HAL open science

Cough sound recognition for COVID-19 risk detection

Marine Diniz, Vincent Zalc, Dan Istrate

To cite this version:

Marine Diniz, Vincent Zalc, Dan Istrate. Cough sound recognition for COVID-19 risk detection. JETSAN 2021 - Colloque en Télésanté et dispositifs biomédicaux - 8ème édition, Université Toulouse III - Paul Sabatier [UPS], May 2021, Toulouse, Blagnac, France. 10.34746/v7mv-4x55 . hal-03501195

HAL Id: hal-03501195

https://hal.science/hal-03501195v1

Submitted on 23 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cough sound recognition for COVID-19 risk detection

M. DINIZ[1], D. ISTRATE[1], V. ZALC[1]

[1] Université de technologie de Compiègne, CNRS, Biomechanics and Bioengineering,

Centre de recherche Royallieu

CS 60 319 - 60 203 Compiègne Cedex

marine.diniz@etu.utc.fr

{dan.istrate, vincent.zalc}@utc.fr

Abstract

The COVID-19 outbreak prompts the need for new ways to detect and prevent epidemics. Since cough is one of the COVID-19 symptoms, our work proposes a sound recognition system based on our previous works which are able to detect different type of cough through a continuously home sound analysis. It is a part of EpiSemioWatch AMI COVID UTC funded project which aims to develop a detection system of epidemic risk in retirement home using sound and movement sensors. This paper presents the sound part of the global system and more explicitly the differentiation of cough from other environmental sound. The proposed system detects automatically the sound activity through a wavelet based system, realize an acoustic segmentation speech/sound and classify the cough sounds. A first evaluation of the system in our lab is presented.

Cough sound recognition, Covid-19, online sound classification, ivectors

1 Introduction

Since January 2020, the pandemic led by the COVID-19 disease has infected more than two millions people in France and has killed 50 000 of them [1]. The median age of death is 77 years old which makes this disease especially virulent for elderly people. Resident homes are especially vulnerable because their group living setting and the age of their resident. Therefore, there is an imperative need to detect early COVID-19 symptoms to minimize the spread of the virus in those homes but also to aid clinical testing.

2 State of Art

2.1 Clinical testing

During this year, many tests have been proposed to expand the diagnosis of COVID-19. The gold standard clinical test is the Reverse Transcription Polymerase Chain Reaction (RT-PCR), it has the best sensibility and specificity but is not easy to realise, needs healthcare staff and laboratory and does not allow a continuous control [2]. Thus the deployment of antigenic tests has made it possible to broaden the detection of the disease with its response time of 15 minutes, unfortunately this test is also known not to have the best specificity [3].

Finally, X-ray and CT-scan are also capable of observing a COVID-19 infection obstruction in the Patient lungs' with a good accuracy. However, they still require a visit to a well-equipped clinical facility and the need of a radiologist.

2.2 Sound based detection system

Additionally, cough detection has gained interest in regard of this issue: it is a common and early symptom of COVID-19 and prior works have shown that this sound is detectable. Apps like AI4COVID19 have been deployed to detect instantly COVID-19 Cough. AI4COVID19 uses the Mel-Spectogramme image of the cough sound and deep learning. Unfortunately, cough is a common symptoms in numerous pathologies. Therefore to build such a model, the data needs for deep neural networks are substantial. Their solution was to put part of the system in the cloud even if it could makes the latter less secure [2].

One can also take as a reference the work of ROBIN et al. [4]. Indeed, the aim of it was to detect distress signs and different activities of daily living (ADL) with sound. And his work tackled the cough recognition but the cough recognition rate is far from being good (35,48%).

What makes the solutions presented above unsuitable for detection in community homes is that they require the active participation of a fragile population, which also mean that the person must have all their locomotor functions and not forget to get tested. Another problem is that these solutions require specific equipment, the movement of the resident or the intervention of a specialist.

In this paper an improved system capable to detect cough sounds based on Robin et al. work [4] is presented.

3 Proposed System

Sound at home contains speech, every day life sounds and human sounds (cough, sneeze, yawn, cries, cry,...). The most developed part are the speech recognition and speaker recognition. For a few years now the everyday life sounds are recognized. Thus, the acoustic characteristics of coughing have the potential to be recognised. The originality of our system is that the recognition is embedded on a system consisting of a Raspberry PI 4, which thus makes it possible to detect sounds continuously without user intervention, while respecting the consent and privacy of the persons being monitored. This system could possibly be coupled with motion and door sensors to detect other symptoms of COVID-19.

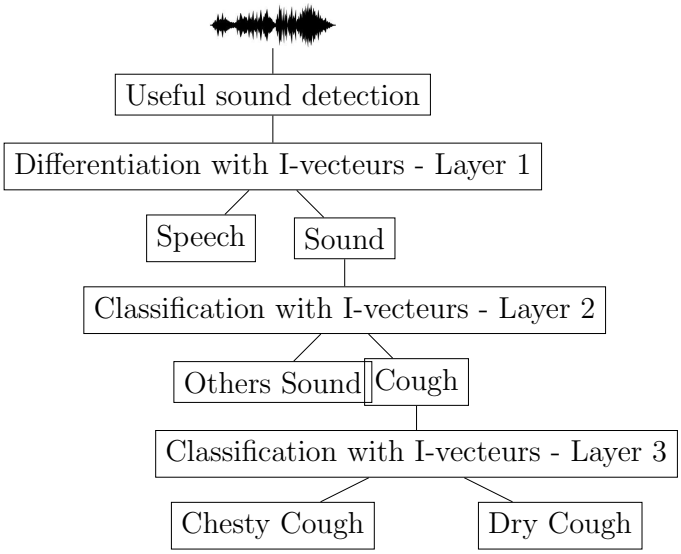

Figure 1: Flow chart of the proposed system

The purpose of this paper is to present our cough recognition system for epidemic. The idea is to detect multiple times the same kind of coughing. We decided to work on a statistical recognition model in order to obtain a portable and low resource consuming system. This choice is also based to our experience in sound environment analysis for distress situations.

A statistical recognition system is generally made of the extraction of vectors form the raw signal which are called features, the generation of a statistical model from those features and the test

file matching with the model using probabilistic methods [5].

To do this, the idea is to detect an interesting sound event and to recognise whether this event is a sound or a speech (layer 1) then to identify whether it is a coughing sound or not (layer 2) as presented in the flowchart above (figure 1). Finally, we would like to classify two types of cough: dry and chesty (layer 3) based on the hypothesis that COVID-19's cough is special.

3.1 Acoustic features

Sounds can not be easily interpret by their waveform, most of the time we extract their acoustic features to identify the sounds event. Mel-frequency cepstral coefficients (MFCC), inspired by the human hearing system, are one of the popular spectral acoustic parameters to extract form waveforms. They are known for their performance in speech and sound recognition. Others cepstral decomposition can be obtained like Linear Frequency Cepstral Coefficients (LFCC), which are used for speaker recognition [4]. The number of coefficients extracted most used is 24 because allows an accuracy spectral decomposition. It is also possible to use the MFCCs' derivatives to add a temporal component but our tests have shown that they did not improve the detection of coughing.

With this features, we decided to add three others parameters : Spectral Roll Off-Point (RoP), Spectral Centroid (SC) and Remarkable Energy Rate (RER). They can quantify the distribution of energy in the waveform and the shape of its envelope.

Spectral RoP is the frequency point where already 95% of the energy have been distributed. SC represent the center of gravity of the power spectrum of the waveform. RER represents the high energy envelope of the signal [4].

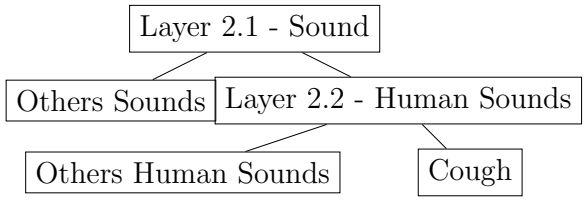

Figure 2: Hierarchical Test for layer 2

3.2 Classification model based on iVectors

I-vectors were initially introduced for speaker recognition but they have become with time very reliable for speech and sound processing. They can be seen as a more compact representation of GMM (the vector is projected in a low dimensionality space named Total Variability Space) :

$$M = m + Tw \tag{1}$$

Where M is the GMM channel-dependant super-vectors, m the channel-independant super-vectors of an Universal Background Model (UBM), T is the matrix containing both channel and sources variability and w is the resulting i-vector [6].

3.3 Hierarchical classification for performances amelioration

Older works have shown that Human vocalized are less recognised than other sound and our tests reinforces this result. Therefore, we decided to break down layer 2 into two sub layers (named 2.1 and 2.2) which recognize Human sounds before the Cough (figure 2).

There are different ways to realize the hierarchical classification. Several performances evaluation was made and two configurations seem to be the best.

4 Evaluation

4.1 Data bases

All our audio files used are coded in 16-bits format with a sampling rate of 16kHz and uncompressed (wav extension). To test our system we first choose to work with the same data base used in Maxime ROBIN works [4]. It contains 1049 files classified in 18 clusters. To separate sound from speech, we add another data base of speech of 2646 speech wav files.

However this data base is flawed since the Cough samples doesn't have known characteristics attached to them (like dry or chesty). We decided to contact a general practitioner to record cough sound with their tendency, we have collected cough of 34 patients (a total duration of 4 min) and some of them did have the COVID-19 infection at the moment of recording. The average age in this database is 49 with a standard deviation of 20 years. There are also 7 people over 70 out of the 34 patients. Table 1 present a example of the samples we collected.

Table 1: Example of samples

File name	Cough Tendency	Age	Sex	Smoking
patient_1.wav	dry	52	M	No
patient_2.wav	dry	72	F	No
patient_3.wav	chesty	31	M	Yes

We have computed the spectral composition of a COVID cough and a non COVID example in the Figures 3 and 4.

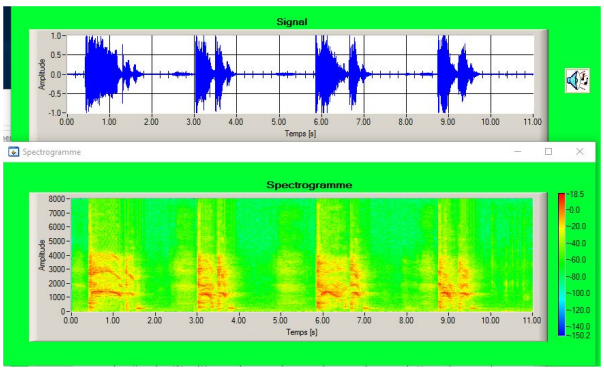

Figure 3: COVID cough example specgram

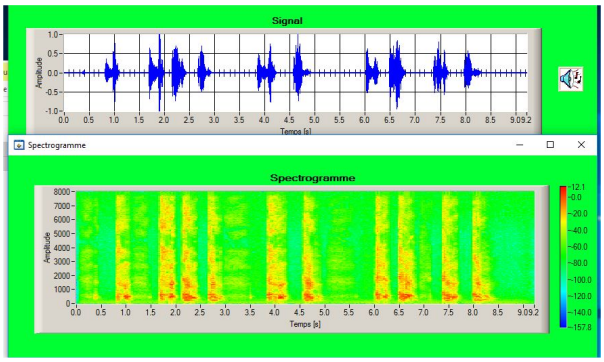

Figure 4: Non COVID cough example specgram

4.2 Evaluation method

To build a complete i-vector system, we have used the ALIZE library. ALIZE is an open-source platform for speaker recognition, it contain a toolkit of different sound recognition method such as Join Factor Analysis, Support Vector Machine, I-vector modelling and Probabilistic Linear Discriminant Analysis. It is also kept updated on the latest development in sound recognition [5].

In order to have statistically correct results despite the reduced number of files, we used the k-fold method: we divide the database into k sub equal bases, then give k-1 sub bases for learning and the last one for testing which is executed k times to test the whole base (figure 5).

We also decided to use 3 indicators to mesure and compare the results : precision, recall and F-measure.

4.3 First classification: sound/speech differentiation

We have evaluated different type of acoustic features that in literature has good results. The results were conclusive and adding the parameters RER, ROP and Centroid did not improve anything. We still tested both MFCC and LFCC to see which one was more interesting if there was a difference.

Parameters	24 MFCC	24 MFCC+ RER+SRF+SC	24 LFCC
TP	1031	932	1003
FP	43	29	65
FN	18	117	46
TN	2603	2617	2581
Precision	96,00%	96,98%	93,91%
Recall	98,28%	88,85%	95,61%

The better recognition was obtained by the MFCC which was chosen for the following.

4.4 Second subdivision: layer 2-1 - Separating Human sounds from other everyday life sounds

Two subdivisions have caught our interest: Human sounds against others classes of sounds and Human sounds against one merged class of others sounds. The strictest tests (with more than 2 classification classes) such often have the best precision (i.e. there are fewer false alarms), whereas

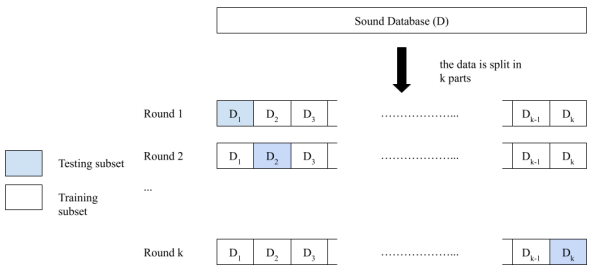

Figure 5: The k-folds method

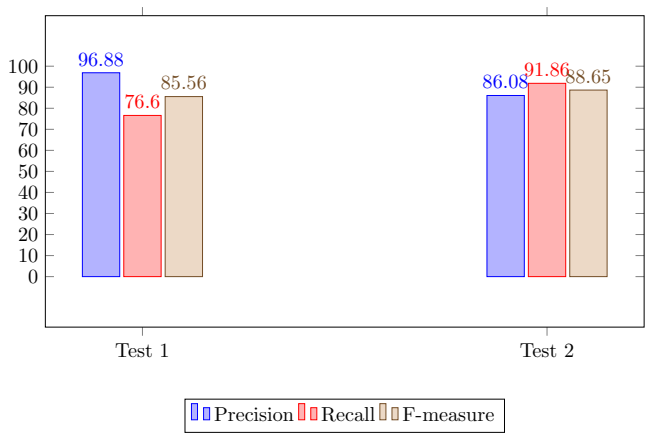

Figure 6: Results for layer 2-1: Human Sounds vs. Other sounds. Test 1 = Human vs. Other Classes; Test 2 = Human vs. Other Sound Class

the most flexible tests (with 2 classification classes) often have the best recall.

We have tested different sets of parameters, here are the two tests we decided for :

- **Test 1**: *Human* vs. Other Classes; Parameters: 24MFCC + RER + SRF
- **Test 2**: *Human* vs. Other Sound Class; Parameters : 24MFCC + RER + SRF + SC

The precision is better for the test 1 but overall is better for the second one (figure 6). The choice has made to use the configuration Test 2.

4.5 Third subdivision : layer 2-2 - Detecting Cough

The layer 2.2 shape is Cough against Sneeze against one other human merged class. We kept Sneeze sounds separated since Sneezing can also be a sign of respiratory illness.

Test:

- Layer 2-2: *Cough* vs. *Sneeze* vs. Class *Other*; Parameters: 24MFCC + RER + SC

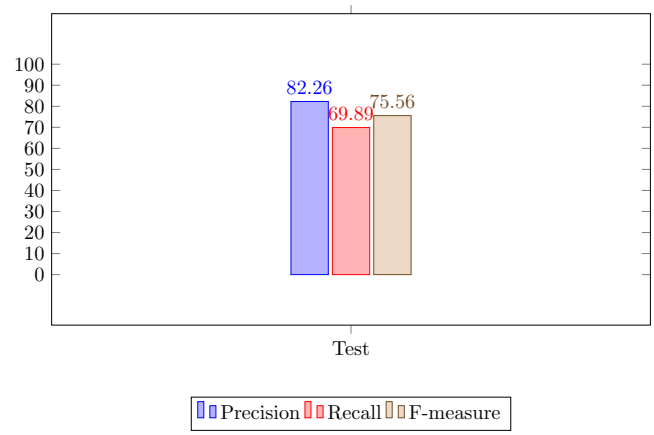

Figure 7: Results for layer 2-2: Cough vs. Other Human sounds

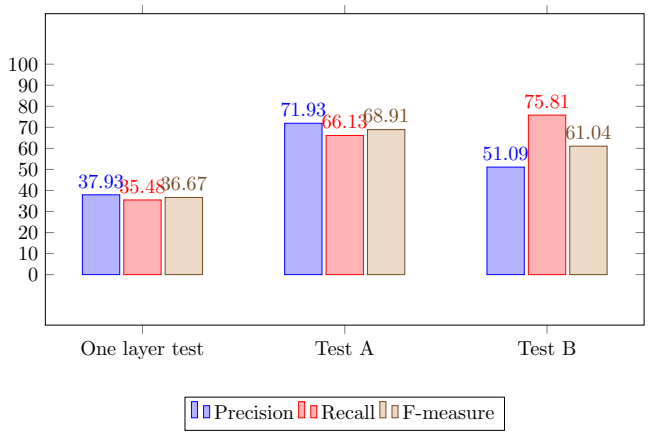

Figure 8: Comparison between single-layer and two-layer tests. Test A = Human vs. Other Classes followed by Cough vs. Sneeze vs. Class other; Test B = Human vs. Other Sound Class followed by Cough vs. Sneeze vs. Other

This give us the reference results for the next classification step (figure 7).

Using the full hierarchical classification steps we present the two best tests:

Test A:

- Layer 2-1: *Human* vs. Other Classes; Parameters: 24MFCC + RER + SRF
- Layer 2-2: *Cough* vs. *Sneeze* vs. Class *Other*; Parameters: 24MFCC + RER + SC

Test B:

- Layer 2-1: *Human* vs. Other Sound Class; Parameters: 24MFCC + RER + SRF + SC
- Layer 2-2: *Cough* vs. *Sneeze* vs. Other Human Sound Class; Parameters: 24MFCC + RER + SC

The final results of the tests is shown on the figure 8. We can assume a good results for overall for the test A with the best recall in the test B.

4.6 Fourth subdivision : layer 3 - Detecting the Cough type

We decided to use the "leave-one-out cross-validation method" (LOOCV) to test our system because of the number of cough collected, which is a special case of the k-folds method. With LOOCV, we test a single file and leave the rest of the database for learning, we repeat the operation as many times as there are files. The result shown in the figure 9 is with the best combination of parameters (24LFCC + SRF). The percentage of good recognition remains quite low (66.66% for dry cough and 40% for chesty cough), this could be explained by the fact that there was more dry cough than chesty cough in the recordings.

5 Conclusions and Perspectives

While our results are encouraging, they are still not as reliable as would be required for a system without user intervention. We are hoping that the complete system using also motion sensors and AI would improve the results. However, this system is light by using I-vecteurs models and was implemented on a Raspberry PI card.

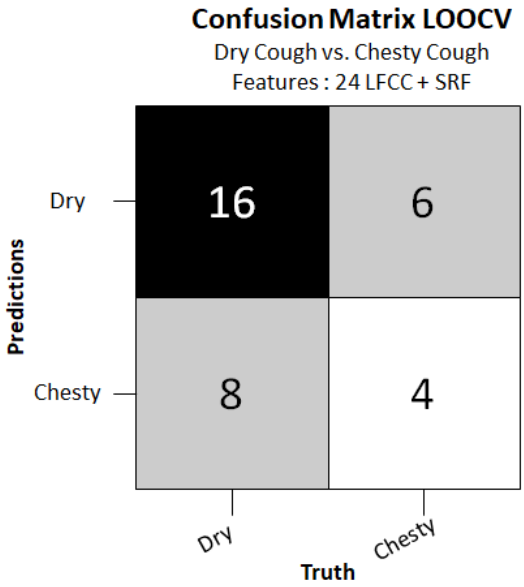

Figure 9: Results for the 34 cough files (24 dry Cough and 10 chesty Cough)

Acknowledgments

We would like to thank the medical practice of Chouilly (France) for providing us with Cough sounds. This project was carried out thanks to the funding of UTC-AMI-COVID-19

References

[1] Santé Publique France. Point épidemiologique en france du 5 novembre 2020. `https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-5\-novembre-2020`, 2020.

[2] Ali Imran, Iryna Posokhova, Haneya N. Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N. John, MD Iftikhar Hussain, and Muhammad Nabeel. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20:100378, 2020.

[3] DALOUR Suzie. Revue rapide sur les tests RT-LAMP sur prélèvement salivaire (hors système intégré de type EasyCoV). 2020.

[4] Maxime Robin, Dan Istrate, and Jerome Boudy. Remote monitoring, distress detection by slightest invasive systems: Sound recognition based on hierarchical i-vectors. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2744–2748, Seoul, July 2017. IEEE.

[5] Anthony Larcher, Jean-François Bonastre, Benoît Fauve, Kong Aik Lee, Christophe Levy, Haizhou Li, John Mason, and Jean-Yves Parfait. ALIZE 3.0-Open Source Toolkit for State-of-the-Art Speaker Recognition. page 6, 2018.

[6] Anthony Larcher, Pierre-Michel Bousquet, Kong Aik Lee, Driss Matrouf, Haizhou Li, and Jean-Francois Bonastre. I-vectors in the context of phonetically-constrained short utterances for speaker verification. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4773–4776, Kyoto, Japan, March 2012. IEEE.