Grenze International Journal of Engineering and Technology, June Issue

Voice Assistance for Visually Impaired using ChatGPT and IOT

Ragavapriya R K[1], Ashwin Raj R[2], Deepika B[3] and Deepika Jaiswal R[4]
[1-4]Sri Ramakrishna Engineering College, Department of Electrical and Electronics Engineering, Coimbatore, Tamil Nadu, India
Email: ragavapriya.krishnasamy@srec.ac.in, ashwinraj.2003010@srec.ac.in, deepika.2003013@srec.ac.in, deepikajaiswal.2003014@srec.ac.in

***Abstract*— According To The World Health Organization (Who), A Significant Number Of People Worldwide Experience Visual Impairments, Which Can Have Profound Personal Implications As A Vast Majority Of Our Perception, Learning, And Daily Activities Rely On Vision. Consequently, Individuals with Visual Disabilities Encounter Challenges In Education And Professional Settings Due To Inaccessible Technologies And Barriers To Basic Tasks In Daily Life. To Address This Issue, We Introduce A Speech-Based Assistive System That Enables Visually Impaired And Blind Individuals To Interact With Chatgpt Through Natural Conversation. This System Integrates Automatic Speech Recognition (Asr), Text-To-Speech (Tts), And Ai-Iot To Create A Seamless Interface With Chatgpt. Through Initial Hardware Designs, We Have Demonstrated The Potential Of This System As An Assistive Technology. The Project Utilizes A Custom-Designed Chatgpt Hardware System To Provide Assistance To Individuals With Disabilities. A Microphone Is Used To Capture The Voice Input, Which Is Then Sent To The Server. In Our Server, The Chatgpt Api Processes The Message, Searches For The Necessary Response, And Displays The Message. The Response Message Is Sent To A Voice Processor Module To Convert It Into Speech, Which Is Then Played Through A Speaker Module, Allowing The Individual**
To Hear The Result In Voice Form And Respond Accordingly. Additionally, A Slave Microcontroller Is Employed To Receive Command Details From The Chatgpt Processor And Control A Display, Indicating Course Changes And Displaying Start And Stop Statuses.

***Index Terms*— Visually Impaired, ChatGPT, Microcontroller, Processor, Automatic speech recognition, Assistance.**

I. INTRODUCTION

In the realm of assistive technology, the development of an intelligent assistant system tailored to the needs of individuals with visual impairments represents a significant stride towards fostering inclusivity and empowerment. This innovative framework seamlessly integrates ChatGPT, a cutting-edge language model, with advanced hardware such as the ESP32 microcontroller, creating a holistic solution that harnesses the power of natural language processing, object recognition, and tactile feedback. Engineered to provide comprehensive support, the assistant enables users to interact effortlessly using voice commands, seamlessly converting spoken language into text and vice versa through the ESP32's voice-to-text and text-to-speech capabilities. Beyond language processing, real-time object recognition enhances environmental awareness, describing and identifying objects to aid in

Grenze ID: 01.GIJET.10.2.96_1

navigation and obstacle avoidance. Furthermore, the inclusion of a vibrating motor as a tactile feedback component enhances the user experience by offering tactile alerts based on object detection. This multi-faceted assistant system stands poised to redefine accessibility, offering a dynamic and responsive interface that caters to the diverse needs of the visually impaired community, ultimately fostering independence and enhancing daily experiences.

II. Proposed System

INPUT SIDE

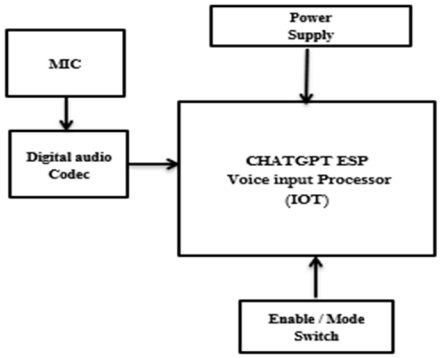

Fig.1 Input Block Diagram

OUTPUT SIDE

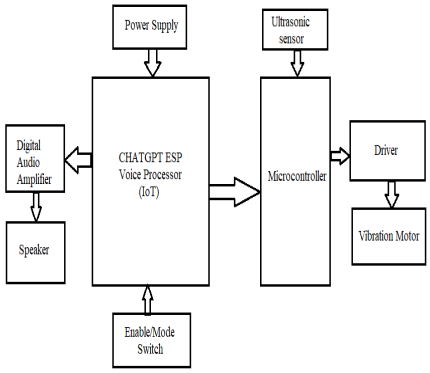

Fig.2 Output Block Diagram

In this project, an innovative feature has been introduced to ChatGPT, enabling the system to accept voice inputs seamlessly. By integrating state-of-the-art speech recognition technology, users can now interact with ChatGPT through spoken commands or queries, significantly enhancing the accessibility and usability of the platform. The proposed system operates by capturing spoken input from users, transcribing it into text format using advanced speech recognition algorithms, and then feeding it into ChatGPT for processing and response generation. This intuitive and user-friendly interface empowers individuals to engage with ChatGPT in a more natural and conversational manner, without the need for traditional text-based input methods. The system's architecture ensures robust performance in accurately transcribing diverse speech patterns and accents, while ChatGPT's advanced natural language processing capabilities enable it to generate contextually relevant and coherent responses in real-time. Overall, this project represents a significant advancement in the field of conversational AI, offering a seamless and efficient solution for voice-based interaction with ChatGPT.

This project focuses on the development of an innovative feature integrating voice input and output capabilities into ChatGPT, aimed specifically at aiding visually impaired individuals. This novel feature serves as a virtual guide, facilitating seamless communication through voice commands and responses. By harnessing the power of natural language processing, users can interact with ChatGPT simply by speaking, receiving synthesized voice responses that provide information, answer queries, and offer guidance tailored to their needs. This technology empowers the blind community by offering them a reliable and accessible means of accessing information and assistance in various contexts, thereby enhancing their independence and quality of life. Through this project, we have contributed to the advancement of inclusive technology solutions that prioritize accessibility and inclusivity for all individuals

III. List Of Components And Specifications

HARDWARE COMPONENTS

A. NODE MCU

The ESP8266EX offers a comprehensive and independent WiFi networking solution, capable of starting up directly from an external flash when hosted by ESP8266EX. To optimize performance in such applications, the system incorporates cache implementation. Moreover, it enables the introduction of wireless internet access to any microcontroller-based device with basic connection interfaces (SPI/SDIO or I2C/UART), effectively serving as a WiFi adapter. Renowned for its integration, the ESP8266EX stands out as one of the most compact WiFi chips in the industry, requiring minimal additional hardware. Its front-end module and overall design are tailored to occupy

minimal space. Beyond WiFi capabilities, the ESP8266EX boasts an advanced 32-bit Tensilica L106 Diamond series CPU, complete with on-chip SRAM, enhancing its functionality and performance.

B. Power Supply

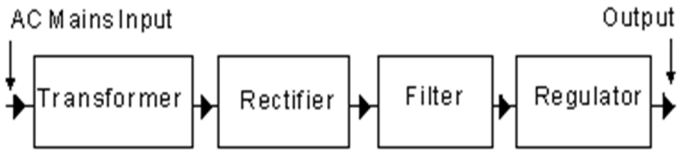

Fig.3 Power Supply Block Diagram

An AC voltage, typically at 220V rms, is initially transformed to the desired DC output level by a transformer. Subsequently, a diode rectifier generates a full-wave rectified voltage, which is then filtered through a series of capacitors to produce a DC voltage. Despite efforts to smooth the waveform, some fluctuations may still persist in the resulting DC voltage. To ensure stability and consistency in the output, a controller circuit is employed to eliminate any remaining waves and uphold the previous DC value. Typically, a voltage controller integrated circuit (IC) is utilized for this purpose, providing precise control and regulation of the DC voltage.

C. INMP441

The INMP441 stands out as an exceptional low-power, computerized output, omni directional MEMS receiver featuring a base port. Its comprehensive setup includes a MEMS sensor, signal conditioning, an analog-to-digital converter, power management circuitry, anti-aliasing filters, and an industry-standard 24-bit I²S interface. Notably, the absence of a built-in sound codec enables direct connection to digital processors like DSPs and microcontrollers, leveraging the I²S interface. With its high signal-to-noise ratio (SNR), the INMP441 is particularly well-suited for near-field applications. Furthermore, its flat wideband frequency response ensures the production of natural sound with excellent clarity. Available in a compact surface mount package measuring 4.72 × 3.76 × 1 mm, the INMP441 offers versatility and performance in a small form factor.

D. Text To Speech Speaker

The MAX98357A/MAX98357B represents a user-friendly, cost-effective solution, offering electronic pulse code modulation (PCM) input Class D amplification with outstanding audio performance and efficiency. These amplifiers boast industry-leading Class AB sound quality coupled with Class D efficiency. Notably, their automated audio recognition feature supports up to 35 different PCM and TDM timing schemes, eliminating the need for I2C programming. Additionally, operation is simplified by eliminating the requirement for an external MCLK signal typically used for PCM communication. Users can achieve audio output with just standard power, LRCLK, BCLK, and digital audio inputs, making the setup process straightforward and hassle-free.

E. LCD

The LCD (Liquid Crystal Display) is a versatile electronic display module widely used across various applications. Among its many variants, the 16x2 LCD display holds particular significance and finds extensive use in different circuits and setups. Unlike traditional displays with seven segments or multi-segment LEDs, LCD screens offer several advantages: they are cost-effective, programmable, versatile in displaying various characters and symbols, and offer adjustable brightness levels. The designation "16x2" indicates that the LCD can display 16 characters per line, with two lines available for display. Each character is represented as a matrix of 5x7 pixels on the screen. The LCD module consists of two registers: the command register and the data register. The command register stores instructions or commands given to the LCD, such as clearing the screen, positioning the cursor, controlling display settings, etc. On the other hand, the data register stores the actual information to be displayed on the screen, such as alphanumeric characters, symbols, or other graphical elements.

F. Relay

A relay acts as an electrically operated switch, triggered by an energized circuit creating a magnetic field that attracts its switch mechanism, toggling its contacts. These switches, with two positions, alternate between open and closed states. Relays serve as intermediaries, enabling one circuit to control another, even with vastly different voltage levels. For instance, a low-voltage battery circuit can connect to a 230V AC circuit via a relay. There are no direct electrical connections between the circuits; instead, a mechanical connection is established through magnetic attraction. The control circuit typically carries modest current, around 30mA for 12V systems, or up to

100mA for lower voltage applications. Since most integrated circuits (ICs) cannot supply such currents, a semiconductor device amplifies the small current from the IC to meet the switching circuit's requirements.

G. Ultrasonic Sensor

An ultrasonic sensor is a device utilized to measure the distance to an object by emitting and receiving ultrasonic sound waves. It employs a transducer to emit ultrasonic pulses and receive the reflected signals, providing information about the object's proximity. As high-frequency sound waves bounce off surfaces, they create distinct resonance patterns. Operating within the range of frequencies beyond human hearing, ultrasonic sensors emit sound waves and then detect the echoes using their transducers. Our ultrasonic sensors, like many others, utilize a single transducer for both emitting and receiving ultrasonic pulses. By calculating the time difference between sending and receiving these pulses, the sensor determines the distance to the target object.

H. DC Motor

A DC motor functions as an electromechanical energy conversion device, converting electrical energy input into mechanical energy output. Its operation relies on the principle that when a current-carrying conductor is placed in a magnetic field, a mechanical force acts upon the conductor. This force is quantified by the equation F = BIl, measured in Newtons, where B represents the magnetic field strength, I denotes the current flowing through the conductor, and l represents the length of the conductor in the magnetic field. The direction of this force is determined by Fleming's left-hand rule, which provides a way to predict the direction of motion or force in a current-carrying conductor placed in a magnetic field.

SOFTWARE COMPONENTS

A. Arduino Software (IDE)

The Arduino Integrated Development Environment (IDE), also known as Arduino Software (IDE), encompasses several essential components. It features a text editor designed for writing code, a message area for displaying feedback and notifications, a text console for displaying program output, a toolbar equipped with buttons for common functions, and a series of menus for accessing various tools and settings. The IDE facilitates seamless connectivity with Arduino and Genuino hardware, enabling users to upload programs to the boards and communicate with them effectively.

B. Writing Sketches

Programs created using the Arduino Software (IDE) are referred to as sketches. These sketches are written within the IDE's text editor and are saved with the file extension .ino. The editor is equipped with features for cutting, pasting, searching, and replacing text, providing a user-friendly environment for code editing. The message area serves to provide feedback during tasks such as saving and exporting, as well as displaying errors encountered during compilation or execution. The console section of the IDE displays text output generated by the Arduino Software, including comprehensive error messages and other relevant information. In the bottom right-hand corner of the window, users can find details about the configured board and serial port. Additionally, the toolbar offers buttons for performing various actions such as verifying and uploading programs, creating, opening, and saving sketches, and accessing the serial monitor for debugging and communication purposes.

IV. ADVANTAGES

This voice input-based ChatGPT system aids blind individuals by offering accessible, intuitive guidance through natural language interaction. Leveraging advanced language processing, it delivers personalized, real-time assistance tailored to specific needs. With hands-free voice output, users can multitask easily, enhancing convenience and independence. Overall, this represents a significant leap in assistive technology, empowering visually impaired individuals to navigate the world with confidence and ease.

V. DISADVANTAGES

While this project provides essential aid to blind individuals via voice input-based ChatGPT, it also carries limitations. Firstly, the variability of voice recognition technology may lead to misunderstandings or errors, potentially frustrating users. Additionally, exclusive reliance on voice input/output might hinder accessibility for those with speech impairments or alternative interaction preferences. Furthermore, the system may struggle with complex queries or noisy environments, affecting response accuracy. Moreover, reliance on internet connectivity

could restrict usage in areas with limited network coverage. Acknowledging and addressing these drawbacks is crucial for ensuring the system's effectiveness and inclusivity.

VI. Application

This project, a voice input-based ChatGPT system for guiding blind individuals, holds diverse applications. It aids in navigation, offering real-time directions through voice prompts. In educational settings, it acts as a virtual tutor, answering questions and facilitating learning through interactive voice interactions. In professional environments, it enhances productivity and autonomy for blind individuals. Additionally, it assists in everyday tasks like accessing information online, managing schedules, and socializing via voice-based platforms.

VII. Prototype Model

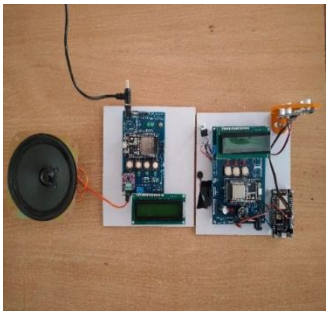

Fig.4 Prototype Model

VIII. Conclusion

The project enabling ChatGPT with voice input/output is a breakthrough in NLP (Natural Language Processing) and HCI (Human-Computer Interaction), enhancing accessibility and engagement. Integrating voice communication opens new avenues for user interaction preferences. Utilizing advanced speech tech, ChatGPT's responsiveness showcases AI (Artificial Intelligence) evolution meeting diverse needs, revolutionizing user experiences.

Acknowledgement

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

References

[1] J. Wu, W. Gan, Z. Chen, S. Wan, and H. Lin, “AI-generated content (AIGC): A survey”, 2023, arXiv:2304.06632.

[2] M. Xu et al., “Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services”, 2023, arXiv:2303.16129.

[3] Y. Cao et al., “A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT”, 2023, arXiv:2303.04226.

[4] H. Du et al., “Enabling AI-generated content (AIGC) services in wireless edge networks”, 2023, arXiv:2301.03220.

[5] Y. Wang et al., “A survey on metaverse: Fundamentals, security, and privacy”, IEEE Commun. Surveys Tuts., vol. 25, no. 1, pp. 319–352, First quarter 2023.