Acoustic properties of different kinds of creaky voice

Patricia Keating, Marc Garellek, Jody Kreiman

UCLA Linguistics, UC San Diego Linguistics, UCLA Head & Neck Surgery

ICPhS
13Aug2015
Glasgow

Introduction

Creaky voice differs acoustically from modal voice, breathy voice, and other phonation types, on several acoustic measures. Different acoustic measures capture different characteristic properties of creaky voice.

Kinds and properties of creaky voice

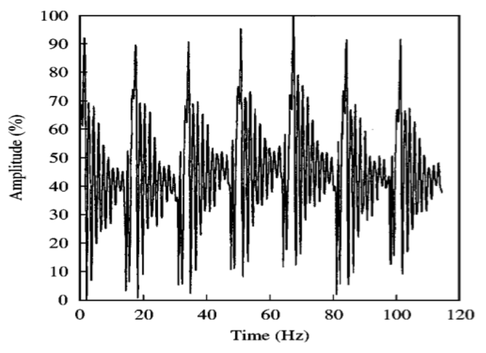
Prototypical creaky voice has these properties:

- Low rate of vocal fold vibration (F0)
- Irregular F0
- Constricted glottis: vocal folds are close together, with a small peak glottal opening and a long closed phase, and so glottal airflow is low

There are other kinds of creaky voice with <u>some but not all</u> of these properties [e.g. 3,15,16, 20, 27, 30]:

Vocal fry:

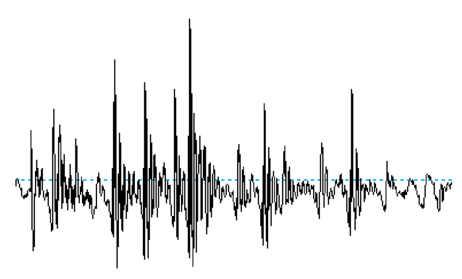
- Low F0
- Regular F0
- Constricted glottis
- Highly-damped pulses – this property, due in part to low F0, makes individual pulses separately audible and enhances the low F0 property



Waveform of vocal fry (Gerratt & Kreiman 2001)

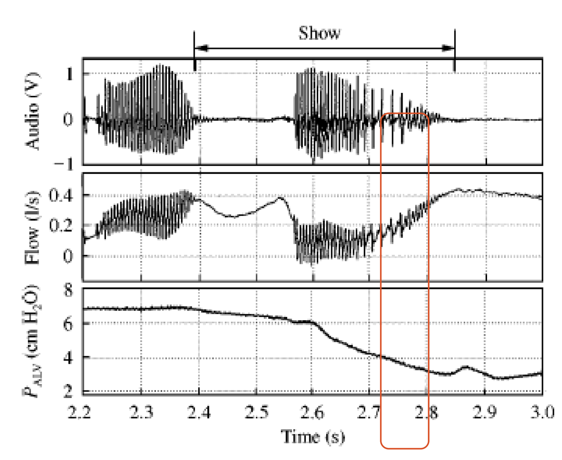
Aperiodic voice:

- Beyond irregular - no periodicity
- No perceived pitch
- Noisy
- Glottal state - ?



Non-constricted voice (called "Slifka voice"):

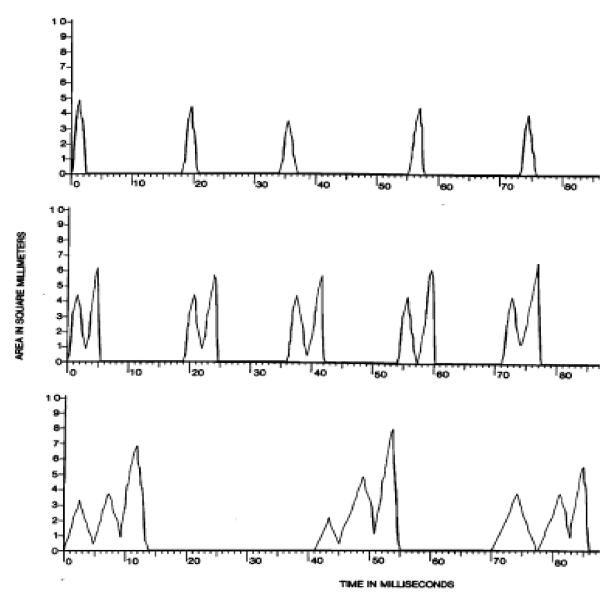
- Low F0
- Irregular F0
- Spreading, not constricted, glottis
- Higher airflow through glottis (mildly breathy)
- Conditions for voicing are not ideal
- Documented by Slifka [33,34] - airflow rises as subglottal pressure falls
- Seen utterance-finally



Version of Fig. 6.18A in Slifka 2000, from Hanson et al. 2001

Multiply pulsed voice:

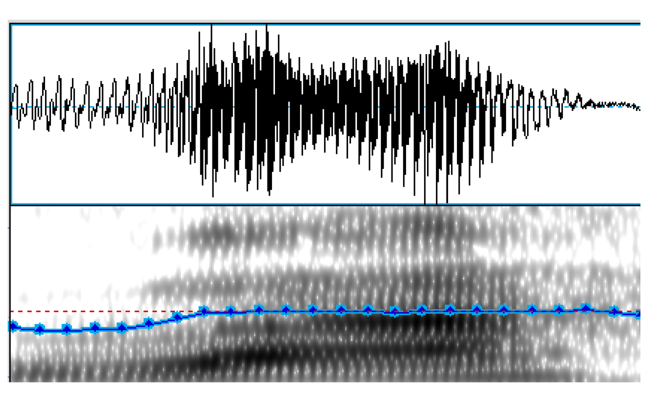
- Alternating longer and shorter cycles
- Multiple or indeterminate F0
- Constricted glottis
- Percept of roughness
- See [16] for review



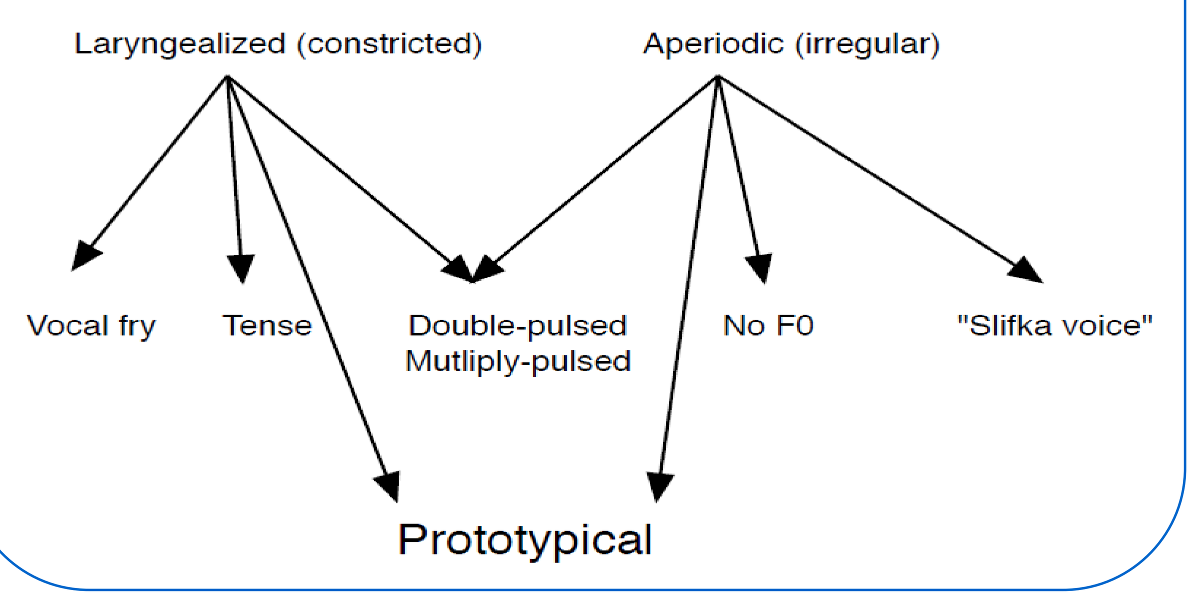
Glottal areas over time (Whitehead et al. 1984)

Tense/pressed voice:

- Mid or high F0
- Regular F0
- Constricted glottis
- E.g.: Mazatec "creaky" voice with high tone [12] – example



Summary: Types of creaky voice



Acoustic measures

What acoustic measures reflect these various properties of creaky voice? Here we describe measures made by **VoiceSauce** [31], a free analysis program from UCLA [32]. Some measurements are of tokens re-synthesized (using the UCLA Voice Synthesizer [24]) from natural vowels by 8 speakers so as to have various combinations of properties of creak.

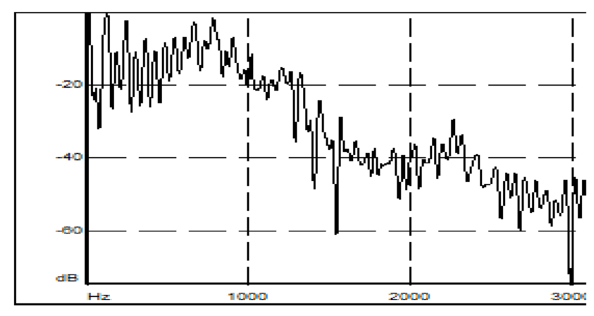
F0: Low in most creaky voice

- ➢ Important correlate of creaky voice in Hmong [9], Mixtec [14]
- The **STRAIGHT pitchtracker** [19] is fairly robust in the face of F0 irregularity. For our re-synthesized tokens with introduced jitter, this was the best pitchtracker in VoiceSauce (where it is the default).
- Sun's method [35] based on the **Subharmonic-to-Harmonic ratio (SHR)**, also available in VoiceSauce, is specifically designed to estimate a perceptual F0 in the face of *subharmonics* (see below), so is appropriate for some multiply pulsed creaky voice. It often gives higher F0s than other methods, however.
- Our re-synthesis indicates that a low-enough F0, around 50-60 Hz, by itself will sound creaky, without affecting H1-H2.
- Our re-synthesis also suggests that lowering the F0 can lower the Cepstral Peak Prominence (see below), and a very low F0 will cause a large increase in HNR 0-500Hz.

F0 regularity: Low in most creaky voice

- ➢ Correlate of creaky voice in Ju|'hoansi [28], Mazatec [12], Hmong [9], English [9,10,11]; Taiwanese [29]
- Often measured as *jitter*, or as SD of the F0. But voicing irregularity is perceived as *noise*, not distinct from other kinds of noise [24]. Therefore VoiceSauce does not measure jitter, but instead, **spectral noise**.
- **Harmonic-to-noise ratios (HNR)** across different frequency bands (0-500, 0-1500, 0-2500, 0-3500 Hz) by de Krom's method [5], or normalized by Hillenbrand's method (**Cepstral Peak Prominence**) [21].
- **Low values indicate less strong periodic excitation** relative to glottal noise – due either to ill-defined harmonics (e.g. with irregular F0) or to prominent glottal noise.
- Our re-synthesis with introduced jitter resulted in lower Energy, HNRs, and CPP. At original F0s, harmonic amplitudes were not affected, but if F0 was also lowered, H1 measures decreased.

Multiple pulsing: A special kind of F0 irregularity in creaky voice



- Two periodicities give two sets of harmonics. Usually one set is stronger and determines the perceived pitch; the other shows as *subharmonics*.
- Spectrum shown here has F0 ~120 Hz, subharmonics between harmonics of the F0
- Sun's **Subharmonic-to-Harmonic ratio (SHR)** measures the strength of the subharmonics; creaky voice tends to have more subharmonics so higher values [35].

Glottal constriction: H1-H2 is lower in most creaky voice

- ➢ Correlate of creaky voice in Zapotec [2,8], Ju|'hoansi [28], Mazatec [4,12], Hmong [1,9], English [13], Trique [7], Taiwanese [29], and of constricted tense voice in Mpi [4], Chong [6] and Yi languages [26]
- Differences in amplitudes of harmonics reflect phonation quality; **H1-H2** is the most commonly used harmonic amplitude measure [17].
- H1-H2 reflects glottal constriction/open quotient, as long as there is no posterior glottal gap [25], with **lower values meaning more constriction.**
- While most creak has lower values of H1-H2, non-constricted creak is notable for its higher values [e.g. 11, 13, 23]
- Our re-synthesis suggests that a smaller H1-H2 also increases HNR measures (i.e. lowers noise).

Other spectral slope measures: Stronger higher-frequency harmonics in most creaky voice

- ➢ Correlates of creaky voice in Mazatec [4,12], English [11], Zapotec [8], Trique [7]
- Creaky voice usually has **strong higher harmonics**, and thus smaller values of most harmonic difference measures. VoiceSauce includes several other harmonic difference measures: **H1-A1, H1-A2, H1-A3, H2-H4, H4-H2k, H2k-H5k**.

Damping: Slightly higher HNR values

- In a small dataset of naturally-occurring fry vs. modal tokens, the various **HNR** measures were 2-6 dB higher in the fry tokens, presumably reflecting sharper, clearer harmonics
- Harmonic amplitude difference measures were also different, as for other constricted creak (fry < modal)

Conclusions

Prototypical creaky voice can be distinguished acoustically by its low F0, by its irregular F0 (which results in lower values of various harmonic-to-noise measures), and by its lower H1 and H1-H2. Just one or two of these prototypical properties apparently suffices to make a sample sound creaky. Some kinds of creak have opposite patterns from the prototypical. Thus **there is no single best acoustic measure for creaky voice**, and no single required pattern of measures for identifying creak. Instead, different patterns of measures point to different sub-types of creak.

References and Acknowledgments

[1] Andruski, J. (2006). Tone clarity in mixed pitch/phonation-type tones. *J. Phon 34.* [2] Avelino, H. (2010) Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *J. Voice 24.* [3] Batliner, A., S. Berger, B. Johne, A. Kiessling (1993). MÜSLI: A classification scheme for laryngealizations. Proc. ESCA Workshop on Prosody, Lund. [4] Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *J. Phon. 30.* [5] de Krom, G. (1993). A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. *J. Sp Hear. Res. 36.* [6] DiCanio, C. (2009). The phonetics of register in Takhian Thong Chong. *JIPA 39.* [7] DiCanio, C. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *J. Phon. 40.* [8] Esposito, C. (2010) Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *JIPA 40.* [9] Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *J. Phon. 40.* [10] Garellek, M. (2014). Voice quality strengthening and glottalization. *J. Phon. 45.* [11] Garellek, M. (2015). Perception of glottalization and phrase-final creak. *JASA 137.* [12] Garellek, M., P. Keating (2011). The acoustic consequences of phonation and tone interactions in Mazatec. *JIPA 41.* [13] Garellek, M., P. Keating (2015). Phrase-final creak: Articulation, acoustics, and distribution. Linguistic Society of America meeting. [14] Gerfen, C., K. Baker (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *J. Phon. 33.* [15] Gerratt, B., J. Kreiman (2001). Toward a taxonomy of nonmodal phonation. *J. Phon. 29.* [16] Gobl, C., A. NiChasaide (2010). Voice source variation and its communicative functions. In Hardcastle et al. (eds.) *The Handbook of Phonetics Sciences (2nd Edition)*, Oxford: Blackwell. [17] Gordon, M., P. Ladefoged (2001). Phonation types: A cross-linguistic overview. *J. Phon.* 29. [18] Hanson, H. (1995). *Glottal characteristics of female speakers*. Ph.D. Dissertation, Harvard. [19] Hanson, H., K. Stevens, H.-K. Kuo, M. Chen, J. Slifka (2001). Towards models of phonation. *J. Phon. 29.* [20] Hedelin, P., D. Huber (1990). Pitch period determination of aperiodic signals. Proc. ICASSP, Albuquerque. [21] Hillenbrand, J., R. Cleveland, R. Erickson (1994). Acoustic correlates of breathy vocal quality. *J. Sp. Hear. Res. 37.* [22] Kawahara, H., H. Katayose, A. de Cheveigné, R. Patterson (1999). Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. *EUROSPEECH '99.* [23] Khan, S., K. Becker, L. Zimman (2015) Acoustic correlates of creaky voice in English (2015). Abstract for Spring ASA meeting. [24] Kreiman, J., B. Gerratt. (2005). Perception of aperiodicity in pathological voice. *JASA 117.* [25] Kreiman J., Y.-L. Shue, G. Chen, M. Iseli, B. Gerratt, J. Neubauer, A. Alwan. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *JASA 132.* [26] Kuang, J. (2103) *Phonation in Tonal Contrasts*. Ph.D. dissertation, UCLA. [27] Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press. [28] Miller, A. (2007). Guttural vowels and guttural coarticulation in Ju|'hoansi. *J. Phon.* 35. [29] Pan, H., M. Chen, S. Lyu (2011). Electroglottograph and Acoustic Cues for Phonation Contrasts in Taiwan Min Falling Tones. INTERSPEECH 2011. [30] Redi, L., S. Shattuck-Hufnagel (2001). Variation in the realization of glottalization in normal speakers. *J. Phon. 29.* [31] Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models*. Ph.D. Dissertation, UCLA. [32] Shue, Y.-L., P. Keating, C. Vicenik, K. Yu (2011). VoiceSauce: A program for voice analysis. *Proc. ICPhS XVII.* [33] Slifka, J. (2000). *Respiratory constraints on speech production at prosodic boundaries*. Ph.D. Dissertation, MIT. [34] Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice 20.* [35] X. Sun (2002), Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. *Proc. ICASSP-2002*, Orlando. [36] Whitehead, R., D. Metz, B. Whitehead (1984). Vibratory patterns of the vocal folds during pulse register phonation. *JASA 75*.

Thanks to NSF grants BCS-0720304 and IIS-1018863 and NIH/NIDCD grant CD 01797 for funding, and to Yen Shue for VoiceSauce.