

Feature Extraction Techniques for Speech Recognition: A Review

Kishori R. Ghule, R. R. Deshmukh

Abstract— Speech is the way of communication between the human. It also defined as it is a process which automatically recognizes the spoken words of person based on given speech signal information. It is also known as Automatic Speech Recognition or computer speech recognition and speech to text conversion. Automatic speech recognition is the very interesting area of research and lots of research work has been done by number of researchers. To recognize the speech feature extraction and word recognition these two steps are followed. After feature extraction feature matching is performed for word recognition. This paper describe the different feature extractions techniques like MFCC,LPC,LGCC,DWT etc. section I gives the introduction. Section II explains the types of speech uttered. Different feature extraction techniques are explained in section III. The section IV gives the classification of speech recognition techniques. Conclusion is given in section V and section V gives acknowledgment.

Index Terms—Speech recognition, word recognition, feature extraction, MFCC, LPC, LGCC, DWT, feature matching.



1 INTRODUCTION

SPEECH Recognition is the ability of machine or program to identify words and phrase from spoken language and convert them in to machine readable format. The main intension of speech recognition area is to evolve techniques and system for speech input to machine. Speech is the way of communication in human beings, and it is the dominancy of this medium that motivates research efforts to allow speech to become a viable human computer interaction [1].

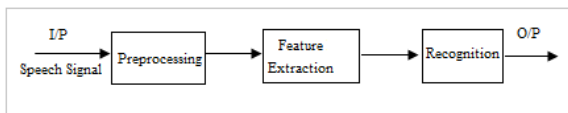


Fig. 1. Speech Recognition Process

There are usually two categories for isolated and continuous speech recognition: speaker-dependent and speaker-independent. Speaker dependent method involves training a system that recognize each of the lexicon words uttered single or multiple times by specific set of speakers , while for speaker independent training methods are generally unworkable and words are recognized by analyzing their inherent acoustical

properties [2].

2 TYPES OF SPEECH UTTERED

The speech recognition system is separated in different classes is based on what type of utterance they have ability to recognize.

i. Isolated speech

Isolated word recognizer requires single utterance at a time. It set necessary condition that each utterance having little or no noise on both sides of sample window.

ii. Connected word

Connected word require minimum pause between utterances to make speech flow smoothly. They are almost similar to isolated words.

iii. Continuous speech

Continuous speech is normal human speech, without silent pauses between words. It is basically computer's dictation. This kind of speech makes machine understanding much more difficult.

iv. Spontaneous speech

Spontaneous speech can be thought of as speech that is natural sounding and no tried out before. An ASR system with spontaneous speech ability should be able to handle a diversity of natural speech features such as words being run at the same time.

3 FEATURE EXTRACTION TECHNIQUES

Feature extraction is most important part of the speech recognition system which distinguishes one speech from another. The goal of feature extraction is to compute saving sequence of feature vectors for providing compact representation of input signal. In three stages feature

- Author name is currently pursuing master degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com
 - Co-Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com
- (This information is optional; change it according to your need.)

extraction is performed. First stage is speech analysis or acoustic front end. This step performs spectro temporal analysis of signal. The second stage compiles extended feature vector composed of static and dynamic features. In the third stage transforms the extended feature vectors into more compact and robust vectors which supplies to the recognizer. [3]

The feature extraction is the process of removing unwanted and redundant information and retains only the useful information in type of speaker independent automatic speech recognition. [4] But in practice this process may lose some important information. [5] The goal of feature extraction is to find out the set of properties called as parameter of utterances by processing of the signal waveform of the utterances. These parameters are the features. After preprocessing the feature extraction is performed. It produce a meaningful representation of speech signal. Feature extraction includes the process of converting speech signals to the digital form and measures important characteristics of signal i.e. energy or frequency and augment these measurement with meaningful derived measurements. [6]

Different feature extraction techniques are LPC, MFCC, LPCC, DWT, WPD, PLP, etc.

3.1 Linear prediction coding (LPC)

LPC is one of the good signal analysis methods for linear prediction in speech recognition process. The feature extraction techniques find out the basic parameters of speech. LPC is the most powerful method for determining the basic parameter and computational model of speech. The idea behind LPC is the Speech sample can be approximated as a linear combination of past speech samples. [7]

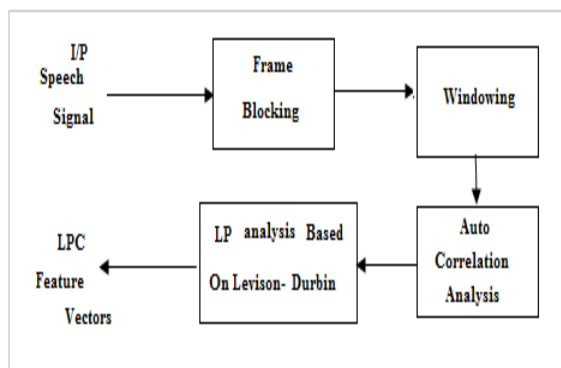


Fig. 2. Feature Extraction by LPC

3.2 Mel frequency Cepstral Coefficient (MFCC)

MFCC is most popular feature extraction technique. Frequency bands are placed logarithmically here so it

approximates the human system response more closely than any other system. Due to its advantage of less complexity in implementation of feature extraction algorithm, only sixteen coefficients of MFCC corresponding to the Mel scale frequencies of speech Cepstrum are extracted from spoken word samples in database.

As shown in below figure 3 the first step is preprocessing in which the signals are preprocessed before feature extraction. In framing the signal are splits in to number of frames in time domain, then on each individual frame the hamming window is applied. DFT is used to convert each frame from time domain to frequency domain.

The filter bank is created by calculating number of picks spaced on Mel-scale and again transforming back to the normal frequency scale. DCT is used to convert mel spectrum coefficient to the time domain. [8]

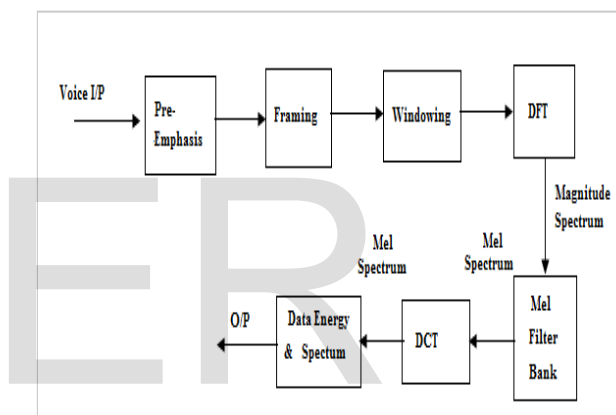


Fig. 3. Block Diagram of MFCC

3.3 Linear prediction cepstral coefficient (LPCC)

The feature extraction is used to demonstrate the speech signals by finite number of measures of the signals. To obtain LPCC coefficients the LPC coding is used. LPCC implemented using autocorrelation method. The main drawback of LPCC is that the LPCC are highly sensitive to quantization noise. [9]

3.4 Discrete Wavelet Transform (DWT)

DWT can be considered as filtering process achieved by a low pass scaling filter and a high pass wavelet filter. The transform decomposition separates the lower frequency contents and higher frequency contents of the original signals. The lower frequency contents provide a sufficient approximation of the signal while the finer details of the variation are contained in the higher frequency contents.

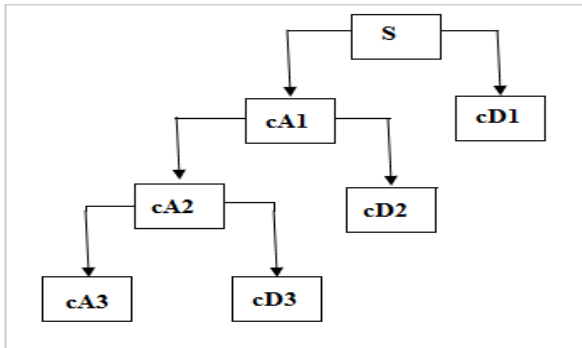


Fig. 4. Decomposition of DWT Coefficient

In the next stage of the decomposition, only the low pass signal is further split into lower and higher frequency contents. The wavelet decomposition can be referred to as a binary tree-like structure, where the left child representing the lower frequency contents. Because of binary tree-like structure the extension is linked to left child. [10][11]

3.5 Wavelet Packet Decomposition(WPD)

The wavelet packet decomposition is the wavelet transform. In WPD the signal is passed through more filters than discrete wavelet transform. Wavelet packets are the linear combination of wavelets. The coefficients in the linear combinations are computed by a recursive algorithm making each newly computed wavelet packet Coefficient sequence the root of its own analysis tree.

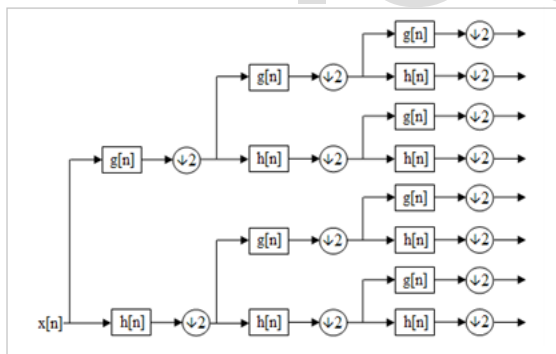


Fig. 5. Wavelet Packet Decomposition Tree

In DWT feature extraction method each level is computed by applying the high pass and low pass filters only on the approximation coefficient. However, in the WPD both the detail and approximation coefficients are decomposed via feature extraction. [12]

In decomposition of WPD, for n level of decomposition it produces 2^n different set of coefficients as shown in the figure. the down sampling process of overall coefficients is still the same and there is no redundancy.

3.6 Perceptual Linear Prediction (PLP)

The Perceptual Linear prediction (PLP) technique is developed by the Hermansky. PLP removes the unwanted information of the speech and thus improves speech recognition rate. PLP is identical to LPC except that its spectral characteristics have been transformed to match characteristics of human auditory system.

4 SPEECH RECOGNITION TECHNIQUES

The speech recognition techniques are classified on the basis of three approaches as follows

- Acoustic Phonetic Approach
- Pattern Recognition Approach
- Artificial Intelligence Approach

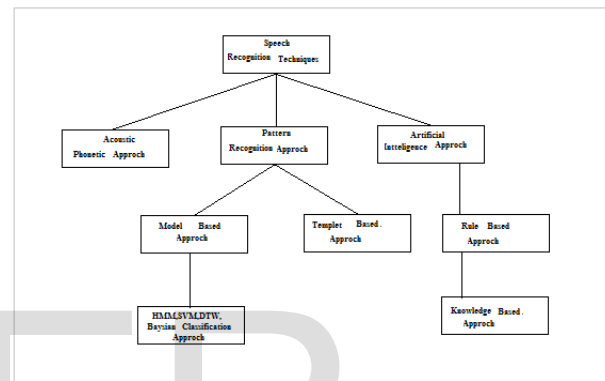


Fig. 6. Speech recognition Technique Classification

4.1 Acoustic phonetic approach

This approach of speech recognition is based on finding speech sound and providing appropriate labels these sounds. The term acoustic is deals with the different sounds in speech and phonetic is the study of phonemes in the language. The basis of acoustic phonetic approach is based on the fact that, there exist finite and exclusive phonemes in spoken language and these phonemes are broadly characterized by a set of acoustic properties that are demonstrated in the speech signal over time.

The acoustic properties of phonetic units are depends on speaker and co articulation effect. Even though the acoustic properties of phonetic units are highly variable, with speakers and with neighboring phonetic units it is also called as co-articulation of sounds. [13]

There are three steps include in to the acoustic phonetic approach to speech recognition are as follow

- 1) Spectral analysis of speech- this step describes the broad acoustic properties of different phonetic units.
- 2) Segmentation and labelling the speech-this step results in a phoneme lattice characterization of the speech

- 3) Determination of string of words- this step determines the string of words from phonetic label sequences brought out by the segmentation to labeling.

This approach has not been most extensively used in most commercial application

4.2 Pattern recognition approach

Pattern Recognition technique is one of the very important and actively searched trait or branch of artificial intelligence. Pattern training and pattern comparison are two steps involves in pattern recognition approach. Using a well formulated mathematical framework and initiates consistent speech pattern representation for pattern comparison, from a set of labeled training samples through formal training algorithm is essential feature of this approach.

In this approach there exist two methods Template base approach and stochastic approach. Stochastic approaches are more suitable approach to speech recognition as it uses probabilistic models to deal with undetermined or incomplete information [14][15].

Stochastic approach having different methods like HMM, SVM, DTW, VQ etc, among all these methods hidden markov model is most popular stochastic approach today.

4.3 Artificial intelligence approach (knowledge based approach)

To study the thought processes of human and deals with representing those processes via machines these two basic ideas involves in Artificial intelligence. Machine behaves like human being so it is called as Artificial intelligence. AI makes machine very smarter and useful. The artificial intelligence approach is the combination of the pattern recognition approach and acoustic phonetic approach so it is called hybrid approach of pattern recognition and acoustic phonetic approach and its recognition procedure is according to how person applies intelligence on set of measured acoustic features. [16]

This approach is not much successful in quantifying skillful knowledge as compare with other two approaches. More reliable method for this type of approach is Artificial Neural Network method. Artificial Neural Network contains large number of simple processing element that is called neurons. These neurons impact each other's performance via a network of excitatory weights.

5 CONCLUSION AND FUTURE WORK

This review paper discussed the speech recognition process which includes the feature extraction and classification techniques. This paper gives the detail study about the feature extraction and different techniques of feature extraction. Every technique has a different recognition rate with different classification method. MFCC is the most popular feature extraction technique and DTW is the more popular classification method. The main aim of this research work is to understand the process of speech recognition to develop ASR system with great accuracy. To develop an ASR system with great accuracy is the future work.

ACKNOWLEDGMENT

The author would like to thank the university authorities for providing the infrastructure to carry out the research. This work is supported by university commission.

REFERENCES

- [1] Smita B. Magre¹, Ratnadeep R. Deshmukh² "Design and Development of Automatic Speech Recognition of Isolated Marathi Words for Agricultural Purpose" Department of Computer Science and IT, Dr. B. A. M. University, Aurangabad - 431004, India, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 16, Issue 3, Ver. VII, May-Jun. 2014
- [2] Bishnu Prasad Das and Ranjan Parekh, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", International Journal of Modern Engineering Research (IJMER), Vol.2, Issue.3, May-June 2012
- [3] W. Ding and G. Marchionini, "A Study on Video Browsing Strategies," Technical Report, University of Maryland at College Park 1997
- [4] Stolcke A., Shriberg E., Ferrer L., Kajarekar S., Sonmez K., Tur G., "Speech Recognition As Feature Extraction For Speaker Recognition" SAFE, Washington D.C., USA pp 11-13, 2007
- [5] Kesarkar M., "Feature Extraction For Speech Recognition" M.Tech. Credit Seminar Report, Electronic Systems Group, EE, Dept, IIT Bombay, 2003
- [6] Bhupinder Singh, Rupinder Kaur, Nidhi Devgun, Ramandeep Kaur "The process of Feature Extraction in Automatic Speech Recognition System for Computer Machine Interaction with Humans: A Review" Dept. of Computer Sc. & Engg., IGCE Abhipur, Mohali (Pb.), India Dept. of Computer Sc. & Engg., Lovely Professional University Jalandhar(Pb.), India, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 2, February 2012
- [7] Leena R Mehta ¹, S.P.Mahajan ², Amol S Dabhade³ "Comparative Study Of MFCC And LPC For Marathi Isolated Word Recognition System" Lecturer, Dept. of ECE,

- CusrowWadia Institute of Technology, Pune, Maharashtra, India 1 Associate Professor, Dept. of ECE, College of Engineering, Pune, Maharashtra, India 2 PG Student [SP], Dept. of ECE, College of Engineering, Pune, Maharashtra, India 3, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 6, June 2013
- [8] Yogesh K. Gedam, Sujata S. Magare, Amrapali C. Dabhade, Ratnadeep R. Deshmukh Dept. of Computer Science and IT, Dr.B.A.M. University, Aurangabad"Development of Automatic Speech Recognition of Marathi Numerals - A Review" International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 9, March 2014
- [9] Nidhi Desai¹, Prof.Kinnal Dhameliya², Prof.Vijayendra Desai³ " Feature Extraction and Classification Techniques for Speech Recognition: A Review" 1M.Tech. [Electronics and Communication] Student, Department Of Electronics and Communication Engineering, C.G.P.I.T, Bardoli, Gujarat, International Journal of Emerging Technology and Advanced Engineering, Website: www.ijetae.com ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 12, December 2013
- [10] Vimal Krishnan V.R "Features of Wavelet Packet Decomposition and Discrete Wavelet Transform for Malayalam Speech Recognition", Babu Anto P School of Information Science and Technology Kannur University, Kerala, India. 670 567, International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009
- [11] Hazrat Ali^{1,2*}, Nasir Ahmad³, Xianwei Zhou², Khalid Iqbal² and Sahibzada Muhammad Ali⁴ "DWT features performance analysis for automatic speech recognition of Urdu" Ali et al. SpringerPlus a SpringerOpen Journal 2014.
- [12] M.A.Anusuya, S.K.Katti, "Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition", International Journal of Computer Science and Information security, Vol.6, No.3, 2010
- [13] Santosh K.Gaikward and Bharti W.Gawali, "A Review on Speech Recognition Technique," International Journal of Computer Applications, vol 10, No.3, November 2010
- [14] Shivanker Dev Dhingra 1, Geeta Nijhawan 2, Poonam Pandit³, "Isolated Speech Recognition Using MFCC And DTW", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 8, August 2013
- [15] M.A.Anusuya, "Speech Recognition by Machine," International Journal of Computer Science and Information security, Vol.6, No.3, 2009
- [16] Nidhi Srivastava and Dr.Harsh Dev"Speech Recognition using MFCC and Neural Networks", International Journal of Modern Engineering Research (IJMER), march 2007