ISSN 2278-7763

Impact of Technology for Discriminating Schizophrenia ,OCD Patient's Speech from Normal Speech Using Spectrum, Cepstrum, Pitch Determination Technique, Feature Extraction & Analysis.

Prof. Syed Mohammad Ali, Prof. Dr. Pradeep Tulshiram Karule, Prof. S.I.Ali, Prof. S.A.Ali

[1]Department of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, India; [2]Department of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Wanadongri, Nagpur, India. [3]Department of Electronics & Telecommunication Engineering, ACET, Nagpur, India;
Email: ali_acet2003@yahoo.co.in; ptkarule@rediffmail.com; irfan_extc@yahoo.co.in; syed_akbar_ali_pro2@yahoo.co.in

ABSTRACT

The automation in voice disorder recognition technologies has developed more and more momentum now days because of the complications in conventional methods. Due to the nature of jobs, unhealthy social habits people are subjected to risk of speech problem [1]. Neurological disorders develops problem of voice disorder. Therefore, voice signal can be a useful tool to diagnose them.
In the present paper, normal &neurological disorder speech samples are taken & a system is designed to differentiate normal from neurological signals. These signals are first preprocessed. Preprocessing techniques involves passing signal through pre-emphasis filter, moving average filter (ma), framing & windowing .The windowed signal is given for spectral analysis. In spectral analysis, various methods like logarithmic spectrum, cepstrum, auto correlation of speech signal, spectrograms are applied to differentiate normal & neurological speech signals. It has been found that with the above method normal & neurological speech signals are differentiated. Pitch of speech signal plays a vital role in analysis of speech signal, so using different techniques, pitch of normal & neurological speech signals are found. More over features like formants extraction and there analysis is done using LPC filter respectively.

Keywords : Neurological speech signal; preprocessing; spectral analysis; pitch; formants estimation; LPC filter; entropies;

1 INTRODUCTION

SPEECH disorder detection has received great momentum in the last decade. Digital signal processing has become an important tool for voice disorder detection [2]. Pathological voice signal & Normal signals are taken. The pathological speech signals were taken from the Government Medical College & Hospital, Nagpur & Dr. Naresh Agarwal's Hospital Nagpur. The signals were recorded keeping the microphone two inch away from the mouth, using voice recorder of Window XP. The sampling frequency is chosen to 11025 samples/sec, 8 bit stereo 21 kb/sec. The patients were told to pronounce vowel 'a', vowel consonant 'ah' & word 'Hello'. Physicians often use invasive techniques like endoscopy to diagnose symptoms of vocal fold disorders however; it is possible to diagnose the disease using certain features like formants of speech signal [2].
Speech signal is a sinusoidal signal having different frequency, different amplitude & different phase. It is given by the expression given below [3].

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

Where $A_i(t)$, $F_i(t)$ & $\theta_i(t)$ are the sets of amplitudes, frequencies & phases respectively, of the sinusoids. Speech production requires close cooperation of numerous organs which from the phonetic point of view may be divided into the following organs.

1. Lungs, Bronchi, Tracheas (producing expiration air steam necessary for phonation)
2. Larynx (amplifying the initial tone)
3. Root of the tongue, throat, nasal cavity, oral cavity (forming tone quality & speech sound) [4].

The use of non-invasive techniques to evaluate the larynx and vocal tract helps the speech specialists to perform accurate diagnosis [5]. Speech signal is non-intrusive in nature & it has the potential for providing quantitative data with reasonable analysis time. So the study of speech signal of pathological voice has become an important topic for research as it reduces work load in diagnosis of pathological voices [6].

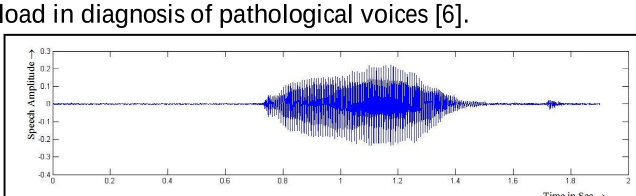

Fig. 1. A speech signal

Figure 1 shows the speech signal of a patient suffering from neurological disorder. The algorithm in figure 2 shows the flow of control. Here, in this paper we have taken speech samples of neurological disorder and normal persons. These speech samples are passed through moving average filter and high pass Filter. The filtered output is framed and then each frame is

ISSN 2278-7763

passed through a window. The output signal, which was framed and windowed, is used for spectral analysis. In spectral analysis, logarithmic spectrum of framed window signal is found, which is then used to get cepstrum. Framed signal is also used for finding autocorrelation of speech signal. From cepstrum, the pitch of the signal is found. Other manual method of pitch estimation, along with formant &entropies estimation is discussed.

2 ALGORITHM FOR SPECTRAL ANALYSIS

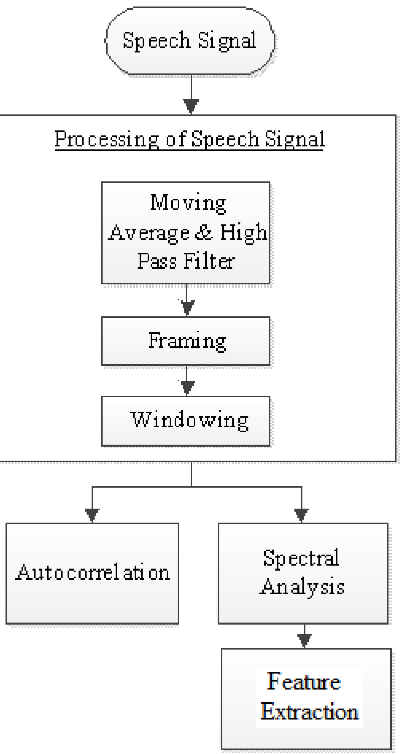

Fig. 2. Flow chart of Spectral Analysis of Speech Signal

The figure 2 shows the algorithm of spectral analysis. The speech signal is passed through the moving average filter, which takes the average of a sample for filtering the noise signal. The expression for output of such a filter is given below [7].

$$Y(n) = \frac{X(n) + X(n-1) + X(n-2)}{3} \tag{2}$$

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

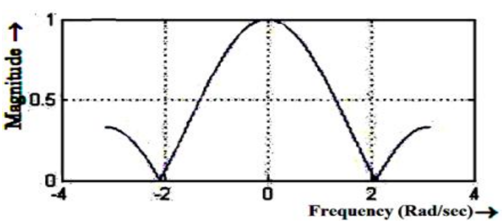

Fig. 3. Flow chart of Spectral Analysis of Speech Signal

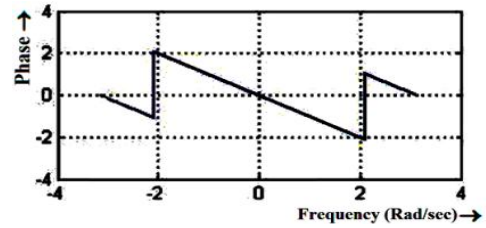

Fig. 4. Phase v/s Frequency plot of moving average filter

The figures 3 & 4 are magnitude and phase plot of the moving average filter. The cut-off frequency of low pass moving average filter is 2 rad / sec. Whenever magnitude changes its sign, a phase jump of $+\pi$ radians on right hand side and $-\pi$ radians on left hand side will occur. Optimum result is obtained by taking average of three samples.

The input to pre-emphasis filter is the output of moving average filter. The Pre-emphasis filter is a 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 [8]. The pre-emphasis filter amplifies the area of spectrum, which improves the efficiency of spectral analysis [9].

The time domain presentation of filter will be

$$Y(n) = X(n) - \lambda X(n-1)$$

Where y(n) is the output, x(n) is the input speech sample & λ is the filter coefficient. With λ = 0.9375 optimum result of filtering is received [10]. The output of this filter is framed & passed through a window. This is done as the speech signals are analyzed for a short period of time (5 msec to 100msec) where the signal is fairly stationary. Windowing is done to avoid problems due to truncation of signal & moreover window helps in the smoothening of signal [11].

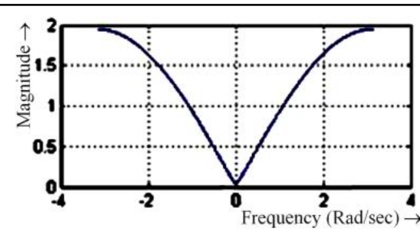

Fig. 5. Magnitude v/s Frequency plot of pre-emphasis filter

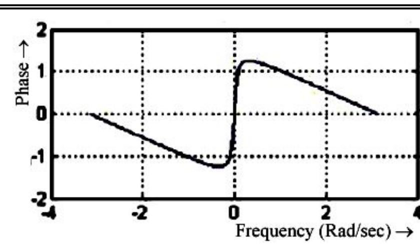

Fig. 6. Phase v/s Frequency plot of Pre-emphasis Filter

The figures 5 & 6 are the magnitude and phase spectrum of pre-emphasis filter. In magnitude plot, the magnitude changes its sign at the origin, so phase jump of $+\pi$ radians is clearly seen in the phase plot.

3 Classification of Speech Using Cepstrum

The name 'cepstrum' was derived by using the first four letter of spectrum [12]. A reliable way of obtaining an estimate of the dominant fundamental frequency for long clean stationary speech signal is to use the 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 & periodicity is searched 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 & the peak in cepstrum is called rahmonics [13].

If X(n) is the speech signal then logarithmic spectrum is given by

$$Y_1(n) = \mathrm{FFT}\{X(n)\}$$

$$Y_2(n) = 20 \times \log_{10}[\mathrm{abs}\ Y_1(n)]$$

The cepstrum is DFT of log spectrum

$$Y_3(n) = \mathrm{FFT}\{\log[\mathrm{abs}(Y_2(n))]\}$$

cepstrum in figures 7 & 8, we have classified the normal and abnormal speech.

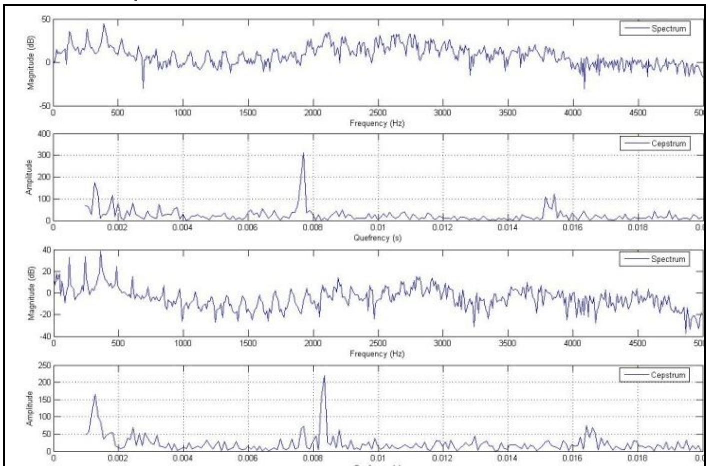

Fig. 7. The spectrum and cepstrum of normal persons

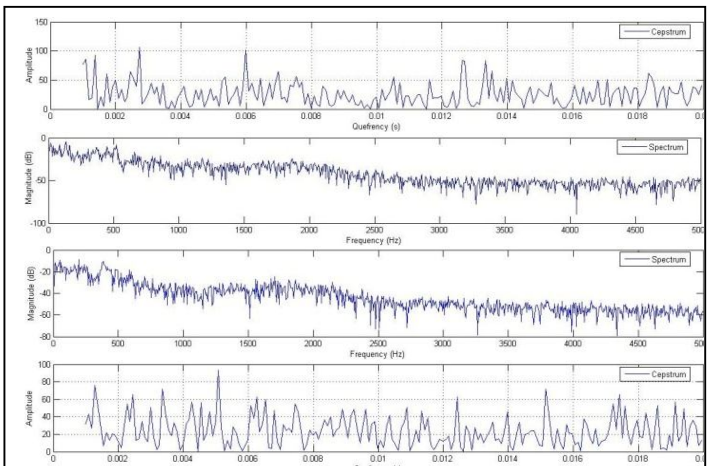

Fig. 8. The spectrum and cepstrum of normal persons

4 Auto Correlation of Pathological and Normal Signal

Another method which is applied for the classification of pathological speech signal from normal signal is the Autocorrelation method. Using this method, one can easily classify the normal and abnormal speech signals. The autocorrelation of discrete time signal X(n) is given by [14].

$$r_{xx}(\ell) = \sum_{n=-\infty}^{+\infty} X(n) \cdot X(n-\ell) \qquad \ell = 0, \pm 1, \pm 2, \cdots$$

The autocorrelation function of a signal is a transformation of signal, which is useful for displaying structure in the waveform [15]. Here it is shown how the autocorrelation function classifies the signals. For the normal signal, the decay of autocorrelation of signal with respect to time is exponential whereas for abnormal, the decay will not be exponential. Their results are shown in figure 9 & 10.

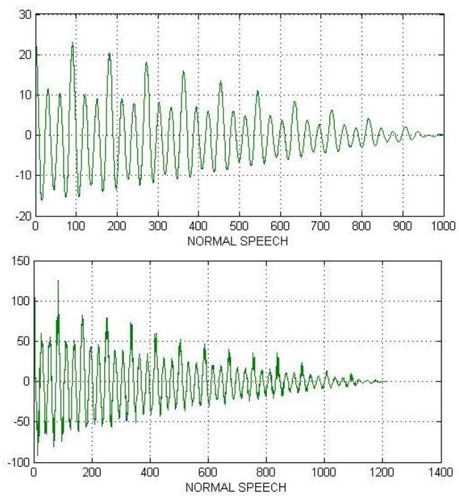

Fig. 9. Autocorrelation of normal speech

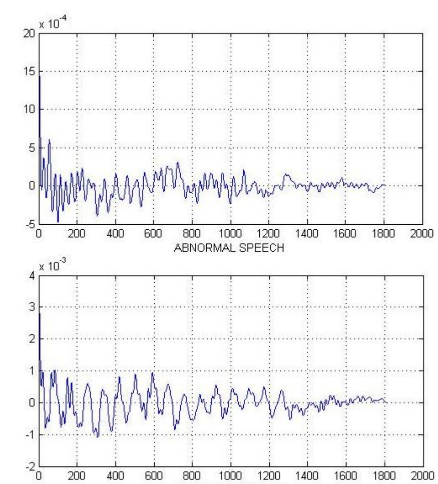

Fig. 10. Autocorrelation of neurological disorder

5 Spectrogram

Among the distinct signal processing techniques employed for voice analysis, the spectrogram is commonly used as it allows the visualization of variation of energy of the signal as function of both time and frequency [16]. The study investigates the use of the global energy of the signal estimated through spectrogram as a tool for discrimination between signals obtained from healthy and pathological subjects.

A spectrogram is a display of frequency content of the signal drawn so that energy content in each frequency region and time is displayed on a colored scale. The horizontal axis of spectrogram is time and the picture shows how the signal develops and changes over time. The vertical axis of the spectrogram is frequency and it provides an analysis of signal into different frequency regions as shown in figure 11 & 12. One can treat each of these signals as comprising a particular kind of building blocks of the signal. If a building block is present in the signal at particular time then highlighted region will be shown at the frequency of building block and time of the event [17].

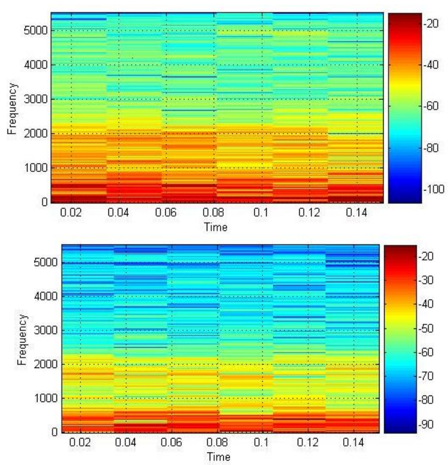

Fig. 11. The spectrogram of normal person

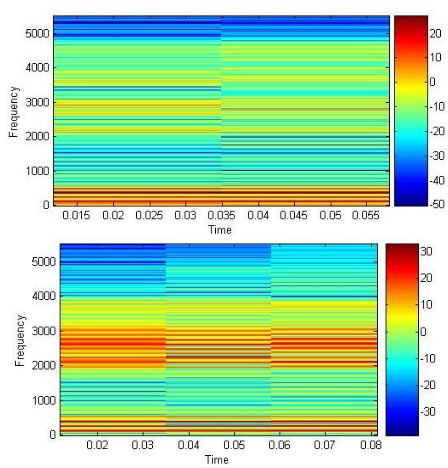

Fig. 12. The spectrogram of neurological disorder

6 Pitch Calculation

Pitch detection is an essential task in a variety of speech processing applications. Although many pitch detection algorithm both in time and frequency domains have been proposed [18]. However, performance improvement in noisy environments is still desired [19].

Here we are proposing cepstrum method, which shows good performance for quasi-periodic signals [20] and manual method to get the pitch of normal and pathological signal. The Figure 13 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 pitches of the signals under consideration are found to be 129.668Hz & 119.33Hz.

The second method which is proposed is manual method here the period of the signal can be calculated by finding the time difference of two successive peaks. Figures 13 & 14 explain the calculation of pitch by manual method.

$$F=\frac{1}{T}=\frac{1}{(T_2-T_1)}=\frac{1}{(0.054-0.0463)}=129.87\text{Hz}$$

$$F=\frac{1}{T}=\frac{1}{(T_2-T_1)}=\frac{1}{(0.05-0.04166)}=119.33\text{Hz}$$

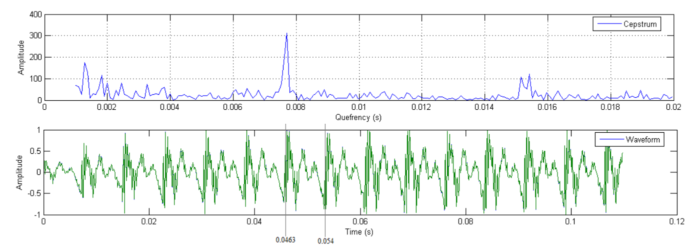

Fig. 13. Calculation of Pitch using Cepstrum & Manual Method

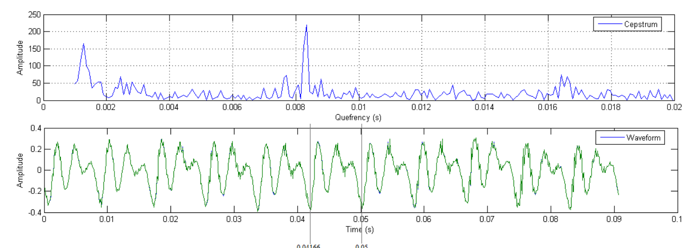

Fig. 14. Calculation of Pitch using Cepstrum & Manual Method

7 Formant Frequency Estimation

Formants are frequencies of resonance for each frame. It is often measured as an amplitude peak in frequency spectrum of the speech formants are resonances of the vocal tract [22]. The formant frequency is calculated using linear predictive coding (LPC). The formant frequency is obtained by finding the roots of prediction polynomial. The LPC finds the best IIR filter from the section of speech signal and then plots frequency response of filter [21].

$$H(z)=\frac{X(z)}{E(z)}=\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}}=\frac{1}{A(z)}$$

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

The two common preprocessing steps applied to speech waveform before linear prediction coding are windowing and pre-emphasis filtering The pre-emphasis is high pass all pole filter. There finding the roots of prediction polynomial return by LPC Because the LPC co-efficient are real valued the roots occur in complex conjugate pair . Retain only the roots on the one sign for imaginary part & determine the angle corresponding to the roots. Convert the angular frequencies in radians to hertz to get formant frequencies [22].

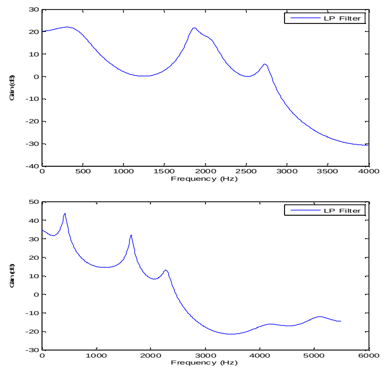

Fig. 15. Formants of neurological patient

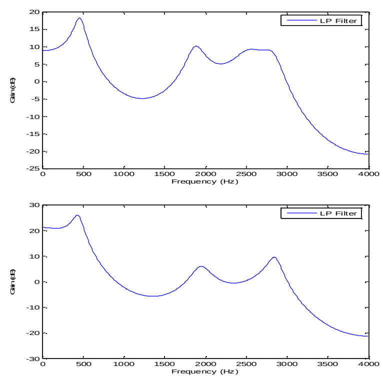

Fig. 16. Formants of normal person

8 RESULTS & DISCUSSION

In this paper, we have taken speech samples of neurological disorder patient and applied various techniques of classifications like logarithmic spectrum, cepstrum, autocorrelation, & spectrogram. Seeing figure from 7 to 12 & figure 15-16, one can easily differentiate the abnormal speech samples from normal. The table shows formants of normal person are close to each other while neurological patients have more number of formants.

Formants of neurological patient	
298.9992	426.4350
438.8949	1.2051e+003
1.8530e+003	1.6433e+003
2.0693e+003	2.2937e+003
2.7427e+003	4.1510e+003
	5.0887e+003
Formants of normal person	
463.3569	443.9122
1.8752e+003	1.9416e+003
2.5162e+003	2.7185e+003
2.8362e+003	2853.7

REFERENCES

[1] M.Hariharan, M.P.Paulraj,Sazali, aacob, "*Time Domain Features and Probabilistic Neural Network For the Detection Of Vocal Fold Pathology*", Malaysian journal Of Computer Science, Vol(23) 2010 pp. 60-67.

[2] Lotfi Salhi,Talbi Mourad,and Adnene Cherif, "*Voice Disorders Identification Using Multilayer Neural Network*", The International Arab Journal of Information Technology Volume 7-No.2, pp.177-185 April 2010.

[3] John G.Proakis, Dimitris, G.Manolakis,"*Digital Signal Processing. principles, Algorithm and Applications*", Prentice Hall India, pp.2 Third Eition.

[4] Orzechowski, A Izworski, R. Tadeusiewiez K. Chmunzynska, P. Radkowski, I Gotkowska, "*Processing of pathological change in speech caused by Dysarthria*", IEEE Proceedings of 2005 International symposium on intelligent signal processing & communication system, pp 49-52, Dec 13-16, 2005.

[5] Marcelo de Oliveira Rosa,Jose Carlos Pereira and Marcos Gellet, "*Adaptive Estimation of Residue Signal for Voice Pathology Diagnose*" IEEE Transaction on Biomedical Engineering , Vol.47, No.1, Jan.2000.

[6] Martinez Cesar E, Refiner Hugo L, "*Acoustic Analysis of speech for detection of Laryngeal pathologies*", IEEE proceeding of the 22nd Annual EMBS international conference chicago IL, pp.2369-2372, 2000.

[7] John G.Proakis, Dimitris, G.Manolakis,"*Digital Signal Processing. principles, Algorithm and Applications*", Prentice Hall India, pp.309 Third Eition.

[8] Antanas Lipeika, Joana Lipeikiene, Laimutis Telksnys, "*Development of Isolated speech Recognition System*", INFORMATICA, 2002, Vol 13, No.1, 37-46, 2002.

[9] J.W.Picone, "*Signal modeling techniques in speech recognition*", Proceedings of the IEEE, Vol.81, No.9, Sept.1993.

[10] Milan Sigmund, "*Voice Recognition By Computer*", Tectum Verlag publication, pp no20-22

[11] Manjot Kaur Gill,Vector, "*Quantization based Speaker identification*", International Journal of Computer Application (0975-8887) Volume 4-No.2, pp1-4 July 2010.

[12] http://en.wikipedia.org/wiki/cepstrum

[13] R.W.Schafer and L.R.Rabiner, "*System for automatic formant analysis of voiced speech*", J.Amer., vol.47, pp 634-648, Feb.1970

[14] John G.Proakis, Dimitris, G.Manolakis,"*Digital Signal Processing. Principles, Algorithm and Applications*", Prentice Hall India, pp.122 Third Eition.

[15] Lawrence R. Rabiner,"*On the use of Autocorrelation Analysis for Pitch Detection*", IEEE Transaction Acoustics, Speech and signal Processing, Vol.ASSP-25, No.1, pp24-30, February 1977.

[16] M.Fernandes, F.E.R.Mattioli, E.A.Lamounier Jr.and A.O.Andrade, "*Assesment of Laryngeal Disorders through TheGlobal Energy of Speech*", IEEE Latin American Transactions, vol.9, No.7, December 2011.

[17] http://www.phon.ucl.ac.uk/resource/sfs/rtgram/

[18] Sasan Ahmadi, Andreas S.Spanians, "*Cepstrum-Based PitchDetection using a new Statistical V/UV Classification Algorithm*" IEEE Transaction On Speech and Audio Processing ",Vol.7,No.3,pp335-337,May 1999.

[19] Tetsuya Shimamura,Hajime Kobayashi, "*Weighted Autocorrelationfor Pitch Extraction of Noisy Speech*", "IEEE TransactionOn Speech and Audio Processing",Vol.9,No.7,October 2001.

[20] Climent Nadeu, Jordi Pascual and Javier Hernando, "*Pitch Determination Using The Cepstrum Of The One-Sided Autocorrelation Sequence IEEE Proceedings*", pp 3677-3680, 1991.

[21] Fatema N Julia, Khan, m. Iftekharuddin, "*Detection of Emotional Expression in Speech*","IEEE Conference Procedings", pp.307-312, 2006.

[22] Snell .R. "*Format location from LPC analysis data* ", IEEE transaction on speech an audio processing, 1(2), PP. 129-134, (1993