

## COUGH CLASSIFICATION TOOL FOR EARLY DETECTION AND RECOVERY MONITORING OF TUBERCULOSIS AND ASTHMA

Dr.Y.PADMA SAI , S.V.N NARAYANA RAO, V NAVEEN KUMAR, V PRIYANKA BRAHMAIAH,D AJAY

Electronics and Communication Department  
VNR VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY  
Bachupally, Nizampet (S.O), Hyderabad, Telangana, India-5000090

### ABSTRACT

Now-a-days there is significant increase in research for acoustic signal analysis in prospect of healthcare domain, one such area of interest is cough signal analysis which is major symptom of chronic diseases like tuberculosis and asthma. Cough can appear sporadically with common illnesses (e.g. cold), but when it becomes chronic in the case of asthma and tuberculosis which severely impair life quality. Though there are existing systems in analyzing cough signal, still there is a need in developing tool which analyses cough signal and capable of detecting the disease at earlier stages as well as monitoring the recovery of patients suffering either tuberculosis or asthma. The software tool proposed will take cough sounds as input and classifies them as wet or dry cough by features MFCC and classification algorithm K-Means with an efficiency of 82%. The classification will further be used as part of the tool being developed for monitoring the recovery of the diseases like tuberculosis, asthma.

**Keywords-** Wet Cough, Dry cough, MFCC K-means, chronic, tuberculosis, asthma

### I.INTRODUCTION

Cough is one of the most common symptoms among all respiratory diseases. The main activity of the cough is to clear the breathing airways from foreign objects, secretions and mucus. Depending on how often it occurs and its severity, cough may persist and become chronic in nature. One third of the pulmonologist consultations are chronic cough diseases.

Primarily coughs are of two kind's wet and dry cough. Differentiation of wet and dry coughs are very subjective, it may be difficult for patients to describe their cough symptoms to doctor, which makes the medication and treatment difficult at times, so there is significant focus in acoustic signal analysis for healthcare in early detection and monitoring. In this paper we discuss about the tool

developed. This tool takes cough signal as input and classifies as wet/dry cough, Further the tool helps in monitoring the disease recovery for tuberculosis and asthma and even the early detection of the disease mentioned above. This paper is organized as follows.

### II.CHARACTERISTICS OF COUGH SOUND

Usually cough sounds consists of three phases: phase1: initial burst, phase 2: noisy air flow and phase 3: glottal closure

#### A. Dry cough characteristics

Dry cough does not contain any mucus or sputum and these three phases of cough are visible in the dry cough as shown in the Figure 1(a), observe that after the phase 1 i.e. initial burst, the energy levels are low in phase at higher frequencies.

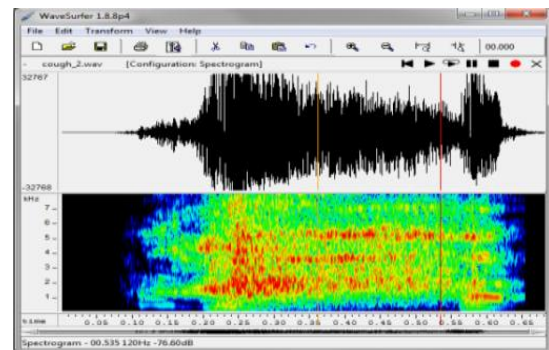


Figure 1. a) Dry cough signal with 3 phases: Phase 1, Phase 2, and Phase 3, b) Spectrogram of dry cough signal.

#### B. Wet cough characteristics

Wet cough is produced by the bacteria or virus causing secretion in the lower airways: bronchitis, asthma etc. therefore the wet cough consists of mucus and sputum. Figure 2(a) represents the wet cough; the wet cough consists of higher energies in the phase 2

i.e. in higher frequencies. 2(b) shows the spectrogram of the wet cough.

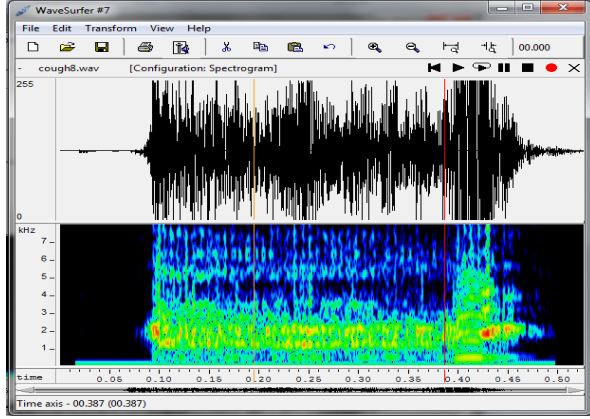


Figure 2. a) Wet cough signal with 3 phases: Phase 1, Phase 2, and Phase 3, b) Spectrogram of wet cough signal.

### III. DATA ACQUISITION

The cough samples from subjects are recorded using mobile application “Easy Voice Recorder” in noise free environment. Samples from 30 subjects were collected 20 were of wet cough samples and 10 are dry cough samples under supervision of medical professional.

### IV. EXPERIMENTATION

#### Training phase

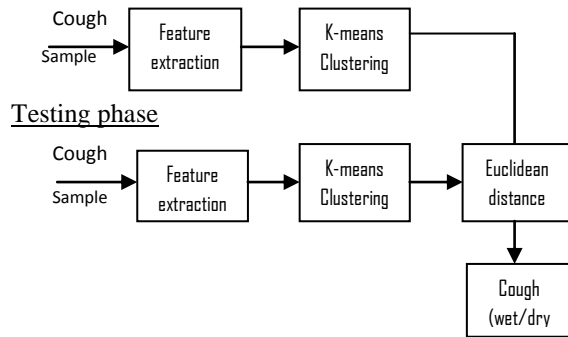


Figure 4: Block diagram of classifying system

Feature extraction is a process of converting the input speech signal in to digital representation for further processing of the signal. Several algorithms are used for this process like as Linear Predictive Coefficients (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP) cepstra.

The algorithm used in this paper for feature extraction is MFCC which are the prominent features extracted in signal and speech processing.

The relation (1) is employed to calculate the Mel's of the signal.

$$\text{mel}(f) = 2595 * \log(1 + f/700) \quad (1)$$

Where ‘f’ is the real frequency and Mel (f) is the perceived frequency.

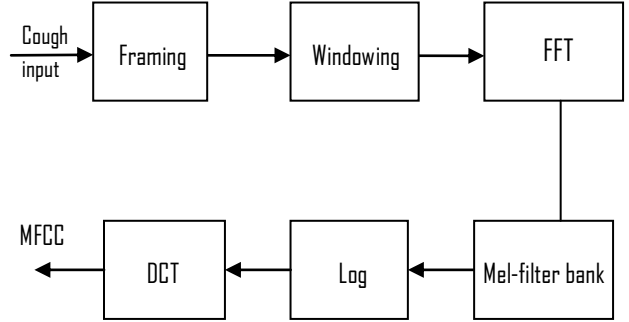


Figure 3: Block Diagram of MFCC computation.

Fig 3 shows block diagrammatic representation of mfcc computation flow. Various steps involved in mfcc calculation are described briefly below.

#### A. Framing

Splitting the signal in to several frames of 10-30ms short length to analyze is known as framing. In this work, the frame length is 20ms with an overlap of 10ms so that the signal is stationary between frames

#### B. Windowing

In this process the framed signal is multiplied with the window function. In frequency domain, this combination becomes the convolution between the spectrum and the transfer function of the window. Hamming window is used to minimize the spectral distortion and the signal discontinuities. Later Fast Fourier Transform showing the relation (2) is employed to the windowed signal to calculate the Mel frequency..

$$X(k) = \sum_{n=0}^{N-1} x[n]W_N^{nk} \quad (2)$$

### C. Mel frequency wrapping

Fourier transformed signal is passed through the filter banks to obtain the Mel frequency wrapping. The MEL filter banks have triangular band pass as shown in Figure 4.

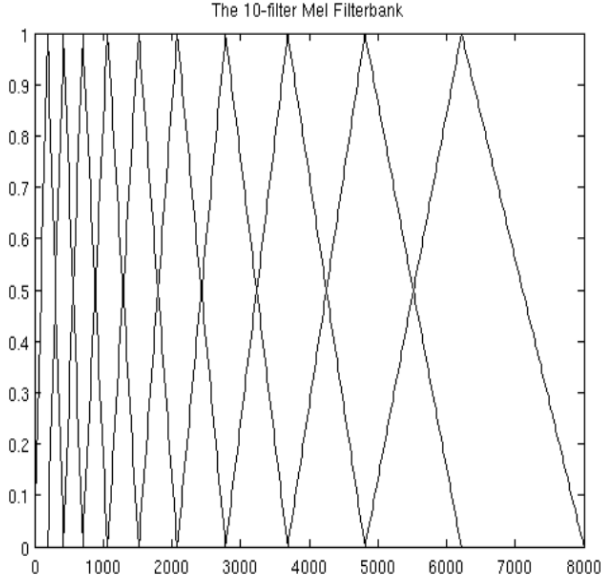


Figure4: MEL scale filter bank.

### D. Discrete Cosine Transform (DCT)

The MFCCs are obtained, by applying discrete cosine transform (DCT) to convert it into time domain. After the feature vectors are extracted the dataset is subjected to classification by k-means algorithm.

### V. K-Means Algorithm

The k-means algorithm is an iterative approach that partitions the given dataset into a user specified number of clusters. The algorithm partitions T vectors into M clusters. Each vector is assigned to nearest centroids by using the Euclidean distance.

The algorithm is implemented through following steps :

- N data vectors are randomly selected out of T vectors and each vector is assigned to a partition set
- Centroids of the partitioned set are calculated.
- Assign objects to the closest cluster based on the distance.
- The process is continued until the
- Centroids value becomes constant.

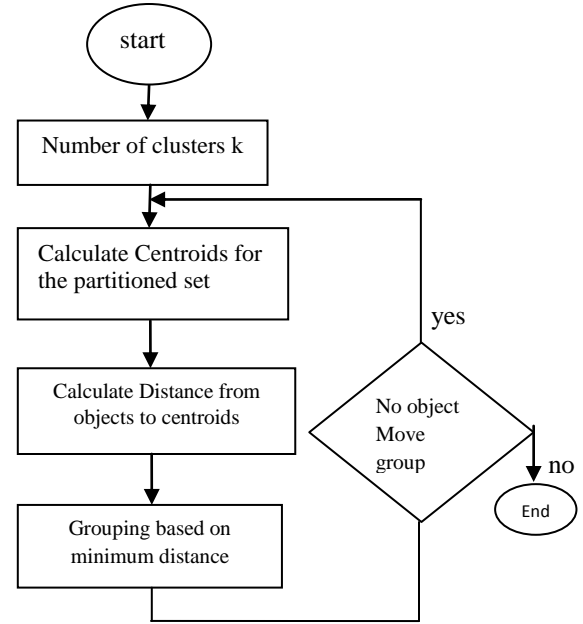


Figure 5: Flow graph of k-means.

The codebooks are created for cough sounds in training and testing phase , to classify given input in to wet or dry cough in testing phase we minimum distance calculation i.e. Euclidian distance .

### VI. Euclidean Distance

Euclidean distance gives the distance between two point vectors a and b.

The Euclidean distance is given as

$$d = |a - b| = \sqrt{\sum_{i=1}^n |a_i - b_i|^2} \quad (3)$$

Distance from the trained data and the test input is calculated. The training data includes both the dry and wet clusters. The minimum distance of the test sample and the cluster will give the classification result.

### VII. Simulation Results

The plot below shows difference between the cepstrum vectors of the both wet and dry cough samples. It is observed from the plot that both these cepstrum are differ substantially from each other.

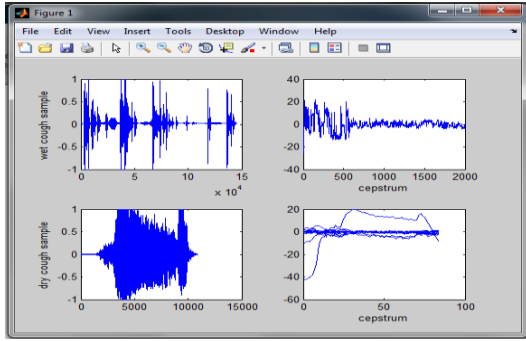


Figure 6: cepstrum Plots for wet and dry cough samples

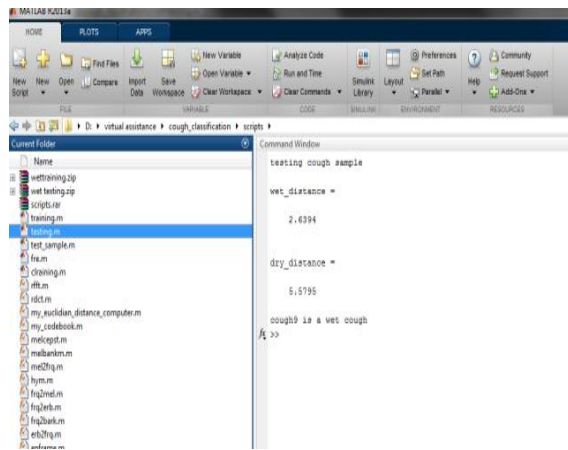


Figure 7: Euclidean distance between test signal and two clusters (Dry and Wet cough)

The figure 7 shows the distance between test sample and clusters (dry and wet cough).our research aims at designing a simple tool for monitoring of certain diseases such as tuberculosis and asthma based on cough patterns and the intensity of the cough sounds. The paper tried to focus on the classification of cough samples into dry and wet categories as the initial step by extracting the MFCC features. The classification will further be used for monitoring the recovery of the diseases like tuberculosis using other features energy along with MFCC. The tool is under progress

## VIII. Conclusion & Future work

This paper proposes classification of cough sounds by MFCC feature vector and K-means algorithm for classification, observed efficiency of the algorithm is 82%.furthermore hybrid algorithms can be used to increase the efficiency of the system. In future work it is proposed to develop the too to provide automatic indicators on recovery levels by the observed intensity of cough signals obtained.

## References

- [1] P. Piirila and A.R.A. Sovijarvi,( Nov. 1995)“Objective assessment of cough,” *European Respiratory Journal*, vol. 8, pp. 1949-1956,.
- [2] K.F. Chung,( May 2002.) “Assessment and measurement of coughs: the value of new tools,” *Pulmonary Pharmacology and Therapeutics*, vol. 15, pp. 267- 272,
- [3] A. Murata, Y. Taniguchi, Y. Hashimoto, Y. Kaneko, Y. Takasaki, and S. Kudoh, (Sep. 1998)“Discrimination of productive and non productive cough by sound analysis,” *Internal Medicine*, vol. 37, no. 9, pp. 732-735,.
- [4] Hanieh Chatzarrin, Amaya Arcelus, Rafik Goubran”,( May 2011) Feature Extraction for the Differentiation of Dry and Wet Cough Sounds”
- [5] Samantha J Barry<sup>1</sup>, Adrie D Dane<sup>1</sup>, Alyn H Morice<sup>\*2</sup> and Anthony D Walmsley<sup>1</sup>,(Sep 2006) “The automatic recognition and counting of cough”:
- [6] Germán Comina, Sandra Larson, Marjory Bravard, José W. López, and Robert H.Gilman, Brian H. Tracey,(2011)Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis
- [7] S.J. Barry, A.D. Dane, A.H. Morice, A.D. Walmsley,( 2006) “The automatic recognition and counting of cough,” *Cough*,; 2:8.
- [8] Matos, S.S. Birring, I.D. Pavord, D.H. Evans, (IEEE Biomed Eng 2006)“Detection of cough signals in continuous audio recordings using hidden Markov models,”; 53: 1078–1083