FEATURE ANALYSIS IN AUDIO RECORDINGS FOR COVID-19 DETECTION

Karan Thakkar, Matthew Blunt, Rohit Kumar
kthakka2@jhu.edu, mblunt3@jhu.edu, rkumar44@jhu.edu

1 ABSTRACT

Current COVID-19 testing processes present many barriers such as high cost, lack of availability, and time delay in the results. While the use of audio signals as biomarkers for COVID-19 detection has shown promise in recent work, there is still room for improvement in both signal analysis and classification performance for COVID-19 audio data. Through feature extraction and analysis of multiple audio modalities, this project attempts to provide a signal-level understanding of COVID-19 biomarkers in audio, allowing for better feature selection and classification performance going forward. To accomplish this, COVID-19 audio samples from breathing, cough, and speech recordings were obtained through the Coswara dataset [Sharma et al. (2020)]. Features outlined in the INTERSPEECH 2013 Computational Paralinguistics Evaluation baseline [Schuller et al. (2013)] were extracted, including a set of low-level features along with their statistical functionals. Features were ranked and selected from each modality according to a two-part selection process. Finally, an in-depth analysis of the optimal features was performed in order to provide further insight into effective features for COVID-19 classification. [1]

2 INTRODUCTION

Acoustic features can serve as effective biomarkers for various conditions, including Parkinson's [Tanaka et al. (2011)], Coronary Artery Disease [Schmidt et al. (2015)], PTSD [Marmar et al. (2019)], and others. Detection of COVID-19 through audio has the following benefits: early diagnosis, affordable, widely available, and unobtrusive. The three widely used modalities for diagnosis of COVID are cough, breathing and speech. Brown et al. (2020) explores the effects of COVID on only breath and cough sounds achieving an AUC score above 80%. As an extension, Chetupalli et al. (2021) combines various modalities along with symptoms information collected from the subject. Quatieri et al. (2020) proposes a signal processing framework based on the neuromotor coordination across speech subsystems involved in activities motivated by COVID. Finally, Ismail et al. (2020) provides a brief analysis of vocal folds for diagnosis of coronavirus. Following this section, we present a signal-level understanding of COVID-19 biomarkers in audio, allowing for better feature selection and classification performance.

3 DATASET

In our project, we have worked with the Coswara Dataset [Sharma et al. (2020)], which is composed of contributions from 1699 participants (after preprocessing and filtering we have a total of 1411 samples with 135 positives). Most of the participants are from India (89%). Additionally, the dataset is highly gender-biased with the majority of participants being male (73%). Some of the samples were collected from crowd sourcing and some with the help of hospitals, where each participant contributed 9 audio recordings, namely, (a) shallow and deep breathing, (b) shallow and heavy cough, (c) sustained phonation of three vowels /ae/, /i/, and /u/, and (d) fast and normal pace 1-20 number counting. Alongside this, each participant also recorded current health status (COVID-19 infection, symptoms, and co-morbidity, if any), gender, age, and broad geographical location. Our project focuses on an analysis of three sound categories: (i) breathing-deep (breathing), (ii) cough-heavy (cough), and (iii) counting-normal (speech).

[1] GitHub Repo-https://github.com/carankt/asp_project

3.1 Preprocessing

Since the dataset was crowd sourced, some preprocessing was necessary prior to performing any analysis. To begin with, all the sound samples were standardized to a sampling rate of 44.1 kHz via re-sampling. Additionally, each sample in the dataset is normalized to ± 1. If any sample was less than 100 msec or peak amplitude less than 10^{-4} are discarded. Finally, any initial and trailing silences in the audio files (greater than 50 msec on either side) are removed using a threshold amplitude value of 10^{-4}.

4 Feature Extraction

This project uses low-level descriptors (LLDs) found in the ComParE2016 Feature Set [Schuller et al. (2016)]. These descriptors broadly quantify the energy, spectral, and voicing attributes in an acoustic signal. We further quantify statistical properties of each of the LLDs over time (and/or frequency) approximating to 100 statistical measures quantified for each LLD. Every sound sample is represented by a 6373 dimensional feature vector. The same set of features are derived for all categories of the acoustic signals namely, cough, breathing and speech. An exact list of all features can be found here[2]. These features were extracted using the open source library called OpenSmile [Eyben et al. (2010)].

5 Feature Selection

An important step in any classification process that utilizes signal features is a robust selection process. This allows for the model to utilize features that provide the most telling information as to whether the subject may be infected with COVID-19 or not. Here we present a two-step process for feature selection that takes advantage of both the difference in the signal feature distributions (bottom-up) as well as the classification performance of only that feature in a logistic regression model (top-down). Through this, we aim to achieve high performance while also maintaining an understanding of the feature differences for COVID and non-COVID subjects.

5.1 Bottom-Up Selection

The first step in the feature selection process involved the comparison between the low-level signal feature distributions of COVID and non-COVID audio recordings. The benefit of this method is that it allows for feature ranking according to feature differences which can be observed and analyzed directly - hence providing a signal-level understanding.

The Z-test was used as the bottom-up selection method, as it takes into account both the mean and spread of the distributions being compared. The Z-test equation is shown below:

$$\mathbf{Z} = \frac{\bar{X_1} - \bar{X_2}}{\sqrt{\sigma^2_{X_1} - \sigma^2_{X_2}}} \tag{1}$$

where $\bar{X}$ is the mean and σ is the standard deviation of the feature distribution. A Z-score was obtained for each low-level feature, and features were ranked according to Z-score magnitude. A higher Z-score indicated a larger difference in the feature set between the COVID and non-COVID condition.

It is important to note that the Z-test relies on two main assumptions: sample independence and normally distributed data. Although there were no formal tests performed to confirm the latter assumption, we accept the Z-score metric as a parameter in the feature selection process but do not accept or reject any hypotheses based on this analysis. We simply use the Z-test as a tool for selection and better analysis of feature distributions.

[2]Extracted Feats: `https://drive.google.com/file/d/1w2lGP3EEDLLJWnkvtZgx3gpOq2l2owKn/view?usp=sharing`

5.2 Top-Down Selection

The next step of the feature selection process involved the comparison between different low-level features (and their summary statistics) on their ability to classify between COVID and Non-COVID audio recordings. The benefit of this method is that it will give us the idea for each modality, which features play a significant role in classification. Therefore, rather than using all the features generated by OpenSmile we can just use the subset of them.

We used the Linear Regression (LR) classifier for the binary classification task, where we computed the area under the receiver operating characteristic curve (ROC). This is also referred to as the area under the curve (AUC) metric. The training dataset is randomly divided into 80-20% split (s.t dev set-1125 samples (106 COVID) and test set-286 (29 COVID)). We perform 5 fold cross-validation to tackle the problem of class imbalance. For best fit we apply hyperparameter tunning, and choose the best hyperparameter with the best AUC. Then using the hyperparameter, the classifier is trained on the entire dev set and evaluated on the test set. To further address the challenge of data imbalance, "balanced loss" was used for training. In the balanced loss setting, the loss value for the positive class samples is weighted by the ratio of the number of negative class samples to positive class samples in the training set.

For each modality, we passed all the features extracted from Opensmile to the classifier to see the performance for all features. Then we grouped each low-level feature and its functionals and passed them through the model, and for each modality selected the top performing low-level feature descriptors.

6 Feature Analysis

In the following sections, analysis of each audio modality is presented, highlighting the best-performing features, signal-level differences, and possible biological driving forces that may explain the COVID and non-COVID feature differences.

6.1 Breathing Analysis

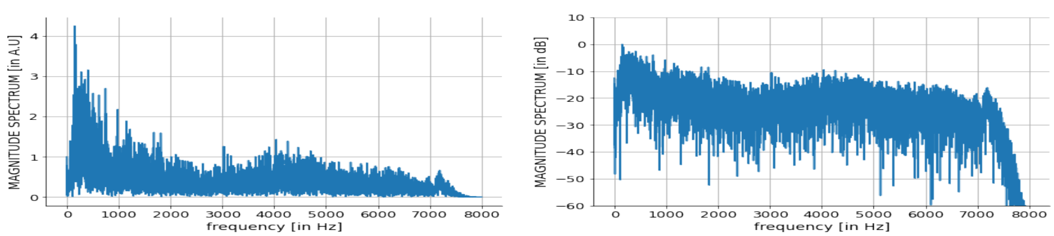

Figure 1: Non-COVID Breathing Spectra

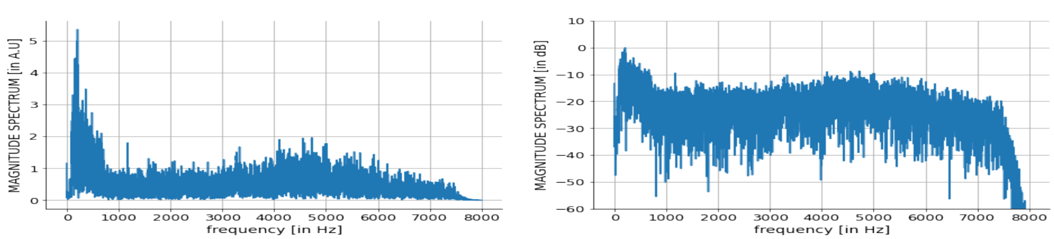

Figure 2: COVID Breathing Spectra

Existing literature discusses the importance of breath duration, an obvious feature as shortness of breath is a common symptom of COVID-19 [O'Horo et al. (2021)]. Signal energy is another common feature that is naturally affected as COVID-19 infection can lead to lower air velocity from the lungs and thus lower acoustic energy [Quatieri et al. (2020)]. Here, the feature selection process yields three additional features that display discriminatory characteristics in the dataset: Spectral Roll-Off, Spectral Centroid, and FFT Magnitude from 1000-4000 Hz.

As discussed in the previous section, bottom-up selection was performed in order to get an idea of each low-level feature distribution. For the breathing modality, spectral features dominated the bottom-up selection Z-score ranking. Further analysis of the average spectrograms for both COVID and non-COVID uncovered visually obvious differences in the way that signal energy was distributed across frequencies. Figure 2 shows a comparison of the average frequency spectrum of 100 breathing recordings for both COVID and non-COVID conditions.

In Figure 2 it can be seen that the average non-COVID spectrum is characterized by a more gradual rolloff in frequency, as well as a flatter distribution of signal energy over frequency. The average COVID spectrum displays a much harsher roll-off around 500-750 Hz, illustrating a key difference that is captured in the three aforementioned features.

This information was also captured in the bottom-up selection process, as the distributions of frequency-related features indicated similar differences. As for the top-down selection, the FFT magnitude from $1000 - 4000$ Hz achieved the best AUC performance of all low-level features with a score of 0.77, followed by spectral roll-off at 0.72 AUC. Thus the classification performance supported the signal-level observations, confirming the effectiveness of spectral features that capture this easily-seen distinction between the two average spectra.

6.2 Cough Analysis

Cough is a powerful reflex mechanism for clearance of the central airways from unwanted materials. Over the years cough has proved to be a valuable modality in the detection of diseases that alter the vocal tract system [Brown et al. (2020), Chetupalli et al. (2021), Quatieri et al. (2020), Avila et al. (2021), Lee et al. (1993)].

Typically, coughing is composed of initial inspiration, leading to glottal closure and development of high thoracic pressure, followed by an explosive expiratory flow of air. Unlike breathing, the glottis vibrates with continued expiratory effort towards the end of the cough [Thorpe et al. (2001)]. The resulting cough sound is a mix of air turbulence and vibration [Polverino et al. (2012)].

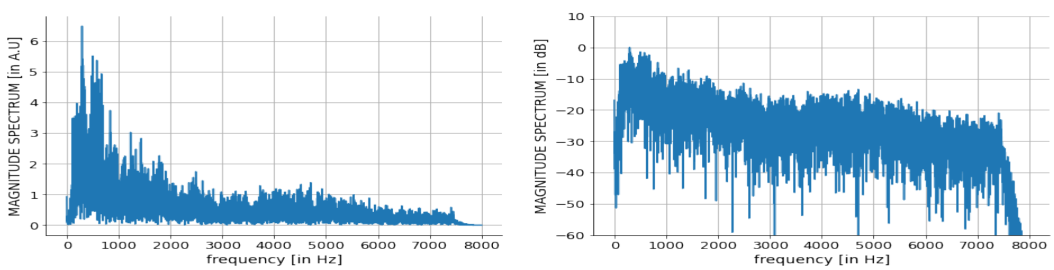

Figure 3: Non-COVID Cough Spectra

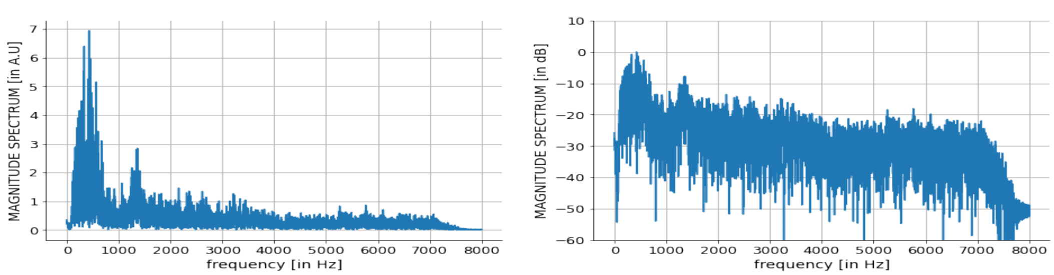

Figure 4: COVID Cough Spectra

In our analysis ZCR and RMS energy features (along with their functionals) came out to be the prominent features. Additionally, as can be seen from Figure 4, a higher proportion of signal energy lies in the higher frequency range for the Non-COVID spectrum as compared to the COVID-19 spectrum. Secondly, the roll-off/decay in the energy of the spectrum was much more prominent in Covid patients compared to Non-Covid patients. Therefore, during feature analysis and selection in cough, all the features which capture this information seemed to be quite prominent like the RASTA-PLP feature, spectral roll-off, spectral centroid.

6.3 Speech Analysis

Studies suggest that lung diseases change lung volumes and breathing patterns for speech production [Lee et al. (1993)]. The problem of vocal fatigue is not uncommon in people with voice disorders. Phonation threshold pressure defined as the minimal lung pressure required to initiate and sustain vocal fold oscillation, is correlated with vocal fatigue [Chang & Karnell (2004)]. It has been suggested that a speaker's perception of increased phonatory effort associated with periods of prolonged voice use is related to increased lung pressure required to initiate and sustain phonation.

Mel Frequency Cepstral Coefficients (MFCCs) are one of the most common features used in speech analysis. During the top-down selection we realized MFCCs play a vital role in classification for COVID and Non COVID. Box plots of 14 MFCCs suggest more concentration of information in lower coefficients. The figure 6 illustrates a higher density of values in the lower frequency range and a less gradual decrease in the spectral roll off. Both approaches suggest MFCCs as an important feature in discriminating between COVID and Non-COVID.

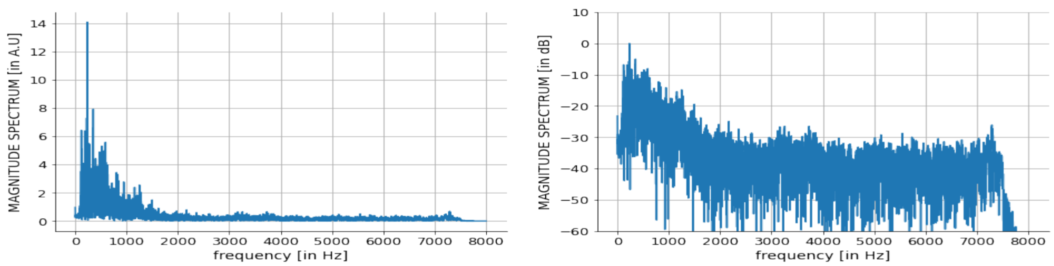

Figure 5: Non-COVID Speech Spectra

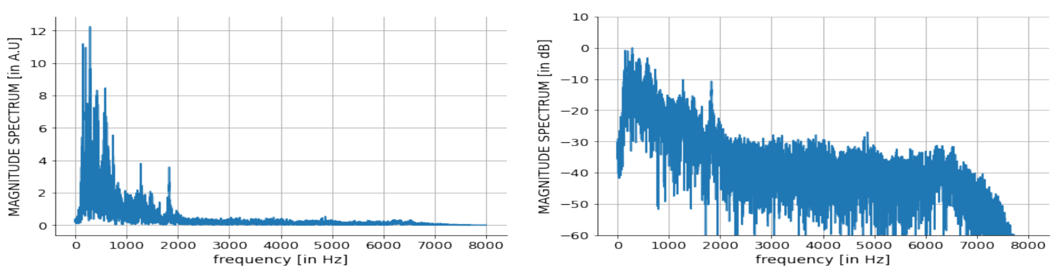

Figure 6: COVID Speech Spectra

Another intuitive feature for discrimination is pitch, however for the task of COVID detection the fundamental frequency did not play an important role as inferred from the bottom-up analysis. As shown in Asiaee et al. (2020), analysis on pitch presented no variation in terms of COVID and non-COVID subjects. In contrast, the variation of pitch across the sample duration as represented by the functionals of the pitch showed potential in classification and performed with an AUC of 0.75. Still, no gender-specific analysis was performed in our study, meaning no significant conclusions can be made.

Features	Breathing	Cough	Speech
Spectral Roll-Off	**0.72**	0.61	0.74
Spectral Centroid	**0.66**	0.51	0.64
FFT Mag. Spectrum 1000-4000 Hz	**0.77**	0.63	0.56
Zero Crossing Rate	0.67	**0.70**	0.67
RASTA Auditory Spectrum	0.71	**0.70**	0.75
RMS Energy	0.67	**0.63**	0.66
MFCC	0.68	0.66	**0.82**
MFCC Delta	0.65	0.63	**0.75**
F0(Pitch)	0.63	0.53	**0.75**

Table 1: Feature-Wise AUC Performance Results Across Modalities

7 DISCUSSION

In this project, we presented a detailed analysis of auditory features across different modalities in order to gain a better signal-level understanding of acoustic biomarkers for COVID-19 detection. This involved the study of three modalities: breathing, cough, and speech. Features were extracted for each modality, and ranked according to a two-part feature selection process. The best performing features were analyzed, including a comparison of the low-level feature distributions and their statistical functionals for COVID and non-COVID conditions.

Overall, we found the results of the study to be relatively encouraging. Analysis for each modality resulted in multiple features that could be useful in a COVID-19 classification task. Interestingly, some features worked well for one modality while not providing significant information for another modality. The resulting table shows the results of highest performing features across the three modalities.

One common trend observed for each of the three modalities was a difference in the spectral envelope of the signal for COVID and non-COVID conditions. Specifically, this phenomenon was characterized by a higher concentration of signal energy in the low frequency bands for the COVID condition when compared to non-COVID observations. As expected, this meant the non-COVID condition was characterized by a flatter, more widely distributed frequency spectrum as compared to the COVID.

Further study involving these features extracted from other datasets could provide information on the extent to which these observations hold true. However, consistency in frequency distribution across modalities shows promise for this type of analysis. Other work in wet/dry cough provides support as well from a biological viewpoint, as COVID-19 is typically dominated by dry cough symptoms which display similar spectral characteristics [Xaviero et al. (2016)].

8 FUTURE WORK

Overall, we see potential in the use of various frequency-related features in the classification of COVID-19 through breathing, cough, and speech. Still, many challenges and unanswered questions remain. While there appeared to be a significant difference in the spectral content of COVID and non-COVID recordings in this study and others, to our knowledge there is no comprehensive study that examines the robustness of features across datasets. Further, our study did not examine differences in features according to gender or geographic location, largely due to an imbalance in the dataset. Analysis of subgroups like these could allow for more specific classification models that take into account more data outside of the recording. Finally, use of audio for COVID-19 classification in any real-world scenario will require studies that examine differences not only between COVID and non-COVID conditions, but also between COVID and other respiratory diseases that may display similar feature differences.

From these observations, future directions of this work include but are not restricted to: Feature Comparison Pre and Post COVID using Novel Methods like i-vectors; Feature Comparison across Datasets; Development of Datasets with More/Cleaner Data; and Deep Learning Methods.

REFERENCES

Maral Asiaee, Amir Vahedian-Azimi, Seyed Shahab Atashi, Abdalsamad Keramatfar, and Mandana Nourbakhsh. Voice quality evaluation in patients with covid-19: An acoustic analysis. *Journal of Voice*, 2020.

Flavio Avila, Amir H Poorjam, Deepak Mittal, Charles Dognin, Ananya Muguli, Rohit Kumar, Srikanth Raj Chetupalli, Sriram Ganapathy, and Maneesh Singh. Investigating feature selection and explainability for covid-19 diagnostics from cough sounds. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, volume 6, pp. 4246–4250. International Speech Communication Association, 2021.

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. *arXiv preprint arXiv:2006.05919*, 2020.

Ann Chang and Michael P Karnell. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. *Journal of Voice*, 18(4):454–466, 2004.

Srikanth Raj Chetupalli, Prashant Krishnan, Neeraj Sharma, Ananya Muguli, Rohit Kumar, Viral Nanda, Lancelot Mark Pinto, Prasanta Kumar Ghosh, and Sriram Ganapathy. Multi-modal point-of-care diagnostics for covid-19 based on acoustics and symptoms. *arXiv preprint arXiv:2106.00639*, 2021.

Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pp. 1459–1462, 2010.

Linda Lee, Robert G Loudon, Barbara H Jacobson, and Roger Stuebing. Speech breathing in patients with lung disease. *American Review of Respiratory Disease*, 147:1199–1199, 1993.

Charles R Marmar, Adam D Brown, Meng Qian, Eugene Laska, Carole Siegel, Meng Li, Duna Abu-Amara, Andreas Tsiartas, Colleen Richey, Jennifer Smith, et al. Speech-based markers for posttraumatic stress disorder in us veterans. *Depression and anxiety*, 36(7):607–616, 2019.

John Charles O'Horo, James R Cerhan, Elliot J Cahn, Philippe R Bauer, Zelalem Temesgen, Jon Ebbert, Andy Abril, Omar M Abu Saleh, Mariam Assi, Elie F Berbari, et al. Outcomes of covid-19 with the mayo clinic model of care and research. In *Mayo Clinic Proceedings*, volume 96, pp. 601–618. Elsevier, 2021.

Mario Polverino, Francesca Polverino, Marco Fasolino, Filippo Andò, Antonio Alfieri, and Francesco De Blasio. Anatomy and neuro-pathophysiology of the cough reflex arc. *Multidisciplinary respiratory medicine*, 7(1):1–5, 2012.

Thomas F Quatieri, Tanya Talkar, and Jeffrey 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*, 1:203–206, 2020.

Samuel E Schmidt, Claus Holst-Hansen, John Hansen, Egon Toft, and Johannes J Struijk. Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering*, 62(11):2611–2619, 2015.

Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al. The interspeech 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*, 2013.

Björn Schuller, Stefan Steidl, Anton Batliner, Julia Hirschberg, Judee K Burgoon, Alice Baird, Aaron Elkins, Yue Zhang, Eduardo Coutinho, Keelan Evanini, et al. The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language. In *17TH Annual Conference of the International Speech Communication Association (Interspeech 2016), Vols 1-5*, pp. 2001–2005, 2016.

Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Prasanta Kumar Ghosh, Sriram Ganapathy, et al. Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis. *arXiv preprint arXiv:2005.10548*, 2020.

Yasuhiro Tanaka, Masaki Nishio, and Seiji Niimi. Vocal acoustic characteristics of patients with parkinsonâs disease. *Folia Phoniatrica et logopaedica*, 63(5):223–230, 2011.

William Thorpe, Miranda Kurver, Greg King, and Cheryl Salome. Acoustic analysis of cough. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*, pp. 391–394. IEEE, 2001.

R Xaviero, A Pramono, S Anas Imtiaz, and E Rodriguez-Villegas. A cough-based algorithm for automatic diagnosis of pertussis. *PLoS One*, 11(9), 2016.