Voice Recognition: A Review Paper

Kunal Thakur, Jagdeep jakhu, Kapil , Yogesh, Bhavnish arora ,Neha

Department of Computer Science & Engineering , Dav institute of Engineering and Technology Jalandhar

Abstract : *Speech is a basic form of communication between human beings. Speech recognition is a process of converting speech sounds into corresponding text. In the last few years, speech recognition technology has been developed extensively. However, there still remain many important research challenges. e.g : differences in speaker and language, environmental sound and word size etc. The purpose of this paper is to present a holistic view of the acceptance of speech that describes various processes and summarizes the various methods used in the standard speech system. Voice signals form the basis for the speaker identification. We can use voice targeting in many application areas such as phone banking, phone shopping, database access and voice mail. One of the most powerful applications of voice recognition for security is where One can enter their voice for verification. Speech is the basic form of communication between people.*

Introduction : A human being has natural traits to communicate with another person by voice communication.

For distant communication where the voice is used to identify the identity of a person, his or her voice pitch is one way. It also helps to identify whether the person is a familiar speaker or not. The demand of automatic speaker recognition in verification of identity in e-commerce usage for general business interactions, forensics applications and much more has considerably increased. Human expert trained to investigative various characteristics of speech such as acoustic, prosodic, and linguistic are referred as prearranged listening.

The human ear is a marvelous **organ**. Beyond our unique human ability to receive and decode spoken language, the ear supplies us with the ability to perform many diverse functions. These include, for example, localization of objects, enjoyment of music, and the identification of people by their voices. **Currently,** along with efforts to develop computer procedures that understand spoken messages, there is also considerable interest in developing procedures that identify people from their voices. Being able to speak to your personal computer, and have it recognize and understand what you say, would provide a comfortable and natural form of communication. It would reduce the amount of work you have to do leaving your hands free. It would also help in some cases if the computer could tell who was speaking .The speech signal conveys information at different levels to the listener. At the primary level, speech conveys a message via words.
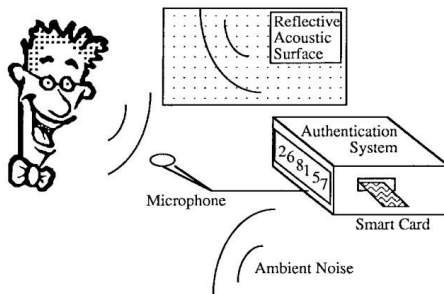
Fig 1 : Typical Voice System

other hand requires a good hand and eye coordination. The physically challenged people find computers difficult to use. Speech which is a natural and very easy way of exchanging the information if used as a medium to interact with the computer and can solve all these problems. Signal processing has made it possible for computers to follow human voice commands and understand human languages as speech can be characterized in terms of signal carrying message information. In automatic speech recognition, an algorithm takes the listener"s role in deciphering speech waves into the underlying textual message. In automatic speaker recognition, an algorithm generates a hypothesis concerning the speaker identity or authenticity.

Speech processing : Speech signal processing could be divided into three different tasks: Analysis, Recognition and Coding. Recognition research fields could be subdivided into three parts: Speech, Speaker and Language recognition systems. While speech recognition aims at recognizing the word spoken in speech, language recognition aims at the detection of language spoken and the goal of speaker recognition systems is to extract, characterize and recognize the information in the speech signal conveying speaker identity. In automatic speech recognition, an algorithm takes the listener"s role in deciphering speech waves into the underlying textual message. In automatic speaker recognition, an algorithm generates a hypothesis concerning the speaker"s identity or authenticity .
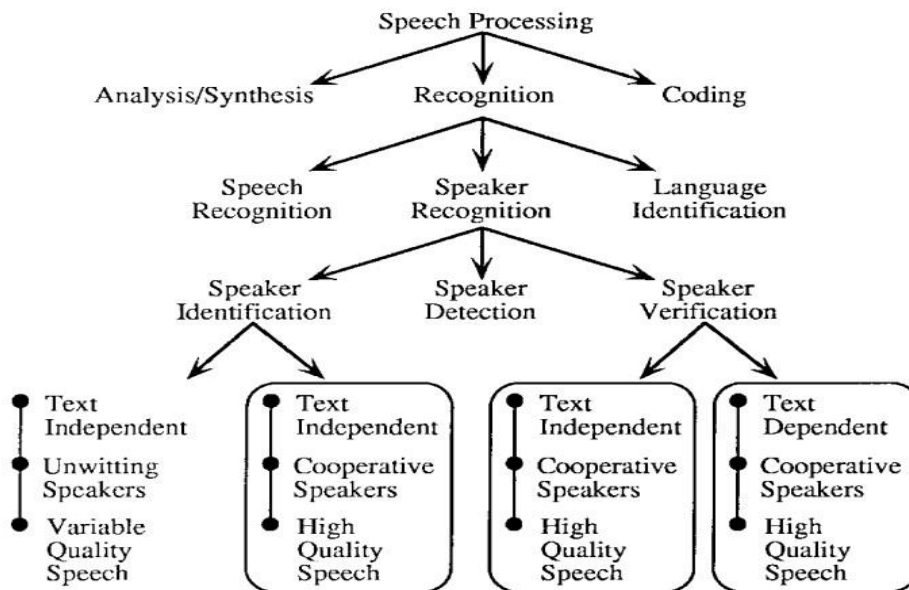
Fig 2 : speech processing

Speaker Identification : The process of identifying a voice of a given speech from the group of given speakers is called speaker identification. The speaker whose maximum voice characteristics are matches with the stored voice is identified & the speaker whose voice characteristics are not matched is eligible for new entry in the database. Speaker identification is the process of finding the identity of an unknown speaker by comparing his/her voice with voices of registered speakers in the database. It is a one to- many comparisons (1: N match where the voice is compared against N templates) .

In Speaker identification System, M speaker models are scored in parallel and the most one of the speaker"s ID in the database, or will be „none of the above" if and only if the matching score is below some threshold and it"s in the case of a open most-likely one is reported, and consequently decision will be open-set Speaker identification system.

The speaker ID problem may further be subdivided into closed set and open set. The closed set refers to a case where the speaker is known and belongs to a set of M speakers. In the open set case, the speaker may be out of the set and hence, a "none of the above" category is necessary. Another distinguishing aspect of speaker ID systems is that they can either be text-independent or text-dependent depending on the application. In the text-independent case, there is no restriction on the sentence or phrase to be spoken, whereas in the text dependent case, the input sentence or phrase is fixed for each speaker. A text-dependent is commonly used in speaker verification systems in which a person"s password is critical for verifying his/her identity
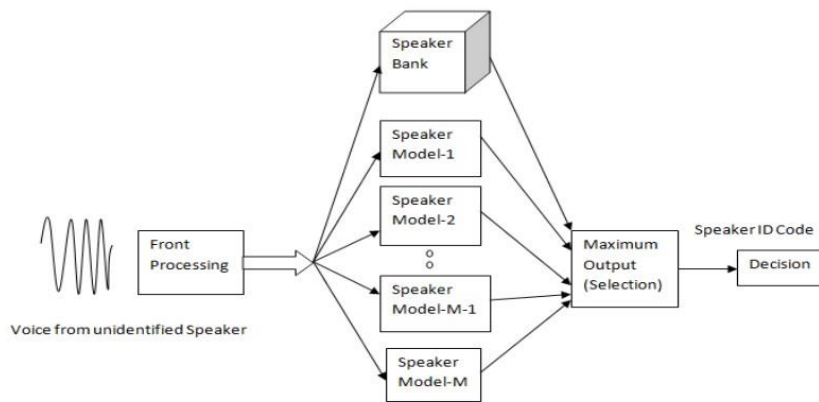
Fig 3 : Speaker identification

Speaker Verification : Speaker verification is used to determine whether a person claims to be according to his/her voice sample. This task is also known as voice verification and speaker detection. Speaker verification is a 1:1 match where one speaker's voice is matched to one template (also called a "voice print" or "voice model") or in other sense Pattern Matching between the claimed speaker model registered in the database and the imposter model will be performed then . If the match is above a certain threshold, the identity claim is verified. Using a high threshold, system gets high safety and prevents impostors to be accepted, but in the mean while it also takes the risk of rejecting the genuine person, and vice versa.

In this process, the system compares two audio samples and then decide by an alternative model that they are spoken by the same speaker or not. Speaker recognition is based on voice system which could be text dependent or independent. This kind of recognition is very significant in applications in which a claimed user conveys specific information like password or some personal identification for getting resource/information. Speaker recognition is a

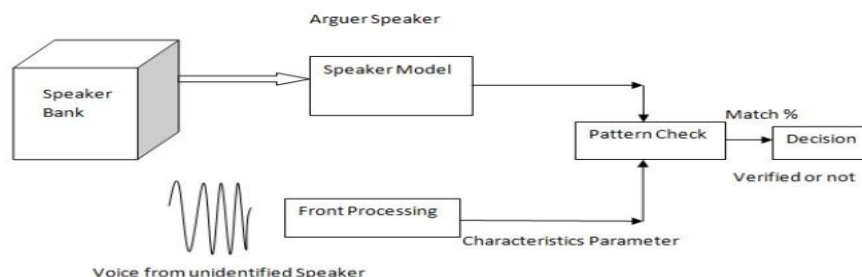text-independent technique for speaker verification, particularly in the handling of automatic systems.

Fig 4 : Speaker verification

Speaker verification is a major task for any recognition system incorporating voice, for this a model-based and features based approach has been developed in previous years.
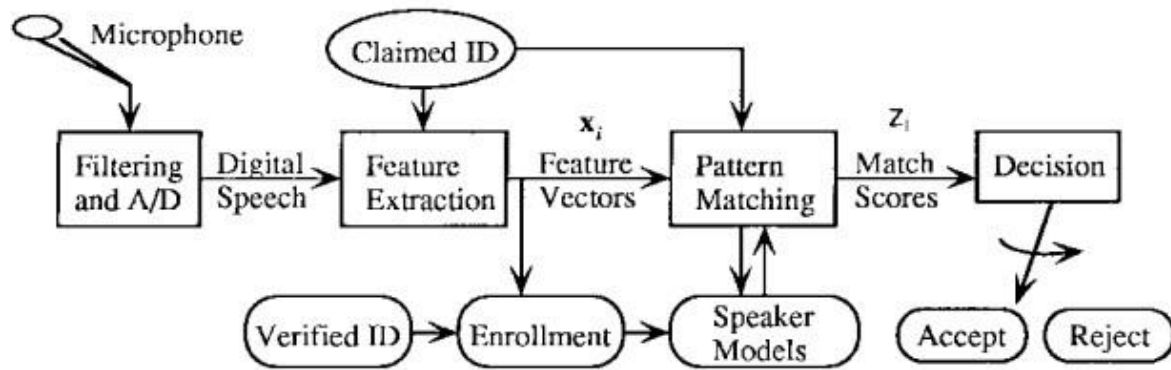

Fig 5 : Generic speaker-verification system

Speaker Recognition System : Speaker Recognition system makes it possible to use the speaker''s voice to verify their identity and control access of the desired services. Speaker recognition system is having three main components: Front -end Processing or Feature Extraction, Speaker Modeling, Pattern Matching or Logical decision (see Figure 3). To get the feature vectors of Incoming voice, front end processing will be performed. Feature vectors are used to create a speaker model. The pattern matching is responsible for comparing the features to speaker models. The decision module analyzes the similarity score (statistical or deterministic) to make a decision. Using a high threshold, system gets high safety and prevents impostors to be accepted, but in the mean while it also takes the risk of rejecting the genuine person, and vice versa.
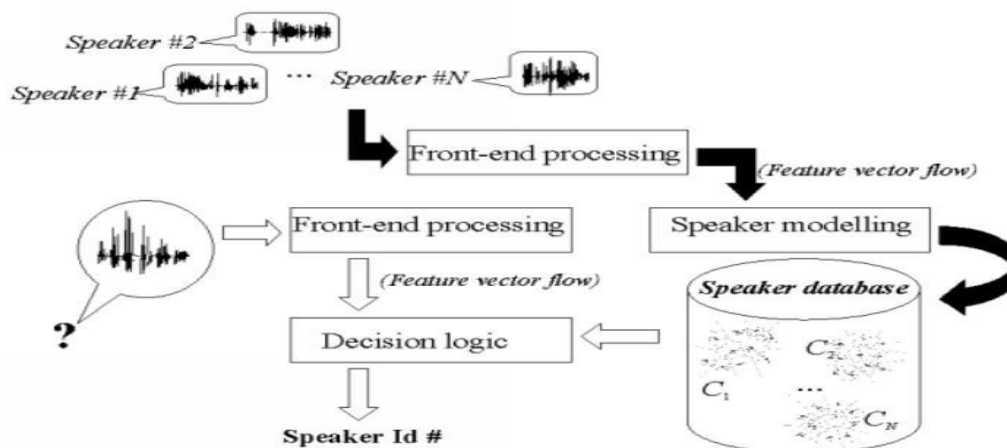

Fig 6 : Speaker Recognition System

Types of Voice Recognition : Speech recognition systems can be of various types depending on types of utterances to be recognized. These various types are classified as follows:

- **Isolated word** recognizers generally obtain each utterance to have quiet on both sides of sample window. These systems usually have two states Listen/Not- Listen states, where speaker has to wait between two utterances. The pauses between utterances are used for processing speech signals.
- **Connected words** are similar to isolated words with only difference of minimal pause between them.
- **Continuous** speech recognition involves almost natural way of speaking. It is difficult to design continuous speech recognizers because it considers special methods to determine utterance boundaries
- **Spontaneous** speech covers the mispronounced, unrehearsed non-words with false statements which are difficult to read . An ASR system under this category is to handle a variety of features such as words being run together such as "ums" and "ahs".

Applications of Voice Recognition : There are several applications of speaker recognition.

- **Forensic Department** Forensics is an important claim for voice recognition. If a voice sample of a criminal was traced during the obligation of crime, the suspect's voice can be compared with this, in order to give a sign of similarity of two voices.
- **Telephony** and other Domain The use of telephony in Automatic Speaker Recognition is very common in today's life whereas in the field of computer gaming and simulation it is becoming more widespread. It is also used in voice mail, e-banking and in various voice control systems.
- **Access Control** Originally various physical facilities like token, finger print, password etc. are used to access the data. But due to the advancement of technologies it is also used in voice recognition system. Voice Recognition system provide access control to various services like automation in cars, homes and mobile phones by voice command, e-banking and telephone shopping etc.
- **Transaction Authentication** Transaction using telephone requires higher level of authenticity to provide access control to various accounts. It is widely used in user verification for e-commerce and m- commerce.

Factors Affecting Voice Recognition : Performance of speaker recognition system is affected by various factors. The interpretation that effect speaker recognition system are described as following [21-22]. In any system background noise is one of the most significant aspects of speaker recognition system accuracy. The accuracy of a system is good if it is a clean sample but it deteriorates rapidly in noisy samples. Babble noise generated by the system has no considerable influence. Another factor that effect speaker recognition is the microphone. Any mismatch in quality of microscope is insignificant and has a minor impact on efficiency. Disguise is one factor that affects accuracy. Quality of Voice signal is other factor on which speaker recognition is dependent. Good quality of microphones leads better results.

Language for training and testing does not have a major impact on the efficiency of speaker recognition in native language speech. Speech modality, it depends on text dependent or text–independent. Speaker population has a very considerable impact on the efficiency of a system.

Conclusion : In this paper the basics are discussed and their recent progress is investigated. Different ways available by building a voice recognition system based on a modified feature translation process and speech recognition language system are compared in this paper. Speech is basic mode of communication between human beings, so a feasible interface is required to connect human with machines. Although this field has gained a wide approval to automate the services and applications but there are several parameters which affect the accuracy and efficiency of speech recognition system. The most of speech variability involves speech rate, environmental conditions, channel and context of utterance. Robustness of speech system depends on some stable parameters/ features of speech signal. To enhance the power of speech recognition system, it is required to design speech recognizers in local languages. Multilingual is new evolving field in area of speech recognition. There is a lot of development and research in the field of foreign languages but to enhance its power and utility for native people, it's essential to use this technology in native languages.

[1] VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad JAFS|ISSN 2395-5554 (Print)|ISSN 2395-5562 (Online)|Vol 1(2)|November 2015

[2] (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009 Speech Recognition by Machine: A Review M.A.Anusuya and S.K.Katti

[3] Volume 5 Issue V, May 2017 IC Value: 45.98 ISSN: 2321-9653 International Journal for Research in Applied

Science & Engineering Technology (IJRASET) ©IJRASET: All Rights are Reserved 262 Voice Recognition Technique: A Review Nisha

[4] Speech Recognition: A Complete Perspective Ashok Kumar, Vikas Mittal international Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019

[5] SPEAKER RECOGNITION USING AUTOMATED SYSTEMS Syed Akhtar Imam1 , Priyanka Bansal2 ,

Vimlesh Singh AGU International Journal of Engineering & Technology 2017, Vol. No. 5, Jul-Dec

[6] J.P. Campbell, "Speaker recognition: A tutorial", Proc. IEEE 85, pp. 1437–1462, 1997

[7] A Review On Speaker Recognition Approaches And Challenges Varun Sharma Dr. P K Bansal M. Tech

(Student) Director MMU,Solan (HP) MMU,Solan (HP) International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 5, May - 2013 ISSN: 2278-0181