IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 15, Issue 2, Ser. II (Mar-Apr 2020), PP 01-06
www.iosrjournals.org

Thyriod Detection using Speech Recognition

Mrs. Dr. S.B Patil[1], Mr. B. S. Patil[2], Mr. A. P. Pande[3]
[1]Department of ETC, Dr.JJMCOE, Jaysingpur
[2]Department of CSE, PVPIT, Budhgaon
Department of CSE PVPIT, Budhgaon

Abstract: *This paper describe an implementation of application which is developed to find malignant & partially malignant thyroid detection using patient's speech signal variations. It detects Thyroid by person's speech samples. Format of Speech sample files are .wav (wave audio file format). This software application inputs speech files of malignant, non-malignant and partially malignant samples of patients. It extracts features using MFCC and HMM algorithms, then train and build models. It can check and detect single and multiple test files for malignant, non-malignant & partial malignant Thyroid.*

Keywords: *MFCC, HMM, SVM, malignant, Matlab*

Date of Submission: 11-04-2020 Date of Acceptance: 26-04-2020

I. Introduction

The main mode of human communication for interaction is speech and it will be preferred mode for human to machine interaction. In this work, we took voice samples of patients who have Thyroid Disease. Basically, the butterfly shaped gland present in the lower anterior of the neck is a thyroid. The thyroid is one of the largest endocrine glands in the body, is found in the neck sits in front of the trachea (also known as the windpipe). When it doesn't function properly, leads to Thyroid disorder. And 15% − 30% of urban people have prone to thyroid disorder due to existing life style and women are four times more prone to thyroid disorders than men. Thyroidal imbalances can make immediate and dramatic effect on the voice. The one method is the acoustic voice analysis which shows significant differences between normal and thyroid voices. With this approach a prospective effort has been made to extract more details about the disease non-invasively. The non-invasive method is cheaper, fast and repeated. The conventional methods for diagnosis of thyroid disease are usually slow, lengthy, expensive and annoying so the purpose of this project is to analyze and classify of voice disorders using voice signal processing.

This Speech Recognition is the method to control some activity by human voices/speech. Here we study a signal voice processing by using MFCC (Mel-Frequency Cepstrum Coefficients) and SVM method. To develop speech recognition is a process to identify speech signal. It is a process to extracting features from speech using Mel-Frequency Cepstrum Coefficient (MFCC) method.

II. Literature Survey

DWT (Discrete Wavelet Transform) -[10] in this paper is discussed about any data which can be turned into a linear Sequence can be analyzed with DWT. And the advantages are increased speed, reduced storing space for the reference template, increased recognition rate. And disadvantages are choosing the appropriate reference template for task is a difficult task. According to this paper threshold can be used in order to stop the process if the error is too great.

Wavelet , According to above paper, method of wavelet is discussed with following points. Better time resolution than Fourier Transform. Advantages is wavelet transform-based features gives recognition accuracy than MFCC and LPC, The WT has a good capability to the unvoiced sound portions for better time resolution. Drawbacks are the cost of computing DWT as compared to DCT is higher. It takes more compression time. So it replaces the fixed bandwidth of Fourier transform with one proportional to frequencies which allow better time resolution at high frequencies than Fourier Transform.

Classifier SVM (Support Vector Machine) -[8], According to above paper, In SVM approach, the main aim of an SVM classifier is obtaining a function, which determines the decision boundary or hyper-plane. This hyper-plane optimally separates two classes of input data points. SVM is a discriminative classifier. It provides good generalization, although it may not be the best for every case. For low-level speech features, the dimensionality of the SVM depends on the duration of speech, which might be problematic for unbalanced speech durations. It also causes a very high dimensional feature vector.

HMM (Hybrid Markov Model) -[8], Use of an integrated HMM/NN classifier for speech recognition. The proposed classifier combines the time normalization property of the HMM classifier with the superior discriminative ability of neural network (NN) classifier. Speech signals display a strong time varying characteristic. In the proposed integrated hybrid HMM/NN classifier, a left-to-right HMM module is used first to segment the observation sequence of every exemplar into a fixed number of states. Porya Saheli et al[10], this paper is analysis and classifies the voice fold disorders. And the classifiers and feature extraction methods provides accuracy and performance of the system.

AVA (Acoustic Voice Analysis) - Here the acoustic voice analysis is recycled for detecting the thyroid disease, which provides the difference in between the normal and thyroid patient's voice. Acoustic Voice Analysis can give significant target data on voice unsettling influences, particularly those with natural and useful changes [4]. Therefore, this method is used for calculating the various voice parameters, which are affected due to thyroid disease. The changes in the voice can be obtained. The voice changes happen due to various reasons like vibrating vocal folds, cough, voice disorders etc. Acoustic voice analysis method checks the various voice parameters like jitter, shimmer, fundamental frequency, harmonic to noise ratio and many more parameters that changes in thyroid patient.

III. MFCC

A. Mel Frequency Cepstrum Coefficient

Feature extraction utilizing Mel Frequency Cepstrum Coefficient (MFCC) strategy - MFCC is a technique for feature extraction of voice signals. Feature extraction is the way toward deciding a worth or vector that can be utilized as an object or an individual character. MFCC is the most utilized strategy in different regions of voice preparing field, since it is viewed as very great in representing signal [12]. Feature is the coefficient of cepstral, the coefficient of cepstral utilized as yet thinking about the impression of the human hearing framework. The operations of MFCC depend on the various frequencies that can be captured by the human ear to represent the sound signals as humans represent them.

During the time spent speech signal pre-emphasis filter is required after the testing procedure. The reason for this filtering is to acquire a smoother unearthly type of speech signal frequency. As such, this filtering procedure is done to decrease clamor during sound capture. Where the ghostly shape is moderately high incentive for low regions and will in general fall strongly to the region of frequency over 2000 Hz [18]. The pre-emphasis filter depends on the info/yield relationship in the time space expressed in the accompanying condition:

$$y(n)=x(n)-ax(n-1) \qquad \text{Eq. I}$$

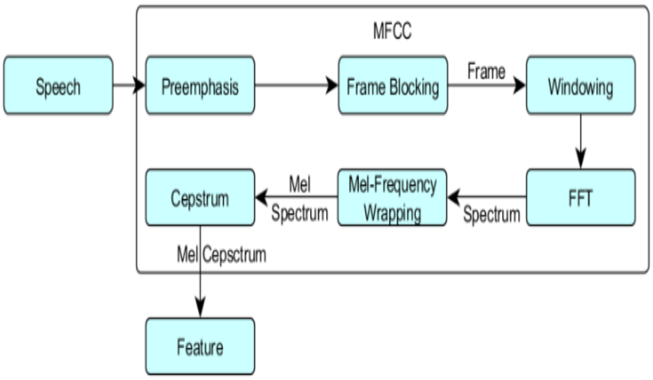

Fig.1 Feature extraction process using MFCC

B. Frame Blocking:

In this process, the sound signal is segmented into multiple overlapped frames, so there is not a single deletion of signals. This process will continue until all signals have entered into one or more frames as illustrated. Voice analysis was done by short-time analysis. The x[n] long voice signal is split into a number of frames. One frame has N voice data sample. Between one frame with another frame overlapping each other a number of M samples of voice data. The value of M is not more than N that is 2xM.

C. Windowing:

Windowing: Windowing is a procedure for breaking down long sound signals by taking an adequately representative segment. Windowing is a Finite Impulse Response (FIR) digital filter approach. This procedure evacuates the associating signal because of the discontinuity of the signal pieces. Discontinuities happen because of the edge blocking process. In the event that we characterize the window as w(n),0 ≤ n ≤ N -1, where N is the number of samples in each frame, the result of windowing is a signal:

$$y1(n)=x1(n)w(n), \quad 0 \le n \le N-1$$

From Equation 2 y(n) is the result signal of the convolution between the input signal and the window function and x(n) represents the signal to be convolved by the window function Where w(n) usually uses window Hamming which has the form:

$$w(n)=0.54-0.46.cos(2\pi nN-1), \quad 0 \le n \le N-1$$

D. FFT:

Fast Fourier Transform (FFT).A work with constrained period can be expressed in Fourier series. Fourier transform is utilized to change over a period series of bounded time domain signals into a frequency range. The casing that has experienced the windowing procedure is changed over into a frequency range. FFT is a fast algorithm of Discrete Fourier Transform(DFT) which is helpful for changing over each edge to N samples from time domain into frequency domain. FFT reduces the repeatable augmentation contained in the DFT.

$$Xn=\sum xkN-1k=0e-2\pi jkn/N$$

Where n = 0, 1,2,..., N-1 and j = sqrt-1. X[n] is the n-frequency pattern generated from the Fourier transform, Wk is the signal of a frame. Result of stage the is called as Spectrum.

II. MEL-FREQUENCY WRAPPING:

Where n = 0, 1,2,..., N-1 and j = sqrt-1. X[n] is the n-frequency pattern produced from the Fourier transform, Wk is the signal of a casing. The aftereffect of this stage is normally called Spectrum. The perception of the human ear against the sound frequency doesn't pursue the linear scale. The genuine frequency scale utilizes units of Hz. The scale that takes a shot at the human ear is known as the frequency Mel scale. The scale of Mel-Frequency is a low frequency that is linear under 1000 Hz and a logarithmic high frequency over 1000 Hz [19]. The accompanying condition shows the connection of the Mel scale to the frequency in Hz.

$$Fmel = \{\ 2595 * [log]10\ (1 + FHZ\ 700),$$
$$FHZ > 1000\ FHZ,\ FHZ < 1000$$

Where Fmel is the Mel scale and f is the frequency shown on above Equation.

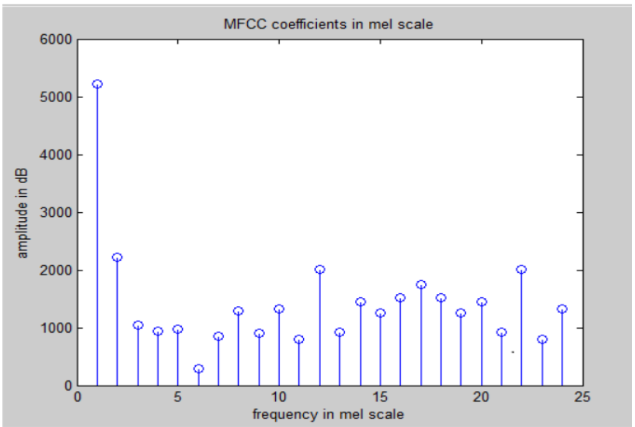

Fig.1: Feature Extraction by using MFCC Method

The frequency spectrum in the Mel scale with the working function of human ear as filter is by Filter Bank. If the F[N] spectrum is the input of this process, then the output is the M[N] spectrum that is the F[N] modified spectrum that contains Power Output of these filters. The spectrum coefficient of Mel is expressed by K, and is specially determined to be 20.

The subsequent FFT signal is assembled into this triangular filter document. The reason for the gathering here is that each FFT esteem is increased against the relating filter gain and the outcome is added. At that point each gathering contains a specific measure of signal vitality weight as expressed as m1… mp. The procedure wrapping to the signal in the frequency domain is finished utilizing the Equation.

$$Xi = log10(\Sigma\ |x(k)|\ N-1\ k=0\ Hi\ (k))$$

Where i = 1, 2, 3, …, M (M is number of triangle filters) and Hi(k) is value of i triangle filter for the acoustic frequency of k.

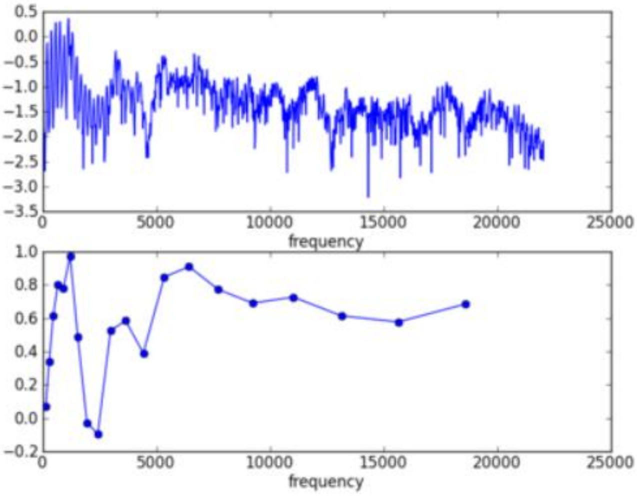

Fig.2 Original Amplitude Spectrum and Mel Bank Filter

Humans listen to voice information based on time domain signals. Mel-spectrum will be converted into time domain by using Discrete Cosine Transform (DCT). The result is called Mel-frequency cestrum coefficient (MFCC).

$cj = \sum Xj\ K\ j=1\ cos\ (j(i-1)/2\ \pi\ K\ )$

above Equation show Cj is the MFCC coefficient, Xj is the power spectrum of Mel frequency, j = 1, 2, 3, …, K (K is the number of desired coefficients) and M is the number of filters.

III. Support Vector Machine (SVM)

SVM can be search for the best hyper plane that serves as a separator of two classes in the input space. SVM used for various machine learning, such as; object recognition, speech recognition, handwritten character recognition speaker recognition and language recognition. SVM is a binary classification algorithm. It is comprised of sums of kernel function

k(xi, xj). [20] $f(x) = \sum\ \alpha i\ ti\ N\ i=1\ K(xi\ ,\ xj + d)$

From above Equation, $\sum\ \alpha i\ ti = 0,\ N\ i=1\ \alpha i > 0$, and ti represent of the outputs either +1 or -1 depends of the class which has sample data.

f(x) value compare with the threshold to decides the output class of certain test sample. Multi-class data problem used a one-vs-all approach adapted usually to achieve classification.

The SVM train by the Gaussian RBF kernel has the data point xi and xj get from Equation 9.

$$K(xi\ ,\ xj) = exp(\gamma \parallel xi\ ,\ xj \parallel)\ 2$$

After multiple iterations on train and test data, the optimal hyper-parameters γ and regularization constant C selected for the SVM.

IV. HMM

Hidden Markov Model is a stochastic process. It has the Markov property, conditional probability distribution of future states of the process i.e conditional on both past and present states depends only upon the present state, not on the sequence of events that lead it.

A process having this property is called as Markov process. Generally, the term “states” are used to refer to the hidden states and “observations” are used to refer to the observed states. HMM emits the observations based on the emission probability distribution from a given state and state transitions occur in a sequence as per the transition probability.

HMM to be generating the observation sequence and the state sequence that has the maximum probability of generating the observed sequence is considered the output of HMM for those problems where we are required to determine the hidden state sequence

V. IMPLEMENTATION

Voice samples are collected from laboratory In pre-processing technique, various methods are used to analyze the signal like framing, Windowing and pre-emphasis. After the pre-processing next step is the feature extraction. In feature extraction, initially we selected features for checking the parameters changes in the patient's voice. Here we used the combined feature extraction method.

In Wavelet and DWT, we firstly generate the wavelet of patients voice and then convert it into discrete form. for better results of the system. MFCC method the Mel scale is depending on the human ear scale and coefficients is depends on the perception system.

The mathematical equation of MFCC is,

$$C(n) = DCT(\log(|FFT(S(n))|))$$

For finding MFCC feature they look hamming window signal by using below equation,

$$W(n,a) = (1-a)-a\cos 2\pi n\ (n-1)$$

Where 0 n N-1 In the DFT transform, complex sequence is given below:

$$Xk = PXn \exp 2\pi i\ Nnk$$

Where k = 0,1,2,N-1.

Then DCT of signal will be,

$$Ci = Pmj\cos[i\pi\ M\ (j-0.5)]\ mj = \log \exp(Yj)$$

Where i = 0, 1, 2, 3, ..N; N¡M

After filter bank equation , mel scale equation can be written as,

$$Mel = 2595 * \log(1 + freq\ 700\ )$$

After feature extraction work further goes to classification stage. During the classification phase, the input characteristic vector data is formed using information relating to known models, and then tested using the test data set. Basically classifier is used for checking the accuracy of the designed system.

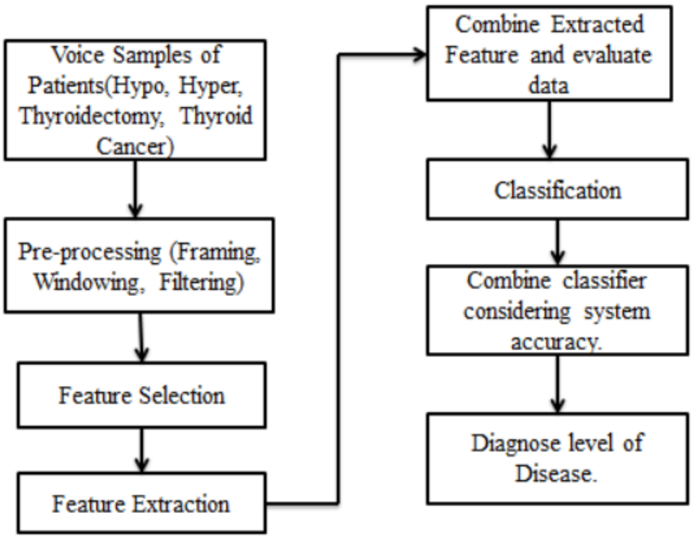
Voice Samples of Patients(Hypo, Hyper, Thyroidectomy, Thyroid Cancer)
Pre-processing (Framing, Windowing, Filtering)
Feature Selection
Feature Extraction
Combine Extracted Feature and evaluate data
Classification
Combine classifier considering system accuracy.
Diagnose level of Disease.

Fig.5: System Flow Diagram

In classification, which method gives higher accuracy that combined for getting higher accuracy. In this proposed system, SVM and HMM classifier is used.

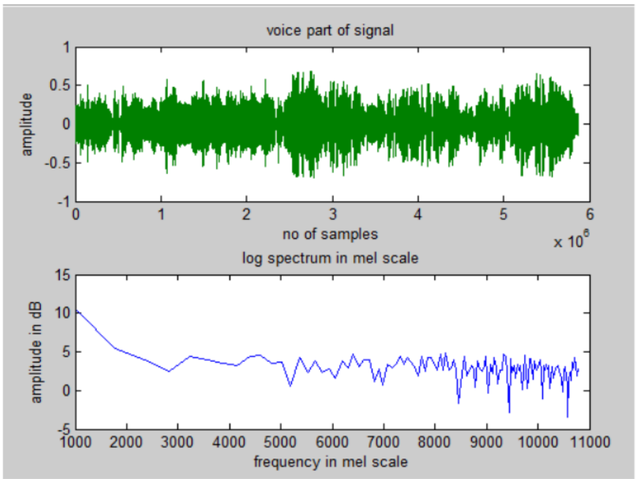

Fig.3 Voice signal VS Power Spectrum of Voice Signal.

VI. CONCLUSION

This work is implemented for diagnosis of thyroid disease. This application can accurately detect malignant & nonmalignant voice signals, which are speech file of patients. The performance of diagnosis of thyroid disease is evaluated by using classification accuracy, k-fold cross-validation, and confusion matrix methods respectively. Experiments are done on 500 voice samples. Overall accuracy recorded is 99.45%.

REFERENCES

[1]. Avci, E. (2007). A new optimum feature extraction classification method for speaker recognition: GWPNN. Expert Systems with Applications, 32(2), 485–498.

[2]. Avci, E., & Avci, D. (2008). A novel approach digital radio signal classification: Wavelet packet energy–multiclass support vector machine (WPE–MSVM).

[3]. Expert Systems with Applications, 34(3), 2140–2147.

[4]. Belousov, A.I., Verzakov, S.A., & von Frese, J. (2001). Support Vector Machines: A versatile and powerful approach to data analysis.

[5]. Çomak, E., Arslan, A., & Turkoglu, I. (2007). A decision support system based support vector machines diagnosis the heart valve diseases. Computers in Biology and Medicine, 37(1), 21–27.

[6]. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. G. R. (1996). Advances knowledge discovery and data mining. Menlo Park, CA: AAAI Press/ The MIT Press.

[7]. Fernández Pierna, J. A., Baeten, V., Michotte Renier, A., Cogdill, R. P., & Dardenne, P.(2004). Combination of support vector machines (SVM) and near-infrared (NIR)imaging spectroscopy for the detection of meat and bone meal (MBM) in

[8]. compound feeds. Journal of Chemometrics, 18(7–8), 341–349.

[9]. Frias-Martinez, E., Sanchez, A., & Velez, J. (2006). Support vector machines vs multi-layer perceptrons efficient off-line recognition. Engineering Applications of Artificial Intelligence, 19(6), 693–704.

[10]. Gunn, S. R. (1998). SVM for classification regression. Technical report, Image Speech and Intelligent Systems Research Group,

[11]. Keles, A., & Keles, A. (2008). ESTDD: Expert system for thyroid diseases diagnosis. Expert Systems Applications, 34(1), 242–246.

[12]. Pasi, L. (2004). Similarity classifier applied to data sets, 2004,10 sivua, Fuzziness in Finland'04. In International conference soft computing, Helsinki, Finland & of Finland & Tallinn, Estonia.

[13]. Polat, K., & Gunes, S. (2007). An expert system based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing, 17(4), 702–710.

Mrs. Dr. S.B Patil, etal. "Thyriod Detection using Speech Recognition." *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* 15(2), (2020): 40-46.