FEATURE EXTRACTION TECHNIQUES IN SPEECH RECOGNITION: A REVIEW

P.V. Naresh
Research Scholar, Department of Computer Science & Engineering, Annamalai University

Dr. R. Visalakshi
Assistant Professor, Department of Computer Science & Engineering, Annamalai University

Abstract

A Speech Recognition is a process to enable Computers to identify and respond to human speech sounds. The paper examines feature extraction techniques applied in speech recognition, even existence of many techniques; the accuracy percentage is a key issue in speech recognition. In this article we present some well-known extraction techniques such as LPC, MFCC, RASTA, PCA, LDA, and PLP identify mostly used feature extraction technique in speech recognition process.

Keywords: Speech Recognition, Feature Extraction, LPC, RASTA, MFCC, PCA, LDA PLP

INTRODUCTION

Speech is the mainly common type of communication for individuals and Speech processing is usually one of the research areas in signal processing. Speech recognition is a process by which human language sounds are identified and answered by a computer. Various methods of extraction and pattern matching are used to improve the quality of voice recognition systems. The techniques of features extraction and pattern matching play a significant role in the voice recognition system in order to maximize the speech recognition of different people.

In this survey paper, we focused on some of the techniques for extracting features mainly used in the field of automatic speech recognition. The main purpose of this survey paper is to provide a brief summary of different techniques for extracting features and to provide an inclusive survey of six techniques for extracting features that help researchers in the domain of ASR Automatic Speech Recognition.

I. SPEECH RECOGNITION SYSTEM

Speech recognition system is a voice recognition system consisting of three main components, namely database preparation, feature extraction and classification. as shown in Fig.1 In ASR, firstly raw speech signal is used for database preparation; secondly the acoustic features are computed using different feature extraction techniques. The extraction of robust features gives the recognition performance of ASR. Therefore, extraction of acoustic features should be opted such that these features give best recognition accuracy with minimum computation. Traditional features extraction techniques are linear predictive codes (LPC), Mel Frequency cepstrum analysis (MFFC), Relative Spectral (RASTA), Linear Discriminate Analysis (LDA), Perceptual Linear Predictive (PLP), Principal Component Analysis (PCA), etc. while hybrid features are

the new scope of this field. Finally, these features are applied as an input to the classifier for testing and training so that one can analyze the recognition performance of ASR.

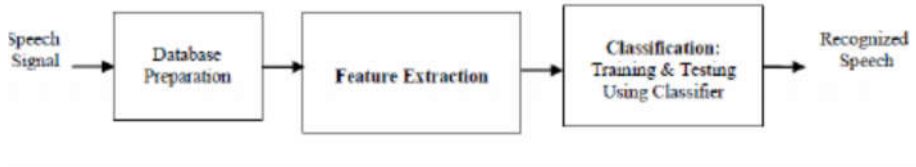

Fig (1) Speech Recognition system

III. FEATURE EXTRACTION TECHNIQUES

Feature extraction acts as an exceptionally important role in speech recognition procedure and as it draws out valuable data from sample speech it is a vital part of research for many years. The key objective of this method is to find out the performance level of different feature extractions techniques and then selecting one of the methods among them. It plays an important role in accurately recognizing speech. The separation of one voice signal from the other, features extraction techniques plays an important role. Since each language has different characteristics implanted in the statement, these characteristics can be extracted from a wide range of feature extractions technique that have been developed and used effectively to recognize speech recognition.

Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) is a different technique that provides a powerful and easy way to extract audio information.The LPC algorithm produces coefficient vectors expressing like a envelope of a breif input signal.The linear prediction method is usually a one of the mainly influential methods for signal analysis. LPC has become the main method of estimating basic voice parameters. It provides an accurate speech parameter estimate and a competent voice calculation model. As a linear combinations of earlier voice samples, the key plan behind LPC is to approximate a voice sample, A single set of parameter or analyst coefficients can be determined over a limited period of time by reducing the addition of square differences among the real voice samples and the predicted values.The predictor coefficients are therefore converted to a more strong set of factors known as Cepstral coefficients.

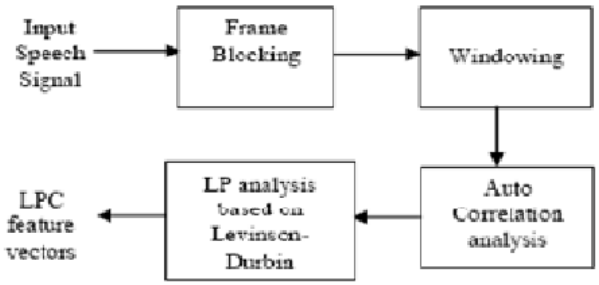

Fig (2) LPC Feature Extraction Technique [1]

Technique	Characteristics	Advantage	Disadvantage
Linear Predictive Coding (LPC)	• Provides auto-regression-based speech features • A Static Technique • The residual sound is very close to the vocal tract input signal	• The advantage of LPC is it has high rate of audio compression • Take short time for training • the redundancy signal could be removed	• Due to its linear calculation nature, LPC could not extract noisy signal at high amplitude. • Take a long time to extract the features

Mel- Frequency Cepstrum (MFCC)

The cepstral coefficients are obtained from speech signals in two stages. The mel-scale filter bank is a technique for spectral estimation. It determines narrowband filter energies. Next, cepstral analysis stage of processing codes the filter energies by using a Fourier transform. A mel-scale filter bank is array of covering triangular filter with center occurrences and bandwidths determined by the Mel-frequency scale. It is based on results from psychophysical learning of humans. MFCC is a renowned technique for speaker recognition which is focused on the speaker recognition vocal tract properties.

$$M = 2595 \log10 \left(1+\frac{f}{700}\right)$$

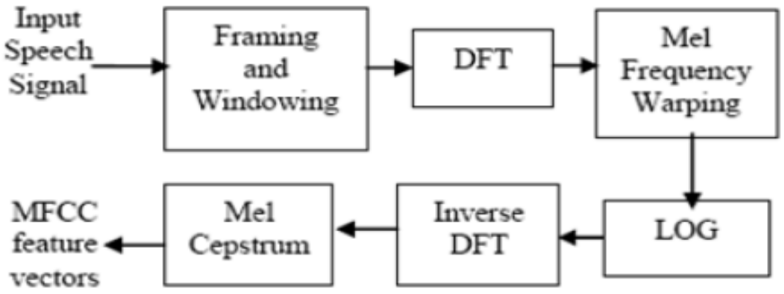

Fig (3) MFCC Extraction Technique [1]

Technique	Characteristics	Advantage	Disadvantage
Mel-Frequency Cepstrum (MFFC)	• Used for speech processing tasks • Mimics the person auditory system.	• The accuracy is high with low complexity • The method is used for find our features • High Performance	• Background noise. • MFCC values are not very robust in the presence of additive noises • Large Computations

		rate	make it difficult to implement

Relative Spectral (RASTA)

The analytical library offers the ability to filter RASTA to reduce for linear distortions of the channel. RASTA filter may be used in log spectral or cepstral areas; Essentially the RASTA channel band passes each element coefficient. Linear channel distortions become visible as an additive constant in both the log spectral and the cepstral domains. The high-pass portion of the band pass filter reduces the effect of convolutionary noise in the channel. The low-pass filtering helps smooth the frame to change the spectrum. In a noisy environment, the RASTA technique is very useful to improve speech quality. The time trajectories in the input voice signal are filtered in RASTA.

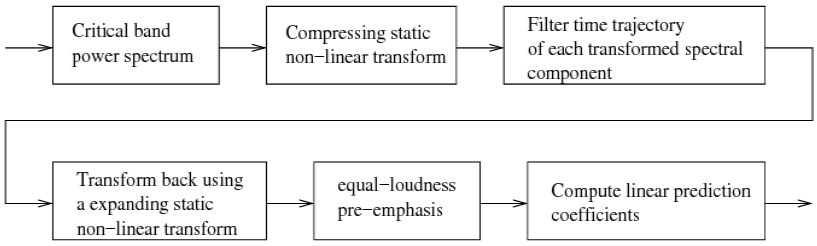

Fig (4) RASTA Extraction Technique

Technique	Characteristics	Advantage	Disadvantage
Relative Spectral (RASTA)	• Designed to reduce noise impact and improve speech. In other words, It is a method widely used for background noise speech signals or just noisy speech. • It is a band – pass filtering technique	• A Robust technique • Removes slow and fast channel distortions	• Need to be used with PLP for better accuracy • Technique cause minor deprivation in performance for clean informaiton.

Principal Component Analysis (PCA)

PCA is thought as a Principle part Analysis – this is often a statistical analytical tool that's used to explore kind and Cluster information. PCA take an oversized variety of

correlate interrelated variables and update this information into a smaller variety of unrelated variables principal components, whereas holding largest quantity of variation, so creating it easier to work the information and build predictions. PCA could be a method of distinguishing information patterns and expressing the information; in this way that it similarities and variations are highlighted. Since a pattern in information is hard to seek out in the information of high dimension, wherever the superior of graphical illustration isn't offered, PCA could be a powerful tool for information analysis

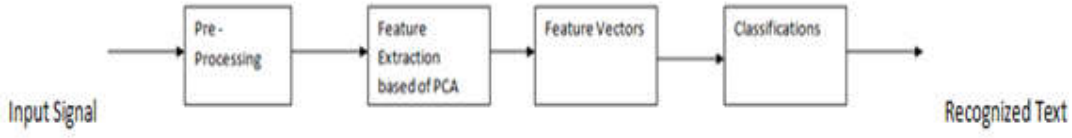

Fig (5) PCA Extraction Technique

Technique	Characteristics	Advantage	Disadvantage
Principal Component Analysis (PCA)	• PCA does not deal with the classification feature • While transformed to different space than the structure and location	• Robust in nature • Retain more significant informaiton and decrease in feature vector's size	• For high dimension data , PCA is expensive

Linear Discriminant Analysis (LDA)

Linear Discrimination Analysis (LDA) is a simplification of the linear discriminant of Fisher's, equation a statistical method for recognition of patterns and machine learning and to find a linear combination of characteristics that Distinguish between two or more object or event classes.LDA is also a linear conversion technique similar to PCA, except that LDA also implicitly finds differences between classes and outside classes. In LDA method, the original feature does not change the location or the structure. LDA works in two steps as shown in Fig. 6.

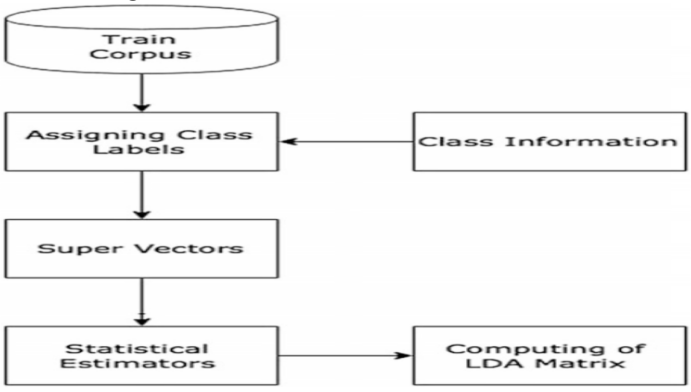

Fig (6) LDA Extraction Technique [22]

Technique	Characteristics	Advantage	Disadvantage
Linear Discriminant Analysis (LDA)	• The location or the structure of the real features does not change • Deals with data classification	• Robust in nature • Within the class, distance is reduced and increase the distance between classes	• Sample distribution is assumed on priority to be Gaussian • It assumes that class samples have equal variance

Perceptual Linear Predictive Cepstrum (PLP)

The Perceptual Linear Predictive (PLP) speech analysis method is based on the short-range spectrum of speech, PLP is a well-accepted acknowledgement of speech and is considered to be started in the smooth spectrum of resonant peaks.PLP parameters are the coefficients that result from standard all-pole modeling which is effective in suppress speaker-specific details of the spectrum. In addition, the PLP order is smaller than is typically needed by LPC-based speech recognition systems. The PLP models human speech on the basis of the concept of hearing psychophysics. The voice spectrum in PLP is changed by a sequence of transformations depend on human auditory system models.The PLP computation steps are critical-band spectral-resolution, The hearing curve of equal-loudness and the hearing power law of intensity-loudness. Once the auditory-like spectrum is estimated, it is converted to autocorrelation values by doing a Fourier transform. The resulting autocorrelations are used as input to a standard linear predictive analysis routine, and its output is perceptually-based linear prediction coefficients.These coefficients are then converted via a standard recursion into cepstral coefficients.

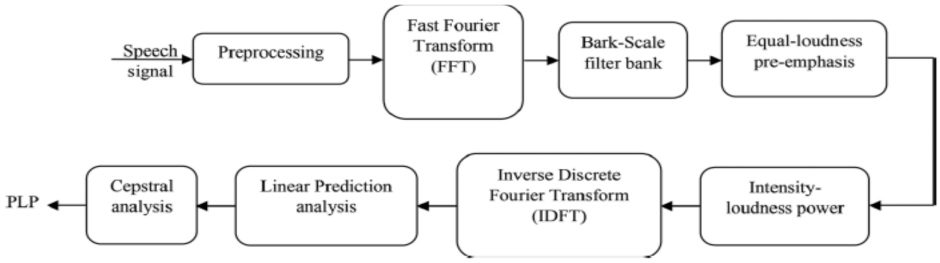

Fig (7) PLP Extraction Technique

Technique	Characteristics	Advantage	Disadvantage
Perceptual Linear Predictive Cepstrum (PLP)	• Similar to LPC except the spectral characteristics • Unwanted information of speech has been discarded	• Reduces the inconsistency between voiced and non voiced signals • It is used in speech signal depend on short term spectrum	• Completely dependent on the spectral balance of formant amplitude • In the spectral balance of the format amplitudes, the result feature vectors are dependent

IV. PERFORMANCE EVALUATION

The speech recognition system performance can be measured in accuracy and speed is two performance measurements most frequently used.

Accuracy

Word Error Rate (WER) is used to assess accuracy, while real time factor is used to measure speed. We can calculate WER by using the equation (1)

$$\text{WER} = \frac{S+D+I}{N}$$

In which,
S: indicate number of substitutions,
D: indicate number of the deletions,
I: indicate number of the insertions and
N: indicate number of words in the reference.

Word Recognition Rate (WRR) is used to report a speech recognition system's performance.

$$\text{WRR} = 1 - \text{WER} = \frac{N-S-D-I}{N} = \frac{H-I}{N}$$

Where H is N-(S+D), the number of properly recognized words

Speed

A parameter to evaluate automatic speech recognition is a real-time factor. If P takes time to process a duration I input, the real-time factor is define as

$$\text{RTF} = \frac{P}{I}$$

V. CONCLUSION

In this article, we summarized some of the extraction techniques used mainly in the area of automatic speech recognition. The major objective of this review paper is to give a brief general idea of different feature extraction techniques. We tried to provide a comprehensive survey of six feature extraction techniques that help automatic speech recognition researchers. Mel Frequency Cepstral Coefficients (MFCC) has also been identified as the widely used extraction technique in the speech recognition system due to the proximity to the actual individual hearing and accuracy

.REFERENCES

[1] Bhuvaneshwari Jolad, Dr. Rajashri Khanai : "Differents Feature Extractions Techniques - Automatic Speech Recognition : A Review," International Journal of Engineering Science and Research Technology, 7(2), February,2018
[2] Divya Gupta, Poonam Bansal and Kavita Choudhary:"The State of the Art of Feature Extraction techniques in speech recognition,"Speech & Language processing for Human machine Communications, Advances in Intelligent Systems and Computing 664.
[3] Lindasalwa Muda, "Voice Recognitions Algorithm using Mel-Frequency Cepstral Coefficients (MFCC) & Dynamic Time Wraping (DTW) Techniques",Journal of Computing, Volume II, Issue III, March 2010
[4] Nidhi Desai, Prof. Kinnal Dhameliya, Prof. Vijayendra Desai : "Features Extraction & Classifications Techniques for speech recognition : A Review," International Journal of Emerging Technology & Advanced Engineering., volume 3, Issue 12, December 2013
[5] Neerja Arora,: "Automatic Speech Recognition System: A Review," International Journal of Computer Applications, volume 151 No.1. October 2016.
[6] Pratik K. Kurzekar, R.Ratnadeep deshmukh, Vishal B Waghmare, P Pukhraj, Shrishrimal, " Comparitive study on feature extraction techniques for speech recognition system."Volume 3, issue 12, December 2014
[7] Swetha khara, shailendra singh:"A Comparative study of the techniques for feature extraction & classification in stuttering , 2nd international conference on inventive communication & computational technologies,2018
[8] S.J.Arora and R.Singh, "Automatic Speech Recognition: A Review, "International Journal of Computer Applications, vol60-No.9, December 2012
[9] Swathy MS, Mahesh KR: "Review of feature extraction & classifications techniques in speaker recognition," International journal Engineerng research & General science, Volume 5, issue 2, March-April 2017
[10]Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh,"A Review on Different Approaches for Speech Recognition System", International Journal of Computer Applications, Volume 115 – No. 22, pp. 23-28, April 2015
[11] Varsha singh, vinay kumar Jain, Dr. Neeta Tripathi: "Comparative study on features extraction technique for language identification," International Journal of engineering research & general science, volume II, issue III May 2014
[12] Wiqas Ghai, Navdeep Sing,: "Literature Reviews on Automatic Speech Recognition," "International Journal of Computer Applications", Volume 41- No.8 March 2012