Acoustic properties of different kinds of creaky voice

Patricia Keating[1], Marc Garellek[2], Jody Kreiman[3]

[1]Dept. Linguistics, UCLA, Los Angeles CA USA 90095; [2]Dept. Linguistics, UCSD, San Diego CA USA 92093, [3]Dept. Head & Neck Surgery, UCLA, Los Angeles CA USA 90095
keating@humnet.ucla.edu; mgarellek@ucsd.edu; jkreiman@ucla.edu

ABSTRACT

There is not one kind, but instead several kinds, of creaky voice, or creak. There is no single defining property shared by all kinds. Instead, each kind exhibits some properties but not others. Therefore different acoustic measures characterize different kinds of creak. This paper describes how various acoustic measures should pattern for each kind of creak.

Keywords: phonation, voice quality, creaky voice

1. INTRODUCTION

The term "creaky voice" (or "creak", used here interchangeably) refers to a number of different kinds of voice production. Early linguistic descriptions of creak (e.g. Laver [32]) enumerated many characteristics: low subglottal pressure and glottal flow, slack, thick, compressed vocal folds with a short vibrating length, ventricular contact with the folds, weak or damped pulses, low F0, irregular F0, period-doubled vibration. Later descriptions (e.g. [7, 20, 28]) added such properties as irregular amplitude, low Open Quotient, skewed glottal pulses, narrow formant bandwidths and sharp harmonics, abrupt closure of the folds, and low spectral tilt. Yet it seems clear that these characteristics are not all seen in each instance of creak, and that (when the full range of types is considered) there is no single defining characteristic shared by all instances.

Indeed, previous studies have argued that there are specific sub-categories of creak, each with its own set of characteristics. Hedelin and Huber [24] distinguished "creak" (or "fry" or "pulse", with low F0 and strong damping), "creaky voice" (with irregular pulses), and "diplophonia" (with period doubling). Batliner et al. [5] used six acoustic properties to distinguish five different types of "laryngealization", but their brief report provides little discussion. Later, a pair of papers from presentations at the 1999 ICPhS proposed similar categories, based primarily on visual inspection of acoustic displays. First, Gerratt and Kreiman [20] described two "supraperiodic" types - one "period doubled" (with interharmonics), and one with "amplitude modulation" - plus a highly aperiodic "noisy" type. They demonstrated that these three types are perceptually distinct to ordinary listeners. They also described "vocal fry", with visibly damped pulses. Second, Redi and Shattuck-Hufnagel [35] distinguished four types of creaky voice: irregular "aperiodicity", damped, low-F0 "creak", "diplophonia" (with any kind of alternating pulse frequency, amplitude, or shape), and the rare "squeak" (with a sudden sustained high F0). Redi and Shattuck-Hufnagel showed that not only do these types vary across speakers, but also across positions-in-utterance for individual speakers.

In this paper we build on these previous studies about different kinds of creak from the perspective of researchers performing varied acoustic analyses of a range of voice samples. If each acoustic measure reflects a specific aspect of creak, and if different kinds of creak exhibit specific combinations of these aspects, then different kinds of creak will be distinguished from modal voice by distinct acoustic measures. Specifically, we attempt to relate different kinds of creak to acoustic measures already used by researchers.

2. PROTOTYPICAL CREAKY VOICE

We begin by describing what we take to be prototypical creaky voice, in line with the brief definitions given in many research papers. Prototypical creaky voice has the following three key properties: (1) low rate of vocal fold vibration (F0), (2) irregular F0, and (3) constricted glottis: a small peak glottal opening, long closed phase, and low glottal airflow.

Figure 1: Waveform showing prototypical creak, phrase-final by a male English speaker, vowel /e/.

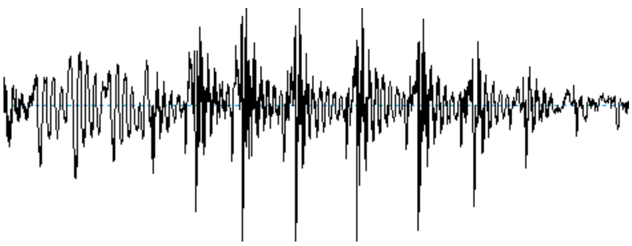


Fig. 1 shows a sample waveform of creaky voice with these properties, from a male speaker of English. F0 is in the range of 70 Hz, but irregular. Glottal constriction is inferred from a high Contact Quotient in the simultaneous EGG signal.

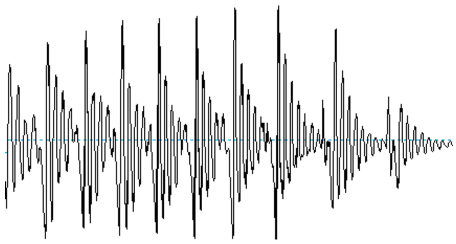
3. OTHER KINDS OF CREAKY VOICE

While prototypical creaky voice is often encountered in speech samples, much of what is called creaky voice – indeed, is perceived as creaky voice – may differ from this prototype in one or more ways. Each of the three properties of prototypical creak can be lacking, yielding several further kinds of creak.

3.1. Vocal fry

Although the term "vocal fry" is often used interchangeably with "creak", vocal fry differs from prototypical creak in a major way: the glottis is constricted and F0 is low, but it is not necessarily irregular. Indeed it is often quite periodic, as in Fig. 2. Its special property is high damping of the pulses – this property, due in part to the low F0, makes individual pulses distinct and separately audible (the "picket fence" effect). Thus the prototypical low-F0 property is enhanced.

Figure 2: Waveform showing phrase-final vocal fry by a female English speaker, with regular F0.



It has been suggested that ventricular incursion, as observed by [1], can be one contributor to vocal fry (though cf. [11]): the ventricular folds contact and mechanically load the vocal folds. This increases the effective mass of the folds, so F0 is extremely low; it can also make vibration irregular. However, as vocal fry was the only kind of creak examined by [1], the incidence of ventricular involvement across kinds of creak is not known.

3.2. Multiply pulsed voice

A very common form of creak involves a special kind of F0 irregularity: alternating longer and shorter pulses. (See [20] for a literature review.) In the case of double pulsing (or period doubling), there are two simultaneous periodicities; higher multiples are also possible. There are thus multiple F0s, usually one quite low and another about (though not exactly) an octave higher, but the resulting percept is usually of an indeterminate pitch, plus roughness. Thus the prototypical low-F0 is not necessarily present. These pulses generally have a very long closed phase, as shown by [41]'s imaging of glottal areas in double- and triple-pulsed creak. See Figs. 3 and 4 for sample waveform and spectrum, the latter showing two sets of harmonics.

Figure 3: Waveform showing double pulsing by a male English speaker on a steady /a/. Note the regular alternation of strong and weak pulses.

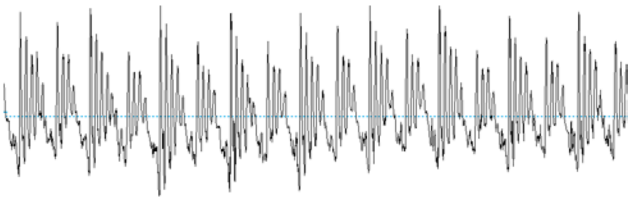
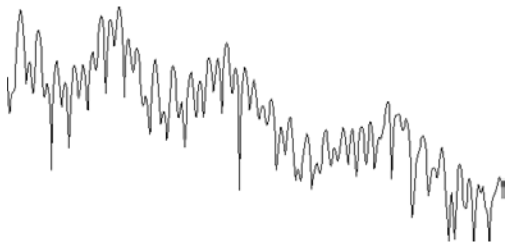


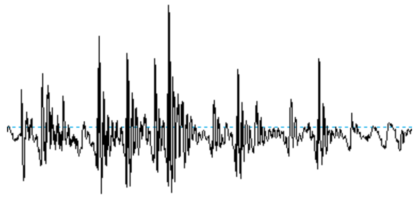
Figure 4: Spectrum of vowel in Fig. 3. Note two sets of harmonics.



3.3. Aperiodic voice

Another variant of F0 irregularity is when it is taken to the extreme – vocal fold vibration is so irregular that there is no periodicity and thus no perceived pitch. See Fig. 5. Like multiply pulsed voice, aperiodic voice lacks the prototypical property of low F0; instead, the property of irregular F0 is enhanced, and the voice is therefore noisy.

Figure 5: Waveform showing extreme aperiodicity, phrase-finally by a female English speaker.

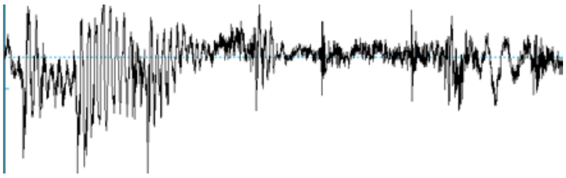


3.4. Nonconstricted creak

This is a voice quality described by Slifka [38, 39]. F0 is low and irregular, as in prototypical creak; but the glottis is spreading, not constricted, and therefore airflow through the glottis is higher, not lower. This kind of creak is attested utterance-finally, with the vocal folds beginning to spread before the utterance is over. The naturally-low subglottal pressure in this position, combined with the spreading glottis, means that conditions for sustaining voicing are not ideal. The slow and irregular vibrations indicate voicing at the edge of failing. See Fig. 6.

While this kind of creak, with its increasing airflow, is necessarily somewhat breathy, it differs from Laver's [32] proposed "breathy creak", said to involve airflow through a posterior (arytenoid) glottal gap, simultaneous with anterior creak.

Figure 6: Waveform showing nonconstricted creak, phrase-final by a male English speaker. The Contact Quotient from EGG is low in this token, indicating little glottal constriction.

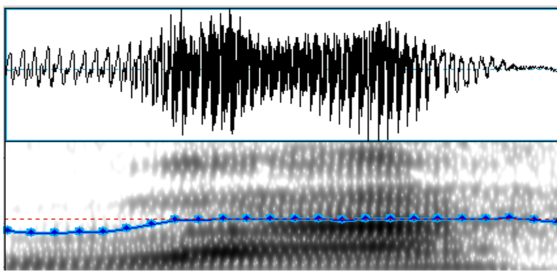


3.5. Tense/pressed voice

When the glottis is constricted, but the F0 is neither low nor irregular, a tense or pressed voice quality is heard. While not always considered a form of creaky voice, it can function phonologically as such in languages in which a creaky (or laryngealized) phonation can co-occur with high tone. Here the constricted glottis is criterial. See Fig. 7.

The discussion above is summarized in Table 1.

Figure 7: Waveform, spectrogram, and pitch track of "creaky" voice with high tone in Mazatec – phonetically a tense or pressed voice quality. Reduced amplitude may be due to constricted glottis.



4. ACOUSTIC MEASURES OF PROPERTIES OF CREAK

These will be considered primarily with reference to the measures provided by our program VoiceSauce ([36, 37]), freely available and often used to study phonation types in languages. In a few cases, exploratory re-synthesis using the UCLA Voice Synthesizer (e.g. [29]) has been compared. The five properties listed in Table 1 are discussed in turn.

Table 1: Properties characterizing different kinds of creak. Check mark means a property characterizes a type; NO means it does not; blank means variable or unknown.

Property >	low F0	irreg F0	glottal constr	damped pulses	sub-harms
Main correlate > Type ∨	low F0	high noise	low H1-H2	low noise; narrow BWs	high SHR
proto-typical	√	√	√		
vocal fry	√		√	√	
multiply pulsed		√	√		√
aperiodic	NO	√	√		
noncon-stricted	√	√	NO		
tense	NO		√		

4.1. Low F0

Creaky voice usually has lower F0 than modal voice. Low F0 has been shown to be a key correlate of creaky voice in Hmong [13] and Mixtec [18]. Yet F0 can be difficult to estimate when irregular; sometimes no F0 can be found. The STRAIGHT pitchtracker [27] is fairly robust in the face of F0 irregularity. Another option, especially appropriate for multiply-pulsed creak, is Sun's method [40], based on his Subharmonic-to-Harmonic ratio measure (SHR, see below). This is specifically designed to estimate a perceptual F0 in the face of competing simultaneous harmonics. See also [24] for additional discussion of methods for tracking irregular F0. In the limit, if no F0 can be extracted, the voice is aperiodic, and thus without the low-F0 property. Our re-synthesis also suggests that lowering the F0 lowers Cepstral Peak Prominence, a measure of noise (see 4.2).

4.2. Irregular F0

Creaky voice usually has less regular voicing than modal voice. This variability can be measured as pulse-to-pulse jitter, or as the standard deviation of the F0, or by autocorrelation [2]. But such voicing irregularity is perceived as *noise*, not distinct from other kinds of noise [29]. Therefore irregular F0 can be measured as spectral noise, by e.g. Harmonic-to-

Noise Ratios (HNR) across different frequency bands, by [8]'s method, or normalized as in [25]. Low HNR values indicate less strong periodic excitation relative to glottal noise – due either to ill-defined harmonics (as with irregular F0) or prominent glottal noise (as with nonconstricted creak). Note, however, that vocal fry will have a relatively high HNR, since in fry the glottal pulses are so sharply defined.

Irregular F0 via low HNR is a correlate of creaky voice in Ju|'hoansi [33], Mazatec [16], Hmong [13], English [13,14,15], and Taiwanese [34]. Our re-synthesis suggests that adding jitter lowers the Cepstral Peak Prominence (i.e., increases noise), but also the amplitude of the higher formants (i.e., increases spectral tilt).

4.3. Constricted glottis

The most common measure of creak is the amplitude difference between the first and second harmonics, H1-H2 - see e.g. [21]. (This is best estimated by the formant-corrected version H1*-H2*, as in [22], [26]). This measure generally reflects glottal constriction, with a lower value indicating greater constriction. [30] used high-speed imaging of the glottis to show that as long as there is no posterior glottal gap, H1-H2 is usually closely related to the glottal Open Quotient. And, [14] and others have found that it is well correlated with Contact Quotient measures from electroglottography. Creaky voice generally has low values of H1-H2, because the glottis is usually constricted. But in non-constricted creak, H1-H2 will have higher, not lower, values than modal voice.

Low H1-H2 has been shown to be a correlate of creaky voice in Zapotec [4, 12], Ju|'hoansi [33], Mazatec [6, 16], Hmong [3, 13], English [15], Trique [10], Taiwanese [34], and of constricted tense voice in Mpi [6], Chong [9] and Yi languages [31].

Constricted glottis may give rise to vibrations that impart more energy to higher-frequency harmonics, perhaps through a more abrupt closure [23]. At the same time, low flow through the glottis means less energy in H1. As a result, various measures of harmonic amplitude differences generally have lower values in creak (i.e., less spectral tilt). Such results have been found for Mazatec [6, 16], English [15], Zapotec [4], and Trique [10]. Our re-synthesis suggests that a smaller H1-H2 also increases HNR measures (i.e., lowers noise). However, none of these are measures of constricted glottis per se.

4.4. Damping

Damping of glottal pulses plays out in two kinds of measures. First, as noted in 4.2 above: unless the F0 is very irregular, the harmonics in damped pulses should be well defined, such that harmonic-to-noise ratios should be high. Second, due to the long closed phase, formant bandwidths should be narrow (e.g. low B1 values). We have so far been unable to demonstrate this through re-synthesis, however.

4.5. Subharmonics in multiple pulsing

As already noted, multiply-pulsed creak has multiple sets of harmonics. Generally one set is stronger and dominates the harmonic spectrum, while the other harmonics ("subharmonics" or "interharmonics") appear between these stronger ones. Sun's Subharmonic-to-Harmonic Ratio SHR [40] measures the relative strengths of the two sets, and has been used by Sun to characterize the strength of period doubling. Multiply-pulsed creak will have more subharmonics, so higher SHR values [17].

5. CONCLUSION

Prototypical creaky voice can be distinguished acoustically by its lower F0, by its irregular F0 (which results in lower values of various harmonic-to-noise measures), and by its lower H1 and H1-H2, and other harmonic difference measures. Just one or two of these prototypical properties apparently suffices to make a sample creaky. Creak that is vocal fry with a regular F0 could instead show *higher* HNR together with lower formant bandwidths. Creak that is multiply pulsed can lack a clear F0 but instead show subharmonics (resulting in higher values of SHR). Non-constricted creak can instead show higher H1-H2, but still with a low and irregular F0. Creak that is more like tense or pressed voice can have a mid or high, and regular, F0.

We hope to convey that there is no straightforward answer to the FAQ, "What is the best acoustic measure for creaky voice?". It entirely depends on what kind(s) of creak the investigator wants to identify. It cannot be expected that measures such as H1*-H2*, or jitter, etc., will *always* characterize creaky voice, since there are special sub-types that are not glottally constricted, or not irregular, etc. It is crucial to keep in mind that when different acoustic measures seem to "disagree" about the creakiness of a speech sample, the set of measures *as a whole* is in fact giving valuable information about the specific voice quality in the sample.

6. ACKNOWLEDGMENTS

We thank NSF grants BCS-0720304 and IIS-1018863, and NIH grant DC01797, for funding.

7. REFERENCES

[1] Allen, E., Hollein, H. 1973. A laminagraphic study of pulse (vocal fry) phonation. *Folia Phon*. 25, 241-250.

[2] Ashby, M., Przedlacka, J. 2014. Measuring incompleteness: Acoustic correlates of glottal articulation. *JIPA* 44, 283-296.

[3] Andruski, J. 2006. Tone clarity in mixed pitch/phonation-type tones. *J. Phonetics* 34, 388-404.

[4] Avelino, H. 2010. Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *J. Voice* 24, 270-280.

[5] Batliner, A., Berger, S., Johne, B., Kießling, A. 1993. MÜSLI: A classification scheme for laryngealizations. *Proc. ESCA Workshop on Prosody,* Lund, 176-179.

[6] Blankenship, B. 2002. The timing of nonmodal phonation in vowels. *J. Phonetics* 30, 163-91.

[7] Childers, D.G., Lee, C.K. 1991. Vocal quality factors: Analysis, synthesis, and perception. *J. Acoust. Soc. Am*. 90, 2394-2410.

[8] de Krom, G. 1993. A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. *J. Sp. Hear. Res.* 36, 254-66.

[9] DiCanio, C. 2009. The phonetics of register in Takhian Thong Chong. *JIPA* 39, 162–188.

[10] DiCanio, C. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. *J. Phonetics* 40, 162-176.

[11] Edmondson, J.A, Esling, J. H. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23, 157-191.

[12] Esposito, C. 2010. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *JIPA* 40, 181-198.

[13] Garellek, M. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *J. Phonetics* 40, 152-161.

[14] Garellek, M. 2014. Voice quality strengthening and glottalization. *J. Phonetics* 45, 106-113.

[15] Garellek, M. (2015). Perception of glottalization and phrase-final creak. *J. Acoust. Soc. Am*. 137, 822-831.

[16] Garellek, M., Keating, P. 2011. The acoustic consequences of phonation and tone interactions in Mazatec. *JIPA* 41, 185-205.

[17] Garellek, M., Keating, P. 2015. Phrase-final creak: Articulation, acoustics, and distribution. Annual Meeting of the Linguistic Society of America, Portland, OR.

[18] Gerfen, C., Baker, K. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *J. Phonetics* 33, 311-334.

[19] Gerratt, B.R., Kreiman, J. 2001. Toward a taxonomy of nonmodal phonation. *J. Phonetics* 29, 365-381.

[20] Gobl, C., Ní Chasaide, A. 2010. Voice source variation and its communicative functions. In: Hardcastle, W., Laver, J., Gibbon, F. (eds), *The Handbook of Phonetic Sciences (Second Edition).* Oxford: Blackwell, 378-423.

[21] Gordon, M., Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *J. Phonetics* 29, 383-406.

[22] Hanson, H. M. 1995. *Glottal characteristics of female speakers*. Ph.D. Dissertation, Harvard.

[23] Hanson, H. M, Stevens, K.N., Kuo, H.-K. J., Chen, M.Y., Slifka, J. 2001. Towards models of phonation. *J. Phonetics* 29, 451-480.

[24] Hedelin, P., Huber, D. 1990. Pitch period determination of aperiodic speech signals. *Proc. ICASSP* Albuquerque, 361-364.

[25] Hillenbrand, J., Cleveland, R., Erickson, R. 1994. Acoustic correlates of breathy vocal quality. *J. Sp. Hear. Res*. 37, 769-778.

[26] Iseli, M., Shue, Y.-L., Alwan, A. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J. Acoust. Soc. Am.* 121, 2283–2295.

[27] Kawahara, H., Katayose, H., de Cheveigné, A., Patterson, R. D. 1999. Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. *Proc. EUROSPEECH* Budapest, 2781–2784.

[28] Klatt, D., Klatt, L. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820-857.

[29] Kreiman, J., Gerratt, B.R. 2005. Perception of aperiodicity in pathological voice. *J. Acoust. Soc. Am.* 117, 2201-2211.

[30] Kreiman J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., Alwan, A. 2012. Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *J. Acoust. Soc. Am.* 132, 2625-2632.

[31] Kuang, J.J. 2013. *Phonation in Tonal Contrasts*. Ph.D. dissertation, UCLA.

[32] Laver, J. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

[33] Miller, A.L. 2007. Guttural vowels and guttural coarticulation in Ju|'hoansi. *J. Phonetics* 35, 56-84.

[34] Pan, H., Chen, M., Lyu, S. 2011. Electroglottograph and Acoustic Cues for Phonation Contrasts in Taiwan Min Falling Tones. *Proc. 12th INTERSPEECH* Firenze, 649-652.

[35] Redi, L., Shattuck-Hufnagel, S. 2001. Variation in the realization of glottalization in normal speakers. *J. Phonetics* 29, 407-429.

[36] Shue, Y.-L. 2010. *The voice source in speech production: Data, analysis and models*. Ph.D. Dissertation, UCLA.

[37] Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proc. 17th ICPhS* Hong Kong, 1846-1849.

[38] Slifka, J. 2000. *Respiratory constraints on speech production at prosodic boundaries*. Ph.D. Dissertation, MIT.

[39] Slifka, J. 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *J. Voice* 20, 171-186.

[40] Sun, X. 2002, Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. *Proc. ICASSP* Orlando, 333-336.

[41] Whitehead, R. L., Metz, D. , Whitehead, B.H. 1984. Vibratory patterns of the vocal folds during pulse register phonation. *J. Acoust. Soc. Am.* 75, 1293-1297