Estimation of vocal fold physiology from voice acoustics using artificial neural networks

Zhaoyan Zhang
Department of Head and Neck Surgery
University of California, Los Angeles, CA, USA

Introduction

Many voice applications require estimating vocal fold physiology from the produced acoustics (e.g., clinical diagnosis of voice disorders)

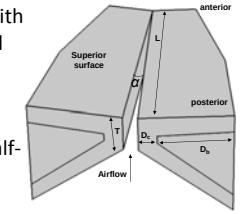
While there have been some previous research on solving the inverse problem in voice production, they are often based on lumped-element models of phonation, whose model parameters are difficult to relate to realistic vocal fold properties.

The goal of this study is to explore the feasibility of using machine learning methods to infer physiologically realistic vocal fold properties, including vocal fold length, thickness, depth, transverse and longitudinal vocal fold stiffness, vocal fold approximation, and the subglottal pressure, from the produced acoustics.

Data

Computational simulations with parametric variations of vocal fold controls (162,000 conditions in total).

Each condition simulates a half-second voice production.



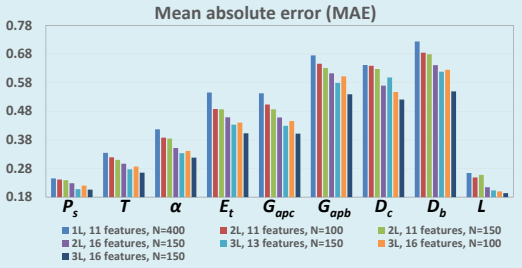
Nine model control parameters and their ranges of parametric variations:

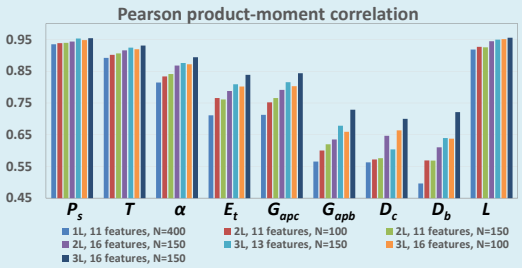
Transverse stiffness	Et = [1, 2, 4] kPa
Longitudinal stiffness in vocal fold cover and body layers	G_{apc} = [1, 10, 20, 30, 40] kPa G_{apb} = [1, 10, 20, 30, 40] kPa
Vertical thickness	T = [1, 2, 3, 4.5] mm
Initial glottal angle (vocal fold approximation)	α = [0, 1.6, 4]°
Vocal fold length	L = [6, 10, 17] mm
Depths of vocal fold cover and body layers	Dc = [1, 1.5] mm Db = [4, 6, 8] mm
Subglottal pressure	Ps = 50 – 2400 Pa (18 values)

Voice inversion training

- For each condition, voice features are extracted from the glottal flow waveform and the output acoustics. Sixteen features are selected for this study, including fundamental frequency, sound pressure level, spectral measures of the voice spectra (H1-H2, H1-H4, H1-H2k, H1-H5k, HNR, CPP, subharmonic to harmonic ratio), and measures of the glottal flow waveform (CQ, mean flow, max. flow, perturbations, max. flow acceleration and deceleration).
- The data are normalized, and randomly divided into three sets, each for training, validation, and testing.
- Feedforward neural network with one (1L), two (2L), or three (3L) hidden layers of N interconnected neurons.

Training results



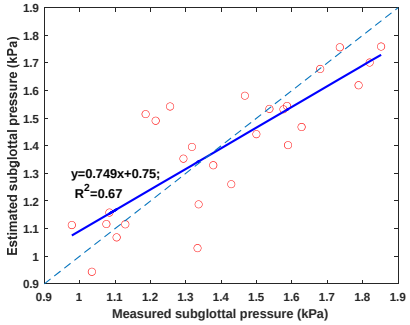


- Similar performance in training, validation, and test data.
- Performance improves with increasing number of features and complexity of neural network.

Best performance

	correlation	MAE	MAE in real unit
Subglottal pressure	0.954	0.206	137.3 (Pa)
Vertical thickness	0.931	0.265	0.32 (mm)
Initial glottal angle	0.894	0.318	0.51 (degree)
Transverse stiffness	0.839	0.403	0.49 (kPa)
Longitudinal stiffness, cover layer	0.844	0.402	5.33 (kPa)
Longitudinal stiffness, body layer	0.729	0.539	7.18 (kPa)
Cover layer depth	0.700	0.521	0.12 (mm)
Body layer depth	0.721	0.549	0.79 (mm)
Vocal fold length	0.956	0.183	0.86 (mm)

Comparison to excised human larynx experiment



- Estimated subglottal pressure follows the experimental trend.
- Estimated geometric parameters are close to their values from MRI measurement.
- No stiffness measurements available from experiment.

Summary

- The subglottal pressure, vocal fold vertical thickness, and vocal fold length can be estimated with very low mean absolute errors
- Vocal fold stiffness and depth, particularly those of the body layer, consistently have large estimation errors.
- The estimation accuracy for the initial glottal angle is somewhere in between.
- Reasonable agreement with experiment.
- Need to identify additional features that improve estimation of vocal fold stiffness and depths.

Acknowledgments

This research was supported by grants R01DC009229 and R01DC011299 from the National Institute on Deafness and other Communication Disorders.

Reference: Zhang, Z. (2020). Estimation of vocal fold physiology from voice acoustics using machine learning, J. Acoust. Soc. Am., 147, EL264-EL270.