A Review on Feature Extraction and Noise Reduction Technique

Smita B Magre
Dr. Babasaheb Ambedkar Marathwada University

Pooja Vinod Janse
Dr. Babasaheb Ambedkar Marathwada University

Ratnadeep R. Deshmukh
Dr. Babasaheb Ambedkar Marathwada University

Volume 4, Issue 2, February 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Review on Feature Extraction and Noise Reduction Technique

S. B. Magre
Dept. Of Computer Science and IT
Dr. B. A. M. University
Aurangabad – 431005, India

P. V. Janse
Dept. Of Computer Science and IT
Dr. B. A. M. University
Aurangabad – 431005, India

R. R. Deshmukh
Dept. Of Computer Science and IT
Dr. B. A. M. University
Aurangabad – 431005, India

Abstract— Speech recognition has created nice strides with the event of digital signal process hardware and software package. This paper provides outline various feature extraction and noise reduction technique. This paper helps in selecting the technique beside on their relative advantages & disadvantages. This paper is concludes with the choice on feature direction for developing technique in human pc interface system.

Keywords— DWT, Filters, LDA, LPC, MFCC, Noise Reduction, wavelet.

I. INTRODUCTION

Speech recognition is also known as automatic speech recognition or computer speech recognition which means understanding voice of the computer and performing any required task or the ability to match a voice against a provided or acquired vocabulary. The task is to getting a computer to understand spoken language. By "understand" we mean to react appropriately and convert the input speech into another medium e.g. text. Speech recognition is therefore sometimes referred to as speech-to-text (STT). Speech processing is one of the exciting areas of signal processing. A speech recognition system consists of a microphone, for the person to speak into; speech recognition software; a computer to take and interpret the speech; a good quality soundcard for input and/or output; a proper and good pronunciation.

A. Human Auditory System

To model a human hearing system, it is important to understand the working of human auditory system. At the linguistic level of communication first the idea is formed in the mind of the speaker. The idea is then transformed to words, phrases and sentences according to the grammatical rules of the language. At the physiological level of communication the brain creates electric signals that move along the motor nerves. These electric signals activate muscles in the vocal tract and vocal cords. This vocal tract and vocal cord movements results in pressure changes within the vocal tract and in particular at the lips, initiates a sound wave that propagates in space. Finally at the linguistic level of the listener, the brain performs speech recognition and understanding [1].

B. Speech Sound and its Categorization

Speech signals are composed of a sequence of sounds and the sequence of sounds are produced as a result of acoustical excitation of the vocal tract when air is expelled from the lungs. There are various ways to categorize speech sounds. Speech sounds based on different sources to the vocal tract. Speech sounds generated with a periodic glottal source are termed voiced. Voiced speech is produced when the vocal cords play an active role (i.e. vibrates) in the production of the sound. Examples /a/, /e/, /i/. Likewise, sounds not generated are called unvoiced. Unvoiced sounds are produced when vocal cords are inactive.

C. Topology of Speech Recognition System

- Speaker Dependent: - Systems that require a user to train the system according to his or her voice.
- Speaker Independent: - Systems that do not require a user to train the system i.e. they are developed to operate for any speaker.
- Isolated Word Recognizers: - Accept one word at a time. These recognition systems allow us to speak naturally continuous.

Connected word systems allow speaker to speak slowly and distinctly each word with a short pause i.e. planned speech. Recognition systems allow us to speak spontaneously Spontaneous.

II. FEATURE EXTRACTION

In speech recognition systems, feature extraction and recognition are two important modules. The primary objective of feature extraction is to find robust and discriminative features in the acoustic data. The recognition module uses the speech features and the acoustic models to decode the speech input and produces text results with high accuracy. A number of speech feature extraction methods have been proposed, such as linear predictive cepstral coefficients (LPCCs), Mel-frequency Cepstral coefficients (MFCCs) and perceptual linear predictive coefficients (PLPs) [2].

In temporal analysis the speech waveform itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis [3]. The main goal of feature extraction step is to compute a saving sequence of feature vectors providing a compact representation of the given i/p signal. First of all, recording of various speech samples of each word of the vocabulary is done by different speakers. After the speech samples are collected; they are converted from analog to digital form by sampling at a frequency of 16 kHz. Sampling means recording the speech signals at a regular interval. The collected data is now quantized if required to eliminate noise in speech samples. The collected speech samples are then passed through the feature extraction, feature training & feature testing stages. Feature extraction transforms the incoming sound into an internal representation such that it is possible to reconstruct the original signal from it. There are various techniques to extract features like MFCC, PLP, RAST, LPCC, PCA, LDA, Wavelet, DTW but mostly used is MFCC.

Basic concept of feature extraction is shown in block diagram of Fig. 1.

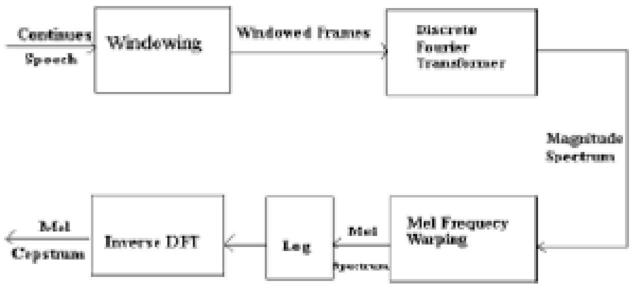

Fig. 1: Feature Extraction Diagram [4].

A. MFCC: Mel-Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCC) is commonly used as feature extraction technique in speech recognition system such as the system which can be automatically recognize which is the task of recognition people from their voice .MFCC are also increasingly finding uses in music information such as gener classification, audio similarity measure etc.Basic concept of MFCC method is shown in block diagram of Fig. 2.

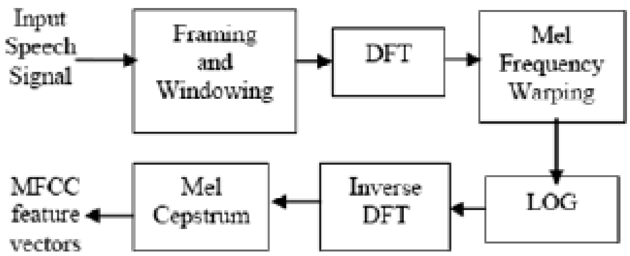

Fig. 2: Steps involved in MFCC Feature Extraction

1) Advantage

As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system.

2) *Disadvantage*

MFCC values are not very robust in the presence of additive noise, and so it is common to normalize their values in speech recognition systems to lessen the influence of noise.

B. Linear Prediction Coefficient

Liner Prediction is a well-known technique used in spectral analysis [4]. LPC (Linear Predictive coding) analyzes the speech signal by estimating the formants, removing speech signal, and estimating the intensity and frequency of the remaining buzz. The process is called inverse filtering, and the remaining signal is called the residue. In LPC system, each expressed as a linear combination of the previous samples. This equation is called a linear called as linear predictive coding. The coefficients of the difference equation characterize the formants. Speech signal recorded using PRAAT and sampled at 16 KHz, is processed features in MATLAB [6]. Basic concept of LPC is shown in block diagram of Fig. 3.

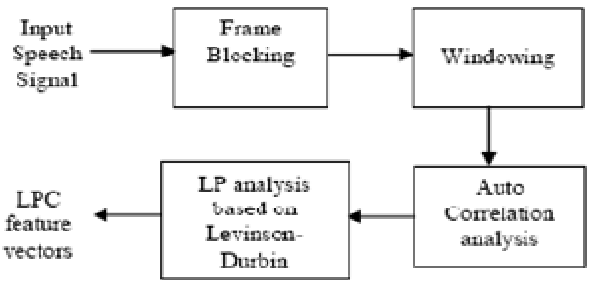

Fig. 3. Block Diagram of LPC.

1) Advantages
   - One of the most powerful signal analysis techniques is the method of linear prediction.
   - LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration.
   - This method provides a robust, reliable, and accurate method for estimating the parameters.
   - LPC provides a good approximation to the vocal tract spectral envelope.

C. Linear Discriminant Analysis(LDA)

LDA are two commonly used techniques for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performance has been observed on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability.

1) Advantages
   - Linear Discriminant Analysis (LDA) is two commonly used techniques for data classification and dimensionality reduction.
   - Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performance has been examined on randomly generated test data.

2) Disadvantages
   - LDA is parametric method since it assumes unimodal Gaussian likehoods.
     - If the distribution are significantly non-gaussian the LDA projection will not be able to preserve any complex structure of the data, which may be needed for classification.

D. Dynamic Time Warping (DWT)

Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW is method that allows a computer to find an optimal match between two given sequence with certain restriction.

1) Advantages
   - Reduced storing space for the reference template.
   - Constraints could be made in finding the optimal path.
   - Increased recognition rate.
   - If the error is too great then it stops the process.
2) Disadvantages
   - To find the best reference template for a certain word. Choosing the appropriate reference template is a difficult task.

E. Wavelet

The Wavelet, with its versatile time-frequency window, is an which tool for the analysis of non-stationary signals like speech fixed have both short high frequency bursts and long quasi-stationary component also.

The wavelet transform provides an improved signal representation with a trade-off between time and frequency resolution. [7] Most of the wavelets used for speech analysis are generic and not particularly designed for the task at hand. Basic concept of wavelet method is shown in block diagram of Fig. 4.

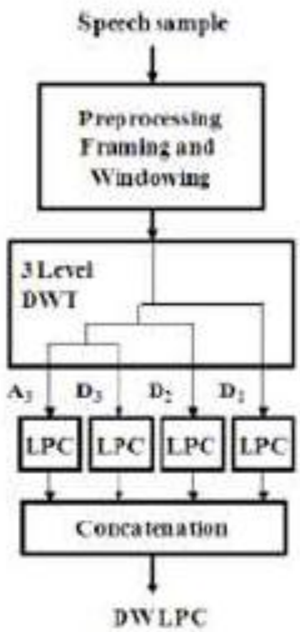

Fig. 4: Block Diagram of Wavelet.

1) Advantage
   - Wavelet transform have been used for speech feature extraction in which the energies of wavelet decomposed sub-bands have been used in place of Mel filtered sub-band energies. Because of its better energy compaction property.
   - Wavelet transform-based features give better recognition accuracy than LPC and MFCC.
   - The WT has a better capability to model the details of unvoiced sound portions.
   - Better time resolution than Fourier Transform.
   - The main advantage of wavelets is that they offer a simultaneous localization in time and frequency domain.
   - Wavelet is that, using fast wavelet transform, it is computationally very fast.
   - Wavelets have the great advantage of being able to separate the fine details in a signal. Very small wavelets can be used to isolate very fine details in a signal, while very large wavelets can identify coarse details.
   - A wavelet transform can be used to decompose a signal into component wavelets.

2) Disadvantages
   - The cost of computing DWT as compare to DCT may be higher.
   - It requires longer compression time.

III. NOISE REDUCTION TECHNIQUES

We are living in a natural environment where noise is inevitable and ubiquitous; speech signals are generally immersed in acoustic ambient noise and can seldom be recorded in pure form. Therefore, it is essential for speech processing and communication systems to apply effective noise reduction techniques in order to extract the desired speech signal from its corrupted observations. Many solutions have been developed to deal with each problem separately. Classically, these solutions have been classified into main areas: speech enhancement and models adaption [8].
Following are the noise reduction techniques:

- Spectral Subtraction.
- Cepstral Mean Subtraction.
- Blind Equalization.
- Adaptive LMS Filtering.
- Adaptive Wiener Filtering.
- Kalman Filtering.

A. Kalman Filtering

Kalman filter normally assumes the process noise and observation noise are both uncorrelated and have normal distributions. This implies that the Kalman based method is best suitable for reduction of white Gaussian noise. The Kalman filter is derived based on two assumptions, linearity and Gaussian noise. The assumption on Gaussian noise holds for many applications and the Kalman filter based method produces high quality enhanced speech signals [9].

Single channel speech enhancement methods, like spectral subtraction and stationary Wiener filtering method are widely applied in practice because their simple hardware structures and efficient algorithms. Spectral subtraction method has drawbacks on "music noise" effects in the enhanced speech. Also, since speech is non-stationary in nature, stationary Wiener filter does not perform very well.

B. Adaptive Wiener Filtering

Wiener filter can be considered as one of the most fundamental noise reduction approaches, which has been delineated in different forms and adopted in various applications [10].

In signal processing, the Wiener filter is a filter used to produce an estimate of a desired or target random process by linear time-invariant filtering an observed noisy process, assuming known stationary signal and noise spectra, and additive noise.

1) Advantages
   - The Wiener filter minimizes the mean square error between the estimated random process and the desired process.
   - It de-noises audio signals, especially speech, as a pre-processor before speech recognition.
2) Disadvantages
   - The estimation criteria are fixed.
   - Simple hardware structures and efficient algorithms.
   - Speech is non-stationary in nature, stationary Wiener filter does not perform very well.

C. Adaptive LMS Filtering

LMS adaptive filtering has become one of the effective and popular approaches for the speech enhancement. An adaptive filter comprises of two basic components [11], these are a digital filter and an adaptive algorithm. The digital filter produces an output in response to an input signal and the adaptive algorithm is responsible for adjusting the coefficients of the digital filter. The adaptive filter is more effective when the carrier frequency of the narrowband interference is offset from the carrier of the spread spectrum signals [12].

1) Advantages
    - It is very effective in rejecting the narrowband interference when the ratio of the narrow band interference bandwidth to the spread spectrum bandwidth is small.
    - It's simple and robustness.
2) Disadvantages
    - High computational complexity.
    - There is no cancellation path transfer function
    - The update of their coefficients is rather heuristic and usually does not come from the minimization of an error.

IV. CONCLUSIONS

We have discussed some features extraction and noise reduction technique and their advantages and disadvantages. Some new methods are developed using combination of more techniques. There is a need to develop new hybrid methods that will give better performance in robust speech recognition area.

ACKNOWLEDGMENT

The authors would like to thank the University Authorities for providing the infrastructure to carry out the research. This work is supported by University Grants Commission.

REFERENCES

[1] S. Therese, C Lingam" Review of Feature Extraction Techniques in Automatic Speech Recognition International Journal of Scientific Engineering and Technology (ISSN: 2277-1581) Volume No.2, Issue No.6, pp : 479-484.
[2] Q. Wu, L. Zhang, and G. Shi, "Robust Multifactor Speech Feature Extraction Based on Gabor Analysis," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 4, May 2011.
[3] M. Kesarkar, P. Rao" Feature Extraction for Speech Recogniton" M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November2003.
[4] S. Gaikwad, B. Gawali, P. Yannawar "A Review on Speech Recognition Technique" International Journal of Computer Applications. Vol. 10, No.3, 2010.
[5] M. Deriche and D. Ning "A Novel Audio Coding Scheme Using Warped Linear Prediction Model and the Discrete Wavelet Transform". IEEE Transactions On Audio, Speech, And Language Processing, Vol. 14, No. 6, November 2006.
[6] L. Deriche and D. Ning "A Novel Audio Coding Scheme Using Warped Linear Prediction Model and the Discrete Wavelet Transform". IEEE Transactions On Audio, Speech, And Language Processing, Vol. 14, No. 6, November 2006.
[7] G. Choueiter and J. Glass" An Implementation of Rational Wavelets And Filter Design For Phonetic Classification" IEEE Transactions On Audio, Speech, And Language Processing, Vol. 15, No. 3, March 2007.
[8] C. Kermorvant "Comparative Study of Noise Reduction Technique" IDIAP Research Report July 1999.
[9] Z. Yu, F. Wu, L. CEN, Z. Gu and Y. Li "Enhance Speech with Temporal Sparse Noise by Robust Kalman Filter" APSIPA ASC 2011
[10] J. Yiteng and S. Doclo, "New Insights into the Noise Reduction Wiener Filter" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, July 2006.
[11] L. Patil, A. Bhalavi, R. Dubey "Adaptive LMS Filtering Approach For Speech Enhancement" Journal Of Information, Knowledge And Research in Electronics And Communication ISSN: 0975 – 6779| Nov 12 To Oct 13 Volume – 02, Issue - 02| Page 884.
[12] W. Milstein" Adaptive LMS Filters For Cellular CDMA Overlay Situations" IEEE Journal On Selected Areas In Communications, 1996, V. 14 N. 8, P. 1548-1559.

