Artificial Intelligence Based Multilingual Voice Assistant

Kale Sushma[1], Dukare Prajakta[2], Meshram Hritik[3], Chinchole Rushikesh[4], Dr. N. P. Kulkarni[5]

Student, Department of Information Technology[1,2,3,4]

Guide, Department of Information Technology[5]

Smt Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *Artificial intelligence technologies are beginning to be actively used in human life, this is facilitated by the appearance and wide dissemination of the Internet of Things (IoT). Autonomous devices are becoming smarter in their way to interact with both a human and themselves. New capacities lead to creation of various systems for integration of smart things into Social Networks of the Internet of Things. One of the relevant trends in artificial intelligence is the technology of recognizing the natural language of a human. New insights in this topic can lead to new means of natural human-machine interaction, in which the machine would learn how to understand human's language, adjusting and interacting in it. One of such tools is voice assistant, which can be integrated into many other intelligent systems. In this paper, the principles of the functioning of voice assistants are described, its main shortcomings and limitations are given. The method of creating a local voice assistant without using cloud services is described, which allows to significantly expand the applicability of such devices in the future. Intelligent Voice Assistants (IVA), like Siri and Alexa, are created to assist their users with simple digital tasks. Here, we propose the steps we have used to develop a voice operated IVA which can process direct commands in two languages: English and Hindi, to perform menial tasks for the users. the language processing is performed by a modified finite state automaton. The IVA also takes advantage of the subject/action structure of commands to reduce the size of the word domain, and utilizes a generalization function to ensure that the language processor can understand multiple languages without undergoing major modification - making this approach suitable when training data is limited.*

Keywords: NLP, NLU, TTS, Voice Assistant, NER, Machine learning, ASR, Internet Of Things, Smart Things

I. INTRODUCTION

Artificial Intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the digital computer in the 1940s, , it has been demonstrated that computers can be programmed to carry out very complex tasks—as, for example, discovering proofs for mathematical theorems or playing chess—with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in performing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, and voice or handwriting recognition.

According to the overall description in the context, the purpose of the project is to develop an Android application that provides an intelligent voice assistant with the functionalities as calling services, message transformation, mail exchange, alarm, event handler, location services, music play service, checking weather, searching engine (Google, Wikipedia), camera, Bing translator, Bluetooth headset support, help menu and Windows azure cloud computing. Many years ago, software programs were developed and run on the computer. Nowadays, smart phones are widely used by all people. About 35 percent of the Americans have some sort of Smartphone. This shows that the market is increasing fast and there are also more capabilities for Smartphone because of this wide use. [2]

301

Currently, the project aims to provide the users a Virtual Assistant that would not only aid in their daily routine tasks like searching the web, extracting weather data, vocabulary help and many others but also help in automation of various activities. In the long run, we aim to develop a complete server assistant, by automating the entire server management process - deployment, backups, auto-scaling, logging, monitoring and make it smart enough to act as a replacement for a general server administrator.

II. BACKGROUND

A. History of Voice Assistants

If we come to the history of Voice Assistants,[2] the first voice activated toy to be released is 'Radio rex' in the year 1911. Foundation of smart virtual assistants was laid by IBM Many companies have used oral dialog system to design such system devices like Amazon Alexa, Microsoft Cortana, Apple's Siri, Google Assistant, etc., [3] General dialogue systems have six components which include Voice recognition, voice language apprehension, dialog manager, natural language generation, text to speech convertor, and knowledge base.

B. Future Applications

In future voice assistants can be used for two developments: First quality of dialogue recognition will increase because broadband allows more complex data processing in powerful data centres. Second, from the users perspective, VAs aid for interaction. In the companies, voice assistants can be used to automate repetitive tasks, for example Amazon's Alexa can open video conferencing and book meeting rooms etc.

C. Aim of this Study

The main aim of our project is that we have created a function, Intelligent Personal Assistant which can perform mental tasks like turning on/off desktop with the help of voice user interface(VUI) which is used to listen and process audio commands.

III. PROPOSED DESIGN

The project will give a fair knowledge about the intelligent assistant which is capable of understanding the commands given by the user. Our assistant can easily understand the commands given by the user through vocal media and responds as required. Our assistant performs the most frequently asked requests from the user and makes their task easier. Our voice assistant listens to the command given by the user through the microphone. After listening it will say "done listening" and displays what the user said and acts accordingly.

In our project we have installed gTTS engine package to make the voice assistant speak like a normal human being. We have defined a function called 'voice assistant speak,' as explained in (1) The gTTs will analyze the command given by the user through microphone and searches in the browser the required response and convert that response into text.

$$tts = gTTS(text=audio\_string, lang='en') \qquad (1)$$

gTTS is basically used to convert the audio string into text. gTTS is basically used to convert the audio string into text. This audio string is nothing but the response which the voice assistant is supposed to give the user. The language of the text is chosen to be English, the code for English is 'en'. We save this entire function into 'tts'. We are saving this text, that is the audio file with the '.mp3' extension. Each audio file is given a random number from 1 to 20000000. The random number can be generated using the command 'random.randint()'. This whole '.mp3' extension file is saved under the name 'audio file'. Finally to save this audio file we have used the command as mentioned in (2).

$$tts.save(audio\ file) \qquad (2)$$

This command (2) saves the audio file in the system.
(Ex-'audio24854.mp3').

IV. TASKS PERFORMED BY THE VOICE ASSISTANT

- These applications make small and smart hand-held devices to continue multiple features.
- They allow you to export and import data.
- Play/download a song or video from YouTube. When user asked 'can you play me a song', 'play movie' the

Impact Factor: 6.252

assistant open YouTube and plays the required content for the user the requested video/song.

- Searches anything from google and tells the required content. If asked 'google search' the assistant searches the content asked from the google and opens the required content in browser.
- Recognize voice commands.
- Control various applications of device.
- Gives the live news around the world. When asked 'news for today', 'tell me the news', 'what's the news', 'what is the news', 'news', the assistant reads the first 5 updated news headlines from the website.
- Reminds you important things on accurate situations or location.

These are some features we have added to our AI-based voice assistant as of now and we are working on many more features to embed into this assistant.

V. ARCHITECTURE

A. Existing Architecture

In this Diagram , the user is giving input / command to the system which can understand only English only and whatever the command or the input is converted into text format and with Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) , the task or the action is executed, the output will be in the form of audio format.
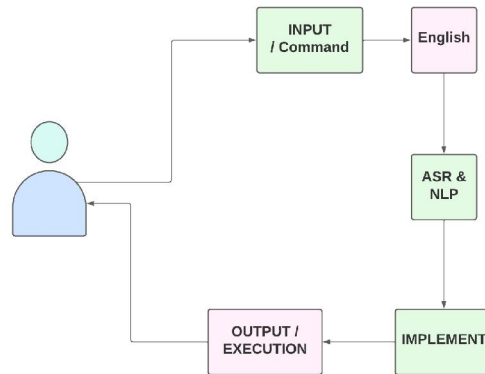

Fig1. Existing Architecture.

VI. METHODOLOGY

A. Proposed Architecture.

In this Proposed Architecture, the user will give the input / command, wheather it is in English or hindi it will convert that audio file or the input/command into the text format the with the Automated Speech recognition (ASR) technology and also with the help of Natural Language Processing (NLP) the command will implement and gives the desired output to the user in audio format with the help of Text-To-Speech(TTS) to the user.
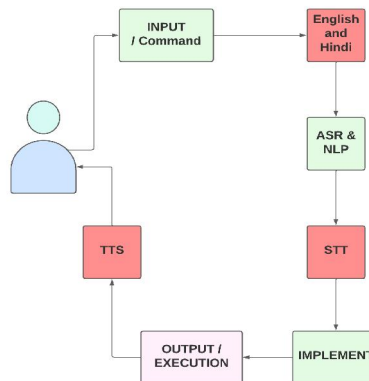

Fig2.Proposed Architecture

VII. METHODOLOGY

In our proposed voice assistant we are working to make it multilingual i.e. make it accessible by using not only one language. To overcome the problem of pronunciation and spellings we will be using greedy search algorithm which just like its name will search for the best word that fits in our given command.

Fig.1 Automatic Speech Recognition

A. Working of ASR

As shown in Fig.1 Automatic Speech Recognization which is termed as ASR is the main principle behind the working the AI based Voice Assistant.[4] ASR systems, at first it records the speech, then the wave file has been created by the device which consist of the words it hears, later the wave file will be cleaned so that the background would get deleted and the volume will be normalized, then it will break down into elements and it will be analysed in sequences, the the ASR software examines this sequences and it impliments statistical probabilities to find out the entire words and then it will get processed into text content.[5] The better method to recognize element is element recognition as it provide better result than method of word decoding.

Fig. 2 Process of ASR

It does not matter what kind of speech recognition software we may use, because all the work happens in its ASR. During a nutshell, at first the method starts with the device gathering audio with the source, where source is microphone, then the Recorded speech waveforms will be sent to acoustic analysis, which will be performed on three different levels, as shown in Fig 3.,

B. Acoustic Analysis

- **Acoustic Modelling:** In this process, it represents that the elements were pronounced or not and what are the words which can complete these elements.
- **Pronunciation Modelling:** That analyses the way, where how these elements are pronounced, it will check whether there is any accent or other peculiarities.
- **Language Modelling:** This is often aimed toward finding contextual probabilities counting on what elements were captured.

All the data which were recorded get processed by Artificial Intelligence without any human interaction, then the speech waveforms data is transmitted to the decoder, where it finally transforms into text for further use like command.
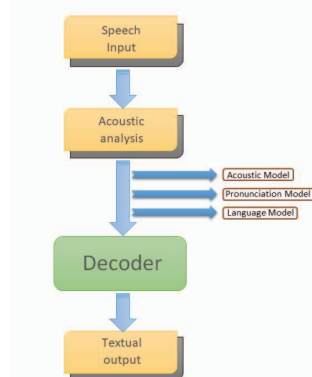

Fig. 3. Acoustic Analysis

VIII. HOW VOICE ASSISTANT IS USEFUL

In this section, explain a set of situational events and it can tell how our voice assistant is useful. For e.g., if want to go to a particular location instead of opening google maps and typing the destination takes a lot of time instead of that with our VA just by command to find the particular location it opens the map and highlight the particular location. In the modern VA's, playing a song on YouTube is just a feature, but they can't download it. For e.g., if you want to download the song "give me some sunshine", in other voice assistants we can't download it but in our VA, the song is searched in the YouTube database and the Video ID is noted, with this the video is downloaded.

If we want to get any information, we have to open google and search for it. But in VA by just giving it in the form of command for e.g., "what are super computers?" it collects the best information available on Wikipedia and gives us and if we want to know about what is happening in the world by giving the command "what are the top five news of the day?" it grabs the information and tells us. From this we can gain knowledge. Concerning the aspect of security, if a girl says "I'm in danger" the Voice Assistant extracts the current location of the girl and sends it to their parents/relatives and the nearby police station.

IX. RESULTS

 The required packages of Python programming language has been installed and the code was implemented using PyCharm Integrated development environment (IDE) and the python code we have developed runs in both Python 2.7 and Python 3.x, and below are the few outputs which we have received in our AI-based voice assistant.

A. Google Search Output

When we ask the voice assistant to search 'MS DHONI', it receives the request and performs the action by searching google.

B. Playing song/video on YouTube

As shown in below Fig 3. When we ask the VA 'Play me a song', it responds by saying 'which song you want me to play' and we ask the VA a particular song/video then the VA will perform the task by playing song/video on YouTube.
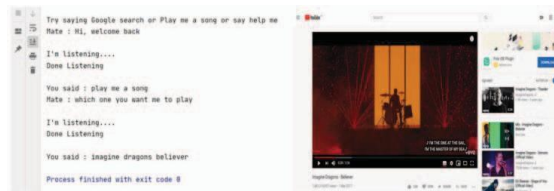

Fig. 4 Output screen of Playing song on YouTube

C. Current News

   When we ask current news to the voice assistant, it receives the request and responds back by giving top 5 news for that day along with website name.

X. CONCLUSION

In our project we have implemented many things compared to other assistants. Now a days it is very useful in human life because it is a hands-free application. It is a very simple application. As well as it is used in a business field also for example in laboratory, the person wears gloves and body suits for their safety purpose so it is difficult to type, through voice assistant they can get any information so that their work becomes easy.

 Voice assistants are useful in many fields such as education, daily life application, home appliances etc. and voice assistant is also useful for the illiterate people they can get any information just by saying to the assistant, luxury is available for people, thanks to AI based voice assistants.

 Voice assistant is developing more and more in daily life. Many companies of voice assistant trying to improve interaction and more features to the next level and many of the youth started using voice assistant in daily life and from many sources the result showing very good feedback. Compared to last 2 years voice assistants have been developed more and more.

REFERENCES

[1]. Yash Mittal, PradhiToshniwal "A voice-controlled multifunctional Smart Home Automation System", 2015 India Conference (INDICON), IEEE.

[2]. PrernaWadikar, Nidhi Sargar, RahoolRangnekar, Prof.PankajKunekar "Home Automation using Voice Commands in the Hindi Language", 2020, IRJET.

[3]. Nagesh Singh Chauhan,"Build Your First Voice Assistant",March,2019 https://towardsdatascience.com/build-your-first-voiceassistant.

[4]. Manohar Swamynathan, "Mastering Machine Learning with Python", Karnataka: Apress, 2017.

[5]. Yu Zhong, T. V. Raman, Casey Burkhardt, FadiBiadsy and Jerey P. Bigham, "JustSpeak: Enabling Universal Voice Control on Android", April 2014.

[6]. SukhadaChokkadi, Sannidhan MS, Sudeepa K B, AbhirBhandary "A Study on various state of the art of the Art Face Recognition System using Deep Learning Techniques", 2019, IJATCSE.

[7]. Steve Joseph, Chetan Jha, Dipesh Jain, Saurabh Gavali, Manish Salvi "Voice based E-Mail for the Blind", 2020, IRJET.

[8]. Dongmahn SEO, Suhyun KIM, Gyuwon SONG, Seung-gil, "Speech-to-Text-based Life Log System for Smartphones", IEEE International Conference on Consumer Electronics (ICCE), 2014.

[9]. Aditi Bhalerao, Samira Bhilare, AnaghaBondade, MonalShingade, AradhanaDeshmukh, "Smart Voice Assistant: a universal voice control solution for non-visual access to the Android operating system", Volume: 04, Issue: 01, Jan-2017.

[10]. Chen-Yen Peng and Rung-Chin Chen "Voice Recognition by Google Home and Raspberry Pi for smart Socket Control" 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI) March 29– 31, 2018, Xiamen, China.

[11]. G. KALYAN KUMAR, K. PAVAM KUMAR REDD "CORTANA (Intelligent Assistant)", Volume 6, Issue 4, April 2017, IJSETR

[12]. John Hearty, "Advance Machine Learning with Python", Birmingham: Packt Publishing, 2016.

[13]. Nikola Smiljković and TatjanaErić, Sandra Ivanović, SunčicaMilivojša, Milica Matić "Voice Control for Smart Home Automation: Evaluation of Approaches and Possible Architectures" 2017 IEEE 7th International Conference on Consumer Electronics - Berlin (ICCEBerlin) 978-1-5090-4014-8/17/$31.00 ©2017 IEEE.

[14]. Mark Summerfield, "Programming in Python 3", Boston: Pearson Education, 2010.

[15]. PiyushVashishtha, Juginder Pal Singh, Pranav Jain, Jitendra Kumar, "Raspberry PI based voice-operated personal assistant", International Conference on Electronic And Communication and Aerospace Technology, ICECA, 2019.