Intelligent Voice Assisstant for Desktop using NLP and AI

Dr. Neha Agrawal[1] | Akash Roshan[2] | Meenu Garg[3]

[1]Ph.D, Mtech & Associate Professor, Information Technology, MAIT, Rohini, New Delhi, India
[2]Student, Department of Information Technology, MAIT, Rohini, New Delhi, India
[3]Ph.D pursuing, Mtech & Associate Professor, Information Technology, MAIT, Rohini, New Delhi, India

To Cite this Article

Dr. Neha Agrawal, Akash Roshan and Meenu Garg, "Intelligent Voice Assisstant for Desktop using NLP and AI", *International Journal for Modern Trends in Science and Technology*, 6(12): 328-331, 2020.

ABSTRACT

In the coming future, virtual assistants will automate nearly all of the manual and time-consuming tasks. This project is an implementation of an intelligent voice assistant for Windows which includes the functionality of security through facial recognition. Until this day, there has not been any good alternative for Windows, so this project aims to implement a voice assistant for the Windows platform while describing the difficulties and challenges that lies in this task.

KEYWORDS: *Virtual assistant, facial recognition, Google Assistant, Google Text-to-Speech API*

I. INTRODUCTION

This project is based on Windows application development using Python and provide personal assistant using voice recognition. This program includes the functions and services of: maintaining user's security using face recognition, mail exchange, WhatsApp message exchange, location services, music player service, checking weather, Google search engine, Wikipedia search engine and performing various other tasks likes opening an application like camera, taking screen shot, restart and shut down your machine. As it integrates most of the desktop services for daily use, it could be useful for getting a more convenient life and it will be helpful for those who have disabilities for manual operations. This is also a part of the reason why it's been chosen because the degree project.

The project originated from a popular application from Windows called "Cortana". This application was released on the date when Windows 10 was launched. The project emphasis on the Windows development by using user's voice command to get control over the machine for the basic functionalities that can be performed by the machine-like Google search and Google map, Wikipedia library and machine references ranging from Speech-To-Text, Text-To-Speech technology. As all these functionalities and services for the program have been explained, the main structure and construction of the project has been basically illustrated with its goals.

Apart from the project itself, there is also some investigation works on the existed products in this area and the tendency of the voice product, personal assistant developing. Two products were mainly investigated that are popular and representative, the Windows product of "Cortana" and the macOS product of "Siri". The main focus behind this investigation was to determine how those ideas originated; what functionalities and services they have; how they provide these services to the customers; test the product and related functions to get the architect, structure, logical algorithms of those products;

how they spread and promote in marketing and how they refine and upgrade the products from different versions.

II. RELATED WORK

Virtual assistants like Cortana and Siri are now a signature feature of machines. Mainly controlled by a user's voice, the digital assistant's primary ability to understand when its spoken to and understanding what's said is important.

1)Cortana

Cortana is a virtual assistant developed by Microsoft which comes standard on Windows machines, but it's also available for download on Android and iOS. Cortana is out there on the $199.95 Harma Kardon Invoke with Cortana speaker, and while there have been rumours of Cortana-enabled HP and Xiaomi speakers for quite a while, there is no concrete evidence that such speakers are still in the works. Whether Cortana are going to be available on other third-party speakers or not remains to be seen, but it's as if the Microsoft AI assistant is losing traction in the face of more sophisticated competition, a minimum for the buyer market.

At Microsoft Build 2018, the tech giant unveiled its long-promised integration with Amazon Alexa. The partnership between Amazon and Microsoft suggests it's going to be early to count Cortana out of the sport entirely, especially if it continues to focus totally on the enterprise market rather than home use.

2)Siri

You can access Siri on nearly any Apple device, including its line of laptops, desktops, phones and tablets, and smartwatches. Apple also sells its own speaker, the Home Pod ($349), and it looks like third-party access is on the way. Sonos, a company that produces and sells high-end home speakers, claims it will roll out a software update in July that will allow users to access Siri (in addition to Alexa and Google Assistant).

Siri offers excellent voice recognition and is sweet at picking what you say-especially when there is relaying some information to the user. Microsoft claims Cortana has a good understanding of context, but it struggles much more than Siri in actually hearing what you're saying.

In terms of response time to queries, Siri consistently provides fast replies to orders, but in case of Cortana, users have reported issues with Cortana understanding even basic requests despite no problems with their microphones, it's hard to measure response time to queries if they are unheard.

For searches and responding to queries for information, Cortana uses Microsoft's own Bing search engine whereas Siri uses Google.

III. PROBLEM STATEMENT

As this program includes the functions and services of: maintaining user's security using face recognition, mail exchange, Whats... es, music
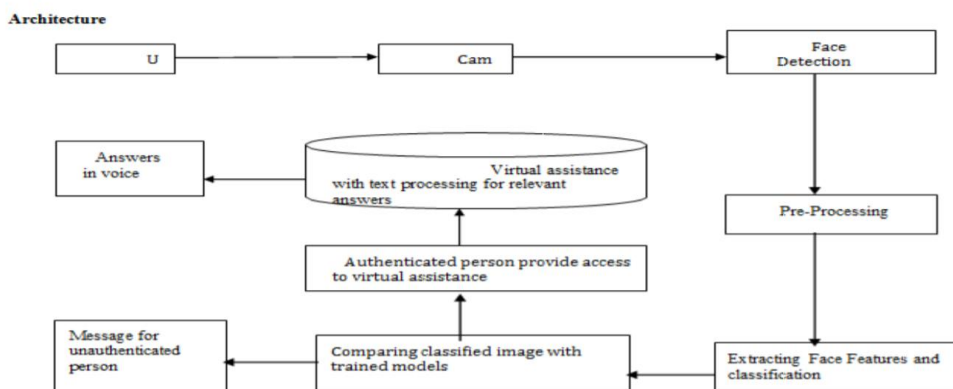
FACERECOGNITION:

Figure 1: Person Identification Virtual Assistant

- Google/Wikipedia search engine, the search engine enables the user to search anything on Google/Wikipedia.
- Camera, the camera function will call the camera on the device to take a picture of the current view.
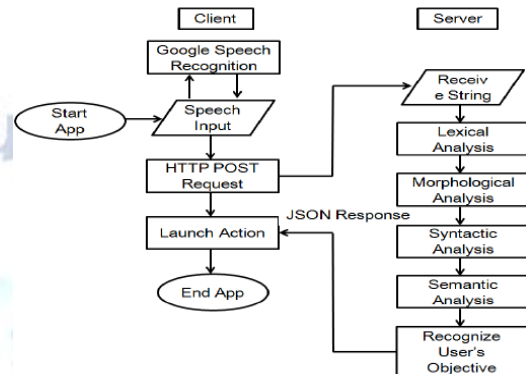

Figure 2: Data Flow Diagram of Speech Recognition

The whole construction of the program mainly cover development using Python libraries, OpenCV library to implement facial recognition, gTTS (Google Text-to-Speech) to interface with Google Translator's text-to-speech API and OpenCV (OpenSource Computer Vision) Library is an opensource computer vision machine learning software and has been added to the program to add the functionality of face detection. It is a cross-platform library used to develop real-time computer vision applications. Its main focus is image processing, video capturing and analysis includes features like face detection and object detection.

gTTS is Google Text to Speech API is a very easy to

Figure 2 Person Identification Virtual Assistant

which can be saved as an mp3 file. The gTTS API supports several languages including English, Hindi, Tamil, French, German and far of more.

checking weather, Google search engine, Wikipedia search engine and performing various other tasks likes opening an application like camera, taking screen shot, restart and shut down your machine.

- User's security, the application first detects whether the user currently giving commands is an authorised user or not using face recognition mechanism.
- Mail exchange, users are able to send the mail to person with mail address in the contacts. Providing request command contains the mail request keyword together with the destination person; the mail should be received by the recipient after it has been sent.
- WhatsApp messaging, users are able to send a WhatsApp message to a specific person in the contacts. This command is carried over Web- WhatsApp platform and hence user needs to ensure that Web-WhatsApp has already been logged in his device.
- Location services, Google maps is used to provide the functions for the user to check the current location or find the direction to a destination.
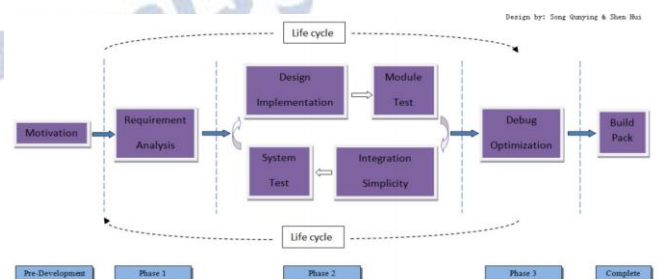
IV. RESULTS

Figure 3: Life Cycle of the proposed model

The Model and Flow Chart describes the development process that include all the phases in the software development life cycle. This chart illustrates very well how the project is carried out and how the development was managed.

The program should first be started on the Windows desktop then the user should have correct voice input "command/request" to form those functions work properly. As this program includes the functions and services of: maintaining user's security using face recognition, mail exchange, WhatsApp message exchange, location services, music player service, checking weather, Google search engine, Wikipedia search engine and performing various other tasks likes opening an application like camera, taking screen shot, restart and shut down your machine.

After the program is completed, the program still needs future maintenance to form it available and stable to execute. The program are going to be tested after a particular period of your time and debug each of the function and possible bugs, whenever a possible bug is detected; the program needs may need to be refined to fix the bugs for better design. Meanwhile, there will be updates and more add on to the database to increase the database capacity. To add more functionality to the program, depending on the new keywords, responses, relevant data found that could be applied to the program; the database will always be improved and can handle more and more cases.

V. CONCLUSION

The project is very useful and owns a large potential use in industries. Although the program primary concerns more about how to do the personal assistant on Windows desktop using the voice, the concept of voice recognition can be applied in different industries as in many situations it is more convenient and saves a lot of time and is also helpful especially for those who have difficulty in working with manual operations. No program has a perfect design without any flaws, it is same here in this program. Even though the program includes all the primary functions implemented and working properly, there are still many things that can be done to improve its performance.

ACKNOWLEDGMENT

I thank my Mentor Dr. Neha Agrawal for analyzing the data and for advising on all aspects related to this paper, along with other concepts too. Later for validating the experimental result and reviewed the paper.

REFERENCES

[1] Efthimios and Constantino in 2017, "Monkey says, Monkey Does-Security and Privacy on Voice Assistants", IEEE journal, ISSN: 2169-3536

[2] Tatiana Ekeinhor-Komi, Jean-Léon Bouraoui, Romain Laroche, Fabrice Lefèvre, 09 February 2017, "Towards a virtual personal assistant based on a user-defined portfolio of multi-domain vocal applications", IEEE Xplore, INSPEC 16657478.

[3] Anurag Mishra, Pooja Makula, Akshay Kumar, Krit Karan and V. K. Mittal, May 28-30, 2015, "A Voice-Controlled Personal Assistant Robot", IEEE journal, Page: 8, INSPEC accession number 15291099.