

Clinically Interpretable Acoustic Meta-Features for Characterising the Effect of Mental Illness on Speech and Voice

Maria K Wolters¹, Kristin K Nicodemus², Alex S Cohen³

1-School of Informatics, University of Edinburgh, UK; 2 - Institute for Genomics and Molecular Medicine, University of Edinburgh, UK; 3 - Department of Psychology, Louisiana State University, USA;

Why Study Speech in Mental Health?

In the recent decade, there is a growing trend to describe psychopathology not in terms of diagnoses, which can be contentious, but in terms of observable behaviours, such as those described in the NIMH's Research Domain Criteria (RDoCS, 2011).

One of the core human behaviours that can be observed is language and speech production and comprehension (Elvevåg et al., 2016). Research has shown that a person's mental health status can be reflected in the acoustics of their speech (e.g. Cohen & Elvevåg, 2014; Cummins et al., 2015). This has led to a large body work on the automatic detection of mental state from parameters of the speech signal.

Why Clinically Interpretable?

Most of the successful algorithms for detecting mental state from speech use Machine Learning, and classifications are based on large numbers of features, many of which covary. In addition, some classification approaches, in particular those based on neural networks, are a black box - it is often not clear which features contribute most, and why.

Research Questions

1. Can we identify meaningful, covarying groups of features in speech data from people with and without mental illness / a history of mental illness?
2. Can those features be used to characterise prosodic and acoustic differences between speakers with different diagnoses and symptoms?

The Feature Set

The Geneva Minimalistic Acoustic Parameter Set (GeMAPS; Eyben et al., 2015) is a set of 62 (extended: 88) acoustic speech features that is optimised for research on voice and the expression of affect on voice. It is a standard data

Features cover: fundamental frequency, energy of the speech signal, the frequency spectrum of the speech signal, and temporal characteristics. Features are computed for a stretch of continuous speech, and characterised by their statistical distribution over that stretch (mean, coefficient of variation, and, if applicable, slope).

The Data Set

We analysed 17949 speech samples collected for three studies at the Affective Sciences and Psychopathology Laboratory, LSU: (1) a baseline sample of college students; (2) a sample of college students recruited to maximise variability in schizotypy scores; (3) a community sample of outpatients and controls. For 141 of the 1077 individual participants, we have a comprehensive DSM history. 58 (41%) report depression and 72 (51%) psychosis. 60 of these patients also have a history of schizophrenia. 38 were healthy controls.

Participants were asked to react to pictures from the International Affective Picture Set (IAPS), speak freely about hobbies or memories, or speak while engaged in another task of varying cognitive load.

Data Reduction (RQ1)

Principal component analysis as implemented in R (psych package; Revelle, 2017). Five rotations were compared, none, varimax, promax, oblimin, and cluster-based. The first five principal components (PC) of the varimax solution are shown in Table 1. Missing values were imputed using medians.

Modelling (RQ2)

For the 141 participants with DSM data, we constructed generalised linear mixed models with participant sex and speech task as group-level predictors, and the five principal components as individual-level predictors.

Principal Component	Description
Loudness / Rate	F1/F2/F3 amplitudes relative to F0, loudness, loudness peaks per second
Loudness / Variation	Loudness level and slope
F0 / Spectrum	Fundamental frequency, formant frequencies
F0 / SD	Variation in fundamental frequency
Spectral Balance	Features related to the spectrum and spectral balance

Table 1: First Five Principal Components

Principal Component	All Depression	Only Depression	Psychosis w/ Schiz.	Psychosis, no Schiz.
N	58	18	60	12
Loudness / Rate	-0.12 (0.05)	0.22 (0.06)	-0.03 (0.04)	-0.49 (0.06)
Loudness / Variation	0.08 (0.03)	-0.07 (0.04)	-0.15 (0.03)	-0.24 (0.06)
F0 / Spectrum	-0.43 (0.08)	0.03 (0.09)	0.47 (0.07)	0.0 (0.011)
F0 / SD	-0.27 (0.04)	-0.72 (0.07)	0.27 (0.04)	0.52 (0.05)
Spectral Balance	0.00 (0.06)	-0.23 (0.07)	0.10 (0.05)	0.28 (0.09)

Table 2: Coefficient estimates and standard deviations for each estimate in the logistic regression, trained on all n=141 speakers. Schiz: Schizophrenia

Results

RQ1: Reliability of principal components

All five analyses yielded comparable solutions. While there is some overlap between components, the first component focuses on speaking rate, and the third on the key frequencies of the spectrum, F0 and the first three vocal tract resonances. The second and fourth component model variation in loudness and fundamental frequency, whereas the fifth covers spectral balance and variation in the spectrum.

RQ2: Modelling

Of the 103 people with mental illness, 20 had a history of only one mental illness (mostly depression, n=18), and the remaining 83 have a history of two or more.

People with a history of depression tend to have a flatter intonation contour, those with a history of psychosis a more variable one. As expected, vocal characteristics changed or were attenuated when including comorbidities.

Discussion

While preliminary results are encouraging, this study needs to be replicated with a larger corpus of speech from people with a range of mental health conditions, to see whether the factor structure persists. Even though GeMAPS was originally defined for automatic classification, we suggest that it should be used more widely for clinical studies.