National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]

Use of Artificial Intelligence in Speech Recognition

Shubham Upadhyay[1], Prof. Pradnya Mhatre[2]

[1]Department of computer Application, University of Mumbai
Viva School of MCA, Shirgaon, Virar (East), Maharashtra, India
Email: shubhamupadhyay109@gmail.com
[2]Department of computer Application, University of Mumbai
Viva School of MCA, Shirgaon, Virar (East), Maharashtra, India
Email: pradnyamhatre@vivamca.org

***Abstract*—**Purpose of this research Paper is to define what is role played by artificial intelligence in *speech recognition. Speech recognition is a term in which computer convert human voice in machine-readable form and act upon that. Humans use speech and gestures during interaction with each other's. Speech recognition makes possible that interact with computer same as humans. Using speech recognition user can perform another task simultaneously.This research paper defines the role of Artificial Intelligence in Speech Recognition. Natural Language Processing used for interaction with computers in natural languages like English. NLP used in speech recognition for understand input and perform right action.*

Keywords—Artificial Intelligence, converting Speech signals, gesture recognition system,Natural Language processing, Speech Recognition.

I. INTRODUCTION

The purpose of this research paper to introduce the use of artificial intelligence in speech recognition. Speech recognition also referred to as ASR (automatic speech recognition) or STT (Speech to text) or CSR (computer speech Recognition).

In today world, everyone wants to do things in easy way and very fast and that is the reason we use technology in every field like industry, business, sports, households, agriculture, war etc.

With the use of AI, we make our technologiesfaster compared to without using AI. Some of the example is some speech recognition technology like amazon Alexa and google home mini; they are smart speakers that is work on human voice commands.

In old days if anyone wants to give public speech in their own language and community does not understand that language then they use a person as a translator who is responsible for translate particular speech in community native language. This process takes a lot of time and efforts now with the use of speech recognition technique this problem solved easily and fast (speaker speeches in their own languages and listener listen in native language) that is the power of AI.

II. BASIC CONCEPT OF SPEECH RECOGNITION

Speech recognition is a complex phenomenon. Most of people does not understand that how this technology works. When two or more people interact with each other that comes in area of telecommunication services, in this paper we focused upon man-machine interaction. Let us try to understand speech recognition process in some steps.

Structure of Speech:speech is a continuous voice stream where rather stable states mixed with dynamically changed states. In this states sequence, one can define less or more similar sounds classes or **phones.**Phone is simply defined as speech sound in a given language. For example the English word kid and kit end with different phenomes, /d/ and /t/ , and swapping one for the other changed one word into different word.

National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]

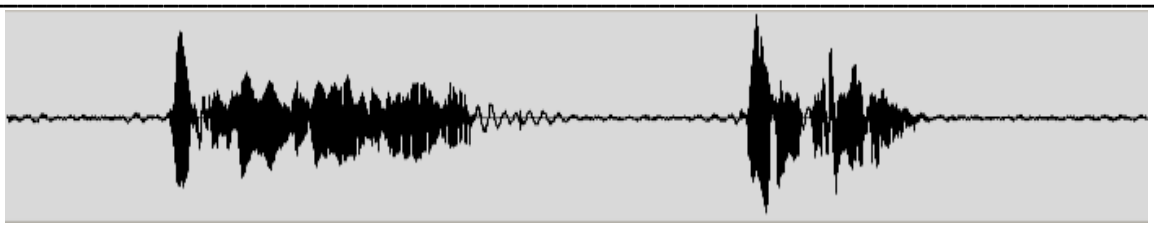


Figure. 1. Structure of Voice Analog signal

a. Recognition Process:

The user interact with application using input devices i.e., a microphone. The simple way of recognize speech is the following:
Microphone receive waveform's, after removing background noise split it at utterances and then try to recognize what's the meaning of each utterances.
To perform that, we should take all possible combinations of words and try to match them with audio and choose the best matching combination.We should always take care some points in matching process.
First, it is **feature's** concept. Since there is large number of parameters, we are trying to optimize it. Number's that computed from speech usually by dividing the speech into frames.
Second, it is **model** concept. Signal modeling for speech recognition is challenging task. The concept of model describes some mathematical object that gathers common attributes of the spoken word.This model of speech known as **Hidden Markov Model** or **HMM**.HMM is describe as the heart of automatic speech recognition system.HMM describes the black-box communication channel**.**In HMM model process described as a sequence of states, which change each other with a certain probability. From the concept of model, the following issues raise:

- How much model describe the reality
- How much model is adaptive if conditions will changed

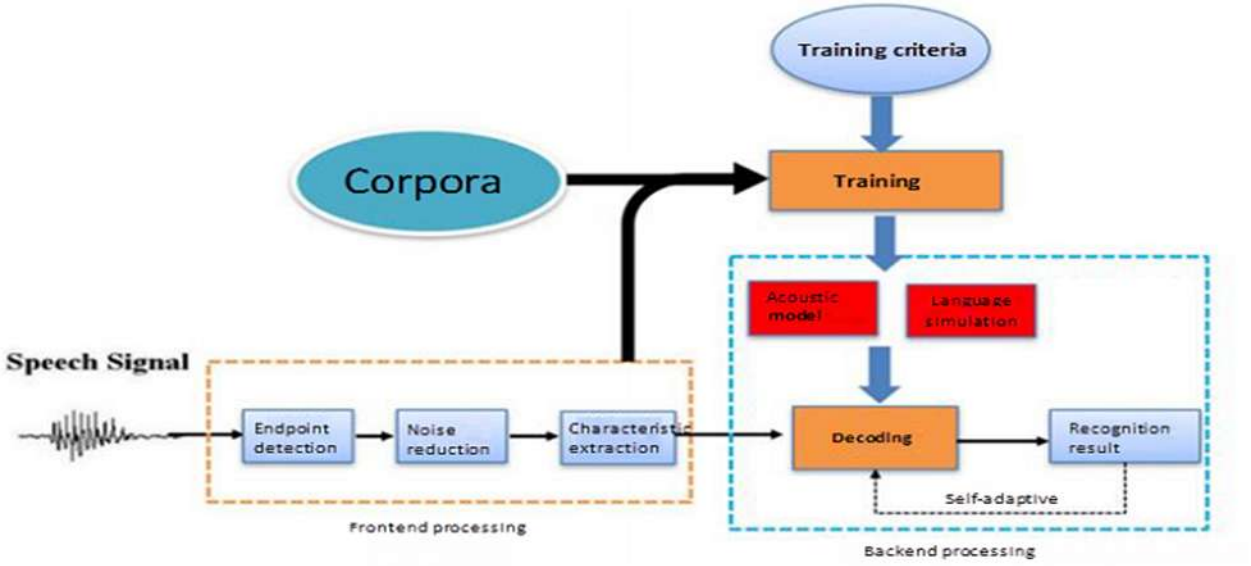


Figure. 2. Recognition Process

The given Figure shows the structure diagram of Automatic Speech recognition System. It is mainly comprises a frontend Processing model, Acoustic model, Language model and decoder.

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]

b. Models: following models are used in speech recognition for matching purpose:

i. Acoustic Model

Acoustic was the originally the study of small pressure wave in air which can be detected by human ear. In addition, the perception of sound is an area of acoustical research. In our present paper, we will define the propagation of signal only in fluids like air and water.Acoustic Model used for represent the relationship between an audio signal and phenomes or other linguistic units that make up speech. The main task of Acoustic model is computing the P(O|W), i.e. the probability of generating a speech waveform by Acoustic model. Acoustic model is backbone of Automatic Speech Recognition.

ii. Language Model

The language model worked for modelling the word sequences in the language.In new technical devices speech recognition systems operate on the audio in small chunks known as frames with an approximate duration of 10ms per frame. Language models used for retrieving the information from query likelihood model.
For building language, model data sparsity is a major problem, most possible word sequences not observed in training. There are only one solution seems good for data separation is find the probability of word which is only depends on previous n words.

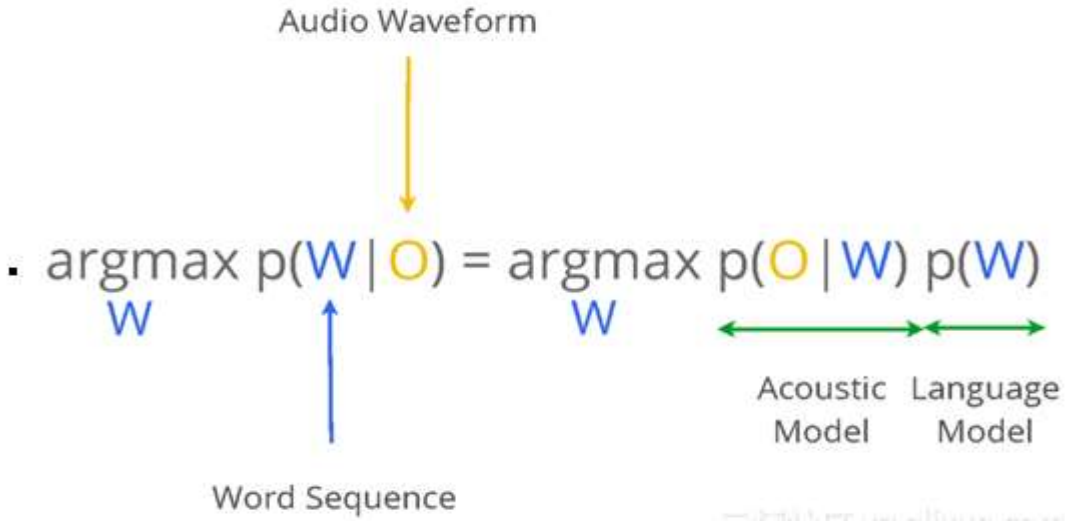
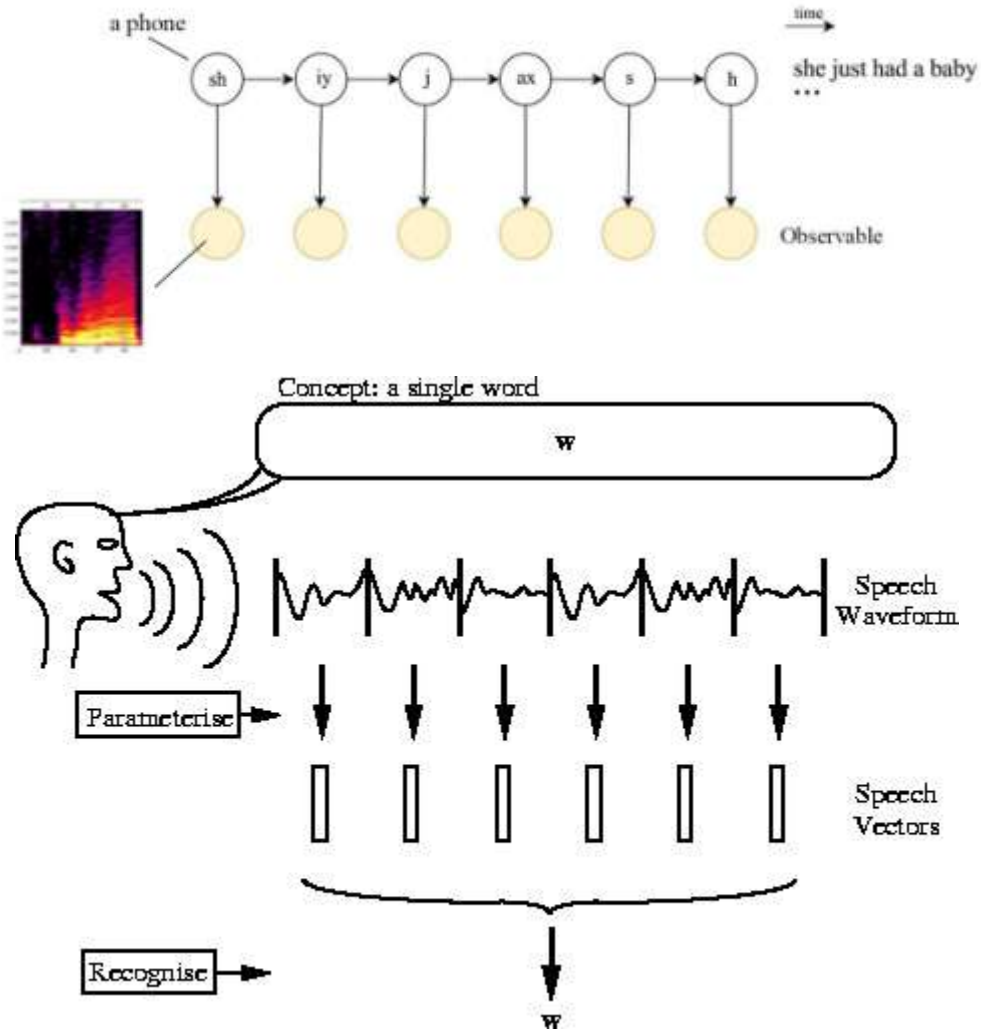


Figure. 3. Formula for calculating Audio Waveform

iii. Hidden Markov Model

Hidden Markov models is especially known for their work in reinforcement learning and temporal pattern recognition e.g. speech recognition, handwriting recognition, gesture recognitionetc. The term Hidden is nothing but a morkov process behind the observation.A Hidden Markov process can be shape as a generalization of the urn problem with replacement. Hidden Markov model is become predominant model from the last serval years. With the use of HMM models we do not require search phone sequences one-by-one. With the use of HMM Model, we can find optimal sequence in polynomial time.That is the reason this method become so popular.HMM is describe as the heart of automatic speech recognition system. HMM describes the black-box communication channel. In HMM model process described as a sequence of states, which change each other with a certain probability.

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]



Today Speech recognition technologies use both acoustic and language model to represent the statistical properties of speech. The main idea of AI is to collect and employ knowledge from a number of sources for solving the problem in question (Rabiner and Juang, 1993).

III. Application areas of Speech Recognition

a. In the Modern office

Speech recognition technologies can improve the throughput of modern office. It helps to minimize the human efforts and do simple task easily and fast. There are some examples of modern offices are as follows:

- Print document on request.
- Start video conferences.
- Schedule meeting
- Make travel arrangements
- Search for reports or documents on your computer etc.

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]

b. In Marketing

With the use of speech recognition, there are new type of data available for marketers to analyses customer needs and requirements.

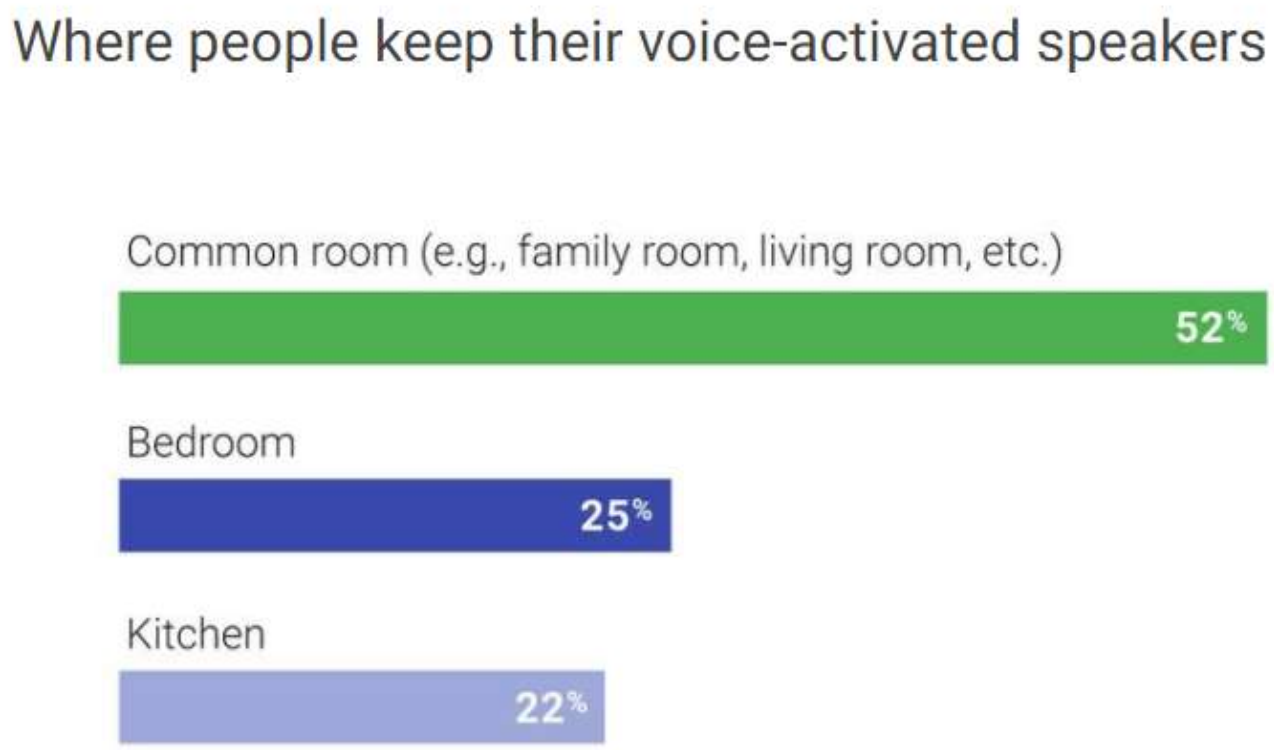
c. In Healthcare

Speech recognition system is more useful thing in health industry. There are following types of benefits available of speech recognition.

- Quickly finding information from medical records
- Less Paperwork
- Less time required for inputting data
- Improved workflow
- ASR systems provides facilities to physically challenged person to command and control a machine.

IV. CURRENT SPEECH RECOGNITION LANDSCAPE

Smartphone are the sole place of residence for digital assistance like google assistance, Siri and Cortana. Recent google research found over 50% of users keep their voice activated speaker in their living room.



V. CONCLUSION

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2020]

Speech recognize will revolutionize the way people conduct business over the web and will, ultimately, differentiate excellent e-business. Voice XML tries to merging speech recognition and telephony together and provides the technology with which business can develop and deploy voice-enabled web solutions today! These solutions can greatly expand the accessibility of web based self-service transactions to customers who would otherwise not have access, and, at the same time, leverage a business' existing web investments. Speech recognition and voiceXML clearly represents the next wave of the web.
Speech Recognition Systems are indefensible part of the ever-advancing field of human computer interaction.

ACKNOWLEDGEMENTS

We would like to express a deep sense of gratitude towards our mentor Mr. Tatwadarshi P. Nagarhalli, Department of Computer Engineering for his constant encouragement and valuable suggestions. The work that we have been able to present is possible because of his timely guidance and support.

REFERENCES

[1] http://murray.newcastle.edu.au/users/staff/speech/ home pages/tutorial acoustic.html
[2] http://www.haskins.yale.edu/haskins/ HEADS/MMSP/acoustic.html
[3] E.Noth and A.Homdasch, "Experiences with Commercial Telephone-based Dialogue Systems", IT Information Technology, 46 (2004) 6, pp.315-321.
[4] https://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/tutorial%20on%20hmm%20and%20applications.pdf
[5] https://medium.com/@jonathan_hui/speech-recognition-gmm-hmm-8bb5eff8b196