



## A Study on Speech Detection

P. Acharya

*PG & Research Department of Computer Science, India*

**Abstract** - 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.

**Keywords** – speech detection

### I. INTRODUCTION

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

Speech Detection (SR) is the inter-disciplinary subfield of computational linguistics which incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields to develop methodologies and technologies that enables the Detection and translation of spoken language into text by computers and computerized devices such as those categorized as Smart Technologies and robotics. It is also known as "automatic speech Detection" (ASR), "computer speech Detection", or just "speech to text" (STT).

### II. AUTOMATIC SPEECH DETECTION

An efficient Automatic Speech Detection system has the major considerations like developing higher Detection accuracy, achieving low word error rate and addressing the issues of variability in the source. Automatic Speech Detection 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 Detection training phase and Detection phase [3].



*Figure 1. Automatic Speech Detection*

#### 2.1. The Hierarchy of Speech Detection

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

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

### III. SPEECH DETECTION METHODOLOGIES

Speech Detection Methodologies have four different stages. Each stage of the methodologies deal 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 Detection.

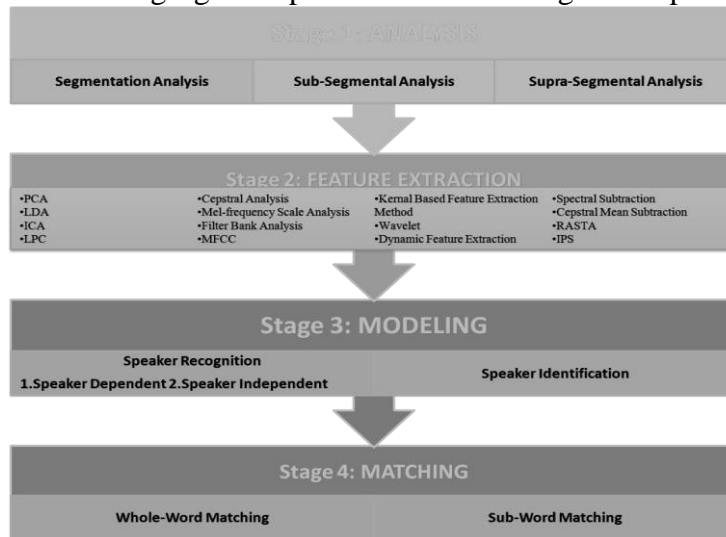


Figure 2. Speech Detection Methodologies

#### 3.1. Analysis

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

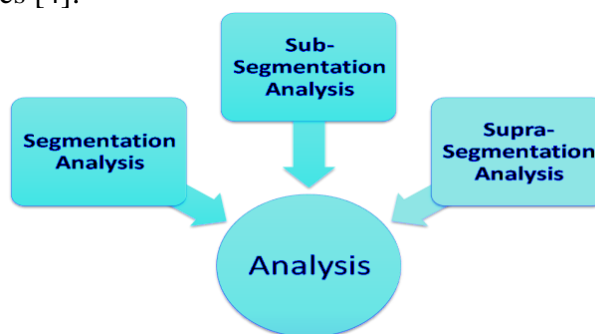


Figure 3. Speech Analysis

- **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 Detection 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.

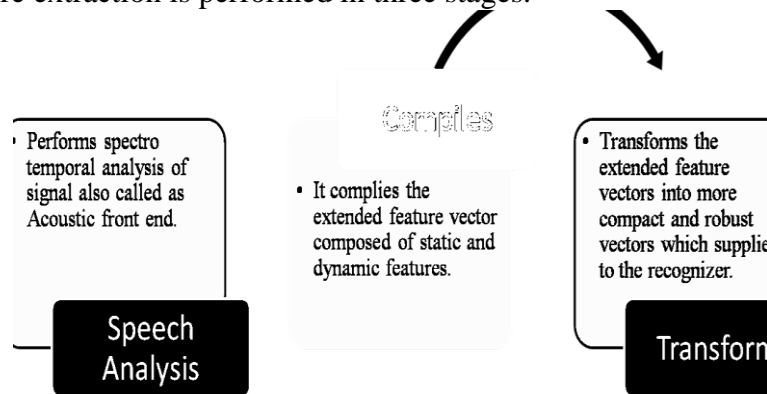


Figure 4. Stages of Feature Extraction

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 Detection. 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].

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>•PCA</li> <li>•LDA</li> <li>•ICA</li> <li>•LPC</li> <li>•Cepstral Analysis</li> <li>•Mel-frequency Scale Analysis</li> <li>•Filter Bank Analysis</li> <li>•MFCC</li> </ul> | <ul style="list-style-type: none"> <li>•Kernal Based Feature Extraction Method</li> <li>•Wavelet</li> <li>•Dynamic Feature Extraction</li> <li>•Spectral Subtraction</li> <li>•Cepstral Mean Subtraction</li> <li>•RASTA</li> <li>•IPS</li> </ul> |
|---|---|

Figure 5. Different types of Feature Extraction Analysis

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

Feature Extraction Techniques	Characteristics
<b>Linear Discriminate Analysis (LDA)</b>	Non linear feature extraction method, Supervised linear map; fast.[3]
<b>Independent Component Analysis (ICA)</b>	Non linear feature extraction method, Linear map, iterative non-Gaussian[3]
<b>Cepstral Analysis</b>	Static feature extraction method, Power spectrum
<b>Mel - Frequency Spectrum</b>	Power spectrum is computed by performing Fourier Analysis
<b>Wavelet</b>	Better time resolution than Fourier Transform
<b>Spectral Subtraction</b>	Robust Feature extraction method , based on Spectrogram
<b>Cepstral Mean Subtraction</b>	Robust Feature extraction
<b>RASTA Filtering</b>	For Noisy speech

### 3.3. Modeling Technique

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

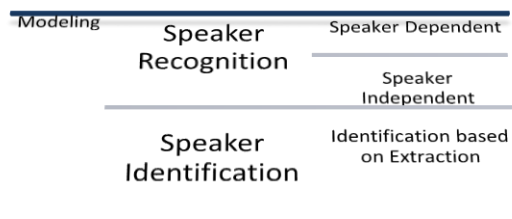


Figure 6 . Modeling Techniques

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

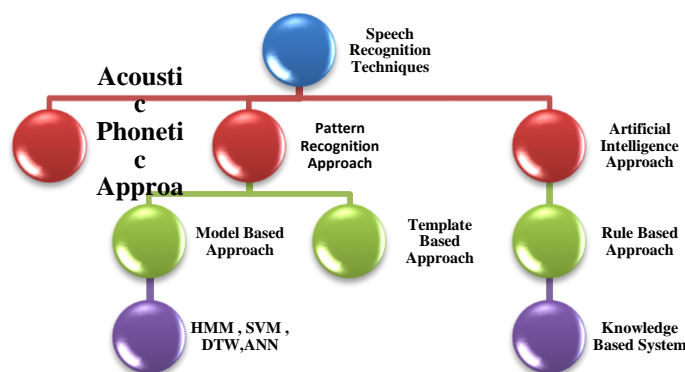


Figure 7. Approaches of Speech Detection

#### 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 Detection

- 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 Detection Approach

Pattern Detection technique is searched trait or branch of artificial intelligence. Pattern training and Pattern comparison are two steps involves in Pattern Detection 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. Detection 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.

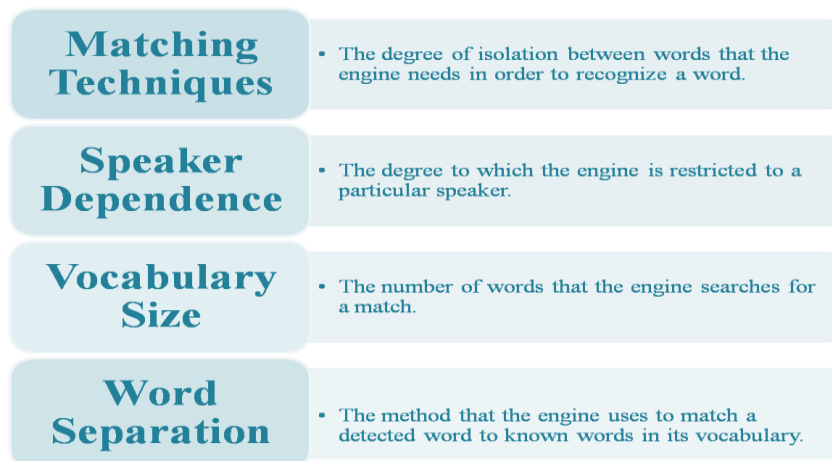
### 3.3.3. Knowledge Based Approach(Artificial Intelligence Approach)

The artificial intelligence approach is the combination of the pattern Detection approach and acoustic phonetic approach so it is called hybrid approach of pattern Detection. 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

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



*Figure 8. Basic Task of Detection Engine*

#### IV. CONCLUSION

In this paper Automatic Speech Detection a small discussion on speech production are noted along with the extraction, modeling and matching. on feature extraction techniques along with their properties and procedure are over fifty decades still we are focusing on the accuracy and hand free applications. The future scope of this study is to progress the complete exact 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, New 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 Detection 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 Detection: 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 Detection 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 Detection by Machine,” International Journal of Computer Science and Information security, Vol.6, No.3, 2009

