Measurements from phase-space trajectories of vocal fold oscillations and their application to COVID-19 detection

Anonymous submission to INTERSPEECH 2023

Abstract

It has been shown previously that the displacements of the vocal folds during phonation can be algorithmically derived from recorded voice signals. A graph of the displacements of the left and right vocal folds represents a *phase-space portrait* of a voice signal. The vocal fold oscillation (VFO) trajectory in the phase space carries a wealth of information, which can be measured from statistical, information-theoretic and other perspectives. In this paper, we use several such measurements for detecting COVID-19 from voice. We are also able to derive insights about the changes that may have occurred in the voice signal as a result of COVID-19 infection.

Index Terms: Vocal fold oscillations, Entropy, Information, Covid-19, Phase-space trajectories

1. Introduction

It is well known that speech changes in response to many factors that affect the vocal production mechanism of the speaker [1, 2]. Respiratory illnesses, for example, form a group of influencing factors that cause significant changes in the voices of individuals with these diseases. Characterizing these changes objectively is useful in building systems that can detect these conditions from voice automatically.

Since 2020 (from shortly after the declaration of COVID-19 as a global pandemic by the World Health Organization), researchers across the globe have attempted to create technologies to detect COVID-19 using automated information processing and multimedia analysis tools, including voice. These efforts continue, and the search for the best methods and tools to detect COVID-19 from voice is still ongoing.

Some groups have approached the problem from a data-based inference perspective [3, 4, 5], hoping to discover representations automatically by pairing audio with labels within machine learning and deep learning frameworks. Others, e.g. [6], opt for an analysis-by-synthesis approach, which aims to understand a phenomenon by creating a model that can reproduce it. Potentially explainable inferences can subsequently be made by analyzing the behavior of the synthesis model.

The work in this paper is based on an analysis-by-synthesis approach. Recently, [6] proposed an algorithm called Adjoint Least Squares (ADLES) for deriving the oscillations of the left and right vocal folds from recorded speech. The derivation was based on a physical model of the vocal folds introduced in [7]. This model emulates the asymmetric vibrations of the left and right vocal folds during phonation. These variables comprise a dynamical system, which we elaborate on in Section 3.1.

We use the ADLES algorithm to directly estimate the solutions to the dynamical system. Based on the solutions of this model, we study the vocal fold oscillations of speakers affected by COVID-19. Recent prior work has added sophistication to the aforementioned vocal fold models, by modeling the interaction between the vocal fold oscillations and the resonances of the vocal tract [8]. To our knowledge, this is the first work that applies this more sophisticated model on real data to study a disease's effect on vocal fold oscillations. It is therefore our goal to use this updated model to detect COVID-19.

In this paper we perform comparative analysis of the phonations from patients who test positive or show symptoms of covid, against the phonations of healthy speakers. Through this analysis we can derive deeper insights into the signatures of COVID-19 in voice. For our analysis we use a collection of statistical and information-theoretic measurements that characterize the phase-space trajectories of the mathematical models used.

The *phase-space* of a dynamical system is the space whose coordinates are its state variables. A single trajectory through this space would correspond to the solutions of the model for one recording. We show that our measurements of the trajectories can not only provide insights into the manner in which COVID-19 affects voice, but can also comprise powerful descriptors that can be used to detect COVID-19 from voice.

We find that some features vary significantly between the positive and negative populations, including the mean displacement and mean velocity of *only one* of the vocal folds. This is a notable result, and intuitively follows our understanding of voice pathology, because it has been shown that diseases can have asymmetric effects on the vocal folds [9]. We also find a large degree of separation between positive and negative patients in the information-theoretic measurements.

In addition to examining populations based on test results, we examine the difference in values for individuals that are symptomatic vs. asymptomatic, as test results do not always reflect the reality of whether an individual has the illness. We find that a number of features differ significantly between the symptomatic and asymptomatic populations, mostly in the time-series of velocity values for the vocal folds. The significant features include mean displacement of right vocal fold, the mean velocity of left vocal fold, the regression intercept of vocal fold velocity, and amplitude ratio of vocal fold velocities.

2. Related work

A growing body of evidence suggests that voice analysis can serve as a reliable and non-invasive tool for COVID-19 detection. Previous studies have shown that COVID-19 can affect the respiratory system, leading to changes in vocal characteristics such as pitch, loudness, and phonation time [10, 11].

At the time of writing this paper, several approaches to detecting COVID-19 from voice have been published. However,

most approaches [12, 13] use data-driven machine learning and deep learning models. The features that have been explored are either standard formats such as spectrograms or standard signal processing-based features obtained from toolkits such as OpenSmile [14].

Research on the use of phonation models to estimate the oscillation of vocal folds from recorded speech is relatively new. In fact, the Adjoint Least Squares Estimation (ADLES) algorithm [15] proposed in 2020 was the first to enable such estimation. In later work [8, 16], the vocal fold model was combined with a vocal tract model with a proposed ADLES-VFT (Adjoint Least Squares—Vocal Folds-Tract) algorithm which was based on joint modeling of the oscillation of the vocal fold and the effect of the vocal tract on it.

The viability of phase-space portraits such as those derived by ADLES has been demonstrated in prior studies for detecting multiple diseases. One study used higher order statistics derived from parameters of the dynamical system, in addition to Lyapunov and Hurst exponents derived from the phase space trajectories of the individualized asymmetric models. These techniques were used to directly detect Amyotrophic Lateral Sclerosis (ALS) from voice with an ROC-AUC of 0.99 [17].

There are also prior studies that have tried to detect COVID-19 from voice using the vocal fold oscillations. The analysis approach, however, has varied. Some works such as [18] have used the VFO trajectory to develop simple machine-learning detectors that achieved an accuracy of 91.6 in detecting COVID-19 from voice signals. In another study, the authors in [3] used the ADLES-estimated glottal flows as features to CNN-based two-step attention neural networks. The neural model detects differences in the estimated and actual glottal flows and predicts two classes corresponding to COVID-19 positive and negative cases. This achieved 0.9 ROC-AUC (normalized) on clinically collected vowel sounds.

Our work is different from any others in that we do not use machine-learning classifiers, and so our work is much more interpretable and involves domain knowledge. However we do not sacrifice the ability to detect COVID-19, which is made more effective through our features. It also differs from earlier attempts at analyzing vocal fold oscillations which applied ML techniques to their phase-space measurements in that, in addition to considering more detailed vocal-tract effects on vocal fold oscillations, captured through the ADLES-VFT algorithm, unlike them, we attempt to explicitly statistically relate these to perceptually meaningful characterizations of voice.

3. Methodology

3.1. Vocal fold oscillation model

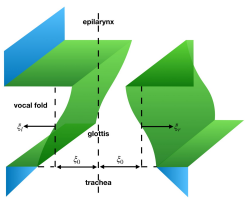

Figure 1: *Diagram of the one-mass body-cover model for vocal folds. The lateral displacements at the midpoint of the left and right vocal folds are denoted as ξ_l and ξ_r, and ξ_0 represents the half glottal width at rest.*

We adopt the specific formulation for the one-mass asymmetric model from [19]. Following Figure 1, we denote the center-line of the glottis as the z-axis. At the midpoint ($z = 0$) of the thickness of the vocal folds, the left and right vocal folds oscillate with lateral displacement ξ_l and ξ_r, resulting in a pair of coupled Van der Pol oscillators:

$$\begin{aligned}\ddot{\xi}_r + \beta(1+\xi_r^2)\dot{\xi}_r + \xi_r - \frac{\Delta}{2}\xi_r &= \alpha(\dot{\xi}_r + \dot{\xi}_l)\\ \ddot{\xi}_l + \beta(1+\xi_l^2)\dot{\xi}_l + \xi_l + \frac{\Delta}{2}\xi_l &= \alpha(\dot{\xi}_r + \dot{\xi}_l)\end{aligned} \tag{1}$$

where β is the coefficient incorporating mass, spring and damping coefficients, α is the glottal pressure coupling coefficient, and Δ is the asymmetry coefficient. For a male adult with normal voice, the reference values for the model parameters (from clinical measurements) are usually approximately set to $\alpha = 0.5$, $\beta = 0.32$ and $\Delta = 0$.

The solution of these models using the vocal folds-tract algorithm suggested in [8] gives us phase space trajectories for the motion of both vocal folds in their displacement and velocity spaces. Examples of such trajectories for COVID-19-positive and normal subjects, derived from the vocalizations of the speakers, are shown in Fig. 2.

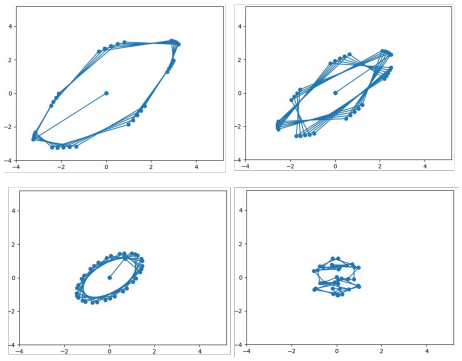

Figure 2: *Examples of displacement phase-space trajectories (right vs. left vocal fold) for patients who tested negative (top row) and positive (bottom row) for COVID-19.*

From these trajectories, we derive various measurements that characterize them in a discriminative manner. These measurements are described below.

3.2. Measurements of phase-space trajectories

3.2.1. Features from a statistical perspective

In addition to the mean displacements and velocities, we use the following features for our analysis. Note that we use displacements ξ_l and ξ_r in the definitions below for brevity, but we can just as easily replace the left and right displacements with velocities $\dot{\xi}_l$ and $\dot{\xi}_r$. We also incorporate velocity measurements in the feature calculations.

1. **Amplitudes of displacement of the left and right vocal fold:** This represents the maximum distance between the equilibrium position and the maximum displacement of the vocal fold during one cycle of oscillation, and is measured in millimeters (mm). Denoting the amplitudes as A_l and A_r, we have $A_l = \frac{1}{2}(\xi_{l,max} - \xi_{l,min})$ and $A_r = \frac{1}{2}(\xi_{r,max} - \xi_{r,min})$ respectively, where $\xi_{r,max}, \xi_{l,max}$ and $\xi_{r,min}, \xi_{l,min}$ are the

maximum and minimum displacement of the right and left vocal folds respectively.

2. **Amplitude ratio**: This represents the relationship between the amplitudes of left and right vocal fold. The ratio is defined as $\frac{A_l}{A_r}$.
3. **Slope and intercept of the regression line fitted to the displacement data points of the left and right vocal folds**: These are computed as:

$$b = \frac{\sum_{i=1}^{n}(\xi_{l_i} - \bar{\xi_l})(\xi_{r,i} - \bar{\xi_r})}{\sum_{i=1}^{n}(\xi_{l,i} - \bar{\xi_l})^2} \tag{2}$$

where b is the slope of the regression line, ξ_{l_i} and $\xi_{r,i}$ are the coordinates of the ith data point for the left and right vocal folds, respectively, and $\bar{\xi_l}$ and $\bar{\xi_r}$ are the sample means of the left and right vocal fold displacements, respectively. Positive values for the slope indicate a positive correlation between the left and right vocal folds. Likewise, negative values indicate negative correlation.

The intercept a of the regression line fitted to the displacement data points is given by $a = \bar{\xi_r} - b\bar{\xi_l}$. It represents the value of the right vocal fold displacement when the left vocal fold displacement is zero (and vice-versa when the dispalcement axes are interchanged).

3.2.2. Features from an information-theoretic perspective

Recall that our data is originally a time-series of continuous values that represents the displacements (or velocities) of the vocal folds. To compute the information-theoretic measurements, we first discretize the data. To compute the thresholds for bins we concatenate all the time-series for the left and right displacements (or velocities), and determine 1000 evenly-spaced bins based on the range of the data. The continuous values for each time-series are then replaced with the ID of the bin that they fall into. To compute a distribution from a time-series, we obtain the counts of observations falling into each bin, and normalize them by the total number of observations. After these steps, we can now calculate the following measures.

1. **Entropy:** This is computed individually for each vocal fold, and indicates the level of uncertainty about a message denoting a value from the distribution of the vocal fold displacements. The computation is computed as:

$$H(\xi_l) = -\sum_{i=1}^{n} p(\xi_{l,i}) \log p(\xi_{l,i}) \tag{3}$$

2. **Mutual Information:** This gives an indication on the amount of information that observing one random variable gives us on the other random variable. In our application this would entail the extent of information contained about one vocal fold trajectory by the other. Given the asymmetry of effects in patients with vocal fold pathology, it is a good candidate to capture differences between individuals in our sub-populations. It is computed as:

$$I(\xi_l; \xi_r) = \sum_{i=1}^{n}\sum_{i=1}^{n} P_{\xi_l,\xi_r}(\xi_{l,i}, \xi_{r,i}) \log \frac{P_{\xi_l,\xi_r}(\xi_{l,i}, \xi_{r,i})}{P_{\xi_l}(\xi_{l,i}) P_{\xi_r}(\xi_{r,i})} \tag{4}$$

where P_{ξ_l,ξ_r}, P_{ξ_l}, and P_{ξ_r} denote the probability mass functions for the joint distribution, marginal of ξ_l, and marginal of ξ_r, respectively.

3. **Kullback-Leibler (KL) Divergence:** This gives us an indication about the difference between the distributions of the left and right vocal fold trajectories. In our application we might expect the KL divergence to be more different in individuals who have COVID-19. It is computed as:

$$D_{KL}(P_{\xi_l} || P_{\xi_r}) = \sum_{i=1}^{n} P_{\xi_l}(\xi_{l,i}) \log \frac{P_{\xi_{l,i}}(\xi_{l,i})}{P_{\xi_r}(\xi_{l,i})} \tag{5}$$

3.3. Detecting differences between populations

We want to determine whether these measurements can differentiate between positive/negative or symptomatic/asymptomatic populations. To test this we use the Student's-t test for difference of means, where the t statistic is

$$t = \frac{\bar{\xi_l} - \bar{\xi_r}}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where the pooled standard deviation s_p is defined as:

$$s_p = \sqrt{\frac{(n_1 - 1)s_{\xi_l}^2 + (n_2 - 1)s_{\xi_r}^2}{n_1 + n_2 - 2}}$$

We determined that the equal variance assumption was appropriate in our situation. Through this test, we can detect whether the features derived from the phase-space trajectories are significantly different between individuals with positive/negative test results, and those who are symptomatic vs asymptomatic.

4. Experiments

4.1. Data

In this paper we use a proprietary dataset from Hematico LLC and Paradigm Biologix, both US-based biomedical and biotechnology companies specializing in healthcare, and another from Merlin, a corporation based out of Santiago Chile. The recordings were obtained from hospitals in Philippines and Chile separately.

The dataset includes recordings of patients speaking the days of the week (DOW), and coughing. All subjects underwent PCR tests. The recordings correspond to 140 positive test results and 100 negative test results. There is also an indication of whether the patient was symptomatic at the time of recording. There was no demographic information of the patients such as age or gender. The breakdown of test results and symptomaticity is as follows: 100 Positive - Symptomatic, 40 Positive - Asymptomatic, 50 Negative - Symptomatic, and 50 Negative - Asymptomatic.

4.2. Implementation Details

The features listed in Section 3.2 were implemented in Python, using the numpy, scipy, and pyinform packages. We ran all experiments on a compute-cluster using 8 x86-64 Intel Xeon CPU's. The VFO trajectories took on the order of 10 hours to run. The calculation of features and the significance testing took on the order of 15 minutes to run in total. Code is available and anonymized[1]. Vocal fold oscillation trajectories were calculated

[1] https://anonymous.4open.science/r/PhonationModeling-041B/README.md

with code released from [6][2].

5. Results

We conducted the t-test as defined above to detect difference of means in positive and negative populations. We ran the significance tests between positive and negative defined in 2 different ways as previously mentioned : a) where positive and negative correspond to the results of the COVID-19 test result, and b) where positive corresponds to whether a patient is symptomatic or asymptomatic regardless of the COVID-19 test result.

In Table 1 we show the results for the significance tests between positive and negative test results. In this table – and all others – the first column indicates the time-series trajectory that we analyze, which is either the displacement or velocity of the vocal folds. The second column shows whether the trajectory was estimated based on a recording of the day-of-the-week (DOW) or a recording of a cough. The third column shows the feature derived over the trajectory. The fourth column shows the p-value for the significance test.

Note that in this table and the following tables, we only report the tests that showed significant results at a level of $\alpha = 0.05$. We performed the tests between all features for all the recording types and time-series types. Not all of these proved to differentiate the populations, so we only report those that were significant.

Recall from the introduction that voice pathology is often characterized by differences in function between the two vocal folds. Our statistical measures reflect this, as the mean values for displacement and velocity of the right fold are significantly different in the positive/negative analysis.

Time-series Type	Recording-Type	Feature	p-value
Displacement	DOW	Mean Right Fold	0.027
Velocity	DOW	Mean Right Fold	0.041

Table 1: *Comparison of statistical measures between positive and negative test result patients*

In Table 2, we look at the differences between the symptomatic and asymptomatic populations. We see slightly different effects compared to the positive/negative test result populations. The mean displacement and velocity of the vocal folds are still relevant, showing that these features are useful in a number of applications. We also see a number of other features derived from the velocity time-series that are more discriminative for the symptomatic sub-population.

Time-series Type	Recording-Type	Feature	p-value
Displacement	Cough	Mean Right Fold	0.037
Velocity	Cough	Mean Left Fold	0.034
Velocity	Cough	Regression Intercept	0.033
Velocity	DOW	Amplitude Ratio	0.038

Table 2: *Comparison of statistical measures between symptomatic and asymptomatic patients*

In Table 3 we show the results for the information-theoretic measurements of the populations differentiated by test results. Our experiments show that the information-theoretic measurements are very discriminative, and almost all the features show the ability to differentiate.

[2]https://github.com/waynezv/PhonationModeling

Time-series Type	Recording-Type	Feature	p-value
Displacement	DOW	Entropy Right Fold	0.001
Displacement	DOW	Entropy Left Fold	0.001
Displacement	DOW	Mutual Information	0.002
Displacement	Cough	Entropy Right Fold	0.008
Displacement	Cough	Entropy Left Fold	0.008
Velocity	DOW	Entropy Right Fold	0.0
Velocity	DOW	Entropy Left Fold	0.0
Velocity	Cough	Entropy Right Fold	0.002
Velocity	Cough	Entropy Left Fold	0.002
Velocity	Cough	Mutual Information	0.05
Velocity	Cough	KL Divergence	0.0

Table 3: *Comparison of information-theoretic measures between positive and negative patients*

Finally in Table 4 we examine the differences in information-theoretic measurements for the symptomatic/asymptomatic populations.

Time-series Type	Recording-Type	Feature	p-value
Displacement	DOW	Entropy Right Fold	0.011
Displacement	DOW	Entropy Left Fold	0.011
Displacement	Cough	Entropy Right Fold	0.0
Displacement	Cough	Entropy Left Fold	0.0
Displacement	Cough	Mutual Information	0.0
Velocity	DOW	Entropy Right Fold	0.012
Velocity	DOW	Entropy Left Fold	0.012
Velocity	Cough	Entropy Right Fold	0.0
Velocity	Cough	Entropy Left Fold	0.0
Velocity	Cough	Mutual Information	0.004

Table 4: *Comparison of information-theoretic measures between symptomatic and asymptomatic patients*

6. Observations and Conclusion

From the tables, we observe some interesting trends. Of the various "simple" statistical measures tested, the simplest – the mean – related significantly to both quantities of interest. However, the correlation between the two trajectories, quantified by the regression fit and amplitude ratio, better matched the presence of symptoms. On the other hand, the entropic measures showed a clear relation to both symptoms and test results. All these show that detailed characterizations of vocal fold oscillations are potentially diagnostic of COVID-19 and its symptoms.

However, more interesting than these results are the *types* of sounds for the results were obtained on. Prior studies on VFO have examined extended vowel sounds, where steady state may be achieved and the vocal folds are likely to take on a steady pattern. In this study, the primary prompts were coughs and recitations of the days of the week, two complex vocal actions that do not allow the vocal folds to fall into steady state movements. Cough sounds, in particular, are unlikely candidates for showing patterns of vocal fold behavior. Nonetheless, our results show that pathologies such as COVID-19 influence vocal folds to a degree that the resulting aberrations in their patterns of movement can be measured directly from the voice signal, even in complex vocal activities such as coughs and chanting. This could extend to potentially any speech or vocal activity, an we now have the computational models and means to measure these aberrations directly from voice.

We believe that this opens up an entirely new avenue of investigation into voice-based analysis of diagnosis of voice-affecting pathologies. We continue to investigate this topic, both through further improving computational models for capturing vocal tract dynamics from voice, and by collaborating with doctors and medical researchers in relating these to illnesses that are known, or suspected to affect any part of the voice-production process as one of their effects.

7. References

[1] B. G. de Boer, "Modelling vocal anatomy's significant effect on speech," *Journal of Evolutionary Psychology*, vol. 8, pp. 351–366, 2010.

[2] H. C. Mahendru, "Quick review of human speech production mechanism," 2014.

[3] S. Deshmukh, M. Al Ismail, and R. Singh, "Interpreting glottal flow dynamics for detecting covid-19 from voice," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1055–1059.

[4] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 3474–3484. [Online]. Available: https://doi.org/10.1145/3394486.3412865

[5] H. Taneja, A. V. S. Abhinav, A. Apoorv, H. Mangal, and N. Agarwal, "Detection of covid-19 using cough sounds," *Fusion: Practice and Applications*, 2022.

[6] W. Zhao and R. Singh, "Speech-based parameter estimation of an asymmetric vocal fold oscillation model and its application in discriminating vocal fold pathologies," 2019. [Online]. Available: https://arxiv.org/abs/1910.08886

[7] J. Lucero, J. Schoentgen, J. Haas, P. Luizard, and X. Pelorson, "Self-entrainment of the right and left vocal fold oscillators," *The Journal of the Acoustical Society of America*, 2015.

[8] W. Zhao and R. Singh, "Deriving vocal fold oscillation information from recorded voice signals using models of phonation," *Information-Theoretic Approaches in Speech Processing and Recognition (Special Issue)*, p. To appear.

[9] B. Pickup and S. Thomson, "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models," *Journal of biomechanics*, vol. 42, pp. 2219–25, 09 2009.

[10] P. Vetter, D. L. Vu, A. G. L'Huillier, M. Schibler, L. Kaiser, and F. A. Jacquerioz, "Clinical features of covid-19," *BMJ*, vol. 369, 2020.

[11] V. Elias, A. A. S. Rabih, S. Bin, H. Aziz, and G. Nassar, "An autonomous acoustic collar to quantify the severity of covid-19 effects by analyzing the vibratory components of vocal and respiratory systems," *2021 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS)*, pp. 1–7, 2021.

[12] A. Yadav, "Predicting covid-19 using random forest machine learning algorithm," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2021.

[13] Z. A. A. Alwaeli and A. A. Ibrahim, "Predicting covid-19 trajectory using machine learning," *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–4, 2020.

[14] V. Despotovic, M. Ismael, M. Cornil, R. Call, and G. Fagherazzi, "Detection of covid-19 from voice, cough and breathing patterns: Dataset and preliminary results," *Computers in Biology and Medicine*, vol. 138, p. 104944, 10 2021.

[15] W. Zhao and R. 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)*. IEEE, 2020, pp. 7344–7348.

[16] W. Zhao, "Computational models for the forensic analysis of human voice," Ph.D. dissertation, Electrical and Computer Engineering Department, Carnegie Mellon University, USA, 5 2022, phD thesis.

[17] J. Zhang, "Vocal fold dynamics for automatic detection of amyotrophic lateral sclerosis from voice," Master's thesis, Computational Biology Department, Carnegie Mellon University, USA, 5 2022, undergraduate thesis.

[18] M. Ismail, S. Deshmukh, and R. Singh, "Detection of covid-19 through the analysis of vocal fold oscillations," 06 2021, pp. 1035–1039.

[19] J. C. Lucero, J. Schoentgen, J. Haas, P. Luizard, and X. 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.