



Survey on Voice Assistant Using Python

Prof. S.A Nalawade

Vaishnavi Wayal

Pradnya Kute

Gautamyi Chougule

Guide

Student, BE-IT

Student, BE-IT

Student, BE-IT

Jayesh Chikane

Student, BE-IT

Department of Information Technology, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

1. ABSTRACT

In this modern era, living has become smarter and easier with help of technologies used in home automation such as Artificial Intelligence (AI), Cloud Computing, Mobile Computing Voice assistant, image and voice processing etc. This project works on the voice input to voice and text output. It's user friendly which'll assist the user to perform the tasks on PC. In such times where advancement is one of the main agenda of development, our Voice Assistant is the assistant which will interact with user as per the regional language like Marathi, English being the default language to access files and websites from the system etc. Via voice the commands are taken as input by the machine and the operations are performed on Windows operating System. AI technology, Python and API is used to build this application.

2. INTRODUCTION

In recent times, automation is in the boom. Wherein the interaction between the machine and the human is initiated via voice assistant. As nowadays there been a new inventions in the field of Artificial

Intelligence(AI) that is Voice Assistant. Speaking which Alexa by Amazon, Siri by Apple etc. Speech recognition such as text-to-speech, speech-to-text, speech-to-speech can be done by using the API's. The implementation of such smart automation system in now constantly increasing as well it's also evolving.

Voice research have dominated over the text search. It's in-hand operation where there's aren't a necessary to use hands. Its been a highlight of this voice automation system that it can be operated from certain radio range from the user. Some of the type of Voice Assistants are:-

- 1.Intelligent Personal Assistant
- 2.Automated Personal Assistant
- 3.Virtual Digital Assistant
- 4.Chat bot

Our voice assistant is designed specifically for Windows operating system using Python. It can perform operations such as opening the file from the file manager, can conversate using regional languages, open the online websites etc. It also has a graphical representation i.e. animated

presentation for voice assistant. It's designed to work more efficiently and improve the way of interaction with machines.

3. Literature Survey

Voice Assistant has the long history. It has been in the phase of evolution since 1880.

In 1880 Alexander Graham Bell implemented further operations over Edison's phonograph, which his Volta Graphophone Company patented in 1886. Instead of foil graphophone was used, which allowed for longer recordings and higher-quality playback. Edison also developed a wax version of the phonograph and both devices were used primarily for dictating letters and other documents.

In 1961 IBM introduced the IBM Shoebox, it's the first digital speech recognition tool. It recognized 16 words and digits 0 to 9. It was able perform mathematical functions and perform speech recognition.

In 1972 Carnegie Mellon completed the Harpy Program. It could able to understand about 1000 words. Harpy processed speech that followed pre-programmed vocabulary, pronunciation and grammar structures.

In 1990 Dragon launched Dragon Dictate, the first speech recognition module for consumers for \$6,000 (Indian currency= 496.27 in current date).

In 1996 Microsoft introduces Clippy. Microsoft Clippy, it's also known as Clippit and officially recognized as Office Assistant, it was an intelligent user interface for Microsoft Office. It assisted the users in a number of interactive ways by appearing as a visualized character on the Office applications and offering help related to various operations of the Office Software. It was made available in the Microsoft

Office for Windows in 1997 and in 2003 it was discontinued.

In 2011 Apple introduced Siri. Voice queries, control based on gesture, focus-tracking and natural language user interface for answering the questions, making recommendations and perform operations by passing on the requests to as set of internet services were used in Siri. With it's continuous use, it adapts to users' individual language usages, searches and preferences, returning individualized results.

In 2012, Google launched Google. Google Now proactively delivered information to users to predict information they might need in the form of informational cards which was based on the users' search habits and other factors. For Android and iOS, Goggle Now was a feature of Google search embedded in Google app. Functionality of Google Now is being used in the Google app and it's discovery tab today also wherein it's branding is no longer used.

In 2013, at annua BUILD developer conference, Cortana was introduced by Microsoft. Cortana is a virtual assistant which uses the Bing search engine to perform tasks such as setting reminders and answering questions for the user. Depending upon the software programs and region in which its used, Cortana is currently available in English, Chinese, French, German or Italian, Portuguese, Spanish and Japanese language editions.

In 2014, Alexa and Amazon Echo, was introduced by Amazon which was available to prime members only. Amazon Alexa is a virtual assistant technology also known as Alexa, it's largely based on a Polish speech synthesizer named Ivona. It's capabilities consists of voice interaction, music playback, playing audiobooks, setting alarms, streaming podcasts, providing weather forecast, traffic news, sports news, making to-do lists and other real-time information, such as news. Using itself as a home automation, it can control numerous smart devices.

In 2015, Cortana on windows 10 desktops and mobile devices was introduced by Microsoft. While in US Amazon officially launched Amazon Echo. Alexa Skills kit was introduced by Amazon.

In 2016, voice-powered virtual assistant app was introduced by SoundHound, HOUND. SoundHound is grandly known for its music recognition app, which listens to songs and identifies them. Amazon launched Amazon Echo Dot and Amazon Tap. Google introduced the Google Assistant as a part of the messaging app, Allo. In the same year the virtual assistant startup Viv was obtained by Samsung.

Google launches Google Home and smart phone Google Pixel. Also Chinese manufacturer Linglong launches Echo competitor DingDong.

In 2017, Samsung introduced Bixby next to Galaxy S8 devices. While thereafter Google Homes was launched in UK. Google introduces multi-user support for Google Homes; it recognizes six different voices. Simultaneously, Amazon introduced Echo look. In China, Baidu unveils its first consumer AI device Xiaoyu. Amazon introduces calling/messaging feature for Echo devices. While on the other hand Apple introduces HomePod and Alibaba launches Genie X1 Smart Speaker.

These Voice Assistants was introduced on only for smart phones, smart homes etc, but also was integrated in cars as well like BMW.AI based Voice Assistants are being evolved in many ways like it can be developed using various languages such as Java and Python.

A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments (2018) by Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita, Giorgio Mario Gtasso stated the software architecture for building lightweight, vision and speech enabled virtual assistant for smart phone and automation application. A complete prototype application was build featuring a realistic graphic assistant able to show facial expression and enabled with

speech synthesis and recognition. The proposed assistant is effective and resource efficient, interactive and customizable and realised prototype runs on low cost, small sized Raspberry PI 3 device.

AI based Voice Assistant Using Python(2019) by Deepak Shende, Ria Umahiya, Monika Raghorte, Aishwarya Bhisikar, Anup Bhangе stated the new insights of natural human-machine interaction, in which machine would learn how to understand the humans language. It also expressed the principles of functioning of voice assistants, it's main shortcomings and limitations, methos of creating local voice assistant without using cloud services is described.

Artificial Intelligent-Based Voice Assistant(2020) by Subhas S, Prajwal N, Siddesh S, Ullas A, Santhosh B stated a voice assistant gathering the audio from the microphone and get converted into text, later it sent through GTTS (Google Text To Speech.). GTTS engine will convert text into audio file in English language , then that audio sound is played using the play sound package of python programming language.

Voice Assistant Using Python (2021) by Nivedita Singh, Dr. Diwakar Yagyasen, Mr. Surya Vikram Singh, Gaurav Kumar, Harshit Agarwal stated a voice assistant using Python which allows the user to run any type of command in linux without interating with keyboard. It performs basic tasks such as weather updates, stream music, search Wikipedia, open desktop applications etc.

Survey On Smart Virtual Voice Assistant(2022) by Manjusha Jadhav, Krushna kalyankar, Ganesh Narkhede, Swapnil Kharose stated natural language processing algorithm that helps machines to engage in communication using natural human language in many forms. It also connects to World Wide Web to provide the results that the user required.

Research Paper On Desktop Voice Assistant(2022) by Vishal Kumar Dhanraj, Lokeshkriplani, Semal Mahajan stated working of a vice assistant without using

cloud services, which will allow the expansion of devices in the future. It can perform any kind of task in exchange of commands given by the user without any error, it will listen to the users 'voice only and will not be activated from environment noise'.

4. Aim

Voice assistants are globally used for performing tasks of Laptops, smartphones, PC's etc. These tasks are performed while using internet which are introduced by software firms like Google, Microsoft etc. By having voice assistant on-board it becomes easier to carry out the tasks without the interference of keyboard and it can help the physically challenged people to operate the mobile devices at ease.

Our aim is to develop the graphically presented(UI) voice assistant for windows using Python programming language at backend to carry out the operations such as accessing the files and applications on users' PC following the web surfing, wherein the accessing of files is independent of internet connection. It can communicate in regional language (Hindi) and English being the default language.

5. System Architecture

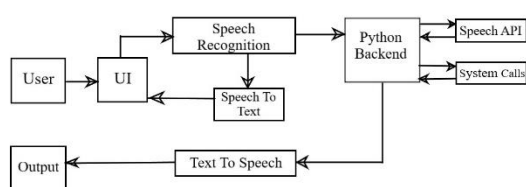


Fig. System Architecture

6. Methodology

1] API- SAPI5

API is an abbreviation for "Application Programming Interface", it's a software intermediary that allows two applications to communicate with each other. An API is a software which can be used by other software, to communicate with other software or even hardware. It acts as a link between different Softwares and devices. There are so many applications using different technologies and programming languages, which uses API's to interact with each other.

This model uses SAPI5 API to communicate. Microsoft developed an API, "Speech Application Programming Interface", which allow the use of speech recognition and speech synthesis within Windows application. It mainly has following features: Shared Recognizer, IN-proc recognizer. Grammar objects, Voice object, Audio interfaces, User lexicon object and Object tokens.

We'll import API SAPI5 for communication between the user and the assistant.

2] Speech Recognition module

Speech recognition, a important feature embedded in the various applications such as AI automation etc. This module helps to obtain the text output from the voice input of the user. It's obtained through speech recognition module which identifies the command is whether it's an API call, System call or Content extraction.

3] Python Backend

The whole program is written in Python backend. Through speech recognition module, Python backend work on obtaining the output in exchange of voice input provided by the user.

4] System Call

System Calls are the programmatic way in which computer program requests a service from the kernel of its' operating system. Provides an important and essential interface between operating system and process.

5] Motion UI

Good design is one of the most important aspect of app or website. It helps to make the app more interactive and innovative. A good user interface helps to build better communication between the user and the app.

Motion UI is a library of Sass which helps in creating flexible UI animations and transitions. It's a library that has control on the transformation effects including components of foundation. It provides set of pre-made effects as CSS package.

We have used Motion UI to build a effective interaction which will run in the backend during the functioning of the instructions or commands given by the user.

7. *Algorithm*

- 1] Take voice commands from user
- 2] Display commands using speech-to-text module
- 3] Mark the key words
- 4] Make API calls System calls
- 5] Display the results in form of text or speech or operation.

8. *References*

1. Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita, Giorgio Mario Gtasso., A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments, '978-1-5386-5024-0/18/\$31.00 ©2018 IEEE'
2. Deepak Shende, Ria Umahiya, Monika Raghorte, Aishwarya Bhisikar, Anup Bhangе., AI based Voice Assistant Using Python, 'Volume 6 Issue 2|JETIR| ©2019'
3. Subhas S, Prajwal N, Siddesh S, Ullas A, Santhosh B., Artificial Intelligent-Based Voice Assistant, '978-1-7281-6823-4/20/\$31.00 ©2020 IEEE'
4. Nivedita Singh, Dr. Diwakar Yagyasen, Mr. Surya Vikram Singh, Gaurav Kumar, Harshit Agarwal., Voice Assistant Using Python, 'Volume 8 Issue 2|IJIRT| © July 2021'
5. Manjusha Jadhav, Krushna kalyankar, Ganesh Narkhede, Swapnil Kharose., Survey On Smart Virtual Voice Assistant, Volume:09 Issue:01|Jan 2022.
6. Vishal Kumar Dhanraj, Lokeshkriplani, Semal Mahajan., Research Paper on Desktop Voice Assistant, Volume 10 Issue 2|2022|PP.15-20