Acoustic analysis of creaky voice

Patricia Keating and Marc Garellek

UCLA and UC San Diego

LSA
Annual Meeting
1/11/15
Portland

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

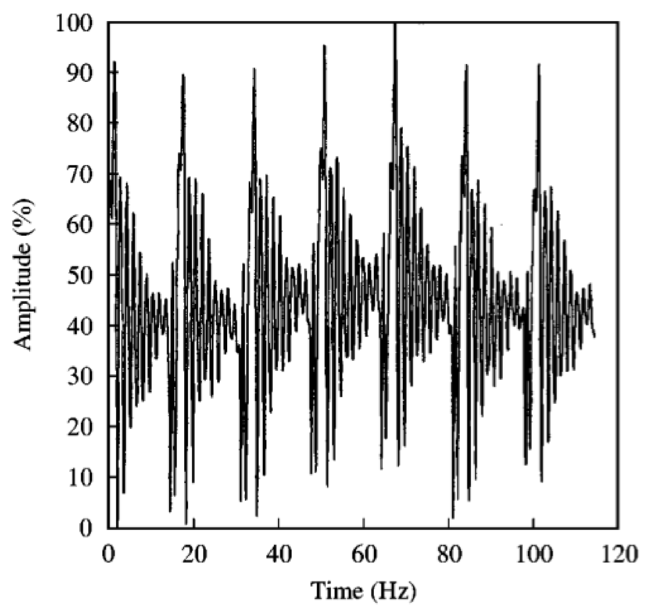
Prototypical creaky voice has these properties:

- 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:

Vocal fry:

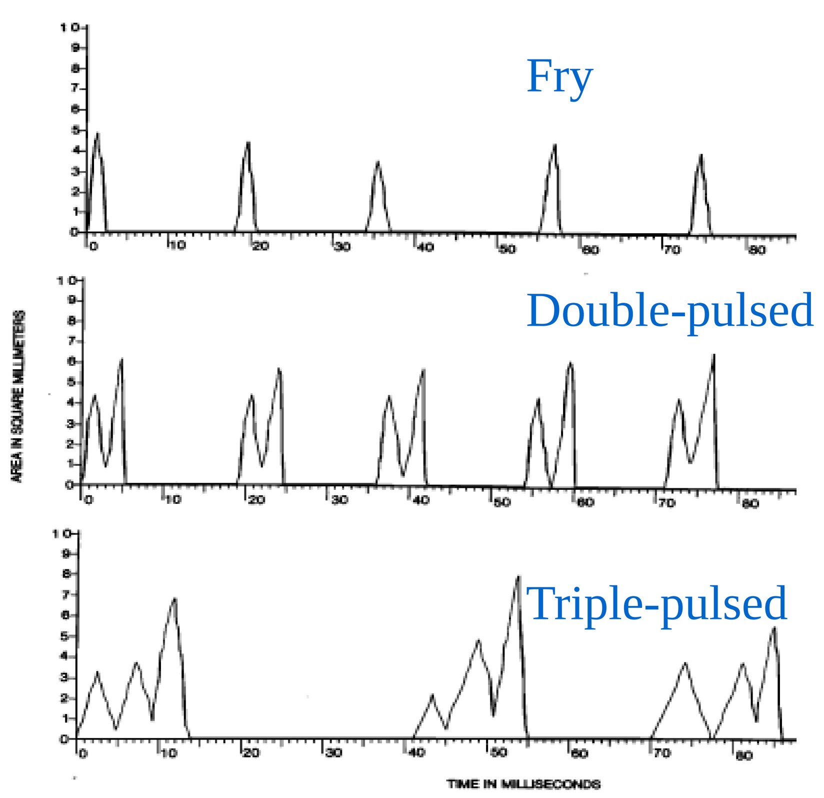
- Low F0
- Regular F0
- Constricted glottis
- Highly-damped pulses – this property combined with low F0 makes individual pulses separately audible



Waveform of vocal fry (Gerratt & Kreiman 2001)

Multiply pulsed voice:

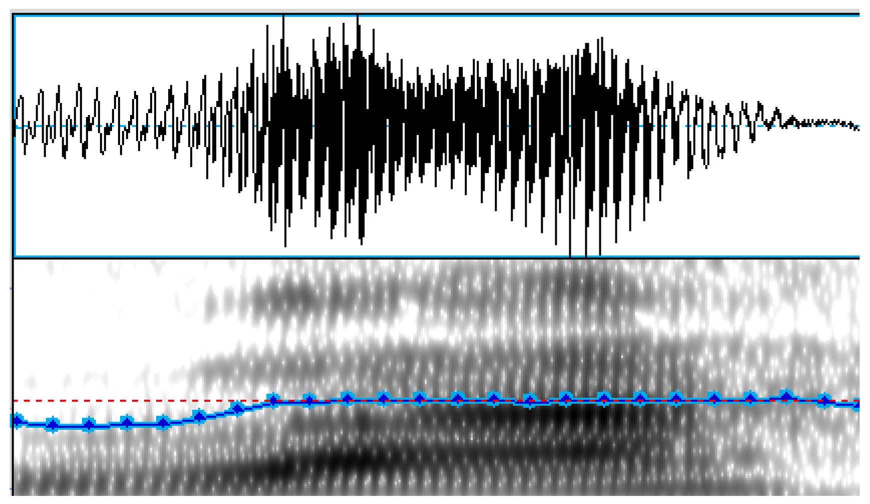
- Alternating longer and shorter cycles
- Multiple or indeterminate F0
- Constricted glottis
- Percept of roughness



Glottal areas over time (Whitehead et al. 1984)

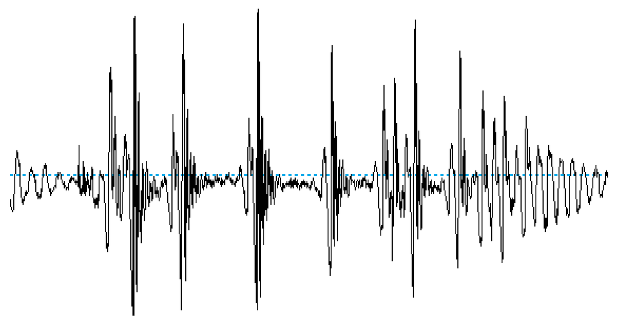
Tense voice:

- Mid or high F0
- Regular F0
- Constricted glottis
- E.g.: Mazatec "creaky" voice with high tone --



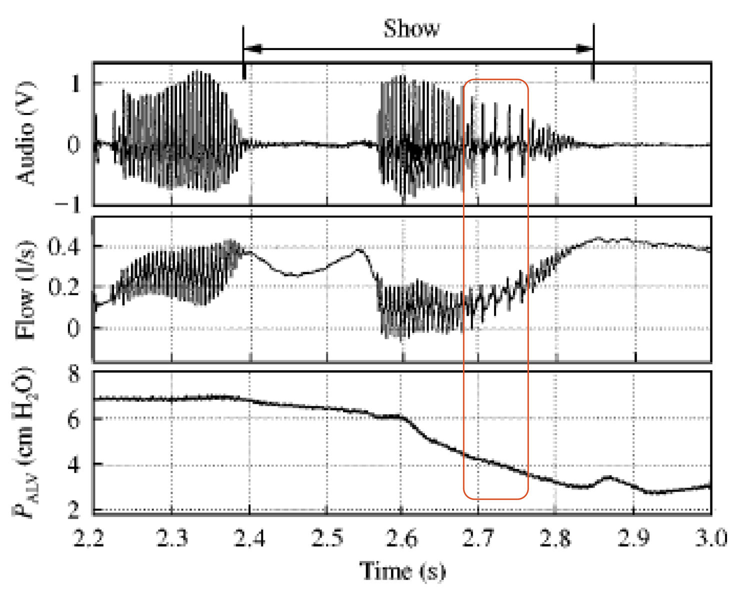
Aperiodic voice:

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



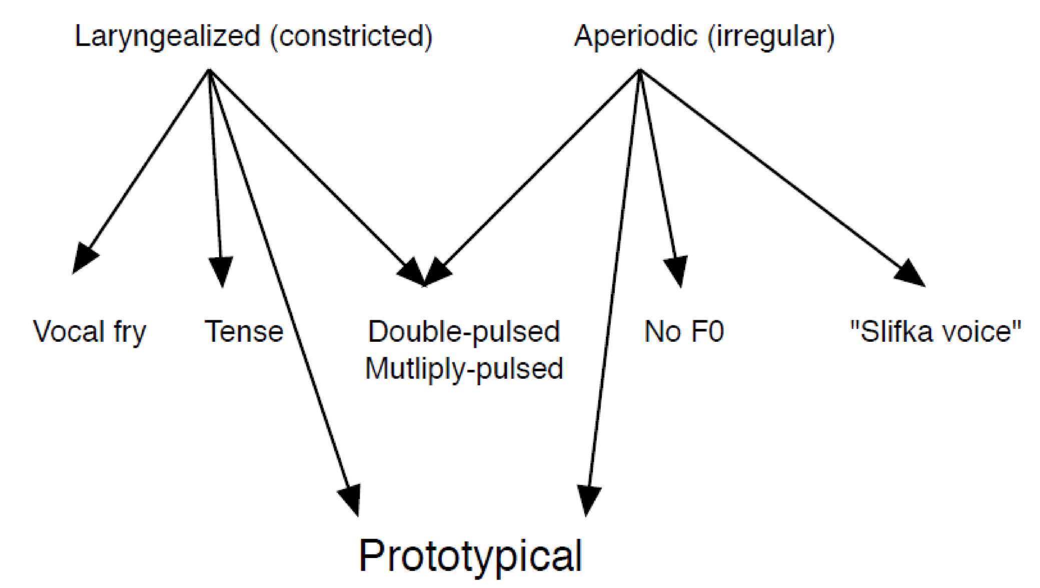
Non-constricted voice (called "Slifka voice"):

- Low F0
- Irregular F0
- Spreading, not constricted, glottis
- Higher airflow through glottis (mildly breathy)
- Conditions for voicing are not ideal
- Documented by Slifka [28,29] - airflow rises as subglottal pressure falls:



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

Summary: Types of creaky voice



Acoustic measures

What acoustic measures reflect these various properties of creaky voice? Here we describe measures made by **VoiceSauce** (Shue 2010), a free analysis program from UCLA described by Shue et al. (2011).

F0: Low in most creaky voice

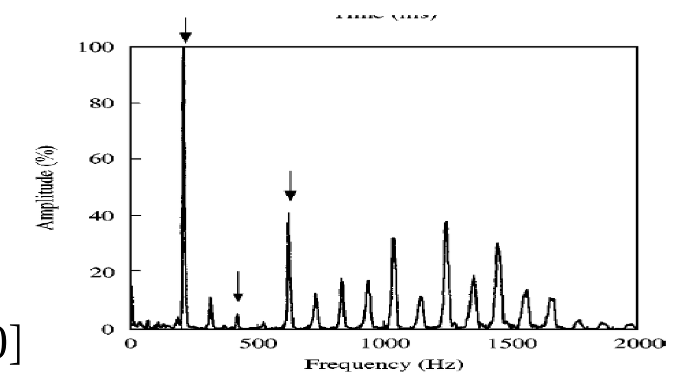
- The **STRAIGHT pitchtracker** [20] is fairly robust in the face of F0 irregularity and is the default in VoiceSauce (**strF0**). In [20], its lowest F0 values for creak were only 18 Hz higher than hand measurements from waveforms.
- Sun's method [30] based on the **Subharmonic-to-Harmonic ratio (SHR)** is specifically designed to estimate a perceptual F0 (**shrF0**) in the face of *subharmonics* (see below), so is appropriate for some creaky voice.
- In VoiceSauce, comparing outputs from different methods can point to the most reliable method for a given dataset, and outliers checked for obvious octave errors. With any method, the "Min F0" parameter should be set very low.
- ➢ Important correlate of creaky voice in Hmong [8], Mixtec [12]

F0 regularity: Low in most creaky voice

- Often measured as *jitter*, or as SD of the F0. But voicing irregularity is perceived as *noise*, not distinct from other kinds of [21]. 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 [4], or normalized by Hillenbrand's method (**Cepstral Peak Prominence**) [17].
- **Low values indicate less strong periodic excitation** relative to glottal noise – due either to <u>ill-defined harmonics</u> (e.g. with irregular F0) or <u>prominent glottal noise</u>. HNR05 is perhaps most sensitive to irregular F0.
- ➢ Correlate of creaky voice in Ju|'hoansi [24], Mazatec [11], Hmong [8], English [8,9,10]; Taiwanese [25]

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*. This spectrum points out H1, H2, H3; there are subharmonics between them:
- Sun's **Subharmonic-to-Harmonic ratio (SHR)** measures the strength of the subharmonics; <u>creaky voice tends to have more subharmonics so higher values</u> [30]



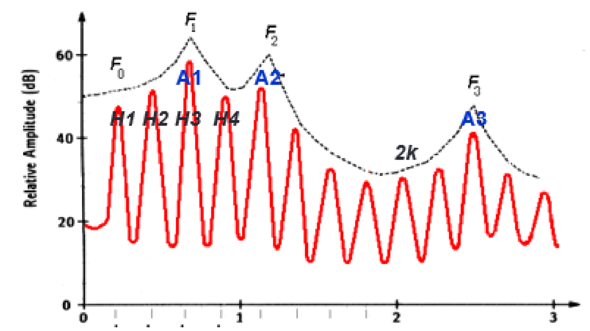
Spectrum of doubling (Gerratt & Kreiman 2001)

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

- Differences in amplitudes of harmonics of the F0 reflect phonation quality; **H1-H2** is the most commonly used.
- H1-H2 reflects glottal constriction/open quotient [4,10], with **lower values meaning more constriction**
- ➢ Correlate of creaky voice in Zapotec [2,7], Ju|'hoansi [24], Mazatec [3,11], Hmong [1,8], English [10], Trique [6], Taiwanese [25], and of constricted tense voice in Mpi [3], Chong [5] and Yi languages [23]

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

- Although the physiology of this is not clear, creaky voice usually has **strong higher harmonics**. VoiceSauce includes several other harmonic difference measures: **H1-A1, H1-A2, H1-A3, H2-H4, H4-2k, 2k-5k**.
- ➢ Correlates of creaky voice in Mazatec [3,11], English [10], Zapotec [2], Trique [6]



On correcting harmonic amplitudes for effects of vocal tract resonances

- Harmonic amplitudes are affected by the vocal tract filter as well as by the source function. [15, 18] provide a method of correcting harmonic amplitudes for local formant frequency and bandwidth influences. Corrected amplitudes are shown with *, e.g. H1*. For H1*-H2*, only F1 and F2 are used in the correction; for H1*-A3*, F1 through F3 are used. The bandwidths used are not those calculated for the tokens, but come from a formula.
- Corrections are not needed for (1) H1-H2 if F1 is very high; (2) any measure if vowel quality is constant.

References and Acknowledgments

1. Andruski, J. (2006). Tone clarity in mixed pitch/phonation-type tones. *J. Phon 34*: 388-404.
2. Avelino, H. (2010) Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *J. Voice 24*: 270-280.
3. Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *J. Phon. 30*: 163-91.
4. 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.
5. DiCanio, C. (2009). The phonetics of register in Takhian Thong Chong. *JIPA 39*: 162–188.
6. DiCanio, C. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *J. Phon. 40*: 162-176.
7. Esposito, C. M. (2010) Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *JIPA 40*: 181-198.
8. Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *J. Phon. 40*: 152-161.
9. Garellek, M. (2014). Voice quality strengthening and glottalization. *J. Phon. 45*: 106-113
10. Garellek, M. (in press). Perception of glottalization and phrase-final creak. *JASA*.
11. Garellek, M. and P. Keating (2011). The acoustic consequences of phonation and tone interactions in Mazatec. *JIPA 41*: 185-205.
12. Gerfen, C. and K. Baker (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *J. Phon. 33*: 311-34.
13. Gerratt, B.R. and J. Kreiman (2001). Toward a taxonomy of nonmodal phonation. *J. Phon. 29*: 365-81.
14. Gordon, M. and P. Ladefoged (2001). Phonation types: A cross-linguistic overview. *J. Phon.* 29: 383-406.
15. Hanson, H. M. (1995). *Glottal characteristics of female speakers*. Ph.D. Dissertation, Harvard.
16. Hanson, H. M, K. N. Stevens, H.-K. J. Kuo, M. Y. Chen, and J. Slifka (2001). Towards models of phonation. *J. Phon. 29*: 451-480.
17. Hillenbrand, J., R. Cleveland and R. Erickson (1994). Acoustic correlates of breathy vocal quality. *J. Sp. Hear. Res. 37*: 769-778.
18. Iseli, M., Y.-L. Shue and A. Alwan (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *JASA 121*: 2283–2295.
19. Kawahara, H., H. Katayose, A. de Cheveigné and R. D. Patterson (1999). Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. *EUROSPEECH '99*: 2781–2784.
20. Keating, P. and G. Kuo (2012). Comparison of speaking fundamental frequency in English and Mandarin. *JASA 132*: 1050-1060.
21. Kreiman, J. and B. R. Gerratt. (2005). Perception of aperiodicity in pathological voice. *JASA 117*: 2201-2211.
22. Kreiman J., Y.-L. Shue, G. Chen, M. Iseli, B. R. Gerratt, J. Neubauer, and A. Alwan. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *JASA 132*: 2625-2632.
23. Kuang, J.J. (2103) *Phonation in Tonal Contrasts*. Ph.D. dissertation, UCLA.
24. Miller, A.L. (2007). Guttural vowels and guttural coarticulation in Ju|'hoansi. *J. Phon.* 35: 56-84.
25. Pan, H., M. Chen, and S. Lyu (2011). Electroglottograph and Acoustic Cues for Phonation Contrasts in Taiwan Min Falling Tones. INTERSPEECH 2011: 649-652.
26. Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models*. Ph.D. Dissertation, UCLA.
27. Shue, Y.-L., P. Keating, C. Vicenik, and K. Hu (2011). VoiceSauce: A program for voice analysis. *Proc. ICPhS XVII*: 1846-1849.
28. Slifka, J. (2000). *Respiratory constraints on speech production at prosodic boundaries*. Ph.D. Dissertation, MIT.
29. Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice 20*: 171-186.
30. Sun (2002), Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. *Proc. ICASSP-2002*, Orlando, 333-336
31. Whitehead, R. L., D. E. Metz and B.H. Whitehead (1984). Vibratory patterns of the vocal folds during pulse register phonation. *JASA 75:* 1293-1297.

Thanks to NSF grants BCS-0720304 and IIS-1018863 for funding, and to Yen Shue for VoiceSauce.

Conclusions

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), by its subharmonics (which result in higher values of the subharmonic-to-harmonic measure), by its relatively weak H1 (due to low airflow through constricted glottis) and by its relatively strong higher-frequency harmonics (which together result in lower values of various harmonic difference measures). But these properties need not all occur in any one token of creak.