International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 6 ǁ 2021 ǁ PP. 20-23

Portable Voice Recognition with GUI Automation

P.Krishnaraj,F.Mohamed Faris,D.Rajesh
Department of Information Technology, A.V.C. College of Engineering, Mannampandal, Mayiladuthurai.

ABSTRACT:
Artificial intelligence technologies are beginning to be actively used in human life. Autonomous devices are becoming smarter in their way to interact with both a human and themselves. 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-computer 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.
KEY WORDS: *Voice assistant, Speech Recognition, Low cost, Internet, Speech Synthesis, Visually Challenged.*

Date of Submission: 25-05-2021 Date of acceptance: 07-06-2021

I. INTRODUCTION

Today the development of artificial intelligence (AI) systems that are able to organize a natural human-machine interaction(through voice, communication, gestures, etc.) are gaining in popularity. One of the most studied and popular was the direction of interaction, based on the understanding of the machine by the machine of thenatural human language. It is no longer a human learns to communicate with a machine, but a machine learns to communicate with a human, exploring his actions, habits, behavior and trying to become his personalized assistant. The work on creating and improving such personalized assistants has been going on for a long time. These systems are constantly improving and improving, go beyond personal computers and have already firmly established themselves in various mobile devices and gadgets. One of the most popular voice assistants are Siri, from Apple, Amazon Echo, which responds to the name of Alex from Amazon, Cortana from Microsoft, Google Assistant from Google, and the recently appeared intelligent assistant. Section I, II presents a brief introduction to the architecture and construction of voice assistants. Section III provides proposed plan of work. Section IV provides methodology of the work of a voice assistant . Section V describes the test results of the voice assistant. Section VI and VII describes the conclusion and future scope of an assistant using various artificial intelligent algorithms, and gives a comparative evaluation of the learning ability of algorithms. The main goal of this work is to build a local voice assistant that does the work of human and the daily task that a human needed to do in daily life. It has some new features like posting comments on the social media websites such as Facebook, Twitter, etc. By just few simple commands. You can also know the weather around you and can get the climate conditions in your region.

II. LITERATURE SURVEY

AbhayDekate [1] presented in the Modern Era of fast moving technology we can do things which we never thought we could do before but, to achieve and accomplish these thoughts there is a need for a platform which can automate all our tasks with ease and comfort. Thus we need to develop a Personal Assistant having brilliant powers of deduction and the ability to interact with the surroundings just by one of the materialistic form of human interaction i.e. Human Voice. The Hardware device captures the audio request through microphone and processes the request so that the device can respond to the individual using in-built speaker module. For Example, if you ask the device 'what's the weather?' or 'how's traffic?' using its built-in skills, it looks up the weather and traffic status respectively and then returns the response to the customer through connected speaker.

Dr.Kshama V. Kulhalli [2] proposed the Most famous application of iPhone is "SIRI" which helps the end user to communicate end user mobile with voice and it also responds to the voice commands of the user. Same kind of application is also developed by the Google that is "Google Voice Search" which is used for in Android Phones. But this Application mostly works with Internet Connections. But our Proposed System has capability to work with and without Internet Connectivity. It is named as Personal Assistant with Voice

Recognition Intelligence, which takes the user input in form of voice or text and process it and returns the output in various forms like action to be performed or the search result is

dictated to the end user. In addition, this proposed system can change the way of interactions between end user and the mobile devices. The system is being designed in such a way that all the services provided by the mobile devices are accessible by the end user on the user's voice commands.

Kishore Kumar R1 [3] presented to develop an economically effective and performance wise efficient virtual assistant using Raspberry Pi for home automation based on the concepts of Internet of Things, Speech Recognition, Natural Language Processing and Artificial Intelligence. People who are using it can give voice inputs and the device itself responds through voice commands by itself. It can fetch the date, time, weather, play your favourite music and fetch search results from the internet along with controlling the home appliances. NodeMCU chips are used to control the appliances which receives the command from the Raspberry Pi. The Raspberry Pi processes the speech inputs online given by the user through the mic and converts it into text and executes the command. The whole project is put in action through a python script which includes online Speech to Text conversion and Text to Speech conversion codes written. The NodeMCU is coded separately using the Arduino IDE to make it control the appliances and allow it to be accessed through its IP address. The device will respond to the user in a casual manner so that the user has a friendly experience with the device and feels it like his or her own assistant. This device makes the day by day

processes easier.

Rutuja V. Kukade [4] proposed there are various communication barriers for people who are blind , and they have to face various challenges. In this paper, we have discussed theimplementation of a personal virtual assistant which can take the human voice commands to perform tasks which otherwise would need the dependence on others. It enables user to receive and send emails, know the weather forecast report, maintain a personal diary/Online Blog, recognize image etc, using Speech to Text Engine, Text to speech Engine, OCR (Optical character recognition) using microphone for the input and speakers for the output.

VetonKëpuska [5] proposed one of the goals of Artificial intelligence (AI) is the realization of natural dialogue between humans and machines. in recent years, the dialogue systems, also known as interactive conversational systems are the fastest growing area in AI. Many companies have used the dialogue systems technology to establish various kinds of Virtual Personal Assistants(VPAs) based on their applications and areas, such as Microsoft's

Cortana, Apple's Siri, Amazon Alexa, Google Assistant, and Facebook's M. However, in this proposal, we have used the multi-modal dialogue systems which process two or more combined user input modes, such as speech, image, video, touch, manual gestures, gaze, and head and body movement in order to design the NextGeneration of VPAs model. The new model of VPAs will be used to increase the interaction between humans and the machines by using different technologies, such as gesture recognition, image/video recognition, speech recognition, the vast dialogue and conversational knowledge base, and the general knowledge base. Moreover, the new VPAs system can be used in other different areas of applications, including education assistance, medical assistance, robotics and vehicles, disabilities systems, home automation, and security access control. In this proposal, we have tested the new VPAs model by using IBM Watson cloud server with Python, Node Red.

Deny Nancy [6] presented in the Modern Era of fast moving technology we can do things which we never thought we could do before but, to achieve and accomplish these thought s there is a need for a plat form which can automate all our tasks with ease and comfort. Thus we humans developed applications like Personal Voice Assistant having the ability to inter act with the surroundings just by one of the materialistic form of human inter action i.e .Human Voice. The most famous application of android mobile phone is "Google Assistant", "Google Voice Search" which is developed by the Google .Various applications like Microsoft Cortana, Amazon Alexa is also used as an voice assistant .The voice application of iphone is"SIRI"which helpsthe end user to communicate end-user mobile with voice and it also responds to the voice commands of the user. We are going to develop a web application were the voice assistant would be available for a particular website. In this proposed system we have took a college website as an example. It can change the way of inter actions between end user and the website. The Application is being designed in such a way that all the services provided by the website are accessible by the end user on the user's voice commands.

Deepak Shende [7] presented 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 humanmachine 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.

Isha S. Dubey [8] proposed about a different combination of a reading machine (OCR), virtual assistant and Domotics system using Raspberry-Pi which will be a combination of a great system. This is a helpful aid for visually impaired people and people with disabilities. OCR stands for optical character recognition where it recognizes the present text and converts them into audio speech using pre and post processing with gTTS (Google Text to Speech). Google is used as its platform for virtual assistant which can be used in day to day life activities like checking mails, weather-forecast, news etc., further using Google Assistant, and python language we implement a voice based home automation. The major objective of this project is to help visually impaired by using various fields of technology. By just voice commands tasks such as reading of document, home automation and personal assistant can be achieved.

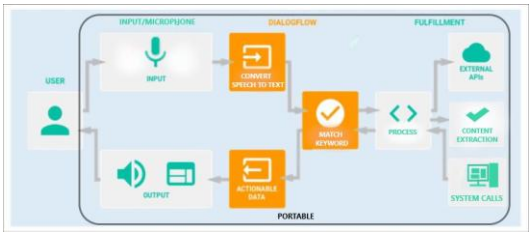
M. A. Jawale [9] proposed in today's world, many artificial intelligence applications developed using programming languages like Python, R and so on. Each language comes with its own programming structure and syntactical forms. Programmers are broadly classified into three categories namely, novice users, knowledge intermittent and expert one. For novice users, it is always a challenge to write a code without typographic errors though users know theoretical knowledge of Programming language, its structure and syntax as well as logic of program. Therefore, this paper explores use of voice recognition technique in the field of programming, specifically for writing program with Python programming language. In experimental analysis, it found helpful for new Python programmers and provide new learning curve for programmers wherein beginner can experience hassle free program writing. This paper adds new way of creating interest in beginners for judging their coding paradigm understanding and explore one of the area for user experience field for better programming Integrated Development Environment Development (IDE).

TusharGharge [10] presented the problem of user while developing a computer program. Developing a computer program is not an easy task it needs hardware resources which user have to handle. While continuous typing the code there may be possibility of injuries to the fingers of the user. To avoid the problems we are designing a system in which the computer program can developed through the voice. The voice will recognized by the system and that recognized words or word will be compared with the stored keywords in the database and if they are matched then that will be printed on editor and after this again by recognizing the specific keywords the program will be compiled and executed. This system will be easy to use, it reduce human efforts and the use of hardware resources. It would be surely useful for blind as well as novice plus knowledge intermittent users.

III. ARCHITECTURE

System architecture involves the high level structure of software system abstraction, by using decomposition and composition, with architectural style and quality attributes.

A software architecture design mustconform to the major functionality and performance requirements of the system



• Collection of data in speech format and convert it to text:

The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information center from the microphone is temporarily stored in the system which is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor

• storing the data and processing : The python backend get the output from the speech recognition module and then identifies whether the command or the speech output is an API Call, Context Extraction, and System Call. The output is then send back to the python backend to give the required output to the user.API stands for Application Programming Interface. An API is a software intermediary that allows two applications to talk to each other. In other words, an API is the messenger that delivers your request to the provider that you're requesting it from and then delivers the response back to you.

• Speech generation and processing output:
Text-to-Speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different languages, dialects and specialized vocabularies are available through thirdparty publishers.the output of each actions are displayed.

IV. RESULTS

This Project work of portable voice recognition started with a brief introduction of the technology and its applications in different sectors The project part of the Report was based on software development for portable voice recognition. After the development of the software finally it was tested and results were discussed, few deficiencies factors were brought in front.

V. CONCLUSION:

This Project work of speech recognition started with a brief introduction of the technology and its applications in different sectors. The project part of the Report was based on software development for speech recognition. At the later stage we discussed different tools for bringing that idea into practical work. After the development of the software finally it was tested and results were discussed, few deficiencies factors were brought in front. After the testing work, advantages of the software were described and suggestions for further enhancement and improvement were discussed.

REFERENCES

[1]. S. Abd El-Moneim, M. I. Dessouky, F. E. Abd El-Samie, M. A. Nassar, and M. Abd El-Naby, "Hybrid speech enhancement with empirical mode decomposition and spectral subtraction for efficient speaker identification," International Journal of Speech Technology, Vol. 18, no. 4, 2015, pp. 555–564.
[2]. O. Kalinli, M. L. Seltzer, J. Droppo, and A. Acero, "Noise adaptive training for robust automatic speech recognition," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, no. 8, Nov 2010, pp. 1889–1901.
[3]. H. F. Pardede, K. Iwano, and K. Shinoda, "Feature normalization based on nonextensive statistics for speech recognition," Speech Communication, Vol. 55, no. 5, 2013, pp. 587–599.
[4]. H. Hirsch and D. Pearce, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," Proc. ISCA ITRW ASR2000, 2000, pp. 181–188.
[5]. Tachibana, Hideyuki, Katsuya Uenoyama, and Shunsuke Aihara.(2018) "Efficiently trainable text-to-speech system based on deep convolutional networks with guided attention." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE.
[6]. Li, Naihan, et al (2019). "Neural speech synthesis with transformer network." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01.
[7]. Yang, Shan, et al. (2020) "On the localness modeling for the self-attention based end-to-end speech synthesis." Neural networks 125: 121-130.
[8]. Valentini-Botinhao, Cassia, et al. (2016) "Investigating RNN-based speech enhancement methods for noise-robust Text-to-Speech." SSW.
[9]. Gurunath, Nishant, Sai Krishna Rallabandi, and Alan Black.(2019) "Disentangling speech and non-speech components for building robust acoustic models from found data." arXiv preprint arXiv:1909.11727.
[10]. Hsu, Wei-Ning, et al.(2019) "Disentangling correlated speaker and noise for speech synthesis via data augmentation and adversarial factorization." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE.