International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Review on Feature Extraction in speech processing

Priyanka R. Kharche[1], Prof. S. S. Bhabad[2]

[2]*E&TC Department, K.K.wagh nashik,priyankakharche214@gmail.com*
[1]*E&TC Department, K.K.wagh nashik,ssb.eltx@gmail.com*

Abstract— Feature extraction is a most important part of speech recognition. Now a days number of researcher focused on area of speech recognition. In this paper we will discuss one of the process of speech recognition namely feature extraction. It plays an important role to separate one speech from other. Because every speech has different individual characteristics. Feature extraction is commonly used spectral analysis, parametric Transform and statistical modeling techniques. In speech recognition system MFCC is most important methods. we discuss the MFCC and observe the feature of some samples of speech.

Keywords- feature extraction; linear predictive coding analysis; Mel frequency cepstral Coefficients; speech recognition;

I. INTRODUCTION

In today's speech recognition system is most commonly used in word. Speech recognition system is a process where speech signals are recognize. Speech recognition is a technique that enables to accept and understand spoken word as an input. User wants their voice, speech signals in to be transcribed into proper way. Isolated word, connected words, continuous speech and spontaneous speech are main classes that are used in speech command recognition. Speech recognition process are divide in three parts as shown in figure.

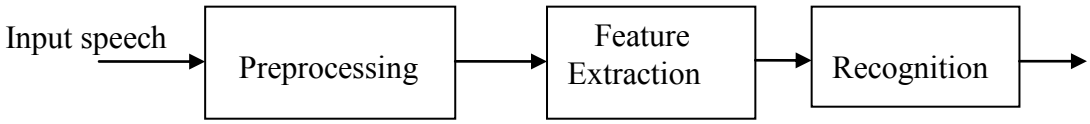

Figure 1 : Speech recognition system

Feature extraction is the main and important part of speech recognition system. In this paper we discuss on feature extraction and their methods. After preprocessing of the input speech signal we extract these speech vectors.

1. We will discuss spectral analysis, parametric transform and statistical modeling technique of feature extraction.
2. We will discuss the method of the feature extraction techniques.

A. Feature Extraction:

Feature extraction is the most important part of speech recognition as it distinguishes one speech from other. Feature extraction can be subdivided into three important operations: spectral analysis, parametric transformation and statistical modeling. Step by step process of feature extraction shown in figure 2.

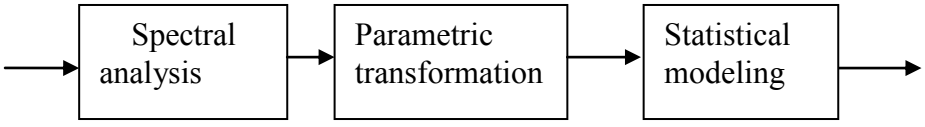

Figure 2: Feature extraction

1.Spectral analysis: Spectral analysis is the major part of the feature extraction method. When speech is produced in the sense of time varying signal then it divide into some analysis techniques. Linear predictive coding gives best results, linear predictive coding provide the estimate of the pole of vocal tract transfer function. The LPC algorithm is a *pth* order linear predictor which attempts to value of any points in a time various linear system based on the value of previous P samples. All pole representation of vocal tract transfer function, H(z) can be represented by following equation.

$$\mathrm{H(z)} = \frac{G}{A(z)} = \frac{G}{1 + a_1 z^{-1} + a_2 z^{-2} ..... + a_p z^{-p}}$$

The value of a(i) is called prediction coefficients. While G represents amplitude or gain associated with the vocal tract excitation. For the speech signal s(n) produce by a linear system, the predicted speech sample $\hat{s}$ (n) is a function of a(i) And prior speech according to:

$$\hat{s}\ (\mathrm{n}) = \sum_{i=1}^{p} a(i) s(n-i)$$

LPC analysis involves the solving for the a(i) term according to lest error criteria. If the error is define as :

E(n) =s(n) - $\hat{s}$ (n)

$$\mathrm{S(n)} = \sum_{i=1}^{p} a(i) s(n-i)$$

Then taking a derivative of square error with respect to coefficient of a(j).and setting it equal to zero gives:

$$\frac{\delta}{\delta a(j)} (s(n) - \sum_{i=1}^{p} a(i) s(n-i))^2 = 0$$

$$\text{Thus} \quad \mathrm{s(n)s(n\text{-}j)} = \sum_{i=1}^{p} a(i) s(n) s(n-j) \quad \text{for j=1,……..,P}$$

For solving the above equation we have two method one is autocorrelation and other is covariance method.

The autocorrelation solution to the equation can be express as

$$\mathrm{R(j)} = \sum_{i=1}^{p} a(i) R(|i-j|) \quad \text{for j=1,……..,P}$$

2.Parameter transforms: Signal parameters are generated from signal measurements through two fundamental operation: Differentiation , Concatenation. The output of this stage is parameter vector which contains raw estimates of signal.

1.Differentiation: It characterize temporal variation in the signal models. The absolute measurements previously discuss can be though of as a *zero*th order derivatives. In digital signal processing these are several ways in which a first order time derivative can be approximated. Three approximations are:

$$\tilde{s}\ (\mathrm{n}) = \frac{d}{dt} s(n) \approx s(n) - s(n-1)$$

$$\tilde{s}\ (\mathrm{n}) = \frac{d}{dt} s(n) \approx s(n+1) - s(n)$$

$$\tilde{s}\,(\mathrm{n}) = \frac{d}{dt}s(n) \approx \sum_{m=-Nd}^{N_a} ms(n+m)$$

The first two equation are known as backward and forward differences respectively. The first equation same as pre-emphasis filter. The third equation represents a linear phase filter approximation to an ideal differentiator. This is often referred to as regression analysis.

2.Concatenation: Most system post process the measurements in such a way that the operations can be easily explain in terms of linear filter theory. It is generalized in the form of matrix operator. The signal measurement matrix usually contains a mixture of measurements i.e power and set of cepstral coefficient.

3.statistical modeling: In this step consider that the signal parameter were generated from some underlying multivariate random process. To learn or discover the nature of this process, it impose a model on the data, optimize the model and then measure the quality of the approximation. Statistical model of speech recognition system is shown in following figure.

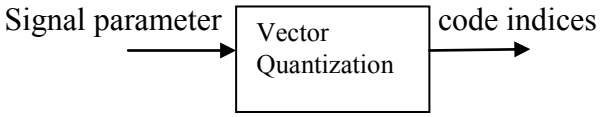

Figure 3. statistical model in speech recognition system

B. Mel frequency cepstral Coefficients (MFCC): MFCC is most evident and popular feature extraction technique for speech recognition. Following figure 3 shows the stepwise process of mfcc feature extraction.

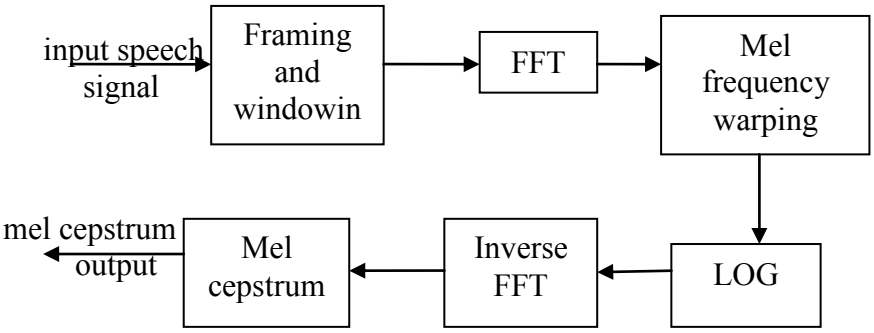

Figure 4. MFCC block diagram

Step-1 Frame blocking: contineous speech signal is devided into frame of N sample. Adjucent frame are being seperated by M(M<N). First frame consist of N sample next frame being M sample it will overlap the first frame by N-M sample.Third frame start with 2M sample after first frame and it will overlap first frame by N-2M and next start 3M and it will overlap N-3M and this procedure is contineous until all input signal

Step2- Windowing: Here each individual frames from step 1 are windowed in order to minimize signal discontinuities and Spectral distortion. The Hamming window is used to decrease the signal to zero at the beginning and end of each
Frame. It is given by
h (n) = 0.54 – 0.46 cos (2πn / N -1), 0≤ n ≤N -1

Step 3 – Fast Fourier Transform (FFT)
Now FFT is used to convert each frame of N samples from the discrete time domain into the frequency domain. Inbuilt MATLAB DFT command is used to obtain DFT.
In that step we can use many operation i.e. DFT,FFT,DWT.

Step 4 – Mel-frequency Wrapping
Here the Mel-frequency Wrapping is used to obtain a mel-scale spectrum of the signal from the step 3.Formula to compute mels for given frequency f:
Mel (f) = 2595 *log(1=f/700)
Mel frequency is linear frequency spacing below 1000 Hz and logarithmic spacing above 1000 Hz. One filter is there for each desired mel frequency component.
Step 5 – Cepstrum: In this final step we use the Discrete Cosine Transform (DCT) to convert the log mel-scale spectrum back to time domain. The result of the conversion is MFCC.
Details of calculating the features based on MFCCs:
Delta and Acceleration Coefficients:
The first order regression coefficients (delta coefficients) are computed by the following regression equation:

$$d_i = \frac{\sum_{n=1}^{N} n(c_{n+i} - c_{n-1})}{2\sum_{n=1}^{N} n^2}$$

Where d_i is the delta coefficient at frame i computed in terms of the corresponding basic coefficients c_{n+1} to c_{n-1}. The same equation is used to compute the acceleration coefficients by replacing the basic coefficients with the delta coefficients.
Cepstral Mean Normalization: This option is aimed at reducing the effect of multiplicative noise on the feature vectors. Mathematically it is:

$$c_i = c_i - \frac{1}{N}\sum_{k=1}^{N} c_{ik}$$

Where c_i is the i^{th} feature element in the feature vector and c_{ik} is the i^{th} feature element at frame k. N is the number of total input frames of data.
Energy and Energy Normalization: Energy is calculated as the log signal energy using the following equation:

$$E = \log \sum_{n=1}^{N} s_n^{\ x}$$

Where s_n is the n^{th} input speech data sample and N is the total number of input samples per frame. The corresponding normalization is to subtract the noise floor from the input data. Note that the silence floor is usually set to 50 dB and the minimum energy value is -1.0e+10.
Liftering: Liftering is applied according to the following equation.

$$c_n = (1 + \frac{N}{2}\sin\frac{\pi n}{N})c_{ns}$$

Pre-emphasis: The first order difference equation: $s'_n = s_n - \alpha s_{n-1}$
is applied a window of input samples.

II. RESULT

In this experiment we have taken 10 samples of speech signals as input. feature extraction is done on this input signal using MFCC techniques. By observing the feature extraction result we found some features. Table 1 shows the some feature selection table using speech sample.

Table 1: Feature selection of speech signal

Features	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	0.002872	0.002289	0.001695	0.002149	0.001901
2	1.90E-05	1.21E-05	7.38E-06	9.28E-06	5.82E-06
3	-2.23E-08	2.15E-08	8.28E-09	7.75E-08	-4.49E-08
4	1.20E-05	6.86E-06	3.31E-06	4.28E-06	1.24E-06
5	2.021471	1.864793	1.713068	1.825034	1.927563

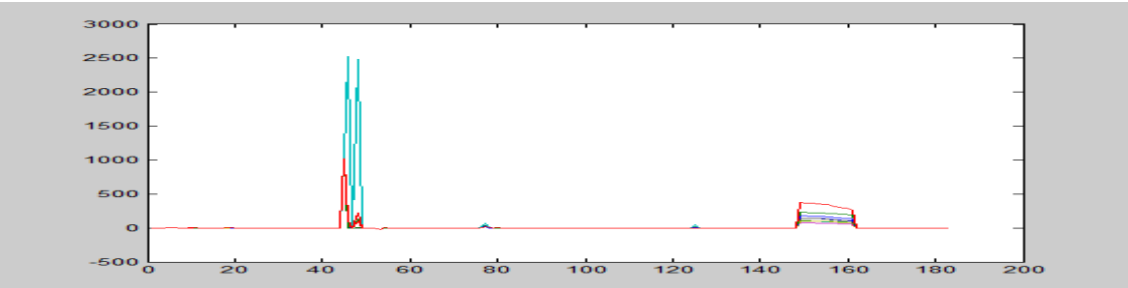

Figure 5.graphical representation of speech feature

III. CONCLUSION AND FUTURE WORK

In this review paper we discussed one of the important part of speech recognition system which is feature extraction and we discuss some steps of feature extraction. By this review we seen that linear predictive coding analysis is used by most of the researcher. Parameter modeling and statistical model also discussed properly.

MFCC is most correct method for speech recognition. we discuss the mfcc block in detail. We observed that some extraction of speech signal and found feature for speech samples also observed the graphical representation of speech features. By using these feature we will going perform the speech recognition system.

REFERENCES

[1] Gerazov, B. and Ivanovski, Z. ``Kernel Power Flow Orientation Coefficients for Noise-Robust Speech Recognition'' Audio, Speech, and Language Processing, IEEE/ACM Transactions on Vol .23 No. 2 FEB 2015.

[2] Wang, K. and An, N. and Li, B. and Zhang, Y. and Li, L. ``Speech Emotion Recognition Using Fourier Parameters'' Affective Computing, IEEE Transactions on 2015.

[3] Sanchez, J. and Saratxaga, I. and Hernaez, I. and Navas, E. and Erro, D. and Raitio, T.`` Towards a Universal Synthetic Speech Spoofing Detection using Phase Information''2015.

[4] H. Wu, Y. Wang, and J. Huang,``identification of electronic disguised voices," ieee transactions on information forensics and security, vol.9,no.3,march 2014.

[5] S.Xue, O.Abdel Hamid,H.jiang, L.dai, and Q.Liu,``fast adaptation of deep neural network based on discriminant codes for speech recognition,"ieee /acm transaction on audio, speech and language processing.vol,22 no.12 December 2014.

[6] J. Grzybowska and M. Kfaczynski, ``computer assisted hfcc-based learning system for people with speech sound disorders.978-1-4799-3700 ieee 2014.

[7] S. Cosentino, T. H. Falk, D. Mcalpine, and T Marquart ``Cochlear implant filter bank design a optimization: simulation study'' ieee/acm transaction on audio, speech and language processing.vol,22 no.2 February 2014.

[8] Z. Ali, M Alsuilaman , G. Mohammad, I. Elamvazuti and T. A. Massallam ``vocal fold disorder detection based on continuous speech by using MFCC and GMM'' ieee gcc conference and exhibition November 17-20,Doha ,Qatar,2013.

[9] A.S and D. P. ``survey about speech recognition and its usage for impaired person,''international journal of scientification engineering research volume 4,issue 2,February 2013 1 ISSN 2013.

[10] S.Sunny, D.P.S,and K.P.Jacob,``performance of different classifier in speech recognition'' IJRET APR 2013.

