# Detection of possibility of Laryngeal Cancer through Mel Frequency Cepstrum Coefficient Analysis

Kavya S, Raghuram Shivram
Dept. of Electronics and Communication Engineering
B.M.S College of Engineering
Bangalore, India
1bm16lvs11@bmsce.ac.in

***Abstract*—Laryngeal Cancer is a condition in which cancerous cells are formed in larynx. Notable symptoms are hoarseness in the voice, sore throat, trouble or pain while swallowing and ear pain. The detection of the modification in voice through Acoustic analysis can serve as a complement to other medical procedures. The novelty of the paper lies in utilizing Mel Frequency Cepstrum Coefficients (MFCC) for detection of Laryngeal cancer. This method for detection of laryngeal cancer was not researched, traditionally the detection of Laryngeal cancer was only possible through clinical examination. In the proposed method, the analysis is carried out by comparing the coefficients of voices of patient affected by Laryngeal Cancer to that of normal human voice. The study involves conversion of voice signal to parametric representation. On further processing, it was observed that there was significant deviation in the vital coefficients of cancerous voice as compared to that of normal voice and the initial results show that the possibility of detection was found to be 100% in a dataset comprising of 60 samples.**

***Keywords— laryngeal cancer, short term energy, fast fourier transform, mel frequency cepstrum coefficients***

## I. Introduction

According to the doctors of All India Institute of Medical Sciences, out of the entire population of cancer patients in India, 25 percent of them suffer from oral or laryngeal cancer [1]. Laryngeal cancer affects the voice box (larynx). The larynx, which is a portion of the throat is found at the entrance of the windpipe (trachea). The symptoms are change in voice, persistent sore throat, difficulty in swallowing, long lasting cough, swelling in the neck and in severe cases, difficulty in breathing. Recognizing the cautious signs of laryngeal cancer and taking immediate course of actions leads to early diagnosis. The risk of being affected considerably increases due to smoking, regular consumption of large quantity of alcohol, unhealthy diet, exposure to asbestos and coal dust. The treatment includes surgery, radiotherapy and chemotherapy. Detection of cancer in its primitive stages immensely boosts the chances of effective treatment. During the early stages, surgery or radiotherapy is used to extract the cancerous cells which often leads to the cure of Laryngeal cancer. The clinical assessment of laryngeal function and hence diagnosis of pathological voice can be achieved through acoustic analysis. The benefit of acoustic analysis is that it is a non-invasive method and has the ability of providing quantitative data.

Different approaches have been followed for the vocal disorder interpretation. The studies conducted so far, primarily focus on extraction of MFCC and their utilization for further analysis. A technique to discriminate normal voice from pathological voice was developed based on the exploitation of MFCC and the discriminatory ability was enhanced by investigation of a projection based Linear Discriminant Analysis (LDA). Classification of the extracted feature combinations was accomplished using Artificial Neural Networks (ANN) [2]. Hypothyroidism in newborns was detected by deriving the features of infant cry using MFCC and F-Ratio analysis was carried out to separate out the vital features [3]. A Fast Fourier Transform (FFT) based MFCC extraction was carried out to distinguish an asthma patient voice from normal voice [4].

This paper proposes a method for detection of laryngeal cancer in its basic stages by utilizing hoarseness of voice as a parameter. The features of normal voice and voice of laryngeal cancer patient are extracted using MFCC. Furthermore, it has been identified that there is a considerable variation in the coefficients of laryngeal cancer patient voice in comparison with that of normal voice.

The novelty of the paper lies in utilizing MFCC for detection of Laryngeal cancer. This method for detection of laryngeal cancer was not researched, traditionally the detection of Laryngeal cancer was only possible through clinical examination. In the process of calculating MFCC, triangular filter are used that are uniformly distributed in Mel scale. This helps achieve high resolution for lower frequency. Since most of the speech signals acquired were concentrated in the lower frequencies, MFCC proved to be the right choice for the study.

## II. Methodology

As outlined in Fig. 1, the first step is the acquisition of voice signals from normal people and people affected by laryngeal cancer. The signals were subjected to windowing using Short Term Energy to separate the voiced signal. The voiced signals were then digitized and converted to frequency domain using FFT. The spectrum was transformed to Mel scale. The logarithm of resulting signal was computed and Discrete Cosine Transform (DCT) was performed on it. This resulted in the generation of MFCCs.

978-1-5386-2440-1/18/$31.00 ©2018 IEEE

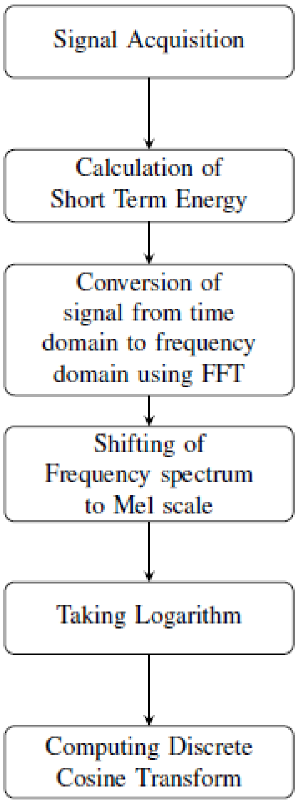


Fig. 1. Design flow of the proposed method

### *A Signal Acquisition*

This study involved collection of thirty voice samples from normal people and thirty from patients of laryngeal cancer each of which comprised of 20 male and 10 female subjects aged between 15 to 60 years. Samples from people suffering from laryngeal cancer was taken in consultation with specialists from Bangalore Medical College and Research Institute (BMCRI). Adhering to the ethical principles of Helsinki declaration [5], the consent of the subjects was taken. The samples were collected from patients in Stage1 of laryngeal cancer. The signals were recorded using the microphone present in a Samsung phone. The signals were sampled at 8 kHz. The subjects were asked to speak the word ”Help”. The word was chosen arbitrarily and can be replaced with any other word. The protocol involved capturing of voice input for a duration of 1 second from each of the subjects. Since the major concentration of the analysis is on the first and second coefficients of MFCC, the duration of 1 second contained the sufficient information required. Thus the signal database comprised of a total of sixty signals.

### *B. Short Term Energy*

Since the energy of speech signal fluctuates with time, it comprises of voiced and unvoiced portions. There is a huge variation in the energies associated with voiced and unvoiced regions [3].

The calculation of short term energy of a speech signal is achieved with the help of (1).This short term energy calculation is used for distinguishing the voiced segments of the speech from the unvoiced.

$$E_m = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (1)$$

The size of the window was chosen to be in the range of 10 to 20ms or 80 to 160 samples with a sampling rate of 8 kHz. This enables output of short term energy to be optimal.

### *C. Fast Fourier Transform*

The conversion of voiced signal in time domain to its equivalent in the frequency domain can be achieved by employing FFT. A 1024 point FFT of the signal is computed for further spectral analysis. The spectral distribution of FFT is wide and linear. Since humans perceive speech in a nonlinear manner, the FFT spectrum obtained is subjected to Mel Frequency Cepstrum analysis that closely relates to human auditory system's reaction [6], [7].

### *D. Mel Frequency Cepstrum Coefficients (MFCC)*

Various experiments conducted on human perception of sound reveal that the human ear focuses on specific regions rather than the whole spectrum. Human ear behaves as a filter, thus concentrating only on particular frequencies. The spacing of these filters is nonlinear in nature, a large number of them are found in the low frequency region and the number decreases with increase in frequency.

In audio processing, Mel-Frequency Cepstum Coefficients (MFCCs) are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale. [8]. MFCC collectively constitute an MFC. The mel frequency scale greatly aligns with auditory features of the human ear. The nonlinear perception of human auditory system is mimicked by MFC where the placement of frequency bands on mel scale are uniform. This greatly improves the representation sound in audio processing applications [9].

The steps involved in deriving MFCCs are as follows:

- Computing Fourier Transform of a signal
- Employing triangular overlapping windows to map the power spectrum obtained in previous step to mel scale
- At each of the mel frequencies, computing log of powers
- Performing Discrete Cosine Transform (DCT) on mel log powers

The power spectrum obtained after computing FFT is mapped onto the mel scale using Mel Filter banks. This enhances the possibility of obtaining non-linear frequency resolution [10], [11]. Corresponding to each tone with frequency in Hz on normal scale, there exists a subjective pitch on mel scale which is calculated using (2).The plot of linear to mel frequency conversion is depicted in Fig. 2.

$$f_{Mel} = 2595 * log(1 + \frac{f_{Hz}}{700}) \quad (2)$$

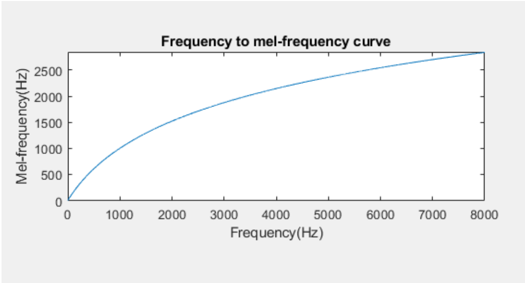


Fig. 2. Plot of linear to mel frequency conversion

Once the frequencies are converted to the mel scale , Mel Filter Banks calculated using (3) are used for interpreting the amount of energy that is present in various frequency ranges [12]. The first filter in the bank is cramped and provides information about energy present around 0 Hz. With increase in frequency, the filters expand in size. The initial filter begins at start point reaches its maximum at second point and comes back to zero at third point. Similarly, the next filter begins at second point reaches its maximum at third point and comes back to zero at fourth point and so on. The mapping of linear to mel frequency and the energies associated with each frequency range is depicted in Fig. 3.

$$H_m(k) = \begin{cases} 0, \text{if } k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, \text{if } f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, \text{if } f(m) \leq k \leq f(m+1) \\ 0, \text{if } k > f(m+1) \end{cases} \quad (3)$$

where, the number of filters is denoted by m and f() denotes all the m+2 frequencies spaced uniformly on the mel scale.

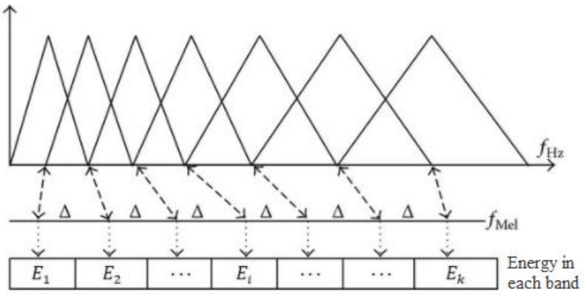


Fig. 3. Linear to Mel frequency conversion along with energy associated with each frequeny range

Once the energies of filter bank are calculated, the logarithm of the energies are computed. This compression process aids in relating the features more intently to actual human hearing. The log filter bank energies are subjected to Discrete Cosine Transform. This is performed to de-correlate the energies which were correlated due to the filter bank overlapping. Amplitudes of the resulting spectrum after taking DCT corresponds to MFCCs. The most valuable information is present in the first two coefficients. The first coefficient contains information regarding the power over all frequency bands and the second coefficient strikes a balance between low and high frequency components of the signal.

## III. RESULTS AND DISCUSSION

The results obtained during the implementation are presented stepwise in this section. The voice sample of a normal person is represented by Fig. 4 . Fig. 5 represents the voice sample collected from a laryngeal cancer patient. In both the cases, the speech input was recorded for a duration of 1 second where the subjects have spoken the word ”Help” (as per section II A).

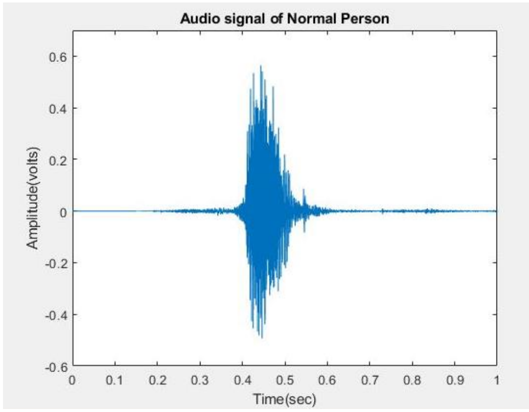


Fig. 4. Audio signal of Normal Person

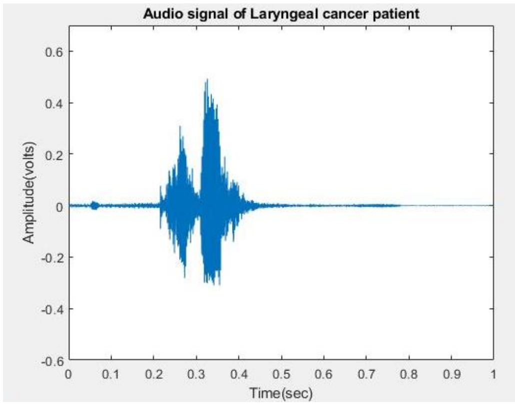


Fig. 5. Audio signal of Laryngeal Cancer Patient

As discussed in section II B, the short term energy calculation is performed on the above signals to obtain the voiced and unvoiced portion of the signals. The unvoiced portion of the signal is discarded to obtain only the voiced signal. The extracted voiced signal is depicted in Fig. 6 and Fig. 7 for normal voice and laryngeal cancer patient voice respectively.

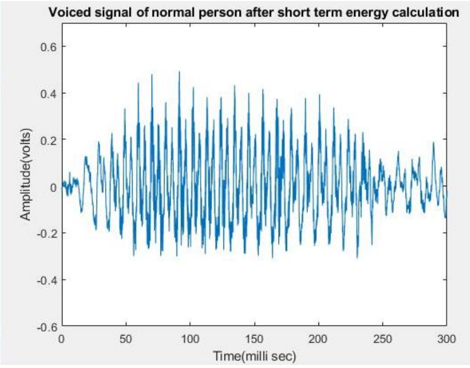


Fig. 6. Voiced signal of Normal Person after Short Term Energy calculation

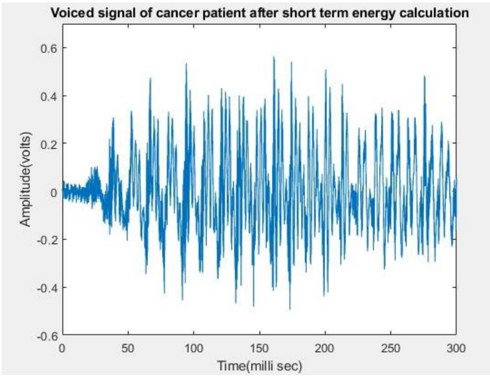


Fig. 7. Voiced signal of Cancer Patient after Short Term Energy calculation

Once the voiced signal was extracted, it was converted into frequency domain using FFT. A 1024 point FFT was computed for the signals (section II C). The FFT coefficients for the normal voice input is depicted in Fig. 8 . Furthermore, the FFT coefficients of cancer patient is represented in Fig. 9.

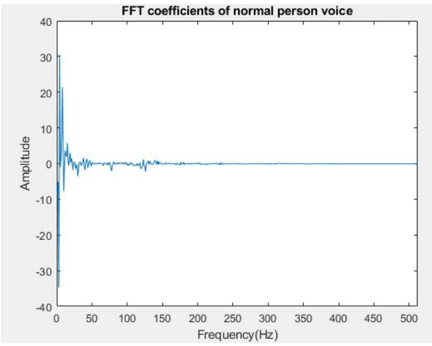


Fig. 8. FFT coefficients of Normal Person voice

In accordance with the discussion in section II D, the signal after performing FFT, is converted to the mel scale using overlapping triangular filter banks. In the analysis, 10 Mel-filterbank depicted in Fig. 10 was implemented.

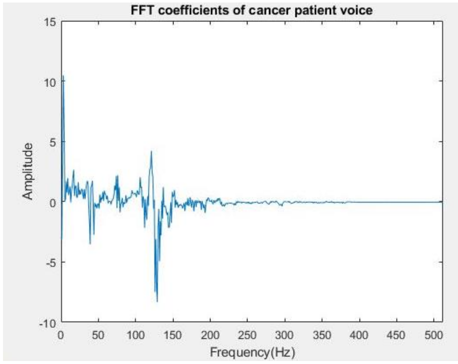


Fig. 9. FFT coefficients of Cancer Patient voice

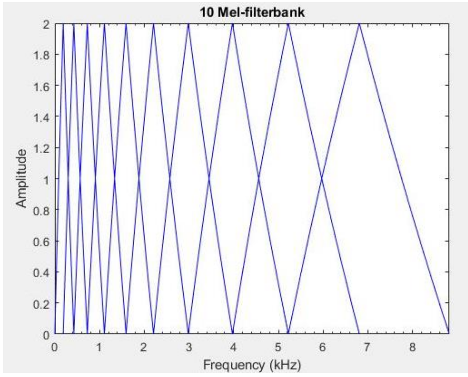


Fig. 10. Ten Mel-filterbank

The MFC coefficients were extracted for both normal and cancer patient voice signals. The MFCCs of 4 normal voice samples that represent the dataset are encapsulated in Table 1. The detailed results can be found at [13]. The plot of MFCCs for the entire dataset of normal voice is represented in Fig. 11.

TABLE 1: MFCC OF NORMAL PERSON VOICE

| MFC Coefficients | Sample_01 | Sample_02 | Sample_03 | Sample_04 |
|---|---|---|---|---|
| 1 | 10.446 | 10.158 | 10.04 | 10.363 |
| 2 | 1.6274 | 1.4925 | 1.6429 | 1.6389 |
| 3 | 0.8043 | -0.4315 | 0.1787 | -0.874 |
| 4 | 0.9059 | -0.8025 | 0.4051 | -0.0163 |
| 5 | 0.1155 | -0.1301 | -0.9616 | -0.47 |
| 6 | -0.0117 | -0.3576 | -0.2139 | -0.4018 |
| 7 | -0.9803 | 0.1569 | 0.1997 | -0.3254 |
| 8 | -0.4094 | 0.6995 | -0.2977 | -0.2289 |
| 9 | 0.626 | 0.201 | 0.1097 | -0.2316 |
| 10 | -0.233 | 0.1469 | 0.0657 | -0.3181 |

Furthermore, the MFCCs for 4 samples that represent the dataset of laryngeal cancer patient voice are summarized in Table 2. The detailed results can be found at [13]. The plot of the MFCCs of the whole dataset of cancer patient voice is illustrated in Fig. 12.

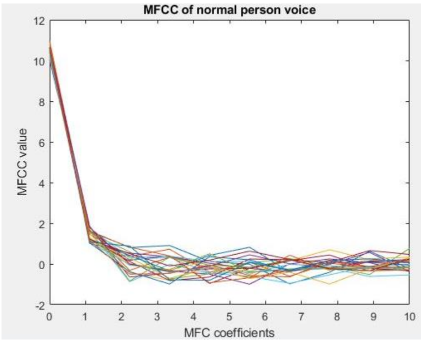


Fig. 11. MFCC of Normal Person voice

TABLE 2: MFCC OF CANCER PATIENT VOICE

| MFC Coefficients | Sample_01 | Sample_02 | Sample_03 | Sample_04 |
|---|---|---|---|---|
| 1 | 12.847 | 12.491 | 12.16 | 12.574 |
| 2 | 3.4532 | 3.6389 | 3.6429 | 3.6274 |
| 3 | 0.3758 | -0.8749 | 0.1787 | 0.8043 |
| 4 | 0.7256 | -0.0163 | 0.4051 | 0.9059 |
| 5 | -0.4094 | -0.47 | -0.9616 | 0.1155 |
| 6 | 0.0803 | -0.4018 | -0.2139 | -0.0117 |
| 7 | -0.6013 | -0.3254 | 0.1997 | -0.9803 |
| 8 | 0.0332 | -0.2289 | -0.2977 | -0.4094 |
| 9 | -0.1553 | -0.2316 | 0.1097 | 0.626 |
| 10 | 0.135 | -0.3181 | 0.0657 | -0.233 |

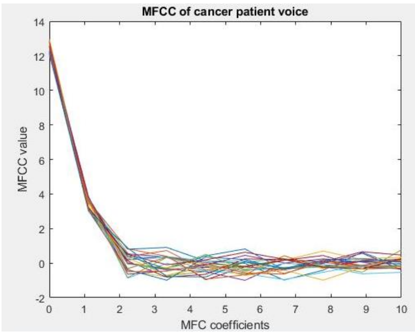


Fig. 12. MFCC of Cancer Patient voice

In order to have results justified, we contacted the specialists from BMCRI, to have the results validated. Our initial findings shows that the proposed method was able to detect voice samples of all patients suffering from Laryngeal Cancer. The detection of laryngeal cancer was found to be 100% in a collected dataset of 60 samples considering various factors like age, gender and location of cancer in the throat.

## IV. CONCLUSION

An approach for detecting the possibility of laryngeal cancer through auditory analysis using hoarseness of the voice as parameter, was proposed in this work. The analysis indicates that short term energy technique greatly facilitates the segregation of the voiced portions of the speech signal from the unvoiced. In addition, the MFCCs derived for normal and cancer patient voice show a considerable deviation in the most valuable coefficients (first and second). In the dataset that was selected (60 samples), the possibility of detection was 100%. The results obtained were also validated by specialists from BMCRI. Therefore, successful detection of cancer patient voice was achieved.

However, the degree of variation in the hoarseness voice due to other medical conditions are not considered. The authors would like to take up the analysis and classification of various medical conditions involving change in voice with considerably large dataset as a part of the future work.

## ACKNOWLEDGMENT

Our sincere thanks to specialists from BMCRI for providing necessary suggestions. We express our gratitude to all the subjects who volunteered for the data acquisition process.

## REFERENCES

[1] https://food.ndtv.com/health/oral-cancer-accounts-for-25-of-all-cancer-cases-aiims-1200459 (Throat Cancer facts and figures)

[2] Nawel Souissi, Adnane Cherif, "Speech Recognition System Based on Short-term Cepstral Parameters, Feature Reduction Method and Artificial Neural Networks", 2nd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016, Monastir, Tunisia, March 21-24, 2016.

[3] Azlee Zabidi, Wahidah Mansor, Lee Yoot Khuan, Rohilah Sahak, Farah Yasmin Abd Rahman, "Mel-Frequency Cepstrum Coefficient Analysis of Infant Cry with Hypothyroidism", 5th International Colloquium on Signal Processing Its Applications (CSPA), 2009.

[4] V. D. Dighore, V. R. Thool, "Analysis of Asthma By Using Mel Frequency Cepstral Coefficient", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, India, May 20-21, 2016.

[5] http://www.wma.net/en/30publications/10policies/b3/ (Declaration of Helsinki)

[6] Dong-Fong Syu, Su-Wei Syu, Shanq-Jang Ruan, Yu-Chang Huang, Chuan-Kai Yang, "FPGA Implementation of Automatic Speech Recognition System in a car environment", IEEE 4th Global Conference on Consumer Electronics (GCCE), 2015.

[7] Jingbo Zhang, Ganggang Ning, Shufang Zhang, "Design of Audio Signal Processing and Display System Based on SoC", 4th International Conference on Computer Science and Network Technology (ICCSNT), 2015.

[8] https://en.wikipedia.org/wiki/Mel-frequency_cepstrum (Mel Frequency Cepstrum)

[9] Yuan Xue Luping Wang, Linxuan Li Zhiqi Liu Jialin Liu, "Matlab-based Intelligent Voiceprint Recognition System", Sixth International Conference on Instrumentation Measurement, Computer, Communication and Control, 2016 IEE.

[10] Othmane El Badlaoui, Ahmed Hammouch,"Phonocardiogram Classification Based on MFCC Extraction", 2017 IEEE

[11] Jennifer C. Saldanha, T. Ananthakrishna, Rohan Pinto, "Vocal Fold Pathology Assessment Using Mel-Frequency Cepstral Coefficients and Linear Predictive Cepstral Oefficients Features", Journal of Medical Imaging and Health Informatics, Vol. 4, 2014.

[12] Kamil Ismaila Adeniyi, Oyeyiola Abdulhamid K., "Comparative Study on the Performance of Mel-Frequency Cepstral Coefficients and Linear Prediction Cepstral Coefficients under different Speaker's Conditions", International Journal of Computer Applications (09758887) Volume 90 No 11, March 2014.

[13] goo.gl/b3suri (Detailed Results of the entire sample set)