ISSN: 0974-5823 Vol. 7 No. 5 May, 2022

International Journal of Mechanical Engineering

SPEECH-BASED VIRTUAL ASSISTANT SYSTEM FOR VISUALLY IMPAIRED PEOPLE

[1]Dr. B. HARICHANDANA, [2]Dr. C. KRISHNA PRIYA, [3]Dr. P. SUMALATHA

[1]Assistant Professor, Department of Computer Science and Engineering, Srinivasa Ramanujan Institute of Technology, Rotarypuram,B.K.Samudram, Anantapur, Andhra Pradesh,India

[2]Assistant Professor, Department of Computer Science and IT, Central University of Andhra Pradesh, Anantapur, Andhra Pradesh,India

[3]Assistant Professor, Vani Institute of Management and Sciences, Anantapur, Andhra Pradesh,India

ABSTRACT: In this world, certain people are visually impaired and they have been facing numerous complications in their daily life activities due to imperfect vision. However the technologies have grown leaps and bounds, the internet but still the websites are inaccessible by visually impaired persons. Hence this work is aiming for developing a device which will help them as their personal assistant. A speech-based virtual assistant system is presented in this paper for visually impaired people. Main aim of this system is providing voice-over assistant to blind-people for doing tasks such as looking for a thing or object, recognition of persons face with emotion, understanding the surroundings and reading, etc. This presented system uses Machine Learning (ML), Artificial Intelligence (AI), Image and Text Recognition for assisting persons who are visually impaired or blind. This concept will be realized by an Android mobile app which includes features like currency recognition, e-book, voice assistant, chat bot and image recognition. This is an effective model for blind individuals for engaging with the world and makes utilize the features of technology. The software reads the contents of websites and then utilizes text to speech and speech to text modules with selenium, it will automate any website. This entire project is dedicated to visually impaired people and is simple and in addition it can make the daily tasks easy for them.

KEYWORDS: virtual assistant, Machine Learning (ML), Artificial Intelligence (AI), Speech Processing, image recognition.

I. INTRODUCTION

In daily life, a person with no disabilities has no issues with his daily activities but a person who has partial blindness or imperfect vision is very hard to do his/her daily activities. They do not perform the actions like identification of an object, reading text, etc, because of their disability. Certain systems have been designed for helping the visually impaired person [1].The most complicated challenge for a blind person is especially, the person with complete blindness is navigating the around places. In this computer vision community, developing the visual aids to help handicapped persons is the most active researches [2]. The mobility aids are intended for explaining the environments on brink of person with surrounding objects appreciation. Such kind of aides is necessary for fine navigation in the environment. Various electronic applications have been developed to help the person who has blindness. As the mobile technologies give the aiding convenience to those persons, at present many of the organizations have been identified visually challenged consumers and developed certain applications for them while using the assistive technologies [3].

However many of these products have been ignored by visually challenged community because of their high cost, accessibility, usability and differences in user requirements. Various developing technologies have been planning to help the visually impaired persons in their daily activities as major interests. But the tasks

identification is one of the major difficulties for visually impaired persons. Several applications are utilized in this task, but certain limitations are required to be improved [4].With accessibility to visual data which has great significance for improving independence and safety of visually impaired and blind persons, since a smart automated system is required for assisting their surroundings navigation especially in unfamiliar places. Due to the availabilities of huge volumes of data and better algorithms, now it is easy for training the computers for multiple objects detection and classification in a image with higher accuracy that helps visually challenged persons [5]. Hence for overcoming these issues which are encountered by the visually impaired persons, this application is developed that will offer the assistance and convenience to visually impaired persons. This application provides object, text recognition and detection of face for the identification of humans, objects and text [6].

II. LITERATURE SURVEY

Dipankar Gupta et al. [7] developed an digital personal assistant with Bangla voice command and detection of face for handicapped persons. The cross-correlation method includes Bangla voice command and executed the tasks as per the defined pre-set command. The user is enabled by a mouse cursor controlling system based on facial movements for accessing the computer system more conveniently. The drawback of this system is that it only performs based on the pre-defined commands. Hence it can't be able for delivering the expected results relying on undefined voice commands.SyedMohidul Islam et. al. [8]discussed about a system which is close to presented system. This introduces a development strategy of Bangla virtual assistant 'Adheetee (Erudite)' for personal computers and smart phones that performs different types of functionalities of smart devices. Various algorithms are utilized for identification and responding as per the commands.

Developing the 'Olivia' is a virtual assistant for transforming a normal home as smart home is the main objective of this paper [9]. This system can be enabled with electronic home appliance integration capabilities in confined range. The limited functionalities such as playing jokes, telling jokes, weather update and performing different calculations based on voice command in English are implemented. In [10], V. Chayapathy presented a personal assistants development is presented that can assist the users in order to interact with home appliance by gesture and speech. This system works based on English speech recognition and enables the users for obtaining certain search results via web scrapping. Prince Bose et al. [11] demonstrated a voice-controlled system in English language for visually challenged persons. The functionalities like News portal reading, weather forecasting, email sending and receiving are implemented. The Google speech API (Application Programming Interface) is utilized to provide the capability of speech recognition.

Maan and Gunawardanaet. al., [12] have investigated the barriers in Ambient Assisted Living Technologies acceptance between the elder Australians. They utilized mixed techniques by the combinations of qualitative models and written questionnaire like focus groups (n=25) for this purpose. The results demonstrated that, various factors which can restrict such technologies usage and elder people may have some preferences while utilizing this technology. These barriers understanding can be achieved for providing the solutions as per the requirements of user. RobiahHamzah et al. [13] presented an application using text to speech technology. The persons who desired for assisting, the blind must type the message that is transformed to speech. The blind person wears an earphone for hearing voice message.

FarzanaJabeenet.al [14] presented technique that can help the blind persons in shopping environments. The web cam may take the snap and images are stored. If the user walks with barefooted and size of his foot is measured by Gait detector using the support of pressure sensor. Person height is measured by laser range finding device. User assistance is given via microphone. Dautenhahn et al. [15] explored the attitudes and perceptions of people interests towards the future robot companion to a home. Here human-centered technique is utilized with human-robot interactions and questionnaires (n = 28 adults). The questionnaires are gathered before and after the interaction sessions with PeopleBot robot. The results indicated that in this study 40% of

people are in the favor of idea of companion robot in home. Many of the subjects observed the major role of this robot in homes as a assistant, servant or machine.

III. SPEECH-BASED VIRTUAL ASSISTANT SYSTEM FOR VISUALLY IMPAIRED PEOPLE

Based on innovation and technology the Android application is promising to academically empower the blind people while freeing them from their dependence on visuals by giving the data via app. The main aim is to offer better functionalities which can make the partially blind person utilize it for recognition, identification, navigation and achieving the information about outer world. The app contains the chatbot and it will ask questions about weather, time or anything for obtaining the information or performing some actions as per user desires. It detects the objects in timely and gives the essential data to the users. In addition this app contains a barcode scanner that helps the user for obtaining information of particular products. Moreover this app helps the user for detecting the faces thus the user may know the presence of human in their surroundings and the number of people presence in that room. This application has a text reader which is utilized for reading the text as louder to a user. This system contains 5 modules of device that are accessed through voice commands and its architecture will be developed by a python interpreter. The arts Deep Learning (DL) states like capturing the image, detection of object and Optical Character Recognition (OCR) would be utilized for developing the functionalities of system. The assistive chatbot is developed using AIML (Artificial Intelligence Markup Language).

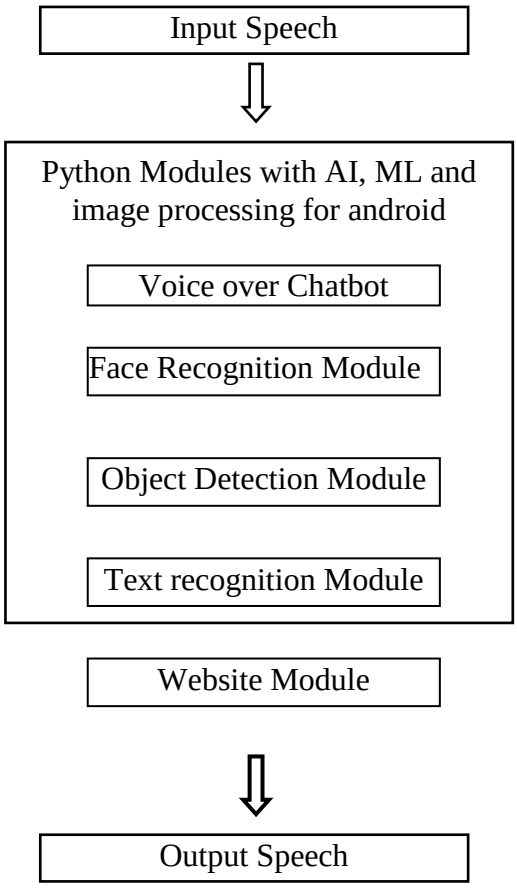

Fig. 1: System Architecture of Speech Based Virtual Assistant System

The Fig.1 illustrates the architecture of speech based virtual assistant system and it illustrates the interfacing of various system modules for voice over conversations. In this system, the modules are image captioning, voice-over chatbot, object detection, face recognition, recognition and reading of text. In this entire system, a user may communicate with the software through the speech-to-text interface module. The speech-to-text Google library (recognition of speech) for Python will be utilized for these purposes. For providing the output

of system to user and confirming the inputs of user, recognized input will be played back to user by the text-to-speech module of Python.

3.1 Voice-over ChatbotModule

The voice-over chatbot module will be utilized for communicating with the user through text-to-speech or text. The backbone of this system is chatbot module since it interacts with the user via voice commands. For making this chatbot as realistic and interactive, AIML (Artificial Intelligence machine learning) will be utilized as backend for writing the dialogues and replies. The gTTS (googleText To Speech) and pyTTSX (Pythontext-to-speech engines) would be utilized for generation of audio outputs and python is used as interpreting language. This module is responsible for voice-based conservations between system and user. The chatbot modules flow is done in 4 parts.

- **Taking input:** As an instance, when the dialogue is "read the text" then its reply will be "reading the text" loud to the user.
- **Categorization:** In this phase, the recognized text is processed for obtaining the wanted dialogue for performing the task. When the dialogue is "detect a object" by AIML pattern matching, the object detecting module will be obtained i.e. "object".
- **Collection:** The system performs a task as per the category after retrieving a text category. So a module performs the task and the data is send back to the chat-bot. Again chat-module processes the collected data for converting the data into meaningful sentences.
- **Reply to the user:** The gTTS module produces the output dialogue out as louder to user. It is the reply from application.

3.2 Face Recognition Module

This module is a computer technology that helps the visually challenged person for human face recognition from images, video or in live video. It is helpful to the person for the identification of people. The recognition of face is done in three different phases.

- **Detecting face in the image:** From a image a face is detected by the OpenCV (Computer Vision Library) and haar_cascade which is a binary file utilized for the detection of particular objects in a given image. The face location is mapped on the image.
- **Extracting the features:** The feature extraction is done after the retrieval of area of interest. The features such as height/weight of a face, nose, lips width, color, etc, will be extracted. Then the features are scaled thus face size in image can't affect the features ratio which is stored in the database. All the features would be saved in XML (Extensible Markup Language) data format.
- **Comparing the database:** The features extracted from phase 2 would be compared here with the presence of faces with their names. If the detected face is matched with the one of the faces in database then labeled name is given to emotion detection. If face is not matched then it asks the user to give aname to that person and it is stored in database.

3.3 Object Detection Module

The main purpose of this module is helping the user to identify the objects which are located in front of him/her. The object recognition is a method for recognition and labeling the detected object from the image, video or live video. The object will be recognized by ML and DL. The algorithm of object algorithms will take the frames as a input from the camera and a specific size bounding box is applied to that image and checks if any object is there in the image. When an object is found then the algorithm recognizes the object in the image. The object recognition is performed in two steps namely classification of image and localization of image. The image classification will predict the object class in an image, while the localization of object will identify that how many objects are there in an image and bounding boxes will be drawn. These tasks are combined by an object detection algorithm and it classifies the object in an image. These classifications may help the system for notifying the user regarding the objects which are located in his/her surroundings. The objects which have high prediction accuracy would be forwarded to chat-bot.

3.4 Text recognition Module

This test recognition module is used for detection and identification of text that is in the formats like oriented, digital or handwritten.This technology transforms the various forms of text into digital form. It can also be known as OCR. Various API will appear for different platforms which are utilized for OCR implementation. In order to recognize the printed or typed text on books or objects, the user needs to open the app on his/her Smartphone and to choose the desired option. Then the application identifies the text and converts it into digital format. Finally the text is read out to a user.

3.5 Website Module

This website module will be written in Python and it utilizes the Selenium for respective module automation and Beautiful Soup to scrap the web page contents. This module "Script" component contains customized code that will entail the website features contained in module. This web site module contains certain modules which are as follows:

- **Google Module:** The Google module contains python script that will automate the website by Beautiful Soup and Selenium. The user might search for any query via the interfaces of speech-to-text, text-to-speech and recognized queries will be searched with the support of Selenium. Resulted search results will be read out to user while scraping the contents of web page by python Beautiful Soup module. The results of search will be indexed that can enable the web page quick access as per the desire of user, hence time will be saved, either the entire search results are read out to user or he can choose the contents to read out.
- **Gmail module:** This Gmail module contains the python script which will starts the Gmail, login the users to their mailboxes and offers support to user for sending or reading the mails. To send a mail, the system is prompted the user for providing related details and after filtered out the noise, the input fields will be filled through selenium and then mail is send with the conformation of user. At every phase, the user can stop and edit any of his inputs. This system will repeat the inputs of recognized user and the input will be finalized only when the user confirms it.

IV. RESULTS

In python, the built-in modules of speech to text and text to speech (pyttsx3) provides better accuracy and offer simple and fast way for text conversion. The speech-to-test module recognized the words with an average of 90% accuracy with 4 different samples where each sample contains 20 different inputs in moderate to quite environments.

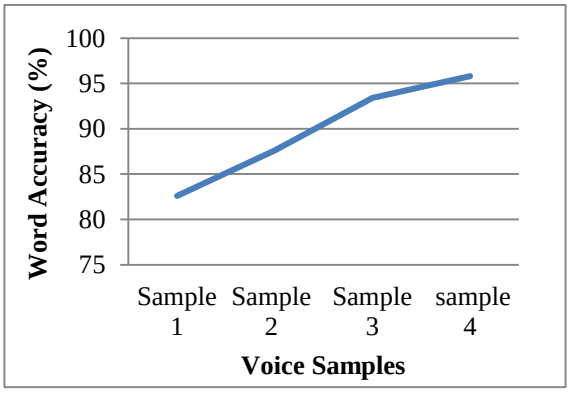

Fig. 2: Word Accuracy of speech-to-text module

The Face Recognition, object recognition and Text Recognition modules have been implemented successfully. Average run time and accuracy of each module function are shown in Fig. 3 and Fig. 4.

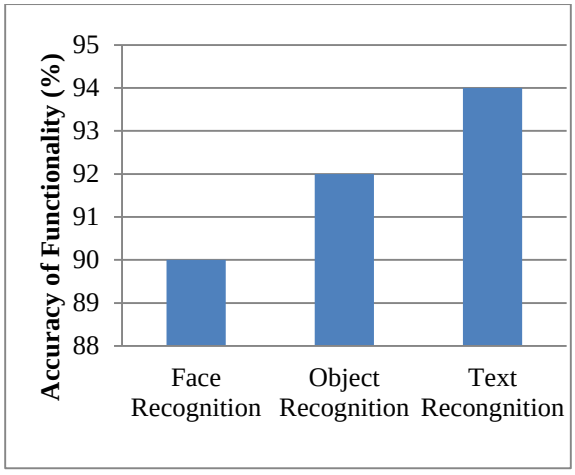

Fig. 3: Accuracy Analysis of Face, Object and Text Recognition Modules Functionality

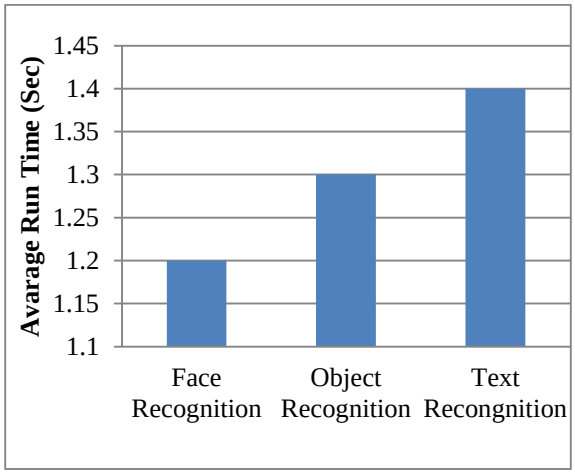

Fig. 4:Average Run time of Face, Object and Text Recognition Modules Functionality

The obtained results demonstrated that it is able for running the software on two most famous sites: Gmail and Google. The software is separately run over each of them. The software will send the email efficiently by the user commands. Google Search Engine is effectively performed search tasks depending on predefined commands or preference commands of user via speech. This operation set will take the minimum seconds compared to all groups. Latency in internet connection is the reason for highest consumption of average time for responding to the Google Search Engine.

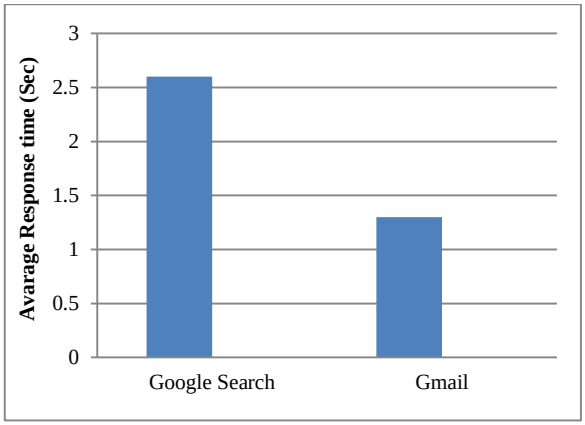

Fig. 5: average response time

Presented Speech-based Virtual assistant system is compared with existed system of ABYS (All By Yours Side) in terms of system average accuracy. The work 'ABYS' proposed by Sultan et. al. [4] presented a similar system with limited functionalities in English language for the Bengali speaking people. Fig. 6 shows Comparison of accuracy in functionalities between ABYS and this speech based virtual assistant system. It can be seen that the speech based virtual assistant system performs much better in terms of accuracy in all functionalities than ABYS system.

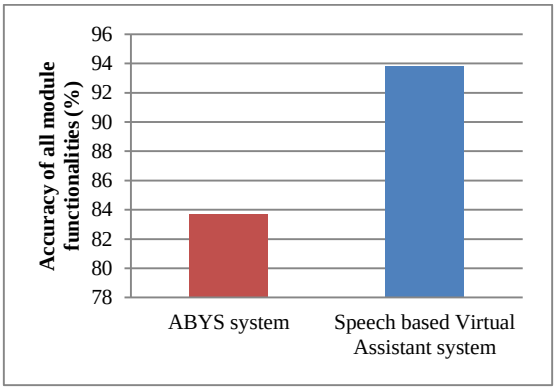

Fig. 6: Comparative analysis of Accuracy

V. CONCLUSION

The android application using speech based virtual assistant system is designed in this work for helping and guiding the visually impaired people for their daily tasks when required. A modular solution is presented in this paper for improving the web based accessibilities for visually impaired persons. The operating system of this virtual assistant is independent and can't depend on the keyboard input from a user for maximizing the ease of usage and aiming for providing a hassle-free experience to user. Using text to speech interfaces and speech to text interfaces, the user communicates and customizes with the system. This system contains five modules namely- Voice based chatbot, face detection, object recognition, text recognition, and Website access that are currently implemented. The software is working as steppingstone towards the Web 3.0 in which everything based on voice commands. The virtual assistant offers a simple way for accessing the website by visually impaired persons. The demonstrated assistant is a greater way for interaction with websites and an effective way to do.

VI. REFERENCES

[1]KanchanPatil, AvinashKharat,PratikChaudhary, ShrikantBidgar, RushikeshGavhane," Guidance System for Visually Impaired People", International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021

[2] V. Iyer, K. Shah, S. Sheth and K. Devadkar, "Virtual assistant for the visually impaired", Proc. Fifth International Conference on Communication and Electronics Systems, pp. 1057-1062, 2020.

[3]AuliaAndinia, IkaNurlailiIsnainiyah, "Design of Learning Application using Trivia Method based on Google Assistant for Vision Impairment Disability", International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)2020

[4] Md. Rakibuz Sultan, MdMoinulHoque, "ABYS (Always By Your Side): A Virtual Assistant for Visually Impaired Persons" 22nd International Conference on Computer and Information Technology (ICCIT), IEEE, December 2019, pp: 1-6.

[5] S. Mondal, S. Paul, S. Ganguly and R. Dey, "Virtual personal assistant with face recognition login system using machine learning and artificial intelligence", International Journal of Engineering Science and Computing (IJESC), vol. 9, no. 4, pp. 21603-21605, 2019.

[6] Laura Burbach, Patrick Halbach, Nils Plettenberg, Johannes Nakyama, MatrinaZiefle and Andre Calero Valdez, "Ok google Hey SiriAlexa. Acceptance relevant of virtual voice assistants", International communication conference IEEE, 2019

[7] Dipankar Gupta et. al, "A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection, International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), IEEE Conference, Dhaka, December 2019.

[8] SyedMohidul Islam et.al, "Adheetee: A Comprehensive Bangla Virtual Assistant", 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, May 2019.

[9] Khattar S, Sachdeva A, Kumar R, Gupta R, "Smart home with virtual assistant using raspberry Pi", 9th International Conference on Cloud Computing, Data Science and Engineering(Confluence), IEEE, January 2019.

[10] V. Chayapathy, G. S. Anitha, and B. Sharath, "IoT based home automation by using personal assistant," International Conference On Smart Technology for Smart Nation(SmartTechCon), IEEE, August 2017.

[11] Prince Bose, ApurvaMalpthak, UtkarshBansal, AshishHarsola, "Digital Assistant for The Blind", 2nd International Conference for Convergence in Technology (I2CT), IEEE, April 2017.

[12]Maan, C., Gunawardana, U.: Barriers in acceptance of ambient assisted living technologies among older australians. In: Life Sciences Conference, pp. 222–225 (2017)

[13] RobiahHamzah and Mohammad IzzatMohamadFadzil “Voice4Blind: The Talking Braille Keyboard to Assist the Visual Impaired Users in Text Messaging” 2016 4th International Conference on User Science and Engineering (iUSEr)

[14] FarzanaJabeen, Aslam Muhammad etal, “Feed Forward Neural Network Training Based Interactive Shopping for Blind”, 2015 12th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico, City. Mexico

[15]Dautenhahn, K., Woods, S., Kaouri, C., Walters, M.L., Koay, K.L., Werry, I.: What is a robot companion – friend, assistant or butler? In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1192–1197 (2005)