ABSTRACT

Voice assistants are software agents that can interpret human speech and respond via synthesized voices. In day to day, life became smarter and interlinked with technology. The voice assistance system, it can act as a basic medical prescriber, daily schedule reminder, note writer, give notification of important events, weather report, calculator, placing orders on online bookings like e- marketing, restaurant, airline booking, railway booking and a search tool. User can ask their assistants questions, control home automation devices like turn on and off lights, fan, AC, and media playback via voice, with verbal commands. Personal Assistants, or conversational interfaces, or chat bots reinvent a new way for individuals to interact with computes. The project works on voice input and give output through voice. The voice assistance takes the voice input through microphone (Bluetooth and wired microphone) and it converts voice into computer understandable language gives the required solutions and answers which are asked by the user. The assistance connects with the world wide web to provide results that the user has questioned. Natural Language Processing algorithm helps computer machines to engage in communication using natural human language in many forms. Python is a suitable language for scriptwriters and developers. The query for the assistant can be manipulated as per users need. Speech recognition is the process of converting audio into text. This is commonly used in voice assistants like Alexa, Siri, etc. Python provides an API called Speech Recognition to allow us to convert audio into text for further processing. The voice assistant using python which allows the user to run any type of command in Linux without interaction with keyboard. The main task of voice assistant is to minimize the use of input devices like keyboard, mouse etc. It will also reduce the hardware space and cost. Personal AI voice assistant that can understand voice command using speech recognition in Python is quite easy. The voiceflow Dialog manager comprises all the conversation and generates an API key from which the input transcript is matched and the output text is extracted using API. In addition to this process, the conversations can be stored in cloud for referring the chat history between the voice assistant and the user.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Over the years, humans have progress in inventing new technologies for reducing human efforts and saving human life. since the development of IOT reduce the human labor to nil. The development of IoT (Internet of Things) had been advanced in several study fields like home automation, personal assistant AI, smart city, smart farming etc. so the personal assistant help reducing the manual effort being input by human in their day-to-day task. The voice controlled personal assistant receive the voice command as input to perform the numerous tasks. Any voice command system needs three basic components which are speech to text converter, query processor and a text to speech converter. Voice has been a very integral part of communication nowadays.

There have been some very good innovations in the field of speech recognition. Some of the latest innovations have been due to the improvements and high usage of big data and deep learning in this field. These innovations have attributed to the technology industry using deep learning methods in making and using some of the speech recognition systems. Text to speech conversion is the process of converting a machine recognized text into any language which could be identified by a speaker when the text is read out loud. It is two step processes which is divided into front end and back end. First part is responsible for converting numbers and abbreviations to a written word format. This is also referred to as normalization of text. Second part involves the signal to be processed into an understandable one.

LITERATURE SURVEY

This gives ideas about the next generation of virtual personal assistants and modifications that can be made to interact with the assistants. Speech recognition has several waves of major innovations. Speech recognition for dictation of voice, search, and voice commands has become a standard feature on smartphones and various other devices. To this aim, a conversational assistant, capable of answering common questions, has been combined with a content discovery engine that is more suitable for finding the proper answers from a collection of heterogeneous sources. Many companies of voice assistants are trying to improve interaction and more features to the next level and many of the youth started using a voice assistant in daily life and from many sources the result showed very good feedback. Smart assistants are useful in many fields such as education, home appliances, etc. and the voice assistant is also useful for blind people. They can get any information just by telling the assistant, and this is possible because voice-based Smart assistants. We are using raspberry pi for SSH and different module connections. Raspberry pi is a low cost and small size computer that plugs into a computer or monitor with the help of connectors and standard keyboard and mouse. It is based on the voice as the research object, it allows the machine to automatically identify and understand human spoken language through speech signal processing and pattern recognition.

PROBLEM STATEMENT

In this project, addressing the problems of data privacy and security, user defined conversation and home automation.

In data privacy and security, data privacy is made to keep the data safe, private, and secure where it wouldn’t be fallen to bad hands and third party will not be able to access it. Data privacy needs to be a top priority for businesses. Data security is a process of protecting data from unauthorized user.

In User defined conversation, when it comes to Siri and Alexa, they have no responding replies if they couldn’t understand what user says, where as in this project, it has a huge advantage of user defined conversation. It makes user comfortable, understandable and creates good impact on fast reply and meaningful sentences as it has in-built intendencies, training phases and google assistant. The user is aware of the sentence to be asked and our chatbot is well trained for answering.

People come home exhausted after a long hard-working day. Some are way too tired that they find it hard to move once they land on their couch, sofa or bed. So, using voice assistant it would help them switch theirs lights on or off, or play their favorite music etc. on a go with their voice with their smart phones would make their home more comfortable.

CUSTOMER SEGMENT

The target of voice assistant is to make things easier and more comfortable. So, the users can be, we may be students, staff, doctors, patients, government, business man and person who need it. Is person without sight cannot learn? No, hearing is enough to learn and gain knowledge especially voice assistant is very useful for them, even the visually challenged people can easily access just by telling what is needed and what they want.

Students often must spend lots of time searching the internet for information about certain topics in order to resolve their doubts. On the other hand, a voice assistant can search the internet for them and provide accurate, relevant results quickly. The voice assistant would be even more useful if it had a special skill to assist students around the school or campus.

At present as pandemic, these people suffer a lot including students, staff, doctors, employees etc. In case of students, it is useful in listening to the audio content and easy to understand and concentrate likely to be clear. Even it prevents from health problems such as eye irritation and students avoid spend time in searching information where lagging of time occurs. When it comes to staff, they feel more relaxed and comfortable due to voice assistant way of pronunciation and communication.

In case of business man, his/her personal data are stored private and kept secured where unknown will not be able to access. The users important dates, meetings, clock, venue, timetable are been recorded and provides notification on time when needed and mandatory remain things.

In general, the user can check weather report, Calendar, breaking news, booking of various airline and railway tickets, orders in restaurant and e-commerce.

PROJECT COMPONENTS

HARDWARE IMPLEMENTATION

Microphone:

Microphone is used to take the audio input of the sound. This audio input when further passed through the system would be searched for keywords. These keywords are essential for the functioning of the voice command system as our modules work on the essence of searching for keywords and giving output by matching keyword.

Keyboard:

Keyboard acts as an input interface mainly for the developers, providing access to make edits to the program code.

Mouse:

Mouse also acts an interface between the system and the developer and does not have direct interaction with the end user.

Raspberry Pi:

Raspberry Pi is the heart of the voice command system as it is involved in every step of processing data to connecting components together. The Raspbian OS is mounted onto the SD card which is then loaded in the card slot to provide a functioning operating system. The Raspberry Pi needs a constant 5V, 2.1 mA power supply. This can either be provided through an AC supply using a micro-USB charger or through a power bank.

Ethernet:

Ethernet is being used to provide internet connection to the voice command system. Since the system relies on online text to speech conversion, online query processing and online speech to text conversion hence we need a constant connection to achieve all this.

Monitor:

Monitor provides the developer an additional way to look at the code and make any edits if any. It is not required for any sort of communication with the end user.

Speakers:

Speakers once the query put forward by the user has been processed, the text output of that query is converted to speech using the online text to speech converter. Now this speech which is the audio output is sent to the user using the speakers which are running on audio out.

HARDWARE SETUP OF VOICE COMMAND SYSTEM

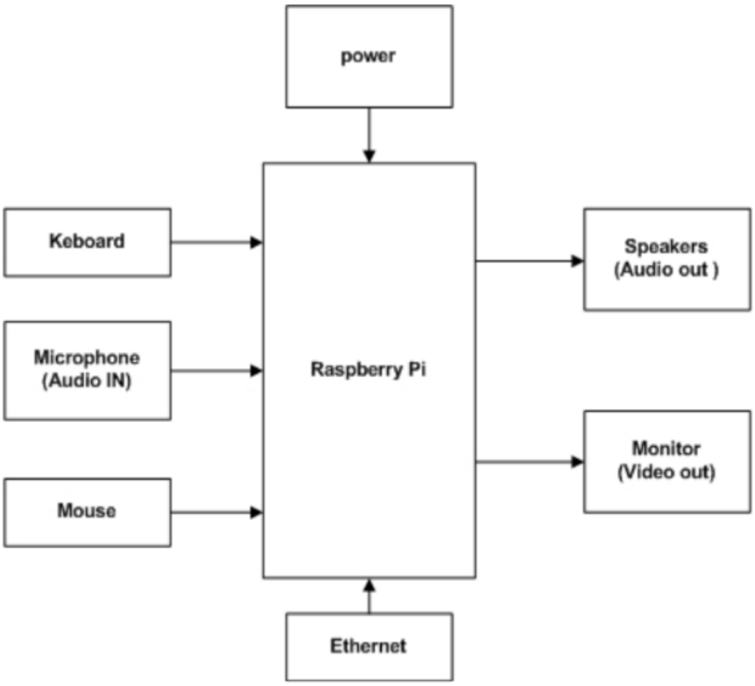


Figure 5.1 Raspberry Pi

In Raspberry pi, to setup the SD card the user needs a computer that has an SD card port whereas most laptop and desktop computers have one. Anything that’s stored on the SD card will be overwritten during formatting. Insert the SD card where the set up with Raspberry Pi OS into the microSD card slot on the underside of the Raspberry Pi. Find the USB connector end of the mouse’s cable, and connect the mouse to a USB port on Raspberry Pi. Connect the keyboard in the same way. Make sure the screen is plugged into a wall socket and switched on. Use a cable to connect the screen to Raspberry Pi’s HDMI port use an adapter if necessary. Connect the screen to the first of Raspberry Pi 4’s HDMI ports, labelled HDMI0. The user can connect an optional second screen in the same way. When the user starts the Raspberry Pi for the first time, the Welcome to Raspberry Pi application will pop up and guide the user through the initial setup.

RASPBERRY PI 4

Merits:

Pi 4 being faster, able to decode 4K video, benefiting from faster storage via USB 3.0, and faster network connections via true Gigabit Ethernet. Raspberry Pi 4 no longer struggles with heavy web pages and apps, and can switch between full online services such as Google's G Suite and today's JavaScript laden sites without lagging.

Pi 4 no longer struggles with heavy web pages and apps, and can switch between full online services such as Google's G Suite and today's JavaScript laden sites without lagging. Raspberry Pi 4 can serve as a learning PC for the kids, a media center, a web server, a game emulation machine or as the brains of a robot or IoT device. Pi 4 has better CPU which results in better processing and performance speeds.

Why Raspberry Pi is used instead of Arduino?

Arduino doesn't have enough processing power to do voice recognition. And by Arduino, meaning the traditional boards like Uno, Nano and Mega. Few of the new boards do come with good enough processing power but on Raspberry Pi it would be much better and simpler since it runs Linux operating system capable enough to run Python and similar programming language. The clock speed of Arduino is 16 MHz while the clock speed of Raspberry Pi is around 1.2 GHz. Raspberry Pi is good for developing software applications using Python, while Arduino is good for interfacing Sensors and controlling LEDs and Motors.

The Raspberry Pi 3 is also a lot faster than the Arduino (1.2 GHz compared to 16 MHz), which gives it the ability to complete everyday tasks that computers do – playing videos, surfing the web, listening to music, etc. This makes the Raspberry Pi 3 an easy choice if you want to use it for voice assistant.

FLOW OF EVENTS IN VOICE COMMAND SYSTEM

First, when the user uses a microphone to send in the input. Basically, what it does is that it takes sound input from the user and it is fed to the raspberry pi to process it further. Then, that sound input if fed to the speech to text converter, which converts audio input to text output which is recognizable by the raspberry pi and can also be processed by it. Then, that text is parsed and searched for keywords.

The voice command system is built around the system of keywords where it searches the text to match by the voice flow. And once key words are matched then it gives the relevant output.

This output is in the form of text. This is then converted to speech output using a text to speech converter which involves using an optical character recognition system. Thus, voice flow identifies the text and then the text to speech engine converts it to the audio output. This output is transmitted via the speakers which are connected to the audio jack of the raspberry pi.

VOICEFLOW

MERITS

Robust context-based input recognition. Easy import and export of chat agent. The interface of Voiceflow is easy to use and it is also easy to set up the voice chatbot. Design contextual conversations easily with Voiceflow. Leverage reusable components, robust context models, interaction model exports. The collection of Integration Steps available will allow you to send and receive data from external sources, and layer in additional functionality to the project that isn't natively supplied by Voiceflow. It creates and re-use components that allow for quick creation and standardization of designs across projects and channels.

DEMERITS

Voice flow does not provide a live chat integration, which is a drawback, because this is the most important integration of any chatbot builder. And flow parts are complicated. When the conversation goes on, the size of the conversation is limited.

FLOW OF EVENTS IN VOICE FLOW

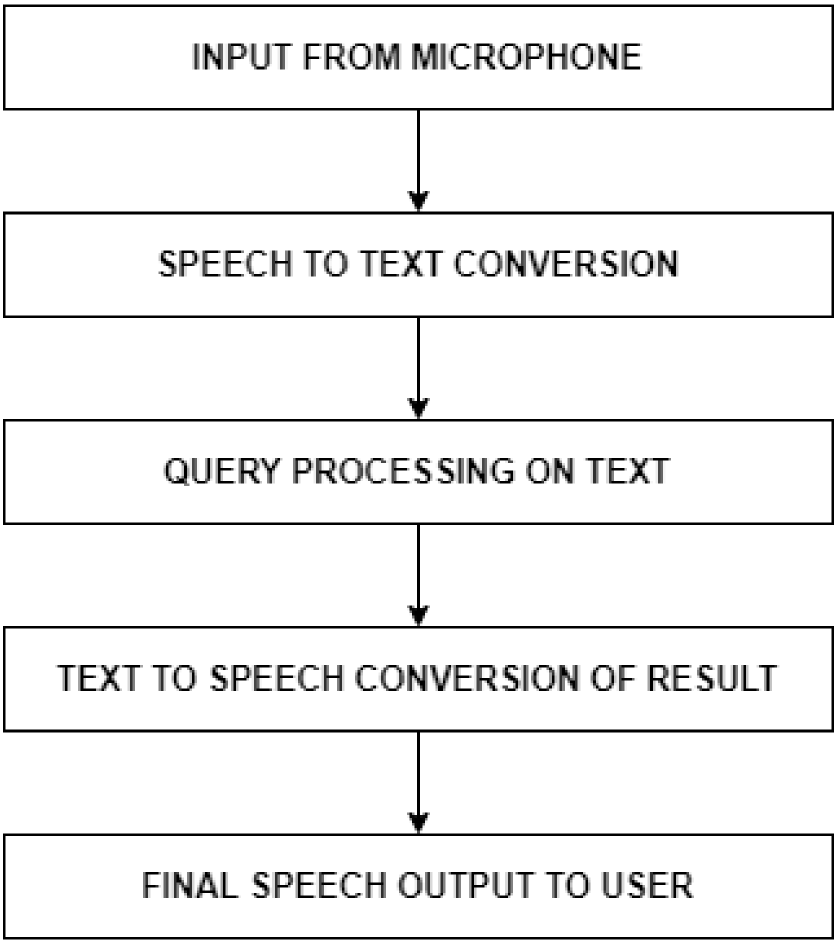

INPUT FROM MICROPHONE
SPEECH TO TEXT CONVERSION
QUERY PROCESSING ON TEXT
TEXT TO SPEECH CONVERSION OF RESULT
FINAL SPEECH OUTPUT TO USER

Figure 5.2 Flow of event in voice flow

MODULES IMPLEMENTED

Speech To Text Engine

Speech-To-Text (STT) engine which is used to convert the commands given by the user in audio input to text form, so that these commands can be interpreted by the modules properly. To use this engine, an application must be created in the Amazon developers console and the generated API key must be used to access the speech engine.

Text To Speech Engine

Text-To-Speech (TTS) engine is used to create a spoken sound version of the text in a computer document, such as a help file or a Web page. TTS can enable the reading of computer display information for the visually challenged person, or may simply be used to augment the reading of a text message. To use engine, an application must be created in the developer's console and the generated API key must be used to access the speech engine.

Porcupine

Porcupine is a highly-accurate and lightweight wake word engine. It enables building always-listening voice-enabled applications. It is compact and computationally-efficient. It is perfect for IoT. It is scalable. It can detect multiple always-listening voice commands with no added runtime footprint.

Query Processor

The Voice Command System has a module for query processing which works in general like many query processors do. That means, taking the input from the users, searching for relevant outputs and then presenting the user with the appropriate output. In this system we are using the site wolfram alpha as the source for implementing query processing in the system. The queries that can be passed to this module include retrieving information.

OS module

OS Module provides operating system dependent functionalities. If we want to perform operations of OS like data reading, data writing, or data manipulate paths then these types of functions are available in an OS module. When these operations raise an error like “OS Error” in case of any error like invalid names, paths, or arguments which may be incorrect or correct but just not accepted by the operating system.

Pyaudio

Pyaudio provides Python bindings for Port Audio, the cross-platform audio I/O library. With Pyaudio, the user can easily use Python to play and record audio on a variety of platforms, such as GNU/Linux, Microsoft Windows, and Apple Mac OS X / macOS.

Use the package manager to install Pyaudio:

“sudo apt-get install python-pyaudio python3-pyaudio”

PYAUDIO SOURCE:

Source is available for download at the Python Package Index (Py PI): pypi.python.org/pypi / Pyaudio.

gTTS

It is Google Text-to-Speech. python and the gTTS module make the converting of text to speech in just a few lines of codes. The installation of the gTTS

After the installation is done, the user can proceed to write a very simple program to understand how exactly user can use this module to convert our typed text into a speech converted output. gTTS (Google Text-to-Speech) is a Python library and CLI tool to interface with Google Translate text-to-speech API. The user will import the gTTS library from the gtts module which can be used for speech translation. The tts variable is used to perform the Google text-to-speech translation on the user’s input. The output of the converted text is stored in the form of speech in the tts variable. The tts.save function allows us to save the converted speech in a format that allows us to play sounds.

WORKFLOW

Start the voice assistant by using the wakeup key words through microphone connected to raspberry pi such as computer, blueberry for detecting and starting it.

If the voice assistant started to recognize the voice of the user, then there will be a response which denotes the user that the voice is being analyzed by raspberry pi through voice flow as the user speaks. After that, there will be speech response from the voice assistant through speaker about the help needed to user to be asked. Now, the user can order the pizza as per their choice by giving the input through microphone. The pizza name will be checked through choices from the menu and save it. If it is stored, the voice assistant gives the response about asking the size required to the user. The user gives the size (small, medium, large) as required through microphone.

Next, the voice assistant asks the choice of the cool drinks to be selected, if needed the user can say through the microphone as per the user's choice.

The order will be placed and the time required to be delivered will be estimated and it will be given as a response to the user through speaker.

Finally, the beep sound from the speaker indicates the end of their conversation.

BLOCK DIAGRAM

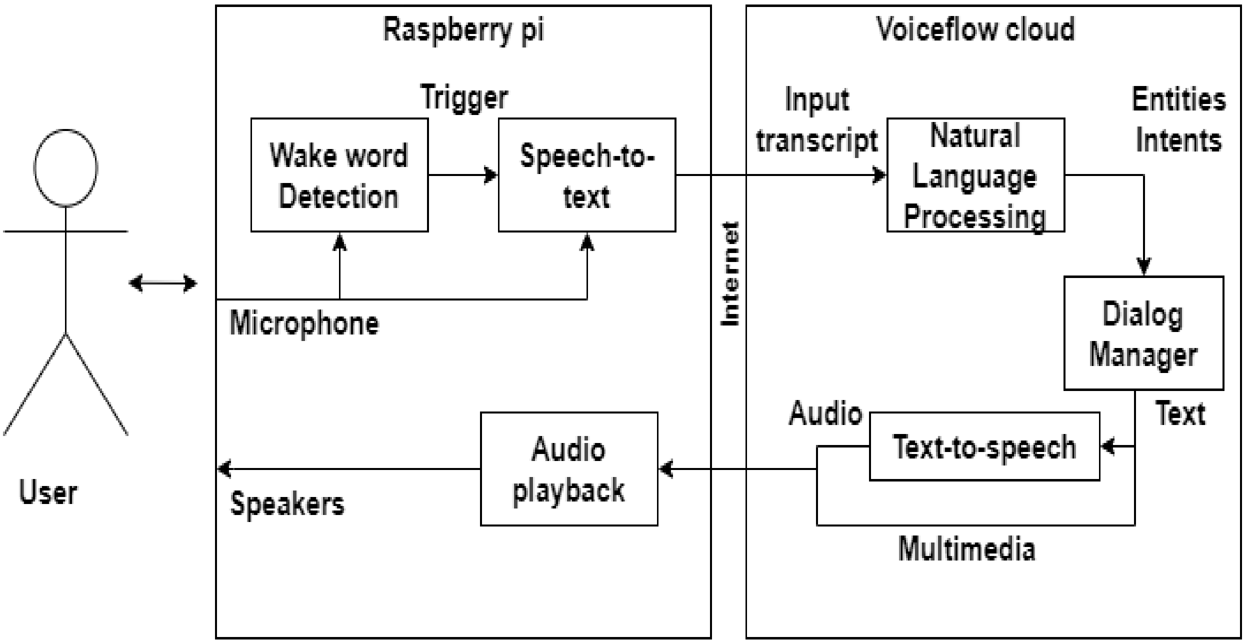
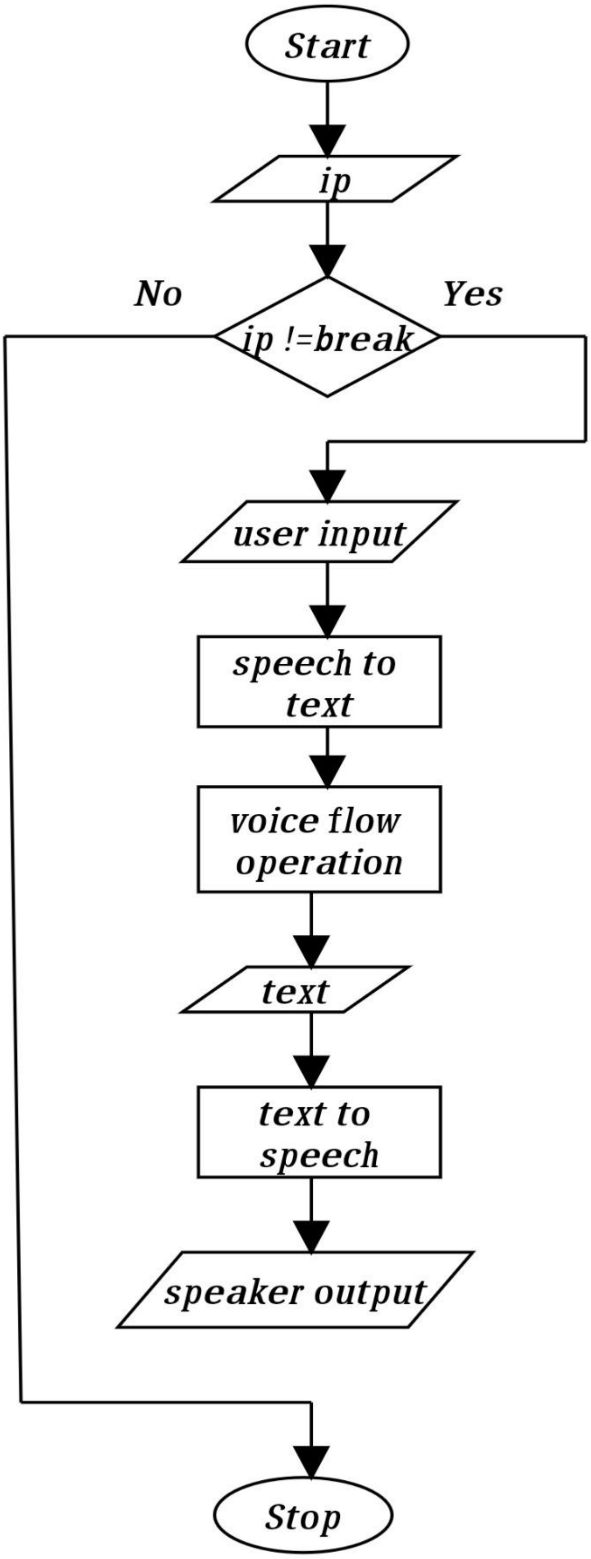
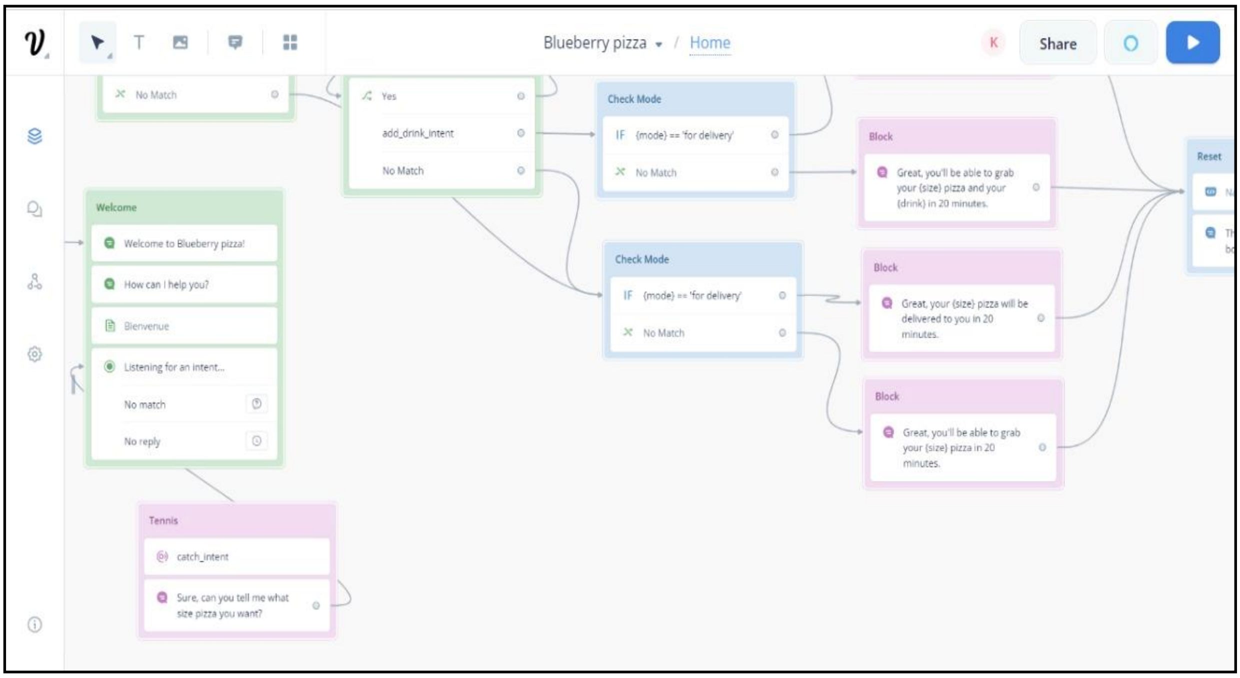


Figure 6.1 Block diagram

FLOWCHART


Start
ip
No
Yes
ip !=break
user input
speech to text
voice flow operation
text
text to speech
speaker output
Stop

VOICE FLOW PIZZA AGENT



Welcome to Blueberry pizza!

How can I help you?

Bienvenue

Great, you'll be able to grab your {size} pizza and your {drink} in 20 minutes.

Great, your {size} pizza will be delivered to you in 20 minutes.

Great, you'll be able to grab your {size} pizza in 20 minutes.

Sure, can you tell me what size pizza you want?

Figure 6.2 Pizza agent Flow diagram

PHOTOGRAPHIC VIEW OF EXPERIMENTAL SETUP

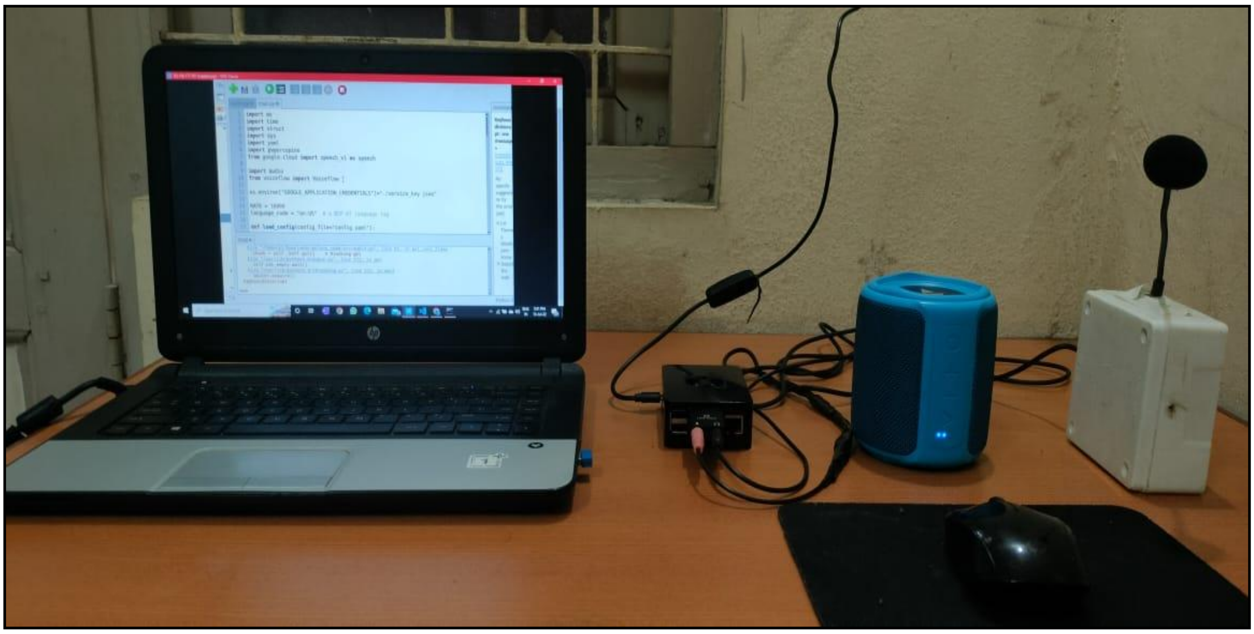


Figure 6.3 Experimental setup

PHOTOGRAPH OF HARDWARE

Figure 6.4 Speaker

Figure 6.5 Microphone

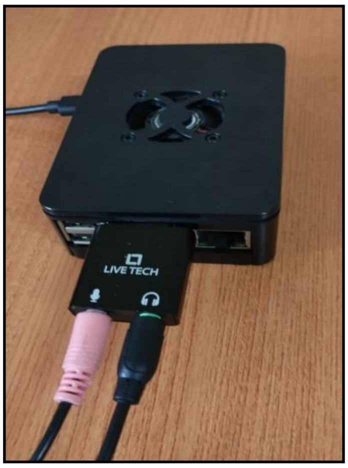
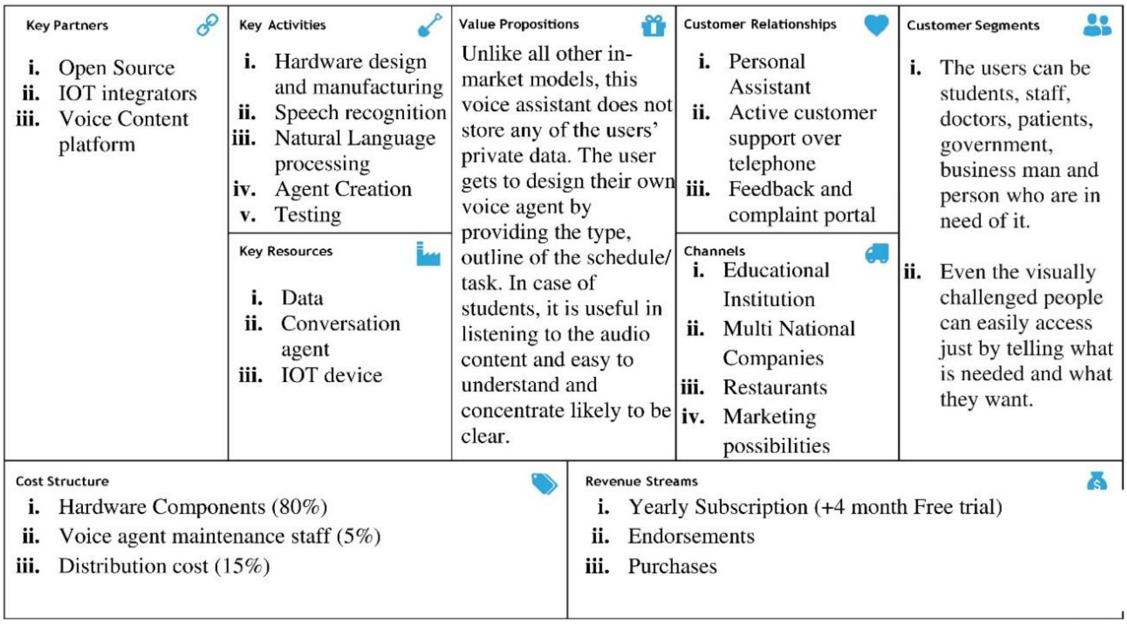


Figure 6.6 Raspberry Pi

Figure 6.7
USB Audio Connector

Figure 6.8
Raspberry pi display connected using VNC to laptop screen

BUSINESS MODEL CANVAS



Key Partners

i. Open Source
ii. IOT integrators
iii. Voice Content platform

Key Activities

i. Hardware design and manufacturing
ii. Speech recognition
iii. Natural Language processing
iv. Agent Creation
v. Testing

Key Resources

i. Data
ii. Conversation agent
iii. IOT device

Value Propositions

Unlike all other in-market models, this voice assistant does not store any of the users' private data. The user gets to design their own voice agent by providing the type, outline of the schedule/ task. In case of students, it is useful in listening to the audio content and easy to understand and concentrate likely to be clear.

Customer Relationships

i. Personal Assistant
ii. Active customer support over telephone
iii. Feedback and complaint portal

Channels

i. Educational Institution
ii. Multi National Companies
iii. Restaurants
iv. Marketing possibilities

Customer Segments

i. The users can be students, staff, doctors, patients, government, business man and person who are in need of it.

ii. Even the visually challenged people can easily access just by telling what is needed and what they want.

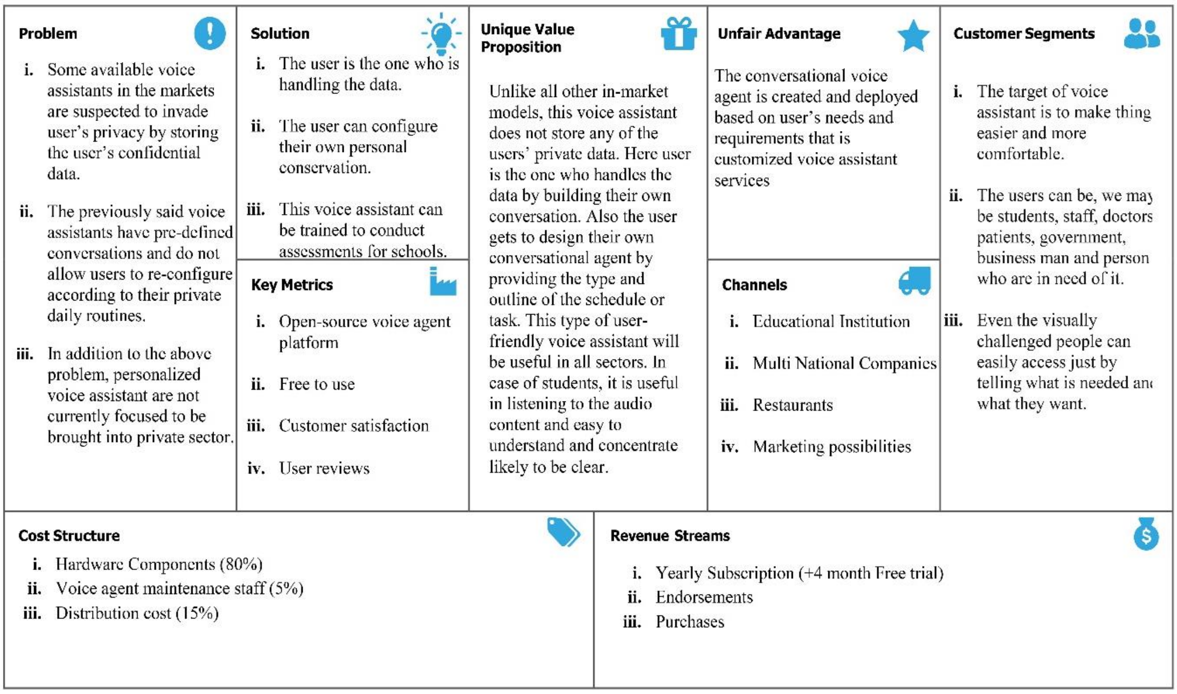
Cost Structure

i. Hardware Components (80%)
ii. Voice agent maintenance staff (5%)
iii. Distribution cost (15%)

Revenue Streams

i. Yearly Subscription (+4 month Free trial)
ii. Endorsements
iii. Purchases

LEAN CANVAS



Problem

i. Some available voice assistants in the markets are suspected to invade user's privacy by storing the user's confidential data.

ii. The previously said voice assistants have pre-defined conversations and do not allow users to re-configure according to their private daily routines.

iii. In addition to the above problem, personalized voice assistant are not currently focused to be brought into private sector.

Solution

i. The user is the one who is handling the data.

ii. The user can configure their own personal conservation.

iii. This voice assistant can be trained to conduct assessments for schools.

Key Metrics

i. Open-source voice agent platform
ii. Free to use
iii. Customer satisfaction
iv. User reviews

Unique Value Proposition

Unlike all other in-market models, this voice assistant does not store any of the users' private data. Here user is the one who handles the data by building their own conversation. Also the user gets to design their own conversational agent by providing the type and outline of the schedule or task. This type of user-friendly voice assistant will be useful in all sectors. In case of students, it is useful in listening to the audio content and easy to understand and concentrate likely to be clear.

Unfair Advantage

The conversational voice agent is created and deployed based on user's needs and requirements that is customized voice assistant services

Channels

i. Educational Institution
ii. Multi National Companies
iii. Restaurants
iv. Marketing possibilities

Customer Segments

i. The target of voice assistant is to make thing easier and more comfortable.

ii. The users can be, we may be students, staff, doctors patients, government, business man and person who are in need of it.

iii. Even the visually challenged people can easily access just by telling what is needed and what they want.

Cost Structure

i. Hardware Components (80%)
ii. Voice agent maintenance staff (5%)
iii. Distribution cost (15%)

Revenue Streams

i. Yearly Subscription (+4 month Free trial)
ii. Endorsements
iii. Purchases

VOICE FLOW AS BUSINESS MODEL

This innovation is suitable not only for personal needs but also for the requirements of a successful business. One of the most meaningful business benefits of voice assistants is the automated processes that they introduce in the business setup.

Consumers are embracing voice technology at unprecedented levels, and businesses must remain up to competition, or risk missing out. Many companies see more and more customers turning to voice assistance. But what they do not understand is just how easily a customer behaviour change happens. Since the smartphone, the adoption of consumer voice assistants has been faster than any product. If user fully grasp how, where, and why your customers are using voice, user can uncover new links between your company and their needs — and provide crucial assistance. Since voice tech is so fresh, the challenge is to leverage data from various sources for accurate, actionable insights. The user can then develop personalized, targeted, and frictionless experiences when people really need them.

With the speed at which this transition is happening, marketers need to start experimenting in this space. At its core however, it’s no different from any other marketing campaign. Smart speaker adoption has moved very quickly. Within only four years, 1 out of every 4 US adults use those voice-enabled intelligent speakers. This was two years ago in 2018. The pace with which its adoption has been growing; it is safe to say it would have achieved 50% market penetration right now. Although smart speakers were a catalyst for voice adoption, they are not the only drivers of growth. Far more customers have voice-enabled smartphones than smart speakers. Lately, a surge in adoption is seen in voice assistants in cars, watches, headphones, televisions, appliances, toilets, and many other devices.

CONCLUSION

The above-mentioned hardware and software components are integrated and tested for their working according to the trained conversation phrases in the voiceflow agent. Further number of agents can be integrated to this project to increase its working range thus making this project an all-in-one ecommerce management device. Many modules are of open-source systems and have customized those modules according to the presented system. This helps get the best performance from the system in terms of space time complexity. The Voice Command System has an enormous scope in the future. Like Siri, Google Now and Cortana become popular in the mobile industry. This makes the transition smooth to a complete voice command system. Additionally, this also paves way for a Connected Home using Internet of Things, voice command system and computer vision. It is greeting the way the person feels greater comfortable and to interact with the voice assistant.

REFERENCES

1] Dahl, George E., et al. "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition." Audio, Speech, and Language Processing, IEEE Transactions on 20.1 (2012): 30-42.

[2] Chelba, Ciprian, et al. "Large scale language modeling in automatic speech recognition." arXiv preprint arXiv:1210.8440 (2012).

[3] Schultz, Tanja, Ngoc Thang Vu, and Tim Schlippe. " GlobalPhone: A multilingual text & speech database in 20 languages." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on IEEE, 2013.

[4] Tokuda, Keiichi, et al. "Speech synthesis based on hidden Markov models" Proceedings of the IEEE 101.5 (2013): 1234-1252.

[5] Bibi Asma Desai. Smart Voice Assistant for IOT: Researchgate.net – 2019
https://www.researchgate.net/publication/336421080_Smart_Voice_Assistant_for_IOT.

[6] Frank Gu. Transforming your Raspberry PI into a simple voice assistant: voiceflow.com – 2021
https://www.voiceflow.com/blog/transforming-your-raspberry-pi-intoa-simple-voice assistant.