Classification of Human Speech for Neurological disease from laryngeal using Pitch, Formants, MFCC, HFCC as features and SVM as classifier

Prof. S.M. Ali  
Electronics and Telecommunication dept.  
Anjuman College of Engineering and Technology  
Nagpur, India  
e-mail: ali_acet2003@yahoo.co.in

Dr. P. T. Karule  
Electronics Engineering dept.  
Yeshwantrao chavan College of Engineering, Wanadongri  
Nagpur, India  
e-mail: ptkarule@rediffmail.com

Prof. S.I Ali  
Electronics and Telecommunication dept.  
Anjuman College of Engineering and Technology  
Nagpur, India  
e-mail: irfan_extc@yahoo.co.in

Prof. S.A.Ali  
Electronics and Telecommunication dept.  
Anjuman College of Engineering and Technology  
Nagpur, India  
e-mail: akbar_acet@yahoo.com

Abstract—Since past few decades speech processing has achieved a great interest in the field of research. Speech processing yields in various applications such as classification of speech, speaker, language, age, speech-to-text converter, emotions, and now a day’s disease. This research includes classification between disease using speech signal, which includes its preprocessing, feature extraction and classification using SVM. The preprocessing includes passing of speech signal through filter, framing and windowing. This is followed by spectral analysis wherefrom various features are extracted. The two groups of features which we have extracted are pitch+formants and pitch+formants+MFCC+HFCC. These features are used to differentiate between Neurological disease from laryngeal using network like support vector machine. This research is very helpful in disease diagnosis. It has been observed from the experiments that the second group with SVM gives 100% accuracy.

Keywords—MFCC; HFCC; Pitch; classification; pre-processing; filtering; framing; windowing;; SVM; support vector machine; speech processing;

I. INTRODUCTION

Automation has become the significant domain in the field of research and technologies. Disease diagnosis has gained momentum these days. Here is the need of classifier that should classify between Neurological diseases from laryngeal so that clinicians can get second opinion.

Speech processing has already achieved the attraction of research students and proved as an excellent tool for classification between various parameters.

For analysis, speech samples were collected from Hospitals using sound recorder of smartphone and putzer voiceldsorder database. Patients were requested to say ‘a’, ‘ah’ within two seconds. These samples are filtered from noise, framed and windowing before giving for feature extraction.

Basically speech signal has three main characteristics namely: amplitude, frequency and phase and it is represented by the following expression.

\[
\sum_{i=1}^{N} A_i(t) \sin[2\pi F_i(t)t + \theta_i(t)]
\]

Where, \(A_i(t), F_i(t), \text{and} \theta_i(t)\) are amplitudes, frequencies & phases respectively of the speech signal. Speech generation is the simultaneous process of numerous organs like Lungs, Bronchi, Tracheas (producing expiration air steam necessary for phonation), Larynx (amplifying the initial tone), Root of the tongue, throat, nasal cavity, oral cavity (forming tone quality & speech sound) [1].

Figure 1 shows speech signal. Figure 2 shows the flow of
control. Here in this paper we have collected speech samples of man, women & child patients which are passed through moving average filter and high pass filter, the filtered output is framed and then each frame is passed through window. The output signal which is framed and windowed is used for feature extraction. The features which we have found are pitch, formant frequency, MFCC and HFCC.

II. SPEECH ACQUISITION
Speech samples of men, women and children are taken at sampling frequency 50000Hz using window recorder. Usually the speech signal frequency of man, woman and child has different frequency band ranging from 80 Hz to 170 Hz, 170 Hz to 300 Hz and 300 Hz to 1000 Hz respectively. Thus, spectral features are extracted.

III. PREPROCESSING
A. Moving Average Filter Design
This includes passing of speech signal through filter like moving average filter to remove the noise from signal. It takes the average of samples and expressed in the following form [5].

\[ Y(n) = \frac{X(n)+X(n-1)+X(n-2)}{3} \]

Where, X(n) is the input speech sample.

B. Pre-Emphasis Filter Design
The output of moving average filter is given to pre-emphasis filter, which is high pass filter. This filter is used to flatten the speech signal spectrum & to make the speech signal less sensitive to finite processing effects later in speech signal processing [4]. The pre-emphasis filter amplifies the area of spectrum. Thus improving the efficiency of spectral analysis [2].

The time domain representation of filter will be

\[ Y(n) = X(n) - \lambda X(n-1) \]

Where y(n) is the output, x(n) is input speech sample & λ is the filter coefficient with \( \lambda = 0.9375 \) optimum result of filtering is received [3]. The output of this filter is framed & passed through window where speech signals are analyzed for short period of time (5 msec to 100msec). The signal is fairly stationary & windowing is done to avoid problem due to truncation of signal & window helps in smoothening of signal [1].

IV. FEATURE EXTRACTION
The spectral features which are calculated are pitch, formant frequency, MFCC & HFCC.

A. Pitch
Pitch is a dominant frequency in speech signal. Usually speech signal has lot of frequencies but the frequency which appears dominantly is called fundamental frequency or frequency of speech of person. There are various methods to calculate the pitch like cepstrum, autocorrelation, manual method and many more. Here we have discussed cepstrum method.

The word ‘cepstrum’ was derived by using the first four letter of spectrum [7]. A reliable way of obtaining an estimate of the dominant fundamental frequency for long clean stationary speech signal is to use cepstrum. The cepstrum is a Fourier analysis of the logarithmic amplitude spectrum of the signal. If the log amplitude spectrum contains many regularly spaced harmonics, then Fourier analysis of the spectrum will show a peak corresponding to the spacing between the harmonics i.e. fundamental frequency. Here signal spectrum is treated as another signal, then looking for periodicity in the spectrum itself. The cepstrum is so called because it turns the spectrum inside out. The X axis of cepstrum has unit of quefrency & peak in cepstrum is called rahmonics [6]. If \( X(n) \) is the speech signal then logarithmic spectrum is given by

\[ Y(n) = FFT[X(n)] \]

\[ Y(n) = 20 \times \log_{10}[\text{abs}(Y(n))] \]

The cepstrum is DFT of log spectrum

\[ Y(n) = FFT[\text{log}(\text{abs}(Y(n)))] \]

Figure 3 shows spectrum and cepstrum of a speech signal. From the cepstrum pitch can be calculated.

B. Pitch Calculation
Although many pitch detection algorithm both in time and frequency domains, have been proposed [8]. However, performance improvement in noisy environments is still desired [9].

Here we are proposing cepstrum method, which shows good performance for quasi-periodic signals [10]. And manual method to get the pitch of speech signal.

The Figure 4 shows the calculation of pitch using proposed method of cepstrum. As we know that x-axis of cepstrum has unit of quefrency & peaks in cepstrum (which relates the periodicity in the spectrum) are called rahmonics. To obtain an estimate of the pitch from the cepstrum we look for the peak in the quefrency region corresponding to typical speech fundamental frequencies (1/quefrency). The pitch of the signal under consideration is found to be 129.668Hz.

The second method which is proposed is manual method here the period of the segment can be calculated by finding the time difference of two successive peaks of figure 4 and doing the calculation of pitch by manual method.

\[ F = \frac{1}{T} = \frac{1}{T_2 - T_1} = \frac{1}{0.054 - 0.0463} = 129.87 \text{ Hz} \]
C. Formant Frequency Estimation

The formant frequency are calculated using linear predictive coding (LPC). The system function and difference equation of LPC filter is given below.

\[
H(z) = \frac{X(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^{P} a_k z^{-k}} = \frac{1}{A(z)}
\]

\[
x[n] = \sum_{k=1}^{P} a_k x[n-k] + e[n]
\]

D. Mel Frequency Cepstral Coefficient (MFCC)

Firstly, the signals are framed into short frames, then for each frame period gram estimate of the power spectrum is calculated. Then this power spectrum is given to the Mel filter bank this sums the energy in each filter. Thereafter, logarithm followed by DCT of the filter bank energies is taken.

DFT for \(i^{th}\) frame is given by

\[
S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-j2\pi kn/N}
\]

Where, \(h(n)\) is hamming window and \(k\) is the length of DFT

Power spectrum of \(i^{th}\) frame is given by

\[
P_i(k) = \frac{1}{N}|S_i(k)|^2
\]

Design of mel filters separates MFCC from HFCC. The N-point DCT of a sequence of N samples is defined as:

\[
X[k] = \frac{2}{N} a[n] \sum_{n=0}^{N-1} x[n] \cos \left( \frac{\pi k (2n + 1)}{2N} \right)
\]

\[0 \leq k \leq N - 1\]

Where, \(a[n]\) is given by

\[
a[n] = \begin{cases} 
1 & \text{if } k = 0 \\
1/\sqrt{2} & \text{if } 1 \leq k \leq N - 1 \\
0 & \text{otherwise}
\end{cases}
\]

II. CLASSIFICATION USING SVM

A Support Vector Machine (SVM) is a machine learning technique that has generated a lot of interest in speech processing in recent years. The greatest benefit of an SVM is its skill to construct nonlinear decision regions in discriminative fashion. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. A support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Figure 5 shows how classes are separated.
A linear support vector machine is composed of a set of given support vectors \( z \) and a set of weights \( w \). The computation for the output of a given SVM with \( N \) support vectors \( z_1, z_2, \ldots, z_N \) and weights \( w_1, w_2, \ldots, w_N \) is then given by:

\[
F(x) = \sum_{i=1}^{N} w_i k(z_i, x) + b
\]

**RESULT** We have taken 100 samples, i.e., 33 Neurological, 33 Laryngeal, 34 Normal samples. Training was done on 75% of samples and testing was done on 25% of random samples. Confusion matrix for different sets of features are shown in table I and II and accuracy in table III. Accuracy is 100% with second set of feature.

**REFERENCES**


**TABLE I. CONFUSION MATRIX FOR FIRST SET OF FEATURES**

<table>
<thead>
<tr>
<th>Output/Desired</th>
<th>Neurological</th>
<th>Laryngeal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laryngeal</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE II. CONFUSION MATRIX FOR SECOND SET OF FEATURES**

<table>
<thead>
<tr>
<th>Output/Desired</th>
<th>Neurological</th>
<th>Laryngeal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laryngeal</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III. PERFORMANCE TABLE AFTER TESTING**

<table>
<thead>
<tr>
<th>Features</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch+Formants</td>
<td>88%</td>
</tr>
<tr>
<td>Pitch+Formants + MFCC+HFCC</td>
<td>100%</td>
</tr>
</tbody>
</table>

IJRITCC | February 2015, Available @ http://www.ijritcc.org