

Voicing of glottal consonants and non-modal vowels

Marc Garellek 

UC San Diego
mgarellek@ucsd.edu

Yuan Chai

UC San Diego
yuc521@ucsd.edu

Yaqian Huang

UC San Diego
yah101@ucsd.edu

Maxine Van Doren

UC San Diego
mvandore@ucsd.edu

Variation in voicing is common among sounds of the world's languages: sounds that are analyzed as voiceless can undergo voicing, and those analyzed as voiced can devoice. Among voiceless glottal sounds in particular, voicing is widespread: linguists often expect the voiceless glottal stop [ʔ] and fricative [h] to be fully voiced, especially between vowels. In this study, we use audio recordings from Illustrations of the International Phonetic Alphabet published in the *Journal of the International Phonetic Association* to explore the extent to which glottal consonants and non-modal (breathy and creaky) vowels differ in terms of percentage voicing and voicing intensity in three phrasal positions. We find that voiceless [h] is only slightly less voiced than voiced [ɦ] in initial position. Between two vowels, both [h] and [ɦ] are as voiced as breathy vowels. Glottal stops and creaky vowels are both characterized by high percentages of voicing, but they differ in voicing intensity: in all phrasal positions, glottal stops generally have periods of strong and weak voicing, whereas creaky vowels are strongly voiced. In contrast, vowels described as 'rearticulated', 'checked', or 'glottalized' show similar drops in voicing intensity to glottal stops. We interpret these results through an articulatory lens: glottal consonants and non-modal vowels are both modulations in phonation resulting from laryngeal constriction and vocal fold spreading. We argue further that, because voicing during [ʔ] and [h] is largely predictable from respiratory and prosodic constraints, many cases of [ʔ] and [h] can be considered phonetically underspecified for voicing.

me ã:fɛ̃, si (h) pø ɛt vokalik sã sɛ:sɛ d ɛ:trə (h), kɛl ɛ la difɛ:rãs ã:trə (h) ɛ (ɦ)?¹
 [But then, if (h) can be voiced without ceasing to be (h), what is the difference between (h) and (ɦ)?]
 (Passy 1900)

In the great majority of languages we have heard, glottal stops are apt to fall short of complete closure, especially in intervocalic positions.
 (Ladefoged & Maddieson 1996: 75)

1 Introduction

For about as long as the International Phonetic Alphabet (IPA) has existed, phoneticians (e.g. Meyer 1900, Passy 1900) have acknowledged that glottal consonants show large variation in voicing. And when glottal consonants are voiced, they look very similar to non-modal (breathy and creaky) vowels. In this paper, we seek to better understand how glottal consonants and non-modal vowels differ in terms of voicing.

The IPA defines several sounds as ‘glottal’. Among the consonants, there are those with a glottal place of articulation, meaning that the primary place of articulation for vocal tract constriction is at the glottis, or space between the vocal folds: the voiceless glottal stop [ʔ] involves occlusion at the glottis, whereas the voiceless glottal fricative [h] and the (breathy) voiced glottal fricative [ɦ] involve spreading. Additionally, there are consonants whose primary place of articulation is supraglottal, but which secondarily involve the larynx; Ladefoged & Maddieson (1996: Section 3.1) refer to this secondary articulation as a sound’s ‘laryngeal setting’. Such sounds are generally called ‘aspirated’ (e.g. [t^h]) if they have vocal fold spreading, or ‘glottalized’ if they have laryngeal constriction (e.g. [ʔt]).

Spreading and constriction in the larynx are also associated with non-modal voicing, independent of whether they occur on a voiced consonant or vowel (Garellek 2019). Voicing with more laryngeal constriction is called ‘creaky’ (or ‘laryngealized’) in the IPA (e.g. [b̤ a]), whereas voicing with more vocal fold spreading is ‘breathy’ (e.g. [b̥ a]).

When glottal consonants that are transcribed as voiceless [ʔ h] undergo voicing, they are realized as creaky and breathy voice, respectively, on an adjacent sound. For example, researchers investigating prosodic strengthening in American English have demonstrated that /ʔV hV/ are often realized as [V̤V V̤V] in prosodically weak positions, such as word-medially (Pierrehumbert & Talkin 1992, Garellek 2014). Intervocalic voicing of [ʔ h] to creaky and breathy voice is known to occur in many languages (Meyer 1900, Ladefoged & Maddieson 1996: 75); see also recent work on the degree of voicing for /h/ in Estonian (Teras 2018) and Hungarian (Deme et al. 2020) and /ʔ(:), h(:)/ in Maltese (Mitterer 2018).

Devoicing of the sound transcribed as voiced [ɦ] is less frequently discussed, but this is likely because the sound is much rarer than voiceless glottal consonants. Still, in languages that are analyzed as having [ɦ], devoicing is attested. For example, DiCanio (2010: 229) analyzed Itunyoso Trique as having /h/ that is usually (partially) voiced, but in later work re-analyzed the sound as a voiced /ɦ/ that frequently varies with [h] in careful speech (DiCanio 2014: 885). And in Nepali, Khatiwada (2009) reports cross-speaker differences in the realization of the language’s glottal fricative, with some speakers producing [ɦ] and others [h].

Similarly, non-modal voicing often has a period of voicelessness. For example, Gujarati breathy vowels can be fully voiced (e.g. [V̤]) or have a period of voiceless aspiration during the vowel (e.g. [V̥V̥]), particularly in careful speech (Khan 2012). The creaky tone in White Hmong tends to end with a voiceless glottal stop, which has led some researchers to analyze it as a ‘checked’ tone; that is, as a (modal) vowel ending in a glottal stop (Huffman 1987;

¹ In French orthography: *Mais enfin, si (h) peut être vocalique sans cesser d’être (h), quelle est la différence entre (h) et (ɦ)?* The term *vocalique* means ‘voiced’ in this context; e.g. in 1902 the IPA differentiated voiceless [s] (*soufflé*) from voiced [z] (*vocalique*) (IPA 1902: 158). Today the French-language IPA chart distinguishes voiceless vs. voiced sounds as *non-voisé* vs. *voisé* (https://www.internationalphoneticassociation.org/IPAcharts/IPA_Kiel_2019_full_fra.pdf; accessed July 2020).

see also discussion in Garellek et al. 2013). Variation in voicing is also common in Zapotec languages, where ‘laryngealized’ vowels typically vary between a weakly creaky (that is, strongly voiced) articulation and a ‘rearticulated’ one with a voiceless glottal-stop-like gesture in the middle of the vowel (Ariza-García 2018).

The quotations included at the start of the paper highlight how the theoretical issues regarding voicing of glottal consonants are as old as modern theories of phonetic notation, and that such issues involve most, if not all, languages with a glottal consonant. Yet despite the known variable realization with respect to voicing, the five main symbols for glottal sounds in the IPA – [ʔ h fi ʋ ʋ] – are assumed to be specified phonetically for their voicing. The goal of this paper is to explore how systematically sounds transcribed as [ʔ h] are voiceless, and how systematically sounds transcribed as [fi ʋ ʋ] are voiced. Our primary hypothesis is that, consistent with the choice of symbols used, voiceless [ʔ h] should in general have less voicing, and be more weakly voiced, than their voiced counterparts [fi ʋ ʋ]. But given previous reports of variation in voicing of glottal consonants by phrasal position, we also expect more voicing of [ʔ h] when they occur word-medially between vowels, compared to word-initially and word-finally. We investigate this by analyzing voicing during glottal sounds in words found in Illustrations of the IPA published in this journal. In the remainder of the Introduction, we discuss why voicing is so variable in glottal consonants (Section 1.1), as well as the implications of voicing variation in glottals for phonological contrasts (Section 1.2) and the theory of the IPA (Section 1.3). We then briefly review the goals and hypotheses of the current study (Section 1.4).

1.1 Causes of voicing variation in glottal sounds

Several related factors are involved in the voicing of voiceless glottal consonants and in the devoicing of [fi] and non-modal vowels: the first involves the changes in subglottal pressure over the course of an utterance; the second is the shared articulatory space for voicing and glottal consonants; the third relates to processes of increased articulatory force and undershoot, such as prosodic strengthening and lenition.²

Voicing is possible as long as there is a sufficient pressure difference across the glottis, and as long as the vocal folds are sufficiently approximated. If the vocal folds are too spread apart, or are too tightly constricted, their vibration will cease (for a recent overview, see Kreiman & Sidtis 2011: Section 2.3). Over the course of an utterance, the subglottal pressure starts low before reaching a higher-pressure plateau, which falls very little until the end of the utterance, at which point the subglottal pressure quickly drops (Ladefoged 1963). Because of the low subglottal pressure, and because voicing initiation requires a higher subglottal–oral pressure differential than voicing maintenance, utterance onsets are either voiceless or weakly voiced. Spreading or constricting the vocal folds in utterance-initial position thus leads to unstable voicing or complete voicelessness, because voicing is weak to begin with, and because spreading and constriction modulate the ‘sufficient approximation’ required of the vocal folds to initiate vibration. But utterance-medially, voicing is generally favored, even with full occlusion of the vocal tract (Westbury & Keating 1986). This is because the subglottal pressure is high, and because it is relatively easier to sustain voicing once it is initiated. In order to inhibit voicing in utterance-medial position – as for a voiceless stop [t] – speakers therefore spread or constrict their vocal folds (Seyfarth & Garellek 2020), and/or retain the oral constriction for as long as needed in order for the oral and subglottal pressure to equalize (Kingston & Diehl 1994). In utterance-final position, voicing continues as long as the subglottal pressure is higher than the oral one. At the very end of an utterance, where the subglottal pressure is low, even weak spreading or constricting

² There are of course other factors that might influence voicing, such as coarticulation from adjacent voiceless vs. voiced sounds. These are not discussed further because of insufficient data in the corpus to adequately address them.

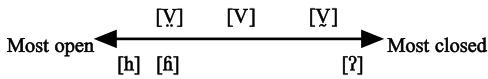


Figure 1 Continuum model of glottal stricture; after Ladefoged (1971) and Gordon & Ladefoged (2001).

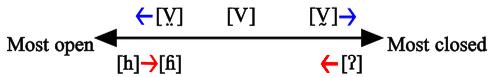


Figure 2 (Colour online) Illustration of reduction of voiceless glottal consonants (inward-pointing red arrows) and strengthening of non-modal vowels (outward-pointing blue arrows) along a glottal continuum. The red arrows correspond to pressures of being in utterance-medial position leading to voicing of glottal consonants, whereas the blue arrows correspond to pressures of being adjacent to an utterance boundary, leading to devoicing of non-modal vowels.

of the vocal folds for [h] and [ʔ] could result in rapid devoicing. But if changes in vocal fold approximation occur earlier, before the rapid fall in subglottal pressure, [h] and [ʔ] can still be largely voiced except at the end of the sound.

Vocal fold spreading and constriction therefore inhibit voicing when sufficiently large in magnitude. If they are weaker in magnitude, however, non-modal voicing can be sustained, particularly when the subglottal pressure is high: utterance-medially, and before the fall in subglottal pressure that characterizes the end of an utterance. Ladefoged (1971) and Gordon & Ladefoged (2001) modeled the relationship between vocal fold spreading, non-modal voice, modal voice, and constriction along a continuum of glottal stricture, shown in Figure 1 (adapted from Garellek 2019).

The Continuum model refers only to the average opening between the vocal folds (or the arytenoid cartilages) in accounting for differences across voice qualities and glottal consonants. Modal voice has an average opening; breathy voice has a greater average opening; creaky voice a smaller average opening; [h] has the greatest opening, and [ʔ] the smallest opening. Voiced [ɦ] is assumed to be similar to breathy voice, with a more open glottis than modal voice, but a less open glottis than for [h]. However, as we discuss in Section 1.3 below, the model ignores much additional detail about articulation, particularly for constricted sounds (Esling et al. 2019: Section 2.5). Still, it provides a straightforward and unidimensional way to account for the often gradient distinctions between glottal consonants and different types of voicing.

A primary reason then for variation in voicing of glottal sounds is the shared articulatory space between glottal consonants and non-modal vowels, which can be modeled along a continuum: voicing inhibition is achieved via greater spreading or constriction of the vocal folds (as well as other laryngeal structures), whereas breathy and creaky voice qualities are achieved through weaker instantiations of those very same gestures. Yet articulatory gestures vary in magnitude, notably when they are produced with greater or lesser articulatory force. For example, gestural magnitude often increases in strong prosodic positions, such as domain-initially, as a form of prosodic strengthening (Fougeron & Keating 1997, Fougeron 2001, Keating et al. 2003, Cho & Keating 2009); gestural magnitude in turn decreases due to reduced speech effects, such as effort reduction and hypo-articulation (Lindblom 1990).

For [ɦ] and breathy vowels, the gestural target is a moderate degree of vocal fold spreading; for creaky vowels, it is a moderate degree of vocal fold (and, more broadly, laryngeal) constriction. Moderate spreading or constriction allows for both voicing and a spread or constricted quality to the voicing. But if these gestures are STRENGTHENED, such as domain-initially, the degree of spreading or constriction could increase to the point of losing voicing. For [h] and [ʔ], however, the gestural targets are higher degrees of the same underlying articulations found for breathy and creaky vowels: spreading and constriction. When these gestures are WEAKENED, such as when they occur word-medially, the degree of spreading/constriction is lessened, leading to non-modal voicing, or at the extreme, even modal voicing. This is represented in Figure 2. Across languages, we do find many reports

of voicing of voiceless glottal sounds, and devoicing of non-modal voice, discussed in the previous section. Therefore, even if there are systematic differences between voiceless vs. voiced glottal targets, we still expect a strong effect of phrasing: voicelessness is favored at utterance edges due to the lower subglottal pressure; in initial position, voicelessness is further favored due to prosodic strengthening, which usually leads to articulatory gestures of greater magnitude. In medial position, however, both respiratory factors (namely, the high subglottal pressure) and prosodic factors (namely, the lack of prosodic strengthening, and hypo-articulation) favor voicing.

1.2 Voicing contrasts among glottal sounds

As we have discussed, voiceless glottal sounds are prone to undergoing voicing, and voiced glottal sounds to devoicing. But in spite of this, languages can contrast voiceless vs. voiced glottals, as well as different degrees of vocal fold spreading and constriction. For example, in Danish, both fortis and lenis stops /t ɖ/ are voiceless in word-initial position, with fortis stops being realized as voiceless aspirated [t^h], and lenis ones as voiceless unaspirated [t~ɖ] (Grønnum 1998). Using fiberoptic imaging of the glottis, Hutters (1985) showed that both stop series involve vocal fold spreading, but that the lenis stops have a weaker and shorter spreading gesture than fortis stops. Moreover, many Indic languages contrast voiceless vs. voiced post-aspirated stops, though they also often show voicing variation during their aspiration (Seyfarth & Garellek 2018, Schertz & Khan 2020).

As for glottal consonants themselves, contrasts between /h/ vs. /ɦ/ are rare (Mielke 2018) and often of unclear phonetic realization. For example, the Lyon–Albuquerque Phonological Systems Database (LAPSyD; Maddieson et al. 2014–2020) notes only two languages with /h/ vs. /ɦ/: Umbundu and Hainanese Min. Umbundu contrasts oral /h/ vs. nasalized /ɦ̃/, but the phonetic nature of the contrast is not well established (Schadeberg 1982). Hainanese Min is said to contrast onset /h/ vs. /ɦ/ (Woon 1979), but phonetically the contrast is between [h] vs. ø (Liu & Wang 2008). Ladefoged & Maddieson (1996: 326) additionally cite two Chadic languages, Lame and Musey, that contrast /h/ vs. /ɦ/. But for Musey they add that Shryock (1995) found that both sounds involve voicing.

The Nguni languages of Southern Africa (e.g. Xhosa, Zulu) often contrast /h/ vs. /ɦ/, where the latter belongs to the set of (often voiced) ‘depressor’ consonants that lower the pitch of the following vowel. However, the realization of depressor contrasts is phonetically multidimensional (Downing 2018), and systematic differences in the presence of voicing between /h/ vs. depressor /ɦ/ – such as less voicing for the former than the latter – have yet to be established. Finally, among Illustrations of the IPA published before 2021, a contrast (either underlying or surface) between [h] and [ɦ] is reported for only six languages: Basaá, Lizu, Shanghainese Wu, Lili Wu, Lower Xumi, and Upper Xumi.

Languages with both /h ɦ/ and /ʔ ʕ/, or with both /ʔ/ and /ʕ/, are also attested. In LAPSyD for example, of the 12 languages with breathy or creaky vowels, seven also have at least one glottal consonant reported in their inventory. This implies that many languages with non-modal phonation will also have glottal consonants in their inventory, but that the converse is not true: most languages with glottal consonants, at least two thirds of spoken languages, have no non-modal vowels. However, because glottal consonants and non-modal vowels occupy different positions within a word, it is extremely rare to have a surface contrast between voiceless glottals vs. non-modal vowels. Zapotec languages frequently contrast ‘laryngealized’ (weakly creaky or rearticulated) vowels from checked ones ending in a glottal stop, but this vocalic contrast also differs in phasing relative to the vocalic gesture (Ariza-García 2018). The Papuan language Gimi provides a rare counterexample to the tendency for glottal consonants and non-modal vowels to occupy different positions: it allows for a contrast between /ʔ/ and a ‘creaky voiced glottal approximant’ between two vowels (Ladefoged & Maddieson 1996: 76–77). An overview of patterns is shown in Table 1.

With respect to voicing contrasts for glottal sounds then, two generalizations emerge: on the one hand, languages can contrast voiceless vs. voiced glottal sounds, such as [h] vs. [ɦ],

Table 1 Types of reported voicing contrasts for glottal sounds, with example languages.

Type of contrast	Language examples
T ^h vs. D ^h	Many Indic languages
h vs. ħ	UMbundu, Hainanese Min, some Chadic languages (Lame, Musey), several Nguni languages (Xhosa, Zulu), Wu languages (Lili Wu, Shanghainese), and other Tibeto-Burman languages (Upper and Lower Xumi)
ʔ vs. ʕ	Gimi
V ^ʔ vs. V̥	Many Zapotec languages

[h] vs. [V̥], or [ʔ] vs. [V̥], even though these sounds frequently vary phonetically in terms of voicing. On the other hand, these contrasts are rare, and are frequently part of a multidimensional phonological contrast, such as nasalization contrasts in UMbundu and depressor contrasts in Nguni languages. This calls into question whether the voicing distinction is a necessary, regularly occurring, component to the contrast.

1.3 Glottal sounds and the IPA

There are many well-known controversial aspects to the theory of the IPA and phonetic notation in general (Mielke 2009, Esling 2010, Ladd 2014). But for glottal sounds it is no exaggeration that ALL conventions – those having to do with their place and manner of articulation, as well as their status as segments vs. suprasegments – are controversial. In terms of place of articulation, the glottal fricatives clearly involve vocal fold spreading (and thus an articulation at the glottis), with minimal engagement of other laryngeal structures. However, glottal stops usually involve incursion of the ventricular folds (Garellek 2013, Esling et al. 2019: 50), and quite often – at least in careful and hyperarticulated speech – constriction of the epilaryngeal area and laryngeal raising (Esling et al. 2019: Section 2.3.5).³

In the years following the 1989 Kiel Convention, there was some published discussion about the controversies regarding the three glottal consonants' manner of articulation (Catford 1990, Ladefoged 1990, Kloster-Jensen 1991, Laufer 1991, Iivonen 1992; see also Ladefoged & Maddieson 1996: 73–77, 325–236). Ladefoged (1990) argued that the glottal stop and fricatives should be relabelled as approximants, and that they should be removed from the main chart of pulmonic consonants chart to join the 'Other symbols', given that they involve no appreciable constriction in the vocal tract. Indeed, glottal consonants cannot be conceived as segments in the way that other supralaryngeal consonants typically are. Glottal consonants involve a change in phonatory source (Catford 1964), but the filtering of the source, and thus place of articulation, depends on whatever the adjacent supralaryngeal sounds are; that is, they are underspecified for (vocal tract) place (Keating 1988). For these reasons, Ladefoged (1990) proposed removing them from the main pulmonic consonants chart and renaming the voiceless glottal fricative [h] as a 'voiceless approximant', with [ħ] being renamed a 'breathy-voiced approximant'.

³ An argument could be made, then, that glottal sounds should be called 'laryngeal' instead. Indeed, the voiceless ones historically were: the 1900 IPA chart referred to [h ʔ] as 'laryngeal' (in French), though the terms 'glottal stop' and 'glottal catch' were already in use before that time (e.g. Sweet 1874; Groth 1894: 7, 26), with 'glottal stop' becoming the more popular term after about 1910 (Google Books Ngram Viewer; Michel et al. 2011). In contrast, voiced [ħ] was called 'guttural' in the 1900 chart, but by 1904 appeared under the column of 'throat' sounds along with [h ʔ]. Since 1930 these three sounds have been referred to as 'glottal' within the chart; see https://www.internationalphoneticassociation.org/IPAcharts/IPA_hist/IPA_hist_2018.html (accessed July 2020).

The close phonetic relationship between consonants [h ɦ ʔ] and non-modal vowels has long been acknowledged in the IPA. For example, Firth used both superscript [V^h] (Firth 1957) and post-vocalic [Vh] (Firth 1961) to refer to breathy voice or ‘murmur’ in Gujarati. In contrast, Pandit (1957) used subscript [V_h] for murmured vowels in Gujarati, but also argued that these can be thought of as underlying /Vh/. The diacritics currently used for breathy and creaky voice were adopted only recently: in 1975, then-Secretary of the International Phonetic Association and *JIPA* editor John Wells suggested using [Vh] and [Vʔ] for these (Wells 1975), but the breathy-voice diacritic was approved in 1976 (along with superscript [h] for aspiration).⁴

Another issue regarding voicing distinctions with glottal sounds relates to their use in secondary articulations. For sounds that are aspirated or glottalized, whether or not the aspiration or glottalization is voiced is often dependent on the voicing of the primary articulation. Thus, [t^h] refers to a voiceless post-aspirated consonant whose alveolar constriction and aspiration are both voiceless, whereas [d^h] refers to a voiced post-aspirated consonant where both are voiced. (Officially the IPA does not symbolically distinguish voiceless vs. voiced aspiration, but in practice many researchers symbolize a voiced post-aspirated consonant as [d^h] instead of [d^h].) Voiced aspirated stops are sometimes thought to be necessarily voiced throughout their aspiration (Ladefoged 1971), though it is clear that the aspiration in voiced aspirated consonants is frequently voiceless (see recent discussion by Schertz & Khan 2020 and Faytak, Steffman & Tankou 2020).

For glottalized sounds, additional confusion arises due to the use of both ‘glottal stop’ and ‘creaky voice’ symbols. Phoneticians generally call a sound ‘glottalized’ if it has any amount of laryngeal constriction, and symbolize it with a superscript [ʔ] (Keating, Wymark & Sharif 2021). Thus, a pre-glottalized alveolar stop is usually represented as [tʔ], but the common assumption is that the glottalization may be realized on a continuum of creaky voice to full glottal occlusion on a preceding voiced sound (Ashby & Przedlacka 2014, Garellek 2015, Seyfarth & Garellek 2020).

In sum, the place and manner of articulation of glottal sounds have long been sources of debate for phonetic notation and the IPA. Debates regarding the segmental idealization of glottal consonants, and the phonetic differences between voiceless vs. voiced glottal sounds that the IPA promotes, are of particular interest here. The segmental and voicing idealizations can be useful for phonological analysis (Catford 1990), but belie an important fact about the phonetic reality: glottal consonants and non-modal vowels exist on a shared continuum that ranges from more voiceless and consonant-like, to more voiced and vowel-like.

1.4 Goals and hypotheses

Our main research question is to explore the extent to which voiceless glottal consonants [ʔ h] differ in voicing from the voiced glottal fricative [ɦ] and breathy and creaky vowels. Our hypotheses are the following: given the choice of symbols used to transcribe these sounds, we expect that VOICELESS glottals [ʔ h] will have less voicing, and weaker voicing, than VOICED glottals – [ɦ] and breathy and creaky vowels. Second, all sounds (regardless of the choice of symbols) should be more voiced utterance-medially, where gestural magnitude is expected to be reduced, and where the high subglottal pressure should favor voicing; conversely, all sounds should be less voiced in utterance-initial position, where low subglottal pressure should favor voicelessness.

⁴ The current diacritics for breathy voice and creaky voice had been in use at least since Ladefoged (1971). Previously, the diacritic that is currently used for creaky voice had been used for pharyngealized vowels (e.g. Smalley 1968: 441; note that this is still often the convention used for the pharyngealized vowels in !Xóð: e.g. Traill 2018, see also Garellek 2020). The diacritic for creaky voice was officially approved only at the Kiel Convention (Pullum & Ladusaw 1996). There are currently no official IPA symbols for weaker forms of breathy and creaky voice, such as ‘slack’ and ‘stiff’ (or ‘lax’ and ‘tense’) voice; see Esposito & Khan (2020) for discussion of weaker forms of non-modal phonation.

2 Method

2.1 The corpus: Illustrations of the IPA

We investigated voicing of glottal sounds in Illustrations of the IPA published in this Journal. The Illustrations provide several benefits for addressing this research question: they provide data from a wide variety of languages; they include words in isolation, thus providing careful pronunciation of glottal sounds; they frequently provide minimal pairs, minimizing differences in the environment in which glottal sounds occur; they provide peer-reviewed, narrow phonetic transcriptions, minimizing transcriber bias; the audio files are freely available for research purposes. However, there are also drawbacks to using Illustrations: there are few tokens per target sound in a given language, and tokens usually come from one speaker for each language. We also do not know about the different types of recording setups and equipment used (such information is typically not included in Illustrations), as well as the post-processing of the audio.

We analyzed laryngeal sounds from the 201 Illustrations of the IPA published from 1989 through 2020.⁵ Forty-six languages were excluded from the analysis because they have no glottal consonants [ʔ h fi] and no non-modal vowels. An additional 24 languages with target sounds were excluded because they did not have sound files available for the particular target words in isolation, or because the recordings were deemed too noisy to ascertain the boundaries of target sounds; for example, strong broadband noise made it difficult to segment aspiration, and low-frequency noise made it difficult to segment glottalization.

In total, laryngeals from 131 languages and approximately 153 speakers were analyzed. (For Illustrations with multiple speakers, it was sometimes impossible to determine which of the speakers uttered a given word. In such cases, we marked the speaker of a particular token as ‘NA’, effectively collapsing across speakers.) The complete list of languages can be found in Supplementary materials at <https://osf.io/rs89n/>.

2.2 Target sounds

Target sounds included glottal consonants [ʔ h fi] and non-modal vowels [ʏ ʊ]. We generally used the transcriptions provided in the Illustrations to determine if a sound should be targeted. Consonants were only targeted if their aspiration or glottalization were adjacent to a vowel or a glide [j ɥ uɥ w]. For instance, we included utterance-initial tokens of form [fia hwa], but excluded those of form [fi.a ʔma] to avoid cases where the non-glide consonant would cause a drop in intensity.

We only targeted words in isolation; neither words from the passage (usually ‘The North Wind and the Sun’) nor words uttered within a larger phrase were included. This was to control as much as possible for the role of phrasal position on laryngeal voicing. Target sounds that were described as (possibly) having supra-laryngeal constriction were excluded, as such constriction promotes devoicing (Esling et al. 2019) and lowers voicing intensity (Chong et al. 2020, Garellek 2020). Thus, we excluded epiglottal stops and fricatives [ʔ ɳ ʕ], as well as [ʔ h] if they were reported to have pharyngealization or velarization, or sounded that way to us (due to trilling and/or lowered vowel quality). We also excluded breathy and creaky vowels with reports of harsh or pharyngealized voice qualities.

⁵ Illustrations published from 1989–1999 later re-appeared in the 1999 *Handbook of the IPA* (IPA 1999). The sound files associated with Illustrations from the *Handbook* and subsequent work are available for IPA members at <https://www.internationalphoneticassociation.org/content/ipa-handbook-downloads> and <https://www.internationalphoneticassociation.org/member/audio-files-illustrations-ipa> (accessed July 2020). The recordings are available for research purposes from the IPA’s webpage free of charge and without further permission: <https://www.internationalphoneticassociation.org/content/IPA-recordings-copyright> (accessed July 2020).

Target sounds could be either phonemic in the language or described as categorical allophones in the language. Categorical allophones include, for example, coda /k/ being realized as [ʔ] in varieties of Malay. We excluded prosodic glottal stops if they were described as occurring variably ('optionally, sometimes') before word-initial vowels (e.g. Lusoga; Nabirye, de Schryver & Verhoeven 2016), but included them if they were described as always occurring predictably before word-initial vowels (e.g. Bohemian Czech; Šimáčková, Podlipský & Chládková 2012), even if they were not transcribed. We avoided variable glottal stops as these frequently occur at utterance edges, and thus could be a product of phrasal creak instead of glottalization; i.e. they might not be due to a segment-like laryngeal constriction gesture.

We included glottal sounds that were prosodically derived. For example, in Chicahuaxtla Triqui, Elliott, Edmondson & Sandoval Cruz (2016: 355) analyze some instances as [h] and [ʔ] as derived from certain lexical tones, but state that these prosodic glottal consonants otherwise 'appear to be identical' to their segmental counterparts. Thus, we analyzed all instances of [h] and [ʔ] as the same target sound, even if phonologically they are analyzed as having either segmental or prosodic sources. The complete list of languages, including notes on transcription, can be found in Supplementary materials at <https://osf.io/rs89n/>.

2.3 Coding of corpus

In addition to segmenting the portions of glottalization and aspiration, we also logged various properties of the target words: language, speaker, the target word and its gloss, as well as the target laryngeal sound's position within the word. For the purposes of the analysis below, we coded the following types of laryngeal sounds: voiceless [h], voiced [ɦ], voiceless [ʔ], as well as breathy and creaky vowels [V̤ V̥]. The target types were based on the intended phoneme or categorical allophone.

We ignored allophonic voicing of [h] and [ʔ] (and conversely, devoicing of [V̤ V̥]), even if it was transcribed in the paper, thus relying on the symbol used for the underlying representation (when the sounds are phonemic) or for the broad transcription of the allophone (when the sounds are derived phonologically). We did this to test our hypothesis that voiceless glottal targets are distinct in their voicing from voiced [ɦ] and non-modal vowels.

We generally agreed with and followed the transcriptions provided by the authors. But in a few cases, we changed the transcriptions because of the descriptions provided by the authors. In Itunyoso Trique for example, DiCanio (2010: 229) posited a voiceless /h/ that is usually (partially) voiced, but in later work transcribes the sound as a voiced /ɦ/ that frequently varies with [h] in careful speech (DiCanio 2014: 885). Thus, we considered these to be voiced /ɦ/. Conversely, we changed Nepali voiced [ɦ] to [h], because Khatiwada (2009: footnote 2) remarks that the speaker in the Illustration only has voiceless [h], and that both [h] and [ɦ] are used for the same sound in the language. We will return to the possible implications of these decisions in the discussion in Section 4 below.

'Creaky' vowels [V̤] were defined as those which were labeled as one of the following: creaky, laryngealized, glottalized, rearticulated, checked. The analysis in the following section differentiates three groupings of creaky vowels, based on the description of the phasing of irregular voicing: (i) 'creaky/laryngealized' vowels were creaky throughout the vowel; (ii) 'rearticulated/glottalized' vowels were creakiest in the middle of the vowel; and (iii) 'checked' vowels were creakiest towards the end of the vowel. For example, Grønnum (1998) analyzes Danish glottalization (*stød*) as a kind of creaky voice that usually occurs towards the end of the vowel; thus, we considered this to be a type of 'checked' vowel. In some cases, a language's 'creaky' vowel varies in terms of the phasing of non-modal phonation. For example, Gutiérrez (2019) typically transcribes 'glottalized' vowels in Nivačle using a close phonetic transcription that reflects their variable realization as [V̤ V̥^Y]. Because in the Illustrations the creaky voicing is usually phased to the middle of the vowel, we paired these with other 'rearticulated' vowels whose creaky voice is also generally phased towards the

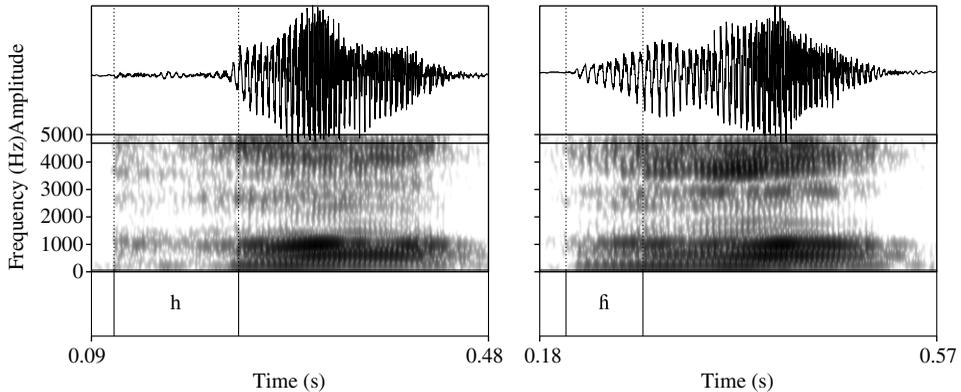


Figure 3 Segmentation of aspiration for two tokens of Lower Xumi (Chirkova & Chen 2013). The left panel is [ha] 'vegetable', whereas the right panel is [fi] 'pigeon'. The aspiration interval corresponds to where F1 and/or F2 energy is excited by noise rather than voicing.

middle of the vowel. The complete list of languages, including notes on coding, can be found in Supplementary materials at <https://osf.io/rs89n/>.

2.4 Segmentation procedure

As mentioned in the previous section, we only segmented aspiration or glottalization that was adjacent to at least one vowel. We segmented the target aspiration/glottalization using Praat (Boersma & Weenink 2020). Aspiration (either from [h fi] or a breathy vowel) was segmented as the portion of the word whose F1 and/or F2 were excited by noise, rather than voicing. Note that the aspiration interval could include aspiration that is completely voiceless, fully voiced (if there are glottal pulses but where either F1 or F2, or both F1 and F2 are excited by noise), or variably voiced. But because of the specific research questions of this study, our segmentation criteria made no reference to the presence of voicing during aspiration. In this respect, the aspiration interval used here is similar to the combined 'voiceless aspiration' and 'voiced aspiration (murmur)' interval used by Schertz & Khan (2020), except that our intervals were segmented exclusively based on F1/F2 rather than the voice bar and/or aspiration above F2. An example of our segmentation and annotation for aspiration is shown in Figure 3.

The segmentation procedure for glottalized targets (that is, for both glottal stops and the different types of creaky vowels) was similar to that used for targets with aspiration. The glottalized portion was assessed by irregular glottal pulses (pulse-to-pulse irregularity in energy or period). We also included any silence as part of the constriction, as a sign of full glottal/laryngeal occlusion. These criteria coincide with those used by Dilley, Shattuck-Hufnagel & Ostendorf (1996) and Garellek (2013) to identify irregular pitch periods and glottal stops. An example of our segmentation and annotation for glottalization is shown in Figure 4.

For breathy and creaky vowels, the non-modal phonation was segmented using the same criteria used for consonantal aspiration and glottalization. If their aspiration/glottalization overlapped with the entire vowel (e.g. if F1 and/or F2 were excited by aspiration noise for the duration of a breathy vowel), then the entire target vowel was segmented. For vowels labeled as non-modal but with only weak breathy/creaky voice – e.g. lax voice – that didn't result in much visible noise, we also segmented the entire vowel. If, however, only part of the vowel was breathy or creaky, then we segmented the sub-section of the vowel that was excited by noise; see Figure 5.

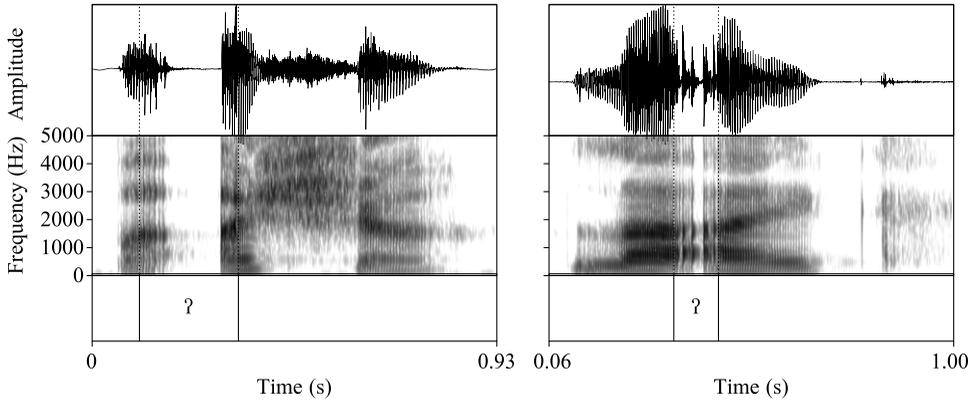


Figure 4 Examples of glottalization segmentation for two tokens in Northwest Sahaptin (Hargus & Beavert 2014). The left panel is [paʔaʃa] 'they entered', whose glottal stop shows full occlusion surrounded by irregular voicing; the right panel is [ʔaʔajk] 'sit relaxed', whose glottal stop is realized only with irregular voicing.

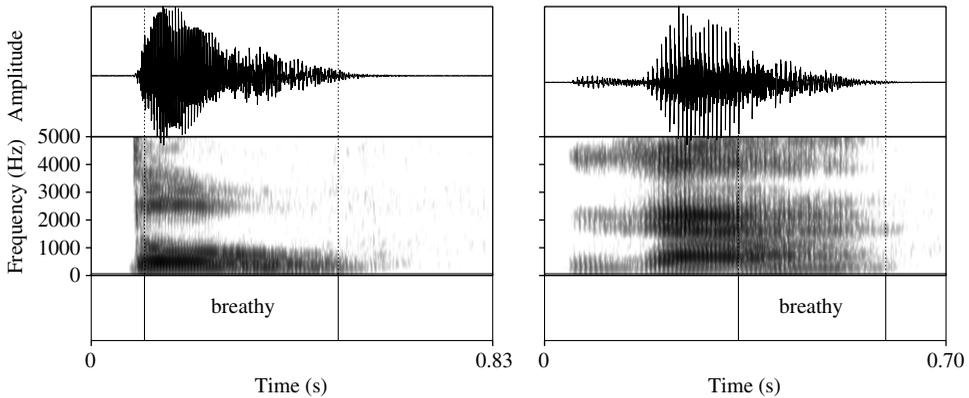


Figure 5 Examples of segmentation for two tokens of Hanoi Vietnamese bearing the mid-falling breathy (A2) tone (Kirby 2011). The left panel is [tu ʔ] 'prison', whose vowel is weakly breathy throughout. The right panel shows [ya ʔ] 'chicken', with voiced aspiration concentrated in the second half of the vowel.

In order to rule out cases where devoicing might be due to sources other than phonemic or allophonic glottalization and aspiration, we excluded tokens whose vowels might be affected by phrasal creak or breathy voice, or whose target sounds could be devoiced due to the presence on the same vowel of multiple laryngeal sounds. Two examples of such exclusions are the following: Upper-Chambira Urarina (Elias-Ulloa & Muñoz Aramburú 2021) has a fairly regular process of utterance-final vowel devoicing, so we excluded any tokens with a target glottal that appeared adjacent to an utterance-final vowel; Mah Meri (Kruspe & Hajek 2009) has breathy vowels in its lax register, but in the Illustration they always occurred before glottalized stops, and so they were not segmented for this study.

2.5 Acoustic measures and analyses

Our hypotheses outlined in Section 1.4 above make reference to the presence and degree of voicing. Therefore, our analysis below focuses on percentage voicing during aspiration/glottalization, and voicing intensity.

Percentage of voicing was calculated using the Voice Report in Praat (Boersma & Weenink 2020), which calculates the percentage of frames during which there are pulses and an f_0 . We used the default settings for voicing estimation, with an f_0 estimated between 50–600 Hz using the cross-correlation algorithm. Given that the f_0 can be lower than 50 Hz during glottalization and at utterance edges, the percentage of voicing that we report here is rather conservative. Glottalization or aspiration intervals shorter than 50 ms were excluded for the calculation of the percentage of voicing due to inaccurate identification of voicing over very short windows.

Voicing intensity was measured using strength of excitation (SoE), a measure of the relative amplitude of the impulse-like excitation at the point of significant excitation or ‘epoch’ during voicing. As such, SoE measures the relative amplitude of voicing, independent of noise in the signal (Murty & Yegnanarayana 2008). This measure has been used in studies of the role of vocal tract constriction (Mittal, Yegnanarayana & Bhaskararao 2014, Chong et al. 2020) and source aspiration or glottalization on the amplitude of voicing (Faytak et al. 2020, Garellek 2020). SoE was calculated using VoiceSauce (Shue et al. 2011) over the labeled intervals of aspiration and glottalization. VoiceSauce outputs an SoE value at every millisecond. In the results that follow, the SoE values have been normalized within language, to avoid potential differences in voicing intensity that may arise as a result of disparate recording equipment and conditions across Illustrations. Normalization proceeded as follows: first, since SoE values are log-normal with a right-tailed distribution, we removed any SoE outliers whose z -scored $\log(\text{SoE})$ values were greater than 2.5 standard deviations from the mean. Next, we calculated the maximum and minimum $\log(\text{SoE})$ values for a given language. Finally, the $\log(\text{SoE})$ values at a given time point were subtracted from the minimum value, and divided by the difference between the maximum–minimum values. Thus, a normed $\log(\text{SoE})$ value equal to 1 refers to the loudest voicing for a given language; values equal to 0 are voiceless.

In the analyses below, we characterize the data using descriptive rather than inferential statistics. Our study is exploratory, and given the asymmetries in the number of tokens per language and the fact that most languages present data from only one speaker, the patterns that we describe are primarily meant to provide predictions for future confirmatory analyses of larger cross-language and within-language phonetic studies of glottal sounds.

3 Results

A summary of the data analyzed can be found in Tables 2 and 3. As expected, given the prevalence of these target sounds across languages, there are many more tokens of voiceless glottal consonants [ʔ h] (and from more languages) than there are tokens of [fɪ] or non-modal vowels. Note that, because non-modal vowels are always preceded by a consonant in the languages sampled, there are no non-modal vowels that occur in utterance-initial position.

In Figure 6, we see that voiceless glottal consonants [ʔ h] in the data set come from many languages spoken in many different regions. In contrast, voiced [fɪ] and non-modal vowels are less frequently attested in Illustrations, and the languages with these sounds are more restricted geographically (as well as genetically). We will return to this point in the discussion in Section 4 below.

3.1 Results for aspirated sounds

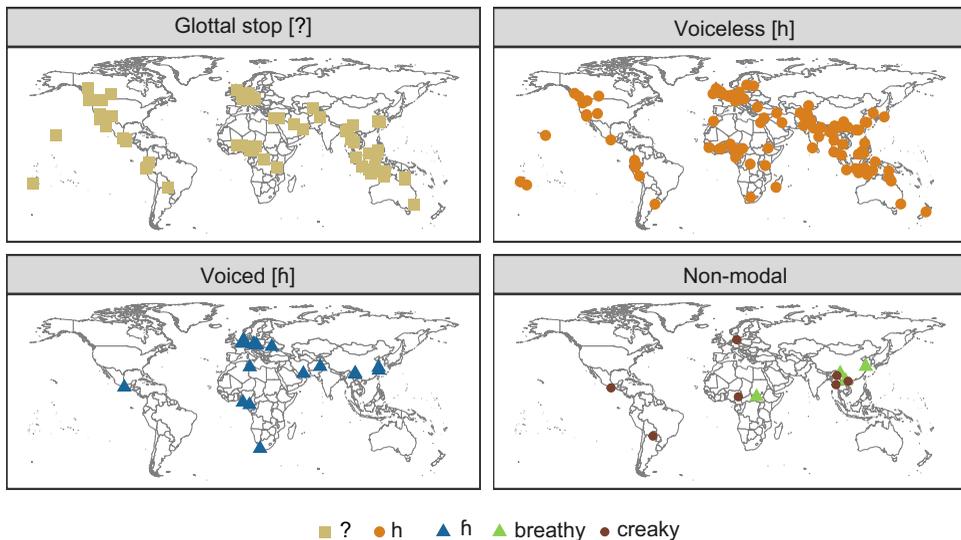
The percentage voicing results for aspirated sounds reveal several interesting patterns (Figure 7). As expected, there is a clear difference by position in utterance: most tokens in medial intervocalic position are voiced, whereas in final and (especially) initial positions there are more voiceless tokens. In initial position, there is only a slight difference in mean percentage of voicing, with [h] being voiced for an average of 30.14% (95% confidence interval

Table 2 Summary of aspirated tokens analyzed, by type of sound [h fɪ ʏ] and word/utterance position.

	Initial		Medial			Final		
	hV	fɪV	VhV	VfɪV	CʏC	Vh	Vfɪ	Cʏ
N tokens	335	35	84	8	41	58	33	31
N languages	91	15	40	4	3	16	3	3

Table 3 Summary of glottalized tokens analyzed, by type of sound [ʔ ʏ] and word/utterance position. As discussed below, vowels called 'creaky' here were labeled using various terms in the Illustrations.

	Initial	Medial		Final	
	ʔV	VʔV	CʏC	Vʔ	Cʏ
N tokens	177	152	53	232	62
N languages	36	35	5	25	6

**Figure 6** (Colour online) Map of sound types in the analysis. Each dot represents a language. Other Illustrations with glottal sounds do not appear here, if no tokens from those languages were analyzed.

(CI): 27.13%–33.15%) of its duration, and [fɪ] being voiced for an average of 38.61% (95% CI: 27.69%–49.54%) of its duration. In medial and final positions, most tokens are voiced for more than 50% of their duration, regardless of whether they are tokens of voiceless [h], voiced [fɪ], or breathy vowels. In these positions, some tokens of [h] are voiced for less than 50% of their duration, with voicelessness usually occurring in the very middle of intervocalic [h] and at the very end of word-final [h]. This could in principle support our hypothesis that [h] is more likely to be (partially) voiceless than voiced [fɪ] or breathy vowels. But given the relatively small sample size of voiced sounds, we hesitate to claim that this is the case. Future studies, ideally of languages with both voiceless and voiced glottal sounds, are needed to confirm this.

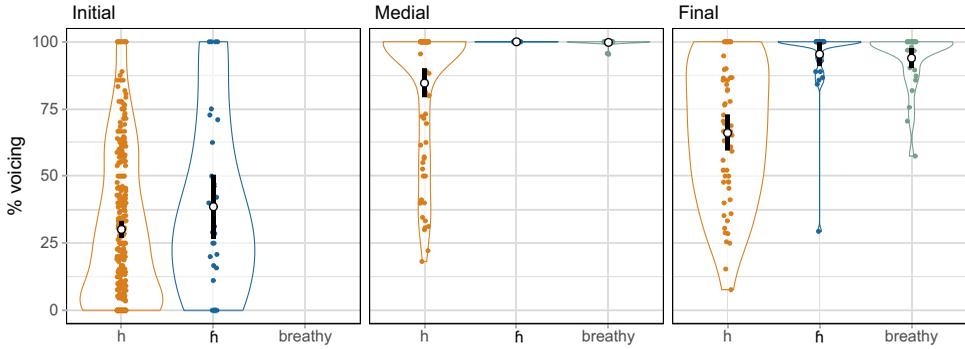


Figure 7 (Colour online) Percentage of voicing during aspiration for [h], voiced [ɦ], and breathy vowels, in the three phrasal positions. Each dot represents a token, and the violin plot represents concentration of tokens. The modeled means and 95% confidence intervals by position are shown in dark grey.

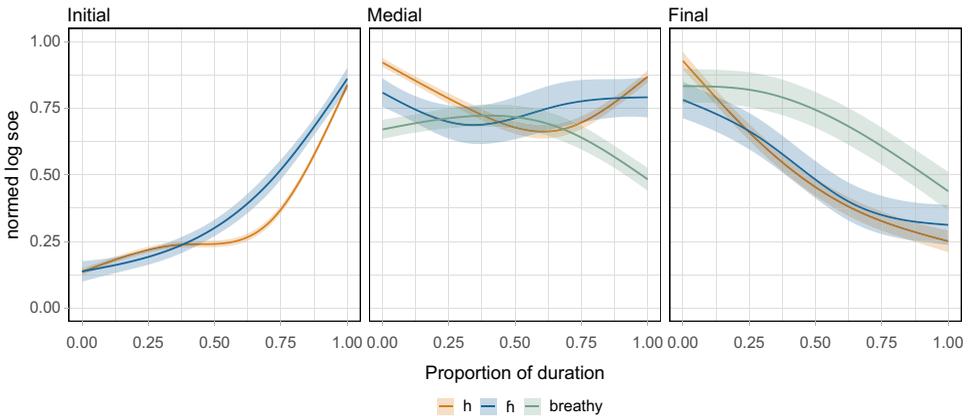


Figure 8 (Colour online) Time course of voicing intensity during aspiration for [h], voiced [ɦ], and breathy vowels, in the three phrasal positions. The values are estimated by cubic spline regression, with lines showing estimated means, and shaded areas showing one standard error above and below the estimated means.

The results for voicing intensity, shown in Figure 8, reveal similar patterns.⁶ As expected, voicing is strongest in medial intervocalic position compared to initial and final positions. In initial prevocalic position, voicing during [h] is slightly weaker than voicing during [ɦ], with the mean normed log SoE of [h] being 0.37 (95% CI: 0.36–0.39), while that of [ɦ] is 0.42 (95% CI: 0.36–0.48). Though the means do not differ appreciably, time course data also suggest that [ɦ] has stronger voicing than [h] during the latter half of aspiration. This is presumably because voiceless [h] involves a wider glottal opening than voiced [ɦ].

In medial position, all sounds are strongly voiced. Word-medial breathy vowels end in slightly weaker voicing because these vowels all happen to be utterance-final vowels (though followed by a coda consonant, and therefore not immediately utterance-final in position). In final position, voiceless [h], voiced [ɦ], and breathy vowels all show drops in voicing

⁶ We used splines models to estimate the predictions of normed log-transformed SoE based on values from subsets of different positions. For example, for aspirated sounds, the model is: $lmer(normed.log.soe \sim h/f/breathy * ns(time, df=3) + (1 + ns(time, df=3)) | filename) + (1 | language)$ for each position (initial, medial, final) subset.

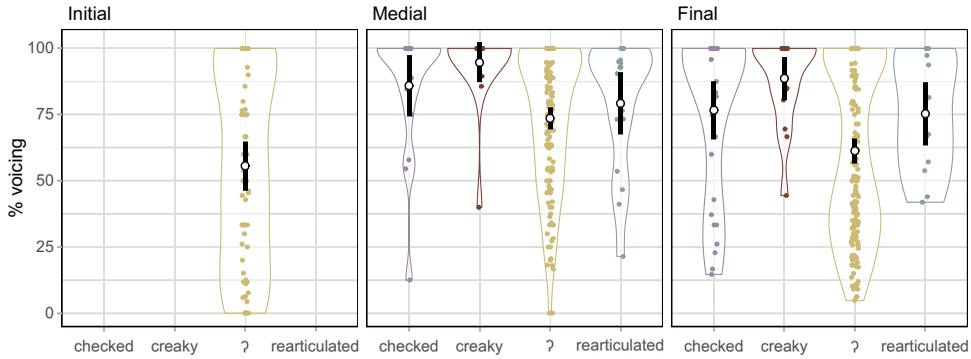


Figure 9 (Colour online) Percentage of voicing during glottalization for [ʔ] and different kinds of glottalized vowels, in the three phrasal positions. Each dot represents a token, and the violin plot represents concentration of tokens. The modeled means and 95% confidence intervals by position are shown in dark grey.

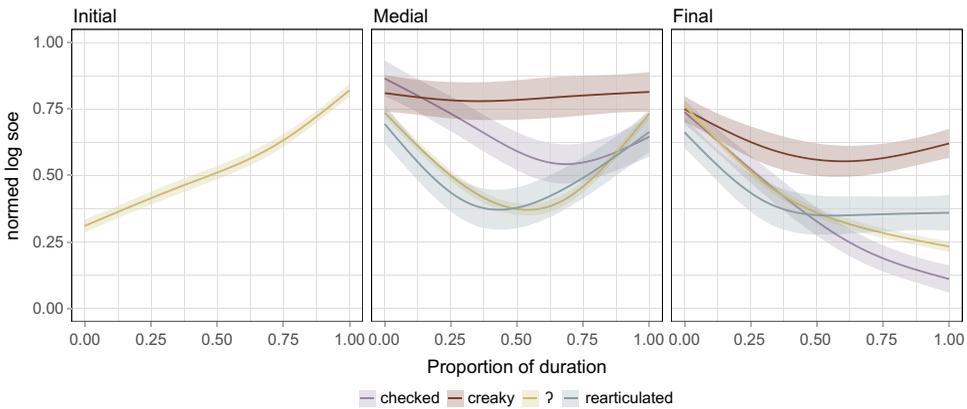


Figure 10 (Colour online) Time course of voicing intensity during glottalization for [ʔ] and different types of glottalized vowels, in the three phrasal positions. The values are estimated by cubic spline regression, with lines showing estimated means, and shaded areas showing one standard error above and below the estimated means.

intensity, though this drop is steepest for [h] and [ɦ]: the mean normed log SoE for [h] is 0.55 (95% CI: 0.51–0.58); for [ɦ] it is 0.61 (95% CI: 0.57–0.64); for breathy vowels it is 0.7 (95% CI: 0.66–0.74).

3.2 Results for glottalized sounds

The results for percentage voicing of glottalized sounds are shown in Figure 9. Glottal stops do not show much difference by position in utterance: although tokens in medial intervocalic position have higher percentages of voicing, tokens in initial and final positions still average at least 50% voicing throughout the glottalization. This finding, and how it contrasts with findings for aspirated sounds, is discussed more in Section 4.2 below. In non-initial positions, vowels labeled as ‘checked, creaky/laryngealized, rearticulated’ show more voicing than glottal stops, but little variation by phrasal position. But given the relatively small sample size of creaky vowels, future studies, ideally of languages with both glottal stops and creaky vowels, are needed to confirm this.

The results for voicing intensity for glottalized sounds are shown in Figure 10. Voicing intensity for [ʔ] and creaky vowels does not vary much by phrasal position. We also see that

vowels labeled as ‘creaky/laryngealized’ have stronger voicing than glottal stops or other glottalized vowels; e.g. the mean normed log SoE of word-medial creaky vowels is 0.63 (95% CI: 0.58–0.69) vs. for word-medial glottal stops the mean is 0.51 (95% CI: 0.49–0.54). The timecourses of voicing intensity show that glottal stops generally have periods of weaker and stronger voicing; in contrast, vowels labeled as ‘creaky/laryngealized’ have more temporally stable voicing intensity with flatter slopes. As expected, ‘rearticulated’ vowels have an earlier drop in voicing intensity than ‘checked’ vowels; unexpectedly, the drop in voicing intensity is stronger for medial rearticulated vowels than for medial checked ones. In utterance-final position, checked vowels have a similar SoE trajectory to final postvocalic glottal stops, whereas rearticulated vowels have a shallower drop in voicing intensity that plateaus over the latter half of the vowel.

4 Discussion

4.1 Glottal fricatives and breathy vowels

The results showed that voiceless [h] and voiced [ɦ] appear to differ very little in their voicing, and any differences are likely to be significant only in utterance-initial positions. Initial prevocalic [h] has slightly less voicing, and is more weakly voiced, than [ɦ] in that position. In initial position, the slight differences in voicing intensity between voiceless vs. voiced glottal fricatives are strongest during the second half of aspiration; during the first half, both sounds are likely to be weakly voiced or voiceless. Between two vowels, both glottal fricatives are as strongly voiced as breathy vowels; in final postvocalic position, [h] is less voiced than [ɦ] or breathy vowels, but most tokens of [h] are nonetheless voiced for over 50% of their duration. In final position, both glottal fricatives and breathy vowels show similar weakening in their voicing intensity. But while the magnitude of voicing at the phrase edges does not appear to be distinct for [h] vs. [ɦ] vs. (utterance-final) breathy vowels, potential differences in the slope of the voicing intensity over the course of the segment is worth exploring in future work, particularly for languages that contrast these sounds.

These results help address the question posed by Passy (1900) cited at the beginning of our paper: we find little evidence that voiceless [h] and voiced [ɦ] are distinct – at least in terms of presence and strength of voicing – between two vowels and in final postvocalic position. This is illustrated in Figure 11, which shows two tokens (one initial, the other intervocalic) of voiceless [h] and voiced [ɦ] in two related languages: Upper Sorbian (which has /h/) and Czech (which has /ɦ/). In initial position, /h/ is mostly voiceless whereas /ɦ/ is mostly voiced, though the voicing increases steadily in intensity. However, between two vowels both sounds are consistently voiced.

Among Illustrations, five of the six languages that contrast [h] and [ɦ] do so only word-initially in the papers; Illustrations tend to be structured in ways that highlight consonantal contrasts in initial position. The only language illustrating a non-initial contrast between phonemic /h/ and /ɦ/ is Basaa, but the Illustration provided no examples of the contrast in intervocalic or final postvocalic position. (There was one pair of post-vocalic /h/ and /ɦ/ followed by sonorants, and both are voiced throughout.) Nevertheless, some languages that do not yet have Illustrations, particularly Nguni languages like Zulu, do in fact contrast /h/ vs. /ɦ/ between vowels, as part of a broader set of ‘depressor’ consonant contrasts. Thus, if /h/ undergoes voicing between vowels, as we would predict, the contrast could presumably still be maintained if the depressor contrast is phonetically multidimensional, involving f_0 among other parameters (Traill, Khumalo & Fridjhon 1987, Downing 2018). Research investigating the correlates of intervocalic /h/ vs. /ɦ/ in languages that contrast these two sounds would therefore be particularly insightful to this question.

As we note in Section 2, for some languages we chose to change the symbols for aspirated sounds from those that were described. We do not believe that these changes would have a

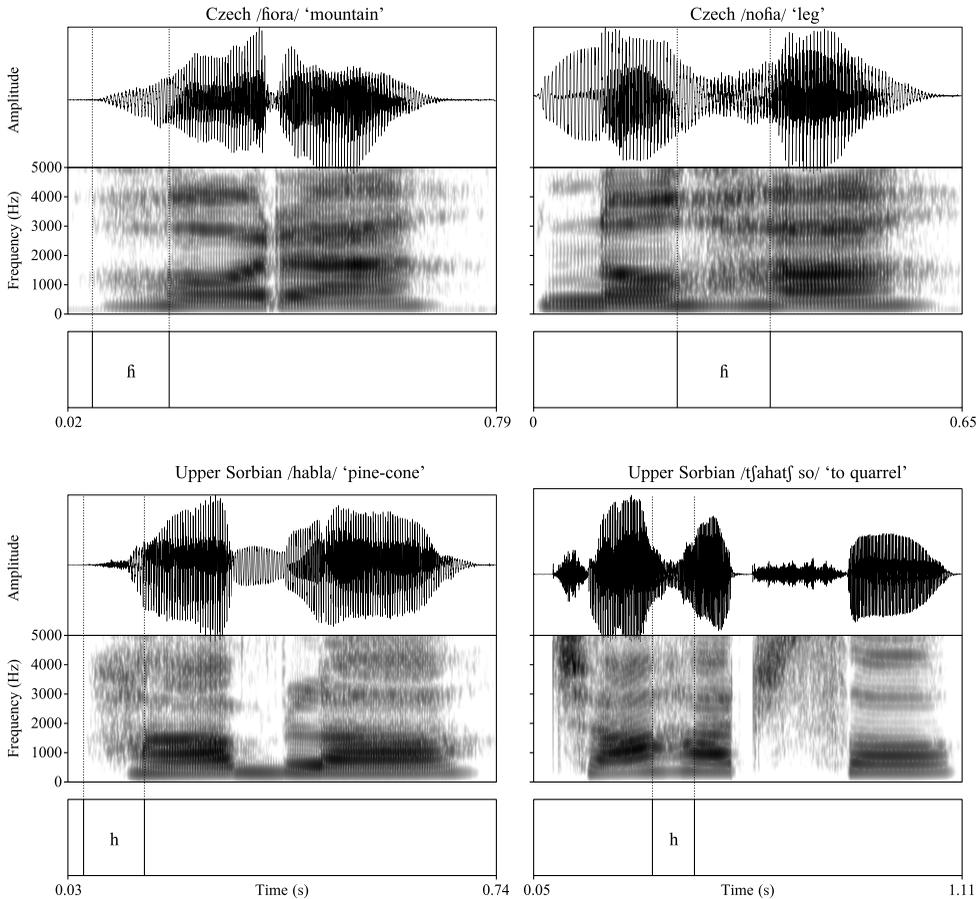


Figure 11 Examples of /f̥/ from a speaker of Czech (Dankovičová 1997) and /h/ from a speaker of Upper Sorbian (Howson 2017), two West Slavic languages. Both sounds are voiced between two vowels.

substantial influence on our results. One transcription change is that in the few Illustrations where intervocalic voicing of /h/ → [f̥] was described, such tokens were considered here to be [h]. As shown in Figure 7 above, the vast majority of tokens of intervocalic [h] are voiced for greater than 50% of their duration. Thus, it is likely that this ‘rule’ of intervocalic voicing applies to most languages, as a form of ‘automatic phonetics’ (Kingston & Diehl 1994) resulting from high subglottal pressure.

Another change was that, in keeping with the in-text description and later work on the language by the same author, Itunyoso Trique /h/ was considered here to be [f̥]. This may in principle affect the results for final position, as the majority of tokens of utterance-final [f̥] in this dataset come from that language. Note, however, that if Trique were considered to have [h], then any slight differences that could be inferred between final [h] vs. [f̥] (in terms of percentage voicing) would be minimized even further. Indeed, in Section 4.4 we discuss how, for many languages, the analysis probably need not specify voicing during aspiration at all.

4.2 Glottal stops and (different types of) creaky vowels

In no phrasal position are glottal stops fully voiceless; on average, glottal stops are characterized by voicing for about 50–75% of the glottalization interval, with a somewhat higher

percentage of voicing in intervocalic position compared to initial or final positions. The results reveal another different pattern for glottal stops compared to glottal fricatives: in every position analyzed, glottal stops show a period of weaker voicing: initial prevocalic [ʔ] begins voiceless or weakly voiced, but shows a gradual rise in voicing intensity throughout the glottalization; intervocalic [ʔ] begins and ends with strong voicing, but the middle portion of the glottalization shows a marked drop in voicing intensity leading to full glottal occlusion; final post-vocalic [ʔ] begins strongly voiced and ends with weaker voicing. In all positions, [ʔ] shows a wide range of voicing, with some tokens being mostly voiced, other tokens mostly voiceless, and many tokens partially voiced. These results confirm the impression of Ladefoged & Maddieson (1996: 75) that glottal stops are often voiced, but they further show that glottal stops are partially voiced in ALL positions, and not especially between vowels.

An interesting difference between the two ‘voiceless’ glottals, [ʔ] and [h], is that voicing during [ʔ] weakens intervocalically, but not substantially for [h]. We suspect that this is due to the physiological mechanism involved in the production of glottal stops: though their degree of constriction varies on a continuum, glottal stops generally involve constriction of supraglottal structures resulting in the postero-anterior narrowing of the epilaryngeal tube (Garellek 2013, Esling et al. 2019: Section 2.3.5). This laryngeal constriction facilitates voicelessness by helping to keep the vocal folds adducted. In contrast, [h] has spread vocal folds and abducted surrounding articulators, which should favor voicing by keeping the supraglottal pressure low (Westbury 1983). Therefore, the voicelessness during the onset of utterance-initial prevocalic [h] and offset of utterance-final [h] might be a result of low subglottal pressure, rather than the spreading gesture itself: subglottal pressure is lowest adjacent to a breath, i.e. at utterance edges (Ladefoged 1963), and low subglottal pressure favors voicelessness (Westbury & Keating 1986). Elsewhere, where the subglottal pressure is higher, glottal fricatives are strongly voiced. This implies that vocal fold spreading for aspiration does not, by itself, cause a substantial weakening of voicing (see Garellek 2020 for similar results for breathy vowels in !Xóǝ). This has important implications for the representation of voicing in [ʔ] and [h], which we will return to in Section 4.4 below.

The results also reveal a difference between creaky vowels labeled as ‘creaky/laryngealized’ compared to those labeled as ‘rearticulated, glottalized’, or ‘checked’: whereas all such categories are largely voiced, creaky vowels show little drop in voicing intensity, yet the other types behave as glottal stops with respect to their patterns of voicing intensity. Rearticulated vowels have an earlier drop in voicing than checked vowels, consistent with constriction that is phased earlier in the rearticulated vowel. In final position, checked vowels and glottal stops have similar phasing of the drop in voicing intensity. Thus, constricted vowels labeled as ‘creaky’ or ‘laryngealized’ in Illustrations tend to be constricted in a way that does not strongly influence voicing intensity; those labeled as ‘rearticulated, glottalized’, or ‘checked’ behave like glottal stops that occur at earlier vs. later points during the vowel, in that they weaken voicing. This distinction is close to that made by Ladefoged & Maddieson (1996: 48), only between ‘stiff’ voice vs. ‘creaky/laryngealized’ voice: they characterize the former as having slightly lower airflow than modal voice, but the latter as having ‘considerably’ lower airflow. Overall then, the results show a phonetic difference between ‘creaky’ vowels that are weakly creaky throughout the vowel, compared to those which, like glottal stops, tend to devoice during the final half of the vowel. The implications of this are discussed in more detail in the following sections. But it should also be noted that, among Illustrations, the languages with vowels labeled as ‘creaky’ additionally have at least one other category of vowels that are either checked (Isthmus Zapotec, Burmese) or rearticulated (Hanoi Vietnamese). Consequently, the difference in voicing intensity found for creaky vowels vs. other constricted categories in this sample might also reflect pressures of contrast maintenance for these particular speakers/languages. It would be worthwhile then to determine whether stronger voicing during creaky voice (meaning

Table 4 Illustrations of the IPA with [h ? ɦ] or non-modal vowels. Counts for the five most illustrated language families are also provided, as well as for languages in LAPSyD (Maddieson et al. 2014–2020). Percentages within each cell are relative to total number of languages (in the second column).

	Number of languages	Number (%) with [h]	Number (%) with [ʔ]	Number (%) with [ɦ]	Number (%) with non-modal vowels
LAPSyD	605	376 (62%)	272 (45%)	14 (2%)	12 (2%)
Illustrations:	201:	131: (65%)	89: (44%)	22: (11%)	15: (7%)
Indo-European	82	42 (51%)	23 (28%)	11 (13%)	1 (1%)
Austronesian	22	20 (91%)	18 (82%)	0 (0%)	0 (0%)
Niger-Congo	17	9 (53%)	7 (41%)	4 (24%)	1 (6%)
Sino-Tibetan	18	13 (72%)	10 (56%)	5 (28%)	5 (28%)
Afro-Asiatic	11	10 (91%)	6 (55%)	1 (9%)	0 (0%)

weaker constriction) is generally found in languages with creaky vowels, even when those creaky vowels do not contrast with another category that has laryngeal constriction.

4.3 Improving phonetic description of glottal consonants and non-modal vowels

Our analysis of Illustrations of the IPA also reveals several ways in which they can be improved as sources of phonetic documentation and description, particularly with regard to glottal consonants and non-modal vowels.

Between 1989 and 2020, there were 201 published Illustrations (including those which appeared in Part 2 of the *Handbook of the IPA*). Among these, the percentage of languages with either [h] or [ʔ] is comparable to that in LAPSyD (Maddieson et al. 2014–2020); see Table 4. However, [ɦ] and non-modal vowels are more prevalent in Illustrations of the IPA than in LAPSyD. Compared with a database like LAPSyD then, Illustrations have one drawback – the over-representation of Indo-European, especially Germanic, languages – and one benefit, namely the fact that Illustrations contain more phonetic detail.

As Whalen, DiCanio & Dockum (published online 28 July 2020) and Baird, Evans & Greenhill (published online 7 June 2021) have recently described, Illustrations are currently heavily biased in terms of the languages' genetic affiliations: nearly three quarters of languages with an Illustration belong to just five families: Indo-European (41% of Illustrations), Austronesian (11%), Niger-Congo (9%), Sino-Tibetan (9%), and Afro-Asiatic (5%). (In LAPSyD, these same language families represent the following percentages of the entire database: Indo-European (7%), Austronesian (8%), Niger-Congo (7%), Sino-Tibetan (3%), and Afro-Asiatic (6%.) Additionally, one third of the Indo-European languages with Illustrations come from the Germanic branch. Clearly then, Germanic languages, and Indo-European ones more generally, are over-represented among Illustrations of the IPA. The prevalence of [ɦ] among languages with Illustrations is probably due to the over-representation of Indo-European, particularly Germanic, languages: half of the languages with [ɦ] are Indo-European, and six of those are Dutch varieties or related languages like Afrikaans and Hamont Limburgish. Moreover, Illustrations with [ɦ] are found for languages belonging by and large to the five most commonly illustrated language families, in which [ɦ] is over-represented relative to LAPSyD. Thus, in our analysis of [ɦ], the sound is both genetically and geographically more restricted than the voiceless glottal consonants; see also Figure 6. However, non-modal vowels occur more often in Illustrations of the IPA than in LAPSyD. This is because the Illustrations (usually) contain more phonetic detail. Many of the non-modal vowels described in Illustrations have breathy or creaky voice as secondary correlates of other contrasts, including tones (e.g. Hanoi Vietnamese), registers (e.g. Bai, Mah Meri) or ATR contrasts (Shilluk); such sub-phonemic detail tends to be under-described

in phonological inventories, where the focus is on the (assumed) primary correlate of a category. It should also be noted here that more than half of the Illustrations describing non-modal vowels are for languages that belong to families other than the five most commonly illustrated ones. Therefore, greater diversity of Illustrations can only help improve our understanding of how non-modal phonation is used across spoken languages.

There are additional ways of improving coverage of non-modal vowels and glottal consonants in Illustrations. While reviewing sound files for this study, non-modal phonation that appeared to us to occur predictably with certain tones and at phrase edges frequently went undescribed in the Illustration. We also regularly encountered glottal stops that occurred predictably before vowel-initial words, at the ends of vowel-final words, between vowels, as well as before implosives and coda stops. (Wanting to conform to the transcriptions and descriptions provided by the Illustration authors, we ignored such cases for the present study.) For instance, glottal stops are described in only four of the eight Illustrations of varieties of English, even though vowel-initial glottalization and coda-stop glottalization are widespread phenomena across World Englishes (see e.g. Foulkes & Docherty 1999; Docherty, Hay & Walker 2006; Garellek 2014; Penney, Cox & Szakay, published online 16 April 2019; Seyfarth & Garellek 2020). Ideally, Illustration authors should describe these for improved phonetic documentation and description of laryngeal events.

The results of this study also have important implications for the ways in which glottal consonants are transcribed. First, researchers should assume, and therefore need not necessarily transcribe, voicing of [ʔ h]. This is particularly true in non-initial positions. But a language with non-initial /h/ that is VOICELESS throughout its duration would be of interest to researchers, because the /h/ might be voiceless due to supraglottal constriction. Glottal stops in initial and final positions show a gradual rise and fall, respectively, in voicing intensity. Between vowels, glottal stops show a dip in voicing intensity, but generally voicing does not weaken to the levels found at the extreme edges of utterances. Therefore, a language in which glottal stops are generally entirely voiceless – that is, a language in which glottal stops are characterized by silent closure, surrounded by modal voicing – would be of interest, as these would also suggest supraglottal constriction and/or extra-long duration, and a (phonetic) specification for voicelessness.

Second, we find insufficient evidence that voiceless [h] and voiced [ɦ] differ in terms of their voicing, especially in non-initial word positions. What transcription should the researcher use for a language that has glottal fricatives that only occur in non-initial positions? We would suggest labeling this sound as [h], because our results show that this sound is expected to be voiced in these environments, and because evidence for the existence of voiced [ɦ] is most robust in utterance-initial prevocalic position. At the very least, positing voiced [ɦ] in non-initial positions should not be motivated exclusively by the presence of voiced aspiration. We will return to this point in the following section, which concerns the phonological implications of this work.

Our study has also made clear how the phasing of non-modal phonation with respect to the segment matters (Silverman 1995, Avelino 2016, Esposito & Khan 2020), because it is often related to the strength of laryngeal constriction: vowels labeled as ‘checked’ have similar changes in voicing intensity contours as V+[ʔ] sequences, and word-medially ‘rearticulated’ vs. ‘checked’ vowels show similar changes in voicing as intervocalic glottal stops (though with different phasing). Finally, vowels that are labeled as ‘creaky’ or ‘laryngealized’ tend to have stronger voicing than checked/rearticulated vowels or glottal stops. In all then, it would be beneficial for Illustration authors to describe both the strength of voicing during non-modal phonation (consistently voiced vs. with a tendency to devoice) and its phasing. However, we also recognize that adopting conventions for TRANSCRIBING these differences using the IPA poses certain challenges (see e.g. Keating et al. 2021 for relevant discussion). One possibility is to use the following conventions, some of which are already occasionally used and others which represent a return to previously used conventions discussed in Section 1.3 above:

- (1) Non-modal phonation that has relatively strong non-modal voicing throughout the target segment (e.g. breathy vowels in Luanyjang Dinka, and non-rearticulated ‘laryngealized’ vowels in Tilquiapan Zapotec): [V̤] and [V̥].
- (2) Non-modal phonation that is strongest early in the segment (e.g. ‘pre-glottalized’ consonants, or vowels with early-phased breathy and creaky voice, e.g. Jalapa Mazatec; Garellek & Keating 2011): [ʰV~hV̤] and [ʰV~ʰV̤]. If the language has post-aspirated or post-glottalized consonants, the use of superscripts can be ambiguous, and so tie bars might be preferred.
- (3) Non-modal phonation that is strongest late in the segment (e.g. ‘checked’ vowels or tones found in many Southeast Asian and Mesoamerican languages, as well as Danish *stød*): [V̤h~V̤h̄] and [V̤ʰ~V̤ʰ̄].
- (4) Non-modal phonation that is strongest partway through the segment (as in breathy vowels of Gujarati, ‘rearticulated’ vowels in Zapotecan languages, or the ‘broken’ tone of Hanoi Vietnamese): [V̤ʰV̤] and [V̤ʰ̄V̤]. The latter could also be transcribed as [V̤̣], using a diacritic not present in the IPA but which follows a convention formerly used for laryngealized vowels in Mesoamerican languages, e.g. Aschmann (1946) for Totonac.

The description of breathy and creaky voice can further be improved by making reference to articulatory-acoustic subtypes (Keating, Garellek & Kreiman 2015, Garellek 2019, Tian & Kuang 2021), some of which can be ascertained directly from the waveform and spectrogram. Here too, our current understanding of articulatory-acoustic differences within major groupings of phonation types is well ahead of our conventions used for phonetic transcription.

In sum, both phonetic description and our understanding of phonetic diversity as it concerns glottal and laryngeal sounds can only be improved by more illustrations of more genetically and geographically diverse languages, and with more attention paid to the presence, quality, and phasing of laryngeal events in the speech signal.

4.4 Approaches to phonological analysis of glottal consonants and non-modal vowels

This study confirms that, for many languages and speakers, voiceless glottal consonants are frequently partially voiced, and breathy and creaky vowels are frequently partially devoiced. Strikingly, transcribed [h] is USUALLY mostly voiced over its duration, and strongly voiced, except in utterance-initial position. Thus, in Section 4.2, we argued that aspiration does not by itself cause a substantial weakening of voicing; voiceless aspiration tends to be predictable, being found at utterance edges where the subglottal pressure is low.

Overall then, much of this variation is expected from utterance position and prosodic modulation. For reasons relating to subglottal pressure and voicing initiation, utterance-initial position favors voicelessness, whereas other positions (except for the very end of an utterance-final sound) favor voicing. A strong prosodic position anywhere in the utterance can lead to devoicing, owing to the increased gestural magnitude of both glottal consonants and non-modal vowels.

This poses challenges for segmental phonetic transcription, as we have described in the previous section. And given that the phonetic signal is often used to motivate analytical decisions, such voicing variation can also lead to challenges in establishing a phonological analysis. Researchers wishing to analyze the phonological status and behavior of glottal consonants and non-modal vowels must deal with two phonetic realities. The first is that voiceless glottal sounds will often be voiced, and non-modal phonation will often be partially voiceless; consequently, it can be problematic to base a phonological analysis of glottal sounds primarily on the presence vs. absence of voicing in the phonetic signal. The second is that glottal sounds, even if they pattern phonologically as segments, are never clearly segmental, even taking into account the broader issues with analyzing any sound as a segment (e.g. Ladd 2014: Chapter 2). This is because glottal ‘consonants’ and non-modal ‘vowels’ are

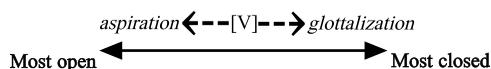


Figure 12 A modified Continuum model with only spreading and constriction specified, instead of quantized targets.

both modulations of PHONATION, rather than modulations of the VOCAL TRACT that occur downstream from laryngeal phonation (Catford 1964, Ladefoged 1990, Gordon & Ladefoged 2001, Esling et al. 2019, Garellek 2019).

The fact that glottal sounds are modulations of phonation can lead to difficulties regarding representation, especially when determining if they are segments or suprasegments, and if glottal consonants are consonantal. For instance, whether to treat surface glottal sounds as segments or suprasegments is widely debated within Oto-Manguean languages (Campbell 2017). To some extent, such analytical issues might be mitigated if evidence gleaned from the phonetic signal is reduced; that is, if the analyst has a prior expectation that modulations of phonation will vary in magnitude, regardless of the phonological analysis or the phonetic symbol that is ultimately chosen.

We argue that, when first approaching a phonetic or phonological analysis of glottal sounds in a language, it can be beneficial to treat them not as segments, but as two phonatory gestures: spreading and constriction (see also related discussion from Pierrehumbert & Talkin 1992: 116). It is further beneficial to assume that these gestures have underspecified magnitudes, at least at the initial stages of an analysis. Such a representation is illustrated using a modified Continuum model shown in Figure 12. In this model, we specify spreading (‘aspiration’) and constriction (‘glottalization’) relative to a modal voicing baseline, but leave these gestures unspecified as to their magnitude. Thus, any sound with aspiration would minimally involve only an increase in vocal fold spreading. If the sound is realized as voiceless, then the spreading gesture was produced with greater magnitude. If the sound is realized instead as breathy voicing, the spreading gesture was produced with lesser magnitude.

According to this model then, a glottal sound is phonetically specified for aspiration or glottalization, as well as the phasing of the gesture relative to adjacent ones, but not necessarily for the magnitude of spreading or constriction gestures. For instance, a vocalic token produced with aspiration phased to its midpoint might (at least initially) be considered a spreading gesture flanked by vocalic ones. It may be realized and analyzed as any one of the following: a breathy vowel [V̤], a voiceless or breathy-voiced rearticulated vowel [V^hV] or [V^hV̤], an intervocalic voiceless glottal fricative [VhV], or an intervocalic voiced glottal fricative [Vh̥V]. Specifying the magnitude of the spreading gesture – whether the aspiration is voiceless or voiced – might be relevant if the researcher wishes to make claims about phonetic patterns that are not PREDICTABLE. Such patterns might include intervocalic aspiration that is usually voiceless, or word-initial aspiration that is usually voiced. Specifying the magnitude of a glottal gesture would also be relevant if the language has a contrast in gestural magnitude (e.g. a contrast between voiced vs. voiceless aspiration), or if the gesture patterns phonologically as a voiced segment along with other voiced segments. But in the absence of contrasts or phonological patterning, arguments in favor of a particular phonological representation are not on very solid ground if they are rooted in the presence of voicing, especially if the role of respiratory and prosodic factors is ignored.

A concrete example is that of English /h/. Linguists usually transcribe this sound as being underlyingly a voiceless glottal fricative. But in fact, its voicing is predictable: owing to low subglottal pressure utterance-initially and initial strengthening, /h/ tends to be voiceless domain-initially; elsewhere it tends to be voiced. And to our knowledge there is no phonological evidence for treating English /h/ as uncontroversially voiceless (or voiced); see Halle & Stevens (1971), Keating (1990), Avery & Idsardi (2001) vs. Davis & Cho (2003) for treatments of /h/ as underspecified for voicing vs. voiceless, respectively. Considering these phonetic and phonological facts, English /h/ is consistent with being merely vocal fold spreading or ‘aspiration’, for which the voicing characteristics (and thus also manner) are

phonetically and phonologically underspecified. Note that similar arguments were used by Keating (1988) to argue, based on the predictable formants of [h], that the sound is underspecified for place of articulation. There are also parallels with vocal fold and laryngeal constriction: Garellek (2014) argued that word-initial glottalization across languages is not clearly either a glottal stop target or a creaky voice target. Thus, word-initial glottalization is consistent with laryngeal constriction, of any degree. This also accords with the commonly held assumption that, across languages, stop pre-glottalization may be realized on a continuum of creaky voice to full glottal occlusion on a preceding voiced sound (Keating et al. 2021).

Researchers might therefore consider whether ‘aspiration’ and ‘glottalization’ – with no mention of their voicing – are sufficiently explanatory for analyzing glottal sounds, at least at a level more abstract than a narrow phonetic transcription. In fact, there already exist phonological frameworks that analyze glottal sounds accordingly (e.g. Halle & Stevens 1971, Avery & Idsardi 2001). Ultimately, the presence of voicing during a glottal sound might tell us little about its phonological status because many cases of laryngeal spreading and constriction appear to be simply that: aspiration or glottalization, of any degree.

5 Conclusions

In this study, we used recordings from *JIPA*'s Illustrations of the IPA to determine the extent to which voiceless glottal consonants differ from voiced ones and non-modal vowels. We find that initial and final voiceless [h] is less voiced than voiced [ɦ] – in terms of both percentage voicing during aspiration and strength of voicing – but that these differences are very slight in these positions; between two vowels, both [h] and [ɦ] are likely to be as voiced as breathy vowels. Glottal stops vary widely in the percentage voicing during glottalization, and generally have periods of strong and weak voicing. In contrast, creaky vowels are strongly voiced. Not all ‘creaky’ vowels are alike, however: vowels described as ‘rearticulated’, ‘checked’, or ‘glottalized’ show similar drops in voicing intensity to glottal stops. We interpret these results through an articulatory lens: glottal consonants and non-modal vowels are both modulations in phonation resulting from aspiration and glottalization at the phonatory source. In many languages, we argue that these could be described and analyzed as without reference to their voicing.

It is worth mentioning here that, in *JIPA* Illustrations, as with any cross-linguistic speech corpus, some language families are better represented than others, and some Illustrations have more recordings of words containing glottal consonants and non-modal vowels. The recordings also generally come from only one speaker. Therefore, all conclusions that we have drawn are possibly more representative of particular speakers’ idiolects than of the languages they speak. We also note that the diverse recording setups used, as well as the post-processing of audio files, could in principle affect the acoustic analysis. Unfortunately, details of the recording and audio are usually absent from Illustrations. Still, we believe that significant effects on the conclusions are unlikely, because only systematic errors, i.e. those which affect the data in a consistent manner, would lead to issues with generalizability of the acoustic results. Systematic errors arising from variable recording setups are unlikely; variable setups likely do not affect voicing percentage or intensity in a single direction (e.g. always weaker voicing), or for a particular category more strongly than the other (e.g. affecting [h] more than [ɦ]). Instead, the variable recording setups are much more likely to introduce unsystematic error, i.e. random noise, to the measures. In order to facilitate future phonetic analyses of audio recordings stemming from Illustrations, authors of *JIPA* Illustrations should therefore ideally provide information as to the recording setup and any post-processing of the audio that was performed.

Nevertheless, there are advantages for using Illustrations for initial exploratory studies like the present one. The recordings which accompany Illustrations are of consistently good

quality, and their transcriptions are peer-reviewed; they often include minimal or close pairs of words that help contrast sounds of interest; words appear in citation form, so target sounds are likely to be maximally distinct from others. Finally, the author often chooses speakers for an Illustration because their speech is deemed ‘representative’ of the larger language community. On the one hand then, we appeal to phoneticians to continue to investigate – with larger data sets of more speakers and of more languages, with more quantification, and with greater detail to the phasing between articulatory gestures – the ways in which glottal consonants and non-modal vowels are realized in terms of their voicing. On the other hand, we also recognize the largely untapped potential that Illustrations of the IPA have for advancing knowledge about sounds of the world’s languages.

Acknowledgments

We are most grateful to the speakers who have lent their voices to Illustrations of the IPA, to the Illustration authors for their careful documentation, and to the *JIPA* editorial teams present and past. We also thank the International Phonetic Association for making the Illustration recordings accessible to researchers.

We also thank Stephanie Luong for her help with file annotation and segmentation. Earlier versions of this work were presented at UCLA, La Trobe, NYU, UPenn, and UCSD, and this paper has benefited greatly from discussions with audience members and other colleagues. In particular, we thank Eric Baković, Lisa Davidson, Janet Fletcher, Bruce Hayes, Pat Keating, Jianjing Kuang, Mark Liberman, Sharon Rose, Stephanie Shih, Will Styler, and Marija Tabain for their helpful comments. We are also grateful to Adam Chong, Christina Esposito, and Sameer ud Dowla Khan for providing feedback on the original manuscript, and to *JIPA* Jody Kreiman, Associate Editor Oliver Niebuhr, Míša Hejtná, and an anonymous reviewer, for their helpful suggestions during the review process. Finally, we thank *JIPA* copy-editor Ewa Jaworska for suggested improvements to the final version.

References

- Ariza-García, Andrea. 2018. Phonation types and tones in Zapotec languages: A synchronic comparison. *Acta Linguistica Petropolitana* 14, 485–517.
- Aschmann, Herman. 1946. Totonaco phonemes. *International Journal of American Linguistics* 12, 34–43.
- Ashby, Michael & Joanna Przedlacka. 2014. Measuring incompleteness: Acoustic correlates of glottal articulations. *Journal of the International Phonetic Association* 44, 283–296.
- Avelino, Heriberto. 2016. Phonetics in phonology: A cross-linguistic study of laryngeal contrast. In Heriberto Avelino, Matt Coler & W. Leo Wetzels (eds.), *The phonetics and phonology of laryngeal features in Native American languages*, 157–179. Leiden & Boston, MA: Brill.
- Avery, Peter & William J. Idsardi. 2001. Laryngeal dimensions, completion and enhancement. In T. Alan Hall (ed.), *Distinctive feature theory*, 41–70. Berlin & New York: Mouton de Gruyter.
- Baird, Louise, Nicholas Evans & Simon J. Greenhill. Blowing in the wind: Using ‘North Wind and the Sun’ texts to sample phoneme inventories. *Journal of the International Phonetic Association*, <https://doi.org/10.1017/S002510032000033X>. Published online by Cambridge University Press, 7 June 2021.
- Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer. Version 6.1.16. <http://www.praat.org/>. [Computer program]
- Campbell, Eric W. 2017. Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass* 11(e12240), 1–22.
- Catford, J. C. 1964. Phonation types: The classification of some laryngeal components of speech production. In David Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott & J. L. M. Trim (eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*: Special issue of *International Journal of American Linguistics* 32, 26–37.
- Catford, J. C. 1990. Glottal consonants . . . another view. *Journal of the International Phonetic Association* 20, 25–26.

- Chirkova, Katia & Yiya Chen. 2013. Xumi (part 1): Lower Xumi, the variety of the lower and middle reaches of the Shuiluo river. *Journal of the International Phonetic Association* 43, 363–379.
- Cho, Taehong & Patricia [A.] Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37, 466–485.
- Chong, Adam J., Megan Risdal, Ann Aly, Jesse Zymet & Patricia [A.] Keating. 2020. Effects of consonantal constrictions on voice quality. *The Journal of the Acoustical Society of America* 148, EL65.
- Dankovičová, Jana. 1997. Czech. *Journal of the International Phonetic Association* 27, 77–80.
- Davis, Stuart & Mi Hui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics* 41, 607–652.
- Deme, Andrea, Márton Bartók, Tekla Etelka Grácsi, Tamás Gábor Csapó & Alexandra Markó. 2020. Intervocalic voicing of Hungarian /h/. In Veronika Hegedűs & Irene Vogel (eds.), *Approaches to Hungarian*, vol. 16: *Papers from the 2017 Budapest Conference*, 48–71. Amsterdam & Philadelphia, PA: John Benjamins.
- DiCanio, Christian T. 2010. Itunyoso Trique. *Journal of the International Phonetic Association* 40, 227–238.
- DiCanio, Christian [T]. 2014. Cue weight in the perception of Trique glottal consonants. *The Journal of the Acoustical Society of America* 119, 3059–3071.
- Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24, 423–444.
- Docherty, Gerry [Gerard J.], Jen Hay & Abby Walker. 2006. Sociophonetic patterning of phrase-final /t/ in New Zealand English. In Paul Warren & Catherine I. Watson (eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, Auckland, 378–383.
- Downing, Laura J. 2018. Grounding Nguni depressor effects. In Wolfgang Kehrein, Björn Köhnlein, Paul Boersma & Marc van Oostendorp (eds.), *Segmental structure and tone*, 109–146. Berlin: de Gruyter Mouton.
- Elias-Ulloa, Jose & Rolando Muñoz Aramburú. 2021. Upper-Chambira Urarina. *Journal of the International Phonetic Association* 51, 137–169.
- Elliott, A. Raymond, Jerold A. Edmondson & Fausto Sandoval Cruz. 2016. Chicahuaxtla Triqui. *Journal of the International Phonetic Association* 46, 351–365.
- Esling, John H. 2010. Phonetic notation. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 678–702. Oxford: Wiley-Blackwell.
- Esling, John H., Scott R. Moisik, Allison Benner & Lise Crevier-Buchman. 2019. *Voice quality: The laryngeal articulator model*. Cambridge: Cambridge University Press.
- Esposito, Christina M. & Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14, e12392.
- Faytak, Matthew, Jeremy Steffman & Rolain Tankou. 2020. True voiced aspirates in Yemba. Poster presented at LabPhon17.
- Firth, J. R. 1957. Phonetic observations on Gujarati. *Bulletin of the School of Oriental and African Studies* 20, 231–241.
- Firth, J. R. 1961. Phonological features of some Indian languages. In *Papers in linguistics, 1934–1951*, 47–53. Oxford: Oxford University Press.
- Fougeron, Cécile. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29, 109–135.
- Fougeron, Cécile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101, 3728–3740.
- Foulkes, Paul & Gerard [J.; Gerry] Docherty (eds.). 1999. *Urban voices: Accent studies in the British Isles*. London: Arnold.
- Garellek, Marc. 2013. *Production and perception of glottal stops*. Ph.D. dissertation, UCLA.
- Garellek, Marc. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45, 106–113.
- Garellek, Marc. 2015. Perception of glottalization and phrase-final creak. *The Journal of the Acoustical Society of America* 137, 822–831.

- Garellek, Marc. 2019. The phonetics of voice. In William Katz & Peter Assmann (eds.), *Routledge handbook of phonetics*, 75–106. Oxford: Routledge.
- Garellek, Marc. 2020. Acoustic discriminability of the complex phonation system in !Xóõ. *Phonetica* 77, 131–160.
- Garellek, Marc & Patricia [A.] Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41, 185–205.
- Garellek, Marc, Patricia [A.] Keating, Christina M. Esposito & Jody Kreiman. 2013. Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America* 133, 1078–1089.
- Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29, 383–406.
- Grønnum, Nina. 1998. Danish. *Journal of the International Phonetic Association* 28, 99–105.
- Groth, Peter Olsen. 1894. *A Danish and Dano-Norwegian grammar*. Boston, MA: D. C. Heath & Co.
- Gutiérrez, Analía. 2019. Nivaçle (*shichaam lhavos* variety). *Journal of the International Phonetic Association* 49, 401–417.
- Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *Quarterly Progress Report: MIT Research Laboratory of Electronics* 101, 198–212.
- Hargus, Sharon & Virginia Beavert. 2014. Northwest Sahaptin. *Journal of the International Phonetic Association* 44, 319–342.
- Howson, Phil. 2017. Upper Sorbian. *Journal of the International Phonetic Association* 47, 359–367.
- Huffman, Marie K. 1987. Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America* 81, 495–504.
- Hutters, Birgit. 1985. Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42, 1–24.
- Iivonen, Antti. 1992. How [h] became a glottal fricative: Comments on the letter by Martin Kloster-Jensen in *JIPA* 21 (1991). *Journal of the International Phonetic Association* 22, 61–64.
- IPA [International Phonetic Association] (ed.). 1902. Alphabet phonétique international: employé dans le maître phonétique. *Le Maître Phonétique* 17, 156–158.
- IPA [International Phonetic Association]. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 5, 275–292.
- Keating, Patricia A. 1990. Phonetic representations in a generative grammar. *Journal of Phonetics* 18, 321–334.
- Keating, Patricia [A.], Taehong Cho, Cécile Fougeron & Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation* (Papers in Laboratory Phonology 6), 143–161. Cambridge: Cambridge University Press.
- Keating, Patricia [A.], Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow.
- Keating, Patricia [A.], Daniel Wymark & Ryan Sharif. 2021. Proposal for superscript diacritics for prenasalization, preglottalization and preaspiration. *Journal of the International Phonetic Association* 51, 75–90.
- Khan, Sameer ud Dowla. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40, 780–795.
- Khatiwada, Rajesh. 2009. Nepali. *Journal of the International Phonetic Association* 39, 373–380.
- Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70, 419–454.
- Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41, 381–392.
- Kloster-Jensen, Martin. 1991. Letter to the Editor. *Journal of the International Phonetic Association* 21, 42.
- Kreiman, Jody & Diana Sidtis. 2011. *Foundations of voice studies*. Oxford: Wiley-Blackwell.
- Kruspe, Nicole & John Hajek. 2009. Mah Meri. *Journal of the International Phonetic Association* 39, 241–248.

- Ladd, D. Robert. 2014. *Simultaneous structure in phonology*. Oxford: Oxford University Press.
- Ladefoged, Peter. 1963. Some physiological parameters in speech. *Language and Speech* 6, 109–119.
- Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago, IL: University of Chicago.
- Ladefoged, Peter. 1990. Some proposals concerning glottal consonants. *Journal of the International Phonetic Association* 20, 24–25.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- Laufer, Asher. 1991. The 'glottal fricatives'. *Journal of the International Phonetic Association* 21, 91–93.
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of H&H theory. In W[illiam] J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–339. Dordrecht: Kluwer.
- Liu, Xinzhong & Maolin Wang. 2008. Zhōnggǔ chúnǚn zì zài hǎinán wénchāng huà zhōng dúyīn de yǔyīn xué fēnxī [Phonetic analysis of the pronunciation of Middle Chinese labial sounds in Hainan Wenchang dialect]. *Proceedings of the 8th Phonetics Conference of China and International Symposium on Frontiers of Speech Science in Honour of Mr. Wu Zongji's 100th Birthday*.
- Maddieson, Ian, Sébastien Flavie, Egidio Marsico & François Pellegrino. 2014–2020. LAPSyD: Lyon–Albuquerque Phonological Systems Databases, Version 1.1. <http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/>.
- Meyer, Ernst A. 1900. Stimmhaftes H. *Die neueren Sprachen: Zeitschrift für den Neusprachlichen Unterricht* 8, 261–283.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331, 176–182.
- Mielke, Jeff. 2009. Segment inventories. *Language and Linguistics Compass* 3, 700–718.
- Mielke, Jeff. 2018. Visualizing phonetic segment frequencies with density-equalizing maps. *Journal of the International Phonetic Association* 48, 129–154.
- Mittal, Vinay Kumar, B. Yegnanarayana & Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America* 136, 1932–1941.
- Mitterer, Holger. 2018. Not all geminates are created equal: Evidence from Maltese glottal consonants. *Journal of Phonetics* 66, 28–44.
- Murty, K. Sri Rama & B. Yegnanarayana. 2008. Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 1602–1613.
- Nabirye, Minah, Gilles-Maurice de Schryver & Jo Verhoeven. 2016. Lusoga (Lutenga). *Journal of the International Phonetic Association* 46, 219–228.
- Pandit, P. B. 1957. Nasalisation, aspiration and murmur in Gujarati. *Indian Linguistics* 17, 165–172.
- Passy, Paul. 1900. Review of Stimmhaftes H by E. A. Meyer. *Le Maître Phonétique* 15, 131.
- Penney, Joshua, Felicity Cox & Anita Szakay. Glottalisation of word-final stops in Australian English unstressed syllables. *Journal of the International Phonetic Association*, <https://doi.org/10.1017/S0025100319000045>. Published online by Cambridge University Press, 16 April 2019.
- Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II*, 90–117. Cambridge: Cambridge University Press.
- Pullum, Geoffrey K. & William A. Ladusaw. 1996. *Phonetic symbol guide*. Chicago, IL & London: University of Chicago Press.
- Schadeberg, Thilo. 1982. Nasalization in UMBundu. *Journal of African Languages and Linguistics* 4, 109–132.
- Schertz, Jessamyn & Sarah Khan. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics* 81, Article 100979. <https://doi.org/10.1016/j.wocn.2020.100979>.
- Seyfarth, Scott & Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71, 425–450.
- Seyfarth, Scott & Marc Garellek. 2020. Physical and phonological causes of coda /t/ glottalization in the mainstream American English of central Ohio. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11, 24, 1–33. <http://doi.org/10.5334/labphon.213>.

- Shryock, Aaron Michael. 1995. *Investigating laryngeal contrasts: An acoustic study of the consonants of Musey*. Ph.D. dissertation, UCLA.
- Shue, Yen-Liang, Patricia A. Keating, Chad Vicenik & Kristine Yu. 2011. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 1846–1849.
- Silverman, Daniel. 1995. *Phrasing and recoverability*. Ph.D. dissertation, UCLA. [Published in 1997 by Garland.]
- Šimáčková, Šárka, Václav Jonáš Podlipský & Kateřina Chládková. 2012. Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association* 42, 225–232.
- Smalley, William A. 1968. *Manual of articulatory phonetics*. Tarrytown, NY: Practical Anthropology.
- Sweet, Henry. 1874. On Danish pronunciation. *Transactions of the Philological Society* 15, 94–112.
- Teras, Pire. 2018. The phonetic variation of short intervocalic /h/ in Estonian. In Katarzyna Klessa, Jolanta Bachan, Agnieszka Wagner, Maciej Karpiński & Daniel Śledziński (eds.), *Proceedings of the 9th International Conference on Speech Prosody 2018*, Poznań, 883–887. [10.21437/SpeechProsody.2018-178](https://doi.org/10.21437/SpeechProsody.2018-178).
- Tian, Jia & Jianjing Kuang. 2021. The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association* 51(2), 202–228.
- Traill, Anthony. 2018. *A trilingual !Xóǀ dictionary*. Cologne: Rüdiger Köppe.
- Traill, Anthony, James S. M. Khumalo & Peter Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46, 255–274.
- Wells, J. C. 1975. The Association's alphabet. *Journal of the International Phonetic Association* 5, 52–58.
- Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America* 73, 1322–1336.
- Westbury, John R. & Patricia A. Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22, 145–166.
- Whalen, D. H., Christian DiCanio & Rikker Dockum. Phonetic documentation in three collections: Topics and evolution. *Journal of the International Phonetic Association*, <https://doi.org/10.1017/S0025100320000079>. Published online by Cambridge University Press, 28 July 2020.
- Woon, Wee-Lee. 1979. A synchronic phonology of Hainan dialect: Part I. *Journal of Chinese Linguistics* 7, 65–100.