

VOICE ASSISTANT USING PYTHON

Dev Ranjan Raut*¹, Ayush Rai*²

*^{1,2}Department of CSE, Galgotias University, Greater Noida, India.

ABSTRACT

Today technological advances are growing day by day. Previously there was a computer system where we could do only a few tasks. But now machine learning, artificial intelligence, in-depth learning, and a few more technologies have made computer systems so advanced that we can do any kind of work. In this age of development when people are still struggling to work together using various input tools, that is not appropriate. For this reason, we have developed a voice assistant using a python that allows the user to use any type of command in linux without any keyboard contact. The main function of the voice assistant is to reduce the use of input devices such as keyboard, mouse etc. It will also reduce hardware space and costs.

Keywords: Artificial Intelligence, Desktop Assistant, Python, Text to Speech, Visual Assistant, Voice Recognition.

I. INTRODUCTION

In this era of technology everything that human being can do are being replaced by machines. One of the main reasons is change in performance. In today's world we train our machine to think like humans and do their task by themselves. Therefore, there came a concept of virtual assistant.

A virtual assistant is a digital assistant that uses voice recognition features and language processing algorithms to recognize voice commands of user and perform relevant tasks as requested by the user. Based on specific commands given by the user a virtual assistant is capable of filtering out the ambient noise relevant information.

Virtual Assistant are completely software based but nowadays they are integrated in different devices and also some of the assistants are designed specifically for single devices like Alexa.

Due to drastic change in technology now it's a high time to train our machine with the help of machine learning, deep learning, neural networks. Today we can talk to our machine with the help of Voice Assistant. Today every big company is using Voice Assistant so that their user can take the help of the machine with his voice. Therefore, with the Word Assistant we move on to the next level of development where we are able to communicate with our machine.

These types of visual aids are very useful for the elderly, the blind & the physically challenged, children, etc. by ensuring that interaction with the machine is no longer a challenge for people. Even blind people who could not see a machine could communicate with it using only their voice.

Here are some of the basic tasks that can be done with the help of voice assistant: -

- Search on web
- Play a music or video
- Know the current date and time

These are just some of the examples, we can do many things according to our needs.

The Voice Assistant we built for Windows and Linux users. The voice assistant we developed is a desktop-based built using python modules and libraries. This tutorial is just a basic version that can do all the basic functions mentioned above but the current technology is good though good for them it will still be integrated with Machine Learning and Internet of Things (IoT) for better development.

We used python modules and libraries for modeling and used Machine Learning to train our model, other windows and linux commands were also added to the model to make our model work better in this application. Basically, our model will work in three ways: -

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Depending on the application the assistant is required by the user. And this can be achieved with the help of machine learning and Deep Learning.

With the help of a Voice Assistant there will be no need to rewrite instructions for any specific task. Once the model is created it can be used with any number of times for any number of users in simple ways.

Therefore, with the help of a virtual assistant, we will be able to control many things around us in one place.

II. LITERATURE SURVEY

Bassam A, Raja N. et al, wrote about a very important statement and speech. The communication between human-machine programming is done with an analog signal that is converted into a speech signal into a digital wave. This technology is widely used, has unlimited usage and allows machines to respond appropriately and consistently to users' voices, and also provide useful and enjoyable services. Speech Recognition System (SRS) is slowly evolving and has endless applications. Research has revealed a summary of the process; is a simple model [1].

B. S. Atal and L. R. Rabiner et al, explained in terms of speech analysis, and the result is always supplemented with tone analysis. The study explained how to determine the pattern of determining whether a given piece of speech signal should be categorized as pronounced speech, non-verbal expression, or silence, depending on the size of the signal. The main limitation of the strategy is the need to use an algorithm in a set of selected sizes, as well as specific recording circumstance[2].

V. Radha and C. Vimala et al, explained that the most common form of human communication is speech. Since this is a very advanced strategy, people can be like using speech to communicate with machines. As a result, independent speech recognition is highly esteemed. Many speech recognition techniques such as Dynamic Time Warping (DTW), HMM. The Mel Frequency Cepstrum Coefficients (MFCC) terminology is used which provides a group of vectors for the speech wave structure. Previous research has revealed the MFCC to be very accurate and realistic than techniques that are a factor in digging into speech recognition. The effort has been completed in MATLAB and the results of the investigation indicate that the system is able to identify words with great accuracy in a satisfactory manner [3].

T. Schultz and A. Waiel et al, describing the worldwide distribution of speech technology products, the immobility of the new languages being addressed is a practical concern. As a matter of fact, the study emphasizes the question of how to implement vocabulary incessant speech recognition (LVCSR) programs in a fast and systematic manner. In particular research is required to evaluate acoustic models of the target language using speech information from different source languages, but only limited data from the target language identification results using language-based, independent and intentionally defined acoustic models in the framework of the Global Phone project that explores LVCSR methods in 15 languages. [4].

J. B. Allen and others explain that Language is the most important form of communication and speaking is its main communication. The human visual interface to the machine, the speech signal was converted into analog and digital waveform as understood by the machine. [10] The most widely used technology and has unlimited applications. Speech technology allows machines to respond appropriately and consistently to human conversations and provides valuable and respected resources. The study provides a summary of the speech recognition process, its basic model, its application, techniques and describes a comprehensive study of several techniques used in the speech recognition system. SRS is slowly evolving and has endless applications. [5]

Mughdha Bapat, Pushpak Bhattacharyya et al, described a morphological analyst of multiple NLP applications for Indian Languages. [11] During the work they describe and evaluate the morphological analyzer of the Marathi language. They started by somewhat modifying the encryption method for the homomorphism "boos trappable" which works during the f-task function to extract each encryption. Studies have shown great Marathi accuracy that travel is subject to varying degrees in the integration of Finite State Systems to reflect language in a more complex way. Collection of positions and growth of FSA is one of the most important resources as Marathi has complex morphotactics [6].

G. Muhammad, M. N. Huda et al, introduced the digital digital ASR model for Bangla. Although Bangla is one of the most widely spoken languages in the world, some of the few works of the Bangla ASR can be seen in

collections of works, most notably the Bangla translation into Bangla. During the study, a number were collected from the Bangladeshi community. The dependent components of the Mel-frequency cepstral coefficients (MFCCs) and the dependent components of the Markov model (HMM) are used for identification. Dialectical variability makes it possible to partially reduce performance. In the case of gender-based experiments, female spoken digits had a greater degree of accuracy than male spoken digits [7].

Sean R Eddy et al work on Hidden Markov models that are a common method of designing statistics for 'specific' problems such as sequence or time series and have been widely used in speech recognition applications for two decades. Within the legitimacy of HMM, it is possible to link legal strategies, which are entirely possible with profiles and gaps in systems. [12] Profiles based on the Markov hidden model have addressed many of the concerns associated with standard profile analysis. HMMs offer a consistent theory of notching and subtraction, as well as a consistent structure for compiling data structure and sequencing. HMM based on multiple sequence programs is immediately filtered. HMM-based Homologic recognition has a pre-existing influence on HMM techniques to be positively correlated with more complex protein-binding mechanisms [8].

III. PROBLEM FORMULATION

This section explains the definition of problem-solving.

As we know each person has their own characteristics and each engineer uses his own method and techniques to improve the product. One facilitator can compile the speech correctly, another can perform tasks more accurately and without further explanation and correction, others are able to perform a smaller range of tasks, but more accurately and as the user wishes.

Therefore, no such helper can do all the work and work equally well. The set of features a helper has depends on the location where the developer pays more attention. Since all systems are based on machine learning and application in the creation of large amounts of data collected from various sources and trained in them, an important role is played by the source of this data.

Aside from the different ways of learning different algorithms, the goal of building a voice assistant remains the same. Technologies used to create a voice assistant that can communicate with people to recognize speech, Teach-Speaking, voice biometrics, chat manager, natural language comprehension and named entity recognition.

VOICE TECHNOLOGY	BRAIN TECHNOLOGY
Voice Activation	Voice Biometrics
Automatic Speech Recognition (ASR)	Dialog Management
Teach-To-Speech (TTS)	Natural Language Understanding (NLU)
	Named Entity Recognition (NER)

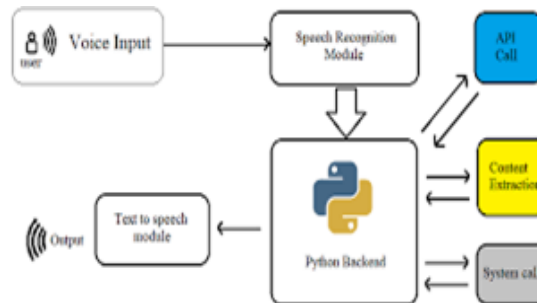
IV. PROPOSED APPROACH

The proposed system will have the following functions:

- (a) The system will continue to obey the instructions and the listening time is flexible which can be adjusted according to the needs of the user.
- (b) If the system is unable to collect information on user input it will continue to request repetition until the desired number of times.
- (c) The system may have male and female voices according to user requirements.
- (d) Features based on current version include playing music, emails, texts, searching on Wikipedia, or opening installed apps, opening anything in a web browser, etc.
- (e) The system will continue to obey the instructions and the listening time is flexible which can be adjusted according to the needs of the user.

(f) If the system is unable to collect data from user input it will keep asking for a repeat until the desired no. of times.

(g) The system may contain both male and female voices according to user requirements [9].



V. RESULT AND ANALYSIS

This section describes a brief description of our outcome on the basis of a comparison and analysis of our proposed work. We have used this concept with Python, Machine Learning and AI. Our main goal is to assist users in their tasks with the help of their voice commands. This can be done in two stages. First, take the audio input from the user and convert it to an English phrase with the help of the Speech Recognition. Second is to search for the user of the task they want to perform and redirect it to the server with the help of HTTP Protocol and display the result in a web browser.

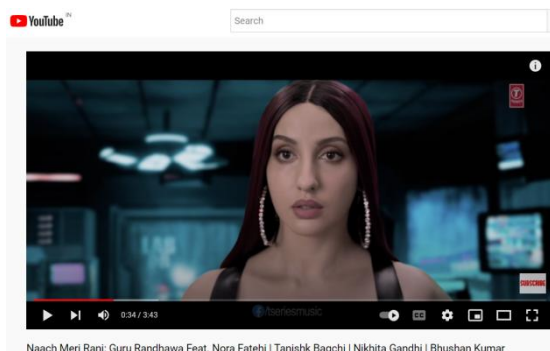
When the Windows Code is executed the first output which will be displayed is to start listening. After the user has to the voice command.

```
C:\Users\dev\Desktop\alexa\venv\Scripts\python.exe C:/Users/dev/Desktop/alexa/main.py
Listening...
hello
hello
Listening...
```

This screen will be visible when user has given voice command and the Google Speech Recognition has translated it into an English Phrase.

```
C:\Users\dev\Desktop\alexa\venv\Scripts\python.exe C:/Users/dev/Desktop/alexa/main.py
Listening...
open youtube and play naach meri rani
Listening...
```

After translation the command which the user has given will be displayed on the web browser.



VI. CONCLUSION

In this paper we have discussed the Voice Assistant built using an anaconda. This assistant currently works as a based app and performs basic functions such as weather updates, streaming music, Wikipedia search, open desktop apps, etc. Current system performance is limited to app operation only. Upcoming updates for this assistant will have machine learning included in the program that will lead to better suggestions for IoT to control nearby devices similar to what Amazon Alexa is doing.

VII. REFERENCES

- [1] 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 Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.
- [2] B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 3, pp. 201–212, 1976.
- [3] V. Radha and C. Vimala, "A review on speech recognition challenges and approaches," *doaj. org*, vol. 2, no. 1, pp. 1–7, 2012.
- [4] T. Schultz and A. Waibel, "Language- independent and language adaptive acoustic modeling for speech recognition", *Speech Communication*, vol. 35, no. 1, pp. 31–51, 2001.
- [5] J. B. Allen, "From lord rayleigh to shannon: How do humans decode speech," in *International Conference on Acoustics, Speech and Signal Processing*, 2002.
- [6] 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 Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.
- [7] 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] S. R. Eddy, "Hidden Markov models," *Current opinion in structural biology*, vol. 6, no. 3, pp. 361–365, 1996.
- [9] Excellent style manual for science writers is "Speech recognition with flat direct models," *IEEE Journal of Selected Topics in Signal Processing*, 2010.
- [10] Srivastava S., Prakash S. (2020) Security Enhancement of IoT Based Smart Home Using Hybrid Technique. In: Bhattacharjee A., Borgohain S., Soni B., Verma G., Gao XZ. (eds) *Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science*, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_44
- [11] S. Srivastava and S. Prakash, "An Analysis of Various IoT Security Techniques: A Review," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 355-362, doi: 10.1109/ICRITO48877.2020.9198027
- [12] Saijshree Srivastava, Surya Vikram Singh, Rudrendra Bahadur Singh, Himanshu Kumar Shukla, "Digital Transformation of Healthcare: A blockchain study" *International Journal of Innovative Science, Engineering & Technology*, Vol. 8 Issue 5, May 2021