Performance Analysis of MFCC and LPCC Techniques in Automatic Speech Recognition

Dr. Madhu Goel[1] Anu Gupta[2]
[1]HOD & A.P., [2]P.G. Scholar
[1, 2]KITM, Kurukshetra, Haryana, India

Abstract—*As a result of ample development of computers, the various types of information exchange between man and computer are discovered. At present, inputting the data in computer by speech and converting the data into another form for eg. Text. with the help of automatic speech recognition system and its recognition by the computer is one of the developed scientific fields. As each language has its specific feature, the various speech recognition systems are investigated for the different languages. In this paper, we have taken two algorithms known as MFCC and LPCC. These two algorithms are used for feature extraction. The performances of the two algorithms are compared to achieve better performance with high recognition rate and low computational complexity and the major advantage of comparing these two algorithms is that they improves the reliability of the system.*

Keywords:- MFFC, LPC, ASR ,DCT.

I. Feature Extraction Techniques

The goal of feature extraction is to represent any speech signal by a finite number of measures of the signal.. Each feature is a representation of the spectrum of speech signal in each window frame. More recently, the majority of the system has conversed to the use of a cepstral vector derived from a filter bank that has been designed according to some model of the auditory systems .MFCC and LPC are two of most commonly used methods.

II. Mel-Cepstrum

MFCC is given by Davis and Mermelstein [2] as a beneficial approach for speech recognition. Figure 1 llustrates the complete process to extract the MFFC vectors from the speech signal. It is to be emphasized that the process of MFCC extraction is applied over each frame of speech signal independent.

Fig.1: MFCC extraction process

The filter bank is a set of overlapping triangular band pass filter, that according to Mel-frequency scale, the centre frequencies of these filters are linear equally-spaced below 1 kHz and logarithmic equally-spaced above. The Mel filter bank is illustrated in Figure 5.3**.** It is interesting to emphasize that these centre frequencies correspond to Mel centre frequencies uniformly spaced on Mel-frequency domain.

Thus, the input to the Mel filter bank is the power spectrum of each frame, X frame[k], such that for each frame a log-spectral-energy vector, Eframe[m], is obtained as output of the filter bank analysis. Such log-spectral-energy vector contains the energies at centre frequency of each filter. So, the filter bank samples the spectrum of the speech frame at its centre frequencies that conform the Mel-frequency scale. Let's define Hm[k] to be the transfer function of the filter m, the log-spectral energy at the output of each filter can be computed as in Eq. (5.1) [9] ; where M (m=1, 2, ..., M) is the number of Mel filter bank channels. M can vary for different implementations from 24 to 40 (Huang et al., 2001)[5].

$$E[m] = \sum_{k=1}^{K-1} \ln[X[k]^2 H_m[k]]$$

m=1, 2, 3 …M (5.1)

Using the Mel filter bank is subjected to two principal reasons:

- Smooth the magnitude spectrum such that the pitch of a speech signals is generally not presented in MFCCs.
- Reduce the size of the features involved.

The last step involved in the extraction process of MFCC is to apply the modified DCT to the log-spectral-energy vector, obtained as input of Mel filter bank, resulting in the desired set of coefficients called Mel Frequency Cepstral Coefficients. In order to compute the MFCCs for one frame, the DCT-II is applied to the log spectral-energy vector of such frame.

$$C_i=\sqrt{\frac{2}{M}}\sum_{m=1}^{M} E_m \cos\left(\frac{\pi i}{M}(m-\frac{1}{2})\right)$$

(5.2)

Cepstral coefficients have the property that both the variance and the average numerical values decrease as the coefficient index increases. In this way, the M filter bank channels can be become into only L MFCCs (L < M) used in the final feature vector. The truncation of the cepstral sequence has a general spectral smoothing effect that is normally desirable because it tends to remove phonetically irrelevant detail [5].

III. Linear Prediction

LPC analysis is an effective method to estimate the main parameters of speech signals [8].. The conclusion extracted was that an all-pole filter, H(z) , is a good approximation to estimate the speech signals. Its transfer function was described. In this way, from the filter parameters (coefficients, {ai}; and gain, G), the speech samples could be synthesized by a difference equation. Thus, the speech signals resulting can be seen as linear combination of the previous p samples. Therefore, the speech production model can be often called linear prediction model, or the

autoregressive model. From here, p, indicates the order of the LPC analysis; and, the excitation signal, e[n], of the speech production model can be called prediction error signal or residual signal for LPC analysis.

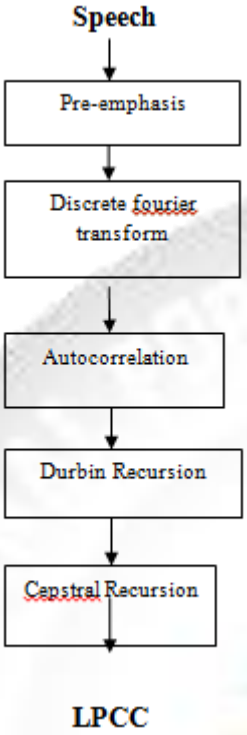

Fig.2: LPC coefficients extraction process

After the LPC analysis, the power spectrum of the speech frame can be calculated from its LPC parameters. Let's define A(z) to be the inverse transfer function of the filter

$$A(z)=1-\sum_{i=1}^{p} a_i z^{-i} \qquad (5.4)$$

From this inverse filter, A(z), a new speech synthesis model is proposed in Figure 4.11**,** which can be considered as inverse model of speech production model

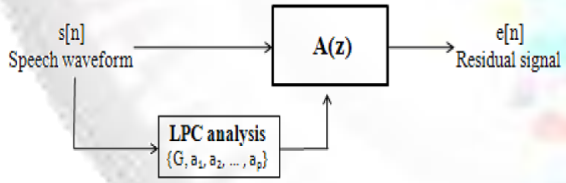

Fig.3: Synthesis LPC filter

The power spectrum of one signal can be obtained by passing one input signal through a filter. If the input signal is the speech signal and the filter is the inverse LPC filter A(z); the power spectrum of the output signal, in this case the residual signal or prediction error signal, can be obtained as:

$$S(\omega)\,A(\omega)^2=\sigma^2(\omega) \qquad (5.5)$$

Then, one can see that the power spectrum of the speech signal can be approximated by the response of a sampled-data-filter, whose all-pole-filter transfer function is chosen to give a least-squared error in waveform prediction . So, in Eq. (5.6), the power spectrum of the speech frame is obtained from its LPC coefficients.

$$S(\omega)=\frac{\sigma_\omega^2}{(1-\sum_{i=1}^{p} a_i e^{-ij\omega})} \qquad (5.6)$$

LPC analysis produces an estimate smoothed spectrum, which much of the influence in the excitation removed. LPC-derived features have been used by many recognition systems, being its performance comparable whit the one obtained from recognizers using filter bank methods [4].

IV. ANALYSIS OF LPC

In order to show the performance of the different steps involved in LPC extraction process, the following figures were executed for save yourself.wav file. In Figure 5.1, the original speech waveform and how is affected after the pre-emphasis filter is illustrated. Figure 5.2 presents the effect of using a Hamming window, and Figure 5.3 shows the Linear Predictor spectrum of one frame as compared with its magnitude spectrum.

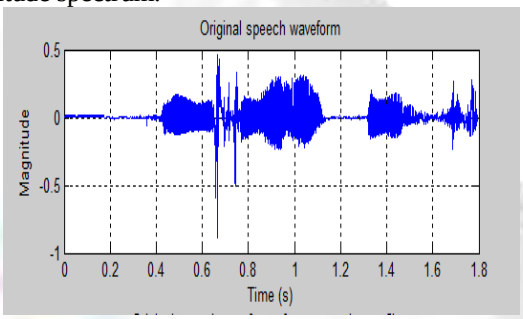

Figure 4.1 (a): Original speech waveform

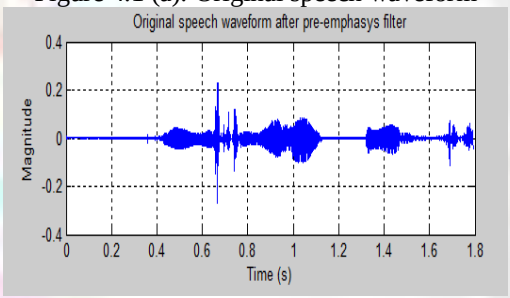

Figure 4.1(b): original speech waveform after the pre-emphasis filter with coefficient equal to 0.97

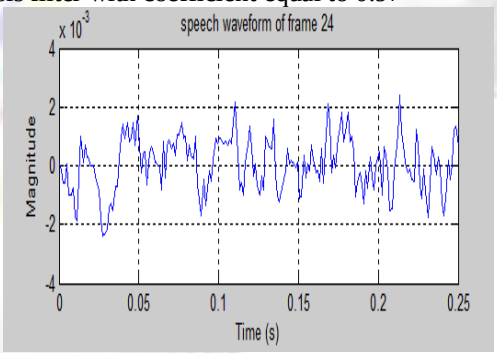

Figure 4.2(a): speech waveform of frame 24

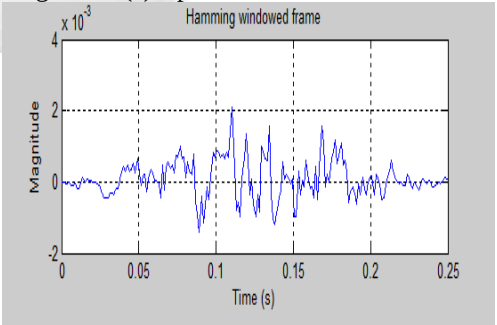

Figure 4.2(b): Effect of multiplying one speech frame by a Hamming window

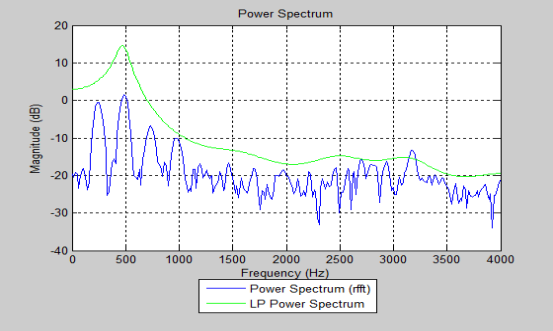

Figure 4.3(a): Comparison of the power spectrum computed from LPC coefficients with the original magnitude spectrum

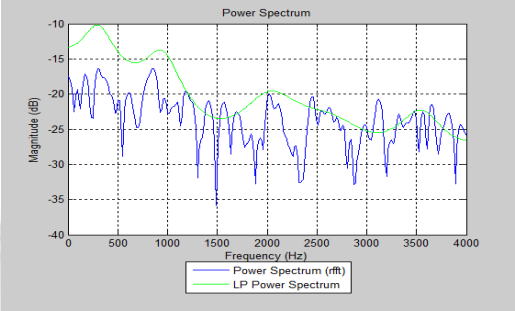

Figure 4.3(b): Comparison of the power spectrum computed from LPC coefficients with the original magnitude spectrum

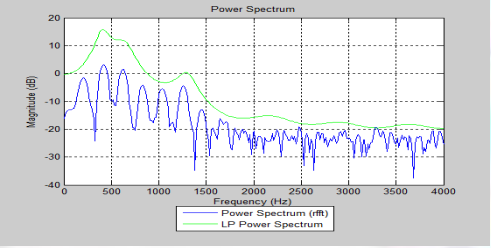

Figure 4.3(c): Comparison of the power spectrum computed from LPC coefficients with the original magnitude spectrum

V. Analysis Of Mfcc

The MFCC coefficients are the DCT-II of the log spectral energies at the centre frequencies of the Mel filter bank. The Fourier Transform of a speech frame is transformed to a Mel-frequency scale by the filter bank analysis with M channels. The output of this process is the M log-spectral-energies at Mel centre frequencies. The DCT-II allows an energy compaction in its lower coefficients. So, the use of the DCT-II makes that the M filter bank channels can be reduced to L (L<M) MFCC coefficients. This truncation into the cepstral components allows recovering a smoothed spectral representation in which phonetically irrelevant detail has been removed.[1]

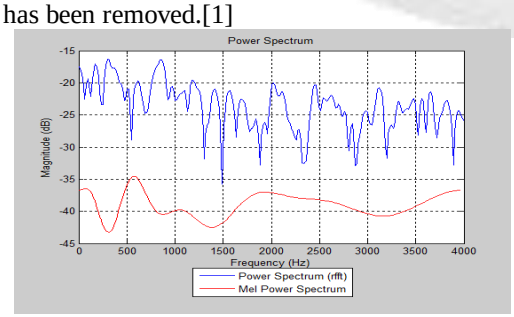

Figure 5.4: Mel power spectrum of one speech frame compared with its magnitude spectrum

Figure 5.4 demonstrates that the Mel power spectrum is the smoothed spectral envelope of the magnitude spectrum of the speech frame. In this case, the harmonics of the speech spectrum are flattened because of, the reduction of the frequency resolution performed with in Mel-filter bank analysis and, the truncation of higher-order coefficients in the DCT-II computation.

VI. Conclusion

The major contribution of this work is the implementation of the Automatic Speech Recognition (ASR) algorithms using MATLAB and evaluates their individual performance. The system model has been developed to compare two algorithms, which is MFCC and LPCC. The performance has been evaluated by considering ten sets of speech signal.
It is shown that MFCC used in Automatic speech Recognition system provide 80 percentage accuracy where as LPCC used in Automatic Speech Recognition provide 60 percentage accuracy.

Results and calculations show that MFCC algorithm provides better result in comparison with LPCC algorithm. From the simulation results we conclude that MFCC algorithm, which require more computation but perform better than LPCC in terms of efficiency and accuracy.

Refrences

[1] B.S.Yalamanchili1, Anusha.K.K2, Santhi.K3, Sruthi.P4, SwapnaMadhavi.B5" Non Linear Classification for Emotion Detection on Telugu Corpus" B.S.Yalamanchili et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2443-2448

[2] Brookes, M., Voicebox: Speech Processing Toolbox for Matlab [on line], Imperial College, London, available on the World Wide Web: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[3] C. Chelba, T.J. Hazen, and M. Saraclar," Retrieval and Browsing of Spoken Content", IEEE Signal Processing Magazine 25 (3), May 2008

[4] Daniel Jurafsky, and James H. Martin ,"Speech and language Processing", Pearson Education, 2000

[5] Douglas O'Shaughnessy, "Interacting with Computer by Voice Automatic Speech Recognition and Synthesis", Proceeding of the IEEE, Vol.91, No.9, pp.1272-1305, Sept 2003

[6] Frederick Jelinek, "The Dawn of Statistical ASR and MT, Computational Linguistics", Vol.35, No. 4, pp. 483-494, Dec 2009.

[7] Guodong Guo and Stan Z. Li,"Content based audio classification and retrieval by SVMs", IEEE trans. Neural Network, Vol.14, pp.209-215, Jan 2003.

[8] Gray Jr., A.H. & Markel, J. D. (1976), "Distance Measures for Speech Processing", IEEE Transactions on Acoustics, Speech and Signal Processing, issue 5, pp. 380-391, Oct 1976.

[9] H.Fletcher," Auditory Pattern Review of Modern Physics", Jan 1940.