

# Voice Assistant Using Artificial Intelligence

1Indukuri Manikanta Sai Varma, 2Kalidindi Pranith Varma, 3Karri Gnana Keerthana, 4Siraparapu Neeraj Kumar, 5Sadi Ram  
Prasad Reddy  
1234Student, 5Associate Professor  
Anil Neerukonda Institute Of Technology And Sciences

**Abstract -** Every procedure is now digitalized. We are already familiar with several voice based assistants, such as Google Assistant and Siri. This voice-assistance device serves as a basic medical prescriber, daily calendar reminder, note writer, calculator, and search engine. This project accepts voice input and outputs text through speech, as well as displaying it on the screen. Our primary goal with our voice assistance is to make folks smarter and to give rapid and measured results. The voice assistance system receives voice input via our microphone (Bluetooth or wired) and converts it to computer-comprehensible English. It then provides the user with the required solutions and replies. This assistance connects with the internet in order to provide answers to the user's inquiries. Genuine Language Processing algorithms enable computers to engage in conversation using a variety of kinds of natural human language. Nowadays, python is the most widely used language for artificial intelligence due to its plethora of AI packages; so, in this study, the speech recognition module (python) is critical for the conversion of voice to text.

**keywords -** Artificial intelligence, natural language processing, python, voice assistant and Bluetooth

## I. INTRODUCTION

Natural language processing and human-machine interaction are the hottest topics in technology right now. One of the most explored and popular directions of interaction was based on the machine's interpretation of real human language. No longer does a human learn to speak with a machine; rather, a computer learns to interact with a human by observing his actions, habits, and behavior and attempting to become his customized assistant. Virtual assistants are computer applications that aid you with daily duties such as displaying weather data, preparing reminders, and building grocery lists, among others. They are capable of receiving commands via text (online chat bots) or voice.

Intelligent assistants that use voice commands require an inciting phrase or wake word before the listener is activated. So many virtual assistants are available to us, such as Siri, Alexa, and Cortana from Apple, Amazon, and Microsoft, respectively. This system was created with desktops in mind. User productivity is increased by personal assistant software, which takes care of the user's everyday duties and feeds them data from an internet source. A virtual assistant with the ability to make intelligent judgments about normal user behaviours may be built using publicly available data and information, which is why this project was begun in the first place. The creation of the intelligent assistant is affected by the methodologies and approaches used by each company's developer. While a few assistants are able to execute a wide range of activities flawlessly, others are more limited in scope but are nonetheless capable of completing them precisely and to the user's satisfaction. No matter how hard you try, you'll never be able to find a universal helper who can do everything. The attributes of an assistant are determined solely by the developer's focus on a certain area. Due to the fact that all systems employ machine learning methods and rely on large quantities of data acquired from a variety of sources and then trained on them, the source of this data plays a significant role. The assistant's personality is shaped by the quantity of data it has access to from various sources. Building such systems, despite the variety in methods to learning as well as algorithm and technique, has remained largely unchanged over time. A human-to-human communication system based on a person's natural language. These include voice activation, speech recognition, Teach-To-Speech, voice biometrics, conversation management, natural language comprehension, and named entity recognition.

## II. LITERATURE SURVEY

The "Text-to-speech technique in human-computer interaction" was tested by Lehlohonolo Mohasi et al. This system has been implemented using text-to-speech and human-computer interaction approaches.

"The voice enabled personal assistant for PC using python" was noticed by Geetha.v et al. They've employed text to speech, speech recognition, conversational interfaces, and chat bots to do this. To converse with them, they've created a personal assistant.

"AI-Based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives" by Farzaneh nasirian et al. They have examined the quality of interaction, trust, and technological adoption in this location.

Tae-kook kim et al. Conducted a "Short research on voice control system based on Artificial Intelligence assistance". Here they have done their research using IFTTT, Raspberry pi, and Google assistant.

Lanu Xu et al. Conducted a research on "CHA: A Caching Framework for Home-based Voice Assistant Systems". They have done this using caching, Edge computing and voice assistant system.

### III. METHODOLOGY

#### 3.1 Proposed system

The proposed system can able to launch the applications directly by the voice commands. The proposed system also able to send E-Mails and whatsapp messages. It has ability to read out the Documents and accessing and Modifying the Files in the system. In this paper python is used for speech recognition module. There are some important modules incorporated in this system.

##### 3.1.1 Pywhatkit module

Python has a wide range of built-in libraries that make our work easier. pywhatkit is a Python module that, among other things, allows users to schedule the delivery of WhatsApp messages.

Python's pywhatkit module has the following features:

1. Send messages with WhatsApp.
2. a link to a video on YouTube will be played.
3. Conduct a search on Google.
4. Find out more about a specific subject. In order to create handwritten text graphics, the pywhatkit module may be utilised.

##### 1. Send Whatsapp Messages:

The pywhatkit module, which uses the WhatsApp web in the browser and able to send the messages to any of the whatsapp registered mobile number .As long as the browser is up and running ,you're good to go.

##### 2. Play a YouTube video:

When you use the function pywhatkit.playonyt(), YouTube is opened in your default browser and the video you specified is played. Plays a random video based on the topic name if the option is supplied. As a parameter, you can provide the URL of the video you want to watch.

##### 3. Google Search:

In order to conduct a Google search, simply enter the following command into your terminal. Your code instructs it to open your browser and search for the topic you specified.

##### 4. Getting Information

We can acquire a short overview of a specific issue. We may also limit the amount of lines that appear on the screen at once. Also, be careful to limit your search to Wikipedia-available subjects.

#### 3.1.2 SMTPLIB

Mail is sent and received between mail servers through the Simple Mail Transfer Protocol (SMTP).Any Internet system having an SMTP or ESMTP listener daemon can receive mail from the smtplib module, which specifies an SMTP client session object.

#### 3.1.3 Python Sending Email using SMTP

Python's email transfer mechanism makes use of SMTP, or Simple Mail Transfer Protocol. Emails are routed through it between different email servers. Sending email is made possible by this application layer protocol. POP (Post Office Protocol) and IMAP (Internet Message Access Protocol) are used by the recipient to retrieve email (Internet Message Access Protocol).

#### 3.1.4 Functools module in Python

Functools module is for higher-order functions that work on other functions. It includes functions for interacting with other functions and callable objects so that they can be used or extended without having to be totally rewritten.. This module has two classes –in which one is partial and the other one is partialmethod.

#### 3.1.5 Partial class

A partial function is a function that only applies to a specific set of arguments. The functools library's "partial" may be used to generate them in Python. Because these properties are not formed automatically, the programmer is responsible for creating them. The three read-only characteristics of partial() objects are:

#### 3.1.6 Tkinter Programming

Tkinter is the Python GUI library. The combination of Python with Tkinter makes it simple to develop graphical user interfaces. The Tk GUI toolkit's object-oriented interface is provided by Tkinter.

- Tkinter makes it simple to create a graphical user interface for your programme. All you have to do is follow these instructions.
- The Tkinter module should be imported.
- Create the primary window for the GUI programme.
- Incorporate one or more of the aforementioned widgets into your GUI application's interface.
- Activate the main event loop to respond to each user-initiated event.

#### 3.1.7 Python | Text to Speech by using pyttsx3

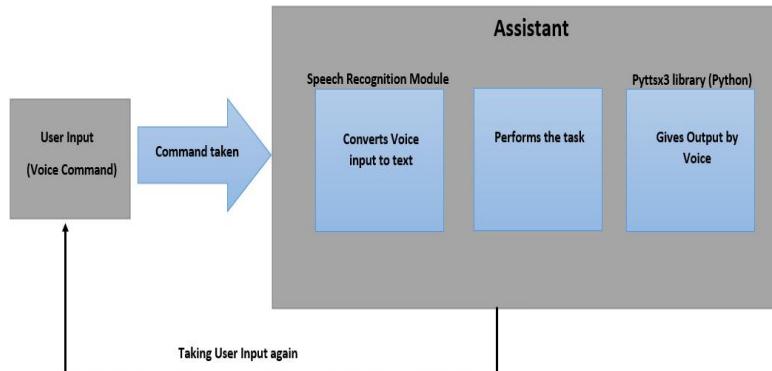
A Python package for converting text to spoken language is called TTS. It works offline, unlike other libraries, and its backwards compatible with Python 2 and 3. This factory method is used by an application to obtain a pyttsx3 Engine object instance. It's a simple-to-use programme that turns any text you type into spoken word.

Using the pyttsx3 module, two voices may be played simultaneously. "sapi5" for Windows has a feminine and a male version.

### 3.2 System design and architecture

#### 3.2.1 Architecture of the proposed system

Analyzing the user's microphone commands was the first step in the process. This might include everything from retrieving data to managing the contents stored within a computer. Reading and testing the cases from the literature cited above, this is an empirical qualitative research. In order to better comprehend the Voice Assistant depicted in Fig. 1, tests are carried out by programming in accordance with books and internet resources. Three components make up this Assistant. First, the user's voice is accepted by the assistant. Secondly, analysing the user's input and translating it to the appropriate function. Secondly, analysing the user's input and translating it to the appropriate function and intent. And the third is that the assistant provides the user with the outcome through voice as well.

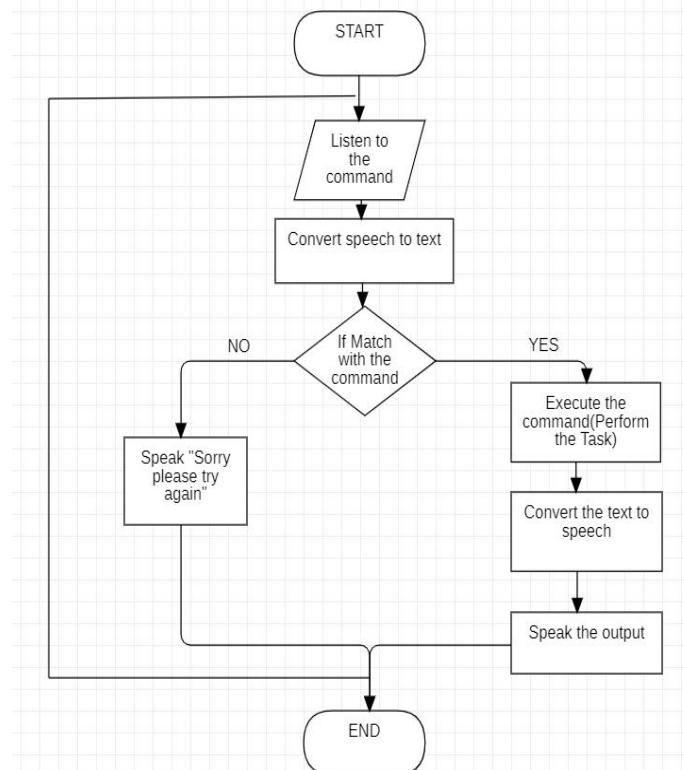


*Fig.1 Architecture diagram*

The fundamental method of the voice assistant is depicted below. To translate spoken input into text, voice recognition software is employed. The central processor uses this text to assess the command's type and then summons the appropriate script to execute it. However, there is much more to it than that. No matter how much time you put in, the programme may still not be able to comprehend what you're saying.

A speech-recognition system can be thrown off course by background noise. You can't tell a dog's barking or a helicopter flying overhead from your voice since it doesn't have the capacity to discern between them. Engineers have to "teach" the gadget to filter out these background noises by collecting data and programming it to do so. Another element is the way people instinctively alter their voice pitch in order to adapt to a loud environment; speech recognition algorithms are capable of detecting these pitch variations.

#### 3.2.2 System flow chart



*Fig.2 System flow chart diagram*

The system pops up the GUI and start with the listening mode. The system recognize the user's voice input and convert the speech to the text by the speech recognition module. Once the speech is converted to the text , then the system verify whether that the command is associated with any action/task.

If the command is associated with some task ,then the system will execute the task for that certain command.For example if the command is to search something in Wikipedia ,then the output for that search is converted from text to speech by the Pytsxs3 library in Python.And the system will speak the output. As shown in the fig 2

If the command is not associated with any of the task, then the system will reflect a speech message “Sorry,Please try again” . System will be redirected to the listening mode and continue the process

## IV. IMPLEMENTATION

### 4.1 Speech recognition module

To translate spoken input into text, the system makes advantage of Google's online voice recognition technology. The audio input When a user speaks into the microphone at an information centre, the system stores the audio and sends it to Google Cloud for voice recognition, where it is transcribed into text. The central processor then receives and processes the corresponding text.

### 4.2 Python backend

After receiving the speech recognition module output, the python backend determines if the command or voice output is an API Call and Context Extraction by comparing it to the context of the command. Finally, the python code is run on the user's computer to produce what they need.

### 4.3 API Calls

Application Programming Interface is the acronym for API. Two programmes can communicate with each other using an API, a software mediator. As a messenger, an API delivers your request to the service provider you're seeking it from and then returns the result to you.

### 4.4 Content Extraction

Extracting structured information from unstructured and/or semi-structured machine-readable materials is the goal of context extraction (CE). Most of the time, this involves employing natural language processing to process texts in human languages (NLP). Context extraction TEST RESULTS from recent multimedia document processing operations such automated annotation and content extraction from photos, audio, and video may be viewed.

### 4.5 Speech—text tool

A computer's capacity to read text out loud is known as Text-to-Speech (TTS). With a TTS Engine, printed text is transformed into phonetic waveforms that may be played back as audio. Third-party publishers offer TTS engines in a variety of languages, dialects, and specialist vocabularies.

### 4.6 Features

1. Messages may be sent by voice over Whatsapp.

When we initially enter a user's name and phone number, the information is stored in a database. When we speak a person's name, we should remember that the database already has that person's name. To avoid asking for a phone number, instead it will ask for a message and send it to that person's phone number as seen in Fig.3

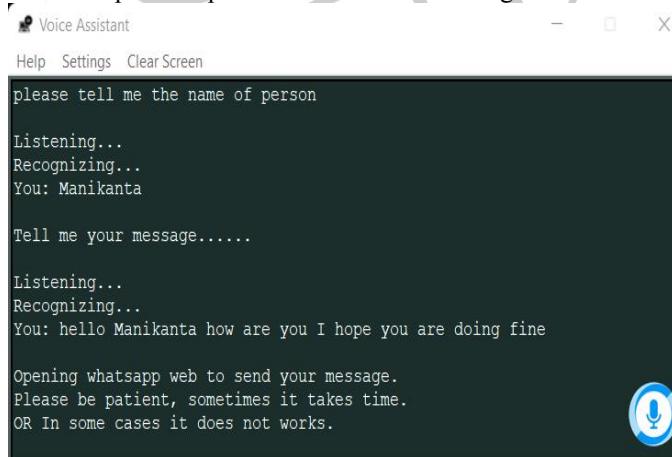


Fig.3 Whatsapp message with voice

2. Send an email by your choice – Voice or Type.

Whether you type or speak, it will ask you. The email.exe file, which was created using python code, will be launched if we select typing. Receiver email, topic, content, and send button will be shown in the gui. Once you've entered your information and clicked "Send," the email will be delivered to the address you provided, as shown in Fig. 4.

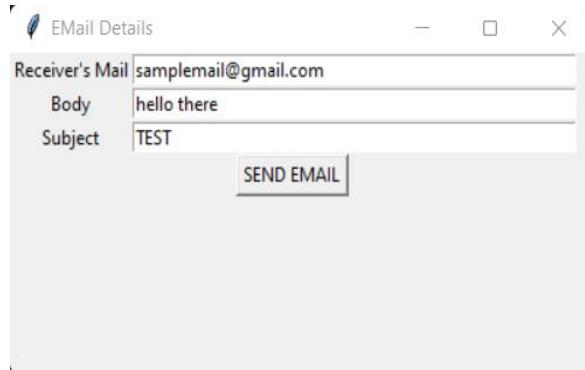


Fig.4 Email by Type

### 3. Record voice with a command

It can record your voice with a voice command as given in the Fig.5.

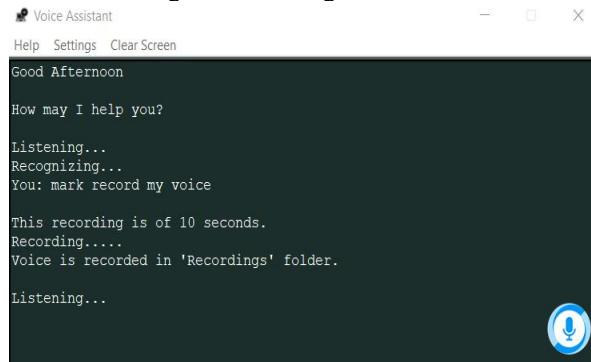


Fig.5 Record voice with a command

### 4. A Note to Yourself

Make a note if you ask your assistance to do so. It will then begin listening to your speech and converting it into text and pasting it into a notepad, as seen in the figure 6.

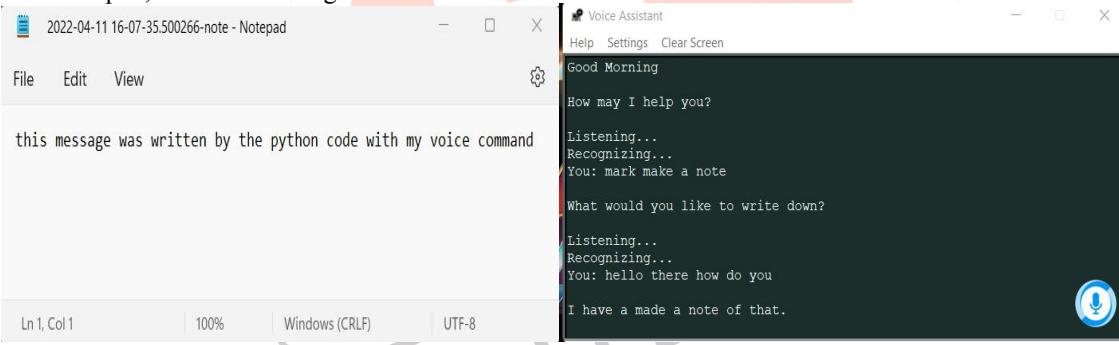


Fig.6 Make a note

### 5. Voice calculator

It can either do mathematical calculations using voice command as given in the Fig.7.

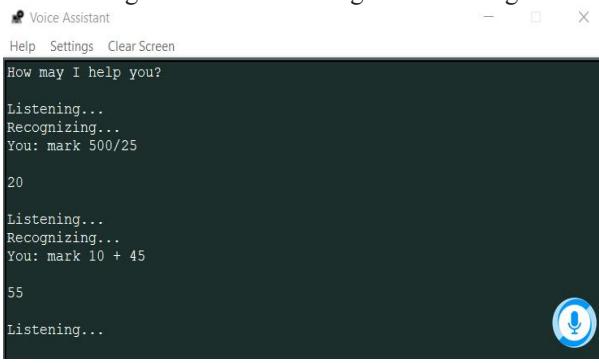


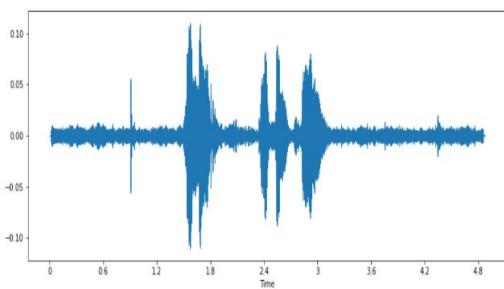
Fig.7 Voice calculator

## V. RESULT AND DISCUSSION

The two different speakers with the command “hello open chrome” is given as the input .For this a male and a female voice is taken as the input with the same command with duration of 4 Sec.

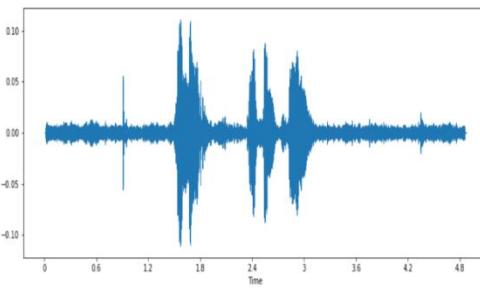
The male input voice signal is shown in the figure 8 and female voice input signal is shown in the figure 9.

▶ 0:00 / 0:04 ━━ ⏪ ⏴



*Fig.8 Male Voice Input Signal*

▶ 0:04 / 0:04 ━━ ⏪ ⏴

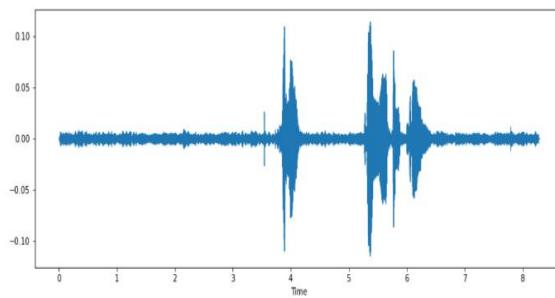


*Fig.9 Female Voice Input Signal*

Both the figures are the voice analysis performance evalution using MFCC. The problem with this approach is that when constant window spacing is used, the lengths of the input and stored sequences are unlikely to be the same. Furthermore, the length of individual words will vary within a word.

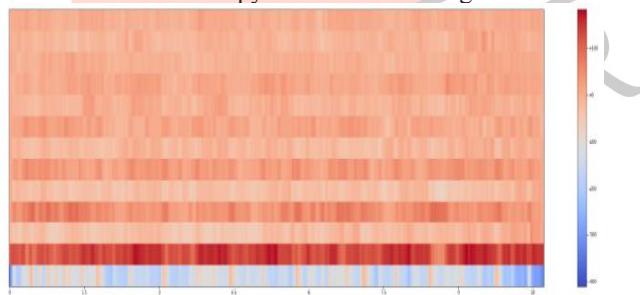
Voice input signal of a male voice with the command “hello can you open chrome”.This WAV format file is of 8 Sec and one can be able to see the difference in the wave as shown in the Fig 10.

▶ 0:08 / 0:08 ━━ ⏪ ⏴



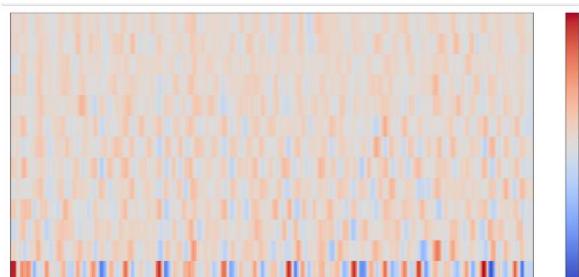
*Fig.10 Male voice input signal with different command*

Fig 11, 12, 13 Shows the Mel frequency cepstrum coefficients (MFCC) of the male voice command “hello open chrome” With the figure size of 25,10 using the librosa module in python for visualizing the MFCC is shown in the figure 11.



*Fig.11 Visualizing the MFCC*

Mel-frequency cepstral coefficients are a common input feature in recognition tasks like phoneme recognition and voice activity detection (MFCCs). They describe the speech signal's instantaneous spectral envelope shape delta 1 in Fig 12 and delta 2 in Fig 13.



*Fig.12 Computing first MFCCs derivatives*

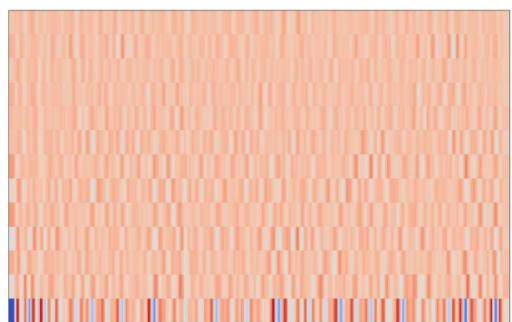


Fig.13 Computing second MFCCs derivatives

We have examined the responses to questions about the photosynthesis with the system that we developed. We posed the questions(commands) with 10 different members which included both the male and female voice .The system recognized the command and gave the relevant output for the 8 voice inputs.

The proposed system is also capable of performing the tasks with the accuracy of about 90%.As all the tasks like sending the messages, making note,playing music etc. can be able to perform by this system which is not possible the existing system like Cortana. The comparison chart in between both the system cortana and VA are shown in Fig 14,where VA represents the system that we developed.

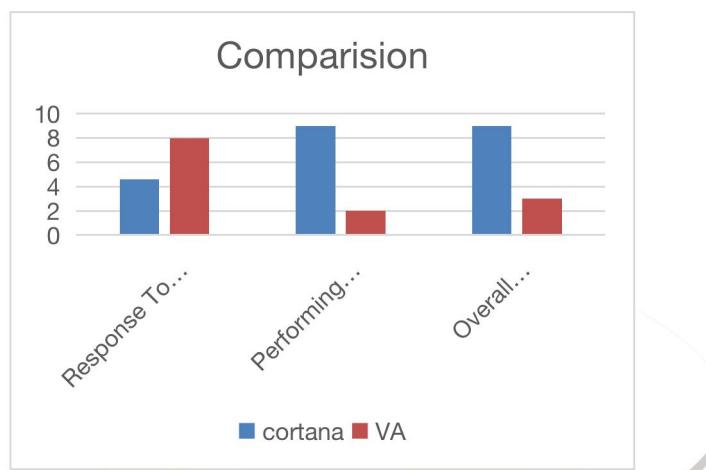


Fig 14 Comparision chart between Cortana and VA

User's input with noise would always be a difficult task for the system to recognize and process. Recognizing the user's input by speech, which it depends upon the hardware of the system which also include the quality of the microphone.

## VI. CONCLUSION

"Voice Assistant Using Python" covered the design and implementation of Digital Assistance in this document. Open source software components are used to build the project, and the PyCharm community is behind it, so it can handle any modifications quickly. Additional features may be added without affecting the present system's functionality because to its modular design. It not only responds to the user's questions and queries, but also opens tasks and processes based on the words said by the user. It's welcoming the user in a way that makes them feel at ease and gives them permission to engage with the voice assistant. In addition, the application should reduce the amount of physical labour required by the user to complete each job. The entire system is based on vocal input, rather than the next one that comes along.

## VII. REFERENCES

- [1] Mohasi L, Mashao D. Text-to-Speech Technology in Human-Computer Interaction. In5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) 2006 (pp. 79-84). [https://www.academia.edu/5906159/Text\\_to\\_speech\\_technology\\_in\\_human\\_computer\\_interaction](https://www.academia.edu/5906159/Text_to_speech_technology_in_human_computer_interaction)
- [2] Elshafei, Moustafa. (2022). Virtual Personal Assistant (VPA) for Mobile Users. [https://www.researchgate.net/publication/228748656\\_Virtual\\_Personal\\_Assistant\\_VPA\\_for\\_Mobile\\_Users](https://www.researchgate.net/publication/228748656_Virtual_Personal_Assistant_VPA_for_Mobile_Users)
- [3] T. -K. Kim, "Short Research on Voice Control System Based on Artificial Intelligence Assistant," 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1-2, doi: 10.1109/ICEIC49074.2020.9051160. <https://ieeexplore.ieee.org/abstract/document/9051160>
- [4] Maedche A, Legner C, Benlian A, Berger B, Gimpel H, Hess T, Hinz O, Morana S, Söllner M. AI-based digital assistants. Business & Information Systems Engineering. 2019 Aug;61(4):535-44. [https://www.researchgate.net/publication/333705636\\_AI-Based\\_Digital\\_Assistants\\_Opportunities\\_Threats\\_and\\_Research\\_Perspective](https://www.researchgate.net/publication/333705636_AI-Based_Digital_Assistants_Opportunities_Threats_and_Research_Perspective)
- [5] Xu L, Iyengar A, Shi W. CHA: A caching framework for home-based voice assistant systems. In2020 IEEE/ACM Symposium on Edge Computing (SEC) 2020 Nov 12 (pp. 293-306). IEEE.

<https://weisongshi.org/papers/xu20-CHA.pdf>

- [6] Shalini S, Levins T, Robinson EL, Lane K, Park G, Skubic M. Development and comparison of customized voice-assistant systems for independent living older adults. InInternational Conference on Human-Computer Interaction 2019 Jul 26 (pp. 464-479). Springer, Cham.  
<https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/72281/ShaliniShradhaResearch.pdf?sequence=1>
- [7] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas and B. Santhosh, "Artificial Intelligence-based Voice Assistant," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.  
<https://ieeexplore.ieee.org/document/9210344>
- [8] Audrey Desjardins, Afroditi Psarra, and Bonnie A. Whiting. 2021. Voices and Voids: Subverting Voice Assistant Systems through Performative Experiments. In Creativity and Cognition (C&C '21). Association for Computing Machinery, New York, NY, USA, Article 29, 1–10. <https://doi.org/10.1145/3450741.3466807>
- [9] Zwakman DS, Pal D, Arpnikanondt C. Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa. SN Comput Sci. 2021;2(1) 28. doi:10.1007/s42979-020-00424-4. PMID: 33458698; PMCID: PMC7798382. <https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC7798382&blobtype=pdf>
- [10] Natale S. To believe in Siri: A critical analysis of AI voice assistants.  
[https://www.kommunikative-figurationen.de/fileadmin/user\\_upload/Arbeitspapiere/CoFi\\_EWP\\_No-32\\_Simone-Natale.pdf](https://www.kommunikative-figurationen.de/fileadmin/user_upload/Arbeitspapiere/CoFi_EWP_No-32_Simone-Natale.pdf)

