Prof. Anamika Nandan, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Voice Assistant Using Gemini

Prof. Anamika Nandan, Suhothra Ks, Subhashith N, Shambu Gr, Rahul Hl
Dept Of CSE-DS,AMCEC,Bengaluru

Abstract - The upgraded Voice Assistant built using Gemini is a compact and intelligent embedded device designed for smooth, voice- driven interaction. A Raspberry Pi acts as the main processor, managing speech recognition, AI communication, and audio playback. A USB microphone captures speech input, while an audio amplifier and mini speaker generate clear vocal responses. The Raspberry Pi interprets the user's query, forwards it to the AI model, and outputs the response in both audio and text form. An ESP32 microcontroller receives the text and displays it on an LCD screen for visual feedback. Powered through standard USB sources, the system offers hands-free assistance capable of answering questions, controlling IoT devices, and enabling real-time AI interactions. With its combination of AI processing, enhanced audio output, and microcontroller-based display, the system fits applications in home automation, education, accessibility, and personal assistance.

Keywords - Voice Assistant, Gemini AI, Raspberry Pi, ESP32, USB Microphone.

I. INTRODUCTION

An AI-based voice assistant is a smart and user-friendly device that interacts with people through spoken commands. It captures a user's voice using a microphone, interprets the input using AI, and responds naturally through a speaker functioning like an intuitive digital companion.

A small display can enhance interaction by showing output text, system messages, or instructions. Since it can operate with modern AI models, the assistant can answer questions, deliver information, and help with day-to-day tasks. Designed for hands-free usage, it integrates intelligent behavior into a compact device, making technology more accessible to everyone.

Voice assistants represent a major step forward in human– computer interaction. Using microphones, speech-to-text systems, natural language understanding, and text-to-speech synthesis, they eliminate the need for physical buttons while providing quick responses. Originally based on advancements in audio processing dating back decades, modern assistants now run efficiently on embedded hardware like the ESP32, enabling cost-effective IoT applications.

II. SYSTEM DESIGN

The proposed system integrates a Raspberry Pi as the main processor that listens, processes, and responds to voice commands. Speech captured through the USB microphone is converted into text and forwarded to the Gemini/ChatGPT model for generating replies.

The output is produced in two forms:

- As spoken audio through an amplifier and mini speaker
- As text sent to the ESP32 microcontroller for LCD display

Communication between Raspberry Pi and ESP32 occurs through serial or Wi-Fi. USB power supplies provide stable operation for all components.

Hardware Components

- ESP32 Microcontroller: A low-power Wi-Fi/Bluetooth enabled board useful for IoT and peripheral control.
- LCD Display: A flat-panel display using liquid crystals and a backlight for visual output.
- USB Microphone: Captures high-quality voice input via USB.
- Speakers: Provide clear audio output.
- Raspberry Pi: A single-board computer handling speech recognition, AI communication, and audio synthesis.

- SD Card: Provides extended storage for configuration files or audio data.

Software Components

- Arduino IDE: Used to program the ESP32 with Embedded C.
- Embedded C: Manages low-level ESP32 operations such as display control and communication.
- Python: Used on the Raspberry Pi for speech processing, API communication, and audio generation.

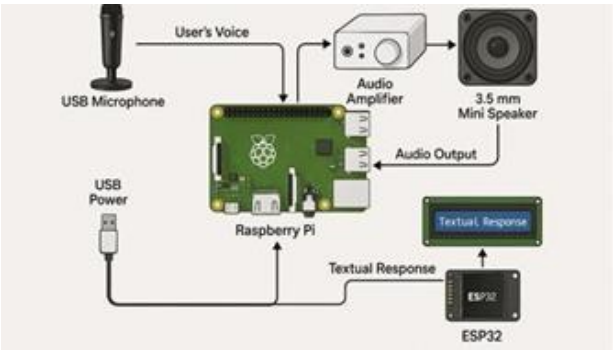

Fig. 1 Diagram of System Architecture

III. METHODOLOGY

The system workflow begins with the Raspberry Pi, which manages voice input, interacts with the AI model, and delivers the output. The microphone captures speech, which is processed into text and sent to Gemini/ChatGPT.

The AI's response is converted to speech and played through a speaker, while the text version is transmitted to the ESP32 for real-time display. Power is supplied through standard USB sources.

Implementation

The solution is implemented as a compact embedded setup using a Raspberry Pi for AI and audio tasks, and an ESP32 for display operations.

The Raspberry Pi continuously monitors the microphone, converts speech to text, communicates with the AI API, and produces speech output via an amplifier-driven speaker.

Simultaneously, the Pi sends text responses to the ESP32 using serial or Wi-Fi communication.

The ESP32 updates the LCD with the output message. Power is provided through USB connections, ensuring stable performance.

This divided architecture—processing on the Pi, display on the ESP32 makes the assistant suitable for educational use, home automation, and personal assistance.

Results

The developed system successfully performs voice input recognition and generates AI-based responses. During testing, the Raspberry Pi consistently detected commands, processed queries, and produced clear, natural-sounding audio replies within a few seconds.

The ESP32 and LCD reliably displayed the textual output, allowing users to view responses even when audio was not

Education

Students can ask questions, receive explanations, or listen to study materials. The system is especially helpful for visually impaired learners.

Healthcare

The assistant can provide reminders for medications, appointments, and daily health routines. It can also help elderly users communicate with caregivers.

Personal Assistant

It can manage reminders, alarms, calendars, and provide general information such as weather and news updates.

IoT Control

The assistant acts as a voice-controlled hub for smart sensors and actuators, enabling the monitoring and control of IoT devices in real time.

Future Scope

The voice assistant can be extended to support multimodal interaction, combining voice, touch, and camera input so that it can recognize objects, read documents, or provide context aware assistance in

real time. With more powerful on device models or edge accelerators, parts of the Gemini/ChatGPT pipeline could run locally, reducing latency, improving privacy, and enabling offline or low connectivity operation.

Future enhancements may include multimodal interaction by integrating voice, touch, and camera inputs. On-device AI models or edge accelerators could reduce latency and allow offline operation.

The system can also expand into a multi-node assistant network with multiple ESP32-based units sharing sensors and microphones. Personalized learning, emotional-aware responses, and support for regional languages can further improve user experience.

IV. CONCLUSION

The AI Voice Assistant demonstrates an effective fusion of modern AI with embedded hardware to deliver interactive, hands-free assistance. By combining speech input, intelligent processing, and audio-visual output, it simplifies user interaction and showcases the potential of embedded AI applications. With its flexible hardware design and real-time capabilities, the system serves as a practical example of daily- life AI integration.

REFERENCES

1. L. Lazzaroni et al., "An embedded end-to-end voice assistant," Engineering Applications of Artificial Intelligence, 2024.
2. F. L. I. Dutsinma et al., "A Systematic Review of Voice Assistant Usability: An ISO 9241 Perspective," 2022.
3. C. Jose et al., "Accurate Detection of Wake Word Start and End Using a CNN," Amazon Alexa Research, Interspeech, 2020.
4. Implementation and Applications of WakeWords Integrated with Speaker Recognition — A Case Study," arXiv / ResearchGate preprint, 2024.
5. M. Pavan et al., "TinySV: Speaker Verification in TinyML with On-device Learning," IEEE, 2025.
6. S. Heydari et al., "Tiny Machine Learning and On-Device Inference: A Survey," MDPI Sensors, 2025.
7. S. Chittepu et al., "Empowering voice assistants with TinyML for user-centric applications," Nature Scientific Reports, 2025.
8. "ESP32-Based Voice Assistant Integrations and Project
9. Reports," collection of project and tutorial papers, 2024–2025.
10. J. R. Patel and K. V. Wong, "Privacy Perceptions and User Trust in Voice-Activated Systems: A Systematic Review," 2022.
11. R. Thakur and H. Sen, "Error Recovery in Conversational Interfaces: A Review Based on ISO Usability Standards," 2022.
12. V. Narang and H. Silva, "Emotional Intelligence and Conversational Tone in AI Voice Assistants," 2021.