On H1–H2 as an acoustic measure of linguistic phonation type

Yuan Chai[a] and Marc Garellek[b]

Department of Linguistics, University of California San Diego, La Jolla, 92093, USA

(Dated: 13 October 2022)

[This paper is part of a special issue on Reconsidering Classic Ideas in Speech Communication.]

The measure H1–H2, the difference in amplitude between the first and second harmonic, is frequently used to distinguish phonation types and to characterize differences across voices and genders. While H1–H2 can differentiate voices and is used by listeners to perceive changes in voice quality, its relation to voice articulation is less straightforward. Its calculation also involves practical issues with error propagation. This paper highlights some developments in the use of H1–H2 and proposes a new measure that we call "residual H1." In residual H1, the amplitude of the first harmonic is normalized against the overall sound energy (as measured by Root-mean-square Energy) instead of against H2. Residual H1 may mitigate some of the issues with using H1–H2. The current study tests the correlation between Residual H1 and electroglottographic contact quotient (CQ) and compares the ability of residual H1 vs H1–H2 to differentiate statistically across phonation types in !Xóõ and utterance-level changes in phonatory quality in Mandarin. The results show that residual H1 has a stronger correlation with CQ and differentiates contrastive and allophonic phonatory quality better than H1–H2, particularly for more constricted phonation types.

[a]yuc521@ucsd.edu

[b]mgarellek@ucsd.edu

I. INTRODUCTION

The acoustic measure H1–H2, also known as L_1–L_2 (Titze *et al.*, 2015), refers to the difference in amplitude between the first and second harmonics. It is probably the most widely-used voice quality measure in linguistic phonetic research, and correlates with changes in phonation type (e.g. breathy vs. modal vowels) in many languages (Esposito and Khan, 2020; Gordon and Ladefoged, 2001), as well as non-phonemic changes in phonation (Hanson *et al.*, 2001; Li *et al.*, 2020; Ní Chasaide and Gobl, 1993). In terms of aerodynamics, H1–H2 reflects the amount of airflow through the glottis (Sundberg and Gauffin, 1979); in terms of voice articulation, the measure is related to vocal fold (and, perhaps more broadly, laryngeal) constriction vs. spreading. Generally, lower values of H1–H2 are associated with lower glottal Open Quotient (OQ), more constriction, and increased medial vocal fold thickness (Kreiman *et al.*, 2012; Samlan and Story, 2011; Zhang, 2016b).

The relationship between H1–H2, aerodynamics, and voice articulation is better studied than for any other acoustic measure of phonatory quality. Nevertheless, researchers occasionally find that phonation types are not distinguished by H1–H2, even when non-modal phonation is perceptually strong (Esposito, 2012; Garellek and Esposito, 2021). This suggests that H1–H2 may not be ideally suited for indexing changes in vocal fold constriction as generally thought, and/or that the measure can be refined in some way. In this paper, we review the history and use of H1–H2, particularly in linguistic studies of phonation type, and discuss the possible reasons for a lack of effect of H1–H2 when distinguishing modal vs. non-modal phonation types. We then motivate the use of a related measure – residual

H1 – to compare modal vs non-modal phonation. Residual H1 is a measure of H1 controlled for overall sound pressure level (SPL). We show that it is as effective as H1–H2 at discriminating modal vs. non-modal phonation in contrastive and allophonic uses of phonation type; further, its use mitigates certain issues inherent to the use of H1–H2.

Moving forward, we first highlight some notes on terminology. We will use "H1–H2" instead of "L_1–L_2" because the former is more widely known, particularly in linguistic phonetic research. "H1–H2" will be used to refer to any measure that compares the amplitude of the fundamental (H1) to that of the second harmonic (H2). But in any given study, H1–H2 can be measured in different ways. If estimated from the voice source (e.g. using inverse filtering), we will refer to the measure as "source H1–H2." If calculated from the audio output without any correction, we will call the measure "uncorrected H1–H2." Finally, if the measure has been obtained from the audio output but corrected for formant frequencies and bandwidths, we will refer to the measure as H1*–H2*, which is in line with current practice. In what follows, we will discuss the reasons for the existence of these different versions of the measure. Our paper also motivates a new measure called "residual H1." This measure will always be corrected for formant frequencies and bandwidths and so will appear with an asterisk as "residual H1*" unless discussed more abstractly. In early work that included uncorrected H1 from the audio output, we will refer to that measure as "uncorrected H1." Finally, H1 estimated at the voice source will be called "source H1," and when referring to both H1 and H1–H2, we will occasionally use "H1(–H2)."

A. The origins of measuring H1(–H2)

We have known since the 1960s that the roll-off or "tilt" of the harmonic spectrum varies as a function of different phonation types. In her pioneering study, Fischer-Jørgensen (1967) described the acoustic differences between modal ("clear") and breathy ("murmured") vowels in Gujarati. Through visual inspection of audio spectra, she found that the most important acoustic distinction between these phonation types is the amplitude of the first harmonic (the fundamental, i.e. uncorrected H1). She found that, for Gujarati breathy vowels, uncorrected H1 is generally stronger than for modal vowels.

Fischer-Jørgensen was surprised by this finding: "I had expected to find some extra noise in the breathy vowels, but instead I found a reinforcement of the fundamental" (p. 71). Further, she knew that to measure H1 on its own would present a confound between phonation differences and differences in sound intensity; for instance, a stronger uncorrected H1 may be due to increased breathiness, but it can also result from an overall higher sound pressure level. The way to disambiguate between these hypotheses is to normalize for SPL in some way. Fischer-Jørgensen did so by subtracting H2 from H1. If the overall signal is relatively strong, this should affect both H1 and H2 equally; thus, H1–H2 can index the strength of the fundamental while normalizing for any differences in SPL across tokens.

The choice to normalize for SPL using H2, rather than another spectral landmark, was not motivated *a priori*; indeed, Fischer-Jørgensen also normalized the fundamental relative to other uncorrected harmonics like H3 and the amplitudes of formants 1–4. Interestingly, she found that the spectral tilt differences between breathy and modal vowels were not consistently found across all measures: uncorrected H1 was indeed stronger in breathy vowels compared to modal ones when normalized against uncorrected H2 or the amplitudes of F1, F2, and F4 but not when normalized against uncorrected H3 or the amplitude of F3. At any rate, the implications of her decision to normalize H1 against H2 remain today: even though H1–H2 involves two harmonic amplitudes, the assumption – usually tacit – is that what we wish to compare across phonation types is a difference in the strength of the fundamental, that is to say H1.

B. H1–H2 and its relation to articulation and perception

While Fischer-Jørgensen (1967) established the importance of H1 and spectral tilt measures like H1–H2 as correlates of a breathy-modal contrast, it remained unclear precisely why breathy vowels have a relatively stronger fundamental than modal ones. This question, though still not fully resolved, has been addressed since the 1970s. Stevens (1977) used models of the transglottal area to schematize overall spectral tilt differences as a function of the degree of inter-arytenoid space, predicting that creaky phonation should have the lowest spectral tilt and breathy phonation the highest. Yet H1 and H1–H2 were not explicitly discussed. In fact, his depictions of differences in spectral tilt suggest that he would not have predicted differences in H1(–H2) for modal vs. creaky phonation; see Figure 5 in that paper, where the increased tilt is schematized in the higher frequencies only. Still, this work is important in highlighting how different vocal fold configurations, and in particular how changes to the cartilaginous glottis, could affect spectral tilt. Around the same time, Sundberg and Gauffin (1979) found that source H1 is related to overall airflow through the glottis. They showed how differences in source H1 are related to changes in overall SPL, and can be regulated by changes in the degree of vocal fold contact during voicing.

Another landmark study about H1(–H2) is the MIT Speech Communication working paper by Bickley (1982), who built on Fischer-Jørgensen's findings for Gujarati, as well as the preliminary analysis by Ladefoged (1981) of phonation types in !Xóõ. Bickley noted (p. 74–76) that the inverse-filtered glottal source in Gujarati breathy vowels had more symmetrical pulses than that of modal vowels, with less abrupt closure and shorter closed intervals [i.e., with higher glottal open quotient (OQ)]. Further development in voice source models, such as the LF model of Fant *et al.* (1985), also showed how differences in overall pulse shape relate to changes in spectral tilt. For example, Fant and Lin (1988) described how changes in various LF model parameters modulate source H1 relative to H2 and to other harmonics; for a recent overview and reassessment, see Gobl and Ní Chasaide (2019).

Bickley (1982) also conducted what is likely the first perceptual assessment of H1–H2, though it is preliminary by today's standards. In one experiment, she presented two listeners (one a native speaker of English, the other of Gujarati) with ten tokens of breathy vowels from !Xóõ, and asked them to rate the tokens on a four-point scale from "very breathy"-sounding to "not breathy"-sounding. The tokens that sounded very breathy had uncorrected H1–H2 values greater than 10 dB; the two tokens that sounded not breathy had uncorrected H1–H2 values of -4 and 4 dB. She also resynthesized a continuum from modal to breathy vowels [i, a, o] using an earlier version of the Klatt synthesizer (Klatt and Klatt, 1990), in which uncorrected H1 of the "breathy" vowel was equal to that of the "modal" vowel, or was higher by 9, 12, and 15 dB. The amplitude of spectral noise also varied (orthogonally to uncorrected H1) in 5-dB increments over a range of 20 dB. The resynthesized vowels were then spliced onto natural tokens of CV(C) words. Four native speakers of Gujarati were presented with the stimuli, and were asked to do a two-alternative forced-choice task with minimal pairs; that is, they chose whether they heard a word with a breathy vowel (e.g. [bi̤] "be afraid") or one with a modal vowel (e.g. [bi] "seed"). The results showed, perhaps surprisingly, that the level of aspiration noise did not affect listeners' choice. However, an increase in uncorrected H1 was associated with an increase in breathy vowel responses.

Klatt and Klatt (1990) outline several studies relating to H1(–H2) and the relationship between articulation, acoustics, and perception. They explicitly argue that source H1 is related to increased OQ, which in turn is related to the size of the posterior glottal opening (i.e., the cartilaginous glottis). Holmberg *et al.* (1995) found that uncorrected H1–H2 was correlated with airflow and electroglottographic (EGG) in English speakers. Similar relationships between H1–H2 and OQ have been found in studies using natural speech (Sundberg *et al.*, 1999) and resynthesized/simulated data (Stevens and Hanson, 1995). In studies of linguistic phonation, EGG contact quotient (CQ) (sometimes also quoted as "OQ") is also reasonably well correlated with H1–H2; for example, DiCanio (2009) reports adjusted R^2 values ranging from 0.3 to 0.46 between EGG OQ and uncorrected H1–H2 across the registers (phonation types) of Takhian Thong Chong, and Kuang (2011) reports an R^2 of 0.3 between EGG CQ and H1*–H2* in Southern Yi.

The relationship between H1–H2 and OQ, though robust, is often found to be weak and/or non-linear. Using inverse filtering of oral airflow, Hanson (1995) obtained glottal waveform and measured its OQ. Of four speakers total, three speakers showed a trend whereby larger

OQ was related to higher H1*–H2* (pp. 81, 85–86). Kreiman *et al.* (2012) found that, although source H1–H2 is closely correlated to OQ, this relationship varies considerably across speakers. H1–H2 is also correlated with other articulatory or aerodynamic parameters, including increased vocal fold process separation (Samlan *et al.*, 2013), increased medial vocal fold thickness (Zhang, 2016a), and glottal skew or symmetry (Doval and d'Alessandro, 1997; Doval *et al.*, 2006; Henrich *et al.*, 2001; Kreiman *et al.*, 2012; Shue *et al.*, 2010; Swerts and Veldhuis, 2001). The effect of glottal open quotient on source H1–H2 further interacts with glottal skew: Gobl *et al.* (2018) and Gobl and Ní Chasaide (2019) found that, when glottal OQ was high, source H2 was affected by glottal skew, while source H1 was mostly independent of glottal skew. More skewed pulses were related to higher source H2. And computational simulations have shown that the relationship between H1*–H2* and vocal fold process separation is non-linear, such that H1*–H2* first increases but then decreases with increasingly large separation (Samlan and Story, 2011).

As mentioned earlier, sometimes H1–H2 does not "behave" as expected (e.g. by not showing a difference across phonation types). The findings from the aforementioned studies imply that the reason for the occasional unexpected behavior of H1–H2 may be that it is affected by other factors that are unrelated to glottal open quotient. Still, support for the continued use of H1–H2 comes from Kreiman *et al.* (2007), who tested the correlation between spectral shape and glottal pulse shape and 78 acoustic measures (e.g. H1–H2, H2–H4, and slope of spectrum at different frequency intervals). Using correlation and principal component analyses, they found that the 78 acoustic measures can be reduced to just four independent ones: source H1–H2, source H2–H4, overall spectral slope, and high-frequency noise. Importantly, they also found that source H1-H2 is related to variability in spectral and glottal pulse shape, and stated that its measured values did not differ appreciably as a function of different glottal source models, implying that H1–H2 is robust to measurement artifacts.

Additional support for using H1–H2 comes from perceptual studies. Since Bickley (1982) [confirmed by Klatt and Klatt (1990)], researchers have found that changes in H1–H2 correlate with perceived changes in breathiness. More recently, Esposito (2010) found that uncorrected H1–H2 correlated with perceived breathiness, regardless of whether listeners' native language was Gujarati (with contrastive breathiness), English (with allophonic breathiness), or Spanish (with no breathiness), though the language groups relied on uncorrected H1–H2 to differing degrees. Garellek *et al.* (2013) systematically manipulated source H1–H2 (as well as other spectral tilt measures) in White Hmong, a language with contrastive breathy voice on a particular lexical tone. They found found that, controlling for all other parameters, an increase in source H1–H2 or source H2–H4 led to more "breathy tone" responses. Finally, studies of the perceptual sensitivity to the harmonic source spectrum have shown that listeners of various languages are sensitive to changes in source H1–H2, though this varies by language (Kreiman and Gerratt, 2010; Kreiman *et al.*, 2010); the just-noticeable differences for source H1–H2 were comparable to those for source H2–H4 and source H4–2 kHz, suggesting that listeners are particularly sensitive to changes in the lower-frequency harmonic source spectral slope (Garellek *et al.*, 2016b).

C. Filter correction and H1*–H2*

In the audio output, harmonic amplitudes from the voice source are significantly affected by the filter function. This leads to a problem with using uncorrected H1–H2, especially when comparing tokens with different vowel qualities: if uncorrected H1–H2 for [i] = H1–H2 for [a], is this because the two vowels' H1–H2 values are the same at the voice source, or could it be because the filtering effects of the vocal tract have resulted in the same output uncorrected H1–H2? Thus, the use of H1–H2 – indeed of all spectral tilt measures – becomes less informative of phonatory quality when compared across different vowel categories, because any effect of phonatory quality could be obscured by the influence of the filter. Hanson (1995, 1997) proposed the formula $20 * log_{10}[F1^2/F1^2 - f^2]$, where f refers to the harmonic frequency that need a formant correction. The product of the formula is subtracted from H1 and H2. The value after subtraction reflects the amplitude of H1 and H2 before its being affected by a formant of similar frequency.

Until Hanson (1995, 1997), uncorrected H1–H2 was measured from the audio signal generally for tokens with low vowels of the same quality. Alternatively, some studies relied on inverse filtering to subtract the effects of the vocal tract filter, thereby approximating source H1–H2 (e.g. Bickley, 1982; Huffman, 1987). Other studies (rightfully) avoided using uncorrected H1-based measures because formant correction was not widely used at the time. For example, in their study of tongue root contrasts in Maa, Guion *et al.* (2004) used A1–A2 (the difference in amplitude between F1 and F2) instead of uncorrected H1–H2 because the vowels differed in quality, with some vowels of interest having low F1 (which would interfere with H1 and H2 estimation); see discussion in Section 2.3.2 of that paper. Hanson's correction and subsequent versions (e.g. Iseli *et al.*, 2007) have enabled researchers to correct for differences in formant frequency and bandwidths without the need of specialized equipment, such as a Rothenberg mask, for inverse filtering. Today, corrected spectral tilt measures, denoted with asterisks (as in "H1*–H2*"), are the norm in linguistic research on phonation types because they enable researchers to compare tokens with differing formant structures, even within vowel category (Garellek, 2022).

D. The effectiveness of H1(–H2) in distinguishing phonatory qualities

Studies have found that H1–H2, and sometimes also H1, is an effective measure at distinguishing differences in phonation between women and men, between contrastive phonation types, and between lexical stress and phrasal accent. In this section, we review the pioneering studies in this area. After the study of Fischer-Jørgensen (1967) on Gujarati modal and breathy vowels, Ladefoged (1981) conducted a preliminary analysis of breathy ("murmured") vs modal ("clear") vowels in !Xóõ, in which he found that the amplitude of uncorrected H1 was higher for breathy vowels. Bickley (1982) measured uncorrected H1–H2 from spectra of breathy and modal vowels produced by ten speakers of !Xóõ (based on recordings made by Tony Traill and Peter Ladefoged) and by four speakers of Gujarati. She found that, in both languages, breathy vowels consistently had higher H1–H2 than modal vowels. There were also large cross-speaker and between-language variations in these H1–H2 comparisons, leading to the important assumption that phonation differences should be measured not in absolute but instead in relative terms, ignoring the raw values of H1–H2. We note here that it remains an open question whether absolute values are informative for H1–H2, but they likely can be. (For example, while voice onset time (VOT) is also compared relatively – voiceless unaspirated stops have lower VOT than voiceless aspirated ones – it is also the case that we generally don't expect unaspirated stops to have a VOT greater than about 35 ms; see discussion by Cho and Ladefoged (1999) and Chodroff *et al.* (2019).) Finally, Bickley called attention to the presence, though variable, of increased spectral noise for breathy vowels. We now know that spectral noise is a very important component to distinguishing phonation types (Garellek, 2019; Gordon and Ladefoged, 2001).

Maddieson and Ladefoged (1985) showed that, in four Tibeto-Burman languages with "tense" (more constricted) vs. "lax" (breathier) vowels (Hani, Jingpho, Yi, and Wa), lax (breathier) phonation had higher uncorrected H1–H2 than "tense" (more constricted) vowels. This was perhaps the first journal article making use of H1–H2 to characterize phonation types that are more constricted than modal voice. Although earlier work did investigate spectral tilt differences between modal and constricted phonation types, we are not aware of a previous study that specifically measured H1–H2; for example, the investigation by Ladefoged (1983) and Kirk *et al.* (1984) of breathy, modal, and laryngealized vowels in !Xóõ and Mazatec measured H1–A1.

In what may have been the first study to measure H1–H2 for consonants, Traill and Jackson (1988) investigated the acoustic differences between breathy and modal nasals in Tsonga, and their effects on following vowels. They measured uncorrected H1–H2, as well as other spectral tilt measures, during the nasal consonant as well as the vowel onset. Generally they found large differences within and across speaker gender on all spectral tilt measures, but reported that breathy vs. modal nasals were more effectively distinguished by two higher spectral tilt measures (the difference in slope between H1 and the harmonic nearest 1400 Hz, and between H1 and the strongest harmonic above 2000 Hz) than by H1–H2 (cf. more recent discussion of the acoustics of breathiness and nasality by Garellek *et al.* 2016a; Simpson 2012; Styler 2017; Tabain *et al.* 2022). Few differences in H1–H2 at vowel onset were found. Since the 1980s, H1–H2 has been used to quantify sex/gender-based differences in phonatory quality. For example, in their investigation of male vs. female voice differences among speakers of British English, Henton and Bladon (1985) used uncorrected H1–H2 to measure whether female speakers of two dialects (Received Pronunciation and Modified Northern British English) differ from male speakers in terms of voice quality. They found that, in both dialects, H1–H2 was higher for women than for men. Hanson and Chuang (1999) compared the spectral tilt in the production by male speakers with the data of female speakers collected from Hanson (1995, 1997), and found that female speakers had higher (by about 3dB) and larger standard deviation of H1*–H2* than male speakers. They suggested that such differences in H1*–H2* indicated that, in terms of voice articulation, female speakers have a larger OQ than male speakers.

In the 1990s there was also work investigating spectral tilt as a correlate of lexical stress and phrasal accent and other phonological contrasts. Sluijter and van Heuven (1996) were perhaps the first to investigate this (for Dutch), but they measured energy in four frequency bands and not H1–H2. In their extension of that work, Campbell and Beckman (1997) measured uncorrected H1–H2 (labeled there as "H2–H1") but did not find it to be a reliable correlate of lexical or phrasal prominence. They suggested (p. 70) that this might be due to the fact that two (of four) speakers produced the low intonation tone with creaky voice. Subsequent work on other languages has confirmed that H1–H2 can indeed correlate with lexical/post-lexical prominence (Caballero and Carroll, 2015; Garellek and White, 2015; Guion *et al.*, 2010).

E. Issues with H1–H2

The association between H1–H2 and glottal OQ, as well as the fact that listeners readily use H1–H2 to perceive changes in voice quality (Esposito, 2010; Garellek *et al.*, 2013, 2016b; Kreiman and Gerratt, 2010; Kreiman *et al.*, 2010), have contributed to the popularity of H1–H2 as a measure of phonatory quality and voice quality more broadly (Garellek, 2022). However, using H2 to normalize for SPL is theoretically arbitrary. Many studies also rely on other landmarks; see Garellek (2019) for an overview. Further, Sundberg (2022) found that H1 and H2 (denoted there as L_1 and L_2) were affected differently by the influence of subglottal pressure: H2 is more sensitive to the pulse amplitude increase than H1, resulting in an inverse relationship between H1–H2 and pulse amplitude. Yet as we mentioned earlier, the use of H2 to normalize the SPL of the signal is based on the assumption that SPL affects H1 and H2 equally. The conclusions from Sundberg (2022) thus provide more support for avoiding the use of H2 as a normalizing landmark.

Compared with H1, measuring H1–H2 is also likely to be more prone to error propagation; that is, the transferring of uncertainties in the input variable(s) to the output variable(s) (Arras, 1998). H1–H2 involves the calculation of two measures – H1 and H2 – whereas H1 involves only one. The correct estimation of H1–H2 thus requires the correct estimation of both H1 and H2. Error propagation in H1–H2 can be attributed to two sources. First, H2 is estimated based on H1, which requires accurate estimation of f0. Yet H1–H2 is often used to measure non-modal voices. Occasionally the aspiration noise in breathy voice, and frequently the decrease in periodicity in creaky voice, can make the correct tracking of f0 difficult or impossible (Garellek, 2019; Keating *et al.*, 2015; Kuang, 2017). Any error in the estimation of f0 will influence both H1 and H2, leading to more calculation error than if H1 alone were estimated.

Another issue arises in the application of the filter correction for H1*–H2*. Although corrected H1*–H2* is now the norm when measuring H1-H2 across vowels qualities, errors in estimating formant frequencies and amplitude inevitably arise. For example, when a token has a high f0 and a low F1 (i.e. for a high vowel), it is possible for the tracking algorithm to mistake f0 for F1 and F1 for F2. Another common error for formant estimation occurs when F1 and F2 are similar in frequency (e.g. for back vowels); in such cases, F1 and F2 are likely to be mistaken for a single formant, with the real F3 consequently being mistaken for F2. When formant frequencies and amplitudes are thus miscalculated, the corrected H1* and H2* are highly likely to be erroneous as well. We hypothesize that H1* is less likely to have such errors than H1*–H2* because erroneous formant tracking will influence H1* only once, but will affect H1*–H2* twice– when estimating each harmonic level.

A further issue arises with (even slight) nasality: the first nasal pole (P0) can increase the amplitude of either H1 or H2, depending on the f0 (Dang and Honda, 1996; Simpson, 2012; Styler, 2017). Nasal zeroes will further attenuate the oral resonances (Dang and Honda, 1996; Simpson, 2012; Styler, 2017). P0 is usually in the range of 200–450 Hz (Styler, 2017), so for typical adult male speakers with an f0 of 120 Hz, P0 is more likely to influence H2 (240 Hz) and H3 (360 Hz). But for typical adult female speakers with an f0 above 200 Hz, P0 is more likely to influence H1 (Simpson, 2012). As a result, when a token contains nasalization, H1–H2 will inevitably be influenced by P0 in unpredictable, f0-dependent ways. Admittedly, measuring H1 alone does not fully avoid these issues. We advise then that f0 should be used as a control variable when analyzing either H1 or H1–H2.

II. ADDRESSING ISSUES WITH H1(–H2)

A. Meta-analysis of H1(–H2) in studies of linguistic phonation type

To review the effectiveness of H1–H2 at distinguishing phonation types, we conducted a meta-analysis of studies that compare H1(–H2) between contrastive phonation types in a given language. We focus here on contrastive phonation types, because we expect the H1(–H2) differences to be relatively large and consistent across speakers. In addition to H1–H2, we also include studies that measure H1, a measure that is closely related to the new measure that we elaborate on below.

Our survey focused on journal articles, particularly from *Journal of Phonetics*, *Journal of the International Phonetic Association*, and *JASA*, though we also included some theses that focused on linguistic phonation type. The earliest study was published in 1985. We include data from 39 languages and 76 comparisons of contrastive phonation types (e.g. breathy vs modal vowels). The languages in the survey come from several families (including Otomanguean, Mayan, Indo-European, Kx'a, Taa, Niger-Congo, Austroasiatic, Hmong-Mien, and Sino-Tibetan) spoken in various parts of the world, but especially from Mesoamerica, Southern Africa, and Southeast Asia, where phonation contrasts are more prevalent.

The spectral measures used in these comparisons include uncorrected H1–H2, H1*–H2*, and H1*. Most studies published before c. 2010 included uncorrected measures that were estimated manually, whereas those published after the advent of Praat scripts and VoiceSauce (Shue *et al.*, 2011) generally include corrected measures that were estimated automatically. Of the 76 phonation type comparisons, there are 15 creaky (i.e. more constricted) vs. modal and 35 breathy vs. modal comparisons that included a quantitative analysis of whether the differences between these phonation types were statistically significant. The languages included in the survey for each measure and for the comparisons of breathy and creaky vs. modal phonations are listed in Table I. We summarize the number of comparisons that showed significant differences, partially-significant differences, and non-significant differences for each contrast and each measure in Table II. The detailed results of the survey, including the language names and the corresponding references, are in supplementary material S1, available at `https://doi.org/10.17605/OSF.IO/QGBKA`. We define a difference as "significant" when the p value of the comparison is smaller than 0.05. We define a difference as being "partially significant" either when the p value of the comparison is between 0.05 and 0.1, or when a significant difference was found only for a subgroup of the speakers or in a subset of the stimuli. We define a difference as "non-significant" when the p value of the comparison was larger than 0.1 for all speakers and all stimuli. Generally, the results in Table II show that, for both breathy–modal and creaky–modal contrasts, the majority show significant differences in either uncorrected H1–H2 or H1*–H2*. This implies that H1–H2 is indeed a robust index of phonation differences, and supports the findings of Kreiman *et al.* (2007) that source H1–H2 is resistant to measurement artifacts. However, there exist cases in which H1–H2 does not distinguish contrastive phonation types: uncorrected H1–H2 did not distinguish breathy from modal phonation in Mon or in Tamang in Esposito (2006); uncorrected H1–H2 did not distinguish creaky from modal phonation in Mpi or in Jalapa Mazatec in Pennington (2005) (cf. Blankenship 2002); and H1*–H2* did not distinguish creaky from modal phonation in White Hmong in Esposito (2012) (cf. Garellek 2012). And while there are fewer studies that use H1–H2 for creaky vs modal comparisons, there are relatively more cases where H1–H2 does not significantly distinguish creaky vs. modal vowels than cases where the measure does not distinguish breathy vs. modal ones (3/15 vs. 2/35). This indicates that H1–H2 may be more effective at capturing breathiness than creakiness, when these phonation types are compared to modal voice. Table II also shows that H1* is rarely used as a measure to distinguish phonation types. To our knowledge, the only existing study that made a quantitative comparison of H1* between contrastive phonation types is Esposito (2012). We therefore need more data to test the effectiveness of H1* as a measure of phonatory quality.

B. Error simulations

Next we verify our hypothesis that H1–H2 is more prone to error propagation than H1, particularly when these measures are corrected as H1*–H2* and H1*. We created simulations of two circumstances: when f0 is wrongly estimated and when formants are

TABLE I. Languages in the survey of spectral differences between phonation types

Contrast	Measure	Language
Breathy vs. Modal	H1–H2	Suai; Ju|'hoansi; Jalapa Mazatec; White Hmong; Krathing Chong; Shanghainese; Ningbo Wu; Changyinsha Wu; Wenzhou Wu; Green Mong; Takhian Thong Chong; !Xóõ; Fuzhou Min; Green Mong; SADV Zapotec; SLQ Zapotec; Tlacolula Zapotec; Gujarati; Tsonga; Mon; Tamang
	H1*–H2*	Jalapa Mazatec; Gujarati; Chichimec; White Hmong; !Xóõ; Black Miao; Khmer; Shanghainese; Green Mong; Chrau
	H1*	White Hmong
Creaky vs. Modal	H1–H2	Ju|'hoansi; Coatzospan Mixtec; Jalapa Mazatec; Takhian Thong Chong; Green Mong; Mpi
	H1*–H2*	Jalapa Mazatec; White Hmong; Chichimec; !Xóõ; Black Miao; Green Mong
	H1*	White Hmong

wrongly estimated. We then determined how uncorrected H1, H1*, uncorrected H1–H2,

TABLE II. Summary of the survey results of spectral differences between phonation types; The numbers represent the number of studies in each category.

Contrast	Measure	Significant	Partially significant	Non-significant
Breathy–Modal	H1–H2	19	2	2
	H1*–H2*	8	3	0
	H1*	1	0	0
Creaky–Modal	H1–H2	4	1	2
	H1*–H2*	5	1	1
	H1*	1	0	0

and H1*–H2* are affected in both circumstances. [See also Simpson 2012 for a simulation of how nasality affects H1–H2.]

We synthesized a token of [u] as the base token using the Klatt synthesizer (Klatt and Klatt, 1990). The values of f0 and formants are shown in Table III. The segment duration, bandwidth fraction, and formant frequency interval are 0.4s, 0.05, and 1000 Hz. We manually entered the correct values of f0, F1, F2, and F3 for the base token [u] (as in Table III) in VoiceSauce (Shue *et al.*, 2011) and used those values to estimate the values of H1, H2, and H1–H2. To simulate cases where f0 is mistracked, we manually changed the f0 between 180 Hz to 300 Hz in six 20-Hz increments and then re-estimated the same spectral energy values. The results of the six f0 conditions are shown in Figure 1. We see that, when f0 is mistracked, neither H1 nor H1–H2 consistently outperforms the other in terms of being closer to the true spectral value. However, there is a tendency for H1–H2 to have a larger deviation than H1 when f0 is incorrectly estimated. Table IV lists the absolute deviation of the estimated value from the true value, for different mistracked f0 values. The mean deviation for H1–H2 is nearly twice that for H1, for both corrected and uncorrected values of these measures. Therefore, when f0 is incorrectly estimated (as is common during creaky voice), H1 appears more resistant to error than H1–H2.

TABLE III. Formant frequencies and bandwidths (in Hz) for synthesized [u].

f0	F1	B1	F2	B2	F3	B3	F4
240	453	50	944	18	2899	593	3778

Next, we stimulated two common formant tracking errors for [u] using the same stimuli and method as the f0 error demonstration. The first error type is when f0 is mistaken as F1, and F1 as F2. The second type is when F1 and F2 of [u] are mistracked as just one formant (F1), and F3 is mistaken as F2. We used the mean of F1 and F2 as "mistracked F1" for the second type of error. All other parameters were held constant when comparing those two scenarios with the correctly estimated values. The formant values used to illustrate formant tracking errors are presented in Table V. As with the f0 manipulation, we manually entered the values of F1, F2, and F3 in Table V into VoiceSauce and calculated the corrected and uncorrected H1 and H1–H2 values for the different conditions. The results are presented in Figure 2. Uncorrected values of H1, H2, and H1–H2 did not change, as expected. But for

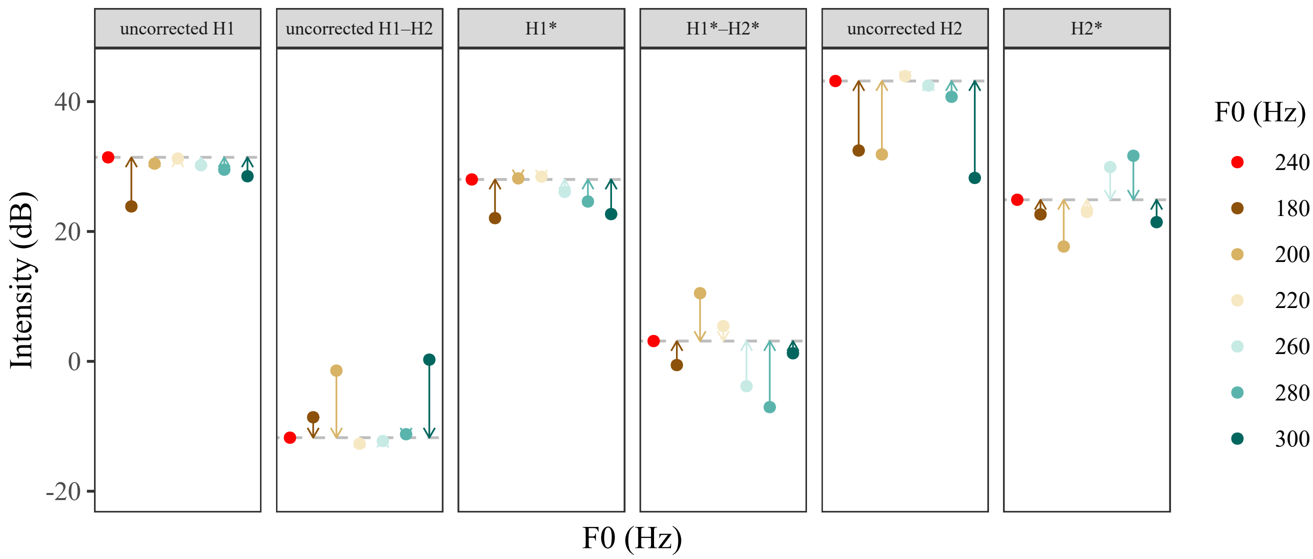

FIG. 1. (color online) Spectral energy with different f0 estimations. The leftmost red dot of each panel and the dotted line represent the true spectral (slope) value when f0 is correctly estimated at 240 Hz. The arrows represent the deviation between the true spectral value and the estimated spectral value when f0 is wrongly estimated, as is the case for all f0 values not equal to 240 Hz.

both conditions with formant tracking errors, H1* shows a smaller deviation from the true value than H1*–H2*. The mean deviation of H1*–H2* is nearly three times that of H1*. Summary statistics appear in Table VI.

C. Current proposal for estimating Residual H1

As an alternative to measuring H1–H2, we propose factoring out the effect of root-mean-squared (RMS) energy (henceforth referred as Energy) from H1, whether uncorrected H1 or H1* corrected for formant frequencies and bandwidths. We call this "residual H1." Of course, differences in recording conditions across speakers and studies will affect energy. We

TABLE IV. Deviation (Δ) of estimated uncorrected and corrected H1 and H1–H2 from their true values, for various miscalculated f0s

f0	Δuncorrected H1	Δuncorrected H1–H2	ΔH1*	ΔH1*–H2*
180	7.559	3.146	5.956	3.701
200	0.974	10.331	0.166	7.368
220	0.184	0.933	0.425	2.304
260	1.220	0.515	1.918	6.956
280	1.901	0.535	3.400	10.178
300	2.911	12.018	5.334	1.884
Mean	2.458	4.580	2.867	5.399

TABLE V. Formant values for two different types of formant tracking errors

Condition	F1	F2	F3
True	453	944	2899
f0 taken as F1	240	453	944
F1 & F2 collapsed into F1	699	2899	3778

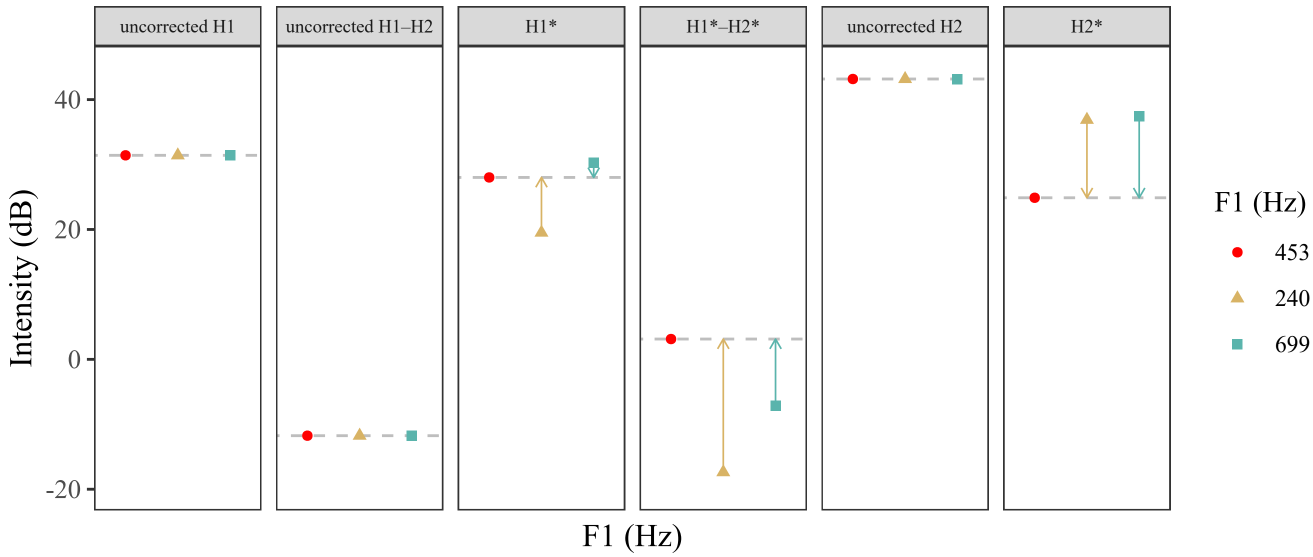

FIG. 2. (color online) Spectral energy with different formant estimations. The leftmost red dot of each panel and the dotted line represent the true spectral value when formants are correctly estimated. The arrows represent the deviation between the true spectral value and the estimated spectral value when formants are wrongly estimated. Only F1 values are listed in the legend. The corresponding F2 and F3 values can be found in Table V.

don't view this as problematic for residual H1, because such differences also affect H1: for example, a quieter signal will result in a lower RMS energy as well as a lower H1. And as we show below, residual H1 controls for the energy of an individual token on the H1 value on that same token.

Residual H1 avoids some of the issues raised in Section II B regarding the estimation of H1–H2. In practice, precisely how we control for the effect of energy on H1 varies depending on whether when H1 is a dependent variable or an independent variable. When H1 is a dependent variable, energy can be added to the model as a covariate: i.e. $H1 \sim$

TABLE VI. Deviation (Δ) of estimated H1* and H1*–H2* from their true values, for two types of mistracking

Condition	F1 (Hz)	ΔH1* (dB)	ΔH1*–H2* (dB)
f0 taken as F1	240	8.496	20.503
F1 & F2 collapsed into F1	699	2.261	10.293
Mean		5.379	15.398

$main\,factor(s) + Energy$. By adding Energy as a covariate, the effect of SPL on H1 is controlled for, and the coefficients of the main factors reflect the independent effects of those factors on H1.

When H1 is an independent variable, the effect of Energy on H1 should be calculated first, and then subtracted from H1, as shown in (1) and (2):

$$\text{Step 1: Get the coefficient of energy } (\mathrm{b}_1)\text{: } H1 \sim\; b_0 + b_1 * Energy\,(logged) \tag{1}$$

$$\text{Step 2: Calculate Residual H1: } Residual\,H1 = H1 - b_1 * Energy\,(logged) \tag{2}$$

In step 1, the coefficient of energy in Model (1) represents how strongly H1 is correlated with energy in a given token. Energy is first log-transformed because it is bounded at zero at the lower end and unbounded at the upper end. In step 2, we multiply the coefficient of energy with the actual value of energy. We then subtract the product from H1. The residual of H1 after subtraction represents the value of H1 after controlling for the SPL of the recordings.

If the data come from multiple speakers, H1 and logged energy can be transformed to z-score to reduce inter-speaker variation.

Compared with normalizing against H2, using energy has certain advantages. First, H1 requires an accurate estimation of only one spectral value, whereas H1–H2 requires two. H1 is thus less likely to be affected by error propagation than H1–H2. Second, H1 is more resistant to the influence of nasalization (Simpson, 2012). Thus, we hypothesize that H1 normalized for energy (i.e. residual H1) should reflect the degree of constriction or breathiness to the same extent as H1–H2, only with less variability. In Section III, we test the relation between H1 and OQ, to investigate whether Residual H1 has an articulatory basis of vocal fold constriction. We also use two case studies to compare residual H1 with H1–H2 in terms of their effectiveness of representing changes in phonatory quality.

III. CONTACT QUOTIENT IN RELATION TO RESIDUAL H1* VS. H1*–H2*

Previous work has shown that there is a positive, if sometimes weak and nonlinear, relationship between H1*–H2* and OQ (Kreiman *et al.*, 2012; Samlan *et al.*, 2013). In this section, we test whether residual H1* has a similar or better correlation with OQ, as indexed by electroglottographic CQ, than H1*–H2*. We used data from the "Production and Perception of Linguistic Voice Quality" project at UCLA (`http://www.phonetics.ucla.edu/voiceproject/voice.html`). The data and R code for data processing and analysis are available in supplementary material S3 at `https://doi.org/10.17605/OSF.IO/QGBKA`.

The corpus includes data from eight languages, 68 speakers, and 9,101 words in total after exclusions, (see summary in Table VII).[1] Each word was measured by nine equal time

TABLE VII. Language data from the UCLA Voice Project used to assess the relationship between CQ and Residual H1* vs. H1*–H2*.

Language	Family	Speakers	Phonation types
Bo	Sino-Tibetan	6 (3 F, 3 M)	Tense, Lax
Gujarati	Indo-European	10 (7 F, 3 M)	Breathy, Modal
Luchun Hani	Sino-Tibetan	9 (4 F, 5 M)	Tense, Lax
White Hmong	Hmong-Mien	11 (2 F, 9 M)	Breathy, Modal, Creaky tones
Mandarin	Sino-Tibetan	11 (5 F, 6 M)	Modal, Creaky tones
Black Miao	Hmong-Mien	8 (0 F, 8 M)	Breathy, Modal, Creaky tones
Southern Yi	Sino-Tibetan	7 (4 F, 3 M)	Tense, Lax
Zapotec	Otomanguean	6 (2 F, 4 M)	Breathy, Modal, Creaky

intervals, resulting in 81,909 data points in total. This data set included acoustic data of H1*–H2* and H1*, calculated using VoiceSauce (Shue *et al.*, 2011). The H1*, H1*–H2*, and f0 values were z-scored by speaker to reduce the variation between speakers. Tokens with an absolute z-score value larger than 3 were considered as outliers and were excluded from analyses. Within each vowel category, we calculated the Mahalanobis distance on the F1-F2 panel. For tokens with a Mahalanobis distance larger than 6, we regarded their formant values as outliers, similar to what has been done in our previous work (Chai and Ye, 2022; Garellek and Esposito, 2021; Seyfarth and Garellek, 2018). Time points whose f0, F1, or F2 values were outliers were also excluded from H1* and H1*–H2* analyses, because H1* and H1*–H2* are calculated based on f0, F1, and F2. For energy, we first excluded tokens with a value of zero, then log-transformed to normalize its right-skewed distribution, and then z-scored the logged energy and excluded tokens with a z-score larger than 3. The outlier detection process for the acoustic measures is the same for the following case studies in Sections IV A and IV B.

The glottal open quotient was estimated using electroglottographic (EGG) CQ calculated using the hybrid method (Howard, 1995; Orlikoff, 1991).[2] For CQ, there were 6,943 points with a value of zero; these were first excluded. The remaining CQ values were then z-scored and those with a z-score larger than 3 were considered outliers and therefore were excluded. After outlier exclusion, there were 76,196 valid data points for H1*, 76,570 for H1*–H2*, 81,222 for f0, 73,363 for CQ, and 81,473 for energy.

To assess the relationship between CQ and H1*–H2* and residual H1*, we regressed CQ on both H1*–H2* and residual H1*, as in Models (3) and (4). Since H1* was an independent variable in the model, we factored out the effect of energy from H1* and calculated residual H1* using Equations (1) and (2). The coefficient of energy on H1* was 0.682. The statistics are shown in Table VIII. We use R^2 to represent the effect size of the models. The R^2 value is defined as the percentage of variance of the dependent variable that is explained by the independent variables in the model. For linear mixed-effect models, we calculate the **marginal** R^2 of the model, which is defined as the percentage of variance of the dependent variable that is explained by the **fixed** variables in the model (Johnson, 2014). The R^2 (for linear models) and marginal R^2 (for linear mixed-effect models) are calculated using the *multilevelTools* package (Wiley, 2020) in *R*. A scatter plot of all data points and the correlation line between CQ and the two spectral tilt measures are shown in Figure 3.

As the results show, residual H1* had a larger absolute coefficient, slightly higher standard error (0.005 vs. 0.003), and higher t-value than H1*–H2*. Model 4 with H1* as the predictor had a higher marginal R^2 value than Model 3 with H1*–H2* as the predictor (0.102 vs. 0.060), suggesting that H1* can explain more variance of CQ than H1*–H2*. Figure 3 illustrates that the regression line for residual H1* is steeper than that for H1*–H2*. This indicates that H1*, after controlling for energy, has a stronger correlation with CQ than H1*–H2*. By extension, this also confirms the articulatory basis of H1* as an acoustic correlate of vocal fold approximation.

$$CQ \sim \; H1^* - H2^* + (1|Speaker) \tag{3}$$

$$CQ \sim \; Residual\, H1^* + (1|Speaker) \tag{4}$$

TABLE VIII. Correlation between CQ and H1*–H2* and H1*

Model	β	Std. Error	t value	p	R^2
CQ $\sim$ H1*–H2*	-0.230	0.003	-66.090	$< .001$	0.060
CQ $\sim$ Residual H1*	-0.449	0.005	-88.700	$< .001$	0.102

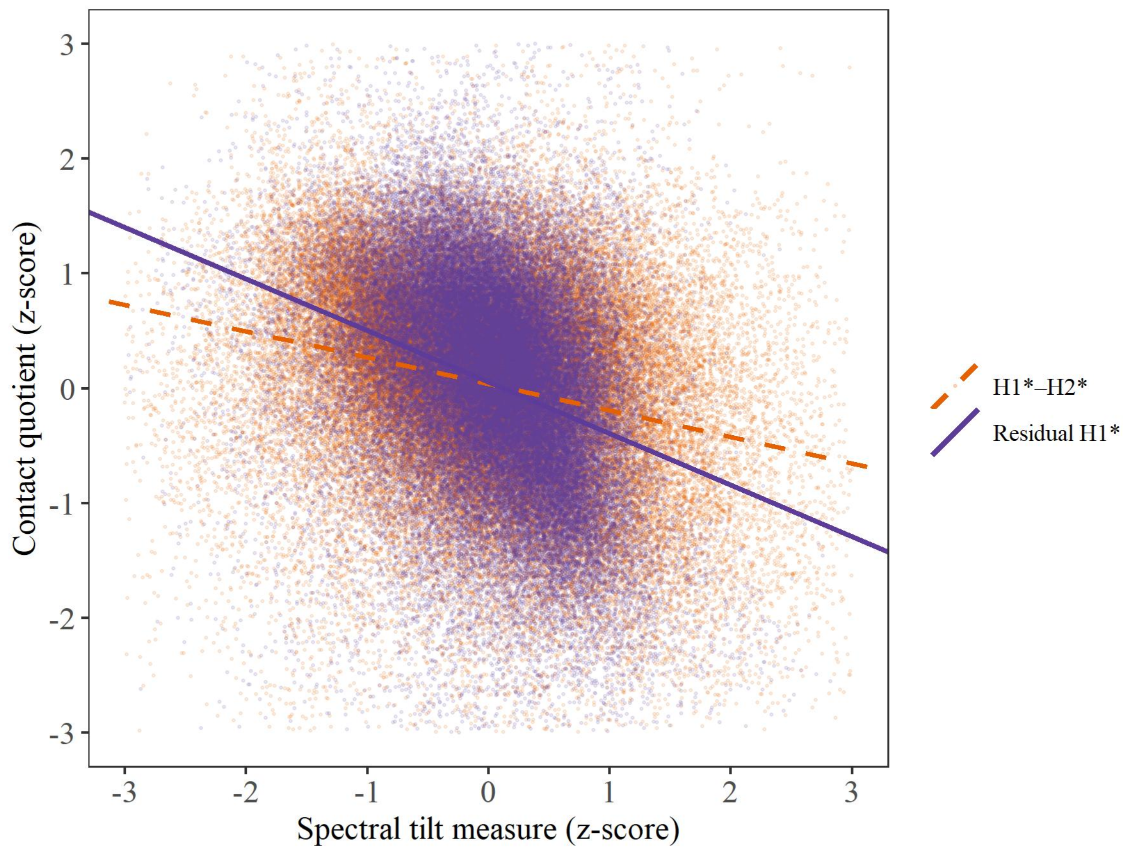

FIG. 3. (color online) Relationship between CQ and H1*–H2* and residual H1*. The CQ values have had the random intercept of subjects subtracted. Regression lines were based on results from Model (3) and (4).

IV. CASE STUDIES

In this section, we provide two case studies where we compare residual H1* to H1*–H2* and their ability to track changes in phonation in two languages, !Xóõ and Mandarin. We use datasets that have previously been analyzed for phonation: Garellek (2020) on !Xóõ phonation types and Chai (2019, 2021) on Mandarin utterance-level changes in voice quality. In neither paper did we look specifically at H1, so for these case studies we were particularly interested in seeing if the phonation differences are better differentiated acoustically using residual H1* instead of H1*–H2*.

A. Phonation types in !Xóõ (Taa)

1. Corpus

!Xóõ (also known as Taa) is a Tuu language spoken in Botswana, whose phonation types have recently been analyzed acoustically by Garellek (2020). That study only measured H1*–H2*; here, we compare H1* (with energy as a covariate) to H1*–H2* for three of the phonation types: breathy, modal, and creaky.

The recordings are of the East !Xóõ dialect, and were made in the late 1970s by Peter Ladefoged and Tony Traill. They are available for download from the UCLA Phonetics Lab Archive at `http://archive.phonetics.ucla.edu/Language/NMN/nmn.html`. We used the same data as (Garellek, 2020), and thus we refer the reader to that source for details on the segmentation criteria and data segmentation procedures. All the words had /a/ vowels (which varied considerably in phonetic quality due to coarticulation), and were produced by ten speakers of !Xóõ. The corpus had 369 words, containing six phonation types. We only compared the spectral values of breathy, modal, and creaky phonations, resulting in 175 tokens for analysis (breathy: 83; modal: 54; creaky: 38). The word list of the stimuli is in supplementary material S2 at `https://doi.org/10.17605/OSF.IO/QGBKA`. The data and R code for data processing and analysis are available in supplementary material S3 at the same URL as S2.

The acoustic measures of the recordings were calculated using VoiceSauce every millisecond. Each token was divided into nine equal intervals. The mean of each interval was calculated. We used nine points to represent each token such that the duration of the tokens was normalized. In total, 1,575 data points were measured (175 tokens * 9 time points). The outlier detection method is the same as described in III. After the outlier exclusion, there were 1,351 valid data points for H1*, 1,363 for H1*–H2*, 1,507 for f0, 1,380 for formants, and 1,492 for energy. We calculated the mean acoustic values for each individual words for the statistical analysis in Section IV A 2.

2. *Results*

To compare how effectively H1* and H1*–H2* differentiate the three phonation types, we regressed both H1* and H1*–H2* on phonation type. The model in which H1* was the dependent variable also had energy as an independent variable to control for the SPL. The models for H1*–H2* and H1* are in (5) and (6):

$$H1^* - H2^* \sim \ Phonation \tag{5}$$

$$H1^* \sim \ Phonation + Energy\,(logged) + (1|Speaker) \tag{6}$$

For each model, the modal phonation was set as the baseline for comparison. Random intercepts or slopes by speaker were not included for Model (5), because they resulted in singular fits and did not improve the model. We compared the effectiveness of H1* and H1*–H2* in differentiating creaky and breathy phonation types from modal phonation by looking at the estimate coefficient, standard error, and t-value (estimate/standard error) of the phonation variable. A higher coefficient means that two phonation types have a larger difference in the acoustic measure. A lower standard error indicates that the values of the acoustic measure of each phonation group are less variable. A higher t-value represents a relatively high coefficient and a relatively low standard error, indicating a better separation between two phonation groups.

The statistics of Models (5) and (6) are shown in Table IX. For the differentiation between creaky and modal phonation, the model with H1* as the dependent variable and energy as the covariate had a higher estimate of coefficient, lower standard error, and higher t-value for creaky vs. modal comparison than the model with H1*–H2* as the dependent variable. This indicates that H1* (after controlling for energy) is better at distinguishing creaky from modal phonation than H1*–H2*.

When comparing breathy and modal phonation, H1*–H2* behaved similarly to H1*. H1*–H2* had a higher coefficient estimate and standard error than did H1*. The t-value of H1*–H2* was similar to that of H1* (10.974 vs. 11.085), whereas both models had p-values smaller than 0.001. Thus, in terms of distinguishing breathy from modal phonation, we consider H1*–H2* and H1* to be equally effective.

We also calculated the effect sizes of Models (5) and (6) using R^2, as shown in Table IX[3]. The marginal R^2 of Model (6) is higher than the R^2 of Model (5) (0.838 vs. 0.592), indicating that the model with H1* as the dependent variable has a larger effect size (and thus more variance is explained) than the model with H1*–H2* as the dependent variable.

The distributions of H1*–H2* and H1* for different phonation types in !Xóõ are shown in Figure 4. For the H1* data in Figure 4 to show how H1* distinguished the three phonation types after controlling for energy, the residual H1* was calculated by subtracting the product of the coefficient of energy in Model (6) (b = 0.606) and the z-scored energy from the z-scored H1* value. Comparing H1*–H2* with H1* in Figure 4, we see that for all the three phonation types, the H1* values are less variable within group, and there is less overlap between modal and creaky phonation types in H1* than H1*–H2*. In sum, after controlling for energy, H1* distinguished creaky phonation from modal phonation in !Xóõ better than H1*–H2*, in terms of having a larger effect size and smaller standard errors. However, H1* and H1*–H2* do not differ in the effectiveness of distinguishing breathy phonation from modal phonation in !Xóõ.

TABLE IX. Model comparison between H1*–H2* and H1* in distinguishing !Xóõ phonation types

Phonation contrast	Model	β	Std. Error	t value	p
Creaky – Modal	$H1^* - H2^* \sim \boldsymbol{Phonation}$	-0.462	0.141	-3.278	0.0013
	$H1^* \sim \boldsymbol{Phonation} + Energy$	-0.554	0.091	-6.069	$< .001$
Breathy – Modal	$H1^* - H2^* \sim \boldsymbol{Phonation}$	1.221	0.111	10.974	$< .001$
	$H1^* \sim \boldsymbol{Phonation} + Energy$	0.671	0.061	11.085	$< .001$

TABLE X. R^2 of Model (5) and marginal R^2 of Model (6)

Model		(Marginal) R^2
(5)	$H1^* - H2^* \sim Phonation$	0.592
(6)	$H1^* \sim Phonation + Energy + (1\|Speaker)$	0.838

B. Phrasing in Mandarin

1. *Corpus*

Chai (2019) found that the final position of declarative sentences in Mandarin had more creak than non-final positions, after controlling for f0. They assumed then that vowels in utterance-final position should be more constricted acoustically than non-final positions, but did not find differences in H1*–H2* according to position. In a follow-up study, Chai (2021) increased the sample size and found a correlation between low H1*–H2* and utterance-final position in declaratives. This suggests that the discrepancy in findings between Chai (2019) and Chai (2021) was due to noisiness in H1*–H2*, requiring a larger data set for effects to emerge. In the present study, we aim to determine whether utterance-final creak in Mandarin is indeed associated with vocal fold constriction, as measured by residual H1* instead of H1*–H2*.

We combined the data sets from both Chai (2019) and Chai (2021). There were 823 target declarative sentences produced by 64 Mandarin speakers. Phonologically identical words were placed in the initial, medial, and final position of each sentence. The stimuli include

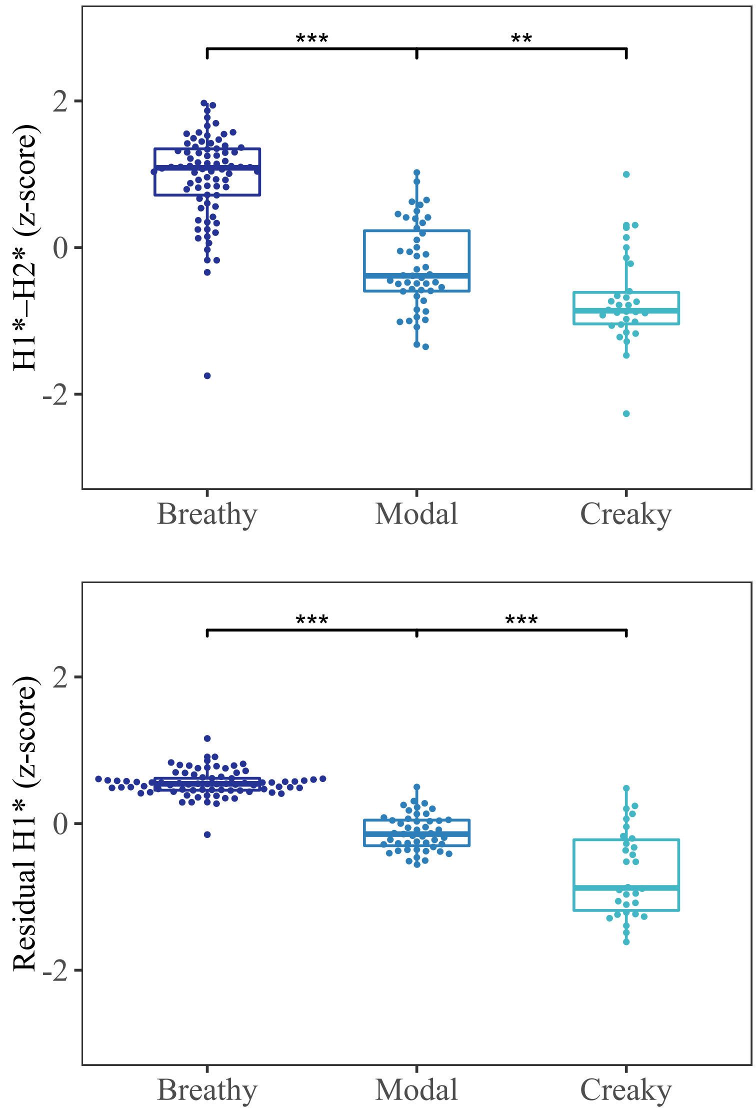

FIG. 4. (color online) H1*–H2* (top) and Residual H1* (bottom) in different phonations in !Xóõ

$* * *\ p < .001;\ * *\ 0.001 < p < 0.01;\ *\ 0.01 < p < 0.05$

Residual H1* = H1* – Energy * Energy coefficient (0.606) – random intercept in Model (6)

1,889 target words in total (initial: 631; medial: 628; final: 630). The stimuli sentence list is in supplementary material S2 at https://doi.org/10.17605/OSF.IO/QGBKA. The data and R code for data processing and analysis are available in supplementary material S3 at the same URL as S2.

The recordings were processed using VoiceSauce, which output a value for H1*, H1*–H2*, energy, and f0 every millisecond. Energy values were first log-transformed. All the acoustic measurements were z-scored by speaker and word. 217,378 data points were generated in total. The outlier detection procedure is the same as described in III. After outlier exclusions, there were 200,505 valid data points for H1*, 204,497 for H1*–H2*, 210,994 for formants, 213,601 for f0, and 214,592 for energy. We calculated the mean value for each individual word for the statistical analysis in Section IV B 2.

2. Results

As with the !Xóõ case study, here two models were fit to test whether H1*–H2* or H1* best distinguishes voice qualities associated with different utterance positions. The models for H1*–H2* and H1* were (7) and (8). Since Chai (2019) suggested that utterance-final position was creakier than non-final positions after controlling for f0, f0 was added to Model (7) and (8). The criteria of a better model were the same as the !Xóõ case study: larger coefficient estimate, smaller standard error, and larger t-value.

$$H1^* - H2^* \sim \ Position + f0 + (1|Speaker) + (f0 + Position|Speaker) \quad (7)$$

$$H1^* \sim \ Position + Energy\,(logged) + f0 + (1|Speaker) + (f0 + Position|Speaker) \quad (8)$$

The statistics of Models (7) and (8) are shown in Table XI. In terms of distinguishing initial position from final position, the coefficient of the position variable in the H1* model was four times larger than in the H1*–H2* model. The standard errors of the two models were similar. The t-value was higher in the H1* model than the H1*–H2* model. Similarly, when distinguishing medial position from final position, the position variable in the H1* model had larger coefficient, similar standard error, and higher t-value than in the H1*–H2* model.

We also calculated the effect sizes of Models (7) and (8) using marginal R^2 of the fixed variables, as shown in Table XII. The marginal R^2 of Model (8) is larger than that of Model (7) (0.805 vs 0.254), indicating that the model with H1* as the dependent variable has a larger effect size than the model with H1*–H2* as the dependent variable; that is, more variance in the dependent variable is explained.

Figure 5 shows residual H1*–H2* and residual H1* in utterance-initial, medial, and final position. The H1*–H2* distributions of the three positions are very similar, whereas in H1* we find that final position has overall lower values than the non-final positions.

In sum, while H1*–H2* did not distinguish the three utterance positions in Chai 2019, an effect of utterance position on H1*–H2* emerged after we increased the number of data points and subjects by adding on data from a subsequent study, Chai 2021. This suggests that the creakier voice quality of utterance-final position in Chai 2019 was indeed produced with more constriction. The effect likely did not emerge in that original study due to a lack of statistical power. In addition, the comparison between the H1* and H1*–H2* models reflects the fact that H1* captures the difference in vocal fold constriction better than H1*–H2* and requires less statistical power.

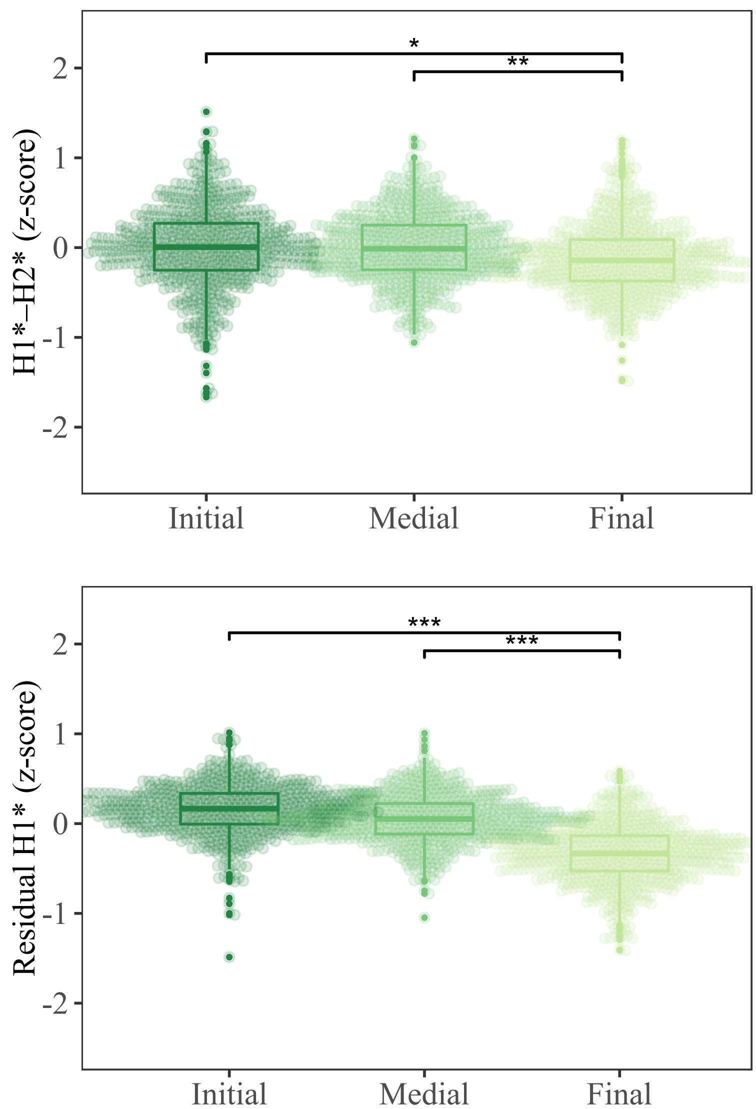

FIG. 5. (color online) Residual H1*–H2* (top) and Residual H1* (bottom) in different positions in Mandarin. $* * *\ p < .001$; $* *\ 0.001 < p < 0.01$; $*\ 0.01 < p < 0.05$

Residual H1*–H2* = H1*–H2* – f0 * f0 coefficient (0.279) – random intercept and slopes in Model (7)

Residual H1* = H1* – f0 * f0 coefficient (0.025) – Energy * Energy coefficient (0.541) – random intercept and slopes in Model (8)

TABLE XI. Model comparison between H1*–H2* and H1* in distinguishing utterance positions in Mandarin

Position comparison	Model	β	Std. Error	t value	p
Initial – Final	$H1^* - H2^* \sim \boldsymbol{Position} + f0$	0.124	0.048	2.565	.013
	$H1^* \sim \boldsymbol{Position} + Energy + f0$	0.494	0.047	10.464	$< .001$
Medial – Final	$H1^* - H2^* \sim \boldsymbol{Position} + f0$	0.129	0.044	2.914	.005
	$H1^* \sim \boldsymbol{Position} + Energy + f0$	0.395	0.037	10.724	$< .001$

TABLE XII. Marginal R^2 of Model (7) and (8)

Model		Marginal R^2
(7)	$H1^* - H2^* \sim Position + f0 + (1\|Speaker) + (f0 + Position\|Speaker)$	0.254
(8)	$H1^* \sim Position + Energy + f0 + (1\|Speaker) + (f0 + Position\|Speaker)$	0.805

V. DISCUSSION AND CONCLUSION

The goals of this paper were to review the history of H1(–H2) as acoustic measures of phonation type. We trace their origin back to the pioneering work by Fischer-Jørgensen (1967) on breathy vs. modal vowels in Gujarati, and highlight later studies that advanced our understanding of these measures in terms of their aerodynamic and articulatory correlates, their perceptibility by listeners, and their use in indexing various phonological contrasts and social factors.

We then highlighted several issues for using H1–H2 as the indicator for vocal fold constriction. First, we reviewed the literature that has made use of H1–H2 in studies of linguistic phonation type. We found that the measure frequently succeeds at distinguishing both breathy and creaky phonation types from modal voice. However, there is a tendency for the measure to be less effective at creaky vs. modal contrasts than breathy vs. modal ones. We attribute this to errors in f0 estimation during irregular voicing associated with creaky voice. We further argue that H1–H2 (and particularly H1*–H2*) is prone to error propagation, in that it measures two spectral values, both of which are affected by f0 and vowel formants. H1–H2 is also affected by other glottal features besides OQ, in part because H2 is affected by other factors like glottal skew. Finally, H1–H2 is affected by nasal poles and zeroes. In the current study, we show that using "residual H1*," for which Energy is used to normalize H1* (instead of H2*), can to some extent mitigate these issues.

Limited previous work has already made use of H1 instead of or in addition to H1–H2. For instance, Esposito (2012) found that H1* differentiated the phonation types in White Hmong better than H1*–H2* (in the sense that significant differences between phonation types were found more often), and H1* had a stronger correlation with CQ than H1*–H2*. However, she did not normalize the amplitude of H1*, meaning that there is a potential for a confound between phonation and SPL differences. The current study used energy to normalize H1* either by adding energy as a covariate of the H1* model or by subtracting the effect of energy from H1*, resulting in a new measurement of residual H1*.

Using three data sets, we also show how residual H1* can be used in practice. A corpus analysis of natural speech data (taken from the "Production and Perception of Linguistic Voice Quality" project at UCLA, which includes EGG and audio recordings from eight languages and dozens of speakers), revealed that residual H1* has a stronger relationship to glottal OQ than H1*–H2*. Second, we showed that residual H1* better differentiated the phonation types in !Xóõ than H1*–H2*, particularly for modal vs. creaky vowels, as expected from our error simulations. Finally, we found that residual H1* better differentiated the changes in phonatory quality by utterance position in Mandarin than H1*–H2*. We therefore suggest that researchers consider using RMS energy to normalize for the amplitude of H1, treating residual H1 as an acoustic correlate of vocal fold constriction– instead of, or in addition to, H1–H2. It is worth noting the one context in which H1–H2 might be preferred: when directly comparing just two tokens. In such a case, the overall effect of energy on H1 cannot be estimated. But given the move towards larger data sets in the phonetic sciences, it is exceedingly rare for researchers to describe a contrast using only two tokens, except for the purposes of general illustration. Certainly, in a phonetic analysis of phonation type that makes use of multiple tokens from several speakers, an estimate of the effect of energy on H1 can be made, and we argue here that it is desirable for researchers to calculate residual H1.

A claim can also be made that residual H1 is better motivated theoretically than H1–H2. As we discussed earlier, early uses of H1–H2 were motivated by observable differences in H1, rather than by any theoretical import assigned to the slope of H1–H2 or to H2 in particular. After all, H1 is the amplitude of the fundamental, which as the primary correlate of vocal pitch clearly matters for overall voice quality perception. Thus, residual H1 is correlated with how loud the fundamental – and thus pitch – is perceived to be relative to the overall loudness of the signal. We also know that SPL is an important component to voice and signal perception, and it is included in psychoacoustic models of the voice (Kreiman *et al.*, 2014). Residual H1 therefore captures information about two important cues: f0 as a cue to pitch, and SPL as a cue to loudness. In contrast, a measure like H1–H2 includes a component of the source spectrum – H2 – that is not known to matter intrinsically for voice quality perception. Clearly, future work is needed to examine how listeners assess H1 as a cue relative to other spectral landmarks and the signal more broadly. This should also include comparisons between H1 and spectral tilt as measured with reference to vowel formants; namely, H1–A1, H1–A2, or other formant-based measures like A1–A2 that make no reference to the fundamental. As Garellek (2019, p. 88) mentioned, the use of formant-based measures carries the assumption that voice quality depends on vowel quality. That assumption may ultimately prove correct, but so far it has gone untested.

The conceptualization of H1 relative to overall energy also has implications for how we model the voice source spectrum. Earlier work in this regard (Garellek *et al.*, 2013, 2016a; Kreiman *et al.*, 2012) explicitly models source H1–H2 as a harmonic slope, in addition to H2–H4, H4–H2kHz (the spectral slope from H4 to the harmonic closest to 2000 Hz) and H2kHz–H5kHz (the spectral slope between the harmonics closest to 2000 and 5000 Hz). But if what matters is H1 and not H1–H2, then perhaps a more suitable model of the harmonic source spectrum could include only H1 instead of H1–H2. Practically this would involve only a minor change to the model: instead of H1–H2 and three additional spectral slopes, the updated harmonic source spectrum would include H1 and those same additional slopes. But there are important theoretical implications to this change, because H1 would not be compared directly to another harmonic or to any other segment of the harmonic source spectrum; its raw amplitude is what would matter, just like its raw frequency (that is, the f0) matters. Of course, to control for overall SPL, H1 (and the other spectral slopes) should be modeled as a sub-component of a larger psychoacoustic model of the voice that includes overall energy, as done already in the psychoacoustic model of Kreiman *et al.* (2014). Ultimately, the choice of whether to include H1 or H1–H2 in a psychoacoustic model of the voice should depend on which of the two measures provides a better link between voice production and voice quality perception. Much more work is therefore needed to determine whether H1 provides a closer link between voice production and perception than H1–H2.

To conclude, we have shown that residual H1 has fewer error propagation issues than H1–H2; using residual H1 can therefore lead to more accurate measurements, and thus better description, of the acoustic correlates of vocal fold constriction. Future studies should investigate what the specific articulatory and aerodynamic correlates of H1 are: does the measure more closely reflect changes in vocal fold constriction, medial fold thickness, or glottal skew? Additionally, future work could investigate the extent to which H1 outperforms H1–H2 when comparing the voice quality across nasal vowels and whether listener judgments of voice quality are better predicted by H1 than by H1–H2.

[1] Words that have nasal vowels; are marked as "do not use"; or have unmatched annotations between the acoustic and EGG results files were excluded.

[2]CQ = 1–OQ.

[3]The effect size of Model (6) is represented by the marginal R^2 of the fixed variables.

Arras, K. O. (**1998**). "An Introduction To Error Propagation: Derivation, Meaning and Examples of Equation Cy= Fx Cx FxT," Technical Report, http://hdl.handle.net/20.500.11850/82620, doi: 10.3929/ETHZ-A-010113668, artwork Size: 22 p. Medium: application/pdf.

Bickley, C. (**1982**). "Acoustic analysis and perception of breathy vowels," MIT Speech Communication Working Papers **1**, 71–81.

Blankenship, B. (**2002**). "The timing of nonmodal phonation in vowels," Journal of Phonetics **30**(2), 163–191, https://linkinghub.elsevier.com/retrieve/pii/S009544700190155X, doi: 10.1006/jpho.2001.0155.

Caballero, G., and Carroll, L. (**2015**). "Tone and stress in Choguita Rarámuri (Tarahumara) word prosody," International Journal of American Linguistics **81**, 457–493.

Campbell, N., and Beckman, M. (**1997**). "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models, and Applications*, International Speech Communication Association, Athens, Greece, pp. 67–70.

Chai, Y. (**2019**). "The source of creak in Mandarin utterances," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Australasian Speech Science and Technology Association Inc, Canberra, Australia, pp. 1858–1862.

Chai, Y. (**2021**). "The source of creak in Mandarin utterances," UC San Diego: San Diego Linguistic Papers **8**, 1–32, https://escholarship.org/uc/item/8mg0x5pb.

Chai, Y., and Ye, S. (**2022**). "Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min," Languages **7**(1), 47, https://www.mdpi.com/2226-471X/7/1/47, doi: 10.3390/languages7010047.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," Journal of Phonetics **27**(207-229).

Chodroff, E., Golden, A., and Wilson, C. (**2019**). "Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization," The Journal of the Acoustical Society of America **145**(1), EL109–EL115, http://asa.scitation.org/doi/10.1121/1.5088035, doi: 10.1121/1.5088035.

Dang, J., and Honda, K. (**1996**). "Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation," The Journal of the Acoustical Society of America **100**(5), 3374–3383, http://asa.scitation.org/doi/10.1121/1.416978, doi: 10.1121/1.416978.

DiCanio, C. T. (**2009**). "The phonetics of register in Takhian Thong Chong," Journal of the International Phonetic Association **39**(2), 162–188, https://www.cambridge.org/core/product/identifier/S0025100309003879/type/journal_article, doi: 10.1017/S0025100309003879.

Doval, B., and d'Alessandro, C. (**1997**). "Spectral correlates of glottal waveform models: an analytic study," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE Comput. Soc. Press, Munich, Germany, Vol. 2, pp. 1295–1298, http://ieeexplore.ieee.org/document/596183/, doi: 10.1109/ICASSP.1997.596183.

Doval, B., d'Alessandro, C., and Henrich, N. (**2006**). "The spectrum of glottal flow models," Acta Acustica united with Acustica **92**(6), 1026–1046, https://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000006/art00021.

Esposito, C. M. (**2006**). "The Effects of Linguistic Experience on the Perception of Phonation," Ph.D. Dissertation, University of California, Los Angeles, Los Angeles, CA, USA, http://phonetics.linguistics.ucla.edu/research/Esposito_diss.pdf.

Esposito, C. M. (**2010**). "The effects of linguistic experience on the perception of phonation," Journal of Phonetics **38**(2), 306–316, https://linkinghub.elsevier.com/retrieve/pii/S0095447010000203, doi: 10.1016/j.wocn.2010.02.002.

Esposito, C. M. (**2012**). "An acoustic and electroglottographic study of White Hmong tone and phonation," Journal of Phonetics **40**(3), 466–476, https://linkinghub.elsevier.com/retrieve/pii/S0095447012000174, doi: 10.1016/j.wocn.2012.02.007.

Esposito, C. M., and Khan, S. u. D. (**2020**). "The cross-linguistic patterns of phonation types," Language and Linguistics Compass **14**(12), 1–25, https://onlinelibrary.wiley.com/doi/10.1111/lnc3.12392, doi: 10.1111/lnc3.12392.

Fant, G., Liljencrants, J., and Lin, Q.-g. (**1985**). "A four-parameter model of glottal flow," STL-QPSR **26**, 1–13.

Fant, G., and Lin, Q.-g. (**1988**). "Frequency domain interpretation and derivation of glottal flow parameters," STL-QPSR **4**, 1–21.

Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Linguistics **28**, 71–139.

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://linkinghub.elsevier.com/retrieve/pii/S0095447011000969, doi: 10.1016/j.wocn.2011.10.003.

Garellek, M. (**2019**). “The phonetics of voice,” in *Routledge Handbook of Phonetics*, edited by W. Katz and P. Assmann (Routledge, Oxford), pp. 75–106.

Garellek, M. (**2020**). “Acoustic Discriminability of the Complex Phonation System in !Xóõ,” Phonetica **77**(2), 131–160, https://www.degruyter.com/document/doi/10.1159/000494301/html, doi: 10.1159/000494301.

Garellek, M. (**2022**). “Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality,” Journal of Phonetics **94**, 101155, https://linkinghub.elsevier.com/retrieve/pii/S0095447022000304, doi: 10.1016/j.wocn.2022.101155.

Garellek, M., and Esposito, C. M. (**2021**). “Phonetics of White Hmong vowel and tonal contrasts,” Journal of the International Phonetic Association 1–20, https://www.cambridge.org/core/product/identifier/S0025100321000104/type/journal_article, doi: 10.1017/S0025100321000104.

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

Garellek, M., Ritchart, A., and Kuang, J. (**2016**a). “Breathy voice during nasality: a cross-linguistic study,” Journal of Phonetics **59**, 110–121.

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

Garellek, M., and White, J. (**2015**). "Phonetics of Tongan stress," Journal of the International Phonetic Association **45**, 13–34.

Gobl, C., Murphy, A., Yanushevskaya, I., and Chasaide, A. N. (**2018**). "On the relationship between glottal pulse shape and its spectrum: correlations of open quotient, pulse skew and peak flow with source harmonic amplitudes," in *Proceedings of Interspeech 2018*, Hyderabad, India, pp. 222–226.

Gobl, C., and Ní Chasaide, A. (**2019**). "Time to Frequency Domain Mapping of the Voice Source: The Influence of Open Quotient and Glottal Skew on the Low End of the Source Spectrum," in *Interspeech 2019*, ISCA, pp. 1961–1965, https://www.isca-speech.org/archive/interspeech_2019/gobl19_interspeech.html, doi: 10.21437/Interspeech.2019-2888.

Gordon, M., and Ladefoged, P. (**2001**). "Phonation types: a cross-linguistic overview," Journal of Phonetics **29**, 383–406.

Guion, S. G., Amith, J. D., Doty, C. S., and Shport, I. A. (**2010**). "Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language," Journal of Phonetics **38**, 137–166.

Guion, S. G., Post, M. W., and Payne, D. L. (**2004**). "Phonetic correlates of tongue root vowel contrasts in Maa," Journal of Phonetics **32**, 517–542.

Hanson, H. M. (**1995**). "Glottal characteristics of female speakers," PhD Thesis, Harvard University.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," Journal of the Acoustical Society of America **101**, 466–481.

Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," The Journal of the Acoustical Society of America **106**(2), 1064–1077, http://asa.scitation.org/doi/10.1121/1.427116, doi: 10.1121/1.427116.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., and Slifka, J. (**2001**). "Towards models of phonation," Journal of Phonetics **29**, 451–480.

Henrich, N., d'Alessandro, C., and Doval, B. (**2001**). "Spectral Correlates of Voice Open Quotient and Glottal Flow Asymmetry : Theory, Limits and Experimental Data," in *Proceedings of EUROSPEECH-2001*, Aalborg, Denmark, pp. 47–50, https://www.isca-speech.org/archive_v0/eurospeech_2001/e01_0047.html.

Henton, C. G., and Bladon, R. A. W. (**1985**). "Breathiness in normal female speech: Inefficiency versus desirability," Language and Communication **5**, 221–227.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," Journal of Speech and Hearing Research **38**, 1212–1223.

Howard, D. M. (**1995**). "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," Journal of Voice **9**, 163–172.

Huffman, M. K. (**1987**). “Measures of phonation type in Hmong,” The Journal of the Acoustical Society of America **81**(2), 495–504, http://asa.scitation.org/doi/10.1121/1.394915, doi: 10.1121/1.394915.

Iseli, M., Shue, Y.-L., and Alwan, A. (**2007**). “Age, sex, and vowel dependencies of acoustic measures related to the voice source,” Journal of the Acoustical Society of America **121**, 2283–2295.

Johnson, P. C. (**2014**). “Extension of Nakagawa & Schielzeth’s R^2_{glmm} to random slopes models,” Methods in Ecology and Evolution **5**(9), 944–946, https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12225, doi: 10.1111/2041-210X.12225.

Keating, P., Garellek, M., and Kreiman, J. (**2015**). “Acoustic properties of different kinds of creaky voice,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, https://idiom.ucsd.edu/~mgarellek/files/Keating_etal_2015_ICPhS.pdf.

Kirk, P., Ladefoged, P., and Ladefoged, J. (**1984**). “Using a spectrograph for measures of phonation types in natural language,” UCLA Working Papers in Phonetics **59**, 102–113.

Klatt, D. H., and Klatt, L. C. (**1990**). “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” The Journal of the Acoustical Society of America **87**(2), 820–857, http://asa.scitation.org/doi/10.1121/1.398894, doi: 10.1121/1.398894.

Kreiman, J., Gerratt, B., and Antoñanzas-Barroso, N. (**2007**). “Measures of the glottal source spectrum,” Journal of Speech, Language, and Hearing Research **50**, 595–610.

Kreiman, J., and Gerratt, B. R. (**2010**). “Perceptual Assessment of Voice Quality: Past, Present, and Future,” Perspectives on Voice and Voice Disorders **20**, 62–67.

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

Kreiman, J., Gerratt, B. R., and Khan, S. u. D. (**2010**). "Effects of native language on perception of voice quality," Journal of Phonetics **38**(4), 588–593, https://linkinghub.elsevier.com/retrieve/pii/S0095447010000641, doi: 10.1016/j.wocn.2010.08.004.

Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., and Alwan, A. (**2012**). "Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation," Journal of the Acoustical Society of America **132**, 2625–2632.

Kuang, J. (**2011**). "Production and Perception of the Phonation Contrast in Yi," Master's thesis, University of California, Los Angeles, Los Angeles, CA, USA.

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, http://asa.scitation.org/doi/10.1121/1.5003649, doi: 10.1121/1.5003649.

Ladefoged, P. (**1981**). "The relative nature of voice quality," The Journal of the Acoustical Society of America **69**(S1), S67–S67, http://asa.scitation.org/doi/10.1121/1.386168, doi: 10.1121/1.386168.

Ladefoged, P. (**1983**). "Cross-linguistic studies of speech production," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 177–188.

Li, S., Gu, W., Liu, L., and Tang, P. (**2020**). "The Role of Voice Quality in Mandarin Sarcastic Speech: An Acoustic and Electroglottographic Study," Journal of Speech, Language, and Hearing Research **63**(8), 2578–2588, http://pubs.asha.org/doi/10.1044/

2020_JSLHR-19-00166, doi: 10.1044/2020_JSLHR-19-00166.

Maddieson, I., and Ladefoged, P. (**1985**). "“Tense” and “lax” in four minority languages of China," Journal of Phonetics **13**(4), 433–454, https://linkinghub.elsevier.com/retrieve/pii/S0095447019307880, doi: 10.1016/S0095-4470(19)30788-0.

Ní Chasaide, A., and Gobl, C. (**1993**). "Contextual Variation of the Vowel Voice Source as a Function of Adjacent Consonants," Language and Speech **36**(2-3), 303–330, http://journals.sagepub.com/doi/10.1177/002383099303600310, doi: 10.1177/002383099303600310.

Orlikoff, R. F. (**1991**). "Assessment of the dynamics of vocal fold contact from the electroglottogram," Journal of Speech and Hearing Research **34**, 1066–1072.

Pennington, M. (**2005**). "The phonetics and phonology of glottal manner features," Dissertation, Indiana University, Bloomington, https://www.proquest.com/docview/304986742?pq-origsite=gscholar&fromopenview=true.

Samlan, R. A., and Story, B. H. (**2011**). "Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computational modeling," Journal of Speech, Language, and Hearing Research **54**, 1267–1283.

Samlan, R. A., Story, B. H., and Bunton, K. (**2013**). "Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling," Journal of Speech, Language, and Hearing Research **56**, 1209–1223.

Seyfarth, S., and Garellek, M. (**2018**). "Plosive voicing acoustics and voice quality in Yerevan Armenian," Journal of Phonetics **71**, 425–450, https://linkinghub.elsevier.com/retrieve/pii/S0095447017302711, doi: 10.1016/j.wocn.2018.09.001.

Shue, Y.-L., Chen, G., and Alwan, A. (**2010**). “On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures,” in *Proceedings of Interspeech 2010*, pp. 34–37.

Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K. M. (**2011**). “VoiceSauce: A program for voice analysis,” in *Proceedings of the 17th International Congress of Phonetic Science*, Hong Kong, pp. 1846–1849.

Simpson, A. (**2012**). “The first and second harmonics should not be used to measure breathiness in male and female voices,” Journal of Phonetics **40**, 477–490.

Sluijter, A., and van Heuven, V. (**1996**). “Acoustic correlates of linguistic stress and accent in Dutch and American English,” in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP ’96*, IEEE, Philadelphia, PA, USA, Vol. 2, pp. 630–633, http://ieeexplore.ieee.org/document/607440/, doi: 10.1109/ICSLP.1996.607440.

Stevens, K. N. (**1977**). “Physics of laryngeal behavior and larynx modes,” Phonetica **34**, 264–279.

Stevens, K. N., and Hanson, H. M. (**1995**). “Classification of glottal vibration from acoustic measurements,” in *Vocal fold physiology: Voice quality control*, edited by O. Fujimura and M. Hirano (Singular Publishing Group, San Diego, CA), pp. 147–170.

Styler, W. (**2017**). “On the acoustical features of vowel nasality in English and French,” Journal of the Acoustical Society of America **142**, 2469–2482.

Sundberg, J. (**2022**). “Objective Characterization of Phonation Type Using Amplitude of Flow Glottogram Pulse and of Voice Source Fundamental,” Journal of Voice **36**(1), 4–14, https://linkinghub.elsevier.com/retrieve/pii/S0892199720301107, doi: 10.

1016/j.jvoice.2020.03.018.

Sundberg, J., Andersson, M., and Hultqvist, C. (**1999**). “Effects of subglottal pressure variation on professional baritone singers’ voice sources,” The Journal of the Acoustical Society of America **105**(3), 1965–1971, http://asa.scitation.org/doi/10.1121/1.426731, doi: 10.1121/1.426731.

Sundberg, J., and Gauffin, J. (**1979**). “Waveform and spectrum of the glottal voice source,” in *Frontiers of Frontiers of speech communication research, Festschrift for Gunnar Fant*, edited by B. Lindblom and S. E. J. Öhman (Academic Press, London), pp. 301–320.

Swerts, M., and Veldhuis, R. (**2001**). “The effect of speech melody on voice quality,” Speech Communication **33**, 297–303.

Tabain, M., Garellek, M., Hellwig, B., Gregory, A., and Beare, R. (**2022**). “Voicing in Qaqet: Prenasalization and language contact,” Journal of Phonetics **91**, 101138, https://linkinghub.elsevier.com/retrieve/pii/S0095447022000134, doi: 10.1016/j.wocn.2022.101138.

Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Kreiman, J., Kob, M., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., and Wolfe, J. (**2015**). “Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization,” The Journal of the Acoustical Society of America **137**(5), 3005–3007, https://asa.scitation.org/doi/citedby/10.1121/1.4919349, doi: 10.1121/1.4919349 publisher: Acoustical Society of America.

Traill, A., and Jackson, M. (**1988**). "Speaker variation and phonation type in Tsonga nasals," Journal of Phonetics **16**(4), 385–400, https://linkinghub.elsevier.com/retrieve/pii/S0095447019305170, doi: 10.1016/S0095-4470(19)30517-0.

Wiley, J. F. (**2020**). "Multilevel and mixed effects model diagnostics and effect sizes" https://github.com/JWiley/multilevelTools.

Zhang, Z. (**2016**a). "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, http://asa.scitation.org/doi/10.1121/1.4944754, doi: 10.1121/1.4944754.

Zhang, Z. (**2016**b). "Mechanics of human voice production and control," Journal of the Acoustical Society of America **140**, 2614–2635.