iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67172

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

ChatGPT Terminal Talks Using Google TTS

Prof. C.M Maind[1], Tanishka Jadhav[2], Siddhi Godse[3]
[1]*Lecturer,* [2,3]*Student, Department of Electronics and Telecommunication Engineering, AISSMS, Polytechnic, Pune*

***Abstract**: This research explores the development of a voice-enabled ChatGPT terminal that integrates Google Text-to-Speech (TTS) technology, enhancing user interaction by allowing spoken input and output. Building upon the ChatGPT terminal introduced in the November 2023 issue, this project advances the concept by incorporating an audio output system using a 125 sound module, such as the MAX95557A, coupled with a 4-ohm loudspeaker. The implementation ensures high-quality, distortion-free speech synthesis, significantly improving accessibility and user experience. This paper discusses the hardware integration, software implementation, and potential applications of a voice-interactive AI assistant in various domains, including assistive technology, smart devices, and hands-free computing.*

***Keywords**: ChatGPT terminal, Google Text-to-Speech (TTS), VoiceGPT, Speech synthesis, Audio output system, MAX95557A, 4-ohm loudspeaker, Distortion-free audio, AI assistant*

I. INTRODUCTION

The evolution of artificial intelligence (AI) has led to the development of conversational agents like ChatGPT, which have revolutionized human-computer interactions. While these AI models have proven highly effective in text-based communication, integrating speech synthesis has become a crucial next step. This project builds upon the ChatGPT terminal introduced in the November 2023 issue, enhancing it with voice output functionality using Google Text-to-Speech (TTS). The new design incorporates an advanced 125 sound module, such as the MAX95557A, and a 4-ohm loudspeaker to deliver clear and distortion-free speech output. The ability to audibly relay questions and responses enhances accessibility and user engagement.

This research explores the integration of the ESP32 board, Google TTS, and other key hardware components to create a ChatGPT terminal that speaks aloud. The implementation of Google TTS is crucial, as existing free sound libraries such as ESP8266SAM.h and AudioGeneratorRTTTL.h are either of poor quality or unsuitable for extended speech. The system requires only three GPIO pins and an internet connection to function, making it an efficient and cost-effective solution. This paper discusses the circuit design, working principles, software implementation, and practical applications of this enhanced ChatGPT terminal.

II. LITERATURE REVIEW

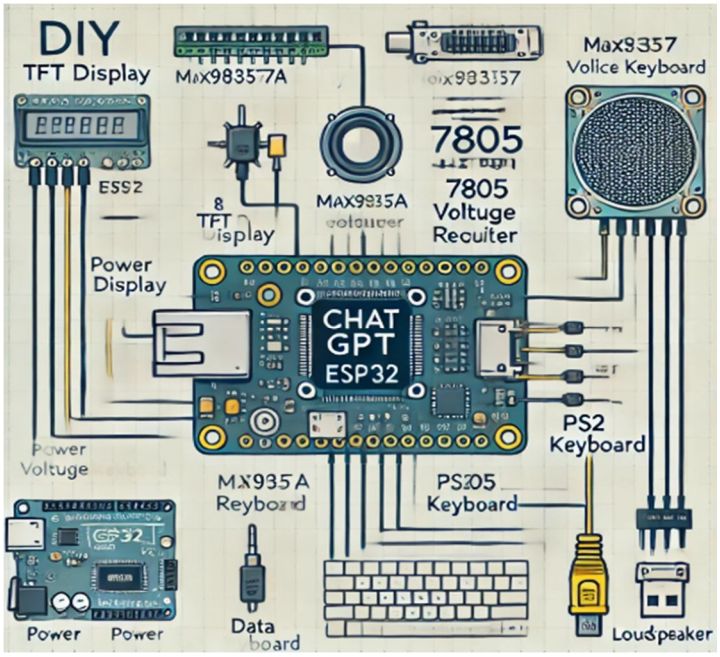
Speech synthesis has significantly evolved over the past decades, with various technologies improving the clarity, naturalness, and accessibility of computer-generated speech. Early developments in text-to-speech (TTS) systems relied on rule-based phoneme synthesis, which often resulted in robotic and unnatural speech. However, modern advancements in neural TTS, such as Google's WaveNet, have enabled near-human-quality speech synthesis.

The ESP32 microcontroller has gained widespread adoption due to its affordability, Wi-Fi capabilities, and support for various peripherals. Several studies have explored its integration with speech synthesis, such as the use of ESP8266SAM.h for limited voice output. However, this library suffers from poor sound quality and limited vocabulary. Research comparing different speech synthesis solutions for microcontrollers suggests that cloud-based solutions like Google TTS offer the best balance of quality and flexibility.[1]

Prior implementations of voice-enabled AI assistants have relied on powerful computing devices such as Raspberry Pi and desktop computers. This project, however, achieves similar functionality using an ESP32 board, making it a low-cost and energy-efficient alternative. The integration of MAX95557A ensures high-quality sound output, overcoming limitations seen in previous designs that struggled with low power and distortion.[2]

Furthermore, existing voice-interactive AI assistants often require complex setups with multiple components, increasing cost and power consumption. By leveraging Google TTS, the system ensures clear speech synthesis without the need for extensive local processing. The project's uniqueness lies in its ability to achieve high-quality voice output while maintaining low hardware requirements, thus making it an ideal solution for DIY enthusiasts, accessibility applications, and cost-sensitive implementations.

III. PROPOSED SYSTEM



A. Visual Diagram Of ChatGpt Terminal That's Talk To Speech Using Google Assistant

The proposed ChatGPT terminal with voice output is designed to provide an enhanced interactive experience by integrating Google TTS with an ESP32-based system[1]. The hardware components include the ESP32 board, a MAX95557A sound module, a 4-ohm loudspeaker, a 3.5-inch TFT display, and a PS2 keyboard for text input. The system is powered by a 5V voltage regulator and can function within a voltage range of 5V to 9V.

The key aspect of this design is the integration of Google TTS, which is responsible for converting textual responses from ChatGPT into spoken output. Unlike previous implementations that relied on local audio synthesis libraries, which had significant limitations, this system uses an internet-based TTS engine to achieve high-quality speech output. The speech synthesis is triggered when the user inputs a query via the PS2 keyboard, which is then sent to ChatGPT via an API request. The received response is displayed on the TFT screen and simultaneously spoken aloud through the loudspeaker.

A critical enhancement in this system is the use of MAX95557A, which amplifies the speech output, ensuring clear and distortion-free audio[5]. Additionally, the gain control of the amplifier is grounded to maximize output power. The system also incorporates Google TTS API constraints, limiting spoken output to 200 characters at a time while displaying the full response on the screen.[6]

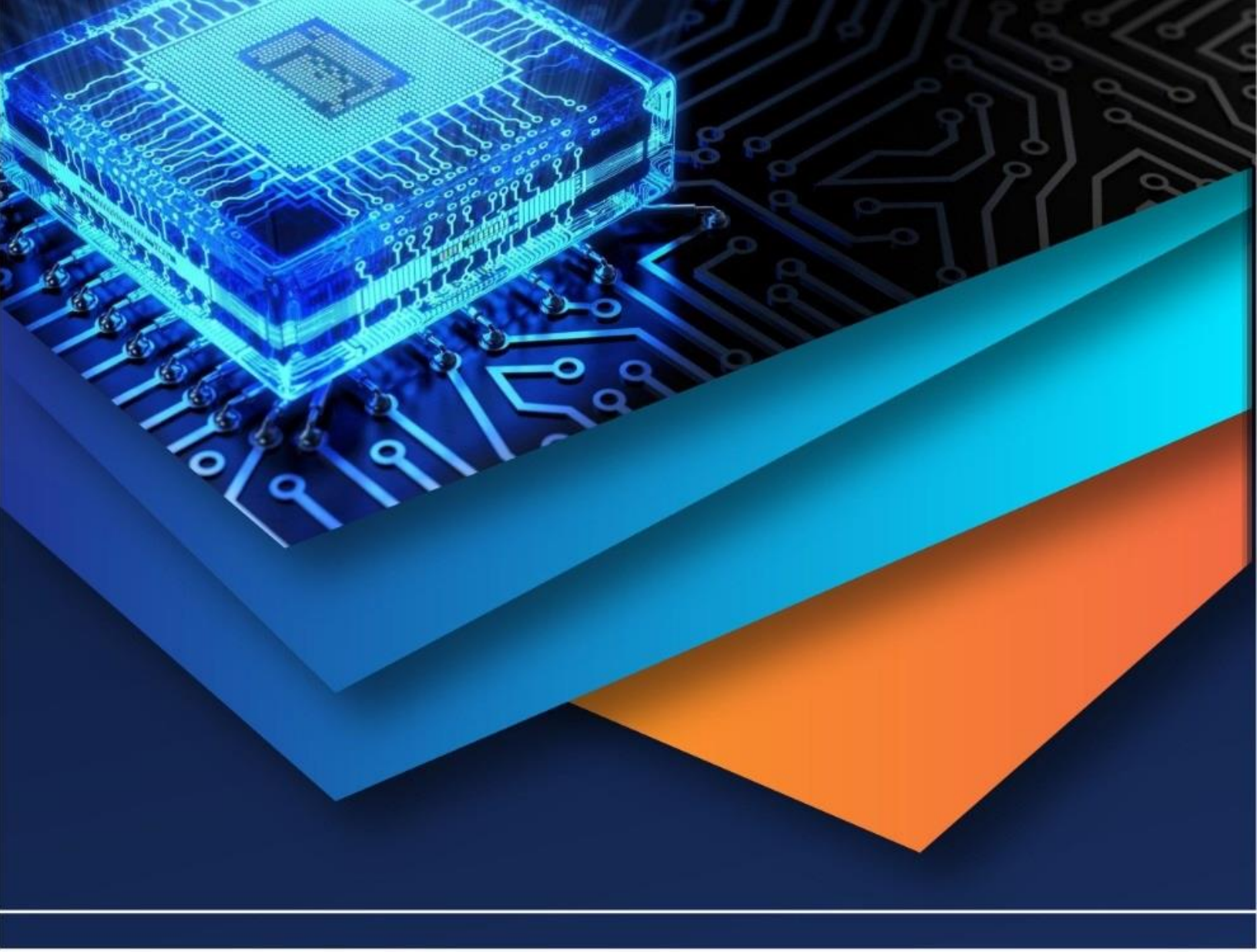
The software implementation is carried out in Arduino IDE, where necessary configurations for Wi-Fi connectivity, API authentication, and GPIO interfacing are programmed. The open-source nature of the implementation makes it accessible for further modifications and improvements, paving the way for future enhancements such as full voice interactivity, where the terminal can listen to spoken queries and respond accordingly.

IV. CONCLUSION

This research presents a novel approach to enhancing the ChatGPT terminal by integrating Google TTS for high-quality speech output. The project effectively demonstrates how an ESP32-based system, combined with a MAX95557A sound module and a 4-ohm [5]loudspeaker, can produce clear and intelligible voice responses. Unlike previous implementations that suffered from poor sound quality or complex hardware requirements, this system achieves an optimal balance of performance, simplicity, and cost-effectiveness. By leveraging cloud-based speech synthesis, the system ensures flexibility and scalability, making it suitable for various applications, including assistive technology, hands-free computing, and AI-driven voice assistants. Future work will focus on improving voice interactivity, allowing the system to process spoken queries, thereby creating a fully functional conversational AI device. This development represents a significant step forward in making AI-driven voice assistants more accessible and practical for a wide range of users.

REFERENCES

[1] https://bookauthority.org/books/best-esp32-books
[2] https://www.elektormagazine.com/articles/esp32-personal-ai-terminal-chatgpt
[3] https://uk.pinterest.com/pin/diy--183943966024673662/
[4] https://www.sciencedirect.com/science/article/pii/S2950162823000176
[5] https://www.analog.com/media/en/technical-documentation/data-sheets/max98357a-max98357b.pdf
[6] https://www.instructables.com/Absolute-Beginners-Guide-to-TFT-LCD-Displays-by-Ar/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY