Abstract

Background/Motivation:

- Estimated that 6.6% of US working population affected by voice disorders; most commonly associated with chronic vocal abuse/misuse (vocal hyperfunction)
- Precise role of vocal hyperfunction in the etiology of voice disorders not well understood making diagnosis and treatment less precise
- Need the capability to do long-term ambulatory monitoring of voice use and to extract features that differentiate hyperfunctional from normal/healthy voice production

Procedures:

- We explored techniques to recognize vocal hyperfunction by inferring vocal cord (fold) movement based on non-invasive neck-skin acceleration signal
- Non-acoustic voice features derived from acceleration signal used to classify voice qualities in healthy speaker who mimicked vocal hyperfunction
- Applied 79 traditional features of dysphonia that included variants of HRN, signal jitter, shimmer, and spectral shape
- SVMs were trained using soft margin radial basis functions with parameters chosen based on the best mean AUC

Results:

- Using single-class SVM, we found a best crossvalidated F-score of 0.843 for modal phonation detection
- Further investigating unsupervised clustering of segmented data based on most significant chosen features to characterize underlying mechanisms associated with vocal hyperfunction

Experimental Setup

A healthy adult female speaker was instructed to produce sustained vowel sounds (a, e, i, o, u) mimicking modal (normal), breathy, rough, and pressed voice qualities. Features were calculated over non-overlapping 100-msec windows, generating a total of 1,836 data points (64% modal, 13% breathy, 13% rough and 10% pressed). We removed all features with a cross-correlation coefficient of more than 0.9 with any other feature, leaving a total of 55 features.

From Lab to Field

- Working with real patients on the examination of continuous speech over 7 days (over 15 GB per patient!)
- Many types of hyperfunction; monitor patients pre-therapy/post-therapy and pair with a vocally-normal control
- Near term goal is prediction of vocal hyperfunction episodes

Acknowledgments

This study was made possible with help from Athanasios Tsanas (University of Oxford), and funding from:
- Grant T15LM007092 from the National Library of Medicine, NIH
- Quanta Research Cambridge
- Grant R21-R33 DC011588 from the National Institute on Deafness and Other Communication Disorders, NIH

Ambulatory Voice Monitor

Noninvasive measurement of vocal cord vibration using a neck-mounted accelerometer connected to a smartphone.

Results

Classifying modal in a vocally-normal subject mimicking disordered voice qualities

- Single class (using modal only)
- Two-class (using modal versus breathy/rough/pressed)

From Lab to Field

Data from MTD patient and paired control:

- Normal
- Pathological
- MTD patient fatigues voice after heavy misuse

Data from modal patient and paired control:

- Pathological soft vowels near Normal soft/breathy vowels
- Pathological vowels near Normal modal vowels
- Pathological soft vowels near Normal soft/breathy vowels

Heat map of correlation for 55 least correlated features

Visualization of exemplar data in 3 PCA dimensions

Nonmodal voice quality

Pathological vowels near Normal modal vowels

Pathological soft vowels near Normal soft/breathy vowels

Day 1

Day 7

Window Size SVM Type F-score Sens. Spec. PPV NPV

100 ms Single Class 0.843 0.848 0.421 0.838 0.440
Two Class 0.835 0.832 0.498 0.838 0.486
Single Class 0.827 0.821 0.476 0.833 0.455
Two Class 0.827 0.821 0.500 0.833 0.478