Cough Sound Classification Based on Similarity Metrics

Nikos Petrellis
Electrical and Computer Engineering Department
University of the Peloponnese
Patra, Greece
npetrellis@uop.gr

George K. Adam
Department of Digital Systems
University of Thessaly
Larissa, Greece
gadam@uth.gr

2021 44th International Conference on Telecommunications and Signal Processing (TSP) | 978-1-6654-2933-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/TSP52935.2021.9522595

***Abstract*— Cough and respiratory sound processing can assist in the early diagnosis of infections such as Covid-19. Even asymptomatic Covid-19 patients can be diagnosed early enough if appropriate speech modeling and signal-processing is applied. Covid-19 affects various speech subsystems that are involved in respiration, phonation and articulation. Based on a symptom tracking platform that was recently presented by the authors (Coronario), we focus on the sound processing subsystem that is capable of classifying cough or respiratory sounds in multiple categories. Specifically, we attempt to classify a cough sound file in one of the following 5 categories: male dry or productive, female dry or productive and child's cough. The classification is performed using Pearson Correlation Similarity, in frequency domain. Several alternative methods that employ averaging and Principal Component Analysis have been tested to estimate their recall and precision/accuracy metrics. The average precision/accuracy achieved is about 75% and 88%, respectively. The sound processing platform used is extensible allowing researches to experiment with several different classification methods applied on the anonymized data exchanged during symptom tracking.**

***Keywords*— sound processing; classification; cough sound; symptom tracking; embedded systems**

I. Introduction

Covid-19 pandemic showed how valuable remote symptom tracking is, since primary health care services had to rely on contactless diagnosis. Morecular tests include Nucleic Acid Amplification such as Reverse Transcription - Quantitative Polymerase Chain Reaction (RT-qPCR)), Immunoassays and Sequencing and are the only way to confirm an infection [1]. However, it is difficult to apply these tests massively.

In the simplest case, symptom tracking can be based on the answers given by the patient in questions about whether he has fever, headaches, cough, fatigue, anosmia, etc. Ten major telemedicine applications are reviewed in [2] that minimize the need to visit a doctor. More advanced techniques search for lesions in pneumonia chest Coaxial Tomography (CT) [3] or X-ray scans [4]. A review on imaging techniques for the detection of Covid-19 can be found in [5].

The coordination of speech neuromotors of the respiration, phonation and articulation is altered if a person is infected by Covid-19. Thus, sound processing and speech models are presented in [6] for tracking both asymptomatic and symptomatic phases of Covid-19. Sound processing of either respiratory or cough sounds can also be employed for Covid-19 diagnosis [7-8].

The purpose of the Coronario platform introduced in [9] is to reduce the traffic in primary healthcare units through remote symptom tracking. It supports real-time screening of vulnerable population with weak immune system while it can be used to monitor the patients in pre- or post-hospitalization phase. It consists of a mobile user and a supervisor application exchanging information through cloud services. It is also combined with eHealth sensors for more reliable diagnosis.

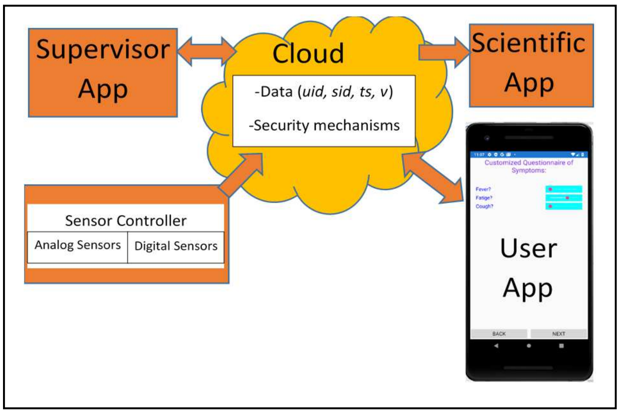

Fig. 1. The architecture of the Coronario platform.

The anonymous medical data that are used in the Coronario platform can also be exploited in the context of research on Covid-19, after patient consent. For this reason, an extensible sound processing platform (ScientificApp) has been developed. We focus on this specific platform in this paper and test several classification methods to demonstrate how it can be used for distinguishing cough sounds that may indicate Covid-19 infection. Since there aren't available public databases with recorded cough sounds from Covid-19 patients we examine a more general problem which is also slightly more difficult: we try to classify cough sounds in one of 5 categories depending on the age, the gender and the dry or productive nature of the cough sound. We do not intend to test all possible sound classification methods. Instead, our purpose is to demonstrate how the developed Coronario ScientificApp can be customized to support and compare different cough or respiratory sound analysis methods. However, the achieved average precision of 75% and accuracy of 88% are quite high taking into consideration that our classification concerns 5 categories and not just a binary decision (e.g., dry or productive cough).

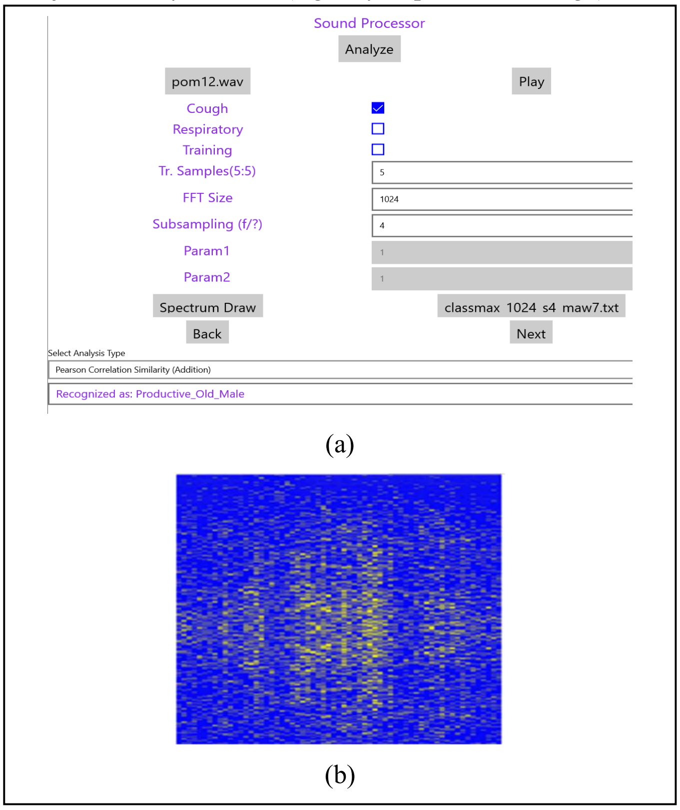

Fig. 2. The main ScientificApp window (a) and an example of visualized sound file in the frequency domain (b) [9].

II. Coronario Architecture

The architecture of the Coronario platform [9] appears in Fig. 1. It consists of 4 modules connected to the Cloud: the user mobile application (UserApp), the supervisor (SupervisorApp) and the scientific (ScientificApp) desktop applications as well as a number of medical IoT sensors that are connected to the Cloud through a Sensor Controller. The information uploaded to the Cloud from the UserApp includes symptom descriptions, the user location and processed or unprocessed recorded sounds (e.g., cough or respiratory) and images. The sensor data and the information from the UserApp are uploaded to the Cloud and can be accessed by the supervisor doctor. Researchers can access anonymized data (e.g., sound files) through ScientificApp after patient consent. The supervisor doctor is responsible for the diagnosis results. All of the information stored in the Cloud can be represented in the following format: (*uid, sid, ts, v*). The parameters *uid, sid, ts* are the user and sensor identity, and the timestamp, respectively. The parameter *v* is string type and it can store a single or multiple integer or floating point values, whole text files such as the Medical Protocols (MP) or even sounds and images encoded as text. Additional meta-data (e.g., sender-recipient of a message can also be incorporated in the *v* field.

Coronario offers a flexible MP file format that allows the dynamic definition of sensor sampling strategies, alert rules and questionnaire formats. Thus, it can support several medical cases with different requirements. Additional services can also be offered by the Coronario platform including user tracking, localization and social distancing. In this paper, we focus on the study of the Coronario sound processing platform in the ScientificApp. We will explain how several cough classification methods can be incorporated and tested in order to demonstrate how the research on Covid-19 or similar infections can benefit from this sound processing platform.

III. Sound Processing and Classification

A. ScientiricApp Description

The main window of the ScientificApp is shown in Fig. 2a. In this page, the researcher can select a sound file (e.g., pom12.wav in Fig. 2a and play it. This sound file can be analyzed through the corresponding button either in the context of training or testing. In training phase, the number of training samples has to be initially defined in the "Tr. Samples (5:5)" field of Fig. 2a. The similarity metric used for the classification in this paper, is applied in the frequency domain. For this reason, Fast Fourier Transform (FFT) is supported and its size is defined in the corresponding field ("FFT size").

The researcher can select to apply the FFT to a subsampled input in order to cover a longer time interval with a single FFT. The number S_m set in the "Subsampling (f/?)" field indicates that only 1 out of S_m consecutive samples will be used as input to the FFT.

The drop down menu "Select Analysis Type" selects the similarity method that will be used. Extra parameters that may be needed by certain similarity methods can be set in the reserved fields "Param1" and "Param2". The magnitude of the FFT output values will serve as the features checked for similarity. They are displayed at the bottom of Fig. 2a when the analysis is completed. These values can be exported for statistical processing outside the application. The researcher can select the file with the reference features of the supported classes ("classmax_1024_s4_maw7.txt" in the case of Fig. 2a) and proceed with the classification. The set of supported classes can be extended if the system is trained with sounds belonging to a new category. The classification results are displayed in the field starting with "Recognized as:" at the bottom of the Fig. 2a page).

B. Similarity Metrics and Classification Methods

Analyzing a sound file may require the application of FFT to multiple segments (depending on the selected subsampling scheme). The output of different FFT segments of the same sound file are either a) averaged (*Add* method) or b) the maximum power of the same frequency in different segments is used (*Max* method). More specifically, let x_i be the *N*-point FFT input vector after subsampling, X_i, be the corresponding output vector and *Nf,* be the number of frames that a single sound file *F* consists of (0≤i<Nf). The file *F* can be represented in time domain as $F_{TD} = x_0 x_1 \dots x_{Nf}$ and in frequency domain as $F_{FD} = X_0 X_1 \dots X_{Nf}$. Although each FFT frame may have different significance we will combine them with equal weights when the sound file signature F_S is estimated:

$$F_S = S(F_D) = S\left(X_0, X_1, \dots, X_{Nf}\right) \tag{1}$$

where *S* in the *Add* method is the averaging of the corresponding elements of the X_i vectors. In the *Max* method the largest value of the corresponding elements of the X_i vectors is used in F_S. During training the signatures F_S of the training sound files belonging to the same class *Ci* are used to generate its feature vector F_{SCi}. In our implementation all the signatures are simply averaged.

When a new file *F* is analyzed, its signature F_S is compared for similarity with all the F_{SCi} reference feature vectors. Pearson Correlation Coefficient (P_C) [10] is used as a similarity metric and is defined as follows:

$$P_c = \frac{N\sum F_S F_{SCi} - \sum F_S \sum F_{SCi}}{\sqrt{N\sum F_D^2 - (\sum F_S)^2}\sqrt{n\sum F_{SCi}^2 - (\sum F_{SCi})^2}}, \qquad (2)$$

Each sum in eq. (2) implies the sum over all the elements of its vector operand. The P_C values are in the range [-1,+1]. If $P_C = 1$ then the compared vectors match while if $P_C = 0$ then, they are totally uncorrelated. If $P_C = -1$, then the minimums of one vector match the maximums of the other and vice versa. We select the class that a sound file belongs to, using the highest P_C value.

We may also apply a smoothing function to the elements of F_S and F_{SCi} in order to compare the fundamental patterns of these vectors. The smoothing functions we have tested are the following two: a) Moving Average Window of 7 or 15 elements (called maw7 or maw15, respectively) and b) Principal Component Analysis (PCA) using *m* out of *n* eigenvectors (called pca2 and pca5 for m=2 or 5 respectively).

PCA [11] is used to smooth or compress data by discovering the fundamental patterns hidden in the high-dimensional data. Let us assume that we want to apply PCA to the F_{FD} vector transformed into a $N \times Nf$ matrix F_{FDM} as (n=Nf):

$$F_{FDM} = X_0^T X_1^T \dots X_{Nf}^T \qquad (3)$$

The covariance matrix R_i with each X_i is:

$$R_i = \sum_{j=1}^{N} X_{ij} X_{ij}^T \qquad (4)$$

where X_{ij} is the j-th element of vector X_i. With Singular Value Decomposition (SVD), R_i can be expressed as:

$$R_i = USV^T \qquad (5)$$

V is a matrix having as columns the right singular vectors of R_i and *S* is the diagonal matrix with the singular values of R_i. Finally, U=[U_1 U_2 … U_{Nf}] is a feature vector with columns being the left singular vectors of R_i.

If we preserve only the first *m* of the *Nf* eigenvectors from the vector *U*, ordered by magnitude then *U* can be reduced to Ur=[U_1 U_2 … U_m], m<Nf. Those U_i vectors correspond to the principal components while *m* represents the level of smoothing or compression. We previously used n=Nf. However, *Nf* is not constant and depends on the sound file duration. Thus, another approach is to split the resulting file signature F_S in *n* segments and create F_{FDM} from these N/n-sized segments. The various combinations of the aforementioned parameters that have been tested are listed in Table I.

TABLE I. ALTERNATIVE SIMILARITY METHODS TESTED

Method Name	***Add/Max, N, S_m, Filtering***
1024_s4	Add, 1024, 4, No filtering
256_s1	Add, 256, 1, No filtering
pca2	Add, 1024, 4, PCA with m=2
pca5	Add, 1024, 4, PCA with m=5
maw15	Add, 1024, 4, MAW with W=15
maw7	Add, 1024, 4, MAW with W=7
max1024_s4	Max, 1024, 4, No filtering
max_pca5	Max, 1024, 4, PCA with m=5
max_maw7	Max, 1024, 4, MAW with W=7

IV. EXPERIMENTAL RESULTS

The sound files were retrieved from SoundSnap (soundsnap.com) and split into 5 categories: child's cough, male dry, male productive, female dry, female productive. The number of sound files is not constant across all categories depending on their availability. The average duration of the sound files was 3 seconds. Ten representative sound files of each category were used for training. The size of the training set ranges between 8% and 33% of the class population. Had a larger training set been used, the achieved precision would be even higher. We use the recall and precision metric to compare the alternative similarity methods listed in Table I. The recall is the fraction of the sound files of a class that were correctly recognized. The precision is the fraction of those files that indeed belong to the class *C* from all the files that were labeled as *C*.

TABLE II. RECALL COMPARISON

Recall	**1024_s4**	**256_s1**	**pca2**	**pca5**	**maw15**	**maw7**	**max1024_s4**	**max_pca5**	**max_maw7**
Child	0.81	0.80	0.87	0.67	0.73	0.80	0.60	0.73	0.67
Male Dry	0.42	0.36	0.48	0.45	0.25	0.25	0.50	0.36	0.52
Fem. Dry	0.56	0.05	0.56	0.61	0.33	0.56	0.72	0.44	0.67
Male Prod.	0.68	0.08	0.52	0.61	0.51	0.60	0.88	0.91	0.82
Fem. Prod.	0.48	0.48	0.41	0.48	0.48	0.48	0.52	0.48	0.65
Average	0.59	0.35	0.57	0.56	0.46	0.54	0.64	0.59	0.66

Table II lists the recall measured for all the alternative methods presented in Table I for the 5 categories of cough defined earlier. Similarly, Table III lists the corresponding precision. The combinations that were listed in Table I, were selected based on the following facts: initially, the unfiltered data were tested using either (N=1024, S_m=4) or (N=256, S_m=1). As can be seen from Table II, 1024_s4 achieves better precision than 256_s1 in all of the cases. Consequently, the rest of the combinations tested, used: N=1024 and S_m=4. With *Add* method, when either moving average or PCA was applied, the results were worse when heavy smoothing was applied (maw15, pca2). After these observations, the *Max* was applied only to the methods that achieved the best results with the *Add* i.e.: 1024_s4, pca5 and maw7.

TABLE III. PRECISION COMPARISON

Precisions	1024_s4	256_s1	pca2	pca5	maw15	maw7	max1024_s4	max_pca5	max_maw7
Child	0.72	0.21	0.46	0.53	0.38	0.44	0.53	0.79	0.67
Male-Dry	0.74	0.32	0.44	0.50	0.35	0.36	0.84	0.63	0.77
Fem.-Dry	0.33	0.25	0.48	0.46	0.21	0.33	0.62	0.50	0.60
Male Productive	0.64	0.38	0.68	0.63	0.67	0.70	0.68	0.65	0.68
Fem. Productive	0.58	0.26	0.50	0.58	0.48	0.58	1.00	0.85	0.79
average	0.60	0.28	0.51	0.54	0.42	0.48	0.73	0.68	0.70

TABLE IV. CONFUSION MATRIX FOR MAX_1024_S4

	Child	**Male Dry**	**Female Dry**	**Male Productive**	**Female Productive**
Child (20)	18	0	2	10	0
Male Dry (66)	4	34	0	28	0
Female Dry (36)	2	0	26	8	0
Make Productive (132)	8	4	4	116	0
Female Productive (46)	2	2	10	8	24

As can be seen from Table III the highest average precision is achieved with *Max* method without filtering although MAW7 is quite close and achieves better average recall. Table IV is the confusion matrix for the case of max_1024_s4. In the first column of this table the population of each class is listed in parentheses.

In the literature, the problem of distinguishing cough from noise is handled with a recall of about 80% [12]. Distinguishing wet from dry cough is achieved in [13] with recall and accuracy: 75% and 76%, respectively. The accuracy metric differs from precision: it takes into account the fraction of both the samples that were correctly recognized as positive or negative. The corresponding accuracy in our case is 88%. It has to be stressed that both [12] and [13] address a binary decision problem which is a much easier task than the classification in 5 categories.

V. CONCLUSION

An extensible sound processing platform was described that analyzed and classified cough sounds in 5 categories with average recall, precision and accuracy of 64%, 73% and 88% respectively. Several parameters of this sound processing platform can be configured including classification methods, similarity metrics, FFT size, filtering methods, etc. Several combinations of these parameters have already been tested to demonstrate the extensibility of this platform.

Future work will focus on testing the platform on the classification of Covid-19 cough sounds. Proper classification methods will be investigated for this case study.

REFERENCES

[1] K.K.-W. To, et al. “Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study,” The Lancet Infectious Diseases 2020, doi:https://doi.org/10.1016/S1473-3099(20)30196-1.

[2] S. Bahl, R.P. Singh, M. Javaid, I.H. Khan, R. Vaishya, R. Suman. “Telemedicine technologies for confronting COVID-19 pandemic: A review,” Journal of Industrial Integration and Management, 2020, vol. 5(4), pp. 547–561, doi: https://doi.org/10.1142/S2424862220300057

[3] G. Wang, et al. “A Noise-robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images,” IEEE Transactions on Medical Imaging, 2020, doi: 10.1109/TMI.2020.3000314.

[4] M. Abdel-Basset, R. Mohamed, M. Elhoseny, R.K. Chakrabortty, and M. Ryan. “A Hybrid COVID-19 Detection Model Using an Improved Marine Predators Algorithm and a Ranking-Based Diversity Reduction Strategy,” IEEE Access 2020, vol. 8, pp. 79521-79540, doi: 10.1109/ACCESS.2020.2990893.

[5] D. Dong et al., "The role of imaging in the detection and management of COVID-19: a review," IEEE Reviews in Biomedical Engineering, doi: 10.1109/RBME.2020.2990959.

[6] T. Quatieri, T. Talkar, J. Palmer. “A Framework for Biomarkers of COVID-19 Based on Coordination of Speech-Production Subsystems,” IEEE Open Journal of Engineering in Medicine and Biology 2020, doi: 10.1109/OJEMB.2020.2998051.

[7] Z. Jiang, M. Hu, G. Zhai. “Portable Health Screening Device of Respiratory Infections,” In Proceedings of the 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, United Kingdom, 2020, pp. 1-2, doi: 10.1109/ICMEW46912.2020.9105969.

[8] C. Uysal, A. Onat, T. Filik. Non-Contact Respiratory Rate Estimation in Real-Time With Modified Joint Unscented Kalman Filter. IEEE Access 2020, vol. 8, pp. 99445-99457, doi: 10.1109/ACCESS.2020.2998117.

[9] N. Petrellis. “A COVID-19 Multipurpose Platform,” Digit Biomark 2020; vol. 4, pp. 89-98. doi: 10.1159/000511704

[10] K.A.S. Immink, J.H. Weber. “Minimum Pearson Distance Detection for Multilevel Channels With Gain and/or Offset Mis-match,” IEEE Transactions on Information Theory, 2014, vol. 60, no. 10, pp. 5966-5974, doi: 10.1109/TIT.2014.2342744.

[11] I.T. Jolliffe. Principal Component Analysis. Springer Series in Statistics. New York: Springer-Verlag. 2002. doi:10.1007/b98835. ISBN 978-0-387-95442-4

[12] S.J. Barry, A.D. Dane, A.H. Morice, et al. “The automatic recognition and counting of cough,” Cough 2006, vol. 2, no. 8, https://doi.org/10.1186/1745-9974-2-8 [7]

[13] Y.A. Amrulloh, D.A.R. Wati, F. Pratiwi, R. Triasih. “A novel method for wet/dry cough classification in pediatric population,” In Proc. of the 2016 IEEE Region 10 Symposium (TENSYMP), Bali, 2016, pp. 125-129, doi: 10.1109/TENCONSpring.2016.7519390.