A CROSS-LANGUAGE ACOUSTIC SPACE FOR
VOCALIC PHONATION DISTINCTIONS

PATRICIA KEATING  JIANJING KUANG  MARC GARELLEK

University of California, Los Angeles  University of Pennsylvania  University of California, San Diego

CHRISTINA M. ESPOSITO  SAMEER UD DOWLA KHAN

Macalester College  Reed College

Many languages use phonation types for phonemic or allophonic distinctions. This study examines the acoustic structure of the phonetic space for vowel phonations across languages. Our sample of eleven languages includes languages with contrastive modal, breathy, creaky, lax, tense, harsh, and/or pharyngealized phonations, and languages with allophonic nonmodal phonation on particular tones. In compiling and analyzing this sample we address related issues such as contrast vs. allophony, phonetic similarity across languages, and understanding complex contrasts of several multidimensional phonetic categories via data reduction. Based on extensive acoustic analysis, all of the languages' phonations were mapped into a single phonetic space, which exhibits dispersion (languages with more categories use more of the space). The space is largely two-dimensional, with dimensions that can be interpreted phonetically (e.g. dimension 2 is like a traditional breathy-to-creaky continuum) and also can be related back to the acoustic measures that structure them, thus indicating which acoustic measures are most important across languages.*

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, that is, 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 phonatory quality. Each individual speaker can manipulate their vocal folds during speech so as to produce a range of pitches and qualities; we do this for both linguistic purposes (e.g. prosody, coarticulation) and paralinguistic purposes (e.g. emotion) (Laver 1980, Garellek 2012, Yanushevskaya, Gobl, & Ní Chasaide 2018, Davidson 2021, and references therein).

Phonatory quality in turn contributes to overall voice quality, which differentiates individuals’ voices and accents. It is also the case that languages may differ in the pitch ranges and voice qualities they typically use (Esling et al. 2019). For example, studies have shown that bilinguals can use two different voice qualities when speaking their two different languages (Bruyninckx et al. 1994, Engelbert 2014) and that pairs of languages can have measurably perceptually different phonatory qualities (Yiu et al. 2008). That is, voice quality is one of the many ways in which languages can sound different from

* This work was supported by NSF grant BCS-0720304, and we thank co-PIs Abeer Alwan and Jody Kreiman for years of work together on voice. We also thank Yen Shue for VoiceSauce, Spencer Lin for data formatting, Yuan Chai for testing the Shiny app, Professor Jiangping Kong of Peking University for use of field recordings, and Sherry Yang, Joe Ptacek, Victoria Thatte, and the Hmong American Partnership for assistance with other data collection. Earlier versions of this work were presented in talks at the National Chiao Tung University in Hsinchu Taiwan, UC Berkeley, Cornell University, McGill University, University of Oregon, ICPhS-2011, the Acoustical Society of America, and Laboratory Phonology. Finally, we thank the editors and referees for helping us clarify our thinking and writing; all errors remain our responsibility.

Printed with the permission of Patricia Keating, Jianjing Kuang, Marc Garellek, Christina M. Esposito, & Sameer ud Dowla Khan. © 2023.
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 phonatory qualities in order to distinguish word meanings. Phonation 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 phonatory 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 multidimensional phonetic space for phonation, within which different languages locate their sets of phonation categories.

The main question we address in this article is thus the 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 phonation 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 multiple phonation types on vowels (rather than on consonants, or as coarticulation from adjacent consonant contrasts). Our sample includes languages with contrastive phonation types on vowels, as well as those with allophonic phonations associated with particular tones.

In addressing our main research question, we delve into other, broader, issues that arise, such as phonological contrasts with multiple phonetic correlates (including but not limited to phonation type); equating sounds across languages as members of the same phonetic category (phonetic categories are never exactly the same across languages, but how similar are they, and what properties do they share?); the usefulness of data reduction in mapping complex contrasts to a low-dimensional phonetic space, whose dimensions could map to phonological features; and phonetic dispersion of contrasting categories, both within any one dimension of contrast and in terms of the number of dimensions needed to support a given number of contrasts.

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 as those described 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, for example, a glottal stop [ʔ]. Modal voice is the phonation type that occurs most frequently in languages, producing sound efficiently by alternating complete closure with opening of the vocal folds. The other voicing categories fall between the 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 glottis\(^1\)): 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 & Ni Chasaide 2012, Esling et al. 2019). For example, two kinds of phonation with breathier voice qualities, that is, those with greater glottal opening and higher airflow, can be distinguished: whispery voice vs. breathy voice (narrowly defined) (Moisik et al. 2021), where whispery voice has more spread vocal folds but greater epilaryngeal constriction. Here we begin with a single phonation category for ‘breathy voice’, but can use our cross-language analyses to explore whether finer linguistic distinctions can be made. Subtypes of creaky voice qualities likewise are possible, as reviewed in Keating et al. 2015. Again, the linguistic relevance of such subtypes is not understood. Therefore, we again 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 in Ladefoged & Maddieson 1996 is strident (or harsh/epiglottalized) phonation, for example, in languages within the ‘Khoisan’ group. In addition to changes in vocal-fold vibration, characterized by Esling as whispey voice due to vocal-fold spreading, this phonation is also said to involve epilaryngeal constriction, aryepiglottic trilling, pharyngeal narrowing, and tongue retraction (Traill 1986, Hess 1998, Esling 2005, Edmondson & Esling 2006, Miller 2007, Moisik et al. 2021). Below we describe a further distinction made between ‘harsh’ and ‘pharyngealized’ phonations.

In their survey article on phonation contrasts (for both vowels and consonants) across languages, Gordon and Ladefoged (2001) use the continuum of glottal opening to describe and classify phonation types in a wide variety of languages, and they suggest measures of the glottal source and filter that can be used to quantify differences across phonation categories. There have since been many 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 sixty-one languages that have phonation contrasts on vowels, including all of the languages with contrasts in the present study. (The list of languages shown here is given in the appendix, with further information and references.) This sample is not exhaustive, since we do not include all known examples from within a given subfamily; for example, we include only five of the many Zapotec languages with contrastive creaky voice on vowels.

Languages with nonmodal phonation on vowels come from a wide variety of families, but 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 García 2018, Kelterer & Schluppler 2020). Nilotic languages tend to contrast breathy vs. modal vowels,

\(^1\) In the recent modeling 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.
though some languages in the family 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 (Traill 1985, Miller 2007). For example, !Xóõ has two phonation types involving pharyngealization: ‘pharyngealized’ vowels are characterized phonetically as having both pharyngealization (higher F1) and creaky voice (glottal constriction, irregular f0) throughout; ‘harsh’ (or ‘strident’) vowels are also pharyngealized but have breathy voice initially during the vowel. But by the middle of the vowel, harsh vowels are characterized by strong glottal and epilaryngeal constriction, resulting in a very creaky and pharyngealized quality (Garellek 2020).

The high occurrence of vowel phonation contrasts among certain Sino-Tibetan, Hmong-Mien, and Austroasiatic 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, Garellek et al. 2021, Tạ et al. 2022); these kinds of contrasts are defined below.

Outside of 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, such as Udihe (Tungusic, Russia). Gujarati (Indo-European, India), however, contrasts breathy vs. modal vowels and has no contrastive creak.

Finally, languages with phonation contrasts on vowels tend to also have tonal contrasts. In our sample, only 25% of the languages (mostly Austroasiatic) lack lexical tone; see the list in the appendix. Many of these have 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 nonmodal phonation can serve as a primary or secondary cue to a particular register (Brunelle & Finkeldey 2011, Brunelle 2012, Tạ et al. 2022), as is argued for (White) Hmong (Hmong-Mien; Southeast Asia) and (Black) Miao (Hmong-Mien; China) (Garellek et al. 2013, Kuang 2013). For instance, in Takhirian Thong Chong two of the four 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). See also Esposito & Khan 2020 for discussion of mixed tone-phonation contrasts; other examples of phonation contrasts can also be found in Esposito & Khan 2020 and in Garellek et al. 2021.

In the present study, we draw from this sample a small but relatively diverse set in order to explore the acoustic differences among phonation categories.

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 and 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, the harmonic nearest 5000 Hz is not shown. The frequency of the first harmonic (H1) is the fundamental frequency ($f_0$).

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 in 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 slope 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 subranges of frequencies as well, especially at the low-frequency end of the spectrum, where the amplitude of the first harmonic is very important. Sundberg and Gauffin (1979) found that the absolute amplitude of the first harmonic reflects the amount of air flowing through the glottis (‘[t]he greater the maximum airflow amplitude in the flow glottogram, the stronger the voice source fundamental’). A less constricted glottis during voicing means greater airflow and a stronger H1. At the same time, less glottal constriction results in a more sinusoidal vibratory pulse, which has relatively more energy in the fundamental than in higher harmonics. Thus with a more open glottis, H1 both absolutely and relatively has a higher amplitude. The higher-frequency part of the spectrum, in contrast, is thought to depend more on the abruptness of the 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, 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, Kreiman et al. 2021) 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. In terms of articulation, harmonic source spectral-tilt measures correlate with glottal open/contact quotient and medial vocal-fold thickness (for a recent overview of the relationship between articulation and commonly used acoustic measures of phonation, see Garellek 2019.)

A source spectrum may also include inharmonic components. Both turbulent flow in the glottis (essentially, aspiration) and aperiodic vibration of the folds (e.g. jitter or shimmer) add noise in the sense of reducing the signal-to-noise ratio (Kreiman & Gerratt 2005). Noise in the source can also 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 fully established how the noise spectrum should be modeled and measured (though see Kreiman et al. 2021 for one model of the noise spectrum). 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 has 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

2 These 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/pratausace).
The proliferation of acoustic measures, especially those that can be made easily and automatically by VoiceSauce, Praat, or other programs, can be confusing. Brunelle and 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. Some researchers have directly compared measures with regard to 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 (CPP), a harmonics-to-noise ratio measure. Using automatic measures provided by VoiceSauce, Esposito (2012) found that CPP, H1*, H2*, and H1*−H2* all show statistically significant differences between phonation types in Hmong (see §2.3 for more on these measures). Nonetheless, in our study we include almost all of the measures provided by VoiceSauce; we can then 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. How is the acoustic space of phonation types structured across languages? One possible answer is that the space is composed 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 proposes four frequency bands for harmonic energy differences, plus glottal noise (measured in four frequency bands in Kreiman et al. 2021, thus suggesting eight dimensions total). 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 vocalic phonation types across languages, with the expectation that fewer than five (or eight) dimensions are likely to suffice.

However many dimensions are involved, we can then ask which acoustic properties are most crucial in defining this phonetic space. As noted above, there is a pressing need to better understand the relative importance of the many acoustic parameters that have been shown to vary across phonation types. Analyzing the various phonation categories of several languages simultaneously is one approach to this important issue.

Once we have an overall acoustic space for phonation types across languages, we can next 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 phonatory 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, consonant, and tone spaces (Lindblom & Maddieson 1988, Schwartz et al. 1997a, Becker-Kristal 2010, Kuang 2013). Here we investigate whether the less common contrasts of phonation types pattern like
more-familiar contrasts, thus helping to clarify what sorts of contrasts are subject to dispersion principles.

Finally, we can ask how similar the ‘same’ categories or contrasts are across languages. The IPA distinguishes only three categories of phonation: modal, breathy, and creaky voice, with the latter two marked by diacritics. 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 phonation types.) 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 ten 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 (though see Kim et al. 2020 for similarities in creak realization in Mandarin, Cantonese, and Hmong).

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 voice onset time (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. Our study thus contributes to understanding phonetic similarity vs. uniqueness across languages.


2.1. Languages and phonation categories. In this study we compare the contrasting or allophonic phonations of eleven languages from five language families across different continents, many of which are understudied. In part the choice of languages was opportunistic, based on the availability of suitable audio recordings. Other languages were then selected so that the sample would better represent the typology of phonation contrasts described above, as follows.

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). We wanted our sample to include languages with modal/breathy/creaky contrasts, as well as languages with tense/lax contrasts. In addition, to ensure that noisy phonation types are represented, one of the languages in our sample is !Xôô, with harsh (‘strident’) and pharyngealized voice qualities (Traill 1985, 1986, Ladefoged & Maddieson 1996:311–12, Garellek 2020). Thus the sample includes examples of seven phonation categories. We also included two languages with no phonation contrast.

3 But see the extended clinical VQ system of Ball, Esling, & Dickson 2018.
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 prosodic, paralinguistic, and/or sociolinguistic functions (Podesva & Callier 2015, Crowhurst 2018), at least in the simple utterances analyzed here. Breathy 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).\textsuperscript{4} In Mazatec and !Xóõ, a phonologically 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 most of these languages are included in the supplementary materials.\textsuperscript{5}

<table>
<thead>
<tr>
<th>LANGUAGE (Abbrev.)</th>
<th>PHONATION CATEGORIES</th>
<th># SPEAKERS (F, M)</th>
<th># REPS</th>
<th># TOKENS BY GENDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bo (Bo) Sino-Tibetan</td>
<td>Contrastive: lax, tense</td>
<td>10 (4, 6)</td>
<td>2</td>
<td>M (117)</td>
</tr>
<tr>
<td>English (En) Indo-European (eng-US)</td>
<td>Modal</td>
<td>18 (11, 7)</td>
<td>3</td>
<td>M (42)</td>
</tr>
<tr>
<td>Gujarati (Gj) Indo-European (guj)</td>
<td>Contrastive: breathy, modal</td>
<td>10 (7, 3)</td>
<td>3–6</td>
<td>M (344)</td>
</tr>
<tr>
<td>(White) Hmong (Hm) Hmong-Mien (mww)</td>
<td>Contrastive: breathy, modal</td>
<td>23 (8, 15)</td>
<td>1</td>
<td>M (474)</td>
</tr>
<tr>
<td>(Luchun) Hani (Lu) Sino-Tibetan (lni)</td>
<td>Contrastive: lax, tense</td>
<td>9 (4, 5)</td>
<td>2</td>
<td>M (351)</td>
</tr>
<tr>
<td>(Black) Miao (Mi) Hmong-Mien (hea)</td>
<td>Contrastive: breathy, modal</td>
<td>8 (0, 8)</td>
<td>2</td>
<td>M (350)</td>
</tr>
<tr>
<td>Mandarin (Mn) Sino-Tibetan (cmn)</td>
<td>Modal</td>
<td>31 (16, 15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Jalapa) Mazatec (Mz) Otomanguean (maj)</td>
<td>Allophonic: creaky</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>!Xóõ (Xo) Tiu (nmn)</td>
<td>Contrastive: breathy, creaky, harsh, modal, pharyngealized</td>
<td>12 (0, 12)</td>
<td>1</td>
<td>M (306)</td>
</tr>
<tr>
<td>(Southern) Yi (Yi) Sino-Tibetan (nsv)</td>
<td>Contrastive: lax, tense</td>
<td>12 (6, 6)</td>
<td>2</td>
<td>M (122)</td>
</tr>
<tr>
<td>(Valley) Zapotec (Zp) Otomanguean (zab)</td>
<td>Contrastive: breathy, creaky, modal</td>
<td>6 (2, 4)</td>
<td>2</td>
<td>M (166)</td>
</tr>
</tbody>
</table>

Table 1. Languages, phonation categories, speakers, and numbers of tokens in this study. #REPS: number of repetitions of each word by each speaker.

It would be possible to limit this sample to phonation categories that are clearly contrastive in each language, but instead we also include some cases (labeled ‘allophonic’) in which the phonation property seems to be a secondary correlate of a contrast,

\textsuperscript{4} Esling (2005) describes the tense/lax contrast in another Yi language as a difference in vowel quality; see Kuang & 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.

\textsuperscript{5} The supplementary materials referenced here and elsewhere can be accessed at http://muse.jhu.edu/resolve/188.
predictable from the primary correlate. We include these cases because the phonation appears fairly consistently and is clearly audible (and even salient to native speakers), and because the weighting of a primary vs. a secondary correlate of a contrast is gradient—in one language a secondary correlate might be quite minor, but in another language it might be almost as informative as the primary correlate. (See Kuang & Cui 2018 and studies by Abramson and colleagues cited elsewhere here for cases of shifting cue weights in register contrasts, and see Brunelle & Kirby 2016 for general discussion of multiple correlates of register and tone contrasts.) That is, it is not always clear when a phonetic feature or correlate is being used contrastively in a language, so our choice of cases was liberal.

Nonetheless, we label each case as either contrastive or allophonic, so here we discuss our use of these labels. In many 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. In such cases it is clear that all of the phonation differences are contrastive, since they cannot be predicted from the tone or pitch.

Furthermore, if a language has two phonations that occur on the same pitch level or pitch contour, or one phonation that occurs on two different pitch levels or contours, then the phonations likewise are contrastive, since again they cannot be predicted from the pitch. Our examples of this are thebreathy phonations of Miao and Hmong (in Miao, there are modal and breathy mid tones; in Hmong there are modal and breathy high-falling tones), and the creaky phonation of Zapotec (there are creaky low-falling and high-falling tones).

The four phonation instances classified as allophonic in Table 1 involve covariation with pitch as part of a tonal category. 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. For example, this is known to be the case in English (e.g. Kuang 2013). As a result, tones with extra-low or extra-high pitch often have a concomitant phonatory quality, predictable from the pitch. A well-known example is that in Mandarin the low-pitched portions of tone 2 (pitch contour 213) and tone 4 (pitch contour 51) are usually creaky, though occasionally breathy (Kuang 2017:1700, Moisik et al. 2014). Since this creaky voice is fairly consistent, we code these two Mandarin tones as a separate creaky category; and since it is predictable from the low pitch, we consider it to be allophonic. Similarly, in White Hmong (Esposito 2012, Garellek 2012) and Black Miao (Kuang 2013) the lowest-pitched tone is usually creaky (though in Xinzhai Hmu (Miao) the low tone is described as breathy; Liu et al. 2020), and these low tones also are listed as instances of allophonic creak. Whether one, the other, or both of two correlates carry a contrast for listeners is an empirical question that must be answered separately for each language. It is noteworthy in this regard that 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; this accords with our classification of that language’s breathy phonation as contrastive, but its creaky phonation as allophonic.

Finally, in Black Miao, a language with five level tones (Kwan 1966, Kuang 2013), the pitch range is expanded on the high end as well as the low end; the highest tone, with an extra-high pitch, has a tense phonatory quality. We consider this to be an instance of allophonic tense phonation. Thus Black Miao, with its large number of level tones and
its expanded pitch range, is treated as using allophonic creaky voice for its lowest tone and allophonic tense voice for its highest tone, in addition to contrastive breathy and modal voice on its mid tones.

2.2. Word lists 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 supplementary materials include complete word lists for each language sample.) 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 languages without phonation contrasts, English and Mandarin, did not follow these constraints, because available corpora used voiceless fricative and nasal onset consonants. These corpora consisted of one or two CV sequences rather than many different words: for English, the word *sure*, and for Mandarin, minimal sets formed with [ma] and [ʂɘ] with the four tones. This second Mandarin vowel has a first formant frequency well above the fundamental frequency, with widely spaced F1/F2/F3.

For the tone languages, all of their level tones, plus some or all of their contour tones, were included in the word lists, 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 the languages other than Mazatec and !Xóõ, speakers were recorded by the authors using either a computer soundcard or PCQuirer (Scicon Research & Development Inc.) with its external D/A box, all at a sampling rate of 22 kHz, 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.6 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 sound booth. Our Mandarin corpus combines recordings made in Beijing and in our laboratory sound booth. All of the original recordings are available at the project website, http://www.phonetics.ucla.edu/voiceproject/voice.html. 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. In this kind of sample, differences in recording conditions are unavoidably confounded with language differences. We mitigate this confound by standardizing the resulting measurements by speaker, rather than analyzing the raw data.

Table 1 above also shows the number of speakers analyzed for each language, along with the number of times target words were repeated by speakers, and the total number of tokens from men and women speakers. While nine languages have both men and women speakers, two languages have only men, and overall it was not possible to evenly balance across gender. In general, we expect women’s voices to be breathier than men’s voices, but specifically with respect to the phonetic realization of phonation categories, large gender effects are not expected for these languages. This is because

---

6 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.
previous studies have found no gender effects on phonation in Mazatec, Gujarati, and Hmong (Garellek & Keating 2011, Khan 2012, and Esposito 2012, respectively), while in our Yi languages (Bo, Hani, Southern Yi) gender differences are seen only in secondary aspects of the physiological electroglottographic signal (Kuang & Keating 2014), and Kuang (2011) found no acoustic effects in Southern Yi. Notably, however, gender differences have been found in Otomanguean languages other than Mazatec: by Esposito (2010b) in Santa Ana del Valle Zapotec, and by Kelterer and Schuppler (2020) in Chichimec; but we do not have a large enough sample here to know if those differences hold across our Zapotec speakers. Thus the general expectation is that gender differences will be ones of scaling (with women breathier overall), and this kind of difference is controlled for by standardizing measurements within speakers and comparing only these standardized measures across speakers. For this reason, we do not think that the different numbers of men vs. women for the different languages in our sample will account for apparent differences between languages.

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 used 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 to produce them as many times as possible within a ten-second window.

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

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. VoiceSauce is then 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 one-millisecond (ms) 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. (Note that $f_0$ was calculated by VoiceSauce 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 preliminary analysis revealed that, not surprisingly, it modeled tone categories instead of phonation categories.) For voiceless intervals, which can sometimes occur in nonmodal vowels with particularly strong constriction or aspiration, VoiceSauce outputs no $f_0$ or harmonic amplitudes.

VoiceSauce then uses the Snack Sound toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at one-millisecond 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 filtering effects of the vocal tract, using Iseli et al.’s 2007 extension
of Hanson’s 1995 method.\(^7\) (In uncorrected measurements, the formant frequencies boost the amplitudes of any nearby harmonics, and since different vowel qualities have different formant frequencies, this very large effect of the vocal-tract filter can dwarf any harmonic amplitude differences in the voice source.) Corrected harmonic amplitudes are indicated by an asterisk, for example, H1*.\(^8\) There may also be interactions between vowel quality and phonation type on harmonic amplitudes, even for corrected measures; Esposito et al. (2019) found that vowels with higher formant values were associated with lower H1* – H2*. Overall, however, these corrections allow us to combine data from different vowel qualities.

In the subsequent analysis, we also include measures of the individual harmonic amplitudes, because these can also be associated with differences across phonation types; for example, Esposito (2012) found that H1* was better than H1* – H2* at distinguishing modal and creaky phonation types. Chai and Garellek (2022) describe a different way to normalize H1 measurements, but that is not included in our data set.

Other VoiceSauce parameters index periodicity and/or noise using cepstral analysis.\(^9\) 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’s (1993) 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 ms and the maximum quefrency; VoiceSauce uses Hillenbrand et al.’s algorithm. Sun’s (2002) 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 is known to occur in creaky voice and with broader laryngeal constriction. (See Herbst 2021 for concerns about this parameter.)

Finally, pitch-synchronous energy and strength of excitation (Murty & Yegnanarayana 2008) are calculated. Strength of excitation (SoE; 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. With greater glottal constriction, voicing weakens. Thus, Garellek 2020 found

\(^7\) Local correction of harmonic amplitudes has come to replace inverse filtering as the way to minimize filter effects; see Gobl & Ni Chasaide 2012 for review. The traditional alternatives to some such correction are: (i) match the vowel qualities (and thus formant frequencies) of the stimulus sets to be compared, or (ii) 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). Even with harmonic amplitude corrections, we do try to avoid high vowels in our sample; in automatic formant estimation, the very low F1 of most high vowels may not be correctly separated from the fundamental for women speakers, and corrections made with faulty formant estimates will give faulty harmonic amplitude estimates.

\(^8\) Simpson (2012) shows that H1–H2 can pattern quite differently for men’s vs. women’s voices independently of, for example, 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.

\(^9\) Traditional 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.
that the constricted phonation types in !Xóõ (creaky, pharyngealized, and harsh) are characterized by lower SoE relative to phonation types (like breathy and modal) with little constriction.

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. We note that languages with phonation contrasts can have distinctive timing patterns, with nonmodal phonations occurring on only portions of the vowels (Silverman 1997, Blankenship 2002, Garellek et al. 2021). In this article, we focus on the phonation types that occur in the middle third of the vowel, since preliminary analysis showed that this interval was most informative across the languages, presumably because it avoids coarticulatory effects from neighboring consonants, and because nonmodal phonation is generally strong over this interval (even if it is stronger earlier or later).

All values from VoiceSauce were then z-score standardized within speaker. Standardizing first within speaker avoids modeling speaker-specific (idiosyncratic) or gender-specific patterns instead of language × phonation categories. The mean standardized values (averaging across individual tokens) of the middle-third interval of each acoustic measurement were then calculated for each language × phonation category, combining speakers. As will be seen below, one of our analyses uses this reduced language × phonation data set, while the other analysis uses the individual data tokens. The data file included in the supplementary materials comprises the individual raw data tokens, not the language × phonation category means.

2.4. Statistical analysis: multidimensional scaling and classification and regression trees. The multiple acoustic measures made in this study can be thought of as defining a multidimensional acoustic space within which different phonations can be located. However, many of these measures are intercorrelated, 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. MULTIDIMENSIONAL SCALING (MDS; Kruskal & Wish 1978) is one method for reducing many individual measures to a smaller number of independent dimensions.10 MDS is a class of analysis that aims at reducing distances between categories in a multidimensional space, with the specific goal of mapping distances or dissimilarities between categories along a smaller number of independent dimensions. MDS is particularly suited to dealing with data with strong multicollinearity, as is found among our voice measures. And, because MDS does not assume linearity, it is preferred over PRINCIPAL COMPONENTS ANALYSIS, another method of data reduction. However, MDS analysis of hundreds or thousands of data points, as we have, does not produce meaningful results. Therefore this analysis uses the language × phonation means because we are interested in the distances between the phonation categories across the languages, not in the individual speakers or tokens.

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.

10 For a quite different approach to comparing phonetic categories across languages, in which failures of an automatic classifier to distinguish categories index their similarity, see Thomson, Nearey, & Derwing 2009.
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 × phonation categories, and these distances were input into Kruskal’s NONMETRIC MULTIDIMENSIONAL SCALING algorithm (performed in R using the ‘metaMDS’ function in the ‘vegan’ package). Nonmetric MDS is used for non-Euclidean distances, as we have here. Our data file and R scripts for MDS and CLASSIFICATION AND REGRESSION TREE (CART) analyses (Breiman et al. 1984) are included in the supplementary materials for this article; in addition, our data file can be reanalyzed directly using a Shiny app.\footnote{The Shiny app for this project is available at https://pennplab.shinyapps.io/MDS_v1/. This app opens our data file, allowing online boxplot and MDS analysis of any of that file’s variables and also allowing new data (formatted appropriately) to be added to ours, for new analyses. Our data file and some R scripts are also included in the supplementary materials for this article for inspection and use outside of the Shiny app. The supplementary materials also include a key to abbreviations used in the data file (along with word lists and audio examples).}

The algorithm 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 the fit of the reproduced space. A scree plot (Figure 3) was plotted 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 Fig. 3, a two-dimensional space is able to sufficiently reproduce the original distance matrix. However, the third dimension adds 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.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{stress_dim.png}
\caption{Scree plot for the MDS solution.}
\end{figure}
In order to have a better understanding of how phonation categories are classified by voice parameters, taking advantage of measurements of all individual tokens, a CART analysis (Breiman et al. 1984) was fitted to the full data set of standardized measures from the vowel middle thirds, using the ‘rpart’ package in R, complementing our MDS analysis. The depth of the tree was determined by a complexity parameter, which was optimized through a 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 kept at the default, restrictive, value of 0.01.

3. Results.

3.1. Two-dimensional space. The MDS two-dimensional (2-D) 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 data point 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 in shades of blue (creaky: dark blue, tense: light blue), modal phonation in green, and the other two phonations in purple (pharyngealized) and pink (harsh).

First, we consider the overall organization of this space. !Xóõ, the language with the most phonation contrasts in this data set, defines the extreme edges of the space: each dimension makes a two- or three-way distinction in !Xóõ, with its modal (green) and harsh (pink) 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 !Xóô. 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 continuum of modal (on the left) to nonmodal (on the right): 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.

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 data point 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 Fig. 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 !Xóô 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 is most prominent during the initial third of the vowel, not during the middle third that was analyzed here (Garellek & Keating 2011).

Otherwise, the obvious outliers are the Gujarati breathy (dark orange) and the Zapotec creaky (dark blue). There is greater similarity between the Gujarati breathy and the lax (light orange) phonation in other languages, and greater similarity between the Zapotec creaky and tense (light blue) phonations.

It is entirely possible that these characterizations are phonetically accurate: the Gujarati modal-breathy contrast could perhaps be characterized as a modal-lax contrast, while the Zapotec modal-creaky contrast could perhaps be characterized in part as a modal-tense contrast. Indeed, the relative distances between these categories in this 2-D space are fairly short. In the Zapotec case, this could be because the creaky category includes some creaky high tones, which are likely phonetically tense rather than creaky.

Fourth, we can compare the eleven languages in terms of how well their phonation categories are separated in the space. Visually, we can think of this as connecting all of
the data points for a language, making a virtual triangle/rectangle and so forth, in much the same way as vowel spaces are described. How big is this language-specific space? We might expect a dispersion factor (e.g. Lindblom & Maddieson 1988, Schwartz et al. 1997b) 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 Fig. 4. As already noted, !Xôô (brown points) is almost optimal, distinguishing four of its categories in a big square that uses both dimensions of the space. (The remaining !Xôô category distinction is discussed below.)

Fig. 5. Two-dimensional MDS solution for the twenty-nine language × phonation categories from the eleven languages, each category a set of acoustic measures from the middle third of all vowels. Same as Fig. 4 except that here colors code languages rather than phonation categories. Color coding: brown: !Xôô, red: Bo, black: English, yellow: Gujarati, light green: Hmong, blue: Luchun Hani, fuchsia: Mandarin, gray: Mazatec, purple: Miao, dark green: Yi, turquoise: Zapotec.

Two languages have three or four categories (allophonic or contrastive), which visually form virtual 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.

Two languages have three categories falling on a virtual 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 (gray) 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. By contrast, 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.** Pairwise two-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 × phonation codes are the same as in Fig. 4.

What about !Xóõ’s fifth category? Pharyngealized and creaky are not distinguished in the 2-D space. It was noted above that dimension 3 of the MDS solution does add some additional information beyond what the first two dimensions provide; the three dimensions together distinguish all five !Xóõ categories. Figure 6 shows the two pairwise 2-D spaces involving this third dimension 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 data points, which !Xóõ pharyngealized is close to. This may reflect the different phasings of creaky voice during !Xóõ creaky (with creaky voice strongest at the end of the vowel) vs. !Xóõ pharyngealized (with creaky voice strongest in the middle of the vowel). Our samples were taken during the middle third interval of the vowel, during which the creaky voice for !Xóõ pharyngealized vowels is strongest. This could result in !Xóõ pharyngealized’s placement among the other creaky data points. Dimension 3 also mildly enhances the difference between tense (light blue) and lax (light orange) data points 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 twenty-one acoustic measures from VoiceSauce relate to each of the three dimensions. Dimension 1, which looks like a nonmodal-to-modal continuum, is most strongly based on measures of periodicity and energy in the excitation or the speech signal—SHR, SoE, 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 largely complementary to those for dimension 1: unsurprisingly, $H_1^* - H_2^*$ and $H_1^* - A_1^*$, which are measures of low-frequency spectral tilt, 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. nonmodal distinction, like dimension 1 above.

<table>
<thead>
<tr>
<th>ACoustic MEASURE</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1^* - H_2^*$</td>
<td>0.07</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>$H_2^* - H_4^*$</td>
<td>1.49</td>
<td>0.09</td>
<td>0.38</td>
</tr>
<tr>
<td>$H_1^* - A_1^*$</td>
<td>1.29</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>$H_1^* - A_2^*$</td>
<td>1.15</td>
<td>1.70</td>
<td>0.14</td>
</tr>
<tr>
<td>$H_1^* - A_3^*$</td>
<td>1.62</td>
<td>1.31</td>
<td>0.36</td>
</tr>
<tr>
<td>$H_4^* - H_2kHz^*$</td>
<td>2.24</td>
<td></td>
<td>0.36</td>
</tr>
<tr>
<td>$H_2kHz^* - H_5kHz^*$</td>
<td>0.63</td>
<td>0.92</td>
<td>0.36</td>
</tr>
<tr>
<td>$H_1^*$</td>
<td>1.01</td>
<td>1.65</td>
<td>0.28</td>
</tr>
<tr>
<td>$H_2^*$</td>
<td>0.12</td>
<td>0.44</td>
<td>0.59</td>
</tr>
<tr>
<td>$H_4^*$</td>
<td>0.10</td>
<td>0.31</td>
<td>0.50</td>
</tr>
<tr>
<td>$A_1^*$</td>
<td>0.49</td>
<td>0.57</td>
<td>0.28</td>
</tr>
<tr>
<td>$A_2^*$</td>
<td>0.40</td>
<td>0.19</td>
<td>0.43</td>
</tr>
<tr>
<td>$A_3^*$</td>
<td>0.49</td>
<td>0.14</td>
<td>0.80</td>
</tr>
<tr>
<td>CPP</td>
<td>1.58</td>
<td>1.13</td>
<td>0.11</td>
</tr>
<tr>
<td>HNR &lt; 500 Hz</td>
<td>1.43</td>
<td>1.44</td>
<td>0.51</td>
</tr>
<tr>
<td>HNR &lt; 1500 Hz</td>
<td>1.40</td>
<td>1.27</td>
<td>0.91</td>
</tr>
<tr>
<td>HNR &lt; 2500 Hz</td>
<td>1.57</td>
<td>1.46</td>
<td>0.93</td>
</tr>
<tr>
<td>HNR &lt; 3500 Hz</td>
<td>1.41</td>
<td>1.61</td>
<td>0.92</td>
</tr>
<tr>
<td>Strength of excitation (SoE)</td>
<td>2.34</td>
<td>0.72</td>
<td>0.39</td>
</tr>
<tr>
<td>SHR</td>
<td>2.90</td>
<td>2.28</td>
<td>1.69</td>
</tr>
<tr>
<td>Root mean squared (RMS) Energy</td>
<td>2.08</td>
<td>0.29</td>
<td>0.87</td>
</tr>
</tbody>
</table>

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

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 ǃXoõ creaky is distant from the other creaky categories while ǃXoõ pharyngealized is near them, suggests that most creaky categories and ǃXoõ pharyngealized show period doubling, while ǃXoõ 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.

3.2. How subsets of languages use the space. From the solution for the eleven 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 rerunning the MDS analysis on a subset. Figure 7 shows the 2-D spaces that result when each of the eleven languages is left out in turn. That is, each MDS solution here is for ten of the eleven languages.
A cross-language acoustic space for vocalic phonation distinctions

We have already seen that !Xöö is the main determinant of the structure of the eleven-language space. When that language is left out (bottom left graph in Fig. 7), 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 eleven-language space, but also in all of the ten-language subsets. However, in many spaces derived from smaller subsets (not seen in Fig. 7), English appears in odd places—in a corner with other contrastive modal categories, or with other categories entirely.

When MDS is run only on the languages with tense-lax contrasts, the resulting space (shown in the lower right corner in Fig. 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 eleven 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 phonation space. A complementary analysis that focuses on the relative importance of the various acoustic measures, and which uses all of our individual data points rather than means, comes
from a CART analysis (Breiman et al. 1984). The classification tree obtained from the rpart package in R is shown in Figure 8. The classification tree is then paired with box-plots of distributions of values on each measure that appears in the tree, here separated by the twenty-nine language × phonation categories. These are shown in Figures 9a–c.

Figure 8. Classification tree of phonation categories derived from the full set of acoustic measures for all individual tokens in the data set labeled as modal, breathy, or creaky. Abbreviations used in this figure are: HNR05_means002: harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz for the middle third of each vowel; SHR_means002: subharmonic-to-harmonic ratio for the middle third of each vowel; H1H2c_means002: H1* − H2* for the middle third of each vowel; B: breathy, M: modal, and C: creaky phonation categories. See the text for a more detailed explanation of the figure. The colors here are arbitrary, unrelated to those in previous figures.

Figure 8 shows how different acoustic parameters are able to divide the entire data set into the phonation categories that label the individual tokens. The top of the figure shows that the most informative acoustic parameter for dividing the data set in two is ‘HNR05’, which is the VoiceSauce abbreviation for the harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz. The ‘means002’ portion of the parameter label indicates that this measure was calculated over the middle third of each vowel. The split in the tree shows that the tokens were divided into two sets: one set, with values above 0.48, which consisted only of modal tokens (61% of the data set), and the other a mixed set, with values below 0.48, which included tokens with all three kinds of phonations. That the clearest split among the tokens is between most of the modal tokens and the nonmodal tokens is reminiscent of dimension 1 in the MDS analysis, which provided a modal-to-nonmodal continuum. Figure 9a shows the distributions of values on this measure across language × phonation categories. It can be seen that modal categories generally have higher values on this measure, though there is a spread, and that it seems to best distinguish modal from creaky categories.

In Fig. 8, the mixed set that resulted from the first split in the data is divided into two subsets, with the most informative acoustic parameter being ‘SHR’, the VoiceSauce
abbreviation for subharmonic-to-harmonic ratio, again calculated over the middle third of each vowel. This parameter divided the tokens into a breathy set with values under 0.0024 (11% of the data set) vs. a mixed set of creaky and modal tokens with values equal to or greater than 0.0024. As was seen in Table 2, SHR 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 subharmonic energy), in contrast to breathy voice, which has very little subharmonic (or even harmonic) energy.

Finally, the mixed set that resulted from the second split is further divided into two subsets, with the most informative acoustic parameter being ‘H1H2c’, the VoiceSauce abbreviation for $H_1^* - H_2^*$, again calculated over the middle third of each vowel. This parameter divided the tokens into creaky tokens with values less than 0.45 (15% of the data set) vs. modal tokens with values greater than or equal to 0.45 (12% of the data set). 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 Fig. 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. These other categories would be shown only if we went beyond the optimized tree depth; thus the CART results suggest that there are four major clusters of phonation categories, including two kinds of modal. The distributions of our seven phonation categories across these four clusters are shown in the list of numbers below each terminal category label; these represent the proportion contributed by each of the seven categories, in alphabetical order—thus, for example, the leftmost terminal, labeled ‘B’, is composed of 44% breathy tokens, 9% creaky tokens, 2% harsh tokens, 16% lax tokens, 24% modal tokens, 0% pharyngealized tokens, and 6% tense tokens. As noted above, these tokens together comprise 11% of the total data set.

![Figure 9c. H1* – H2* z-scores, middle third of vowels. Colors and language × phonation codings as in Fig. 4 above.](image)

It can be seen in Figs. 9a–c that some phonation × language categories are much more variable (across speakers and tokens) than others on a given measure. Most strikingly, Gujarati in all three panels, Hmong modal in Fig. 9b, and Hmong breathy and modal in Fig. 9c show many outlier data points. Our best guess as to why is that these are the two
languages where the test words were recorded in sentences, which could well contribute more variability. In contrast, there is no apparent pattern in variability based on contrastive vs. allophonic phonation, or number of contrastive categories. Notably, the English modal shows fairly tight distributions, even though with no contrast it is free to cover a range of phonations, as indeed the Mandarin modal does.

4. Discussion and Conclusions.
4.1. Dimensions of the Space. The analyses presented above indicate that the space for the phonations of our sample of eleven languages is largely two-dimensional, as presented in Fig. 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. We note that a separate MDS analysis of just the male speakers from each language (not presented here) shows a very similar result, with slightly better clustering of the phonation categories, but the same overall organization.

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 nonmodal 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 nonmodal edge of the modal categories, and breathy similar to tense and lax.

The second dimension of the space, by contrast, 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 does not use this dimension for at least some phonemic contrast, which is not the case for dimension 1.

Nonetheless, !Xôô’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. !Xôô 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-D 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 a single dimension (dimension 2) and are not very separated even along that one. The English noncontrastive 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/nonmodal 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 phonation 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 subtypes. We noted in the introduction that subtypes of both creaky and breathy phonations are attested, but the linguistic significance of these subtypes 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 nonmodal phonation matters.

Recall that we classified four categories as allophonic, as they are related to tone contrasts: creaky in Mandarin (with no phonation contrasts) and in Hmong and Miao (with contrastive breathy and modal), plus tense in Miao (also with contrastive breathy and modal). The allophonic tense category of Miao patterns like the contrastive tense categories of the three Yi languages. However, the results for the creaky voice categories are different. The three tone-related allophonic creaky cases (Mandarin, Hmong, Miao) do pattern differently from the contrastive cases (Mazatec and !Xóõ). These 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. (This difference cannot be due to Mandarin creaky being mostly modal, since if anything Mandarin tone 3 and tone 4, especially for Beijing speakers, have more creak mid-vowel than do the Miao and Hmong low tones.) 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, even though both of these languages locate the creak more at vowel edge. As discussed in the results (§3 above), all five creaky categories lie in a separate part of the overall space (in the upper right corner), but we can speculate that allophonic (or tone-based) vs. contrastive status (Mandarin, Hmong, Miao vs. Mazatec and !Xóõ) might 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 (2020) study of the role of individual phonation cues in Mandarin tone perception. Huang found that Mandarin perception depends on the low \( f_0 \) and/or irregular phonation properties of creak, but not on its harmonic spectrum properties. In our analysis, irregular phonation (periodicity) is most strongly related to dimension 1, while the harmonic spectrum is related more to dimension 2; that is, the information about Mandarin creak is in dimension 1, not dimension 2. In other words, production and perception are aligned in Mandarin; speakers use the acoustic cues that listeners attend to.

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 in !Xóõ is whispery because the vocal folds are more spread while the epilarynx is constricted (as shown in Traill 1986 and argued recently by Garellek 2020), 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. Data reduction, followed by correlations of acoustic parameters with the dimensions of the reduced space, shows which acoustic measures are more important in defining the low-dimensional phonetic space. We found that for dimension 1 (roughly modal to nonmodal), the most important measures were SHR, SoE, and Energy; for dimension 2 (roughly glottal constriction), \(H_1^* - H_2^*\), \(H_1^* - A_1^*\), and again SHR; for dimension 3 (a different modal-to-nonmodal spread), again SHR. By contrast, the splits on the classification tree (where all individual data points are used) are made first by HNR 0–500 Hz (for modal vs. others), SHR (for breathy vs. others), and \(H_1^* - H_2^*\) (for creaky). SHR and \(H_1^* - H_2^*\) are thus the parameters that are important in both analyses, and HNR 0–500 Hz is most important in the classification tree analysis, where it seems to take the place of the SHR of MDS dimension 1. These three parameters reflect different aspects of the voice source spectrum, and therefore different aspects of phonation. Higher \(H_1^* - H_2^*\) is correlated with decreased vocal-fold contact during voicing (DiCanio 2009) as well as greater medial fold thickness (Zhang 2016). SHR measures the strength of subharmonics in the spectrum, and thus higher values index irregular vocal-fold vibration (Sun 2002, Keating et al. 2015). A lower HNR is correlated with increased aspiration and, especially in the lower frequencies as here, voicing irregularity. Of the other measures here, SoE is a measure of the strength of voicing; increased constriction at the glottis or in the larynx is associated with weaker voicing (Chong et al. 2020, Garellek 2020), whereas higher \(H_1^* - A_1^*\) may additionally reflect greater posterior glottal opening (Hanson et al. 2001).

A major, and surprising, result is the dominance of the SHR measure 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. In future work we will take a closer look at this parameter, especially in light of Herbst’s recent (2021) concerns.

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 vs. modal vowels even after the aspiration of a preceding breathy consonant; Esposito and Khan (2012) demonstrated that both categories of vowels in this position have much stronger acoustic and electroglottographic cues to breathiness, suggesting that, in comparison, contrastive breathy vowels in Gujarati are characterized by a subtler degree of breathiness. Notably, this result is seemingly at odds with Tian and Kuang’s (2021) study of breathy voice in Shanghainese and three of the languages from our sample (Hmong, Gujarati, Southern Yi). They 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 and Kuang (2021) did not include SHR in their model, yet in our results, that one parameter is important on every dimension of the 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 inclusion 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*, SoE, and Energy, as a small set of the most informative parameters. These measures also capture differences in the way phonation types are articulated: H1*—H2* and H1*—A1* are associated with vocal-fold contact, medial fold thickness, and posterior glottal opening; SHR is associated with period-doubled phonation, and HNR 0–500 Hz with low-frequency noise (such as from irregular voicing); SoE is associated with voicing intensity, and Energy with overall sound intensity. Other measures apparently either are not very different between phonation categories or are so highly correlated with one of the measures just listed that they provide little additional information.

We conclude this section with a reminder about the limitations of this study. While a corpus of eleven languages goes well beyond the typical single-language study of phonation, it is nonetheless a very small sample for a typological or general-phonetic study. We look forward to future research that will expand cross-language comparisons to dozens, or hundreds, of languages. To this end, we have developed and posted a Shiny app for this project (see n. 11), which facilitates MDS and boxplots of a data set. It opens our data set, which can then be reanalyzed with different subsets of languages or measures; furthermore, anyone can add their own new data to the data set and see how the results change.

As mentioned earlier, differences among our language samples with respect to recording conditions, speaker gender balance, and the words recorded (e.g. different vowel qualities, different tones, in a sentence or not) could confound what appear to be language differences. Our results also depend on the validity and accuracy of our measurements, and as with any automated measurements, the data here are noisy, especially the corrected harmonic amplitude measures. All of these factors have the potential to introduce errors and random variation into the estimation of acoustic measures and consequently into the statistical analysis. However, such errors are likely to be unsystematic, such that they could affect data from any phonation type, speaker, or language. Therefore, we believe that such errors are unlikely to change the phonation spaces that were obtained in this study.

4.4. Implications for phonological representations. We have shown that, not surprisingly, when many acoustic measures are made of a phonatory distinction, multiple measures from among these will contribute to making the distinction. However, at the same time, these many measures can be reduced to a low-dimensional phonetic space, due to overlap among the measures. That is, the high-dimensional measurement space, in which each acoustic measure is a dimension, is so full of redundancy that most of the measures play little independent role in defining the reduced space. The kind of data reduction illustrated here can help both phoneticians and phonologists make sense of complex phonetic categories and contrasts. Acoustic measures that play a larger role—that is, are primary acoustic correlates of the distinctions—are likely to also be primary perceptual cues for listeners. Nonetheless, languages will differ somewhat in the relative importance of primary vs. other acoustic correlates; we know that in general, languages differ in the extent to which they deploy multiple cues to contrasts.
In a low-dimensional phonetic space like the one presented here, acoustic similarity among phonation types can be visualized—in MDS solutions, closer categories are more similar, and more distant categories are more dissimilar. In this sense, our phonetic space for phonatory categories is similar to other low-dimensional acoustic representations of more familiar distinctions like VOT and vowel quality. Thus a VOT continuum, from long closure voicing to long aspiration, provides a range in which languages can fit up to three categories, though with different category boundaries seen across languages (Cho & Ladefoged 1999). Similarly, a vowel space defined by the lowest two formant frequencies can fit many vowel categories (Ladefoged & Maddieson 1996). It seems likely that linguistic contrasts in general function in low-dimensional phonetic spaces, and that identifying such spaces for seemingly more complex contrasts through data-reduction methods is a useful step in understanding sound categories. Keating 1984 emphasized phonetic categories as an interface between phonological contrasts and detailed phonetic realizations. With VOT, the number of possible phonetic categories is an upper limit on the number of contrasting categories within any one language; some languages use all three of them, with other languages that use fewer categories choosing from among this larger set. This is not the case for vowel systems, where there are many more phonetic categories available than any one language uses. Nor is it the case here, for phonation. No language contrasts all of the categories available even in the 2-D phonation space; for example, no language uses all five of the phonations along dimension 2, and even !Xôô seems to make only four distinctions in the 2-D space. That is, Keating’s proposal at best holds only for the simplest category systems, like VOT.

The dimensions of phonetic spaces derived by data reduction could also indicate not just the important acoustic measures, but also the phonological features involved. If we take 2-D vowel formant plots for comparison, we can see that the two dimensions of such plots correspond to sets of phonological features crucial to vowel behavior: one dimension shows vowel height, which corresponds to two or more phonological height/tension features, while the other dimension shows vowel front-backness, which in most languages corresponds to a single phonological backness feature. The more categories lie along a phonetic dimension, the more features are needed to distinguish those categories; for example, if a language also has vowel rounding contrasts, then there will be more phonetic categories along the (primarily) front-back dimension, requiring an additional feature.

Our 2-D plot of the vowel phonation space can, speculatively, be viewed in a similar way. We have interpreted dimension 1 in this space as distinguishing modal from nonmodal categories. This suggests a possible feature [±modal], which would group together seemingly disparate categories like creaky and breathy as laryngeally marked. We do not know if these categories pattern together as a class in opposition to modal in any languages with vowel phonation contrasts. However, Gallagher (2012 and elsewhere) provides a similar case from Quechua, where ejective and aspirated stops (with constricted and spread glottis, respectively) pattern together in opposition to plain stops. For vowel phonation, we can offer the hypothesis that such a pattern is also possible.

We have interpreted dimension 2, by contrast, as a phonetic continuum that ranges over several categories from creaky to breathy. This then suggests features like those for vowel height, perhaps [±spread glottis] and [±constricted glottis], with modal being [−spread, −constricted]. The status of the tense and lax phonetic categories is not yet
clear—it is possible that these are not phonological categories at all, but rather language-specific realizations of creaky and breathy in the absence of a modal category. If they do not need to contrast with modal, or with any other categories, then there would be little dispersion pressure on them, and they could occupy edges of the modal area, with minimal creakiness and breathiness. It is thus an open question whether tense and lax phonations would be [+constricted] and [+spread], respectively, with a reduced phonetic realization, or whether an additional feature is needed. That will depend in part on whether any languages are found to contrast breathy vs. lax, and/or creaky vs. tense.

4.5. 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 two-dimensional space, though the most complex system uses three 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 sixty-one languages with phonation contrasts on vowels mapped in Fig. 1 is given below. Key: t/p cont: tonal/pitch contrast; B: breathy, C: creaky, H: harsh, L: lax, M: modal, P: pharyngealized, T: tense.

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>FAMILY</th>
<th>PHONATION TYPES</th>
<th>T/P CONT</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-Hmao, Western</td>
<td>Hmong-Mien</td>
<td>B, M</td>
<td>Y</td>
<td>Johnson 1999</td>
</tr>
<tr>
<td>Bai</td>
<td>Sino-Tibetan</td>
<td>B, L, M, T, H</td>
<td>Y</td>
<td>Edmondson et al. 2001</td>
</tr>
<tr>
<td>Bo</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Kuang &amp; Keating 2014</td>
</tr>
<tr>
<td>Cham</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>Y</td>
<td>Brunelle 2012</td>
</tr>
<tr>
<td>Chichimeco</td>
<td>Otomanguean</td>
<td>B, M, C</td>
<td>Y</td>
<td>Kelterer &amp; Schluppler 2020</td>
</tr>
<tr>
<td>Danish</td>
<td>Indo-European</td>
<td>M, C</td>
<td>N</td>
<td>Fischer-Jørgensen 1989</td>
</tr>
<tr>
<td>Dinka, Bor</td>
<td>Nilotic</td>
<td>B, M, H</td>
<td>Y</td>
<td>Edmondson &amp; Esling 2006</td>
</tr>
<tr>
<td>Halang</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Cooper &amp; Cooper 1966</td>
</tr>
<tr>
<td>Hani, Haoni</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Maddieson &amp; Ladefoged 1985</td>
</tr>
<tr>
<td>Hani, Luchun</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Kuang &amp; Keating 2014</td>
</tr>
<tr>
<td>(h)Mong, Green</td>
<td>Hmong-Mien</td>
<td>B, M, C</td>
<td>Y</td>
<td>Andruski &amp; Ratliff 2000</td>
</tr>
<tr>
<td>Jingpho</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Maddieson &amp; Ladefoged 1985</td>
</tr>
<tr>
<td>Ju’hoansi</td>
<td>Kx’a</td>
<td>B, M, C, H</td>
<td>Y</td>
<td>Miller 2007</td>
</tr>
</tbody>
</table>

(Table continues)
<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>FAMILY</th>
<th>PHONATION TYPES</th>
<th>T/P</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kedang</td>
<td>Austronesian</td>
<td>B, M</td>
<td>N</td>
<td>Samely 1991</td>
</tr>
<tr>
<td>Kri</td>
<td>Austroasiatic</td>
<td>T, L</td>
<td>N</td>
<td>Enfield &amp; Difloth 2009</td>
</tr>
<tr>
<td>Kui</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Thongkum 1987a</td>
</tr>
<tr>
<td>Lisu</td>
<td>Sino-Tibetan</td>
<td>M, C</td>
<td>Y</td>
<td>Tabain et al. 2019</td>
</tr>
<tr>
<td>Livonian</td>
<td>Uralic</td>
<td>M, C</td>
<td>N</td>
<td>Balodis 2018</td>
</tr>
<tr>
<td>Mah Meri</td>
<td>Austroasiatic</td>
<td>T, L</td>
<td>N</td>
<td>Kruspe &amp; Hajek 2009</td>
</tr>
<tr>
<td>Mamaindé</td>
<td>Nambiquaran</td>
<td>M, C</td>
<td>Y</td>
<td>Eberhard 2009</td>
</tr>
<tr>
<td>Mambay</td>
<td>Niger-Congo</td>
<td>M, C, P</td>
<td>Y</td>
<td>Anonby 2006</td>
</tr>
<tr>
<td>Maya, Yucatec</td>
<td>Mayan</td>
<td>M, C</td>
<td>Y</td>
<td>Frazier 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miao, Xinzhai</td>
<td>Hmong-Mien</td>
<td>B, M</td>
<td>Y</td>
<td>Liu et al. 2020</td>
</tr>
<tr>
<td>Mixtec, Coatzospan</td>
<td>Otomanguean</td>
<td>M, C</td>
<td>Y</td>
<td>Gerfen &amp; Baker 2005</td>
</tr>
<tr>
<td>Mixtec, Ixpantepec Nieves</td>
<td>Otomanguean</td>
<td>M, C</td>
<td>Y</td>
<td>Carroll 2015</td>
</tr>
<tr>
<td>Mon, Ban Nakhonchum</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Thongkum 1987b, Abramson et al. 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mpi</td>
<td>Sino-Tibetan</td>
<td>M, T</td>
<td>Y</td>
<td>Blankenship 2002</td>
</tr>
<tr>
<td>Mundurukú</td>
<td>Tupian</td>
<td>M, C</td>
<td>Y</td>
<td>Picanço 2005</td>
</tr>
<tr>
<td>Nuer</td>
<td>Nilotic</td>
<td>B, M</td>
<td>Y</td>
<td>Monich 2017</td>
</tr>
<tr>
<td>Nyah Kur</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Thongkum 1987a</td>
</tr>
<tr>
<td>Oeshi, Louma</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Lew &amp; Gruber 2016</td>
</tr>
<tr>
<td>Phalok</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Thongkum 1988</td>
</tr>
<tr>
<td>Reel, Thok</td>
<td>Nilotic</td>
<td>B, M</td>
<td>Y</td>
<td>Reid 2010</td>
</tr>
<tr>
<td>Sedang</td>
<td>Austroasiatic</td>
<td>M, C</td>
<td>N</td>
<td>Smith 1975</td>
</tr>
<tr>
<td>Shanghainese</td>
<td>Sino-Tibetan</td>
<td>B, M, C</td>
<td>Y</td>
<td>Tian &amp; Kuang 2016</td>
</tr>
<tr>
<td>Suai, Kuai</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>Y</td>
<td>Abramson et al. 2004</td>
</tr>
<tr>
<td>Tamang</td>
<td>Sino-Tibetan</td>
<td>B, M</td>
<td>Y</td>
<td>Mazaudon &amp; Michaud 2008</td>
</tr>
<tr>
<td>Ticuna, Cushillococha</td>
<td>Ticuna</td>
<td>M, C</td>
<td>Y</td>
<td>Skilton 2016</td>
</tr>
<tr>
<td>Trique</td>
<td>Otomanguean</td>
<td>M, T</td>
<td>Y</td>
<td>DiCanio 2010</td>
</tr>
<tr>
<td>Udihe</td>
<td>Tungusic</td>
<td>M, C</td>
<td>N</td>
<td>Nikolaeva &amp; Tolskaya 2001</td>
</tr>
<tr>
<td>Vietnamese, Northern</td>
<td>Austroasiatic</td>
<td>B, M, C</td>
<td>Y</td>
<td>Pham 2003, Brunelle et al. 2010</td>
</tr>
<tr>
<td>Wa, Lin Tsang</td>
<td>Austroasiatic</td>
<td>T, L</td>
<td>N</td>
<td>Maddison &amp; Ladefoged 1985</td>
</tr>
<tr>
<td>Wa, several varieties</td>
<td>Austroasiatic</td>
<td>B, M</td>
<td>N</td>
<td>Watkins 1999, 2002</td>
</tr>
<tr>
<td>Yi, Luquan Nasu</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Maddison &amp; Ladefoged 1985</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yi, Northern</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Edmondson et al. 2001</td>
</tr>
<tr>
<td>Yi, Southern</td>
<td>Sino-Tibetan</td>
<td>T, L</td>
<td>Y</td>
<td>Kuang &amp; Keating 2014</td>
</tr>
<tr>
<td>Zapotec, Choapam</td>
<td>Otomanguean</td>
<td>M, C</td>
<td>Y</td>
<td>Lyman &amp; Lyman 1977</td>
</tr>
<tr>
<td>Zapotec, San Juan Guelavia</td>
<td>Otomanguean</td>
<td>M, C</td>
<td>Y</td>
<td>Jones &amp; Knudson 1997</td>
</tr>
<tr>
<td>Zapotec, San Lucas Quiravini</td>
<td>Otomanguean</td>
<td>B, M, C</td>
<td>Y</td>
<td>Munro &amp; López 1999</td>
</tr>
<tr>
<td>Zapotec, Santa Ana Del Valle</td>
<td>Otomanguean</td>
<td>B, M, C</td>
<td>Y</td>
<td>Esposito 2010b</td>
</tr>
<tr>
<td>Zapotec, Yalálag</td>
<td>Otomanguean</td>
<td>M, C</td>
<td>Y</td>
<td>Avelino 2010</td>
</tr>
<tr>
<td>Hloon, Eastern</td>
<td>Kx’a</td>
<td>M, C, P</td>
<td>Y</td>
<td>Honken 2013</td>
</tr>
</tbody>
</table>
A cross-language acoustic space for vocalic phonation distinctions


Garellek, Marc; Yuan Chai; Yaqian Huang; and Maxine Van Doren. 2021. Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*, First view. DOI: 10.1017/S0025100321000116.


Iseli, Markus; Yen-liang Shue; and Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121.2283–95. DOI: 10.1121/1.2697522.


A cross-language acoustic space for vocalic phonation distinctions


SIMPSON, ADRIAN P. 2012. The first and second harmonics should not be used to measure breathiness in male and female voices. Journal of Phonetics 40.477–90. DOI: 10.1016/ j.wocn.2012.02.001.


Keating
Department of Linguistics
3125 Campbell Hall
UCLA
Los Angeles, CA 90095-1543
[keating@humnet.ucla.edu]
[kuangj@ling.upenn.edu]
[mgarellek@ucsd.edu]
[esposito@macalester.edu]
[skhan@reed.edu]

[Received 13 November 2019; revision invited 23 February 2020; revision received 1 December 2021; accepted pending revisions 1 January 2022; revision received 18 December 2022; accepted 18 December 2022]