ISSN: 2456-9348
Vol (01) _Issue (05) Impact Factor: 4.520

iJETRM
International Journal of Engineering Technology Research & Management

A STUDY ON DIFFICULTIES AND EXCEPTIONS TO RECOGNIZE SPEECH IN NOISY MEDIUM

Ashok Kumar Shrivastava
Dept of CSE- ASET
Amity University Madhya Pradesh, Gwalior
akshrivastava1@gwa.amity.edu

ABSTRACT
The most important & obvious mode of exchanging information among the human beings is the voice. Human can instruct machine using speech, thus education industry, military and medical sectors, uses this technique. However recognition of speech is not the new area, researchers are engaged for accurateness in voice recognition system, from last few decades. The sketch of that system care considerable challenges like set of speech, mode of speech, word list, transducers, illness and medium; because of all this necessity the component of noise in automatic speech recognition is at a great distant. Many researchers have put their efforts to sort out above challenges. This paper provide brief summary of the latest work in the field of speech recognition and through some light on virtuous and pandemonium databases of pieces of voice.

Keywords:

Pandemonium And Virtuous Database; Feature Extraction; Feature Recognition.

INTRODUCTION
Speech identification is an exercise to recognize said words and translate it into machine legible and comprehensible layout. There are so many application areas, where the science of digital signal processing (DSP) is used to process the speech such as speech signal compression, enhancement, synthesis, and recognition [1, 2]. Speech identification has big role in many applications [4] which is depicted in figure.

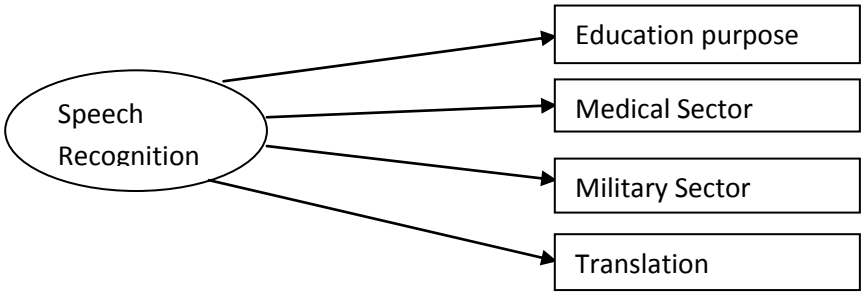


Fig. 1.1 Application areas of Speech recognition

Speech recognition helps people to correct pronunciation of vocabulary while learning foreign language, health care, accomplishment in fighter aircraft, helicopter, military fights administration and conversion of one language into another. Although speech identification has been a well developed technology yet there are some topic to discuss that are responsible in making fewer exact system which are tabulated below.

Table 1.1 Topic responsible for making fewer exact system.

iJETRM
International Journal of Engineering Technology Research & Management

S. No.	Topic	Description
1	Environment	Weather conditions causes change in voice due to presence of noise.
2	Transducer	Responsible in change in range of frequency between caller and receiver, while calling.
3	Channel	In electronic communication distortion echoes because of band amplitude.
4	Vocabulary	Due to typical feature of applicable practice data, particular or common language of person.
5	Speech styles	Due to pitch and pace of voice.
6	Illness	Cough and fever make difference.

Noisy environment is responsible for the above tabulated topics. To work on the tabulated topics, there is a requirement to develop a system which could be able to identify voice exactly in the noisy environment. The whole paper is being segmented into 5 parts, segment 1 having the introduction part, segment 2 is based on review on Automatic Speech Recognition (ASR). Segment 3 reveals associated techniques used in ASR. Section 4 emphasis on work examination. Eventually section 5 giving conclusion remark.

LITRATURE REVIEW

In human communication speech plays an important role and is the best and legitimate form of communication that is why in present era processing of speech has becomes one of the biggest areas of research in signal processing. Although there are major advances in statistical modeling of speech, Today, Automatic Speech Recognition (ASR) has broad range of applications that needs human machine interface. In this section discussed compressed study of considerable memorable part in the research and development of ASR.

It was the digits that pave the way to speech recognition by recognizing the digits. The spoken digits were recognizing by Audrey system in 1952 at bell laboratories [1]. IBM introduces Shoe Box in 1962, which recognizes 16 English worlds. Speech recognition has been expanding day by day and to recognize about a few hundred words. That has the potential to recognize an unlimited number of words because of new statistical method known as the Hidden Markov Model (HMM)[1]. In 1990's the methods for statistical learning of acoustic, language models for stochastic language understanding and the methods for implementation of large vocabulary speech understanding systems was introduces. After that, the computer speech recognition systems was introduced and performed well. To overcome these issues within arbitrary environment they still had a problem with the pitch level i.e. low, high, among similar-voice sample words. In 2008 the key technologies developed have to recognize very large vocabulary for limited task within arbitrary environment. Recent researchers are working on unlimited vocabulary using ASR for unlimited task and for many languages. SantoshK. Gaikwad, et al.[2] discussed the brief overview of all speech recognition techniques. An efficient algorithm for extracting speech was proposed by Wei HAN et al.[3], which gives computation power 53% and accuracy claimed was 92.93%, where FFT filter has been used for enhancement. M.A. Anusuya et al. [4] discussed about all speech recognition techniques and gives better future scopes for implementation of speech recognition system. Mohammad A. M. Abushariah et al. [5], used Hidden Markov Model (HMM) as classifier. MFCC has been used for feature extraction by using English digits database, with recognition rate up to92%.Hui Jiang et al.[6] proposed a method for estimating continues density HMM for ASR by the principle of maximizing the minimum multiclass separation margin and gave significant recognition error rate. Spectral subtraction approach and signal subspace approach had drawback of residual noise therefore, M. A. Abd El-Fattah etal. [7], used adaptive wiener filtering approach which gave better accuracy. Leena R Mehta et al.[8] compared the Mel Frequency Cepstral Coefficients(MFCC) and Linear Predictive coding(LPC) methods .They found MFCC is better than LPC.

The techniques used in speech recognition system are discussed in section 3 according to their working nature.

iJETRM
International Journal of Engineering Technology Research & Management

RESEARCH METHODOLOGY

Speech identification has four essential steps as shown in Fig3.1. In first step voice samples database is created. Voice samples contain noise available in environment. Secondly, removal of noise is performed by doing pre-processing on speech ie. Speech enhancement and normalization is done to remove undesirable noise from voice. In third step, feature extraction and lastly is matching is performed, that based on speech percentage being recognized.

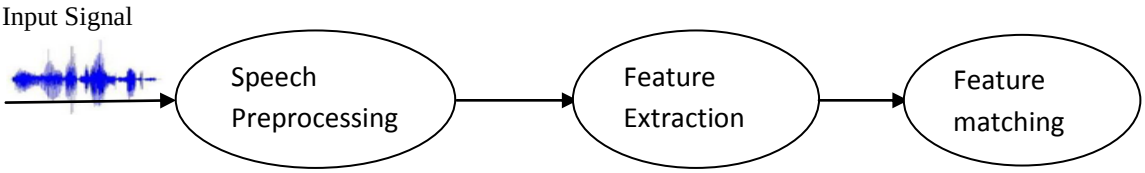


Fig 3.1 System of speech recognition

Each of these steps discussed below.

DATABASE OF VOICE SAMPLE

There is ready to use common database is available for speech recognition system, but according to problem definition's need, every author device their own voice sample database. There are some important points regarding frequency of voice, should be considered at voice recording time. An audio signal has voice frequency as part of it. Range of voice frequency band is relatively 300 Hz to 3400 Hz. The bandwidth of single voice-frequency transmission channel is usually 4 kHz that including guard bands, and sampling frequency is 8 kHz. The fundamental frequency of adult male is 85 to 180 Hz, and adult female is 165 to 255 Hz. There should be proper mix of speech samples of various speakers and from male or female of different age groups. Sample recording should be in blank environment by using specialized Visual- Audio studio. For real environment some samples can be recorded in noisy environment, like faculty class rooms and students' rooms in the university's hostels.

PRE-PROCESSING OF SPEECH

Pre-processing is the task to remove noise from speech signal. Speech is generally a random, so for the same speaker; the same words may have different frequency bands. This is due to the different vibrations in vocal cords. Thus, the shapes of frequency spectrum generated may be different. But, the similarity between these spectrums determines the degree of recognition between the speech signals.

There were many speech pre-processing techniques available to improve the recognition performances. Speech enhancement is one of the most important topics in speech signal processing. Several techniques have been proposed for this purpose like the spectral subtraction approach [7], the signal subspace approach [9] and adaptive wiener filtering approach [10]. The performances of these techniques depend on quality of the processed speech signal. The improvement of the speech Signal-to- Noise (SNR) ratio is the target of most techniques.

FEATURE EXTRACTION

In speech recognition, the main goal of the feature extraction step is to compute a mean sequence of feature vectors which provide a compact representation of the given input signal. Previously the various techniques were used for feature extraction invoice recognition such as Linear discriminant analysis (LDA), Linear Predictive coding (LPC), Mel Frequency Cepstral Coefficients (MFCC).These techniques are briefly discussed next.

Linear discriminant analysis(LDA)

Linear Discriminant Analysis (LDA) is originally used for classification. The LDA has been used to improved recognition performance for small-vocabulary or considerable large vocabulary[12]. The LDA consider two feature

iJETRM
International Journal of Engineering Technology Research & Management

vector X and Y. Find a linear transformation of feature vectors X from an n-dimensional space to vectors Y in an m-dimensional space (m<n)[9].

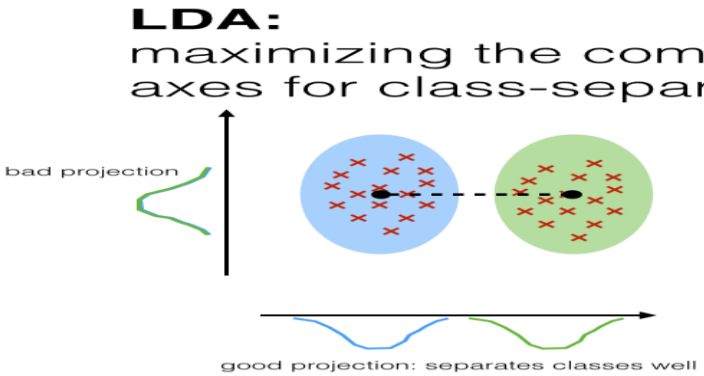


Fig. 3.1 Linear Discriminant Analysis

Linear Predictive coding (LPC):

Linear Predictive coding (LPC) is a computational mathematical operation that analyses the speech signal by estimating the formants which remove their effects from the speech signal and estimate the frequency and intensity. In LPC, each sample of the signal is expressed as a linear combination of the previous signal frames. This is called a linear predictor and hence it is called as linear predictive coding .Fig2 shows the steps involved in LPC.

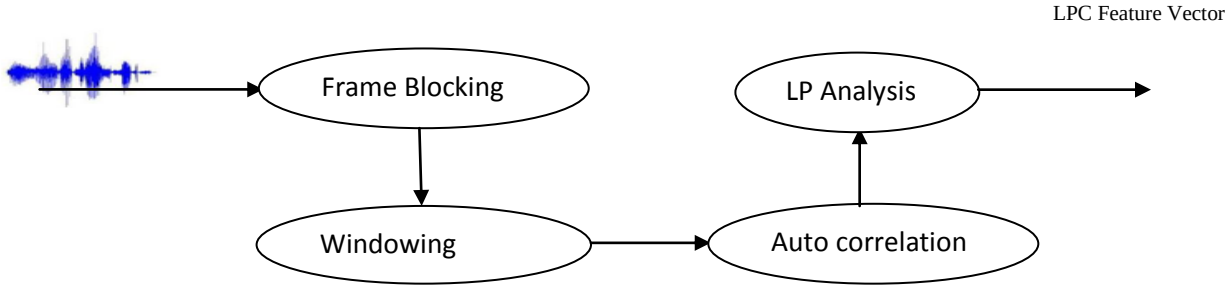
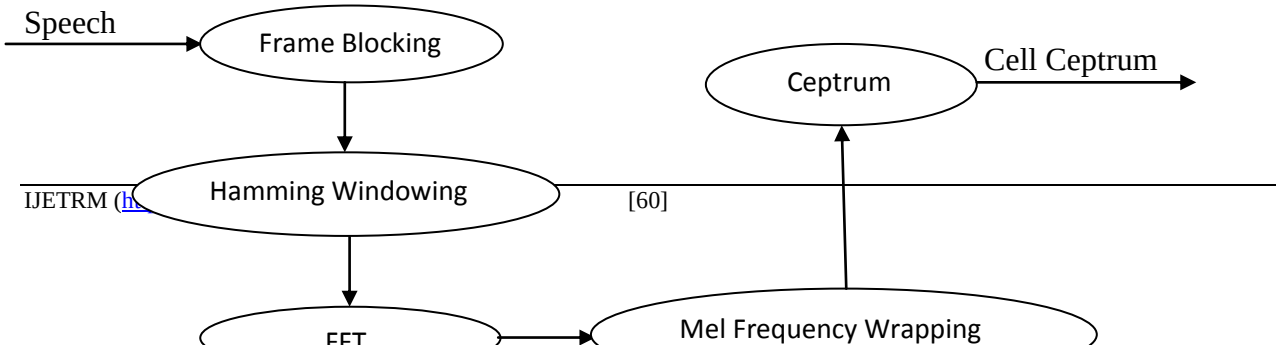


Fig. 3.2 Steps Involved in LPC

Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstral Coefficient (MFCC) s is one of the most accurate feature extraction method used in automatic speech recognition. Fig 3 shows the steps of MFCC.



iJETRM
International Journal of Engineering Technology Research & Management

Mel Ceptrum

Fig. 3.3 Steps Involved in MFCC

Frame Blocking

It is called frame blocking because it is literally using a single frame of the clip to block the previous shot and reveal the next one.

Hamming Windowing

Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame (to be detailed in the next step)

FFT

It converts each frame of N samples in time domain to frequency domain.

Mel-Frequency Wrapping:

It convert the frequency spectrum to Mel spectrum.

Cepstrum

It convert the log Mel spectrum back to time.

FEATURE RECOGNITION TECHNIQUE:

There were two methods used for Feature recognition, are briefly discussed in this paper.

Hidden Markov Model (HMM)

Hidden Markov Model (HMM) is a statistical Markov modelin which the system being modeled is assumed to be aMarkov process with unobserved (i.e. hidden) states. Thehidden markov model can be represented as the simplest dynamic Bayesian network. The mathematics behind theHMM were developed by L. E.

Vector Quantization (VQ)

Vector quantization (VQ) is a classical quantization technique from signal processing that allows the modeling of probability density functions by the distribution of prototype vectors. It was originally used for data compression. It works by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms.

***Table1: Comparison Of Recognition Rates*.**

Author	Techniques of Feature Extraction	Techniques of Feature Classification	Rate of Recognition
B. Milner	Mel Frequency Cepstral	Vector Quantization	88.88%

iJETRM
International Journal of Engineering Technology Research & Management

	Coefficients		
S.K.Podder	Linear Predictive coding	Vector Quantization and Hidden Markov Model	62% to 96%
S.M. Ahadi	Mel Frequency Cepstral Coefficients (Clean) Mel Frequency Cepstral Coefficients (Noisy)	Hidden Markov Model (Clean) Hidden Markov Model (Noisy)	86% 28% to 78%

CONCLUSION

In this review paper different techniques of speech recognition are discussed. Performance of the ASR system based on the feature extraction technique and their accuracy is compared. In coming years, the large vocabulary speaker independent continuous speech has gained more importance. Based on this review, the advantage of MFCC features is more suitable which reduces the complexity of the calculation and offers good recognition result. It also achiever education in time consumption.

REFERENCES

[1] Reddy, D.Raj. "Speech Recognition by Machine: A Review"Proceedings of the IEEE, vol. 64, no. 4, pp:501-531, April 1976.
[2] Santosh Gaikwad, Bharti Gawali, PravinYannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications, vol. 10, no.3, pp"16-24, Aurangabad, November 2010.
[3] Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong- Pang PUN, "An Efficient MFCC Extraction Method in Speech Recognition", In Circuits and Systems, (ISCAS) Proceedings. IEEE International Symposium on pp: 4, May 2006.
[4] M.A.Anusuya, S.K.Katti,"Speech Recognition by Machine: A Review"International Journal of Computer Science and Information Security, vol. 6, no. 3, 2009.
[5] Mohammad A. M. Abushariah, "English Digits Speech Recognition System Based on Hidden Markov Models", IEEE International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia,11-13May 2010.
[6] Hui Jiang, Xinwei Li, and Chaojun Liu, "Large Margin Hidden Markov Models for Speech Recognition", IEEE transactions on Audio, speech, and language processing, vol. 14, no. 5, September, pp:1584-1595, 2006.
[7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust., Speech, Signal Processing,vol 27,no 2,pp. 113-120, 1979.
[8] Leena R Mehta, S.P.Mahajan, Amol S. Dabhade, "Comparative study of MFCC and LPC for Marathi Isolated Word Recognition system", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering ,vol. 2, no. 6,pp:2133- 2139, June 2013.
[9] Y. Ephriam and H. L. Van Trees," A signal subspace approach for speech enhancement", in Proc. International Conference on Acoustic, Speech and Signal Processing, vol.2, pp. 355-358, Detroit, MI, U.S.A, May 1993.
[10] M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd Elsamie, "Adaptive wiener Filtering Approach for speech Enhancement" ,Ubiquitous Computing and Communication Journal, vol 3,no 2.pp;23-31,2010.
[11] A. Rezayee and S. Gazor," An adaptive KLT approach for speech enhancement", IEEE Trans. Speech Audio Processing, vol. 9, pp. 87- 95 February. 2001.

iJETRM
International Journal of Engineering Technology Research & Management

[12] R. Haeb-Umbach, H. Ney "Linear discriminant analysis for improved large vocabulary continuous speech recognition"Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on. vol. 1. IEEE, 1992.
[13] Ujwalla Gawande, "An efficient iris recognition system based on Efficient Multialgorothmic Fusion technique", IJCA proceeding on international conference and workshop of emerging tread in technology (ICWET), no 13, 2011.
[14] B. Milner, "A Comparison of Front-End Configurations for Robust Speech Recognition". ICASSP, vol. 1, IEEE 2002.