ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Review Paper on Speech Recognition Technology and its Application in Intelligent Systems

Vidhi Patel[1], Nirali Dodhia[2], Purav Chheda[3], Aarti Sahitya[4]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[1]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[2]

U.G. Student, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[3]

Assistant Professor, Department of Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India[4]

ABSTRACT: The concept of speech recognition technology is one of the fastest growing engineering technologies. Speech recognition technology is a translation of spoken words into text and then the text is converted into speech.Some fields that include speech recognition technology are physiology, psychology, linguistics, computer science, and signal processing. The main aim of the speech recognition technology is to allow natural language communication between man and machine. This paper narrates the development of automatic speech recognition, its basic principle, applications, methodology etc.

KEYWORDS: Speech Recognition, Hidden Markov Model, Applications, Natural Language Communication..

I. INTRODUCTION

The process of taking spoken words as an input to a computer program can be called speech recognition. Speech Recognition is the technology wherein spoken words, sounds by humans or phrases spoken are converted into electrical signals, and these signals are then transformed into coding patterns to which meaning has been assigned. Certain speech recognition systems require training and they can be classified as "speaker -dependent " systems. Speech Recognition technology can be a difficult technology to implement because of the facts: The production and transition of phonemes vary from person to person. Different people speak differently, even a single word or phrase may differ from time to time.

Innovation has assumed control over the market for a decade ago. These days, for the most part, every single individual in the general public have smart devices. Among all the innovations speech recognition has taken over the market. Speech recognition system can be used in various domains. It has vast exposure in all the emerging technologies. One of the advantages of using speech recognition technology is that it reduces the use of text and other types of input also help in minimizing the calculations required for the process. A few years back it was difficult to implement the speech recognition system in any of the devices but with advancement in technologies it is now highly possible to implement it When trying to cook something new, people usually tend to watch videos or read upon recipes online. There is usually a bridge between what we intend to cook and what we actually end up doing. People might lose track of the recipes or

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

misunderstand the context of it. The main aim of this idea is to bridge this gap by proposing a smart cooking assistant.It will be an interactive system using a Raspberry Pi toolkit.

II. LITERATURE REVIEW

1. The paper was created on a real-time speech recognition system. The design of a bidirectional nonstationary Kalman filter was used to enhance the speech recognition ability. They have created their own database for its flexibility and TIDIGIT database for its accuracy comparison with HMM-based speech recognition. MFCC technique is used.
2. In this paper, the MFCC technique is used. Minimum distance classifier and Support vector machine(SVM) methods are used for speech classification. The Machine Learning concepts of classification have been used. Languages such as English, Marathi, and the combination of both is used. It uses training data as well as testing data.
3. Voice controlled smart home systems has now become the latest trend in the field of Speech recognition. The general idea of smart home design is to monitor and control the household appliances wirelessly and through voice. They have used various services provided by Amazon. They have used the Raspberry Pi module, Alexa skill kits, Ngrok service, Electronic components, and sensors, etc. They have divided the implementation into two parts- Alexa voice service and Raspberry pi module.They have written the main python script in Raspberry Pi to execute the commands that are processed by Alexa Voice Service. If the execution of the main python script is done and Ngrok online, Alexa is ready to work.
4. This paper gives an overview of major technological perspectives of the fundamental progress of speech recognition. In this system, they are going to develop an online speech-to-text engine. It uses a technique based on the HMM (Hidden Markov Model). In this paper, the topics relevant to STT are discussed. The ASR systems work in two phases- training phase and recognizing phase. In the training phase, the system represents the different speech sounds that constitute the vocabulary of the application. In recognizing phase, an unknown input pattern is identified by considering the set of references. This paper concludes with the decision for developing techniques in human-computer interface system in different mother tongues
5. They have given the general idea of HMM based speech synthesis. The HMM based speech synthesis is considered to be very effective in synthesizing speech. It is very flexible in terms of changing speaker identities, emotions, accents, tones, etc. They have implemented it based on the source filter theory of voice production. It is called as voice filter model

IJIRSET

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

III.SYSTEM DESIGN

A. RASPBERRY PI

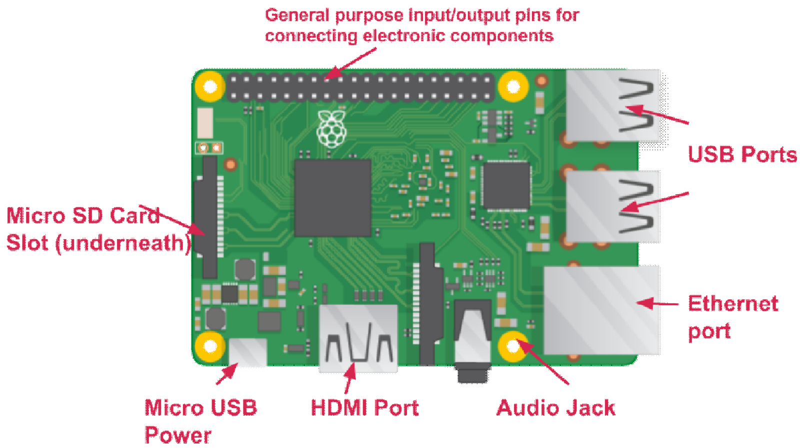

Figure 1. Raspberry Pi 3

The Raspberry Pi is a human palm size microcomputer. It is considered as the brain of the entire Speech Recognition system, capable of processing programming codes and controlling a huge number of devices. Some of the specifications of Raspberry Pi are that it processes at 1.2GHz with a Broadcom BCM2837 64bit. It has 40-pin extended GPIO, 4 USB 2 ports, a full-size HDMI port, Micro SD port for loading your operating system and storing data, upgraded switched Micro USB power source up to 2.5A. Additionally, it also has CSI camera port for connecting a Raspberry Pi camera, DSI display port for connecting a Raspberry Pi touchscreen display For its wireless communication, the Raspberry Pi has both wifi and Bluetooth module on board. For the Raspberry Pi device to be programmed has to connect with a keyboard, mouse and a monitor. Raspberry Pi can also be connected and controlled by a smartphone or computer via the Secure Shell (SSH). Thus Raspberry Pi is a device that takes input from the processed voice commands and various sensors, processes them with the written Python codes and executes the corresponding commands specified by the codes. Figure 1 shows the front schematic of a Raspberry Pi 3 Model B.

B. HIDDEN MARKOV MODEL

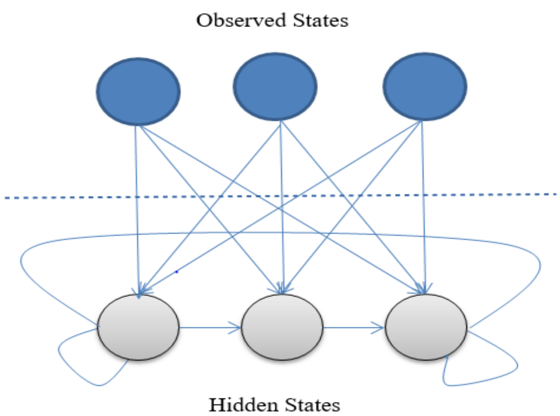

Fig 2.Process of Hidden Markov Model

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Hidden Markov model(HMM) is a statistical Markov Model in which the system being modelled is assumed to be a Markov process with unobserved (i.e. hidden) states. The simple example of a Hidden Markov Model is giving the output by cancelling the surrounding noise when given the human voice input (hidden)and providing the user only with desired voice output(observed). An HMM can be viewed as a Bayes Net unrolled through time with observations made at a sequence of time steps being used to predict the best sequence of hidden states. The construction of an inference model based on the assumptions of a Markov process reasons its name, the Hidden Markov Model. The Markov process assumption is simply that the “future is independent of the past given the present”[9].Fig 2 a shows the process between observed states and the hidden states.

There are three basic algorithms linked with Hidden Markov Models:
1.The forward algorithm, useful for isolated word recognition;
2.The Viterbi algorithm, useful for continuous speech recognition.
3.The forward-backward algorithm, useful for training an HMM. [6]

There are some of the inherent limitations of Hidden Markov Model in speech recognition they are as follows

- The assumption that successive observations areindependent.
- The Markov assumption itself.
- The distribution of individual observation parameterscan be well represented as a mixture of Gaussian or autoregressive densities.
- Constant length of observation frames.
- Trial and error method for choosing a model topology.
- The number of parameters needed to set up an HMM is huge.
- Amount of data required to train an HMM is very large.[10]

IV. WORKFLOW

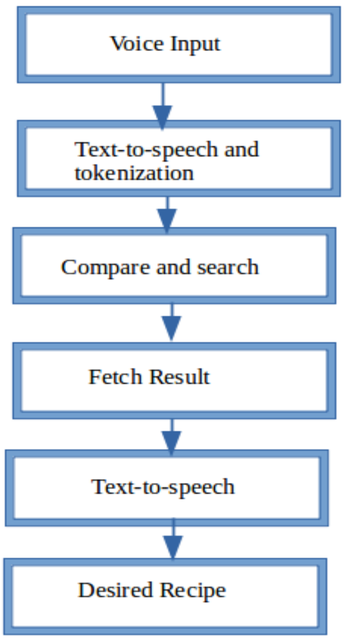

Fig 3. Flow Diagram

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Here, voice is captured using google API, which involves packages such as PyAudio, SpeechRecognition.It is based on voice commands. The device recognizes the user voice and then it processes the voice by amplifying and clearing background noise. After processing it, the voice issplitted and most common or frequent words are removed using tokenization. Then the required data is searched and compared if there is a match the information is fetch. the fetched information is then passed to text to speech class. this class convert the text data into speech format and desired output is given out through speaker. inthis speech to text , TTS text to speech and searching algorithm has important roles.One of the important implementation of the intelligent systems is cooking assistant where in speech recognition technology is used where the desired output is obtained only with the input of human voice and output is obtained as an audio file. Fig. above shows the whole process of conversion from speech to text.

V. CONCLUSION

In the near future, smart cooking assistants will be essential to everyone with the ever-improving intelligent systems. Thus, the proposed system will effectively and efficiently bridge the gap that is created due to the tedious task of collecting knowledge and information from various sources about various recipes. This system will only be taking the mere input of voice from the user asking for various recipes, and the system cooking assistant will be doing all the tasks of searching for the recipe that the user has asked for and providing the required content. The system will also the problem of language barrier

REFERENCES

[1] N. Sharma and S. Sardana, "A real time speech to text conversion system using bidirectional Kalman filter in Matlab," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, 2016, pp. 2353-2357. doi: 10.1109/ICACCI.2016.7732406

[2] N. Sharma and S. Sardana, "A real time speech to text conversion system using bidirectional Kalman filter in Matlab," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, 2016, pp. 2353-2357. doi: 10.1109/ICACCI.2016.7732406.

[3] Yue, Chan Zhen, and Shum Ping. “Voice Activated Smart Home Design and Implementation.” 2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST), IEEE, 2017, pp. 489–92. Crossref, doi:10.1109/ICFST.2017.8210563.

[4] PrachiKhilari, Bhope V.P, “A review on speech to text conversion methods,”

[5] Tokuda, Keiichi, et al. “Speech Synthesis Based on Hidden Markov Models.” Proceedings of the IEEE, vol. 101, no. 5, May 2013, pp. 1234–52. Crossref, doi:10.1109/JPROC.2013.2251852.

[6] Pahini A. Trivedi, “Introduction to various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks”,IJEDR, 3590-3596, 2014

[7] Yang, Fan. “Research on Speech Visualization Technology and Its Application in English Listening Teaching.” 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE, 2018, pp. 355–58. Crossref, doi:10.1109/ICITBS.2018.00097.

[8] V. Ajantha Devi, V Suganya, “An analysis on types of Speech Recognition and Algorithms”, IJCST

[9] https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab

[10] Chandralika Chakraborty, P.H. Talukdar, Issues and Limitations of HMM in Speech Processing: A Survey, International Journal of Computer Applications (0975 – 8887 Volume 141 – No.7