CROSS-LINGUISTIC VARIATION IN THE PHONETIC REALIZATION OF "BREATHIER VOICE"

Jia Tian[1], Yipei Zhou[2], Jianjing Kuang[3]

[1] [3]University of Pennsylvania, U.S.A [2]Peking University, China
jiatian@sas.upenn.edu, zhyp15happy@sina.com, kuangj@sas.upenn.edu

ABSTRACT

Breathier voices are non-modal voices produced with a relatively less constricted glottis. Although three major subtypes of breathier voice have been identified, namely slack/lax voice, breathy voice and whispery voice, and distinct IPA symbols have been proposed to transcribe them, actually there is no consensus on how many subtypes should be identified, and how each subtype should be defined. Most studies simply use "breathy voice" as a general term to cover all its subtypes. In this study, we propose that different subtypes of breathier voice can be distinguished by evaluating the relative importance of the glottal constriction and the noise component. Drawing data from Gujarati, White Hmong, Southern Yi, and Shanghainese, we show that our proposed method successfully identified each of the three subtypes found in these four languages. We will discuss the implication for future voice quality studies.

Keywords: phonation, voice quality, breathy, whispery, slack/lax

1. INTRODUCTION

Generally speaking, breathy voice (more accurately, breathier voice) is the non-modal phonation produced with relatively less constricted glottis along the voicing continuum [22, 24]. However, as noted by a number of studies [16, 22–24], the phonetic realization of the 'breathier voice' substantially vary across languages. Particularly, three major subtypes of breathier voice have been recognized, namely slack/lax voice, breathy voice and whispery voice [3, 7, 15, 24, 25], and distinct IPA symbols have been proposed to transcribe them [1]. The breathier voice in languages has been classified into different subcategories. For instance, the breathier voice in Jingpho, Javanese, Yi, Wa, and Mpi were denoted as "slack/lax voice", while the breathier voice in Chong, Gujarati, Hindi, Sindhi, and Jalapa Mazatec were classified as "breathy voice" [16, 24]. Whispery voice is relatively less documented, but the breather voice in Zhenhai, Tamang, and Mon were considered to be this subtype [26, 27, 30].

However, although theoretical studies (e.g., [3, 22, 25]) generally recognize the cross-linguistic variation in the phonetic realization of "breathier voice", there is no consensus on how many subtypes should be identified, and how to define define these different types of breathy voice acoustically. For example, Ladefoged [16, 22, 24] does not distinguish between whispery voice and breathy voice; Catford [3] and Laver [25] do not distinguish between lax voice and whispery voice. One important reason for this challenge is that, as [15, 22, 25] point out, these subtypes essentially form a continuum, with no clear borderline between them. Moreover, there is often considerable individual variation in the realizations of any given type of voice quality [15]. Therefore, in practice, different types of "breathy voice" are generally not distinguished and collapsed in the literature.

However, even though no language appears to contrast different types of "breathier voice" phonologically, cross-linguistic variation has been widely observed. Different articulatory strategies in phonation production lead to different interactions with other phonological structures such as tones, and also develop different paths of sound change. It is therefore meaningful to better understand the variation of "breathier voice." The question then is: since different types of "breathier voice" vary rather continuously, is it possible to quantify the variation with acoustic measures? This study aims to provide some insight into this question.

Generalizing across previous studies [3, 7, 15, 22, 24, 25], the three subtypes can be defined as follows. 1) Compared with modal voice, slack/lax voice is produced mainly with less adductive tension and medial compression in the vocal folds. The arytenoid cartilages are not drawn apart as they are in breathy voice, so the turbulence noise from the glottis is relatively minor. 2) Breathy voice, by contrast, is produced with a considerable glottal aperture, and there is some audible noise. Compared with modal voice, the glottal pulse is more symmetrical for breathy voice. 3) Whispery voice differs from both breathy voice and slack/lax voice in that it is produced with more aperiodic noise by maintaining a higher degree of medial compression (i.e., more constricted than breathy voice, but still less constricted than modal voice). Whispery voice has a more skewed glottal flow pulse compared with breathy voice.

We propose that the three subtypes of "breathier voice" essentially vary in the contribution of two articulatory aspects: how much medial compression in the vocal folds is involved and how large the posterior glottal aperture is. The former is acoustically related to the slope of the voice spectrum, and the latter is acoustically related to the amount of turbulence noise generated in the glottis. Therefore, subtypes of "breathy voice" essentially vary in the relative importance of cues correlated to glottal constriction and noise component.

Therefore, by modeling the relative importance of these two aspects of voice measures, we should be able to tease apart different types of breathier voice. This paper will test this hypothesis by looking into the cross-linguistic variation of four languages that involve some kind of "breathier voice."

2. METHODS

2.1. Languages

In this study, we compare four languages that involve some type of breathier voice: Gujarati (Indo-European, Indo-Iranian; atonal), White Hmong (Hmong-Mien, Hmongic; tonal), Southern Yi (Sino-Tibetan, Tibeto-Burman; tonal), and Shanghainese (Sino-Tibetan, Sinitic; tonal). Typologically, these four languages come from different language families and cover both tonal and non-tonal languages. Importantly, Gujarati and White Hmong are both considered to have a typical "breathy voice" that contrasts with modal voice [9, 10, 16, 18, 19, 24]. Gujarati is atonal, while White Hmong has seven phonemic tones. White Hmong distinguishes three phonation types (modal, breathy, and creaky). Two of White Hmong tones are associated with non-modal phonation. Southern Yi has a three-tone system (low, mid, and high) and two phonation registers (tense vs. lax) [16, 18, 20, 21]. Both tense and lax phonations occur in syllables with both mid and low tones. The high tone, on the other hand, only occurs in syllables with lax phonation. In the literature, the number of languages that have been reported to have a "whispery voice" is small. In this study, we examine Shanghainese, the most well-known Chinese Wu dialect. All Chinese Wu dialects have a upper vs. lower register contrast in which pitch and phonation contrasts co-occur. Breathier voice is associated with lower register tones in Wu [2, 4, 5, 11, 12, 29, 33, 35]. The breathier voice in Zhenhai, a closely related Chinese Wu dialect, has been reported to be "whispery voice [30]."

2.2. Materials

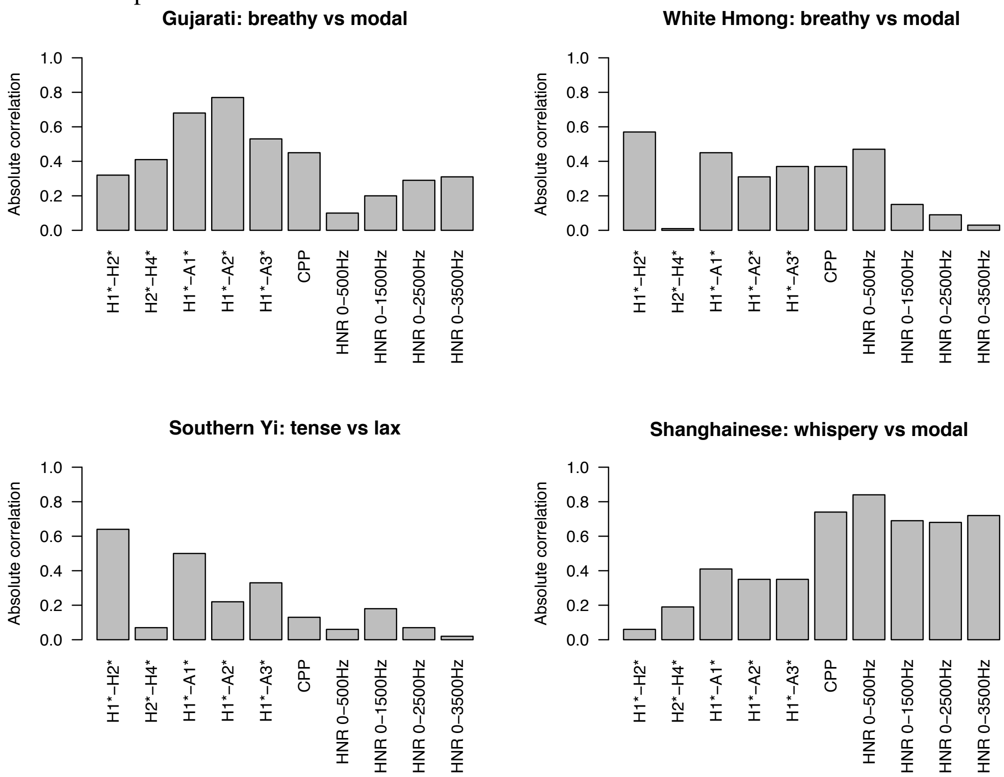
Acoustic measurements of Gujarati, White Hmong, and Southern Yi were retrieved from the "Production and Perception of Linguistic Voice Quality" project at UCLA[1]. Recordings of Shanghainese were made by the first author in Shanghai. Shanghainese is gradually losing its breathier phonation [11, 33, 35], so we only selected speakers born before 1980 in this study, because they were previously shown to maintain the breathier phonation [33]. Since the recordings of all four languages were collected with the same recording setup (e.g., Shure SM10A dynamic microphone and Glottal Enterprises EG2 electroglottograph), and all measurements were obtained using the same tools (i.e., VoiceSauce [31] and EggWorks [32]), the results of the statistical modeling are comparable across the four languages.

2.3. Statistical modeling

In this section, we test the relative importance of spectral tilt and noise measures for each of the four languages. Linear Discriminant Analysis (LDA), a procedure that determines the relative importance of different cues between two or more groups [6], was conducted in R using the lda() function from the MASS package [28, 34]. This method was chosen because it works better than logistic regression when the predictors are highly correlated with each other (like in our case, where all the spectral tilt measures are highly correlated). This method has been found to be effective in evaluating the relative importance of different acoustic cues in various linguistic contrasts (e.g., Mazatec phonation contrast [8, 13]; Korean tonogensis [17]; Tongan stress [14]; Shanghainese phonation contrast [11], to name a few). Spectral tilt measures included were: H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, and H1*-A3*. Noise measures included were: CPP and HNR in four regions of the spectrum (0-500 Hz, 0-1500 Hz, 0-2500 Hz, and 0-3500 Hz). Mean values over the entire vowel were used in the LDA analysis.

The Gujarati model included all breathy and modal syllables in the dataset. A total of 3055 observations (target words embedded in a short sentence that the subject immediately thought of upon seeing the target words) from 10 speakers were included. The White Hmong model included all sylla-

Figure 1: Relative importance of acoustic measures to the phonation contrast in Gujarati (top left), White Hmong (top right), Southern Yi (bottom left), and Shanghainese (bottom right): Linear Discriminant Analysis. Higher bars indicate more importance.



bles with modal falling (52) and breathy falling (42) tones. A total of 195 observations (monosyllables embedded in carrier sentences) from 11 speakers were included. The Southern Yi model included all tense and lax syllables on the low tone and the mid tone, where there are contrastive phonations. A total of 929 tokens (isolated monosyllables) produced by 12 speakers were included. The Shanghainese model included syllables with low rising (23) and high rising (34) tones. A total of 1347 tokens (isolated monosyllables) produced by 52 speakers were included.

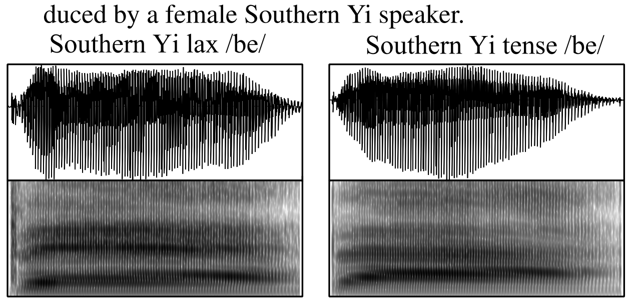
Given that there are two phonation types, LDA produced one discriminant function. The relative importance of each measure was estimated by the Pearson's r correlation between the values generated by the discriminant function and the acoustic measure. A predictor with more importance should show a larger absolute correlation.

3. RESULTS

The LDA results are visually represented in Fig. 1. Measures with higher absolute correlation to the discriminant function (i.e., higher bars in Fig. 1) are of greater importance.

At least three different patterns can be observed in Fig. 1. Compared with the other three languages, noise measures (i.e., CPP and HNRs) make very little contribution to the phonation contrast in Southern Yi, and spectral cues (especially H1*-H2*) play the dominant role. Based on the definition laid out in section 1, the breathier voice in Southern Yi can be categorized as "slack/lax voice." The contrast between tense vs. lax phonation in Southern Yi is mainly realized by the difference of the adductive tension in the vocal folds; since the arytenoid cartilages are not drawn apart, the turbulence noise from the glottis is relatively minor. Indeed, the pair of tense vs. lax syllables shown in Fig. 2 shows that both tense and lax vowels are highly periodic, but they differ in the energy of high frequency components.

Figure 2: Audio signals and spectrograms for /be 21/ (breathy, left) vs /be 21/ (modal, right), produced by a female Southern Yi speaker.



Among the three languages that more heavily rely on noise cues, the cue-weighting pattern for Shanghainese is clearly different from that for Gujarati and White Hmong, suggesting that they should be categorized into different subtypes of breathier voice.

For Gujarati and White Hmong, although both spectral cues and noise cues contribute to the phonation contrast, spectral cues overall play more important roles than noise cues. By contrast, noise cues are the primary cues for the phonation contrast for Shanghainese. Among the three subtypes of breathier voice, whispery voice has the strongest aperiodic noise, so it is more appropriate to categorize the breathier voice in Shanghainese as whispery voice. Since the breathier voices in Gujarati and White Hmong lie between the two more extreme cases, they can be categorized as breathy voice.

Figure 3: Audio signals and spectrograms for /ba/ (breathy, left) vs /ba/ (modal, right), produced by a female Gujarati speaker.

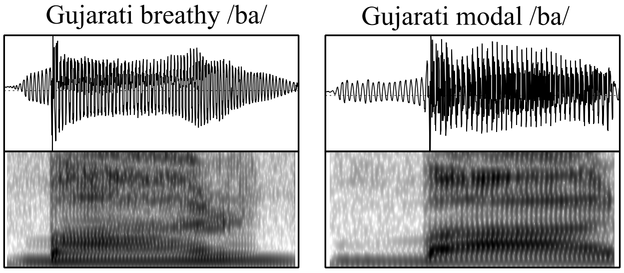


Figure 4: Audio signals and spectrograms for /po 42/ (breathy, left) vs /po 52/ (modal, right), produced by a female White Hmong speaker.

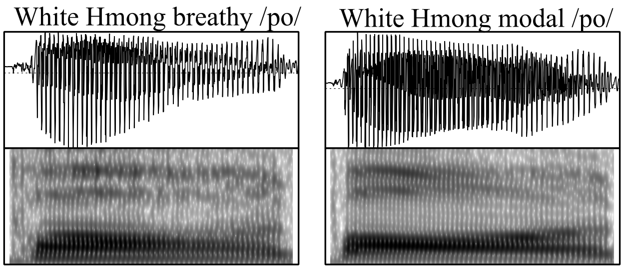
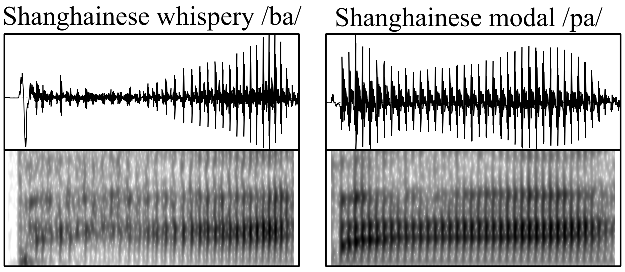


Figure 5: Audio signals and spectrograms for /ba 23/ (whispery, left) vs /pa 34/ (modal, right), produced by a female Shanghainese speaker.



Again, the LDA results can be validated by the acoustic signals. As can be observed in Fig. 3 and Fig. 4, breathy vowels in White Hmong and Gujarati have weakened formant structure and increased aperiodic noise. But the aperiodic noise is not as strong as that in Shanghainese (Fig. 5). The case of Shanghainese is particularly interesting, since whispery voice is especially under-documented among languages. Ladefoged and Maddieson [24] previously suggested that the breathier voice in Shanghainese is slack voice, but our LDA results clearly suggest that Shanghainese and Southern Yi do not pattern together. It should be noted that Rose [30] had reported that the breathier voice in Zhenhai, a closely related Northern Wu dialect, is also produced with a substantial amount of aperiodic noise and relatively constricted glottis. It is likely that this voice quality is a common property shared by some Wu dialects.

4. CONCLUSION

In this study, we demonstrated that there is cross-linguistic variation in the phonetic realization of the so-called “breathier voice”, and that examining the relative importance of aperiodic noise and glottal constriction is an effective way to tease apart the different subtypes of “breathier voice.” In particular, our LDA models distinguished three cue-weighting patterns that are consistent with the subtypes of “breathier voice” proposed in the literature [3, 15, 24, 25], and therefore the three subtypes can be defined more clearly with acoustic cues: slack/lax voice is produced with spectral cues as the dominant cues but little noise cues; breathy voice is produced with dominant spectral cues but noise cues also play important roles; whispery voice is produced with noise cues as the dominant acoustic cues.

This study has important implications for future studies on “breathier” type of voice: first of all, to achieve a comprehensive understanding of the “breathier voice” in languages, it is important to investigate both the degree of glottal constriction and the presence of aperiodic noise. Moreover, although our findings generally support the subcategorization of “breathier voice”, it is important to bear in mind that these subtypes of “breathier voice” essentially form a continuum in the phonetic space, a space defined by both glottal constriction and noise component. We do not intend to claim that there are clear cut-offs among these subtypes. Therefore, the approach proposed in this paper has the advantage of successfully capturing the cross-linguistic variation of “breathier voice” without over-assuming the categoricity of the subtypes. Finally, the cue weighting patterns found in this study should be further validated through cross-linguistic perception experiments.

5. REFERENCES

[1] Ball, M. J., Esling, J. H., Dickson, B. C. 2018. Revisions to the VoQS system for the transcription of voice quality. *Journal of the International Phonetic Association* 48(2), 165–171.

[2] Cao, J., Maddieson, I. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20, 77–92.

[3] Catford, J. 1977. *Fundamental Problems in Phonetics*. Bloomington: Indiana University Press.

[4] Chao, Y. R. 1928. *Studies in the Modern Wu-Dialects*. Peking: Tsing Hua College Research Institute.

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

[6] Duda, R. O., Hart, P. E., Stork, D. G. 2012. *Pattern classification*. John Wiley & Sons second edition.

[7] Esling, J. H., Harris, J. G. 2005. States of the glotis: An articulatory phonetic model based on laryngoscopic observations. In: *A Figure of Speech: A Festschrift for John Laver*. 347–383.

[8] Esposito, C. M. 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38, 306–316.

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

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

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

[12] Gao, J., Hallé, P. 2017. Phonetic and phonological properties of tones in Shanghai Chinese. *Cahiers de Linguistique Asie Orientale* 46, 1–31.

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

[14] Garellek, M., White, J. 2015. Phonetics of Tongan stress. *Journal of the International Phonetic Association* 1–25.

[15] Gobl, C., Ní Chasaide, A. 2012. Voice source variation and its communicative functions. In: Hardcastle, W. J., Laver, J., Gibbon, F. E., (eds), *The Handbook of Phonetic Sciences*. Wiley-Blackwell 378–423.

[16] Gordon, M., Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29, 383–406.

[17] Kang, Y., Han, S. 2013. Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua* 134, 62–74.

[18] Keating, P., Esposito, C. M., Garellek, M., Khan, S. u. D., Kuang, J. 2010. Phonation Contrasts Across Languages. *UCLA Working Papers in Phonetics* 108, 188–202.

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

[20] Kuang, J. 2011. *Production and Perception of the Phonation Contrast in Yi*. Master's thesis UCLA.

[21] Kuang, J., Keating, P. 2014. Vocal fold vibratory patterns in tense versus lax phonation contrasts. *Journal of the Acoustical Society of America* 136(5), 2784–2797.

[22] Ladefoged, P. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

[23] Ladefoged, P., Disner, S. F. 2012. *Vowels and Consonants*. Wiley-Blackwell third edition.

[24] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

[25] Laver, J. 1980. *The phonetic description of voice quality*. Cambridge University Press.

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

[27] Michaud, A. 2012. Monosyllabicization : patterns of evolution in Asian languages. In: Nau, N., Stolz, T., Stroh, C., (eds), *Monosyllables: from phonology to typology*. Berlin: Akademie Verlag 115–130.

[28] R Core Team, 2018. R: A language and environment for statistical computing.

[29] Ren, N. 1992. *Phonation types and stop consonant distinctions: Shanghai Chinese*. Ph.d. dissertation The University of Connecticut.

[30] Rose, P. 1989. Phonetics and phonology of the yang tone phonation types in Zhenhai. *Cahiers de Linguistique Asie Orientale* 18(2), 229–245.

[31] Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences* Hong Kong. 1846–1849.

[32] Tehrani, H. 2010. EGGWorks.

[33] Tian, J., Kuang, J. 2016. Revisiting the Register Contrast in Shanghai Chinese. *Tonal Aspects of Languages 2016* 147–151.

[34] Venables, W. N., Ripley, B. D. 2002. *Modern Applied Statistical with S*. New York: fourth edition.

[35] Zhang, J., Yan, H. 2018. Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. *Journal of the Acoustical Society of America* 144(3), 1293–1308.

[1] http://www.phonetics.ucla.edu/voiceproject/voice.html