The phonetics of voice[1]

Marc Garellek, University of California San Diego

Chapter in *The Routledge Handbook of Phonetics* (W. Katz and P. Assmann, editors)

Revised 14th June 2018

1 Introduction

This chapter focuses on the phonetics of the voice. The term 'voice' is used to mean many different things, with definitions varying both within and across researchers and disciplines. In terms of voice articulation, definitions can vary from the very narrow – how the vocal folds vibrate – to the very broad, where 'voice' is essentially synonymous with 'speech' – how the vocal folds and all other vocal tract articulators influence how we sound (Kreiman and Sidtis, 2011). In this chapter, I will use the term 'voice' to refer to sound produced by the vocal folds, including but not limited to vocal fold vibration. I have chosen to focus only on a narrow conception of the voice in order to constrain the discussion; as we will see, the phonetics of voice – even when it concerns only vocal fold articulation – is remarkably complex and of great relevance to phonetic and linguistic research. In contrast, I will use the term 'voice quality' to refer to the percept resulting from the voice: in other words, different vocal fold configurations have specific perceptual ramifications, which we will call changes in voice quality. The distinction between voice and voice quality adopted here is therefore analogous to that made between 'fundamental frequency (f0)' and 'pitch'.

Why should we be interested in the phonetics of the voice? Linguists are interested in how specific forms contribute to linguistic meaning; for spoken languages, phonetic and phonological research addresses this goal from the point of view of how sounds contribute to meaning. Because the vocal folds are part of the speech apparatus, a complete understanding of the sound-to-meaning relationship requires knowledge of how sounds produced by the vocal folds contribute to linguistic meaning. There are two main contributions of the vocal folds: first, their movements can be used contrastively in languages. That is, in some languages changes in (a) the presence vs. absence of vocal fold vibration, (b) the rate of vibration, and (c) the quality of vibration can signal a change in lexical meaning: compare (a) English /ˈslɒpi/ 'sloppy' vs. /ˈslɒbi/ 'slobby'; (b) White Hmong /tɔ ˧/ 'pierced' vs. /tɔ ˩/ 'wait'; and (c) Jalapa Mazatec breathy-voiced [ⁿd̤æ̤ ˥] 'horse', creaky-

voiced [ⁿdæ̰˥] ‘buttock’, and modal-voiced [ⁿtʰæ˥] ‘seed’ (Silverman et al., 1995; Garellek and Keating, 2011).[2]

Second, because of their position downstream from the trachea and upstream from the remainder of the vocal tract, all speech is first modulated by the vocal folds. This may seem counterintuitive at first, given that phoneticians tend to speak of two sources of sound excitation – voicing (produced by the vocal folds) and noise (produced in the vocal tract in all obstruents but [ʔ], [h], and [ɦ]). But even for sounds characterized by acoustic noise generated in the vocal tract (such as [t] and [s]), the vocal folds still assume a particular articulatory target; for example, voiceless obstruents (even when unaspirated) often show some degree vocal fold spreading (Munhall and Löfqvist, 1992), presumably to inhibit voicing and/or to facilitate the high airflow needed to generate turbulence for voiceless fricatives. The voice is thus part of the production of *all* speech sounds, regardless of whether it is used to make phonological contrasts. Aside from phonological meaning, the voice contributes to changes in prosody, syntax, discourse, as well as to speaker identity and emotional state (Choi et al., 2005; Zhuang and Hasegawa-Johnson, 2008; Gobl and Ní Chasaide, 2010; Esling and Edmondson, 2010; Kreiman and Sidtis, 2011; Yanushevskaya et al., 2011; Podesva and Callier, 2015; Park et al., 2016; Yanushevskaya et al., 2016). Changes in vocal fold vibration are also associated with differences in singing registers and with voice disorders (Sundberg, 1987; Titze, 1994; Kempster et al., 2011; Sapienza et al., 2011).

This chapter therefore takes as a starting point that the voice is ultimately used to convey information, which is transmitted between speakers and hearers in stages commonly known as the ‘speech chain’ (Denes and Pinson, 1993). Although there have been many advancements over the years in understanding how the vocal folds are innervated, how they move and vibrate, and how different voice settings are manifested acoustically (for instance, see overviews in Titze, 1994; Baken and Orlikoff, 2000; Stevens, 2000; Gobl and Ní Chasaide, 2010; Hirose, 2010; Story, 2015), it is still unclear how these stages of the speech chain interact with one another to influence voice quality. Yet, ultimately the main goals of the study of the voice should be to answer two fundamental questions: (1) When we perceive a change in the voice, what caused that change? and (2) What are the acoustical and perceptual results of a change in voice production? In my research with colleagues (e.g. Kreiman et al., 2014), we address these questions by modeling how information about the voice is transmitted from speaker to hearer. Ideally, a unified theory of voice production,

acoustics, and perception should be able to model any type of information that the voice can convey, including phonological, prosodic, discourse, sociolinguistic, and paralinguistic information, as well as talker identity. In this chapter, I focus largely on linguistic (especially phonological) meaning: how the voice is used to create sounds of the world's languages. Thus, the first goal of the phonetic study of the voice can be restated with respect to phonological meaning: (1′) When we perceive a *phonologically-relevant* change in the voice (such as a switch between two sounds), what caused that change?

The past decade has seen a rapid increase in research on the role of the voice in sound systems, much of which will be reviewed in this chapter. This research will help outline the primary voice dimensions, defined here in articulatory terms, that are used in languages of the world to convey phonological meaning (Section 2). Readers will see that the study of the voice is essential for our understanding of phonetics and phonology, because *every* sound that we make involves an articulation of the vocal folds, which has specific acoustic attributes that listeners can hear and use in language.

1.1 What is covered in this chapter?

This chapter is largely theoretical in nature. The focus is on how we can model (within a unified framework) the three stages of the speech chain as it concerns the vocal folds and their use in sounds of the world's languages. Therefore, we will not review fine-grained details pertaining to voice anatomy and physiology (though see Titze, 1994; Stevens, 2000; Titze, 2006; Reetz and Jongman, 2008; Hirose, 2010; Kreiman and Sidtis, 2011; Gick et al., 2013; Story, 2015; Zhang, 2016*b*), as well as voice source modeling (Stevens, 2000; Gobl and Ní Chasaide, 2010; Story, 2012; Samlan et al., 2013; Kreiman et al., 2015; Moisik and Esling, 2014; Moisik et al., 2014; Story, 2015). Neither will we review instrumentation used to measure muscular and articulatory properties of the voice; I refer readers to Baken and Orlikoff (2000); Hirose (2010); Gick et al. (2013), among others. Given that this chapter focuses on linguistic meaning, readers who are especially interested in how the voice varies according to specific voice disorders, emotions, individuals, and singing styles should consult Laver (1980); Sundberg (1987); Titze (1994); Esling and Edmondson (2010); Gobl and Ní Chasaide (2010); Kreiman and Sidtis (2011); Sapienza et al. (2011), among others. However, this chapter should also be of use to these readers, in that we will

review a theory of the voice that links vocal production, acoustics, and voice quality perception more generally. Moreover, we will discuss how the speech chain, as it concerns the vocal folds, relates to meaningful categories that must be accurately perceived in order for communication to be effective.

In the following Section 2, I classify vocal fold articulations according to their primary phonological dimensions. In Section 3, I review the psychoacoustic voice model by Kreiman et al. (2014) and how parameters of this model are perceived by listeners and relate to the phonological dimensions outlined in the previous section. In Sections 4 and 5, I discuss recent work showing how the parameters of Kreiman et al. (2014)'s voice model relate to voice source acoustics and vocal fold articulation as parameterized in a recent model by Zhang (2015, 2016*a*). In Section 6, I conclude with discussion of outstanding questions and areas for future research.

2 Primary linguistic voice dimensions

Vocal fold movements, despite being very complex, can be organized along two articulatory dimensions that are especially important in language (Table 1): how far apart the folds are from each other, and whether they are vibrating. (Further details on these articulations are presented in Section 5.) In this chapter, I refer to this first dimension as *vocal fold approximation*, though it could likewise be called 'abduction/adduction' and 'spreading/constriction'. The minimal vocal fold approximation in speech can be found during voiceless aspiration (e.g. for [h] and aspirated stops like [t^h]), and the maximal vocal fold approximation, when the vocal folds are in full contact, can be found for a glottal stop [ʔ] as well as glottalized sounds and ejectives (e.g. [$\widehat{ʔt}$, t']). Note that 'minimal vocal fold approximation' is understood within the context of speech sounds; for example, during active breathing the vocal folds abduct even more than during aspirated speech sounds (Gick et al., 2013). Other voiceless sounds can be described as having incomplete vocal fold approximation that is nonetheless greater than that found for aspiration; this state of vocal fold approximation is sometimes called 'prephonation' (Harris, 1999; Esling and Harris, 2003; Edmondson et al., 2011). Likewise, voiced sounds require some degree of vocal fold approximation, which can also vary in its degree. Thus, all sounds, regardless of whether they are 'plain' voiceless sounds, aspirated, glottalized, or voiced, involve some degree of vocal fold approxima-

tion. A schematic of different states of vocal fold approximation for voiceless sounds is shown in Figure 1.

[Figure 1 about here.]

The second dimension, vocal fold vibration, is also often called *voicing*. This dimension is dependent on vocal fold approximation: voicing can only be initiated with some amount of vocal fold approximation, and only certain degrees of vocal fold approximation can sustain voicing once it has begun (Titze, 1992; Kreiman and Sidtis, 2011). Consequently, all sounds with vocal fold vibration necessarily make use of both dimensions. Voicing is generally treated as being categorical: sounds produced with vocal fold vibration are 'voiced', those without are 'voiceless'. (Voicing is sometimes also called 'phonation', but this term is also used more generally to describe any sound generated in the larynx or the remaining components of the vocal tract.) Voicing is very important in languages' phonologies; voiced sounds, including vowels, sonorant consonants, and voiced obstruents, are found in every spoken language, and about 80% of languages have voicing contrasts in obstruents (Maddieson et al., 2016).

[Table 1 about here.]

Voicing can be further characterized along two dimensions that are very important for linguistic meaning. The first is *rate* (or 'frequency') of vocal fold vibration, a continuous dimension that determines the fundamental frequency of the voice and is used to convey lexical tone and intonation, and is one of the main correlates of lexical and phrasal stress (Gordon and Applebaum, 2010; Gordon, 2014; Garellek and White, 2015). The second dimension to voicing is the *quality* or 'manner' of vocal fold vibration. Voice quality is also important for stress, tone, and intonation (Kreiman, 1982; Sluijter and van Heuven, 1996; Campbell and Beckman, 1997; Garellek and White, 2015; Mooshammer, 2010; Lancia et al., 2016); moreover, it is the primary dimension used for contrastive voice quality ('phonation type') and is important for voice registers, a multidimensional linguistic contrast involving a change in voice quality and other (laryngeal and supralaryngeal) changes (DiCanio, 2009; Brunelle, 2012; Abramson et al., 2015; Brunelle and Kirby, 2016; Tian and Kuang, 2016).

The most common voice qualities that are used in language are 'modal', 'breathy', and 'creaky', and we will focus on these for the remainder of the chapter. They are defined relative to one another; thus, there is no 'absolute' breathy voice, but certain voice qualities are 'breathier' than others. It is partly due to their relative differences that many names for voice qualities exist: breathy voice qualities are sometimes called 'lax, slack, murmured, aspirated' while creaky qualities are 'stiff, tense, laryngealized, glottalized, (vocal) fry, pressed', to name but a few. Other terms, such as 'rough', 'strident', 'sphincteric', 'epiglottalized' or 'harsh' voice, tend to be used for voice qualities that also necessarily involve supraglottal constriction (Laver, 1980; Traill, 1985; Gerratt and Kreiman, 2001; Edmondson and Esling, 2006; Miller, 2007; Moisik and Esling, 2011; Moisik, 2013) and thus fall outside the narrow definition of 'voice' used here.

[Figure 2 about here.]

Although these different terms can sometimes refer to the same articulation, often researchers will use different terms to refer to distinct manners of vocal fold vibrations, acoustic characteristics, or perceptual voice qualities (Batliner et al., 1993; Gerratt and Kreiman, 2001; Redi and Shattuck-Hufnagel, 2001; Slifka, 2006; Kane et al., 2013; Kuang, 2013*b*; Keating et al., 2015). From an articulatory perspective, differences between breathy, modal, and creaky voice can minimally be described using a one-dimensional model of vocal fold approximation (Ladefoged, 1971; Gordon and Ladefoged, 2001): as we have already discussed, voiceless sounds can be made with minimal or maximal vocal fold approximation (as in [h] or [ʔ], respectively). In between these extremes, there is voicing. Voicing with less vocal fold approximation is 'breathy', voicing with more approximation is 'creaky', and voicing that is neither breathy nor creaky is 'modal'. ('Lax/slack' voice is sometimes considered intermediate to breathy and modal, and 'tense/stiff' voice to creaky and modal; Keating et al., 2011; Kuang and Keating, 2014.) This one-dimensional model of vocal fold approximation is schematized in Figure 2. It is conceptually simple and useful for describing the phonologically-relevant relationship between voice qualities and categories of voiceless (consonantal) vocal fold approximation (Lombardi, 1991). However, it suffers from certain drawbacks. First, 'modal' voice is defined in articulatory terms, relative to other states of vocal fold approximation; however, many researchers also use this term to refer to a speaker's default or 'normal' voicing. Defined thus, if a speaker's normal voice quality is quite creaky, that speaker's 'modal'

voice would involve quite a bit of vocal fold approximation, and thus should be captured under 'creaky' in this model. The difference between these two definitions of 'modal' (one articulatory and relative to non-modal voice, the other with reference to a speaker's normal quality) is generally thought to be of little practical importance, because we assume that speakers can always become creakier or breathier than their normal voice quality; for instance, a Jalapa Mazatec speaker whose normal voice quality is very creaky should be able to produce even creakier voice quality than her default in order to say a word with contrastive creaky voice in the language. But given that we still know very little about individual differences in voice quality – for example, if the degree of vocal fold approximation in a particular speaker's default voice quality has an effect on how that same speaker will produce non-modal voice – there may be an important distinction to be made between 'modal' as defined articulatorily vs. speaker-dependently.

Second, although breathy and creaky voice are (generally) produced with different degrees of vocal fold approximation, creaky voice can involve additional supraglottal constriction (e.g. of the ventricular and aryepiglottic folds). The additional constrictions – and how these relate to linguistically-relevant relationships between voice quality and supralaryngeal articulations (like tongue root advancement and vowel quality) – are captured more straightforwardly in a 'valves' model of voice quality (see Table 2, after Esling and Harris, 2005 and Edmondson and Esling, 2006). However, only Valve 1 from this model is involved in voice quality if the 'voice' is narrowly defined as including only articulation of the vocal folds.

[Table 2 about here.]

The third drawback to both continuum and valves models of the voice is that 'creaky' voice – even if it is defined narrowly as a manner of vocal fold vibration with no additional supraglottal articulation – represents a cluster of voice qualities that share some perceptual attributes. In order to be perceived as creaky, the voice must be minimally low in pitch, irregular in pitch, or constricted-sounding (we will discuss what it means to be 'constricted-sounding' in the following section), though not necessarily all three (Keating et al., 2015). For example, there are cases of creaky voice that are irregular and low in pitch, but also unconstricted (Slifka, 2000, 2006), which I call 'unconstricted creaky voice'. Further, there are two distinct types of creaky voice that are constricted in quality but can be regular in pitch: vocal fry (which is low and regular in pitch) and

tense (or 'pressed') voice, which is constricted but with a high and (usually) regular pitch (Keating et al., 2015). Tense voice is often regular in pitch, but the increased constriction could result in sudden and irregular changes in voicing amplitude. Lastly, a very low-pitched voice (e.g., below 60 Hz) can also be perceived as creaky, despite not having any constriction or irregularity (Keating et al., 2015). Figure 3 illustrates the differences between prototypical creaky voice, unconstricted creaky voice, tense voice, and vocal fry according to pitch height, pitch regularity, and constricted quality.

[Figure 3 about here.]

Therefore, either a low pitch, an irregular pitch, or a constricted quality (regardless of pitch) is alone sufficient for listeners to perceive a voice as creaky, even if the articulatory origins and acoustic attributes underlying this percept can differ. However, much more work is needed to determine how perceptually distinct these subtypes can be (cf. Gerratt and Kreiman, 2001; Garellek, 2015), and how linguistically-relevant they are. We have evidence that 'prototypical' creaky voice (low-pitched, irregular in pitch, and constricted) is linguistically important, since it is commonly found for contrastive and allophonic creaky voice (Gordon and Ladefoged, 2001; Esposito, 2012; Garellek and Keating, 2011). Both other types of constricted voices listed in Figure 3 are of linguistic relevance: vocal fry can be used to perceive a glottal stop [ʔ] (Hillenbrand and Houde, 1996; Gerfen and Baker, 2005), and tense voice can be used for the phonetic realization of contrastive creaky voice on a vowel with a high lexical tone, and is found more generally with higher-pitched lexical tones (Garellek and Keating, 2011; Kuang, 2013*b*, 2017*a*). Unconstricted creaky voice is found as a form of phrase-final creak (Kreiman, 1982; Redi and Shattuck-Hufnagel, 2001; Garellek, 2015) in utterance-final position when the subglottal pressure is low, at least for some speakers of American English (Slifka, 2000, 2006). If unconstricted creaky voice occurs due to low subglottal pressure, it should also occur in similar environments in other languages.

Given that creaky voice represents a cluster of vocal fold articulations and voice qualities, one might be inclined to ask whether phoneticians should retain the more general term 'creaky' at all. I believe we should, because it is useful to have a word for the abstract phonological category, which may be realized (in different phonological environments and/or by different speakers) using different articulations. For instance, it is useful to have a 'creaky' phonological category in

Jalapa Mazatec, since this phonation type contrasts with more modal and breathy voice qualities. However, we know that the 'creaky' voice type varies phonetically by lexical tone and preceding consonant in the language (Garellek and Keating, 2011; Kuang, 2013*a*), and likely also by prosodic position, talker, and other factors.

In sum, we can analyze languages' use of voice primarily in terms of vocal fold approximation and voicing, which in turn may differ by rate and quality. There are other dimensions of the voice, including whisper and vocal intensity, but these generally play smaller roles in phonologies of the world's languages (cf. Fulop and Golston, 2008 for the role of whispery voice in White Hmong). Vocal intensity, which is important for stress (Gordon and Applebaum, 2010), is controlled mostly by the subglottal pressure, but also by vocal fold and supralaryngeal adjustments (Sundberg, 1987; Zhang, 2016*b*). Having reviewed the primary dimensions of the voice used to make sounds of the world's languages, we will now turn to a model of how language users produce and perceive these dimensions.

3 A psychoacoustic model of the voice

One of the biggest challenges with modeling the voice is that voice articulation, acoustics, and perception are inherently multidimensional. The vocal folds have a complex structure and vibrate in a three-dimensional space. Acoustically, one can model the voice source in both temporal and spectral domains, and within each domain there are many different attributes to the voice that can be parameterized (Cumming and Clements, 1995; Gobl and Ní Chasaide, 2010; Kreiman, Gerratt and Khan, 2010; Kreiman et al., 2015). Voice perception is also extremely complex; there are dozens of ways of characterizing the voice (Kreiman and Sidtis, 2011), with terminologies and relevant taxonomies varying by discipline. For example, a whispery voice quality might not be a primary voice dimension in sounds of the world's languages, but it is common in speech and is associated with certain speech disorders (Laver, 1980; Sapienza et al., 2011; Gick et al., 2013). And, as discussed in Section 1, models of voice articulation and acoustics may not be able to account for all perceptual changes in voice quality, or may have articulatory and acoustic parameters that are irrelevant for perception (Zhang et al., 2013; Kreiman et al., 2014; Garellek, Samlan, Gerratt and Kreiman, 2016). Since our main goal in studying the voice is to link what speakers do with their voices with what listeners hear, models of the voice should link to voice

perception; acoustic attributes are important only insofar as they are perceptible, and articulatory attributes only insofar as they result in perceptible acoustic changes.

For these reasons, Kreiman et al. (2014) propose a psychoacoustic model of the voice, shown in Table 3, which has the following parameters: four pertaining to the harmonic structure of the voice source spectrum (the sound produced by the vocal folds, before being filtered by the vocal tract), another parameter for modeling the inharmonic component of the source spectrum (i.e., the noise), two temporal components of the voice source (f0 and amplitude), and the vocal tract transfer function (which models the filtering from the vocal tract). Only the first three groups are relevant for the voice as it concerns the vocal folds; though resonances and anti-resonances affect overall voice quality, they are here considered independent of the voice source parameters (cf. Cumming and Clements, 1995). This model makes several other important assumptions, notably that its parameters are both necessary and sufficient to model voice quality; thus, with one of these parameters missing, voice quality cannot be faithfully modeled, and no other measures are needed to model the voice.

[Table 3 about here.]

The temporal parameters (the f0 and amplitude tracks) relate to the presence of voicing, its rate of vibration, and its amplitude. The harmonic spectral slope parameters represent differences in harmonic amplitudes; thus, H1–H2 refers to the difference in amplitude between the first and second harmonic (Bickley, 1982), H2–H4 is the difference in amplitude between the second and fourth harmonics (Kreiman et al., 2007), H4–H2 kHz is the difference in amplitude between the fourth harmonic and the harmonic closest to 2000 Hz, and H2 kHz–H5 kHz is the difference in amplitude between the harmonic closest to 2000 Hz and the one closest to 5000 Hz (Kreiman et al., 2011). Together, they characterize the voice's 'spectral tilt' in various harmonic and frequency bands.

The assumptions regarding the harmonic spectral slope model and its parameters warrant further discussion. The first main assumption is that the spectral tilt parameters can be articulatorily and perceptually independent of f0, even though it is clear that f0 changes are associated with changes in spectral tilt (Kuang, 2013*a*; Garellek, Samlan, Gerratt and Kreiman, 2016; Kuang, 2017*a*). Thus, with a constant f0 = 100 Hz, a measure like H2–H4 will have a frequency bandwidth

of 200 Hz (between 200 Hz and 400 Hz), whereas with a constant f0 of 200 Hz, the frequency bandwidth of H2–H4 will be 400 Hz (between 400 and 800 Hz). Frequency bandwidth can vary for H1–H2 and H2–H4, as well as well as H4–H2 kHz, though the latter parameter is bounded at the high end by a particular frequency of 2000 Hz. (The source spectrum model is therefore used only for f0 values below 500 Hz; higher than that, and H4 will be equal to or surpass 2 kHz!) The final parameter H2 kHz–H5 kHz depends on frequency bandwidth alone, not harmonic bandwidth. Overall then, we assume in this model that voice quality is perceptually dependent on spectral tilt between fixed *harmonics* (regardless of their frequency) in the lower-frequency end of the spectrum, and that voice quality is perceptually dependent on spectral tilt between harmonics of fixed *frequencies* at the higher-frequency end of the spectrum. It should be noted here that spectral tilt measures over fixed frequency bands have also been shown to correlate with changes in quality (de Krom, 1995; Hartl et al., 2003; Samlan et al., 2013; Samlan and Kreiman, 2014).

The second main assumption is that we need four spectral tilt parameters. Kreiman et al. (2014) and Garellek, Samlan, Gerratt and Kreiman (2016) motivate a model of the harmonic source spectrum with four tilt parameters for several reasons. First, spectral tilt can vary independently according to frequency band; thus, spectral tilt can be negative in one portion of the spectrum, but positive in another (Kreiman et al., 2007; Garellek, Samlan, Gerratt and Kreiman, 2016). Another reason for modeling spectral tilt in terms of multiple components is that distinct articulations might be responsible for different tilt components (Zhang et al., 2013). Moreover, source spectra share certain inflection points; for example, it is common for the harmonic slope to change abruptly around 2000 Hz. Finally, listeners are not equally sensitive to every component of spectral tilt (Garellek, Samlan, Gerratt and Kreiman, 2016), and indeed sometimes the slope of one component can cancel out the perceptual effect of another (Garellek et al., 2013). We will also discuss the relevance of these parameters for linguistically-relevant dimensions of the voice in the following section.

This model further assumes that it is the slope between harmonics (rather than amplitudes of individual harmonics), that is perceptually relevant for quality (Kreiman et al., 2014; Garellek, Samlan, Gerratt and Kreiman, 2016). For instance, we assume that it is of no importance whether H3 is louder than either H2 or H4; the parameter H2–H4 depends only on the amplitudes of the harmonics adjacent to H3. In our use of this model, we therefore alter the amplitude of the inter-

mediate harmonics to conform to the slope made between the two harmonic endpoints, as shown in Figure 4. Note also that the harmonic spectral slope model is based on our analysis of adult voices, and therefore it is unclear whether children's voices could and should be modeled similarly.

The final parameter of the model, the harmonics-to-noise ratio (HNR), refers to the difference in amplitude between the harmonic and inharmonic components of the source spectrum, as measured in the cepstral domain (de Krom, 1993). The choice of HNR as a means of measuring noise in this model assumes that spectral noise is psychoacoustically important (Kreiman and Gerratt, 2005; Shrivastav and Sapienza, 2006; Zhang et al., 2013; Garellek, Samlan, Gerratt and Kreiman, 2016). Although there are many time-domain noise measures like jitter and shimmer, these are not included because they are not perceptually relevant independently of HNR (Kreiman and Gerratt, 2005). The current version of the model also assumes that HNR over the entire frequency range is sufficient to model the inharmonic component of the source spectrum. However, it is clear that noise interacts with harmonic components in different frequency bands in distinct ways (Kreiman and Gerratt, 2012; Garellek, Samlan, Gerratt and Kreiman, 2016), and that changes in linguistic uses of voice quality are expressed with more narrowband noise measures (Garellek, 2012). Therefore, much more work is needed to determine how best to model noise and its role in causing changes in voice quality.[3]

[Figure 4 about here.]

A note on how to obtain measurements for these model parameters: Kreiman et al. (2014) and Garellek, Samlan, Gerratt and Kreiman (2016) calculate these model parameters using inverse filtering of audio recordings followed by analysis-by-synthesis. Although this allows for very accurate source and filter estimation, the main disadvantages of this process are that it is time-consuming and difficult to do on conversational speech (see Gobl and Ní Chasaide, 2010; Kreiman, Antoñanzas-Barroso and Gerratt, 2010 for more discussion on the matter). However, these model parameters can be estimated from the audio signal if formant correction is used, e.g. with a program like VoiceSauce (Shue et al., 2011). This practice is often used in phonetic studies, but requires accurate formant measurements, which can be problematic with certain sounds. This is discussed in more detail in Section 4.2.2.

In the following sections, I will show how the primary linguistic dimensions of the voice can be expressed and measured using this psychoacoustic model, and how vocal fold articulation as it concerns these dimensions relates back to the model's parameters.

4 Acoustic properties of the primary phonological voice dimensions

The primary phonological voice dimensions (vocal fold approximation, voicing, rate of vibration, and quality of voicing) have clear acoustic ramifications that can be measured using the parameters of the psychoacoustic voice model outlined in Section 3.

4.1 Vocal fold approximation

Vocal fold approximation during voiceless sounds has both direct and indirect acoustic consequences. Direct consequences include the absence of periodic energy, and, for minimal vocal fold approximation (i.e., maximal abduction/spreading), the presence of aspiration noise, which in Kreiman et al. (2014)'s model can be measured using HNR (higher values of the measure indicate less noise). Aspiration noise can also be seen in spectra and spectrograms as broadband noise (though filtered by the shape of the vocal tract); in waveforms, aspiration noise is hard to distinguish from other sources of noise, such as frication from the vocal tract. Indirect consequences of vocal fold spreading can be seen on adjacent voiced sounds, which will be breathier, as discussed in Section 4.2.2.

Complete vocal fold approximation, on the other hand, is hard to measure using parameters in Kreiman et al. (2014)'s model, and is hard to see directly in either temporal or spectral domains; if the vocal folds are constricted but not vibrating, then there is no acoustic energy that is produced during the constriction. However, as vocal folds transition from voicing to and from complete constriction, there are indirect consequences present in form of creaky voice on the adjacent sounds. In sum, vocal fold approximation can be measured indirectly through its effects on adjacent voicing, which we will discuss more below.

4.2 Voicing

Vocal fold vibration usually produces a complex (quasi-)periodic wave; that is, a wave with many frequency components (Stevens, 2000). The slowest component is the fundamental frequency (f0),

and the faster components are whole-integer multiples of the f0 (see Figure 4). Thus, if voicing is present, there must be an f0. Figure 5 illustrates how an f0 track in the sequence [àhá] aligns with the presence of glottal pulses (the vertical striations) in the waveform and spectrogram. The intensity of voicing (roughly, 'how loud' the voice is) can be measured by the psychoacoustic model using the amplitude track, also shown in Figure 5.

[Figure 5 about here.]

4.2.1 Rate of vibration

Differences in rate (or frequency) of vocal fold vibrations are reflected in the acoustic signal primarily through changes to the fundamental frequency and its change over time. In the temporal domain, this means that the periodic wave produced during voicing will recur more quickly or slowly, resulting in more closely- or distantly-spaced glottal pulses. This can be seen in Figure 5: where the glottal pulses are closer together (i.e. in the second vowel), the f0 track is higher. In the spectral domain, f0 corresponds to the frequency of the first harmonic, though here temporal change in f0 is not calculable. Using Kreiman et al. (2014)'s model, we can measure f0 change using the f0 track.

4.2.2 Voice quality

As discussed above, the three main voice qualities (breathy, modal, and creaky voice), as well as subtypes of creaky voice, can be described in terms of their degree of constriction (or spreading) and noise: relative to modal voice, breathy voice is both more spread and noisier (because of the presence of aspiration noise), whereas prototypical creaky voice is more constricted and noisier because of irregular pitch.

These two basic dimensions to voice quality – spreading/constriction and noise – have specific acoustic attributes in the psychoacoustic model proposed by Kreiman et al. (2014). The most reliable correlate of increased spreading or constriction during voicing is through increased spectral tilt for spreading and decreased spectral tilt for constriction (Klatt and Klatt, 1990; Gordon and Ladefoged, 2001; Hanson et al., 2001; Kreiman et al., 2012; Chen et al., 2013; Samlan et al., 2013; Keating et al., 2015; Zhang, 2016*a*). For the four spectral tilt parameters discussed earlier (H1–H2, H2–H4, H4–H2 kHz, and H2 kHz–H5 kHz), greater spectral tilt due to breathy voice would be

indicated by higher values of these measures. Of these four spectral slopes, it is clear that at least H1–H2 and H2–H4 are relevant phonologically. For example, native speakers of White Hmong can use changes in either of these slopes to perceive their contrastive breathy-voiced tone (Garellek et al., 2013), and many other studies on a variety of languages have shown that breathy and creaky voice correlate with changes in H1–H2 (Bickley, 1982; Gordon and Ladefoged, 2001; Blankenship, 2002; Miller, 2007; DiCanio, 2009; Brunelle and Finkeldey, 2011; Garellek and Keating, 2011; Garellek, 2012; Esposito, 2012; Khan, 2012; Berkson, 2013; DiCanio, 2014; Yu and Lam, 2014; Abramson et al., 2015; Misnadin et al., 2015; Zhang and Yan, 2015; Tian and Kuang, 2016). Listeners of languages with contrastive or allophonic non-modal phonation show differing degrees of sensitivity to H1–H2 (Kreiman, Gerratt and Khan, 2010; Kreiman and Gerratt, 2010), and H2–H4 is also relevant for listeners' identification of speaker sex (Bishop and Keating, 2012). On the other hand, it is still unclear whether the higher-frequency slopes H4–H2 kHz and H2 kHz–H5 kHz contribute to linguistically-relevant voice distinctions, though recent work suggests that they, along with H1–H2 and H2–H4, help differentiate creaky vowels from non-creaky ones in American English (Garellek and Seyfarth, 2016).

Other measures of spectral tilt over different frequency bands, such as H1–A1, H1–A2, and H1–A3 (differences in amplitude between the first harmonic and the harmonic closest to the first, second, and third formants) have proved useful for distinguishing voice qualities or registers in languages of the world (Ní Chasaide and Gobl, 1993; Hanson, 1997; Gordon and Ladefoged, 2001; Hanson et al., 2001; Wayland and Jongman, 2003; Gerfen and Baker, 2005; Andruski, 2006; DiCanio, 2009; Esposito, 2010*a*,*b*; Avelino, 2010; Garellek and Keating, 2011; Brunelle, 2012; Berkson, 2013; Kirby, 2014; Abramson et al., 2015; Tian and Kuang, 2016). There are several important remarks to make on the similarities and differences between these measures and the spectral tilt parameters proposed by Kreiman et al. (2014). First, both types of measures use harmonic-based bandwidths, rather than fixed frequency ranges (see discussion in Section 3). Moreover, the harmonics closest to a formant (i.e., A1, A2, and A3) are not defined in a source model. One consequence of this is that measures like H1–A1, H1–A2, and H1–A3 are correlated and can even overlap with the source spectral tilt measures like H1–H2. For instance, the vowel [i] spoken by the average adult male speaker of American English has an F1 at around 340 Hz (Hillenbrand et al., 1995). So if an adult male speaker of American English says [i] with an f0 of 150 Hz, the harmonic

closest to F1 is H2, which means that, for this token, H1–H2 would be equal to H1–A1 (see the left panel of Figure 6). Thus, depending on context, it might not make sense to measure both H1–H2 and H1–A1 for the same token. However, this issue would not occur when using source spectral parameters like the kind found in Kreiman et al. (2014)'s model, because the model parameters can never overlap in frequency. Of more theoretical relevance is the fact that, for a measure like H1–A1, what counts as 'A1' will vary by context. For instance, if the same adult male speaker from the previous example were to say the vowel [ɑ] with an f0 = 150 Hz, then H1–A1 would be equal to the difference in amplitude between the first harmonic and the harmonic closest to roughly 770 Hz, which is H5 (at roughly 750 Hz, see the right panel of Figure 6). Measures like H1–A1, H1–A2, and H1–A3 vary in their harmonic bandwidth depending on the formant frequencies; on the other hand, *source* spectral tilt measures like H1–H2 and H2–H4 have fixed harmonic bandwidths and therefore do not depend on vowel quality. Thus, use of measures like H1–A1 assumes that the ways in which spectral tilt can determine voice quality necessarily depend on the vowel quality. For instance, if I compare Mazatec breathy vs. modal vowels of different qualities using H1–A1 (as in Garellek and Keating 2011), then I assume that a relevant way of distinguishing these voice qualities is by comparing the spectral tilt between H1 and the first formant, regardless of the harmonic that is most affected by that formant. On the other hand, using measures such as H1–H2, H2–H4, H4–H2 kHz, and H2 kHz–H5 kHz implies that voice source characteristics are relevant independent of the filter; although Kreiman et al. (2014) assume that both the source and filter can influence voice quality, they further assume that measuring spectral tilt within fixed harmonic or frequency bands is a relevant way of distinguishing voice quality. It is still unclear which method of representing spectral tilt more closely reflects perception of voice quality; i.e., whether listeners perceive changes in quality more as a function of formant- or harmonic-based differences in spectral tilt.

[Figure 6 about here.]

The second important dimension to voice quality is noise, which in the psychoacoustic voice model is calculated in the spectral domain as harmonics-to-noise ratio (HNR). Since breathy voice has aspiration noise (due to the increase in vocal fold spreading or lower vocal fold thickness, see Section 5), this will lower the HNR. Numerous studies have shown that HNR measures are

useful for distinguishing breathy vs. non-breathy voice qualities used in language (Gordon and Ladefoged, 2001; Blankenship, 2002; Brunelle, 2012; Berkson, 2013; Esposito, 2012; Garellek, 2012; Khan, 2012; Kuang, 2012; Simpson, 2012; Tian and Kuang, 2016). Creaky voice also lowers HNR, but usually this is because of its irregular pitch: if the f0 is not regular, the signal's noise will increase. Various studies also provide evidence for the use of HNR in distinguish creaky vs. non-creaky voice qualities in language (Blankenship, 2002; Esposito, 2012; Garellek, 2012, 2015; Keating et al., 2015; Garellek and Seyfarth, 2016). 'Harsh' voice qualities, which are not reviewed in detail here because they necessarily involve supraglottal constriction (Edmondson et al., 2001; Edmondson and Esling, 2006), can also have lower HNR (Miller, 2007), due at least in part to the supralaryngeal noise.

Phoneticians often use acoustic measures to categorize different voice qualities. For example, the lower values of H1–H2 in Category A compared with Category B might be used to justify an analysis in which Category A is creaky and Category B is modal. There are, however, two important things to keep in mind with regard to this practice. First, spectral tilt measures like H1–H2 vary continuously between more constricted creaky voice qualities (which have lower H1–H2) and less constricted or breathier voice qualities (which have higher H1–H2); modal voice's spectral tilt is somewhere in between that of constricted creaky and breathy voice. But raw values of spectral tilt measures do not index a precise voice quality; one person's creaky voice can have an average H1–H2 of -2 dB while another person's creaky voice averages 5 dB. Thus, if Category A has a higher H1–H2 than Category B, we cannot know whether the difference between A and B is one between more modal vs. creaky voice, between more breathy vs. modal voice, or between more breathy vs. more creaky voice. This is why spectral tilt measures are often interpreted with respect to noise measures like HNR (Blankenship, 2002; Garellek, 2012; Simpson, 2012; Garellek and White, 2015): if Category A has both a higher H1–H2 and a higher HNR than Category B, then we can assume A is more modal than B, because modal voice generally has higher H1–H2 and HNR values than creaky voice. But if Category A′ has a higher H1–H2 and a *lower* HNR than Category B, then we can assume A′ is breathier than B′, because breathy voice generally has higher H1–H2 and lower HNR values than modal voice (see Table 4). Figure 7 illustrates the relationship between breathy, modal, and (prototypical) creaky voice in a two-dimensional acoustic space consisting of spectral tilt and HNR.

[Table 4 about here.]

[Figure 7 about here.]

The second caveat pertains to the fact that lower spectral tilt measures correlate with increased constriction. As we have discussed earlier, not all subtypes of creaky voice are more constricted or noisy than modal voice. For instance, tense voice is constricted, high in pitch, and regular in pitch, which means that we would not expect to find a decrease in HNR for tense voice relative to modal voice. The acoustic characteristics (in terms of both spectral tilt and noise) of these types of creaky voice are shown in Table 5.

[Table 5 about here.]

Another difficulty with measuring voice quality is that when speech sounds travel through air, they bear evidence both of voice characteristics and of the supralaryngeal modulations of the vocal tract. It can sometimes be challenging to infer whether a particular acoustic property is due to voice or supralaryngeal vocal tract articulations: for instance, nasalized vowels (caused by velum lowering) and breathy vowels (caused by vocal fold spreading during voicing) both have higher spectral tilt (Klatt and Klatt, 1990; Simpson, 2012; Garellek, Ritchart and Kuang, 2016). There are several options for researchers who wish to disentangle confounding effects of manner of voicing and supralaryngeal articulation. First, one can use 'inverse filtering' to remove the acoustic effects of other articulators, and measure manner of voicing from the source waveform (often with the additional step of modeling the waveform using a variety of voice source models, Ní Chasaide and Gobl, 1993; Epstein, 2002; Gobl and Ní Chasaide, 2010; Kreiman et al., 2015). Another option is to measure spectral tilt and noise from the audio output spectrum, but to 'correct for' or undo the effects of the supralaryngeal articulators (Hanson, 1995, 1997; Iseli et al., 2007). However, it should be noted that formant corrections are meant to undo the effects of vowel formants specifically; thus, they cannot remove the effects of nasalization or consonantal resonances (Simpson, 2012; Garellek, Ritchart and Kuang, 2016). Of course, formant corrections will also fail whenever the formants are mistracked. This is especially common with high-pitched and breathy voices where the high-frequency and high-energy of the first harmonic can be misidentified as a formant. It is thus recommended that researchers check the f0 and formant frequencies to ensure that the

spectral slope measures are not mistracked due to inaccurate tracking of f0 or (for 'corrected' measures) of the formants . Spectral slope measures whose harmonic amplitudes have been corrected for effects of formants and bandwidths are usually denoted with asterisks, e.g. H1*–H2*. Readers should therefore be aware that different researchers will use a label such as 'H1–H2' to refer either to uncorrected spectral tilt derived from the output *audio* spectrum, or to the measure calculated directly from the *source* spectrum.

As mentioned at the start of this section, it is impossible to measure voice quality during a voiceless sound, which makes use of vocal fold approximation as its primary voice dimension. However, differences in vocal fold approximation can have ramifications for voice quality on an adjacent voiced sound. Numerous linguistic studies have shown that vowels adjacent to aspirated sounds are breathier than vowels adjacent to non-aspirated ones; conversely, vowels adjacent to glottalized sounds (including ejectives) tend to be creaky compared to vowels adjacent to non-glottalized ones (Löfqvist and McGowan, 1992; Ní Chasaide and Gobl, 1993; Blankenship, 2002; Vicenik, 2010; DiCanio, 2012; Esposito and Khan, 2012; Garellek, 2012; Gallagher, 2015; Misnadin et al., 2015; Garellek and Seyfarth, 2016). This relationship between voiceless consonants and voice quality during vowels is captured easily in Ladefoged's continuum model, because both voiceless vocal fold spreading/constriction and voice quality are modeled along a single dimension (analogous to our 'vocal fold approximation' dimension, see Figure 2). For instance, in the sequence [ah], the vocal folds will have to transition from modal voicing to minimal vocal fold approximation, necessarily 'passing through' breathy voice. This transition can be seen during the first vowel of [àhá] from Figure 5. Using VoiceSauce (Shue et al., 2011) to measure the four spectral tilt parameters from the audio spectra (but correcting for vowel formants), as well as harmonics-to-noise ratio (HNR), it is clear that the vowel is breathier in the second half, nearest the [h] (see Figure 8): H1*–H2*, H2*–H4*, and H4*–H2 kHz* are higher, whereas HNR is lower (indicating greater noise). On the other hand, H2 kHz*–H5 kHz* is *lower* in the second half, which is likely due to interactions with high-frequency noise (Kreiman and Gerratt, 2012; Garellek, Samlan, Gerratt and Kreiman, 2016).

[Figure 8 about here.]

Therefore, we can use various parameters of the psychoacoustic model described in Section 3 to measure acoustic changes in the voice associated with vocal fold approximation, voicing, and its rate and quality. The crucial parameters are f0 (its presence vs. absence, and its value when present) as well as the source spectral tilt parameters and HNR. The formant parameters of Kreiman et al. (2014)'s model are also important for correcting for spectral tilt parameters like H1*–H2*, which are measured from the output audio spectrum, and when measuring a parameter like H1–A1, which makes reference to formants. In the next section, we show how a model of voice articulation can account for changes in these parameters.

5 Voice production

Voice production is remarkably complex and depends on lung pressure and several muscles that, working together, alter the shape and stiffness of the vocal folds, the distance between them and other laryngeal structures, and the position of the larynx as a whole. There currently exist numerous excellent sources on laryngeal anatomy and physiology as they pertain to speech (e.g., Titze, 1994; Stevens, 2000; Reetz and Jongman, 2008; Hirose, 2010; Kreiman and Sidtis, 2011 and Gick et al., 2013), but in this section we will focus on the articulations of the vocal folds that are associated with the primary dimensions of the voice that are used in language (vocal fold approximation and voicing, the latter of which can be further characterized by its rate and manner), and how these relate back to the psychoacoustic voice model discussed in Section 3.

Measuring vocal fold articulation is fraught with challenges. The first main challenge is that the vocal fold dynamics are multidimensional, which makes it hard to determine what aspects of vocal fold articulation we should be measuring in the first place. For instance, many researchers have noted (via scoping) that individuals with voice disorders have asymmetric vocal fold vibration. Crucially though, asymmetric vibration is also very common in individuals with no voice disorders (Bonilha et al., 2012), and not all vocal fold asymmetries produce relevant changes in voice quality (Zhang et al., 2013; Samlan et al., 2014). The second main challenge is methodological: it is very hard to see and thus measure the vocal folds. Direct observation (e.g. via laryngoscopy) is invasive and limits the types of sounds speakers can make while being scoped and the types of sounds we can observe during imaging. And while direct observation of the vocal

folds provides extremely important information about vocal kinematics, it also only allows for a two-dimensional bird's-eye view of the superior part of the vocal folds.

Because of these challenges, there is a longstanding tradition (e.g., van den Berg and Tan, 1959) of using physical and, more recently, computational models of vocal fold articulation, which enable researchers to lower the degrees of freedom and determine how individual parameters affect vocal production, acoustics, and quality (Flanagan, 1972; for a recent overview, see Zhang, 2016*b*). In the remainder of this section, I will outline one such model, and show how its parameters can be used to understand how vocal articulation leads to linguistically-relevant changes in voice quality.

5.1 Modeling voice articulation

Although the vocal folds are anatomically and physiologically complex, we can describe their linguistically-relevant dimensions using a simplified model (Figure 9, after Zhang (2015, 2016*a*, 2017)). This three-dimensional model of the vocal folds has been used to simulate computationally the effects of various vocal fold parameters on voicing. Although there are many other models of vocal fold vibration (Isogai et al., 1988; Titze et al., 1995; Titze, 2006; Samlan and Story, 2011; Story, 2012; see also a recent overview in Zhang, 2016*b*), Zhang's model systematically relates model parameters to acoustic ones, some of which crucially appear in the psychoacoustic voice model described earlier. Therefore, it is particularly useful for assessing the cause-and-effect relationship between voice articulation, acoustics, and perception.

The relevant parameters are vocal fold stiffness in the front-back dimension (represented by the oblique arrow in Figure 9), medial surface thickness in the vertical direction, the angle between the vocal folds (the horizontal arrow), and subglottal pressure. This model makes several assumptions (which are described in detail in Zhang 2015, 2016*a*, 2017), and at present does not include interactions between the subglottal and supraglottal tracts. And though the vocal folds have multiple layers (Hirose, 1997; Kreiman and Sidtis, 2011; Gick et al., 2013; Zhang, 2016*b*) and are often simplified as a two-layered structure composed of a body and a cover (Hirano and Katika, 1985), Zhang (2016*a*) models the folds as a one-layer structure because the vocal fold body and cover rarely differ in stiffness (see also discussion for a two-layered version of the model in Zhang 2017). Moreover, Zhang (2016*a*) models only front-back stiffness because different degrees of muscular activation had strong effects on this dimension but much smaller effects on transverse

stiffness (Yin and Zhang, 2013). Zhang (2017) also models transverse stiffness (set to identical values for body and cover layers), but its effects are mostly limited to phonation pressure threshold, where increasing transverse thickness results in an increase in the phonation pressure threshold, especially when the glottal width is large.

[Figure 9 about here.]

All of these parameters can vary in degree with changes in activation of laryngeal muscles as well as respiration. In the remainder of this section, I will review how these model parameters have been shown to, or might eventually be shown to, produce the primary linguistic dimensions of the voice that are used in sounds of the world's languages. A summary is shown in Table 6.

5.2 Vocal fold approximation

Zhang (2015, 2016*a*) models vocal fold articulation during voicing, and not voiceless vocal fold approximation used for aspirated and glottalized sounds. Nonetheless, we know that vocal fold approximation can be described by a continuum of glottal width, following the continuum and valves models described in Section 2. This dimension is indeed parameterized in Zhang (2015, 2016*a*)'s model as the angle between the two folds, or 'glottal width' (the green arrow in Figure 9). This should therefore be the primary parameter responsible for changes in vocal fold approximation that are used to make voiceless aspirated and glottalized consonants.

As we reviewed in Section 2, other voiceless sounds can be described as being in a 'prephonation' state, during which the vocal folds are close together but not completely adducted, nor spread enough to produce much aspiration noise (Harris, 1999; Esling and Harris, 2003; Edmondson et al., 2011). This too can be modeled articulatorily using the glottal width parameter in Zhang (2015, 2016*a*)'s model. Yet we also know that voiceless unaspirated (and voiceless aspirated) stops are often followed by a rise in f0 (Hombert et al., 1979). This may imply that the mechanisms involved in f0 control (Section 5.3.1) can also be involved in the production of voiceless stops. On the other hand, increased activation of the cricothyroid muscle (which results in increased vocal fold stiffness) is not clearly associated with production of voiceless stops (Hirose and Gay, 1972; Hombert et al., 1979); thus, Hombert et al. (1979) speculate that raising of the whole larynx, which indirectly affects vocal fold stiffness and thus f0, can be responsible for this effect (see also Honda et al., 1999; Stevens, 2000; Brunelle, 2010).

5.3 Voicing

The mechanism of voicing is usually characterized by the myoelastic aerodynamic theory (van den Berg, 1958) and its more recent extensions (Titze, 2006). According to this theory, the combination of tissue elasticity (e.g. altering the stiffness of the intrinsic laryngeal muscles) and aerodynamic forces is responsible for initiating, sustaining, and ending the vibration of the folds. Vibration usually cannot start until after the vocal folds are brought together or nearly so, in order to build up a subglottal pressure (3-5 cm H_20 for modal voice, Titze, 1992). In his model simulations, Zhang (2016*a*) found that the most important parameter to voicing onset is the angle between the vocal folds: the greater the angle between the folds, the higher the pressure must be to initiate voicing (see also Titze, 1992). Vocal fold thickness also matters: the thinner the folds, the higher the pressure needed to initiate voicing, which Zhang (2016*a*) attributes to the difficulty of very thin vocal folds to maintain a degree of glottal opening that is conducive to voicing against the subglottal pressure. When the vocal folds are thin, their prephonatory glottal opening is much larger than the resting glottal opening, which makes it difficult to initiate voicing. But as vocal fold thickness increases, the resting glottal opening is easier to maintain. This enables voicing to begin at lower pressure, unless the vocal folds are so thick that additional pressure is needed to keep them from remaining closed.

The findings regarding voicing initiation pressure are also important for sounds of language, because at edges of utterances (when the subglottal pressure is low), voicing is accompanied by specific vocal fold changes. For instance, stressed word-initial vowels in English and other languages are often glottalized (as in saying the word 'after' with a glottal stop [ʔæftɚ]), especially phrase- and utterance-initially and when stressed (Nakatani and Dukes, 1977; Umeda, 1978; Pierrehumbert and Talkin, 1992; Dilley et al., 1996; Davidson and Erker, 2014; Garellek, 2013, 2014). Since glottalization involves vocal fold approximation, it would be parameterized with a smaller angle between the vocal folds in this model. The smaller angle between the vocal folds is also associated with lower voicing initiation pressure, which would be beneficial for utterance-initial stressed vowels; these must be strongly voiced (to mark stress) despite the low subglottal pressure (Garellek, 2014).

5.3.1 Rate of vibration

Consistent with earlier work (e.g. Stevens, 2000), Zhang (2016*a*, 2017) found that rate of vibration depends on three parameters; f0 increases with greater vocal fold stiffness, subglottal pressure, and vocal fold approximation. The role of increased vocal fold approximation in achieving higher f0 is interesting for linguistic tonal patterns, because sounds with increased vocal fold adduction are often accompanied by higher f0 (e.g., Korean fortis stops), and high tones are often accompanied by increased constriction (Kingston, 2005; Kuang, 2013*b*).

However, the articulatory parameters interact with one another in complicated ways that affect f0 (and voice quality, which we discuss in the following section), because in human voices the vocal fold stiffness will covary with the other model parameters. For example, Zhang (2016*a*) finds that the f0 can be raised by decreasing the glottal width (with or without an increase in the subglottal pressure), but that this is likely to come with greater vocal fold thickness and *low* front-to-back stiffness. The acoustic result is a higher f0 with more constriction and decreased spectral tilt – essentially, the characteristics of tense voice. On the other hand, just raising the vocal fold stiffness parameter will likely be accompanied by decreased vocal fold thickness, which Zhang (2016*a*, p.1506) says will be falsetto-like in quality. Falsetto voice is used as a singing register (Sundberg, 1987), but can also be used in language to index various emotions and types of sociolinguistic meaning (Callier, 2013; Stross, 2013; Zimman, 2013; Podesva and Callier, 2015; Starr, 2015). It may also be used for phonological purposes, e.g. as the phonetic implementation of very high-pitched tones such as the high level tone in Black Miao and the falsetto tones of Hubei Chinese (Kuang, 2013*b*; Wang and Tang, 2012; Wang, 2015).

5.3.2 Voice quality

In Zhang (2016*a*)'s model, changes in voice quality are driven by vocal fold approximation, subglottal pressure, vocal fold thickness, and their interactions. Not surprisingly, lower vocal fold approximation (i.e., a greater angle of glottal width) is associated with higher noise (as measured by HNR); within limits, the more the vocal folds are spread, the more turbulent airflow is generated at the glottis. Lower subglottal pressure is also associated with higher noise, because voicing is weaker and less regular in this condition; not surprisingly, languages often have breathy and/or irregular creaky voicing at utterance edges, where the subglottal pressure is low (Rodgers, 1999;

Ogden, 2001; Redi and Shattuck-Hufnagel, 2001; Slifka, 2003, 2006; Esling and Edmondson, 2010; Garellek, 2014, 2015; Di Napoli, 2015; Garellek and Seyfarth, 2016; Kuang, 2017*b*).

Interestingly, increasing vocal fold thickness has a strong influence on spectral tilt. Zhang (2016*a*) measures tilt using H1–H2, as well as larger frequency bands with reference to H1 (H1–H4, H1–H2 kHz, and H1–H5 kHz); aside from H1–H2, the other bands are not equivalent to the parameters from Kreiman et al. (2014)'s psychoacoustic model, though his results should be comparable to some extent. In his simulations, increasing vocal fold thickness results in lower spectral tilt in all frequency bands. Therefore, we would expect measures like H2–H4, H4–H2 kHz, and H2 kHz–H5 kHz to also be lower with increasing thickness. The combination of thick vocal folds with tight approximation, and very low stiffness and subglottal pressure, produces a voice quality that Zhang (2016*a*) describes as vocal fry-like, with a low f0, spectral tilt, and noise.

Of the articulatory model's four parameters, only vocal fold thickness had a sizable effect on spectral tilt measures. This has implications for our understanding of linguistic voice quality, because breathy, modal, and creaky voices are usually analyzed and modeled in terms of glottal width (see the continuum and valves models in Section 2). On the other hand, Zhang (2016*a*)'s results imply that vocal fold thickness should matter more than width in producing changes in voice quality associated with changes in spectral tilt. But because glottal width influences f0, this parameter might be especially relevant for voice quality changes associated with specific lexical tones (Kingston, 2005; Kuang, 2013*b*). And even though vocal fold thickness might contribute more to changes in spectral tilt in model simulations, in human voices this parameter is likely to covary with others in ways which we have yet to fully understand.

[Table 6 about here.]

6 Summary of chapter and future work

This chapter provides an overview of the phonetics of voice, as defined narrowly by the activity of the vocal folds. I took as starting points an assumption and descriptive fact regarding the voice: first, I assume that the study of voice should be driven by what humans can *hear* (rather than what we can *do*); and second, that the multidimensionality of voice production, acoustics, and perception necessitates a unified model of the voice that is driven by what we can hear.

In this chapter I narrowed considerably the discussion of 'what we can hear' by focusing exclusively on linguistic properties of the voice, especially those that we know matter for languages' sound systems: vocal fold approximation, voicing, rate of vibration, and quality of voicing. A combination of these four dimensions can be found in all sounds of language. Using the psychoacoustic model of the voice developed in Kreiman et al. 2014 and further (regarding the source spectrum) in Garellek, Samlan, Gerratt and Kreiman, 2016, I showed how these linguistically-relevant vocal dimensions can be modeled and measured. Recent articulatory modeling, such as that proposed by Zhang (2015, 2016*a*), also bring us closer to understanding the links between voice production, acoustics, and voice quality perception. By demonstrating how different aspects of vocal fold dynamics condition changes in the acoustic signal, articulatory and psychoacoustic models also enable us to better understand sound changes involving the different voice dimensions of language (Kirby, 2013, 2014; Kuang and Liberman, 2015; Ratliff, 2015; Brunelle and Kirby, 2016).

The articulatory and psychoacoustic models of the voice reviewed here are continuously being refined and improved; as I mention at various points in this chapter, many of their assumptions, and the predictions they make, have yet to be confirmed. Moreover, we are still far from knowing the relevant articulatory and psychoacoustic properties of all linguistically-relevant aspects of the voice. For example, Keating et al. (2015) discuss other subtypes of creaky voice not reviewed here, and modeling the temporal characteristics of voice quality is still in its early stages, though temporal properties of the voice are extremely important in language (Nellis and Hollenbach, 1980; Silverman, 2003; DiCanio, 2009; Brunelle et al., 2010; Esposito and Khan, 2012; Garellek, 2012; Remijsen, 2013; Garellek and Seyfarth, 2016; Yu, 2017). Much further work is also needed to determine how to integrate other laryngeal and supralaryngeal structures with these models, since it is clear that the phonetics of the voice and its linguistic patterns cannot be separated completely from the rest of the vocal tract, or even the rest of the larynx (Edmondson and Esling, 2006; Moisik and Esling, 2011; Kuang, 2011; Brunelle, 2012; Story, 2012; Moisik and Esling, 2014; Samlan and Kreiman, 2014; Brunelle and Kirby, 2016; Garellek, Ritchart and Kuang, 2016; Kuang and Cui, 2016; Carignan, 2017). For instance, recent modeling work by Story (2012) involves synthesis of voice (including aspiration) and supralaryngeal articulations, along with their time-varying characteristics. Moreover, the role of the voice in indexing non-phonological meaning (including

differences based on age, sex, and gender) is also an important component to linguistic studies of the voice (Iseli et al., 2007; Zhuang and Hasegawa-Johnson, 2008; Esling and Edmondson, 2010; Kreiman and Sidtis, 2011; Mendoza-Denton, 2011; Callier, 2013; Zimman, 2013; Podesva and Callier, 2015; Starr, 2015; Park et al., 2016). Finally, through interdisciplinary research we stand to learn much more about how our voices convey different types of meaning.

Notes

[1]I thank editors William Katz and Peter Assmann, an anonymous reviewer, Adam Chong, Pat Keating, Jody Kreiman, Yaqian Huang, and Robin Samlan for their thoughtful comments and suggestions on earlier versions of this chapter.

[2]The transcriptions in Hmong and Mazatec include "Chao letters" (Chao, 1930) for differences in lexical tone: the vertical bar reflects the pitch range from lowest (level 1) to highest (level 5) in one's normal speaking range, and the horizontal lines perpendicular to the vertical bar represent pitch values along that range from the beginning to the end of the tone.

[3]In fact, some current versions of this psychoacoustic voice model parameterize the spectral slope of the inharmonic noise in four frequency bands (Kreiman et al., 2016).

References

Abramson, A. S., Tiede, M. K. and Luangthongkum, T. (2015), 'Voice register in Mon: acoustics and electroglottography', *Phonetica* **72**, 237–256.

Andruski, J. E. (2006), 'Tone clarity in mixed pitch/phonation-type tones', *Journal of Phonetics* **34**, 388–404.

Avelino, H. (2010), 'Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation', *Journal of Voice* **24**, 270–280.

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

Batliner, A., Burger, S., Johne, B. and Kießling, A. (1993), MÜSLI: A classifcation scheme for laryngealizations, *in* 'Proceedings of ESCA workshop on prosody', Lund, pp. 176–179.

Berkson, K. H. (2013), Phonation Types in Marathi: An Acoustic Investigation, PhD thesis, University of Kansas.

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

Bishop, J. and 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.

Blankenship, B. (2002), 'The timing of nonmodal phonation in vowels', *Journal of Phonetics* **30**, 163–191.

Bonilha, H. S., Deliyski, D. D., Whiteside, J. P. and Gerlach, T. T. (2012), 'Vocal fold phase asymmetries in patients with voice disorders: A study across visualization techniques', *American Journal of Speech-Language Pathology* **21**, 3–15.

Brunelle, M. (2010), The role of larynx height in the Javanese tense $\sim$ lax stop contrast, *in* R. Mercado, E. Potsdam and L. d. Travis, eds, 'Austronesian and Theoretical Linguistics', John Benjamins, Amsterdam, pp. 7–24.

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.

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

Brunelle, M. and Kirby, J. (2016), 'Tone and phonation in Southeast Asian languages', *Language and Linguistics Compass* **10**, 191–207.

Brunelle, M., Nguyễn, D. D. and Nguyễn, K. H. (2010), 'A laryngographic and laryngoscopic study of Northern Vietnamese tones', *Phonetica* **67**, 147–169.

Callier, P. (2013), Linguistic context and the social meaning of voice quality variation, PhD thesis, Georgetown University.

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

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.

Chao, Y. R. (1930), 'A system of "tone-letters"', *Le Maître Phonetique* **45**, 24–27.

Chen, G., Kreiman, J., Gerratt, B. R., Neubauer, J., Shue, Y.-L. and Alwan, A. (2013), 'Development of a glottal area index that integrates glottal gap size and open quotient', *Journal of the Acoustical Society of America* **133**, 1656–1666.

Choi, J.-Y., Hasegawa-Johnson, M. and Cole, J. (2005), 'Finding intonational boundaries using acoustic cues related to the voice source', *Journal of the Acoustical Society of America* **118**, 2579–2587.

Cumming, K. E. and Clements, M. A. (1995), 'Glottal models for digital speech processing: A historical survey and new results', *Digital Signal Processing* **5**, 21–42.

Davidson, L. and Erker, D. (2014), 'Hiatus resolution in American English: the case against glide insertion', *Language* **90**, 482–514.

de Krom, G. (1993), 'A cepstrum-based technique for determining harmonics-to-noise ratio in speech signals', *Journal of Speech and Hearing Research* **36**, 254–266.

de Krom, G. (1995), 'Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments', *Journal of Speech and Hearing Research* **38**, 794–811.

Denes, P. B. and Pinson, E. N. (1993), *The speech chain: The physics and biology of spoken language*, 2nd edn, W. H. Freeman, New York.

Di Napoli, J. (2015), Glottalization at phrase boundaries in Tuscan and Roman Italian, *in* J. Romero and M. Riera, eds, 'The Phonetics–Phonology Interface: Representations and methodologies', John Benjamins, Amsterdam.

DiCanio, C. (2014), 'Cue weight in the perception of Trique glottal consonants', *Journal of the Acoustical Society of America* **119**, 3059–3071.

DiCanio, C. T. (2009), 'The phonetics of register in Takhian Thong Chong', *Journal of the International Phonetic Association* **39**, 162–188.

DiCanio, C. T. (2012), 'Coarticulation between tone and glottal consonants in Itunyoso Trique', *Journal of Phonetics* **40**, 162–176.

Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (1996), 'Glottalization of word-initial vowels as a function of prosodic structure', *Journal of Phonetics* **24**, 423–444.

Edmondson, J. A., Chang, Y., Hsieh, F. and Huang, H. J. (2011), Reinforcing voiceless finals in Taiwanese and Hakka: Laryngoscopic case studies, *in* 'Proceedings of the 17th International Congress of Phonetic Sciences', Hong Kong.

Edmondson, J. A. and 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.

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

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

Esling, J. H. and Edmondson, J. A. (2010), Acoustical analysis of voice quality for sociophonetic purposes, *in* M. D. Paolo and M. Yaeger-Dror, eds, 'Sociophonetics: A student's guide', Routledge, London, chapter 11.

Esling, J. H. and 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.

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

Esposito, C. M. (2010*a*), 'The effects of linguistic experience on the perception of phonation', *Journal of Phonetics* **38**, 306–316.

Esposito, C. M. (2010*b*), 'Variation in contrastive phonation in Santa Ana Del Valle Zapotec', *Journal of the International Phonetic Association* **40**, 181–198.

Esposito, C. M. (2012), 'An acoustic and electroglottographic study of White Hmong phonation', *Journal of Phonetics* **40**, 466–476.

Esposito, C. M. and Khan, S. (2012), 'Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong', *Journal of the International Phonetic Association* **42**, 123–143.

Flanagan, J. L. (1972), *Speech Analysis, Synthesis, and Perception*, Springer, Berlin.

Fulop, S. A. and Golston, C. (2008), 'Breathy and whispery voice in White Hmong', *Proceedings of meetings on acoustics* **4**, 060006.

Gallagher, G. (2015), 'Natural classes in cooccurrence constraints', *Lingua* **166**, 80–98.

Garellek, M. (2012), 'The timing and sequencing of coarticulated non-modal phonation in English and White Hmong', *Journal of Phonetics* **40**, 152–161.

Garellek, M. (2013), Production and perception of glottal stops, PhD thesis, UCLA.

Garellek, M. (2014), 'Voice quality strengthening and glottalization', *Journal of Phonetics* **45**, 106–113.

Garellek, M. (2015), 'Perception of glottalization and phrase-final creak', *Journal of the Acoustical Society of America* **137**, 822–831.

Garellek, M. and Keating, P. (2011), 'The acoustic consequences of phonation and tone interactions in Jalapa Mazatec', *Journal of the International Phonetic Association* **41**, 185–205.

Garellek, M., Keating, P., Esposito, C. M. and Kreiman, J. (2013), 'Voice quality and tone identification in White Hmong', *Journal of the Acoustical Society of America* **133**, 1078–1089.

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

Garellek, M., Samlan, R., Gerratt, B. R. and Kreiman, J. (2016), 'Modeling the voice source in terms of spectral slopes', *Journal of the Acoustical Society of America* **139**, 1404–1410.

Garellek, M. and Seyfarth, S. (2016), Acoustic differences between English /t/ glottalization and phrasal creak, *in* 'Proceedings of Interspeech 2016', San Francisco, pp. 1054–1058.

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

Gerfen, C. and Baker, K. (2005), 'The production and perception of laryngealized vowels in Coatzospan Mixtec', *Journal of Phonetics* **33**, 311–334.

Gerratt, B. R. and Kreiman, J. (2001), 'Toward a taxonomy of nonmodal phonation', *Journal of Phonetics* **29**, 365–381.

Gick, B., Wilson, I. and Derrick, D. (2013), *Articulatory Phonetics*, Wiley-Blackwell, Oxford.

Gobl, C. and Ní Chasaide, A. (2010), Voice source variation and its communicative functions, *in* W. J. Hardcastle, J. Laver and F. E. Gibbon, eds, 'The Handbook of Phonetic Sciences', 2nd edn, Wiley-Blackwell, Oxford, chapter 11, pp. 378–423.

Gordon, M. (2014), Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels, *in* H. van der Hulst, ed., 'Word Stress: Theoretical and Typological Issues', Oxford University Press, Oxford, pp. 83–118.

Gordon, M. and Applebaum, A. (2010), 'Acoustic correlates of stress in Turkish Kabardian', *Journal of the International Phonetic Association* **40**, 35–58.

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

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

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

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

Harris, J. G. (1999), States of the glottis for voiceless plosives, *in* 'Proceedings of the International Congress of Phonetic Sciences', San Francisco, pp. 2041–2044.

Hartl, D. M., Hans, S., Vessière, J. and Brasnu, D. F. (2003), 'Objective acoustic and aerodynamic measures of breathiness in paralytic dysphonia', *European Archives of Oto-Rhino-Laryngology* **260**, 175–182.

Hillenbrand, J., Getty, L. A., Clark, M. J. and Wheeler, K. (1995), 'Acoustic characteristics of American English vowels', *Journal of the Acoustical Society of America* **97**, 3099–3111.

Hillenbrand, J. M. and Houde, R. A. (1996), 'Role of F0 and amplitude in the perception of glottal stops', *Journal of Speech and Hearing Research* **39**, 1182–1190.

Hirano, M. and Katika, Y. (1985), Cover-body theory of vocal fold vibration, *in* R. G. Daniloff, ed., 'Speech science: recent advances', College-Hill Press, San Diego, pp. 1–46.

Hirose, H. (1997), Investigating the physiology of laryngeal structures, *in* W. J. Hardcastle and J. Laver, eds, 'The Handbook of Phonetic Sciences', Blackwell, Oxford, pp. 116–136.

Hirose, H. (2010), Investigating the physiology of laryngeal structures, *in* W. J. Hardcastle, J. Laver and F. E. Gibbon, eds, 'The Handbook of Phonetic Sciences', 2nd edition edn, Blackwell, Oxford, pp. 130–152.

Hirose, H. and Gay, T. (1972), 'The activity of the intrinsic laryngeal muscles in voicing control', *Phonetica* **25**, 140–164.

Hombert, J.-M., Ohala, J. J. and Ewan, W. G. (1979), 'Phonetic explanations for the development of tones', *Language* **55**, 37–58.

Honda, K., Hirai, H., Masaki, S. and Shimada, Y. (1999), 'Role of vertical larynx movement and cervical lordosis in F0 control', *Language and Speech* **42**, 401–411.

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

Isogai, Y., Horiguchi, S., Honda, K., Aoki, Y., Hirose, H. and Saito, S. (1988), A dynamic simulation model of vocal fold vibration, *in* O. Fujimura, ed., 'Vocal Physiology: Voice Production, Mechanisms and Functions', Raven Press, New York, pp. 191–206.

Kane, J., Drugman, T. and Gobl, C. (2013), 'Improved automatic detection of creak', *Computer Speech and Language* **27**, 1028–1047.

Keating, P., Esposito, C., Garellek, M., Khan, S. and Kuang, J. (2011), Phonation contrasts across languages, *in* 'Proceedings of the International Congress of Phonetic Sciences', Hong Kong, pp. 1046–1049.

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

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J. and 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.

Khan, S. (2012), 'The phonetics of contrastive phonation in Gujarati', *Journal of Phonetics* **40**, 780–795.

Kingston, J. (2005), *The phonetics of Athabaskan tonogenesis*, John Benjamins, Amsterdam, pp. 137–184.

Kirby, J. (2013), The role of probabilistic enhancement in phonologization, *in* A. C. L. Yu, ed., 'Origins of sound change: approaches to phonologization', Oxford University Press, Oxford, pp. 228–246.

Kirby, J. P. (2014), 'Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies', *Journal of Phonetics* **43**, 69–85.

Klatt, D. H. and 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.

Kreiman, J. (1982), 'Perception of sentence and paragraph boundaries in natural conversation.', *Journal of Phonetics* **10**, 163–175.

Kreiman, J., Antoñanzas-Barroso, N. and Gerratt, B. R. (2010), 'Integrated software for analysis and synthesis of voice quality', *Behavior Research Methods* **42**, 1030–1041.

Kreiman, J., Antoñanzas-Barroso, N. and Gerratt, B. R. (2016), 'The UCLA voice synthesizer, version 2.0', *Journal of the Acoustical Society of America* **140**, 2961.

Kreiman, J., Garellek, M., Chen, G., Alwan, A. and Gerratt, B. R. (2015), 'Perceptual evaluation of voice source models', *Journal of the Acoustical Society of America* **138**, 1–10.

Kreiman, J., Garellek, M. and Esposito, C. (2011), 'Perceptual importance of the voice source spectrum from H2 to 2 kHz', *Journal of the Acoustical Society of America* **130**, 2570.

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

Kreiman, J. and Gerratt, B. R. (2005), 'Perception of aperiodicity in pathological voice', *Journal of the Acoustical Society of America* **117**, 2201–2211.

Kreiman, J. and Gerratt, B. R. (2010), 'Perceptual sensitivity to first harmonic amplitude in the voice source', *Journal of the Acoustical Society of America* **128**, 2085–2089.

Kreiman, J. and Gerratt, B. R. (2012), 'Perceptual interaction of the harmonic source and noise in voice', *Journal of the Acoustical Society of America* **131**, 492–500.

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

Kreiman, J., Gerratt, B. R. and Khan, S. (2010), 'Effects of native language on perception of voice quality', *Journal of Phonetics* **38**(4), 588–593.

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

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

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

Kuang, J. (2012), 'Registers in tonal contrasts', *UCLA Working Papers in Phonetics* **110**, 46–64.

Kuang, J. (2013*a*), Phonation in tonal contrasts, PhD thesis, UCLA.

Kuang, J. (2013*b*), 'The tonal space of contrastive five level tones', *Phonetica* **70**, 1–23.

Kuang, J. (2017*a*), 'Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice', *Journal of the Acoustical Society of America* **142**, 1693–1706.

Kuang, J. (2017*b*), Creaky voice as a function of tonal categories and prosodic boundaries, *in* 'Proceedings of Interspeech 2017', Stockholm, pp. 3216–3220.

Kuang, J. and Cui, A. (2016), Relative cue weighting in perception and production of a sound change in progress. Talk presented at LabPhon 2016, Cornell, NY.

Kuang, J. and Keating, P. (2014), 'Vocal fold vibratory patterns in tense versus lax phonation contrasts', *Journal of the Acoustical Society of America* **136**, 2784–2797.

Kuang, J. and Liberman, M. (2015), Influence of spectral cues on the perception of pitch height, *in* 'Proceedings of the 18th International Congress of Phonetic Sciences', Glasgow, UK.

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

Lancia, L., Voigt, D. and Krasovitskiy, G. (2016), 'Characterization of laryngealization as irregular vocal fold vibration and interaction with prosodic prominence', *Journal of Phonetics* **54**, 80–97.

Laver, J. (1980), *The Phonetic Description of Voice Quality*, Cambridge University Press, Cambridge.

Löfqvist, A. and McGowan, R. S. (1992), 'Influence of consonantal envelope on voice source aerodynamics', *Journal of Phonetics* **20**, 93–110.

Lombardi, L. (1991), Laryngeal features and laryngeal neutralization, PhD thesis, University of Massachusetts.

Maddieson, I., Flavier, S., Marsico, E. and Pellegrino, F. (2016), 'LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0.', `http://www.lapsyd.ddl.ish-lyon.`

cnrs.fr/lapsyd/. (Last checked 2017-07-22).
URL: *http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/*

Mendoza-Denton, N. (2011), 'The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona', *Journal of Linguistic Anthropology* **21**, 261–280.

Miller, A. L. (2007), 'Guttural vowels and guttural co-articulation in Ju|'hoansi', *Journal of Phonetics* **35**, 56–84.

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

Moisik, S. R. (2013), 'Harsh voice quality and its association with blackness in popular American media', *Phonetica* **69**, 193–215.

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

Moisik, S. R. and Esling, J. H. (2014), 'Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: a low-dimensional model of vocal-ventricular fold coupling', *Journal of Speech, Language, and Hearing Research* **57**, S687–S704.

Moisik, S. R., Lin, H. and 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.

Mooshammer, C. (2010), 'Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German', *Journal of the Acoustical Society of America* **127**, 1047–1058.

Munhall, K. and Löfqvist, A. (1992), 'Gestural aggregation in speech-laryngeal gestures', *Journal of Phonetics* **20**, 111–126.

Nakatani, L. H. and Dukes, K. D. (1977), 'Locus of segmental cues for word juncture', *Journal of the Acoustical Society of America* **62**, 714–719.

Nellis, D. G. and Hollenbach, B. E. (1980), 'Fortis versus lenis in Cajonos Zapotec phonology', *International Journal of American Linguistics* **46**, 92–105.

Ní Chasaide, A. and Gobl, C. (1993), 'Contextual variations of the vowel voice source as a function of adjacent consonants', *Language and Speech* **36**, 303–330.

Ogden, R. (2001), 'Turn transition, creak and glottal stop in Finnish talk-in-interaction', *Journal of the International Phonetic Association* **31**, 139–152.

Park, S. J., Sigouin, C., Kreiman, J., Keating, P., Guo, J., Yeung, G., Kuo, F.-Y. and Alwan, A. (2016), Speaker identity and voice quality: Modeling human responses and automatic speaker recognition, *in* 'Proceedings of Interspeech 2016', San Francisco, pp. 1044–1048.

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

Podesva, R. J. and Callier, P. (2015), 'Voice quality and identity', *Annual Review of Applied Linguistics* **35**, 173–194.

Ratliff, M. (2015), Tonoexodus, tonogenesis, and tone change, *in* P. Honeybone and J. Salmons, eds, 'The Oxford Handbook of Historical Phonology', Oxford University Press, Oxford, pp. 245–261.

Redi, L. and Shattuck-Hufnagel, S. (2001), 'Variation in the realization of glottalization in normal speakers', *Journal of Phonetics* **29**, 407–429.

Reetz, H. and Jongman, A. (2008), *Phonetics: transcription, production, acoustics, and perception*, Wiley-Blackwell, Oxford.

Remijsen, B. (2013), 'Tonal alignment is contrastive in falling contours in Dinka', *Language* **89**, 297–327.

Rodgers, J. (1999), 'Three influences on glottalization in read and spontaneous German speech', *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* **34**, 177–284.

Samlan, R. A. and Kreiman, J. (2014), 'Perceptual consequences of changes in epilaryngeal area and shape', *Journal of the Acoustical Society of America* **136**, 2798–2806.

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

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

Samlan, R. A., Story, B. H., Lotto, A. J. and Bunton, K. (2014), 'Acoustic and perceptual effects of left–right laryngeal asymmetries based on computational modeling', *Journal of Speech, Language, and Hearing Research* **57**, 1619–1637.

Sapienza, C., Hicks, D. M. and Ruddy, B. H. (2011), Voice disorders, *in* N. B. Anderson and G. H. Shames, eds, 'Human Communication Disorders: An Introduction', 8th edition edn, Pearson, Boston, pp. 202–237.

Shrivastav, R. and Sapienza, C. M. (2006), 'Some difference limens for the perception of breathiness', *Journal of the Acoustical Society of America* **120**, 416–423.

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

Silverman, D. (2003), Pitch discrimination between breathy vs. modal phonation, *in* J. Local, R. Ogden and R. Temple, eds, 'Phonetic Interpretation (Papers in Laboratory Phonology 6)', Cambridge University Press, Cambridge, pp. 293–304.

Silverman, D., Blankenship, B., Kirk, P. and Ladefoged, P. (1995), 'Phonetic structures in Jalapa Mazatec', *Anthropological Linguistics* **37**, 70–88.

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

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

Slifka, J. (2003), 'Respiratory constraints on speech production: Starting an utterance', *Journal of the Acoustical Society of America* **114**, 3343–3353.

Slifka, J. (2006), 'Some physiological correlates to regular and irregular phonation at the end of an utterance', *Journal of Voice* **20**, 171–186.

Sluijter, A. M. C. and van Heuven, V. J. (1996), 'Spectral balance as an acoustic correlate of linguistic stress', *Journal of the Acoustical Society of America* **100**, 2471–2485.

Starr, R. L. (2015), 'Sweet voice: The role of voice quality in a Japanese feminine style', *Language in Society* **44**, 1–34.

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

Story, B. H. (2012), 'Phrase-level speech simulation with an airway modulation model of speech production', *Computer Speech and Language* **27**, 989–1010.

Story, B. H. (2015), Mechanisms of voice production, *in* M. A. Redford, ed., 'Handbook of Speech Production', Wiley-Blackwell, Oxford, pp. 34–58.

Stross, B. (2013), 'Falsetto voice and observational logic: Motivational meanings', *Language in Society* **42**, 139–162.

Sundberg, J. (1987), *The Science of the Singing Voice*, Northern Illinois University Press, DeKalb, IL.

Tian, J. and Kuang, J. (2016), Revisiting the register contrast in Shanghai Chinese, *in* 'Proceedings of Tonal Aspects of Languages 2016', pp. 147–151.

Titze, I. R. (1992), 'Phonation threshold pressure: A missing link in glottal aerodynamics', *Journal of the Acoustical Society of America* **91**, 2926–2935.

Titze, I. R. (1994), *Principles of Voice Production*, Prentice Hall Inc., Engelwood Cliffs, NJ.

Titze, I. R. (2006), *The Myoelastic Aerodynamic Theory of Phonation*, National Centre for Voice and Speech, Iowa City, IA.

Titze, I. R., Schmidt, S. S. and Titze, M. R. (1995), 'Phonation threshold pressure in a physical model of the vocal fold mucosa', *Journal of the Acoustical Society of America* **97**, 3080–3084.

Traill, A. (1985), *Phonetic and Phonological Studies of the !Xóõ Bushmen*, Buske, Hamburg.

Umeda, N. (1978), 'Occurrence of glottal stops in fluent speech', *Journal of the Acoustical Society of America* **64**, 88–94.

van den Berg, J. (1958), 'Myoelastic aerodynamic theory of voice production', *Journal of Speech and Hearing Research* **1**, 227–244.

van den Berg, J. and Tan, T. S. (1959), 'Results of experiments with human larynxes', *Practica Oto-Rhino-Laryngologica* **21**, 425–450.

Vicenik, C. (2010), 'An acoustic study of Georgian stop consonants', *Journal of the International Phonetic Association* **40**, 59–92.

Wang, C. (2015), Multi-register Tone Systems and Their Evolution on the Jianghan Plain, PhD thesis, Hong Kong University of Science and Technology.

Wang, C.-Y. and Tang, C.-J. (2012), The falsetto tones of the dialects in Hubei province, *in* 'Proceedings of the 6th International Conference on Speech Prosody'.

Wayland, R. and Jongman, A. (2003), 'Acoustic correlates of breathy and clear vowels: The case of Khmer', *Journal of Phonetics* **31**, 181–201.

Yanushevskaya, I., Murphy, A., Gobl, C. and Ní Chasaide, A. (2016), Perceptual salience of voice source parameters in signaling focal prominence, *in* 'Proceedings of Interspeech 2016', San Francisco, pp. 3161–3165.

Yanushevskaya, I., Ní Chasaide, A. and Gobl, C. (2011), Universal and language-specific perception of affect from voice, *in* 'Proceedings of the 17th International Congress of Phonetic Sciences', Hong Kong, pp. 2208–2211.

Yin, J. and Zhang, Z. (2013), 'The influence of thyroarytenoid and cricothyroid muscle activation on vocal fold stiffness and eigenfrequencies', *Journal of the Acoustical Society of America* **133**, 2972–2983.

Yu, K. M. (2017), 'The role of time in phonetic spaces: temporal resolution in Cantonese tone perception', *Journal of Phonetics* **65**, 126–144.

Yu, K. M. and Lam, H. W. (2014), 'The role of creaky voice in Cantonese tonal perception', *Journal of the Acoustical Society of America* **136**, 1320–1333.

Zhang, J. and Yan, H. (2015), Contextually dependent cue weighting for a laryngeal contrast in Shanghai Wu, *in* 'Proceedings of the International Congress of Phonetic Sciences', Glasgow, UK, pp. 1–5.

Zhang, Z. (2015), 'Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control', *Journal of the Acoustical Society of America* **137**, 898–910.

Zhang, Z. (2016*a*), '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.

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

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.

Zhang, Z., Kreiman, J., Gerratt, B. R. and Garellek, M. (2013), 'Acoustic and perceptual effects of changes in body layer stiffness in symmetric and asymmetric vocal fold models', *Journal of the Acoustical Society of America* **133**, 453–462.

Zhuang, X. and Hasegawa-Johnson, M. (2008), Towards interpretation of creakiness in Switchboard, *in* 'Proceedings of Speech Prosody', pp. 37–40.

Zimman, L. (2013), 'Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men', *Journal of Language & Sexuality* **2**, 1–39.

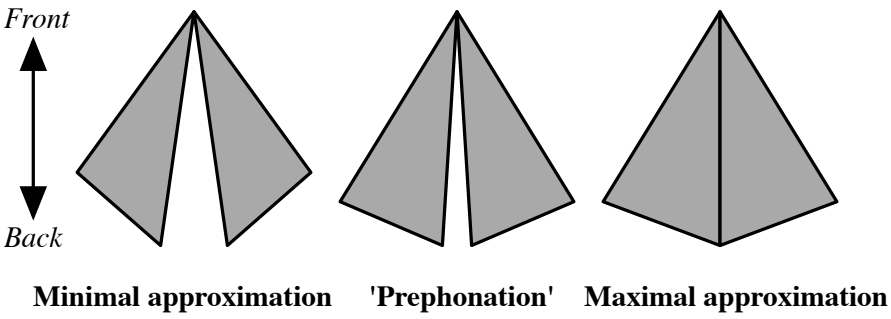


Figure 1. *Schematized laryngoscopic (bird's-eye) view of the vocal folds showing different degrees of vocal fold approximation for voiceless sounds. In between minimal approximation (e.g., for a voiceless glottal fricative) and maximal approximation (e.g., for a glottal stop) is a state sometimes called 'prephonation', which is found for voiceless unaspirated stops (Harris, 1999; Esling and Harris, 2003; Edmondson et al., 2011).*

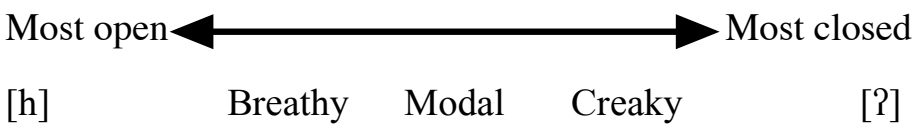


Figure 2. *Simplified one-dimensional space for voice quality (based on Ladefoged, 1971; Gordon and Ladefoged, 2001).*

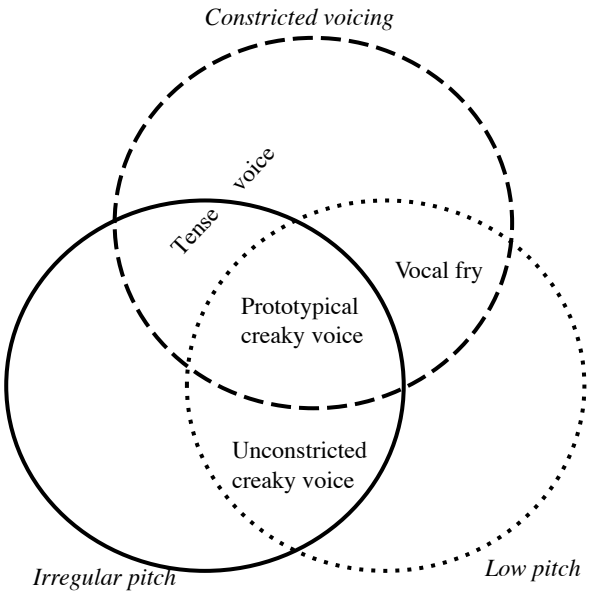


Figure 3. *Three-dimensional space for creaky voice, with some more common subtypes shown (based on Keating et al., 2015).*

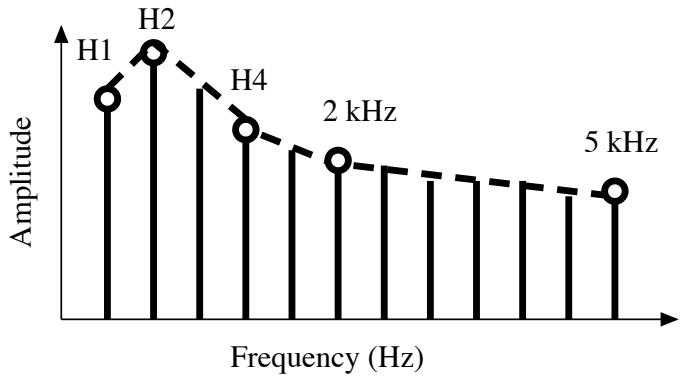


Figure 4. *The four-parameter harmonic source spectrum model, fitted to the spectrum of a natural voice. The voice source was estimated via inverse filtering, and its spectrum was then calculated via fast Fourier transform. Differences in the amplitudes of individual harmonics are altered so that they conform to the slope of the appropriate model segment (based on Kreiman et al., 2014).*

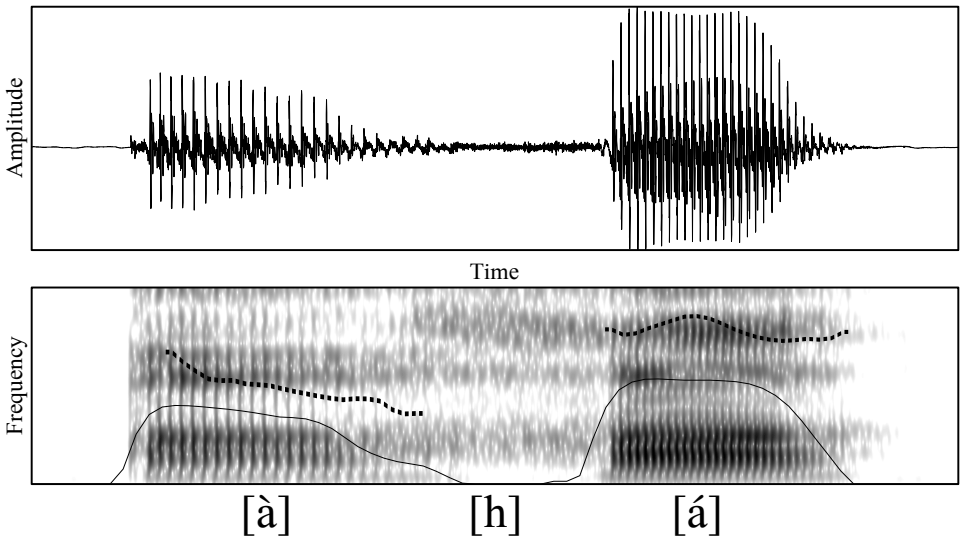


Figure 5. *Waveform and spectrogram of the sequence [àhá]. The presence of voicing is indicated by the glottal pulses in both displays, as well as by the presence of an f0 track (shown in the dashed line). The intensity track (the solid horizontal line in the display) provides information on the amplitude of the signal during voicing.*

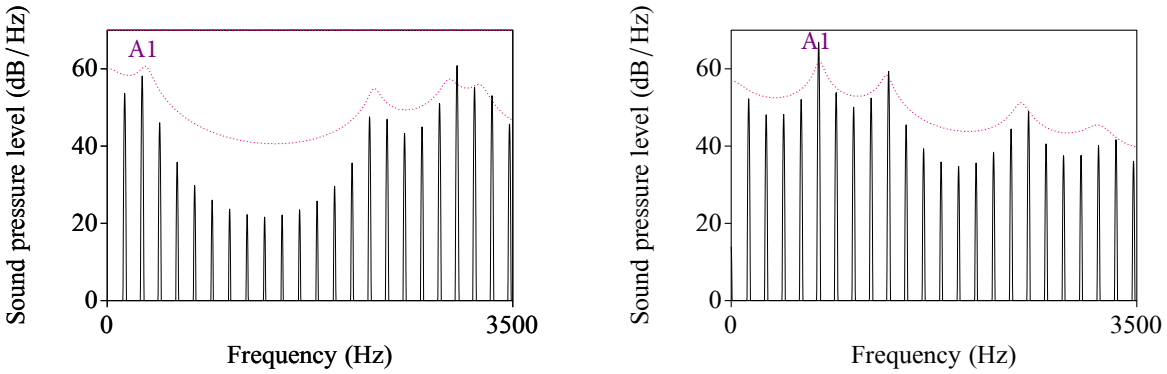


Figure 6. *FFT and LPC spectra for synthetic* [i] *(left) and* [ɑ] *(right) with f0 = 150 Hz and formants equal to that of an average adult male speaker of American English. For* [i] *(left), the second harmonic (H2) is also the harmonic closest to the first formant (A1), so H1–H2 equals H1–A1. For* [ɑ] *(right), H1–A1 equals H1–H5.*

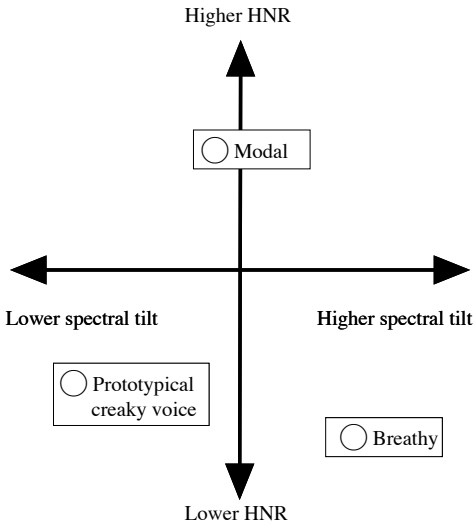
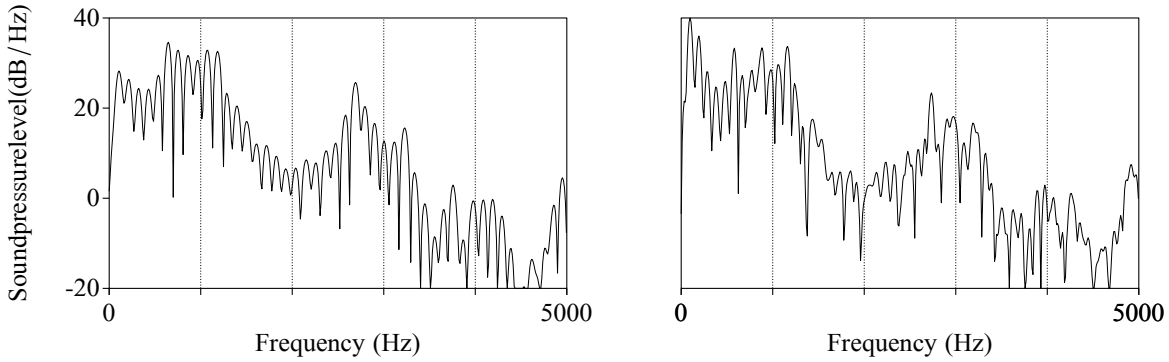


Figure 7. *Relative acoustic differences between breathy, modal, and (prototypical) creaky voice, in terms of both spectral tilt and noise. The difference in position of breathy voice and prototypical creaky voice on the HNR scale is arbitrary; cf. Blankenship 2002 and Garellek 2012.*



	First half	Second half
H1*–H2* (dB)	1.8	2.6
H2*–H4* (dB)	2.3	10.1
H4*–H2 kHz* (dB)	19.1	27.4
H2 kHz*–H5 kHz* (dB)	0.9	–5.7
CPP (a measure of HNR, dB)	21.3	19.0

Figure 8. *Top: audio spectra of the first vowel in the sequence* [àhá]. *The left panel is the spectrum taken over the first half of the vowel; the right panel over the second half. Bottom: spectral tilt parameters and CPP (Cepstral Peak Prominence, a measure of HNR). All acoustic measures but H2 kHz*–H5 kHz* indicate greater breathiness in the second half compared with the first.*

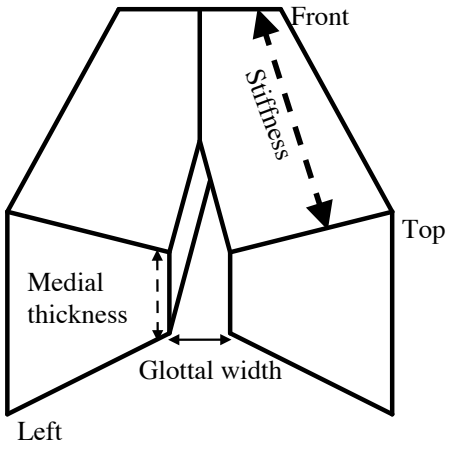


Figure 9. *Simplified model of the vocal folds, after (Zhang, 2015, 2016*a*). The primary voice dimensions are influenced mainly by glottal width (the angle between the folds, thin unbroken arrow), their medial vertical thickness (thin dashed arrow), their stiffness from front to back (thick dashed arrow), and the interactions of these parameters with the subglottal pressure. Transverse stiffness is also included in later models (Zhang, 2017), but is less relevant for the primary voice dimensions in language (as discussed in text).*

Table 1. *Primary vocal fold movements and their use in sounds of the world's languages. The dashed line between rate vs. quality of voicing indicates that these two dimensions are often dependent on each other.*

Dimension	**Articulatory description**	**Relevant sounds**
Spreading-Constriction	How far apart the vocal folds are from each other	All voiced sounds All voiceless sounds, *e.g.* Aspirated sounds Glottalized sounds Fricatives Trills Ejectives
Voicing	Whether the vocal folds are vibrating	All voiced sounds, *e.g.* Sonorant consonants (Voiced) vowels
Rate	Rate of vibration	Tone Intonation Stress
Quality	Constriction of vibration Irregularity/noise	Register Contrastive voice quality ('phonation type')

Table 2. *Valves of the throat and their functioning (from Esling and Harris 2005; Edmondson and Esling 2006; see also Figure 1 in Edmondson and Esling 2006). Valve 1 is similar to the continuum represented in Figure 2. The remaining valves pertain to structures above the vocal folds, and therefore would fall outside the narrow definition of 'voice' used in this chapter.*

Valve	**Description**
Valve 1	vocal fold adduction and abduction
Valve 2	ventricular fold adduction
Valve 3	compression of the arytenoids and aryepiglottic folds
Valve 4	epiglottal-pharyngeal constriction
Valve 5	larynx raising
Valve 6	pharyngeal narrowing

Table 3. *Components of the psychoacoustic model of voice quality and associated parameters (from Kreiman et al. 2014).*

Model component	**Parameters**
Harmonic source spectral slope (Figure 4)	H1–H2 (see description in text) H2–H4 H4–H2 kHz H2 kHz–H5 kHz
Inharmonic source noise	Harmonics-to-noise ratio (HNR)
Time-varying source characteristics	f0 track Amplitude track
Vocal tract transfer function	Formant frequencies and bandwidths Spectral zeroes and bandwidths

Table 4. *Sample H1–H2 and HNR values for two groups of Vowels A vs. B and A′ vs. B′. Even though Vowels A and A′ share the same H1–H2 (as do B and B′), the HNR differences suggest that Vowel B is creakier than Vowel A, whereas Vowel A′ is breathier than Vowel B′.*

	Vowel A	Vowel B	Vowel A′	Vowel B′
H1–H2	10	5	10	5
HNR	20	10	10	20
Interpretation	A has higher tilt, *less* noise than B. A = modal, B = creaky		A′ has higher tilt, *more* noise than B′. A′ = breathy, B′ = modal	

Table 5. *Summary of psychoacoustic voice model's parameters according to primary phonological dimensions of voice.*

Dimension	**Relevant model parameters**
Vocal fold approximation	Absence of f0 track Aspiration noise (if vocal folds are spread) Voice quality changes on adjacent voiced sounds
Voicing	Presence of f0 track
Rate of vibration	Frequency of f0 track
Voice quality (compared with modal)	**Breathy voice:** Higher H1–H2, H2–H4, H4–H2 kHz, H2 kHz–H5 kHz Lower HNR **Unconstricted creaky voice:** Higher H1–H2 H2–H4, H4–H2 kHz, H2 kHz–H5 kHz Lower HNR Lower f0 **Constricted creaky voice qualities** *(Prototypical creaky, tense voice, and vocal fry)*: Lower H1–H2 H2–H4, H4–H2 kHz, H2 kHz–H5 kHz Lower HNR (prototypical creaky voice) Lower f0 (prototypical creaky and vocal fry)

Table 6. *Summary of Zhang (2015, 2016*a*, 2017)'s articulatory voice model's parameters according to primary phonological dimensions of voice. The initiation and sustaining of voicing depends on complex interactions between the four parameters (for more details, see Zhang 2016*a*). The precise model parameters required for some voice qualities are at present speculative (and are marked by an asterisk), but are based on articulatory and acoustic characteristics described in previous studies, as described in Section 2.*

Dimension	**Relevant model parameters**
Approximation	Glottal width
Voicing	Glottal width Medial vocal fold thickness Vocal fold stiffness Subglottal pressure
Rate of vibration	*To raise f0:* Greater vocal fold stiffness Greater subglottal pressure Smaller glottal width
Voice quality (compared with modal)	**Breathy voice:** Less vocal fold thickness Larger glottal width ***Unconstricted creaky voice**: Less vocal fold thickness Larger glottal width Less vocal fold stiffness Lower subglottal pressure **Constricted creaky voice qualities** *(Prototypical creaky, tense voice, and vocal fry)*: More vocal fold thickness Smaller glottal width (Tense voice, vocal fry, and *prototypical creaky) Less vocal fold stiffness (Vocal fry and *prototypical creaky) Lower subglottal pressure (Vocal fry)