Analysis and classification of phonation modes in singing

ISMIR 2016 Conference Submission

Daniel Stoller, Simon Dixon

Centre for Digital Music
Queen Mary University London

SIGMA Meeting 19.12.2016
Technical University Dortmund

- Neutral ♬
- Pressed ♬
- Breathy ♬
- Flow ♬

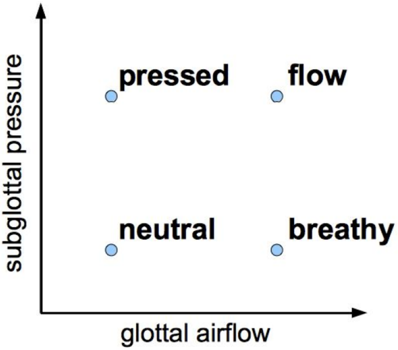


Figure: Four different phonation modes as defined by Sundberg [5]

Related work: Physiology

Laryngeal Resistance as predictor of phonation mode

- Laryngeal Pressure different in each phonation mode
  - Subglottal pressure divided by average airflow
- Difficult: How would this manifest acoustically?

TABLE 1.
Sample Mean, Standard Deviation, and 95% CIs That Could Cover the Population Parameter of μ for Pressed, Normal, Resonant, and Breathy Voice Qualities by LR ($cm\ H_2O/l/s$)

	Sample Mean	Standard Deviation	95% CIs
LR, pressed	766.23	69.77	614.22, 918.24
LR, normal	150.09	9.89	128.55, 171.65
LR, resonant	178.93	16.64	142.68, 215.17
LR, breathy	42.79	6.65	28.31, 57.27

Figure: Laryngeal resistance identified as different in every phonation mode [1]

Dataset

- Dataset with 900 recordings of female professional [2]
- Four phonation modes: Breathy, Neutral, Pressed, Flow
- Nine different vowels
- Pitches between A3 and G5

Related Work

Using the phonation dataset [3]

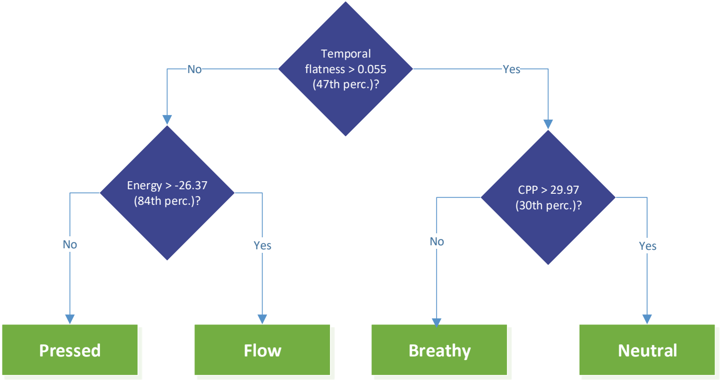
- Source-filter model for simulating the vocal tract [3]
  - Trained on each vowel individually
  - Physiologically interpretable
  - Accuracies between 60% and 75%
- Parallel work [4]
  - Focus on genre and style analysis
  - Accuracy 81.6%
  - No feature analysis

- MFCC
- Temporal flatness
- Spectral flatness
- Zero-crossing rate
- Spectral flux mean/dev.
- Spectral centroid
- High-frequency energy
- F0 mean/dev.
- Harmonic 1-6 amp.
- Harmonic-Noise Ratio
- Formant 1-4 amp./bandw./freq.

- Cepstral peak prominence (CPP)
- Normalised amplitude quotient (NAQ)
- Maxima Dispersion Quotient (MDQ)
- (Glottal) Peak Slope

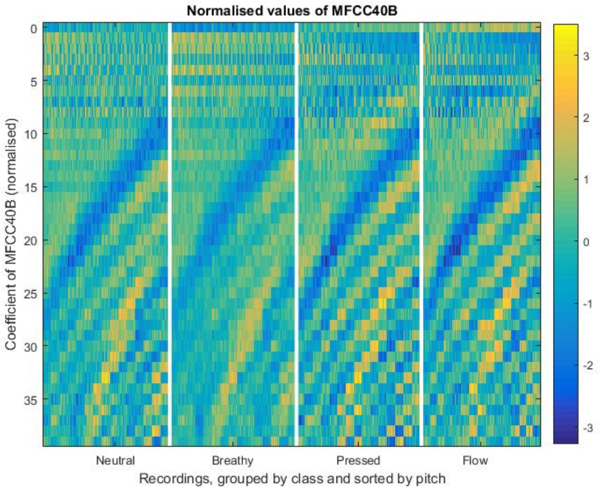
Feature analysis
Simple explanation for phonation modes

- Pruned decision tree on whole dataset:



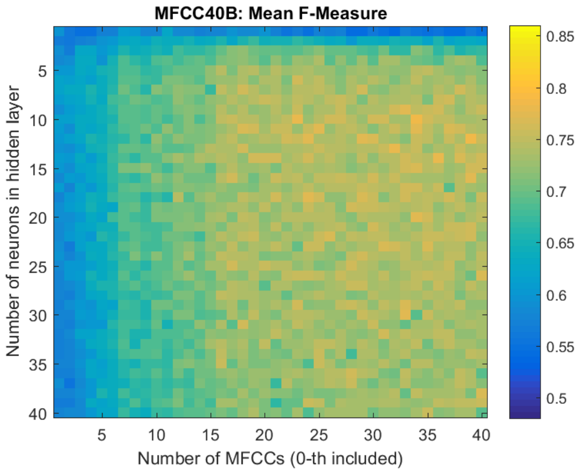
- Categorises 78% of recordings correctly

Feature Analysis
MFCC Visualisation

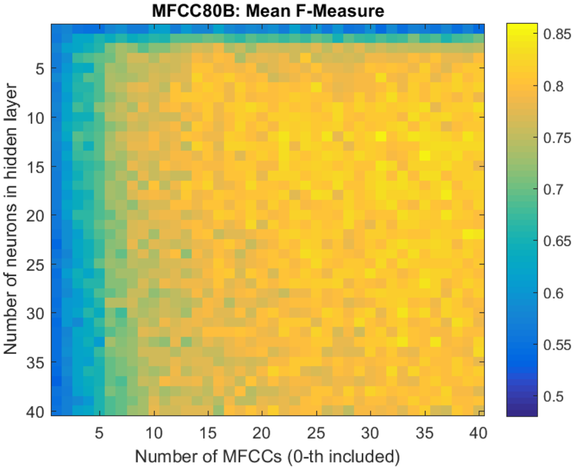


- Single-layer neural network with N hidden nodes
- 10-fold cross-validation
- Grid search optimisation: Run with $N \in \{1, \ldots, 40\}$ neurons and the first $\{1, \ldots, 40\}$ MFCCs

Classification
MFCC40B



Classification
MFCC80B



- Feature analysis to explain phonation modes as acoustic phenomenon
    - Surprisingly variable performance of specialised features
    - Simple decision tree
- MFCCs: Lowest coefficients important, higher number of bands better $\Rightarrow$ Hyperparameter
- Improved classification accuracy, but: Loudness-based classification?
- More comprehensive dataset needed

E. U. Grillo and K. V. and.
Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects.
Journal of Voice, 22(5):546 – 552, 2008.

P. Proutskova, C. Rhodes, T. Crawford, and G. Wiggins.
Breathy, resonant, pressed automatic detection of phonation mode from audio recordings of singing.
Journal of New Music Research, 42(2):171–186, 2013.

P. Proutskova, C. Rhodes, G. A. Wiggins, and T. Crawford.
Breathy or resonant - A controlled and curated dataset for phonation mode detection in singing.
In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, pages 589–594, 2012.

J.-L. Rouas and L. Ioannidis.

Automatic classification of phonation modes in singing voice: towards singing style characterisation and application to ethnomusicological recordings.
In *interspeech*, volume 2016, pages 150–154, 2016.

J. Sundberg.
The science of the singing voice.
Illinois University Press, 1987.