Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 1; January-March, 2015 pp. 18-20
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/acsit.html

Speech Feature Extraction and Classification Techniques

Kamakshi[1] and Sumanlata Gautam[2]

[1,2]Department of Software Engineering ITM University Gurgaon, India
E-mail: [1]kamakshi338@yahoo.com, [2]sumanlatagautam@itmindia.edu

Abstract—*Using feature extraction we are able to reduce the variability in speech by eliminating the unwanted voices, noise and different sources of speech signal information. Speech signals are highly variable. There are varieties of techniques to extract feature from speech. In this paper we have presented most used and exposed techniques and their benefits and importance.*

1. INTRODUCTION

Speech is most common form by which humans communicate their feelings and necessities. It is one of the most ancient ways to express ourselves. There are two types of speech, voiced speech and unvoiced speech. When there are glottal pulses being created by periodic opening and closing of vocal fold, it is called as voiced speech and when there is continuous air flow pushed by lungs is called as unvoiced speech. Speech signals are of high variability depends on various factors and features of speech, which includes rate of Speaking(words uttered per minute),Contents spoken, Acoustic conditions, Tone, Pitch(frequency of speech) , Accent and Pronunciation.

Speech signals are studied by speech processing, where valid information is extracted about the content and the speaker. Phonemes are the elementary object of a speech signal –the smallest unit of speech sound, Syllable-can be defined as one or more Phoneme, while Word- is composition of one or more syllable.

Using feature extraction we are able to reduce the variability in speech by eliminating the unwanted voices, [1] background noise and many different sources of speech signal information which may occur from multiple speakers. Feature extraction is used in speech processing and speech recognition systems, which find vast scope of utility in security, military, law and medical sciences. Speech processing is basically study and analysis of speech signals and thereby processing them through various methods to extract valid information about the content and the speaker.

But there are several problem which occur in Speech Processing like Acoustic variability, Noise, Different types of microphones used, Speaking variability if the person shouts, whispers, or is suffering from cold, [2] Speaker variability, Linguistic variability when a sentence is pronounced in different ways, Pronunciation of the speaker, Some people tend to speak in a louder tone.

Speech recognition or voice recognition is conversion of speech signal into computer readable format or just Speech text.

Applications of Speech Recognition are that they are used in vehicle navigation systems, human computer

Interaction, Pronunciation evaluation, field of robotics, fields of gaming, [3] Transcription of speech into mobile

Texts, field of disabilities which occur in people.

Some SR systems use “Speech independent speech recognition” whereas in some individual speaker reads a section of text into SR system. These systems are capable of analyzing the person’s voice, fine-tune it to extract the specific features, compare it accurately with the different original speaker’s voice samples to recognize the person. These are called “speaker dependent systems”.

Voice recognition or speaker identification refers to “who” is speaking rather than “what” is being spoken. It can be use for various security purposes such as authentication or verification of the identity of a person. There are 2 types in which voice recognition can be classified. Speaker dependent system is used for dictation software. No specific training is required and the working of software is based on learning unique characteristics of a person’s voice and the other is speaker independent Is commonly found in telephone application, here in this technique user have to read few pages.

For a successful feature extraction we must follow some speaking etiquettes. There should be No Mimicry, Speech signal should be balanced all time, Speech signal should occur normally and naturally, Speech signal should be easy to measure, there should be lesser variation and least amount of noise.

2. TECHNIQUES

There are number of techniques available for speech feature extraction like PLP (Perceptual Linear Predictive Coefficient), LPC (Linear predictive coder analysis),LPCC(Linear Predictive Cepstral Coefficient), MFCC(Mel-Frequency Cepstral Coefficient),FFT (Power Spectral Analysis),MEL (Mel Scale Cepstral Analysis),RASTA (Relative Spectra Filtering of Log Domain Coefficient),DELTA (First order Derivative).

2.1 LPC (Linear Predictive Analysis)

LPC is most profound and powerful method used for encoding quality speech at low bit-rate.LPC is tool used mostly used in audio signal processing and speech processing. Here specific speech samples can be approximated at current time as linear combination of previous speech samples.

LP model is based on production of human speech. Utilizing a conventional filter source model, where glottal vocal tract and the lip radiation transfer function is interpret into one all-pole filter. [4] Its principle is just to minimize sum of the squared differences between speech (original) and estimated speech signal over a finite duration.

LPC is simple to implement & mathematically precise, it is a powerful speech analysis technique .LPC is used in the electronic music field as well, With all such application LPC has a disadvantage of having highly correlated feature components.

Type of LPC filters are Voice excitation LPC, Residual Excitation LPC, Pith Excitation LPC, Multiple Excitation LPC (MPLPC), Regular Pulse Excitation LPC (RPELPC), Coded Excited LPC (CELP).

2.2. MFCC (Mel Frequency Cepstral Coefficient)

This can be considered as of the standard method for feature extraction. Through many years of research and Analysis over recognizer, vast variety of speech signals feature representations have been experimented. MFCC is most popular and accurate amongst all. MFCC works by reducing frequency information of the speech signals into smaller number of coefficients .It is simplified model of the auditory processing of signals which is relatively fast and easy to compute. In the below Fig. MFCC can be considered as its noise sensitivity.

MFCC are commonly derived as [5]:
1. Take Fourier transform of a signal.
2. Powers of the spectrum obtained are mapped above onto the Mel scale by using triangular overlapping windows.
3. Log of the powers is taken for each Mel Frequencies
4. Discrete cosine transform is taken which has the list of Mel Log powers just as of some signal.
5. MFCCs are amplitude of resulting spectrum.

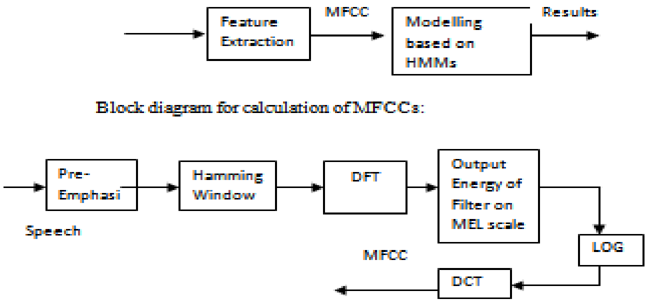

Fig. 1: Block Diagram for Calculation of MFCCs

MFCC provides good discrimination and is very simple, fast, efficient technique in signal processing.

Problems faced in MFCC are it has low robustness to noise, Limited representation of speech signals, its Sensitivity to noise and Information of 2 phonemes instead of 1may occur in a frame in continuous speech environment.

2.3 LPCC (Linear Predictive Cepstral Coefficient)

Linear prediction coding is an alternative method for spectral envelop estimation. This method is also known by names all-pole model or auto-regressive model.

In LPCC the only difference is that MFCC has window but the warping step is deferred

In LPCC, the feature components are de-correlated because of cepstral analysis. It has better robustness in comparison to LPC. In LPCC, Linear scales are not adequate for representation of speech production or perception.

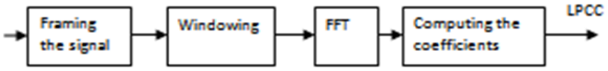

Fig. 2: Computation of LPC Coefficient

2.4 PLP (Perceptual Linear Prediction)

PLP model was developed by Hermensky 1990, the goal of which was to elaborate the understanding of psychophysics of the human auditory and hearing accurately as the features are extracted by extraction process. PLP cepstral coefficient has their computation which uses PLP functions that are already defined in an analysis library. The method mentioned is vulnerable when short-term [6] spectral values have been modified by frequency response of communication channels. Before we compute frame which is based on PLP analysis, we define guidelines which will govern this computation process.

PLP works in similar fashion as that of LPC analysis based on short term spectrum of the speech signals by some transformations based on the psychophysics.

PLP has low-dimensional resultant feature vector. PLP peaks are independent to the length of the vocal tract. Disadvantage of PLP is that due to the communication channel, noise and equipment used causes alteration of the spectral channel.

2.5 FFT (Power Spectral Analysis)

One of the common techniques of carrying out studies on a spectral signal is through power spectrum. The frequency component of signal over time is being described in power spectrum of speech signal.

During the analysis a major question arises that “power of a signal is contained in which specific frequency?” Eureka to answer is the Power Spectra. It is typically in a form of distribution of power value as frequency function, where the average of speech signal is considered as “power”. In frequency domain, it’s the FFT´s magnitude squared.

Power spectra computations can be performed for the entire signal in a single and simple method by averaging together segments of periodgrams of the time signal which gives output to be the "power spectral density"(PSD).

This averaging of long-duration signal periodgrams segments further assigns power to correct and relevant frequencies with utter accuracy and reduces noise fluctuations in power amplitudes.[7] The reduction in frequency resolution is because of lesser data points that are now available for each and every FFT calculation.

Spectral windowing (Windowing of each segment) is most effective method of improving PSD accuracy, but this eliminates contribution of the speech signal near end of the segment. The solution obtained is overlapping of the segments.

First step for computation of power spectrum is to perform DFT(Discrete Fourier Transform) which further computes frequency information of equivalent time domain signal here we use real-points FFT for increased frequency. Both the information regarding magnitude and phase of any of original signal in the time domain are being contained in resulting output.

2.6 MEL (Mel Scale Cepstral Analysis)

MEL is very similar to PLP. In MEL, the spectrum is always warped in accordance to MEL scale, while in PLP it is performed in accordance to the BARK scale coefficients. Mel scale analyses also have an option similar to that of PLP of using a RASTA filter for compensation of linear channel distortions. All pole model is being used in PLP, this helps in smoothening the modified power spectrum whereas cepstral smoothing is used in MEL to smoothen the modified power spectrum.

2.7 RASTA (Relative spectra Filtering)

RASTA filtering is for removing of all distortions. It can be used either with log spectra or cepstral domains. RASTA is a technique in which a band pass filter is applied to the energy in each frequency sub band for smoothening over short term noise variations and also to smoothen spectral changes over frame to frame. [8] For smoothing per frame’s spectral changes low-pass filtering is very helpful. Equivalent to band pass filter’s, high-pass portion performs the task of alleviating effect of convolution noise that is introduced in the channel.

Application of RASTA is that its robust along with that the spectral component that change slower or quicker than the rate of the speech signal are suppressed. Disadvantage is that it gives a poor performance in clean speech environment.

3. ACKNOWLEDGEMENTS

I would like to extend my heartiest thanks with a deep sense of gratitude and respect to all those who provides me immense help and guidance during my project.

I would also like to thank my guide Sumanlata Gautam for providing a vision about the system. I have been greatly benefited from her regular critical reviews and inspiration throughout my work.

I would like to express my sincere thanks to our Head of Department Dr.Latika Singh.

REFERENCES

[1] Automatic Speech Recognition: A Deep Learning Approach By Dong Yu, Li Deng

[2] Multilingual Speech Processing (Recognition and Synthesis)

[3] Understanding Computers in a Changing Society By Deborah Morley

[4] Urmila Shrawankar, Dr. Vikas Thakrey “TECHNIQUES FOR FEATURE EXTRACTION IN SPEECH RECOGNITION SYSTEM: A COMPARATIVE STUDY”

[5] Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013) By Veena S. Chakravarthi, Yasha Jyothi M. Shirur, Rekha Prasad

[6] A Graphical Framework For The Evaluation Of Speaker Verification Systems, Nikolaos Mitianoudis

[7] Http://www.wavemetrics.com/products/igorpro/dataanalysis/signalprocessing/powerspectra.htm

[8] http://www.cslu.ogi.edu/toolkit/old/old/version2.0a/documentation/csluc/node5.html