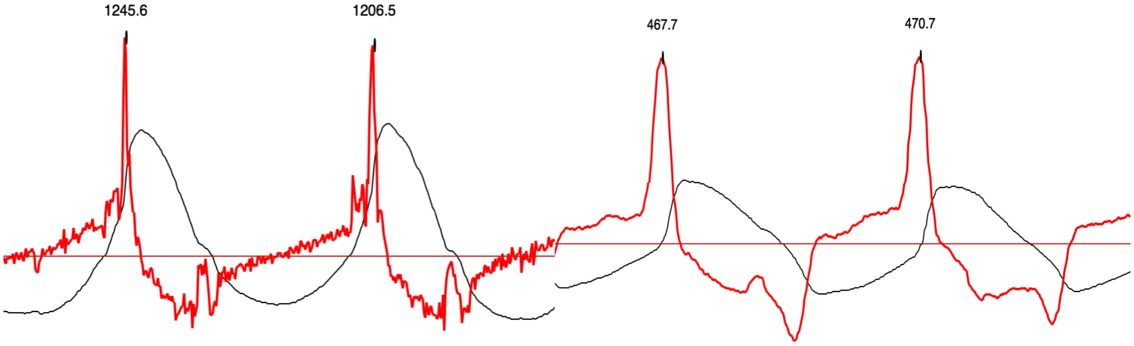


FIG. 4. Sample EGG pulses from tense and lax phonations in southern Yi. Left= Lax; Right= tense. Vertical scale for EGG signals is normalized between -1 and 1; note that the dEGG display scales differ between left and right panels. PIC in the left panel ≈ 1200; PIC in the right panel ≈ 400. Darker lines = EGG signal; lighter lines = dEGG signal.

As expected, the tense phonation consistently has a greater CQ than the lax phonation, suggesting a smaller open quotient. However, the differences are not large; the mean CQ values for both phonation categories lie between 0.4 and 0.5. The tense phonation also has relatively shorter contacting intervals and earlier moments of peak change in contact (PIC) (and thus also smaller SQ ratios) than the lax phonation. However, as seen in Figure 4, the tense phonation also has smaller peaks in the dEGG signal (PIC and PDC, i.e. lower peak rates of change in contacting and decontacting).

In sum, almost all the EGG parameters differ between the two phonations, with CQ and contacting duration the most important. However, CQ is ambiguous, in that changes of CQ can be caused by many different glottal articulations. For example, Esposito and Khan (2012) found that Gujarati and White Hmong have different strategies in producing breathy phonation, although they both consistently have a smaller CQ. Because of such variation in articulations, focusing on CQ, as most studies do, provides only partial information, and does not reflect every aspect of the underlying articulation. As shown in this section, so many parameters are different in phonation contrasts, that it is very probable that the entire pulse shapes differ. Therefore, in the following section, the overall shapes will be analyzed, and related to the EGG parameters.

B. EGG pulse shape components

At first separate FDA analyses for the individual languages were performed, and the patterns were very similar (though the pattern for Hani is the clearest, as the Hani speakers have overall better SNR compared to the speakers from the other two languages). Therefore, the languages were collapsed together and a single FDA analysis was performed for tense vs. lax contrasts across all three languages. This analysis gives functional Principle Components. Figure 5 shows the first four principle components (PC1 to PC4) with their factor scores, 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. It can be seen that for each factor, either the positive or the negative factor value reaches the maximum normalized amplitude of 1.0, while the other factor value does not[11].

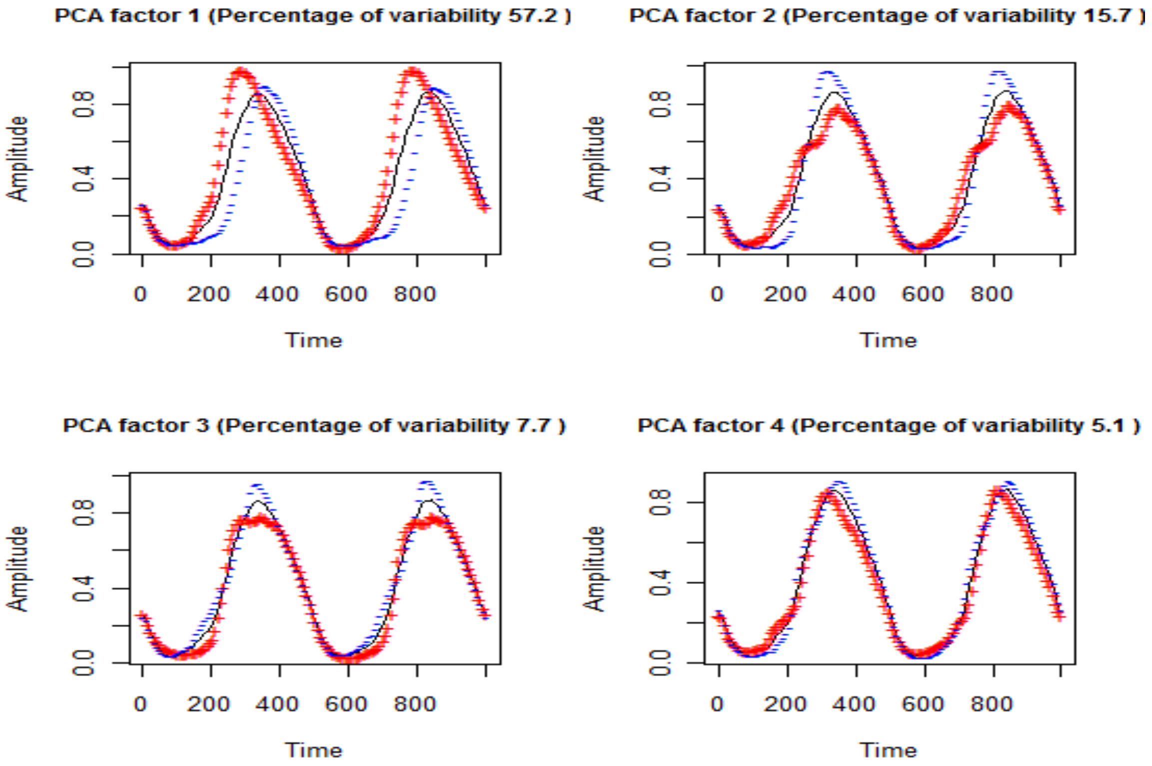


FIG.5. (Color online) The shapes of the EGG signals. Negative factor values are indicated by minus signs, positive factor values by plus signs, and the mean curve by a solid line, for the first four factors.

To help understand the meaning of the positive/negative deviations from average pulses seen in Figure 5, Figure 6 provides a graphical representation of the relevance of the experimental factors on each of the four principal components of the shapes, following Mooshammer (2010). For each sex 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 5. PC1 shows a strong phonation effect: positive values in Figure 6 are the tense phonation and negative values are the lax phonation. Therefore, the plus line for PC1 in Figure 5 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 sex effect in Figure 6, 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 5 can be taken to mostly represent the male pulses, and the minus line the female pulses. Finally, PC4 in Figure 6 singles out 33T. Therefore, the plus line for PC4 in Figure 5 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 sex effect is mostly seen in PC2 and PC3; and an effect of 33T is seen in PC3 and PC4. These interpretations are confirmed quantitatively by repeated ANOVAs between the component scores and the experimental factors (language, sex, phonation and tone) given in Table II. By far the strongest relation is between PC1 and the phonation categories. The forward stepwise logistic regression with EGG parameter measures and Principal Components was also replicated. Unsurprisingly, PC1 is the most significant predictor (p =7.41-06), and it is more significant than CQ (p=5.24-05).

(TABLE II about here)

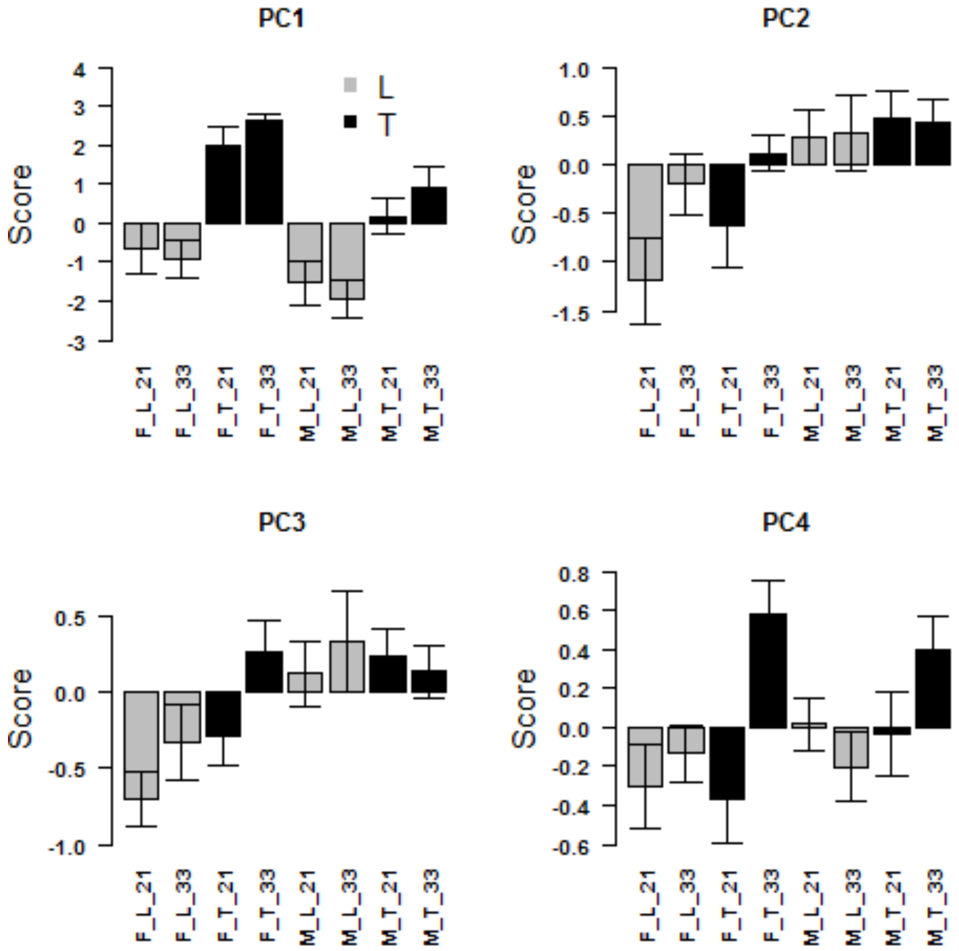


FIG. 6. Mean scores and standard errors of factors in each principle component for each sex/phonation/tone condition. F=female, M=male, L=lax phonation (light bars), T=tense phonation (dark bars), 21=low tone, 33=mid tone.

Because PC1 is the most important component (by definition), and because the phonation factor has the largest effect on PC1, it follows that phonation explains most of the variability in the EGG pulse shapes, more important than tone or speaker sex. It can be seen in Figure 5 that the main difference between the tense and lax phonations lies in the contacting part of the pulse: relatively speaking, the tense phonation (the plus line in

PC1) has a greater change in contact, larger contact quotient, and shorter contacting duration (steeper overall positive slope). The decontacting part of the pulse patterns the other way, with tense phonation having a longer duration, resulting in a more skewed pulse; this can be seen not only in PC1, but in PC4 (associated with the tense 33 tone). The tense phonation appears to show a more abrupt contacting phase in PC1 (for further discussion, see Kuang and Keating, 2012). Abruptness of contact is a very important property of vocal fold vibration, as it has a significant impact on glottal flow and acoustics, but it has been hard to capture in EGG signals by annotating landmarks (Marasek, 1996, 1997). Certainly the PIC landmark measure does not capture this property; indeed, it patterns in quite the opposite direction.

C. Correlations among EGG measures

It has been shown that phonation category affects nearly all the traditional EGG parameters (Table I), and that the first principal component from FDA is most strongly related to phonation (Table II). Thus it would be expected that the traditional EGG parameters should be correlated with PC1 (and to a lesser extent with PC4, which is weakly affected by phonation). The relations between the traditional EGG parameter measures and the FDA principal components were tested by Spearman correlation coefficients calculated between factor scores (from the FDA of pulse shapes) and EGG parameters (from signal landmarks), and are shown in Table III. The expected correlations were generally found. The strongest correlation is between CQ and PC1, the two measures most clearly related to the phonation contrasts. CQ is also more weakly correlated with PC4, which singles out the tense mid tone. PIC and PDC, two of the other parameters found to be important for the phonation contrasts, also have their strongest correlations with PC1, and Contacting Duration is also correlated with PC1, though the PIC and Contacting Duration correlations are not very strong. Less expected is that SQ is also most strongly related to PC1. F0 is only weakly correlated with any of the principal components; its strongest correlation (r=0.24) is with PC1; it is more strongly correlated with the EGG parameter measure Decontacting Duration (r=-0.45).

(TABLE III about here)

The fact that multiple EGG parameters are correlated with the principal components reflects that the components capture more holistic aspects of the pulse shapes than any one parameter does. As described above, PC1 shows a relatively greater extent of contact (more complete contact), larger contact quotient (wider pulse), and shorter contacting duration (steeper positive slope) for the tense phonation. These can all affect CQ, though in different ways; PC1 helps in teasing these apart.

V. RELATION OF PRODUCTION TO PERCEPTION

In the previous section, it was shown that there is one contacting pattern underlying the portion of the variance related to the phonation contrast (PC1), and FDA successfully factored out the variations caused by sex/speakers (PC2 and PC3). In order to validate that these different articulatory components are “real” to native listeners, it is necessary to show that these principal components of articulations also have different perceptual consequences. Specifically, it is important to know which of these aspects of articulation 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 Southern Yi (Kuang, 2011), and examine the correlations between the perception responses obtained in that earlier study and the gestural measures obtained in the present study, as well as with the acoustic measures from Kuang and Keating (2012).

A. Summary of earlier perception results

The Kuang (2011) study tested a set of eight Southern Yi words which varied along three parameters: vowels ([bu] vs. [be]), phonations (tense vs. lax), and tones (21 vs. 33). These words had been produced in isolation by six speakers (3 females, and 3 males), giving 48 test stimuli. The perception experiment was an AXB identification task in which the A and B items served as standards of the tense and lax categories to which the X stimulus was matched. A male and a female speaker each produced half the A and B items; both speakers were chosen because they seemed to maintain a good contrast between the phonation categories. With repetitions, a total of 80 AXB stimuli were included, and the experiment yielded 2160 (80 stimuli x 3 repetitions x 9 listeners) responses for tense vs. lax identification. See Kuang (2011) or Kuang and Keating (2012) for further details.

Kuang's listeners were largely able to perceive the phonation contrasts across the vowels and tones in this stimulus set. By vowel, the average accuracy rate for low vowels was 73.5% compared to 57% for high vowels, while by tone, the average accuracy rates for the two tone conditions were nearly identical: 66% for the low tone and 64.5% for the mid tone. Such less-than-perfect, but above chance, perceptual performance is well-suited to statistical testing of listeners' attention to different potential cues in the stimuli.

B. Replications of previous analyses

The 48 items tested in the perception experiment are a subset of the larger corpus analyzed in the present study. Comparisons of the EGG and audio recordings for just this subset with those for the whole corpus confirm that the subset is representative of the whole: unpaired t-tests on all EGG and acoustic measures show no reliable differences between the subset and the whole (all p-values > 0. 05); FDA on just the subset yielded PCA components very much like those in Figure 5, accounting for similar variance (74.4% for the first two components); correlation analyses likewise show that only PC1 is dominated by phonation, and that the phonation-related shape components (PC1 and PC3) are most related to CQ. SQ and contacting duration are also important EGG parameters that can capture aspects of the phonation-related pulse shapes. F0, in contrast, shows a strong correlation (r=-0.6) with decontacting duration.

C. Production-perception relations

Given these analyses of the EGG signals of the 48 test stimuli, the correlations of the mean perceptual accuracy scores from Kuang (2011) with the different characterizations of the EGG pulses, and with the acoustic measures, can now be calculated.

Considering the FDA analysis, significant correlations with perception accuracy are found only for PC1 (r=0.49, p<0.001) and, more weakly, for PC3 (r=0.13, p<.05). Specifically, as listeners had to attend to the phonation contrast in order to classify the stimuli, the identification accuracy rates are only highly correlated to PC1, while PC2, a shape component related to sex, is successfully factored out by listeners. Although PC3 and PC4 are not major components here, the sensitivity of detecting phonation-related gestures/acoustic correlates is still noticeable, as there is a weak correlation with PC3 (related in part to phonation).

Crucially, this study is especially 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), retrieved from our earlier acoustic study (see section II.C above), (2) EGG parameters (CQ, SQ and duration of contacting 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. This result agrees with the result from the acoustic analysis of the full dataset, which showed that the most important acoustic parameters were H1*-H2* and H1*-A1*, and with the result that PC1 is correlated with the various H1* measures. The correlation coefficient between perception accuracy and H1* is 0.28 ($p < .01$). Among the EGG parameters, CQ is the best predictor, and SQ and contacting duration make significant contributions as well. The correlation between CQ and perception accuracy is 0.35 ($p < .001$).

All of these relations are weaker than the relation already presented 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, listeners' perception varies in relation to small changes in the aspects of EGG pulse shape seen in the first principle component of pulse shape (PC1). This single glottal pulse shape factor is a 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.

VI. DISCUSSION AND CONCLUSION

The goal of this study has been to clarify the nature of one type of phonation contrast: tense vs. lax phonation contrasts in three Yi languages of the Tibeto-Burman family. EGG has the benefits of being non-invasive and not interfering with speech, and thus is 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 well understood, especially in the face of large inter-speaker and inter-sex variability. This study therefore presented a new analysis of the articulatory EGG patterns for the tense vs. lax contrasts, exploiting multiple electroglottographic analysis methods in order to better understand both the production of tense vs. lax phonations in the Tibeto-Burman family of languages, and the EGG signal itself. It explored 1) the important landmark-based parameters of the EGG pulses for tense vs. lax phonations; 2) the relative shape components of the EGG pulses for tense vs. lax; 3) how these EGG properties relate to each other, and 4) how listeners' perception of the contrast relates to all the available measures, both EGG and acoustic.

A. Summary of EGG results

Consistent cross-linguistic contrast patterns were found for EGG parameters. CQ and contacting duration are the most important properties of the EGG waveforms, and the amplitudes of positive and negative peaks of dEGG are also important parameters. That there are so many significantly different landmark measures suggests that the overall pulse shapes differ. Therefore, following Mooshammer (2010), this study adopted Functional Data Analysis, 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. This analysis was applied to a corpus of pulses that differ in phonation type (tense vs. lax), lexical tone (low vs. mid), and speaker sex (male vs. female). As shown in Figure 5, despite the great variability among speakers, there is essentially only *one* principal component of pulse shape involved in the tense vs. lax contrast in the three languages. As indicated by PC1, which accounts for 57% of the variation in pulse shape from all sources, these two phonations mainly differ in the relative shape of their contacting phase: tense phonations show a larger contact quotient (wider pulse), and shorter contacting duration (overall steeper positive slope). Decontacting durations are also somewhat longer for tense phonations, and their contacting/decontacting phases are therefore less symmetric. This asymmetry may be part of strategy for varying the contacting phase without varying the overall pulse length, which has been shown to remain (nearly) constant for a given lexical tone. (One should bear in mind that, as mentioned in the Method section, due to EGG signal artifacts the shapes discussed in this paper should not be regarded as the absolute contact patterns.)

This overall pattern in the pulse shapes is correlated, consistently across speakers and languages, with a variety of EGG parameters which distinguish the phonations, especially CQ, SQ, and EGG contacting duration. Because PC1 by definition reflects the greatest variation in the pulse shapes, it is not surprising that it is more informative than any single parameter-based measure. What is perhaps surprising is that this first component relates specifically to the phonation category, rather than to pitch or any other variable. In at least some languages, e.g. English (Iseli et al., 2007), Mandarin (Kuang, 2013) and Hmong (Esposito, 2012), CQ is highly correlated with f0. Even when pulses are time-normalized so that pulse duration is not varying, pulse shapes will vary strongly as a function of the pitch. (Far from removing the effects of f0 on pulse shape, time-normalization makes such effects visually more salient.) Such a relation might be expected also in the Yi languages. Nonetheless, Yi pulse shapes do not vary much with f0; they vary instead primarily with phonation category.

Crucially, abruptness of contact, a property seen in the FDA component, is not captured by any of the landmark measures. As abruptness of closing is an important aspect of the glottal flow waveform and the resulting acoustics, this aspect of the FDA component is important. Baken and Orlikoff (2000) discuss the presumed close relation between EGG contact-increase measures and the Maximum Flow Declination Rate (MFDR), which correlates with vocal effort and especially SPL intensity. The PIC landmark measure was originally intended to capture differences in rate of change in contact, and thus perhaps indirectly abruptness of contacting, along the lines of MFDR. If this were the case, then PIC should be higher for the tense phonations, with more abrupt contact. However, in our study, as in previous studies, the opposite pattern is found: PIC is reliably lower for the tense phonations. Thus it can be concluded that PIC (at least as measured from typical EGG signals) is apparently not related to abruptness or impact of contacting, and cannot be an EGG analog of MFDR. The FDA component seems better able to reflect this particular aspect of vibration in EGG signals.

B. Comparisons with other languages and phonation types

The first principal component, related most strongly to the tense/lax phonation contrast, is surprisingly similar in shape to Mooshammer (2010)'s PC1, though the difference between our positive and negative curves is larger. Mooshammer compared lexical stress and vocal effort in German, and found similar Contact Quotients and first principal components, with the stress differences smaller than the vocal effort differences. In the lexical stress conditions, acoustic energy was not correlated with PC1. This is like the result here (acoustic energy is weakly correlated with sex). Thus it seems that Yi-language tense phonation is like a bigger version of the shape component for German lexical stress that is related to prominence but not loudness. However, Yi tense phonation is apparently not like emphasis/focus in the Yi languages studied by Michaud (2004) – he found that PIC is higher under emphasis/focus, as long as the pitch is not high, but here, PIC was consistently lower for tense phonation.

A starting point for the present study was our earlier finding that acoustically, tense and lax phonations seem to be fairly similar, and more like modal phonation than like creaky or breathy phonations. Here it is seen that the Yi phonation categories are well-distinguished by their EGG Contact Quotients, but at the same time, these values all fall within a modal-like range: neither the more extreme high values seen in creaky voice in other languages, nor the more extreme low values seen in breathy voice in other languages. Nor are the EGG pulse shapes for tense and lax phonations too far from those of modal phonations in other languages. Lax phonation EGG pulses, though more symmetrical than the tense pulses, are nonetheless not sine-like as they are in breathy voice. Tense phonation pulses, though showing a more abrupt contacting interval than the lax pulses, at the same time have a clear interval of minimum contact, unlike in creaky voice. In addition, tense phonation pulses are periodic, unlike many creaky voice pulses. Thus the EGG results help us understand the acoustic results: tense and lax pulses are different from each other, but not as different from modal as are creaky voice and breathy voice pulses. At the same time, our results indicate that Yi tense phonations share properties with creaky voice, and Yi lax phonations share properties with breathy voice, in other languages. Tense voice is like creaky in having higher CQ (Hmong) and/or lower PIC (Hmong, Zapotec), and lax voice is like breathy in having lower CQ (Hmong, Gujarati) and/or higher PIC (Hmong).

Esposito and Khan (2013) suggested that Gujarati and Zapotec differ in their use of CQ vs. PIC because only Zapotec is tonal. They reasoned that since f0 is free to vary with phonation in Gujarati, the more open portion of the pulse is increased in breathy voice at the expense of elongating the entire pulse. Tone languages, in contrast, make a PIC distinction because f0 cannot vary with phonation. But our results show that tone languages can certainly vary CQ – that CQ can be the parameter most correlated with phonation – while keeping f0 fixed. Maybe this is possible here because these lax phonations are closer to modal phonation than are the breathy phonations of Gujarati and Zapotec.

C. Perception

This study also took advantage of the results available from a perception study in southern Yi to compare the relative strengths of the various kinds of information, EGG and acoustic, about the phonation categories. Listeners' responses were most strongly related to the first principal component from the Functional Data Analysis of the EGG pulses, the component most closely related to the phonation contrast. The next strongest correlation was with the Contact Quotient measure from EGG, which is the best landmark measure for distinguishing the phonation contrast. The strongest correlation with an acoustic measure, H1*, was weaker still. Thus one way in which FDA is valuable is that the shape components it produces do a better job at capturing aspects of the stimuli that are important to listeners.

D. Conclusions

Tense/lax contrasts are a hallmark of many Tibeto-Burman languages. Previously, there was only limited direct evidence that these are glottal contrasts. The present study, with its extensive EGG data, provides strong evidence for this claim. Analysis of data from a perception study also shows that listeners pay attention to information about glottal articulations in perceiving the tense/lax contrast. The differences between tense and lax phonations are neither large nor extreme, but apparently consistent enough, and with robust enough acoustic consequences, to support this linguistic contrast. The present study thus expands our understanding of the possibilities in phonation: that languages can control modest differences in phonation consistently enough to provide a linguistic contrast.

ACKNOWLEDGMENTS

This work was supported by NSF grant BCS-0720304 to the second author. Thanks to Yen-Liang Shue for VoiceSauce and Henry Tehrani for EggWorks. Fieldwork was greatly helped by Prof. Jiangping Kong and Feng Wang of Peking University: special thanks go to Prof. Wang for sharing the 2000-word corpus and to Prof. Kong for sharing the recordings of Hani. We also thank Christina Mooshammer, Mark Liberman, Stephen Isaad, the editor, and the reviewers for comments on earlier versions of this paper.

ENDNOTES

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

2 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 or 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.

3 Diplophonia and other complex periodicities may also be seen in creaky voice, but in linguistic usage, unlike Titze (2000)'s usage, are not criterial for creaky voice.

4 The * indicates that a harmonic amplitude has been corrected for the influence of formant frequencies and bandwidths (Iseli *et al.*, 2007).

5 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.

6 Marasek (1996) did not find this, but this discrepancy might be in part due to their different methods of defining CQ, an issue addressed elsewhere (e.g. Childer *et al.*, 1990; Henrich *et al.*, 2004).

7 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. For example, although SQ of glottal flow has been found to consistently increase in higher vocal effort, SQ from EGG signals has not replicated this result (Dromey et al., 1992).

8 The corpus was a collection of high frequency words across various Tibeto-Burman languages, built by Feng Wang in Peking University for the purpose of historical comparison among related languages.

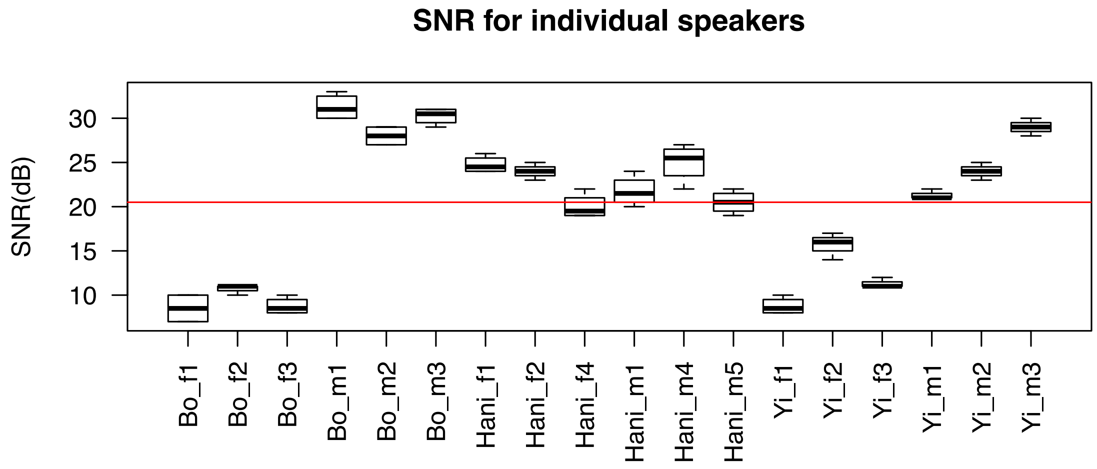
9 The phase distortion of high-pass filtering is discussed in detail in Rothenberg (2002). For the extreme comparison of 2 Hz vs 40 Hz cutoff frequencies, the difference in pulse

shapes is apparent only in the decontacting phase; the contacting phase seems unaffected. This conclusion is reinforced by the finding that Rothenberg's phase compensation algorithm (not applied in the present study) affects only the decontacting phase.

10 There were also no consistent interactions between tone and phonation. The lax mid tone (33L) had a slightly greater CQ than the lax low tone (21L), but reached statistical significance only in Hani.

11 Mooshammer (2010) shows this too. One possible reason for this could be averaging of pulses at each timepoint – if each pulse has its 1.0 peak at a different time, then the average pulse will peak below 1.0. Furthermore, actual pulse shapes are combinations of all of the factor contributions. Possibly the peak normalized amplitude in actual pulse shapes (which is always 1.0) is not due to the same factor across shapes, but instead arises from a variety of contributing factors.

Appendix 1 Signal to Noise Ratio (SNR) for all the speakers used in the pulse analysis. The line in the middle is the median of the SNRs.



Appendix 2 Means and standard deviations of EGG parameter measures of tense and lax phonations in three languages.

	YI		**Bo**		**Hani**	
	Lax	*Tense*	*Lax*	*Tense*	*Lax*	*Tense*
CQ	0.4 (0.08)	0.46 (0.07)	0.4 (0.06)	0.45 (0.06)	0.45 (0.06)	0.51 (0.06)
SQ	0.51 (1.94)	0.43 (0.35)	0.49 (2.93)	0.39 (0.2)	0.52 (1.68)	0.3 (0.17)
Contact_dur	1.31 (2.54)	1.21 (0.69)	1.44 (1.8)	1.14 (0.67)	1.32 (2.11)	1.03 (0.4)
Decontact_dur	1.12 (1.62)	1.31 (1.31)	0.48 (3.24)	0.63 (3.11)	1.62 (1.36)	1.16 (3.15)
PIC	1111.16 (488.94)	1009.937 (452.24)	959.47 (354.46)	864.42 (323.59)	963.53 (345.83)	815.67 (372.38)
|PDC|	512.58 (258.86)	401.31 (177.23)	457.49 (200.80)	354.61 (152.57)	405.61 (171.66)	302.22 (156.37)
PIC_time	373.72 (145.97)	353.92 (138.41)	454.23 (125.65)	440.28 (125.89)	372.20 (90.93)	349.13 (77.03)
PDC_time	375.95 (145.91)	356.64 (138.38)	456.58 (125.56)	442.85 (125.82)	375.10 (90.77)	352.41 (76.83)
F0	166.84 (39.78)	166.72 (39.58)	169.05 (54.89)	174.75 (55.26)	156.82 (42.46)	156.02 (42.18)

Appendix 3 Stepwise Logistic regression models of EGG parameter measures for individual languages: CQ, closing duration, PIC and PDC are the most important predictors.

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	
Decontact_dur	1	1640.6	1644.6	0.63	0.42718	
Contact_dur	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	
Decontact_dur	1	1462.9	1466.9	0.943	0.3315606	

Contact_dur	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**	***
Decontact_dur	1	2238.4	2242.4	17.062	3.62E-05	***
Contact_dur	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

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

Baayen, R. H. (**2010**). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge, UK), pp. 263–327.

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), pp. 394–451.

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.

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.

Ethonologue (**2012**). Ethonologue: http://www.ethnologue.com, accessed Dec. 5th 2012.

Fabre, P. (**1957**). “Un procédé électrique percutané d’inscription de l’accolement glottique au cours de la phonation (A percutaneous electrical method for recording glottal contact during phonation),” Bull. Nat. Méd.**141**, 66-99.

Fourcin, A. J. **(1974)**. “Laryngographic examination of vocal fold vibration,” in *Ventilatory and Phonatory Control Systems*, edited by B.Wyke (Oxford University Press, London), pp. 315-326.

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

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

Gordon, M., and Ladefoged, P. (**2001**). “Phonation types: a cross-linguistic overview,” Journal of 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. (**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., 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.

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

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.

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

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

Kong, J. (**2001**). *On Language Phonation* (in Chinese, The Central University of Nationalities Press, Beijing), pp. 1–353.

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

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

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

Kuang, J. **(2013)**. "Phonation in tonal contrasts," Ph.D. thesis, University of California, Los Angeles.

Ladefoged, P. (**1964**). *A phonetic study of West African languages* (Cambridge University Press, Cambridge), pp.1–74.

Ladefoged, P. (**1971**). *Preliminaries to linguistic phonetics* (University of Chicago Press, Chicago), pp.1–123.

Ladefoged, P., Maddieson, I. (**1996**). *The sounds of the world's languages* (Blackwell, Oxford), pp. 1- 425.

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), pp.1–186.

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," Ph.D. thesis. Institute für Maschinelle Sprachverarbeitung, Stuttgart.

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

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

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), pp.1–326.

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**, 3718–3727.

Ramsay, J. O., and Silverman, B. W. (**1997**). *Functional Data Analysis* (Springer, New York), pp.1–426.

Ramsay, J. O., and Silverman, B. W. (**2002**). *Applied Functional Data Analysis: Methods and Case Studies* (Springer, New York), pp. 1–181.

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

Rothenberg, M. **(1992)**. "A multichannel electroglottograph," Journal of Voice **6**, 36-43.

Rothenberg, M. **(2002)**. "Correcting low-frequency phase distortion in electroglottograph waveforms," Journal of Voice **16**, 32-36.

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

Sundberg, J., and Gauffin, J. (**1979**). "Waveform and spectrum of the glottal voice source," in *Frontiers of speech communication research*, edited by B. Lindblom, S. Ohman (Academic Press, London), pp. 301-322.

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

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.

Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech), pp. 1- 407.

TABLE I. Main effects of phonation in Yi, Bo and Hani on 8 EGG parameter measures. CQ=Contact Duration; SQ=Skew Quotient; Contact_dur = Duration between minimum contact and maximum contact; Decontact_Dur = Duration between maximum contact and minimum contact; PIC = Peak Increase in Contact (positive peak of dEGG); PDC = Peak Decrease in Contact (negative peak of dEGG) values are presented as absolute values, as onlythe amplitude of the peak is of interest.

	Yi	**Bo**	**Hani**
CQ	T > L	T > L	T > L
SQ		33T<33L	T < L
Contact_Dur	T < L	T < L	T < L
Decontact_Dur		T > L	T < L
PIC	T < L	T < L	T < L
|PDC|	T < L	T < L	T < L
PIC_time	T < L		T < L
PDC_time	T < L		T < L
F0			

TABLE II. *F*-values of repeated measures ANOVA for PCs and the factors (phonation, sex, tone and language). Bold indicates p <0.05, and italic means p <0.1.

PC	**Df.**	**Phonation**	**Sex**	**Tone**	**Language**
1	1,70	**20.4**	5.45	0.08	2.57
2	1,70	0.53	**4.85**	0.97	1.26
3	1,70	0.58	**3.15**	0.85	0.78
4	2,69	1.88	0.19	*1.94*	1.87

TABLE III. Correlation coefficients between factor scores from FDA of EGG pulse shapes, and EGG parameters. Only significant coefficients ($p < .05$) are reported here, and highly significant ones ($p < 0.001$) are highlighted in bold.

	PC1	**PC2**	**PC3**	**PC4**
CQ	**0.696**			-0.224
PIC	0.251	-0.140	0.174	
PIC_time			-0.289	0.140
PDC	0.410			0.166
PDC_time			-0.290	0.139
Contact_Dur	-0.282		**-0.369**	
Decontact_Dur		-0.223		**-0.252**
SQ	**-0.472**		-0.310	-0.191
F0	0.24	0.1	-0.17	0.02

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, Luchun Hani, Mandarin, Miao, Mazatec, Yi, Zapotec; the third letter, plus the plotting symbol and the color (online only), indicates the phonation (breathy/ open triangles, creaky/ open circles, lax/ filled triangles, modal/ filled diamonds, tense/ filled circles); Eng indicates English. The circled phonation types are the tense vs. lax. They are similar to each other, and near the modal range.

FIG. 2. (Color online) Illustration of EGG measures from EggWorks. The darker line is the EGG pulses, and the lighter line is the dEGG signal. The left pulse illustrates the hybrid method of estimating CQ (Howard, 1990), and PIC and PDC from dEGG, and the right pulse demonstrates measures related to skewness of the pulse: contacting duration and decontacting duration (SQ=T-cont/T-decont).

FIG. 3. (Color online) EGG pulses before amplitude normalization (left panel) and after amplitude normalization (right panel); in both panels time is normalized into 1000 intervals.

FIG. 4. (Color online) Sample EGG pulses from tense and lax phonations in southern Yi. Left= Lax; Right= tense. Vertical scale for EGG signals is normalized between -1 and 1; note that the dEGG display scales differ between left and right panels. PIC in the left panel ≈ 1200; PIC in the right panel ≈ 400. Darker lines = EGG signal; lighter lines = dEGG signal.

FIG.5. (Color online) The shapes of the EGG signals. Negative factor values are indicated by minus signs, positive factor values by plus signs, and the mean curve by a solid line, for the first four factors.

FIG. 6. Mean scores and standard errors of factors in each principle component for each sex/phonation/tone condition. F=female, M=male, L=lax phonation (light bars), T=tense phonation (dark bars), 21=low tone, 33=mid tone.