Jarvis Voice Assistance

SAI MADHAVI D[1], SUDARSHAN REDDY R[2], SHIVAKUMAR MEDA[3], SIDDESH GODINAL[4], HARSHIT B[5]

[1]*Professor & Head of the Department of cse(Ai&Ml), Rao Bahadhur Y. Mahabaleshwarappa Engineering College, Bellary, Karnataka, India*

[2,3,4,5] *Department of computer science and engineering, Rao Bahadhur Y. Mahabaleshwarappa Engineering College, Bellary, Karnataka, India*

Abstract— Developing a personal assistant is the primary objective of the project especially in view of windows-based operating system. Personal assistant namely JARVIS has been an idea inspired from existential virtual assistance like Cortana and Siri for the respective windows and iOS platforms, which carries out various tasks using the command line making it user friendly interface. Synchronous process which has speech recognition and processes and analyze the artificial speech dependent on humans as the base in this personal voice assistance, this software is capable of performing various tasks and provide services to the end user according to the requirement. Few purposes of the assistant developed, was to adapt automation to perform certain tasks having a wide variety beginning from mailing to management of tasks. With the development of technology human dependency on the computers as increased drastically, people or users want their gadgets to be easy and more practical hence there has been a shift even in providing commands or input from text to speech. This shift has marked the beginning of speech recognition. Thereby the need for not just taking the input through speech arose even performing the task and command trough speech recognition has been the idea behind the Personal voice Assistant.

I. INTRODUCTION

Everything that a human being can do is being transformed by machines in this era. One of the key causes is the performance exchange. In today's society, we teach our machines to think like people and complete their missions on their own. As a result, the concept of a digital assistant emerged. A digital assistant that understands voice commands and does relevant tasks as requested by the user using voice reputation features and language processing algorithms. A digital assistant is capable of filtering out ambient noise and returning relevant facts based on particular commands supplied by the person. Virtual Assistants are entirely software-based, but they've recently been integrated into a variety of devices, and some, such as Alexa, are created specifically for standalone devices Because of the rapid change in era, it is now more important than ever to educate our machines using machine learning, deep learning, and neural networks. With the help of Voice Assistant, we can now speak to our machines. Nowadays, every significant company uses Voice Assistant so that their customers may use their voice to interact with their devices. As a result, with the Voice Assistant, we're progressing to the next level of development, where we'll be able to converse with our device.

These types of digital assistants are extremely beneficial to the elderly, the visually and physically challenged, children, and others by ensuring that interacting with the system is no longer a difficult task for humans. Even blind people who are unable to see the system can engage with it by using their voice.

Here are a few of the most important tasks that can be completed with the help of a voice assistant: -

1. newspaper reading
2. receive email updates
3. search the internet
4. Listen to a song or watch a video.
5. Setting an alarm and a reminder
6. Start any program or utility.

These are just a few samples; we will carry out many more tasks in accordance with our requirements.

The Voice Assistant that we've developed is suitable for both Windows and Linux users. The voice assistant we've developed is a computer-based gadget built with Python modules and libraries. This assistant is only a basic model that can execute all of the basic tasks listed above, but modern generation, while true in

some ways, still has to be combined with gadget mastery and the internet of things (IoT) for greater results We used python modules and libraries to create the version, and we used device learning to school our version. Some windows and Linux instructions were also added to the version so that it might operate smoothly on this running computer. Our approach will essentially work in three modes: -

1. Supervised mastering
2. Getting to know each other without being supervised
3. Reward-based learning

Depending on the purpose for which the aid is required for the individual. These can be accomplished with the help of machine learning and deep learning. With the help of Voice Assistant, there may be no need to write down instructions for completing a specific activity over and over again. Once a model is built, it can be utilised an unlimited number of times by an unlimited number of users in the most basic of ways. As a result, with the help of a virtual assistant, we will be able to control numerous things around us on a single platform.

1.1. Speech Recognition :

Speech recognition, frequently known as speech- to-textbook, is the capacity of a machine or program to fete and transfigure spoken words into comprehendible textbook. The vocabulary of rudimentary voice recognition software is confined, and it can only fete words and rulings when pronounced easily. More advanced software can deal with natural speech, multitudinous accentuations, and several languages Computer wisdom, linguistics, and computer engineering exploration are all used in speech recognition. Speech recognition functions are included into numerous current widgets and textbook-concentrated program to make using them easier or hands-free. Speech and voice recognition are two distinct technologies that mustn't be confused

- Speech recognition is a technology that recognizes words in spoken language.
- Voice recognition is a biometric technology for relating an existent’s voice.

Speech recognition systems process and assay spoken words before converting them to text using computer algorithms. Following these four processes, a software programme converts the sound a microphone records into written language that computers and humans can understand

1. dissect the audio;
2. break it into corridor;
3. digitize it into a computer- readable format; and
4. use an algorithm to match it to the most suitable textbook representation.

1.2 Artificial-intelligence:

Artificial intelligence (AI) refers to a computer's or a computer-controlled robot's ability to negotiate tasks normally performed by intelligent beings.The expression is considerably used to relate to a design aimed at creating systems with mortal- suchlike cognitive capacities, analogous as the capability to reason, discern meaning, generalize, and learn from formerly exploits. Since the invention of the digital computer in the 1940s, it has been proved that computers can be programmed to perform extremely complicated jobs with ease, analogous as chancing evidences for fine theorems or playing chess. Despite ongoing increases in computer processing speed and memory capacity, no program can yet match mortal severity across broader fields or in conditioning taking a great deal of common knowledge. still, certain programes have surpassed the performance situations of mortal specialists and professionals in executing specific tasks, and artificial intelligence in this limited sense can be set up in operations as different as medical opinion, computer quest machines, and voice or handwriting recognition.

1.3 Virtual assistance:

A virtual adjunct is an independent contractor that works for a customer and provides executive support while working from a position other than the client's office. A virtual adjunct generally works from home, but may pierce important planning accoutrements similar as participated timetables from anywhere. Virtual sidekicks constantly have times of experience working as an executive adjunct or office director. Virtual sidekicks with chops in social media, content operation, blog post jotting, graphic design, and online marketing are chancing new jobs. The demand for educated virtual sidekicks is projected to rise as

working from home becomes further accepted by both workers and businesses.

- A virtual assistant is a self-employed professional who provides administrative help to clients from a remote location, usually a home office.
- Scheduling appointments, making phone calls, planning vacations, and managing email accounts are all common responsibilities of a virtual assistant.
- Graphic design, blog authoring, bookkeeping, social media management, and marketing are some of the specialties of virtual assistants.

One benefit of hiring a virtual assistant for an employer is the ability to contract for only the services they require.

As small businesses and startups rely on virtual offices to cut expenses, and organizations of all sizes boost their use of the internet for day-to-day operations, virtual assistants have become more popular. A virtual assistant is an independent contractor, therefore a company does not have to give the same benefits or pay the same taxes as a full-time employee.

II. LITREATURE SRUVEY

1.B.A.R.AL-Hashemy and S.M.R.Taha demonstrate how to use a sample recognition method to determine whether a given phase of a speech signal is voiced. depending on measurements taken at the sign, unvoiced words, or stillness Five one-of-a-kind measurements are taken on the speech phase to be categorised in this technique. The speech phase is assigned to a certain elegance using a minimal-distance criterion, which is based on the assumption that the measured parameters are distributed in a multi-dimensional Gaussian probability density function. Two statistical approaches are used to calculate the method and covariances for the Gaussian distribution. With speech samples as short as 10ms, the approach has been proven to give reliable classification.

2.Tanja Schultz and Alex Waibel provide an explanation. With speech technology solutions being distributed all over the world, portability to new target languages has become a practical concern. As a result, our research is focused on how to quickly and efficiently port large vocabulary continuous speech recognition (LVCSR) systems. We wish to use voice data from a variety of source languages to estimate acoustic models for a new target language, but have limited data from the target language. We offer various ways for multilingual acoustic model combining as well as a polyphone decision tree specialization procedure for this purpose. In the context of our GlobalPhone project, which examines LVCSR systems, recognition results utilizing language dependent, independent, and language-adaptive acoustic models are shown and analyzed.

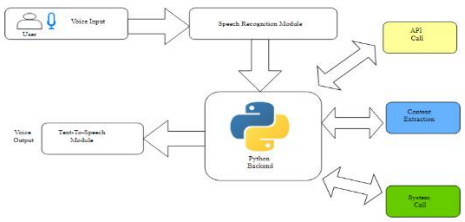
3. Speech, or verbal communication, is one of the most essential qualities that distinguishes humans from other animals, according to Lehlohonolo Mohasi and Daniel Mashao. Researchers in voice technology are still trying to figure out how to get machines to communicate with humans in the same way that humans do. Human-computer interaction (HCI) is a field that studies the development, Interactive computer systems for human use: assessment and implementation [2]. This article looks into how text-to-speech technology is used and how to make it more acceptable and user-friendly on an interactive basis. This can be accomplished by creating a design that considers user expectations. Evaluation is also necessary to obtain input and determine whether the design fits the user's expectations. A conclusion is reached as a result of this process.

4. Foreign Language Learning (FLL) students often have few opportunities to utilise their target language, according to Luke Fryer and Rollo Carpenter. Teachers in FLL scenarios try their hardest to generate opportunities during class through pair or group work, but a range of issues, from a lack of time to shyness or a lack of quality feedback, thwart this. The possible role of online chatbots in meeting this requirement is discussed in this study. Students could use chatbots to practise their language skills at any time and from nearly anyplace.

5. Ming Zhou, Jizhou Huang, and Dan Yang explained a This study describes a novel method for extracting high-quality pairs as conversation knowledge from online discussion forums in order to aid in the development of a chatbot for a certain domain. A

cascaded architecture is used to extract high-quality pairs from a forum. First, an SVM classifier is used to extract the replies logically relevant to the thread title of the root message from all the replies, based on correlations such as structure and content. After that, a ranking SVM is used to rate the extracted pairs based on their content attributes. Finally, the Top-N pairs are chosen as knowledge for chatbots. Experiments conducted in a movie forum have shown that the proposed method is effective.

III. METHODOLOGY



A. Speech Recognition :
The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information center from the microphone is temporarily stored in the system which is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor.

B. Python Backend :
The python backend parses the voice recognition module's output to determine whether the command or speech output is an API Call, Context Extraction, or System Call. The output is then sent back to the python backend to provide the user with the desired results.

C. API Calls :
The acronym for Application Programming Interface (API) is Application Programming Interface. An API is a software interface that allows two apps to communicate with one another.To put it another way, an API is the messenger that sends your request to the provider and then returns the result.

D. Context Extraction :
Context extraction (CE) is the process of obtaining structured data from machine-readable materials that are unstructured or semi-structured. The majority of the time, this activity entails using natural language processing to process human language texts (NLP). TEST RESULTS for context extraction can be seen in recent activities in multimedia document processing, such as automatic annotation and content extraction from images/audio/video.

E. System Calls :
The mechanism through which a computer software requests a service from the kernel of the operating system on which it is running is known as a system call. Hardware-related services (for example, accessing a hard disc drive), the creation and execution of new processes, and communication with core kernel services such as process scheduling are all examples of this. A process's interface with the operating system is provided by system calls.

F. Text-To-Speech :
The capacity of computers to read text aloud is referred to as text-to-speech (TTS). Written text is converted to a phonemic representation, which is subsequently converted to waveforms that can be generated as sound by a TTS Engine. Third-party publishers offer TTS engines in a variety of languages, dialects, and specialist vocabularies.

IV. RESULTS

Speech recognition software uses natural language processing(NLP), Speech recognition breaks down into bits it can interpret, converts it into a digital format, and analyzes the pieces of content. From there, the software makes determinations based on programming and speech patterns, after determining what users said, the software transcribes the conversation into text, the text is saved in the variable, as our voice assistance is text dependent the , the given command in the form of text searches for the similar text written in the form of functions, if the text is matched it performs the particular task, if not found it throws an exception.

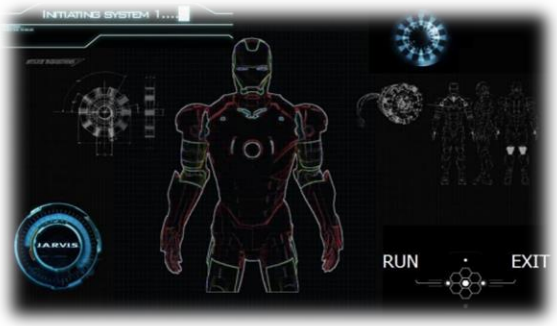
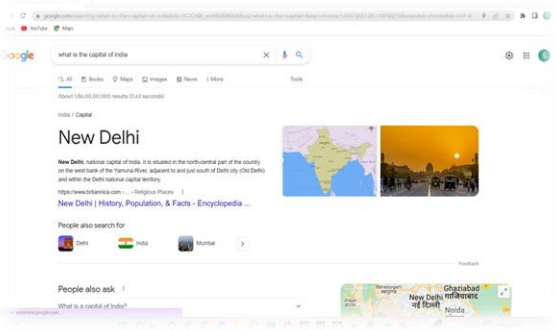


Fig.2.Jarvis main UI



Fif.3.Input:Open Google

V. CONCLUSION

In this paper we have discussed about Jarvis voice assistance for windows using python. Jarvis voice assistance makes life easier to humans. As like Google assistance and Cortana we make Jarvis voice assistance to be available to all the windows version, We use Artificial intelligence technology for this project, Jarvis voice assistance be able to do all the tasks like other assistance including some special functions like restarting the devices, locking the device, sleeping the device for some particular time and shutdown the device with our voice input. We can expect this Jarvis voice assistance to be permanent.

FUTURE WORK

Future work includes Making the voice assistance to work on all the versions of the python in all the windows version and Linux versions and including some function that helps in doing payments while online shopping.

REFERENCES

[1] Vishwakarma, Aditya, Ankit Sahu, Nadeem Sheikh, Peeyoosh Payasi, Saurabh Kumar Rajput, and Laxmi Srivastava. "IoT based Greenhouse monitoring and controlling system." In *2020 IEEE Students Conference on Engineering & Systems (SCES)*, pp. 1-6. IEEE, 2020.

[2] Nivedita Singh, Dr. Diwakar Yagyasen, Mr. Surya Vikram Singh. Gaurav Kumar, Harshit Agrawal Department of CSE. Babu Banarasi Das National Institute of Technology and Management, Lucknow India.

[3] Monika Raghorte, 4Aishwarya Bhisikar, 5Anup Bhange 1Student, 2Student, 3Student, 4Student, 5Assistant Professor 1Computer Technology, 2Computer Technology, 3Computer.

[4] G. Bohouta, V. Z. Këpuska, "Comparing Speech Recognition Systems (Microsoft API Google API And CMU Sphinx)", Int. Journal of Engineering Research and Application 2017, 2017.

[5] Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior, 49, pp.245-250.

[6] Huang, J., Zhou, M. and Yang, D., 2007, January. Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI(Vol. 7, pp. 423-428).

[7] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp. 79-84).

[8] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology