Conversational System, Intelligent Virtual Assistant (IVA) named- DIVA using Raspberry Pi

1st Mr. Divij Bajaj
MBA-Data Sciences and Data Analytics
Symbiosis International University (Deemed)
Symbiosis Center for Information
Technology, Pune, Maharashtra

Dr. Dhanya Pramod
MBA-Data Sciences and Data Analytics
Symbiosis International University (Deemed)
Symbiosis Center for Information
Technology Pune, Maharashtra

***Abstract*— We are living in an era where we are interacting with machines day in day out. In this new era of the 21st century, a virtual assistant (IVA) is boon for everyone. It has opened the way for a new world where devices can interact their own. The human voice is integrated with every device making it intelligent. These IVAs can also be used to integrate it with Business Intelligence Software's such as Tableau and PowerBI etc. to give dashboards the power of voice and text insights using NLG (Natural Language generation). This new technology attracted almost the entire world like smart phones, laptops, computers, smart meeting rooms, car InfoTech system, TV etc. in many ways. Some of the popular voice assistants are like Mibot, Siri, Google Assistant, Cortana, Bixby and Amazon Alexa. Voice recognition, contextual understanding and human interaction are some of the issues, which are continuously improving in these IVAs and shifting this paradigm, towards AI research. This research aims at processing Human Natural Voice and give a meaningful response to the user. The questions, which it is not able to answer, are stored in a database for further investigation.**

***Keywords*— Voice Assistant, Google Home, Alexa, Conversational System, Text-to-Speech, Speech-to-Text IOT, Sensor Data, RaspberryPi3.**

I. INTRODUCTION

People have become busy as they have ever been in this changing global sphere, and virtual assistants can be used for anything for trivial things like answering the phone or replying to mails or extensive research, businesses are rapidly adopting conversational AI mainly for reasons such as improving their productivity, efficiency, and accuracy. Platforms as a case in point— like WhatsApp or Telegram — will have chatbots sending customized notifications in the future.

According to market intelligence firm Gartner, voice-based digital assistants will grow to 1.9 billion by the end of 2022, from more than 380 million worldwide users in 2017. The Adobe Analytics survey shows that 71 percent of owners of smart devices like Amazon Echo and Google Home use at least daily voice assistants, and 44 percent say they use them multiple times a day. If you're a business and you're looking for better performance and help your customers much faster when it comes to customer queries and quick conflict resolution; you're likely starting to think about AI-powered Intelligent Virtual Assistants.

Besides the desire to converse more naturally with technology, people can also see the benefits voice conversation offers when interacting with more complex interfaces such as home automation, BI tools, dashboards etc. which often takes several mapping steps to accomplish even simple tasks. A full spoken sentence, by person, can often convey all the information needed to achieve the desired result. The same applies to in-car applications where more than 64 per cent of people will welcome the opportunity to use their voice to perform tasks while on the run. In addition, to simplify workflow processes, utility services, travel and banking were all seen as key areas that would benefit from voice interaction Since the Turing Test, developed by Alan Turing in 1950, was first conceived, chatbots have tried to fool humans into thinking that they too are one with same capabilities. As it turns out, 58 percent of people don't mind talking to a bot than dealing with a human customer service agent or would prefer it. Thus, there is negligible scope of getting hurt emotionally due to non-human interaction. As human / machine conversations become more sophisticated with increased capabilities, it is

expected that the number of people deliberately seeking a digital assistant to cope with their request and set to increase significantly for reasons of speed, convenience and ease, currently 25 per cent. Intriguingly, 87 percent of people would prefer to know that they talk to a digital assistant. This may be because it enables users to repeat questions any number of time without seeming imprudent or that they know they can dispense with niceties and get straight down to the question they need answering Either way, it's important that businesses believe that people like to know who they interact with and don't purport inadvertently as if they're trying to mislead their clients. By providing a corporate person, often an avatar, sometimes just a name, that customers recognize as an automated support contact, many companies solve this problem. They are typically available through the basic web, mobile and email channels, but increasingly digital assistants are being used with good effect in call centers, and through social networks and messenger services, providing a consistent, 24/7, multilingual service.

II. APPLICATIONS

- To integrate IVA with most of BA tools and give Dashboard Voice Insights.
- IOT Application- Controlling Electrical Appliances on voice Command.
- To understand Google Cloud Console Platform Analytics and monitor API's usages.
- Integrating health sensors with the virtual assistant to track various body parameters on the voice commands.
- Drafting and sending mail to the desired user using voice dictation.

III. LITERATURE REVIEW

The study of literature for Voice Controlled Assistant – DIVA revealed efforts made by the researchers/scholars in the different disciplines mentioned in reference section. Also, this project is first implemented and demonstrated with all the functionalities working to make this voice assistant smart and personalized by triggering it with custom hot word detection- "Hey Diva", followed by user query.

The continuous efforts are made in adding emotions and human like touch in these IVAs. There is a huge impression that voice assistants would be used more if the capacity for conversation were improved, with 68 percent of people saying that when it works as expected it is fantastic and extremely useful. Google Assistant, Siri, Cortana, Alexa etc. and building our own assistant called "DIVA" using Google Voice Assistant APIs and Cloud Console Platform on the top of RaspberryPi3 and integrating it with BI software's and electrical appliances to control everything on voice.

We aim to integrate the power of NLG (Natural Language Generation) to get voice as well as text insights inside the dashboard itself. This technology focus in worldwide like smart phone, laptop, computer, etc. The objective of this paper is to test voice recognition and contextual understanding between user and human interaction in order to process the voice recognition and human interaction analysis. It was necessary to know to recognize the voice that the Virtual Assistant regularly understood the words that the users referred to the idea of giving feedback or estimation. In this survey, users tried to identify voices in gadgets and various differentiations along with them Changing the volume of the sounds of the base. According to the reports [5], Google and Siri understood better.

Google Assistant is good at understanding natural language and is less prone in understanding human voice whereas Alexa music delivery is good but, Alexa isn't easy with basic questions. This was the first human interaction issue that the survey found that a hands-free connection to a neglected user was a real use case.

In these special circumstances, the disruption ability was a huge obstacle. When they were given requests that they were asked to select or select an option by dragging the touch screen using the sermon on the outside of the intelligent virtual assistive screen or by using the discourse. It essentially intervened without handling the hands of the IVA and was particularly thought to be revolutionary in the situation, for example, driving Maintaining the discourse as basic information and the need to link all activities through the link in order to ensure the future of the IPA, with a specific guarantee that the Hands Free Association is complete and that the activities do not interfere with the process of collaboration[1]. This paper includes voice-activated smart home design and implementation.

Voice is the only source of data generation in these IVAs. The Voice Command System is largely a system that intakes and processes voice as input, decodes or understands the meaning of the input, processes it with finding key-words using N-Gram technique, formerly unigram and subsequently generates an appropriate voice output. Every system of voice commands needs three basic components, namely a speech-to-text converter, a query processor and a text-to-speech converter. Voice was a very significant part of today's communication. Since sound and voice processing is faster than written text processing, voice command systems are omnipresent in computer devices. In terms of speech recognition there have been some very good innovations. Some of the new developments were attributed to the development and heavy utilization of big data and profound learning in this field [5]. The technology sector used deep-learning techniques for the development and use of some speech recognition systems, and Google was able to reduce the word-error rate from 19 percent up to 54 percent by 6 to 10 percent in comparison with the system [1]. Text conversion is a process by which the recognized text machine is converted to any language that the speaker may identify when the text is read loud. It's a two-stage process divided into the front and rear ends [2]. The first aspect is to translate numbers and abbreviations into written text. The second section concerns the message to be transformed into a comprehensible one. It is called standardization of the text. Speech recognition is the computer's ability to recognize words and phrases in any language [5]. Then these words or phrases are converted into a language that the consumer may comprehend. In general, vocabulary systems [6] are used to recognize speech. The voice recognition system can be a small vocabulary system for many users or a large vocabulary system for small users.

IV. LITRATURE RESEARCH ON EXISTING VOICE ASSISTANT:

Amazon only created Alexa [4] to act as a personal assistant for Amazon Echo products. Alexa can have a casual conversation, play music, read emails, build to - do-lists and more. Alexa skills are not only developed by Amazon Developers but also, the community which uses amazon development account and add customized learnable skills to the package which anyone can download. Alexa also has an advantage because it can be used in home automation. Installing skills or installing apps can merge Alexa to increase its potential. The development of innovative applications can always be based on different tools available in Amazon. As it has not incorporated the batteries, Alexa always requires an apoer source for its operation. IBM has extended its limits with the DeepQA super-computer Watson [5], which can answer natural language queries. It provided a platform for all developers and start-ups to develop apps on the IBM Watson Development Cloud, not launching its own AI interface that interacts with consumers. Similarly, Google Cloud also facilitate these services with some free credits up to $300. It has a dashboard that can be customized to the user's liking. Bixby [7] the smartest IVA by Samsung has gone one step ahead of everyone. It is only for Samsung's flagship handsets a rebooted version of "S Voice" and is customizable. It includes Bixby Voice and Bixby Vision. Devices can be activated if the keyword "Bixby" is called out, which is part of Bixby Voice. Bixby Vision enables user to experience augmented reality. Samsung also has its own AR glasses which together makes the best user experience. The unique feature, which only Bixby VA provide, is it can be trained on original human voice with same annotations and syllables. They have started this to help MDP cancer patients who lost their loved ones at very early stage with the mission of restoring their voice forever.

Mobvoi [8] is a leading Chinese personal assistant in the field of AI, founded in 2012, the Chumenwenwen (mobile voice search application). Google is a small corporate shareholder. It provides an exact answer by vertical search. Your watch could be activated by "Hey Watch." The app is for wearable products like Google Glass and Android on iOS and android. The company has also developed a "Tic watch" smartwatch to work with iOS and Android devices. A new AI- powered assistant developed by Gatebox in Japan, and she is ‘Kawaii’. She is not just an assistant but a virtual companion for single depressed people. In Japan people date, have romantic chats and go out with her to make her feel special. Their machine learning program is set based on user daily schedule pattern and its master preferences and tastes. Hound [15] by Sound hound is notable for its grasp of the natural language and its ability to handle complex orders. Houndify offers additional functionality for the identification of feelings, the analysis of iris and fingerprints.

Mycroft [11] is the first independent virtual assistant on the worldwide platform. It is basically free (open source) software that can be modified on the basis of users ' needs, combined with other projects or incorporated with Raspberry pi desktops. Its main aim is to help physically challenged and handicapped people to make their life easier. An intelligent virtual voice-based helper for visually disabled users was presented to Aditya Sinha et al. [14]. This process is the following: voice input is first recognized in speech after speech synthesis. Afterwards the information is removed, after which the consumer is returned the message. This project uses Java Sphinx's speech analysis library, and MaryTTS is used to perform text-to-speech parts. It also uses neural networks to improve task performance through its ability to learn. A chatbot based on NLP has been suggested by Rishab Shah [13]. This paper shows educational systems that require natural learning. This system has solved the problem because of inaccessible education. The system includes tokenizing sentences and extracting queries based on the algorithm of the N-gram division. This metadata is searched into your database and information is retrieved and sent to the user if a match is found [14]. The application proposed by Sirius, which includes speeches and photos. Sirius is an application associated with the database. It emphasizes server architecture design space and also stresses the use of FPGAs, CPLDs and GPUs. ASR, IMM and QA are some of the new systems. Speech is translated into text using the statistical models. Sirius is particularly interested in his computer vision methods, which try to fit the image input in his image database and the relevant image information into his image database and return relevant image information. The supported questions fall under the arithmetic, logic and general categories

V. CONCEPTUAL MODEL/FRAMEWORK

A. System Architecture

The figure represents system architecture required to build this intelligent virtual assistant. It mainly consists of following requirements such as Raspberry pi 3 development board, USB mic sensor, audio speaker connected to 3.5mm audio jack, a google voice assistant installed on top of it and a relay module to control any electrical appliance.

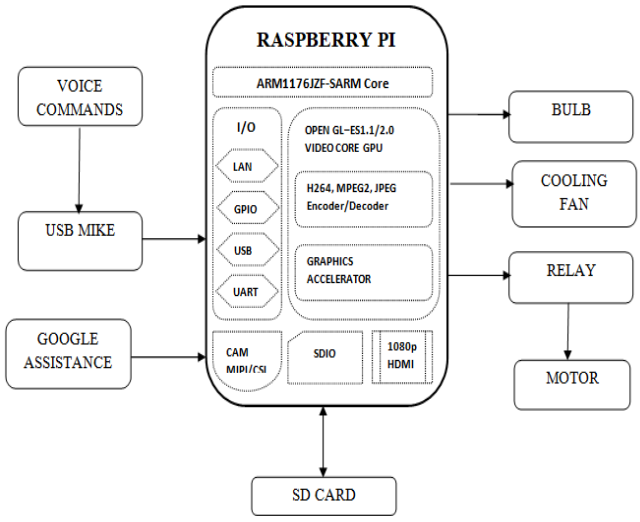


Fig1: System Architecture

B. Flow of Events in Voice Command System

First, the user uses a microphone to trigger it with our custom hot word written in python script- 'Hey Diva', followed by the query. It takes the user's sound input and is fed to the computer to continue processing it. The sound input is then transmitted to the speech to the text converter, which converts the audio input to that text output which the computer can recognize and manage. This all happens in the background as we have enabled Google APIs from Google Cloud Console platform. The text will then be checked and keywords are parsed using N-Gram technique. N-gram is the fields of computational linguistics and probability, an n-gram is a contiguous sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The most common N-gram pattern we use in this process is unigram or bigram to have maximum of two keywords parsed together. Our voice command system consists of a system of keywords that searches the text to match keywords. And the output is specified when the keywords are matched. The result is in a text form which will be displayed on raspberry pi terminal screen. Then the screen output is converted to a voice text converter with a system of recognition of optical character. OCR classifies and recognizes the text and converts it to the audio output by the speech engine text. This output is transmitted through speakers connected to the raspberry pi 3.5.mm audio jack.

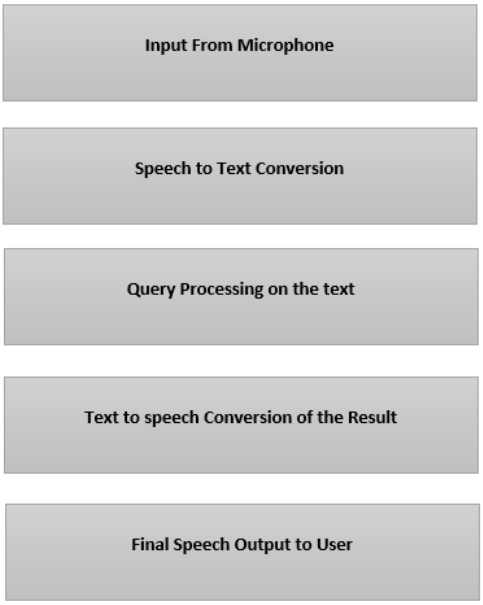


Fig2: Flow of events

C. *Speech-to-Text Engine*

As the title indicates, the module speech-to-text engine needs to convert the user's speech to a text sequence that can be processed by the logic engine. This involves recognizing the voice of the user, capturing the words specially keywords from the input for example-"Hey Diva play song name". So here keywords will be song name and the platform from where it has to be played. Other important function is to cancel any noise and fixing distortions in the process, and then using natural language processing (NLP) to convert the recording to a text string.

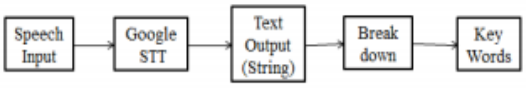


Fig3: STT Module

D. *Text-to-Speech Engine*

This module receives the output from the Query Processing System and converts the string to the voice for full user-to-user interaction. Text-to-Speech, in particular compared to text-based confirmation, is important to make the virtual assistant humane. The voice can be both male and female as natural as the human being, and is now mostly available in all languages.

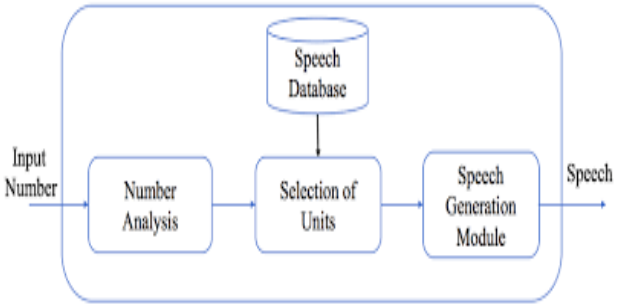


Fig4: TTS Module

E. *Query Processor*

A query-processing module in the Voice Command System operates usually as many database processors. This means that users ' input is taken, that the output is searched and that the user is given the correct result. In this procedure, we use the Alpha Wolfram Website as a source for the processing of queries. audio jack.

VI. RESEARCH DESIGN

In building virtual voice assistant, the primary source of data generation or collection is voice commands. There are instances when the query asked is not correctly interpreted by the Google Assistant or wrong interpretation of the words used. These speech-to-text data is collected and stored to refine the model to improve accuracy. This speech-to-text includes data from general queries, software integration queries, special fed data queries, social media queries etc.

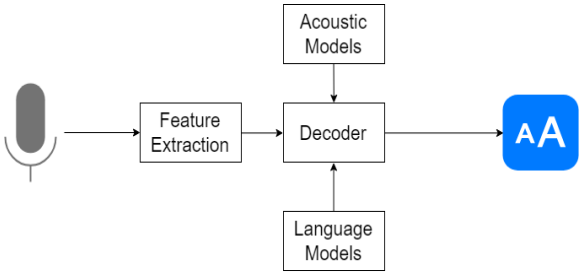


Fig5: System Process

The research also aims in collecting the data which is wrongly interpreted by the raspberry pi installed voice assistant. As the speech-to-text conversion takes place we use N-gram technique preferably unigram in python to fetch all the individual words from a sentence and see the

wrongly interpreted word. Then we can train the model with the correct word recognition with correct spelling. In addition, google assistant is trained on mostly all English words available in a dictionary.

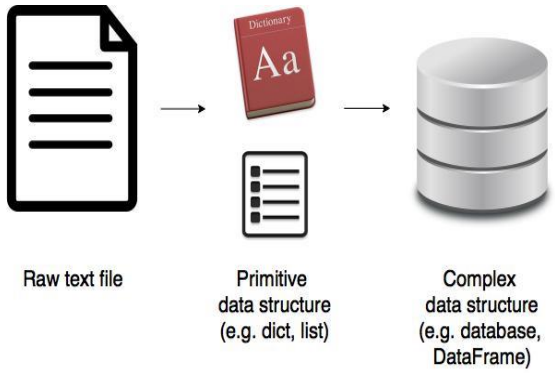


Fig6: Data Sourcing

VII. HARDWARE IMPLEMENTATION

- **USB Microphone** The microphone is used to take the sound input mainly voice. This audio input, when passed through the system, detects the words that would be searched for keywords.
- For example- 'Hey Diva, please switch on kitchen lights. So here keywords are 'Switch-on' and 'Kitchen-Lights'. These keywords are essential for the functioning of the voice command system as these modules are triggering words for our module and work on the essence of searching for keywords and output by matching keywords as shown in Figure- Hardware Setup (above) of the voice command system.

- **Keyboard** The keyboard acts as an input interface primarily for developers, providing access to the program code that makes edits. The mouse also acts as an interface between the system and the developer and does not interact directly with the end-user. If connected through VNC server to your laptop or desktop wirelessly you do not need additional keyboard and mouse.

- **Raspberry Pi** the Raspberry Pi is the foundation of the voice control system because it is used to connect the components together in each phase of the data processing. Raspbian OS is installed on the SD card and loaded to the operating system onto the card slot. We can also use NOOBS installer which is direct installation from raspberry pi foundation site. For this project, we have used Raspberry pi 3 model B+ with 32 GB memory card and Raspbian Buster, the latest OS installed in it. The Raspberry Pi needs a constant 5V, 2.2 mA power supply. This can either be provided through an AC supply using a micro USB charger or through a power bank.

- **Ethernet/WIFI** The Ethernet / Wi-Fi is used to provide internet connection to the voice-activated device. It cannot run without internet as we have interfaced so many APIs so they have to call it for proper functioning. Since the system relies on online text to speech translation, online query processing and online speech to text conversion, we therefore need a constant connection to accomplish this.

- **Monitor** is also optional in case user has connected raspberry pi through VNC server and VNC viewer is installed in system which enables it to imitate the screen in laptop screen only. If a user or developer is using HDMI port of raspberry pi then it is required to connect to a monitor screen to access raspberry pi desktop for running terminal commands and other python installation scripts.

- **Speakers,** After the user has asked query by triggering it with hot word detection 'Hey Diva' , the text output of the query is displayed on terminal and is converted to speech using the Google API text-to-speech. Now this converted audio speech is sent as output to the speaker connected to the 3.5mm audio jack of raspberry pi 3.

VIII. DATA ANALYSIS

Enabling all the APIs required for the project by log in to Google Cloud Console Platform and creating new project. Here my project name is- 'bivoiceai'.

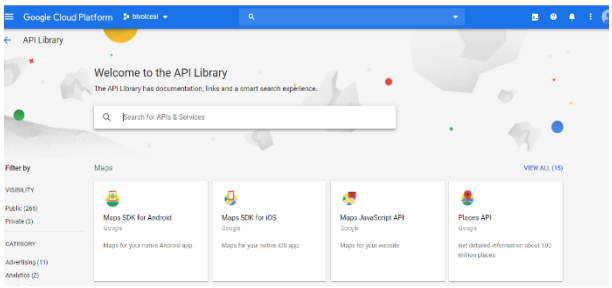


Fig7: Google Cloud Console

The list of enabled APIs which are under dashboard section of google cloud platform are mentioned below. These APIs helps to integrate google assistant with raspberry pi3. The user can subscribe for the usage for long run operations.

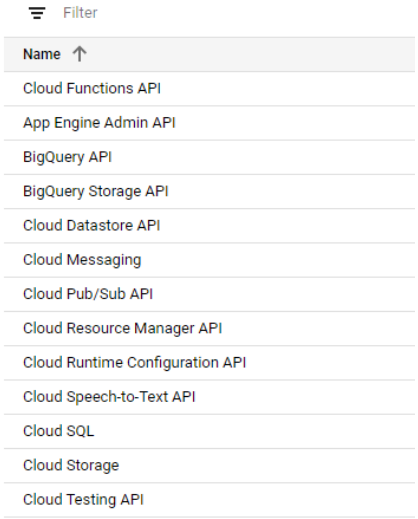
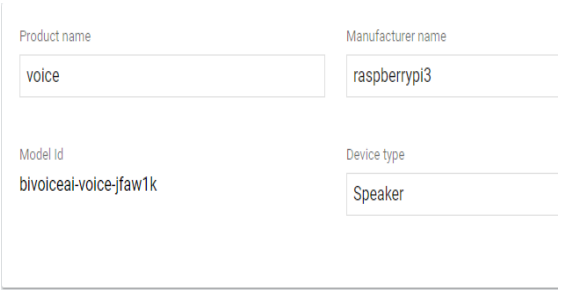


Fig8: List of APIs

Now login to actions.consolegoogle.com and click on device registration. Give name to your project and



generate model-id which will be used at the time of installation.

Fig9: Google Cloud Console

Open the google instructions for installing google voice assistant on raspberry pi and run the installation script in raspberry pi terminal and add all the details of your project such as project-id, OAuth Credentials and model-id when prompted.

Once the installation gets complete, we use snow boy hot word detection tool which train the voice assistant to

get triggered on custom wake words such as I used 'Hey Diva' in my use case. One can select any name

and can train on various voice sample and later execute the file to get name file installed inside assistant.

IX. RESULT

It is the idea and the logic it was developed by the Voice Command System. Our staff member uses the button to receive a order. The names of the modules in the program codes will match each of the commands given. When the command name coincides with any number of keywords, the Voice Command System performs that set of actions. Google Voice Assistant's main advantage in Raspberry Pi is that we can control GPIO and operate any electronics on voice controls If the system cannot match any of the command's keywords, the system apullizes it cannot accomplish the task and saves the wrongly interpreted question for N-grams and trains or automatic learning. Furthermore, it can be connected to many API connection software. Finally, all the features initially proposed are operated on the expected lines Moreover, the device is also a great success in the future, as new modules can be implemented at all times without interrupting the operation of existing modules. The working project video link is mentioned as a reference at last for deeper understanding.

CONCLUSION

We can conclude the following observations:

i) Voice search to get anything from google.

ii) Controlling home appliances using voice commands.

iii) Controlling Tableau dashboard using voice commands.

iv) Using IFTTT we can connect this IVA to mobile phone and other social media applications such as twitter and facebook.

REFRENCES

[1] Anup Kumar, Ranjeeta Chauhan, “Voice Controlled Robot”, 2014 IJIRT Volume 1 Issue 11 ISSN: 2349-6002.

[2] AkifNaeem, Abdul Qadar, WaqasSafdar, “Voice Controlled Intelligent Wheelchair using RaspberryPi”, International Journal of Technology and Research.

[3] Jonathan Gatti, Carlo Fonda, LivioTenze, Enrique Canessa, “Voice-Controlled Artificial Handspeak System”.

[4] Emad S. Othman, Senior Member IEEE – Region 8, High Institute for computers and information Systems “Voice controlled personal assistant using Raspberry pi”. International Journal of scientific and engineering Research volume (2017)

[5] G. Ashwini1, M. Nithish Reddy2, R. Paramesh3, P. Akhil4, an intelligent virtual assistant using raspberry pi.

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

[7] 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.

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

[9]. Chan Zhen Yue and Shum Pig, “Voice activated Smart home design and implementation”.2nd international 2017 conference on frontiers of sensor technology 978-1-5090-4860-1/17/$31.00 ©2017 IEEE

[10] Amrita s. Tulshan and Sudhir Namdeorao Dhage “Survey on Virtual Assistant: google Assistant, Siri, Cortana, Alexa”. © Springer 981-13-5758-9_17

[11] Achal S Kaundinva,Nikhil S P Atreyas,Smrithi Srinivas,Vidhya Kehav,Naveen Kumar “Voice enabled home automation using Amazon echo” “International Research Journal of Engineering and Technology (IRJET)Volume:04Issue:08August-2017”.eISSN:2395-0056 p-ISSN:2395-0072 [12] Ass. Prof. Emad S. Othman “Voice Controlled Personal Assistant Using Raspberry Pi” International Journal of Scientific & Engineering Research Volume 8, Issue 11, November-2017 ISSN 2229-5518

[13] Veton Kepuska and Gamel Bohouta “Next generation of virtual personal assistant (Microsoft cortana) Apple siri, Amazon alexa and Google home”. 978-1-5386-4649-6/18/$31.00 ©2018 IEEE.

[14] Amazon. Amazon Lex is a service for building conversational interfaces https://aws.amazon.com.

[15] Google: Google Assistant. https://assisatnt.google.com.

[16] Ali ZiyaAlkar, “An Internet Based Wireless Home Automation System for Multifunctional Devices”,

[17] G. Dizon, “Using intelligent personal assistants for second language learning: A case study of alexa,” TESOL Journal, vol. 8, no. 4, pp. 811–830, 2017.

[18] Dhiraj Kalyankar, Dr.P.L.Ramteke ," Review On IoT Based Automation By Using Personal Assistant For Visually Impaired Person"., vol. 3, no. 3, pp. 164–173, 2018.

[19] Madhusudhanan. R & Divya Subramaniyan In Paper “Personal Assistant and Intelligent Home Assistant Via Artificial Intelligence Algorithms-(Raspberry Pi/Pineapple)- Personal Assistant”, International Journal of Research in Engineering & Technology (IMPACT: IJRET) ISSN(P):2347-4599; ISSN(E):2321-8843, V ol. 4, Issue 6, Jun 2016, 9-14.

[20] Melissa Ram´ırez, Miguel Sotaquir´a, Alberto De La Cruz, Esther Mar´ıa, Gustavo Avellaneda, Ana Ochoa,“An Automatic Speech Recognition System for Helping Visually Impaired Children to Learn Braille”, 2016 IEEE.

[21] Project Video Link- https://bit.ly/2JYYJ9C