Interpreting glottal flow dynamics for detecting COVID-19 from voice*

Soham Deshmukh[1] Mahmoud Al Ismail[1] Rita Singh[2]

[1]Microsoft [2]Carnegie Mellon University

*The work was done at Carnegie Mellon University

Motivation

Figure: COVID-19 cases worldwide over time

- 100 million infections, 2 million casualties spanning across 200 countries
- Critical to identify people who are in need of care in timely fashion with subsequent isolation steps
- Testing still lacking
- **Needed:** A scalable testing method to provide timely results

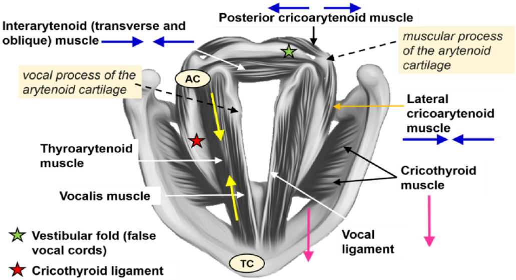
Motivation- wide spread accessible testing

Recent studies have shown vocal fold motion is adversely affected in symptomatic patients [Al Ismail et. al., 2020].
However, these studies are able to identify broad-level anomalies by visual comparisons between oscillation patters of healthy and *symptomatic* COVID-19 positive people.

In this work

We focus on developing a non-invasive scalable way to detect and analyse vocal fold motion during voice production and utilise it for downstream tasks like COVID-19 detection

Phonation



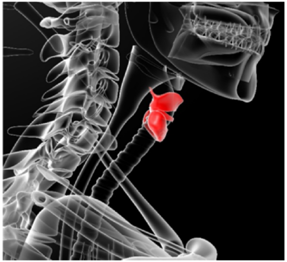


Figure: Laryngoscopic view of the vocal folds

During phonation (such as in producing the sound /a:/), the movements of the vocal folds are self-sustained

- Vibration of vocal folds is driven by physical and aerodynamic forces in the glottis
- The forces are governed by biomechanical properties of vocal folds
- The motion of vocal fold is generally emulated by asymmetric body-cover model

Ascribing the pathologies to deviation in vocal fold motion

Assumption

The 1-mass asymmetric body mass physical model emulated by Van der Poll oscillator is able to explain the vocal fold motions of an individual

Practically

The 1d asymmetric body mass model can best *model healthy persons* with certain degree of asymmetry in vocal fold motion

Approach

Capture the vocal fold oscillation impairment as discrepancy surfacing in the form of differences between:

- Glottal flow waveform obtained from inverse filtering
- Glottal flow waveform estimated from 1d asymmetrical body mass model

Vocal fold parameter estimation

ADLES primary formulation

$$\min \int_0^T (u_0(t) - u_0^m(t))^2 dt$$
$$\Leftrightarrow \min \int_0^T \left(\tilde{c} d(2x_0 + x_l(t) + x_r(t)) - \frac{A(0)}{\rho c} \mathcal{F}^{-1}(p_m(t))\right)^2$$

$$\text{s.t.} \quad \ddot{x}_r + \beta(1 + x_r^2)\dot{x}_r + x_r - \frac{\Delta}{2} x_r = \alpha(\dot{x}_r + \dot{x}_l)$$
$$\ddot{x}_l + \beta(1 + x_l^2)\dot{x}_l + x_l + \frac{\Delta}{2} x_l = \alpha(\dot{x}_r + \dot{x}_l)$$
$$x_r(0) = C_r, x_l(0) = C_l, \dot{x}_r(0) = 0, \dot{x}_l(0) = 0$$

Notation

$u_0(t)$: Measured glottal flow
$u_0^m(t)$: Estimated glottal flow
$\tilde{c}$: Air particle velocity
A: Vocal tract area function
$\mathcal{F}^{-1}$: Inverse filter
α, β, Δ: Model parameters where

- α is the coupling coefficient between the supra- and sub-glottal pressure
- β incorporates the mass, spring and damping coefficients of the vocal folds
- Δ is an asymmetry coefficient.

Use ADLES to iteratively estimate the model parameters α, β, Δ

Feature extraction

The classification network

$$\{X_i\}_{i=1}^{T} = \{[u_i(t), u_i^m(t)]\}_{i=1}^{T} \in \mathbb{R}^{2\times T}$$

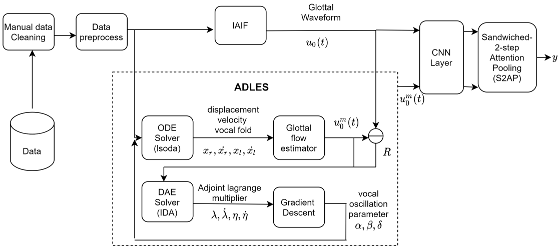
For each frame window X_i

- A pattern detector finds the similarities and differences between $u_i(t), u_i^m(t)$ at each time step
- Aggregates the outputs of the second stage to yield a single prediction for each frame.

- To make classifier decision interpretable, we use Multiple Instance Learning (MIL)
- In this MIL setup, the first stage is CNN based pattern detector and second stage is pooling function
- Specifically, we use two step attention pooling in the second stage

Experiment setup- COVID-19 detection

COVID-19 analysis system setup



Covid analysis setup

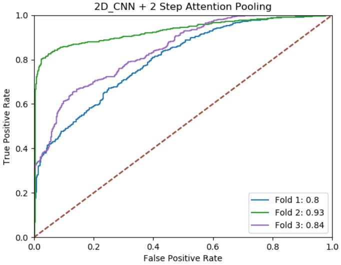
Experiment setup

- The data was collected under clinical supervision and curated by Merlin Inc., a private firm in Chile.
- The data included recordings of extended vowels /a/, /i/, /u/ of COVID-19 positive and negative (medically tested) individuals.
- We used 19 recordings that were collected within 7 days of testing. These comprised 10 females (5 positive) and 9 (4 positive) males. The recordings comprised 8kHz sampled signals recorded on commodity devices.
- For each file we use a window $\tau = 50ms$ with shift $o = 25ms$, resulting in total of 3835 frames

Results: classification

Feature extractor	Pooling	ROC-AUC	STD
-	2AP	0.6611	0.0978
-	2SAP	0.7925	0.1073
CNN (1,3,32)	2AP	0.8009	0.1009
CNN (2,3,32)	2AP	0.8248	0.0790
CNN (2,3,32)	2SAP	0.8330	0.0745
CNN (2,5,64)	2SAP	**0.8520**	**0.0577**

Table: Classifier performance on 3-fold cross validation

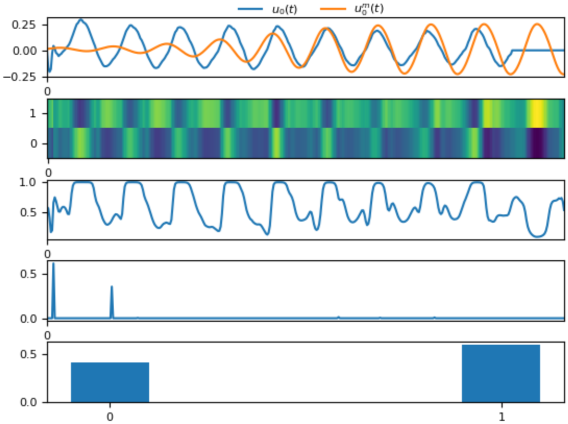


	/a/	/i/	/u/	/a/+/i/	/a/+/u/	/i/+/u/
AUC	0.57	0.839	0.896	0.690	0.804	**0.900**
STD	0.119	0.102	0.067	0.064	0.074	**0.062**

Table: Performance on individual extended vowels and their combinations

The results show the effectiveness of the method in detecting upper respiratory track illness

Results: Visualisation



- Easy to visualize
- Add expert human intervention
- Discard erroneous results

Summary and future work

Summary

- Develop and verify a hypothesis for ascribing upper respiratory tract pathologies to vocal fold motion
- A computationally scalable and interpretable method to use estimated parameters for COVID-19 detection
- Extendable framework to detection of any upper respiratory tract pathologies

Future work

- Large scale experiments on data with symptomatic patients having multiple pathologies
- Explore more sophisticated voice production models than the 1-body mass model being used
- Investigate projected gradient descent and/or Neural ODES

References

- Wikipedia: COVID-19 pandemic data
  COVID-19 pandemic data
- Gauri Deshpande and Björn W. Schuller
  Audio, Speech, Language, & Signal Processing for COVID-19: A Comprehensive Overview (2020)
- T. F. Quatieri and T. Talkar and J. S. Palmer
  A Framework for Biomarkers of COVID-19 Based on Coordination of Speech-Production Subsystems
- Brown, Chloë and Chauhan, Jagmohan and Grammenos, Andreas and Han, Jing and Hasthanasombat, Apinan and Spathis, Dimitris and Xia, Tong and Cicuta, Pietro and Mascolo, Cecilia
  Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data
- Huang, Ying hui et. al.
  The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods
- Lucero, Jorge C and Schoentgen, Jean
  Modeling vocal fold asymmetries with coupled van der Pol oscillators
- Wenbo Zhao, Rita Singh
  Speech-Based Parameter Estimation of an Asymmetric Vocal Fold Oscillation Model and its Application in Discriminating Vocal Fold Pathologies
- Mahmoud Al Ismail, Soham Deshmukh, Rita Singh
  Detection of COVID-19 through the analysis of vocal fold oscillations

Thank you for your attention