Voice Control Device using Raspberry Pi

Pooja Singh[1], Pinki Nayak[2], Arpita Datta[3], Depanshu Sani[4], Garima Raghav[5], Rahul Tejpal[6]

[1,2,3,4,5,6]Dept. of Computer Science & I.T.,
Amity School Of Engineering & Technology, Affiliated by GGISPU, Delhi
New Delhi, India
[1]wizkid_pooja@yahoo.com, [2]pinki_dua@yahoo.com

Abstract: This paper shows the working of a device based on implementation of a voice command system as an intelligent personal assistant. The services provided by the device depends on the input given in the form of voice command by the user and ability to access information from a variety of online sources such as weather, telling time or accessing online applications to listen to music.

This Voice driven device uses Raspberry Pi as its main hardware. Speech to text engine is used to convert the voice command to simple text. Query processing is then applied using natural language processing (NLP) onto this text to interpret the intended meaning of the command given by the user. After interpreting the intended meaning, text to speech conversion is used to give appropriate output in the form of speech.

This device might provide a platform to visually impair to do their day to day tasks more easily like listening to music, checking weather conditions, checking current time or even doing a simple mathematical calculation. Many experiments and results were accomplished and documented.

Keywords: Virtual Personal Assistant, Natural Language Processing, Query Processing, Raspberry Pi.

I. INTRODUCTION

A virtual personal assistant can be seen as a software agent that understands natural language voice commands and completes tasks for the user. Such tasks were earlier performed by a personal assistant or a secretary that included tasks like dictation, reading text or email messages aloud, searching for contacts, scheduling, making phone calls and setting reminders for appointments. But in today's world all of these tedious tasks that had to be managed by a single person earlier have been made easy, effective and more efficient by using a device. This devicehas edge over the people that were hired for the tasks that needed to be performed for some simple and clear reasons,which are, it doesn't get tired so it can work efficiently all day long, it only needs a one-time investment during its purchase and needs no salary thus saving the cost of the user, it is compact and easy to carry anywhere, it can easily save much more data and there's vey less probability of data loss, it can also be used for personal use at home like for listening to music or setting a timer or setting an alarm. Currently most popular personal assistants are Alexa by Amazon, Siri by Apple and Cortana by Microsoft - generally used for Windows 8.1 and Windows 10.

Personal assistants like Alexa and Google Home have inbuilt hardware components like microphones and speakers. Some of these assistants also have inbuilt LED displays. As interpreting sounds takes up a lot of computational power, devices like Alexa and Google Home use their own server for speech recognition and task identification. There is a huge dataset on their server which helps them attain high accuracy and efficiency.

Meanwhile in this device hardware components like microphones and speakers needs to be attached externally. This device uses Google API to convert speech into text. This text is further processed through NLP. Processing is performed on the device itself as there are limited dataset resources and no server accessibility.

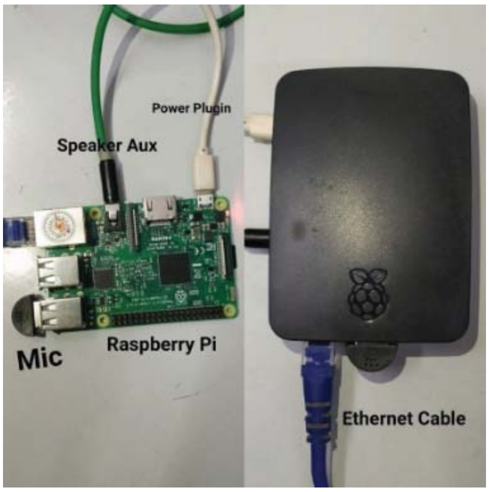

Fig. 1. Roomie

The device that is used as virtual personal assistant uses Raspberry Pi as its major component. Raspberry Pi can be thought of as a small and affordable computer. It has ports on it through which other components of the device like speaker and mic can be connected to it. These components are then used to take in command from the user as input and then give out the results as outputs.

978-1-5386-9346-9/19/$31.00 ©2019 IEEE

The device takes in the command from the user through the microphone being plugged into the raspberry pi in the form of input. The user gives its command in his natural language to the device. The vocal command is converted to plain text by speech to text synthesis .The device then performs query processing using natural language processing (NLP) on the plain text derived from the command given by the user. During query processing the device tries to interpret the intended meaning of the instruction given by the user. The keywords are searched in the sentence and the data that seems irrelevant to the device is ignored. Device figures out intentions of the user by using these keywords from the sentence. After the command is interpreted device simply performs the given task that user asked it to do. The performance of the device varies in different situations. Sometimes the input that the user gives through his vocal commands is not taken in or considered by the device. One reason for this is that the distance between the mouth of the user and the microphone is very large. Another reason maybe that the environment in which the user is giving the command to the device is very noisy. Also sometimes the device may give out some output that the user was not expecting, this problem may arise due to the unusual accent used by the user.

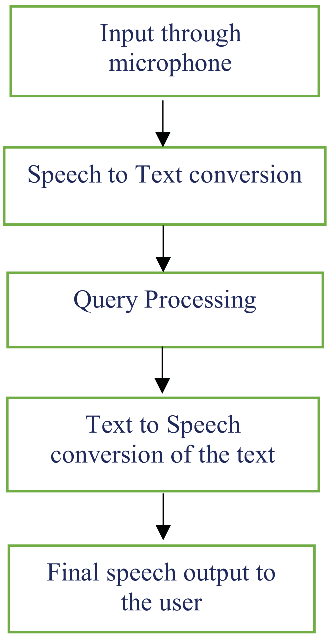
Input through microphone
Speech to Text conversion
Query Processing
Text to Speech conversion of the text
Final speech output to the user

Fig. 2. Working

II. HARDWARE IMPLEMENTATION

A. Microphone

The vocal commands given by the user which is used as input is given in through the microphone that is connected to the device. This vocal command is then later converted to simple text and keywords are searched through this text which helps the device to perform its functions and give out the expected results.

B. Raspberry Pi

Raspberry Pi is the major component of the device. It acts as a mini computer. It is indulged in all the activities since the beginning when the user gives the input till the end when the output is presented to the user. It sorts of binds all the components together. All the processing of the data takes place here.

C. Ethernet

The Ethernet cable helps us to provide the internet connection to the device. Internet plays a very important role in the operation of the device as it helps the device to do speech to text conversion, query processing through NLP and text to speech conversion. All these processes take place online that’s why the internet connection is very essential.

D. Speakers

Speaker performs the last function in this process. The speaker helps the device to give out the output in the form of speech that is being converted from the text online. The speaker can be connected to the device through an AUX cable.

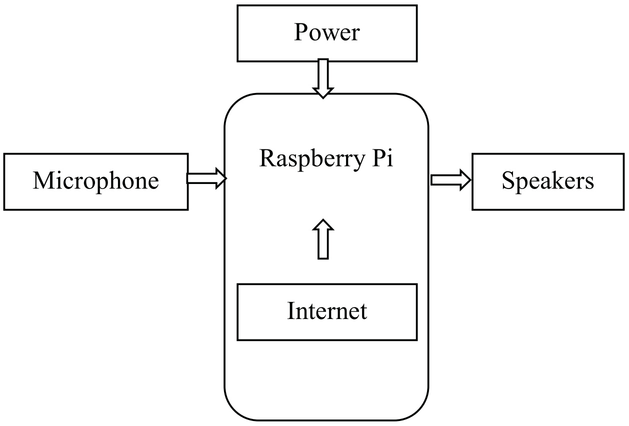

Fig. 3. Hardware setup

III. SOFTWARE IMPLEMENTATION

A. Input through microphone

The user gives in his command verbally to the device through the microphone in the form of input that the device later will process on. But before using the microphone it needs to be configured properly.

B. Speech to text conversion

The input that is given by the in the form of a vocal command is first converted to the plain text. This action is performed by

using the Google speech recognition API. This API can recognize around 120 languages and variants so that it can support global user base. It's easy to use and very effective.

C. Query Processing

This part of the whole process is the most important one. This process uses natural language processing (NLP) to operate. The input that is converted to the text by the device is studied in this step. Whole text is analysed, thereafter the tokens are identified from the text being received. By considering the tokens selected by the device it tries to interpret that what could be the action that the user wants it to perform. Once it analyses the intended action that user wants it to do it then performs the action that the user is expecting it to perform. Natural Language Toolkit (NLTK) is being used here to implement the natural language processing.

D. Text to speech conversion

Once the device interprets the intended meaning of the command given by the user it then performs the action and gives its output in form of text. For the user this text is then converted to the speech. For this conversion of the text to speech python text to speech package (pyttsx) is used.

E. Output through speakers

Once all the above steps are performed then it's time to give the output to the user in the form of speech. This action is achieved by using the speaker. Speaker can be connected to the device by using an AUX cable.

IV. QUERY PROCESSOR

To extract the intended meaning of the command given by the user the device uses the query processor to process the text performing the following steps:

1) Convert the raw text (sentence) into a list of words, known as tokens, by splitting the sentence by white spaces and storing them in a list "words".

2) Remove the punctuations from each item in "words" by replacing them with nothing.

3) Normalize the case by converting all the words in "words" to lower case.

4) Collect all the stop words from the English dictionary and storing them in a list "stopWords".

5) Remove the tokens from "words" that are present in the list "stopWords".

6) Remove the tokens from "words" that have a length less than 4.

7) Remove the inflectional endings from each token to return the base or dictionary form of a word, known as lemma.

8) Find the feature that is nearest to all the tokens by calculating the similarity of each feature, provided by a previously defined list "features", with each token.

9) Return the feature with maximum similarity threshold.

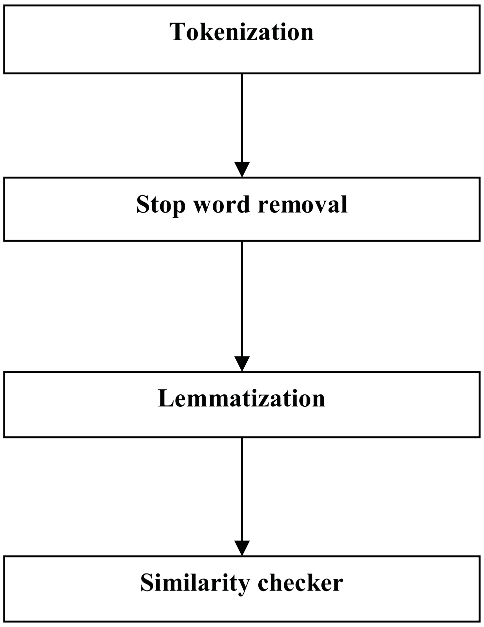

Fig. 4. Query Processor

ALGORITHM

```
1) Set words to the list of words that are split by white spaces.
2) Repeat the following steps for each token in words:
   a. For each character in token repeat :
         i. If character is a punctuation :
               Replace punctuation with nothing, i.e. ‘’.
        ii. Else :
               Continue
3) Change case of token to lower case for each token in words.
4) Set stopWords to the set of words that are considered to be stop words in English dictionary.
5) Set filteredSentence to an empty list.
6) Repeat the following steps for each token in words:
   a. If token is not in stopWords :
         Append token to filteredSentence.
   b. Else :
         Continue
7) Set words to filteredSentence.
8) Repeat the following steps for each token in words:
   a. If length(token) is greater than or equal to 4:
         Append token to filteredSentence.
        b. Else :
               Continue
    9)  Set words to filteredSentence.
    10) Create an instance of WordNetLemmatizer known as lemmatizer.
    11) Set lemmas to an empty list.
    12) Repeat the following steps for each token in words:
        a. Lemmatize the token (as a noun) and store it in temp.
        b. Append temp to lemmas.
    13) Set similarity to an empty list.
    14) Repeat the following steps for each feature in features :
        a. Set w1 to a set of all the synonyms of the feature.
        b. Set sim to an empty list.
        c. Repeat the following steps for each lemma in lemmas :
            i.  Set w2 to a set of all synonyms of the lemma.
            ii. Compute the similarity of w1 with w2 and append it to sim.
        d. Find the maximum from the list sim and append it to similarity.
    15) Return the feature having the maximum threshold of similarity.
```

V. IMPLEMENTATION AND TESTED MODULES

A. Setting timer

This module is used to set the timer for a particular number of minutes. After setting the timer the device will make a beeping sound when the number of minutes for which the timer is set passes by.

ALGORITHM:
Timer (minutes)

```
1) Initialize the mixer module of PyGame.
2) Load the alarm alert tone that is to be played when timer
   counts to zero.
3) Play the alarm tone that was loaded in the previous step.
4) Return
```

A. Current time

This module helps the user to check the current time by asking the device.

ALGORITHM:
Time ()

```
1) Create an instance of datetime.
2) The now() method of instance object is invoked to get the
   current date and time and the result is stored in time.
3) Current time is extracted from time and the result is stored
   back to time.
4) Set time_list to a list of all the items that are split by colon
   (:)
5) Set hour to 1st item in time_list.
6) Set minute to 2nd item in time_list.
7) Set x to “AM”.
8) If hout is greater than 12
   a. Set hour to hour mod 12
   b. Set x = “PM”
9) Set current_time to string: hour + “ ” + minute + “ ” + x.
10) Return current_time
```

B. Weather

When the user wants to know the weather of his location then it can ask the device to update him with the weather conditions. User can also ask about the weather conditions of some other place.

ALGORITHM:
Weather (location)

```
1) Provide a valid API key to allow responses from Open
   Weather Map (OWM).
2) Set weather to the current weather conditions at location.
3) Extract the current temperature in degree Celsius from the
   set of different conditions, i.e. from the set weather.
4) Return temperature.
```

C. Sending emails

User can send emails using this device. At the user end we need to provide the device with the receiver’s email id, subject of the message and the message that needs to be sent.

ALGORITHM:
Mail (sender, receiver, password, subject, body)

```
1) Create a SMTP object which is going to be used for
   connection with server by setting the server to
   “smtp.gmail.com” and port number to “587”.
2) Use the starttls() methodfrom the object which is required
   by gmail.
3) Login to the server using the sender and password
   variables.
4) Set message to a formatted string: ‘Subject:
   {}\n\n{}’.format(subject, body).
5) Send the email.
6) Return.
```

D. Calculator

This device can also be used to perform simple arithmetic calculations.

ALGORITHM:
Calculator (expression)

```
1) Set words to list of all the words that are split by white
   spaces.
2) Set i to 0.
3) Type cast the first item in words to float and store it in
   solution.
4) Repeat the following steps for each word in words :
    a.  If word is equal to “multiplied”
        Type cast the (i + 2)th item to float and multiply it with solution storing the result back in solution.
    b.  If word is equal to “*” or “x”
        Type cast the (i + 1)th item to float and multiply it with solution storing the result back in solution.
    c.  If word is equal to “divided”
        Type cast the (i + 2)th item to float and divide it from solution storing the result back in solution.
    d.  If word is equal to “upon” or “/”
        Type cast the (i + 1)th item to float and divide it from solution storing the result back in solution.
    e.  If word is equal to “raise” or “raised” or “power” or “^”
        Type cast the (i + 1)th item to float and compute the intended power of solution storing the result back in solution.
    f.  If word is equal to “by”
     i. If words[i-1] is not equal to “divided” and “multiplied”
        Type cast the (i + 1)th item to float and divide it from solution storing the result back in solution.
    g.  If word is equal to “plus” or “+”
        Type cast the (i + 1)th item to float and add it to solution storing the result back in solution.
    h.  If word is equal to “minus” or “-”
        Type cast the (i + 1)th item to float and subtract it from solution storing the result back in solution.
    i.  Increment i.
5)  Return solution.
```

E. Playing song

This gadget can be used to play music from an external device. The external device is searched for the queried song and if the song exists in the device, it is played using the omxplayer.

ALGORITHM:
PlayMusic (songs)

```
1)  List all the files and folder in the storage device and save them in songs.
2)  Traverse the list songs and extract all the files having .mp3 suffix.
3)  If the list songs have a folder having query in it.
     a. Set query = song
4)  Set url as the path to the song.
Run command line omxplayer.bin to play the song.
```

VI. RESULTS AND DISCUSSION

In this section of the paper we discuss the different test cases that are performed on implementation of natural language processing. The results show the success rates of the modules on which we work. To determine the success rate we test each module for certain number of times. Each time a command is given to use the same module the structure of the sentence changes. For example, if the user is asking the device, “Solve this Mathematical expression for me” or “Calculate this problem” then the device should give the same and expected results, that is, open calculator. This success rate is presented in the paper with the help of a graph where the X-axis depicts the modules and the Y-axis depicts the percentage of success received after testing for certain number of times.

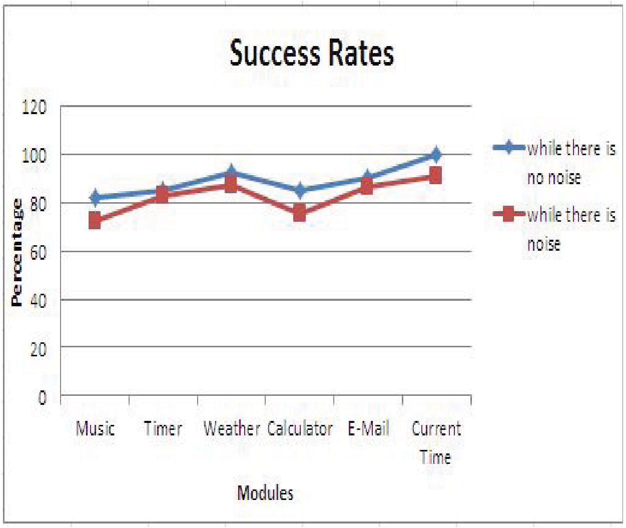

Fig. 5. Graph depicting success rate

VII. APPLICATION

A. In homes and daily use

This device can be used in homes for doing some simple and useful tasks. User can use it send an email to someone. This device can be used to listen to music. Also the user can set timer on the device.

B. For visually impaired people

Visually impaired people have difficulty in accessing basic services. For them this device changes the scenario as they can just access some basic services just by giving a vocal command to the device. They can easily sit anywhere and ask the device to tell the current time or current weather of their location or listen to their favourite music.

VIII. FUTURE SCOPE

In the future we can develop the device and take it a step further by attaching a camera to it and train it to detect different actions and perform according to those actions. This feature maybe used for security reasons as for instance if the device detects any person other than the user then it will send a notification on user’s mail with an image of the room at that point of time. If the device detects that no one is in the room then it may switch off all the electrical appliances like tube lights, fans, television and air conditioner. Also the device can be trained to react on several hand gestures like if the user waves his hand then the device may tell the current to the user.

IX. CONCLUSION

Development of Virtual Personal Assistant (VPA) is discussed in this paper. The vocal commands that the user gives are first converted into the text using speech to text synthesis. Next task of the device is to analyze the text converted from the speech. The device performs query processing using natural language processing (NLP). Query processing helps the device to identify the tokens from the text formed. Using these tokens the device can interpret what the user wants it to do. The device then performs the required action.

The tests for success rates are performed for each module. The measure is taken for each and every module whether it is performing in accordance to our command or not. The result is depicted through a graph. The distance between the microphone connected to the device and mouth of the user may sometimes give the results we don't expect as the device may not consider command of the user if he is very far from the microphone. The microphone may also not take the command from the user if the environment is very noisy. On the other hand device may give the results that user didn't expect, this may happen due to the unusual that the user has.

It can be used both in the official places and domestic places. People operating this device in both these places can use it for various purposes like setting a timer or for calculating an arithmetic expression. This device can really be very useful for visually impaired people as they have to just give the commands through their voice. They can use this device to find out the current time for any location, or the current weather of any location or just to listen to the music.

This device holds the potential to implement home automation. A camera can be included in the device. Many objectives can be achieved through this like for example security surveillance, or when no one is in the room then automatically turn off electrical appliances like tube lights, fans, air conditioner and television, or the device can be further trained to perform certain functions by observing hand gestures.

REFERENCES

[1] M. R. and D. Subramaniyan, "Personal Assistant and Intelligent Home Assistant via Artificial Intelligence Algorithms-(Raspberry PI/Pineapple)", *Impact: International Journal of Research in Engineering & Technology (IMPACT: IJRET)*, vol. 4, no. 6, pp. 9-13, 2016.

[2] Ass. Prof. Emad S. Othman, "Voice Controlled Personal Assistant using Raspberry PI," *International Journal of Scientific & Engineering Research* , vol. 8, no. 11, pp. 1611–1615, Nov. 2017.

[3] 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.

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

[5] 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.

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