Voice Assistant Using Python and AI

Divisha Pandey[1], Afra Ali[2], Shweta Dubey[3], Muskan Srivastava[4], Shyam Dwivedi[5], Md. Saif Raza[6]

[1, 2,3,4,] *Student of B. Tech fourth year, Department of Computer Science and Engineering, Rameshwaram Institute of Technology & Management, Lucknow, India*
[5]*Assistant Professor and Head of Department CSE, Rameshwaram Institute of Technology and Management, Lucknow, India*
[6]*Assistant Professor, Department of CSE, Rameshwaram Institute of Technology and Management, Lucknow, India*

---***---

Abstract – *Today's era is the era of digitalization. Having smart phones and desktops is no less than having the world on our fingertips. Our lifestyle is involving being busy day by day. That busy, that people even find it a load to even type something to perform a task. So here comes virtual assistant at rescue. Just speak to it and the task is done. From sending a hello on WhatsApp to your friend to sending a full fleshed email to your boss virtual assistant will do it all for you. With time voice search is dominating over text searching. But what are virtual assistants? A software program that helps us perform our daily task just by speaking to it is a virtual assistant. A waking word is necessary to activate the software. This system can be used efficiently on desktops. The premise behind starting this project was that the data present on the web is sufficient and is openly available that can be used to build a virtual assistant that can make and perform intelligent decision for the user.*

Index Terms – Python, **Artificial Intelligence, Natural Language Processing, Speech Recognition.**

1. INTRODUCTION

We are living in the era of technology where the era is replacing human beings by machines. Lifestyle and productivity are the main reason behind this performance change and will also evolve with coming time. We need machine that think like humans and perform the task given to them by human beings, and to do so we are training them. And as a result of one of these training came the concept of virtual assistant.

A virtual assistant is self-employed software who is specialized in offering administrative services to clients from remote location, usually a home office. Scheduling appointments, making phone calls, booking tickets, sending messages and what not a virtual assistant can perform them all. It uses voice recognition features and language processing algorithms to perform a task by recognizing the voice command of users. Filtering out irrelevant noise and background disturbances are ignored by the assistant itself and give out relevant information as per the user requirement. This is a software-based technology but companies nowadays are creating special devices integrated with this system that perform tasks. Amazon Alexa is one such example.
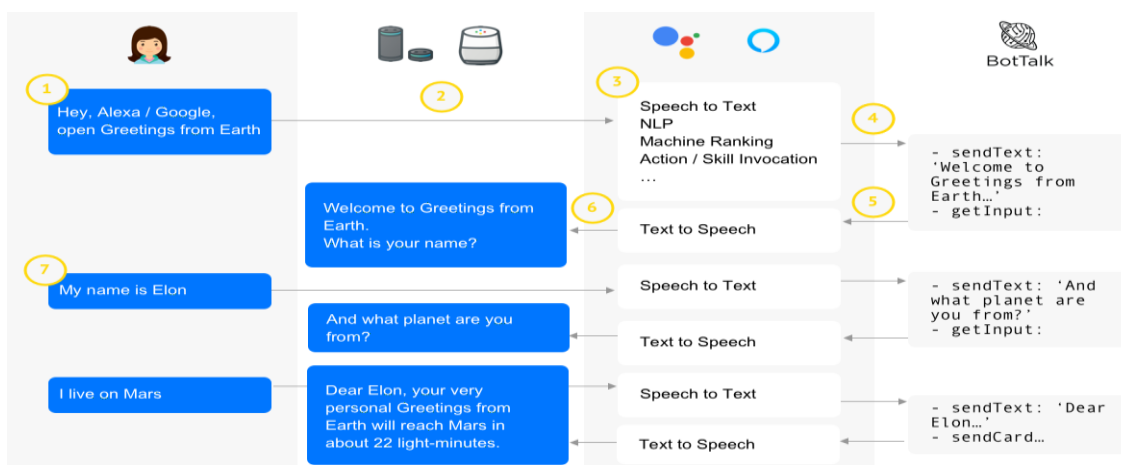

Fig -1: Backend Working of Virtual Assistant

Day by day drastic changes are forming out in technologies. These changes are making it necessary to train our machines with advancement. Deep learning, machine learning and neural network are some of the current technologies that involve in the training of machines for their advancement. Voice assistant have made possible human and machine conversation.

Basically, we can say that these assistants are next level of advancement in development. The main privileged parts of the society who are benefiting from these assistants are old age, blind, physically challenged, and children. Blind people who cannot see can even interact with the machine with their voice only. Following are few tasks that can be performed by virtual assistant:-

1. Reading out newspaper
2. Sending emails
3. Searching among web
4. Playing music
5. Playing YouTube video
6. Making notes
7. Setting up alarm
8. Giving weather updates
9. Run any application
10. Checking stock price
11. Playing game

These listed examples are only few task of the assistant. It can perform many more task as per the demand of the user. The voice assistant developed by us is for the Windows user. This voice based module is desktop based which is built using python modules and libraries. It is a basic version that can perform the entire basic day to day task assigned to them by the user operating it. Few of the tasks to be performed by our assistant is listed above. The current technology is good in many aspects but still can be improved by merging it with Machine Learning and Internet of Things (IoT). Python modules and libraries have been used by us along with artificial intelligence and machine learning for training our model. Some windows command has also been used by us in our model for making it to run smoothly on window operating system. Basically, there are three working modes of our model:-

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

It can be used according to the requirement of the user. Machine learning and Deep learning along with natural language processing concepts help us in achieving our goal and performing our desired task. With assistant we don't need to type the command again and again for performing the particular task. After creation the model can be used any number of times by any number of users easily. Basically, this virtual assistant we can control many things on a single platform.

2. LITERATURE SURVEY

[1.] Bassam A, Raja N. et al, have wrote about statement and speech for communication between humans and machines analog signals are used which is converted by speech signal to digital wave. The technology is massively utilized and has unlimited uses and also permit machines to reply accordingly to users command and voices. Speech recognition system is growing day by day and also has unlimited uses.

[2] B.S. Atal and L.R. Rabiner et al, has explained regarding speech analysis, and the theory is getting evolved day by day. The research performed describes a pattern recognition technique for the determination of voice. It determines that the voice input is weather voiced speech, unvoiced, or silence. It completely depends upon the dimensions finishing on the signal. The system although comes with restrictions and the main restriction here is the requirement for exercising the algorithm on the exact set of dimensions picked, and also for recording circumstances.

[3.] V. Radha and C. Vimala et al, explained about the most suitable way of communication between humans is speech. Since speech recognition is an utmost technique of recognition, hence it makes human beings identical and makes it easier for machines to recognize them. This helps in autonomous speech recognition and also has a lot of reputation. Some of the most used speech recognition techniques are Dynamic Time Warping (DTW), HMM. For feature mining of speech Mel Frequency Cepstrum Coefficients (MFCC), it offers a group of characteristic vectors of speech waveform. Studies have revealed that MFCC is more precise and real than other mining approaches in speech recognition. The research has been done on MATLAB and the outcomes on investigation depict that the system is capable in identification of words at a great satisfactorily accuracy.

[4.] T. Schultz and A. Waielet al, explained about the spreading of speech technology products around the world. The research tells about the query on how to port huge vocabulary incessant speech recognition (LVCSR) systems in a fast and well-organized manner. However, there is a need to evaluate the acoustic models for novel destination language by means of speech information from different source languages. But the restricted data from destination language identification outcomes using language dependent, independent and language adaptive acoustic models are deliberated in the framework of Global Phone project which examines LVCSR methods in 15 languages.

5. J. B. Allen et al has described Language as the utmost and significant means of communication and speech is its major interface. For the interface creation between humans and machines, the speech signals were converted into analog and digital wave shape as for the machine to understand. Speech technologies today permit the machines to react appropriately according to human speeches and offers valuable and appreciated services. The carried out research gave the result in terms of speech identification procedure, its basic model, its application, and techniques and also describe several other research techniques that are necessary for speech recognition system. SRS is an emerging technology and is increasing its vitality day by day gradually and also has infinite applications.

6. Mugdha Bapat, Pushpak Bhattacharyya et al, described morphological analyzer for almost of the Indian languages. At the starting phase the planning was about some extent homomorphism "boos trappable" encryption technique. The research proved out to be a great success for Marathi language that resulted in engagement of the Finite State Systems for the demonstration of language in a sophisticated way. Since Marathi has a really difficult morphotactics hence the growth of FSA is one of significant assistances.

7. G. Muhammad, M.N. Huda et al, presented an ASR model for the Bangla digits. To carry out this research the information was gathered for general Bangladeshi public. For identification purpose Mel-frequency cepstral coefficients (MFCCs) and hidden Markov model (HMM) were used. In the trial it was discovered that female spoken digits have higher accuracy than male spoken digits.

8. Sean R Eddy et al researched on Hidden Markov Models. They are basically a common statistical designing approach for issues like sequences or time series. These methods are extensively being used in the process of speech recognition. With the help of HMM formalism, it is possible to create a relation between formal, completely probabilistic techniques to profiles and gapped structure arrangements. Steady theory for insertion and deletion, constant structure for joining structural and sequence data are some of the popular offerings of HMM. It also makes sequence arrangements more refining. It also makes satisfactorily arrangements for difficult threading techniques for protein reverse fold.

3. FEATURES OF VOICE ASSISTANT

TASK PERFORMANCE

A task is a piece of work to be done or undertaken. It can be occurring once or on repetition. A task that is occurring on repetition is known as recurring task. Its repetition can occur at some certain intervals or at a pre appointed time to the system in some cases. Let us understand it better with an example, suppose our team lead wants the progress of our work on every Thursday, so we will add it to the recurring task list. Once we mark the current week task as done at the desired time we will start getting reminders about the task of the upcoming week   . Similarly, Task Request can also be created by the user. With the help of task request a user can assign task to different users. Another feature that is a task list is associated to task request. This list contains information like who assigned the task, who are assigned the task, date of assigning, and followed by reassigning of the task.

INTERNET SOLICITATION

The assistant allows the person to engage with the internet for accessing of information like weather, directions, schedules, stock performance, news etc, and that also just using simple voice command. The growth of internet is creating a vast new network - a Voice Web – that help in accessing internet content just by the use of human voice. It can be called as a voice portal to access the web. It creates a platform for users with natural language interface to access the web content.

SYSTEM ARCHITECTURE

The system architecture of this project shows flow of control through the system. The hardware and software specifications are also depicted here. The architecture diagram is as follows
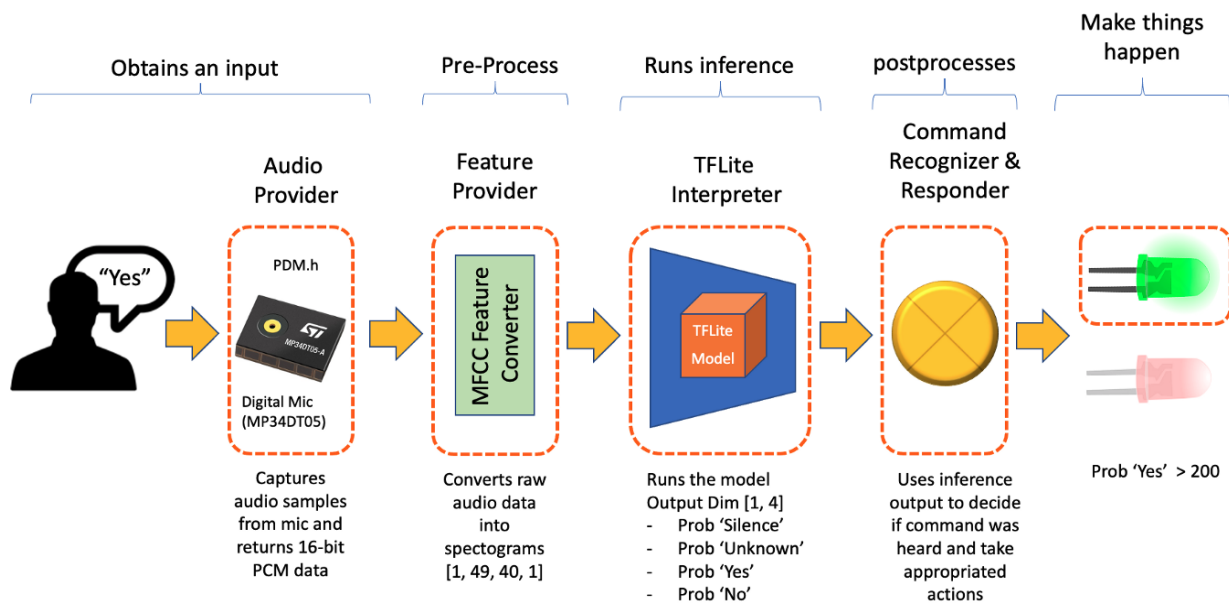
Fig -2: Architecture Of Virtual Assistant

HARDWARE AND SOFTWARE REQUIREMENTS

> **HARDWARE**
> - A desktop / laptop
> - Minimum 512 MB RAM
> - Internet connectivity
> - USB debugging mode for development and testing
> - Pentium-pro processor or later

> **SOFTWARE**
> - Windows 8 and higher
> - Selenium Web Automation
> - SQLite

4. SYSTEM DESIGN AND IMPLEMENTATION

EXISTING MODEL

Out of all the existing projects in the market most of them only use speech recognition using neural network. Although their system give result based on moderate accuracy. Few of the techniques used by them are-

> **CONTEXT AWARE COMPUTING**

Context-aware computing is a style of computing in which situational and environmental information about people, places and things is used to anticipate immediate needs and proactively offer enriched, situation-aware and usable content, functions and experiences. The main use of this technique is to recognise the word spoken by the peoples and also presuppose the mispronounced words.

> **MEL-FREQUENCY CEPSTRAL COEFFICIENTS**

MFCC is the collection of coefficients; this technique aims to develop the features from the audio signal which can be used for detecting the phones in the speech. It is widely used technique for extracting the features from the audio signal.

> **NATURAL LANGUAGE PROCESSING**

NLP is the branch of computer science more widely it is the branch of artificial intelligence that helps in the interaction between humans and machines. It is due to the existence of NLP only that makes possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

♣ PROPOSED MODEL

➢ SPEECH TO TEXT

It is software that enables the recognition of human language and also convert in into the language understood by machines using computer linguistics. It is also known as speech recognition.

➢ TEXT ANALYZING

- Inputs provided are just letters for computer.
- Software converts the speech into machine understood language.
- Commands are understood by the computers, virtual assistants convert this text to command.
- Virtual assistants convert or relate the words to functions and parameters for the creation of a command to be understood by the computer.

The major milestone of our project is trying to increase the accuracy of speech to text software. The model will basically be able to convert any speech with modulations or different accents with a higher accuracy on the day to day basis. The given model is combines voice recognition with neural network to increase the precision.

5. WORKING PRINCIPLES

The virtual assistant involve following principles for working:

♣ Natural Language Processing:

A method used in artificial intelligence for communicating with the machines or an intelligent system is known as natural language processing (NLP). Processing of natural language is required when humans want to make machines like robots to follow their command and also respond to them in human language. Five steps of natural language processing are-
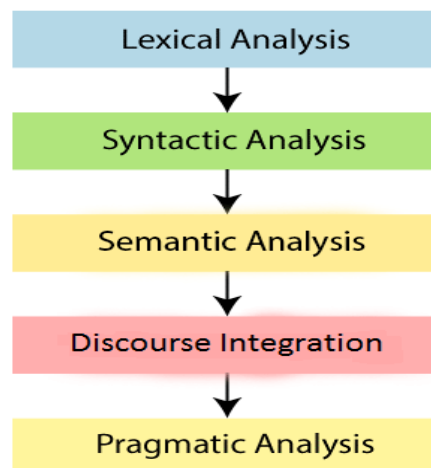

Fig -3: Working Model Of NLP

♣ AUTOMATIC SPEECH RECOGNITION

This feature helps the machine to understand the command as per user's input. The architecture of speech recognition system is given below:
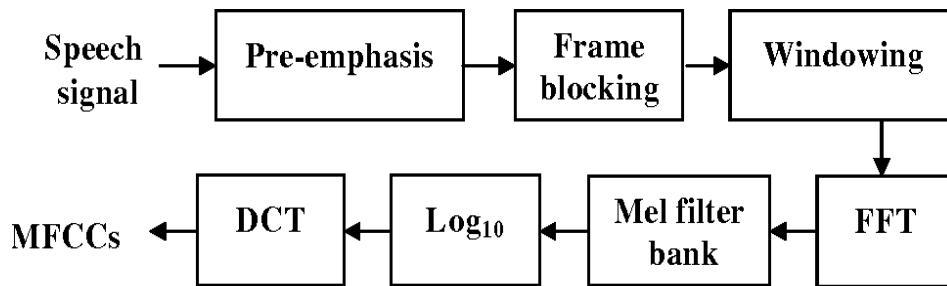
Fig -4: Architecture of Automatic Speech Recognition

ARTIFICIAL INTELLEGENCE

Artificial Intelligence (AI) is a way of making the machines performs the tasks given by humans in a way a human will do it. A machine can calculate, perceive analogies, learn from experiences, store and retrieve information in its memory, solve problems, use natural language, classification, and generalization and even adapt to environment and many more, this all has been made possible due to the presence of artificial intelligence.
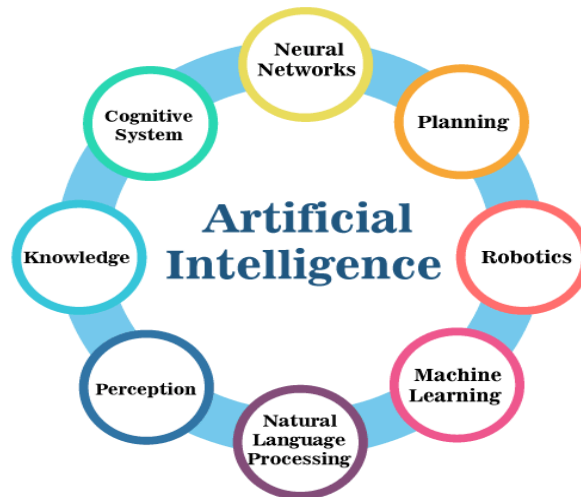

Fig -5: Branches of Artificial Intelligence

INTER PROCESS COMMUNICATION

Inter process communication between operating system and the undergoing or ongoing processes.

6. CONCLUSION

The paper tells about the new emerging technology for the desktop users. The virtual assistant provides a smart working experience for the desktop user over the web. This new service is based on internet of things, speech recognition and various other modern technologies like artificial intelligence, natural language processing and deep learning. Virtual Assistant reduces the interruption of user, reduce the working time performance, and provide single platform for doing all sort of work such as sending messages, contacting, and various other information. The system has become an ideal platform for millions of user around the globe. It also overcomes many of the drawbacks of the existing system. It is basically more efficient than various other existing software in the market. Although it has some of its own limitations. Though it has high efficiency and also may have higher time consumption for task completion. Also the algorithms used make it quite a challenge to tweak it in the near future.

7. REFERENCES

[1]. G.O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics,* 2nd ed. Vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-24.

[2]. M. Bapat, H. Gune, and P. Bhattacharyya, "A paradigm-based finite state morphological analyzer for marathi," in Proceedings of the 1st Workshop on South and Southeast Asia Natural Language Processing (WSSANLP), pp. 26-34, 2010.

[3]. Knote, R., Janson, A., Eigenbrod, L. and Sollner, M., 2018. The What and How of Smart Personal Assistants: Principles and Application Domains for Is Research.

[4]. V. Radha and C. Vimala, "A review on speech recognition challenges and approaches," doaj. Org, vol. 2, no. 1, pp. 1-7, 2012.

[5]. G. Muhammad, Y. Alotaibi, M.N. Huda, et al, pronunciation variation for asr: A survey of the"Automatic speech recognition for bangla digits, literature" Speech Communication, vol. 29, no. in Computers and Information Technology, 2009.2, pp. 225-246, 1999.

8. BIOGRAPHIES

Divisha Pandey - She is currently a student of B. Tech fourth year , Dept. of Computer Science and Engineering , Rameshwaram Institute of Technology & Management, Lucknow and working on Virtual Assistant using Python and AI.

Shweta Dubey - She is currently a student of B. Tech fourth year, Dept. of Computer Science and Engineering , Rameshwaram Institute of Technology & Management, Lucknow and working on Virtual Assistant using Python and AI.

Afra Ali - She is currently a student of B. Tech fourth year, Dept. of Computer Science and Engineering , Rameshwaram Institute of Technology & Management , Lucknow and working on Virtual Assistant using Python and AI.

Muskan Srivastava - She is currently a student of B. Tech fourth year, Dept. of Computer Science and Engineering , Rameshwaram Institute of Technology & Management , Lucknow and working on Virtual Assistant using Python and AI.

Shyam Dwivedi - He is currently Working as an Assistant Professor and of Head of Department in Rameshwaram Institute of Technology and Management , Lucknow ,   India . He is M.TECH – 2012 BIT Mesra , Ranchi , he has a teaching experience of 10 years and 1 – year in TCS Industrial experience.