Journal of Engineering Sciences

Vol 11, Issue 11, NOV/2020
ISSN NO:0377-9254

Offline Voice Recognition With Low Cost Implementation Based Intelligent Home Automation System

M.Kamala[1](Associate Professor)
Department of Computer Science and Engineering
St. Peter's Engineering College, Dullapally, Maisammaguda,
Medchal, Hyderabad, Telangana 500043
kamalam@stpetershyd.com

Gone Bhanusree (UG Student)
Department of Computer Science and Engineering
St. Peter's Engineering College,
Dullapally, Maisammaguda,
Medchal, Hyderabad, Telangana
500043

Priyank Thakkar (UG Student)
Department of Computer Science and Engineering
St. Peter's Engineering College,
Dullapally, Maisammaguda,
Medchal, Hyderabad, Telangana
500043

Rajeshwar Patel (UG Student)
Department of Computer Science and Engineering
St. Peter's Engineering College,
Dullapally, Maisammaguda,
Medchal, Hyderabad, Telangana
500043

***Abstract*—Sound intelligence is added to a home automation based on acoustics for sophistication of physically challenged people as a broad perspective of the thesis. Home automation already in practice is by switching on or off a device via wired networks. But this is inefficient for people with impairment in mobility or spinal cord disability due to ageing factors. The easiest way of controlling a device is through human voice. Existing products are expensive and also speech recognition is available with usage of internet (online). This paper is presented with the concept of speech recognition being implemented using Ivona in an offline manner (without internet). Human voice is converted to text using Ivona and it is wirelessly Trans received using GSM modems. According to the received texts appliances can be controlled. The module contains a secured speech recognizer for automatic door opening/closing and a general voice recognizer to control appliances like television, music player, fan, light, etc. All the above are implemented in a low cost Raspberry Pi board. Thus a goal of producing an automation device has been designed at low cost using offline speech recognition**

Keywords*—*RASPBERRY PI , Ivona

I. INTRODUCTION

Home automation can be understood as the uptake of technology in residential environment in order to enhance the quality of life of its inhabitants by providing services such as telehealth, multimedia entertainment, and power saving. In this context, this technology focuses on investigating the means to improve the quality of life of inhabitants, in particular disabled people. The home automation industry has been growing rapidly, which is mainly fueled up by the need to provide systems to support elderly and disabled people, mostly automation must adapt to the needs of these citizens. The population structure of industrialized countries is changing; there will be less people to take care of the growing number of elderly citizens. This problem has encouraged a large number of studies involving robotics and automation Rather than live in an elderly care institution, it is more feasible to inhabit home. In addition, disabled people help from someone to do these tasks. In the absence of someone to help, their lives become highly difficult. Therefore, the development of a system to assist them to turn on and off their home devices is crucial. The breakthrough in cutting edge voice command technology has an enormous potential in solving this problem. 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. Speech Recognition is the ability of machine for instance a computer to understand words and sentences spoken in any language. These words or sentences are then converted to a format that could be understood by the machine. Speech recognition is basically implemented using vocabulary systems. A speech recognition system may be a

Small Vocabulary-many user system or a Large Vocabulary-small user system.

II. System Architecture

Existing Systems

The existing systems suffer from the drawback that only predefined voices are possible and it can store only limited voices. Hence, the user can't get the full information coherently

Proposed Systems

The proposed system is such that it can overcome the drawback of the existing system. The project design involve text to speech. Here whatever the system receives as input after the command the output will get in the form of voice means speech.

Hardware Implementation

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 as shown in Figure 1.

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

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

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 is being used to provide internet connection to control switch system. Since the system relies on offline text to speech conversion, offline query processing and offline speech to text conversion hence we need a don't need a constant connection to achieve all this, but for controlling any switch's/equipment's there is a requirement for a ethernet connection/WIFI module.

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.

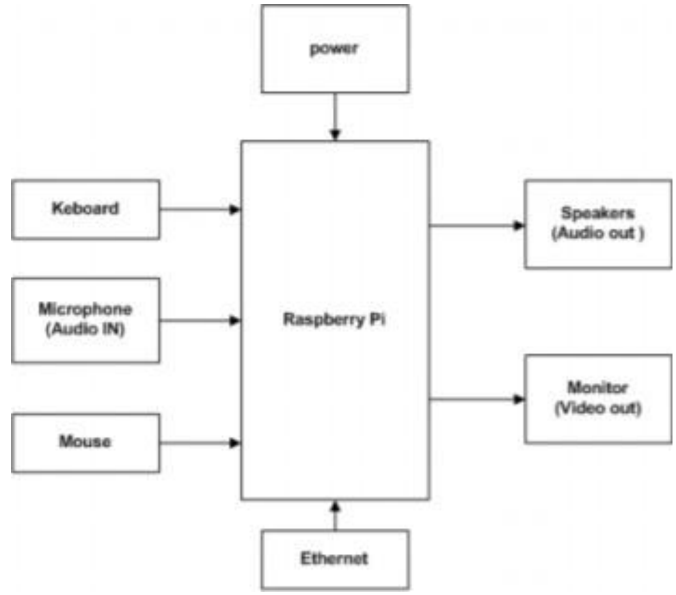

Figure 1. Hardware setup of voice command system

S.no	Component	Cost
1.	Raspberry Pi 4 Zero	₹ 2199.
2.	USB Microphone	₹ 199
3.	USB Speaker	₹ 150
Total Cost		₹ 2548

Table I . Cost analysis of total hardware model

Flow of Events in Voice Command System

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 as shown in Figure 2.

Vol 11, Issue 11, NOV/2020
ISSN NO:0377-9254

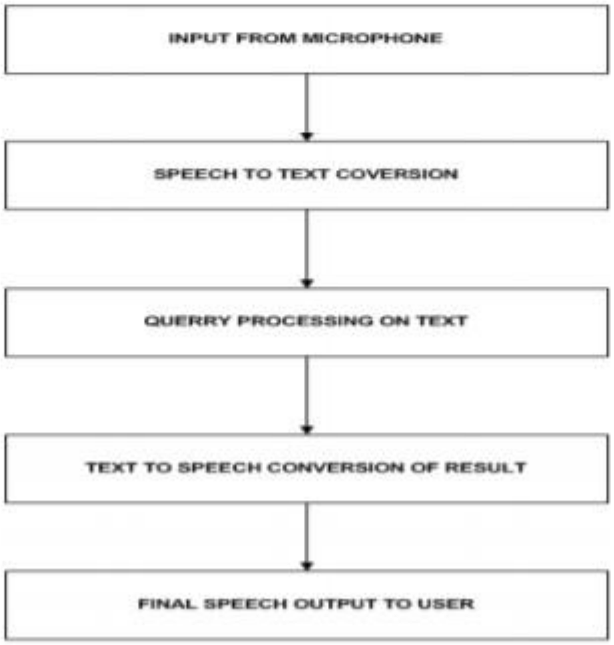

Figure 2. Flow of events in voice command system

III. Modules Implemented

Speech To Text Engine

SpeechRecognition package 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 which does not require any internet connection to compute.

Text To Speech Engine

Ivona 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.

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.

Wikipedia

This module works on the keyword of “wiki”. The system asks for what you would like to learn about. Then the request is made to the Wikipedia API for the required query. It generates the summary of the information regarding the query and the data is output through the microphone to the listener in audio form. In case of failure, the error message is generated saying “unable to reach dictionary of wiki”.

News

News module can be executed by using the keyword “news”. The headlines of top articles are retrieved from the internet using Google news. The system tells the user all these headlines and asks the user if any of these articles should be sent to the user’s email address. If the user specifies the number of the article to be sent, the article is sent to the specified email address. Otherwise, no further action is taken. If any failure occurs in retrieving headlines or sending articles, a corresponding error message is generated.

Weather

This module tells the user about the weather conditions of the location whose station identifier is specified in the profile of the user. This module can be executed by using the keyword “weather”. The weather information is taken from the weather underground service which includes the details of temperature, wind speed and direction etc. It generates an error message, if the information cannot be retrieved for the specified location.

Movies

The voice command system searches for relevant movie by the keyword “movie”. It is implemented using a function which asks the user, the movie they want to know more about. Then a function is called to ask for the movie the user needs to know about. It searches for the top five results regarding the movie name. It confirms the movie among the listed names. On confirmation of the movie, it gives the detailed information including the rating, producer, director, cast, runtime and genres. In case of the failure, it generates the error of “Unable to find information on the requested movie”.

Define

Define module is used to fetch the definitions of word specified by the user. The execution of this module is started by the keyword “define”. Then, the user is asked for the word whose definition has to be provided. The module checks for the definitions of the specified word using the Yandex Dictionary API. All the available definitions are provided as a response from the API and the same is given as output to the user in audio form. If the connection to API cannot be established properly, then the error message “Unable to reach dictionary API” is generated for the user.

Find My IPhone

Find my iPhone module lets the users find their iPhone by their voice. This is done using the user’s iCloud id and password to make the connection to the server. After the authentication of the password it checks the status of the iPhone. After checking, it connects to iPhone using the iCloud. After the connection, the command for ringing is sent to the iPhone to start a sound and notification on the phone for detection. If a failure occurs, the system generates an error that the iPhone is not found. An error is also generated when there are multiple iPhones associated with a single apple id.

Joke

Joke module can be used for entertainment purposes by the user. This module works on the keywords “joke” or “knock knock”. The jokes used in this module are predefined in a text file from which the jokes are read in a random order. A start and end line is present in every joke to differentiate it from others present in the file. All the lines of a joke are spoken by the system in the specified order only.

Unclear

This module is used to generate an error message if the keyword specified by the user does not match the keywords present in any module of the system. This module has an empty array of keywords and is allocated the least priority, so that it is executed only after the keyword has been matched with all the other modules of the system. The messages generated by this module include statements like “I beg your pardon”, “Say that again”, “I’m sorry, Could you repeat that” and “My apologies, Could you try saying that again” which are selected in a random order each time this module is executed.

Other Command Specific Modules

The Voice Command System also has some command specific modules like fetching hacker news, email and current time. Each of these modules is related to the system using keywords like “hacker news”, “email” and “time” respectively. Whenever any of this keyword is said to the system, it fetches that module and launches the contents of that module thereby providing the appropriate response to the user.

IV. Conclusion

In this paper, a speech-based home automation is done at a low cost as shown in Table I. The system contributes towards empowering the privacy, aged, impaired to meet their needs. Accuracy can be improved by incorporating various filters. Thus, an offline low-cost speech recognition hard model has been implemented using Raspberry Pi board.

V. Future Scope

One of the most important improvement which can be done is the addition of native languages. It is a hectic process yet it will be useful to all the people who do not know English. It makes this device usable by almost all of the people in the world. By the addition of native languages, the device becomes much more user friendly and easily accessible. Machine Learning should be implemented completely into this system. By that way, the system will be able to learn new processes by itself and adapt to the user based on its past experiences. This makes it easier for the user to interact with the assistant as well. Memory can be improved and let the system store the new information gathered from the user and use it in the future if required. Much more modules can be created based on the necessities.

VI. Acknowledgment

This research paper has been equipped as a result of discussion among the faculty members of my college. The authors like to thank Mrs. Kamala (Associate Professor) of St. Peter’s Engineering College, Hyderabad, Telangana for their constant guidance and support during the research work time period.

VII. References

[1] Rouse, M. Speech Recognition, Available: https://searchcrm.techtarget.com/definition/speech-recognition.
[2] Boyd, C. The Past, Present and future of Speech Recognition technology, Available: https://medium.com/swlh/the-past-present-andfuture-of-speech-recognition-technology-cf13c179aaf.
[3] A short history of Speech Recognition, Available: https://sonix.ai/history-of-speech-recognition.
[4] van der Valde, N. Speech Recognition technology overview, Available: https://www.globalme.net/blog/the-present-future-of-speechrecognition.
[5] Bennett, I.M., Babu, B.R., Morkhandikar, K., Gururaj, P. 2015. Distributed Real-time Speech Recognition, Naunce Communication Inc., Patent No. US 9,076,448.
[6] Bennett, I.M., Babu, B.R., Morkhandikar, K., Gururaj, P. 2014. Speech Recognition System Interactive Agent, Naunce Communication Inc., Patent No. US 8,762,152 B2, 24 June 2014
[7] Deng, L., Hinton, G., Kingsbury, B. 2013. New type of Deep Neural Network learning for speech recognition and related applications: an overview, IEEE International Conference on Acoustics, Speech and Signal Processing (13886524), ISBN-(978-1-4799-0356-6), 26-31 May 2013.
[8] Chadha, N., Gangwar, R.C., Bedi, R. 2015. Current Challenges and Applications of Speech Recognition Process using Natural Language Processing: A Survey, International Journal of Computer Applications (0975-8887), 131(11), 28-31.
[9] Petkar, H. 2016. A review of Challenges in Automatic Speech Recognition”, International Journal of Computer Applications (0975-8887), 151(3), 23-29.
[10] Ramírez, J., Górriz, J.M., Segura, J.C. 2007. Voice Activity Detection, Fundamentals and Speech Recognition Systems Robustness, University of Granada, Spain.