Factors for Improved Classification of Creaky Voice

Theresa Kettelberger
University of Rochester

Abstract

Very few algorithms exist to identify vocal fry, or voice creak, in speech. Despite this, creaky voice (CV) is an important paralinguistic feature across languages and its detection would benefit many systems. Existing algorithms are reliant on periodicity of the pulses of CV despite the fact that CV is often aperiodic. Newer algorithms propose alternate measures, but these are generally limited in the environments where they are effective. This paper investigates possible heuristics for CV, several taken from these newer methods and some new ones, and integrates the most effective into a support vector machine.

Background

What is creaky voice?
- Unusual vocal effect that sounds lazy and croak-like
- Caused by compressed, slack vocal folds
- Irregular and strong glottal pulses surrounding quiet, damped voicing

Why detect it?
- Indicator for analyzing English speech prosody and emotion
- Phoneme and tone contrast in other languages
- Interference with pitch detection

Existing approaches:
- Classify by aperiodicity, periodicity, and “very short term” power peak detection [3]
- Apply a resonator to speech to detect extra glottal pulses [1]
- Detect sudden changes in the number of harmonics [4]
- Ratio between the first and second harmonics [5]

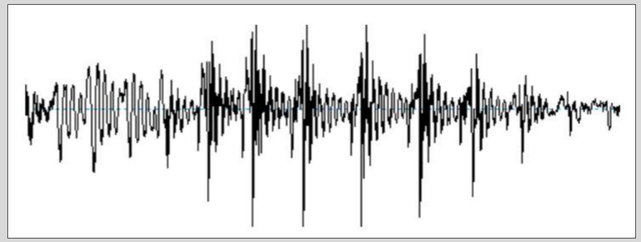

Figure 1. Wave with creaky voice

Data and Annotation

Data collection
- 8 speakers, 4 male and 4 female
- Hour and a half of recording on tired speakers (more likely to display CV)

Annotation
- Time-aligned annotations of selected speech segments [CITE]
- Used audio and visual cues to identify creaky (1) or modal (0) voice
- Segments average about 0.9 seconds
- Modal segments can include anything: silence, voiced, and unvoiced
- 60% modal and 40% creak
- Processed into annotated segments by a script

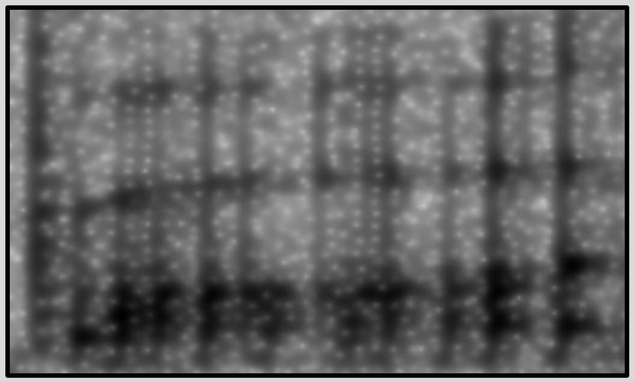

Figure 2. Spectrogram with creaky voice

Measures of Creak

Score
- Change in spectral power: ΔPower
- Good candidates have high changes in power due to irregular amplitude and glottal pulses
- Measured by difference of total sum of spectrum at t and t-1
- Change in formants or harmonic composition: ΔFrequency
- Good candidates have little change in formant structure
- Mitigates false positives from change in power due to change in phoneme
- Measure top k frequencies at each window and find difference in position of these frequencies by window
- Apply a heavier weighting to the frequency change

$$score = \Delta Power/\alpha\Delta Frequency$$

RMS
- RMS is useful as a measure of noisiness
- Creaky segments and voiceless segments share similar RMS measures.

Learning and Results

Support Vector Machine
- I trained on 50,000 examples (spectrogram windows)

Testing
- I tested on the remaining 2,020 examples from my data
- Mean accuracy on classifying test samples was 56.11%
- Not entirely negative results, but could be improved significantly
- Low score on training data (62.97%) evidence for not overfitting, just flawed approach

Questions

Challenges
- Noisiness of data due to multiple categories of measurement with 1 label
- Large spaces between glottal pulses
- Variant pitch range for speakers

Future
- Data annotated for voicedness should be more easily classified by RMS
- Separate training for male and female voices

Conclusions

- SVM may not be the optimal classifier for this task
- Classifiers are still a useful tool as they can train to specific voices
- RMS and score/spectral flux are meaningful but need refinement

References

Drugman, Thomas. Kane, John. Gobl, Christer. Resonator-based Creaky Voice Detection. (2012). INTERSPEECH.
Acoustic properties of different kinds of creaky voice (2015). 18th International Congress of Phonetic Sciences, Glasgow, Scotland.
Ishi, Carlos T. Sakakibara, Ken-Ichi, Ishiguro, Hiroshi. (2011). A Method for Automatic Detection of Vocal Fry. IEEE Transactions on Audio, Speech, and Language Processing.
Martin, Phillipe. Automatic detection of voice creak. (2012). CLILLAC-ARP, EA 3967, UFR Linguistique.
Ishi, Carlos T. Analysis of Autocorrelation-based Parameters for Creaky Voice Detection. JUST/CREST.
Paul Boersma and D Weenik. Praat: a system for doing phonetics by computer. Report of the institute of phonetic sciences of the University of Amsterdam. Amsterdam: University of Amsterdam, 1996.