

A CROSS-LANGUAGE ACOUSTIC SPACE FOR PHONATION DISTINCTIONS

Patricia Keating
UC Los Angeles
keating@humnet.ucla.edu

Jianjing Kuang
University of Pennsylvania
kuangj@ling.upenn.edu

Marc Garellek
UC San Diego
mgarellek@ucsd.edu

Christina Esposito
Macalester College
esposito@macalester.edu

Sameer Khan
Reed College
skhan@reed.edu

Mailing address for first (corresponding) author:

Patricia Keating
Dept. Linguistics
3125 Campbell Hall
UCLA
Los Angeles, CA 90095-1543

A CROSS-LANGUAGE ACOUSTIC SPACE FOR PHONATION DISTINCTIONS

ABSTRACT

Many languages use multiple phonation types for phonemic contrasts. This study examines the acoustic structure of the phonetic space for vowel phonations across languages. Our sample of 11 languages, from five language families, includes languages with contrastive phonation types on vowels, as well as those with allophonic non-modal phonation associated with particular tones, and English as a single-category case. Together these 11 languages provide 29 instances of Modal, Breathy, Creaky, Lax, Tense, Harsh, and/or Pharyngealized phonation categories.

Acoustic measures of vowel samples were made using VoiceSauce, and Multi-Dimensional Scaling was used to obtain from these measures a low-dimensional acoustic space within which the 29 categories can be compared. This space is largely two-dimensional; English lies in the middle of the space. !Xóõ, with 5 contrastive phonation categories, uses the largest space, and requires a third dimension to fully distinguish all five. The first dimension varies from modal to non-modal phonations, and is important for distinguishing allophonic creaky voice from modal.

The second dimension is like a traditional Breathy-to-Creaky continuum, and seems to be basic for phonation contrasts: all languages with a contrast use it, and languages with just two categories make them on this dimension. The dimensions can be related back to the acoustic measures that structure them, indicating which measures are most important across languages. From our results, we can recommend $H1^*-H2^*$, Subharmonic to Harmonic Ratio, Harmonic to Noise Ratio 0–500 Hz, and optionally $H1^*-A1^*$, Strength of Excitation, and Energy, as a small set of the most informative parameters that could be included in future studies.*

*Acknowledgments to follow in non-anonymous version.

Keywords: Phonetic typology, phonation types, voice quality, phonetic categories, breathy voice, creaky voice

1. INTRODUCTION. PHONATION is the production of sound in the larynx. Often this term is used in a narrow sense to refer only to the production of voicing, i.e. vibration of the vocal folds inside the larynx, but more broadly, it refers to the production of aperiodic noise as well. Human voices can vary in the rate (frequency) of vibration of the vocal folds, which we hear as changes in voice pitch; but they can also vary in the spatial pattern of the folds' vibratory movements. We hear these kinds of variations as differences in VOICE QUALITY. Each individual speaker can manipulate their vocal folds during speech so as to produce a range of voice pitches and qualities; we do this for both linguistic purposes (e.g. prosody, coarticulation) and paralinguistic purposes (e.g. emotion) (Laver 1980, Garellek 2012, Podesva & Callier 2015, Yanushevskaya, Gobl, & Ní Chasaide 2018, and references therein).

It is also the case that languages may differ in the pitch ranges and voice qualities that they typically use. For example, studies have shown that bilinguals can use two different voice qualities when speaking their two different languages (Bruyninckx 1994, Engelbert 2014), and that pairs of languages can have measurably perceptually different voice qualities (Yiu et al. 2008). That is, voice quality is one of the many ways in which languages can sound different from one another; put another way, having a native accent in a language involves using an appropriate range of voice qualities for that language.

At the same time, many languages use multiple PHONATION TYPES (OR PHONATION CATEGORIES) for phonemic contrasts. In such languages, each speaker must produce a set of distinctive laryngeal voice qualities in order to distinguish word meanings. Voice variation then cannot depend only on prosody or emotion, and it is not entirely up to the speaker how much variation to show. Instead, each speaker has to consistently employ a range of voice qualities, and do so in a way similar to other speakers of their language. What is the space of possibilities available to languages for such phonation contrasts? What laryngeal articulations, what acoustic dimensions, what auditory qualities are accessible to populations of speakers and give reliably contrasting categories? We envision a multi-dimensional phonetic space for voice, within which different languages locate their sets of phonation categories.

The main question we address in this paper is the acoustic structure of this phonetic space for phonation types, across languages. There are two prerequisites to this research. First, we need to know about the languages of the world and their linguistically-relevant phonation types, from

which we can select a (small) sample for our study. Second, we need appropriate tools for semi-automatic phonetic analysis at this scale, since even a small sample of languages will involve thousands of individual speech sound tokens. Such analysis is much more practical in the acoustic domain, and therefore we direct our research to questions about the acoustic phonetic space for phonations. And, because acoustic voice analysis is more reliable when limited to vowel sounds, we limit our research to the acoustic phonetic space for phonation in vowels, and thus to languages with non-modal phonation on vowels (rather than on consonants, or as coarticulation from consonant contrasts). Our sample includes languages with contrastive phonation types on vowels, as well as those with allophonic non-modal phonation associated with particular tones.

1.1. VOWEL PHONATION IN LANGUAGES. Ladefoged and Maddieson, in their 1996 compendium of segmental contrasts in the world's languages, discuss phonation types as a minor feature of vowel contrasts. Their phonation types are the same ones as they describe for consonants: breathy voice, slack or lax voice, modal voice, stiff or tense voice, and creaky voice. These are said to form a CONTINUUM in terms of the airflow through the glottis: at one extreme of the continuum is voicelessness with maximum glottal airflow, and at the other extreme is voicelessness with zero glottal airflow, e.g. a glottal stop [ʔ]. The voicing categories fall between these voiceless extremes: breathy and lax voice with greater airflow than modal voice, and creaky and tense voice with less. Ladefoged (1971) also describes a subset of these categories as forming a continuum in terms of glottal opening (specifically, distance between the arytenoid cartilages at the rear of the glottis¹): voiceless aspiration – breathy – modal – creaky – glottal stop. Lax and tense voice, in contrast, are distinguished by tension in the vocalis muscle.

In a different tradition, Laver and others distinguish many more phonation types, based on the possibilities of the human larynx rather than attested linguistic contrasts (e.g. Laver 1980, Esling & Harris 2005, Gobl & Ní Chasaide 2012). For example, two kinds of phonation with breathier voice qualities, i.e. those with greater glottal opening and higher airflow are distinguished: whispery voice vs. breathy voice (narrowly defined) (Moisik et al. 2019). It has long been recognized that glottal opening can occur between either the vibrating vocal folds themselves (often called the membranous or ligamental glottis), or the arytenoid cartilages at the rear of the

larynx (often called the cartilaginous glottis); these two locations of airflow have different effects (see Zhang 2016 for review). Most recently, Tian et al. (2019) suggested that languages can differ with respect to whether they use whispery voice or breathy voice: Shanghainese uses the former in its low register, while Gujarati and White Hmong use the latter. Here, we will begin with a single phonation category for “breathy voice”, but can use our cross-language analyses to explore whether finer distinctions can be made. Sub-types of creakier voice qualities likewise are possible, as reviewed by Keating et al. (2015). Again, the linguistic relevance of the sub-types is not understood. Therefore, again we begin with a single category of “creaky voice”, but can examine our own data for evidence of language differences in this regard.

An additional phonation type described separately by Ladefoged & Maddieson (1996) is strident (or harsh/epiglottalized) phonation, e.g. in languages within the “Khoisan” group. In addition to changes in vocal fold vibration, characterized by Esling as whispery voice due to vocal fold spreading, this phonation is also said to involve aryepiglottic vibration and trilling, pharyngeal narrowing, and tongue retraction (Traill 1986, Hess 1998, Esling 2005, Edmondson & Esling 2006, Miller 2007, Moisik, Lin, & Esling 2014).

In their survey paper on phonation contrasts (for both vowels and consonants) across languages, Gordon & Ladefoged (2001) use the continuum of glottal opening to describe and classify phonation types in across a wide variety of languages, and suggest measures of the glottal source and filter that can be used to quantify differences across phonation categories. There have since been many recent instrumental studies of vowel phonation contrasts in individual languages. We now have basic descriptions of such contrasts in dozens of languages, providing a foundation for cross-language comparisons. Figure 1 displays a sample of 60 languages that have phonation contrasts on vowels, including all of the languages with contrasts in the present study. (The list of languages, with further information and references, is shown in the Appendix.) To build a more balanced and better-representative sample, we chose not to sample ALL languages within the same sub-family. For instance, there are many Zapotec languages with creaky vowels, but we included only five in this sample. Instead, the languages in the sample are meant to highlight several generalizations that can be made about the distribution of vowel phonation contrasts across languages of the world. Also included were languages with mixed tone–phonation or REGISTER systems, where contrastive categories are distinguished by

phonation as well as pitch, vowel quality, and duration; in these languages, the non-modal phonation can serve as a primary or secondary cue to a particular register (Brunelle & Finkeldey 2011, Brunelle 2012). For instance, two of the four Takhian Thong Chong registers (breathy and breathy-tense) have similar pitch contours, while the other two registers (modal and tense) are fairly well-separated from these and from each other in pitch (DiCanio 2009). However, the sample does not include languages with purely allophonic phonation differences; thus the sample does not include all of the languages in the present study.

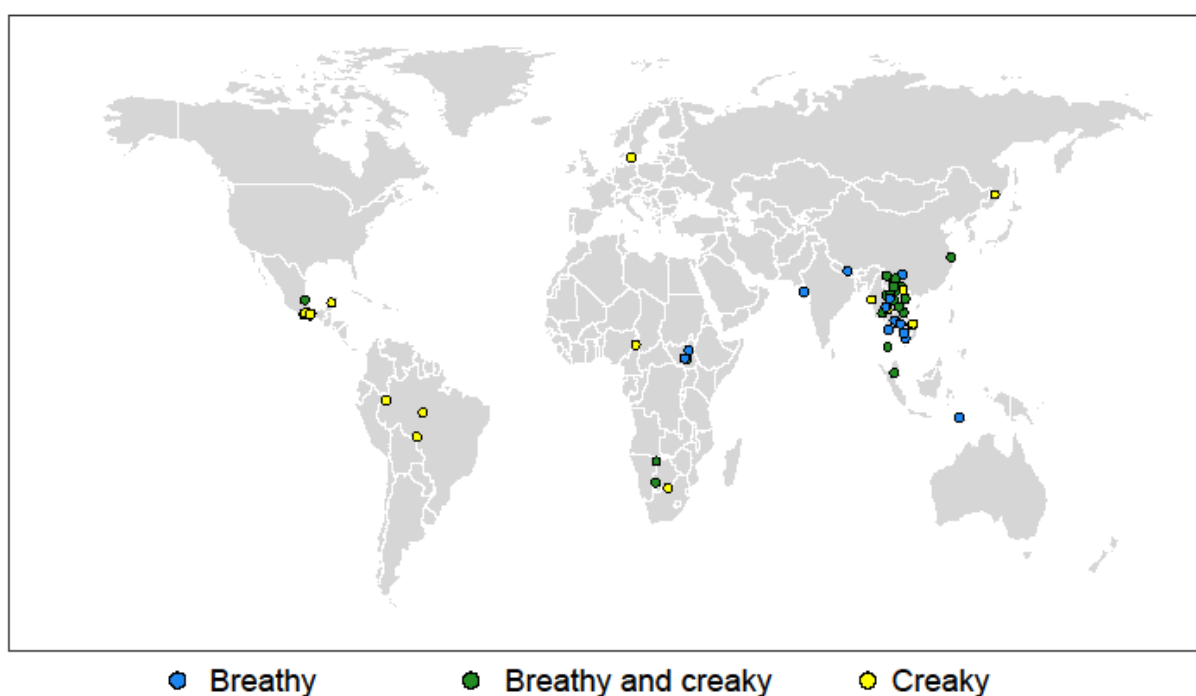


Figure 1. Geographical distribution of a sample of 60 languages with contrastive non-modal phonation on vowels. Each circle represents a language (some circles cannot be distinguished due to overlap). Blue circles are languages that have breathy (or lax) vowels; yellow circles are languages that have creaky (or tense) vowels; green circles are languages with both breathy and creaky vowels. The figure includes the languages with phonation contrasts presented in this paper, as well as others from sources such as Gordon & Ladefoged (2001) and the PHOIBLE 2.0 online repository (Moran & McCloy 2019).

Languages with non-modal phonation on vowels come from a wide variety of families. Languages from certain families – Otomanguean, Nilotic, Kx'a, and Tuu – regularly exhibit vowel phonation contrasts. Of these, Otomanguean languages (especially those in the Zapotec and Mixtec branches) tend to contrast creaky vs. modal vowels, though some Valley Zapotec languages have a three-way creaky vs. modal vs. breathy distinction in vowels (Munro & Lopez 1999, Esposito 2010b, Ariza Garcia 2018). Nilotic languages tend to contrast breathy vs. modal vowels instead, though some languages in the family (such as Bor Dinka) have breathy and harsh vowels (Edmondson & Esling 2006). Languages in the Kx'a (e.g. Ju|'hoansi) and Tuu (e.g. !Xóõ) families often have breathy and creaky vowels, which further contrast with pharyngealized and/or harsh vowels, with aryepiglottic constriction (Traill 1985, Miller 2007, Garellek 2019).

The high occurrence of vowel phonation contrasts among certain Sino-Tibetan, Hmong-Mien, Austroasiatic, and Austronesian languages is largely geographically restricted to the Southeast Asian Sprachbund, where languages with register and tense–lax contrasts abound (Maddieson & Ladefoged 1985, Zhu 2012, Brunelle & Kirby 2016).

Outside the aforementioned groups, isolated cases of languages with vowel phonation contrasts can be found across the world, and tend to contrast creaky vs. modal vowels; e.g., Udihe (Tungusic, Russia). However, Gujarati (Indo-European, India) and Shanghainese (Sino-Tibetan, China) contrast breathy vs. modal vowels, and have no contrastive creak.

Finally, languages with phonation contrasts on vowels tend to also have tonal contrasts. In our sample, only 25% of languages (mostly Austroasiatic) lack lexical tone; see the list in the Appendix. As with register languages, non-modal phonation in tone languages can serve as a secondary and also primary cue to the contrast, as is argued for (White) Hmong (Hmong-Mien; Southeast Asia) and (Black) Miao (Hmong-Mien; China) (Garellek et al. 2013, Kuang 2013).

We note that languages with phonation contrasts can have distinctive timing patterns, with non-modal phonations occurring on only portions of the vowels (Silverman 1997, Blankenship 2002). In this paper, however, we look only at vowels as wholes, focusing on the phonation qualities.

1.2. ACOUSTIC TOOLS. Since at least Klatt & Klatt (1990; see also Baken & Orlikoff 2000, Ch. 7) a variety of acoustic measures have been identified that reflect aspects of voice production and

voice quality variation. Gordon & Ladefoged (2001) list measures of periodicity, energy, spectral tilt, pitch, and duration. Of these, SPECTRAL TILT measures have been the most popular, and have been shown to characterize phonation contrasts in many languages. We use spectral tilt measures based on differences in amplitudes between individual harmonics in the spectrum.

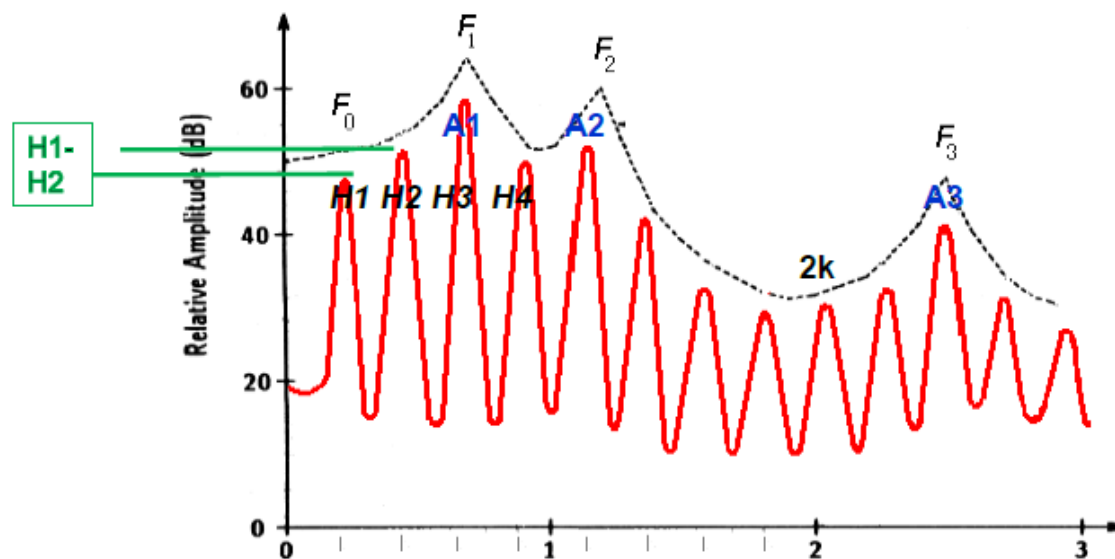


Figure 2. Sample harmonic spectrum with labels for the four lowest-frequency harmonics (H1, H2, H3, H4), the harmonic nearest 2000 Hz (“2k”), and harmonics nearest the formant frequencies F1, F2, and F3 (A1, A2, A3). Since the spectrum shows only up to 3000 Hz (3 kHz), the harmonic nearest 5000 Hz is not shown. The frequency of the first harmonic (H1) is the fundamental frequency (F0). The difference in amplitude between the first and second harmonics is H1–H2. This figure is based on one by Mammano and Nobili at <http://147.162.36.50/cochlea/cochleapages/theory/sndproc/sndcomm.htm> and is used with permission.

Figure 2 shows a harmonic spectrum in which several individual harmonics are labeled (H1 etc.). (The lowest-frequency harmonic is the fundamental component, and its frequency is the fundamental frequency, f_0 .) The amplitude of each harmonic is its height on the vertical

dimension. In the spectrum of voicing at the larynx, without the influence of the vocal tract, harmonic amplitude is expected to decline with frequency. Spectral tilt, or roll-off, refers to the steepness of that decline. A spectrum with strong higher-frequency harmonics has little tilt, and the high-frequency energy makes the voice sound brighter and tenser/more pressed. In contrast, a spectrum with weak higher-frequency harmonics has a greater tilt, and the lack of high-frequency energy makes the voice sound breathier and weaker. Spectral tilt matters not only over the spectrum as a whole, but within particular sub-ranges of frequencies, especially at the low-frequency end of the spectrum. The amplitude of the first harmonic is very important, and reflects the amount of air flowing through the glottis. A less constricted glottis during voicing means greater airflow and a stronger H1. The higher-frequency part of the spectrum, in contrast, is thought to depend more on the abruptness of closing of the vocal folds as they vibrate.

While the amplitudes of the individual harmonics can be calculated, these measures depend in large part on the overall amplitude of a sound, and so most research instead uses amplitude differences between pairs of harmonics. For example, we work with the measures H1–H2, H2–H4, H4–H2k, H2k–H5k, H1–A1, H1–A2, and H1–A3. The first four of these measures together approximate, piecewise linearly, the gross overall shape of the harmonic spectrum between the F0 and 5000 Hz. Kreiman and colleagues (e.g. Kreiman & Gerratt 2011, Garellek et al. 2016) have shown that these four measures characterize well the spectral differences across individual voices, that they can vary independently of one another, and that all of them matter perceptually to listeners.

A source spectrum may also include noise, where noise can be the result of either turbulent airflow in the glottis (essentially, aspiration noise), or aperiodic vibration of the folds (e.g. jitter). Such noise components of the source can vary in overall spectral shape, but at this time it is not known how sensitive listeners are to such spectral differences, and therefore it is not known how the noise spectrum should be modeled and measured. Some common measures of noise in the source spectrum compare the amplitudes of the harmonic and noise (inharmonic) components: these are called HARMONICS-TO-NOISE RATIO measures. The value of a harmonics-to-noise ratio is affected by the levels of both the harmonics and the noise. A high value indicates that the vibration is strongly periodic, with strong harmonics, and that there is relatively little noise of any kind. A low value indicates either that the harmonics are weak — the vibration is not

strongly periodic, or with little overall energy — or that there is significant noise of some kind. Thus, modal phonation will show the highest values, while both prototypically creaky and breathy phonations will show lower values.

In our work we use the VoiceSauce analysis program (Shue 2010, Shue et al. 2011; available for free download at <http://www.phonetics.ucla.edu/voicesauce/>), to make these and other measures automatically over many audio files. Only with such batch processing is extensive cross-language comparison possible. Possible advantages of using VoiceSauce rather than Praat for this purpose are discussed in Shue et al. (2011).²

The proliferation of acoustic measures, especially those that can be made easily and automatically by VoiceSauce, Praat, or other programs, can be confusing. Some researchers have directly compared measures in how well they distinguish different contrastive phonation categories. For example, Blankenship (2002) compared many measures (mostly made by hand) and found three that were the most useful (H1–H2, H1–A2, and Cepstral Peak Prominence, a harmonics-to-noise ratio measure). Esposito (2012) includes a table giving check marks for all the measures that showed statistically significant differences between phonation types in Hmong. Brunelle & Kirby (2016:200) consider the “variable utility of different voice quality measures” to be a pressing phonetic issue for the study of tone in languages. In our study, we can determine which acoustic measures are most useful in defining the cross-language space using data reduction and correlation techniques, as explained in §2.4.

1.3. RESEARCH QUESTIONS. What acoustic properties structure the space of phonation types across languages? One possible answer is that the space is comprised of a single dimension, perhaps corresponding to a Ladefoged-style continuum of glottal airflow, as described above (Ladefoged 1971, Ladefoged & Maddieson 1996). Yet as also noted above, Ladefoged’s continuum did not include strident/harsh phonation types, which presumably require an additional dimension.

In contrast, a more recent suggestion in the literature about individual speaker voice differences is that at least five dimensions are needed. The voice source model of Garellek et al. (2016) mentioned above proposes four frequency bands for harmonic energy differences, plus glottal noise. While these dimensions clearly underlie differences between voices of individual

speakers, it is not necessarily the case that the same dimensions characterize different types of phonation found across languages. Indeed, it has been suggested that the acoustic space for linguistic phonation might be simpler than the cross-speaker acoustic space (Garellek et al., 2013); after all, the entire population of speakers, with their individual physiological differences, must be able to produce linguistic contrasts. In this study, we focus on the cross-language space for different phonation types across languages, with the expectation that fewer than five dimensions are likely to suffice.

If we have an overall acoustic space for voice across languages, we can then ask how different languages use this space. One possibility is that in every language, distinct phonation types (whether contrastive or allophonic) are maximally dispersed, always using the full range of available voice qualities, regardless of the number of phonation categories. Alternatively, languages with more categories could use an expanded phonetic space, beyond the range used by languages with fewer categories. These alternative scenarios are already familiar in phonetics from considerations of cross-language vowel and consonant spaces (Becker-Kristal 2010, Schwartz et al. 1997a, Lindblom & Maddieson 1988).

Finally, we can ask how similar are the “same” categories or contrasts across languages. The IPA treats only three categories of phonations: modal, breathy, and creaky voice, with the latter two marked by diacritics.³ There are also strong traditions describing lax/slack voice, tense/stiff/pressed voice, and strident voice, even though the IPA does not include diacritics for these. (Unofficial extensions of the IPA, as well as ad hoc transcription systems for individual languages, include several proposals for how to represent these voice qualities.) Whenever a small number of phonetic categories is applied to many different languages, these realizations are bound to differ. We know that, for example, breathy vs. modal contrasts are made in different languages in slightly different ways acoustically: for example, Esposito (2010a) compared breathy and modal vowels in 10 languages on seven different acoustic measures and found that no two languages made the contrast in exactly the same way. Similarly, Keating et al. (2015) provide acoustic criteria that distinguish several kinds of creaky voice, and it is possible that different languages could use different subsets of these kinds. In general, it would not be surprising if different languages have somewhat different realizations of their phonation types; phonetic categories of all kinds are never exactly the same across languages. For example,

probably no two languages with vowels transcribed with the same phonetic symbols have exactly the same phonetic vowel qualities; and no two languages have exactly the same distributions of VOT values. It is very unlikely that phonation categories would be any different from these more familiar kinds of phonetic categories. Nonetheless, just as with these more familiar categories, we expect instances across languages that have been given the same phonetic label to cluster together in the larger phonetic space. The question then is how tightly they will cluster — that is, how similar the categories are when looking across languages.

2. METHODS

2.1. LANGUAGES AND PHONATION CATEGORIES. In this study we compare the contrastive or allophonic phonations of 11 languages, from five language families across different continents. Many of the languages in our sample are understudied. As described above, the basic glottal-airflow phonation continuum of voicing is taken to comprise three major categories — modal, breathy, and creaky voice — plus two intermediate categories, tense voice (between creaky and modal) and lax voice (between breathy and modal). In addition, so that noisy phonation types are represented, one of the languages in our sample is !Xóõ, with harsh (‘strident’) and pharyngealized voice qualities (Traill 1985, 1986, Ladefoged & Maddieson 1996:311-312).

Modal voice is by definition the most common phonation type in the world, and it is also the most common in this language sample: all but the three Yi languages (i.e., Southern Yi, Bo, Hani) have this category. English here is coded only for this category; no effort was made to identify individual English tokens with creaky voice, which in this language is a feature with primarily paralinguistic and/or sociolinguistic variation (Podesva & Callier 2015) rather than phonemic or allophonic. Breathily voice occurs contrastively in Gujarati, Hmong, Miao, Mazatec, !Xóõ, and Zapotec. Creaky voice is contrastive in Mazatec, !Xóõ, and Zapotec, and occurs allophonically with some lower-pitched tones in Hmong, Mandarin, and Miao. The three Yi languages contrast tense vs. lax voice, and Miao has an allophonically tense-voiced high tone (Kuang 2013).⁴ In Mazatec and !Xóõ, a creaky-voice high tone vowel sounds tense rather than creaky, but is coded here as creaky. The categories distinguished within each language, contrastively and allophonically, are given in Table 1. Sample audio files for some languages are included in the Supplemental Materials.

Table 1. Languages, phonation categories, and speakers in this study.

Language (Abbrev.) <i>(Family)</i>	Phonation categories	# Speakers (F, M)
Bo (Bo) <i>Sino-Tibetan</i>	Contrastive: Lax, Tense	10 (4, 6)
English (En) <i>Indo-European</i>	Modal	18 (11, 7)
Gujarati (Gj) <i>Indo-European</i>	Contrastive: Breathy, Modal	10 (7, 3)
(White) Hmong (Hm) <i>Hmong-Mien</i>	Contrastive: Breathy, Modal Allophonic: Creaky	23 (8, 15)
(Luchun) Hani (Lu) <i>Sino-Tibetan</i>	Contrastive: Lax, Tense	9 (4, 5)
(Black) Miao (Mi) <i>Hmong-Mien</i>	Contrastive: Breathy, Modal Allophonic: Creaky, Tense	8 (0, 8)
Mandarin (Mn) <i>Sino-Tibetan</i>	Modal Allophonic: Creaky	31 (16, 15)
(Jalapa) Mazatec (Mz) <i>Otomanguean</i>	Contrastive: Breathy, Creaky, Modal	12 (6, 6)
!Xóõ (Xo) <i>Tuu</i>	Contrastive: Breathy, Creaky, Harsh, Modal, Pharyngealized	12 (0, 12)
(Southern) Yi (Yi) <i>Sino-Tibetan</i>	Contrastive: Lax, Tense	12 (6, 6)
(Valley) Zapotec (Zp) <i>Otomanguean</i>	Contrastive: Breathy, Creaky, Modal	6 (2, 4)

While English and Gujarati are non-tonal, most of the languages in the sample are lexical tone languages. Across languages, phonation and pitch show various relationships. In general, phonation is expected to vary somewhat across a voice's pitch range: to reach the voice's lowest or highest pitches, the vocal folds must vibrate somewhat differently. This is known to be the

case in English (e.g. Kuang 2013). In Mandarin, the low-pitched portions of tones 3 (213) and 4 (51) are often allophonically creaky (Kuang 2017), so in the present study these two tones are coded as Creaky, while the other two Mandarin tones are coded as Modal. Miao, a language with five level tones (Kwan 1966, Kuang 2013), uses an expanded pitch range aided by allophonic phonation variation: the low tone is coded as Creaky, and the high tone is coded as Tense. In contrast, breathy phonation combines more freely with different pitches: here the Miao mid tone and the Hmong high falling tones are Breathly (both of these contrasting with similar-pitched Modal voice), as is the Zapotec low-falling tone (contrasting with a Creaky high-falling tone).

In some languages, phonation–pitch combinations (i.e. registers) are conventionally referred to as “tones”. For example, in Hmong creaky voice occurs with low-falling pitch, while breathy voice occurs with high-falling pitch (Esposito 2012, Garellek 2012); all other tones have modal voice. Andruski (2006) showed that in the closely related language Green Mong, clear speech involves enhancing the phonation cues of these tones, suggesting that they are important to listeners. However, Garellek et al. (2013) showed that only breathy voice is criterial for native White Hmong listeners’ tone recognition, while creaky voice is not attended to. Therefore, Hmong breathy voice is listed as contrastive in Table 1, while creaky voice is listed as allophonic.

In other languages tone and phonation cross-classify in the lexicon (that is, combine more or less freely on a single vowel). A well-known and striking example is Mpi as exemplified by Harris and Ladefoged (<http://www.phonetics.ucla.edu/vowels/chapter12/mpi.html>); here, our examples are the Yi languages (Bo, Hani, Southern Yi), Mazatec, and !Xóõ.

2.2. WORDLISTS AND AUDIO RECORDINGS. For each language with phonation contrasts, a set of words was compiled to provide several minimal- or near-minimal pairs for the contrasts. The numbers of such pairs differed across languages. The words were mostly constrained to be monosyllables (or disyllables, in the case of !Xóõ) and to exclude high vowels, nasalized vowels, nasal consonants, and voiceless fricatives, as these segment types present challenges to automatic acoustic analysis of voice quality. The exceptions were English and Mandarin, where available corpora used voiceless fricative and nasal onset consonants.

For the tone languages, all of their level tones, plus some or all of their contour tones, were included in the wordlists, though their numbers were not balanced. To provide some pitch variation, the English recordings selected included both statement and question intonations. Because the Gujarati words were recorded in unique sentences (see below), they also exhibit some pitch variation.

For languages other than Mazatec and !Xóõ, speakers were recorded using either a computer soundcard, or PCQuirer (Scicon Research & Development Inc.) with its external D/A box, all at a sampling rate of 22kHz, with a head-mounted Shure SM10A unidirectional microphone (50–15000 Hz frequency response) close to the corner of the mouth. Recordings of four indigenous languages of southwestern China (i.e. Bo, Hani, Miao, Yi) were made in the field.⁵ Recordings of Hmong and Zapotec were made in language community centers in the US, and recordings of English and Gujarati were made in our laboratory soundbooth. Our Mandarin corpus combines recordings made in China and in our laboratory soundbooth. For Mazatec and !Xóõ, existing recordings were accessed from the online public UCLA Phonetic Archive (<http://archive.phonetics.ucla.edu/>); the recordings available there are by Peter Ladefoged with Paul Kirk and Tony Traill, respectively.

Table 1 also shows the number of speakers analyzed for each language. While nine languages have both men and women speakers, two languages have only men, and in general it was not possible to evenly balance across gender.

For most of the languages, test words were spoken in isolation; in Hmong, test words were spoken in a short carrier sentence. Gujarati speakers are prone to produce spelling pronunciations of words with breathy vowels in formal settings and when reading, so we followed the elicitation procedure described in Khan (2012), in which written prompts were shown only briefly, after which speakers were asked to quickly create their own sentences beginning with the test words and produce them as many times as possible within a 10-second window.

In some of the languages, simultaneous electroglottographic recordings were made for the larger project; however, these will not be discussed in the present paper.

2.3. ACOUSTIC MEASURES. Audio recordings were converted to WAV format for analysis; Version 23 (July 5, 2015) of VoiceSauce, precompiled for standalone use on PCs, was used. A

preliminary step with VoiceSauce is to segment and label the vowels of interest from the test utterances, using Praat to make TextGrid files. A Praat script can be used to help with this task, but to some extent this is done manually. Then VoiceSauce is run to estimate a set of acoustic parameters for each of these labeled segments.

In the version of VoiceSauce used here, the original Kawahara STRAIGHT algorithm (Kawahara et al. 1998) estimates the fundamental frequency at 1-msec intervals. Harmonic spectra are computed pitch-synchronously over windows of three pitch pulses. Given the f_0 estimate, VoiceSauce uses an optimization function to locate the harmonics of the spectrum, and finds their amplitudes. This method greatly reduces variability compared to methods that use a fixed-length window. VoiceSauce then uses the Snack Sound Toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at 1-msec intervals. The harmonics nearest to these formant frequencies are located, and their amplitudes are taken as the amplitudes of the formants. Finally, the formant frequencies, along with stored estimates of their bandwidths, are used in an algorithm that corrects harmonic amplitudes for the effects of the vocal tract, using Iseli et al.'s (2007) extension of Hanson's (1995) method.⁶ Corrected harmonic amplitudes are indicated by an asterisk, e.g. H1*.⁷ (Note that f_0 is calculated by VoiceSauce here only as the basis for finding the harmonics; f_0 itself is not included in any of our analyses of phonation quality shown here, as it turns out to contribute little information.)

Other VoiceSauce parameters index periodicity and/or noise using cepstral analysis.⁸ Harmonics-to-noise ratios (HNRs) over four frequency bands (0–500 Hz, 0–1500 Hz, 0–2500 Hz, 0–3500 Hz) have high values for very periodic signals with strong harmonics, and/or when spectral noise is low; in VoiceSauce they are computed using de Krom (1993)'s algorithm. Cepstral Peak Prominence (CPP; Hillenbrand, Cleveland, & Erickson 1994) is an instance of HNRs over the entire frequency range, in which the cepstral peak is normalized relative to a regression line in the cepstrum between 1 msec and the maximum quefrequency; VoiceSauce uses Hillenbrand et al.'s algorithm. Sun (2002)'s Subharmonic-to-Harmonic ratio (SHR) describes the relative strength of any subharmonics (interharmonics) in the spectrum. Subharmonics in the spectral domain correspond to alternating periods in the time domain, and thus SHR indexes period doubling, which often occurs in creaky voice.

Finally, pitch-synchronous Energy, and Strength of Excitation (Murty & Yegnanarayana 2008), are calculated. Strength of Excitation (a measure available in VoiceSauce since 2015 but not frequently used) reflects the relative amplitude of impulse-like excitation in each pitch pulse. It is generally correlated with overall Energy, but depends more on the glottal excitation and less on vocal tract effects or noise.

VoiceSauce produces an output text file that gives the mean value of each parameter for each labeled segment; in this study we also output means over each third of each labeled segment interval.

All values from VoiceSauce were then z-score standardized within speaker and the mean values (across individual tokens) of the middle-third interval of each acoustic measurement were then calculated for each language x phonation category. The middle third was chosen because it was the most informative across the languages in the sample.

2.4. STATISTICAL ANALYSIS: MULTI-DIMENSIONAL SCALING. The multiple acoustic measures made in this study can be thought of as defining a multi-dimensional acoustic space within which different phonations can be located. However, many of these measures are inter-correlated, so the dimensions of the space are not necessarily independent, and can readily be reduced in number without much loss of information. Our goal is a map of the low-dimensional space that best fits the acoustic data, a map in which phonation categories can be located such that distances between the categories reflect their (dis)similarities. Multi-Dimensional Scaling (MDS; Kruskal & Wish 1978) is one method for reducing many individual measures to a smaller number of independent dimensions.⁹ Clopper (2012) gives examples of MDS in linguistic phonetic research; generally, these involve perception data, either confusion matrices or similarity judgments. Our use is somewhat different, since our data are from production; we take the sizes of acoustic differences as distance (dissimilarity) measures.

MDS uses measured distances between items to define a map in which those distances are preserved in a lower-dimensional space. The Manhattan (or city-block) distances on the set of acoustic measurements were used as the basis for estimates of the physical distances between all pairs of language x phonation categories, and these distances were input into Kruskal's Non-

metric Multidimensional Scaling algorithm (performed in R using the *metaMDS* function in the *vegan* package).¹⁰

This yielded solutions with different numbers of dimensions, where more dimensions typically do better at preserving the original distances, but too many dimensions offer diminishing returns in data-fitting, and can be hard to interpret and visualize. To evaluate how well a particular low-dimensional space reproduces the original distance matrix, stress values were calculated as a measure of goodness-of-fit. The smaller the stress value, the better is the fit of the reproduced space. A scree plot (Figure 3) was plot to visualize the stress values against different numbers of dimensions. The “elbow” point of the curve was identified as the sufficient number of dimensions. As suggested by Figure 3, a two-dimensional space is able to sufficiently reproduce the original distance matrix. However, the third dimension adds some additional information that helps to highlight some of the subtler phonation contrasts. The dimensions of an MDS solution can then be explored by correlating them with the original acoustic measures.

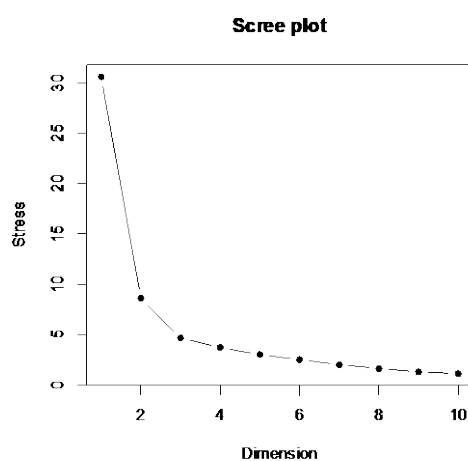


Figure 3. Scree plot for the MDS solution.

In order to have a better understanding of how phonation categories are classified by voice parameters, a Classification And Regression Tree (CART) (Breiman et al. 1984) was fitted to the dataset, using the *rpart* package in R. The depth of the tree was determined by the complexity parameter, which was optimized through the process of cross-validation and grid search. The

purpose of this procedure is to find the tree that has the best predicative accuracy. The complexity for the final model was 0.01.

3. RESULTS

3.1. 2-D SPACE. The MDS 2-dimensional solution can be viewed as a spatial map, a kind of voice space for phonation distinctions. This solution is shown in Figure 4, in which each datapoint is an average phonation category in one language. Each phonation category is plotted in a different color, the same across languages: breathier phonations are shown in shades of orange (breathy = dark orange, lax = light orange), creakier phonations are shown in shades of blue (creaky = dark blue, tense = light blue), modal phonation is shown in green, and the other two phonations are purple (pharyngealized) and pink (harsh).

First, we consider the overall organization of this space. !Xóõ, the language with the most phonation contrasts in this dataset, defines the extreme edges of the space: Each dimension makes a two- or three-way distinction in !Xóõ, with its Modal (green) and Creaky (dark blue) maximally distant on Dimension 1, and its Breathy (dark orange) maximally distant from Creaky (dark blue) on Dimension 2. The categories of the other languages lie within the diamond-shaped space defined by !Xóõ. English Modal (green), the only category coded for that language, lies right in the center of the space.

Second, we consider the two dimensions individually. Dimension 1 (x-axis) seems to give a rough Modal (on the left) to non-Modal (on the right) continuum: Modal (green) points lie mostly to the left, while the points with other colors lie mostly to the right. Modal (green) and Harsh (pink) points are well-distinguished on this dimension. Tense (light blue) and Lax (light orange) points lie together in the middle, where Modal (green) and Breathy (dark orange) points also overlap. Dimension 2 (y-axis) seems to roughly represent a harmonic-spectral-tilt continuum, from more high-frequency energy (orange-family points, at the bottom) to less high-frequency energy (blue-family points, at the top). Breathy (dark orange) and Creaky (dark blue) points are well-distinguished on this dimension. Tense (light blue), Lax (light orange), and Modal points (green) lie in the middle, overlapping with some of the Creaky (dark blue) points. Harsh phonation (pink) lies with the Breathy (dark orange) points, but also near the Creaky (dark

blue) points. Pharyngealized phonation (purple) is like Creaky (dark blue) here, on both dimensions.

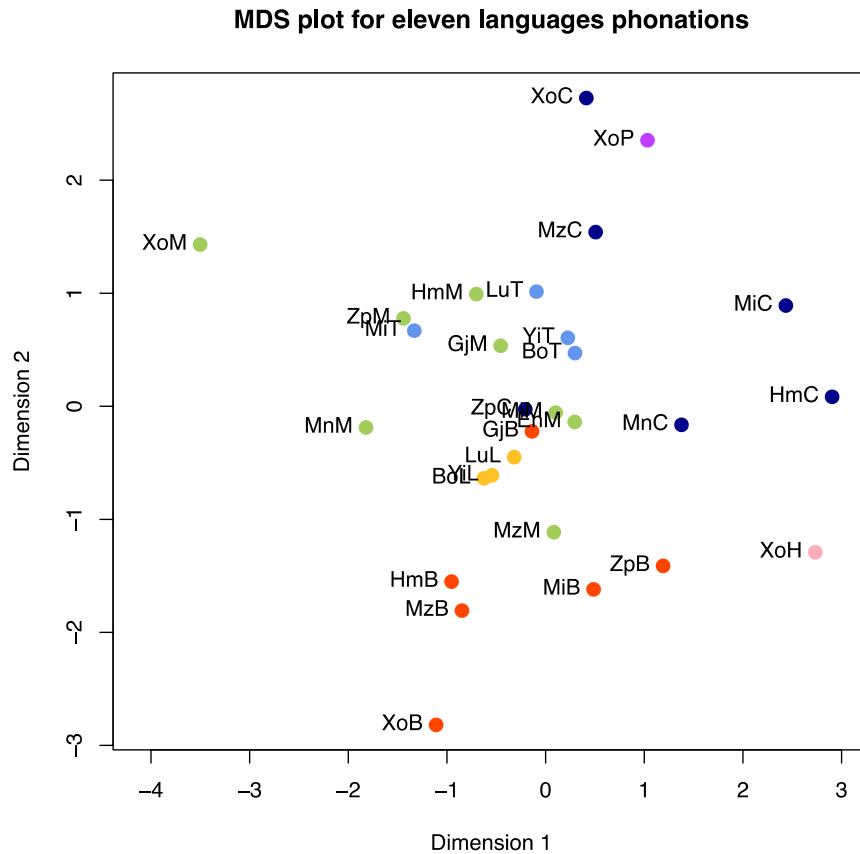


Figure 4. 2-dimensional MDS solution for the 29 language x phonation categories from the 11 languages, each category a set of acoustic measures from the middle third of all vowels. Color coding: green = Modal (M), dark orange = Breathy (B), dark blue = Creaky (C), light blue = Tense (T), light orange = Lax (L), purple = Pharyngealized (P), pink = Harsh (H). Language coding: Bo = Bo, En = English, GJ = Gujarati, Hm = Hmong, Lu = Luchun Hani, Mi = Miao, Mn = Mandarin, Mz = Mazatec, Xo = !Xóǎ, Yi = Yi, Zp = Zapotec

Third, we consider whether the various instances of the “same” category tend to cluster together. We expect this to be more or less the case because the same category names are used

across languages, but it is certainly possible that any single datapoint might look more like instances of some other category, because the voice space could be divided up and labeled differently in descriptions of different languages.

In Figure 4, three major clusters can be seen by focusing on the different-color points. In the upper-to-middle right corner are the Creaky (dark blue) and Pharyngealized (purple) points. At the bottom are the Breathy (dark orange) and Harsh (pink) points. The Tense (light blue), Lax (light orange), and Modal (green) lie mostly in the center of the space, but extend out to the far left. The Lax (light orange) points are closest to Breathy (dark orange), while the Tense points (light blue) are closest to Creaky (dark blue). The Tense (light blue) and Lax (light orange) points mostly form tighter clusters within this middle region, though Miao Tense lies a bit apart.

Modal (green), in contrast, is more spread out: there is a general Modal area, but !Xóõ Modal lies at some distance to the left. Also, some instances of Modal are more like other categories. Notably, Mazatec Modal patterns with the Breathy tokens, but this is presumably because the Mazatec Breathy–Modal contrast lies in the final third of the vowel, not in the middle third as shown here (Garellek & Keating 2011).

Otherwise, the obvious outliers are Gujarati Breathy (dark orange), which patterns with Lax (light orange) phonation in other languages, and Zapotec Creaky (dark blue), which patterns with Tense (light blue) phonation in other languages. It is entirely possible that these characterizations are phonetically accurate: the Gujarati Modal–Breathy contrast could be characterized as a Modal–Lax contrast, while the Zapotec Modal–Creaky contrast could be characterized as a Modal–Tense contrast. Indeed, the distances between these categories in this 2-D space are fairly short.

Fourth, we can compare the 11 languages in terms of how well their phonation categories are separated in the space. Visually, we can think of this as connecting all the datapoints for a language, making a triangle/rectangle etc. How big is this language-specific space? We might expect a dispersion factor (e.g. Lindblom & Maddieson 1988, Schwartz et al. 1997) to be at work here: the more categories a language contrasts, the more the language should use this 2-D space, until that space is saturated. Figure 5 facilitates such comparisons by using a different color for each language; other than the color-coding, the information is exactly the same as in Figure 4. As already noted, !Xóõ (brown points) is almost optimal, distinguishing four of its categories in a

big square that uses both dimensions of the space. (The remaining category distinction will be discussed below).

Two languages have three or four categories (allophonic or contrastive), which form triangles within the 2-D space. Hmong (light green), with three categories, is well-dispersed. Three of Miao's four categories (purple), including its allophonic Creaky and Tense along with its contrastive Breathy, cover an area similar in size to Hmong's, but then Miao's Modal category sits in the middle of this triangle.

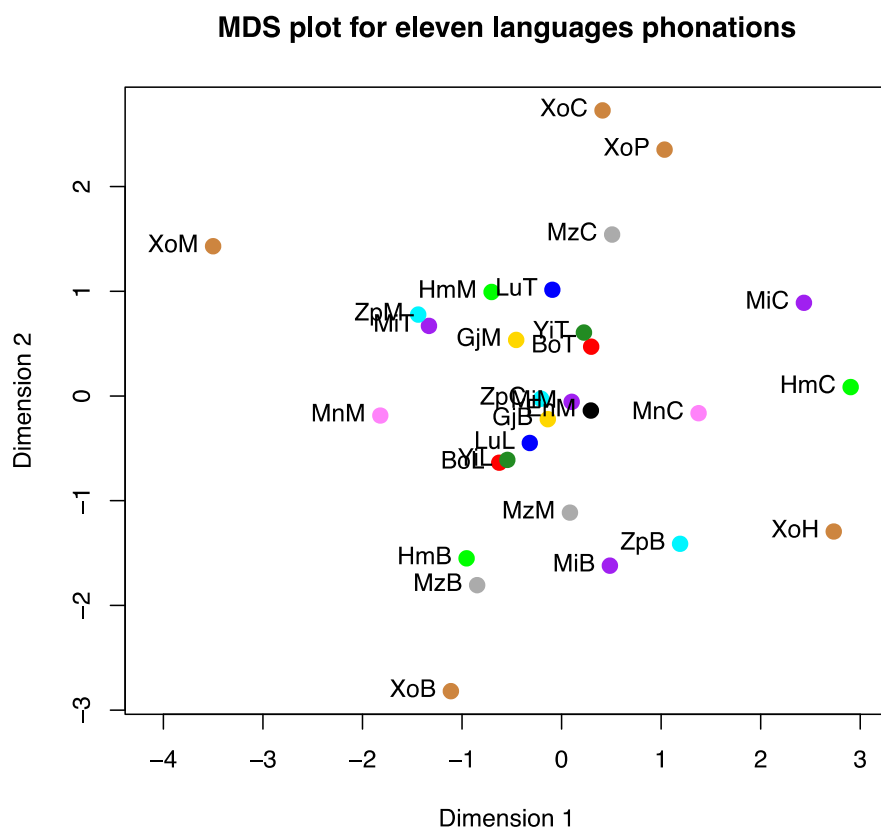


Figure 5. 2-dimensional MDS solution for the 29 language x phonation categories from the 11 languages, each category a set of acoustic measures from the middle third of all vowels. Same as Figure 4 except that here, colors code languages rather than phonation categories. Color coding: brown=!Xóõ, red=Bo, black=English, yellow=Gujarati, light green=Hmong, blue=Luchun Hani, fuschia=Mandarin, grey=Mazatec, purple=Miao, dark green=Yi, turquoise=Zapotec

Two languages have three categories falling on a line. Zapotec’s points (turquoise) lie along a line defined by the two dimensions together, but the categories are not as well-separated as in other languages, due to Zapotec Creaky lying in the Modal area. The three categories of Mazatec (grey) lie on a line defined mostly along only Dimension 2.

The languages with two categories do not use the 2-D space as fully. The Yi languages (red, blue, and dark green) and Gujarati (yellow) — which in fact also patterns like the Yi languages — use primarily Dimension 2 for their contrasts, with relatively small distances between their categories. On the other hand, Mandarin (fuchsia), with its two allophonic categories, keeps them well separated along only Dimension 1. This is the only language that uses Dimension 1 as its primary dimension. Also, as previously noted, English (black) lies in the middle of the space, making no contrasts.

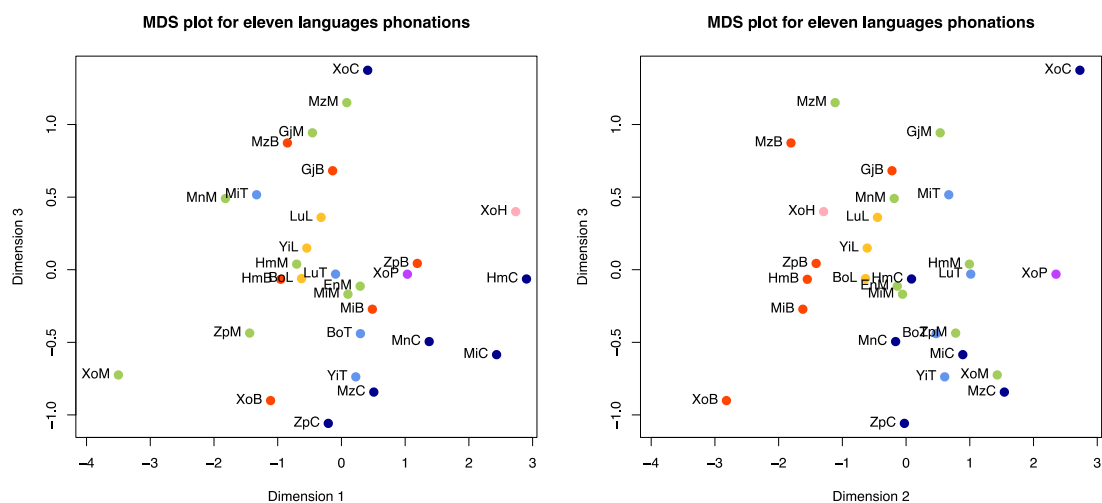


Figure 6. Pair-wise 2-dimensional spaces involving the third dimension of the MDS solution. The left panel shows Dimension 1 vs. Dimension 3, while the right panel shows Dimension 2 vs. Dimension 3. The color coding of the phonation categories and the language x phonation codes are the same as in Figure 4.

What about !Xóõ’s fifth category? It was noted above that Dimension 3 of the MDS solution does add some additional information beyond what the first two dimensions provide. Figure 6

shows the two pairwise 2-D spaces involving the third dimensions of the solution. Here, Dimension 3 is on the y-axis. It can be seen that on this dimension, !Xóõ Creaky (dark blue) and Pharyngealized (purple) are well distinguished, though in a way that enhances the difference between !Xóõ Creaky and the cluster of other Creaky datapoints, which !Xóõ Pharyngealized is close to. Dimension 3 also mildly enhances the difference between Tense (light blue) and Lax (light orange) datapoints on Dimension 2.

Last, we can test which acoustic measures are making the dimensions of the space, by checking the weights of the measures on each dimension. Table 2 shows how 21 acoustic measures from VoiceSauce relate to each of the three dimensions. Dimension 1, which looks like a non-modal to modal continuum, is most strongly based on measures of periodicity and energy in the excitation or the speech signal – SHR, the Strength of Excitation, and Energy – but also on $H4^*-H2k^*$, a mid-frequency spectral tilt measure. CPP and the other HNRs – measures of periodicity and noise – also contribute to Dimension 1, but not as strongly. Dimension 2, which looks like a glottal airflow continuum, is based on measures complementary to those for Dimension 1: unsurprisingly, $H1^*-H2^*$ and $H1^*-A1^*$ – measures of low-frequency spectral balance – but also, perhaps surprisingly, SHR, a measure of periodicity. Dimension 3, which does relatively little work in distinguishing the categories, is also based on periodicity – SHR – though presumably in a way different from on Dimension 2.

This division of labor between the parameters across the dimensions is also made clear by running MDS analyses on subsets of parameters. When only the harmonic-amplitude measures are used, the first dimension of the MDS solution gives a spectral-tilt continuum, like Dimension 2 above. When the other measures are used instead, the first dimension of that MDS solution gives a modal vs. non-modal distinction, like Dimension 1 above.

It is interesting that SHR, a measure proposed by Sun (2002) to reflect period-doubling in creaky voice, is strongly weighted on all three dimensions. The pattern on Dimension 3, where !Xóõ Creaky is distant from the other Creaky categories while !Xóõ Pharyngealized is near them, suggests that most Creaky categories and !Xóõ Pharyngealized show period-doubling, while !Xóõ Creaky, like most Modal categories, does not. Nonetheless, it turns out that, overall, removing SHR from the set of acoustic measures submitted to MDS does not change the overall shape of the solution and its dimensions.

Table 2. Weight of each acoustic measure on each dimension of the 3-D MDS solution (D1, D2, D3). Parameters that have higher weights than other parameters on each dimension are in boldface (weights > 2.0 for D1 and D2, and the highest-weighted parameter for D3).

	D1	D2	D3
H1*-H2*	0.07	2.25	0.92
H2*-H4*	1.49	0.09	0.38
H1*-A1*	1.29	2.30	0.35
H1*-A2*	1.15	1.70	0.14
H1*-A3*	1.62	1.31	0.36
H4*-H2kHz*	2.24	0.58	0.36
H2kHz*-H5kHz*	0.63	0.92	0.36
H1*	1.01	1.65	0.28
H2*	0.12	0.44	0.59
H4*	0.10	0.31	0.50
A1*	0.49	0.57	0.28
A2*	0.40	0.19	0.43
A3*	0.49	0.14	0.80
CPP	1.58	1.13	0.11
HNR < 500 Hz	1.43	1.44	0.51
HNR < 1500 Hz	1.40	1.27	0.91
HNR < 2500 Hz	1.57	1.46	0.93
HNR < 3500 Hz	1.41	1.61	0.92
Strength of Excitation (SoE)	2.34	0.72	0.39
SHR	2.90	2.28	1.69
Root mean squared (RMS) Energy	2.08	0.29	0.87

3.2. HOW SUBSETS OF LANGUAGES USE THE SPACE. From the solution for the 11 languages presented above, it is not clear how the individual languages in the sample, and the phonation

categories each contributes to the overall set, influence this overall solution. We can get a better idea of this by leaving out one or more languages and re-running the MDS analysis on a subset. Figure 7 shows the 2-D spaces that result when each of the 11 languages is left out in turn. That is, each MDS solution is for 10 of the 11 languages.

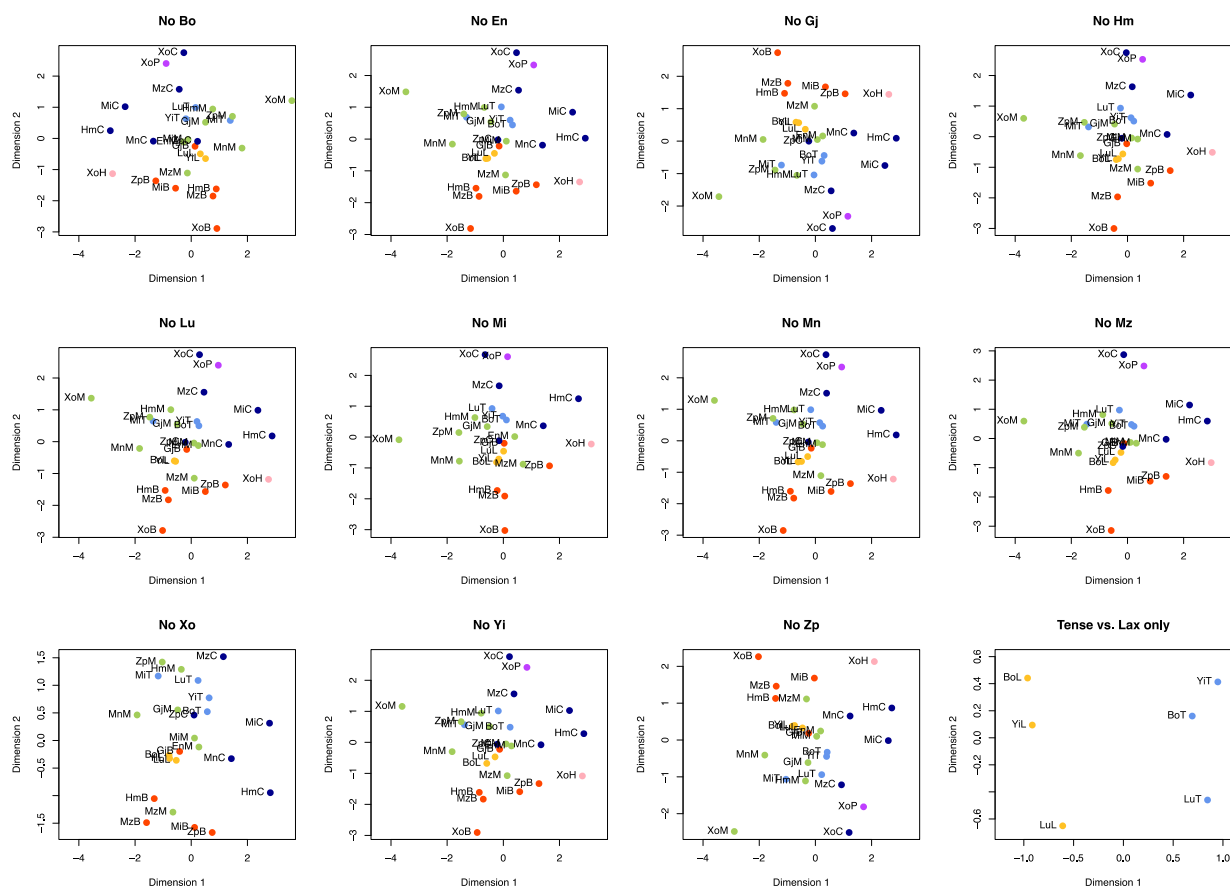


Figure 7. MDS spaces for the 11 leave-one-out subsets of languages. Each graph shows the 2-D MDS solution for a different 10-language subset. The lower right corner of the figure plots the three tense-lax languages. The color coding of the phonation categories and the language x phonation codes are the same as in Figure 4.

We have already seen that !Xoõ is the main determinant of the structure of the 11-language space. When that language is left out (bottom left graph in Figure 6), the resulting space is like a

zoomed-in view of the other languages. Hmong, Mazatec, and Miao Breathy/Creaky/Modal form the edges of this new space. The Miao categories are now better separated, with its Modal joining the English Modal in the center. The Tense–Lax distinctions are clearer.

English remains in the center not only in the 11-language space, but also in all of the 10-language subsets. However, in many spaces derived from smaller subsets (not seen in Figure 7), English appears in odd places – in a corner with other, contrastive Modal categories, or with other categories entirely. It may be that the English Modal category is the most internally heterogeneous of all the language categories, and therefore has some similarity to various categories of other languages; depending on which language categories are included in any given subset, English will most resemble different ones.

When MDS is run on only the languages with Tense–Lax contrasts, the resulting space (shown in the lower-right corner in Figure 7) distinguishes their categories much more clearly. For this subset, the new Dimension 1 is related primarily to $H1^*–H2^*$. Conversely, when these languages are omitted, the solution for the remaining languages is essentially unaffected. That is because the Tense–Lax contrasts are not important in structuring the voice space for all 11 languages.

3.3. CLASSIFICATION TREE. The analyses above, by looking at weights of measures on dimensions, and spaces defined by subsets of measures, give a fair idea of how different measures contribute to making different category distinctions within a voice space. A different way of looking at the relative importance of the various acoustic measures comes from a classification tree analysis. A Classification And Regression Tree (CART) (Breiman et al. 1984) was fitted to the dataset, using the *rpart* package in *R*. The depth of the tree was determined by the complexity parameter, which was optimized through the process of cross-validation and grid search. The purpose of this procedure is to find the tree that has the best predicative accuracy without overfitting the data. The complexity for the final model was found to be 0.01. The resulting classification tree is shown in Figure 8. The classification tree is then paired with boxplots of distributions of values on each measure that appears in the tree, here separated by the 29 language x phonation categories. These are shown in Figures 9a-c.

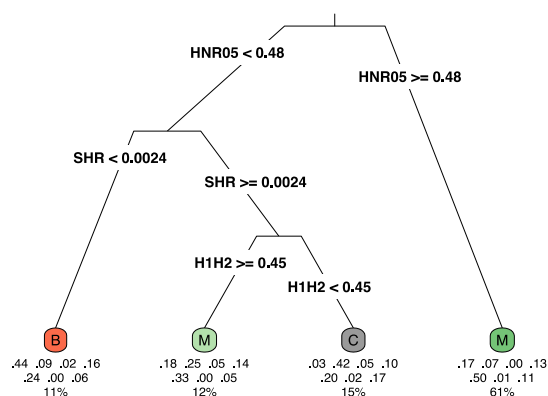


Figure 8. Classification tree of phonation categories derived from the full set of acoustic measures (middle thirds). The colors here are arbitrary, unrelated to those in previous figures.

Figure 8 shows that the first division is made by the HNR between 0–500 Hz, which splits off most of the Modals from the non-Modals. Figure 9a shows the distributions of values on this measure across language x phonation categories. It can be seen that Modals generally have higher values on this measure, though there is a spread, and that it seems to best distinguish Modal from Creaky categories.

In Figure 8, the non-Modals are then split by SHR into Breathy vs. non-Breathy. As was seen in Table 2, this measure weights strongly on all three dimensions of the MDS voice space. Figure 9b shows the distributions for this measure. Breathy categories mostly have the lowest values, while most of the Creaky categories, and Pharyngealized, have higher values. It is interesting here that SHR mostly contributes to the distinction between Breathy and other phonation types, even though (as reviewed earlier) it was developed as a way to track period-doubled Creaky voice. It seems likely that Modal voice often is somewhat creaky (which contributes to the presence of stronger sub-harmonic energy), in contrast to Breathy voice, which has very little sub-harmonic (or even harmonic) energy.

Finally, in Figure 8 the non-Breathy remainder is divided into Creaky vs. (residual) Modal by H1*–H2*. Figure 9c shows generally declining values from left to right, and thus reflects a spectral tilt continuum. (Obvious exceptions are that Gujarati Breathy again is more like Lax or Modal than like the other Breathy categories, as is Harsh voice compared to Creaky

phonation.) With respect to the classification tree, we can see in Figure 9c that this measure is generally lower for Creaky. This classification tree does not show enough splits to divide Tense or Lax phonations from Modal, or Harsh or Pharyngealized from other phonations.

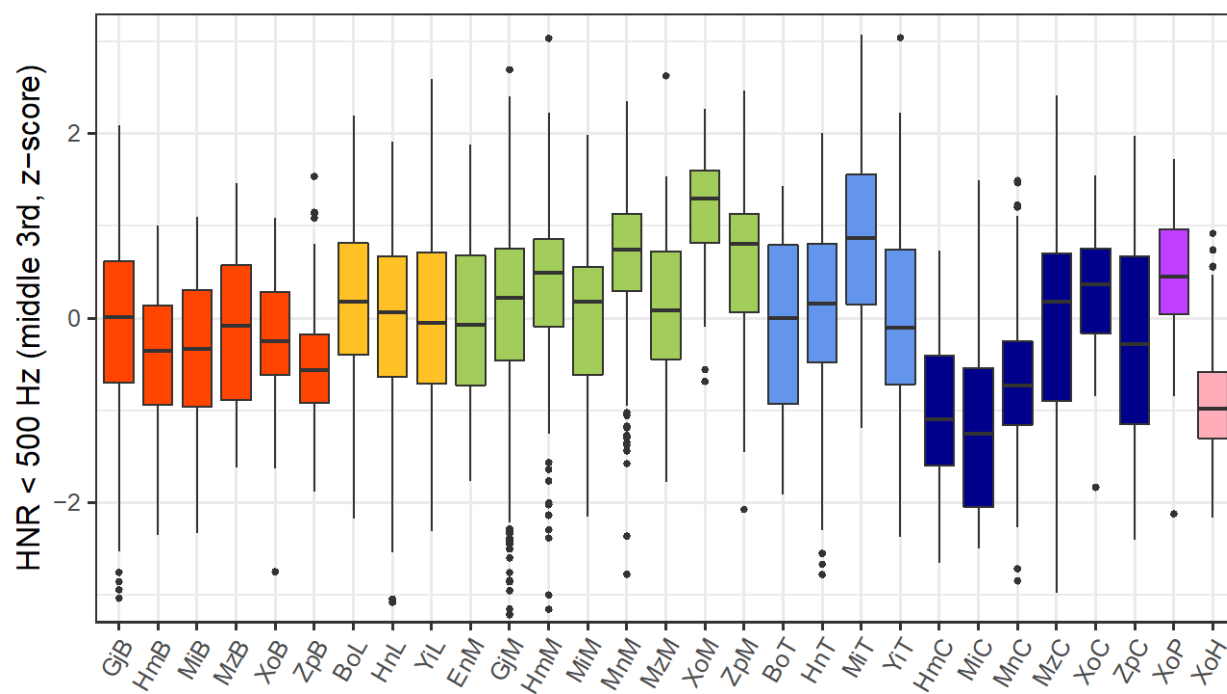


Figure 9a. HNR 0–500 Hz z-scores, middle third of vowels. Colors and language x phonation codings as in Figure 4 above.

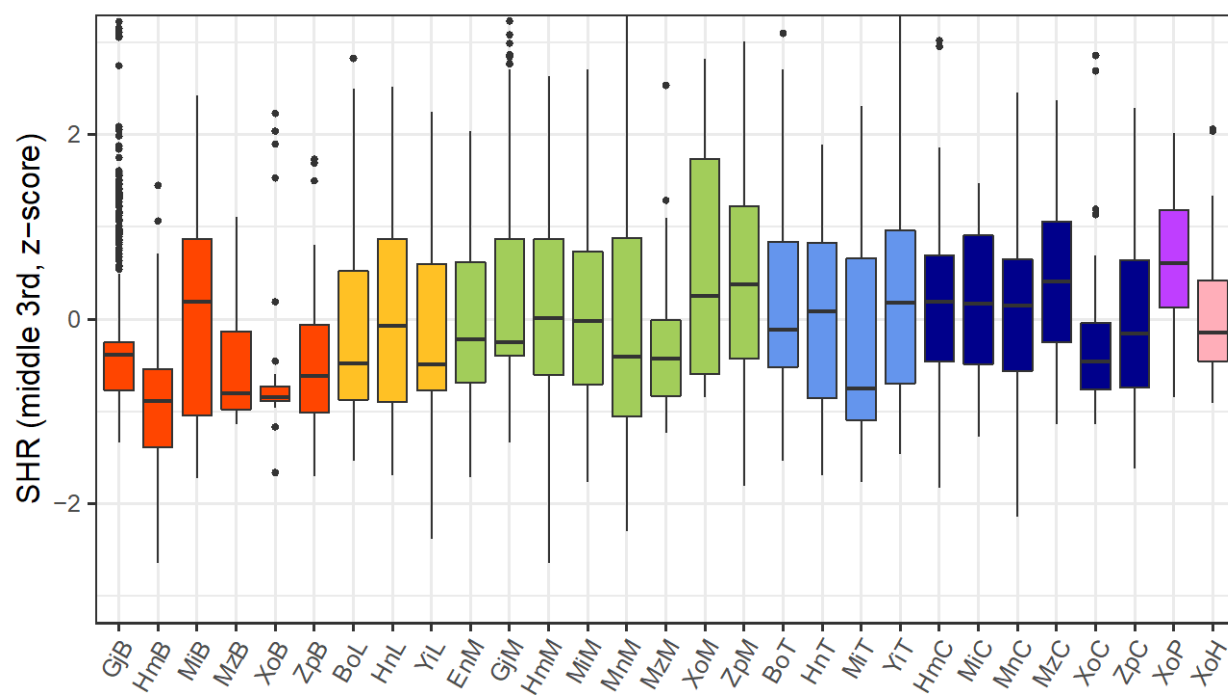


Figure 9b. Subharmonic-to-Harmonic Ratio z-scores, middle third of vowels. Colors and language x phonation codings as in Figure 4 above.

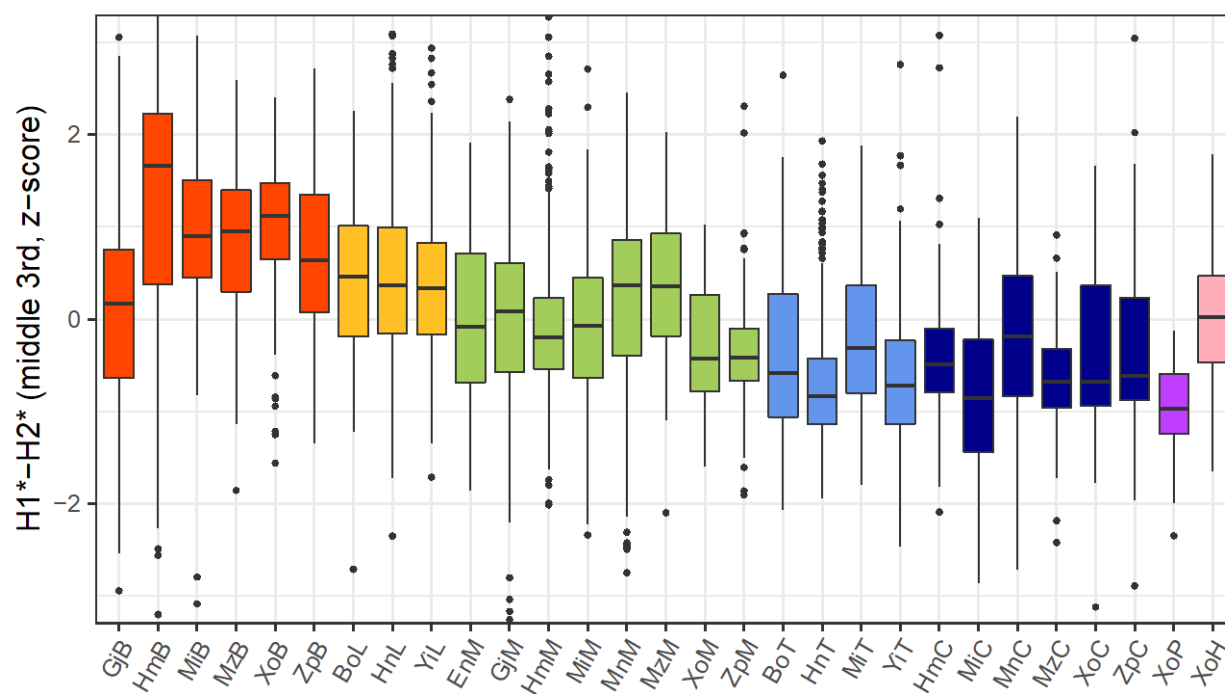


Figure 9c. $H1^*-H2^*$ z-scores, middle third of vowels. Colors and language x phonation codings as in Figure 4 above.

4. DISCUSSION AND CONCLUSIONS.

4.1. DIMENSIONS OF THE SPACE. The analyses presented above indicate that the voice space for the phonations of our sample of 11 languages is largely 2-dimensional, as presented in Figure 4. Within this space, Creaky and Pharyngealized categories cluster together in one region, Breathy and Harsh categories cluster together in an adjacent region, and Modal (and Modal-like) categories are spread out in the rest of the space.

The structure of the space bears on the question of the status of “Modal” as a phonation category: Is there indeed a distinction between modal and non-modal phonations? Our results indicate that there is, but it is a continuous rather than a binary distinction. The first (and therefore by definition most important) dimension of the space distinguishes Modal phonations from Creaky, Pharyngealized, and Harsh, with Tense and Lax at the non-modal edge of the Modal categories, and Breathy similar to Tense and Lax.

On the other hand, the second dimension of the space bears on the question of the status of a phonation type continuum, as popularized by Ladefoged. This dimension is indeed structured like a traditional Breathy-to-Creaky glottal airflow continuum. Here Modal phonations are in the middle; Tense is between Modal and Creaky, while Lax is between Modal and Breathy. This dimension appears to be basic to phonation contrasts. First, the languages contrasting just two categories (the Yi languages and Gujarati) mainly use this dimension for their distinctions. Second, even Mazatec, with three contrasting categories, mostly uses only this dimension. Third, all languages with three or more categories use this dimension. That is, there is no language that doesn't use this dimension for at least some phonemic contrast.

Nonetheless, !Xóǀ's contrasting categories are distinguished only when all three dimensions are considered. This case provides an example of dispersion within the voice space as a function of the number of contrasts. !Xóǀ has more contrasts than the other languages in our sample, and it not only needs all three dimensions to make these contrasts, but it also makes the most use of the 2-dimensional space. Lindblom and Maddieson (1988) proposed that increasing numbers of contrasting segments first lead to expanding and filling a basic phonetic space, up to a point where the space itself must become more complex by incorporating additional dimensions of contrast. Other languages are not as dispersed in the 2-D space, and the languages with just two categories mostly use just a single dimension (Dimension 2), and are not very separated even along that one. The English non-contrastive Modal category is in the middle of the space, not pushed to the modal edge. We have already suggested that the Breathy-to-Creaky phonation continuum is the most basic space for phonation contrasts; the modal/non-modal distinction is the first expansion of the space beyond that, and a third dimension is also available for even further expansion.

Thus, our results indicate that the phonetic space for voice is more complex (of higher dimensionality) than the Ladefoged-style continuum, but at the same time it is less complex than the voice source model of Garellek et al. (2016). As previously suggested, it does seem that the space for linguistic contrasts is simpler than the space for individual speaker differences.

4.2. FURTHER SUB-TYPES. We noted in the Introduction that sub-types of both Creaky and Breathy phonations are attested, but the linguistic significance of these sub-types is unknown. Do

our results contribute to this issue? In our data, the answer to this question seems partly related to the answer to another question we posed: whether the contrastive vs. allophonic status of the use of non-modal phonation matters.

We classified four categories as allophonic: Creaky in Mandarin (with no phonation contrasts) and in Hmong and Miao (with contrastive Breathy), plus Tense in Miao (also with contrastive Breathy). The allophonic Tense category of Miao patterns just like the contrastive Tense categories of the three Yi languages. However, the results for the Creaky voice categories are different. The three allophonic Creaky cases (Mandarin, Hmong, Miao) do pattern differently from the contrastive cases (Mazatec and !Xóõ). The allophonic distinctions are made along Dimension 1, with these Creaky categories in the Modal region on Dimension 2. The Miao and Hmong cases are more extreme – they lie at one end of Dimension 1, while Mandarin Creaky is closer to Modal on that dimension. In contrast, Mazatec and !Xóõ Creaky are Modal-like on Dimension 1 but lie at an extreme on Dimension 2, the Breathy-to-Creaky dimension. As discussed in the Results section, all five Creaky categories lie in a separate part of the overall space (in the upper right corner), but it seems that contrastive vs allophonic status could determine the kind of Creaky voice used in each language. Other than these differences, we see no evidence for different kinds of Creaky voice across languages (in the sense of Keating et al. 2015).

A possible reason for the importance of Dimension 1 in some languages' Creaky category is found in Huang's (2019) study of the role of individual phonation cues in Mandarin tone perception. Huang found that Mandarin perception depended on the low f_0 and/or irregular phonation properties of creak, but not on its harmonic spectrum properties. Irregular phonation (periodicity) is most strongly related to Dimension 1, while the harmonic spectrum is related more to Dimension 2. Our results suggest that contrastiveness plays a role in this distinction.

With respect to distinguishing breathy from whispery voice within our broad Breathy category, things are less clear. Hmong, Mazatec, Miao, and Zapotec are fairly close together in the space, while !Xóõ is more extreme and Gujarati less extreme on Dimension 2. If we accept the idea that Harsh voice is somewhat whispery, then we look for Breathy cases lying near our one Harsh case in order to identify a whispery part of the space. Miao and Zapotec are less

modal on Dimension 1, nearer to the Harsh case, than are Hmong and Mazatec; but the difference is so small that it is at best suggestive.

4.3. ACOUSTIC PARAMETERS. Our results have also shown which acoustic parameters seem most important in structuring this space. Certain measures are most strongly correlated with the dimensions of the MDS solution: for Dimension 1 (roughly modal to non-modal), SHR, Strength of Excitation, and Energy; for Dimension 2 (roughly glottal constriction), $H1^*-H2^*$, $H1^*-A1^*$, and again SHR; for Dimension 3 (a different modal to non-modal spread), again SHR. On the other hand, the splits on the classification tree are made by HNR 0–500 Hz (for Modals vs. others), SHR (for Breathy vs. others), and $H1^*-H2^*$ (for Creaky). SHR and $H1^*-H2^*$ are thus the parameters that are important in both analyses, and HNR 0–500 Hz is important in the classification tree analysis. These three parameters reflect very different aspects of the voice source spectrum.

A major, and surprising, result here is the dominance of the SHR across analyses and dimensions. Its role here clearly goes well beyond indexing period-doubled creaky voice. Indeed, in these analyses of our language sample, it is not especially connected to creaky voice at all.

We have noted that the Gujarati Breathy category lies with the Yi languages' Lax categories, rather than the Breathy categories of other languages. This may be related to the fact that Gujarati contrasts Breathy vowels from Modal vowels preceded by the aspiration of a Breathy consonant; Esposito & Khan (2012) demonstrated that vowels in this latter category have much stronger acoustic and electroglottographic cues to breathiness, suggesting that truly Breathy vowels in Gujarati are characterized by a subtler degree of breathiness in comparison. Notably, this result is seemingly at odds with Tian et al.'s (2019) study of breathy voice in Shanghainese and three of the languages from our sample (i.e. Hmong, Gujarati, Southern Yi). Tian et al. (2019) compared the importance of different acoustic parameters in characterizing breathy vowels in each language, and found that Gujarati was more like Hmong than like Southern Yi: while all three languages favored harmonic amplitude correlates, only Southern Yi had virtually no noise correlate. (Shanghainese was more different yet, with an almost entirely noise-based category.) A possible explanation for these opposite results lies with the more extensive set of acoustic parameters used in the present study. Tian et al. (2019) did not include SHR in their

model, yet in our results, that one parameter is important in structuring the overall phonetic space. We know very little about this parameter, and relatively little voice research that uses VoiceSauce has included it. Clearly, however, it merits detailed future study aimed at understanding exactly what it tells us about phonation.

These results, then, offer a perspective on the acoustic parameters that could be selected for including in linguistic studies of phonation. Researchers may find the array of parameters in VoiceSauce overwhelming. In our own past practice, we have tended to select one or two harmonic measures, plus CPP as a noise measure. Now, with this new, more inclusive, analysis, we can recommend $H1^*-H2^*$, SHR, HNR 0–500 Hz, and optionally $H1^*-A1^*$, Strength of Excitation, and Energy, as a small set of the most informative parameters.

4.4. CONCLUSIONS. In sum, our study has demonstrated an overall acoustic space for voice quality. We looked at a diverse range of languages with contrasting or allophonic phonation categories, and found that most such categories can be accommodated within a 2-dimensional space, though the most complex system uses 3 dimensions. Just as with the more familiar phonetic category distinctions of vowel quality and VOT, phonation categories that are given the same labels in different languages are found in similar parts of the overall space, but nonetheless differ across languages.

APPENDIX

Further information about the 60 languages with phonation contrasts on vowels mapped in Figure 1.

Language	Family	Phonation types	Tonal/pitch contrast	References
A-Hmao, Western	Hmong-Mien	B, M	Y	Johnson (1999)
Bai	Sino-Tibetan	B, L, M, T, H	Y	Edmondson et al. (2001)
Bo	Sino-Tibetan	T, L	Y	Kuang & Keating (2014)

Burmese	Sino-Tibetan	M, C	Y	Watkins (2001), Gruber (2011)
Cham	Austronesian	B, M	Y	Brunelle (2012)
Chichimeco	Otomanguean	B, M, C	Y	Kelterer (2011)
Chong	Austroasiatic	B, M, C	Y	Thongkum (1987a); Blankenship (1997); DiCanio (2009)
Danish	Indo-European	M, C	N	Fischer-Jørgenson (1989)
Dinka, Bor	Nilotic	B, M, H	Y	Edmondson & Esling (2006)
Gujarati	Indo-European	B, M	N	Fischer-Jorgenson (1967); Pandit (1957); Nara (2017); Khan (2012); Esposito & Khan (2012); Nelson et al. (2016)
Halang	Austroasiatic	B, M	N	Cooper & Cooper 1966
Hani, Haoni	Sino-Tibetan	T, L	Y	Maddieson & Ladefoged (1985)
Hani, Luchun	Sino-Tibetan	T, L	Y	Kuang & Keating (2014)
Hmong, Green	Hmong-Mien	B, M, C	Y	Andruski & Ratliff (2000)
Hmong, White	Hmong-Mien	B, M, C	Y	Huffman (1987); Esposito (2012); Garellek (2012), Garellek et al. (2013), Esposito & Khan (2012)
Jingpho	Sino-Tibetan	T, L	Y	Maddieson & Ladefoged (1985)
Ju'hoansi	Kx'a	B, M, C, H	Y	Miller (2007)
Karen, Sgaw	Sino-Tibetan	B, M, C	Y	Sun (2016); Brunelle & Finkeldey (2011)
Kedang	Austronesian	B, M	N	Samely (1991)
Khmer	Austroasiatic	B, M	N	Thongkum (1988); Wayland & Jongman (2003); Kirby (2014)
Kri	Austroasiatic	T, L	N	Enfield & Diffloth (2009)
Kui	Austroasiatic	B, M	N	Thongkum (1987a)

Lisu	Sino-Tibetan	M, C	Y	Tabain et al. (2019)
Mah Meri	Austroasiatic	T, L	N	Kruspe & Hajek (2009)
Mamaindê	Nambiquaran	M, C	Y	Eberhard (2009)
Mambay	Niger-Congo	M, C, P	Y	Anonby (2006)
Maya, Yucatec	Mayan	M, C	Y	Frazier (2009)
Mazatec, Jalapa	Otomanguean	B, M, C	Y	Blankenship (1997); Kirk et al. (1993); Silverman et al. (1995); Garellek & Keating (2011)
Miao, Xinzhai	Hmong-Mien	B, M	Y	Liu et al. (2018)
Mixtec, Coatzacoapan	Otomanguean	M, C	Y	Gerfen & Baker (2005)
Mixtec, Ixpantepec Nieves	Otomanguean	M, C	Y	Carroll (2015)
Mon, Ban Nakhonchum	Austroasiatic	B, M	N	Thongkum (1987b); Abramson et al. (2015)
Mpi	Sino-Tibetan	M, T	Y	Blankenship (2002)
Mundurukú	Tupian	M, C	Y	Picanço (2005)
Nuer	Nilotic	B, M	Y	Monich (2017)
Nyah Kur	Austroasiatic	B, M	N	Thongkum (1987a)
Oeshi, Louma	Sino-Tibetan	T, L	Y	Lew & Gruber (2016)
Phalok	Austroasiatic	B, M	N	Thongkum (1988)
Reel, Thok	Nilotic	B, M	Y	Reid (2010)
Sedang	Austroasiatic	M, C	N	Smith (1975)
Shanghainese	Sino-Tibetan	B, M, C	Y	Tian & Kuang (2016)
Suai, Kuai	Austroasiatic	B, M	Y	Abramson et al. (2004)
Tamang	Sino-Tibetan	B, M	Y	Mazaudon & Michaud (2008)
Ticuna, Cushillococha	Ticuna	M, C	Y	Skilton (2016)

Trique	Otomanguean	M, T	Y	DiCanio (2010)
Udihe	Tungusic	M, C	N	Nikolaeva & Tolskaya (2001)
Vietnamese, Northern	Austroasiatic	B, M, C	Y	Pham (2003); Brunelle et al. (2010)
Wa, Lin Tsang	Austroasiatic	T, L	N	Maddieson & Ladefoged (1985)
Wa, several varieties	Austroasiatic	B, M	N	Watkins (1999, 2002)
Yi, Luquan Nasu	Sino-Tibetan	T, L	Y	Maddieson & Ladefoged (1985)
Yi, Northern	Sino-Tibetan	T, L	Y	Edmondson et al. (2001)
Yi, Southern	Sino-Tibetan	T, L	Y	Kuang & Keating (2014)
Zapotec, Choapam	Otomanguean	M, C	Y	Lyman & Lyman (1977)
Zapotec, San Juan Guelavia	Otomanguean	M, C	Y	Jones & Knudson (1997)
Zapotec, San Lucas Quiaviní	Otomanguean	B, M, C	Y	Munro & López (1999)
Zapotec, Santa Ana Del Valle	Otomanguean	B, M, C	Y	Esposito (2010)
Zapotec, Yalálag	Otomanguean	M, C	Y	Avelino (2010)
Zhuang, Du'an	Tai-Kadai	M, C	Y	Perkins et al. (2016)
ꨀHoan, Eastern	Kx'a	M, C, P	Y	Honken (2013)
!Xóô (Taa)	Tuu	P, H, B, M, C	Y	Trail (1985); Naumann (2016); Garellek (2019)

REFERENCES

- ABRAMSON, ARTHUR S.; L. THERAPHAN; and P. W. NYE. 2004. Voice register in Suai Kuai: An analysis of perceptual and acoustic data. *Phonetica* 61.147–171.
- ABRAMSON, ARTHUR S.; MARK K. TIEDE; and T. LUANGTHONGKUM. 2015. Voice register in Mon: Acoustics and electroglottography. *Phonetica* 72.237–256.
- ANDRUSKI, JEAN E. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34.388–404.
- ANDRUSKI, JEAN E., and MARTHA RATLIFF. 2000. Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association* 30.37–61.
- ANONBY, ERIK JOHN. 2006. Mambay. *Journal of the International Phonetic Association* 36.221–233.
- ARIZA GARCIA, ANDREA. 2018. Phonation types and tones in Zapotec languages: A synchronic comparison. *Acta Linguistica Petropolitana* 14.485–516.
- AVELINO, HERIBERTO. 2010. Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *Journal of Voice* 24.270–280.
- BAKEN, R. J., and ROBERT F. ORLIKOFF. 2000. *Clinical Measurement of Speech and Voice*. 2nd edition. San Diego, CA: Singular Publishing Group.
- BALL, MARTIN J.; JOHN H. ESLING; and B. CRAIG DICKSON. 2018. Revisions to the VoQS system for the transcription of voice quality. *Journal of the International Phonetic Association* 48.165–171.
- BECKER-KRISTAL, ROY. 2010. *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. Los Angeles, CA: UCLA dissertation.
- BISHOP, JASON, and PATRICIA KEATING. 2012. Perception of pitch location within a speaker's own range: fundamental frequency, voice quality and speaker sex. *Journal of the Acoustical Society of America* 132.1100–1112.
- BLANKENSHIP, BARBARA. 1997. *The time course of breathiness and laryngealization in vowels*. Los Angeles, CA: UCLA dissertation.
- BLANKENSHIP, BARBARA. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30.163–191.

- BREIMAN, LEO; JEROME FRIEDMAN; CHARLES J. STONE; and RICHARD A. OLSHEN. 1984. *Classification and Regression Trees*. Boca Raton, FL: Taylor and Francis Group.
- BRUNELLE, MARC. 2012. Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *Journal of the Acoustical Society of America* 131.3088–3102.
- BRUNELLE, MARC, and JOSHUA FINKELDEY. 2011. Tone Perception in Sgaw Karen. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 372–375.
- BRUNELLE, MARC, and JAMES KIRBY. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10.191–207.
- BRUNELLE, MARC; DUY DUONG NGUYÊN; and KHAC HUNG NGUYEN. 2010. A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica* 67.147–169.
- BRUYNINCKX, MARIELLE; BERNARD HARMEGNIES; JOAQUIM LLISTERRI; and DOLORS POCH-OLIVÉ. 1994. Language-induced voice quality variability in bilinguals. *Journal of Phonetics* 22.19–31.
- CARROLL, LUCIEN SERAPIO. 2015. *Ixpantepec Nieves Mixtec Word Prosody*. San Diego, CA: UCSD dissertation.
- CLOPPER, CYNTHIA G. 2012. Clustering and classification methods. *The Oxford Handbook of Laboratory Phonology*, ed. by A. C. Cohn, C. Fougeron, and M. K. Huffman, 678–691. Oxford: Oxford University Press.
- COOPER, JAMES, and NANCY COOPER. 1966. Halang phonemes. *Mon-Khmer Studies* 2.87–98.
- DE KROM, Guus. 1993. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research* 36.224–266.
- DICANIO, CHRISTIAN T. 2009. The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association* 39.162–188.
- DICANIO, CHRISTIAN T. 2010. Itunyoso Trique. *Journal of the International Phonetic Association* 40.227–238.
- EBERHARD, DAVID M. 2009. *Mamaindê Grammar: A Northern Nambikwara language and its cultural context*. Amsterdam: Netherlands Graduate School of Linguistics.

- EDMONDSON, JEROLD A., and JOHN H. ESLING. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23.157–191.
- EDMONDSON, JEROLD; JOHN ESLING; JIMMY G. HARRIS; LI SHAONI; and LAMA ZIWO. 2001. The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. *Mon-Khmer Studies* 31.83–100.
- ENFIELD, NICHOLAS J., and GERARD DIFFLOTH. 2009. Phonology and sketch grammar of Kri, a Vietic language of Laos. *Cahiers de linguistique Asie orientale* 38, v–69.
- ENGELBERT, ANA PAULA PETRIU FERREIRA. 2014. Cross-linguistic effects on voice quality: A study on Brazilians' production of Portuguese and English. *Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers in Applied Linguistics* 5. 157–170.
- ESLING, JOHN H. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 50.13–44.
- ESLING, JOHN H., and JIMMY G. HARRIS. 2005. States of the glottis: An articulatory phonetic model based on laryngoscopic observations. *A figure of speech: A Festschrift for John Laver*, ed. by W. J. Hardcastle and J. Mackenzie Beck, 347–383. Mahwah NJ: Lawrence Erlbaum Associates.
- ESPOSITO, CHRISTINA M. 2010a. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 28.71–139.
- ESPOSITO, CHRISTINA M. 2010b. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40.181–198.
- ESPOSITO, CHRISTINA M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics* 40.466–476.
- ESPOSITO, CHRISTINA M., and SAMEER UD DOWLA KHAN. 2012. Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association* 42.123–143.
- FISCHER-JØRGENSEN, ELI. 1967. Phonetic analysis of breathy murmured vowels. *Indian Linguistics* 28.71–139.

- FISCHER-JØRGENSEN, ELI. 1989. Phonetic analysis of the stød in standard Danish. *Phonetica* 46.1–59.
- FRAZIER, MELISSA. 2009. *The production and perception of pitch and glottalization in Yucatec Maya*. Chapel Hill, NC: University of North Carolina at Chapel Hill dissertation.
- GARELLEK, MARC. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40.152–161.
- GARELLEK, MARC. 2017. The phonetics of voice. *The Routledge Handbook of Phonetics*, ed. by W. F. Katz and P. F. Assmann, 75–106. London: Routledge.
- GARELLEK, MARC. 2019. Acoustic discriminability of the complex phonation system in !Xóö. *Phonetica* 1–30. <https://doi.org/10.1159/000494301>.
- GARELLEK, MARC; CHRISTINA M. ESPOSITO; PATRICIA KEATING; and JODY KREIMAN. 2013. Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America* 133.1078–1089.
- GARELLEK, MARC, and PATRICIA KEATING. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41.185–205.
- GARELLEK, MARC; ROBIN SAMLAN; BRUCE R. GERRATT; and JODY KREIMAN. 2016. Modeling the voice source in terms of spectral slopes. *Journal of the Acoustical Society of America* 139.1404–1410.
- GERFEN, CHIP, and KIRK BAKER. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics* 33.311–334.
- GOBL, CHRISTER, and AILBHE NÍ CHASAIDE. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11.481–490.
- GOBL, CHRISTER, and AILBHE NÍ CHASAIDE. 2012. Voice source variation and its communicative functions. *The handbook of phonetic sciences* (2nd edition), ed. by William J. Hardcastle, John Laver, and Fiona E. Gibbon, 378–423. John Wiley & Sons.
- GORDON, MATTHEW, and PETER LADEFOGED. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29.383–406.
- GRUBER, JAMES. F. 2011. *An articulatory, acoustic, and auditory study of Burmese tone*. Washington, DC: Georgetown University dissertation.

- HANSON, HELEN M. 1995. *Glottal characteristics of female speakers*. Cambridge, MA: Harvard University dissertation.
- HANSON, HELEN M.; KENNETH N. STEVENS; HONG-KWANG JEFF KUO; MARILYN Y. CHEN; and JANET SLIFKA. 2001. Towards models of phonation. *Journal of Phonetics* 29.451–480.
- HUANG, YAQIAN. 2019. The role of creaky voice attributes in Mandarin tonal perception. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, ed. by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren, 1465–1469. Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- HILLENBRAND, JAMES RONALD A. CLEVELAND, and ROBERT L. ERICKSON. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37.769–778.
- HONKEN, HENRY. 2003. Eastern #Hoan. *The Khoesan Languages*, ed. by Rainer Vossen, 218–223. New York: Routledge.
- HUFFMAN, MARIE K. 1987. Measures of phonation type in Hmong. *Journal of the Acoustical Society of America* 81.495–504.
- ISELI, MARKUS; YEN-LIANG SHUE; and ABEER ALWAN. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America* 121.2283–2295.
- JOHNSON, MICHAEL. 1999. Tone and phonation in Western A-Hmao. *SOAS working papers in linguistics* 9.227–251.
- JONES, TED, and LYLE KNUDSON. 1977. Guelavía Zapotec phonemes. *Studies in Otomanguean Phonology*. ed. William R. Merrifield, pp. 163–80. Dallas: Summer Institute of Linguistics.
- KAWAHARA, HIDEKI; ALAIN DE CHEVEIGNÉ; and ROY D. PATTERSON. 1998. An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite. *Proceedings of the International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, paper 0659.
- KEATING, PATRICIA; MARC GARELLEK; and JODY KREIMAN. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow.
- KELTERER, ANNELIESE 2017. Non-modal voice quality in Chichimeco - "Hablamos más con la garganta". M.A. thesis, Lund University.

- KHAN, SAMEER UD DOWLA. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40.780–795.
- KIRBY, JAMES P. 2014. Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics* 43.69–85.
- KIRK, PAUL L.; PETER LADEFOGED; and JENNY LADEFOGED. 1984. Using a spectrograph for measures of phonation types in a natural language. *UCLA Working Papers in Phonetics* 59.102–113.
- KLATT, DENNIS H., and LAURA C. KLATT. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87.820–857.
- KREIMAN, JODY; BRUCE R. GERRATT; and NORMA ANTOÑANZAS-BARROSO. 2007. Measures of the glottal source spectrum. *Journal of Speech Language and Hearing Research* 50.595–610.
- KREIMAN, JODY, and BRUCE R. GERRATT. 2005. Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America* 117.2201–2211.
- KREIMAN, JODY, and BRUCE R. GERRATT. 2011. Modeling overall voice quality with a small set of acoustic parameters. *Journal of the Acoustical Society of America* 129.2529.
- KREIMAN, JODY; BRUCE R. GERRATT; MARC GARELLEK; ROBIN SAMLAN; and ZHAOYAN ZHANG. 2014. Toward a unified theory of voice production and perception. *Loquens* 1.e009.
- KREIMAN, JODY; MARC GARELLEK; GANG CHEN; ABEER ALWAN; and BRUCE R. GERRATT. 2015. Perceptual evaluation of voice source models. *Journal of the Acoustical Society of America* 138.1–10.
- KRUSKAL, JOSEPH B., and MYRON WISH. 1978. *Multidimensional scaling*. Newbury Park, CA: Sage.
- KRUSPE, NICOLE, and JOHN HAJEK. 2009. Mah Meri. *Journal of the International Phonetic Association* 39.241–248.
- KUANG, JIANJING. 2013. The tonal space of contrastive five level tones. *Phonetica* 70.1–23.
- KUANG, JIANJING. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America* 142.1693–1706.
- KUANG, JIANJING, and ALETHEIA CUI. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71.194–214.

- KUANG, JIANJING, and PATRICIA KEATING. 2014. Vocal fold vibratory patterns in tense versus lax phonation contrasts. *Journal of the Acoustical Society of America* 136.2784–2797.
- KWAN, JULIA C. 1966. A phonology of a Black Miao dialect. M.A. thesis, University of Washington, Seattle.
- LADEFOGED, PETER. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.
- LADEFOGED, PETER, and IAN MADDIESON. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- LAVER, JOHN. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- LEW, SIGRID, and JAMES GRUBER. 2016. An acoustic analysis of tone and register in Louma Oeshi. *Proceedings of the Linguistic Society of America* 33.1–14.
- LINDBLOM, BJORN, and IAN MADDIESON. 1988. Phonetic universals in consonant systems. *Language, speech, and mind: Studies in Honour of Victoria A. Fromkin*, ed. by Larry Hyman and Charles N. Li, 62–78. London: Routledge.
- LIU, WEN; YOU-JING LIN; ZHENGHUI YANG; and JIANGPING KONG. 2018. Hmu (Xinzhai variety). *Journal of the International Phonetic Association* (First View online).
- LYMAN, LARRY, and ROSEMARY LYMAN. 1977. Choapan Zapotec phonology. *Studies in Otomanguan phonology* 54.137–161.
- MADDIESON, IAN, and PETER LADEFOGED. 1985. "Tense" and "lax" in four minority languages of China. *Journal of Phonetics* 13.433–454.
- MAZAUDON, MARTINE, and ALEXIS MICHAUD. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65.231–256.
- MILLER, AMANDA L. 2007. Guttural vowels and guttural co-articulation in Jul'hoansi. *Journal of Phonetics* 35.56–84.
- MOISIK, SCOTT R.; HUA LIN; and JOHN H. ESLING. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44: 21–58.
- MONICH, IRINA V. 2017. Vowel length in Nuer. *Proceedings of the Annual Meeting on Phonology* 4 (AMP 2016), New York, 1–6.

- MORAN, STEVEN, and DANIEL MCCLOY (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History.
- MUNRO, PAMELA, and FELIPE H. LOPEZ. 1999. *Di'csyonaary X: tè'n Dìi'zh Sah Sann Lu'uc: San Lucas Quiavini Zapotec Dictionary*. Chicano Studies Research Center.
- MURTY, K. SRI RAMA, and BAYYA YEGNANARAYANA. 2008. Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16.1602–1613.
- NARA, KIRANPREET. 2017. Acoustic and electroglottographic study of breathy and modal vowels as produced by heritage and native Gujarati speakers. *Proceedings of INTERSPEECH*, Stockholm, 1054–1058.
- NAUMANN, CHRISTINE. 2016. The phoneme inventory of Taa West !Xoon dialect. *Lone Tree: Scholarship in the service of the Koon: Essays in memory of Anthony T. Traill*, ed. by Rainer Vossen and Wilfred H. G. Haacke, 311–351. Cologne, Germany: Rüdiger Koppe.
- NELSON, MAX; KELLY BERKSON; and CHRISTINA ESPOSITO. 2016. Perception of breathy phonation in Gujarati. *Formal Approaches to South Asian Languages* 2.14–22.
- NIKOLAEVA, IRINA, and MARIA TOLSKAYA. 2011. *A grammar of Udihe*. Berlin: Walter de Gruyter.
- PANDIT, PRABODH B. 1957. Nasalization, aspiration and murmur in Gujarati. *Indian Linguistics* 17.165–172.
- PERKINS, JEREMY; SEUNGHUN LEE; and JULIÁN VILLEGAS. 2016. An interplay between F0 and phonation in Du'an Zhuang tone. *Proceedings of Tonal Aspects of Languages 2016*, Buffalo, 56–59.
- PHAM, ANDREA H. 2003. *Vietnamese tone: A new analysis*. New York: Routledge.
- PICANÇO, G. L. 2005. *Mundurukú: Phonetics, phonology, synchrony, diachrony*. Vancouver, BC: University of British Columbia dissertation.
- PODESVA, ROBERT J., and PATRICK CALLIER. 2015. Voice quality and identity. *Annual review of applied linguistics* 35.173–194.
- REID, TATIANA. 2010. *Aspects of phonetics, phonology and morphophonology of Thok Reel*. Edinburgh, UK: University of Edinburgh dissertation.
- SAMELY, URSULA. 1991. *Kedang Eastern Indonesia: Some aspects of its grammar*. Hamburg: Helmut Buske.

- SCHWARTZ, JEAN-LUC; LOUIS-JEAN BOË; NATHALIE VALLÉE; and CHRISTIAN ABRY. 1997a. Major trends in vowel system inventories. *Journal of Phonetics* 25. 233–253.
- SCHWARTZ, JEAN-LUC; LOUIS-JEAN BOË; NATHALIE VALLÉE; and CHRISTIAN ABRY. 1997b. The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25.255–286.
- SHUE, YEN-LIANG. 2010. *The voice source in speech production: Data, analysis and models*. Los Angeles, CA: UCLA dissertation.
- SHUE, YEN-LIANG; PATRICIA KEATING; CHAD VICENIK; and KRISTINE YU. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 1846–1849.
- SILVERMAN, DANIEL. 1997. Laryngeal complexity in Otomanguean vowels. *Phonology* 14.235–61.
- SILVERMAN, DANIEL; BARBARA BLANKENSHIP; PAUL KIRK; and PETER LADEFOGED. 1995. Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* 37.70–88.
- SIMPSON, ADRIAN P. 2012 The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics* 40.477–490.
- SJÖLANDER, KÅRE. 2004. Snack sound toolkit. KTH Stockholm, Sweden. Online: <http://www.speech.kth.se/snack>.
- SKILTON, AMALIA. 2016. *Contrastive voice quality in Cushillococha Ticuna*. M.A. thesis, University of California, Berkeley.
- SMITH, KENNETH D. 1975. *Phonology and syntax of Sedang, a Vietnam Mon-Khmer language*. Philadelphia, PA: University of Pennsylvania dissertation.
- SUN, XUEJING. 2002. Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, I-333–I-336.
- SUN, XIAOLE. 2016. Interaction between breathy tones and aspirated consonants in S'gaw Karen. *Proceedings of Tonal Aspects of Languages 2016*, Buffalo, 121–124.
- TABAIN, MARIJA; DAVID BRADLEY; and DEFEN YU. Central Lisu. *Journal of the International Phonetic Association* 49.129–147.
- THOMSON, RON I.; TERRANCE M. NEAREY; and TRACEY M. DERWING. 2009. A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *Journal of the Acoustical Society of America* 126.1447–1460.

- THONGKUM, THERAPHAN L. 1987a. Phonation types in Mon-Khmer languages. *UCLA working papers in phonetics* 67.29–48.
- THONGKUM, THERAPHAN L. 1987b. Another look at the register distinction in Mon. *UCLA working papers in phonetics* 67.132–165.
- THONGKUM, THERAPHAN L. 1988. Phonation types in Mon-Khmer languages. *Vocal physiology: Voice production, mechanisms and functions*, ed. by Osamu Fujimura, 319–333. New York: Raven Press.
- TIAN, JIA, and JIANJING KUANG. 2016. Revisiting the register contrast in Shanghai Chinese. *Proceedings of Tonal Aspects of Languages*, Buffalo, 147–151.
- TIAN, JIA; YIPEI ZHOU; and JIANJING KUANG. 2019. Cross-linguistic variation in the phonetic realization of “breathier voice”. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, ed. by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren, 1450–1454. Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- TRAILL, ANTHONY. 1985. *Phonetic and Phonological Studies of the !Xóõ Bushmen*. Hamburg: Buske.
- TRAILL, ANTHONY. (1986). The laryngeal sphincter as a phonatory mechanism in !Xóõ Bushman. *Variation, culture and evolution in African populations: Papers in honour of Dr. Hertha de Villiers*, ed. by Ronald Singer and John K. Lundy, 123–131. Johannesburg: Witwatersrand University Press.
- WATKINS, JUSTIN W. 1999. CQ of laryngeal gestures and settings in Wa. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1017–1020.
- WATKINS, JUSTIN W. 2001. Burmese. *Journal of the International Phonetic Association* 31.291–295.
- WATKINS, JUSTIN W. 2002. *The phonetics of Wa: Experimental phonetics, phonology, orthography and sociolinguistics*. Pacific linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- WAYLAND, RATREE, and ALLARD JONGMAN. 2003. Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics* 31.181–201.

- YANUSHEVSKAYA, IRINA; Christer GOBL; and AILBHE NÍ CHASAIDE. 2018. Cross-language differences in how voice quality and f0 contours map to affect. *Journal of the Acoustical Society of America* 144.2730–2750.
- YIU, EDWIN M. L.; BRUCE MURDOCH; KATHRYN HIRD; POLLY LAU; and ELAINE M. HO. 2008. Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals. *Folia Phoniatica et Logopaedica* 60.107–119.
- ZHANG, ZHAOYAN. 2016. Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *Journal of the Acoustical Society of America* 139.1493–1507.
- ZHU, XIAONONG. 2012. Multiregisters and four levels: A new tonal model. *Journal of Chinese Linguistics* 40.1–17.

NOTES

¹In the recent modelling work by Zhang (2016), variation in glottal opening is less important than medial vocal fold thickness in causing changes to voice acoustics. The distance between the arytenoids is important only with respect to noise generated in the glottis.

²These advantages of VoiceSauce hold for ports from Matlab to other languages, such as Octave or Python (<https://github.com/voicesauce>), but not for analyses done entirely within Praat, such as by PraatSauce (<https://github.com/kirbyj/praaatsauce>)

³But see the extended clinical VQ system of Ball, Esling, & Dickson 2018.

⁴ Esling (2005) describes the tense/lax contrast in another Yi language as a difference in vowel quality; see Kuang and Cui (2018) for discussion of ongoing sound changes in the Yi languages, and for evidence that the speakers in our sample produce a phonation contrast.

⁵ We thank Professor Jiangping Kong of Peking University for his assistance with these fieldwork trips and recordings, and for permission to use the Hani recordings made during one trip.

⁶ Local correction of harmonic amplitudes has come to replace inverse filtering as the way to minimize filter effects; see Gobl and Ní Chasaide (2012) for review. The traditional alternatives to some such kind of correction are: (1) match the vowel qualities (and thus formant frequencies) of the stimulus sets to be compared, or (2) analyze only vowels with a high first-formant frequency, using only acoustic measures below this high F1 (e.g. H1-H2 measured for low vowels).

⁷Simpson (2012) shows that H1-H2 can pattern quite differently for men's vs women's voices independently of e.g. breathiness level, because the sexes tend to differ in degree of nasality, and nasality affects harmonic amplitudes. H1*-H2* is equally problematic for this reason, since the effects of nasality are not estimated in the all-pole LPC filter model and thus not corrected for. Simpson thus cautions against using H1-H2 measures to compare the sexes.

⁸Traditional measures of voicing irregularity measured in the time domain, such as jitter and shimmer, are not included in VoiceSauce because these are not perceived independently of spectral noise; see Kreiman & Gerratt 2005.

⁹For a quite different approach to comparing phonetic categories across languages, in which failures of an automatic classifier to distinguish categories indexes their similarity, see Thomson, Nearey and Derwing (2010).

¹⁰We have developed a Shiny app for this project, available from https://pennplab.shinyapps.io/MDS_v1/, which facilitates MDS and boxplots of a dataset. Our own dataset, available from https://www.dropbox.com/s/m6bz9mhrnh5jviq/lg11_all_color.csv?dl=0, can be re-analyzed with different subsets of languages or measures; anyone can add their own new data to the dataset, and see how the results change.