



# Innovative Creation of Virtual Personal Assistant for New Era

Prof. Sumanth V<sup>1</sup> Bindushree P M<sup>2</sup>, Likitha Y<sup>3</sup>, Hemalatha R<sup>4</sup>

<sup>1</sup>Assistant Professor, Dept. of CSE, R.R Institute of Technology, Bengaluru, India

<sup>2,3,4</sup>UG Student, Dept. of CSE, R.R Institute of Technology, Bengaluru, India

**ABSTRACT:** 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. Example for dialogue system includes Apple's Siri, Microsoft Cortana etc. In this proposal, we have used the multi-modal dialogue systems which process two or more combined user input modes, such as speech, image and manual gestures in order to design the Next- Generation 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 recognition, speech recognition, the vast dialogue and conversational knowledge base, and the general knowledge base.

**KEYWORDS:** Virtual Personal Assistants; Multi-modal Dialogue System; Gesture Recognition; Speech Recognition; Image Recognition.

## I. INTRODUCTION

Spoken dialogue systems are intelligent agents that are able to help users finish tasks more efficiently via spoken interactions. Also, spoken dialogue systems are being incorporated into various devices such as smart-phones, smart TVs, in car navigating system. Also, Dialogue systems or conversational systems can support a wide range of applications in business enterprises, education, government, healthcare, and entertainment. Personal assistants, known by various names such as virtual personal assistants, intelligent personal assistants, digital personal assistants, mobile assistants, or voice assistants. Many companies have used the spoken dialogue systems to design their dialogue system device, such as Microsoft's Cortana, Apple's Siri, Amazon Alexa, Google Assistant. There are many techniques used to design the VPAs, based on the application and its complexity.

Moreover, there are some companies and researchers that have attempted to improve their applications by using the Multi-modal dialogue technique to design the Next-Generation of dialogue systems. The Multi-modal dialogue process two or more combined user input modes, such as speech, pen, touch, manual gestures, gaze, and head and body movement. For example, In the Ford Model U Concept Vehicle, this system, including a touch screen and a speech recognizer, is used for controlling several non-critical automobile operations, such as climate, entertainment, navigation, and telephone. The prototype implements a natural language spoken dialog interface integrated with an intuitive graphical user interface, as opposed to the traditional, speech only, command-and-control interfaces deployed in some of the vehicles currently on the market. Also, Semio is a part of research at the University of Southern California: "Semio is developing a cloud-based platform to allow humans to use robots through natural communication-speech and body language. The platform allows developers to create and deploy speech/gesture-based applications to be executed by robots and allows non-expert users to access and use those robot applications through natural communication".

In this proposal, we propose an approach that will be used to design the Next-Generation of Virtual Personal Assistants, increasing the interaction between users and the computers by using the Multi-modal dialogue system with techniques including the gesture recognition, image/video recognition, speech recognition, the vast dialogue and conversational knowledge base, and the general knowledge base. Moreover, our approach will be used in different tasks including education assistance, medical assistance, robotics and vehicles, disabilities systems, home automation, and security access control.

## II. RELATED WORK

In [1] Optical Character Recognition (OCR) is a piece of software that converts printed text and images into digitized form such that it can be manipulated by machine. The early OCR systems were criticized due to errors and slow recognition speed. In this paper, an overview of various techniques of OCR has been presented. An OCR is not an atomic process but comprises various phases such as acquisition, pre-processing, segmentation, feature extraction,



classification and post-processing. In [2] the authors developed a model called Lizzy, which is fully working Virtual Personal Assistance, which can perform task in offline condition. This model converted text to speech and speech to text. In online condition Lizzy gets more resources to work with. Also, any peripherals connected to Raspberry pi can be control with this Virtual Personal Assistance, just by giving command. In [3] This paper tells us about a handy keyboard that is used to type in character by recognizing hand gesture keyboard are currently the most universally accepted computer input device. They may be wired, wireless or virtual, but the chances are that you are within a few centimetre of a keyboard right now. In this project, we describe aptly named the gesture keyboard which as the name suggest converts gesture into keystrokes. Gesture keyboard sending signals to the computer, together with a python library using scikitlearn's SVM (support vector machine) algorithm the system captures motion and transform them into character that appear on screen. In [4] This paper presents a usability of four virtual personal assistant voice-based and contextual-text (Google Assistant, Cortana, Alexa, Siri). In this paper a survey was made on virtual personal assistants to validate the real potential of all the assistants. In comparison to all virtual assistant, google assistant was found to be more accurate with 59.08%. In [5] In this paper they have presented the design of the voice-based navigation system for blinds using voice recognition module ang GPS module implemented on Arduino board. The blind people give input destination as input to voice recognition module. GPS module continuously receives latitude and longitude of the current location. The blind person receives the pronounced directions which he needs to follow to reach the destination.

[6] In thia a Hand Gesture Recognition System uses a sign language in the form of hand gestures to type character on computer screen. Blinds face problem in reading and writing. They cannot access computer independently because of difference in communicating medium. By using this system user can write the letter on the screen in both English and braille font together efficiently using flex sensors attached to gloves. To recognize errors an error recognition module is added along with auto correct module.

### III. PROPOSED ALGORITHM

- We propose a system which is capable of controlling the device with not only speech but also image input and gesture input and as an add on demonstration we will be adding up home automation.
- Image input is done by OCR which converts the text on image to command.
- The gesture-based input is a wireless device which connects via Bluetooth and if we write on air it reads and understands the image.

#### A. Design Considerations:

We have two operating devices. One is the device controlling by raspberry pi and the other is the gesture remote controlled by Arduino. Here we have 3 modes of input given to the raspberry pi for personal assistant. Those three modes are Camera, Microphone and gesture. Mode switch is connected to the raspberry pi. Mode Switch is to say to raspberry pi, which mode of input we will give out of 3 inputs. After we select the input mode as camera mode then camera will activate and capture image convert image to text and analyse, then switch the appliances on/off. Similar way, if voice input mode is selected then voice is given via microphone and analyse the voice and convert it to text and, then switch the appliances on/off. If gesture input mode is selected, then based on gestures on Arduino with accelerometer transmits texts to raspberry pi via Bluetooth and the text is analysed and appliances is turned on/off.

### IV. PSEUDO CODE

We have two operating devices. One is the device controlling by raspberry pi and the other is the gesture remote controlled by Arduino. Here we have 3 modes of input given to the raspberry pi for personal assistant. Those three modes are Camera, Microphone and gesture. Mode switch is connected to the raspberry pi. Mode Switch is to say to raspberry pi, which mode of input we will give out of 3 inputs. After we select the input mode as camera mode then camera will activate and capture image convert image to text and analyze, then switch the appliances on/off. Similar way, if voice input mode is selected then voice is given via microphone and analyze the voice and convert it to text and, then switch the appliances on/off. If gesture input mode is selected, then based on gestures on Arduino with accelerometer transmits texts to raspberry pi via Bluetooth and the text is analysed and appliances is turned on/off.

### V. SIMULATION RESULTS

In this proposal, we have presented the approach and tested this approach by using many tools, technologies, and speech corpus. Also, we have worked on both hardware and software sides at the same time. On the hardware side, we started our work by collecting all tools and devises that would be used in the system, the mini board with CPU and LCD, the camera and microphone. The software side, we have tested each model with different cloud sever, such as



Google cloud, then we will connect all models together to test the system in the final stage. After testing the system and according to the results of the models (ASR Model, Gesture Model, Graph Model, Interaction Model), we found that the whole concept of this system is the best solution for Next-Generation of Virtual Personal Assistants by adding some improvements in the hardware and software in the final stage. To achieve the final stage and all these improvements to the new system with high accuracy, we need funding from an organization that will work with us to improve the system by funding the new hardware devices that have high accuracy, as well as the tools and cloud servers that we will need for testing the new system.

## VI. CONCLUSION AND FUTURE WORK

This proposal introduces the structure of Next-Generation of Virtual Personal Assistants that is a new VPAs system designed to converse with a human, with a coherent structure. This VPAs system has used speech, graphics, video, gestures and other modes for communication in both the input and output channel. Also, the VPAs system will be used to increase the interaction between users and the computers by using some technologies such as gesture recognition, image/video recognition, speech recognition, and the Knowledge Base. Moreover, this system can enable a lengthy conversation with users by using the vast dialogue knowledge base. Moreover, this system can be used in different tasks such as education assistance, medical assistance, robotics and vehicles, disabilities systems, home automation, and security access control. Also, it can be a satisfactory solution that can be used by applications, such as responding to customers, customer service agent, training or education, facilitating transactions, online shopping, travelling information, counselling, tutoring system, ticket booking, remote banking, travel reservation, Information enquiry, stock transactions, taxi bookings, and route planning etc. In the end, to achieve the final stage and all these improvements to the new system with high accuracy, we need funding from an organization that will work with us to improve the system by funding the new hardware devices that have high accuracy, as well as the tools and cloud servers that we will need for testing the new system.

## REFERENCES

- [1] Noman Islam, Zeeshan Islam, Nazia Noor "A Survey on Optical Character Recognition System" Journal of Information & Communication Technology-JICT Vol. 10 Issue. 2, December 2016
- [2] Aditya K, Biswadeep G, Kedar S and S Sundar "Virtual personal assistance" IOP Conf. Series: Materials Science and Engineering 263 (2017) 052022 doi:10.1088/1757-899X/263/5/052022
- [3] Bharathi M, Bhavani R, Keerthana Priya B, R.Paavaikarasi M.E., "Implementation of Gesture Keyboard Using Machine Learning" International Journal of Innovative Research in Science, Engineering and Technology Vol. 8, Special Issue 2, March 2019
- [4] Amrita S. Tulshan(&) and Sudhir Namdeorao Dhage "Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa" Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai 400058, India Springer Nature Singapore Pte Ltd. 2019 S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 190–201, 2019.
- [5] Manisha Bansode<sup>1</sup>, Shivani Jadhav<sup>2</sup>, Anjali Kashyap<sup>3</sup> Assistant Professor, Electronics, SPIT, Mumbai, India<sup>1</sup> Student, Electronics, SPIT, Mumbai, India<sup>2,3</sup> "Voice Recognition and Voice Navigation for Blind using GPS" INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING Vol. 3, Issue 4, April 2015
- [6] Prof. Devika Deshmukh, Ashlesha Bhandarkar, Shyam Chandak, Sarthak Patil, Shubhangi Gaikwad, Shivani Bhise "Hand Gesture Recognition System for Blinds" National Conference on Advances in Engineering and Applied Science (NCAEAS) 16th February 2017 In association with International Journal of Scientific Research in Science and Technology
- [7] Veton Këpuska & Gamal Bohouta "Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)" 978-1-5386-4649-6/18/\$31.00 ©2018 IEEE
- [8] S. Arora, K. Batra, and S. Singh. Dialogue System: A Brief Review. Punjab Technical University.
- [9] R. Mead. 2017. Semio: Developing a Cloud-based Platform for Multimodal Conversational AI in Social Robotics. 2017 IEEE International Conference on Consumer Electronics (ICCE).
- [10] Y. Nung, A. Celikyilmaz. Deep Learning for Dialogue Systems. Deep Dialogue.
- [11] Amazon. Amazon Lex is a service for building conversational interfaces. <https://aws.amazon.com>.
- [12] Microsoft. Cortana Intelligence. <https://azure.microsoft.com>.
- [13] B. Marr. The Amazing Ways Google Uses Deep Learning AI. <https://www.forbes.com>.
- [14] K. Wagner. Facebook's Virtual Assistant 'M' Is Super Smart. It's Also Probably a Human. <https://www.recode.com>.