Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

An Empirical Study on Feature Extraction Methods for Speech Recognition

Easwari.N[1], Dr.P.Ponmuthuramalingam[2]
[1,2]*PG & Research Department of Computer Science, Government Arts College*

Abstract - Speech Recognition is a main idea of human machine interaction which led to this research. Speech recognition refers to the ability of listening spoken words and identifying various sounds present in it, and recognizing them as words of some known language. Speech recognition aims to extract the lexical information from the speech signal independently of the speaker by reducing the inter speaker variability. It is concerned with extracting the identity of the person. This paper describes the technical overview of the feature extraction in Speech Recognition such as their property, advantages and disadvantages. The most important part of the speech recognition system which distinguishes one speech from another is Feature Extraction. Brief notes on Automatic Speech Recognition, Methodologies of Speech Recognition and Word Separation are also discussed.

Keywords – phonemes; analysis; modeling; matching; acoustic; word separation.

I. INTRODUCTION

Speech is a natural form of human communication. Speech sounds have a rich and multi-layered temporal-spectral variation that convey words, intention, expression, intonation, accent, speaker identity, gender, age, style of speaking, state of health of the speaker and emotion. Speech is also a sequence of elementary acoustic sounds or symbols known as phonemes that convey the spoken form of a language. About 40-60 phonemes in the English language from which a very large number of spoken words can be constructed. Speech signals convey much more than spoken words. The information conveyed by speech is multi-layered and includes time, frequency, and modulation of information as formants and pitch. Formants are the resonances of vocal tract and pitch is the sensation of the fundamental frequency of the opening and closings of the glottal folds [1] [2].

II. AUTOMATIC SPEECH RECOGNITION

An efficient Automatic Speech Recognition system has the major considerations like developing higher recognition accuracy, achieving low word error rate and addressing the issues of variability in the source.Automatic Speech Recognition is the process by which a computer maps an acoustic speech signals to text .The process of converting a speech signals into a sequence of words, by means of an algorithm implemented as a computer program. There are two phases in automatic speech recognition training phase and recognition phase [3].

Figure 1. Automatic Speech Recognition

2.1. The Hierarchy of Speech Recognition

Speech recognition has a hierarchy of classes which is based on the type of utterance which they have ability to recognize. They are

- Isolated Word Recognition
- Connected Word Recognition
- Continuous or fluent Speech Recognition
- Speech Understanding System
- Spontaneous Conversation System

III. SPEECH RECOGNITION METHODOLOGIES

Speech Recognition Methodologies have four different stages. Each stage of the methodologies deals with various analyses of speech signals, extracting algorithms, identification of signal and related word matching. Each stage deals with various segments and its algorithm which focuses highly on the accurate results. The following figure explains the methodologies of Speech Recognition.

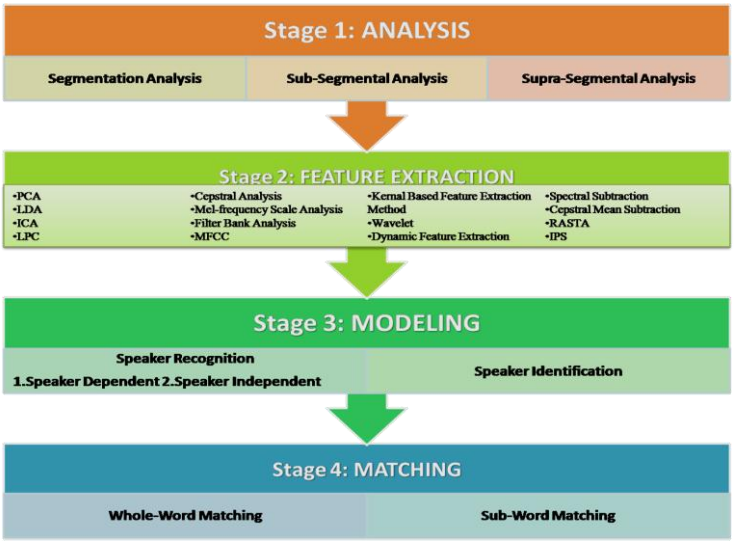

Figure 2. Speech Recognition Methodologies

3.1. Analysis

Speech analysis stage deals with the selection of suitable frame size. Speech Analysis can be further classified into three analyses [4]:

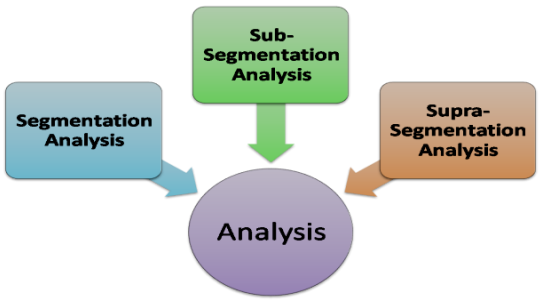

Figure 3. Speech Analysis

- **Segmentation Analysis:** In segmentation analysis, the testing to extort the information of speaker is done by utilizing the frame size as well as the shift which is in between 10 to 30 milliseconds (ms).
- **Sub-Segmental Analysis:** In this analysis technique, the testing to extract the information of speaker is done by utilizing the frame size as well as the shift which is in between 3 to 5 milliseconds (ms). The features of excitation state are analyzed and extracted by using this technique.
- **Supra-Segmental Analysis:** In Supra-segmental analysis, the analysis to extract the behavior features of the speaker is done by utilizing the frame size as well as the shift size that ranges in between 50 to 200 milliseconds.

3.2. Feature Extraction

This is the most important part of the speech recognition system which distinguishes one speech from another. 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 produces a meaningful representation of speech signal. Feature extraction is performed in three stages.

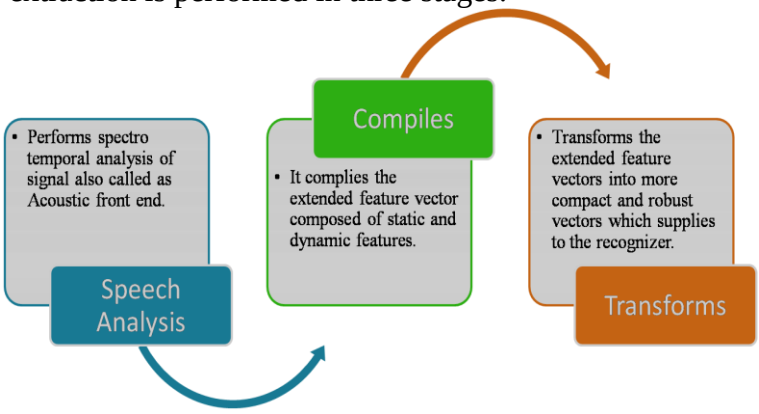

Figure 4. Stages of Feature Extraction

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. 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 measurements with meaningful derived measurements [5].

- PCA
- LDA
- ICA
- LPC
- Cepstral Analysis
- Mel-frequency Scale Analysis
- Filter Bank Analysis
- MFCC
- Kernal Based Feature Extraction Method
- Wavelet
- Dynamic Feature Extraction
- Spectral Subtraction
- Cepstral Mean Subtraction
- RASTA
- IPS

Figure 5.Different types of Feature Extraction Analysis

These are the most commonly used techniques in many applications for feature extraction especially in speaker recognition, speech recognition, biometric systems etc. The following table provides a brief overview of the above analysis.

Sr. No	Feature Extraction Techniques	Characteristics	Advantages	Disadvantages
1	Principal Component Analysis(PCA)	Non linear feature extraction method, Linear map, fast and tradition.[3]	Reduction techniques are highly dimensional	No proper model , limited by particular derivation .
2	Linear Discriminate Analysis (LDA)	Non linear feature extraction method, Supervised linear map; fast.[3]	Highly accurate recognition and flexible.	Based only on assumption.
3	Independent Component Analysis (ICA)	Non linear feature extraction method, Linear map, iterative non- Gaussian[3]	Easy to use, simple and consume little amount of memory space	Performance done one by one element only.
4	Linear Predictive Coding (LPC)	Static feature extraction method,10 to 16 lower order coefficient,	Low bit rate for encoding.	Does not represent the desired spectral information to be modeled.
5	Cepstral Analysis	Static feature extraction method, Power spectrum	Provides a best methodology for separating the features from their vocal tract.	Limited for formant localization especially at high frequency.
6	Mel - Frequency Scale Analysis	Static feature extraction method, Spectral analysis	Low complexity, high accuracy and high performance rate.	False accuracy in background noise and performance may affected by various number of filters.
7	Filter Bank Analysis	Filters tuned required frequencies	Offers perfect reconstruction , directional selectivity and efficient structure.	Increased aliasing and phase distortion. High algorithmic complexity and Signal delay.
8	Mel - Frequency Spectrum	Power spectrum is computed by performing Fourier Analysis	Best method to find features	Time to time certain dropouts in accuracy.
9	Kernel Based Feature Extraction Method	Non linear transformations , optimization done separately.	Better classification in dimensionality reduction and improved in classification error.	Combined optimization provides better results.
10	Wavelet	Better time resolution than Fourier Transform	Fast computation and simultaneous localization in time and frequency domain.	Requires long compression time.
11	Dynamic Feature Extraction i)LPC ii)MFCCS	Acceleration and delta coefficients i.e. II and III order derivatives of normal LPC and MFCCs coefficients . It is used by dynamic or runtime Feature[3]	High performance rate and high accuracy than others.	Lack of robustness.
12	Spectral Subtraction	Robust Feature extraction method , based on Spectrogram	Provides a compact representation of vocal tract system information.	Limitations occur in removal of artifacts.
13	Cepstral Mean Subtraction	Robust Feature extraction	Low complexity, high accuracy and high performance rate.	Limited for Non-linear noise effects.
14	RASTA Filtering	For Noisy speech	Removes slow variation, does not depends on microphone.	Loss of clean information performance and cut off only minimum error.
15	Integrated Phone Subspace Method (PCA+LDA+ICA)	A transformation based on PCA+LDA+ICA	High accuracy , high reduction techniques , flexible, consumption of low memory space and	Accurate acoustic modeling method for components independently along with nonlinear transformation

			ease of use.	method.

Table 1. An Overview on Feature Extraction Analysis

3.3. Modeling Technique

To produce the speaker model is the main goal of modeling technique by the use of extracted features.

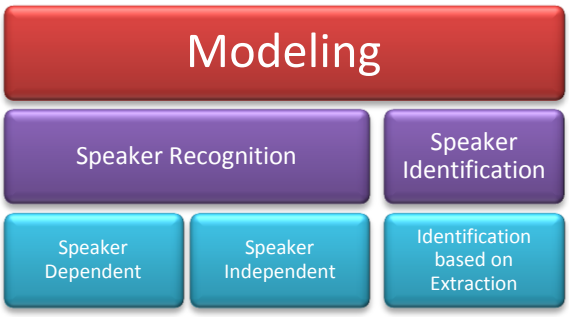

Figure 6 . Modeling Techniques

In speech recognition process the following modeling approaches can be used.

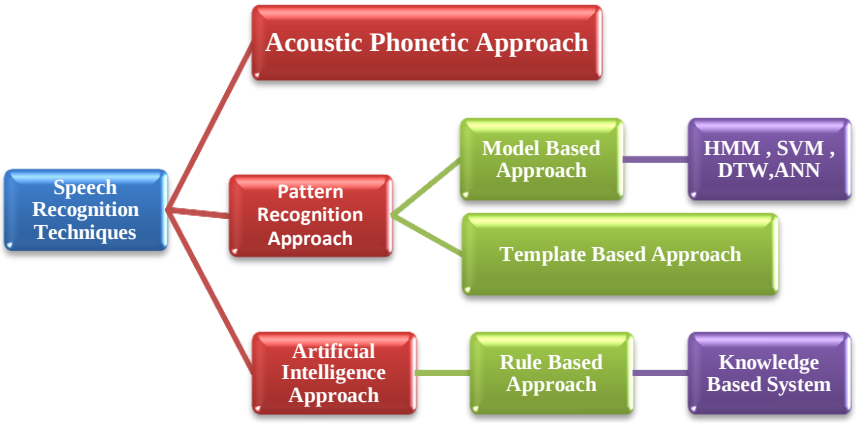

Figure 7. Approaches of Speech Recognition

3.3.1. Acoustic Phonetic Approach

Acoustic means different sounds in speech whereas Phonetic means 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. Also highly variable phonetic units of acoustic properties, with speakers and with neighboring phonetic units it is also called as co-articulation of sounds [3].

Three steps in the acoustic phonetic approach to speech recognition

- Spectral analysis of speech - describes about the broad acoustic properties of different phonetic units.
- Segmentation and labeling the speech - results in a phoneme lattice characterization of the speech.

- Determination of string of words - the string of words from phonetic label sequences for segmentation to labeling.

3.3.2. Pattern Recognition Approach

Pattern Recognition technique is searched trait or branch of artificial intelligence. Pattern training and Pattern comparison are two steps involves in Pattern Recognition Approach. The essential feature of this approach is using a well formulated mathematical framework along with initiates' consistent speech pattern representation for pattern comparison, from a set of labeled training samples through formal training algorithm [6] [7].
There are two methods Model Based approach or stochastic approach and Template Based approach.

- **Model Based Approach** compares of different methods like HMM, SVM, DTW, VQ etc, among all these methods hidden markov model is most popular stochastic approach today.
- **Template Based Approach** a collection of prototypical speech patterns those are stored as reference patterns representing the dictionary of candidate words. Recognition is then carried out by matching an unknown spoken utterance with each of these reference templates and selecting the category of the best matching pattern.[8]

3.3.3. Knowledge Based Approach (Artificial Intelligence Approach)

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. 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 [10].

These are the different types of modeling technique which is used to produce the speaker models for the extracted features.

3.4. Matching Technique

All speech recognition involves detecting and recognizing words. Detecting and Recognition of recorded words is the major functionality of Speech Recognition. The characterizations of the Speech Recognition engines are referred by four different tasks. They are Matching Technique, Speaker Dependence, Size of the vocabulary and the Word Separation. Most speech recognition engines can be categorized by how they perform these basic tasks [9]

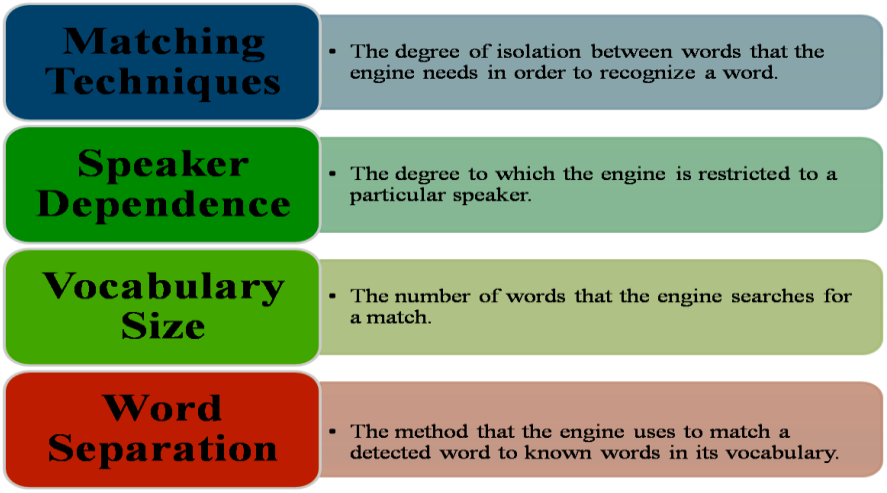

Figure 8. Basic Task of Recognition Engine

Using Whole word matching and Sub-word Matching techniques the detected words for the Speech Recognition engines are classified in the following table.

Whole Word Matching	Sub-Word Matching
Compares recorded digital audio signals with prerecorded template of the word.	Concentrate on Sub-Word like Phonemes and implement them into Pattern Recognition.
Less Processing	More Processing
Require large storage space	Require less storage space
Per word 50 – 512 bytes	Per word 5 – 10 bytes
Need user hand for prerecord each word	No need of user hand
Possible only for the known Vocabulary	Performs based on the pronunciation of the word from English Text.

Table 2. Detecting the words

IV. WORD SEPARATION

Speech recognition engines typically require a specific type of verbal input in order to detect words. The following table explains the word separation engine.

Discrete Speech	Word Spotting	Continuous Speech
Isolation of words by a pause about ¼ of a second.	Continuous utterance without any discrete pauses and recognize only one word	Continuous utterance without any discrete pauses and recognize all words
Low Processing , less natural and user friendly	Process done only for the limited number of commands	Most natural speaking style , best technology of usability.

Table 3 . Word Separation

V. CONCLUSION

In this paper a brief description of Automatic Speech Recognition techniques are reviewed. As a first step a small discussion on speech production are noted along with the phases like training and recognition of the Automatic Speech Recognition followed by a summary of methodologies like analysis, feature extraction, modeling and matching. A detailed study on feature extraction techniques along with their properties and procedure are sorted out. Based on the study of feature extraction it has been concluded that each technique has various advantages and disadvantages which is used for various purposes. However Speech Recognition has been researched over fifty decades

still we are focusing on the accuracy and hand free applications. The future scope of this study is to develop the complete accurate applications and will focus on the hearing impaired also.

REFERENCES

[1] L. R. Rabiner and R. W. Schafer. Digital Processing of Speech Signals. Prentice Hall, Englewood Cliffs, Newb Jersey, 1978.
[2] http://speechc.blogspot.in/ Understanding the basic concept of speech communication.
[3] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar , "A Review on Speech Recognition Technique" International Journal of Computer Applications (0975 – 8887)Volume 10– No.3, November 2010.
[4] Shreya Narang, Ms. Divya Gupta" Speech Feature Extraction Techniques: A Review" International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015, pg. 107-114
[5] Kishori.R.Ghule, R.R.Deshmukh, "Feature Extraction Techniques for Speech Recognition: A Review" International Journal of Scientific & Engineering Research, Volume 6, Issue 5, May-2015 143 ISSN 2229-5518
[6] Shivanker Dev Dhingra, Geeta Nijhawan , Poonam Pandit, "Isolated Speech Recognition Using MFCC And DTW", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.2 , Issue 8, August 2013
[7] M.A.Anusuya,"Speech Recognition by Machine," International Journal of Computer Science and Information security, Vol.6, No.3, 2009
[8] Ms. Savitha and S Upadhya" Digit Recognizer Using Single and Average Template Matching Techniques", International Journal of Emerging Technologies in Computational and Applied Sciences, 3(3), Dec.12-Feb.13, pp. 357-362
[9] http://www.o2a.com/SpeechPrimer.htm
[10] Nidhi Srivastava and Dr.Harsh Dev "Speech Recognition using MFCC and Neural Networks", International Journal of Modern Engineering Research (IJMER), March 2007

