Research Article

Phonetica
DOI: 10.1159/000494301

Received: October 20, 2017
Accepted: October 2, 2018
Published online: February 8, 2019

Acoustic Discriminability of the Complex Phonation System in !Xóõ

Marc Garellek

Department of Linguistics, University of California, San Diego, San Diego, CA, USA

Abstract

Phonation types, or contrastive voice qualities, are minimally produced using complex movements of the vocal folds, but may additionally involve constriction in the supraglottal and pharyngeal cavities. These complex articulations in turn produce a multidimensional acoustic output that can be modeled in various ways. In this study, I investigate whether the psychoacoustic model of voice by Kreiman et al. (2014) succeeds at distinguishing six phonation types of !Xóõ. Linear discriminant analysis is performed using parameters from the model averaged over the entire vowel as well as for the first and final halves of the vowel. The results indicate very high classification accuracy for all phonation types. Measures averaged over the vowel's entire duration are closely correlated with the discriminant functions, suggesting that they are sufficient for distinguishing even dynamic phonation types. Measures from all classes of parameters are correlated with the linear discriminant functions; in particular, the "strident" vowels, which are harsh in quality, are characterized by their noise, changes in spectral tilt, decrease in voicing amplitude and frequency, and raising of the first formant. Despite the large number of contrasts and the time-varying characteristics of many of the phonation types, the phonation contrasts in !Xóõ remain well differentiated acoustically.

Introduction

The number of acoustic studies of contrastive voice quality has increased rapidly in the last 15 years. This has allowed for a clearer understanding both of how to measure changes in voice quality acoustically, and of how different languages make use of a multidimensional acoustic space for realizing phonation contrasts. For example, the measure H1–H2 (Bickley, 1982), the difference in amplitude between the first and second harmonics, has been shown to differentiate the three most common contrastive phonation types found across languages: modal voice, breathy voice qual-

KARGER

E-Mail karger@karger.com
www.karger.com/pho

Marc Garellek
Department of Linguistics, University of California
9500 Gilman Drive #0108
La Jolla, CA 92093-0108 (USA)
E-Mail mgarellek@ucsd.edu

Table 1. Components of the psychoacoustic model of voice quality, their associated parameters, and their articulatory correlates

Model component	Parameters	Some articulatory correlates
Harmonic source spectral slope (or "tilt")	H1–H2 and spectral tilt in 3 other bands (described below)	Medial vocal fold thickness, glottal width
Inharmonic source noise	Harmonics-to-noise ratio	Glottal width, irregular voicing
Time-varying source characteristics	F0 track Amplitude track	Rate of voicing Intensity of voicing
Vocal tract transfer function	Formant frequencies and bandwidths Spectral zeroes and bandwidths	Vocal tract configuration

Based on Kreiman et al. (2014), Garellek et al. (2016), Zhang (2016), and Garellek (to appear).

ities (including "lax" or "slack" voice), and creaky voice qualities (including those called "laryngealized," "glottalized," or "tense" voice) (Blankenship, 2002; Thurgood, 2004; Miller, 2007; DiCanio, 2009; Brunelle & Finkeldey, 2011; Garellek & Keating, 2011; Kuang, 2011; Garellek, 2012; Esposito, 2012; Khan, 2012; Berkson, 2013; DiCanio, 2014; Abramson et al., 2015; Misnadin et al., 2015).

Nonetheless, because of the acoustic multidimensionality of voice quality (which itself is derived in large part from the complex nature of vocal fold vibrations), there are many different acoustic measures that the researcher may choose to investigate when describing phonation types. Which measures should the researcher include? The first goal of this paper addresses this question by investigating the complex phonation system of East !Xóõ (Taa), a Tuu language spoken in Botswana. This language has perhaps the highest number of contrastive phonation types among the languages of the world, with four basic phonation types that can be coproduced to form six (and possibly up to eight) phonation type contrasts (Traill, 1985, 1994b). Given the complexity of the phonation contrasts in !Xóõ, one can assume that an acoustic analysis of the phonation system of the language will require either an unusually large number of measures, or measures that incorporate temporal changes, in order to differentiate the phonation types.

Answering the question of which measures to include in an analysis of phonation is of course important for methodological reasons: using the same set of measures across languages and research groups facilitates cross-study comparisons; it also avoids investigating measures which are less likely to be perceptible (such as "jitter" and "shimmer"; see Kreiman & Gerratt, 2005) and focusing on a particular measure simply because a statistical difference between two phonation types is found.

The question of "which measures to include" is also related to a theoretical one: how can we model voice quality? To answer this, Kreiman et al. (2014) propose a psychoacoustic model of the voice, illustrated in Table 1. The model assumes that voice quality changes can be derived from manipulations of both the vocal folds and supraglottal structures. It includes only acoustic parameters that are both necessary and sufficient to distinguish any perceivable change in voice quality, but these parameters should also link back to physiological changes in the vocal folds or vocal tract. For instance, the spectral slope parameter H1–H2 is related physiologically to how open the vocal folds are when they vibrate, as well as their medial thickness (Kreiman et al., 2008; Samlan et al., 2013; Zhang, 2016). The harmonics-to-noise ratio (HNR), which in more recent versions of the model is measured in multiple frequency bands (Kreiman et al., 2016), is related to the presence of aspiration noise from a more open glottis, as well as irregular voicing (Zhang et al., 2013; Keating et al., 2015; Zhang, 2016). In addition, prior work has shown that listeners are sensitive to both H1–H2 and the HNR (Kreiman & Gerratt, 2010; Kreiman et al., 2010; Kreiman & Gerratt, 2012; Garellek et al., 2013, 2016).

Still, there has yet to be an investigation of whether (and how well) the entire model by Kreiman et al. (2014) succeeds at distinguishing a given language's phonation types. In this paper, I use linear discriminant analysis (LDA), a statistical classification tool, to determine how accurately six phonation types of !Xóõ can be classified using measures based on the parameters in the psychoacoustic model. Assuming that the model by Kreiman et al. (2014) includes parameters that are sufficient to describe any perceptible change in voice quality, it should therefore be able to successfully differentiate many phonation types from a very complex system like the kind found in !Xóõ.

A secondary goal of this study is to characterize the acoustics of relatively understudied phonation types. Researchers investigating the acoustic differences across phonation types have shown that spectral tilt and noise measures are successful at distinguishing breathy, modal, and creaky phonation types, which minimally involve engagement of the vocal folds (Gordon & Ladefoged, 2001; Garellek & Keating, 2011; Keating et al., 2011). Yet languages can contrast other phonation types, notably vowels with a harsh quality, which can show engagement of the ventricular and aryepiglottic folds, tongue retraction, and pharyngeal narrowing, in addition to irregular vocal fold vibration (Edmondson et al., 2001; Edmondson & Esling, 2006; Miller, 2007; Moisik & Esling, 2011; Moisik et al., 2014). Indeed, !Xóõ has "pharyngealized" vowels, which are phonetically both pharyngealized and creaky. It also has "strident" vowels, which are characterized by their harsh quality (Ladefoged, 1983; Traill, 1985; Ladefoged & Maddieson, 1996; Naumann, 2016).

The acoustic characteristics of the pharyngealized and harsh "strident" vowels in !Xóõ are still unclear. Based mostly on visual inspection of spectrograms, pharyngealized vowels have been characterized as having irregular voicing (and even sustained vocal fold closure), raising of F1 and F2, lowering of F3, and a diminution of energy around 400–700 Hz (Traill, 1985; Ladefoged & Maddieson, 1996). Based on spectrograms of words produced by the author himself, Traill (1986) shows that the harsh vowels (i.e., those with the "sphincteric" mechanism in its most constricted state) can be produced by a sequence of weak breathy voicing followed by voiceless noise and low-frequency pulsing. Investigating the spectrograms from recordings of four speakers saying one word, Ladefoged & Maddieson (1996, pp. 310–313) describe the "strident" vowels as being irregular and noisy, with a particularly raised second formant (relative to the pharyngealized vowels) but lowered F3.

Similar phonation types in other Khoisan languages have been more thoroughly studied. Like creaky voice, "epiglottalized" or "pressed" voice in Ju|'hoansi has lower

DOI: 10.1159/000494301

H1–H2 and higher noise relative to modal voice (Miller, 2007), as well as F1 raising, higher-formant lowering, a creaky "interrupted" (i.e., rearticulated) quality, and longer duration (Snyman, 1977a, 1977b). We might therefore expect the pharyngealized and "strident" vowels in !Xóõ to share these attributes. However, articulatory configurations associated with harsh voice qualities – mainly constriction of the ventricular folds and epilaryngeal tube, as well as aryepiglottic trilling – have shown other acoustic attributes, including higher H1–H2, a decrease in noise (in the case of aryepiglottic constriction with no trilling), and formant raising (Laver, 1994; Ladefoged & Maddieson, 1996; Gordon & Ladefoged, 2001; Samlan & Kreiman, 2014). Overall, then, it remains unclear how the pharyngealized and "strident" vowels in !Xóõ are realized acoustically.

Phonation Types and Other Suprasegmental Contrasts of !Xóõ

!Xóõ (also spelled !Xoon) is a Tuu language with about 4,000 speakers in Botswana and Namibia (Vossen, 2013; Bradfield, 2014). The Tuu language family has in the past been classified as Southern South African Khoisan (Greenberg, 1963), whose status is controversial (Honken, 2013; Güldemann, 2014; Naumann, 2014). Researchers increasingly refer to the language as Taa, which is used by native speakers (Traill, 1985) and which follows established naming conventions (e.g., those by Haspelmath, 2017). Here I follow earlier work in referring to the language as !Xóõ (Traill, 1985, 1991, 1994a, 1994b; Bradfield, 2014; Güldemann, 2014; Naumann, 2014). The main varieties of the language are West !Xóõ and East !Xóõ (Traill, 1985, p. 10; Naumann, 2011); most research on East !Xóõ, the variety studied here, comes from work by Anthony Traill (e.g., Traill, 1985, 1986, 1991, 1994b; Traill & Vossen, 1997).

Although famous for its consonantal complexity, especially with regard to clicks (Ladefoged & Traill, 1984; Traill, 1985, 1994a, 1994b; Bradfield, 2014; Güldemann, 2016; Naumann, 2016), !Xóõ also shows great suprasegmental complexity on vowels. The language contrasts the five vowels /i, e, a, o, u/, all of which except for /i/ can be either oral or nasal. Traill (1985, 1991, 1994b) analyzes the East !Xóõ dialect as having four tones (high-falling, mid-falling, mid-level, and low-falling), which are only assigned to the root. The four tones were reanalyzed as four level tones by Miller-Ockhuizen (1998), but there have also been suggestions that they could be reduced to bimoraic sequences of H versus L tones (Traill, 1985, p. 50; Elderkin, 1989, pp. 248–252; Naumann, 2008).

Vowels in the language distinguish four main phonation types: modal, breathy ("murmured"), creaky ("laryngealized, glottalized"), and pharyngealized ("pressed") (Ladefoged, 1983; Traill, 1985, 1986, 1994b).[1] Although the breathy, creaky, and pharyngealized phonation types have been called various other names in previous work, I choose to use these terms in keeping with current research on voice quality and phonation types (more on this regarding the harsh "strident" phonation type below).

Modal, breathy, and creaky phonation types are found on all five vowels; pharyngealized voice is only contrastive on the non-front vowels /a, o, u/ (see also Table 2). Traill (1985) describes the licensing of phonation types in terms of morae, where non-modal phonation types can only appear on the initial mora. It is perhaps simpler to describe where non-modal phonation can occur in terms of initial versus non-

[1] Sample modal, breathy, and pharyngealized vowels, as well as harsh vowels, can be found on the website for Ladefoged (2005): http://www.phonetics.ucla.edu/vowels/chapter14/_xoo.html.

DOI: 10.1159/000494301

initial vowels, obviating the need to appeal to an additional prosodic structure: lexical items in !Xóõ can be of the form CVV (with identical or non-identical vowels), CVC, or CVCV. Non-modal phonation can only occur on the initial vowel, even in monosyllabic words with identical vowels (or a bimoraic vowel). Thus, in a word like /qâˤa/ "long ago," only the initial vowel is pharyngealized; the second one is modal.

Pharyngealization is typically considered a secondary articulation, rather than a phonation type per se; however, in !Xóõ it functions phonologically as a phonation type, and moreover often involves irregular voicing (Traill, 1985, p. 75), in addition to tongue retraction and lowering (Traill, 1985; Hess, 1998; see also the spectrograms and X-ray tracings in Ladefoged & Maddieson, 1996, p. 309). It is well known that so-called pharyngealized vowels in many languages show additional constrictive effects on the vocal folds (see Catford, 1983; McCarthy, 1994; Esling, 1996; Hess, 1998; Zawaydeh, 1999; Esling, 1999; Moisik & Esling, 2011; Moisik, 2012). Such "pharyngealized" vowels might today best be described as "epilaryngeal" (Moisik & Esling, 2011) following the Laryngeal Articulator Model (Esling, 2005; Edmondson & Esling, 2006). They have constriction of the epilaryngeal tube, which entails laryngeal raising and tongue retraction, in addition to a non-modal voice quality such as creaky voice.

Like pharyngealization, nasalization is often treated in the literature on Khoisan as a phonation type. However, nasalized vowels are excluded in the present study because, unlike pharyngealization, nasalization does not involve a specific laryngeal articulation. Thus, it can be considered a separate phonetic dimension from the other phonation types. Moreover, unlike the other phonation types, which can occur only on initial vowels, nasalization is regarded as a property of the second vowel; initial vowels are nasalized only if the second vowel is nasalized, regardless of the stem structure. This suggests that initial vowels are nasalized by leftward spreading (Traill, 1985, pp. 88–89).

Although the phonation types are contrastive on initial vowels, they nonetheless differ in terms of their phasing. Breathy vowels are breathy throughout their duration. On the other hand, both creaky and pharyngealized vowels tend to be creakiest between the middle and end of the initial vowel, giving them a rearticulated quality. (The same is attested for Jul'hoansi; Snyman (1977a) calls these vowels "interrupted" and "juxtaposed".) Creaky vowels can also end with sustained vocal fold closure – that is, a glottal stop [ʔ], without an "echo" vowel associated with rearticulation. An example of a creaky vowel with sustained closure is shown in the rightmost example of Figure 1. I choose to represent creaky vowels with a superscript glottal stop (e.g., [aˀ]) because the International Phonetic Alphabet (IPA)'s "creaky" diacritic is often used to represent pharyngealization in !Xóõ (Traill, 1985). The pharyngealized quality in pharyngealized vowels is usually strongest in the middle of the initial vowel (Traill, 1985). Figure 2 illustrates the variation in (epi-)glottal realization seen on pharyngealized vowels. Although phonologically pharyngealized vowels can also be phonetically creaky, there exist acoustic differences between them and phonologically creaky vowels, as will be seen in this study.

According to Traill, the four main phonation types can also be combined with one another, yielding vowels that are breathy-creaky, pharyngealized-creaky, and breathy-pharyngealized ("strident"). Only /a, o, u/ are attested with breathy-creaky and breathy-pharyngealized phonation, whereas only /a, u/ are attested with pharyngealized-creaky phonation. And though not discussed explicitly in earlier work (e.g., in Traill, 1985), Traill's dictionary includes at least eight words that have a combina-

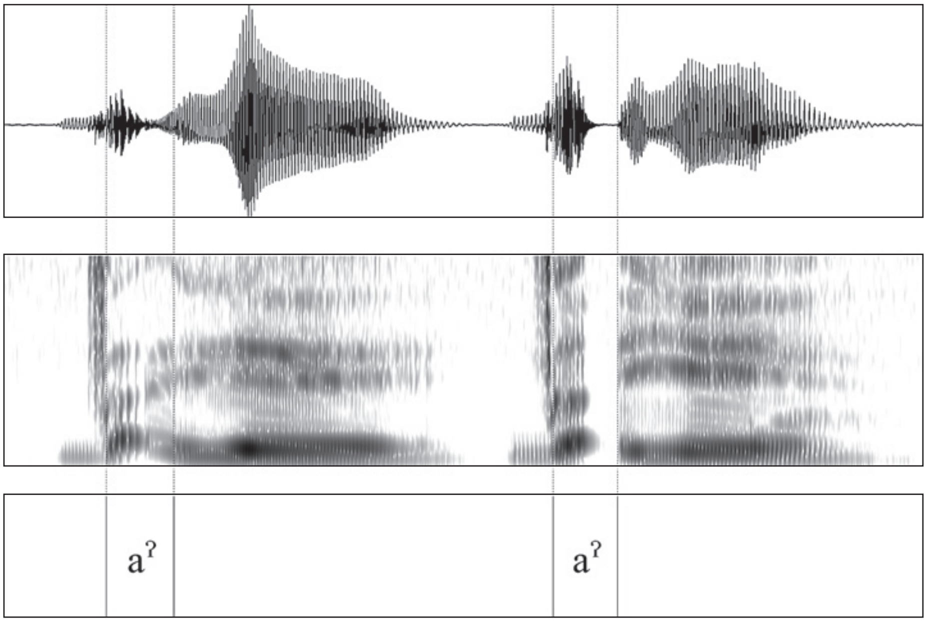

Fig. 1. Sample waveforms and spectrograms of two instances of /ᶢ|àˀje/ "bend" from two speakers in the data set from this study. The vertical axes show up to 5,000 Hz. The first vowel is phonologically creaky in both cases, yet can be realized phonetically with only creaky voice (left), or with both creaky voice and sustained glottal closure (right).

tion of all three non-modal phonation types: breathy-pharyngealized-creaky words with /a/ and /o/ (Traill, 1994b).[2] There are additional phonotactic restrictions that determine where non-modal phonation types occur (see Traill, 1985; Miller, 2010; Güldemann, 2013). Combinations of phonation types are also only found on initial vowels.

Breathy-creaky and pharyngealized-creaky phonation types are described as having their creaky voice located at the very end of the vowel; the "strident" breathy-pharyngealized vowels are characterized by their harsh voice, with a rough, growling quality. According to X-ray and fiberscopic data, these vowels are produced with a lowered tongue body, forward movement of the posterior pharyngeal wall, laryngeal raising, extreme constriction in the upper part of the larynx, and a posterior gap between the vocal folds (Traill, 1985, 1986; Hess, 1998; see also the spectrograms and X-ray tracings from native speakers and Tony Traill in Ladefoged & Maddieson,

[2] Words with breathy-pharyngealized-creaky vowels include the following (using the orthography in Traill, 1994b): *ǀqá̰h'le* "phalanges of an ungulate" (p. 61); *!gà̰h'm* "erythema" (p. 80); *!gō̰h'u-ka* "juvenile Bushveld lizard" (p. 81); *ǁnà̰h'ɲa* "turn over" (p. 124); *ǂā̰h'a* "restrain" (p. 130); *tà̰h'a* "young of […] ostrich" (p. 155); *tsā̰h'li* "split moist […] pods" (variant of pharyngealized-creaky *tsā̰'li*; p. 162); and *dzā̰h'nu* "Fork-marked sand snake" (p. 164). Only some of these words could plausibly be considered breathy-pharyngealized vowels followed by a glottal stop; see also a similar discussion of breathy-creaky vowels by Naumann (2016).

DOI: 10.1159/000494301

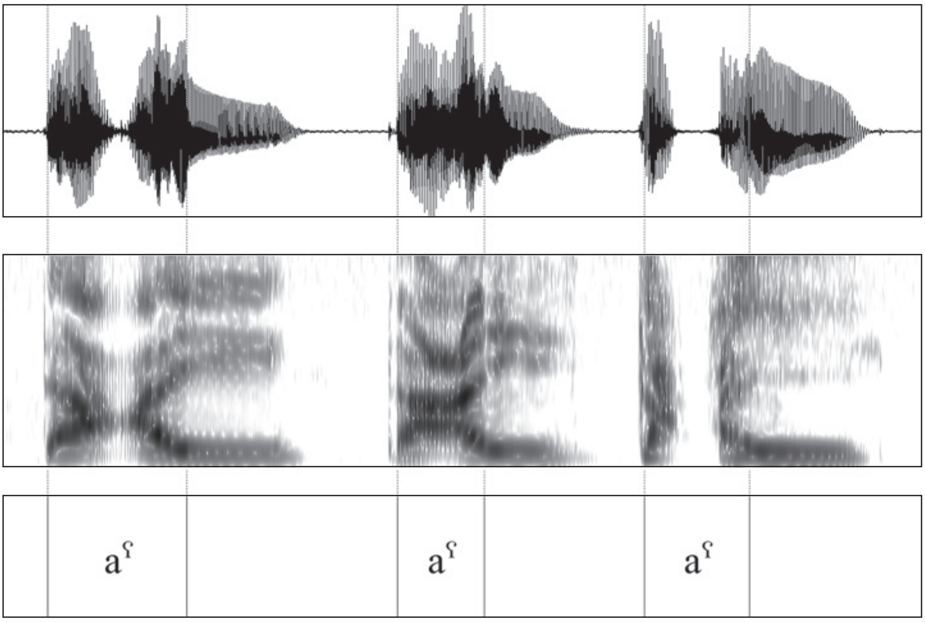

Fig. 2. Sample waveforms and spectrograms of three instances of /táˤi/ "far" from three speakers in the data set from this study. The vertical axes show up to 5,000 Hz. The first vowel is phonologically pharyngealized in each case, and shows phonetic signs of pharyngealization (convergence of F1 and F2, lowering of F3). But phonologically pharyngealized vowels can also be realized phonetically with creaky voice (sudden drop in intensity and F0) in the middle of the vowel (left), with little visible creaky voice (middle), and with sustained glottal closure associated with either a pharyngealized glottal stop ([ʔˤ]) or an epiglottal stop ([ʡ], which necessarily involves glottal closure; Esling, 2005) (right).

1996, pp. 311–312). Additionally, there appears to be what Traill (1985, 1986) calls vibration of the arytenoid cartilages and/or of the epiglottis.[3] Traill (1985, pp. 84–86) states that the strident vowels may be considered breathy-pharyngealized from a phonological viewpoint, mainly due to analytic parsimony, but also because they share some phonetic features with both pharyngealized vowels (notably pharyngeal constriction and tongue lowering) and breathy ones (including higher airflow and noise). Still, their articulation is so unlike both breathy and pharyngealized vowels in other ways that Traill later claims that the articulatory mechanism is distinct phonetically from both pharyngealized and breathy vowels (Traill, 1986). He calls this articulatory mechanism "sphincteric," which is consistent with laryngeal constriction and trilling of the aryepiglottic folds (Moisik & Esling, 2011). From this point onward, I choose to call vowels exhibiting this type of phonation "harsh," rather than "strident," or by their phonological status as "breathy-pharyngealized," because "harsh" is now commonly used to describe such voice quality in a variety of languages (Edmondson et

[3] Most likely, it is the aryepiglottic folds that are vibrating, given that these bodies connect the arytenoids to the epiglottis.

Table 2. Attested phonation types in East !Xóõ and their combinations, based on Traill (1985)

Combination → Base phonation ↓	(None)					Creaky			Pharyngealized (harsh)		
Breathy	i̤	e̤ e̤ẽ	**a̤** a̤ã	o̤ o̤õ	ṳ ṳũ	**a̤ˀ** a̤ˀã	o̤ˀ o̤ˀõ	ṳˀ	**a^{ʢ}** a^{ʢ}ã	o^{ʢ}	u^{ʢ}
Modal	i	e ẽ	**a** ã	o õ	u ũ						
Creaky	iˀ	eˀ eˀẽ	**aˀ** aˀã	oˀ oˀõ	uˀ uˀũ						
Pharyngealized			**aˤ** aˤã	oˤ oˤõ	uˤ uˤũ	aˤˀ		ṵˤˀ			

Note that Traill (1994b) also includes words with harsh-creaky /a^{ʢʔ}, o^{ʢʔ}/ (see Footnote 2). The six phonation types analyzed with the vowel /a/ in this study are bolded and in squares. Nasalization can combine with the other phonation types, but it is considered a property of the second mora that spreads leftward on to the initial mora (Traill, 1985, pp. 88–89).

al., 2001; Gerratt & Kreiman, 2001; Edmondson & Esling, 2006; Moisik & Esling, 2011; Moisik, 2012; Moisik et al., 2014). I also represent harsh voice using a superscript voiced epiglottal trill (e.g., [a^{ʢ}]).[4]

The harshest quality of harsh vowels generally appears in the middle of the vowel's duration, and it can be realized with a combination of breathy voice, voiceless breath (i.e., voiceless aspiration noise), irregular creaky voicing, and sustained glottal closure in addition to pharyngealization (Fig. 3). We will not analyze vowels that are pharyngealized-creaky or breathy-pharyngealized-creaky (harsh-creaky); the latter is unattested in the corpus, and the former is found on only one lexical item, which was not recorded for all speakers. It remains unclear then how pharyngealized and harsh vowels – which often show creaky voice – differ systematically from pharyngealized-creaky or harsh-creaky ones. Sample spectrograms comparing pharyngealized and pharyngealized-creaky words are shown online in supplementary materials.[5]

A summary of attested phonation types and their combinations is shown in Table 2. Waveforms and spectrograms of the six phonation types to be analyzed acoustically are shown in Figure 4. It is also worth mentioning that the suprasegmental contrasts in West !Xóõ are simpler than the ones described here for East !Xóõ: West

[4] Traill (1986, p. 129) states that the "sphincteric" phonation responsible for the quality of these vowels differs specifically from Laver's (1980) view of "harsh voice," which involves vibration of the ventricular folds. However, Traill (1985, p. 80) stated earlier that ventricular fold vibration "probably" occurs with strident vowels. And as J. Esling (pers. commun.) notes, both views are consistent with how the laryngeal articulator performs as a sphincter, compacting posteroanteriorly and vertically to compress the ventricular folds in harsh voice and engaging vibration of the aryepiglottic folds in a more constricted state, as is the case with "sphincteric" phonation in !Xóõ (Esling et al., 2005).
[5] Supplementary materials can be found at http://idiom.ucsd.edu/~mgarellek/files/SupplementaryFiles_!Xoo.html.

DOI: 10.1159/000494301

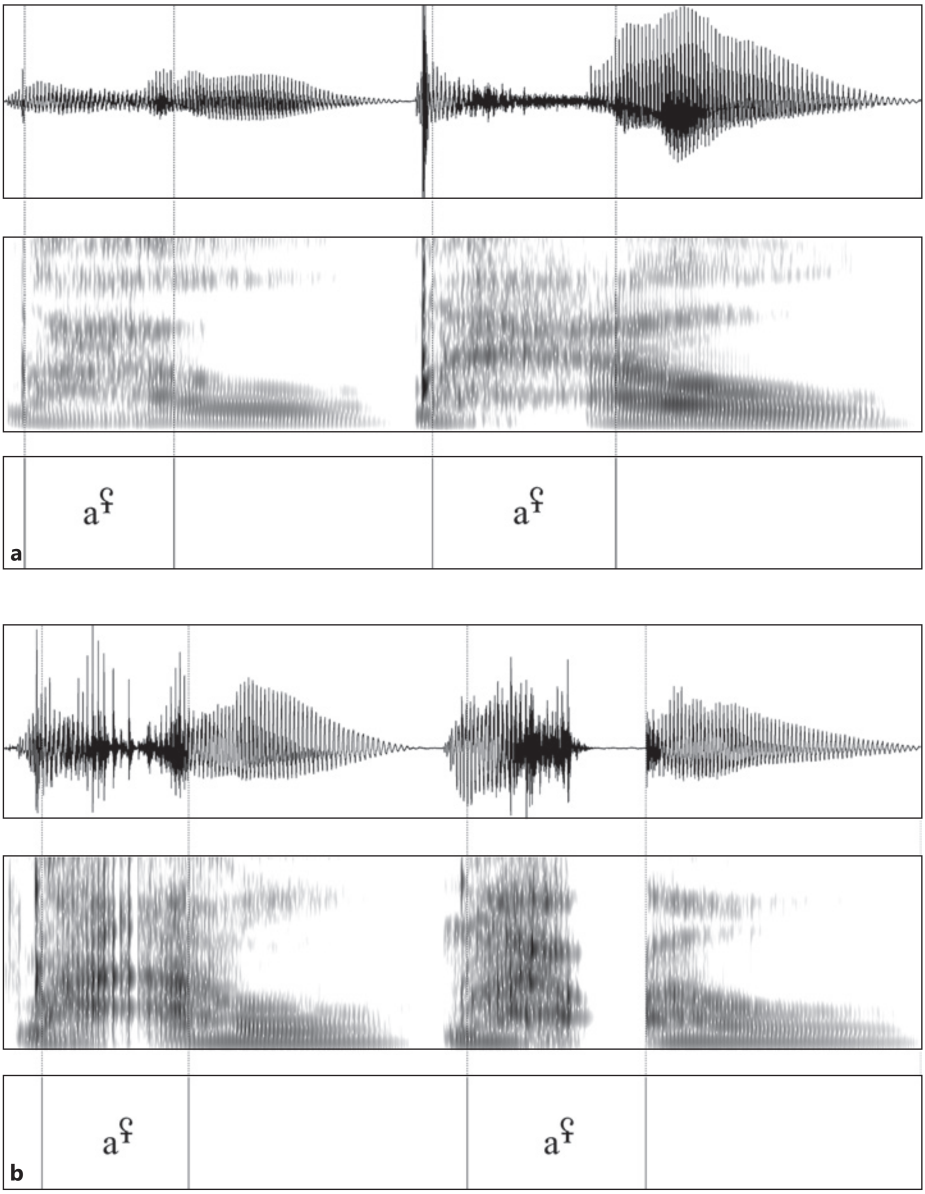

Fig. 3. Sample waveforms and spectrograms of four instances of /ˀᵑ!àʢo/ “base” from four speakers in the data set from this study. The vertical axes show up to 5,000 Hz. The first vowel is phonologically breathy-pharyngealized (harsh) in each token, yet can also be realized phonetically with breathy voice and pharyngealization in the middle of the vowel (**a**, left), voiceless breath and pharyngealization (**a**, right), breathy voice followed by creaky voice and pharyngealization (**b**, left), and even sustained glottal closure (**b**, right).

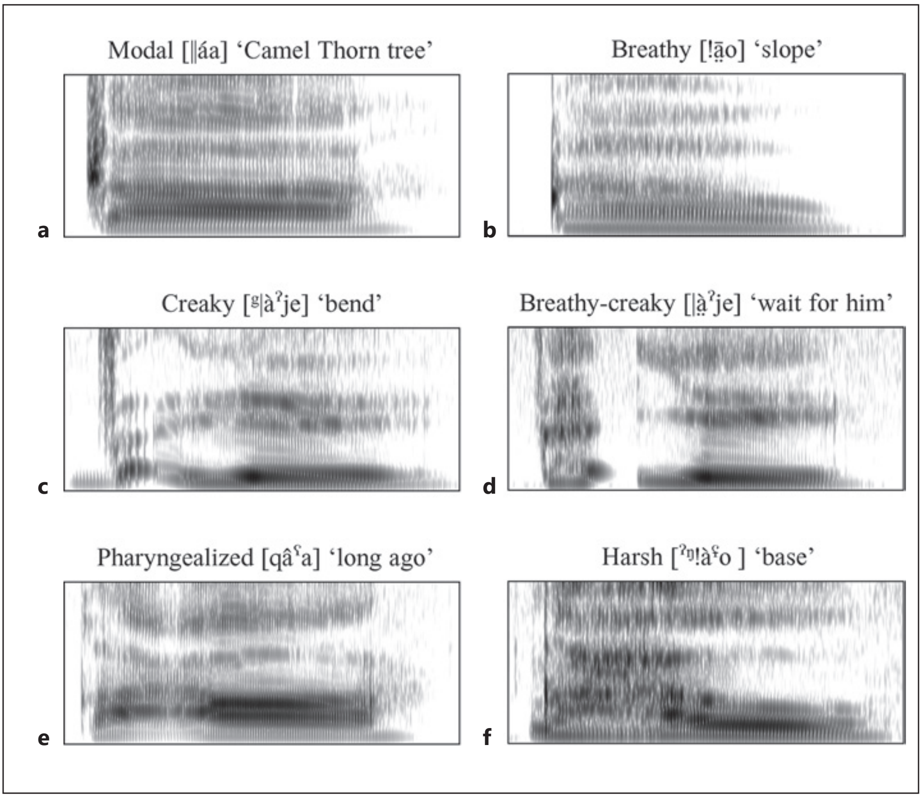

Fig. 4. a–f Sample spectrograms of the six phonation types analyzed in this study. The samples are from the same speaker. The vertical axes show up to 5,000 Hz.

!Xóõ has been analyzed as having only two tones, high and low (Güldemann, 2013; Naumann, 2016). Nasalization is contrastive on /i, a, u/ only. There are five phonation types: modal, tense/glottalized, breathy, pharyngealized, and strident, which correspond to those in East !Xóõ; however, non-modal phonation types can only occur on back vowels, and there are no combinations of breathy or pharyngealized voice with creaky voice (Naumann, 2016).

Language Materials

Recordings

The sample words come from three field recordings from Lokalane, in southwestern Botswana. The recordings were made by Peter Ladefoged and Tony Traill in July and August 1979. In the first recording, words without a carrier sentence were spoken by 10 male speakers. The second recording followed the same procedure with a subset of the wordlist used in the first recording, but including 2 different male speakers. The third recording had the same speakers as the first recording and followed the same procedure, but the speakers said a subset of the words

DOI: 10.1159/000494301

Table 3. Structure of the wordlist, illustrating the number of lexical items by phonation and tone

	Breathy	Modal	Creaky	Pharyngealized	Breathy-creaky	Harsh
High-toned		2		1		
Mid-toned	2				2	
Low-toned	1	1	1	1	1	3
Falling-toned			1	1		

from the first recording, and in a different order. Although some of these recordings were used to outline a few characteristics of harsh voice which were later discussed in Ladefoged & Maddieson (1996) (and further in Gordon & Ladefoged, 2001), the recordings have yet to be systematically analyzed acoustically. The three analog recordings were digitized at a sampling rate of 44.1 kHz and with 16-bit depth, and are available on the website of the UCLA Phonetics Lab Archive at http://archive.phonetics.ucla.edu/Language/NMN/nmn.html.

Sample Words

The sample words chosen for this study all had /a/ as their initial vowel; phonation contrasts are limited to the first vowel. Only /a/ was chosen because of its greater proportion in the wordlists. The sample words could have any of the four lexical tones. Figure 4 illustrates a sample of the !Xóõ words chosen (the complete wordlist is listed in the Appendix and online in the supplementary materials; see Footnote 5). Table 3 shows the number of words by phonation type and tone. All six phonation types are found on words with low tones, whereas only a subset of the phonation types is found on words with the other tones. No phonation type is attested with all four tones (the same is true if one assumes a two-tone analysis, where high and falling tones are H and mid and low tones are L). The onsets are also unbalanced; for example, uvular onsets are attested mostly with pharyngealized and harsh vowels, and Miller (2010) considers the [^{g}!χ] onset (found here on the word for "udder" with a creaky vowel) to be a guttural, which could affect the creaky voice on the following vowel. Differences by lexical item for each acoustic measure are shown in the online supplementary materials (see Footnote 5); however, since the number of lexical items per phonation type is very small, it is impossible in this study to determine whether any differences by lexical item are due to onset effects or lexical idiosyncrasies.

Tokens with audible background noise were excluded. All words with aspirated, postglottalized, or ejective clicks preceding the target vowel were excluded from the analysis to reduce the effect of a consonant's laryngeal setting on the target vowel's voice quality. Two lexical items transcribed as having creaky vowels were excluded because of questions regarding the accuracy of the transcription (see the online supplementary materials for more detail; see Footnote 5).

A total of 17 words were sampled across all speakers: 3 lexical items for each of the six phonation types, with the exception of creaky voice, which had 2 lexical items. Because three different recordings were used, the number of words that were uttered varied across speakers; of the 17 words, 2 were recorded by only 10 speakers, and the remaining 15 words were recorded by all speakers. In a few cases (e.g., if a speaker repeated the same word), multiple tokens of a word were analyzed, for a total of 369 tokens of the 17 words.

Acoustic Measures

The vowel portion of each target word was labeled in Praat (Boersma & Weenink, 2015); the vowel onset was set to the second glottal pulse following the onset (to avoid high-frequency noise due to noisy click releases), and the vowel offset was set to the last glottal pulse before the drop in amplitude of the following consonant. For following consonants that were not characterized as having a sudden drop in amplitude (notably [j]), the start of the F2 change towards the consonantal target was chosen as the vowel offset. VoiceSauce (Shue et al., 2011) was used to obtain a variety of acoustic measures over the vocalic interval. The acoustic measures were calculated ev-

Table 4. Measures in VoiceSauce that are related to the psychoacoustic model of the voice (Kreiman et al., 2014)

Measure	Explanation
Spectral slope measures (dB)	
Higher values associated with increased vocal fold spreading	
H1*–H2*	Difference in amplitude between the first and second harmonics
H2*–H4*	Difference in amplitude between the second and fourth harmonics
H4*–H2* kHz	Difference in amplitude between the fourth harmonic and the harmonic nearest 2,000 Hz
H2* kHz–H5* kHz	Difference in amplitude between the harmonic nearest 2,000 Hz and the harmonic nearest 5,000 Hz
Harmonics-to-noise ratio (HNR) measures (dB)	
Higher values associated with decreased noise (aspiration noise and/or irregular voicing)	
HNR05	HNR <500 Hz
HNR15	HNR <1,500 Hz
HNR25	HNR <2,500 Hz
HNR35	HNR <3,500 Hz
Formants and bandwidth measures (Hz)	
F1, F2, F3, and F4	First through fourth formant calculated using Snack (Sjölander, 2004)
B1, B2, B3, and B4	First through fourth bandwidth calculated using Snack (Sjölander, 2004)
Other measures	
F0	Fundamental frequency (Hz) calculated using STRAIGHT (Kawahara et al., 1998)
SoE	Strength of glottal excitation (amplitude of voicing, regardless of noise) calculated following Murty & Yegnanarayana (2008)

The analysis includes both mean measures (over the entire vowel) and delta measures between the first and second halves of the vowel. Measures with asterisks are corrected for vowel formants, following Iseli et al. (2007).

ery millisecond, and for the analysis we used mean measures (over the entire interval) as well as changes in measures between the first and second halves of the vowel. We included these "delta" measures because several of the phonation types are highly dynamic. A complete list of the acoustic measures is shown in Table 4. Four spectral tilt measures were included, following Kreiman et al. (2014) and Garellek et al. (2016). These measures are typically lowest during constricted voice qualities like creaky voice and highest during breathy voice (Gordon & Ladefoged, 2001; Garellek & Keating, 2011; Garellek et al., 2013; Kuang, 2013b, 2017). Four noise measures (in different frequency bands) were also included. Voice qualities characterized by increased noise (either from aspiration or from irregular voice quality), such as breathy, creaky, and harsh voice, typically have lower HNR values than periodic voice qualities like modal voice (Blankenship, 2002; Miller, 2007; Garellek, 2012).

Also included were the first four formants' frequencies and bandwidths. Breathy voice sometimes shows a lower F1 frequency and higher F1 bandwidth than modal/creaky voice (Gordon & Ladefoged, 2001; Hanson et al., 2001; Garellek & Keating, 2011). Pharyngealized and epi-

Table 5. Confusion matrix from the linear discriminant analysis

Actual → Predicted ↓	Breathy	Modal	Creaky	Pharyngealized	Breathy-creaky	Harsh
Breathy	78	2	0	0	1	2
Modal	1	45	0	0	0	0
Creaky	0	0	23	0	4	0
Pharyngealized	0	0	1	49	0	3
Breathy-creaky	0	0	5	0	42	3
Harsh	0	0	0	2	1	78

glottalized vowels often show increased formant frequencies, particularly for F1 (Alwan, 1989; Laver, 1994; McCarthy, 1994; Miller-Ockhuizen, 2003; Jongman et al., 2007; Moisik, 2012).

Finally, we also included the fundamental frequency (F0, in Hz) and strength of excitation (SoE). SoE measures the relative amplitude of impulse-like excitation during voicing, and thus represents the amplitude of voicing independent of the amplitude of noise in the signal (Murty & Yegnanarayana, 2008; Mittal et al., 2014). The greater the constriction in the larynx or vocal tract, the lower the SoE, assuming increased constriction leads to weaker voicing (Mittal et al., 2014; Risdal et al., 2016).

Because the accuracy of the spectral tilt measures depends on correct F0 and formant estimation, several steps were taken to identify mistracked tokens. Six tokens were excluded because their mean F0 was more than 2.5 standard deviations from the speaker's average. An additional 23 tokens were excluded because their F1 and/or F2 value(s) was/were judged to be mistracked via visual inspection. These exclusions resulted in a reduced data set of 340 tokens.

Results

Linear Discriminant Analysis

An LDA was conducted in R (R Core Team, 2014)[6] to determine which measures are most important for distinguishing the six phonation types in the data set. All the acoustic parameters listed in Table 4 were included. For each acoustic parameter, three measures were included: the average over the entire vowel, the change in the parameter from the beginning to the middle of the vowel, and the change in the parameter from the middle of the vowel to the vowel's end. These "delta" measures were included because several phonation types are known to involve dynamic changes (Traill, 1985). All measures were standardized by speaker.

Given that there are six phonation types, the LDA produced five functions. The first discriminant function (LD1) accounted for 34.0% of the explained variance; the second (LD2) accounted for 31.6%, and the third accounted for 23.2%. The remaining two LDs contributed 7.3% and 3.9%, respectively. These last two functions will not be discussed further, because they contributed relatively little to the analysis. The confusion matrix from the LDA is shown in Table 5; overall, the accuracy of the model's classification was very high, at 92.7%. (The category occurring most frequently in

[6] The LDA was conducted using the lda() function available from the MASS package (Venables & Ripley, 2002).

Table 6. Correlations between the first three discriminant functions and the acoustic measures

LD1	LD2	LD3
SoE (r = –0.85)	F1 (r = 0.75)	H1*–H2* (r = –0.70)
ΔSoE (beg) (r = –0.69)	ΔF0 (end) (r = 0.57)	HNR <500 Hz (r = 0.62)
HNR <500 Hz (r = –0.46)	HNR <3,500 Hz (r = –0.51)	HNR <3,500 Hz (r = 0.63)

Only the three measures with the highest absolute correlations are shown. LD, linear discriminant; SoE, strength of excitation; HNR, harmonics-to-noise ratio; beg, change in the parameter from the beginning to the middle of the vowel; end, change from the middle to the end of the vowel.

the data set was harsh voice, with a 25.3% prevalence.) Breathy vowels were the most accurately discriminated (98.7%), whereas creaky vowels were the least accurately discriminated (79.3%). Breathy vowels are most confusable with modal ones (and vice versa), creaky vowels with breathy-creaky ones (and vice versa), pharyngealized vowels with creaky ones, and harsh vowels with breathy-creaky and pharyngealized ones.

The mean predicted value of the first three discriminant functions was calculated for each phonation type, and the strongest correlations between the first three discriminant functions and the acoustic measures are shown in Table 6. Figure 5 shows the confidence ellipses around the group means for LD1 and LD2 (Fig. 5a) and for LD1 and LD3 (Fig. 5b). The first discriminant function is used to separate breathy and modal vowels from all other categories, and is most strongly correlated with mean SoE, change in SoE during the first half of the vowel, and HNR <500 Hz. The second discriminant function is used to separate harsh and pharyngealized vowels as distinct from breathy and modal vowels, which in turn are distinct from breathy-creaky and creaky vowels. LD2 is most strongly correlated with mean F1, change in F0 in the last half of the vowel, and mean HNR <3,500 Hz. The third discriminant function is used to group breathy versus modal and pharyngealized versus harsh vowels. LD3 correlates most strongly with mean H1*–H2*, HNR <500 Hz, and HNR <3,500 Hz.

Acoustic Measures Contributing to the LDA

The LDA produced five discriminant functions, three of which contributed substantially to the discriminability of the six phonation types of !Xóõ. In this section, linear mixed-effects models were run on the average measures that were most strongly correlated with the first three discriminant functions, to determine whether specific phonation types differed statistically from each other on a particular measure.[7] The specific p values (derived from the lmerTest package, which uses Satterthwaite approximations; Kuznetsova et al., 2015) and other model details can be found in the

[7] Mixed-effects models were run in R using the lmer() function in the lme4 package (Bates et al., 2014).

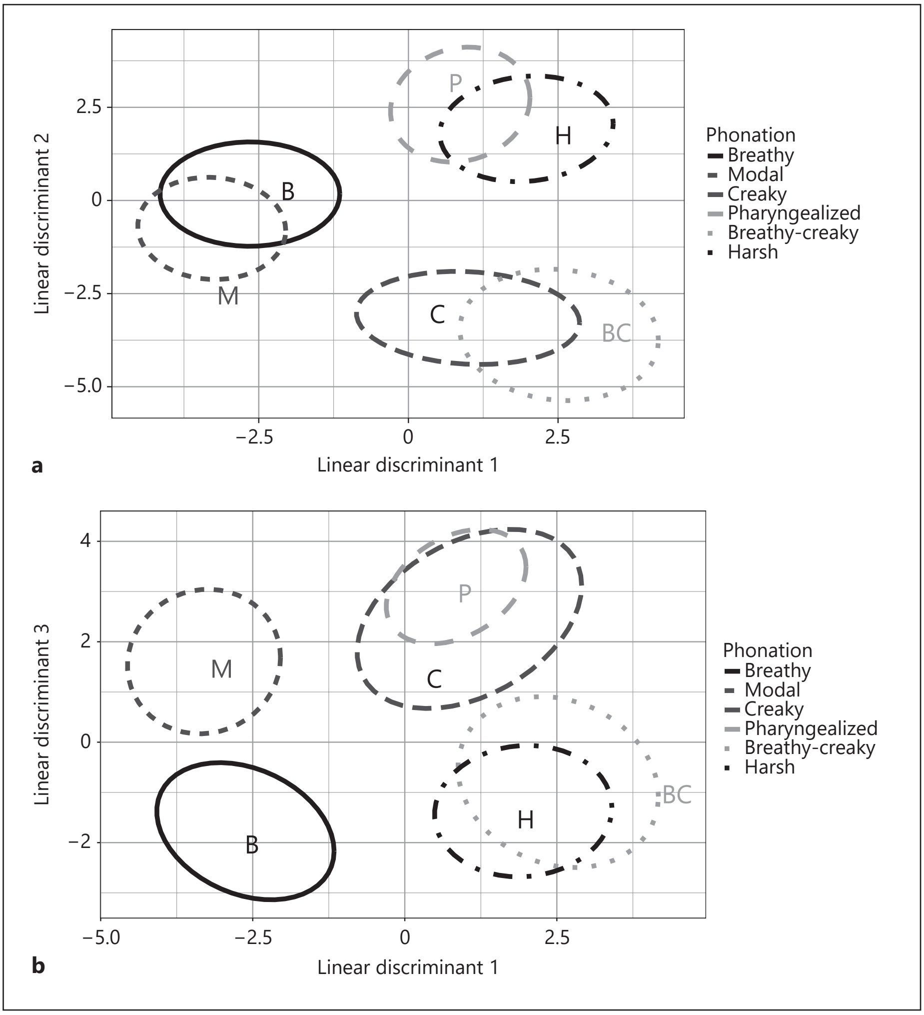

Fig. 5. LD1–LD2 (**a**) and LD1–LD3 (**b**) spaces. The ellipses represent 75% confidence intervals around the mean of each group. Note that pharyngealized versus harsh vowels, breathy-creaky versus creaky vowels, and breathy versus modal vowels are poorly discriminated in the LD1–LD2 space (**a**), but are better discriminated by LD3 (**b**). LD, linear discriminant; B, breathy; M, modal; P, pharyngealized; H, harsh; BC, breathy-creaky; C, creaky.

summary tables. Time courses of the relevant measures are also illustrated and discussed qualitatively. Additional time course figures for other measures, as well as for the measures below but separated according to lexical item, can be found in the online supplementary materials (see Footnote 5).

The first discriminant function LD1 discriminates breathy and modal vowels from other phonation types. It is most strongly correlated with SoE (mean and change in first half) and HNR <500 Hz. Recall that SoE measures the intensity of voicing, re-

DOI: 10.1159/000494301

Table 7. Results of the linear mixed-effects models for mean measures correlated with LD1

	SoE			HNR <500 Hz		
	β	*t*	*p* value	β	*t*	*p* value
Modal	0.07	16.47	<0.001*	37.09	20.94	<0.001*
B	0.00	0.23	>0.05	–13.87	–7.34	<0.001*
C	–0.03	–5.61	<0.001*	–9.80	–4.86	<0.001*
P	–0.03	–5.60	<0.001*	–6.15	–3.01	<0.05*
BC	–0.03	–6.90	<0.001*	–12.81	–4.98	<0.001*
H	–0.03	–5.53	<0.001*	–22.59	–9.09	<0.001*
Breathy	0.07	14.57	<0.001*	23.22	13.26	<0.001*
M	–0.00	–0.23	>0.05	13.87	7.34	<0.001*
C	–0.03	–5.91	<0.001*	4.06	1.75	>0.05
P	–0.03	–5.42	<0.001*	7.71	3.25	<0.01*
BC	–0.03	–7.36	<0.001*	1.05	0.37	>0.05
H	–0.03	–5.87	<0.001*	–8.72	–4.74	<0.001*

The five phonation types are all compared to a baseline phonation type (the intercept, indicated in italics). All models are of the structure *DV* ~ *Phonation* + (*Phonation*|*Speaker*) + (1|*Item*). LD, linear discriminant; SoE, strength of excitation; HNR, harmonics-to-noise ratio; B, breathy; C, creaky; P, pharyngealized; BC, breathy-creaky; H, harsh; M, modal.

gardless of the presence of noise; it is thus lower when there is constriction at the glottis or in the vocal tract (Mittal et al., 2014). The linear mixed-effects models show that the mean SoE is significantly higher for modal voice than for all other phonation types, with the exception of breathy voice (Table 7).

Breathy vowels in turn have a significantly higher mean SoE than all other phonation types, except for modal vowels. These results are expected, since neither modal nor breathy voice has appreciable constriction of the vocal folds or in the vocal tract. The linear mixed-effects models also show that the mean HNR <500 Hz is significantly higher for modal voice than for all other phonation types (Table 7). This is expected, given that modal vowels should be more periodic and less noisy than vowels with non-modal phonation. Breathy vowels in turn have a significantly higher mean HNR <500 Hz than harsh vowels, but a lower HNR <500 Hz than pharyngealized ones.

Time courses for SoE are illustrated in Figure 6a. The measure is highest for breathy and modal voice, as these are the least constricted phonation types. Harsh, pharyngealized, and creaky vowels show a sharp drop in SoE centered around the midpoint of the vowel, where the strongest constriction for these phonation types occurs. Breathy-creaky vowels, on the other hand, show a sharp drop in SoE by the end of the vowels' duration, where they are most constricted. Time courses for HNR <500 Hz are shown in Figure 7a, and they are discussed below with reference to the more broadband noise measure HNR <3,500 Hz.

The second discriminant function LD2 discriminates harsh and pharyngealized vowels from creaky and breathy-creaky ones. It is most strongly correlated with mean F1, as well as with mean HNR <3,500 Hz and the change in F0 from the midpoint to

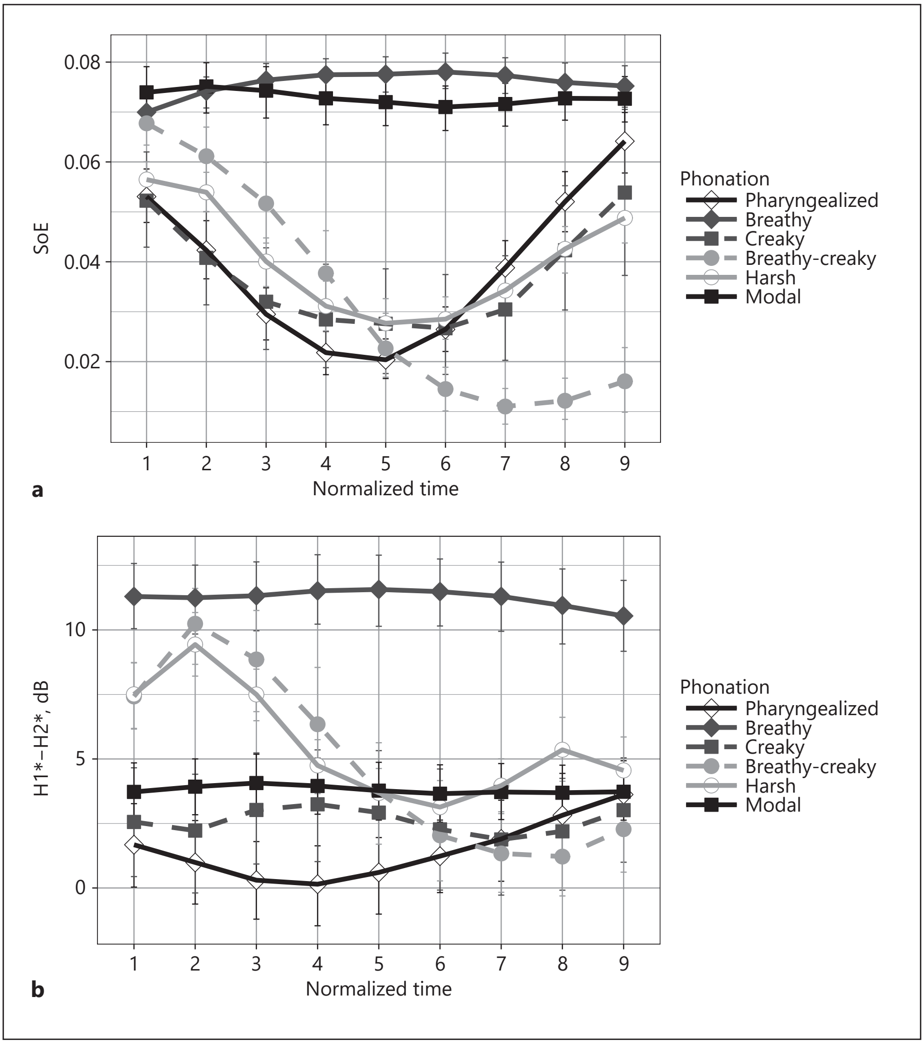

Fig. 6. Time courses of strength of excitation (SoE) (**a**) and H1*–H2* (**b**) by phonation type. Error bars indicate bootstrapped 95% confidence intervals of the means. Average SoE and changes in SoE at the beginning and end of the vowel are strongly correlated with linear discriminant (LD) 1; H1*–H2* is strongly correlated with LD3.

the end of the vowel. The linear mixed-effects models show that breathy-creaky and creaky vowels have significantly lower mean F1 values of than harsh and pharyngealized vowels (Table 8). The higher F1 for pharyngealized and harsh vowels in !Xóõ was previously discussed by Traill (1985), and is cross-linguistically common with pharyngealized vowels (Alwan, 1989; McCarthy, 1994; Miller-Ockhuizen, 2003; Jongman et al., 2007; Moisik, 2012; Al-Tamimi, 2017). The time courses for F1 are shown in Figure 8b. Harsh and pharyngealized vowels are characterized by a higher F1 over-

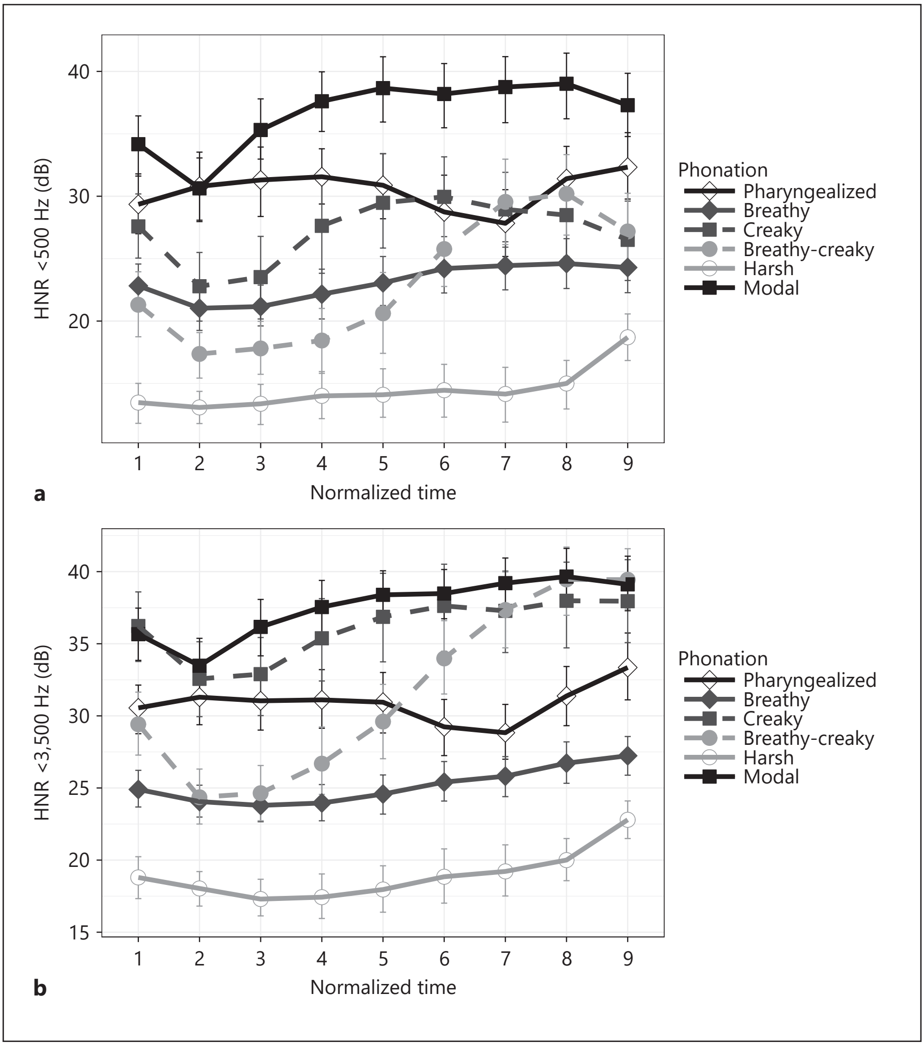

Fig. 7. Time courses of harmonics-to-noise ratio (HNR) <500 Hz (**a**; correlated with linear discriminant [LD] 1 and LD3) and HNR <3,500 Hz (**b**; correlated with LD2 and LD3) by phonation type. Error bars indicate bootstrapped 95% confidence intervals of the means.

all compared with the other phonation types. Creaky and breathy-creaky vowels show a drop in F1 by the end of the vowel, presumably because these categories had lexical items whose creaky and breathy-creaky vowels tended to be followed by [j].

The time courses for F0 are shown in Figure 8a. Harsh and pharyngealized vowels are characterized by an overall rise in F0 from the vowel midpoint onward. The depressive effect of pharyngealization and harshness on F0 is a known feature of these phonation types in !Xóõ (Traill, 1985). It should be noted that the distribution of

DOI: 10.1159/000494301

Table 8. Results of the linear mixed-effects models for mean measures correlated with LD2

	F1			HNR <3,500 Hz		
	β	*t*	*p* value	β	*t*	*p* value
Breathy-creaky	675.69	26.30	<0.001*	32.18	16.41	<0.001*
C	–75.14	–2.00	0.055	3.74	1.83	>0.05
H	170.55	5.49	<0.001*	–12.75	–6.58	<0.001*
P	169.68	5.37	<0.001*	–0.83	–0.47	>0.05
M	17.13	0.50	>0.05	6.08	2.62	<0.05*
B	79.37	2.16	<0.05*	–6.79	–2.90	<0.01*
Creaky	600.55	21.13	<0.001*	35.92	16.76	<0.001*
BC	75.14	2.00	0.056	–3.74	–1.83	>0.05
H	245.69	6.22	<0.001*	–16.50	–6.83	<0.001*
P	244.82	6.49	<0.001*	–4.58	–2.35	<0.05*
M	92.26	2.56	<0.05*	2.34	1.04	>0.05
B	154.50	3.57	<0.01*	–10.54	–4.65	<0.001*

The five phonation types are all compared to a baseline phonation type (the intercept, indicated in italics). All models are of the structure *DV* ~ *Phonation* + (*Phonation*|*Speaker*) + (1|*Item*). LD, linear discriminant; HNR, harmonics-to-noise ratio; B, breathy; C, creaky; P, pharyngealized; BC, breathy-creaky; H, harsh; M, modal.

lexical tones in the wordlist is unbalanced (Table 3); moreover, some of the lexical items have different tone markings in the UCLA Phonetics Lab Archive compared to Traill's 1994 dictionary, and – more problematically – there are some words whose F0 contours analyzed here suggest that their tone marking should be further revised. I highlight and elaborate on these issues in the online supplementary materials (see Footnote 5). Nevertheless, the depressive effect of harsh and pharyngealized vowels on F0 is found regardless of the tone.

The third discriminant function LD3 largely helps discriminate the categories which were overlapping in the LD1–LD2 space (Fig. 5). The crucial comparisons are harsh versus pharyngealized voice, and breathy versus breathy-creaky voice; these two pairs of phonation type are well discriminated by LD3, but not by LD1 or LD2. The third discriminant function is most strongly correlated with mean H1*–H2*, mean HNR <500 Hz, and HNR <3,500 Hz. The linear mixed-effects models show that, compared with modal vowels, breathy vowels have higher H1*–H2* (indexing less vocal fold constriction) but lower HNR (both <500 Hz and <3,500 Hz), presumably due to the increase in aspiration noise. On the other hand, compared with harsh vowels, pharyngealized vowels have lower H1*–H2* (indexing greater vocal fold constriction) but higher HNR (indexing more periodicity) (Table 9).

The time courses of H1*–H2* are shown in Figure 6. As expected, breathy vowels show the highest values of H1*–H2* overall. Breathy-creaky and harsh vowels also have high values of the measure in the first third, after which the measure decreases, presumably due to increased constriction. Pharyngealized and creaky vowels do not differ significantly along this measure (Table 9), probably because pharyngealized vowels are also constricted, especially by the vowel midpoint.

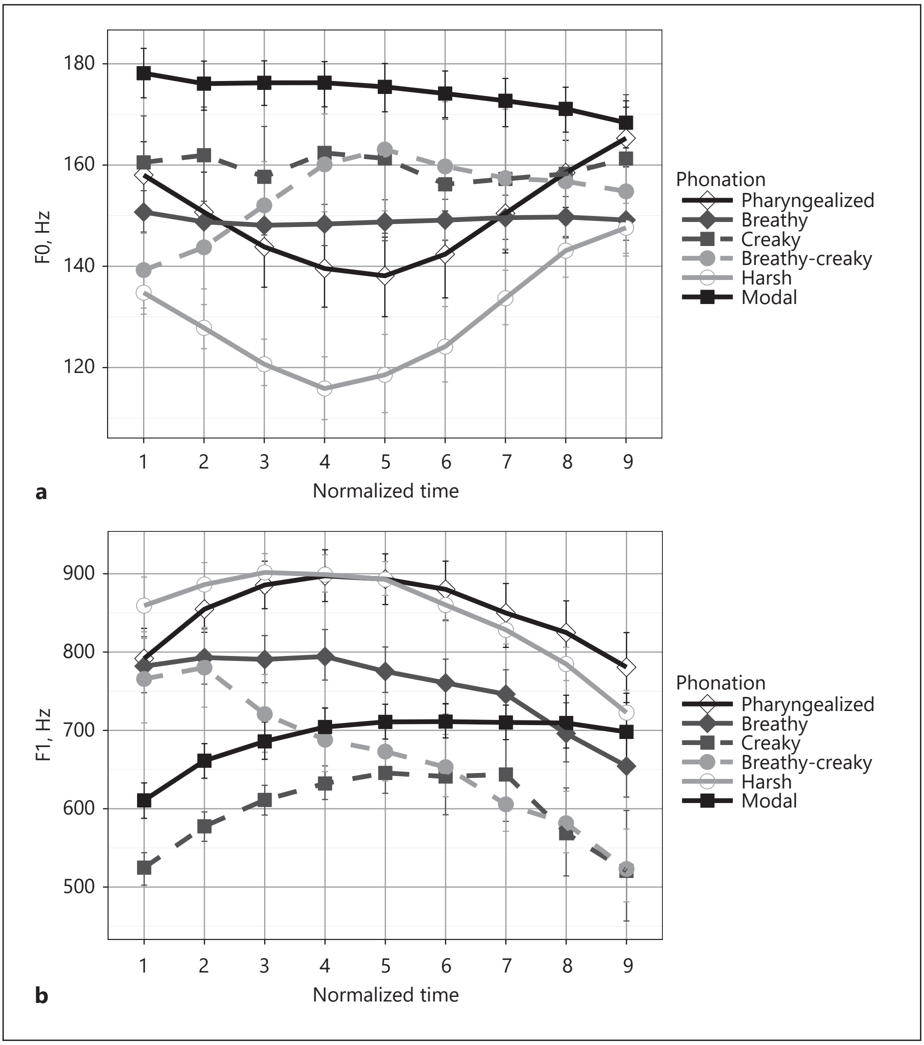

Fig. 8. Time courses of F0 (**a**) and F1 (**b**) by phonation type. Error bars indicate bootstrapped 95% confidence intervals of the means. Both measures are associated with linear discriminant (LD) 2, which differentiates pharyngealized and harsh vowels from other phonation types.

The time courses for the two noise measures that are closely correlated with the linear discriminants, HNR <500 Hz (correlated with LD1 and LD3) and HNR <3,500 Hz (correlated with LD2 and LD3), are shown in Figure 7. Both show low HNR values (i.e., high levels of noise) for harsh vowels, as well as for breathy vowels. Creaky vowels have lower HNR <500 Hz than modal vowels, but do not differ from modal vowels in terms of HNR <3,500 Hz (Tables 7, 8). This indicates that creaky vowels are characterized by more low-frequency noise, as opposed to more broadband noise

Table 9. Results of the linear mixed-effects models for mean measures correlated with LD3

	H1*–H2*			HNR <500 Hz			HNR <3,500 Hz		
	β	*t*	*p* value	β	*t*	*p* value	β	*t*	*p* value
Breathy	11.11	6.49	<0.001*	23.22	13.26	<0.001*	25.38	19.24	<0.001*
C	–8.72	–7.83	<0.001*	4.06	1.75	>0.05	10.54	4.65	<0.001*
BC	–6.11	–6.23	<0.001*	1.05	0.37	>0.05	6.79	2.90	<0.01*
H	–5.44	–5.30	<0.001*	–8.72	–4.74	<0.001*	–5.96	–3.09	<0.01*
P	–9.30	–10.10	<0.001*	7.71	3.25	<0.01*	5.95	2.73	<0.05*
M	–6.79	–5.52	<0.001*	13.87	7.34	<0.001*	12.88	6.09	<0.001*
Pharyngealized	1.81	1.26	>0.05	30.93	15.83	<0.001*	31.34	18.13	<0.001*
B	9.30	10.10	<0.001*	–7.71	–3.25	<0.01*	–5.95	–2.73	<0.05*
C	0.57	0.80	>0.05	–3.64	–1.63	>0.05	4.58	2.35	<0.05*
BC	3.18	3.66	<0.01*	–6.65	–3.25	<0.01*	0.83	0.47	>0.05
H	3.85	4.42	<0.001*	–16.43	–6.72	<0.001*	–11.91	–5.63	<0.001*
M	2.50	2.92	<0.05*	6.15	3.01	<0.05*	6.92	3.37	<0.01*

The five phonation types are all compared to a baseline phonation type (the intercept, indicated in italics). All models are of the structure *DV* ~ *Phonation* + (*Phonation*|*Speaker*) + (1|*Item*). LD, linear discriminant; HNR, harmonics-to-noise ratio; B, breathy; C, creaky; P, pharyngealized; BC, breathy-creaky; H, harsh; M, modal.

(similar results were found for English and Hmong in Garellek, 2012). Interestingly, compared to creaky vowels, pharyngealized vowels have lower degrees of low-frequency noise (i.e., they have higher HNR <500 Hz) but higher degrees of broadband noise (i.e., lower HNR <3,500 Hz). Thus, even though pharyngealized vowels are phonetically creaky (and are creaky at the midpoint, just like phonologically creaky vowels), it appears that phonologically creaky vowels and phonologically pharyngealized vowels differ in the way that their phonetic creakiness is realized acoustically. When laryngeal and pharyngeal constriction co-occur in !Xóõ, as with pharyngealized vowels, more high-frequency noise is generated; on the other hand, when laryngeal constriction occurs on its own, as for the creaky vowels, that creaky quality is dominated by low-frequency noise components. This also implies that different types of creaky voice can be differentiated in the voice model of Kreiman et al. (2014). Further, if the phonetic implementation of the creaky voice found in creaky vowels differs from that of the creaky voice found in pharyngealized vowels, this could help account for the very high classification accuracy for creaky versus pharyngealized vowels, in addition to the inclusion of F1.

Discussion

The results of the discriminant analysis reveal that the complex phonation system of !Xóõ is well differentiated acoustically, using parameters from the voice model of Kreiman et al. (2014). The average classification accuracy across the six phonation types that were analyzed was over 92%, well above chance. The phonation types are best distinguished by discriminant functions that are correlated in particular with HNR measures, low-frequency harmonic spectral slope (H1*–H2*), SoE, F1, and F0. Thus, all four classes of parameters from the voice model of Kreiman et al. (2014) were important in the LDA (Table 1): harmonic spectral slope (H1*–H2*), inharmonic noise (HNR), voicing frequency and amplitude (F0 and SoE), and vocal tract parameters (F1) all interact to help discriminate the six phonation types studied here.

Interestingly, measures averaged over the entire vowel – rather than changes in a measure over intervals of the vowel – were most strongly correlated with the discriminant functions. This is in spite of two important facts regarding the phonation system of !Xóõ: its complexity (having eight contrastive phonation types, six of which were analyzed in this study) and the dynamic realization of several phonation types; only modal and breathy vowels show fairly stable acoustic realizations of their respective voice qualities. By contrast, harsh and pharyngealized vowels have their non-modal phonation phased with the middle of the vowel, and creaky vowels with the middle or end of the vowel. Breathy-creaky vowels are inherently dynamic, transitioning from a breath-like quality to a creaky one. Thus, even if researchers perceive a highly variable realization of a particular phonation type, it is not necessarily the case that a temporal analysis is warranted to discriminate the contrasts statistically. Nonetheless, as is the case for !Xóõ as well as for any language, perception studies are needed to confirm that the measures which statistically help discriminate the phonation types are in fact the ones used by listeners to perceive the contrasts.

The Acoustics of Harsh Voice Associated with "Strident" Vowels

Another goal of the paper was to provide a multidimensional description of the acoustics of harsh voice as it occurs for the breathy-pharyngealized "strident" vowels in !Xóõ. Unlike the acoustics of breathy and creaky voice qualities, there has been relatively little work on how voice qualities with pharyngeal constriction and/or aryepiglottic trilling are realized acoustically (cf. Miller, 2007). Given the multidimensional articulatory attributes of harsh voice (which may include glottal spreading leading to breathy voice and sometimes voicelessness, along with aryepiglottic constriction and trilling, as well as pharyngeal narrowing), we expected that it could be measured acoustically by means of spectral tilt, noise, and resonance measures. This is indeed the case in !Xóõ: relative to modal vowels, harsh vowels begin with a higher spectral tilt (measured here using H1*–H2*; see Fig. 6b), but end with a lower spectral tilt, with a trajectory of the measure similar to that found for breathy-creaky vowels. That H1*–H2* is correlated with glottal opening (Kreiman et al., 2008; Samlan et al., 2013; Zhang, 2016) implies that vocal fold spreading is utilized early in harsh vowels (Traill, 1985, 1986). It also helps differentiate harsh from pharyngealized vowels, which in the language show low spectral tilt throughout their duration.

Harsh vowels are noisier than vowels of any other phonation type according to the HNR measures, which are strongly correlated with the first three linear discriminant functions. This is likely driven by the irregularity due to aryepiglottic trilling, in addition to the aspiration noise at the beginning of the vowel (Traill, 1985, 1986). The middle of harsh vowels, like pharyngealized, creaky, and breathy-creaky vowels, also shows a sharp decrease in voicing energy (as indexed by SoE), consistent with strong constriction produced either at the vocal folds or in the epilarynx.

Finally, harsh vowels (like pharyngealized ones) are higher in F1, which is consistent with epilaryngeal constriction and laryngeal raising (Traill, 1985; Laver, 1994;

Ladefoged & Maddieson, 1996). Taken together, these results provide phonetic support for the phonation type's phonological analysis as breathy-pharyngealized (Traill, 1985, 1986). Yet it is still unclear whether these effects are also found for other harsh vowels in !Xóõ – recall, only /a/ was investigated here.

Acoustic Multidimensionality of Phonation Contrasts

Voice quality differences can be measured using only parameters related to the "voice source" – the periodic energy derived from vocal fold vibration. In ongoing work (e.g., Keating et al., 2011, 2012), my colleagues and I have sought to characterize (using source measures like spectral tilt, as well as noise measures) the low-dimensional acoustic space for cross-linguistic uses of voice quality. Thus, in analyses of four (Keating et al., 2011) and subsequently ten (Keating et al., 2012) languages with non-modal phonation, we excluded vocal tract filter measures like F1, though it is clear that formant frequencies and bandwidths are important correlates of phonation contrasts in at least two of the languages studied, Mazatec and Southern Yi (Garellek & Keating, 2011; Kuang, 2011; Kuang & Cui, 2018). It is of course worthwhile to determine what low-dimensional space is needed for characterizing the harmonic and inharmonic *source* characteristics of voice quality, but it is also important to bear in mind that *filter* characteristics play a role in both phonation and register contrasts (Brunelle, 2012; Abramson et al., 2015; Brunelle & Kirby, 2016). Certainly, the results of this study show that first-formant frequency is a key parameter for distinguishing voice qualities that involve supraglottal sources of energy, like harsh and pharyngealized voice qualities in !Xóõ, from those with phonatory sources (modal, breathy, and creaky).

In the analysis of ten languages with non-modal phonation (Keating et al., 2012), only spectral tilt measures (H1*–H2*, as well as higher-frequency tilt measures) were important for the low-dimensional space, as determined by multidimensional scaling; surprisingly, noise measures were not significant correlates of the dimensions, even though it is clear that noise can statistically differentiate the phonation types in the same languages (Garellek & Keating, 2011; Keating et al., 2011; Kuang, 2011; Esposito, 2012; Garellek, 2012; Khan, 2012; Kuang, 2013a, 2013b). The results from this study suggest that once !Xóõ is included in the cross-linguistic analysis, noise measures like HNR <500 Hz and HNR <3,500 Hz should also emerge as relevant to the low-dimensional space of voice quality.

Mapping of Acoustic Parameters to Articulation

For researchers investigating the acoustic characteristics of voice quality differences, it is beneficial to make use of measures that relate back to parameters in the model by Kreiman et al. (2014). For one, this study shows that the parameters can be used to discriminate as many as six phonation types. More importantly, these measures are motivated a priori by a theory of voice that relates articulatory movements to perceptible acoustic measures (Garellek et al., 2016; Garellek, to appear).

What do the acoustic measures that emerged as important for distinguishing six phonation types in !Xóõ tell us about how these phonation types are articulated? H1*–H2* is higher with a more open glottis and/or decreased medial vocal fold thickness, and is lower with a more closed glottis and/or increased medial vocal fold thickness (Kreiman et al., 2008; Samlan et al., 2013; Zhang, 2016). Thus we see higher values of H1*–H2* for breathy vowels in !Xóõ, implying thinner and more open vocal

Table 10. Summary of acoustic measures that were important for distinguishing the six phonation types in !Xóõ and their articulatory correlates

Phonation type	Acoustic correlates	Articulatory correlates
Modal	High HNR (both measures) High SoE	Negligible noise, periodic voicing Strong voicing
Breathy	Low HNR (both measures) High SoE High H1*–H2*	Aspiration noise Strong voicing Open glottis
Creaky	Low HNR <500 Hz High HNR <3,500 Hz Low SoE Low H1*–H2*	Irregular voicing Negligible aspiration/supraglottal noise Weak voicing Glottal constriction
Pharyngealized	Moderate HNR <500 Hz Moderate HNR <3,500 Hz Low SoE Low H1*–H2*, drop in F0 High F1	Somewhat irregular voicing Some supraglottal noise Weak voicing Glottal constriction Laryngeal raising
Harsh	Low HNR <500 Hz Low HNR <3,500 Hz Low SoE High-to-low H1*–H2*, drop in F0 High F1	Irregular voicing Aspiration/supraglottal noise Weak voicing Open glottis to glottal constriction Laryngeal raising

Breathy-creaky vowels are excluded because they are defined by the acoustic characteristics of breathy vowels followed by the acoustic characteristics of creaky ones. HNR, harmonics-to-noise ratio; SoE, strength of excitation.

folds during voicing. For creaky and pharyngealized vowels, we see lower values of the measure, which implies thicker and more tightly adducted vocal folds, which are likely also the result of more general laryngeal constriction, especially for the pharyngealized vowels (Edmondson & Esling, 2006; Moisik & Esling, 2011; Moisik, 2012). For both breathy-creaky and harsh vowels, we see a change from high to low H1*–H2*, likely as vowels transition from an open-glottis configuration to one with a more closed glottis, which surely in the case of harsh vowels is a by-product of greater laryngeal constriction (Traill, 1985, 1986). The increased vocal fold adduction and laryngeal constriction for creaky, pharyngealized, breathy-creaky, and harsh vowels also cause weaker voicing, as seen by the lowered SoE, as well as irregular voicing, as seen by the decreased HNR <500 Hz for these phonation types. Finally, we see higher F1 for pharyngealized and harsh vowels, which indexes laryngeal raising. The extra-low HNR measures (both broadband and low-frequency) found for harsh vowels likely index aryepiglottic trilling and/or frication (i.e., pharyngeal noise), which is attested in the language (Traill, 1986). A summary of the relevant acoustic measures and their likely articulatory origins is shown in Table 10. The harsh vowels in !Xóõ are therefore probably produced in a manner similar, if not identical, to the “whispery growled harsh voice" described by Moisik (2012), with aryepiglottic trilling, irregular and weak voicing, and high airflow. Moreover, if creaky and breathy-creaky vowels involve general laryngeal constriction (rather than simply increased vocal fold thickness or vocal fold constriction), then it is also likely that breathy-creaky vowels are produced as whispery-creaky, since whisper involves an open glottis with laryngeal constriction (Esling & Harris, 2003). For both harsh and breathy-creaky vowels, more articulatory work is needed to determine how the open-glottis configuration in these phonation types is phased with respect to laryngeal constriction. But assuming that phonologically breathy-creaky vowels are produced as whispery-creaky (meaning that the laryngeal constriction co-occurs with an open glottis), it is unclear how breathiness and whisper are differentiated using the acoustic measures in this model; the HNR measures show that breathy vowels and breathy-creaky vowels have similar degrees of noise in both the lower and the higher end of the spectrum. Instead, they differ in the timing of the noise (Fig. 7). In sum, articulatory studies of the phonation types in !Xóõ, as well as further cross-linguistic work on articulatory-acoustic modeling of phonation, are still needed to confirm how particular acoustic behavior relates back to specific articulations.

Conclusion

The first goal of this study was to determine whether the acoustic differences between the six phonation types of !Xóõ can be adequately modeled using the psychoacoustic model of the voice elaborated by Kreiman et al. (2014) and Garellek et al. (2016). The results of the LDA showed that, despite the large number of contrastive phonation types and their dynamic complexity, the phonation system is very well discriminated using parameters from the model. Further, all classes of parameters – harmonic source spectral slope, noise, voicing frequency and amplitude, and formants – mattered for differentiating the phonation types, illustrating the utility of both source and filter measures in describing complex voice qualities.

The second goal of this study was to determine how best to characterize harsh voice, a voice quality whose acoustic attributes remain largely understudied. In !Xóõ, this voice quality occurs in the phonation type traditionally called "strident." The results indicate that harsh voice in !Xóõ has the following characteristics: (1) higher F1, like the pharyngealized phonation type in the language (consistent with laryngeal raising and pharyngealization); (2) higher H1*–H2* at the beginning of the vowel, like the breathy and breathy-creaky phonation types (consistent with its being analyzed phonologically as breathy-pharyngealized; Traill, 1985); and (3) more acoustic noise than the other phonation types, likely due to aryepiglottic trilling. Vowels with harsh voice also show a sharp drop in voicing rate and amplitude where constriction is strongest.

Together, these findings show how a psychoacoustic model of the voice can distinguish many different voice qualities, and how using such a model to analyze voice quality is generally recommended for the study of phonation types: it provides an a priori set of parameters to investigate; moreover, the parameters which distinguish voice qualities acoustically are also likely to be used by listeners, given that they are both perceptible and are systematically related to voice articulation.

DOI: 10.1159/000494301

Acknowledgements

I am especially grateful to John Esling, Bonny Sands, and Associate Editor Wentao Gu for their comments and suggestions on earlier versions of the manuscript. I would also like to thank Timothy Ressler for his help with corpus segmentation and annotation, as well as Adam Chong, Patricia Keating, Jianjing Kuang, and members of the UC San Diego Phonetics-Phonology group for helpful advice and feedback.

Disclosure Statement

There were no conflicts of interest during data collection, analysis, or publication of this article.

Appendix: Complete Wordlist

	Transcription	Spelling (Traill, 1994b)	Gloss
Breathy			
	!ā̤o	!āho	“slope”
	!ā̤le	!āhle	“stand on”
	!à̤la	!àhla	“Peeling-bark Ochna tree”
Modal			
	ᶢǂàa	ǂgàa	“exploit”
	ǂába	ǂába	“peg”
	ǁáa	ǁáa	“Camel Thorn tree”
Creaky			
	ᶢ!ᵡâˀje	g!xâ’je	“udder”
	ᶢ|àˀje	|gà’je	“bend”
Pharyngealized			
	qâˤa	qâ̰a	“long ago”
	táˤi	tá̰i	“far”
	ǂqàˤn	ǂqà̰n	“smooth”
Breathy-creaky			
	|à̤ˀje	|àh’je	“wait for”
	ᶢ|ā̤ˀbe	|gāh’be	“hide”
	ᵑ|ā̤ˀbe	|nāh’be	“surround it”
Harsh			
	ˀᵑ!àʢo	’!nà̰ho	“base”
	ˀᵑ!àʢle	’!nà̰hle	“lower it”
	ᴳǂàʢli	ǂɢà̰hli	“Bladethorn tree”

The transcriptions and glosses have been modified from those appearing on the UCLA Phonetics Lab Archive, reflecting the updated versions in Traill (1985) as well as the author’s transcriptions for distinguishing the six phonation types.

DOI: 10.1159/000494301

References

Abramson, A. S., Tiede, M. K., & Luangthongkum, T. (2015). Voice register in Mon: Acoustics and electroglottography. *Phonetica, 72*(4), 237–256. https://doi.org/10.1159/000441728

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 8*(1), 28. https://doi.org/10.5334/labphon.19

Alwan, A. (1989). Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *The Journal of the Acoustical Society of America, 86*(2), 549–556. https://doi.org/10.1121/1.398234

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using "Eigen" and S4* [Computer software program]. R package version 1.0-6.

Berkson, K. H. (2013). *Phonation types in Marathi: An acoustic investigation* (Unpublished Ph.D. thesis), University of Kansas, USA.

Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. In *MIT Speech Communication Working Papers* (Vol. 1, pp. 71–81). Cambridge, MA: MIT.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics, 30*(2), 163–191. https://doi.org/10.1006/jpho.2001.0155

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* [Computer program]. Version 5.4.04. Retrieved January 5, 2015 from http://www.praat.org/

Bradfield, J. (2014). Clicks, concurrency and Khoisan. *Phonology, 31*(1), 1–49. https://doi.org/10.1017/S0952675714000025

Brunelle, M., & Finkeldey, J. (2011). Tone perception in Sgaw Karen. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 372–375).

Brunelle, M. (2012). Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *The Journal of the Acoustical Society of America, 131*(4), 3088–3102. https://doi.org/10.1121/1.3693651

Brunelle, M., & Kirby, J. (2016). Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass, 10*(4), 191–207. https://doi.org/10.1111/lnc3.12182

Catford, J. C. (1983). Pharyngeal and laryngeal sounds in Caucasian languages. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 344–350). San Diego, CA: College-Hill Press.

DiCanio, C. (2014). Cue weight in the perception of Trique glottal consonants. *The Journal of the Acoustical Society of America, 135*(2), 884–895. https://doi.org/10.1121/1.4861921

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association, 39*(02), 162–188. https://doi.org/10.1017/S0025100309003879

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology, 23*(2), 157–191. https://doi.org/10.1017/S095267570600087X

Edmondson, J. A., Esling, J. H., Harris, J. G., Li, S., & Ziwo, L. (2001). The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. *Mon-Khmer Studies, 31*, 83–100.

Elderkin, E. D. (1989). *The significance and origin of the use of pitch in Sandawe* (Unpublished Ph.D. thesis), University of York, York, UK.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26*(2), 65–88. https://doi.org/10.1017/S0025100300006125

Esling, J. H. (1999). The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech, 42*(Pt 4), 349–372. https://doi.org/10.1177/00238309990420040101

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics, 50*(1), 13–44. https://doi.org/10.1353/cjl.2007.0007

Esling, J. H., Fraser, K. E., & Harris, J. G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics, 33*(4), 383–410. https://doi.org/10.1016/j.wocn.2005.01.003

Esling, J. H., & Harris, J. G. (2003). An expanded taxonomy of states of the glottis. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1049–1052). Barcelona.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong phonation. *Journal of Phonetics, 40*(3), 466–476. https://doi.org/10.1016/j.wocn.2012.02.007

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics, 40*(1), 152–161. https://doi.org/10.1016/j.wocn.2011.10.003

Garellek, M. (to appear). The phonetics of voice. In W. Katz & P. Assmann (Eds.), *The Routledge handbook of phonetics*. Abingdon-on-Thames, UK: Routledge.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association, 41*(2), 185–205. https://doi.org/10.1017/S0025100311000193

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America, 133*(2), 1078–1089. https://doi.org/10.1121/1.4773259

Garellek, M., Samlan, R., Gerratt, B. R., & Kreiman, J. (2016). Modeling the voice source in terms of spectral slopes. *The Journal of the Acoustical Society of America, 139*(3), 1404–1410. https://doi.org/10.1121/1.4944474

Gerratt, B. R., & Kreiman, J. (2001). Toward a taxonomy of nonmodal phonation. *Journal of Phonetics, 29*(4), 365–381. https://doi.org/10.1006/jpho.2001.0149

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics, 29*(4), 383–406. https://doi.org/10.1006/jpho.2001.0147

Greenberg, J. H. (1963). *The languages of Africa*. Bloomington, IN: Research Center in Anthropology, Folklore, and Linguistics, Indiana University.

Güldemann, T. (2013). Phonetics and phonology: Southern Khoesan (Tuu). In R. Vossen (Ed.), *The Khoesan languages* (pp. 75–83). New York, NY: Routledge.

Güldemann, T. (2014). “Khoisan” linguistic classification today. In T. Güldemann & A. M. Fehn (Eds.), *Beyond ‘Khoisan’: Historical relations in the Kalahari Basin* (pp. 1–41). Amsterdam, The Netherlands: John Benjamins.

Güldemann, T. (2016). Phonological regularities of consonant systems in genetic lineages of Khoisan. In R. Vossen & W. H. G. Haacke (Eds.), *Lone Tree: Scholarship in the service of the Koon* (pp. 159–207). Cologne, Germany: Rüdiger Köppe.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics*, 29(4), 451–480. https://doi.org/10.1006/jpho.2001.0146

Haspelmath, M. (2017). Some principles for language names. *Language Documentation and Conservation, 11*, 81–93.

Hess, S. A. (1998). *Pharyngeal articulations* (Unpublished Ph.D. thesis), UCLA, USA.

Honken, H. (2013). Genetic relationships: an overview of the evidence. In R. Vossen (Ed.), *The Khoesan languages* (pp. 13–23). New York: Routledge.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America, 121*(4), 2283–2295. https://doi.org/10.1121/1.2697522

Jongman, A., Herd, W., & Al-Masri, M. (2007). Acoustic correlates of emphasis in Arabic. In *Proceedings of the International Congress of Phonetic Sciences Saarbrücken* (pp. 913–916). Saarbrücken, Germany.

Kawahara, H., de Cheveigné, A., & Patterson, R. D. (1998). An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT-suite. In *ICSLP-1998*, paper 0659.

Keating, P., Esposito, C., Garellek, M., Khan, S., & Kuang, J. (2011). Phonation contrasts across languages. In *Proceedings of the International Congress of Phonetic Sciences, Hong Kong* (pp. 913–916). Hong Kong, China.

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Keating, P., Kuang, J., Esposito, C., Garellek, M., & Khan, S. (2012). *Multi-dimensional phonetic space for phonation contrasts*. Poster presented at LabPhon 2012.

Khan, S. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics, 40*(6), 780–795. https://doi.org/10.1016/j.wocn.2012.07.001

Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. *Loquens, 1*(1), e009.

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America, 117*(4 Pt 1), 2201–2211. https://doi.org/10.1121/1.1858351

Kreiman, J., & Gerratt, B. R. (2010). Perceptual sensitivity to first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America, 128*(4), 2085–2089. https://doi.org/10.1121/1.3478784

Kreiman, J., & Gerratt, B. R. (2012). Perceptual interaction of the harmonic source and noise in voice. *The Journal of the Acoustical Society of America, 131*(1), 492–500. https://doi.org/10.1121/1.3665997

Kreiman, J., Antoñanzas-Barroso, N., & Gerratt, B. R. (2016). The UCLA voice synthesizer, version 2.0. *The Journal of the Acoustical Society of America, 140*(4), 2961. https://doi.org/10.1121/1.4969149

Kreiman, J., Gerratt, B. R., & Khan, S. U. (2010). Effects of native language on perception of voice quality. *Journal of Phonetics, 38*(4), 588–593. https://doi.org/10.1016/j.wocn.2010.08.004

Kreiman, J., Iseli, M., Neubauer, J., Shue, Y.-L., Gerratt, B. R., & Alwan, A. (2008). The relationship between open quotient and H1*–H2*. *The Journal of the Acoustical Society of America, 124*(4), 2495. https://doi.org/10.1121/1.4782813

Kuang, J. (2011). *Production and perception of the phonation contrast in Yi* (Unpublished Master's thesis), UCLA, USA.

Kuang, J. (2013a). *Phonation in tonal contrasts* (Unpublished Ph.D. thesis), UCLA, USA.

Kuang, J. (2013b). The tonal space of contrastive five level tones. *Phonetica, 70*(1-2), 1–23. https://doi.org/10.1159/000353853

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America, 142*(3), 1693–1706. https://doi.org/10.1121/1.5003649

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214. https://doi.org/10.1016/j.wocn.2018.09.002

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models. R package version 2.0-20*. http://CRAN.R-project.org/package=lmerTest

Ladefoged, P. (1983). The linguistic use of different phonation types. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 351–360). San Diego, CA: College-Hill Press.

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages* (2nd ed.). Malden, MA: Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, P., & Traill, A. (1984). Linguistic phonetic descriptions of clicks. *Language, 60*(1), 1–20. https://doi.org/10.2307/414188

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166621

McCarthy, J. J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III* (pp. 191–233). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.012

Miller, A. L. (2007). Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics, 35*(1), 56–84. https://doi.org/10.1016/j.wocn.2005.11.001

Miller, A. L. (2010). *Guttural co-occurrence restrictions: Articulatory goals and adjacency*. Poster presented at LabPhon 12.

Miller-Ockhuizen, A. (1998). Towards a unified decompositional analysis of Khoisan lexical tone. In M. Schladt (Ed.), *Language, identity, and conceptualization among the Khoisan* (pp. 913–916). Cologne, Germany: Rüdiger Köppe.

Miller-Ockhuizen, A. L. (2003). *The phonetics and phonology of gutturals: a case study from Ju|'hoansi*. New York, London: Routledge.

Misnadin, Kirby, J. P., & Remijsen, B. (2015). Temporal and spectral properties of Madurese stops. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Mittal, V. K., Yegnanarayana, B., & Bhaskararao, P. (2014). Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America, 136*(4), 1932–1941. https://doi.org/10.1121/1.4894789

Moisik, S. R. (2012). Harsh voice quality and its association with blackness in popular American media. *Phonetica, 69*(4), 193–215. https://doi.org/10.1159/000351059

Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association, 44*(1), 21–58. https://doi.org/10.1017/S0025100313000327

Moisik, S. R., & Esling, J. H. (2011). The 'whole larynx' approach to laryngeal features. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong, China.

Murty, K., & Yegnanarayana, B. (2008). Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(8), 1602–1613. https://doi.org/10.1109/TASL.2008.2004526

Naumann, C. (2008). High and low tone in Taa ǂaan (!Xóõ). In S. Ermisch (Ed.), Khoisan languages and linguistics: *Proceedings of the 2nd International Symposium, Quellen zur Khoisan-Forschung* (Vol. 22, pp. 279–302). Cologne, Germany: Rüdiger Köppe.
Naumann, C. (2011). A preliminary classification of Taa dialects. In *20th International Conference for Historical Linguistics*. Osaka, Japan.
Naumann, C. (2014). Towards a genealogical classification of Taa dialects. In T. Güldemann & A. M. Fehn (Eds.), *Beyond 'Khoisan': Historical relations in the Kalahari Basin* (pp. 283–301). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/cilt.330.11nau
Naumann, C. (2016). The phoneme inventory of Taa (West !Xoon dialect). In R. Vossen & W. H. G. Haacke (Eds.), *Lone Tree: Scholarship in the service of the Koon: Essays in memory of Anthony T. Traill* (pp. 311–351). Cologne, Germany: Rüdiger Köppe.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Risdal, M., Aly, A., Chong, A. J., Keating, P., & Zymet, J. (2016). *On the link between glottal vibration and sonority*. Poster presented at LabPhon 15.
Samlan, R. A., & Kreiman, J. (2014). Perceptual consequences of changes in epilaryngeal area and shape. *The Journal of the Acoustical Society of America, 136*(5), 2798–2806. https://doi.org/10.1121/1.4896459
Samlan, R. A., Story, B. H., & Bunton, K. (2013). Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *Journal of Speech, Language, and Hearing Research: JSLHR, 56*(4), 1209–1223. https://doi.org/10.1044/1092-4388(2012/12-0194)
Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1846–1849). Hong Kong, China.
Sjölander, K. (2004). *Snack sound toolkit*. Stockholm, Sweden: KTH.
Snyman, J. W. (1977a). The interrupted juxtaposed vowels of Žhu|'hõasi. In A. Traill (Ed.), *Khoisan linguistic studies* (Vol. 3, pp. 93–103). Johannesburg: Witwatersrand University Press.
Snyman, J. W. (1977b). The pressed vowels of Žhu|'hõasi. In J. W. Snyman (Ed.), *Bushman and Hottentot linguistic studies: Papers of seminar held on 25 October 1975 Pretoria* (pp. 48–73). Pretoria, South Africa: University of South Africa Press.
Thurgood, E. (2004). Phonation types in Javanese. *Oceanic Lingustics, 43*(2), 277–295. https://doi.org/10.1353/ol.2005.0013
Traill, A. (1985). *Phonetic and phonological studies of the !Xóõ bushmen*. Hamburg, Germany: Buske.
Traill, A. (1986). The laryngeal sphincter as a phonatory mechanism in !Xóõ Bushman. In R. Singer & J. K. Lundy (Eds.), *Variation, culture and evolution in African populations: Papers in honour of Dr. Hertha de Villiers* (pp. 123–131). Johannesburg, South Africa: Witwatersrand University Press.
Traill, A. (1991). Pulmonic control, nasal venting, and aspiration in Khoisan languages. *Journal of the International Phonetic Association, 21*(01), 13–17. https://doi.org/10.1017/S002510030000596X
Traill, A. (1994a). The perception of clicks in !Xóõ. *Journal of African Languages and Linguistics, 15*(2), 161–174. https://doi.org/10.1515/jall.1994.15.2.161
Traill, A. (1994b). *A !Xóõ dictionary*. Cologne, Germany: Rüdiger Köppe.
Traill, A., & Vossen, R. (1997). Sound change in the Khoisan languages: New data on click loss and click replacement. *Journal of African Languages and Linguistics, 18*(1), 21–56. https://doi.org/10.1515/jall.1997.18.1.21
Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://doi.org/10.1007/978-0-387-21706-2
Vossen, R. (2013). Introduction. In R. Vossen (Ed.), *The Khoesan languages* (pp. 1–12). New York, NY: Routledge. https://doi.org/10.4324/9780203084465
Zawaydeh, B. A. (1999). *The phonetics and phonology of gutturals in Arabic* (Unpublished Ph.D. thesis), Indiana University, USA.
Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America, 139*(4), 1493–1507. https://doi.org/10.1121/1.4944754
Zhang, Z., Kreiman, J., Gerratt, B. R., & Garellek, M. (2013). Acoustic and perceptual effects of changes in body layer stiffness in symmetric and asymmetric vocal fold models. *The Journal of the Acoustical Society of America, 133*(1), 453–462. https://doi.org/10.1121/1.4770235