

Desktop Voice Assistant System with the Help of Natural Language Processing

Vishal Bisht

Bachelors of Technology, Galgotias University, Greater Noida, U.P., India
Email: vishal30bisht@gmail.com

Abstract—In the era of Technology, everyone aspires to live a more comfortable, convenient, and tech-savvy. In this hustle, everyone requires a personal assistant to help them with their everyday tasks. A personal voice assistant is software that can carry out various tasks and give services to a person based on their spoken commands. This is accomplished by a synchronous process that involves the speech patterns recognition and subsequent creation of synthetic speech, followed by the operations of command given. These assistants allow users to automate operations such as texting, mailing, emergency contact dialling, giving home remedies for diseases and media playback etc. Also, they might be emotionally active. People are developing reliant on technology as it advances, and the computer or laptop is one of the most widely used platforms. We all want to make using these computers more pleasant and easy to use. The usual way to provide a command to the computer is to type it in, but a more convenient approach is to speak it.

Giving voice input is useful not just for regular individuals, but also for those who are visually handicapped. unable to provide input using a keyboard. Only India has 18.4 million blind individuals, accounting for one-fourth of all the people worldwide. India is home to one in every three blind people. This necessitates the use of a voice assistant that can not only receive commands via speech but also execute the required instructions and provide output by voice or any other method.

Index Terms— Virtual assistant, Voice recognition, NLP (Natural language processing), Voice assistant.

I. INTRODUCTION

In the era of Technological Development and Advancement, it's our duty to make technology more simplified and usable. Many new features are being added to mobile phones and desktop computers as technology advances. We need a quicker and more reliable mode of input to use them in a more convenient and enjoyable manner. In our voice assistant project, we use voice commands to enter data into the system, which requires the usage of a microphone to convert acoustic energy to electrical energy. Our voice assistant is a computer program that can conduct tasks and perform operations for an individual based on human verbal instructions, such as using human voice commands and the voice assistant responding with synthetic voice. Users may use voice commands to ask their assistants questions, control home automation devices, and playback media, as well as perform other basic tasks like as email, to-do lists, open or stop any programme, and send messages to anyone on WhatsApp. The following are some of the fundamental tasks that may be accomplished with the assistance of a voice assistant: -

- Surfing on internet: Just by asking assistant search <Topic>
- Setting up Alarm clock
- Emergency calling
- Listening to music or watching a video
- Running any software or application.

These are only instances; we can do a lot more integration depending on our needs.

With this project our goal is not only to make things easier for ordinary people but also for the disabled. This computer application will help many visually impaired people, helping them while performing some basic tasks on their desktop. According to the World Health Organization, 1 billion people are visually impaired due to uncorrected refractive error (89 million), glaucoma (7.7 million), corneal opacities (4.2 million), diabetic retinopathy (3.9 million), cataract (94 million) and trachoma (2 million), and visual impairment due to irreversible presbyopia (826 million). Thus, this project will help and assist many such people with disabilities and will make their professional life easier.

II. RELATED WORK

The construction of voice assistants began in 1962 at the World's Fair in Seattle, when IBM introduced a gadget called the Shoebox. IBM Box that can detect spoken numbers and perform calculations and then illuminate them by illuminating numbers labelled 0-9 in them. It was able to detect a total of 16 words. [1]According to J. B. Allen, language is the most important medium of communication, and voice is its primary interface. As a human-machine interface, voice signals were transformed into analogue and digital waveform that a machine could understand. [2]Mulyana and Hakimi (2018) proposed LTKA-Bot, a virtual assistant based on Artificial Intelligence chatbots. It is a virtual assistant that provides a wide range of course-related services such as session data recording and attendance, task management, and so on. [3]Sheppard (2017) lists the following characteristics that distinguish voice assistants from others: Natural Language Processing is capable of comprehending and processing human language. The ability of machine learning to adapt to new situations by identifying patterns is required to bridge the communication gap between humans and machines. [4]Song (2019) conducted an online survey with 433 adult participants to assess user satisfaction, usefulness and ease of use, behaviour toward voice assistants, and intention for using them. The findings revealed that perceived usability has a significant impact on people's attitudes toward voice assistants and intentions to use this technology. Additionally, consumers prefer to purchase devices that are simple to use. [5]Miss. Priyanka V. Mhamunkar and colleagues presented a technology that would allow individuals to retrieve word meanings using synthesized sounds. [6]In 2019, Lovato et al. investigated the interactions of children aged 5 to 6 and their parents with a smart speaker. The research, which lasted two weeks, included 18 families. Each family was interviewed to get their perspective on the smart speaker experience. The study found that around 89 percent of the children's questions were correctly transcribed, but only 50 percent of the children's questions received a full answer. Children and their parents told them that provided answers were either too long or difficult to understand. The majority of the children's questions concerned about their own world.

[7]According to Abdolrahmani et al. (2018), voice interaction is convenient for visually impaired people and can play a significant role in instilling feelings of freedom, liberation and empowerment in them. For people from a wide range of disabilities, including hearing and speech difficulties. Their study also confirmed that voice interplay is practically convenient for especially abled people and allows them to perform day-to-day tasks. [8]Tong Lai Yu and Santhrushna Gande have created an Android solution that uses open-source services to assist programmers who face physical issues while developing. On the client side, they employed the producer-consumer paradigm to integrate operations that may be performed using an Android hand-set. [9]Omyonga Kevin and Kasamani Bernard Shibwabo claim to have created an app that can carry out spoken orders without the need for an internet connection, saving us money on data. Their technique is faster than many strong engines like Alexa and Cortana since it does not require an internet connection. [10]According to Malkin et al. (2019), users of voice assistants and smart speakers face significant security and privacy concerns. In a survey of 116 smart speaker owners' beliefs, behaviours, and concerns about the recordings made and shared by their voice assistants, 56 percent were totally unaware of fact that their recordings were being stored and could be reviewed if needed for personal benefits. While some participants did not consider their own recordings to be important or sensitive, but they were strong against the third-party use of their data. Only 3% of participants acted smartly and went over their recordings and deleted them. Furthermore, only 5% of participants muted their device, and only 4% disconnect their gadget to stop listening. It comes to the conclusion that confidentiality regulations are underutilised. [11]Hoy (2018) takes privacy into account as well, because anyone with access to a voice-

activated device can violate its users' integrity and confidentiality by asking it questions, obtaining information about the user's accounts linked with the device, and requesting tasks. [12] A morphological analyser for most Indian language NLP applications was described by Mugdha Bapat, Pushpak Bhattacharyya, and others. Through- out the research, they described and estimated a Marathi morphological analyser. The study found that Marathi has a high level of correctness, with consistent inflectional standards when using Finite State Systems to demonstrate language in a sophisticated fashion. Because Marathi has complicated morphotactic, grouping of post places and the formation of FSA is one of the most important aids. [13]Sutar Shekhar and a group of academics have collaborated on an application that uses speech to execute the majority of the system's functions, as well as the ability to send a message using their voice command, to aid visually impaired individuals. They want to keep improving their software so that it can eventually recognize a variety of local languages, such as Bengali and a number of dialects of the common Hindi language.

In my project work I have chosen Python as the programming language since it is one of the most robust, with huge libraries and with the use of pytsx3 and voice recognition API's, development becomes easier and more accurate. To assist visually impaired persons, our software always repeats the command that the user enters into the system, so that the user knows if they entered the correct command or not. On the other hand, it continues to listen and respond to the user's needs until the user decides to stop.

III. PROBLEM FORMULATION

As in introduction I have discussed about product is made to help the visually impaired peoples and also ease the User Experience of the users, while working on the desktop. Earlier using system even for a basic task was bit difficult for the especially abled individuals, they use traditional braille scripted keyboard to these short tasks, but with the help of voice assistant system they can perform all day-to-day tasks very easily and simplified manner.

IV. PROPOSED WORK

Our work is based on NLP (Natural Language Processing) because here have we used the Python Speech Recognition module to identify user speech and convert it into a character unit and we are getting desktop assistant response using the Text to speech pytsx3 module.

Initially, the system will continue to obey commands, and the time spent on listening commands by the system varies and can be adjusted to meet user's needs. If the system is unable to retrieve information about command from the user, it will prompt the user to repeat the process in loop until the required number of times is reached. Playing music, sending texts, searching Wikipedia, opening applications installed, opening websites in a web browser, playing videos on YouTube and other features are supported in the current version.

V. Methodology

A. Speech recognition

The proposed system converted input speech by user's end into text using the Google API which uses Natural language processing (NLP). The speech is sent to Google Cloud server for processing and the system receives the resulting text as an output, then that output strings are taken as input by the system.

B. Command Extraction:

After the user's speech (command) is converted as input string, we extract certain predefined keywords from our conditional statements, which are based on users' commands or queries and try to match with of input string.

C. Project Frontend

For frontend I will be using Python PyQt5 module.

PyQt5 is the most recent version of a graphical user interface widgets toolkit. It's a Python interface to Qt, a sophisticated and widely used cross-platform GUI framework. PyQt5 is a combination of Python and the Qt library.

D. Project backend

We used the Python programming language to complete the work. We used the python module as follows:

- pytsx3 sapi5 API-to convert the output string to speech in order to provide a better user experience.
- Speech recognition- to recognize user speech and convert them into string.

- Datetime- to work on date and time.
- Wikipedia- to extract information from Wikipedia.
- Web browser- to surf on different websites.
- os-to work on operating system.
- Random-for random selection.
- Twilio.rest- api for phone calls.
- Pywhatkit- used for playing YouTube videos.
- Subprocess- used for google surfing. Etc.

We created many functions that will be executed on the user's command.

E. Working Formula

A statistical measure used to quantify the value of a word to a document in a collection is TF-IDF, which stands for term frequency-inverse document frequency.

The TF-IDF score value rises in proportion as the number of times a word occurs in the text, but it is counterbalanced by the number of documents in the collection that includes the word.

Let's look at the formula for determining a TF-IDF score for a certain phrase x within a document y.

For a term an in b

$$W_{a,b} = T_{f_{a,b}} * \text{Log}(N/df_a)$$

$T_{f_{a,b}}$ =number of occurrences a in b

df_a =number of documents containing a

N =Total Number of documents

VI. SYSTEM ARCHITECTURE

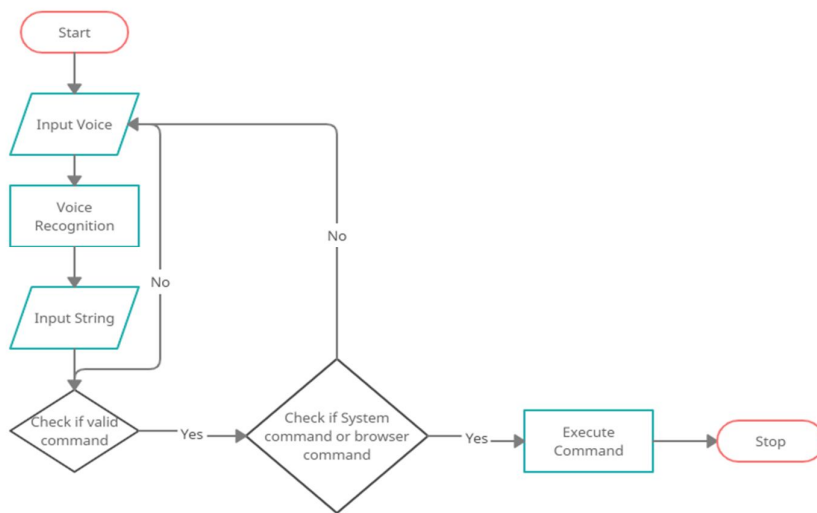


Fig.1. Flowchart of Desktop Voice Assistant

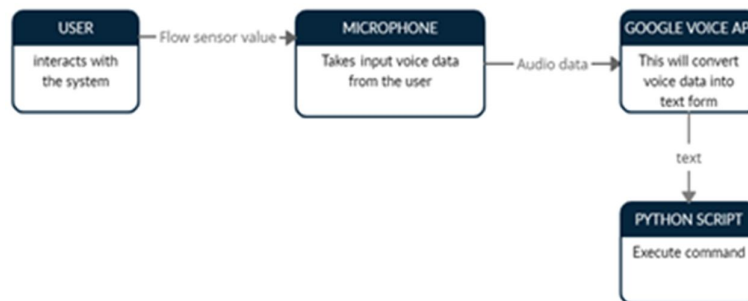


Fig.2. Dataflow diagram (DFD) of Desktop voice assistant

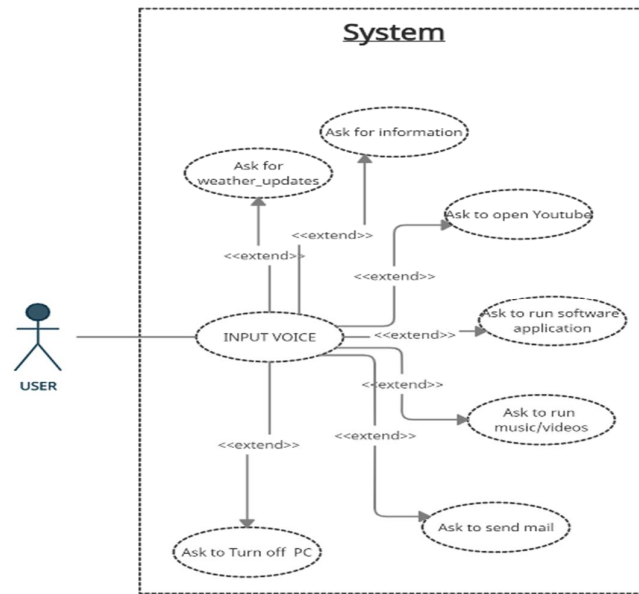


Fig.3. UML Use Case Diagram of Desktop voice assistant

VII. CONCLUSION

This Project is focused to help visually impaired individuals and also to improve User Experience for users while sitting on the desktop. In this I have tried to create multiple useful day to day lifestyle function for the users like playing music, surfing googles, making phone calls etc. In future I will try to make this virtual desktop assistant more and more helpful for the users.

FUTURE SCOPE

The future of voice assistants can be defined in a variety of ways. There is still a lot to be done in terms of interoperability and integration with other devices and systems., Another dimension would be the redundant use of wake words at the start of each command and beautiful Graphic user interface. But, for all intents and purposes, these technologies have a bright future.

AKNOWLEDGEMENT

I would like to thank my project mentor Dr.Dileep Kumar Yadav (Professor at Galgotias University) for donating his precious time to the guidance and information he has shared with us. With guidance he encouraged me to work on this paper. Under his direction we completed this paper.

REFERENCES

- [1] J. B. Allen, "From lord rayleigh to shannon: How do humans decode speech," International Conference on Acoustics, Speech and Signal Processing, 2002.
- [2] E. H. R. Mulyana, "Bringing Automation to the Classroom: A ChatOps-Based Approach," in 4th International Conference on Wireless and Telematics (ICWT),(IEEE), 2018.
- [3] B. Sheppard, "Theological Librarian vs. Machine: Taking on the Amazon Alexa Show (with SomemReflections on the Future of the Profession)," Theological Librarianship, vol. 10, no. 1, pp. 8-25, 2017.
- [4] Y. W. Song, "User acceptance of an artificial intelligence (AI) virtual assistant: an extension of the technology acceptance model," UT Electronic Theses and Dissertations, 2019.
- [5] M. p. v. B. M. k. S. & N. L. Mhamunkarr, "Android application to get word meaning through voice," International journal of Advance Research in Computer Engineering & Technology (IJARCET), vol. 2, no. 2, pp. 572-572, 2013.
- [6] W. A. M. A. S. B. P. Lovato, "Hey Google, Do Unicorns Exist?: Conversational Agents as a Path to Answers to Children's Questions," in Proceedings of the 18th ACM International Conference on Interaction Design and Children , 2019.

- [7] B. S. M. A. K. R. Abdolrahmani, “.Siri Talks at You: An Empirical Investigation of VoiceActivated Personal Assistant (VAPA) Usage by Individuals Who Are Blind.,” in In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, 2018.
- [8] S. T L Yu, “An open -Source Based Speech Recognition Android Application for Helping Handicapped Students Writing Programs,” International Conference on Wireless Networks (ICWN),The Steering Committee of the International Conferennce on Wireless networks., pp. 71-71, 2015.
- [9] K. O. B K Shibwabo, “The application of realtime voice recognition to control critical mobile device operations,” 2015.
- [10] N. D. J. T. A. W. P. E. S. W. D. Malkin, “Privacy Attitudes of Smart Speaker Users,” Proceedings on Privacy Enhancing Technologies, vol. 4, p. 250–271, 2019.
- [11] M. B. (. Hoy, “Alexa, Siri, Cortana, and more: an introduction to voice assistants.,” Medical reference services quarterly ,37(1), pp. 81-88, 2018.
- [12] H. M.Bapat, “A paradigm-based finite state morphological analyzer for marathi,” Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26-34, 2010.
- [13] P. S. S. Shekhar, International Journal of Advance research in computer Science and Management Studies, 2015.
- [14] L. R. R. B S Atal, “A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition,” Acoustics, Speech and Signal Processing,IEEE Transactions, vol. 24, no. 3, pp. 201-2012, 1976.
- [15] V. C Radha, “A review on speech recognition challenges and approaches,” doaj.org, vol. 2, no. 1, pp. 1-7, 2012.
- [16] A. W. T Schultz, “Language independent and language adaptive acoustic modeling for speech recognition,” Speech Communication, vol. 35, no. 1, pp. 31-51, 2001.