ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

A Survey on Feature Extraction Algorithm for the Speech Recognition System

Nomita Ningthoujam[1], Prathima V.R[2], Dr K.Fayaz[3]

PG Student [CSE], Dept. of CSE, Rajiv Gandhi Institute of Technology, Bangalore, Karnataka, India[1]

Assistant Professor, Dept. of CSE, Rajiv Gandhi Institute of Technology, Bangalore, Karnataka, India[2]

Professor, S.K. University, Ananthpur, Andhra Pradesh, India[3]

ABSTRACT: Intelligent system for speech recognition has been a high demand in the recent years. The main motive of shifting the research of language processing from text to speech is to introduce an enhanced system for automatic speech recognition. Automatic Speech Recognition (ASR) is an intelligent system which could provide an easy way of communication with the computer and humans. Few algorithms like – Mel Frequency Cepstral Coefficients (MFCC), Perceptual Linear Prediction (PLP), Linear Predictive Codes (LPC) are used for feature extraction in speech processing. The paper aims to gives an overall survey of these feature extraction algorithms used in speech processing based on its framework and a comparison between the three algorithms.

KEYWORDS: Intelligent System, Automatic Speech Recognition, MFCC, PLP, LPC.

I.INTRODUCTION

Language is the first and foremost way of communication. The development of Natural Language Processing (NLP) system is a way in which the world can be brought closer to each other through an efficient communication system. The motive behind NLP is to educate people which are unable to access the latest technology being developed. NLP includes various analysing and computational processes which made the machine to understand easily [3, 5]. In the present scenario, the Natural Language Processing (NLP) is an emerging domain for researchers to explore the regional language to bring it to the global platform.

Regional languages are less explored and spoken as compared to the languages like English, Germany, Korean, etc. This serves the need for the regional languages to be explored so that it can be brought to par to other popularly spoken languages. Language processing of the regional languages is another challenge [4, 5]. This is because the regional languages are less known to people. Processing and analysing them is a big hurdle to those who are not aware of the language. Also, these regional languages are lack in linguistic study. In other words, it meant to say that these languages are not well defined. Most of them are complex and agglutinative word and processing them is a very tedious work. Also, many of the regional languages do not even have a script of their own, so processing the, through text-to-text or text-to-speech is out of question. So, to have a content independent processing system, the NLP has shifted it's researched to processing of the speech through acoustic features of the audio voice rather than its textual content. Some of the feature extraction algorithms used in NLP is - Linear Predictive Codes (LPC), Mel Frequency Cepstrum Coefficients (MFCC) and Perceptual Linear Prediction (PLP).

The outline of the paper is as follows: the first section gives the introduction to Natural Language Processing (NLP); the second section discusses the challenges in processing the regional languages; the third section gives a brief framework of the Feature Extraction algorithms used in NLP; the fourth section gives the comparison between the three algorithms-MFCC, NLP, PLP and the last section gives the conclusion.

II.CHALLENGES IN PROCESSING REGIONAL LANGUAGES

Regional languages are less popular as compared to the widely spoken languages like English, Korean, French, etc. It is constraint by the number of people speaking the language considering the landscape and the diversity [4, 5]. The regional languages are usually less computerized and hence less work is being done. Moreover, all the regional language doesn't have the script. So, developing a content dependent language processing system is a tedious task.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

The regional languages are usually characterized by complex agglutinative word and monosyllabic word. Segmenting the word adds to the complexity of the system. Thus, a content independent system is a must to resolve the above mention issues.

III. FRAMEWORK OF THE FEATURE EXTRACTION ALGORITHM

This section of the paper discusses a brief about the algorithms which areused in feature extraction of the audio speech processing. The algorithm namely - Linear Predictive Codes (LPC), Mel Frequency Cepstrum Coefficients (MFCC) and Perceptual Linear Prediction (PLP) are used to extract the acoustic features from the audio which is used in Automatic Speech Recognition System. The feature extraction algorithm is initially design for the speaker dependent system for speech processing, later it is used as a Speaker Recognition System [5,7].

a. LPC Algorithm

The Linear Predictive Coding (LPC) algorithm is based on linear predictive coefficients derived from the estimated predicted speech signal and the original speech signal [8, 9, 10]. It uses the knowledge of the previous sample signal to predict the consecutive signal coefficients value. The difference between the predicted signal value and the original signal value determine the predicted forward error. So, a prediction error filter can be model as an all zero FIR filter whose z-transform relationship is given by:

$$\mathrm{Fp}(z) = \mathrm{Ap}(z)X(z) \qquad (1)$$

whereF_p is the Forward predicted error, A_p is the Predicted value for the next signal, and X is the knowledge of the previous signal.

i. Sampling and windowing:

Input is divided into n-samples and windowed to remove the discontinuity in frames.

ii. Correlation Analysis:

The predictive coding finds the correlations between the expected next signal and the previous signal history and predicts the next signal value. The Forward Predictive error gives the accuracy of the predicted value by finding the difference between the estimated and the original value.

iii. LPC Analysis and parameter conversion:

The predicted values are applied for Z^{-1} transformation to predict the original signals. The coefficients obtained from the transformation are input as parameters for the spectral identification in the LPC system.

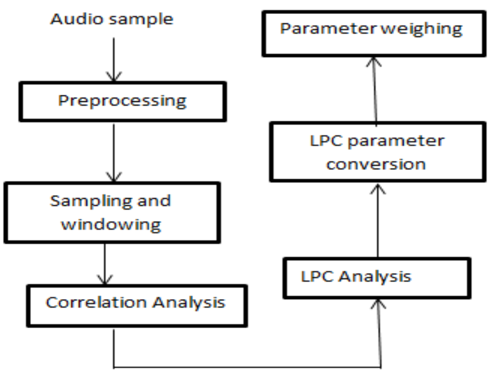

Fig 1: LPC Speech Processing

Fig 1 gives a framework design of the Linear Predictive Codes (LPC) algorithm. The figure depicts the workflow of the LPC algorithm used in feature extraction of the audio voice.

b. MFCC Algorithm

The Mel Frequency Cepstrum Coefficients (MFCC) Algorithm for feature extraction analyses the acoustic features in a speech to determine the Mel coefficients for processing a speech in Automatic Speech Recognition System (ASR) [1, 2, 8, 13]. These Mel Coefficients are the critical parameters for identifying a speech when processing in ASR. Below is the working of the MFCC system.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

i.Sampling and Windowing:

After pre-processing the input audio for any noise removal, the audio voice is divided into frames for further processing. Hamming window is applied to the frames to remove the discontinuity in the frames.

ii.Spectral Analysis:

A Fast Fourier Transform is applied to the N sample frames for estimating the Periodogram power Spectrum of the input speech. The power Spectrum values estimated are taken for further processing.

iii. Calculate Mel Filter bank:

The power spectrum values are converted to the Mel Filter triangular bank by applying the below two equations to the estimated values.

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (2)$$

$$M^{-1}(m) = 700\left(\exp\left(\frac{m}{125}\right) - 1\right) \quad (3)$$

The values received from Mel filter bank are processed for Discrete Fourier transform to convert back to Cepstral coefficients. The coefficients give 26 values where 13 are the Mel coefficients and the remaining are derivatives from the 13 standard vectors.

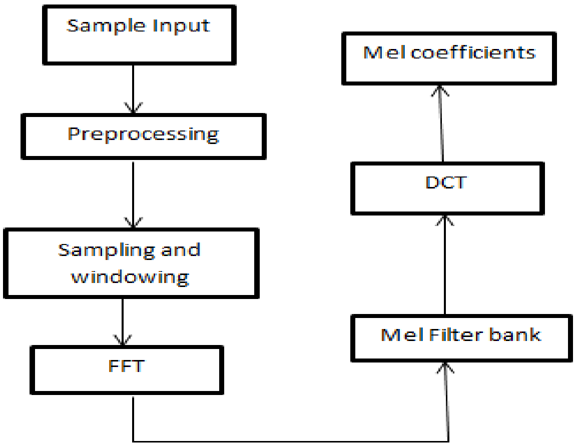

Fig 2: MFCC Feature Extraction

Fig 2 gives a framework design of the Mel Frequency CepstrumCoefficients (MFCC) algorithm. It gives the step by step modelling of the MFCC algorithm for feature extraction.

c. PLP Algorithm

The Perceptual Linear Prediction (PLP) algorithm for feature extraction depends on three factors [8, 11, 12]: i) Critical Band Analysis, ii) the loudness pre-emphasis, and iii) Conversion of intensity loudness.

i. Critical Band Analysis:

The input audio sample is process for spectral analysis after pre-processing for any noise removal. This is to identify the different entities in the audio for further processing.

ii. Loudness pre-emphasis

The loudness pre-emphasis is the process of equalizing the different loudness in the audio voice to a standardized value. This is because the human ear has a perception frequency intensity of 40-db. So, the inequality in the loudness should be standardized so that it can be easily perceived by human ear.

iii. Conversion of intensity loudness

Here, the loudness intensity is processed for amplitude compression for approximation of the power law of hearing. This is required the non-linearity in perception of the sound and the sound intensity.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

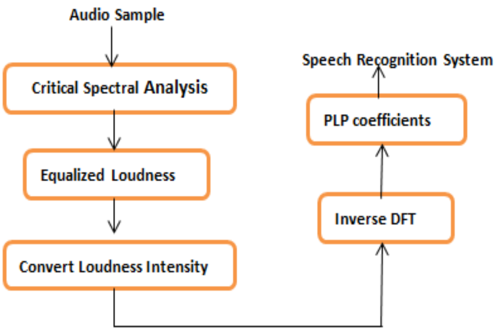

Fig 3: PLP Feature Extraction

Fig 3 gives the modelling design of Perceptual Linear Prediction (PLP) algorithm. The workflow of how feature extraction is carried out in PLP is depicted by the diagram given above.

IV.COMPARISION OF MFCC, LPC, PLP

Table1: Comparison of the three Feature Extraction algorithms used in NLP

Consideration parameters	MFCC	LPC	PLP
Generation of coefficients	Analyse the acoustic features of the sample to generate the coefficients	The consecutive signal is predicted from the previous signal value.	Analyse the acoustic features of the sample to generate the coefficients
No of coefficients	13 feature vectors	10 coefficients	-
Generation of filter bank bands.	Non-linear Frequency is used for calculation of Mel Frequency.	Maximum likelihood of the LPC values is given by Linear combination of the signal value and predicted signal value.	Amplitude compression for approximation of the power law to generate the filter bank.
Spectrum Processing	Short term Spectrum modified based of spectral transformation.	Short term Spectrum transform based on Predictive signals.	Short term Spectrum transformed based on psychological perception.
Accuracy	More accurate.	Less accurate if the next signal is not predicted correctly.	More accurate than LPC.

V. CONCLUSION

In this paper, reviews of the techniques used in speech processing are discussed. These techniques are used in Natural Language processing for designing Automatic Speech Recognition (ASR). The paper also gives a brief framework and working of the Feature Extraction algorithms and how the coefficient parameters are generated for used in the feature identification. Lastly, a comparison is drawn for the three algorithms to help better understanding and differentiation between them.

REFERENCES

[1] Majeed, Sayf A., et al. "Mel Frequency Cepstral Coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study."*Journal of Theoretical & Applied Information Technology* 79.1 (2015).

[2] MATLAB based feature extraction Using MFCC for Automatic speech recognition by Siddhant C. Joshi, Dr. A.N.Cheeran, Engineering and Technology Research (IJSETR), Volume 3, Issue 6, June 2014

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

[3] Harisha, S. B., S. Amarappa, and Dr SV Sathyanarayana. "Automatic Speech Recognition-A Literature Survey on Indian languages and Ground Work for Isolated Kannada Digit Recognition using MFCC and ANN."*International Journal of Electronics and Computer Science Engineering*.

[4] Nongmeikapam, Kishorjit, et al. "Improvement of CRF based Manipuri POS tagger by using Reduplicated MWE (RMWE)." *arXiv preprint* (2012).

[5] Shah, Mohit, et al. "Audio content-based feature extraction algorithms using J-DSP for arts, media and engineering courses." *Frontiers in Education Conference (FIE), 2010 IEEE*. IEEE, 2010.

[6] Improved DTW Speech recognition Algorithm Based on MFCC by Ming-,Ren Li-mian,Li-Xi

[7] Tiwari, Vibha. "MFCC and its applications in speaker recognition."*International Journal on Emerging Technologies* 1.1 (2010): 19-22.

[8] Shrawankar, Urmila, and Vilas M. Thakare. "Techniques for feature extraction in speech recognition system: A comparative study." arXivpreprintar Xiv:1305.1145 (2013).

[9] Atri, Mohamed, FatmaSayadi, and RachedTourki. "Efficient Implementation of LPC Algorithm for Voice Decoding Process." Communications in Information Science and Management Engineering 1.3 (2011).

[10] Atri, Mohamed, et al. "Efficient Hardware/Software Implementation of LPC Algorithm in Speech Coding Applications." (2012).

[11] Alam, Md Jahangir, et al. "Multitaper MFCC and PLP features for speaker verification using i-vectors." Speech communication 55.2 (2013): 237-251.

[12] Hönig, Florian, et al. "Revising Perceptual Linear Prediction (PLP)." INTERSPEECH. 2005

[13] Anusuya, M. A., and S. K. Katti. "Comparison of different speech feature extraction techniques with and without wavelet transform to Kannada speech recognition." International Journal of Computer Applications 26.4 (2011): 19-24.