PHONATION CONTRASTS ACROSS LANGUAGES

Patricia Keating[1], Christina Esposito[2], Marc Garellek[1], Sameer ud Dowla Khan[3], Jianjing Kuang[1]

[1]Dept. Linguistics, UCLA, Los Angeles CA USA 90095; [2]Dept. Linguistics, Macalester College, St. Paul MN; [3]Dept. Cognitive Linguistic & Psychological Sciences, Brown U., Providence RI
keating@humnet.ucla.edu; esposito@macalester.edu; marcgarellek@ucla.edu; sameerudowlakhan@gmail.com; kuangjianjing@ucla.edu

ABSTRACT

This study compares the phonetics of phonation categories within and across four languages: Gujarati (modal, breathy), White Hmong (modal, breathy, creaky), Jalapa Mazatec (modal, breathy, creaky), and Southern Yi (tense, lax). In addition to acoustic measures in all four languages, electroglottographic measures were also compared for Gujarati, Hmong, and Yi. Several measures distinguished phonation categories within each language, although only H1*-H2* and CQ did so in all languages measured. When within-language phonation categories were then compared across languages, they were found to differ across languages on multiple acoustic measures, e.g. Hmong breathy voice is distinct from Gujarati breathy voice. This unexpected result suggests that language/speaker/recording differences are larger than phonation category differences, a claim that finds support in a Multidimensional Scaling analysis of the acoustics. A three-dimensional space turns out to mostly distinguish languages rather than phonations; the phonation categories do not form clusters in this space across languages, but they do occupy separate regions along the dimension of the space correlated with H1*-H2*. Thus, H1*-H2* is seen to be the most important measure of phonation contrasts across languages.

Keywords: phonation, electroglottography, cross-linguistic, acoustic space

1. INTRODUCTION

Across languages with phonation contrasts, the phonation categories are distinguished by a variety of measures [4, 5] but not by every measure in every language. We ask the following questions:

- Which measures distinguish phonation categories within and across languages?
- What are the dimensions of the cross-language acoustic voice quality space?
- How are the categories of different languages arranged in this space?

2. METHODS

Basic information about the four languages we studied is in Table 1. For each language, a wordlist contrasting phonations was compiled. Words with low (Gujarati (G), Mazatec (Mz), Hmong (H)) or low and low-mid (Yi) vowels were selected, with a variety of onset consonant types. In Mz, only level tones were included; in Yi, only non-high tones were included. Words were uttered in isolation for Yi and Mz and in sentences for H and G. All but the G data are from field recordings; Mz is taken from the online UCLA Phonetic Archive.

Table 1: Languages studied, their contrastive phonations, use of tone, and number of speakers recorded. M=modal, B=breathy, C=creaky, L=lax, T=tense, * = has lexical tone. f=female, m=male.

Language	Phona-tion	# speakers (acoustic)	# speakers (EGG)
Gujarati	M, B	10: 7f, 3m	10: 7f, 3m
White **H**mong*	M, B, C	32: 9f, 23m	11: 5f, 6m
Jalapa **Maz**atec*	M, B, C	16: 6f, 10m	none
Southern **Yi***	L, T	12: 6f, 6m	12: 6f, 6m

2.1. Acoustic measures

Seven acoustic measures were made semi-automatically over the entire vowel duration with VoiceSauce [16]: **H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, Energy,** and **Cepstral Peak Prominence (CPP)**. Asterisks indicate that the harmonic amplitudes were corrected for the effects of formants using [11], an extension of [6]. See [2, 5, 7, 12] for descriptions of these measures. Formant values were corrected by hand as needed.

2.2. EGG measures

Two EGG measures were made automatically over the entire vowel using EggWorks [17] for Yi, G, and H: contact quotient using the hybrid method with a 25% threshold (CQ_H) and peak increase in contact (PIC). ("Hybrid": the edges of the glottal cycle's contacting phase are defined using two different methods; see [1, 8, 10, 14, 15]). PIC is the peak positive value in the derivative of the EGG signal, equivalent to DECPA [13].

3. RESULTS

3.1. Analyses of individual languages

Statistical comparisons were made to determine which measures distinguished phonation categories (Table 2). Within-language comparisons were based on means over entire vowels in G, H, and Yi, but over the first third in Mz (where contrasts are strongest). Repeated measures ANOVAs were used in G and H, and linear mixed effects models in Mz (with speaker and item as random effects) and Yi (with speaker as random intercept and tone and phonation as random slopes).

Table 2: Results of within-language tests of significance of contrasts on 8 measures. A checkmark indicates that the measure significantly ($p<.05$) distinguished categories in the expected direction.

Measure	Guajrati	Hmong	Mazatec	Yi
H1*–H2*	✓	✓	✓	✓
H2*–H4*				
H1*–A1*	✓		✓	✓
H1*–A3*	✓		✓	✓
CPP		✓	✓	✓
Energy			✓	
CQ_H	✓	✓	N/A	✓
PIC		✓	N/A	✓

H1*-An* measures do not distinguish Hmong phonations; CPP does not distinguish Gujarati phonations. Energy distinguishes phonations only in Mazatec. H2*-H4* does not distinguish any categories. Only H1*-H2* distinguishes phonations in all four languages, with breathy/lax voice exhibiting the highest values (Fig. 1).

CQ_H distinguished all phonations in all three languages except for the creaky vs. modal distinction in Hmong. PIC did not distinguish phonations in Gujarati. Means for CQ_H and PIC are shown in Figs. 2 and 3, respectively. CQ_H is lower for breathy and lax phonations, and higher for creaky and tense phonations. The results for PIC are contrary to expectations. It might be thought that breathier phonation, typically having a more gradual vocal fold closing, would have lower PIC values (if PIC reflects, even indirectly, the speed of closing of the vocal folds). However, the breathy and lax phonations have higher, not lower, PIC values (Fig. 3). Visual inspection of EGG signals suggests that PIC might follow a principle of "the further, the faster", i.e. the longer the open phase of the glottal cycle, the faster the transition to the closed phase.

Figure 1: H1*-H2* for phonations, grouped by tone. Mazatec "laryngealized" is called "creaky" elsewhere. Hmong tones are grouped into three basic levels. Yi tense phonation and high tone do not co-occur. Values come from means over vowels except in Mazatec.

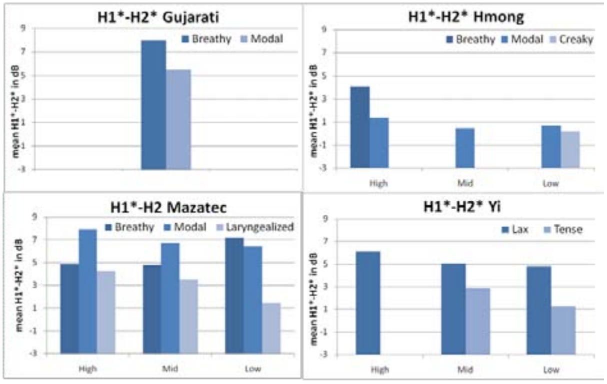


Figure 2: Contact Quotient for phonations, grouped by tone. Hmong tones are grouped into three basic levels. Yi tense phonation and high tone do not co-occur. Values come from means over entire vowels.

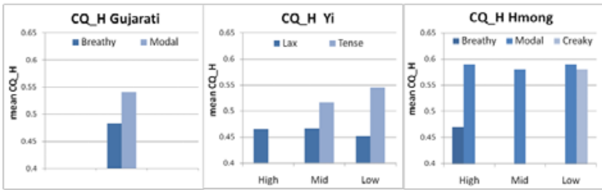
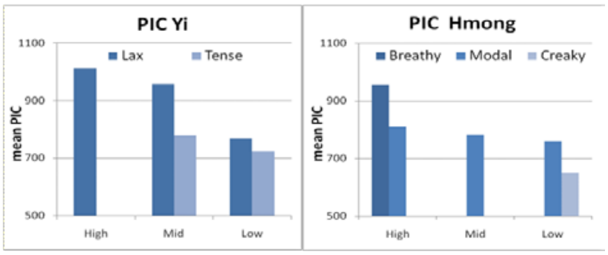


Figure 3: PIC for phonations in the two languages in which significant differences were found, grouped by tone. Hmong tones are grouped into three basic levels. Yi tense phonation does not co-occur with high tone. Values come from means over entire vowels.



All three languages for which we have EGG data distinguish their phonations by both CQ_H and H1*-H2*. (This is true in Hmong for creaky vs. modal phonations when just the end of the vowel is considered.) It is not surprising that these two measures should pattern together, given established literature on their relation [3, 9]. Second, the PIC measure distinguished the phonations in Hmong and Yi, the languages with contrasting creaky/tense phonation. Although speed of vocal fold closure is thought to be related to spectral tilt, there is no obvious relation in Table 2 between PIC and spectral tilt measures (H1*-An*). However, these are the languages in which CPP also distinguished the phonations. The connection between PIC and CPP should be explored further.

3.2. Comparison of all language categories

Next, we examined the ten language-specific phonation categories: 2 Gujarati + 3 Hmong + 3 Mazatec + 2 Yi = 10 total. How many of them are acoustically distinct? At the most conservative extreme, all ten could be different, or they could cluster into as few as three distinct groups, roughly: breathy/lax, modal, and creaky/tense.

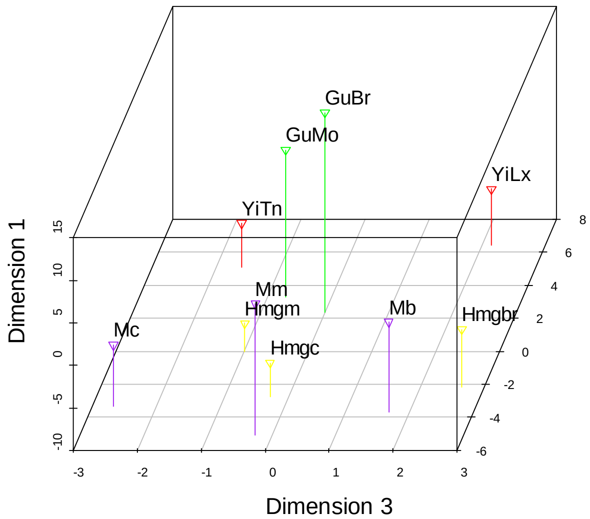
The tokens included in this analysis include aspirated onsets in G and Mz. These comparisons were made with Linear Mixed Effects models, one for each of the acoustic measures, with speaker and item as random effects. Results show that modal, breathy, and lax phonations all differ on a large set of measures. The result is the unexpected extreme: phonation categories with the same descriptive names (e.g. G breathy, H breathy, and Mz breathy) in fact differ significantly along several acoustic dimensions. This suggests that speaker/ language/recording differences are larger than phonation differences. This possibility is examined in another way in the next analysis.

3.3. Cross-language multidimensional scaling

Multidimensional Scaling (MDS) uses measured distances between items to define a map in which those distances are arranged in space. Here, tokens were further controlled for vowel height, tone, and onset aspiration. Acoustic measurements were used as the basis for estimates of the physical distances between all pairs of tokens. A solution is shown in Fig. 4, in which each of the ten within-language phonation categories is plotted in a 3-D space.

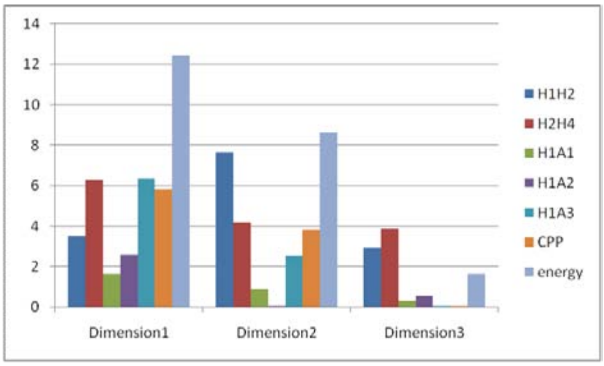
If languages dispersed their phonations within an overall space, we would expect to see the same-colored (within-language) bars spread well apart in the figure. Instead, they cluster together on one or more of the dimensions. That is, the cross-language differences appear greater than the cross-phonation differences. Dim1 (vertical) distinguishes G, and to some extent H, from the other languages. Within G, H, and Yi, the breathy/lax phonation is higher on this dimension; but in Mz the modal phonation is highest. Dim2 (front-to-back) separates the languages into two groups: G and Yi with positive values, H and Mz with negative ones. These two dimensions serve mainly to put each language into its own region of the space, possibly due to differences in voice quality settings, recording conditions, etc.

Figure 4: 3-D MDS solution for the ten categories with Dimension 3 in the foreground. Each color represents a separate language. Heights of bars reflect values on Dimension 1.



Dim3 (horizontal) distinguishes the phonation categories within languages. Mz and Yi phonations are separated along this dimension. H modal and creaky are not so well separated; this is expected since these data are averaged across entire vowels while H creakiness is localized at the ends of vowels. The G phonations are not well separated on any one dimension, but Dim1 and Dim3 together may be crucial. Overall, there is no apparent tendency for all instances of any one kind of phonation to cluster tightly together in any part of the space. However, some phonation-specific patterns are seen: the creaky/tense phonations have negative values on Dim3, the modal phonations are between -1 and 0, and the breathy/lax phonations are ≥ 0. Dim3 seems to provide a continuum of phonation types as suggested by [5]. To better understand the basis for these distinctions, the weighting of each acoustic measure on each dimension is shown in Fig. 5. Dim1 is related to energy, H2*-H4*, H1*-A3*, and CPP; this dimension separates the languages, but it also contributes to distinguishing breathy/lax phonation from other phonations on the basis of spectral tilt and noise. Dim2 is related to energy and H1*-H2*. The strong relations of these first two dimensions to energy (which was largely unsuccessful at contrasting within-language phonations) underscores that these dimensions are mainly characterizing language/speaker/recording differences, not phonation differences. Dim3, which does the most to distinguish the phonations, is related to H2*-H4* and H1*-H2*. The importance here of H2*-H4* is surprising, as it does not distinguish the phonations within languages. It must be further contributing to the language distinctions on this dimension. H1*-H2*, the most important measure for distinguishing the phonations, also contributes to the language differences seen on Dim2. H1*-A1* and H1*-A2* are not strongly related to any of the MDS dimensions.

Figure 5: Weights of acoustic measures on each dimension of the 3-D MDS solution.



4. CONCLUSION

What measures best distinguish phonation categories within and across languages? We found that several acoustic measures differentiate the categories within each language, but only H1*-H2* does so in all four languages. One EGG measure, Contact Quotient, also does so in all three languages with EGG data. Comparing each category across languages, differences were found on several measures. These consistent cross-linguistic differences suggest that language/ speaker/recording differences in voice quality are larger than phonation category differences. This is supported in the MDS analysis of the acoustic measures. A 3-D space mostly distinguishes the languages rather than the phonations. The ten language/phonation categories do not form clusters across languages as might have been expected, but instead occupy separate regions along the third dimension of the space, a dimension correlated with H1*-H2* and H2*-H4*. While H1*-H2* distinguished the phonations in the within-language statistical comparisons, H2*-H4* did not, and therefore it is unlikely that this measure is contributing to distinguishing the phonations in this space. More likely, it differs across the languages, speakers, and/or recordings. On this interpretation, H1*-H2* is again seen to be the most important measure of phonation contrasts across languages.

5. ACKNOWLEDGMENTS

This work was supported by NSF grant BCS-0720304. We thank Y.-L. Shue for VoiceSauce, H. Tehrani for EggWorks, and S. Yang, J. Ptacek, V. Thatte and the Hmong American Partnership for assistance.

6. REFERENCES

[1] Baken, R.J., Orlikoff, R.F. 2000. *Clinical Measurement of Speech and Voice.* San Diego: Singular, 185-187.
[2] Blankenship, B. 2002. The timing of nonmodal phonation in vowels. *JPhon 30*, 163-191.
[3] DiCanio, C. 2009. The phonetics of register in Takhian Thong Chong. *JIPA 39*, 162-188.
[4] Esposito, C.M. 2010. The effects of linguistic experience on the perception of phonation. *JPhon 38*, 303-316.
[5] Gordon, M. & Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *JPhon 29*, 383-406.
[6] Hanson, H.M. 1995. *Glottal characteristics of female speakers*. Ph.D. dissertation, Harvard.
[7] Hanson, H., Stevens, K., Kuo, H.-K., Chen, M., Slifka, J. 2001. Towards models of phonation. *JPhon 29*, 451-480.
[8] Herbst, C., Ternström, S. 2006. A comparison of different methods to measure the EGG contact quotient. *Logopedics Phoniatrics Vocology 31*, 126-138.
[9] Holmberg, E.B., Hillman, R.E., Perkell, J.S., Guiod, P.C., Goldman, S.L. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *JSHR 38*, 1212-23.
[10] Howard, D.M. 1995. Variation of electro-laryngographically derived closed quotient for trained and untrained adult female singers. *JVoice 9*, 163-172.
[11] Iseli, M., Shue, Y.-L., Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *JASA 121*, 2283-2295.
[12] Kreiman, J., Gerratt, B., Antoñanzas-Barroso, N. 2007. Measures of the glottal source spectrum. *JSLHR 50*, 595-610.
[13] Michaud, A. 2004. A Measurement from Electroglottography: DECPA, and its Application in Prosody. Speech Prosody 2004, Nara, 633-636.
[14] Orlikoff, R.F. 1991. Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects. *JSHR 34*, 1066-1072.
[15] Rothenberg, M., Mahshie, J. 1988. Monitoring vocal fold abduction through vocal fold contact area. *JSHR 31*, 338-351.
[16] Shue, Y.-L., Keating, P. Vicenik, C. 2009. VoiceSauce: A program for voice analysis. *JASA 126*, 2221 (abstract).
[17] Tehrani, H. 2009. EggWorks: a program for automated analysis of EGG signals. Available online: *http://www.linguistics.ucla.edu/faciliti/facilities/physiology/EggWorksSetup.exe*.