STM JOURNALS

Journal of
Operating Systems
Development & Trends

ISSN: 2454-9355
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoOSDT

https://journals.stmjournals.com/joosdt

Review JoOSDT

ChatGPT Based Voice Assistant for Blind People

Mohini R. Deore[1], Vaishnavi Premkumar Sangave[2,*], Pratiksha Lahu Padwal[2], Mayuri Bharat Amte[2]

Abstract

The proposed system for converting speech input into text format to facilitate interaction with ChatGPT is a sophisticated integration of hardware and cloud-based services. Utilizing state-of-the-art technologies, it facilitates seamless communication between users and the AI model. At the outset, the microphone serves as the input device, capturing audio signals from the user's speech. These signals are then amplified to ensure clarity and fidelity before being transmitted to the ESP32 microcontroller for further processing. The ESP32, known for its versatility and computational power, plays a central role in signal conversion and transmission. Utilizing its analog-to-digital converter, the ESP32 converts the analog signals from the microphone into digital format. This digital representation of the speech input is then sent to the Google Cloud platform through an API connection. Google Cloud's Speech-to-Text service, powered by advanced machine learning algorithms, analyzes the digital audio data and accurately transcribes it into text. Once the speech input is transcribed into text, it is handed over to ChatGPT for interpretation and response generation. ChatGPT, an AI language model, processes the text input to understand user queries, provide information, or engage in conversation. Its capacity to produce contextually relevant responses through natural language comprehension renders it an ideal conversational companion. To complete the loop of communication, the system employs a text-to-speech (TTS) library on the ESP32 microcontroller. This library converts the text-based responses generated by ChatGPT back into spoken format. The synthesized speech is then amplified and delivered through a speaker, allowing the user to hear the AI-generated responses in a natural and comprehensible manner. By seamlessly integrating speech recognition and text-to-speech synthesis, the system enables users to interact with ChatGPT through both spoken and text-based inputs. The hybrid approach improves accessibility and usability, accommodating a diverse array of users with different communication preferences. Overall, the system represents a significant advancement in conversational AI technology, paving the way for more intuitive and immersive user experiences.

Keywords: ChatGPT, ESP32, Google cloud, text to speech, speech to text, API

***Author for Correspondence**
Vaishnavi Premkumar Sangave
E-mail: sangavevaishnavi@gmail.com

[1]Assistant Professor, Department of Electronics & Telecommunication Engineering, Sinhgad College of Engineering, Pune, Mahrashtra, India
[2]Student, Department of Electronics & Telecommunication Engineering, Sinhgad College of Engineering, Pune, Mahrashtra, India

Received Date: April 24, 2024
Accepted Date: July 12, 2024
Published Date: July 22, 2024

Citation: Mohini R. Deore, Vaishnavi Premkumar Sangave, Pratiksha Lahu Padwal, Mayuri Bharat Amte. ChatGPT Based Voice Assistant for Blind People. Journal of Operating Systems Development & Trends. 2024; 11(2): 23–31p.

INTRODUCTION

Recent advancements have brought about significant improvements in the real-time voice assistant space, specifically customized for each platform [1]. These developments raised experiences to new heights and revolutionized user interactions on various platforms. One of those platforms that has gained recognition recently as a novel means of interacting with the virtual world and bringing convenience to a variety of lives is ChatGPT, in particular, to individuals who are having trouble obtaining services. This initiative addresses the challenges faced by blind individuals trying to use virtual services.

© STM Journals 2024. All Rights Reserved

Individuals who are blind deal with numerous difficulties every day. Controlling their surroundings and gaining access to information are two of the largest obstacles. While screen readers and Braille displays are examples of traditional assistive technology that might be useful, they are frequently costly and challenging to operate. The way blind individuals engage with the world around them could be completely changed by ChatGPT, a voice assistant for the visually impaired [2].

The objective of this project is to develop a personalized artificial intelligence voice assistant tailored for blind individuals, facilitating various functionalities such as informational support and question-answering. The ESP32 microcontroller and ChatGPT, a sizable language model from OpenAI, will serve as the assistant's foundation. To translate the user's speech to text, the assistant will be integrated with the Cloud Speech-to-Text API [3]. Compared to current assistive gadgets, the suggested system offers a number of advantages. To start with, it is inexpensive and simple to use. Secondly, it is movable and may be utilized anywhere. Thirdly, it is adaptable and suitable for a range of jobs. The suggested solution has the ability to significantly improve the lives of blind individuals by giving them a practical and affordable way to access information and control their environment.

OBJECTIVES

- Develop the virtual assistants to understand and process natural Language inputs, enabling seamless communication with Blind users.
- Implement high quality TTS capabilities to convert Text Responses into clear and Natural speech for auditory feedback.
- Enable voice command and voice control for hand-free interaction with the virtual assistance.
- Enhancing blind users access to information and services.

ELECTRONIC COMPONENTS

ESP32 Microcontroller

A versatile and powerful microcontroller ideal for IoT applications and embedded systems (Figure 1).

- The microcontroller is the brain of the system. It is responsible for controlling all of the other components, such as the Mic, Speaker and LCD display.
- A powerful and versatile microcontroller with built-in Wi-Fi and Bluetooth connectivity (Table 1).

Figure 1. ESP32 Module.

Table 1. Specification microcontroller.

S.N.	Specification	Rating
1	Operating Voltage	2.3–3.6 V
2	Frequency	2.4 GHz
3	Memory	4 MB
4	Operating current	80 mA

© STM Journals 2024. All Rights Reserved

Microphone

A device that converts sound into electrical signals for recording or amplification (Figure 2).

The INMP441, featuring an omnidirectional MEMS microphone with a bottom port, delivers exceptional performance while operating at low power consumption. This comprehensive solution encompasses an industry-standard 24-bit I2S interface, an analog-to-digital converter, an anti-aliasing filter, power management, and a MEMS sensor. The INMP441 can link directly to digital processors, including ESP32 Microcontrollers, thanks to the I2S interface, negating the requirement for an audio codec within the system. Because of its high SNR, the INMP441 is a great option for near field applications (Table 2).

Amplifier

The I2S digital audio amplifier board is designed for microcontrollers, converting I2S digital audio to amplified analog output for speakers (Figure 3). It integrates an I2S D/A converter with an audio amplifier, suitable for ESP32 or Raspberry Pi projects. Ideal for audio playback, game sounds, alarms, and I2S microphone interfacing, this compact stereo amplifier delivers up to 3.2 W on 4 Ω speakers. It operates efficiently on 2.7 to 5.5 V, making it great for portable and battery-powered projects. The I2S input supports 3.3 and 5 V levels, with adjustable gain from 3 to 15 dB (Table 3).

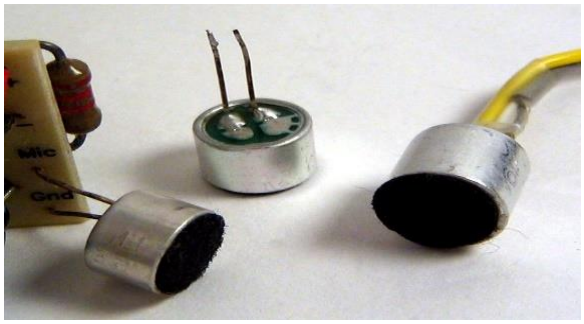

Figure 2. Microphone.

Table 2. Microphone specification.

S.N.	Specification	Rating
1	Operating Voltage	12 V
2	Sensitivity	–26 dB FS
3	SNR	61 dB
4	Power consumption	1.4 mA

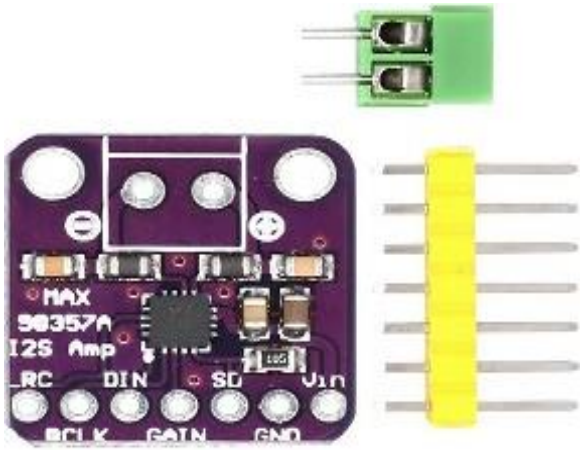

Figure 3. MAX98357 i2s 3 W Class D Amplifier.

© STM Journals 2024. All Rights Reserved

Speaker

Speakers are transducers that convert electromagnetic waves into sound waves (Figure 4). The speaker may receive digital or analog input. To create sound waves, analog speakers only amplify analog electromagnetic signals. Signals in analogy are sound waves. Digital speakers must initially convert the digital input into an analog signal before generating an output sound wave (Table 4).

IR Sensor

Infrared sensors are electronic devices designed to emit light for the purpose of detecting objects within their surroundings (Figure 5). These sensors detect infrared radiation, which is imperceptible to the human eye. They are specialized motion sensors utilizing infrared light for detection. In addition to detecting motion, infrared sensors are capable of discerning proximity, motion, and a broad spectrum of physical attributes. Furthermore, they possess the capacity to measure the temperature of objects. These radiations are detectable by infrared sensors. They work well for detection at distances of 100 to 500 cm. Working similarly to an object detecting sensor is the infrared sensor. Whether a sensor is active, or passive determines how it functions (Table 5).

Table 3. Amplifier Specification.

S.N.	Specification	Rating
1	Operating Voltage	2.7–5.5 V
2	Output power	3.2 W
3	Selectable Gain	3, 6, and 9 dB
4	Impedance	4 Ω

Figure 4. Speaker.

Table 4. Speaker Specification.

S.N.	Specifications	Rating
1	Peak power	3 W
2	Impedance	8 Ω

Figure 5. IR Sensor.

© STM Journals 2024. All Rights Reserved

Table 5. IR sensor specification.

S.N.	Specifications	Rating
1	Supply voltage	2.5–5 V
2	Supply Current	20 mA
3	Detection Range	2–30 cm
4	Active output level	0 (When output detected)

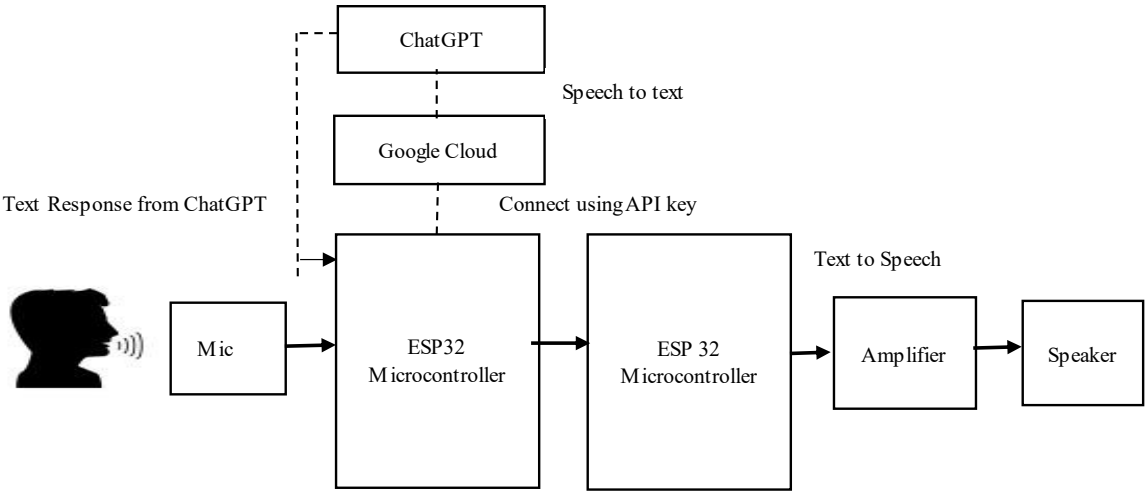

Figure 6. Block Diagram of Text Response from ChatGPT.

PROPOSED SYSTEM

The main components of the system and how they are interconnected are shown in the Figure 6.

Mic

The system will receive the audio signal through a microphone input. After that, the signal is sent to the amplifier in the next block, which will amplify the low-power audio signal. The ESP32 Microcontroller can be directly connected to the microphone via the I2S interface.

ESP32

"Capable of diverse applications, the ESP32 stands as a robust and versatile platform". The controller will take the input from amplifier. ESP32 has built-in analog to digital converter. With the help of ADC controller will covert analog input to digital which will be used for further processing. The ESP32 will send this digital data to cloud with the help of API [4].

Google Cloud

ChatGPT is platform which accepts the data only in the text format. The input we are having is in the speech format, so, further process we need to make the data convert into text format. To achieve this goal, we utilize Google Cloud services. The Google Cloud Speech-to-Text service delivers a dependable and precise solution for transcribing spoken language into text. Utilizing cutting-edge machine learning algorithms, it delivers top-notch transcription capabilities across numerous languages and dialects. This service is adaptable, accommodating various audio formats.

Speech to Text

Google Cloud's Speech-to-Text service employs cutting-edge machine learning algorithms, with a focus on deep neural networks, to accurately transcribe spoken language into text [3]. The service accepts audio input in the form of files or streaming data. These could be recordings of human speech captured from various sources like microphones, phone calls, or video recordings.

© STM Journals 2024. All Rights Reserved

Before processing the audio, the service may perform preprocessing steps to enhance the quality of the audio signal The audio data is then converted into a format suitable for analysis. This involves extracting relevant features from the audio waveform. At the core of the process lies the utilization of deep learning models, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), which are trained on extensive speech data [5]. These models improve accuracy, the service often employs language models that provide context for the transcribed text. Utilizing these models, the system predicts the most probable sequence of words by considering the conversation context or the grammar of the spoken language. It analyzes the extracted features to identify patterns and deduce the associated text. Finally, the recognized speech is converted into text and returned as output. Depending on the Configuration, the service can return the raw text or provide the additional information like timestamps, speaker identification, or confidence scores for each transcribed word. Overall, Google Cloud's speech-to-text service combines sophisticated algorithm, vast amounts of training data, the extensive language models to deliver accurate and reliable transcription of spoken language into text [6].

ChatGPT

The text data will receive at ChatGPT through API. ChatGPT will process this text data to answer the user's query. ChatGPT has its own database; with the help of its database, ChatGPT will answer the query. The response will then be send to the ESP32 for next operation [7].

Text to Speech

Using an ESP32 microcontroller along with a text-to-speech (TTS) library it is possible to convert given text data to speech. There are several TTS libraries available for ESP32 platforms which includes support for TTS functionalities. Install the necessary libraries for your ESP32 development environment. Depending on the TTS solution you opt for, you might need to download and install additional libraries. Connect a speaker or buzzer to your ESP32 board. Ensure that it is properly wired and powered. Develop the code to interact with the TTS library. This typically involves initializing the TTS engine, providing text input, and playing the synthesized speech through the speaker. Compile your code and transfer it to your ESP32 board using the Arduino IDE or another compatible development environment. Once uploaded, test your system by sending different text inputs to the ESP32 and verifying that the speech synthesis works as expected. Depending on your project requirements, you can expand the functionality by adding features such as adjusting speech parameters (pitch, speed, volume), integrating with external APIs for text input, or incorporating additional sensors for interaction [8].

Speaker

The speech data obtained from ESP32 will be amplified with the help of amplifier; then it is given to the speaker as a response to the user.

Abbreviations and Nomenclature

API: Applications Programming Interface
AI: Artificial Intelligence
ChatGPT: Chat Generative Pre-trained Transformer

LITERATURE REVIEW

A literature review of ChatGPT-based voice assistants for blind people highlights the potential of advanced AI in enhancing accessibility. These assistants can provide efficient, natural language interaction, improving daily tasks and independence for visually impaired users through responsive and intuitive voice commands (Table 6).

Flowchart

Figure 7 outlines a process where an audio signal is captured by a microphone, received by an ESP32 device, and then sent to Google Cloud for processing. The audio is converted to text using speech-to-text technology, then processed by ChatGPT to generate a response. This reply undergoes conversion to audio using text-to-speech technology and is then played through a speaker. Finally, the process ends after receiving the audio response.

© STM Journals 2024. All Rights Reserved

Table 6. Latest literature review [1–3, 8–11].

Name	Part Used	Publication Year	Analytical Method	Constituents	Reference
Subhash	Text to Speech	2020	TTS, AI-Based Voice Assistant	Microphone, TTS	[1]
Kiran *et al.*	Virtual Assistant	2023	AI	AI, Voice Assistant	[9]
Kuzdeuov *et al.*	ChatGPT	2023	Blind people, TTS synthesis	TTS, ChatGPT	[2]
Burbach *et al.*	Virtual voice assistant	2019	Virtual Voice Assistant	Virtual Assistant	[10]
Ghadge and Shelke	Speech to Text	2016	STT	STT	[3]
Marek Babiuch, Petr Foltynek	ESP32 Microcontroller	2019	ESP32 control system	ESP32 control system	[8]
Mondal *et al.*	Chatbot	2018	Question answering	Chatbot	[11]

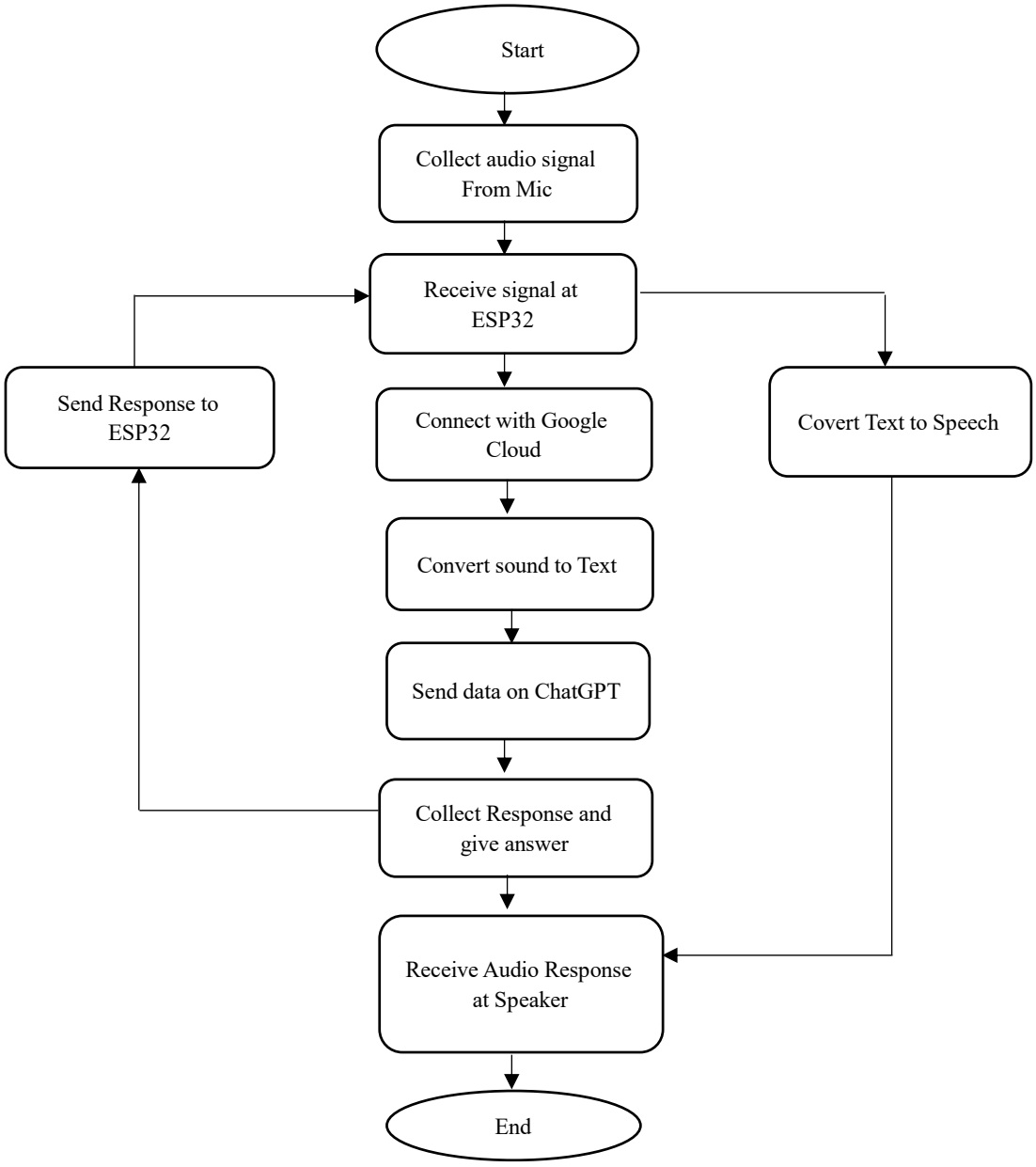

Figure 7. Flowchart.

© STM Journals 2024. All Rights Reserved

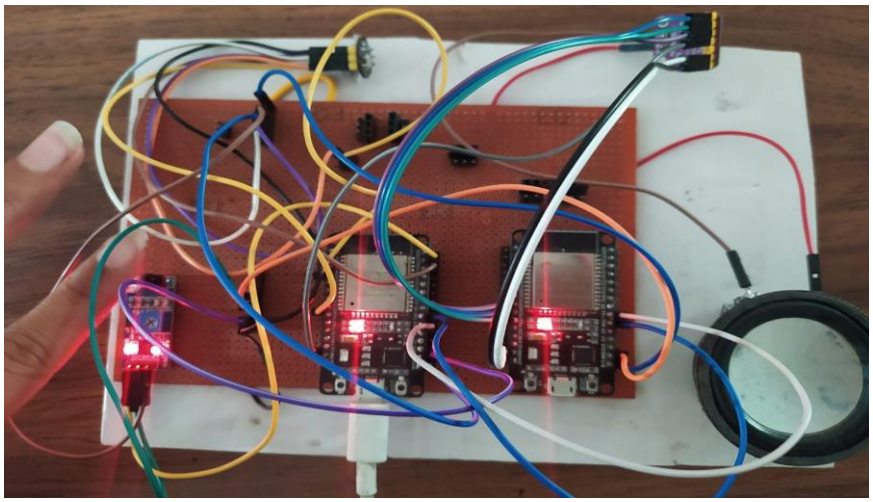

Figure 8. Result.

RESULT

The voice Assistant has shown promising results in enhancing accessibility for blind individuals in Figure 8. On going improvement will continue to address challenges and ensure a more seamless user experience. This research adds to the progress of inclusive technologies designed to empower individuals with visual impairments.

Voice-activated Interaction

Speak to your companion and receive responses generated by ChatGPT.

AI-powered Assistant

Ask questions, get information, or have casual conversations with your AI companion.

CONCLUSION

The ChatGPT-based voice assistant tailored for blind individuals marks a pivotal advancement in accessibility technology. By addressing the distinctive requirements of the blind community, this initiative has the capacity to transform their daily experiences. Through intuitive voice commands and seamless interaction, it offers equal access to information, services, and opportunities previously inaccessible. This innovation holds the promise of empowering blind individuals by fostering independence and enhancing their participation in various spheres such as education and employment. By serving as a reliable companion, it enables smoother navigation of tasks, from reading texts to accessing online resources, thereby enriching their overall quality of life. Furthermore, by leveraging cutting-edge technology, this project underscores the importance of inclusivity and underscores the potential of technology to bridge gaps and create a more equitable society.

Acknowledgement

As we attempt to articulate our profound gratitude for the invaluable support and guidance that have shaped the course of this journey, we are deeply humble. Words seem inadequate to convey the depth of appreciation we feel towards the numerous individuals who have played pivotal roles in this endeavor. First and foremost, we express our heartfelt thanks to our esteemed mentor, Prof. M.R. Deore, whose guidance, patience, and unwavering support have been instrumental in navigating the complexity of this undertaking. Under her mentorship, we have been afforded the opportunity to grow and excel, and for that, we are truly grateful.

REFERENCES

1. Subhash S. Voice Control Using AI-Based Voice Assistant. In 2020 International Conference on Smart Electronics and Communication (ICOSEC), Bangalore, India. 2020; 592–596.

© STM Journals 2024. All Rights Reserved

2. Kuzdeuov A, Mukayev O, Nurgaliyev S, Kunbolsyn A, Varol HA. ChatGPT for visually impaired and blind. In 2024 IEEE International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2024 Feb 19; 722–727.
3. Ghadage YH, Shelke SD. Speech to text conversion for multilingual languages. In 2016 IEEE International Conference on Communication and Signal Processing (ICCSP). 2016 Apr 6; 0236–0240.
4. Babiuch M, Foltýnek P, Smutný P. Using the ESP32 microcontroller for data processing. In 2019 IEEE 20th International Carpathian Control Conference (ICCC). 2019 May 26; 1–6.
5. Van der Zee RA, van Tuijl EA. A power-efficient audio amplifier combining switching and linear techniques. IEEE J Solid-State Circuits. 1999 Jul; 34(7): 985–91.
6. Kumar V, Singh H, Mohanty A. Real-Time Speech-To-Text/Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation and Abstract Meaning Representation. International Journal of Engineering and Advanced Technology (IJEAT). 2020 Apr 3; 9(4): 2361–5.
7. Ye Y, You H, Du J. Improved trust in human-robot collaboration with ChatGPT. IEEE Access. 2023 Jun 1; 11: 55748–54.
8. Babiuch M, Foltýnek P, Smutný P. Using the ESP32 microcontroller for data processing. In2019 20th International Carpathian Control Conference (ICCC) 2019 May 26: pp. 1–6. IEEE.
9. Kiran H, Girish Kumar, Hanumanta DH, Dilshad Ahmad, Lalitha S. Voice Based Virtual Assistant. International journal of Scientific Research in Engineering and Management (IJSREM). 2023; 7(7): 1–5. Available from: https://ijsrem.com/download/voice-based-virtual-assistant/
10. Burbach L, Halbach P, Plettenberg N, Nakayama J, Ziefle M, Valdez AC. "Hey, Siri", "Ok, Google", "Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants. In 2019 IEEE international professional communication conference (Procomm). 2019 Jul 23; 101–111.
11. Mondal A, Dey M, Das D, Nagpal S, Garda K. Chatbot: An automated conversation system for the educational domain. In 2018 IEEE International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). 2018 Nov 15; 1–5.

© STM Journals 2024. All Rights Reserved