



# A Review: Automatic Speech Recognition for Machine

Sudha Sharma

*Department of Computer Science and Engineering, L.R.I.E.T Solan  
Himachal Pradesh Technical University, India*

**Abstract** - Speech Recognition technology is a fast growing engineering technology now days. Speech Recognition is the ability for a device to recognize individual words or phrases from human speech. These words further can be command the operation of a system. This paper describes the basics of speech recognition including its types. There would be number of factors that can affect and does matters in speech recognition, few of these are discussed in this paper with application in numerous fields.

**Keywords** - Speech recognition, voice recognition, speaker identification, speaker verification, isolated words.

## I. INTRODUCTION

There is billions of human being in all over the world. They speaking different languages and yet we are able to recognize someone by listening to someone's conversation or speech as long as we can understand the language. We can usually recall someone's voice even we have not seen that person for years. [3]

Research in speech recognition means the interaction between the human and machine. Automatic speech recognition uses the process and technology for converting a speech signal into a sequence of words. [1] Speech recognition research has been ongoing more than 80 years. By 2001, computer speech recognition had reached 80% accuracy and no further process was reported till 2010. Speech recognition technology development began to edge back into the front with one major event: the arrival of the "Google voice search app for the iPhone ". In 2010 Google added "personalized recognition to "voice search on android phones, so that the software could record users". Voice search and produce a more accurate speech model. It draws its knowledge about the speaker to generate a conceptual reply and responds to voice input. [1]

Speech recognition is technology is one from the fast growing engineering technology. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition system in those particular cases provide a significant help them, so they can share information with people by operating input through voice input.[4] Speech recognition and voice recognition are two different things. Both use the recording of the human voice, but they do different things with it. Speech recognition strips out the personal difference to detect the words. Voice recognition typically disregards the language and meaning to detect the physical person behind the speech. Speech recognition is

language dependent while voice recognition is language independent.

Speech recognition is the ability for a device to recognition individual words or phrases from human speech. These words can be used to command the operation of a system. [4] Voice recognition is also called speaker identification, by checking both voice and phrase authenticity to recognition who said. Voice recognition has two sides, the speaker identity and speaker confirmation. [3]Voice recognition is aimed toward identifying the person who is speaking. Voice recognition works by analyzing the features of speech that differ between individuals. Everyone has a unique pattern of speech stemming from their anatomy (the size and shape of the mouth and throat) and behavioral pattern (their voice's pitch, their speaking style, accent, and so on). The applications of voice recognition are markedly different from those of speech recognition. Most commonly, voice recognition technology is used to verify a speaker's identity or determine an unknown speaker's identity. Speaker verification and speaker identification are both common types of speech recognition.

**Speaker verification** is the process of using person's voice to verify that they are who they say they are. Essentially, a person's voice is used like a fingerprint. Once a sample of there is recorded, a person's speech patterns are tested against a database to see if their voice matches their claims identity. Most commonly, speaker verification is applied to situations where secure access is needed. Such systems operate with the user's knowledge and cooperation.[1]

**Speaker identification** is the process of determining an unknown speaker's identity. Unlike speaker verification, speaker identification usually covert and done without the user's knowledge. [1] For example, speaker identification can be used to identify a criminal solely by their voice. In this situation, a sample of their voice would be checked against a database of a criminal's voice until a match is found. Recently, this technique was unable to identify a South American drug kingpin who had hidden his physical identity by undergoing extensive plastic surgery.

## II. TYPES OF SPEECH RECOGNITION

Divided into the number of classes based on their ability to recognize that words and list of words they have:

- 1) **Isolated speech**:- The isolated speech involve a pause between two utterance ,it doesn't mean that it only

accepts a single word but instead it requires one utterance at a time.[4]

- 2) **Connected speech:-** Connected words or connected speech is similar to isolated speech but allow separate utterance with minimum pause between them.[4]
- 3) **Continuous speech:-** Continuous speech allows the user to speak almost naturally; it is also called the computer dictation.[4]
- 4) **Spontaneous speech:-** Spontaneous speech is natural sounding and not reheard. An automatic speech recognition system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being together “ums” and “ahs” and even slight stutters.[4]

### III. FLOW DIAGRAM OF SPEECH RECOGNITION SYSTEM

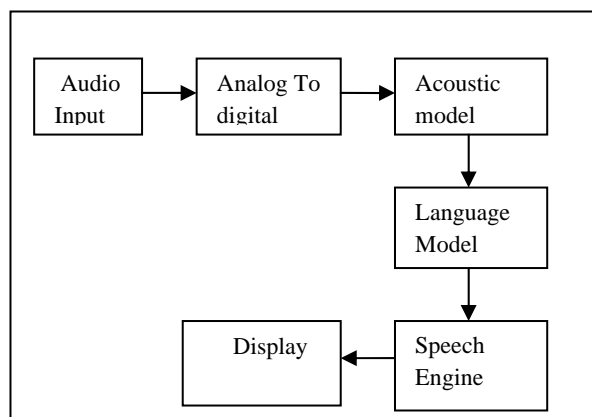


Fig.1 A Speech Recognition Process

- **Audio input:-** An audio input I recorded for the recognition process. The input I in analog form.
- **Analog to Digital conversion:-** The analog input I converted into the digital form by A to D converter.
- **Acoustic Model:-** This model I used to represent the relationship between audio signal and the phonemes or other linguist units used to make a speech. Each of these statistic representations is assigned a label called a phoneme.
- **Language Model:-** The language model represents the ways in which the words of a language are combined.
- **Speech Engine:-** The speech research engine taken an audio stream as input and convert it into text transcription.
- **Display:-** The output provided by speech engine is displayed in words.

### IV. FACTORS AFFECTING SPEECH RECOGNITION

There are five factors used for simplify and control the speech recognition system-

- 1) **Isolated words** : - speech consists of isolated words (short silence between the words) is much easier to recognize than tee continuous speech because word boundaries are difficult to find in continuous speech. [4]

- 2) **Single speaker:-** Speech from a single speaker is also easier to recognize than the speech from a variety of speakers because most parametric representation of speech are sensitive to the characteristics of the particular speaker.[4]
- 3) **Vocabulary size:-** The size of vocabulary of words to be recognized strongly influences recognition accuracy. Large vocabulary is more likely to contain ambiguous words than small vocabulary. Ambiguous words are those words whose pattern matching templates appear similar to the classification algorithm used by the recognizer.
- 4) **Grammar:-** The grammar of the recognition domain defines the allowance sequence of words. A tightly constrained grammar is one in which the number of words that can legally follow any given word is small.
- 5) **Environment:-** Background noise, change in microphone characteristics and loudness can all dramatically affect recognition accuracy. Many recognition systems are capable of very low rates as long as the environment conditions remain quiet and controlled.

### V. APPLICATIONS OF SPEECH RECOGNITION

- 1) **Medical perspective:-** people with disability can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands.
- 2) **For military purpose:-** In air force speech recognition has defined potential for reducing pilot workload. Beside the air force such programs can also be trained to be used in helicopters battle management and their applications. They didn't have to use their hands.
- 3) **For education perspective:-** Individual with learning disability who have problem with thought to paper communication can benefit from system.
- 4) Radiology scanning hundreds of X-rays, ultra sonograms, CT scans and simultaneously dictating conclusion to a speech recognition system connected to word processor. The radiologist can focus his attention on the image rather than writing the text.[5]
- 5) Voice recognition could also be used on computer for making airline and hotel reservation. A user required simply stating his needs to make reservations, cancel a reservation or making enquiry about schedule.[5]

### VI. CONCLUSION & FUTURE SCOPE

Speech recognition is the interaction between the human and the machine. Speech Recognition should be helpful for the people who are suffering from various disability such as in case of blindness or if unable to use their hands effectively. Speech recognition can be helpful by operating input through voice input. It does involve various processes, by using number of method and techniques for having the entire process/ operation to be done. Number of different techniques can be studied and applied for checking the better results in future.

### REFERENCES

- [1] Suma Swamy & K.V. Ramakrishnan, "An Efficient speech recognition system", *CS & Engineering: An International Journal (CSIJ)* Vol.3, No. 4, 2013.
- [2] B.H. Juang & Lawrence R. Rabiner, "Automatic speech Recognition – A Brief History of Technology Development", pp.2.
- [3] Vangie Beal, voice recognition, [Online], Available at: [http://www.webopedia.com/TERM/V/voice\\_recognition.html](http://www.webopedia.com/TERM/V/voice_recognition.html), [Accessed: July 2015]
- [4] Chris Woodford, Voice recognition software, June, 2015, [Online] Available at: <http://www.explainthatstuff.com/voicerecognition.html>, [Accessed: Aug 2015]
- [5] Raghvendra Priyam, Rahmi Kumari & Dr. Prof Videh Kihori Thakur, "Artificial Intelligence Applications for speech Recognition", *Conference on Advances in communications and Control system*, pp.475-476, 2013.