Journal on Electronic and Automation Engineering
Vol: 4(2), June 2025
REST Publisher; ISSN: 2583-6951 (Online)
Website: https://restpublisher.com/journals/jeae/
DOI: https://doi.org/10.46632/jeae/4/2/46

# Design and Implementation of AI Voice Assistant Speaker

***N. Ramya, U. Rahul, S. Vasudev, T. Ashwith, U. Gangasagar**
*Nalla Malla Reddy Engineering College, Hyderabad, Telangana, India.*
*Corresponding Author Email: ramya.ece@nmrec.edu.in

**Abstract:** *As voice interfaces become increasingly essential for smart devices, the demand for efficient, embedded AI solutions that operate reliably in real time is increasing. Audio input is captured using a high-sensitivity I2S microphone, which is processed over the I2S interface to ensure high-quality digital audio streaming. The captured voice data is securely transmitted over Wi-Fi with TTL encryption to the Deepgram cloud-based ASR engine for accurate speech-to-text conversion. The resulting text is then sent to Google Gemini AI for advanced natural language understanding, allowing the system to interpret user intent in context. Based on the interpreted query, a response is generated and sent via a text-to-speech (TTS) engine, either cloud-based or local, before being output via an onboard DAC and a connected audio amplifier. The system uses a hybrid processing model: simple voice commands, such as toggling GPIOs, controlling devices, or accessing local sensor data, are processed locally to reduce latency, while more complex or open-ended queries are handled in the cloud. A lightweight command parser on the ESP32 detects and processes predefined keywords or phrases. The entire communication system is designed to prioritize secure, low-latency performance, ensure fast response times, and protect user data. The modular design allows for easy customization and integration with additional sensors, devices, or third-party services. This voice assistant platform is a versatile solution for edge AI applications, ideal for smart home automation, IoT systems, portable assistants, and voice-controlled embedded devices.*

**Keywords:** *Embedded AI, real-time processing, I2S microphone, I2S microphone encryption, natural language understanding (NLU), text-to-speech (TTS), cloud processing, secure communication, edge AI applications, portable voice assistant.*

## 1. INTRODUCTION

Recent advances in voice-controlled smart assistants have led to natural and interactive human-computer interfaces. For example, Sudarsan et al. introduced a camera-enabled Alexa smart speaker system using a Raspberry Pi and a Reespeaker v2 microphone array. Their prototype focuses on enhancing voice interactions through digital signal processing (DSP) techniques such as beamforming and echo cancellation, along with cloud services for command execution and object detection. However, relying on cloud processing introduces challenges such as latency, power consumption, and privacy issues. To overcome these, the current project presents a compact, AI-powered voice assistant based on the ESP32 microcontroller and Google's Gemini AI. By integrating a hybrid architecture that combines local processing with cloud-based natural language processing (NLP), this approach supports real-time, privacy-aware interactions, while ensuring low power consumption and modularity for open-source development.

In a similar study, Subhash et al. built an AI-based voice assistant using Python, GTTS (Google Text-to-Speech), and the Play Sound package for real-time voice interactions. Their system processes audio input into text and produces audio responses in English, demonstrating various features such as Google searches, media playback, location queries via Google Maps, news reading, and computer commands. While useful in demonstrating core voice assistant functionalities, their system relies on a PC-based environment and lacks edge processing or hardware-level integration. In contrast, the current project focuses on building a small, low-power AI voice assistant with an ESP32 microcontroller, integrating cloud services such as Deepgram and Gemini AI for speech recognition and natural language understanding. This hardware-based approach offers advantages in portability, power efficiency, offline capabilities, and a secure embedded design, making it ideal for IoT applications.

As embedded AI continues to evolve, many studies have focused on building voice-controlled personal assistants on microcontroller-based platforms. A notable example is the Neobot developed by Vashista et al., a Raspberry Pi-based personal assistant capable of performing various tasks via voice commands. Their system integrates IR sensors, a Pi camera, OCR capabilities, and the Google Assistant API, supporting features such as text-to-speech conversion, motion control, and web-based information retrieval. While this system is useful, the reliance on the Raspberry Pi increases power consumption and limits portability. In contrast, the current project uses an ESP32 microcontroller to provide a more power-efficient and portable solution, which is combined with Deepgram for speech recognition and Google Gemini AI for advanced natural language understanding. This hybrid edge-cloud architecture offers faster response times, improved privacy, and flexible integration with IoT devices, making it well-suited for real-time, mobile applications.

Burbach et al. examined the adoption of virtual voice assistants (VAs), and found that privacy was the most important factor in user adoption, even surpassing price and language processing performance. While users appreciated advanced NLP features such as context-aware interaction, concerns about always-on voice assistants that send data were widespread. This highlights a major challenge in VA design – striking a balance between functionality and privacy. Unlike traditional, always-connected systems like Alexa or Siri, the current work proposes a hybrid AI assistant built on an ESP32 microcontroller, where basic commands are processed locally, and only complex queries are sent to secure cloud services like Deepgram and Google Gemini AI. This system ensures user data privacy while maintaining responsive, efficient interactions. Addressing the privacy concerns identified by Burbach et al., the proposed system provides users with greater control over their data while maintaining high-quality interactions. Another notable contribution is that of Choudhary et al., who developed a domain-specific intelligent personal assistant (IPA) that can understand bilingual voice commands in English and Bengali. Their system uses the CMU Sphinx-4 speech recognition engine and a deterministic finite state automaton (FSA) for natural language processing. By focusing on domain-specific commands and using a generalization function, the system simplifies the problem by isolating key subject-verb pairs, which leads to efficient interpretation of user input. This approach works well in environments with limited training data and supports multilingual processing, relying on pre-trained static models and lacking real-time adaptive or cloud intelligence. In contrast, the proposed ESP32-based voice assistant incorporates a hybrid architecture that combines on-device processing with cloud-based AI services such as Deepgram for speech recognition and Google Gemini AI for natural language understanding. This not only improves response time and ensures privacy, but also supports dynamic, scalable interactions suitable for IoT and wearable systems. Mishi et al. conducted a study on the capabilities of the ReSpeaker Mic Array 2.0 for real-time speech processing algorithms, including acoustic echo cancellation, direction of arrival (DOA), source counting, and speech separation. Their experiments, which used the built-in XMOS XVF-3000 DSP chip and a Raspberry Pi, demonstrated the effectiveness of multiple microphone arrays in improving speech recognition, especially in noisy environments or at a distance. The study also evaluated Google's speech-to-text (STT) service, highlighting configuration sensitivity such as automatic gain control and noise reduction. Although the system is based on a Raspberry Pi and ReSpeaker 2, the findings directly inform an existing project that uses MEMS microphones with an ESP32 microcontroller, the Deepgram API (ASR) for automatic speech recognition, and Google Gemini AI (NLP) for natural language processing. By adopting an edge-processing architecture with cloud-enhanced features, the project addresses similar challenges related to audio clarity, latency, and environmental robustness, while providing a low-power, more portable alternative to a Raspberry Pi-based system.

## 2. LITERATURE SURVEY

Hearing impairment is a prevalent issue affecting about 6.2% of the global population, severely impacting individuals' quality of life, particularly in areas like education, social engagement, cognitive growth, and mental health. One of the most significant challenges for those with hearing loss is the difficulty in distinguishing a single voice amid background noise, commonly referred to as the cocktail party effect, which those with normal hearing can naturally manage. This issue makes it especially challenging to participate in group discussions or engage in conversations in noisy settings. To address this, the current study introduces an innovative Smart Hearing Aid aimed at improving speech comprehension in noisy environments. The device uses a microphone array placed on a flat surface to capture sound from all directions, utilizing intelligent voice algorithms to isolate and prioritize the speaker’s voice while reducing background noise and competing conversations. It also features a smartphone interface for easy customization of acoustic settings. This paper outlines the design, development, and functionality of the proposed Smart Hearing Aid, aiming to restore more natural and effective auditory experiences for those with hearing impairment.

In automotive settings, robust speech recognition is especially difficult due to various and dynamic background noises, such as engine sounds, road noise, and passenger conversations. To ensure accurate speech recognition, effective front-end processing is required to enhance the speech signal before it enters the recognition system. This paper proposes a comprehensive front-end speech enhancement approach tailored for automotive environments. The method integrates hybrid voice activity detection (VAD), relative transfer function (RTF)-based generalized sidelobe cancelation, and single-channel post-filtering techniques to isolate and enhance the target speech. By training deep neural networks (DNNs) with data from four driving conditions, the system can accurately estimate speech and noise characteristics, allowing for precise speech presence detection. These estimations are fused with traditional energy-based VAD, leading to a robust hybrid VAD that drives the subsequent enhancement steps. Real-world testing in automotive environments showed significant improvements in VAD performance and overall automatic speech recognition (ASR) accuracy, proving the effectiveness of the approach in handling noisy conditions typical in vehicles.

The increasing integration of technology in education has highlighted the potential of companion robots as educational facilitators. However, many current systems fall short due to limited content diversity and insufficient domain expertise. This study addresses these gaps by introducing an advanced educational companion robot designed specifically for primary school subjects. The robot includes voice-based question-answer and control features to enhance interaction and provide a broader and more professional knowledge base. Utilizing a structured FAQ system alongside a Turing robot, the platform offers focused guidance, enriching the learning process. Experimental results show that this approach not only improves the quality of educational content but also enhances the interaction between learners and robots, positioning educational companion robots as valuable tools for primary education. The robot's design and development are aimed at overcoming the limitations of previous systems, significantly improving voice interaction quality and providing effective educational support.

Hearing impairment remains a significant global health issue, affecting over 466 million people and influencing various aspects of life, including education, employment, social interaction, and cognitive health. One of the most common challenges for those with hearing loss is difficulty understanding speech in noisy environments, often due to the loss of the "cocktail party effect" that allows normal-hearing individuals to focus on a single speaker amidst background noise. This can hinder basic interactions, such as participating in group conversations or communicating in crowded public spaces. To address this, this paper presents the development of an advanced Smart Hearing Aid that uses a microphone array and intelligent voice processing algorithms to enhance the primary speaker's voice while minimizing background noise. The device, which is placed on a flat surface like a table, dynamically tracks and switches between speakers as conversations progress. Additionally, a smartphone interface allows users to fine-tune the device's audio settings to their preferences, offering a more personalized and effective auditory experience for individuals with hearing impairment.

The popularity of voice-controlled smart assistants like Amazon Alexa has surged, enabling users to interact with devices using natural language commands. Despite significant advancements, current Alexa devices are limited to audio interactions and lack integrated camera systems, which restricts the development of real-time image and video-based applications. To fill this gap, there is growing demand for a camera-enabled Alexa smart speaker that supports a wider range of AI-driven functionalities. While Alexa has become a leading voice assistant, its inability to process visual data limits its potential for advanced applications like object detection, facial recognition, and visual scene understanding. This paper introduces a novel, camera-enabled Alexa smart speaker platform designed to overcome these limitations. This advancement expands Alexa's capabilities and sets the foundation for the next generation of smart assistants, integrating both auditory and visual intelligence to provide richer, more interactive user experiences.

## 3. EXISTING SYSTEM

**Amazon Alexa:** Amazon's Alexa is one of the most widely recognized and adopted voice assistants around the world. It seamlessly integrates with a wide array of smart home devices, including lights, thermostats, security cameras, and smart TVs. Alexa processes voice commands through cloud-based servers, using Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) algorithms to understand and respond to user requests. Alexa is capable of performing a range of tasks, such as answering questions, playing music, controlling smart home devices, setting reminders, and providing weather updates. While Alexa offers extensive functionality, its reliance on cloud-based processing requires a stable internet connection, and may raise concerns related to data privacy and security.

**Google Assistant:** Google Assistant is another leading voice assistant, found on various devices such as smartphones, Google Home, and smart displays. It utilizes Google’s machine learning algorithms and is closely integrated with

Google Search, making it efficient in answering questions, performing voice searches, and providing personalized responses. Like Alexa, Google Assistant also depends on cloud-based processing to understand speech, interpret context, and generate meaningful answers. A key strength of Google Assistant is its ability to deliver contextualized results by leveraging data from Google services like Gmail, Google Calendar, and YouTube. However, as with Alexa, the reliance on cloud infrastructure introduces potential latency issues and raises concerns about privacy, particularly with data being processed online.

**Apple's Siri:** Apple's Siri is an AI-powered voice assistant integrated into Apple devices such as iPhones, iPads, Macs, Apple TVs, and the HomePod smart speaker. While Siri generally depends on cloud-based processing, Apple has introduced some on-device processing for specific tasks, improving both speed and privacy. For instance, certain actions, such as setting reminders or sending messages, are processed locally without needing an internet connection. Siri is tightly integrated with the Apple ecosystem, enabling users to control HomeKit-enabled smart devices, set reminders, send messages, and interact with Apple services like Apple Music and Apple Maps. Despite its many capabilities, Siri has faced criticism for its limited conversational abilities and occasional inaccuracies in handling complex queries compared to competitors like Google Assistant.

**Microsoft Cortana:** Initially launched as a competitor to Siri and Google Assistant, Microsoft Cortana has evolved into a productivity-focused assistant. It was originally available on a range of devices, including Windows PCs, smartphones, and Xbox, but Microsoft has since shifted its focus to integrating Cortana primarily within productivity tools like Microsoft Office (Word, Excel, Outlook) and Microsoft Teams. Cortana aids with tasks like setting reminders, managing calendars, and responding to emails. While it has been integrated with cloud-based systems and NLP technologies, Cortana lacks the level of integration with smart home devices and third-party applications found in Alexa and Google Assistant. Its reduced presence in the consumer market is attributed to the competitive landscape and shifting priorities.

The dominance of other voice assistants and Microsoft's shift towards a productivity-centric AI strategy has affected Cortana's position in the market.

**Samsung Bixby:** Samsung's Bixby is an AI-powered voice assistant featured on a variety of Samsung devices, including smartphones, smart TVs, refrigerators, and wearables. Bixby sets itself apart by providing advanced voice control for device-specific tasks, allowing users to adjust settings within Samsung's ecosystem. Unlike other voice assistants, Bixby offers deep integration with device hardware and supports contextual actions, enabling users to interact with apps and settings using voice commands. However, Bixby has not achieved the same level of market penetration or user adoption as Alexa or Google Assistant. Despite its advanced voice recognition capabilities, Bixby faces limitations in third-party integrations and struggles with less natural conversational flow, leading to lower user satisfaction compared to its competitors.

**Re Speaker and Microphone Array Systems:** Advancements in voice assistant hardware have led to the development of devices like the Re Speaker v2, which includes a microphone array designed to improve voice capture and noise suppression. These systems are particularly beneficial in noisy environments where traditional microphones struggle to capture speech accurately. The microphone array enables directional sound processing and beamforming, allowing the voice assistant to prioritize the main speaker's voice while filtering out background noise. Such systems are especially useful in smart homes, where voice commands need to be understood from a distance or in noisy rooms. The integration of advanced audio processing and AI algorithms in these microphone arrays represents a significant improvement in the performance of voice assistants in real-world settings.

**Edge AI and On-Device Processing:** The demand for edge AI—processing data locally on the device rather than relying on cloud services—has grown, particularly concerning privacy and real-time performance. Companies like Google and Amazon have incorporated on-device machine learning models to reduce latency and enhance privacy by keeping sensitive data within the device rather than transmitting it to the cloud. Devices such as the Google Nest Hub Max and Amazon Echo, equipped with onboard AI processing, can handle tasks like speech recognition, face detection, and gesture control more efficiently and securely. This shift toward on-device processing is crucial for creating portable AI voice assistants that can function without a constant internet connection, making them particularly beneficial in remote or offline environments where cloud connectivity may be unreliable.

**Smart Home Ecosystem Integration:** As the Internet of Things (IoT) expands, voice assistants are increasingly integrated into a wide variety of smart home devices, enabling users to control and automate their home environments. Systems like Amazon Alexa, Google Assistant, and Apple HomeKit allow users to control lighting, thermostats,

security systems, and even household appliances, such as refrigerators and washing machines, all through voice commands. The integration of voice assistants into IoT devices has revolutionized how users interact with their homes. However, while these systems are widely adopted, they often require users to invest in specific devices within a particular ecosystem, which can lead to compatibility issues between brands and devices. As more companies develop their own voice assistants and smart devices, ensuring cross-compatibility will become a key challenge to prevent fragmentation in the smart home market.

## 4. BLOCK DIAGRAM

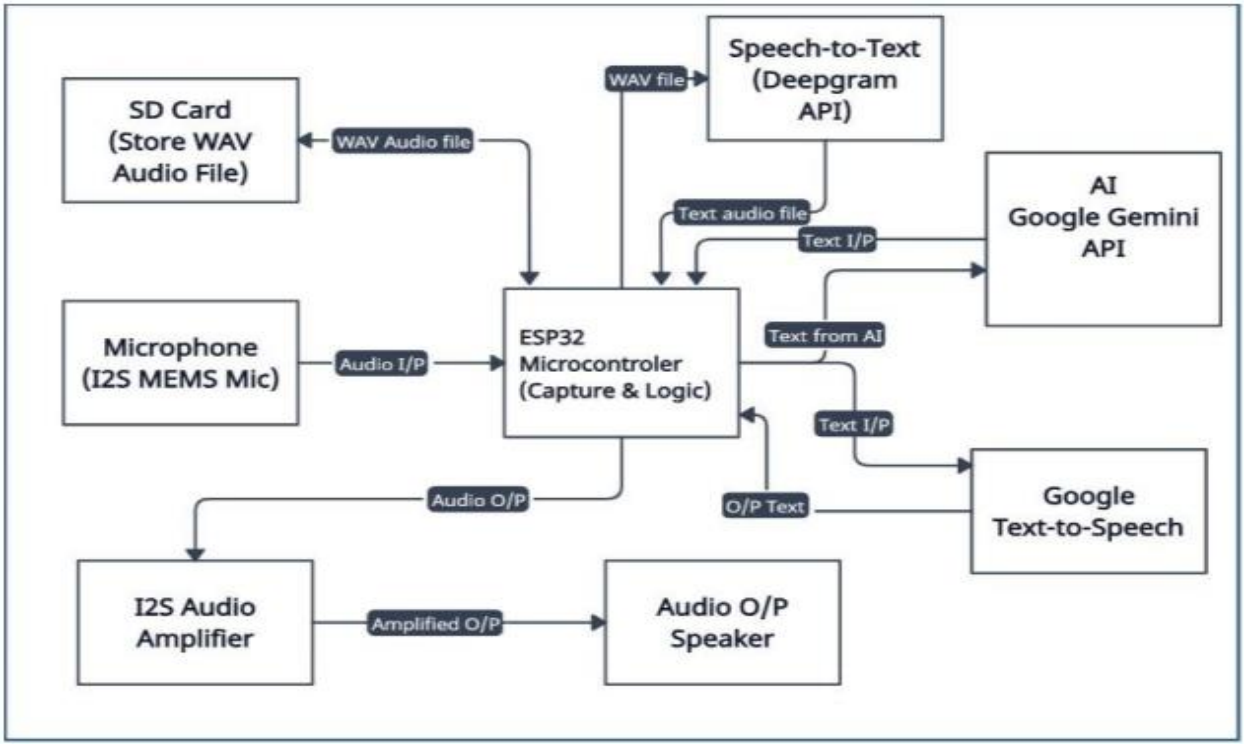


**FIGURE 1.** Block Diagram

The block diagram in Fig. 1 illustrates a portable AI voice assistant system that utilizes an ESP32 microcontroller and Gemini AI for real-time voice recognition and interaction with IoT devices. Below is a description of each block:

1. **Microphone (I2S MEMS Mic)**
   - Captures voice input from the user.
   - Converts the analog sound waves into a digital I2S (Inter-IC Sound) signal.
   - Sends the digital audio signal to the ESP32 microcontroller.
2. **ESP32 Microcontroller (Capture & Logic)**
   - Acts as the central processing unit, handling audio input, conversion, AI processing, and audio output.
   - Captures the audio signal from the I2S MEMS microphone.
   - Converts the WAV audio file and processes it using the Speech-to-Text API (Deepgram API).
   - Sends the transcribed text to the Google Gemini AI API for generating intelligent responses.
   - Sends the AI-generated text to the Google Text-to-Speech API to create a voice response.
   - Manages the storage and retrieval of WAV audio files on the SD card.
   - Outputs the processed audio signals to the I2S Audio Amplifier and speaker.
3. **SD Card (Store WAV Audio File)**
   - Stores recorded audio in WAV format.
   - Provides storage for both processed and raw audio files.
4. **Speech-to-Text (Deepgram API)**
   - Converts the recorded WAV audio files into text.
   - Sends the transcribed text to the Google Gemini AI API for generating a response.

5. **AI - Google Gemini API**
    - Processes the transcribed text and generates a context-aware response.
    - Sends the response text back to the ESP32 microcontroller.
6. **Google Text-to-Speech API**
    - Converts the AI-generated text into speech/audio output.
    - Sends the generated speech back to the ESP32 microcontroller for playback through the speaker.
7. **I2S Audio Amplifier**
    - Receives the audio signal from the ESP32 microcontroller.
    - Amplifies the audio output to enhance speaker performance.
8. **Audio Output Speaker**
    - Produces the final AI-generated speech output for the user.
    - Plays back the AI's response in real-time.

## 5. WORKING

The AI voice assistant system, based on ESP32 and Google Gemini AI, facilitates real-time voice recognition, processing, and response. This system is designed to capture voice input, convert it into text, process the text using Natural Language Processing (NLP), and generate a corresponding audio reply. Below is a step-by-step breakdown of how the system functions.

The process starts with the I2S MEMS microphone, which records the user's voice and converts it into a digital audio signal. Unlike conventional analog microphones, the I2S (Inter-IC Sound) microphone offers a direct digital output, ensuring higher accuracy and minimizing noise interference. The ESP32 microcontroller, acting as the central unit, receives this raw audio signal and processes it. The voice data is then saved as a WAV file on an SD card, providing temporary storage before further processing. Storing the audio in WAV format preserves the original quality, allowing the speech-to-text conversion system to accurately recognize words. At this stage, pre-processing methods like noise filtering and signal enhancement can be applied to improve voice clarity and enhance speech recognition precision.

Once the voice data is stored, it is sent to the Deepgram Speech-to-Text API, which specializes in fast, AI-driven voice transcription. The ESP32 transmits the recorded WAV file over the network to Deepgram’s cloud-based speech recognition service. The API employs advanced deep learning models to convert the spoken words into structured text. This step is essential for making the AI assistant understand the user’s commands in a readable format, facilitating the next processing stage.

After the speech is transcribed, the text is sent back to the ESP32, which serves as a bridge between the speech recognition system and the AI model. With the text data available, the system proceeds to perform Natural Language Processing (NLP) using Google Gemini AI, which interprets the text and generates an appropriate response.

Upon receiving the transcribed text, the ESP32 forwards the data to the Google Gemini AI API, where NLP is used to analyze and comprehend the user's request. This enables the AI to determine the intent, extract keywords, and generate an appropriate response. If the command involves controlling an IoT device, like turning on a fan or adjusting the temperature, the ESP32 processes the request and activates the necessary IoT functions. If a verbal response is required, such as answering a question, the Gemini AI generates a text-based reply, which will be converted to speech in the next phase.

The generated text response is sent to the Google Text-to-Speech (TTS) API, which converts it into a natural-sounding audio file. The resulting audio data is sent back to the ESP32 for playback. To ensure clear, amplified sound, the audio signal is processed through the MAX98357 I2S Audio Amplifier, which enhances the sound quality before it is played through a speaker. This completes the entire voice interaction cycle, and the system resets itself, awaiting the next user command. The ESP32 continuously listens for new voice inputs, enabling real-time and seamless communication between the user and the AI assistant.

## 6. HARDWARE MODULES

The proposed AI voice assistant system is powered by the ESP32 microcontroller, chosen for its dual-core architecture, integrated Wi-Fi and Bluetooth connectivity, and low power consumption. These features make the ESP32 ideal for IoT applications, facilitating smooth communication with smart devices while handling basic AI tasks. For more complex processing, the system offloads tasks to cloud-based services like Gemini AI when necessary. To capture

voice inputs, the system employs a high-sensitivity microphone or microphone array, ensuring accurate speech recognition even in noisy settings. Advanced algorithms for noise reduction and beamforming further enhance the clarity of the voice inputs.

For real-time processing, the system incorporates audio codecs that convert analog voice signals into digital data, ensuring efficient processing of voice information. It also includes a speaker module for providing audio feedback, enabling the assistant to respond to user commands and manage voice-based interactions. Designed with portability in mind, the system uses low-power components and features a rechargeable battery with power management capabilities, ensuring long-lasting battery life and continuous operation.

The system also supports various connectivity modules to interact with different IoT devices. It uses Wi-Fi for cloud communication and Bluetooth or Zigbee for local smart home control, ensuring compatibility with a wide array of smart devices like lights, thermostats, security cameras, and other automation systems. Additionally, the assistant may offer touch-sensitive buttons or LED indicators to provide alternative user interfaces, offering visual feedback or a manual control option when voice recognition is not preferred.

For enhanced security, the hardware integrates secure communication protocols and on-chip encryption, safeguarding voice data from unauthorized access. The overall design emphasizes compactness, energy efficiency, and durability, making the device suitable for both stationary and portable use while ensuring a seamless user experience within smart home and IoT environments.

The AI voice assistant system is constructed using a blend of embedded computing, audio processing, communication, and power management components. Below is a detailed list of the essential hardware components integrated into the system.


**FIGURE 2.** ESP32

The ESP32 is a powerful and versatile microcontroller developed by Espressif Systems. It is a successor to the ESP8266 and has gained popularity due to its combination of low power consumption, integrated Wi-Fi and Bluetooth capabilities, and high performance. The ESP32 is designed for a wide range of applications, including IoT (Internet of Things) projects, wearables, home automation, and voice assistants. It supports a variety of programming languages and development platforms, making it highly accessible to developers and hobbyists alike. With its dual-core processor, rich peripheral options, and ample memory, the ESP32 has become one of the go-to microcontrollers for embedded systems.


**FIGURE 3.** I2S Amplifier

I2S (Inter-IC Sound) amplifiers are essential components in digital audio systems, particularly those involving microcontrollers or digital signal processors (DSPs). They receive audio signals in a digital format and convert them into an amplified analog output that can drive speakers or other audio playback systems. The I2S interface is used for high-quality digital audio transmission, offering a reliable and noise-resistant method for transmitting sound data. I2S amplifiers are commonly used in a variety of applications, including consumer electronics, embedded systems, automotive audio, and voice assistant devices. The main advantage of I2S amplifiers is that they simplify audio system design by eliminating the need for analog-to-digital (ADC) or digital-to-analog (DAC) conversion stages. Since the I2S protocol is already a digital format, the amplifier can directly process the incoming audio signal without additional conversion, preserving audio fidelity and reducing noise interference.

**FIGURE 4.** I2S Microphone Module

I2S (Inter-IC Sound) microphones are specialized audio capture devices that convert sound into a digital signal using the I2S protocol. Unlike traditional analog microphones, which output an analog signal that needs to be converted into a digital format, I2S microphones provide a digital signal directly. This makes them particularly useful in modern embedded systems, where digital processing of audio data is required for tasks such as speech recognition, sound analysis, and communication with other digital systems. The I2S protocol is designed for efficient and noise-resistant

audio data transmission, making I2S microphones ideal for applications like voice assistants, IoT devices, and automotive systems.


**FIGURE 5.** DC-DC Boost Converter module

A DC-DC boost converter is a type of power electronics circuit that is used to convert a lower DC voltage into a higher DC voltage. It is widely used in a variety of applications where the input voltage is lower than the required output voltage, such as battery-powered devices, renewable energy systems, and power supplies for electronic equipment. The main function of a boost converter is to step up the voltage, while maintaining or improving the efficiency of the power conversion process. The boost converter operates on the principle of inductive energy storage. When the switch in the circuit (typically a transistor) is closed, energy is stored in an inductor. When the switch opens, the energy stored in the inductor is transferred to the output through a diode, causing the output voltage to increase. Boost converters are often favored for their simplicity and efficiency, especially in systems where space, weight, and power efficiency are critical factors.

**FIGURE 6.** Push Button

A push button is a simple mechanical switch used to complete or interrupt an electrical circuit when pressed. It is one of the most fundamental components in electronics and is commonly used in various applications, including consumer electronics, industrial machinery, automotive controls, and home automation. Push buttons are designed to momentarily change the state of a circuit—either from open to closed or vice versa— while being pressed. Once released, they typically return to their default state due to an internal spring mechanism. They are available in different shapes, sizes, and configurations, depending on the application.

## 7. RESULTS

The ESP32-based voice assistant project integrates multiple hardware and software components to create a compact, efficient speech processing system. Centered around the ESP32 microcontroller, the device handles voice recording via an external microphone, saves audio using the I2S protocol to an SD card, and processes speech with APIs like

Deepgram for transcription and Google's Gemini for natural language understanding. Noise suppression, button debouncing, and error handling ensure clean input and accurate results. The system also uses an RGB LED to indicate states such as recording, processing, and playback, offering intuitive user feedback without a display. Once the user query is transcribed, it's sent to Gemini for response generation, which is then converted to speech using either Google or OpenAI's TTS services. Audio output is managed through an I2S amplifier, providing high-quality playback, and a repeat function allows users to replay responses. The project emphasizes efficient memory use and process management to maintain reliable performance on the ESP32. With its blend of voice recognition, AI integration, and user-friendly interaction, this voice assistant showcases the potential of microcontroller-based intelligent systems in real-world applications.

(A) Recording

(B) Processing

(C) Playback

**FIGURE 7.** Output

## 8. CONCLUSION

The proposed ESP32-based voice assistant exemplifies the seamless fusion of AI-powered cloud services with low-cost, compact embedded hardware. Utilizing real-time speech recognition, natural language understanding, and text-to-speech synthesis, the system delivers a responsive and interactive voice interface that competes with commercial solutions yet remains budget-friendly and accessible for developers and broader user communities. Its efficient resource management ensures stable performance despite hardware constraints, while features like LED indicators and real-time battery monitoring enhance user engagement and system reliability. This proposal stands out as a scalable, portable, and practical solution for a wide range of voice-controlled applications. the design demonstrates significant potential for real-world deployment. Its low power consumption, ease of use, and adaptability position it as a strong foundation for future development in smart assistant technologies, encouraging further innovation and customization in the field of voice-enabled embedded systems.

## REFERENCES

[1]. B. Sudharsan, S. P. Kumar, and R. Dhakshinamurthy, "AI Vision: Smart Speaker Design and Implementation with Object Detection Custom Skill and Advanced Voice Interaction Capability," 2019 International Conference on Advanced Computing (ICoAC), DOI: 10.1109/ICoAC48765.2019.247125.

[2]. Senthilkumar Meyyappan and N. Selvamuthukumaran, "Network Selection in Heterogeneous Wireless Systems using GRA Method", *Journal on Electronic and Automation Engineering*, Vol. 4(1), March 2025, pp. 127-132.

[3]. R. Kathiresh, V.M. Ramprasath and M. Senthil Kumar, *"A Systematic Approach for Design of Compressed Test Data in SOC"*, CiiT International Journal of Software Engineering and Technology. (Vol.4, No.4, Issue: April 2012, Print: ISSN 0974 – 9748 & Online: ISSN 0974 – 9632).

[4]. R. Kathiresh, P. Kalidass and M. Senthil Kumar, *"A Study of Energy Efficient Embedded Processor and its Reuse"*, International Journal of Modern Engineering Research (IJMER). (Vol.2, Issue 3, May-June 2012, PP 830-833, ISSN: 2249-6645).

[5]. Senthilkumar Meyyappan, A. Bharath Naik, A. Uma Sai and Ch. Keerthi, "Improving Weather Forecasting Accuracy Using Machine Learning", *Journal on Electronic and Automation Engineering,* Vol. 2(4), December 2023, pp. 9-18.

[6]. M. Senthil Kumar and M. Gopinath, "An Efficient Polynomial Pool-Based Scheme for Distributed Heterogeneous WSNs", *International Journal of Modern Engineering Research (IJMER)*. (Vol.3, Issue 6, Nov-Dec.2013, PP 3328-3335, ISSN: 2249-6645).

[7]. Artificial Intelligence Based AI Voice Assistant, 2020 IEEE World Symposium, DOI: 10.1109/WorldS450073.2020.9210344.

[8]. M. Senthil Kumar and L. Praveen, "An Assuring Approach for Tree-Based Routing Topology in WSNs", *International Journal of Emerging Trends in Engineering and Development (IJETED).* (Issue 3, Vol.6, November 2013, ISSN: 2249 – 6149).

[9]. Raspberry Pi Based Voice-Operated Personal Assistant (Neobot), 2019 International Conference on Electronics, Communication and Aerospace Technology (ICECA), DOI: 10.1109/ICECA.2019.8821892.

[10]. Senthilkumar Meyyappan, K. Susmitha, K. Vaishnavi and M. Sai Rao, "Condition Based Monitoring and Maintenance System for Underground Metro Stations", *Journal on Electronic and Automation Engineering*, Vol. 4(1), March 2025, pp. 175-182.

[11]. C.I. Vimalarani and M. Senthil Kumar, "Energy Efficient PCP Protocol for k-Coverage in Sensor Networks", *IEEE International Conference on Computational Intelligence and Computing Research, IEEE Proceedings, 2010.*

[12]. M. Kavitha, T. Maheshwaran and M. Senthil Kumar, "Secure Routing in MANETs with Key Management", *International Journal on Engineering Technology and Sciences (IJETS).* (Vol.1, Issue 6, October 2014, ISSN (P): 2349 – 3968, ISSN (O): 2349 - 3976).

[13]. M. Senthil Kumar, "Energy Efficient Techniques for Transmission of Data in Wireless Sensor Networks", *Journal of Computing Technologies (JCT)*. (Vol.5, Issue 2, February 2016, ISSN: 2278 – 3814).

[14]. "Hey, Siri", "Ok, Google", "Alexa": Acceptance Relevant Factors of Virtual Voice Assistants, IEEE International Professional Communication Conference, 2019, DOI: 10.1109/ProComm.2019.00025.

[15]. Domain-Specific Intelligent Personal Assistant with Bilingual Voice Command Processing, IEEE TENCON 2018, DOI: 10.1109/TENCON.2018.8650203.

[16]. M. Senthil Kumar and Ashish Chaturvedi, "Energy-Efficient Coverage and Prolongs for Network Lifetime of WSN using MCP*", European Journal of Scientific Research (EJSR).* (Vol.95, No.2, January 2013, ISSN: 1450 – 216X / 1450 – 202X).

[17]. On Using ReSpeaker Mic Array 2.0 for Speech Processing Algorithms, 2020 International Symposium on Electronics and Telecommunications (ISETC), DOI: 10.1109/ISETC50328.2020.9301144.

[18]. K. Arutselvan, C. Sridhathan and M. Senthil Kumar *"Unlocking Mobile Devices using Improved Face Recognition and Eye Blinking Technique",* International Journal of Applied Engineering Research (IJAER). (Vol.13, No.24, 2018, PP 16907-16909, ISSN: 0973-4562).

[19]. C. Sridhathan, M. Senthil Kumar and G. Rajesh Krishna, *"Smart and Secure Railway Transport System",* Journal of Computing Technologies (JCT). (Vol.7, Issue 8, August 2018, ISSN: 2278 – 3814).

[20]. J. Chu, G. Zhao, Z. Fu, W. Zhu, and L. Song, "Design and Implementation of Education Companion Robot for Primary Education," 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 1327–1331, DOI: 10.1109/ICCC47050.2019.9064253.

[21]. M. Senthil Kumar and C. Sridhathan, "Impact of Mobility on the Routine of Enhanced – DSDV Protocol in Mobile Ad-hoc Networks", *International Journal of Applied Engineering Research (IJAER)*. (Vol.13, No.14, 2018, PP 11674-11679, ISSN: 0973-4562).

[22]. H. Wang, Z. Ye, and J. Chen, "A Speech Enhancement System for Automotive Speech Recognition with a Hybrid Voice Activity Detection Method," IWAENC 2018, pp. 456–460, DOI: 10.1109/IWAENC.2018.8521410.

[23]. B. Sudharsan, M. Chockalingam, "A Microphone Array and Voice Algorithm Based Smart Hearing Aid," International Journal of Computer Applications, vol. 178, no. 41, pp. 1–6, 2019, DOI: 10.5120/ijca2019919295.

[24]. "Understanding the Adoption of Voice-Activated Personal Assistants," International Journal of E-Services and Mobile Applications (IJESMA), vol. 3, no. 9, pp. 1–21, 2017, DOI: 10.4018/IJESMA.2017070101.

[25]. Senthilkumar Meyyappan, G. Lava Kumar, G. Niharika and G. Chakradhar, "Cellular Network Signal Strength Analyser", *Journal on Electronic and Automation Engineering*, Vol. 4(1), March 2025, pp. 165-174.

[26]. M. Senthil Kumar and Ashish Chaturvedi, "A Novel Enhanced Coverage Optimization Algorithm for Effectively Solving Energy Optimization Problem in WSN", *Research Journal of Applied Sciences, Engineering and Technology (RJASET).* (Issue 4, Vol.7, January 2014, ISSN: 2040 – 7459 & e-ISSN: 2040 – 7467).

[27]. Senthilkumar Meyyappan, Kalyan Kasturi, G. Vijaya Lakshmi, J. Srinija Reddy and K. Grace Sampoorna, "Improvement of LEACH Protocol for Enhancing Features of WSN", *Journal on Electronic and Automation Engineering*, Vol. 2(4), December 2023, pp. 19-26.

[28]. Amit Gupta, M. Senthil Kumar, M. Raman Kumar and D. Hemanth Kumar, "Deep Learning Technique Used for Tomato and Potato Plant Leaf Disease Classification and Detection", *2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES),* July 2023.

[29]. M. Kavitha, T. Maheshwaran and M. Senthil Kumar, "Ensure Data Transmission in Mobile Ad-Hoc Networks", *International Journal on Engineering Technology and Sciences (IJETS)*. (Vol.2, Issue 4, April 2015, ISSN (P): 2349 – 3968, ISSN (O): 2349 - 3976).

[30]. Amit Gupta, Abha Dargar, Abdul Majid, Shashi Kant Dargar, M. Senthil Kumar and M. Raju, "Markov Chain Model Used in Agricultural Yield Predictions Utilizing on Indian Agriculture", *2023 IEEE World Conference on Applied Intelligence and Computing (AIC),* July 2023.