Multi-dimensional phonetic space for phonation contrasts

P. Keating[1], J. Kuang[1] , C. Esposito[2], M. Garellek[1], S. Khan[3]

[1]UCLA; [2]Macalester College; [3]Reed College

Introduction

Many languages have phonation contrasts, but the multi-dimensional phonetic space for voice quality is not yet well understood. Here we ask:

• What is a **low-dimension space** (acoustic, physiological) for voice quality?

• How are the phonation categories of different languages **located** in this space?

Language Samples

We compare contrastive and allophonic phonations of 10 languages, 8 with EGG as well as audio recordings:

• **Bo** (Tibeto-Burman)
Tonal; **tense vs. lax** (largely independent of pitch)
12 speakers in Yunnan, China (isolated words)

• **California English** (Indo-European) [NO EGG AVAILABLE]
Non-tonal; treated as all **modal**
22 speakers in Los Angeles, USA (isolated words)

• **Gujarati** (Indo-European)
Non-tonal; **modal vs. breathy**
10 speakers in Los Angeles, USA (sentence-initial words)

• **Luchun Hani** (Tibeto-Burman)
Tonal; **tense vs. lax** (largely independent of pitch)
10 speakers in Yunnan, China (isolated words)

• **White Hmong** (Hmong-Mien)
Tonal; **modal vs. breathy** on H-falling tones; **creaky** low tone
32 speakers in St. Paul, USA (isolated words)

• **Beijing Mandarin** (Sino-Tibetan)
Tonal; Tone 3 coded as **creaky** and other tones as **modal**
20 speakers in Beijing, China (disyllables)

• **Jalapa Mazatec** (Oto-Manguean) [NO EGG AVAILABLE]
Tonal; **modal vs. breathy vs. creaky** (independent of pitch)
16 speakers in Jalapa de Díaz, Mexico (isolated words in online archive)

• **Black Miao** (Hmong-Mien)
Tonal; **modal vs. breathy** mid tones, **creaky** low tone , **pressed** high tone
15 speakers in Guizhou, China (isolated words)

• **Southern Yi** (Tibeto-Burman)
Tonal; **tense vs. lax** (largely independent of pitch)
12 speakers in Yunnan, China (isolated words)

• **Santiago Matatlán and San Juan Guelavia Valley Zapotec**
(Oto-Manguean) *(Two varieties grouped together here)*
Tonal; **modal** H tone, **creaky** H-falling tone and **breathy** L-falling tone
6 speakers in Los Angeles, USA (isolated words)

References and Acknowledgments

•Hanson, H. M. (1995) *Glottal characteristics of female speakers*, Ph.D. Dissertation, Harvard.
• Iseli, M., Y.-L. Shue & A. Alwan (2007) Age, sex, and vowel dependencies of acoustic measures related to the voice source, *JASA 121*, 2283–2295.
• Kawahara, H., H. Katayose, A. de Cheveigné & R. D. Patterson (1999) Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity, EUROSPEECH '99, 2781–2784.
•Michaud, A. (2004) A Measurement from Electroglottography: DECPA, and its Application in Prosody, Speech Prosody, Nara, 633–636.
•Shue, Y.-L., P. Keating , C. Vicenik, K. Yu (2011) VoiceSauce: A program for voice analysis, Proceedings of the ICPhS XVII, 1846-1849.

Thanks to NSF grant BCS-0720304, Yen Shue for VoiceSauce, and Henry Tehrani for EggWorks.

Acoustic Measures

Acoustic measures over time were made semi-automatically from the audio signals by **VoiceSauce** (Shue et al. 2011), a free UCLA program.

Spectral measures analyzed:

•**F0** by the STRAIGHT algorithm (Kawahara et al. 1999) for finding harmonics

•**Harmonic amplitudes and differences** (* indicates Hanson 1995, Iseli et al. 2007 corrections) **:**

- **•H1*, H2*, H4*, A1*, A2*, A3***
- **•H1*-H2*, H2*-H4***
- **•H1*-A1*, H1*-A2*, H1*-A3***

•**Noise measures**

- **•Cepstral Peak Prominence**
- **•Harmonic-Noise ratios**

To minimize differences across speakers and recording conditions, all measures are converted to standardized scores by speaker.

EGG Measures

Electroglottographic signals were recorded with the audio for 8/10 languages. Automated EGG measures were made by **EggWorks**, a free UCLA program.

EGG measures analyzed:

•**CQ_H**: **Contact Quotient**, here using the "hybrid" method with 3/7 threshold

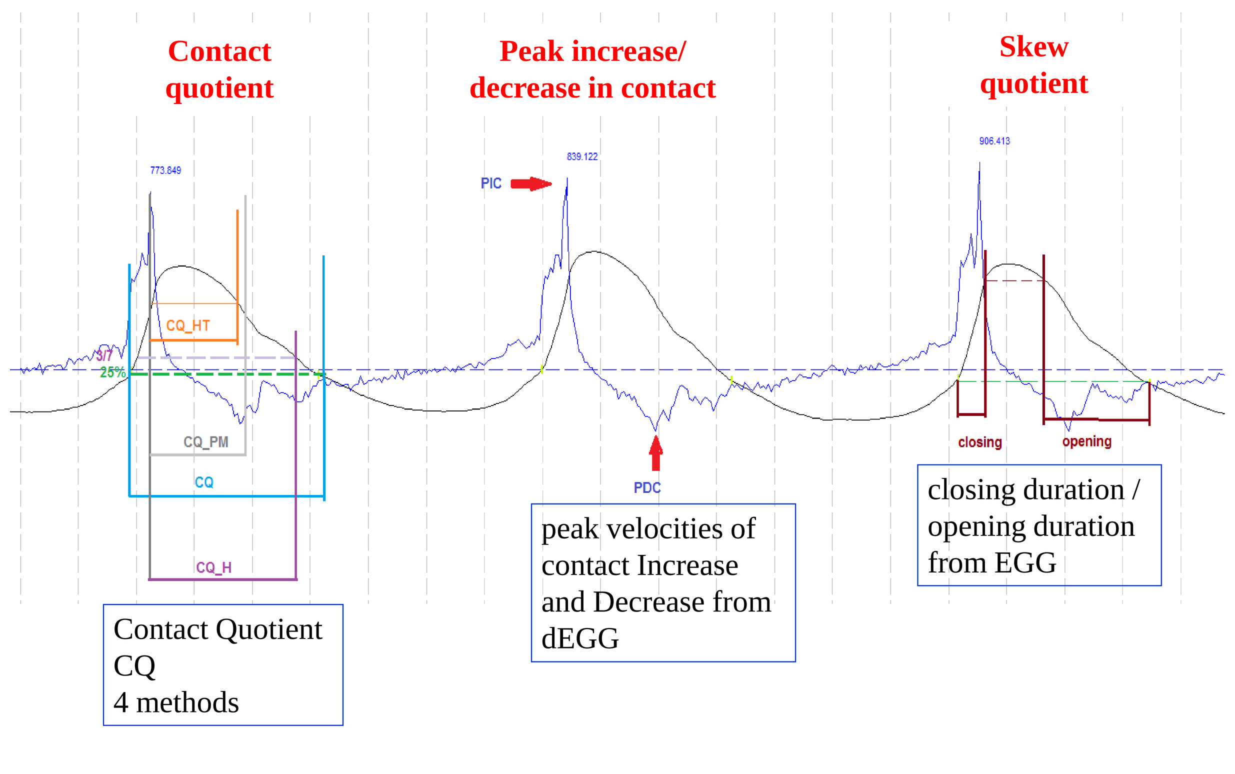
•**PIC: Peak Increase in Contact** (the peak positive value in the EGG derivative, like DECPA (Michaud 2004))

•**PDC: Peak Decrease in Contact** (the peak negative value in the EGG derivative)

•**OP_DUR: Opening duration** (not included)

•**CL_DUR: Closing duration**

•**SQ: Skew quotient** (ratio of CL_DUR/OP_DUR)



Categories within languages

Within-language logistic regressions were used to find the acoustic measures that best predict each pairwise contrast. In every language, one or more **energy or noise** measure, and one or more **harmonic** measure, work well, but exactly which measure(s) of each type varies across languages.

Categories across languages

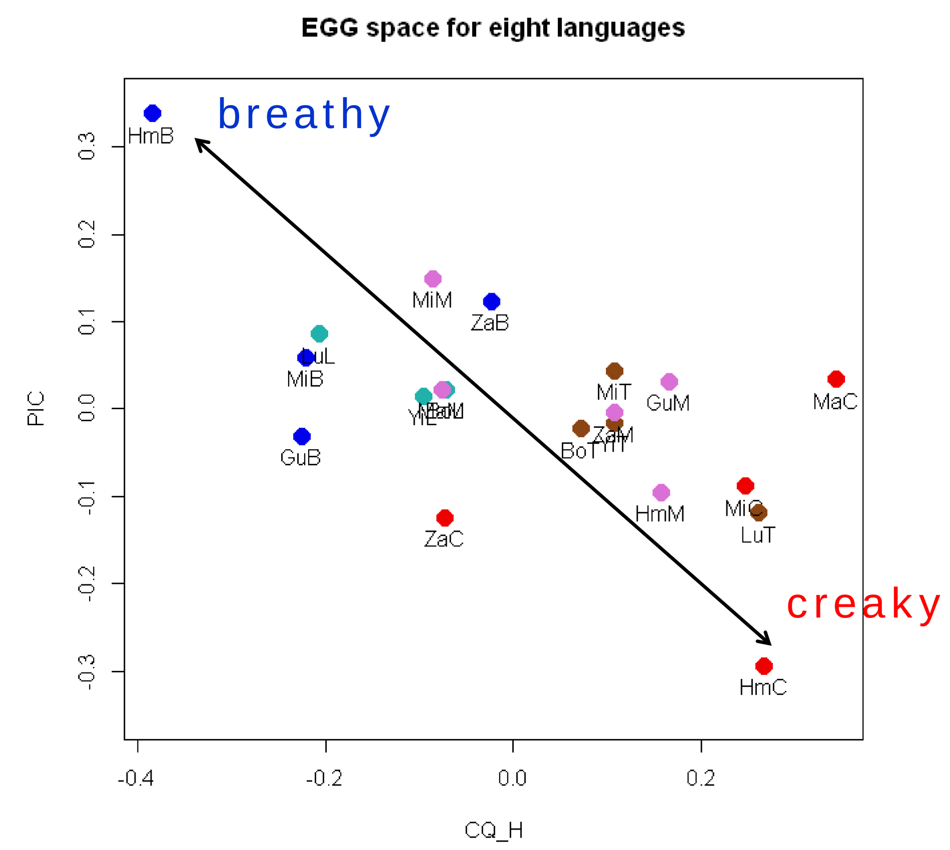
Across-language Linear Mixed Effects models (with several random factors) are being used to compare all possible pairings of the 24 individual-language phonation categories on all the acoustic measures: quantitative tests of differences like those seen in the second plot below. On H1*-H2*, for example, **Breathy** phonations group together, while **Creaky and Tense** phonations group together, and **Modal** phonations vary from **Lax**-like to **Creaky/Tense**-like.

Low-dimension phonetic spaces

For just the non-high, oral, vowels after unaspirated consonants, measures have been standardized by speaker; colors = 5 phonation category labels:

Best measures of **EGG** signals (for 20 phonation categories in 8 languages) are **CQ_H** and **PIC** from timepoint2. These are plotted together.

CQ_H gives a very rough continuum of phonation categories (except for Zapotec), as do the 2 dimensions together, on the diagonal.

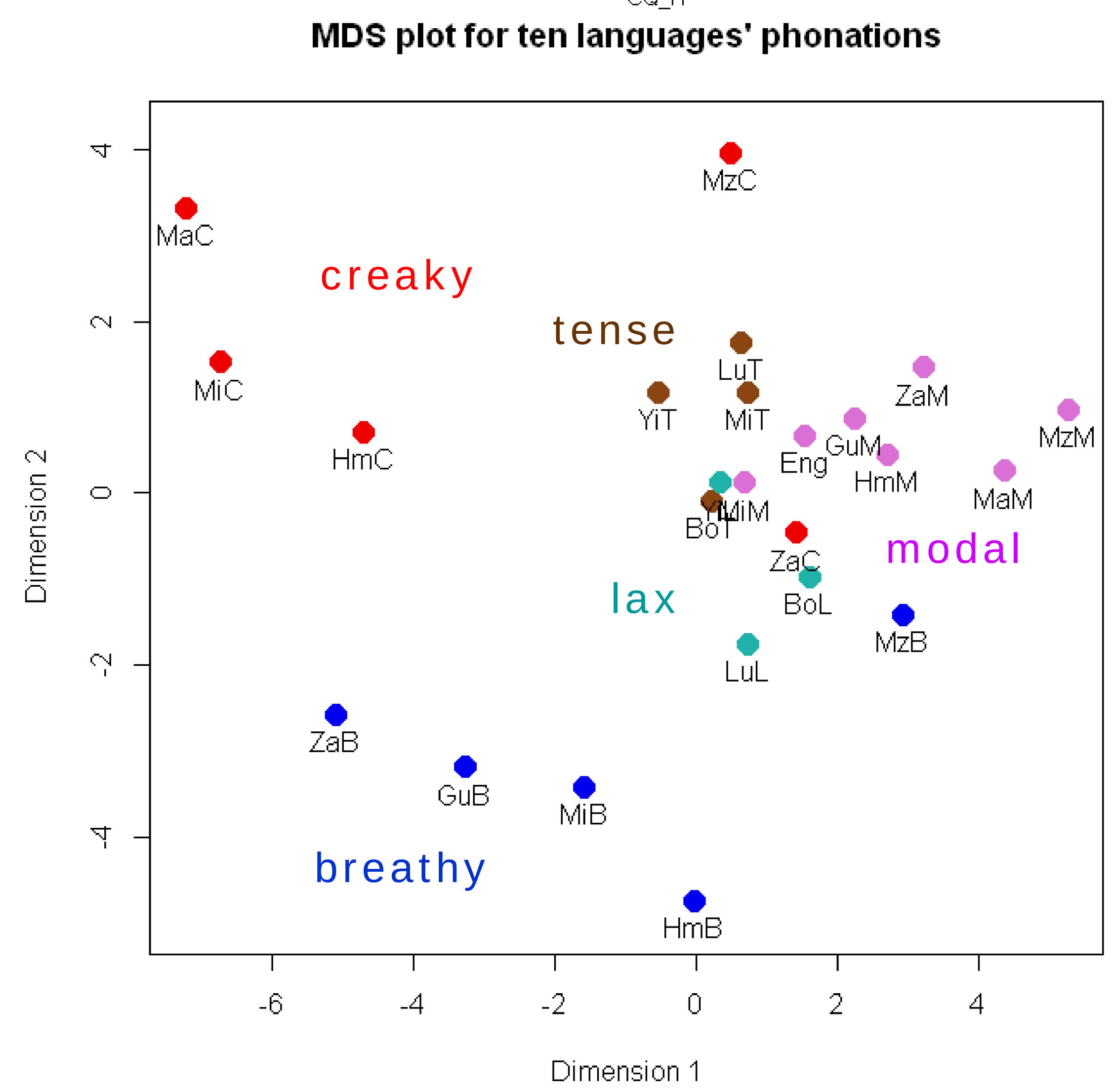


Multi-Dimensional Scaling of **acoustic** measures (from middle-vowel for 24 phonation categories in 10 languages, standardized by measure as well as by speaker). MDS uses differences between items on these measures to define a lower-dimension space of distances between items.

• **Dimension1 (X-axis): H1*-A1*, A3*, H2***
• **Dimension2 (Y-axis): H1*, H1*-H2*, H1*-A1***

Surprisingly, noise measures are not important here.

Dimension1 goes from least to most modal.
Dimension2 is like a glottal constriction continuum.
Lax, **Modal,** and **Tense** are all similar but form sub-clusters. Mazatec's non-modal phonations, which occur on all tones, lie apart from other languages'. Zapotec's **Creaky** is an outlier.



Conclusions

Low-dimension phonetic spaces for phonation can be derived from standardized acoustic and physiological measures of phonation. Phonation categories are somewhat grouped in these spaces, arranged from Breathy to Lax to Modal to Tense to Creaky, but the EGG space shows more overlap.

• The **EGG** space is structured by **Contact Quotient** and **Peak Increase in Contact**

• The **acoustic** MDS space is structured by a dimension of **non-modal to modal** (reflecting mid-frequency amplitudes), and a dimension like **glottal constriction** (reflecting low-frequency amplitudes).