INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Diagnosis of COVID-19 using Auditory Acoustic Cues

Rohan Kumar Das, Maulik Madhavi and Haizhou Li

Department of Electrical and Computer Engineering
National University of Singapore, Singapore
{rohankd, maulik.madhavi, haizhou.li}@nus.edu.sg

Abstract

COVID-19 can be pre-screened based on symptoms and confirmed using other laboratory tests. The cough or speech from patients are also studied in the recent time for detection of COVID-19 as they are indicators of change in anatomy and physiology of the respiratory system. Along this direction, the diagnosis of COVID-19 using acoustics (DiCOVA) challenge aims to promote such research by releasing publicly available cough/speech corpus. We participated in the Track-1 of the challenge, which deals with COVID-19 detection using cough sounds from individuals. In this challenge, we use a few novel auditory acoustic cues based on long-term transform, equivalent rectangular bandwidth spectrum and gammatone filterbank. We evaluate these representations using logistic regression, random forest and multilayer perceptron classifiers for detection of COVID-19. On the blind test set, we obtain an area under the ROC curve (AUC) of 83.49% for the best system submitted to the challenge. It is worth noting that the submitted system ranked among the top few systems on the leaderboard and outperformed the challenge baseline by a large margin.

Index Terms: COVID-19, acoustics, cough, gammatone, constant-Q transform, equivalent rectangular bandwidth

1. Introduction

The COVID-19 pandemic[1] has affected people from more than 200 countries in the recent time. As per present official records, there are more than 125 million cases with over 2.74 million causalities across the globe [1]. Due to its adverse affects in social and professional life, the research investigations for its detection gained wide attention. While the reverse transcription-polymerase chain reaction (RT-PCR) based molecular test is found to be most effective for COVID-19 detection, the rapid antigen test (RAT) is an alternative molecular test to know the result promptly. The latter has a high false negatives, whereas the former is associated with concerns like time, cost and scalability, although it is widely adopted as a standard by most of the countries. This leads various research groups to study different alternatives for the detection of COVID-19.

It has been found that dry cough, difficulty in breathing and chest pains are some of the symptoms shown by the symptomatic COVID-19 infected patients as reported by WHO [1]. Again, the change in anatomy and physiology of the respiratory system is highly related to the generated sounds in terms of cough, breathing and speech as suggested by medical science [2]. Further, cough or difficulty in breathing can affect the continuous speech production for an individual [2, 3]. Due to such reasons, investigating presence of COVID-19 using cough or speech from patients became interest of the community [4,5].

Several attempts have been made for COVID-19 diagnosis using cough and speech from patients [6–10]. The short-term frame-level acoustic features and VGGish embeddings extracted from a pre-trained network are found to be effective as reported in [11]. Again, the usefulness of convolutional neural network (CNN) with mel spectrogram as features and ResNet with short-term magnitude spectrum for investigating COVID-19 from cough samples are also studied [12, 13]. Similarly, an end-to-end system for cough detection followed by classification to find COVID-19 has been explored in [14]. The latest edition of Computational Paralinguistics Evaluation (ComParE) also studies various acoustic front-ends and classifiers for detection of COVID-19 from cough or speech from patients [15].

Along similar directions, the diagnosis of COVID-19 using acoustics (DiCOVA) Challenge[2] aims to promote COVID-19 detection research using cough and speech from individuals to benchmark results of various systems on a common corpus [16]. The challenge runs on two tracks, the Track-1 considers cough sounds, whereas the Track-2 is based on speech from individuals. The Track-1 is a leaderboard style challenge, where the participating teams can upload resultant scores for the blind test set generated using their developed systems. There are three baselines of the challenge that consider widely popular mel frequency cepstral coefficient (MFCC) features as front-end and three different back-end classifiers, which are logistic regression (LR), random forest (RF) and multilayer perceptron (MLP) as shared by the challenge organizers [16]. We participate in the Track-1 of the challenge to detect COVID-19 using cough sounds and report our findings in this work.

Inspired by the success of various acoustic features for detecting COVID-19 from cough or speech [11–13], we focused on exploring a few novel auditory acoustic features during our participation in DiCOVA challenge. We consider long-term transform, equivalent rectangular bandwidth (ERB) and gammatone filterbank based auditory acoustic representations for capturing the artifacts to detect COVID-19 from cough sounds of individuals. The spectrum derived from constant-Q transform (CQT), which is a long-term window transform is used as one of the front-ends. Similarly, the gammatone cepstral coefficient (GTCC) features and the spectrum derived from ERB constitute another two front-ends. We evaluated these representations using the LR, RF and MLP classifiers to submit the best combinations observed from the studies on the validation set of DiCOVA Challenge Track-1 corpus. The contribution of this work lies in exploration of novel auditory acoustic cues for detecting COVID-19 from cough of individuals.

The remainder of the paper is organized as follows. Section 2 describes the different auditory acoustic features considered in this work. The details of the implemented systems and their experimental setup are included in Section 3. Section 4 reports the results and analysis for the systems submitted to the DiCOVA Challenge. Finally, Section 5 concludes the work.

[1] www.who.int/emergencies/diseases/novel-coronavirus-2019
[2] https://dicova2021.github.io/

http://dx.doi.org/10.21437/Interspeech.2021-497

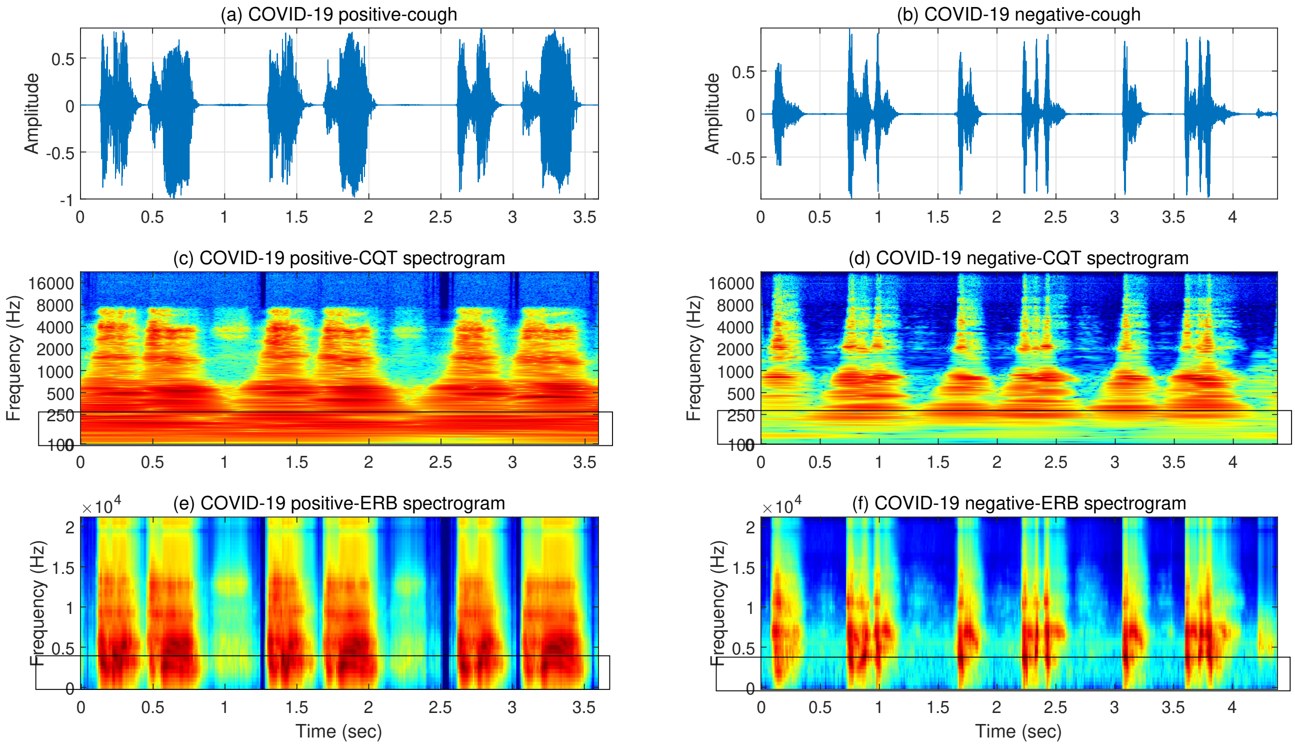

Figure 1: *Spectral analysis for COVID-19 positive vs COVID-19 negative cough.*

2. Auditory Acoustic Cues

In this work, we focus on novel acoustic front-ends for detection of COVID-19 given the fact that the data for the challenge is very limited. We consider long-term transform, gammatone filterbank and equivalent rectangular bandwidth spectrum based auditory acoustic cues for capturing discriminative signal characteristics for detection of COVID-19. We discuss the different auditory acoustic cues considered in this work in the following subsections.

2.1. CQT spectrum

We consider CQT, a long-term window transform [17] to capture acoustic cues for COVID-19 detection. It is different from short-term transform such as Fourier transform, which considers window of few milliseconds. The CQT has a higher frequency resolution for lower frequencies as well as a higher temporal resolution for higher frequencies. Apart from this, it has unique characteristics due to the geometrically distributed center frequencies of each filter. Such salient properties have made CQT based features useful for various classification and detection tasks as reported in the literature [18–22]. Along similar direction, we believe such representations may also be useful for detecting cough speech from COVID-19 positive patients. Therefore, we consider the log power spectrum of CQT as one of the front-ends in this study.

Let us consider a signal $x(n)$, then its long-term transform CQT $Y(k, n)$ is computed as follows

$$Y(k,n) = \sum_{j=n-\lfloor \frac{N_k}{2} \rfloor}^{n+\lfloor \frac{N_k}{2} \rfloor} x(j) a_k^*\left(j - n - \frac{N_k}{2}\right) \qquad (1)$$

where $k = 1, 2, \ldots, K$ represents frequency bin index, N_k are the variable window lengths, $a_k^*(n)$ denotes the complex conjugate of $a_k(n)$, and $\lfloor \bullet \rfloor$ denotes rounding towards negative infinity. The basic functions $a_k(n)$ are complex-valued time-frequency atoms and are defined in [18].

2.2. ERB spectrum and GTCC

Literature shows that the impulse response associated with basilar membrane vibrations are highly correlated with gammatone signals [23, 24]. The mechanical displacement on basilar membrane is regarded as band pass filter and the frequency response can be modeled as gammatone. The ERB frequency scale is a psychoacoustic measure of the auditory filter width on different locations of cochlea [25]. Therefore, the ERB frequency scale is used in the gammatone filterbank design. The spectral energies for different filters are used as erbSpectrum.

The impulse response of the gammatone filter for center frequency f_c is given as follows [23, 26]:

$$g(t) = Kt^{(n-1)} e^{-2\pi Bt} \cos(2\pi f_c t + \phi) \qquad (2)$$

where K is the amplitude factor, n is the order of filter, ϕ is the phase shift and B is the ERB associated with the duration of impulse response, which is $B = 1.019ERB(f_c)$. The center frequencies f_c of gammatone filterbank are spaced equally on ERB frequency scale. The relation between ERB and frequency in Hz, i.e., $ERB(f)$ is as follows :

$$ERB(f) = 21.4 \log_{10}(0.00437f + 1) \qquad (3)$$

The gammatone filterbanks have been widely used for many sound classification research problems [27]. They have some advantages over the conventional mel filterbanks used in

MFCC feature extraction [28]. The gammatone filters have smoother behaviour than mel filters and also used for cough sound recognition study previously [29]. Furthermore, the analysis study on Croup diagnosis using cough sound (which is also a respiratory tract infection that commonly appears in children) presented in [30] observed that the cochleagram representation obtained via gammatone filterbank was found to be effective. Therefore, we believe that ERB spectrum as well as GTCC derived using ERB specturm can be the potential auditory acoustic cues to discriminate the cough sounds from COVID-19 positive and negative patients.

2.3. Spectral analysis observations

Here, we have considered two audio samples from the training set of DiCOVA Challenge database, one from COVID-19 positive and another from COVID-19 negative cough. We are interested to analyze the artifacts using CQT spectrogram and the ERB spectrogram on these samples. Figure 1 shows the spectral analysis comparison of CQT and ERB for the considered samples. We observe that the spectral energy trajectories have relatively more intensity in COVID-19 positive cough than the COVID-19 negative cough sample in case of both CQT and ERB spectrum. The rectangular boxes on Figure 1 show several regions in the low frequency bands showing high discrimination on the spectrum, indicating useful spectral cues for COVID-19 detection. In addition, the discrimination between both classes can also be observed from comparing the contents of high frequency regions. This observations support the motivation of considering CQT and ERB spectrum as potential auditory acoustic cues in this work.

3. System Description

This section describes the details of the database, evaluation protocol and experimental setup. Next, we discuss them in the following subsections.

3.1. Database and evaluation protocol

The database released for DiCOVA Challenge is derived from Coswara corpus[3], which is collected using a crowd sourcing platform from COVID-19 positive and negative individuals [4]. Each participant provided 9 audio recordings that include shallow and heavy cough, shallow and deep breathing, some sustained phonation of vowels and 1 to 20 number counting in fast as well as normal speaking rate using the web-application[4]. The collected data is then divided into two tracks for DiCOVA Challenge. The Track-1 includes the cough sound recordings from the individuals, whereas the Track-2 is composed of deep breathing, vowels and normal speed number counting [16]. We focus only on the Track-1 database of the challenge in this work.

The audio examples of the database are given in 44.1 kHz sampling rate and in FLAC format. The average duration of the Track-1 examples is 4.72 seconds. The database has two subsets, namely train/validation set and test set. A summary of the Track-1 database of DiCOVA Challenge is shown in Table 1. The train/validation set data has to be used for cross-validation and hence, have a split for that. The organizers provided fixed lists for a 5-fold cross validation, which has to be followed by various teams while developing their systems. The average performance across the 5-folds are considered as the final result on

[3]https://github.com/iiscleap/Coswara-Data

[4]https://coswara.iisc.ac.in/

Table 1: *Summary of examples in DiCOVA Challenge Track-1.*

Class	Train/Val	Test
COVID-19 positive	75	blinded
COVID-19 negative	965	blinded
Total	1,040	233

the validation set. On the other hand, the test set of the challenge is blinded and the participants of the challenge can upload their scores on the challenge leaderboard[5]. All participating teams are given maximum of 25 score submissions to the leaderboard during the challenge. It is to be noted that the participants are also required to share the validation set score while submitting the test set scores to the leaderboard. The challenge considers area under the receiver operating characteristic curve (AUC) as a performance metric for evaluation of submitted systems.

3.2. Experimental setup

The DiCOVA Challenge organizers provide three baselines for the challenge. All the three baseline systems consider 39-dimensional MFCC features extracted using `librosa` python library [28]. It is noted that all the samples undergo an amplitude normalization to ± 1 and sample level sound activity detection [16]. The extracted MFCC features are then used by three different classifiers, namely, LR, RF and MLP. The classifiers are implemented using `scikit-learn` python library and their details can be viewed from [16].

We now focus on experimental setup related to our auditory acoustic features used in this study. The CQT spectrum is extracted using `librosa`[6] python library. The logarithm of the CQT power spectrum is then saved as extracted feature for each example. The parameters related to CQT extraction follows our previous work given in [31]. On the other hand, the ERB spectrum (erbSpec) is extracted with 43 number of gammatone filters using `audioFeatureExtractor` of MATLAB audio toolbox[7]. Consideration of 43 gammatone filters make the dimension of erbSpec features as 43. To expand the dynamic range of feature vectors, we applied logarithm to the spectral features. Further, the frames containing very smaller values are expected to be the part of silence and hence removed for feature computation. Similarly, the GTCC features along with dynamic features delta and delta-delta are also extracted using `audioFeatureExtractor` of MATLAB audio toolbox. We note that the window duration is 1024 samples (23.22 ms) and shift is 512 samples (11.61 ms) for 39-dimensional (13-static+13-Δ+13-$\Delta\Delta$) GTCC computation.

Our extracted features are also evaluated using the three classifiers, namely, LR, RF and MLP that are part of the challenge baselines. For this purpose, once our features are extracted, we follow the baseline recipe given by the organizers for training the model using fixed 5-fold cross-validation [16]. During testing, we adopt a slightly different strategy than the organizers, as they utilized all the 5-fold models to generate 5 set of scores for the test set and then finally considered the average score from them. In contrast to this, for our studies, we build a new model considering all the labelled examples from the train/validation set, that we used to test the blinded test set examples. This strategy is adopted as the train set examples are

[5]https://competitions.codalab.org/competitions/29640results

[6]https://librosa.org/

[7]https://www.mathworks.com/help/audio/ref/audiofeatureextractor.html

Table 2: *Performance of various single systems submitted to DiCOVA Challenge.*

System (Feature-Classifier)	Validation AUC (%)	Test AUC (%)
MFCC-LR	64.04	60.83
MFCC-RF	67.71	66.77
MFCC-MLP	69.36	66.09
CQT-LR	65.48	-
CQT-RF	71.95	68.85
CQT-MLP	68.71	71.79
GTCC-LR	64.84	-
GTCC-RF	68.65	72.68
GTCC-MLP	68.61	78.61
erbSpec-LR	67.54	-
erbSpec-RF	**73.41**	**81.89**
erbSpec-MLP	68.17	65.38

very less and we expected that pooling all the examples from train/validation set can lead to an improved model for unknown test set once it is confirmed that respective features perform well on the validation set. Next, we discuss the results and analysis of our submitted systems to the challenge.

4. Results and Analysis

In this section, we report the results and analysis of our submitted systems in DiCOVA Challenge. We group the studies into two parts based on single systems and fusion of multiple systems. The following subsections describe the details based on this grouping.

4.1. Single system studies

We first evaluate the performance of the baseline systems using MFCC features with three different classifiers on the test set. Then we evaluate the three auditory acoustic features explored in this work. Table 2 shows the performance with different combinations of acoustic features and classifiers. On comparing the performances of various features on validations set, we find that our three auditory acoustic features perform better than baseline MFCC features with LR and RF classifiers. Among them, erbSpec-RF achieves the highest AUC of 73.4% on the validation set, which shows its strong potential for COVID-19 detection from cough sounds. In addition, we observed that all the features performed better with RF and MLP classifiers than LR on validation set. It is noted that as the test set results are available for only a limited number of submissions on the leaderboard for each team, we could not report results for a few systems on the test set. Therefore, we submitted the systems using RF and MLP classifiers for our investigated features as they performed better than LR classifier in most of the cases on the validation set.

We now focus on the results obtained on the blind test set that we submitted to the challenge. It is observed that erbSpec features with RF classifier not only performs the best on the validation set, but also on the evaluation set. Further, the GTCC feature with MLP classifier performs as the second best system on the test set. It is worth to be noted that both these single systems outperforms the MFCC feature based baselines by a large margin as can be viewed from Table 2.

Table 3: *Performance for score-level fusion of erbSpec-RF and GTCC-MLP systems submitted to DiCOVA Challenge.*

Weight (α)	Validation AUC (%)	Test AUC (%)
0.5	69.98	82.42
0.9	71.95	**83.49**

4.2. Fusion studies

It is a known fact that combination of multiple systems can contribute towards achieving an improved results [32, 33]. Therefore, we performed some analysis for score-level fusion using the two best single systems discussed above. A weighted score-level fusion is adopted for this study, where we fused the scores of each example obtained from erbSpec-RF and GTCC-MLP as follows:

$$S_{\text{fusion}} = \alpha S_{\text{erbSpec-RF}} + (1-\alpha) S_{\text{GTCC-MLP}} \tag{4}$$

where α is the weighted ratio and $S_{\text{erbSpec-RF}}$, $S_{\text{GTCC-MLP}}$ and S_{fusion} represent the scores from erbSpec-RF, GTCC-MLP systems and the final fused score.

Table 3 shows the analysis for the fusion studies carried out with the two best single systems. We consider two weighted combination cases for this study. For the first one, we consider equal weights for both systems and for the second one, we put a much higher weight for erbSpec-RF system as it performed the best for both validation as well as test set among all our single systems. From Table 3 we observe that the fusion of the two systems improve the performance for both cases, which is more significant when a higher weightage is given for erbSpec-RF system. Thus, our best system submitted to the challenge achieves an AUC of 83.49% on the test set, which lists among the top systems submitted to the challenge. The future work will focus on investigating the considered auditory acoustic features on Track-2 of the DiCOVA Challenge to know their importance for detecting COVID-19 using speech instead of cough sounds.

5. Conclusion

This work devotes on exploring novel auditory acoustic cues for diagnosis of COVID-19 using cough sounds for our participation in Track-1 of DiCOVA Challenge. We considered a few novel auditory acoustic features based on long-term transform, equivalent rectangular bandwidth spectrum and gammatone filterbank. These acoustic representations are used in the framework of three different classifiers, namely, LR, RF and MLP. The developed systems using the auditory acoustic cues outperformed the challenge baseline and the score-level combination of the best two systems among them reported an AUC of 83.49%, which lists one among the top performing systems in the DiCOVA Challenge leaderboard.

6. Acknowledgement

This research work is partially supported by Programmatic Grant No. A1687b0033 from the Singapore Government's Research, Innovation and Enterprise 2020 plan (Advanced Manufacturing and Engineering domain), Human-Robot Interaction Phase 1 (Grant No. 192 25 00054) by the National Research Foundation, Prime Minister's Office, Singapore under the National Robotics Programme. The authors would also like to acknowledge the DiCOVA Challenge organizers for having the interesting tracks and learderboard style of the challenge.

7. References

[1] "WHO Coronavirus Disease (COVID-19) Dashboard," https://covid19.who.int/, 2021, [Online; accessed 27-Mar-2021].

[2] J. E. Huber and E. T. Stathopoulos, *Speech Breathing Across the Life Span and in Disease*. John Wiley & Sons, Ltd, 2015, ch. 2, pp. 11–33.

[3] A. Chang and M. P. Karnell, "Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue," *Journal of Voice*, vol. 18, no. 4, pp. 454–466, 2004.

[4] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, N. R., P. K. Ghosh, and S. Ganapathy, "Coswara – a database of breathing, cough, and voice sounds for COVID-19 diagnosis," in *Proc. Interspeech 2020*, 2020, pp. 4811–4815.

[5] J. Han, K. Qian, M. Song, Z. Yang, Z. Ren, S. Liu, J. Liu, H. Zheng, W. Ji, T. Koike, X. Li, Z. Zhang, Y. Yamamoto, and B. W. Schuller, "An early study on intelligent analysis of speech under COVID-19: Severity, sleep quality, fatigue, and anxiety," in *Proc. Interspeech 2020*, 2020, pp. 4946–4950.

[6] "Cambridge University, UK - COVID-19 Sounds App," https://covid-19-sounds.org/en/, 2021, [Online; accessed 27-Mar-2021].

[7] "Cough Against COVID - Wadhwani AI Institute," https://coughagainstcovid.org/, 2021, [Online; accessed 27-Mar-2021].

[8] "NYU Breathing Sounds for COVID-19," https://breatheforscience.com/, 2021, [Online; accessed 27-Mar-2021].

[9] "EPFL Cough for COVID-19 Detection," https://coughvid.epfl.ch/, 2021, [Online; accessed 27-Mar-2021].

[10] "CMU sounds for COVID Project," https://node.dev.cvd.lti.cmu.edu/, 2020, [Online; accessed 07-Aug-2020].

[11] 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 *Proc. ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery, 2020, p. 3474–3484.

[12] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020.

[13] P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, S. Rane, N. Agarwal, and R. Panicker, "Cough against COVID: Evidence of COVID-19 signature in cough sounds," *arXiv preprint arXiv:2009.08790*, 2020.

[14] W. Wei, J. Wang, J. Ma, N. Cheng, and J. Xiao, "A real-time robot-based auxiliary system for risk evaluation of COVID-19 infection," in *Proc. Interspeech 2020*, 2020, pp. 701–705.

[15] B. W. Schuller, A. Batliner, C. Bergler, C. Mascolo, J. Han, I. Lefter, H. Kaya, S. Amiriparian, A. Baird, L. Stappen, S. Ottl, M. Gerczuk, P. Tzirakis, C. Brown, J. Chauhan, A. Grammenos, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, M. R. Leon J. J. Zwerts, J. Treep, and C. Kaandorp, "The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates," in *Proc. Interspeech 2021*, 2021.

[16] A. Muguli, L. Pinto, N. R., N. Sharma, P. Krishnan, P. K. Ghosh, R. Kumar, S. Ramoji, S. Bhat, S. R. Chetupalli, S. Ganapathy, and V. Nanda, "DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics," 2021. [Online]. Available: https://arxiv.org/abs/2103.09148

[17] J. C. Brown, "Calculation of a constant Q spectral transform," *Journal of Acoustical Society of America*, vol. 89, pp. 425–434, 1991.

[18] M. Todisco, H. Delgado, and N. Evans, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech & Language*, vol. 45, pp. 516–535, 2017.

[19] R. K. Das, J. Yang, and H. Li, "Long range acoustic features for spoofed speech detection," in *Interspeech 2019*, 2019, pp. 1058–1062.

[20] R. K. Das and H. Li, "Instantaneous phase and long-term acoustic cues for orca activity detection," in *Interspeech 2019*, 2019, pp. 2418–2422.

[21] R. K. Das, J. Yang, and H. Li, "Long range acoustic and deep features perspective on ASVspoof 2019," in *Automatic Speech Recognition and Understanding (ASRU) Workshop*, 2019, pp. 1018–1025.

[22] R. K. Das and H. Li, "Classification of speech with and without face mask using acoustic features," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) 2020*, 2020, pp. 747–752.

[23] X. Valero and F. Alías, "Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification," *IEEE Trans. Multim.*, vol. 14, no. 6, pp. 1684–1689, 2012.

[24] Y. Shao, Z. Jin, D. Wang, and S. Srinivasan, "An auditory-based feature for robust speech recognition," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2009*, 2009, pp. 4625–4628.

[25] R. D. Patterson and J. Holdsworth, "A functional model of neural activity patterns and auditory images," *Advances in speech, hearing and language processing*, vol. 3, no. Part B, pp. 547–563, 1996.

[26] M. Slaney *et al.*, "An efficient implementation of the Patterson-Holdsworth auditory filter bank," *Apple Computer, Perception Group, Tech. Rep*, vol. 35, no. 8, 1993.

[27] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," *Expert Systems with Applications*, vol. 136, pp. 252–263, 2019.

[28] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[29] J. Liu, M. You, G. Li, Z. Wang, X. Xu, Z. Qiu, W. Xie, C. An, and S. Chen, "Cough signal recognition with gammatone cepstral coefficients," in *IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP*, 2013, pp. 160–164.

[30] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, "Automatic Croup diagnosis using cough sound recognition," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 485–495, 2019.

[31] R. K. Das, J. Yang, and H. Li, "Data augmentation with signal companding for detection of logical access attacks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2021*, 2021, pp. 6349–6353.

[32] R. K. Das, Abhiram B., S. R. M. Prasanna, and A. G. Ramakrishnan, "Combining source and system information for limited data speaker verification," in *Interspeech 2014*, 2014, pp. 1836–1840.

[33] R. K. Das and S. R. M. Prasanna, "Exploring different attributes of source information for speaker verification with limited test data," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 184–190, 2016.