DETECTION OF COVID-19 THROUGH THE ANALYSIS OF VOCAL FOLD OSCILLATIONS

Mahmoud Al Ismail   *Soham Deshmukh*   *Rita Singh*

Carnegie Mellon University, Pittsburgh, PA, USA
{mahmoudi,sdeshmuk,rsingh}@andrew.cmu.edu

ABSTRACT

Phonation, or the vibration of the vocal folds, is the primary source of vocalization in the production of voiced sounds by humans. It is a complex bio-mechanical process that is highly sensitive to changes in the speaker's respiratory parameters. Since most symptomatic cases of COVID-19 present with moderate to severe impairment of respiratory functions, we hypothesize that signatures of COVID-19 may be observable by examining the vibrations of the vocal folds. Our goal is to validate this hypothesis, and to quantitatively characterize the changes observed to enable the detection of COVID-19 from voice. For this, we use a dynamical system model for the oscillation of the vocal folds, and solve it using our recently developed ADLES algorithm to yield vocal fold oscillation patterns directly from recorded speech. Experimental results on a clinically curated dataset of COVID-19 positive and negative subjects reveal characteristic patterns of vocal fold oscillations that are correlated with COVID-19. We show that these are prominent and discriminative enough that even simple classifiers such as logistic regression yields high detection accuracies using just the recordings of isolated extended vowels.

Index Terms— COVID-19 detection, Vocal fold oscillations, Phonation models, Voice based detection, Voice profiling

1. INTRODUCTION

The vibration of the vocal folds is the primary source of voicing (or *phonation*) in humans [1]. The membranes that comprise the vocal folds are partially tethered by the muscles, cartilage and ligaments surrounding them, allowing them to open and close the glottal area, and to vibrate in response to the passage of air through the glottis. As a result of their structure and physical placement in the larynx, they have characteristic eigen-modes of vibration, or eigen-frequencies at which they can independently vibrate. These are a function of the biophysical properties of the vocal folds, such as their length, thickness, elasticity etc. During phonation, the vibrations of the two vocal fold membranes *synchronize* or *lock* at one of their many eigen-frequencies. Both, the oscillations of the vocal folds during phonation, and this *entrainment* (or synchrony during vibration), result from an intricate balance of aerodynamic forces across the glottis. These forces are directly dependent on the respiratory functions of the speaker, among other factors [2], and are highly sensitive to changes in them. The oscillation patterns of the vocal folds, the symmetry of their motion as the glottis opens and closes, the frequencies at which they synchronize (or the extent of their synchrony), can all be very easily compromised by fine fluctuations in the airflow dynamics of the upper respiratory tract, or even by slight impairments of any of the laryngeal muscles. Disturbances in any of these factors can cause the vocal folds to vibrate in an asymmetrical and asynchronized fashion, and to fail to lock due to unstable eigen-modes.

Clinical observations of symptomatic patients of COVID-19 have so far revealed that this virus moderately or often seriously impairs the functions of the lower and mid respiratory tract, including that of the lungs, airways and musculature of the respiratory tract. Patients who are symptomatic and have tested positive for COVID-19 as the underlying cause have not only reported changes in their voice, but also a general inability to *produce* voice normally. This leads us to hypothesize that the vocal folds of these persons are likely to exhibit anomalies in their oscillation patterns during phonation, and that these can be used to detect COVID-19 from voice. The goal of this paper is to validate this hypothesis.

1.1. Related Work

As of now, literature on detecting COVID-19 from voice, coughs and other respiratory sounds is recent and sparse [3]. One study [4] has attempted to detect COVID-19 by analyzing the speech envelope, pitch, cepstral peak prominence and the formant center-frequencies. This study observes high-rank eigen-values tending toward relatively lower energy in post-COVID-19 cases, but does not provide strict interpretations. Researchers have also used crowd-sourced data [5, 6] with data-driven end-to-end deep learning methods for this purpose. However, the data remain scarce, and deep learning models are prone to over-fitting – there is no guarantee that the network will specifically learn only COVID-19 related characteristics, and not speaker-specific characteristics.

A controlled medical study that is of special relevance to our work is reported by Huang *et. al.* [7], which uses stethoscope data from lung auscultation to analyze the breathing patterns of COVID-19 patients. In this study, recorded audio

signals were analyzed by six independent physicians. All COVID-19 patients were observed to have abnormal breath sounds like crackles, asymmetrical vocal resonances and indistinguishable murmurs. These results were reported to be consistent with CT scans of the 9th intercostal cross-section of the corresponding patient. The study found concrete evidence of the association of abnormal breath sounds, and asymmetries in vocal resonances with COVID-19 infection. This study suggests that COVID-19 affects the source signal that excites the vocal tract, which implicates abnormalities in vocal fold oscillations. While it supports our hypothesis that observing vocal fold oscillations may yield information relevant to detection of COVID-19, it is infeasible to make such direct observations of patient symptoms (using a stethoscope, or using high-speech videography of vocal fold motion) at scale for widespread diagnostic purposes.

In our work, we use a much more scalable and accessible approach of computationally deducing the oscillations of the vocal folds directly from recorded speech signals. The algorithmic details of this approach are given in Sec. 2. Experiments on clinically curated data reveal the presence of clear bio-markers of COVID-19 in the vocal fold oscillation patterns, in the estimated glottal flow, and in the residuals obtained. In Sec. 3 we discuss these, and analyze their usefulness in detecting COVID-19 using multiple classifiers.

2. ESTIMATING VOCAL FOLD DISPLACEMENTS

2.1. The vocal fold oscillation model

Of the several mathematical models of phonation proposed in the past decades [8, 9, 10, 11, 12, 13], the 1-mass asymmetric body-cover model [8] is of particular interest to us due to its ability to capture asymmetry in the oscillation of left and right vocal folds. We briefly describe this model below.
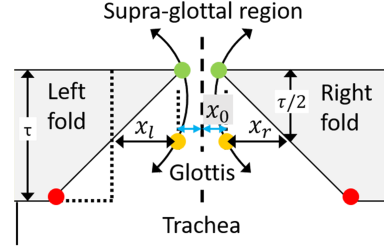
Fig. 1 shows a schematic diagram of the vocal folds. As they vibrate, the horizontal displacements of the left and right vocal folds (x_l and x_r) are measured with reference to the center of the glottis (central dashed line). x_0 represents displacements at rest. The model measures the displacements at the location (yellow dots) where the folds are half their maximum thickness (τ). The length of the vocal folds d is normal to the plane of the figure and not shown.

The asymmetric 1-mass body-cover model is described by the set of coupled non-linear differential equations:

$$\ddot{x}_r + \beta(1 + x_r^2)\dot{x}_r + x_r - \frac{\Delta}{2}x_r = \alpha(\dot{x}_r + \dot{x}_l) \quad (1)$$

$$\ddot{x}_l + \beta(1 + x_l^2)\dot{x}_l + x_l + \frac{\Delta}{2}x_l = \alpha(\dot{x}_r + \dot{x}_l) \quad (2)$$

where α is the coupling coefficient between the supra- and sub-glottal pressure, β incorporates mass, spring and damping coefficients of the vocal folds, and Δ is an asymmetry coefficient. For a male adult with normal voice, their values

Fig. 1: Schematic diagram depicting a cross sectional (frontal) view of the vocal folds. The folds have both horizontal and vertical (curved arrows) modes of oscillation.

(calculated from actual videographic measurements), average to around $\alpha \approx 0.25$, $\beta \approx 0.32$ and $\Delta \approx 0$.

The solution of the dynamical system above yields the displacement, velocity and acceleration of the vocal folds as a set of time-series. The time-series corresponding to x_r and x_l represent the oscillations of the vocal folds. To obtain these, the *forward problem* of estimating the time series must be jointly solved with the *inverse* problem of estimating the parameters of the dynamical system themselves. In [14], we introduced the ADLES algorithm that achieves this by minimizing the error between the glottal flow waveform obtained by inverse filtering, and the vocal fold oscillations predicted by the model as its parameter space is sampled. This joint estimation algorithm is briefly explained in the section below.

2.2. Solving the forward and inverse problems jointly

During phonation, the vocal tract (of length L) acts as a filter that modulates the pressure wave produced by the airflow through the glottis: $\mathcal{F} : p_0(t) \mapsto p_L(t)$. $p_0(t)$, the pressure at the glottis, can be deduced from $p_L(t)$, the pressure sensed by a microphone close to the lips, through inverse filtering: $p_0(t) = \mathcal{F}^{-1}(p_L(t))$. If $A(0)$ represents the cross-sectional area of the vocal channel at the glottis, then the volume velocity of airflow at the glottis, $u_0(t)$, can be deduced from $p_0(t)$ at the glottis as $u_0^m(t) = \frac{A(0)}{\rho c}p_0(t)$, where c is the speed of sound and ρ is the ambient air density. The superscipt m denotes that $u_0^m(t)$ is estimated from the pressure wave measured by a microphone near the mouth.

The volume velocity $u_0(t)$ can *also* be estimated from the solution to the model in Eqns. 1 and 2: $u_0(t) = \tilde{c}d(2x_0 + x_l(t) + x_r(t))$, where d is the length of vocal folds, and $\tilde{c}$ is the air particle velocity at the midpoint of the vocal fold.

We derive our model parameters such that the glottal flow $u_0(t)$ predicted by the model matches the measured flow $u_0^m(t)$ as closely as possible. We define the *residual* $R(t) = u_0(t) - u_0^m(t)$ as the difference between the predicted and actual glottal flows, and the residual energy as

$$\mathcal{E} = \int_0^T R(t)^2 dt \quad (3)$$

We estimate our model parameters to minimize the residual energy $\mathcal{E}$ subject to Eqns. 1 and 2, and boundary constraints:

$$x_r(0) = C_r,\ x_l(0) = C_l,\ \dot{x}_r(0) = 0,\ \dot{x}_l(0) = 0 \quad (4)$$

where C_r and C_l are constants. To solve the above functional least squares, we define the Lagrangian:

$$\mathcal{L} = \mathcal{E} + \int_0^T (\lambda_r E_r + \lambda_l E_l)dt + \nu_l \dot{x}_l(0) + \nu_r \dot{x}_r(0)$$
$$+ \mu_l(x_l(0) - C_l) + \mu_r(x_r(0) - C_r) \quad (5)$$

where E_r encodes the constraint of Eq. 1:

$$E_r = \ddot{x}_r + \beta(1 + x_r^2)\dot{x}_r + x_r - \frac{\Delta}{2}x_r - \alpha(\dot{x}_r + \dot{x}_l)) \quad (6)$$

and E_l is similarly obtained from Eq. 2. $\lambda_l, \lambda_r, \mu_r, \mu_l, \nu_r$ and ν_l are Lagrangian multipliers. Differentiating $\mathcal{L}$ w.r.t. the model parameters and simplifying, we get, for λ_r:

$$\ddot{\lambda}_r + (2\beta x_r \dot{x}_r + 1 - \frac{\Delta}{2})\lambda_r + 2\tilde{c}dR = 0$$
$$\beta(1 + x_r^2)\lambda_r - \alpha(\lambda_r + \lambda_l) = 0 \quad (7)$$

and a similar pair of equations for λ_l as well. At the end of the recording we also have:

$$\lambda_r(T) = 0, \dot{\lambda}_r(T) = 0, \lambda_l(T) = 0, \dot{\lambda}_l(T) = 0$$

Substituting into the Lagrangian and simplifying we get the derivatives of $\mathcal{L}$ w.r.t. the model parameters:

$$\mathcal{L}_\alpha = \int_0^T -(\dot{x}_r + \dot{x}_l)(\lambda_r + \lambda_l)dt \quad (8)$$

$$\mathcal{L}_\beta = \int_0^T ((1 + x_r^2)\dot{x}_r\lambda_r + (1 + x_l^2)\dot{x}_l\lambda_l)dt \quad (9)$$

$$\mathcal{L}_\Delta = \int_0^T \frac{1}{2}(x_l\lambda_l - x_r\lambda_r)dt \quad (10)$$

Using gradient descent to optimize objective (3), we get the following update rules:

$$\alpha^{k+1} = \alpha^k - \delta\mathcal{L}_\alpha$$
$$\beta^{k+1} = \beta^k - \delta\mathcal{L}_\beta$$
$$\Delta^{k+1} = \Delta^k - \delta\mathcal{L}_\Delta \quad (11)$$

where δ is the step-size and k refers to k^{th} iteration.

3. EXPERIMENTS AND RESULTS

The algorithm described above is used to solve for the model parameters α, β and Δ. These parameters are then substituted in the model to iteratively obtain x_r and x_l. The time series corresponding to x_r and x_l comprise the vocal fold oscillations. The behavior of their trajectories is studied in the model's phase space. The behavior can also be located on a bifurcation diagram that maps the behavior types in the model's parameter space. However, we do not extend our study to bifurcation diagrams in this paper.

Data used: For our study we used a data set collected under clinical supervision and curated by Merlin Inc., a private firm in Chile. The dataset included recordings from 512 individuals who were tested for COVID-19, and turned out either COVID-19 postive or negative. Of these, we chose the recordings from only those individuals who had been recorded within 7 days of being medically tested. Only 19 individuals satisfied this criterion. Of these, 10 were females and 9 were males. 5 females and 4 males had been diagnosed with COVID-19, and the rest had tested negative. The speech signals were sampled at 8 kHz, and recorded over microphones on commodity devices. Each individual was asked to utter multiple sounds, including the vowels /a/, /i/ and /u/.

Experiments performed: We performed two studies. In one, we estimated the vocal fold oscillations of the subjects in our dataset, observed the differences in the patterns of phase space trajectories of the model. Only the recordings of extended vowels /a/, /i/ and /u/ were used for this purpose. Each recording was sectioned into segments of 50ms duration, with an overlap of 25ms, generating 3835 sets of oscillation time-series in all. We used the value of the residual $R(t)$ in Eq. 3 to gauge our model's sufficiency in modeling extreme asymmetry in vocal fold motion. The value of $R(t)$ inversely relates to the accuracy with which the model is likely to estimate the vocal fold oscillations.

In the second study, we used the residuals and the coefficients α, β and Δ as features, and investigated the use of several classifiers to discriminate between COVID-19 positive and negative individuals. The classifiers tested in this binary classification task were Logistic regression (LR), Support vector machine with a nonlinear radial basis function kernel (NL-SVM), Decision tree (DT), Random forest (RF) tree and AdaBoost (AB). 3-fold cross validation experiments were done using recordings of the vowels /a/, /i/ and /u/.
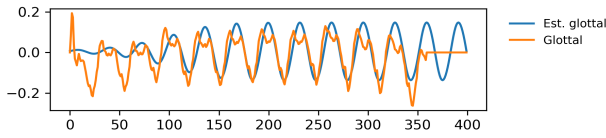
Results of Study 1: The results of the first study are shown in Figs. 2 and 3. Fig. 2 shows the phase space trajectories of the model on a displacement vs. velocity plane for each vocal fold, for COVID-19 positive and negative patients of both genders. We see a significant difference in the phase space behaviors of COVID-19 positive and negative individuals (with a very small number of outliers the need to be investigated in further studies). The phase space trajectories for COVID-19 negative individuals are limit cycles or slim toroids, indicating a greater degree of synchronization in the eigenmodes of vibration, and greater symmetry of motion. For COVID-19 positive patients, the trajectories are more complex, indicating a higher degree of both asynchrony and asymmetry and the *range of motion is reduced*. The vocal folds are unable to maintain the natural self-sustained vi-

(a) Female, negative   (b) Female, positive   (a) Male, negative   (b) Male, positive
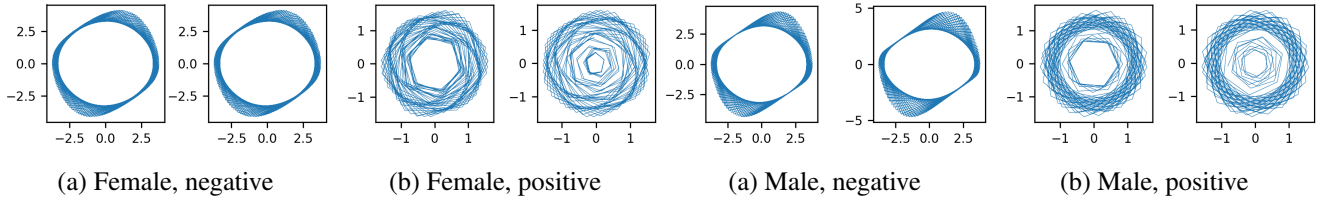
Fig. 2: Phase space trajectories for the left and right vocal folds for COVID-19 positive and negative individuals for the vowel /i/. **Left panels:** x_l (x-axis) vs. $\dot{x}_l$ (y-axis) **Right panels:** x_r vs. $\dot{x}_r$ for each pair.

brations required for vocalization, and thier range of motion is restricted by an order of magnitude relative to normal. Although measures of divergence may be used to quantify these, e.g. Lyapunov exponents [15], we have not used these yet.

Fig. 3 shows a comparison of the estimated oscillations of the vocal folds to the glottal flow waveform obtained by inverse filtering. Note that in reality, the two are not the same. The former are the actual displacements of the vocal folds during phonation, the latter is the airflow volume velocity values across the glottis. Their strong correlation is however reflected in the example shown in Fig. 3.

Fig. 3: Estimated vocal fold oscillations compared to the estimated glottal flow waveform of a subject

Results of Study 2: The results of the second study are shown in Tables 1 and 2, In all experiments, performance was evaluated using the coresponding Reciever Operating Characteristics (ROC) curve. Tables 1 and 2 report the area under this curve (ROC-AUC) and its standard deviation (STD) for each experiment.

Table 1 presents the ROC-AUC and STD obtained for the vowels - /a/, /i/ and /u/. The segments used in the 3-fold cross-validation experiment were stratified – the speakers in the training set were not included in the test set. We observe from Table 1 that all the classifiers achieve a comparable performance of ≈ 0.8 ROC-AUC. The statistical significance was tested for all classifiers and all were found to be significant, with p-values better than $1e^{-5}$. This strongly indicates that the features (residual values and vocal fold oscillation coefficients) can indeed capture the anomalous vibrations of COVID-19 patients without using sophisticated modeling techniques such as neural networks.

In order to gain further insight into the importance of these features, we examined the splits within the decision tree classifier specifically. We found that the residual R is consistently the most important feature, indicating that the vocal fold displacements themselves are highly discriminative for

Classifiers	LR	NL-SVM	DT	RF	AB
ROC-AUC	0.825	0.789	0.803	0.794	0.812
STD	0.032	0.037	0.081	0.060	0.064

Table 1: Performance of different classifiers in a stratified 3-fold cross-validation experiment.

COVID-19. We point out here that while high residual values are discriminative, extreme values may occur because of the inability of the simple model used to model abnormally deviant oscillations. More sophisticated models must be used to to oversome this shortcoming, for better accuracy.

	/a/	/i/	/u/	/a/+/i/	/a/+/u/	/i/+/u/
AUC	0.653	**0.912**	0.877	0.728	0.784	0.901
STD	0.119	0.023	0.035	0.089	0.038	0.023

Table 2: Performance of logistic regression on extended vowels and their combinations.

Table 2 shows the performance of logistic regression on different vowels and their combinations. We observe that the vowel /i/ (a high front vowel) consistently yields the best performance, followed by /u/ (a high back vowel) then /a/ (a low back vowel). This indicates that the ability to reach the higher frequency energy peaks during phonation is compromised due to COVID-19 infection.

4. CONCLUSIONS

While vocal fold oscillation patterns can be indicative of COVID-19, two caveats **must** be noted: a) they are likely to be useful *only* in symptomatic patients, and b) the exclusiveness of the anomalies observed to other respiratory conditions has **not** been tested. We can only say that COVID-19 disrupts the entrainment of the vocal folds during phonation, and causes asymmetries in their motion, and that these characteristics can yield discriminative features that can be used to detect COVID-19 with even simple classifiers. Furthermore, it seems possible to achieve a high ROC-AUC using just a single phonated sound (e.g. the vowel /i/). We hope that the techniques presented in this paper can help facilitate future work towards a simple and cheap alternative for the rapid detection of COVID-19, using more sophisticated models to better capture pathological vocal fold oscillations.

5. REFERENCES

[1] Ingo R Titze, "Nonlinear source–filter coupling in phonation: Theory," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 1902–1915, 2008.

[2] Rita Singh, "Production and perception of voice," in *Profiling Humans from their Voice*, pp. 27–83. Springer, 2019.

[3] Gauri Deshpande and Björn W. Schuller, "Audio, speech, language, signal processing for covid-19: A comprehensive overview," 2020, preprint arXiv, https://arxiv.org/abs/2011.14445.

[4] T. F. Quatieri, T. Talkar, and J. S. Palmer, "A framework for biomarkers of covid-19 based on coordination of speech-production subsystems," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 203–206, 2020.

[5] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo, *Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data*, p. 3474–3484, Association for Computing Machinery, New York, NY, USA, 2020.

[6] Ali Imran, Iryna Posokhova, Haneya N. Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N. John, MD Iftikhar Hussain, and Muhammad Nabeel, "Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, pp. 100378, 2020.

[7] Ying hui Huang, Si jun Meng, Yi Zhang, Shui sheng Wu, Yu Zhang, Ya wei Zhang, Yi xiang Ye, Qi feng Wei, Nian gui Zhao, Jian ping Jiang, Xiao ying Ji, Chun xia Zhou, Chao Zheng, Wen Zhang, Li zhong Xie, Yong chao Hu, Jian quan He, Jian Chen, Wang yue Wang, Chang hua Zhang, Liming Cao, Wen Xu, Yunhong Lei, Zheng hua Jian, Wei ping Hu, Wen juan Qin, Wan yu Wang, Yu long He, Hang Xiao, Xiao fang Zheng, Yi Qun Hu, Wen Sheng Pan, and Jian feng Cai, "The respiratory sound features of covid-19 patients fill gaps between clinical data and screening methods," *medRxiv*, 2020.

[8] Jorge C. Lucero, Jean Schoentgen, Jessy Haas, Paul Luizard, and Xavier Pelorson, "Self-entrainment of the right and left vocal fold oscillators," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 2036–2046, 2015.

[9] Jorge C Lucero and Jean Schoentgen, "Modeling vocal fold asymmetries with coupled van der pol oscillators," in *Proceedings of Meetings on Acoustics ICA2013*. Acoustical Society of America, 2013, vol. 19, p. 060165.

[10] Kenzo Ishizaka and James L Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell system technical journal*, vol. 51, no. 6, pp. 1233–1268, 1972.

[11] Anxiong Yang, Michael Stingl, David A Berry, Jörg Lohscheller, Daniel Voigt, Ulrich Eysholdt, and Michael Döllinger, "Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 948–964, 2011.

[12] Fariborz Alipour, David A Berry, and Ingo R Titze, "A finite-element model of vocal-fold vibration," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 3003–3012, 2000.

[13] Ingo R Titze, "The physics of small-amplitude oscillation of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1536–1552, 1988.

[14] Wenbo Zhao and Rita Singh, "Speech-based parameter estimation of an asymmetric vocal fold oscillation model and its application in discriminating vocal fold pathologies," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 7344–7348.

[15] Alan Wolf, Jack B Swift, Harry L Swinney, and John A Vastano, "Determining lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285–317, 1985.