Phonation type and Praatsauce

Christian DiCanio - cdicanio@buffalo.edu
Thea Knowles - theaknow@buffalo.edu

Department of Linguistics
Department of Communicative Disorders and Sciences

4/27/21

What is phonation type?

Phonation type, or voice quality, refers to the source signal produced by varying the relative constriction between the vocal folds while expelling air through the glottis.

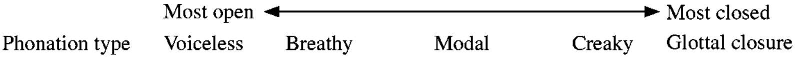
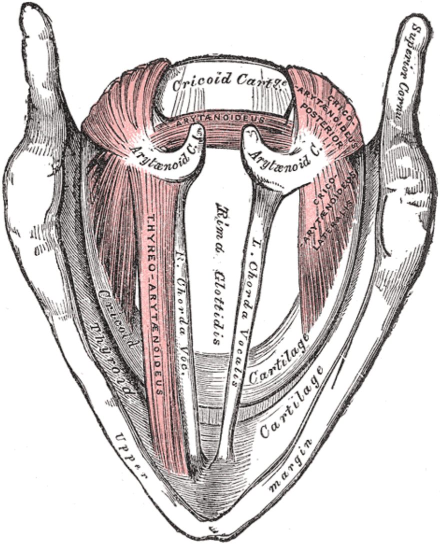


Figure 1. Continuum of phonation types (after Ladefoged, 1971).

By varying glottal aperture different voice qualities are produced (Gordon and Ladefoged, 2001).

Constriction occurs asymmetrically with the vocal folds, more at the anterior end near the thyroid notch than at the posterior end where the arytenoid cartilages lie.


Cricoid Cart.ge
ARYTÆNOIDEUS
CRICO-ARYTÆNOIDEUS POSTERIOR
Arytænoid C.
Arytænoid C.
Superior Cornu
CRICO-ARYTÆNOIDEUS LATERALIS
THYREO-ARYTÆNOIDEUS
R. Chorda Voc.
Rima Glottidis
L. Chorda Vocalis
Cricoid Cartilage
Thyroid Cartilage
Upper margin

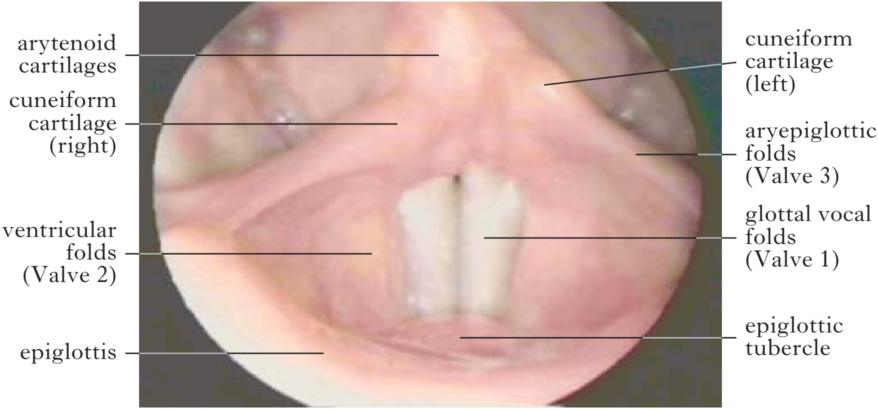
Constriction

In addition to changing glottal aperture, changes in tension also occur at the vocal folds. Generally speaking, phonation types with a narrower aperture involve greater medial compression of the vocal folds.

However, the longitudinal tension of the vocal folds may also vary with phonation types. We also control longitudinal tension for producing changes in F_0.

It is important to remember that we are referring to changes in the temporary resting state of the vocal folds here; they still open/close with voicing.

Modal voicing (Edmondson and Esling, 2006)



Breathy voicing (Edmondson and Esling, 2006)

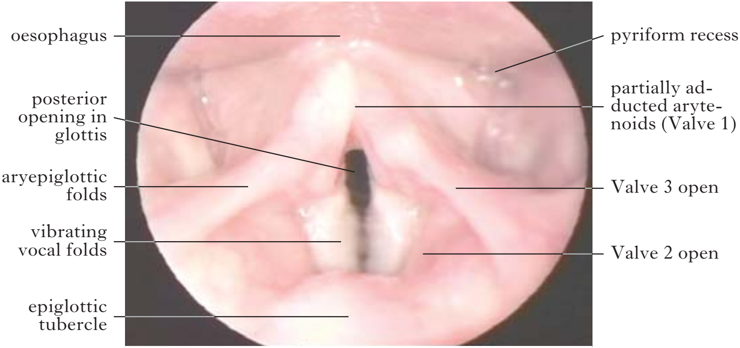
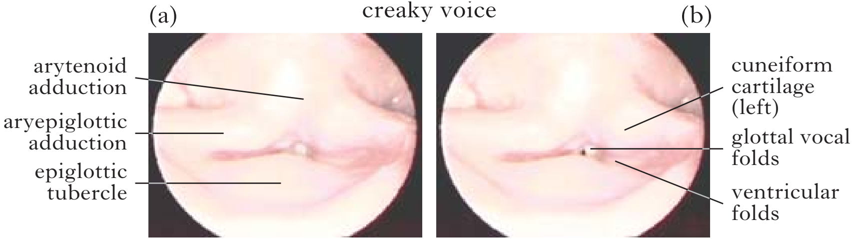


Figure 8

Canonical Breathy Voice showing a partial closure of Valve 1, leaving a large aperture between the arytenoid cartilages. The glottal vocal folds can oscillate anteriorly along the ligamental glottis, while remaining open posteriorly.

Creaky voicing (Edmondson and Esling, 2006)



Note the narrowing of the entire cavity here and the greater aryepiglottic constriction. Only the posterior portion of the vocal folds permits airflow and vocal fold vibration due to this greater overall constriction.

Uses of phonation type (c.f. Davidson (2020))

Phonation type can be used for linguistic contrast. There are languages which distinguish word meaning with phonation type changes (next slide). This is **contrastive voice quality**.

Phonation type can also occur as a result of coarticulation between voiceless and voiced speech segments. This is **coarticulatory voice quality** or allophonic voice quality. English final /t/ - [hɪt͡ʔ] often involves some creaky phonation on the vowel, [hɪ̰t͡ʔ].

Phonation type can also occur at prosodic boundaries. Creaky phonation is common in English at the ends of phrases as speakers trail off (Dilley et al., 1996; Garellek, 2014). This is **prosodic use of voice quality**. (Though, it is qualitatively distinct from coarticulatory creak, c.f. Keating et al. (2015).)

Contrastive voice quality

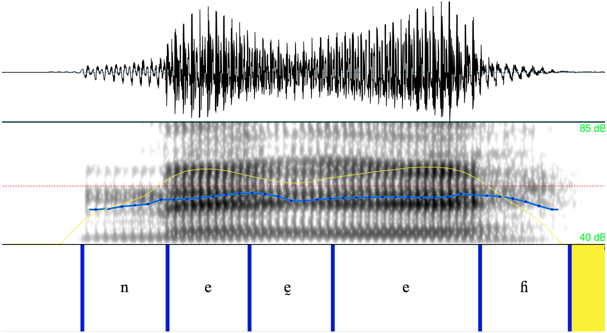
Chong, spoken in Thailand, contains a four-way contrast in what is called "voice register." Words are distinguished mainly by phonation type and secondarily by pitch and other cues (DiCanio, 2009).

Modal	Breathy	Tense	Breathy-Tense
tɔɔŋ	ra̤a̤j	tɔ̰ɔ̰ŋ	pa̤a̰j
'*six*'	'*ten*'	'*fear*'	'*two*'
kəlɔɔŋ	lɔ̤ɔ̤ŋ	lɔ̰ɔ̰n	cɕɔ̤ɔ̰ŋ
'*stride*'	'*husband*'	'*navel*'	'*Chong*'
ceew	cṳṳn	pḛḛw	ro̤o̰j
'*to go*'	'*to send*'	'*dinner*'	'*melon*'

Coarticulatory voice quality

Glottal stops are also rarely realized with full glottal closure across languages (Garellek et al, to appear), but instead with creaky phonation.

Itunyoso Triqui (Mexico)
'they/them' [neɦ^{3}] vs. 'baby' [ne^{3}ˀeɦ^{3}] (DiCanio, 2012)



Sociolinguistic use of phonation type

Different social cues can be marked with changes in voice quality.

Creaky voice (also known as vocal fry) is a characteristic of many speakers' voices. It is associated with more dominant personality among English speakers (alongside lower f0) and is found in both men and women equally (Davidson, 2020).

Breathy phonation may be used to express more intimacy (Gussenhoven, 2016).

In Peninsular Spanish speakers, hoarse voice quality is associated with a "tough girl" persona (Armstrong et al., 2015).

Many more studies on this issue! It's a hot area.

Articulation and acoustic consequences

- The vocal folds are open for longer during breathy phonation, so glottal frication is produced along with voicing. This has the effect of reducing the amplitude of the signal as less subglottal pressure is powering voicing.
- Greater constriction of the vocal folds with creaky or tense phonation causes changes in how fast the vocal folds close. They close much more quickly and stay closed for longer than they do with modal or breathy phonation.

Acoustic effects - breathy phonation

As a consequence of noise in breathy phonation, there is much more aperiodic energy across the spectrum and the formant structure is less clear.

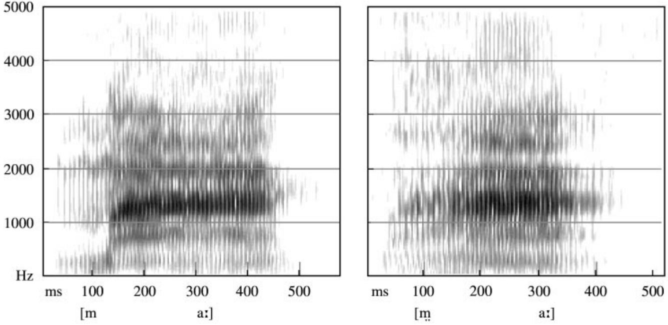
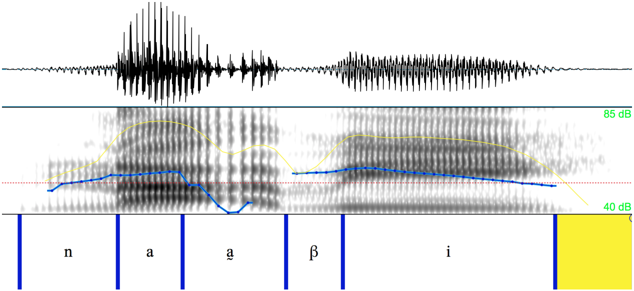


Figure 2. Spectrograms and waveform excerpts of modal and breathy voiced nasals in the Newar words /maː/ "garland" and /m̤aː/ "be unwilling" (male speaker).

(Gordon and Ladefoged, 2001)

Acoustic effects - creaky phonation

Creaky phonation is characterized with irregular glottal periods (jitter) but with clear formant structure. As a consequence of this irregularity, F_0 is not (usually) calculated so accurately.



Itunyoso Triqui [na^{3ʔ}βi^{32}] '*orange*' (DiCanio et al., 2020)

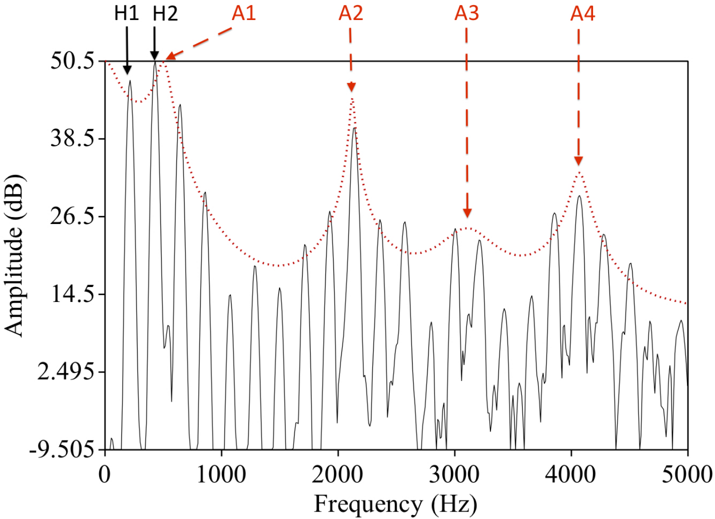
Spectral tilt

The vocal folds come together more quickly in the production of creaky or tense phonation.

As a result of this, vocal fold closure produces a higher amplitude transient. This excites a greater number of harmonics and resonances across the spectrum, resulting in a flatter overall **spectral tilt** than one finds for modal voice (Kirk et al., 1993).

By contrast, the closing phase of the vocal folds is slower for breathy phonation. The resulting weaker transients fail to excite higher resonances in the vocal tract and the spectrum is steeper.

To estimate spectral tilt, we must examine the amplitudes of the different harmonics (H1, H2... HN) and formants (A1, A2... AN) in the spectrum.



Suppose the amplitude of H1 is 60 dB and the amplitude of H2 is 55 dB. The measure H1-H2 would be 5 dB (negative slope!).

Suppose the amplitude of H1 is 60 dB and the amplitude of A3 is 35 dB. The measure H1-A3 would be 25 dB (negative slope!).

Generally speaking, we expect *lower* spectral tilt values (flatter or positive slope) for tense/creaky phonation type and higher spectral tilt values (more of a negative slope) for breathy phonation type.

On measurement

A number of acoustic features correlate with changes in voice quality.

Aside from spectral tilt measures, cepstral peak prominence (CPP) - a measure of periodicity, HNR (harmonics-to-noise ratio), jitter, shimmer, f0, F1, and intensity may all vary with voice quality changes.

There is no "one" measure that is most informative, but narrower spectral tilt (H1-H2, H2-H4), broader spectral tilt (H1-A3, A1-A3), CPP, and HNR (or some measure of spectral noise) are each often quite useful.

Consult Garellek (2019); Kreiman et al. (2007); Pennington (2005) for a discussion of different measures.

Formants impact spectral tilt measurement

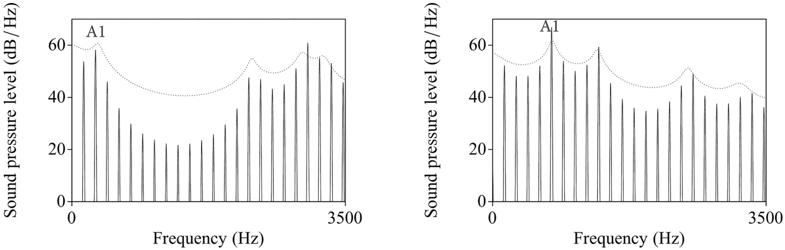


Figure 4.6 FFT and LPC spectra for synthetic [i] (left) and [ɑ] (right) with f_0 = 150 Hz and formants equal to that of an average adult male speaker of American English. For [i] (left), the second harmonic (H2) is also the harmonic closest to the first formant (A1), so H1–H2 equals H1–A1. For [ɑ] (right), H1–A1 equals H1–H5.

(Garellek, 2019, 88)

If formant frequencies impact these measurements, what do you do?

You apply an inverse filtering algorithm to determine the impact of the formants on the source spectrum.

Then you subtract this effect from the observed harmonic and formant amplitudes. This results in a **corrected** measure: H1-H2*, H2-H4*, H1-A3*, etc.

Praatsauce (Kirby, 2019) and its precursor, VoiceSauce (Shue et al., 2009) do this for us automatically.

They also each provide pitch-synchronous spectral tilt measurement, which permits a closer examination of the temporal correlation between different acoustic measures of voice.

Armstrong, M., Henriksen, N., and DiCanio, C. (2015). Sociophonetic analysis of young Peninsular Spanish women's voice quality. In Klassen, R., Liceras, J., and Valenzuela, E., editors, *Hispanic linguistics at the crossroads: Theoretical linguistics, language acquisition, and language contact*. Amsterdam: John Benjamins Publishing Company.

Davidson, L. (2020). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *WIREs Cognitive Science*, e1547; https://doi.org/10.1002/wcs.1547.

DiCanio, C., Martínez Cruz, B., Cruz Martínez, B., and Martínez Cruz, W. (2020). *Diccionario bilingüe - Triqui de Itunyoso y Castellano*. `http://www.acsu.buffalo.edu/~cdicanio/Diccionario_Triqui_01-08-20.xhtml`.

DiCanio, C. T. (2009). The Phonetics of Register in Takhian Thong Chong. *Journal of the International Phonetic Association*, 39(2):162–188.

DiCanio, C. T. (2012). Coarticulation between Tone and Glottal Consonants in Itunyoso Trique. *Journal of Phonetics*, 40(1):162–176.

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

Edmondson, J. A. and Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology*, 23(2):157–191.

Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics*, 45(1):106–113.

Garellek, M. (2019). The phonetics of voice. In Katz, W. and Assman, P., editors, *The Routledge Handbook of Phonetics*, pages 75–106. Routledge.

Garellek, M., Chai, Y., Huang, Y., and Van Doren, M. (to appear). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*.

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

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8:425–434.

Keating, P., Garellek, M., and Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015, editor, *Proceedings of the 18th International Congress of the Phonetic Sciences*, pages 1–5, Glasgow. University of Glasgow. Available online at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0821.pdf.

Kirk, P. L., Ladefoged, J., and Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. In Mattina, A. and Montler, T., editors, *American Indian Linguistics and Ethnography in honor of Lawrence C. Thompson*, pages 435–450. University of Michigan, Ann Arbor.

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

Pennington, M. (2005). *The Phonetics and Phonology of Glottal Manner Features*. PhD thesis, Indiana University.

Shue, Y.-L., Keating, P., and Vicenik, C. (2009). VOICESAUCE: A program for voice analysis [computer program]. *Journal of the Acoustical Society of America*, 126(2221(A)).