Journal of Phonetics 94 (2022) 101155

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality

Marc Garellek

Department of Linguistics, University of California San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, USA

ARTICLE INFO

Article history:
Received 11 September 2021
Received in revised form 28 April 2022
Accepted 2 May 2022
Available online 29 June 2022

Keywords:
Voice
Voice quality
Phonation
Laryngeal
Glottal

ABSTRACT

Twenty years after the publication of a special issue in this journal on non-modal phonation (*JPhon* 2001: 49(4)), the phonetic study of voice quality has shown impressive progress. Here I focus on what we have learnt over these years about the linguistic sources of voice quality modulation. I stress how voice quality has a role to play in all of speech: among its many functions, the voice is involved in the articulation of all sounds, and voice quality is worth investigating as much for "modal" consonants and vowels as for contrastive phonation type. The voice also encodes structure at many prosodic levels: sub-lexical, lexical, and post-lexical. I further highlight some of the important technological developments and refinement of various voice quality models that have led to progress in the phonetic study of voice quality. Reviewing all of the above, one can only conclude that human voice has a central role to play in the phonetician's pursuit towards understanding spoken language.

1. Introduction

Given the current pace of scientific developments, it is certainly appropriate to look back and assess the progress made in a field after some twenty years. And although this special issue's temporal landmark – the beginning of the 21st century of the Gregorian calendar – is arbitrary, it is particularly appropriate for the phonetics of voice quality. Twenty years ago, we saw the publication of another special issue in this journal that was devoted entirely to the topic (*JPhon* 2001: 49(4)). That issue, entitled 'Nonmodal phonation' and edited by Stefanie Shattuck-Hufnagel, Bruce Gerratt, and Jody Kreiman, in turn stemmed from a workshop that had been held at the 1999 International Congress of Phonetic Sciences in San Francisco. Its five papers represent an excellent, if forcibly limited, sample of the many studies of the phonetics of voice quality that were being undertaken at the dawn of this century, and several of them can now be considered classic references.

Since the publication of that special issue, the following twenty years have seen a rapid increase in research on voice quality, and have led to the publication of a number of review papers devoted to the topic (Esposito & Khan, 2020; Garellek, 2019; Gobl & Ní Chasaide, 2010; Hirose, 2010), as well as seminal books from both the interdisciplinary discipline of voice studies (Kreiman & Sidtis, 2011) and linguistic phonetics (Esling, Moisik, Benner, & Crevier-Buchman, 2019). Now is as good a time as any to take stock of the theoretical and empirical landscape.

As described recently by Esling et al. (2019, pp. 1-2) and Garellek (2019), 'voice quality' can be defined differently, from the broadest sense of the term (how we identify an individual's voice) to the narrowest (the kind of voicing produced at the glottis– i.e., phonation type). In this paper I will discuss both conceptions of voice quality, but focus on changes in perceived voice quality which are generally attributed to articulations in the larynx and that are associated with individual speech sounds or larger prosodic constituents. I do so because laryngeal articulations have figured very strongly in the broader theoretical discussions of linguistic phonetics. For example, in the special issue of *JPhon* devoted to phonetic representations (1990:18(3), edited by Mary Beckman), laryngeal articulations were frequently brought up (e.g. in Keating, 1990; Ladefoged, 1990; Nearey, 1990; Pierrehumbert, 1990). But I will conclude the paper with a brief discussion of how a more complete theory of the voice will need to go beyond phonation type and prosodic variation of the voice, and also account for habitual and socially-indexed voice quality characteristics, as well as the relationship between affect and the voice.

E-mail address: mgarellek@ucsd.edu

https://doi.org/10.1016/j.wocn.2022.101155
0095-4470/Published by Elsevier Ltd.

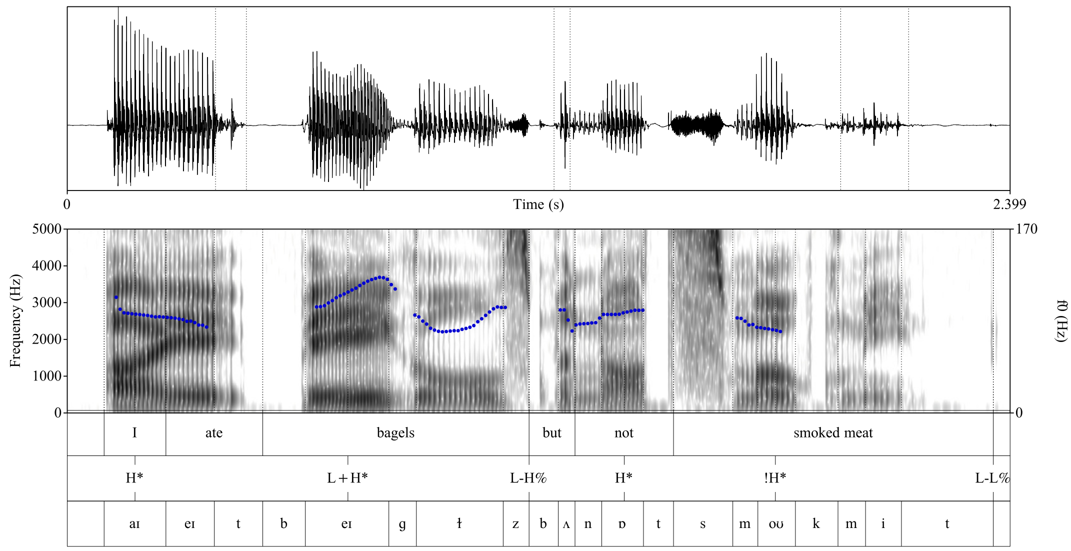

Fig. 1. Waveform, spectrogram, and f0 track (in blue) of the utterance "I ate bagels, but not smoked meat" with phonemic segmental and tonal (ToBI) transcriptions. H*, L+H*, and !H* refer to high, rising, and downstepped high pitch accents associated with prominent words; L-H% refers to a low phrase accent followed by a high phrase-final boundary tone, and L-L% refers to a low phrase accent followed by a phrase-final low boundary tone. Dashed vertical lines in the waveform enclose portions of irregular voicing around coda /t/; see relevant discussion in the body of the text. Note that the /t/ of "but" is not segmented on the phoneme tier because, as I discuss below, the /t/ is realized as glottalization that overlaps entirely with the preceding vowel.

The reason why the voice often takes centerstage in phonetic research is because it is everywhere and matters for everything in the phonetic signal. Laryngeal articulations are involved in the production of all speech sounds; in addition to serving as the source for sound phonation, laryngeal articulations can serve as standalone segments, lexically-contrastive phonation types, phrase boundary markers, and idiosyncratic characteristics of a speaker's voice. We have known this for quite some time. But it is only since the beginning of this century that we have really begun to understand all the ways in which laryngeal articulations can be derived from different components of the grammar (as well as paralinguistic and extralinguistic factors), and how these different sources interact.

2. Structure and the voice: a case study of my English

To illustrate how involved the voice is in speech, and how its quality connects to different levels of linguistic structure, throughout this paper I will refer to a single audio recording that I made of my own voice. Let us imagine the scenario in which, after returning from a visit to my hometown, my colleagues ask me "Did you have a nice trip?" A possible reply from me, in my Montreal-accented English, could be something like "I ate bagels, but not smoked meat" (implying the trip was, on culinary grounds, only partially successful).

What would my larynx be doing during the articulation of this utterance? To answer this question, my inclination would be to start with a phonemic transcription of the utterance: /aɪ eɪt ˈbeɪɡl̩z bʌt nɒt ˈsmoʊkˌmit/.[1] Note how the phonemic transcription is devoid of any glottal sounds. Still, the phonemes can also be divided up according to whether they are voiceless vs. voiced, a crucial distinction made by the voice. The waveform, spectrogram, and f0 track for this recording are shown in Fig. 1, along with the words and phonemes segmented. I also include a phonemic intonational contour using ToBI annotations for American English (Beckman & Ayers Elam, 1997), because we will later see how important the prosodic structure is for answering the question posed at the start of this paragraph. I remind the reader here that the intonational melody and modulations in voicing intensity also depend on the voice; f0 depends in large part on anterior-posterior stiffness of the vocal folds, and intensity varies with vocal fold properties, though to a much lesser extent than with changes in subglottal pressure (Zhang, 2016).

The phonemic transcription of the sample utterance initially suggests, rather deceptively, that the voice is involved only in the modulation of voicing– its presence vs. absence, and the rate and amplitude of voicing when present. This is clearly not the case: even a quick glance at the waveform will show that voicing is at times less periodic, and that these instances of irregular voicing are associated with some tokens of coda /t/ (those in "ate, but, meat") but not all of them ("not" is realized without irregular voicing). The signal thus shows evidence of **coda-/t/ glottalization**, a common phenomenon in North American English (Pierrehumbert, 1995; Zue & Laferriere, 1979), and indeed, across varieties of English spoken all over the world. In Section 5.5, I describe in some detail the progress made in our understanding of coda-/t/ glottalization. A narrower transcription in which coda-/t/ glottalization is transcribed could be something like [aɪ eɪ͡ʔt ˈbeɪɡl̩z bʌ͡ʔt nɒt ˈsmoʊkˌmi͡ʔt], excluding still many details about how the glottalization is realized phonetically. But since this example serves as an exercise in understanding layered, simultaneous, sound structure, I will avoiding modifying the IPA transcription further, preferring instead to segment on a separate tier the glottalization interval coinciding (approximately) with irregular voicing; I label the intervals simply as "ʔ" in Fig. 2, because differences in degree

[1] Despite the orthography, I pronounce the compound "smoked meat" as "smoke meat."

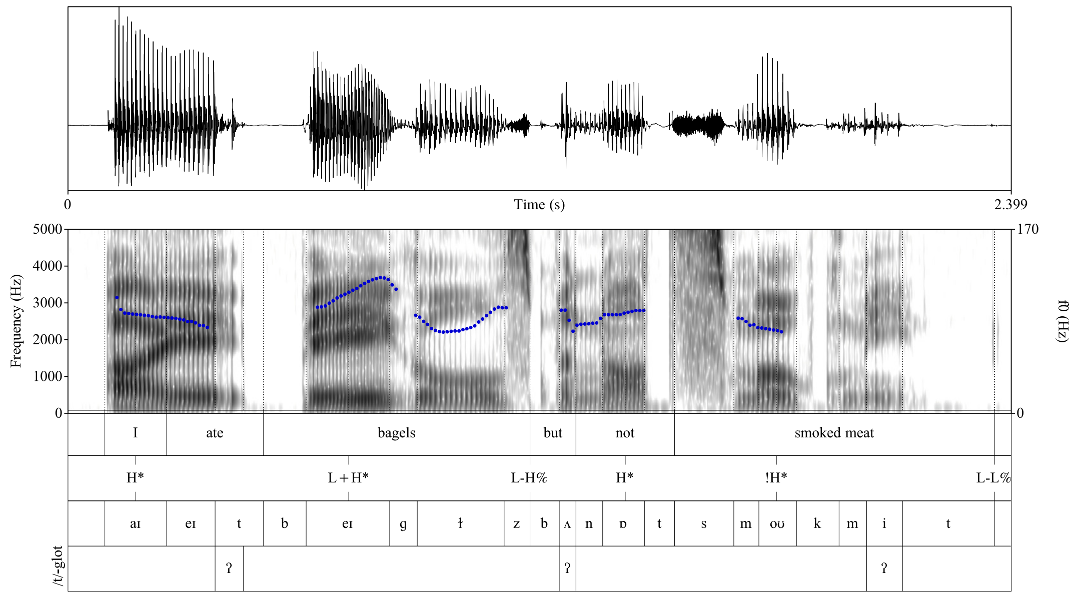

Fig. 2. Inclusion of bottom tier for coda-/t/ glottalization, with irregular voicing intervals labeled as "ʔ."

of glottalization are not particularly relevant here. Note also that, in the terminology adopted here, 'glottalization' refers to the phenomenon whereby /t/ is associated with glottal constriction – the term is not used here to refer to a type of voicing. Glottalization can thus be realized phonetically as a glottal stop or as creaky voice, itself often called 'glottalization.'

The visible onset of the utterance consists of a weak glottal pulse, after which the voicing of "I" is strong; I assume that pulse is the release of a glottal stop or a 'hard glottal attack'; if no glottal stop were present, we would instead expect a 'soft glottal attack,' where voicing intensity increases more smoothly (Baken & Orlikoff, 2000; Orlikoff, Deliyski, Baken, & Watson, 2009). Therefore, we have an example in this signal of **word-initial** or **vowel-initial glottalization**, another common phenomenon in English (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Nakatani & Dukes, 1977; Pierrehumbert & Talkin, 1992; Umeda, 1978), as well as across the world's languages. We have seen much progress in our understanding of this source of glottalization; I will address this in Section 5.4. In Fig. 3, I label the interval corresponding to vowel-initial glottalization simply as "ʔ." (Note that, as we only see evidence in the waveform and spectrogram of its release, the onset boundary of the glottal stop is arbitrarily placed.).

We haven't yet captured all instances of visible irregular voicing. At the end of the utterance, all voicing during the /mit/ of "smoked meat" appears to be creaky, so much so that

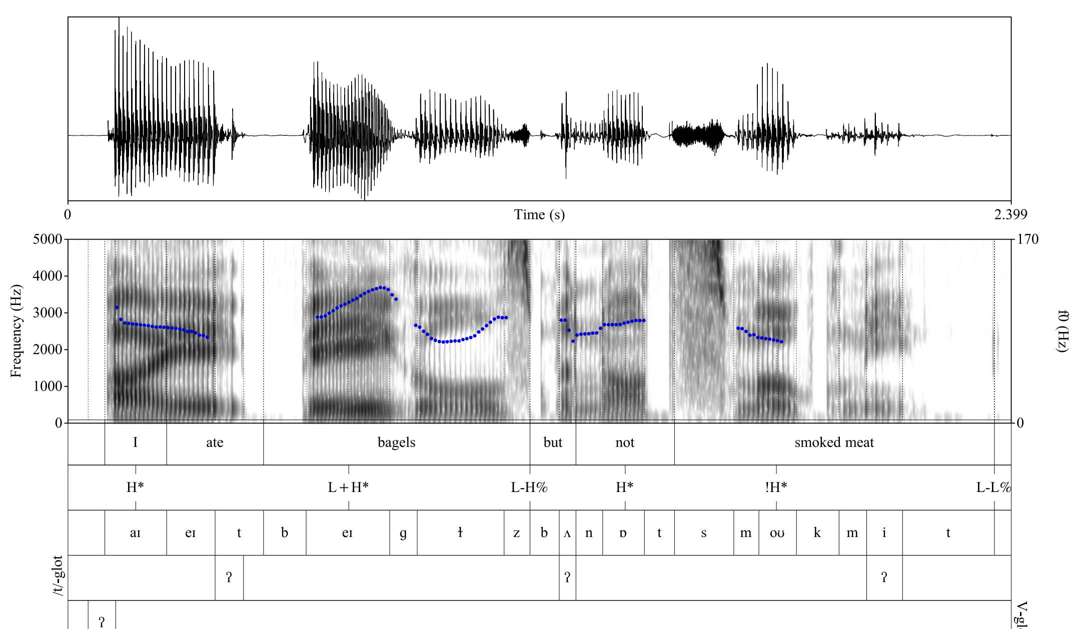

Fig. 3. Inclusion of a tier for vowel-initial glottalization, with the glottal stop labeled as "ʔ." Note that the onset of the interval is arbitrary. As discussed in Section 5.5, other transients in the signal are all attributed to the formation and release of oral stops.

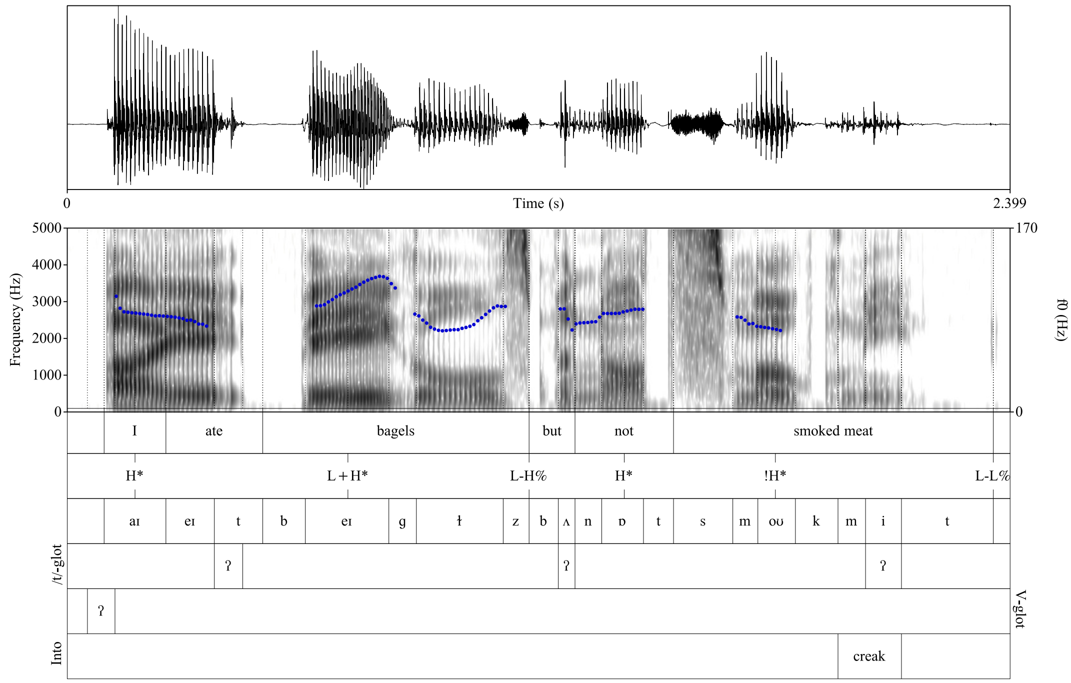

Fig. 4. Inclusion of a tier for phrase-final creak. It is labeled 'Into' for 'intonation,' because phrase-final creak is assumed to be due in large part to the low f0 associated with ends of phrases.

the f0 track from Praat (Boersma & Weenink, 2020) failed during that interval. We might at first attribute this long stretch of creaky voice to coarticulation from the coda-/t/ glottalization, but the source of creaky voice here is in fact ambiguous: it is a product of either coda-/t/ glottalization or **phrase-final creak**. Phrase-final creak refers to the interval of creaky voice occurring at the end of a phrase, due in large part to phrase-final low f0 (Redi & Shattuck-Hufnagel, 2001); thus unlike coda-/t/ or vowel-initial glottalization, its domain is that of phrases and thus often extends over multiple segments and words. It is found in North American English (Kreiman, 1982), but also in other varieties of English as well as in other languages (Henton & Bladon, 1988; Huber, 1988; Local, Wells, & Sebba, 1985). In Section 5.3 I discuss the research progress made in the past two decades in terms of our understanding of phrase-final creak. Assuming that some of the irregular voicing at the end of the sample utterance can be attributed to the presence of phrase-final creak, in Fig. 4 I now include an interval corresponding to its approximate domain, labeled simply as "creak."

So far, we have seen three distinct linguistic sources of voicing irregularity: coda-/t/ glottalization, vowel-initial glottalization, and phrase-final creak. We will soon see that Fig. 4 grossly underestimates what else my voice is or might be doing over the course of this utterance: it is likely that I am collapsing two distinct causes of irregularity under the label of 'coda-/t/ glottalization,' and for now I am ignoring several other sources of voice quality modulation.

We have long known about coda-/t/ glottalization, vowel-initial glottalization, and phrase-final creak; note that the references in this section are generally from the 20th century. But our understanding of voice quality as a product of distinct, sometimes overlapping, sources of linguistic structure has improved as a result of only recent developments in the phonetics of voice, and thus remains rather shallow today. In the next sections, I review some of the advances we have made in the last two decades, starting with an overview of technological innovations and voice models, and then returning to progress made in understanding the specific linguistic sources of voice quality modulation.

3. Technological innovations

3.1. Articulatory methods

This century has seen a rise in number of electroglottography (EGG) studies for an increasingly wide set of languages and phenomena, as well as the more in-depth quantification and analysis of EGG waveforms. The electroglottograph (or laryngograph) and its use in phonetic research are not new (Abberton, 1972; Fourcin & Abberton, 1971; Lecluse, Brocaar, & Verschuure, 1975), but until recently EGG studies have been limited in number and in depth of analysis. For example, between 1980–2000 in this journal, I count only a handful of papers focusing on characterizing differences between phonetic/phonological categories using EGG: these include Flege (1982) and Kuijpers (1993) on stop voicing contrasts in English and Dutch, Esling (1984) on phonation type and broad voice qualities, Traill and Jackson (1988) on phonation contrasts in Tsonga nasals, and Smith (1997) on /z/ devoicing in English. EGG analyses were often impressionistic and/or did not involve the kind of quantification of pulse shape that we use today.

In contrast, between 2000–2020, *JPhon* published at least fourteen EGG studies on a wider range of topics in linguistic phonetics, including studies of contrastive phonation type (Esposito, 2012; Khan, 2012), phonation changes as a result of phrasing, prominence, and vowel-initial glottalization

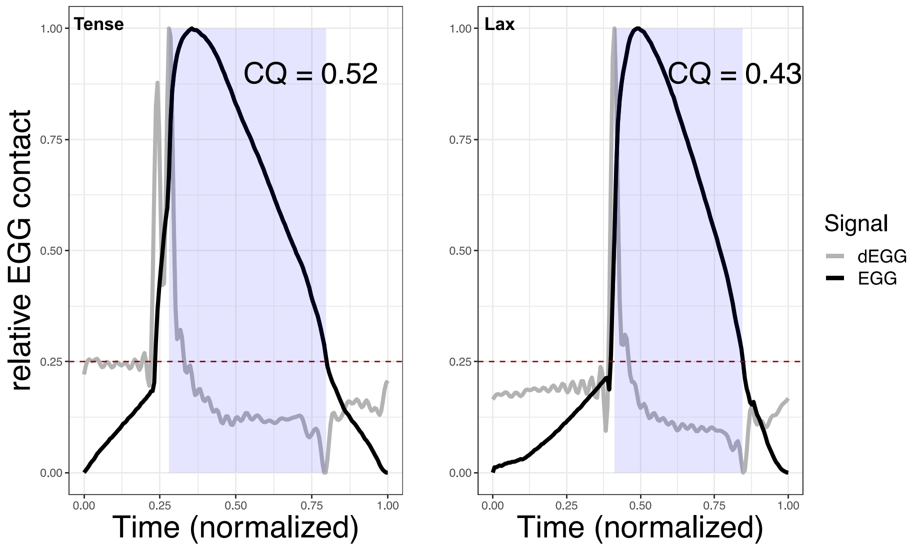

Fig. 5. Comparison of CQ for one pulse of a tense-voiced vowel from Bo /nạ 33/ 'deep, a drop of oil' (left) vs. one pulse of lax-voiced /nạ 33/'black' (right). CQ is measured from the positive peak in the derivative EGG (dEGG) signal (in grey) to a 25% (of maximal amplitude) threshold of the raw EGG signal (in black). The shaded area of each facet corresponds to the contact interval, with higher CQ for the tense token than for the lax one.

(Garellek, 2014; Lancia, Voigt, & Krasovitskiy, 2016), tongue-root contrasts and harmony (Guion, Post, & Payne, 2004; Olejarczuk, Otero, & Baese-Berk, 2019), sound change from phonation to vowel contrasts (Kuang & Cui, 2018), breathiness as correlates of contrastive and allophonic nasality (Carignan, 2017; Garellek, Ritchart, & Kuang, 2016), among others. (And as can be seen from the sample citations, most EGG studies in this journal have appeared only in the last decade.) These EGG studies also show more quantification of characteristics of the EGG pulses using pulse-internal landmarks, driven in part by the availability and ease of use of (Praat, Matlab) scripts and programs like EggWorks.[2]

By far the most commonly-used EGG measure is contact quotient (CQ), the proportion of a cycle during which the vocal folds are assumed to be in contact. CQ can be measured in different ways; the onset and offset of contact are usually defined by the maximum and minimum peak in the derivative signal or using an amplitude threshold. In Fig. 5, I illustrate how CQ can be measured for two tokens of tense (slightly creaky) vs. lax (slightly breathy) voice in the Tibeto-Burman language Bo, using data collected by Jianjing Kuang (and previously analyzed in several papers, including Garellek, Ritchart, & Kuang, 2016; Kuang & Keating, 2014; Keating, Kuang, Garellek, Esposito, & Khan, 2021; Kuang, 2011). CQ here is measured using the popular 'Hybrid' method (Howard, 1995), in which the onset of contact is defined by the positive peak in the derivative signal and the offset by an amplitude threshold in the raw EGG signal, here chosen as 25%, which also happens to align closely in both cases with the negative peak in the derivative signal. See Herbst (2020) for a recent detailed review of the history of EGG and quantification of EGG pulses according to various landmarks.

Aside from EGG, there have been advances in other articulatory methods. These include the use of ultrasound for measuring larynx height, and (high-speed) laryngoscopy for phonetic analysis of (epi-) laryngeal articulations such as larynx height, vocal fold vibration, ventricular incursion, aryepiglottic trilling, among others (Brunelle, 2010; Garellek, 2013; Moisik, 2013b; Moisik, Lin, & Esling, 2014; Moisik, Hejná, & Esling, 2019). In particular, these methods have played an important role in the Laryngeal Articulator Model of the voice (Esling et al., 2019), which I discuss in Section 4.

3.2. Acoustic measures

In a recent overview, I reviewed the acoustic measures used to investigate voice quality (Garellek, 2019, pp. 85–92), highlighting some of the theoretical assumptions and issues involved in their estimation and use. In the spirit of this special issue, here I focus instead on how voice acoustics has advanced in the last twenty-odd years. Some of the most important developments for voice acoustics have been the development of algorithms to correct for the frequencies and bandwidths of formants, and the development of software for automatic estimation of voice parameters.

By far the most common acoustic measures of voice quality are H1–H2 and other spectral tilt measures, which were proposed and used to distinguish contrastive phonation types and different voice qualities more generally (Bickley, 1982; Fischer-Jørgensen, 1967; Gordon & Ladefoged, 2001; Hanson, Stevens, Kuo, Chen, & Slifka, 2001); typically, breathy voice has a higher spectral tilt than modal voice, which in turn has a higher spectral tilt than creaky voice. These measures are also widely used because they robustly distinguish individual voices (Kreiman, Gerratt, & Antoñanzas-Barroso, 2007; Lee, Keating, & Kreiman, 2019), and because they are clearly correlated with important aspects of vocal fold physiology and glottal pulse shape; see Table 1. But there is a big catch with using spectral tilt measures: in the audio output, vowel formant structure affects harmonic amplitudes. So prior to the development of formant correction, the voice researcher had to control for vowel quality in order to measure spectral tilt appropriately. For this reason, acoustic studies of phonation

[2] http://phonetics.linguistics.ucla.edu/sales/software.htm

Table 1
Classes of commonly-used acoustic measures of voice quality. Excluded from the list are f0 and measures of the vocal tract filter characteristics (e.g. formant frequencies and bandwidths), which are also known correlates to voice quality. Measures in bold are recommended by Keating et al. (2021) because they emerged as important in structuring the cross-linguistic acoustic space for vocalic phonation types.

Spectral tilt	Spectral noise	Energy
H1–H2	CPP	**Root-mean-squared energy**
H2–H4	**HNR < 500 Hz**	**Strength of excitation**
H4–H2 kHz	HNR< 1500 Hz	
H2 kHz–H5 kHz	HNR< 2500 Hz	
H1–A1	HNR < 3500 Hz	
H1–A2	**Subharmonics-to-harmonics ratio**	
H1–A3		

type before the advent of the formant correction tended to be composed of small wordlists with mostly low vowels as targets; the amplitudes of H1 or H2 are generally less affected by F1 when, as for low vowels, F1 has a relatively high frequency. The alternative at the time was to inverse-filter the signal and measure spectral tilt at the (estimated) source, which requires specific tools not available to most researchers (Gobl & Ní Chasaide, 2010). The formant correction algorithms (Hanson, 1997; Iseli, Shue, & Alwan, 2007) essentially serve as “the poor man’s inverse filter,” allowing the researcher to estimate spectral tilt at the source by correcting for local effects of the formant frequencies and bandwidths without the challenges (with inverse filtering) of correcting for the entire spectrum, and without any special (and costly) equipment. With the formant correction, spectral tilt across different vowel qualities can be directly compared– though vowel quality may still affect voice quality (Esposito, Sleeper, & Schäfer, 2021).

Automated estimation of voice measures took hold in our community around 2008, thanks to scripts in Praat (Boersma & Weenink, 2020) and the development of the standalone program for voice acoustics, VoiceSauce (Shue, 2010; Shue, Keating, Vicenik, & Yu, 2011), which is a product of Pat Keating’s UCLA Voice Project[3] and her 2007–2012 National Science Foundation grant (with Jody Kreiman and Abeer Alwan) called “Production and Perception of Linguistic Voice Quality”.[4] Before VoiceSauce, voice quality measures were largely calculated manually– a task that was time-consuming, prone to human error, and that in practice limited how many tokens could be analyzed. (See also the analysis in Shue et al. 2011 for measures obtained from VoiceSauce, compared to those obtained from a Praat script and by-hand measurement.) Together, formant correction and automated estimation of acoustic measures of voice quality through programs like VoiceSauce have facilitated broader application of acoustic measures for voice quality analysis, enabling researchers to look more deeply at bigger datasets using more measures of voice quality.

Some researchers who are interested in undertaking an acoustic study of voice quality for the first time have told me that they feel overwhelmed by the sheer number of measures one can look at. Using VoiceSauce, for example, one can choose among several spectral tilt and harmonics-to-noise ratio measures, formant frequencies and bandwidths, subharmonics-to-harmonics ratio, RMS energy and strength of voicing, among others. (A summary of some of the main acoustic measures of voice quality can also be found in Table 1.) Confirmatory research should have measures that are chosen a priori, so which measures to choose? I refer the reader here to recent papers that have specifically addressed this issue: Garellek, 2019, 2020; Seyfarth & Garellek, 2018; Keating et al., 2021. Now, the measures we choose will certainly continue to change with time, especially as our models of voice phenomena evolve and are refined (see Section 4). In my own work, I have used different sets of measures to address different questions, and/or because I have adopted different voice models. For instance, Seyfarth and Garellek (2018) and Garellek and Esposito (2021) argue for minimally using two acoustic measures of voice quality, H1*–H2* and CPP, because H1*–H2* is correlated with vocal fold spreading and constriction and CPP with source noise; together, the two measures are assumed to be both necessary and sufficient for distinguishing the three most common phonation types (breathy, modal, and creaky voice). In contrast, in my recently analysis of phonation types of !Xóõ (Garellek, 2020), I used four spectral tilt measures and four noise measures, specifically because they are components of Kreiman, Gerratt, Garellek, Samlan, & Zhang (2014)’s psychoacoustic model of voice quality, which I set out to assess in that paper (and which I describe in more detail in Section 4.2.2).

I would also argue that the use of many acoustic measures has its benefits: researchers have acknowledged all along that voice quality, including contrastive phonation type, is phonetically multidimensional, and therefore could never be reduced to just one measure. In contrast, phonation type’s suprasegmental siblings – contrastive length, lexical tone, and lexical stress – are generally assumed to have either one “primary” acoustic correlate (duration for length, f0 for lexical tone), or three (f0, duration, and intensity for lexical stress). But these are, of course, convenient simplifications for our abstract phonological categories; we know for instance that lexical tone is not just about pitch, and that pitch is not just f0 (Kuang, 2013; Kuang & Liberman, 2018). In any phonetic study of spoken language, it is our models of the relationship between linguistic units and speech that determine which measure, or measures, we use. And as we’ll see in Section 4, our models of voice quality are (almost) always multidimensional. From a practical viewpoint, we also have more statistical tools for dealing with multidimensionality than ever before (Gubian, Torreira, & Boves, 2015; Roettger, Winter, & Baayen, 2019); many such tools – in particular linear discriminant analysis (e.g. Schertz & Khan, 2020), multidimensional scaling (e.g. Esposito, 2010), and principle components analysis (e.g. Kuang & Cui, 2018) – are being used to reduce and explain the voice’s acoustic multidimensionality.

3.3. Perception studies

Until the 21st century, there were very few studies of how linguistic voice quality is perceived; for example, in her 2010 paper, Christina Esposito counted only three previous studies of the perception of contrastive phonation type (Esposito, 2010, p. 306), one of which also appeared in the current cen-

[3] http://www.phonetics.ucla.edu/voiceproject/voice.html

[4] For a list of papers that came out of that grant, see https://www.nsf.gov/awardsearch/showAward?AWD_ID=0720304.

tury (Gerfen & Baker, 2005). Therefore, studying the perception of linguistic voice quality itself represents progress of the 21st century. For a recent survey of voice quality perception, I refer the reader to Esposito & Khan, 2020 (§4).

There has also been more recent progress due to the increase in paradigms and tools used in perception research. Eye-tracking has now been used to investigate the online perception of vowel-initial and coda-stop glottalization in English and Maltese (Chong & Garellek, 2018; Steffman, 2020; Mitterer et al., 2019). The development of the UCLA Voice Synthesizer (Kreiman, Antoñanzas-Barroso, & Gerratt, 2016), which allows for resynthesis of individual harmonics, provides researchers with the means of investigating various questions of interest, including: the effects of linguistic background on sensitivity to H1–H2 (Kreiman & Gerratt, 2010; Kreiman, Gerratt, & Khan, 2010); which specific components of the harmonic source spectrum are used by Hmong listeners to perceive contrastive breathiness (Garellek, Keating, Esposito, & Kreiman, 2013); and cue shifting between f0 and H1–H2 among listeners of English, Gujarati, Mandarin, and Vietnamese (Yang & Sundara, 2019). Also in the last decade, the KLSYN88 formant 'Klatt' synthesizer (Klatt & Klatt, 1990) has been used for perceptual studies of the phonetics of register (Brunelle, 2012), coda-stop glottalization (Chong & Garellek, 2018), phonation and nasality (Simpson, 2012), as well as paralinguistic affect (Yanushevskaya, Gobl, & Ní Chasaide, 2018).

4. Linguistic models of the voice

In the previous section I have highlighted just how much technological progress we have made during two decades of phonetic research on voice quality. Those advances have also coincided with changes in how we model the voice. In this section, I focus on advances in what may be called *linguistic* models of the voice; that is, models whose primary goals include differentiating sounds of the world's languages.

4.1. Glottal states

A very influential model of phonation types has been the continuum model of glottal states, first suggested by Ladefoged (1971, pp. 6–22) in his *Preliminaries to linguistic phonetics*. It is useful to review some of Ladefoged's insights in that reference to better understand how our conception of a glottal continuum has evolved over the years. Before we review its history and development, it is important to make certain things clear about what the continuum model is meant to model. Unlike the other models described in this section, the continuum model was always intended to be a model of the relationship between phonation types and glottal consonants, and of how these linguistic categories are articulated *specifically* at the glottis; the continuum does not model supraglottal articulation or aerodynamics, not because these are deemed to be irrelevant for phonation types, but because a unidimensional representation of glottal states is sufficiently explanatory, both for accounting for some of the articulatory differences between phonation types and glottal consonants, and for describing some of the relationships between these sounds that are found to occur within and across languages.

Table 2
Reproduction of Ladefoged (1971)'s 'Some states of the glottis' (Table 1, p. 8)

Phonetic term	Brief description	Symbols
Voice	Vibration of the vocal cords	m z b a
Voiceless	Vocal cords apart	m̥ s p h
Aspiration	A brief period of voicelessness during and immediately after the release of an articulatory stricture	sʰ pʰ
Murmur	"Breathy voice" – arytenoids apart, ligamental vocal cords vibrating	m̤ z̤ b̤
Laryngealization	"Creaky voice" – arytenoids tightly together, but a small length of the ligamental vocal cords vibrating	m̰ z̰ b̰
Glottal stop	Vocal cords held together	ʔ
Whisper	Vocal cords together or narrowed except between the arytenoids	(no symbol)

The continuum model came in response to earlier work by Catford (1964), who in a preliminary analysis had included over ten states of the vocal folds that were deemed to be linguistically significant. Based on phonological contrasts, Ladefoged reduces that number down to six (seven, if whisper is included). The original description of those states is reproduced in Table 2.

It is worth briefly reviewing several of these 'states of the glottis' to see how our conceptions of these states and terms used to describe them have changed. The reader should note how 'voiceless' here refers not just to the absence of voicing, but to the absence of voicing specifically via vocal fold spreading; aspiration is a subtype of voicelessness, defined in terms of its phasing with respect to another articulatory constriction or 'stricture'; 'murmur' is taken as synonymous with 'breathy voice,' and these days the use of the former term has largely fallen out of favor; 'laryngealization' is taken as synonymous with 'creaky voice,' as is still often the case today; the articulatory description of 'whisper' is the traditional one (e.g. Sweet, 1877, point 12), but according to Esling et al., 2019 (§2.3.7), whisper is produced with spread vocal folds and epilaryngeal constriction (cf. similar earlier descriptions for "medium whisper" and "strong whisper" in Scripture 1902 (p. 275)).

Ladefoged excluded some of Catford's earlier distinctions because of the lack of clear evidence for their use in linguistic contrasts, for example: "Catford (1964) distinguishes between creak and creaky voice, but I am not sure this distinction is needed for a theory of linguistic phonetics" (p. 15). We will soon see that such a distinction, at least as defined by Laver (1980) and more recently by Esling et al., 2019, remains relevant for certain voice models.

Ladefoged goes on to state that some of the glottal states "are really more like points on a continuum, which can be split up into a greater number of categories" (p. 17). For instance, 'breathy voice' can be subdivided into most breathy (labeled simply 'breathy voice' in his Fig. 4), intermediate 'murmur,' and least breathy 'lax' voice. Weaker forms of creaky voice, labeled 'tense,' are also treated as synonymous with those authors' use of 'stiff.' Ladefoged makes explicit that some categories, such as the breathy voice or 'murmur' of Gujarati, are realized on a 'continuum of glottal stricture,' and that even the glottal stop "itself may have several degrees of tightness" (p. 18), a point which we will return to later. In his *Preliminiaries* and subsequent work (Ladefoged, 1972; Ladefoged, 1973), Ladefoged refers to these glottal states or strictures as falling on a continuum, but Eugénie Henderson may have been the

Fig. 6. Reproduction of Gordon and Ladefoged (2001)'s 'Continuum of phonation types' (after Ladefoged, 1971; Fig. 1).

first to refer, at least in published work, to Ladefoged's model as a "continuum model" (Henderson, 1977).

The formulation of the continuum remained unchanged for thirty years until the publication by Gordon and Ladefoged (2001) in this journal's special issue on non-modal phonation. The authors schematized the continuum as I have reproduced in Fig. 6, and modified its formulation slightly from the original; for example, the continuum was framed solely in terms of the aperture between the arytenoids, with no explicit mention of the ligamental glottis, and weaker forms of breathy ('lax') and creaky ('tense') voice are discussed (pp. 398–399) but are not placed on the continuum. Whisper was excluded from the continuum (and is not mentioned in the paper), presumably because the model sets out to describe phonological contrasts, and evidence is still lacking that whisper may be used phonologically.

Several empirical findings support the idea that there is a continuum of glottal states. Klatt & Klatt (1990, Fig. 1) mention that spectral tilt measured by H1–H2 (and parameterized in the Klatt synthesizer as OQ) lowers continuously from more breathy, to modal, to more creaky ('laryngealized') phonation types. Within linguistic phonetic research, this was also found for contrastive breathy, modal, and creaky ('laryngealized') phonation types in Mazatec (Silverman, Blankenship, Kirk, & Ladefoged, 1995; Blankenship, 2002; Garellek & Keating, 2011), and Ladefoged (1983) provides even earlier support for an acoustic continuum for the phonation types in !Xóõ using another spectral tilt measure, H1–A1. Finally, Keating et al. (2021)'s cross-linguistic comparison of phonation types in eleven languages provides support for the presence of an acoustic continuum from more breathy-like, to more modal, to more creaky-like phonation types. We will return to that work in Section 5.1.

In *Sounds of the World's Languages*, Ladefoged and Maddieson (1996) incorporate the continuum model (there called 'Glottal stricture') within a broader model of phonation types that also incorporates glottal timing (to distinguish e.g. voiceless unaspirated vs. aspirated segments) and glottal movement (to distinguish e.g. voiceless stops with glottal closure vs. ejectives); I refer the reader to Table 11.4 (p. 372) of the 'Coda' chapter. This broader model of phonation types also appears as glottal articulatory features in Ladefoged 2007.

As work on phonation types has increased in the 21st century, the continuum of glottal states has been modified in different ways. For example, Garellek (2019, Fig. 4.2) changed the endpoint term referring to maximal opening from "voicelessness" to [h] to reflect the present-day conceptualization of voicelessness as being the absence of voicing in general, rather than vocal fold spreading in particular. In their review of phonation type patterning across languages, Esposito & Khan (2020, Fig. 3) include other commonly-used terms like 'lax/slack' and 'tense/stiff' on the continuum. And as I discuss in more detail below, Garellek, Chai, Huang, & Van Doren (2021, Fig. 12) argue that for some glottal sounds in particular languages, the continuum can be modeled to include, besides modal voice, only 'aspiration' (vocal fold spreading) and 'glottalization' (vocal fold or laryngeal constriction) without specifying the degree of spreading or constriction and thus without stipulating whether it is phonetically voiceless or voiced. I will return to that point in Section 4.4.

4.2. Broader voice quality

Within phonetic research, there exist two main models of voice quality defined at its broadest, i.e. models that aim to distinguish across voices (different voices for a given individual, and differences across individuals' habitual or long-term voice qualities). The two models are the Laryngeal Articulator Model (LAM), developed by John Esling and colleagues (e.g., Esling et al., 2019) and the psychoacoustic model of the voice developed by Jody Kreiman and colleagues (e.g., Kreiman et al., 2014). And because in these models 'voice quality' is broadly construed, they are also used to describe short-term changes in phonation, such as for contrastive phonation types.

4.2.1. The Laryngeal Articulator Model

The precursors to the LAM have a long history (e.g., see Esling et al., 2019, §1.3). To start, it is useful to go back the work by John Laver from the 1970s and 1980s, in particular his 1980 book entitled *The phonetic description of voice quality*.

Laver introduces his book on the first page with the broadest definition of voice quality: "Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, and not in the more narrow sense of the quality deriving solely from laryngeal activity." To account for various auditory changes in voice quality, Laver divides up the articulatory space into supralaryngeal settings (e.g., labial protrusion, but also including changes in larynx height) and phonatory ones thought then to be intrinsic to the larynx; the articulatory settings that matter are those that can be controlled by speakers. Laver's phonatory settings were inspired by work on pathological and singing voices, and include the neutral 'modal' voice (a descriptive rather than explanatory term borrowed from Hollien (1971) and later Hollien (1974), in the second volume of this journal) and falsetto, also inspired heavily by Harry Hollien's work. Both Hollien and Laver treat modal voice and falsetto to be different modes of vocal fold vibration, but for Hollien these modes are very much dependent on pitch (as in singing voice 'chest' and 'falsetto' registers), whereas for Laver they differ mainly in terms of voice quality. Modal voice and falsetto represent different modes of vibrations, and therefore are mutually exclusive. In contrast, the settings called 'creak' and 'whisper' can combine with modal voice and falsetto, as well as with each other, e.g. as in 'whispery creaky falsetto.' All falsetto voice qualities involve vocal fold vibration, but do not have 'voice' in their names; for example, the use of 'voice' in 'whispery creaky voice' implies non-falsetto voicing. Finally, the settings called 'harshness' and 'breathiness' have a more restricted combinatoriality with the other settings (see discussion pp. 111–118). Both breathiness and harshness involve noise, with harshness being 'raspier' in quality and characterized by f0 irregularity– even more so than creak (p. 138). Breathiness without voice is equivalent to aspiration, much like Catford (1964)'s description of the contrast in Hindi between breath (e.g., in voiceless aspirated consonants) vs. breathy voice (e.g., in voiced aspirated consonants).

Laver's distinction between 'creak' and 'creaky voice' warrants some discussion. For him, 'creak' is akin to the very low-frequency voice register found in singing, which is usually called 'fry' or 'pulse' register. Creak is defined acoustically as a voice quality involving vocal fold vibration within a range below that typically employed for modal voice and with total damping of the glottal pulses; the f0 should be low enough to give off the impression of, to quote Catford (1964, p. 32), "a rapid series of taps" rather than of sustained voicing. Articulatorily, both creak and creaky voice are characterized by adductive tension and medial compression of the vocal folds, as well as by incursion of the ventricular folds, which can also load downward onto the vocal folds to slow down vibration. Laver's 'creaky voice' is articulatorily and auditorily similar to 'creak,' but the inclusion of "voice" implies that 'creaky voice' has the same mode of vibration and similar f0 range as (non-creaky) modal voice.

Laver's broader theory of voice quality has been implemented in the Vocal Profile Analysis protocol (Laver, Wirz, Mackenzie, & Hiller, 1981), which is used, especially in clinical and dialect studies, to assess auditorily speakers' voice quality in terms of different vocal tract configurations (see Sóskuthy & Stuart-Smith, 2020 for a recent dialectal study). His model has further undergone substantial development in work by John Esling (himself a student of Laver's) and colleagues. Laver's auditory voice qualities derived from phonatory settings in the larynx are adopted by Esling et al. (2019) in the Laryngeal Articulator Model; where that model differs most substantially from Laver (1980)'s is that the LAM does not divide up the vocal tract into supralaryngeal vs. phonatory settings; instead, the LAM consists of two parts: the upper vocal tract, and the lower vocal tract or 'laryngeal constrictor,' which is involved in tongue retraction (e.g. for vowel [ɑ]) and pharyngealization in addition to phonation type: "All 'pharyngeal' and 'epiglottal' sounds, pharyngealization effects, epiglottalization effects, laryngealization, glottalization, and an array of lower-vocal-tract volume effects are produced by the action of the aryepiglottic constrictor mechanism together with associated tongue retraction and laryngeal raising" (p. 5). An important reconceptualization of the LAM is that laryngeal sounds are no longer just 'sources' of sound, as commonly framed in classic source-filter theory (e.g., Stevens, 2000); sounds with (epi-)laryngeal constriction function as both phonatory sources (vocal fold vibration, aryepiglottic trilling) and as filters (tongue retraction, vocal tract shortening).

Although they are retained in the LAM, Esling et al. (2019, Ch. 2) reconceptualize Laver's voice qualities accordingly: breath and breathy voice are defined by the lack of engagement of the laryngeal constrictor and a more open glottis than modal voice; harshness and creak both involve engagement of the laryngeal constrictor, but only creak has a low-pitched quality; whisper is a combination of breath and laryngeal constriction. As with Laver, modal voice is a mode of voicing that stands in opposition to falsetto. Esling et al. (2019, p. 64)'s distinction between 'creak' vs. 'creaky voice' is more clearly articulated than Laver's (or Catford's): according to the LAM, creak can be treated as voiceless because the glottal pulses during creak are *aperiodic*, while creaky voice can be treated as voiced because it has periodicity; in contrast, for Laver both are voiced: 'creak' has an f0, although the f0 should be low enough to perceive the glottal pulses discretely (Laver, 1980, p. 124).

The framework elaborated in the LAM by Esling et al. (2019) represents an immense achievement in the modeling of broad voice quality in this century, and its many implications and predictions will surely lead to much phonetic and phonological research for decades. Earlier work along the way include papers by Esling from the 1980s and 1990s (Esling, 1984, 1996, 1999), but the model's development increased rapidly since 2000 in work by Esling and collaborators (e.g., Edmondson & Esling, 2006; Esling, 2005; Esling & Harris, 2003, 2005; Moisik, 2013a; Moisik, Czaykowska-Higgins, & Esling, 2021; Moisik & Esling, 2011), driven to an important extent by technological advancements in articulatory methods discussed in Section 3. As such, the model represents a true 21st-century advancement in phonetics.

The Laryngeal Articulator Model has also been computationally implemented in ArtiSynth, a biomechanical simulation model of the vocal tract (Lloyd, Stavness, & Fels, 2012; Esling et al., 2019, §3.3.4). This allows for simulating how different laryngeal configurations contribute to overall voice quality. And as described above, the LAM relates articulation to auditory voice qualities. Through our current understanding of the relationship between articulation and acoustics, it is certainly possible to make predictions about how particular voice qualities should be realized acoustically; for example, see discussion by Esling et al. (2019) of the acoustic attributes of harshness (p. 15) and raised-larynx voice (p. 17, after Moisik, 2013b, pp. 293–308). But generally speaking, the LAM does not (yet) explicitly model the acoustic characteristics of the discretized voice qualities. The psychoacoustic model of the voice, described next, puts acoustics at the heart of voice quality modeling.

4.2.2. The psychoacoustic model of the voice

Like with the LAM, the psychoacoustic model defines the voice at its broadest, with the goal of being able to differentiate one voice from another as a function of changes that can derive articulatorily from any part of the speech apparatus. To better understand the motivation for the model, it is useful to review the discussion by Gerratt & Kreiman (2000). The psychoacoustic model of the voice is spearheaded by linguist Jody Kreiman, and reflects her longstanding collaboration with Bruce Gerratt, a speech-language pathologist. The primary motivation for this model comes from a persistent issue in clinical research: to assess voice disorders, clinicians have typically relied on rating scales (such as the CAPE-V; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2011) of perceived breathiness and other qualities commonly found in pathological voices; however, such ratings usually suffer from low inter- and intra-rater reliability. And as mentioned in Section 3, voice quality is inherently multidimensional and cannot be reduced to a single (acoustic or auditory) parameter. Individual scales of perceived quality are therefore not effective means of assessing voice quality. So where to go from here? Gerratt and Kreiman argue that a more reliable approach is through copy-synthesis, whereby copies of a pathological (or non-pathological) voice are synthesized using different sets of acoustic parameters; the synthesized copies are then played to listeners, who are asked to judge how similar the copies are to each other and to the original: “This comparative method not only can help determine the parsimonious list of synthesis parameters necessary for adequate modeling of pathological voice samples, but also has the potential to reveal the relative importance of the stimulus attributes that actually underlie listeners’ voice quality perceptions” (p. 340).

With over twenty years of research since the publication of that paper, Kreiman, Gerratt, and their colleagues have arrived at a list of synthesis parameters fitting for a unified theory of voice. These are articulated in various articles; I refer the reader to overviews in Garellek, 2019; Garellek, Samlan, Gerratt, & Kreiman, 2016; Kreiman, Lee, Garellek, Samlan, & Gerratt, 2021; and Kreiman et al., 2014. Very generally, the parameters include acoustic measures of the voice source (including f0, harmonic spectral tilt and noise at various frequency ranges), time-varying f0 and energy, and measures of the vocal tract transfer function like formant frequencies and zeroes, as well as their bandwidths. The model is computationally implemented in the UCLA Voice Synthesizer (Kreiman, Antoñanzas-Barroso, & Gerratt, 2010).[5] For practical reasons the synthesizer assumes a feed-forward source→filter computation of the acoustic signal, but articulatory and aerodynamic source-filter interactions are not ruled out.

The parameters are acoustic, but are chosen specifically because they relate (or should in principle relate) to articulation; I write “in principle,” because the relationship between the parameters and their articulatory sources is in many cases still poorly understood (see discussion in Garellek, 2019). However, recent mechanical and computational modeling by Zhaoyan Zhang has helped understand how different articulatory characteristics of the vocal folds, as well as subglottal pressure and the epilaryngeal area, affect some of the parameters (e.g. Zhang, 2016; Zhang, 2016; Zhang, 2017; Zhang, 2021). The model is also called ‘psychoacoustic’ rather than simply ‘acoustic’ because we know that listeners are sensitive to (most of) these parameters; e.g., see perceptual sensitivity to harmonic source spectral slope in Garellek et al., 2013; Garellek et al., 2016; Kreiman & Gerratt, 2010, 2012; Kreiman et al., 2010; and Kreiman et al., 2021.

Both the LAM and the psychoacoustic model seek to describe how voices, broadly construed, may be similar and different from one another, and both models have also been used to describe phonological contrasts. But the two models differ in other important respects. For the LAM, the main goal of modeling voice quality is to provide an explicit link between articulation and auditory qualities; for the psychoacoustic model, it is to provide an explicit link between acoustics and perception, and between acoustics and articulation. The LAM also discretizes voice qualities into types, e.g. ‘creaky voice’ vs. ‘whispery creaky voice,’ whereas the psychoacoustic model treats quality as inherently gradient, because it is a result of interactions between acoustic measures that are themselves continuous. The psychoacoustic model can be used to determine what parameters matter for differentiating discrete *phonological* categories like the contrastive phonation types of Hmong (Garellek et al., 2013) or !Xóõ (Garellek, 2020), or differences across individuals (Bishop & Keating, 2012; Kreiman et al., 2007; Lee et al., 2019), but (unlike in the LAM) those categories do not stem from an a priori taxonomy of possible voice qualities attested in humans.

Table 3
Summary table comparing how ‘breathy voice’ is modeled relative to modal voice in the LAM, and the possible acoustic correlates using parameters of the psychoacoustic model.

Laryngeal articulator model	Psychoacoustic model
Unconstricted larynx+voicing	Strong voicing intensity
Somewhat turbulent airflow+voicing	Higher spectral noise
Open cartilaginous glottis	Higher spectral tilt
Greater vertical phase difference during voicing	
Relatively free vibration along ligamental glottis	
Lowered larynx	Formant changes (e.g. lower F1)
Low pitch	Low f0

To consider both their compatibility and differences, let’s see how the LAM and psychoacoustic model might treat breathy voice. ‘Breathy voice’ is clearly defined in the LAM as an auditory quality with prototypical articulatory characteristics. However, the LAM doesn’t specify directly the acoustic ramifications of breathy voice; those must be inferred. The psychoacoustic model, for its part, does not discretize voice quality into categories like ‘breathy voice.’ But the category ‘breathy voice’ can still be inferred from relative values of acoustic measures that are parameters of the psychoacoustic model. The result is that some the LAM’s reported articulatory characteristics of breathy voice, relative to modal voice, can be partly converted into a combination of acoustic parameters of the psychoacoustic model, as shown in Table 3.

4.3. Acoustic subtypes of creaky and breathy voice

Very much in the spirit of the psychoacoustic model, the last few years have seen the development of novel classifications for acoustic subtypes of creaky voice (Keating, Garellek, & Kreiman, 2015) and breathy voice (Tian & Kuang, 2021). These classifications make direct reference to fine-grained changes in acoustic measures, many of which cannot be seen during visual inspection of the waveform and spectrogram. As such, they too are direct consequences of the technological innovations discussed in Section 3.

That ‘irregular’ or creaky voice can appear with different acoustic characteristics is well acknowledged (e.g., Batliner, Burger, Johne, & Kießling, 1993), and the taxonomy elaborated by Redi and Shattuck-Hufnagel (2001) in this journal’s special issue on non-modal phonation has been widely used. More recently, Keating et al. (2015) provide a revised taxonomy of subtypes of creaky voice. We argue for treating ‘creaky voice’ as an umbrella term for voice qualities that are often, but not necessarily, irregular. For a voice to be perceived as creaky, it should have at least one of the following properties: constriction, irregular voicing, or low f0 (because when the f0 is very low, at or below about 60 Hz, the voice generally sounds creaky). Each of these can be measured acoustically, via lower spectral tilt, higher noise, and low f0. From there we defined various subtypes according to combinations of these properties. For example, ‘prototypical creaky voice’ has all three characteristics, whereas ‘vocal fry’ is constricted, low in f0, but still periodic. This new classification has been used for describing what kind of creaky voice is found in acoustic studies (Davidson, 2019), but also for testing how subtypes of creaky voice differentially affect tone identification (Huang, 2020).

[5] See also the manual at https://www.uclahealth.org/head-neck-surgery/bga/software.

More recently, Tian and Kuang (2021) take a slightly different approach than Keating et al. (2015) in classifying subtypes of voice qualities that sound ‘breathy-like’ but which can nonetheless be differentiated acoustically. They identify three acoustic subtypes, which they call ‘breathy’ voice (found in Gujarati and White Hmong), ‘whispery’ voice (found in Shanghainese), and ‘lax’ voice (found in Southern Yi). However, these subtypes are gradiently differentiated: they depend on the relative importance of spectral tilt measures vs. spectral noise ones. ‘Breathy’ voice is distinguished from modal voice in Gujarati and White Hmong by a combination of both measures, but spectral tilt matters more; in contrast, ‘lax’ voice in Southern Yi differs from tense voice mostly in terms of spectral tilt, while the ‘whispery’ voice of the lower register in Shanghainese differs from the upper register’s more modal voice quality mostly in terms of noise.

4.4. Laryngeal sounds and the IPA

The 1989 Kiel Convention led to the first major revision of the International Phonetic Alphabet in some forty years. While many changes were made to the alphabet (International Phonetic Association, 1989; Ladefoged, 1990; Pullum, 1990), the only major change for laryngeal sounds was the adoption of the diacritic for creaky voice. But in the years following the Convention, there was much published discussion about the manner of articulation of glottal consonants and where they belong within the IPA chart (Ladefoged, 1990; Catford, 1990; Kloster-Jensen, 1991; Laufer, 1991; Iivonen, 1992). Issues with the representation of glottal consonants [h ʔ ɦ] and glottalized or aspirated consonants within the IPA have abated since the early 1990s but are far from settled. Perhaps because of the primacy of the voicing distinction in the IPA, we can easily forget that producing a phonetic glottal stop generally entails the production of ‘glottalized’ creaky voicing: when a glottal stop is preceded by a modally-voiced sound, for example, the vocal folds pass through a transition of creaky voicing as they constrict. So whether the goal is actually to produce a ‘glottal stop’ or ‘creaky voice,’ both targets will be realized with some amount of voiced glottalization.

This has important implications, starting with what we think the articulatory goals for ‘glottal stops’ and ‘creaky vowels’ might be. In a recent study of voicing during glottal consonants and non-modal vowels from Illustrations of the IPA, Garellek et al. (2021) find that utterance-initial glottal stops average only 50% voicelessness (over a ‘glottalization’ interval that, crucially, did not depend on voicing). This high percentage of voicing is especially surprising, given that utterance-initial position favors voicelessness, as I discuss in more detail in Section 5.5. But the conclusion must be that the phonetic distinction between glottal stops and creaky voice is not really the one traditionally assumed – namely, that glottal stops are voiceless while creaky voicing is voiced. Instead, both targets are at least partially voiced, only glottal stops have more voicelessness than at least some forms of creaky vowels. (The study also found that vowels described as ‘glottalized,’ ‘rearticulated,’ or ‘checked’ likely do not differ from glottal stops in terms of voicing.) Therefore, Ladefoged and Maddieson (1996, p. 75) were right when they famously said “In the great majority of languages we have heard, glottal stops are apt to fall short of complete closure, especially in intervocalic positions.” But while their statement is technically correct, I think it is more precise to state that glottal stops are apt to be *less voiced* in utterance-initial positions. This statement highlights the fact that glottal stops are usually (at least partially) voiced in all positions.

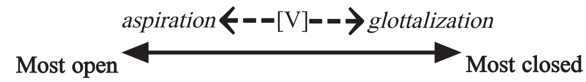

Fig. 7. Reproduction of Garellek et al. (2021)’s simplification of the continuum model (Fig. 12). Modal voice is represented as ‘[V].’ The authors argue that this level of simplification is sufficient for modeling aspiration and glottalization in many languages, given that these articulations vary predictably in terms of their voicing.

More generally, Garellek et al., 2021 argue that the variable voicing during so-called ‘voiceless’ [ʔ h] and ‘voiced’ [ɦ] implies that these sounds can for some languages be represented abstractly as ‘glottalization’ and ‘aspiration’ with unspecified voicing; see their modified continuum model in Fig. 7. We base this recommendation on the fact that voicing of glottal consonants across languages is rarely contrastive and usually predictable from respiratory and prosodic factors. The idea that glottal stop and [h] are better construed as articulatory (or, more accurately, as phonatory) gestures that vary prosodically from voiced to voiceless can be found as early as Pierrehumbert & Talkin (1992)’s work on English. But in our study we show that prosodic variation applies across languages. Surprisingly, we also find little support for a phonetic voicing distinction between ‘voiceless’ [h] and ‘voiced’ [ɦ], except in utterance-initial position. We therefore appeal for more research on this contrast.

The Laryngeal Articulator Model also has major implications for the IPA, from revising the articulatory terms used for the vowel chart (Esling, 2005), to how we model the differences between pharyngeal, epiglottal, and glottal consonants. For example, Esling et al. (2019, p. 79, Fig. 2.18) extend Ladefoged’s continuum model to include the effects of the laryngeal constrictor above the glottis. Not only are [h] and [ʔ] on a continuum of open-to-closed, but so are [ʔ] (moderate engagement of the laryngeal constrictor mechanism) and [ʡ] (full engagement of the laryngeal constrictor mechanism), harkening back to the possibility, as stated by Ladefoged (1971, p. 18) in his original formulation of the continuum model, for glottal stops to have “several degrees of tightness” (see also Section 4.1). All of these developments suggest that the major phonetic dimensions of glottal consonants and non-modal vowels – their place and manner of articulation, voicing, segment-internal phasing of the laryngeal articulation, and their relationship to supraglottal consonants – are not ideally captured in the current IPA chart. Nearly forty years after the Kiel Convention, we should probably expect another major IPA revision in a few years’ time.

5. Linguistic sources of voice quality modulation

The technological progress and continued development of voice quality and laryngeal models have enabled us to better understand the linguistic sources of voice quality. This is especially true for contrastive phonation types: linguistic categories defined primarily by changes in voice quality. I start this section with some review of what we have learnt about these sounds.

5.1. Contrastive phonation types

Given a set of contrastive phonation types, how can we go about characterizing their production, acoustics, and perception? That was the main question of some of the earliest linguistic phonetic work in voice quality. Included among these are the pioneering studies by Fischer-Jørgensen (1967) and Bickley (1982), both on Gujarati breathy vs. modal voice, from which we owe the widespread use of spectral tilt measures for description of contrastive categories; Ladefoged (1983) on several of the phonation types in !Xóõ, including the harsh "strident" vowels; Maddieson and Ladefoged (1985) on tense vs. lax voice in Jingpho, Hani, Yi and Wa; and Silverman et al. (1995) on the breathy vs. modal vs. laryngealized contrast in Mazatec.

In the first two decades of this century, phonetic studies of contrastive phonation type have increased in number; see the recent survey by Esposito & Khan, 2020. They have also increased in depth: the reader can compare older references to newer ones that made use of the same datasets: such as Ladefoged (1983) to Garellek (2020) (for !Xóõ) and Silverman et al. (1995) to Garellek and Keating (2011) (for Mazatec). Garellek and Keating (2011) investigated Mazatec phonation types in a larger set of words, using more acoustic measures at more time points in the vowel, and looked at more factors that can influence phonation type (such as aspiration and tone). The increased depth of analysis in the newer articles is a direct result of some of the technological and modeling progress outlined in the previous sections. Simply put, we now have a much better idea of how to model phonation types phonetically, and we have the tools to do so more quickly and using more measures.

In a longstanding collaboration headed by Pat Keating (Keating, Esposito, Garellek, Khan, & Kuang, 2011; Keating et al., 2012; Keating et al., 2021) and part of the UCLA Voice Project, we have explored the cross-language acoustic space of phonation contrasts, making ample use of VoiceSauce and its ability to measure multiple acoustic parameters related to voice quality. Our analysis of this space, with seven phonation types from eleven languages, reveals some interest findings about how phonation is structured phonetically. The acoustic space is largely two-dimensional, where the dimensions are obtained by multidimensional scaling and thus can correlate with many acoustic measures: one of the dimensions distinguishes modal from non-modal phonation types, and so is correlated closely with acoustic measures of noise; the other dimension distinguishes breathy-modal-creaky phonation types, and is correlated with spectral tilt measures. English, with no phonation contrast, falls in the middle of that two-dimensional space. Nearly all languages with phonation contrasts require just these two dimensions, but make use of a larger area within the space as the number of contrasting categories increases. !Xóõ, which has the largest number of phonation types in our sample (and likely the world), occupies the largest space and requires a third dimension to effectively distinguish all of its contrasts, mirroring the ways in which vowel spaces can be structured based on number of contrasts (Maddieson, 1984; Ladefoged & Maddieson, 1990; Guion, 2003; Becker-Kristal, 2010). And the fact that the second dimension (correlated with spectral tilt measures) differentiates breathy, modal, and creaky phonation types along an acoustic continuum provides acoustic support for a major assumption of the continuum model– namely, that breathy, modal, and creaky phonation types do in fact fall on a continuum.

Also of particular theoretical interest here, given my focus on linguistic structure, is the *domain* of phonation type within languages' sound systems. Typically, phonation types are considered features of vowels or consonants: for instance, Gujarati has a contrast between modal-voiced vs. voiced-aspirated consonants, as well as between modal and breathy vowels (Khan, 2012); vowels in Jalapa Mazatec contrast breathy, modal and laryngealized (creaky) phonation types (Garellek & Keating, 2011). Yet some languages contrast the phasing of phonation within a segment: for example, Zapotec languages may contrast 'rearticulated' vowels (with glottalization phased toward the vowel's midpoint) vs. 'checked' ones ending in glottalization (Ariza-García, 2018). Kehrein and Golston (2004) have claimed that laryngeal contrasts are a property of syllable structure: of onsets, nuclei, and rhymes, but not of segments. In many 'tense-lax' languages, such as Southern Yi, the contrast between slightly-constricted 'tense' voice and slightly-breathy 'lax' voice is thought to be contrastive at the syllable level, but words also tend to be monosyllabic (Kuang, 2011). Some more support for contrastive phonation at the syllable level comes from so-called 'register' languages, where the contrast pervades the entire syllable. For example, the Javanese 'tense-lax' contrast is thought to be contrastive just for stops, yet the lax 'stops' show breathiness into a following vowel, even when there is an intervening liquid (Thurgood, 2004; Matthews, 2015; Seyfarth, Vander Klok, & Garellek, 2019). It is possible then to analyze this contrast as a phonation contrast on syllables, but where the contrast is restricted to syllables with a stop onset. Yet in Javanese and other register languages of Southeast Asia, the contrast tends to be so phonetically multidimensional that it is unclear whether one should treat phonation as the primary feature of contrast; this multidimensionality is one reason why we call such contrasts 'register' ones in the first place. A better understanding of the phonetics of register contrasts is needed, though fortunately much work is currently underway (Brunelle, 2012; Brunelle & Kirby, 2016; Brunelle & Tạ, 2021; DiCanio, 2009). Finally, in Mixtec languages, Danish, and Rarámuri, glottalization can be analyzed as a contrastive suprasegment on morphemes (Caballero, 2008; Fischer-Jørgensen, 1989; Macaulay & Salmons, 1995), even if it is transcribed segmentally. In terms of the lexical phonology then, it appears that phonation contrasts can be found at most levels of sublexical structure (Avelino, 2016). This makes phonation type distinct from its suprasegmental siblings: length is typically thought to be contrastive on segments; lexical stress on syllables; and lexical tones on syllables or moras. In contrast, the prosodic domain of phonation types does not appear to be so constrained.

5.2. Voice quality associated with lexical tones and register

Lexical tones are very often produced with non-modal phonation, and pitch and voice quality interact to such an extent that it can be unclear whether a language's system of contrast is really one of 'tone' vs. 'phonation' (Esposito &

Khan, 2020, §3.2), or neither– as is the case of 'register' languages. Not surprisingly then, the increase in research on phonation in this century has been accompanied by work on the role of voice quality and glottal consonants in tone and register systems, particularly for languages of East and Southeast Asia as well as Mesoamerica; see for example studies by Andruski & Ratliff, 2000; Brunelle, 2009; DiCanio, 2012; Huang, 2020; Tạ, Brunelle, & Nguyễn, 2022; Turnbull, 2017; Yu & Lam, 2014; Zhu, 2012, among many others.

The interplay between pitch and voice quality remains a central topic in research on the phonetics of voice. We have seen much progress in terms of describing the interdependence of tone and phonation for particular languages, but also in developing explanatory theories to account for their close relationship. For example, Jianjing Kuang has shown how voice quality can be used to increase tonal dispersion, and that certain pitch levels are likely to be accompanied by physiologically-predictable changes in quality (Kuang, 2013; Kuang, 2013; Kuang, 2017). The Hmong-Mien language Black Miao has a typologically rare tone system that includes five tone levels and three contour tones. Although this system is quite dense, Kuang found that the tones are rendered more dispersed through the inclusion of voice quality differences. At the extremes of the pitch range, the tones tend to have laryngeal constriction, as evidenced by lower spectral tilt and higher EGG contact. She interprets the highest tone 55 as accompanied by stiff or tense voice, a result of increased longitudinal tension of the folds as f0 increases. The lowest tone 11 is accompanied instead by vocal fry, produced as a result of the vocal fold compression that naturally occurs at low f0. The tendency for tense voice at high f0 levels and for fry-like creaky voicing at low f0 levels is found in many tone languages, supporting Kuang's claim that these are expected consequences of changes in f0. Kuang also found that the level tone 33 in Black Miao is breathier than the other tones; unlike for tones 55 and 11 at the pitch extremes, this voice quality cannot be a consequence of f0. She therefore claims that the occurrence of breathy voice in tone systems is relatively independent of f0: in Black Miao's dense tone system, the breathiness of tone 33 helps increase tonal dispersion.

Kuang's claim that very-low f0 should naturally be accompanied by creaky voice also has important implications for how we study creaky voice in tone languages. For example, Kuang (2017) found that the occurrence of creaky voice for Mandarin's Tone 3 is driven by (tone-independent) f0 more than by lexical tone; other tones may also be creaky if they are produced at a low f0, such as in post-focal environments. Therefore, to disentangle tone-specific vs. f0-specific voice quality adjustments, researchers should ideally study phonation-tone interactions in a broad set of phrasal contexts (Garellek & Esposito, 2021). And as I discuss next, phrasing plays a very important role in accounting for voice quality variation in general.

5.3. Intonational sources of voice quality variation

As I discussed in Section 2, the study of phrase-final creak is not new to this century, though there is only limited work from before the year 2000. Pierrehumbert and Talkin (1992), Dilley et al. (1996), and Redi and Shattuck-Hufnagel (2001) were pioneers in accounting for glottal consonants and non-modal phonation within a prosodic theory of utterances. For example, Redi and Shattuck-Hufnagel (2001) showed (among many other things) that creaky voice is more common phrase-finally, especially at the ends of higher-level phrases. They suggest (p. 427) that creaky voice in English might be a prominence-lending and boundary-lending phenomenon, a hypothesis for which there is recent empirical support (Crowhurst, 2018; Steffman, 2020, 2021).

In later work, Garellek and Keating (2015) found that an important predictor of the occurrence of phrase-final creak (in American English and Mexican Spanish) is low f0, suggesting its domain might coincide with that of intonational low-pitched targets, much like Kuang (2017) showed for creaky voice in Mandarin. (It isn't surprising then that, in our sample utterance in Fig. 4, creak begins immediately after the final pitch accent – that is, at the start of the f0 drop for the L- phrase accent.) And in a word-identification task, Garellek (2015) showed that listeners can distinguish coda-/t/ glottalization and phrase-final creak, likely due to their distinct acoustic characteristics: Garellek & Seyfarth (2016) found that vowel adjacent to coda-/t/ glottalization got progressively noisier; in contrast, phrase-final creak was very noisy throughout (see also results by Peña, Davidson, & Orosco (2021), who compared creak to three types of glottalization). Recent overviews of phrase-final creak can be found in Dallaston & Docherty, 2020; and Davidson, 2021.

One of the most striking findings from this century was made by Janet Slifka in her 2000 dissertation and documented in later published work (Slifka, 2000, 2006): that utterance-final creak (which she called 'glottalization'), need not show increased glottal constriction. She found that utterance-final creaky voice can coincide with a drop in subglottal pressure and an *increase* in glottal opening. This has in turn led to a refinement of how subtypes of creaky voice are classified (Keating et al., 2015). Although creaky voice is prototypically construed as having increased constriction, Slifka's results show that this is clearly not always the case.

Two other higher-level prosodic sources of laryngeal articulations are subtler, in the sense that they are not readily perceived and therefore are under-researched. The first I call **prominence-lending constriction**. Recent research on American English has shown an increase in glottal constriction (as measured by means of EGG contact quotient) for vowels that are pitch accented (Bird & Garellek, 2019). This provides articulatory evidence for the finding by Campbell & Beckman (1997) that English pitch-accented vowels have lower spectral tilt. Similar acoustic results were found for Tongan (Garellek & White, 2015), in which lexical stress and post-lexical accent are confounded, and for both lexical and post-lexical stress-accent in Dutch (Sluijter & van Heuven, 1996). In our sample utterance from Section 2, the pitch-accented vowels on "I," "bagels," "not," and "smoked meat" are therefore likely to have an increase in glottal constriction, even if such a gesture is not strong enough to lead to visible creaky voice. Although perhaps functionally similar to vowel-initial glottalization, prominence-lending constriction is probably always weak; if it were not, we would expect pitch-accented syllables beginning with consonants to have vowels that are occasionally "broken" by a glottal stop.

The final type of higher-level prosodic laryngeal articulations is what I call **phrase-initial breathiness**. Using EGG, Garellek (2014) found that phrases in a read-speech corpus of American English and Mexican Spanish tend to begin with breathy voicing. This breathiness is also subject to domain-initial strengthening (Cho, 2016; Fougeron & Keating, 1997; Keating, Cho, Fougeron, & Hsu, 2003): the higher the prosodic phrase, the more breathiness there is. Vowel-initial glottalization overrides phrase-initial breathiness: phrases beginning with non-prominent (unstressed) vowels show initial breathiness, whereas those beginning with prominent vowels do not. If the phrase begins with a sonorant, phrase-initial breathiness is always found, regardless of whether the sonorant is in a stressed syllable. The larger EGG corpus created by Bird & Garellek (2019) provided additional evidence for phrase-initial breathiness in American English read speech.

The original discovery of breathiness on phrase-initial unstressed syllables was unexpected, and it is still unclear why phrases, especially utterance-medial ones, would begin with less constriction. (Utterance-initially, breathiness is expected if voicing is weak due to the low subglottal pressure, or if the vocal folds default to an abducted articulatory setting during pauses.) In Garellek 2014, I hypothesized that pitch reset may be responsible for its occurrence, but testing this remains a topic of future research.

In the sample utterance, the glottalization at the start of the utterance is expected to inhibit phrase-initial breathiness. But we could expect phrase-initial breathiness at the start of the second intonational phrase; it's unclear whether phrase-initial breathiness is inhibited by the coda-/t/ glottalization on the phrase-initial word. So in Fig. 8, I segment the locations of *expected* phrasal voice quality manipulations: pitch-accented vowels are expected to be accompanied by an increase in constriction, whereas phrase-initially we expect breathier voicing, as long as there isn't vowel-initial glottalization at the onset of the phrase.

5.4. Vowel-initial glottalization

Linguists have long analyzed vowel-initial words as being targets for epenthetic glottalization at the word onset; for instance, this process is frequently (though inconsistently) described in Illustrations of the IPA (Garellek et al., 2021, p. 20). Vowel-initial glottalization is also a prosodic phenomenon, at least in English: it is more likely to occur in hiatus environments, especially if the second word begins with a stressed vowel (Dilley et al., 1996; Garellek, 2012; Davidson & Erker, 2014). In several languages, it is also sensitive to post-lexical prosodic structure: it is more likely to occur under lexical prominence and at phrase onsets; the higher the prosodic phrase, the more likely it is for the glottalization to be realized as a full stop (Pierrehumbert & Talkin, 1992; Dilley et al., 1996; Fougeron, 2001; Garellek, 2012; Davidson & Erker, 2014; Garellek, 2014).

The implications of the analysis of glottalization by Garellek et al., 2021, discussed in Section 4.4, also pertain to vowel-initial glottalization. For example, word-initial vowels produced with creaky voicing, as well as word-initial vowels preceded by glottal stops, can both be analyzed as implementations of vowel-initial glottalization, whose target is glottal/laryngeal constriction that is unspecified for voicing. That is, whether the glottalization target is realized phonetically as (partially) voiceless or fully voiced is a matter of prosodic encoding, not the glottalization target itself; for a similar treatment of Maltese glottal stops as being licensed by prosodic structure, see Mitterer, Kim, and Cho (2021).

In Garellek 2014, I also claimed that vowel prominence (i.e., whether a vowel is lexically stressed but unaccented or both

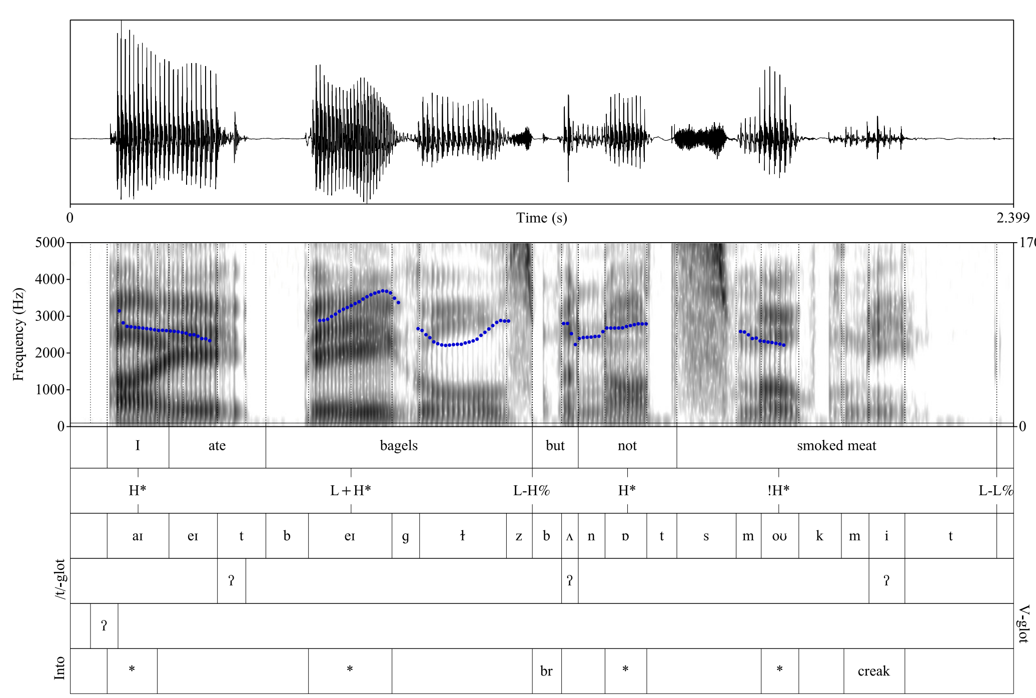

Fig. 8. Inclusion in the 'Intonation' tier of intervals of expected prominence-lending constriction ('*') on pitch-accented vowels and phrase-initial breathiness ('br') on the IP-initial word, up to the point of coda-/t/ glottalization.

lexically stressed and pitch-accented) is likely to matter for vowel-initial glottalization, because the glottalization helps in the realization of prominence through decreased spectral tilt (thus amplifying the formants). Nevertheless, vowel-initial glottalization of *non-prominent* vowels (those not bearing lexical stress or a pitch accent) is sometimes attested at phrase onsets (Dilley et al., 1996; Garellek, 2013), possibly as a means of phonation initiation and/or to mark phrase onsets in general (Umeda, 1978; Redi & Shattuck-Hufnagel, 2001).

Despite how common it is across languages, vowel-initial glottalization is not a universal phenomenon. It is avoided in Tongan, even at utterance onsets, presumably because the language has a contrast between null onsets and phonemic /ʔ/ (Garellek & Tabain, 2020). Yet in other languages with a phonological contrast between null onsets and /ʔ/ (such as Maltese and several Mayan languages; Mitterer et al., 2021; Bennett, 2016), this contrast can in some contexts be phonetically neutralized, because syllables with null onsets can have vowel-initial glottalization. At any rate, because such contrasts are rare, most languages have little functional reason to avoid vowel-initial glottalization. In Garellek 2014, I hypothesized that, if one of its functions in languages is to make initial stressed or pitch-accented vowels sound prominent (particularly phrase-initially, where voicing is breathy), then perhaps vowel-initial glottalization will be more widespread in languages where lexical stress is robustly realized phonetically.

Turning again to our sample utterance, it should therefore come as no surprise that the utterance-initial, pitch-accented "I" is glottalized, and that (because of domain-initial strengthening) that glottalization is realized as a full stop. Note also how "ate" is not obviously glottalized; this might contradict previous findings that hiatus will lead to glottalization on the vowel-initial word if it is lexically stressed (Dilley et al., 1996; Garellek, 2012; Davidson & Erker, 2014). Yet while the token of "ate" here is lexically stressed, it is not pitch-accented. Assuming that the presence of a pitch accent can strengthen glottalization (Garellek, 2014), it is conceivable then that the vowel does indeed have glottalization, only it is so weakly realized so as to not be visible in the waveform or spectrogram; in that case, the vowel onset would be expected to show spectral and articulatory characteristics of increased constriction. Possible vowel-initial glottalization on "ate" is therefore segmented in Fig. 9.

5.5. Laryngeal setting and coda-/t/ glottalization

Among the stops and fricatives, our IPA conventions tend to force a primary distinction in terms of voicing. For many languages this may be true phonologically, but how do speakers implement this distinction phonetically? This topic was explored in quite some detail (for stops) in the 1980s and 1990s, in tandem with the notions of phonetic categories, language-specific phonetics, and the phonetic grammar (Keating, 1984, 1990; Kingston & Diehl, 1994; Cho & Ladefoged, 1999).

One legacy of those years is that the phonetic categories for stops are today often described only as voiced, voiceless unaspirated, or voiceless aspirated (excluding categories like voiced aspirates, ejectives, and implosives), which also correspond to a unidimensional scale from lead voice onset time (VOT), short-lag VOT, to long-lag VOT. But VOT is generally thought to have a laryngeal gesture as its articulatory cause (Cho, Whalen, & Docherty, 2019). Basically, the vocal folds can spread, or they can constrict, and both gestures will result in voicelessness.

In reality, however, voicelessness is very dependent on prosody, especially the position within an utterance. In immedi-

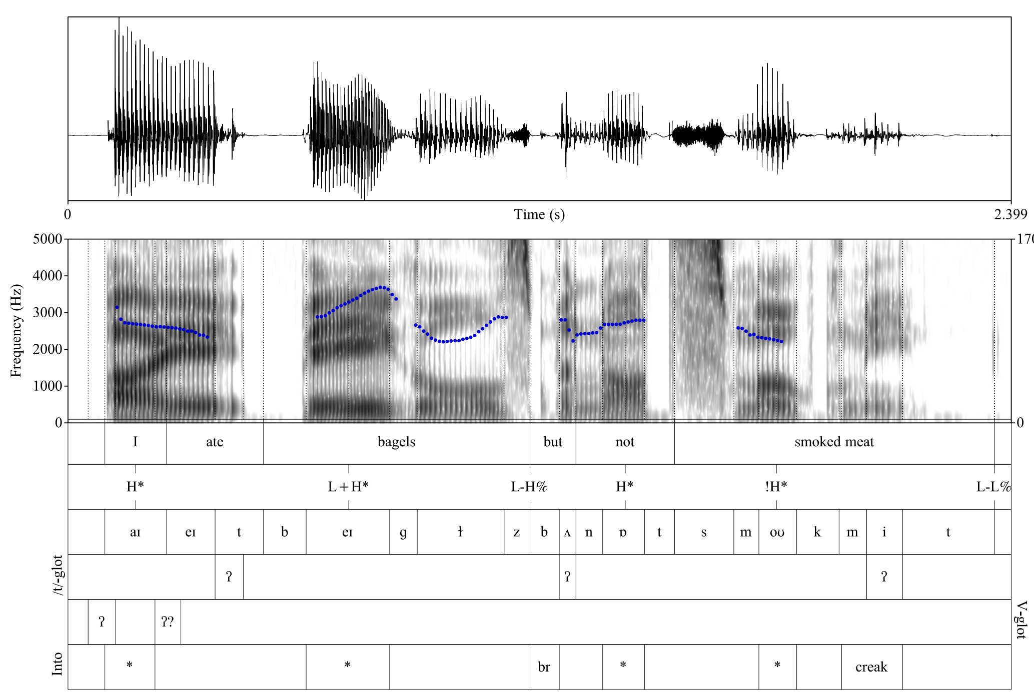

Fig. 9. Possible vowel-initial glottalization on "ate" is segmented and labeled as 'ʔʔ'.

ately utterance-initial and utterance-final positions, the low subglottal pressure can lead to voicelessness in obstruents, even without a laryngeal gesture. In other words, obstruent voicelessness “comes for free” at utterance boundaries. But utterance-medially, the subglottal pressure is high, favoring voicing even during stop closure (Westbury & Keating, 1986). In the middle of an utterance then, speakers need to actively inhibit voicing if obstruents are to be voiceless, and we do this through a laryngeal gesture. The accompanying laryngeal gesture of a consonant has also been called its ‘laryngeal setting’ (Ladefoged & Maddieson, 1996, §3.1).

Phoneticians often describe the laryngeal setting of non-pulmonic and fortis/lenis consonants (Cho, Jun, & Ladefoged, 2002; Miller, 2007; Vicenik, 2010; Gallagher, 2015; Bang, Sonderegger, Kang, Clayards, & Yoon, 2018), but the concept is also very important for consonants said to differ primarily in terms of voicing. Indeed, the study of the laryngeal setting for obstruent voicing has a long history; see Davidson (2016, 2018) and Cho et al., 2019 for recent overviews. And the notion of laryngeal setting is also relevant for our sample utterance, because it allows us to account for the occasional presence of voicelessness in the signal, and to refine our construal of coda-/t/ glottalization in varieties of English.

To illustrate this, let’s start by finding the voiceless portions in the sample utterance, based on the absence of periodicity. Voicelessness is found in six locations: at the start of the utterance; during the /t#b/ interval of “ate bagels”; during the /z#b/ interval of “bagels but”; the /t#s/ of “not smoked meat”; during the /k/ of “smoked meat”; and at the end of the utterance. How do we account for each of these periods of voicelessness?

The voicelessness at the utterance edges is predicted based on the low subglottal pressure. The utterance also begins with glottalization, which would further favor voicelessness. As for the /k/ of “smoke meat,” its release is slightly aspirated, so we can assume that it is produced with a glottal spreading gesture.[6]

The /t#b/ interval of “ate bagels” is voiceless, presumably because the /t/ is phonologically voiceless and phonetically glottalized, but also perhaps because in English initial /b/ can be realized as voiceless unaspirated – that is, with a glottal spreading gesture (Keating, 1984; Davidson, 2016). Devoicing during the /z#b/ interval of “bagels but” is similarly explained, only for /z/ we expect a partial glottal spreading gesture to ensure the high airflow needed for sibilance (Ohala, 1983; Smith, 1997). The intervening phrase boundary may also favor voicelessness due to phrase-initial breathiness, described earlier.[7]

Voicelessness[8] over the /t#s/ of “not smoked meat” is interesting, because it can be explained in several possible ways: it is uncontroversial that the /s/ should have a glottal spreading gesture, for the same reason that the /z/ does (in fact, the spreading gesture for /s/ is expected to be greater in magnitude than that of /z/). The possible explanations pertain to the /t/. Note that the onset of stop closure corresponds neatly to the offset of voicing; this means that the gesture responsible for devoicing is in-phase with the coronal one. Had voicing ceased towards the end of the stop closure, we could have assumed just one spreading gesture phased roughly with the onset of the /s/.

Does the /t/ in “not” have vocal fold spreading or constriction? Until recently, we probably would say that non-glottalized variants of coda /t/ have spreading, whereas glottalized ones have constriction, and that the choice of gesture is variable or “optional” (Huffman, 2005), possibly due to phonetic enhancement (Keyser & Stevens, 2006; Pierrehumbert, 1995; Pierrehumbert, 1994). Under that view, the /t/ in “not” must have vocal fold spreading, because it is perceptually non-glottalized. The glottalized variant of coda /t/ is the one expected under enhancement, yet we don’t expect enhancement before a voiceless sound.

For Australian English, Joshua Penney and colleagues have recently found increasing support for the enhancement hypothesis of coda-/t/ glottalization (Penney, Cox, Miles, & Palethorpe, 2018; Penney, Cox, & Szakay, 2020; Penney, Cox, & Szakay, 2021). For American English, however, results from perception and production studies show little support. In an eye-tracking study, Chong & Garellek (2018) found that the presence of glottalization before coda /t/ did not lead to faster looks to words ending in /t/ vs. /d/, though glottalization before coda /d/ caused delays in looks to the voiced candidate. Further, Seyfarth and Garellek (2020) provided a very detailed analysis of where visible coda-/t/ glottalization occurs in the Buckeye Corpus (Pitt et al., 2007), and argued that the distribution of glottalization can largely be explained by phonetic conditions which favor reduction of the coronal gesture, rather than enhancement.

In the sample sentence, the coronal gesture for the /t/ in “not” is fully realized; it has a clear stop closure and release. According to Seyfarth and Garellek (2020) then, the lack of glottalization is expected. However, we further claim that the data are consistent with the hypothesis that *all* instances of voiceless coda [t] are produced with glottal constriction. If we assume that coda [t] is voiceless because of a glottal constriction gesture, then it follows that this gesture will sometimes be perceivable: visible in waveforms and spectrograms, and audible to listeners. For example, if only the coronal gesture is reduced, listeners should perceive the glottalization more readily. The *expected* laryngeal settings for the obstruents discussed above are now shown in Fig. 10, over the tier that was previously used for *visible* coda-/t/ glottalization.

Reduction of the coronal gesture is not the only factor at play. Seyfarth and Garellek (2020) also argued that coda /t/ is realized as allophonic [ʔ] before sonorants (see also Davidson, Orosco, & Wang, 2021). For our target utterance, this means that we would want to analyze coda-/t/ glottalization in pre-sonorant “but (not)” as a planned glottal stop [ʔ]. Assuming our hypothesis is correct, (North) American English coda-/t/ glottalization would be analyzed not as a single process, but a reflex of several processes, including (but not limited to) gradient coronal reduction and planned allophony.

[6] Note that there is a transient before the /k/ closure. I attribute this to the oral formation of the velar closure, and not to a glottal pulse, on account of its frequency profile that is similar to the release burst.

[7] The visible transient with broadband energy during this interval is attributed to the release burst for /b/. The end of the /z/ has no visible transient, although the dotted line corresponding to the docking site of the L-H% might suggest one visually.

[8] The low-frequency energy seen in the spectrogram during [t] closure is considered to be noise, because the frequency of the single pulse (35 Hz) is much lower than that of adjacent voicing (93 Hz).

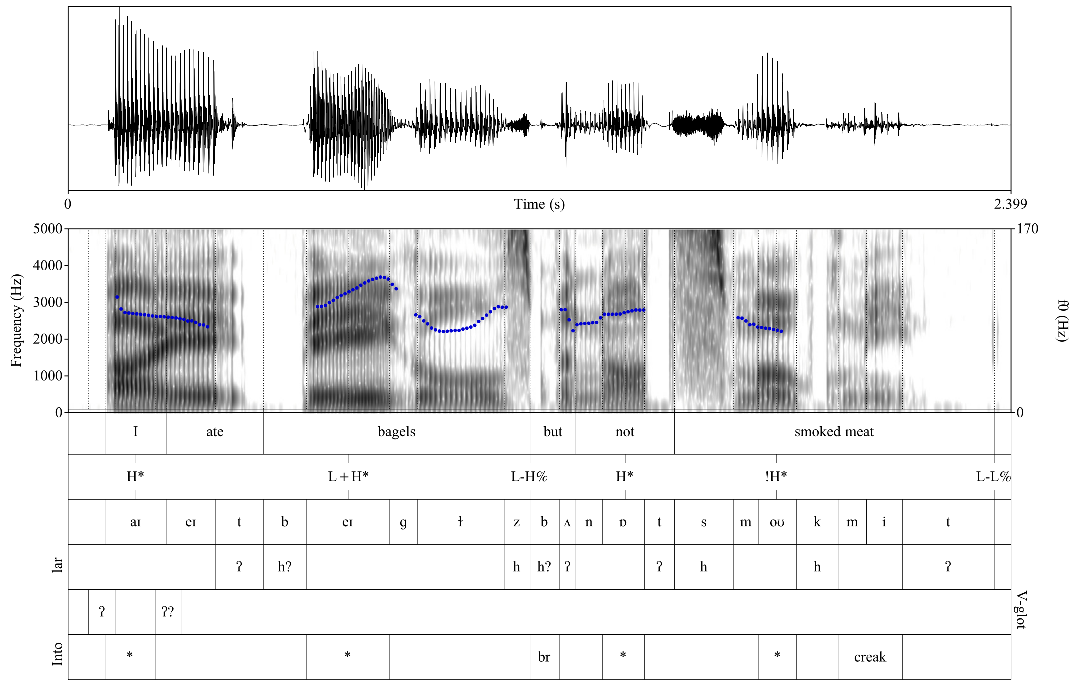

Fig. 10. The original tier for coda-/t/ glottalization is replaced with a 'lar' tier with intervals corresponding to the expected laryngeal setting of all obstruents. Vocal fold spreading is labeled as 'h' and glottalization/constriction as 'ʔ.' Because these are settings over consonants, the glottalization intervals have been adjusted in their duration to refer not (just) to the intervals with irregular voicing, but to the entire /t/ constriction.

6. Conclusions

6.1. Descriptions and explanations in the phonetics of voice quality

Throughout this paper, I have called attention to the fact that very rapid changes in voice quality are expected throughout an utterance, and that these changes result from the many linguistic sources of laryngeal articulations (e.g. Fig. 10). As we study phonation type, glottalization, and other voice phenomena in connected speech, we should therefore be mindful of the many possible sources of voice quality changes, from domains of different sizes, that can interact to affect a single sound. Some of those changes pertain to individual speech segments, whereas others are due to higher-order prosodic structure like phrase boundaries and post-lexical prominence.

As a consequence of the large amount of phonetic research done on the language, this is especially clear for English (e.g. Epstein, 2002, though see, e.g., Aguilar, 2020 for Nahuatl). We have known for quite some time that non-modal phonation in English is associated with initial vowels, coda /t/, and ends of phrases. Pierrehumbert and Talkin (1992), Dilley et. al. (1996), and Redi and Shattuck-Hufnagel (2001) showed that these phenomena are better understood from a theory of prosody, and subsequent research has only further highlighted the importance of understanding changes in voice quality with reference to prosodic structure: prosody modulates the voice quality associated with lower-level structures like segments; other changes to voice quality like prominence-lending constriction, phrase-initial breathiness, and phrase-final creak are a result of higher-order prosodic structure itself. All the while, sociolinguists have been making progress in describing the social factors that affect the occurrence of these phenomena (Dallaston & Docherty, 2020; Foulkes & Docherty, 1999). What remain underdeveloped are explanatory theories– those that explain what phonetic and phonological processes underlie these phenomena.

When I first introduced the sample utterance in Fig. 2, I located instances of creaky voice and associated them with a particular segmental anchor, coda /t/. In other words, I described an association between two phonetic events. It is still common to analyze the factors that predict (the occurrence and type of realization of) creaky voice when it occurs with coda /t/, but without proposing explanations for what underlies the association between coda /t/ and creaky voice. To be clear, I do not mean to imply that such work is of little worth or benefit– on the contrary, it provides a crucial step in developing explanatory theories of coda-/t/ glottalization (e.g. Seyfarth & Garellek, 2020). But we must remember that descriptive associations are not explanations.

And while we can formulate explanatory theories to account for distinct, overlapping sources of voice modulation, there are still practical limitations that remain. For one, there is no single acoustic parameter to provide us with something like a 'voice quality track' to help us find landmarks and anchors, as we do with an f0 or 'pitch' track. Moreover, some changes in voice quality continue to fly below the radar. Looking back on our analysis of non-modal phonation in the sample utterance, we began by reviewing some of the changes in voice quality that could be *seen* in the waveform and spectrogram: the utterance has visible creaky voicing and a glottal stop, as segmented in Figs. 2–4. Utterance-medially, the absence of voicing for certain obstruents can be considered visual evidence for the presence of a glottal spreading or constriction gesture, as segmented in Fig. 10. But we should not expect all changes in voice quality to be visible. Thus in Figs. 8–10, I left open the possibility that one instance of coda-/t/ glottalization leads to visibly hidden – but still measurable – constriction; that there might be phrase-initial breathiness at the onset of the utterance-medial intonational phrase; and that pitch-accented vowels are produced with increased constriction. The technological innovations of this century, notably the possibility for detailed articulatory and acoustic analyses of voice quality, have allowed us to infer and predict these sources of voice quality modulation that are not immediately visible in waveforms and spectrograms, but are nonetheless perceptible and measurable.

For a language with contrastive phonation types, an explanatory theory of the presence of non-modal phonation can easily stem from a theory of phonological contrast. For instance, Jalapa Mazatec has breathy vowels, because words with breathy vowels contrast with those with modal and creaky vowels. But the issue of overlapping sources of voice quality modulation is not unique to English: in Mazatec, any study of contrastive phonation type needs to deal with the fact that all words in citation form, regardless of the word-final vowel's phonation type, end with breathiness (Silverman et al., 1995; Garellek & Keating, 2011). Is this breathiness part of the phonetic implementation of the contrastive phonation type, or is it a result of another, possibly phrasal, source of non-modal voicing in the language?

Any explanation will also want to determine whether voice quality modulations are controlled by speakers, or are automatic consequences of other articulations. We have only just begun to address this. Some findings suggest, for example, that non-contrastive creaky voice of low tones and phrase-final creak can at times be an automatic (unplanned) consequence of speakers' f0 falling to a very low target (Garellek & Keating, 2015; Garellek et al., 2013; Kuang, 2013). At the same time, the fact that phonation type can also be contrastive on falling tones (Esposito, 2010), and that creak can acquire social meaning (Podesva & Callier, 2015), implies that the rapid changes in voice quality throughout an utterance are likely to be a product of both language-specific grammatical and automatic processes.

6.2. "I suspect we can all be busy for many generations"

This review article covers only some of the phonetic and phonological sources of laryngeal modulation. I have barely addressed, or didn't address, other sources that have also received more attention in this century. These include, but are scarcely limited to (with citations just from articles in *JPhon*): covariation between voice quality and pitch, lexical tones, and stress (Guion, Amith, Doty, & Shport, 2010; Kirby, 2014; Yang & Sundara, 2019); voice quality's role in register systems (Tạ, Brunelle, & Nguyễn, 2022; Wayland & Jongman, 2003) and consonant depressor contrasts (Jessen & Roux, 2002); phonation type during nasality and the nasalization of glottal consonants (Carignan, 2017; Garellek, Ritchart, & Kuang, 2016; Johnson, Barlaz, Shosted, & Sutton, 2019; Tabain et al., 2022); voice quality variation by vowel quality and tongue root (Guion et al., 2004; Kuang & Cui, 2018; Olejarczuk, Otero, & Baese-Berk, 2019); and the effects of articulatory constriction (consonants vs. vowels, obstruents vs. sonorants) on the realization of contrastive phonation (Khan, 2012; Berkson, 2019).

There are also many other sources of voice variation, including sentence type and particles (Zhuang & Hasegawa-Johnson, 2008; Chai, 2019), interactional and conversational factors (Local, 2003; Ogden, 2001; Winter & Grawunder, 2012; Lee, 2015), and sociolinguistic factors like gender, sexuality, and region (Clark & Watson, 2016; Eckert & Podesva, 2021; Podesva & Callier, 2015; Sóskuthy & Stuart-Smith, 2020; Zimman, 2013), among others. In the last two decades, we have learned more about these sources as well as voice quality's role in signaling affect (Yanushevskaya et al., 2018) and distinguishing among different speakers' voices (Kreiman et al., 2007; Lee et al., 2019; Park et al., 2016).

Voice quality varies as a function of all these factors, and therefore a complete theory of the voice will need to account for all of them; conversely, a theory of spoken language needs a theory of the voice. So where do we go from here in the next decade or two? Shortly before his death early this century, Peter Ladefoged was asked this very question, only about the future of phonetics. Given how suffused phonetics is with voice quality, it is telling – but not at all surprising – that Ladefoged's answer could just as well have been in response to a question about the future of voice research (from Kaye, 2006, p. 144):

> There will always be many different kinds of phonetic research. Those of us who like wandering around the world looking for previously unrecorded sounds will no doubt continue to do so. (I heard a contrast that was new to me just last month when I was in Ethiopia, a three way contrast in Kambaata between an intervocalic voiced liquid, /r/ or /l/, the same liquid with laryngealized voicing, and the same liquid preceded by a glottal stop.) Other phoneticians will continue with systematizing intonation and tone contrasts, which is much needed work if we are to get natural sounding synthetic speech – which is in itself a great field for phonetic research. There's also still much to learn about how we listen to speech, and new experimental techniques such as MRI are showering us with information about articulatory gestures. I suspect we can all be busy for many generations.

These pursuits – documenting sounds of the world's languages; understanding how the structure of language affects those sounds; recording, measuring, and synthesizing speech – are fundamental to phonetics, and voice quality has an important role to play in all of them. We phoneticians will ask many other questions of relevance to both voice and speech: for instance, questions about the evolutionary origin of speech and language, how these are represented in the brain and acquired by learners, and how they change over time. We still have so much to learn, and will indeed be busy for many generations.

Acknowledgement

I dedicate this paper to Pat Keating, on the occasion of her retirement, for the profound impact she has had on me as a researcher and for all she has done to advance the field of phonetics.

I'd like to thank Taehong Cho and Natasha Warner for the invitation to write this review article. The ways in which I think about voice quality have been influenced by many people over the years; I especially thank Pat Keating and Jody Kreiman. I'm very fortunate to have worked with and learnt much from wonderful collaborators in my research on the voice: Andrés Aguilar, Abeer Alwan, Liz Bird, Gabriela Caballero, Lucien Carroll, Yuan Chai, Gang Chen, Adam Chong, Christina Esposito, Bruce Gerratt, Birgit Hellwig, Yaqian Huang, Sameer Khan, Jianjing Kuang, Yoonjeong Lee, Anna Mai, Mai Moua, Ignatius Nip, Page Piccinini, Robin Samlan, Amanda Ritchart-Scott, Scott Seyfarth, Marija Tabain, Jozina Vander Klok, Maxine Van Doren, Jamie White, and Zhaoyan Zhang. I also thank Joy Iwamoto for her help in preparing to write this review, Stephanie Shih and members of the UCSD Phonetics-Phonology group for their feedback, as well as Adam Chong and Pat Keating for their comments on the initial version of the paper. Finally, I thank Taehong Cho, Lisa Davidson, and two anonymous reviewers for their help in improving this paper.

References

Abberton, E. (1972). Some laryngographic data for Korean stops. *Journal of the International Phonetic Association, 2*, 67–78. https://doi.org/10.1017/S0025100300000517.

Aguilar, A. E. (2020) Ph.D. thesis. *Phonology and phonetics of laryngeal sounds in Chicontepec Nahuatl*. San Diego: University of California.

Andruski, J. E., & Ratliff, M. (2000). Phonation types in production of phonological tone: the case of Green Mong. *Journal of the International Phonetic Association, 30*, 37–61. https://doi.org/10.1017/S0025100300006654.

Ariza-García, A. (2018). Phonation types and tones in Zapotec languages: a synchronic comparison. *Acta Linguistica Petropolitana, 14*, 485–517. https://doi.org/10.30842/alp2306573714220.

Avelino, H. (2016). Phonetics in phonology: a cross-linguistic study of laryngeal contrast. In H. Avelino, M. Coler, & W. L. Wetzels (Eds.), *The phonetics and phonology of laryngeal features in Native American languages* (pp. 157–179). Brill: Leiden/Boston.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*. San Diego: Singular Publishing Group.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005.

Batliner, A., Burger, S., Johne, B., & Kießling, A. (1993). MÜSLI: A classification scheme for laryngealizations. In *Proceedings of ESCA workshop on prosody, Lund.* (pp. 176–179). ISCA.

Becker-Kristal, R. (2010). Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus Ph.D. thesis. *UCLA*.

Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling, Version 3*. The Ohio State University Research Foundation.

Bennett, R. (2016). Mayan phonology. *Language and Linguistics Compass, 10*, 469–514. https://doi.org/10.1111/lnc3.12148.

Berkson, K. H. (2019). Acoustic correlates of breathy sonorants in Marathi. *Journal of Phonetics, 73*, 70–90. https://doi.org/10.1016/j.wocn.2018.12.006.

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

Bird, E., & Garellek, M. (2019). Dynamics of voice quality over the course of the English utterance. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 2406–2410). Canberra: Australasian Speech Science and Technology Association Inc.

Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. *Journal of the Acoustical Society of America, 132*, 1100–1112. https://doi.org/10.1121/1.4714351.

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

Boersma, P. & Weenink, D., (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.16., http://www.praat.org/.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics, 37*, 79–96. https://doi.org/10.1016/j.wocn.2008.09.003.

Brunelle, M. (2010). The role of larynx height in the Javanese tense∼lax stop contrast. In R. Mercado, E. Potsdam, & L. D. Travis (Eds.), *Austronesian and Theoretical Linguistics* (pp. 7–24). Amsterdam: John Benjamins.

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

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

Brunelle, M., & Tạ, T. T. (2021). Register in languages of Mainland Southeast Asia: the state of the art. In P. Sidwell & M. Jenny (Eds.), *The Languages and Linguistics of Mainland Southeast Asia* (pp. 683–706). de Gruyter Mouton: Berlin/Boston.

Caballero, G. (2008). *Choguita Rarámuri (Tarahumara) phonology and morphology*. Ph.D. thesis. University of California, Berkeley.

Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics, 63*, 87–105. https://doi.org/10.1016/j.wocn.2017.04.005.

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

Catford, J. C. (1990). Glottal consonants.. another view. *Journal of the International Phonetic Association, 20*, 25–26. https://doi.org/10.1017/S0025100300004229.

Catford, J. C. (1964). Phonation types: the classification of some laryngeal components of speech production. In: D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, J. L. M. Trim (Eds.), In honour of Daniel Jones: papers contributed on the occasion of his eightieth birthday, 12 September 1961, vol. 32, *International Journal of American Linguistics*, Chicago, 26–37.

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

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*, 120–141. https://doi.org/10.1111/lnc3.12178.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193–228. https://doi.org/10.1006/jpho.2001.0153.

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics, 72*, 52–65. https://doi.org/10.1016/j.wocn.2018.11.002.

Chong, A. J., & Garellek, M. (2018). Online perception of glottalized coda stops in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*, 4. https://doi.org/10.5334/labphon.70.

Clark, L., & Watson, K. (2016). Phonological leveling, diffusion, and divergence: /t/ lenition in Liverpool and its hinterland. *Language Variation and Change, 28*, 31–62. https://doi.org/10.1017/S0954394515000204.

Crowhurst, M. J. (2018). The joint influence of vowel duration and creak on the perception of internal phrase boundaries. *Journal of the Acoustical Society of America, 143*, EL147–EL153. https://doi.org/10.1121/1.5025325.

Dallaston, K., & Docherty, G. (2020). The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review of the literature. *PLoS ONE, 15*, e0229960. https://doi.org/10.1371/journal.pone.0229960.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50. https://doi.org/10.1016/j.wocn.2015.09.003.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*, 207–229. https://doi.org/10.1006/jpho.1999.0094.

Davidson, L. (2018). Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association, 48*, 331–356. https://doi.org/10.1017/S0025100317000330.

Davidson, L. (2019). The effects of pitch, gender, and prosodic context on the identification of creaky voice. *Phonetica, 19*, 235–262. https://doi.org/10.1159/000490948.

Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science, 12*, *e1547*. https://doi.org/10.1002/wcs.1547.

Davidson, L., & Erker, D. (2014). Hiatus resolution in American English: the case against glide insertion. *Language, 90*, 482–514. https://doi.org/10.1353/lan.2014.0028.

Davidson, L., Orosco, S., & Wang, S.-F. (2021). The link between syllabic nasals and glottal stops in American English. *Laboratory Phonology, 12*, 4. https://doi.org/10.5334/labphon.224.

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

DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics, 40*, 162–176. https://doi.org/10.1016/j.wocn.2011.10.006.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*, 423–444. https://doi.org/10.1006/jpho.1996.0023.

Eckert, P., & Podesva, R. J. (2021). Non-binary approaches to gender and sexuality. In J. Angouri & J. Baxter (Eds.), *The Routledge Handbook of Language, Gender, and Sexuality* (pp. 25–36). Oxford/New York: Routledge.

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*, 157–191. https://doi.org/10.1017/S095267570600087X.

Epstein, M. A. (2002). Voice quality and prosody in English Ph.D. thesis. *UCLA*.

Esling, J. H. (1984). Laryngographic study of phonation type and laryngeal configuration. *Journal of the International Phonetic Association, 14*, 56–73. https://doi.org/10.1017/S0025100300002796.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26*, 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*, 349–372. https://doi.org/10.11772/F00238309990420040101.

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics, 50*, 13–44. https://doi.org/10.1017/S0008413100003650.

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

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. M. Beck (Eds.), *A Figure of Speech: a Festschrift for John Laver* (pp. 347–383). Mahwah, NJ: Erlbaum.

Esling, J. H., Moisik, S. R., Benner, A., & Crevier-Buchman, L. (2019). *Voice quality: the laryngeal articulator model*. Cambridge: Cambridge University Press.

Esposito, C. M. (2010). Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association, 40*, 181–198. https://doi.org/10.1017/S0025100310000046.

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

Esposito, C. M., & Khan, S. D. (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass, 14*, e12392. https://doi.org/10.1111/lnc3.12392.

Esposito, C. M., Sleeper, M., & Schäfer, K. (2021). Examining the relationship between vowel quality and voice quality. *Journal of the International Phonetic Association, 51*, 361–391. https://doi.org/10.1017/S0025100319000094.

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

Fischer-Jørgensen, E. (1989). Phonetic analysis of the stød in Standard Danish. *Phonetica, 46*, 1–59. https://doi.org/10.1159/000261828.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics, 10*, 177–192. https://doi.org/10.1016/S0095-4470(19)30956-8.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics, 29*, 109–135. https://doi.org/10.1006/jpho.2000.0114.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728–3740. https://doi.org/10.1121/1.418332.

Foulkes, P., & Docherty, G. J. (Eds.). (1999). *Urban Voices: Accent Studies in the*. London: British Isles, Arnold.

Fourcin, A. J., & Abberton, E. (1971). First applications of a new laryngograph. *Medical and Biomedical Illustration, 21*, 171–182.

Gallagher, G. (2015). Natural classes in cooccurrence constraints. *Lingua, 166*, 80–98. https://doi.org/10.1016/j.lingua.2015.08.013.

Garellek, M. (2012). Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *UCLA Working Papers in Phonetics, 110*, 1–23.

Garellek, M. (2013). Production and perception of glottal stops. Ph.D. thesis. *UCLA*.

Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics, 45*, 106–113. https://doi.org/10.1016/j.wocn.2014.04.001.

Garellek, M. (2015). Perception of glottalization and phrase-final creak. *Journal of the Acoustical Society of America, 137*, 822–831. https://doi.org/10.1121/1.4906155.

Garellek, M. (2019). The phonetics of voice. In W. F. Katz & P. F. Assmann (Eds.), *Routledge Handbook of Phonetics* (pp. 75–106). Oxford: Routledge.

Garellek, M. (2020). Acoustic discriminability of the complex phonation system in !Xóõ. *Phonetica, 77*, 131–160. https://doi.org/10.1159/000494301.

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

Garellek, M., & Keating, P. (2015). Phrase-final creak: articulation, acoustics, and distribution. *Talk presented at the 2015 Annual Meeting of the Linguistic Society of America*.

Garellek, M. & Esposito, C. M., (2021). Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association*. Published online by Cambridge University Press 22 June 2021. doi:https://doi.org/10.1017/S0025100321000104.

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

Garellek, M., & Tabain, M. (2020). Tongan. *Journal of the International Phonetic Association, 50*, 406–416. https://doi.org/10.1017/S0025100318000397.

Garellek, M., & White, J. (2015). Phonetics of Tongan stress. *Journal of the International Phonetic Association, 45*, 13–34. https://doi.org/10.1017/S0025100314000206.

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

Garellek, M., Ritchart, A., & Kuang, J. (2016). Breathy voice during nasality: a cross-linguistic study. *Journal of Phonetics, 59*, 110–121. https://doi.org/10.1016/j.wocn.2016.09.001.

Garellek, M., Chai, Y., Huang, Y., & Van Doren, M. (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*. https://doi.org/10.1017/S0025100321000116.

Garellek, M. & Seyfarth, S. (2016). Acoustic differences between English /t/ glottalization and phrasal creak. In: *Proceedings of Interspeech 2016*, San Francisco, 1054–1058, 2016.

Gerfen, C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics, 33*, 311–334. https://doi.org/10.1016/j.wocn.2004.11.002.

Gerratt, B. R., & Kreiman, J. (2000). Theoretical and methodological development in the study of pathological voice quality. *Journal of Phonetics, 28*, 335–342. https://doi.org/10.1006/jpho.2000.0120.

Gobl, C., & Ní Chasaide, A. (2010). Voice source variation and its communicative functions. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd, pp. 378–423). Oxford: Wiley-Blackwell.

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

Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics, 49*, 16–40. https://doi.org/10.1016/j.wocn.2014.10.001.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica, 60*, 98–128. https://doi.org/10.1159/000071449.

Guion, S. G., Amith, J. D., Doty, C. S., & 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. https://doi.org/10.1016/j.wocn.2009.03.006.

Guion, S. G., Post, M. W., & Payne, D. L. (2004). Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics, 32*, 517–542. https://doi.org/10.1016/j.wocn.2004.04.002.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America, 101*, 466–481. https://doi.org/10.1121/1.417991.

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

Henderson, E. J. A. (1977). The larynx and language: a missing dimension? *Phonetica, 34*, 256–263. https://doi.org/10.1159/000259884.

Henton, C., & Bladon, A. (1988). Creak as a sociophonetic marker. In L. M. Hyman & C. N. Li (Eds.), *Language, Speech, and Mind: Studies in Honour of Victoria A. Fromkin* (pp. 3–29). London: Routledge.

Herbst, C. T. (2020). Electroglottography – an update. *Journal of Voice, 34*, 503–526. https://doi.org/10.1016/j.jvoice.2018.12.014.

Hirose, H. (2010). Investigating the physiology of laryngeal structures. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd, pp. 130–152). Oxford: Blackwell.

Hollien, H. (1971). Three major vocal fold registers: a proposal. In: *Proceedings of the 7th International Congress of Phonetic Sciences*, Montreal, 320–331.

Hollien, H. (1974). On vocal registers. *Journal of Phonetics, 2*, 125–143. https://doi.org/10.1016/S0095-4470(19)31188-X.

Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice, 9*, 163–172. https://doi.org/10.1016/S0892-1997(05)80250-4.

Huang, Y. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *Journal of the Acoustical Society of America, 147*, 1441–1458. https://doi.org/10.1121/10.0000721.

Huber, D. (1988). *Aspects of the communicative function of voice in text intonation: constancy and variability in Swedish fundamental frequency contours* Ph.D. thesis. University of Gothenburg.

Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics, 33*, 335–362. https://doi.org/10.1016/j.wocn.2005.02.004.

Iivonen, A. (1992). How [h] became a glottal fricative: comments on the letter by Martin Kloster-Jensen in JIPA 21 (1991). *Journal of the International Phonetic Association, 22*, 61–64. https://doi.org/10.1017/S0025100300004631.

International Phonetic Association (1989). Report on the Kiel Convention. *Journal of the International Phonetic Association, 19*, 67–80. https://doi.org/10.1017/S0025100300003868.

Iseli, M., Shue, Y.-L., & 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. https://doi.org/10.1121/1.2697522.

Jessen, M., & Roux, J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics, 30*, 1–52. https://doi.org/10.1006/jpho.2001.0150.

Johnson, S. E., Barlaz, M., Shosted, R. K., & Sutton, B. P. (2019). Spontaneous nasalization after glottal consonants in Thai. *Journal of Phonetics, 75*, 57–72. https://doi.org/10.1016/j.wocn.2019.05.001.

Kaye, A. S. (2006). An interview with Peter Ladefoged. *Journal of the International Phonetic Association, 36*, 137–144. https://doi.org/10.1017/S0025100306002519.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286–319. https://doi.org/10.2307/413642.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation (Papers in Laboratory Phonology 6)* (pp. 143–161). Cambridge: Cambridge University Press.

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, 1046–1049.

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

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* (pp. 2–7). Glasgow: ICPhS.

Keating, P., Kuang, J., Garellek, M., Esposito, C. M., & Khan, S., (2021). A cross-language acoustic space for phonation distinctions, Manuscript under revision. Available at https://linguistics.ucla.edu/people/keating/Keating-etal_ms_Nov2021.pdf.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics, 18*, 321–334. https://doi.org/10.1016/S0095-4470(19)30377-8.

Kehrein, W., & Golston, C. (2004). A prosodic theory of laryngeal contrasts. *Phonology, 21*, 325–357. https://doi.org/10.1017/S0952675704000302.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2011). Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. *American Journal of Speech-Language Pathology, 18*, 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017).

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language, 82*, 33–63. https://doi.org/10.1353/lan.2006.0051.

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

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*, 419–454. https://doi.org/10.1353/lan.1994.0023.

Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics, 43*, 69–85. https://doi.org/10.1016/j.wocn.2014.02.001.

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

Kloster-Jensen, M. (1991). Letter to the Editor. *Journal of the International Phonetic Association, 21*, 42. https://doi.org/10.1017/S0025100300006022.

Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics, 10*, 163–175. https://doi.org/10.1016/S0095-4470(19)30955-6.

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

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

Kreiman, J., Lee, Y., Garellek, M., Samlan, R., & Gerratt, B. R. (2021). Validating a psychoacoustic model of voice quality. *Journal of the Acoustical Society of America, 149*, 457–465. https://doi.org/10.1121/10.0003331.

Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies*. Oxford: Wiley-Blackwell.

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

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research, 50*, 595–610. https://doi.org/10.1044/1092-4388(2007/042).

Kreiman, J., Antoñanzas-Barroso, N., & Gerratt, B. R. (2010). Integrated software for analysis and synthesis of voice quality. *Behavior Research Methods, 42*, 1030–1041. https://doi.org/10.3758/BRM.42.4.1030.

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

Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. *Loquens*, 009. doi:https://dx.doi.org/10.3989%2Floquens.2014.009.

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

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

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

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America, 142*, 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.

Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. *Journal of the Acoustical Society of America, 136*, 2784–2797. https://doi.org/10.1121/1.4896462.

Kuang, J., & Liberman, M. (2018). Integrating voice quality cues in the pitch perception of speech and non-speech utterances. *Frontiers in Psychology, 9*, 2147. https://doi.org/10.3389/fpsyg.2018.02147.

Kuijpers, C. T. L. (1993). Temporal aspects of the voiced-voiceless distinction in speech development of young Dutch children. *Journal of Phonetics, 21*, 313–327. https://doi.org/10.1016/S0095-4470(19)31341-5.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

Ladefoged, P. (1972). Phonological features and their phonetic correlates. *Journal of the International Phonetic Association, 2*, 2–12. https://doi.org/10.1017/S0025100300000384.

Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics, 1*, 73–83. https://doi.org/10.1016/S0095-4470(19)31376-2.

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*. San Diego: College-Hill Press. 351–260.

Ladefoged, P. (1990). Some proposals concerning glottal consonants. *Journal of the International Phonetic Association, 20*, 24–25. https://doi.org/10.1017/S0025100300004217.

Ladefoged, P. (1990). Some reflections on the IPA. *Journal of Phonetics, 18*, 335–346. https://doi.org/10.1016/S0095-4470(19)30378-X.

Ladefoged, P. (2007). Articulatory features for describing lexical distinctions. *Language, 83*, 161–180. https://doi.org/10.1353/lan.2007.0026.

Ladefoged, P., & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics, 18*, 93–122. https://doi.org/10.1016/S0095-4470(19)30396-1.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lancia, L., Voigt, D., & Krasovitskiy, G. (2016). Characterization of laryngealization as irregular vocal fold vibration and interaction with prosodic prominence. *Journal of Phonetics, 54*, 80–97. https://doi.org/10.1016/j.wocn.2015.08.001.

Laufer, A. (1991). The "glottal fricatives". *Journal of the International Phonetic Association, 21*, 91–93. https://doi.org/10.1017/S0025100300004448.

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

Laver, J., Wirz, S., Mackenzie, J., & Hiller, S. (1981). *A perceptual protocol for the analysis of vocal profiles* (14. Edinburgh University Department of Linguistics Work in Progress.

Lecluse, F. L. E., Brocaar, M. P., & Verschuure, J. (1975). The electroglottography and its relation to glottal activity. *Folia Phoniatrica et Logopaedica, 27*, 215–224. https://doi.org/10.1159/000263988.

Lee, S. (2015). Creaky voice as a phonational device marking parenthetical segments in talk. *Journal of Sociolinguistics, 19*, 275–302. https://doi.org/10.1111/josl.12123.

Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *Journal of the Acoustical Society of America, 146*, 1568–1579. https://doi.org/10.1121/1.5125134.

Lloyd, J. E., Stavness, I., & Fels, S. (2012). ARTISYNTH: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery* (pp. 355–394). Springer.

Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics, 31*, 321–339. https://doi.org/10.1016/S0095-4470(03)00045-7.

Local, J., Wells, W. H. G., & Sebba, M. (1985). Phonology for conversation: Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics, 9*, 309–330. https://doi.org/10.1016/0378-2166(85)90029-3.

Macaulay, M., & Salmons, J. C. (1995). The phonology of glottalization in Mixtec. *International Journal of American Linguistics, 61*, 38–61. https://doi.org/10.1086/466244.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I., & Ladefoged, P. (1985). Tense and lax in four minority languages of China. *Journal of Phonetics, 13*, 433–454. https://doi.org/10.1016/S0095-4470(19)30788-0.

Matthews, M. (2015). An acoustic investigation of Javanese stop consonant clusters, in: A. Camp, Y. Otsuka, C. Stabile, N. Tanaka (Eds.), Proceedings of AFLA 21, Asia-Pacific Linguistics, Canberra, Australia, 201–217.

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

Mitterer, H., Kim, S., & Cho, T., (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese, Journal of Memory and Language.

Mitterer, H., Kim, S., & Cho, T. (2021). Glottal stops do not constrain lexical access as do oral stops. *PLoS ONE, 16*, e0259573. https://doi.org/10.1371/journal.pone.0259573.

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, 2011.

Moisik, S. R. (2013b). *The epilarynx in speech* Ph.D. thesis. Victoria: University of Victoria.

Moisik, S. R. (2013a). Harsh voice quality and its association with blackness in popular american media. *Phonetica, 69*, 193–215. https://doi.org/10.1159/000351059.

Moisik, S. R., Czaykowska-Higgins, E., & Esling, J. H. (2021). Phonological potentials and the lower vocal tract. *Journal of the International Phonetic Association, 51*, 1–35. https://doi.org/10.1017/S0025100318000403.

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*, 21–58. https://doi.org/10.1017/S0025100313000327.

Moisik, S.R., Hejná, M., & Esling, J. H., (2019). Abducted vocal fold states and the epilarynx: A new taxonomy for distinguishing breathiness and whisperiness, in: S. Calhoun, P. Escudero, M. Tabain, P. Warren (Eds.), Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 220–224.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62*, 714–719. https://doi.org/10.1121/1.381583.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18*, 347–373. https://doi.org/10.1016/S0095-4470(19)30379-1.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association, 31*, 139–152. https://doi.org/10.1017/S0025100301001116.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag.

Olejarczuk, P., Otero, M. A., & Baese-Berk, M. M. (2019). Acoustic correlates of anticipatory and progressive [ATR] harmony processes in Ethiopian Komo. *Journal of Phonetics, 74*, 18–41. https://doi.org/10.1016/j.wocn.2019.01.004.

Park, S.J., Sigouin, C., Kreiman, J., Keating, P., Guo, J., Yeung, G., Kuo, F.-Y., & Alwan, A., (2016). Speaker Identity and Voice Quality: Modeling Human Responses and Automatic Speaker Recognition, in: Proceedings of Interspeech 2016, San Francisco, 1044–1048.

Orlikoff, R. F., Deliyski, D. D., Baken, R. J., & Watson, B. C. (2009). Validation of a glottographic measure of glottal attack. *Journal of Voice, 23*, 164–168. https://doi.org/10.1016/j.jvoice.2007.08.004.

Peña, J., Davidson, L., & Orosco, S. (2021). The independence of phrasal creak and segmental glottalization in American English. *JASA Express Letters, 1*, 075205. https://doi.org/10.1121/10.0005601.

Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics, 66*, 161–184. https://doi.org/10.1016/j.wocn.2017.10.001.

Penney, J., Cox, F., & Szakay, A. (2020). Glottalisation, coda voicing, and phrase position in Australian English. *Journal of the Acoustical Society of America, 148*, 3232–3245. https://doi.org/10.1121/10.0002488.

Penney, J., Cox, F., & Szakay, A. (2021). Effects of glottalisation, preceding vowel duration, and coda closure duration on the perception of coda stop voicing. *Phonetica, 78*, 29–63. https://doi.org/10.1159/000508752.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375–394. https://doi.org/10.1016/S0095-4470(19)30380-8.

Pierrehumbert, J., 1994. Knowledge of variation, in: Papers from the parasession on variation, 30th meeting of the Chicago Linguistic Society, Chicago Linguistic Society, Chicago, 232–256.

Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology: voice quality control* (pp. 39–60). San Diego: Singular Publishing Group.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 90–117). Cambridge: Cambridge University Press.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University.

Podesva, R. J., & Callier, P. (2015). Voice quality and identity. *Annual Review of Applied Linguistics, 35*, 173–194. https://doi.org/10.1017/S0267190514000270.

Pullum, G. K. (1990). Remarks on the 1989 revision of the International Phonetic Alphabet. *Journal of the International Phonetic Association, 20*, 33–40. https://doi.org/10.1017/S0025100300004047.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29*, 407–429. https://doi.org/10.1006/jpho.2001.0145.

Schertz, J. & Khan, S., (2020). Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu, Journal of Phonetics 81. https://doi.org/10.1016/j.wocn.2020.100979.

Roettger, T. B., Winter, B., & Baayen, H. (2019). Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics, 73*, 1–7. https://doi.org/10.1016/j.wocn.2018.12.001.

Scripture, E. W. (1902). *The elements of experimental phonetics*. New York: Charles Scribner's Sons.

Seyfarth, S., & Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics, 71*, 425–450. https://doi.org/10.1016/j.wocn.2018.09.001.

Seyfarth, S., & Garellek, M. (2020). Physical and phonological causes of coda /t/ glottalization in the mainstream American English of central Ohio. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(24), 1–33. https://doi.org/10.5334/labphon.213.

Seyfarth, S., Vander Klok, J., & Garellek, M. (2019). Evidence against interactive effects on articulation in Javanese verb paradigms. *Psychonomic Bulletin & Review, 26*, 1690–1696. https://doi.org/10.3758/s13423-019-01637-2.

Shue, Y.-L. (2010) Ph.D. thesis. *The voice source in speech production: data, analysis and models*. UCLA.

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, Hong Kong, 1846–1849.

Silverman, D., Blankenship, B., Kirk, P., & Ladefoged, P. (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics, 37*, 70–88. https://www.jstor.org/stable/30028043.

Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics, 40*, 477–490. https://doi.org/10.1016/j.wocn.2012.02.001.

Slifka, J. (2000). Respiratory constraints on speech production at prosodic boundaries Ph.D. thesis. *MIT*.

Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice, 20*, 171–186. https://doi.org/10.1016/j.jvoice.2005.04.002.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471–2485. https://doi.org/10.1121/1.417955.

Smith, C. L. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics, 25*, 471–500. https://doi.org/10.1006/jpho.1997.0053.

Sóskuthy, M., & Stuart-Smith, J. (2020). Voice quality and coda /r/ in Glasgow English in the early 20th century. *Language Variation and Change, 32*, 133–157. https://doi.org/10.1017/S0954394520000071.

Steffman, J. A. (2020) Ph.D. thesis. *Prosodic prominence in vowel perception and spoken language processing*. UCLA.

Steffman, J. (2021). Prosodic prominence effects in the processing of spectral cues. *Language, Cognition and Neuroscience, 36*, 586–611. https://doi.org/10.1080/23273798.2020.1862259.

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge: MIT Press.

Sweet, H. (1877). *A handbook of phonetics*. Oxford: Clarendon Press.

Tabain, M., Garellek, M., Hellwig, B., Gregory, A., & Beare, R. (2022). Voicing in Qaqet: prenasalization and language contact, Journal of Phonetics 91 (101138). https://doi.org/10.1016/j.wocn.2022.101138.

Tạ, T. T., Brunelle, M., & Nguyễn, T. Q. (2022). Voicing and register in Ngãi Giao Chrau: Production and perception studies. *Journal of Phonetics, 90*(101115). https://doi.org/10.1016/j.wocn.2021.101115.

Thurgood, E. (2004). Phonation types in Javanese. *Oceanic Linguistics, 43*, 277–295. https://doi.org/10.1353/ol.2005.0013.

Tian, J., & Kuang, J. (2021). The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association, 51*, 202–228. https://doi.org/10.1017/S0025100319000148.

Traill, A., & Jackson, M. (1988). Speaker variation and phonation type in Tsonga nasals. *Journal of Phonetics, 16*, 385–400. https://doi.org/10.1016/S0095-4470(19)30517-0.

Turnbull, R. (2017). The phonetics and phonology of lexical prosody in San Jerónimo Acazulco Otomi. *Journal of the International Phonetic Association, 47*, 251–282. https://doi.org/10.1017/S0025100316000384.

Umeda, N. (1978). Occurrence of glottal stops in fluent speech. *Journal of the Acoustical Society of America, 64*, 88–94. https://doi.org/10.1121/1.381959.

Vicenik, C. (2010). An acoustic study of Georgian stop consonants. *Journal of the International Phonetic Association, 40*, 59–92. https://doi.org/10.1017/S0025100309990302.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics, 31*, 181–201. https://doi.org/10.1016/S0095-4470(02)00086-4.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics, 22*, 145–166. https://doi.org/10.1017/S0022226700010598.

Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics, 40*, 808–815. https://doi.org/10.1016/j.wocn.2012.08.006.

Yang, M., & Sundara, M. (2019). Cue-shifting between acoustic cues: Evidence for directional asymmetry. *Journal of Phonetics, 75*, 27–42. https://doi.org/10.1016/j.wocn.2019.04.002.

Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2018). Cross-language differences in how voice quality and f_0 contours map to affect. *Journal of the Acoustical Society of America, 144*, 2730–2750. https://doi.org/10.1121/1.5066448.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *Journal of the Acoustical Society of America, 136*, 1320–1333. https://doi.org/10.1121/1.4887462.

Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *Journal of the Acoustical Society of America, 139*, 1493–1507. https://doi.org/10.1121/1.4944754.

Zhang, Z. (2016). Mechanics of human voice production and control. *Journal of the Acoustical Society of America, 140*, 2614–2635. https://doi.org/10.1121/1.4964509.

Zhang, Z. (2017). Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model. *Journal of the Acoustical Society of America, 142*, 2311–2321. https://doi.org/10.1121/1.5008497.

Zhang, Z. (2021). Interaction between epilaryngeal and laryngeal adjustments in regulating vocal fold contact pressure. *JASA Express Letters, 1*. https://doi.org/10.1121/10.0003393 025201.

Zhu, X. (2012). Multiregisters and four levels: A new tonal model. *Journal of Chinese Linguistics, 40*, 1–17. https://www.jstor.org/stable/23754196.

Zhuang, X., Hasegawa-Johnson, M., 2008. Towards interpretation of creakiness in Switchboard, in: Proceedings of Speech Prosody, 37–40.

Zimman, L. (2013). Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men. *Journal of Language and Sexuality, 2*, 1–39. https://doi.org/10.1075/jls.2.1.01zim.

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of America, 66*, 1039–1050. https://doi.org/10.1121/1.383323.