ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 1, January 2014

A Proportional Study on Feature Extraction Method in Automatic Speech Recognition System

Dr E.Chandra[1] K.Manikandan[2] M. Sivasankar[3]

Dean, Department of Computer Application, Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, India[1]

Assistant Professor, Department Of Computer Science, P.S.G College of Arts and Science. Coimbatore, India[2]

Research scholar, Department Of Computer Science, P.S.G College of Arts and Science. Coimbatore, India[3]

Abstract: Automatic speech recognition (ASR) has been the focus of many researchers for several years. In speech recognition system is for a computer be able to "hear," understand," and "act upon" spoken information. The speaker recognition system viewed as working in a Analysis , Feature extraction , Modeling , Testing/Matching techniques .speech processing is to convey information about words, speaker identity, tone of voice, expression, style of speech, emotion and the state of health of the speaker. Speech features can be extracted by the parameter as Linear Predictive Codes (LPC), Perceptual Linear Prediction (PLP), Mel Frequency Cepstral Coefficients (MFCC), PLP-Relative Spectra .Neural network training like Vector Quantization (VQ), Dynamic Time Warping (DTW), Support Vector Machine (SVM), and Hidden Markov Model (HMM),Gaussian Mixture Module (GMM) can be used for classification and recognition technique, We have to illustrate some of most important method in neural network like LPC, MFCC, and PLP.

Keywords: LPC, MFCC, PLP, Neural network.

I. INTRODUCTION

Speech is the most important communication method for person interacts with each other. Person also likes to interact with machine via speech. This can be done by developing an Automatic Speech Recognition (ASR) system that is the process of converting a speech signal to a string of words by means of an algorithm applied as a computer program. Automatic Speech Recognition system involves two major phase one for Training phase and another one for Recognition phase. In training phase, means speech is recorded and parametric depiction of the speech is extracted and stored in the speech database [3]. In the recognition phase, for the given input speech signal the features are extracted and the ASR system compares it with the suggestion templates to recognize the speech utterance Speech processing can be processed as three levels for Signal level processing (human auditory system), phoneme level processing(Unit of speech), word level processing (linguistic entity) .

The feature psychoanalysis component of an ASR system plays a essential role in the generally performance of the system. Many feature extraction techniques are available, these include

- LPC-Linear predictive analysis
- LPCC- Linear predictive cepstral coefficients
- PLP-perceptual linear predictive coefficients
- MFCC-Mel-frequency cepstral coefficients
- FFT-Power spectral analysis
- MEL-Mel scale cepstral analysis
- RASTA-Relative spectra filtering of log domain coefficients
- DELTA-First order derivative.

This paper is organized as follows: Section II. Describe Modules of ASR In Section III. Details of the feature extraction techniques like LPC, PLP and MFCC are discussed. That is followed by description of neural network used for speech recognition. Section IV. Conclusions are given based on comparison done on all the three above mentioned methods of speech recognition in last section.

II.MODULES OF ASR

Modules that are identified to develop speech recognition System are [1]

1) Speech Signal acquisition-speech Recording depends on the recording Environment
2) Feature Extraction-providing a compact Representation of the input signal.
3) Acoustic Modeling
4) Language & Lexical Modeling
5) Recognition

III. DETAILS OF THE FEATURE EXTRACTION TECHNIQUES

A) Linear Predictive Coding

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

IJIREEICE

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 1, January 2014

Speech signal processing and linear prediction is often called linear predictive coding (LPC) and can be viewed as a subset of filter theory. Sounds are either voice or unvoiced .In Speech generation, during vowel sound vocal cords vibrations are periodic in time, thus are approximated by an impulse train (pitch). Periodic impulse train and random noise can be measured as excitation source and digital filter as speaking tract. Unvoiced sound excitation is modeled by a White Gaussian Noise source.[3] The speech signal transfer via speech analysis filter to eradicate the redundancy in signal, residual error is generated as an output. Its compare to small number of bit in the original signal. So now, as an alternative of transferring complete signal we can transfer this residual error and speech parameters to generate the original signal. The LPC cepstral coefficients are the features that are extracted from voice signal and these coefficients are used as the input data of Artificial Neural Network [5].

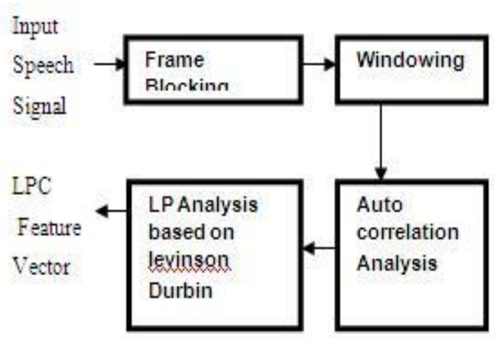

Figure 1: LPC processing diagram

The basic idea is to find set of predictor coefficients that will decrease the mean squared error over a short segment of speech waveform.[6] The resulting parameters are then assumed to be the parameters of the system function in the model for speech production.

B) Mel-Frequency Cepstral Coefficients

The Mel-Frequency cepstral Coefficients is a prevalent method used in a feature extraction techniques. Used in Automatic speech recognition based on frequency domain task. It's more precise than time domain features. MFCC is based on the Short-term Analysis, and so beginning each frame of a MFCC vector is computed. A Mel Frequency filter bank-based method is incorporate personality of the human auditory system. MFCC is broadly used in speech recognition system mainly for three reasons one for) the cepstral features are more or less orthogonal because of the DCT, cepstral mean division second one for eliminates static channel noise, and finally it is a lesser amount of sensitive to additive noise than LPCC. [7]

Methodology:

The presentation of the MFCC may be affected by the quantity of filters, the shape of the way of filters is spaced and then the power spectrum is warped [11]. The segmentation of speech sample is taken as input and then hamming window is applied to decrease the discontinuities of a signal. The filter bank coefficient hamming window length (n) calculating formula is given below:

$$W(n) = 0.54 - 0.46\,cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1$$

N=total number of sample
N=current sample
W (n) = Hamming window

After that Fast Fourier Transformation is considered for each frame to extract the frequency components of a signal in the time-domain. Davis and Mermelstein (D&M) in 1980 [2], describe logarithmic Mel-Scaled filter bank is applied to the Fourier transformed frame used in mfcc by including of non-linear frequency scale used an approximation to the Mel-frequency scale means linear lower 1 KHz and logarithmic higher than 1 KHz. Mel-scale filter bank produce the higher frequency filters have better bandwidth than the lower frequency filters, in same time the temporal resolutions are same. MFCC can be established as by using this formula,

Mel (f) = 2595*log10 (1+f/700) (2)

f- Frequency

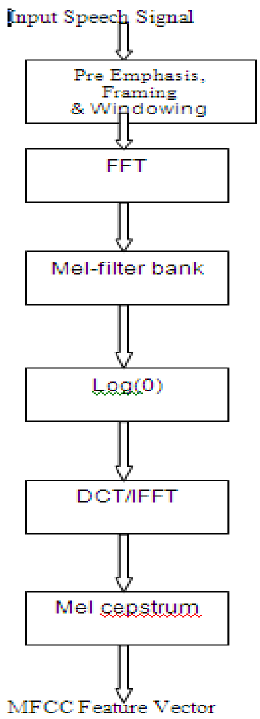

Figure 2: Block Diagram of PLP Processing

The MFCC features are corresponding to the log total energy in a critical band around the center frequency is included. as a final point the Inverse Discrete Fourier Transformer is applied for the cepstral coefficients calculation, finally calculate DCT of the output from the FilterBank.

$$\vec{c}_t[m] = \sum_{n=0}^{M-1} S_t[n] \cos\left(\pi m \left(\frac{n-0.5}{M}\right)\right) \qquad 0 \le m < M.$$

The various MFCC parameters are also included in Feature Extraction technique in ASR system .The following parameters are:

Table 1. Various MFCC parameters

MFCC Parameter	Author	Description
MFCC-Filter Bank-20	Davis and Mermelstein	sampling frequency of 10 kHz; speech bandwidth [0,4600] Hz.
MFCC Filter Bank-24 HTK	Young	Young uses a filter bank of 24 filters for speech bandwidth [0, 8000] Hz
MFCC Filter Bank-40	Slaney	Slaney assumes sampling rate of 16 kHz, and speech bandwidth [133, 6854] Hz.
HFCC-E FB-29	Skowronski	Skowronski and Harris assume sampling rate of 12.5 kHz and speech bandwidth [0, 6250] Hz.

C) Perceptual Linear Prediction

Herman sky was introduced Perceptual Linear Prediction PLP model (1990) .The aim of the PLP model is human vocalizations based on the concept of psychophysics of hearing and then more accurately in the feature extraction process. PLP increase the speech recognition rate and also eliminate the irrelevant information of speech [3].PLP is similar to LPC Analysis so that, it pure LP analysis of speech PLP transform the short-term spectrum of the speech by l psychophysically based transformations.[10] Autocorrelation LP technique is applied to perceptually motivate auditory spectrum by a low order pole function.

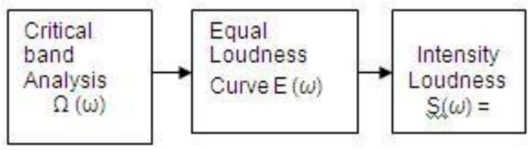

Figure3: Block Diagram of PLP Processing

PLP methods involve two main steps one for approximating the auditory spectrum critical-band filtering, the equal-loudness curve, intensity-loudness power-law relation.[10] PLP parameter computation step as:

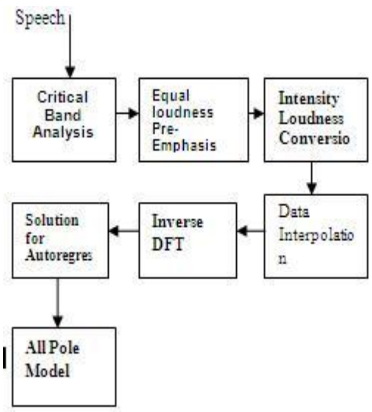

Figure4: PLP computation diagram

The LP all pole mode approximate power distribution equally well at all frequency of the frequency band. Critical band analysis is inconsistent of human hearing [2]. It is more sensitive in the middle frequency range of the audio spectrum. The power spectrum is calculates as given formula as

$$P(\omega) = Re(S(\omega))^2 + Im(S(\omega))^2$$

A frequency warping into the Bark scale is applied. The first step is a conversion from frequency to bark, which is a better representation of the human hearing resolution in frequency. The bark frequency corresponding to an audio frequency is,

$$\Omega(\omega) = 6ln\left[\frac{\omega}{1200\pi} + \left[\left(\frac{\omega}{1200\pi}\right)^2 + 1\right]^{0.5}\right]$$

The auditory warped spectrum is convoluted with the power spectrum of the simulated critical-band masking curve to simulate the critical-band integration of **Ω**. PLP analysis models perceptually forced an auditory spectrum by a low order all pole function, with help of the autocorrelation (LP) technique [4].PLP analysis provide the similar result based on LPC ,and also is order of PLP model is half of LP model . The main difference between PLP and LPC analysis techniques is that the LP model assumes the all-pole transfer function of the vocal tract with a specified number of resonances within the analysis band.

D) Neural network

Artificial neural network provides fantastic imitation of information processing analogues to human nervous system. The common choice of classification and pattern recognition is used as Multilayer Feed Forward Back propagation method in Neural Network [1]. The Audio-Visual Speech Recognizer (AVSR) used is based On HMMs process and was trained for huge vocabulary continuous speech a Neural Network Genetic Algorithm can be used with neural network for performance improvement by optimizing parameter combination.

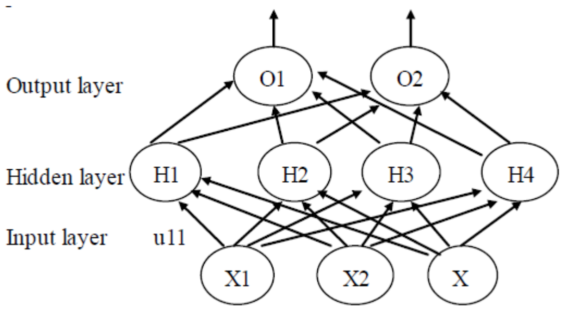

Figure5: Neural Network Basic structure

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 1, January 2014

But in this paper multi-layer feed forward back propagation neural network as shown in Figure 5 with total number of features as number of input neurons in input layer for LPC, PLP and MFCC parameters respectively. As shown in Figure 5 Neural Network consists of input layer, hidden layer and output layer. Variable number of hidden layer neurons can be tested for best results [8] we can train network for different combinations of epochs with target as least amount error rate.

CONCLUSION

In this paper we have discussed some features extraction techniques and their performance. Its need to develop a new hybrid method, so that it will be gives better performance in robust speech recognition. LPC parameter is a linear computation nature but human voice is nonlinear in nature so Linear Predictive Codes are not a good choice for speech estimation. HMM and Neural Network are measured as the most important pattern recognition techniques in the field of speech recognition. MFCC and PLP is the logarithmically spaced filter bank of human auditory system and hence the better Report compare to LPC parameters. In feature we have to develop Fuzzy system with help of feature extraction in speech recognition.

REFERENCES

[1] Urmila Shrawankar and Dr. Vilas Thakare, "Techniques for Feature Extraction In Speech Recognition System: A Comparative Study," (IJCAETS), ISSN 0974-3596, 2010

[2] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition," (ijaret), Volume 1, Issue VI, July 2013

[3] Om Prakash Prabhakar , Navneet Kumar Sahu, " A Survey On: Voice Command Recognition Technique," (ijarcsse)Volume 3, Issue 5, May 2013

[4] Shanthi Therese S., Chelpa Lingam, "Review of Feature Extraction Techniques in Automatic Speech Recognition," (IJSET) Volume No.2, Issue No.6, pp : 479-484

[5] Mark *D. Skowronski and John G. Harris,* "increased mfcc filter bandwidth for noise-robust phoneme Recognition," ICASSP'02, volume No. I, pp 801-4, ISBN: 0-7803-7403-7.

[6] Lee Y. & Hwang K.-W.: "Selecting Good speech Features for Recognition". ETRI Journal, Vol. 18, No. 1, 1996

[7] M. D. Skowronski and J. G. Harris, "Increased mfcc filter bandwidth for noise-robust phoneme recognition," International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 801–4, 2002.

[8] Rabiner, Juang, Fundamentals of Speech Recognition, 1993

[9]Motlíček P.: Feature Extraction in Speech Coding and Recognition, Report, Portland, to research, data, and theory. Belmont, CA: Thomson/Wadsworth, 2003 US, Oregon Graduate Institute of Science and Technology, pp. 1-50, 2002

[10]H. Hermansky. Perceptual linear predictive (PLP) analysis of speech. *Journal of the Acoustical Society ofAmerica*,87(4):1738–1752,1990

[11]Douglas A. Reynolds," overview of automatic speech recognition" JHU 2008 Workshop Summer School

BIOGRAPHIES

Dr.E.Chandra received her B.Sc., from Bharathiar University, Coimbatore in 1992 and received M.Sc., from Avinashilingam University, Coimbatore in 1994. She obtained her M.Phil. In the area of Neural Networks from Bharathiar University, in 1999. She obtained her PhD degree in the area of Speech recognition system from Alagappa University Karikudi in 2007. She has totally 15 yrs of experience in teaching including 6 months in the industry. Presently she is working as Dean, Department of Computer Applications in Dr.SNS Arts and Science, Coimbatore. She has published more than 30 research papers in National, International Journals and Conferences in India and abroad. She has guided more than 20 M.Phil. Research Scholars. Currently 3 M.Phil Scholars and 8 PhD Scholars are working under her guidance. She has delivered lectures to various Colleges. She is a Board of studies member of various Institutions. Her research interest lies in the area of Data Mining, Artificial Intelligence, Neural Networks, Speech Recognition Systems, Fuzzy Logic and Machine Learning Techniques. She is an active and Life member of CSI, Society of Statistics and Computer Applications. Currently she is Management Committee member of CSI Coimbatore Chapter.

K. Manikandan received his Bsc from Bharathidhasan University, trichy in 1998 and received MCA from Bharathiadsan University, Trichy in 2001. He received M.Phil in the area of soft computing in 2004. He has 12 years of experience in teaching. Currently, he is working as Assistant Professor, Department Of Computer Science, Coimbatore and pursuing PhD in Bharathiar University, Coimbatore. Presently he is working Assistant Professor, department of computer science in PSG College of arts and Science. He has presented research papers in National and International Conferences and published a paper in International Journal. His Research Interest is Machine Learning and Artificial Intelligence. He is Life a member of IAENG. He has guided more than 4 M.Phil Research Scholars. Currently 3 M.Phil Scholars are working under his guidance. He has delivered lectures to various Colleges

M. Sivasankar received his integrated post graduate degree in Msc Software Engineering from Anna University, Chennai in 2013. Presently, he is working as a Research Scholar, Department of Computer Science, P.S.G. College of Arts and Science, Coimbatore. He has presented two research papers in National and International Conferences and published a paper in an International Journal. His research interest is speech processing.