

# ANALYSIS OF COUGH SOUND CHARACTERISTICS FOR RESPIRATORY DISEASES

<sup>1</sup>Ravali Tirunagari, <sup>2</sup>Dr.Y.Padma Sai, <sup>3</sup>V.Naveen Kumar, <sup>4</sup>S.V.Narayana Rao

<sup>1</sup>MTECH, <sup>2</sup>Head of the Department, <sup>3</sup>Assistant Professor, <sup>4</sup>Associate Professor

<sup>1</sup>Department of ECE,

<sup>1</sup>VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

**Abstract:** *Respiratory Diseases like Chronic obstructive pulmonary disease (COPD), Asthma, Pneumonia, and Tuberculosis (TB) are emerging as major health threats for many people across the globe. Hence early diagnosis and recovery monitoring of these diseases is obligatory. Cough is found to be the best symptom for the diagnosis of Respiratory Diseases. It has vital information regarding the disease. Cough is a reflex, with a deep inspiration, glottal closure, and an explosive expiration followed by a sound. The objective is to study and analyze cough characteristics and see how they differ among different diseases. This analysis helps in building applications like digital health advisory for Respiratory Diseases. We identified the important characteristics like Duration, Intensity, Type of cough and Frequency of coughs which will vary among the diseases. This paper presents the approach to be followed in deriving these characteristics and provides detailed analysis on how these characteristics are varying and presents results based on data collected covering important parameters like accuracy, specificity and sensitivity. Hence, we can conclude that the proposed methodology is an essential tool for respiratory diseases diagnosis in timely manner.*

**Index Terms –** Cough Characteristics, Duration, Intensity, Dry cough, Wet cough, Frequency of coughs, GMM

## I. INTRODUCTION

Respiratory Diseases are the disorders of airways and lungs which affects the human respiration. Respiratory Diseases such as Asthma, Chronic Obstructive Pulmonary Disease (COPD), Tuberculosis (TB), Pneumonia etc damages any respiratory system organs like nasal cavity, larynx, pharynx, trachea, bronchi, bronchioles, pleura and parenchyma. According to the recent study, it is known that 90% deaths caused by COPD in low and middle income countries, 235 million people suffer from Asthma. World Health Organization (WHO) declares that the deaths due to Respiratory diseases in India were rising to 11 percent of the total deaths and 14.2 in every one million, died of Respiratory disease. The main problem about these diseases is knowledge about the disease. So there is a great need of the tool which gives screening information about the disease. Cough is a classic symptom and one of the earliest symptoms among all respiratory diseases. It is a sign of severity of the disease. Cough is natural protective mechanism which is produced due to irritants or mucus in the parts of respiratory system. Cough sound characteristics like duration, intensity, type of cough and frequency of coughs present at different parts vary differently which adds us a great value in the diagnosis. Cough characteristics analysis helps in knowing the affected part of respiratory system hence can identify the Disease. Wet cough is a cough with mucus associated with lower respiratory system and dry cough is the absence of mucus associated with upper respiratory system.

The previous works shown by Chatzarrin proposed that peaks of the energy envelop and spectral features of the cough sounds. Murata argued that cough sound frequencies can be used to discriminate between wet and dry cough. Swarnkar showed the classification of wet and dry only for pediatrics. The present research work differs from others by dry, wet classification with standard features and GMM used as a classifier along with analysis of duration, intensity and frequency of coughs for all age groups, genders and for different Respiratory diseases with large dataset.

## II. MATERIALS AND METHODS

Because the pattern differs among different diseases, the characteristics which are influencing these patterns are: a) Duration is the distance the signal travels, b) Intensity is the amplitude of a signal, c) Type of cough which indicates the presence of mucus or not. Wet cough is defined as the presence of mucus and Dry cough is defined as the absence of mucus, d) Frequency of coughs is the number of cough events present for a granular time frame. The patterns of different diseases are shown in below figure 2.1

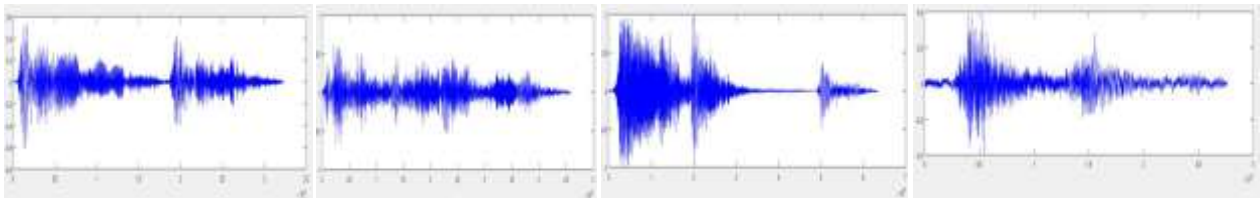


fig.2.1 cough patterns of Asthma, COPD, Pneumonia and TB

The system below presents the methodology we followed in deriving these characteristics and also the analysis we did for different diseases. This is divided into two steps 1) Data Acquisition, 2) Analysis of cough.

### 2.1 Data Acquisition

Data acquisition setup was done from four Chest hospitals in Hyderabad. fig 2.1 shows the process followed for data collection and dataset creation. The patients with different respiratory diseases and with different age groups, genders are considered. Data has

Collected using Zoom H5 handy recorder, with sampling rate 44.1 KHZ, 16 bit resolutions. Recorder is placed at 50cm distance from the patient. Cough samples are collected from Inpatients in the ward and also few out patients suffering with Respiratory diseases came with

severity in their case in the controlled environment. The patient disease label, symptoms are known from doctor. Initially, patient with cough is approached and place recorder. The recording is called as raw sample, which is recorded from minutes to few hours. These raw samples contain cough events and non cough events like noise, speech and other sounds etc. All these recordings are segmented by standard event extraction technique and checked by manually segmenting cough events or sequences by listening to start and end of a sequence. Cough sequence or event is defined as single inspiration followed by multiple expirations. The cough events are disease labeled as wet and dry using praat software based on pattern. Cross validation for labeling is performed using GMM algorithm. After 20 iterations a golden dataset of 100 samples is created. The dataset is further validated with the doctors for correctness. This dataset is used as a Golden standard for training. All the other sequences are labeled using golden dataset. Hence cough dataset has been created with 500 cough samples from 30 patients.

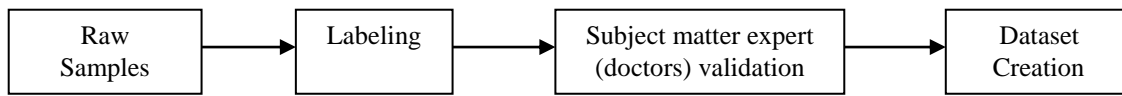


fig.2.2 Data acquisition and dataset creation

**2.2 Analysis of cough characteristics**

Study of cough characteristics is important in identification of disease at a very early stage and even used in patient recovering from the disease. Analysis of cough characteristics such as Duration of cough, Intensity, Type and Frequency of coughs is performed for the diseases Asthma, COPD, TB and Pneumonia.

Let ‘N’ be the number of cough events or sequences

- i. Cough duration analysis is done by calculating length of every cough event for all ‘N’ cough sequences. In case of multiple cough events in a recording, the average duration of cough sequence is calculated.
- ii. Cough sound intensity is calculated and analyzed using histograms for each cough event for all ‘N’ sequences in decibels
- iii. Cough type as dry and wet analysis is performed after its classification. Using machine learning algorithm GMM, with the novel features and good amount of data, classification is performed with positive results.
- iv. Frequency of coughs is other characteristic which helps in analyzing the recovery status of the patient. The frequency of occurrence of cough events i.e., number of wet cough events, number of dry cough events is calculated on minute basis

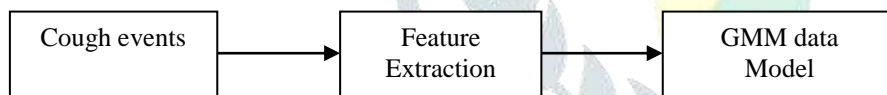
From the findings of all the above characteristics, identification of disease is done and digital healthcare monitoring is possible.

**2.2.1 Classification of cough type**

Cough can be classified into wet and dry cough as shown in fig.2.2. Graphical User Interface (GUI) is designed in Matlab for classification. Training and Testing are two basic steps in machine learning. The data collected is divided as 80% for training and 20% for testing. Cough events which are manually cut using Praat software are given as per the division to Training and Testing.

This is passed to Feature Extraction. Feature extraction gives salient information helpful for classification. GMM supervised learning algorithm builds a model and whence a test cough event is given, feature extraction of test sample is performed and using log likelihood, with the help of the built model. GMM classifies cough as wet cough and dry cough.

**TRAINING**



**TESTING**

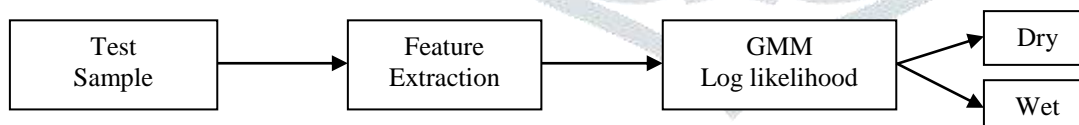


fig.2.3 block diagram for classification of dry and wet

**2.2.1.1 Feature Extraction**

From the ‘N’ cough events, feature vector containing ‘F’ mathematical features is computed and a cough event feature matrix was formed. Computation of ‘F’ features from a cough event is done by following steps

- i. Let x denotes a discrete time sound signal from a cough event.
- ii. Segment x into ‘n’ equal size sub-segments with 20ms duration overlap.
- iii. Compute the features for each sub-segment and form feature vector with Mel-frequency cepstral coefficients (MFCC), Formant Frequencies, log energy, Zero crossings and Kurtosis
- iv. Repeat steps (i)–(iii) for all ‘N’ cough events and form cough event feature matrix.

Hence the required feature set is formed.

**Definitions of Features:**

**Mel frequency Cepstral coefficients (MFCC):**

MFCCs are prominently used in the music and speech audio signal analysis. They represent the short term power spectrum of an acoustic signal based on a cosine transform of a log power spectrum on a non-linear mel-scale of frequency.

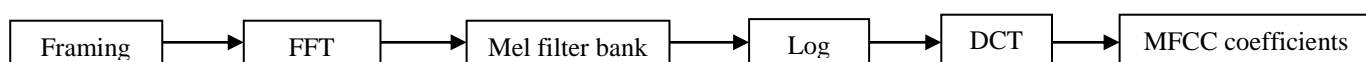


fig.2.4 MFCC flow diagram

1. Frame the signal into short frames say 20ms to 40ms say x.
2. For each frame calculate the FFT.
3. To the power spectra apply the Mel filter bank and add the energy in each filter.
4. Apply logarithm to all filter bank energies.
5. Apply DCT to the log filter bank energies.

**Zero crossings:** The Zero crossings is defined as the total time a signal crosses the zero axis. For each cough event number of zero crossing were found.

#### Formants frequencies:

Formants frequencies are referred as the resonance of the human vocal tract. In case of cough sound, the resonances of the overall airway that causes the cough sound will be represented in the formant structure. These give the highest frequency information present. FF has a lot impact in the classification. In cough first 4 formants F1, F2, F3, F4 are found to be giving the best results. F1–F4 is computed by peak picking the Linear Predictive Coding (LPC) spectrum of cough sounds.

**Kurtosis:** The kurtosis is a measure of sharpness of peak of cough event with a probability distribution. It is calculated by the formula  $\frac{E(X[K]-\mu)^4}{\sigma^4}$

**Log energy:** Energy of a signal mathematically is the square of amplitudes. It gives the information about strength of signal.

#### 2.2.1.2 Classification: GMM Model and Log likelihood

GMM is parametric probability density function represented as a weighted sum of Gaussian component densities. GMM model is built using the probability density function of observed data points using a multivariate Gaussian mixture density for all 'N' cough events. After giving the cough events to GMM, by using expectation-maximization algorithm refines the weights of each distribution.

1. Initialize  $K$  centroids (random centroids;  $k$  data points chose randomly as centroids)

#### Expectation step:

2. Associate each of the data points as belonging to one of the centroids
3. Accumulate the set of data points belonging to centroid
4. Calculate the average of distance between each of the data points and their closest centroid

#### Maximization step:

5. Each of the should be updated to be the centroid of the data points it owns
6. Stopping criteria: Repeat the steps 2-5 until it is less than a specified threshold value.

Hence the model is built. This built model used as a training dataset. Testing is performed by using GMM log likelihood. Test sample is taken and feature extraction is performed. It calculates average distance of centroids for every cough event and checks with the GMM model and then gives the classification result as dry or wet cough.

### III. RESULTS AND DISCUSSION

Cough characteristics analysis is done for four Diseases such as Asthma, COPD, TB and Pneumonia. The table 3.1 below shows the calculated duration, intensity and frequency values. Duration of TB is found to be high among all from the table. This difference in values signifies the area of disease. Cough duration is long for lower respiratory system and low comparatively for upper respiratory system. Intensity values for different diseases is shown in the below table. Intensity analysis is done by histograms. From figures 3.1, 3.2, 3.3 and 3.4 histograms of four diseases are calculated. Low amplitude range is less than 0.3, medium amplitude range is 0.3 to 0.7 and high amplitude ranges greater than 0.7. For asthma and Pneumonia high amplitudes are found and COPD, TB medium amplitudes are found from the graphs. For Pneumonia it is seen that graph decreases abruptly whereas for TB it gradually decreases. The bars in the histogram represent the frequency of occurrence of cough. From the table it is found the frequency of coughs per minute is more in case of COPD. The number of dry cough in TB is absent and for COPD wet coughs are very high. All these findings can be used in early detection of diseases.

Table 3.1 Duration, Intensity and Frequency results

DISEASE	AVERAGE DUARTION(SEC)	INTENSITY(dB)	FREQUENCY OF COUGH EVENTS(no. of coughs per minute)	NO.OF DRY COUGH EVENTS	NO.OF WET COUGH EVENTS
ASTHMA	0.91	67	2-4	2	2
COPD	1.05	63	6-10	2	8
PNEUMONIA	0.8	72	1-5	4	1
TUBERCULOSIS	1.9	65	2-6	0	5

Histogram gives information regarding intensity and frequency of coughs, which is shown in the below figures for different diseases COPD, Asthma, TB and Pneumonia.

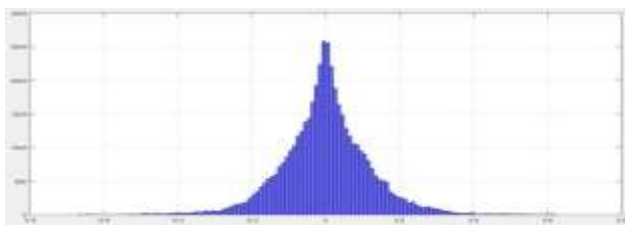


fig 3.1 Histogram of cough for COPD

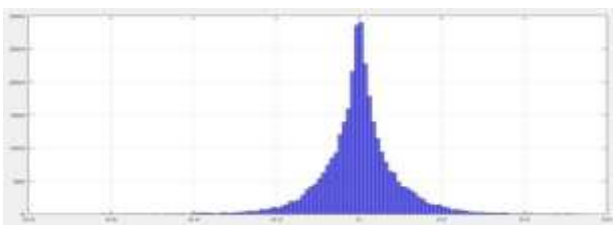


fig 3.2 Histogram of cough for Asthma

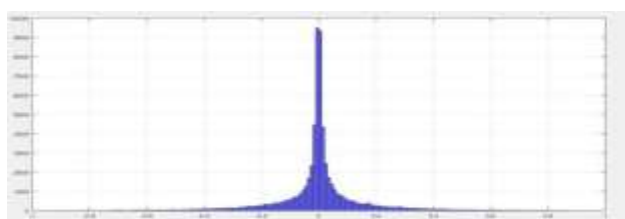


fig 3.3 Histogram of cough for Pneumonia

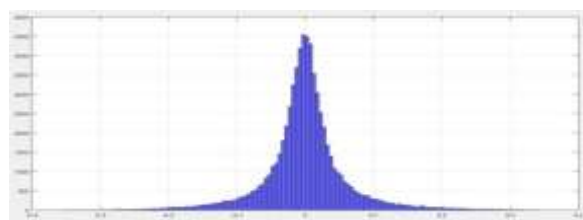


fig 3.4 Histogram of cough for TB

In order to analyze the type of cough as dry and wet, we need to classify. The classification results are shown in table 3.2. The figures 3.5 and 3.6 show that the graphs of dry cough and wet cough with its spectrograms. The table 3.3 shows the considered features and their significance in classifying them. From this classification we can easily get wet and dry cough frequencies for every disease which is used for the identification of the diseases.

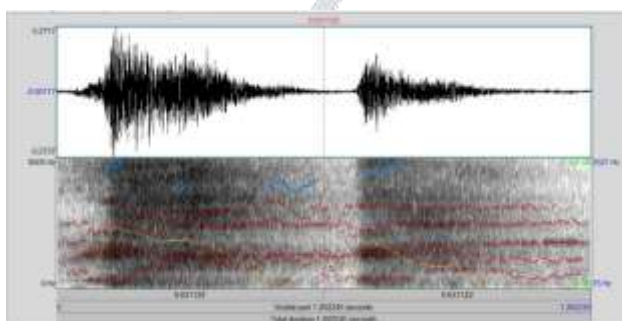


fig 3.5 Dry cough

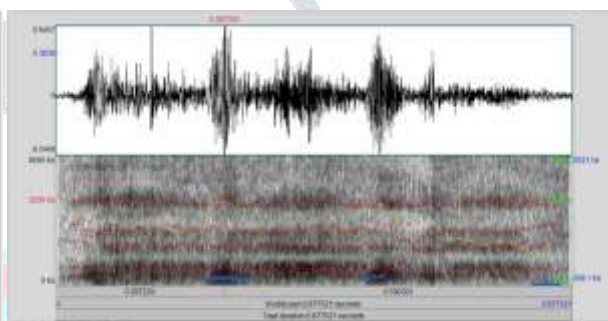


fig 3.6 Wet cough

Table 3.2 Cough Type classification results

WET/DRY CLASSIFICATION	ACCURACY (%)	SPECIFICITY (%)	SENSITIVITY (%)
	93.33	77.7	88

Table 3.3 Showing significance of used features in dry and wet classification

Features	Definition	Wet cough	Dry cough
Zero Crossing Rate	Measure of frequency content of a signal	Wet cough has less number of zero crossings than dry	Dry cough has more number of zero crossings than wet
Log Energy	Signal strength is known	Energy is high at high frequencies in phase-2 of wet cough	Energy is low at high frequencies in phase-2 of dry cough
Formant Frequencies	Vocal cord parameters	Observed low frequencies at phase2.	Observed High frequencies at phase2
Kurtosis	Measure of thickness of tails of distribution.	Range in wet found to be 3 to 6. More frequently occurring are 3,6	Range in dry is found to be 1 to 5. More frequently occurring are 1, 2.

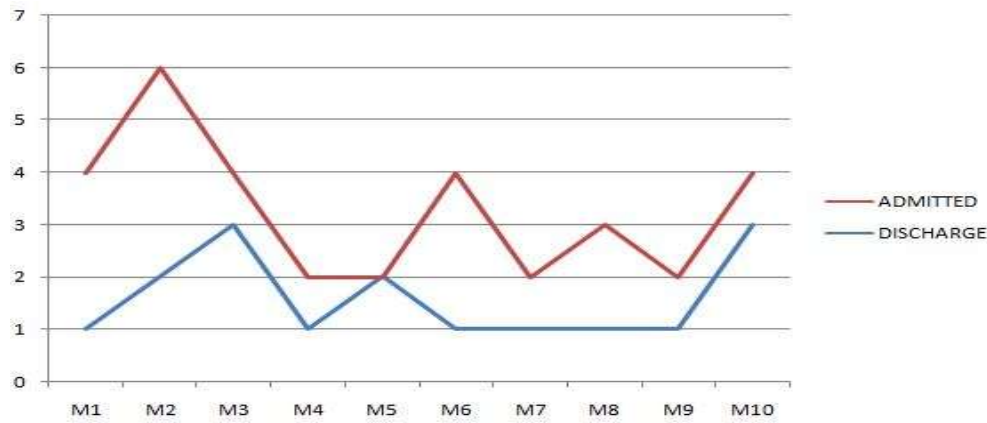


fig.3.7 Plot showing Frequency of coughs/minute for TB

The above fig 3.7 shows the frequency of occurrence of coughs in TB. The frequency is noted on minute basis. For every minute the frequency is noted. The Red line in the graph shows the cough occurrences at the time of admission and blue line shows the discharge time graph. From both the graphs the observation of cough frequency reduction can be made. This indicates the disease decrement level, which is most useful in Recovery Monitoring of patients, mostly useful for doctors. In COPD the irritants are present at small and medium airways, in asthma it is medium airways, for TB irritants present at lower airways, similarly for pneumonia irritants found at alveoli i.e., lower airways. Thus, the above analysis of cough characteristics helps in mapping the location or part of disease and airway in respiratory system, by which we can identify the disease very accurately. This helps in digital monitoring application

#### IV. CONCLUSION

The proposed cough characteristics analysis provides relevant information regarding the disease by showing the uniqueness in different diseases. Hence cough characteristics play a very prominent role. The goal of early diagnosis and recovery monitoring of Respiratory diseases is reached by this analysis. There is no other approach till date which gives us good Performance. Hence we can say that the present work can be used in digital health monitoring application.

#### REFERENCES

- [1] P. Piirilä, A.R.A. Sovijärvi, "Objective assessment of cough," European Respiratory Journal (ERS), 1995
- [2] Jaclyn A Smith, H Louise Ashurst, Sandy Jack, Ashley A Woodcock<sup>1</sup> and John E Earis, "The description of cough sounds by healthcare professionals," BioMed Central, 2006
- [3] Chatzarrin, "Feature Extraction for the differentiation of dry and wet cough sounds," Medical Measurements and Applications Proceedings (MeMeA), 2011
- [4] Akira Murata, Yasuyuki Taniguchi, Yasushi Hashimoto, Yasuyuki Kaneko, Yuji Takasaki and Shoji Kudoh, "Discrimination of Productive and Non-Productive Cough by Sound Analysis," Internal Medicine Vol. 37, No. 9, September 1998
- [5] J. Korpa's, J. Sadlon ova, M. Vrabec, "Analysis of the Cough Sound: an Overview," Acta Medica Mar Tiniana 2013
- [6] Vinayak Swarnkar, Udantha R. Abeyratne, Anne B. Chang, Yusuf A. Amrulloh, Amalia Setyati, Rina Triasih, "Automatic Identification of Wet and Dry Cough in Pediatric Patients with Respiratory Diseases," Annals of Biomedical Engineering, Vol. 41, No. 5, May 2013 pp. 1016-1028
- [7] Dr. Y. Padma Sai, S.V.N Narayana Rao, V Naveen Kumar, V Priyanka Brahmaiah, D Ajay, "Cough Classification Tool For Early Detection And Recovery Monitoring Of Tuberculosis And Asthma," 4th International Conference On Computing Communication And Sensor Network, CCSN2015
- [8] Yusuf Amrulloh, Udantha Abeyratne, Vinayak Swarnkar, Rina Triasih, "Cough Sound Analysis for Pneumonia and Asthma Classification in Pediatric Population," 6th International Conference on Intelligent Systems, Modelling and Simulation, 2015