International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 4

Smart Speaker using Raspberry Pi

Mallesh S Jambagi, Basavaraj H, Gagan L, Prateek R Prabhu
SET, Jain University, Bengaluru, Karnataka, India

ABSTRACT

As the software systems are getting more and more complex each day, the ability of the user to put up with this complexity is reducing. An elegant solution to deal with this issue is to build a personal assistant agent which is capable of determining user's habits, preferences and intentions and helping them accordingly. This work discusses on developing a "Smart Speaker using Raspberry Pi". The work aims at the development of a personal voice assistant which can assist users in performing their personal and professional tasks using speech commands and provide more interactive and user-friendly experience. The assistant also helps in controlling home appliances and IoT devices and reduce the workload of the user. The major element of the Smart Speaker is the Raspberry Pi. The Raspberry Pi collects speech input and interprets it to manage certain tasks.

Keywords: *IPA, A.I., AVS, IoT, NLU, TTS, STT smart speaker, personal assistant, etc.*

1. INTRODUCTION

Our digital life is determined by innovations. Especially in recent years, more innovative technologies were developed to facilitate our professional and daily life. Intelligent Personal Assistants (IPA) are an important achievement, which have become an essential part of the universal digitalization process. These assistants are now available in all gadgets such as smartphones, tablets and even smart watches. Advancements in the field of Machine Learning, Artificial Intelligence and Natural Language Processing has led to the development of IPAs. The increasing competition in this area has made the IPAs more advanced and interactive.

Google Home which was introduced in 2016 is a hands-free speaker to control with voice commands. The device connects to the Google Assistant API to perform various tasks such as play music and instantly provide information such as news, weather and sports scores. It was observed that personal assistants like Google Home failed to provide the user with a sense of control as it sometimes remained unresponsive when given voice commands that are not valid and these personal assistant devices are expensive. Furthermore, integrating them to household appliances like lighting requires one to purchase lights such as the Philips Hue which add to the overall cost of automating a house.

2. SMART ASSISTANTS

The core issue behind the need of Smart Assistance (SA), regardless of the domain in which it is applied for, is the user's lack of knowledge. The user does not have a complete knowledge that would help him/her to achieve his/her goals, and therefore, assistance is needed to update the user's knowledge to achieve these goals. With the expansion of cyberspace, and the enormous process in computing and software applications, technology is covering every aspect of our lives, and therefore, many of our tasks and goals are now technology driven. Consequently, the problem of lack of knowledge has increased; as the user now might be required to work with many complicated applications to achieve his/her goal. Therefore, a form of smart assistance, beyond the user interface, is essential.

Gabriela Czibula Et al. discusses two of the most important issues that personal assistant agent have to deal with are learning and adapting to the users preferences and to solve this the assistant has to continuously improve its behavior based on the experience of the actions taken by users that successfully achieved a specific task, this way the agent has to be endowed with the learning capability,

thus becoming able to adapt itself to its dynamic environment [4].

Ke-Jia Chen Et al. proposes a memory mechanism for personal assistant agents in order to enhance agent intelligence while working with the user or with other agents. His work focuses on an attempt to improve the competence of a PA agent by making it more intelligent [5].

3. IMPLEMENTATION

The final goal of this work is categorized into two parts:

The focus of this work is to assist users in their personal and professional life and also to control home appliances with speech commands.

The work also aims at building an inexpensive personal assistant which is achieved using a Raspberry Pi that used Amazon Alexa Voice Services (AVS) to convert spoken text, picked up using a microphone to written text.

The work helps in consumers' access to a hands-free personal assistant that uses speech or gesture commands to interact with appliances in a house at 1/3rd the cost of devices like the Google Home and Apple Home.

To implement the above goals, the following methodology is followed:

- Understanding the dynamics of each part of the system.
- Determination of required project hardware components.

3.1 Hardware Implementation

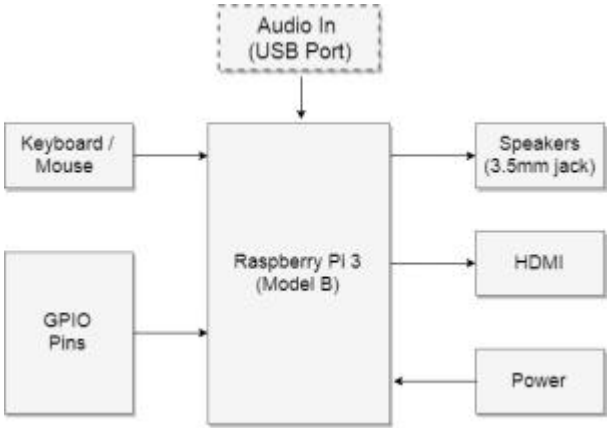

Figure-1: Hardware Setup.

Microphone is interfaced to raspberry pi using one of the USB 2.0 ports. The voice input to the raspberry pi is given through microphone and then it is further passed to the system where specific keywords are identified.

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.

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, 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.

GPIO Pins are one powerful feature of the Raspberry Pi which is located along the edge of the board. These pins are a physical interface between the Pi and the outside world. These act as switches that can be turned on or off (input) or that the Pi can turn on or off (output).

3.2 System Events Flow

First, when the user starts the system, he 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 computer 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 computer and can also be processed by it.

Then that text is parsed and searched for keywords. Our voice command system is built around the system of keywords where it searches the text for key words to match. 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. OCR categorizes and 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.

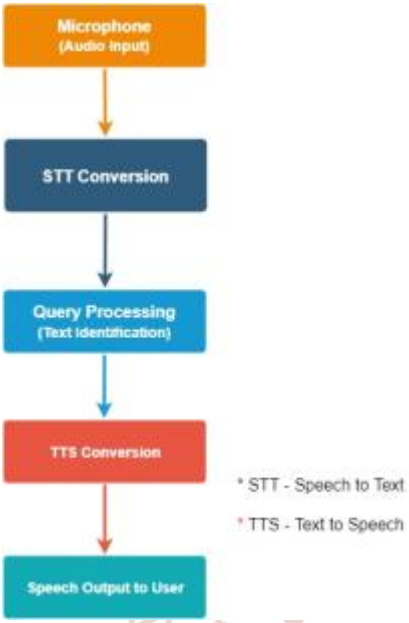

Figure-2: Event Flow Diagram.

4. SOFTWARE MODULES

Speech To Text Engine

(AVS) is a 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 (AVS) engine, an application has to be created in the Amazon developers console and the generated API key has to be used to access the speech engine. It requires continuous internet connection as data is sent over the Amazon servers.

Text To Speech Engine

(AVS) is a 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 (AVS) engine, an application has to be created in the Amazon developers console and the generated API key has to be used to access the speech engine. It requires continuous internet connection as data is sent over the Amazon servers.

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 about famous personalities, simple mathematical calculations, description of any general object etc.

5. RESULTS

The Smart Speaker System works on the idea and the logic it was designed with. Our personal assistant uses the button to take a command. Each of the commands given to it is matched with the names of the modules written in the program code. If the name of the command matches with any set of keywords, then those set of actions are performed by the Voice Command System. The modules of Find my iPhone, Wikipedia and Movies are based upon API calling. We have used open source text to speech and speech to text converters which provide us the features of customizability. If the system is unable to match any of the said commands with the provided keywords for each command, then the system apologizes for not able to perform the said task. All in all, the system works on the expected lines with all the features that were initially proposed. Additionally, the system also provides enough promise for the future as it is highly customizable and new modules can be added any time without disturbing the working of current modules.

Figure-3: Initial Model

Figure-4: Final Model

6. CONCLUSION

In this paper, introduced the idea and rationale behind the Voice Command System, the flaws in the current system and the way of resolving those flaws and laid out the system architecture of the presented Voice Command System. 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.

REFERENCES

1. Dr. V. Chayapathy, Dr. Anitha G S, Sharath B, "IoT Based Home Automation By Using Personal Assistant", International Conference On Smart Technologies For Smart Nation (SmartTechCon), IEEE 2017.
2. Ass. Prof. Emad S. Othman, "Voice Controlled Personal Assistant Using Raspberry Pi", International Journal of Scientific & Engineering Research Volume 8, 2017.
3. Ali Hussain, "Personal Smart Assistant for Digital Media and Advertisement", Western Libraries, The University of Western Ontario, 2013.
4. Gabriela Czibula, Adriana-Mihaela Guran, Istvan Gergely Czibula and Grigoreta Sofia Cojocar, "IPA - An Intelligent Personal Assistant Agent For Task Performance Support", IEEE 2009.
5. Ke-Jia Chen and Jean-Paul Barthès, "MemoPA: Intelligent Personal Assistant Agents with a Case Memory Mechanism", ICIC 2007.