ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 6, June 2017

Providing Voice Enabled Gadget Assistance To Inmates of Old Age Home Including Physically Disabled People

[1] Akshatha P, [2] Abhirami Balaraman, [3] Dr.Rathna.G.N
[1] IBM, Bangalore, [2] Abhirami Balaraman Wipro Technologies, Bangalore
[3] Principal Research Scientist Indian Institute of Science, Bangalore

***Abstract*— Speech recognition is one among the most recently developing field of research, at both industrial and scientific levels. In this paper we have developed a gadget assistance system for an old age home, where the operational commands for a TV was provided as inputs and tested for its performance. This system is configured for both online and offline speech recognition. The online speech recognition uses DNN algorithm, whereas the offline speech recognition uses HMM algorithm. This gadget is also used for automation of devices such as television with the help of a LIRC module. Various speech recognition modules for playing Music, Videos, Games and Television control were written and executed. Accuracy of speech recognition was observed.**
]

I. INTRODUCTION

Speech recognition is currently considered to be one of the most recent and upcoming field in technology. Automation by speech can help not only daily activities on an individual but also has a larger scope when it comes to industrial applications [1]. Our project portrays the use of speech automation to provide information and entertainment to otherwise solitary people; it acts as a personal assistant. Speech recognition can aid people with disabilities like the blind. It can also be a means of relief to people who have difficulties in mobility like the old-aged. It can be used for educational purposes as well as to teach those who have these shortcomings to control and manage computers; since everything in today's world is online, technical knowledge has gained utmost importance. Though there are many challenges in the domain of speech automation, due to the fact there are many languages internationally and in India alone we have 14 languages and different dialect for each. Training a software on all these phonetics is not an easy task [2]. Yet, there has been much progress in the field, and many famous recognition systems are now in market. For Example: IBM's Watson, Google Speech recognition and so on [3].

In, this paper we have added functionality to not only recognize a word but also to perform actions such as play music, videos, games and control Television. Further implementations are feasible and are discussed later in this paper

II. SPEECH RECOGNITION ALGORITHMS

a) Offline Speech Recognition

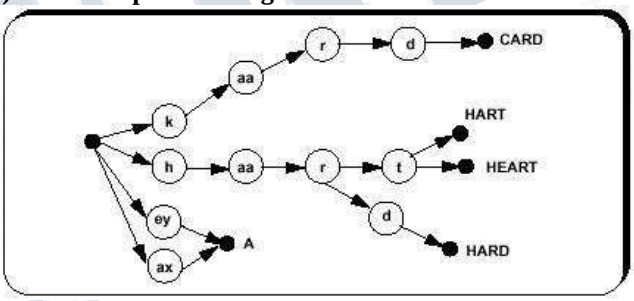

Figure 1

The offline speech recognition is implemented using the HMM Algorithm as illustrated in Figure 1 above.

b) Online Speech Recognition

The online speech recognition is implemented using the DNN Algorithm (Deep Neural Networks), as shown in Figure 2 below.

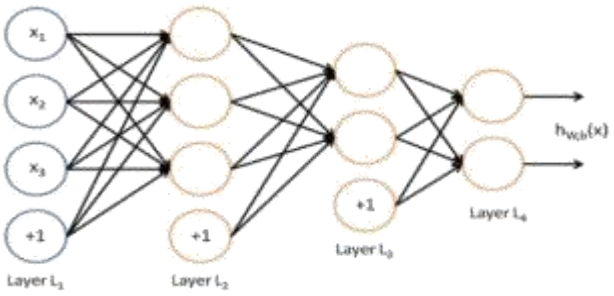

Figure 2

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 6, June 2017

HMM

A general-purpose speech recognition system is based on Hidden Markov Models (HMM), being statistical models that output a sequence of symbols or quantities.

DNN

Most current speech recognition systems use hidden Markov models (HMMs) to deal with the temporal variability of speech and Gaussian mixture models (GMMs) to determine how well each state of each HMM fits a frame or a short window of frames of coefficients that represents the acoustic input.

A DNN (Deep Neural Network) is a feed-forward, artificial neural network, that has more than one layer of hidden units between its inputs and its outputs.

III. EXPERIMENTAL SETUP AND IMPLEMENTATION

The schematic of the setup is shown in blocks as in figure 3 along side.

Raspberry pi Board

The major component required is the raspberry pi board as shown in Figure 4. This is a palm sized computer that can be plugged into a computer monitor or TV. This could be used to program the speech recognition code in python. In a Shut down state the pi utilizes a power of 0.05A = 0.26 Watts, 5.23 V

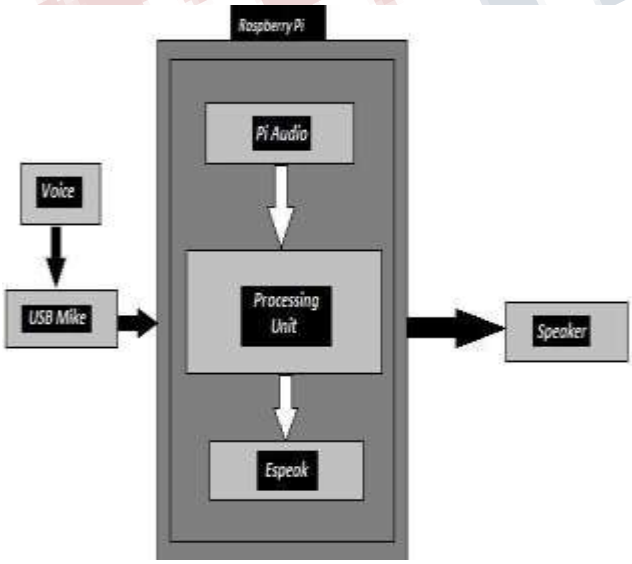

Figure 3: Block Diagram of Speech Recognition.

a) First the keyword configured is said, we will hear a high beep, which means the unit is awaiting response.

b) The command given, is decoded using the HMM algorithm.

c) The block diagram of speech recognition consists of the main unit; i.e. the processing unit which implements the deep learning algorithm.

d) Modules have been coded for various functions.

e) Match is found to mentioned words in modules and the appropriate function is executed. Which can play a song or a video or reading a book or changing TV channel, it also has an additional feature of playing a quiz game.

f) The song and video database can have any regional language songs as well.

g) The output of the system is then heard through the speakers.

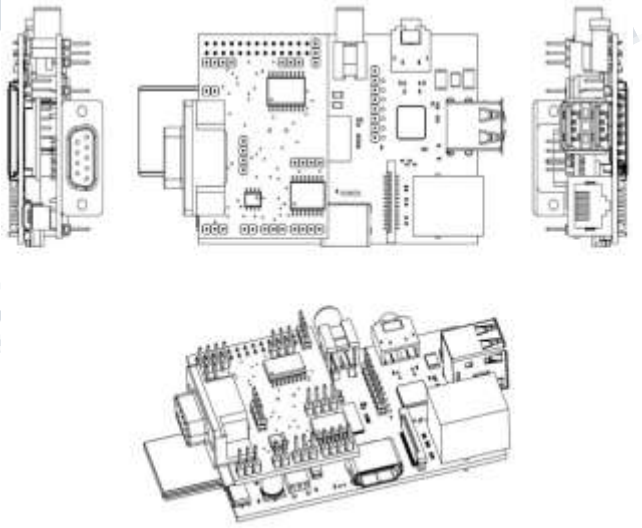

Figure 4

The raspberry pi board with raspbian OS was used to implement, the speech recognition HMM algorithm. The programming language used for the coding was Python 2.7, as the raspberry pi has an inbuilt python IDE. Python is also a very popular language used in many speech recognition systems. The code was written to not only use the HMM algorithm to recognize the words, but also to perform certain automated operations. Different modules were written for each function. The modules, the code consisted were Music, Video, TV, and Quiz .These modules could be considered as a pioneer for further possible operations.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 6, June 2017

The flowchart of the program is shown in Figure 5.

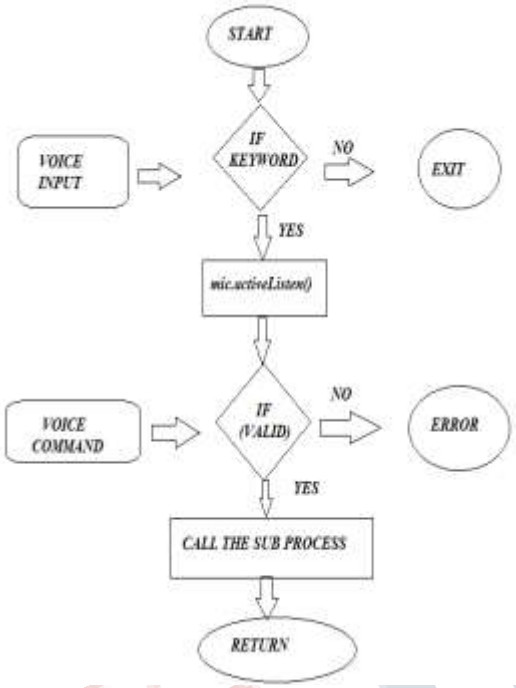

Figure 5

The speech signal that was input through the microphone, was passed through a noise removal program, then sampled into chunks of data .The gender of the speaker and varying pitch of human voice was not considered for simplication of the recognition program .This is then passed into the HMM code, decodes the speech signal into text using phenomes of the uttered word. The text is then differentiated into two types: command or name.

Command: If the recognized word was a command the respective module would start. For Example: Music would start the Music module. Command words are also used to control what happens in a module or when to stop it.
Different modules have their own set of command words.

Name: After the program has entered a module it requests for name of the song or video or TV channel. The word spoken in response to this is the Name of the file to be accessed.

Song or Video: When a song or video is requested python opens a pipe to run the music player also with the main thread. The control commands are used now to "Stop", "Volume up" or "Volume down".

TV module: When in the Television control module, LIRC is used to connect to the IR sensor and IR transmitter in the circuit as illustrated in Figure 6 below, to communicate with the TV set-top box.

The code would use LIRC module to send out IR codes of the digits of the channel number individually, with no delay so that it resembles an action of entering the number manually. The channel name and number were hardcoded. This mechanism over-rode the need to remember the numbers of each channel you need to watch, the task is simplified as you just have to say the name of the channel for that particular channel, and the module drives to that particular channel. LIRC is software that allows you to decode and send infra-red signals of many (but not all) commonly used remote controls. We can use LIRC with Raspberry Pi.But before that, we need to connect some parts that are capable to send and receive/decode IR signals. If IR data needs to be sent, then an IR Led would be required.

The schema illustrated in Figure 6 is required for Connecting the IR Led.

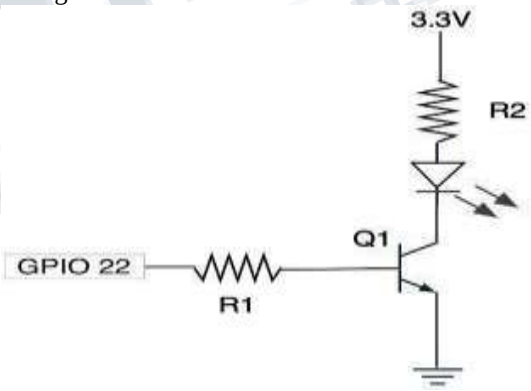

Figure 6

- Q1 = BC547 it's a transistor used to give the IR LED
enough power.
- R1=220 ohm
- R2 can be skipped
- 940nm IR LED 40 deg

Using the IR transmitter and receiver we have programmed to switch ON or change the TV Channels with a voice input.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 6, June 2017

An IR LED, i.e the .IR transmitter, is a special purpose LED that transmits infrared rays in the range of 760nm wavelength. They along with IR receivers are commonly used as sensors. When IR light falls on the photodiode, its resistance and correspondingly its output voltage, change in proportion to the magnitude of the IR light received. This is the underlying principle of working of the IR sensor. There could be improvements done on both hardware and programming. The words would be recognized by the algorithm only if they are present in the dictionary hardcoded with each module, since we are going with the no internet approach. The constraints due to the memory and processing speed can be changed with the hardware used. The programming can be altered to consider not only the phenomes but also the pitch, amplitude, external noise origin of speech signal. This could considerably improve the recognition accuracy. The files accessed by the modules are the ones present in the memory of the board. But modules can be written to download requested songs or videos when the program is expanded to include online support. The Television control module could be expanded to automate any other devices with remote controls, for example Computer, AC .With further research, modules could also be written to control any device with an intelligent response system

In conclusion, the project is an initial prototype to portray the possibilities of speech based automation as an instrument for aiding our society and its needs.

IV. ONSITE TESTING

We chose to test the performance of our device, by demonstrating a prototype to inmates of an old age home, and gathered their experiences. Based on the feedback, we made changes to our programs to recognize more familiar keywords, rather than the complicated ones already in place. We were able to obtain acceptable accuracy with the recognitions, with limited database. The device played songs, videos and changed the channels based on their names.

V. CONCLUSION

- The device performs well on a limited dictionary of words, but on increase of number of words the accuracy drops down.
- Being functional offline, there is no need for network connectivity at all times.
- This makes the device very useful in areas without internet access.

VI. ACKNOWLEDGMENT

We would like to thank the members of Hiteshi Mahila Maneyangala Old Age Home, Bangalore for giving us an opportunity to test our device. We would like to thank them for their time, patience and valuable feedback to improve the project.

REFERENCES

[1] D. Yu and L. Deng"Automated Speech Recognition: A Deep Learning
Approach, " Springer Pub., 2014
[2] Claudio Becchetti and KlucioPrinaRicotti. "Speech Recognition :
Theory and C++ implementation," Wiley-Blackwell pub., 1999
[3] Douglas A Renoylds and Richard C Rose, "Robust text- independent speaker identification using gaussian mixture speaker models," IEEE Transactions on Speech and Audio processing, Vol. 3, No.1, pp72-83,
1995