International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 22, Issue 11 (November 2026), PP 94-98

AI Enabled Voice Interaction System for Audio Visual Feedback and Smart Device Control

Adnan Khan[1], Arjun Sonkeshariya[2], Dewansh Rudrakar[3], Prof. Tanuja Kashyap[4]
[1,2,3]*Student, Department of Electronics & Telecommunication Engineering* [4]*Guide, Department of Electronics & Telecommunication Engineering Bhilai Institute of Technology, Durg, Chhattisgarh, India*
[1]*300102823004,* [2]*300102823013,* [3]*300102823025*

Abstract*—Voice-based human–machine interaction is becoming central to smart-home and Internet of Things (IoT) applications, yet most commercial voice assistants depend on continuous cloud connectivity, add response latency, and offer limited audio-visual feedback or customization for embedded use. This paper presents a low-cost, modular, AI-enabled voice interaction system built around the ESP32-S3 microcontroller. The system captures speech through an INMP441 I^2S digital microphone, processes commands on-board and through a cloud AI service, and delivers audio feedback through a MAX98357A I^2S amplifier alongside visual feedback on an SSD1306 OLED display. Wi-Fi connectivity enables the platform to interpret spoken commands and wirelessly control connected smart devices such as lights, fans, switches, and other IoT-enabled appliances. The proposed architecture combines I^2S, I^2C, and wireless communication protocols into a single extensible platform intended for real-time operation, affordability, and future expansion toward smart-home and educational IoT applications. The Phase-I prototype establishes the complete system architecture and hardware integration, forming the basis for full functional evaluation in the next phase of the project.*

Index Terms*—ESP32-S3, voice interaction, IoT, smart home automation, audio-visual feedback, embedded systems.*

Date of Submission: 01-11-2026 Date of acceptance: 10-11-2026

I. INTRODUCTION

Voice interaction has emerged as one of the most natural modalities for human–machine communication, and it is now a core component of smart-home and IoT ecosystems. Commercial voice assistants such as Amazon Alexa and Google Assistant have popularized hands-free device control, but they are largely closed, cloud-dependent platforms that are difficult to customize for educational or research use and that introduce latency whenever network connectivity is poor or unavailable.

This paper proposes an AI-enabled voice interaction system designed for real-time human–machine communication and smart-device control. The system uses an ESP32-S3 microcontroller as its processing core to capture and interpret voice commands. An INMP441 I^2S digital microphone provides voice input, while a MAX98357A I^2S audio amplifier and an SSD1306 OLED display together provide combined audio and visual feedback to the user. Inter-module communication is handled through the I^2S and I^2C protocols, while Wi-Fi is used both to reach an external AI/cloud service for speech processing and to issue control commands to connected IoT devices.

The resulting platform is intended to be low-cost, modular, and easily expandable, so that it can serve as a foundation for further smart-home and IoT applications rather than a closed, single-purpose device. The remainder of this paper is organized as follows. Section II reviews related work on voice-controlled embedded and IoT systems. Section III identifies the specific problems this work addresses. Section IV describes the proposed methodology and hardware implementation. Section V discusses the results expected from the Phase-I prototype.

Sections VI and VII present the conclusion and future scope, respectively.

II. LITERATURE REVIEW

A survey of twelve recent studies (2022–2026) on ESP32-based and AI-driven voice-controlled smart systems was carried out to identify existing approaches, their capabilities, and their limitations. Table 1 summarizes the key findings from this survey.

Across the reviewed studies, ESP32/ESP32-S3 recurs as the dominant low-cost embedded platform for voice control, and cloud-based speech-to-text and language models are commonly used to interpret natural

language commands. Reported recognition accuracies range from roughly 80% to 90%, and end-to-end response latency of several seconds is common when cloud AI services are involved. Several works point to edge AI processing, multimodal interaction, and privacy protection as open research directions.

III.PROBLEM IDENTIFICATION

Based on the gaps observed across the reviewed literature, the following problems are identified:

1) Existing voice assistants often rely on cloud-based processing, causing response delay and requiring continuous internet connectivity.
2) Many systems focus mainly on voice response, with limited integration of audio-visual feedback alongside smart-device control.
3) Commercial voice assistants can be costly and less customizable for educational and embedded application needs.
4) There is a clear need for a low-cost, modular, and expandable embedded platform that combines voice interaction, AI processing, and IoT device control within a single system.

- ESP32-S3 — main processing unit and Wi-Fi connectivity module.
- INMP441 — I^2S digital MEMS microphone used for voice capture.
- MAX98357A — I^2S class-D audio amplifier driving the output speaker.

Table 1: Summary of reviewed literature on voice-controlled embedded and IoT systems

#	Study (Author, Year)	Key Findings
1	Dai *et al.* — Speech Interaction & Vision Sensing in Smart Homes (IEEE ISCAS)	Integrated speech interaction with vision sensing to improve voice recognition, visual accuracy, and device control.
2	Agrawal *et al.* — ESP32-S3 AI Voice Assistant (IJLRP, 2026)	Combined ESP32-S3, INMP441, Deepgram STT/TTS, and an LLM; achieved 6–8 s end-toend response latency.
3	Chen *et al.* — MIST: Multimodal Speech-Based Tool-Calling Assistants (2026)	Introduced a multimodal IoT benchmark covering speech, contextual reasoning, device states, and API/tool calling.
4	Pusatkar *et al.* — Voice-Activated Laboratory Control System (IJIRT, 2026)	Delivered voice, ofline-AP, and Firebase-based control of 16 laboratory devices using ESP32 and relays.
5	Suryawanshi *et al.* — Voice Assistance Using ESP32 & Conversational AI (2026)	Combined ESP32 with conversational AI for voice interaction and smart-device control.
6	Syamsuri *et al.* — Smart Room Using ESP32 & Google Assistant (2025)	Achieved up to 90% response accuracy in ESP32/Google-Assistant-based laboratory room automation.
7	Lakshmi *et al.* — Innovations in Voice-Controlled Smart Home Technology (IJRASET, 2025)	Demonstrated ESP32-based voice automation; highlighted edge AI, privacy, and multimodal interaction as future directions.
8	Wahid *et al.* — Speech Recognition in IoT Smart Home Automation (2025)	Implemented ESP32-S3, INMP441, and Deepgram, achieving approximately 80–90% recognition accuracy for appliance control.
9	Lin *et al.* — VoiceTalk: No-Code VoiceControlled Smart Home Applications (IEEE OJ-CS, 2025)	Used ASR and LLMs to simplify development and improve command interpretation for voice-controlled IoT applications.
10	Bhaganagare *et al.* — The Vocal Cortex of the Smart Home: A Review (IJARSCT, 2025)	Reviewed ESP32 voice-control architectures, comparing cloud, edge, and hybrid processing approaches.
11	Nayak — Home Automation Using Google Assistant (IJFMR, 2024)	Integrated ESP32, Google Assistant, Blynk, and IFTTT for voice-based appliance control.
12	Torad *et al.* — AI/IoT Voice-Controlled Smart Home System (TELKOMNIKA, 2022)	Developed an AI/NLP-based voice-controlled IoT home automation system for appliance control.

IV.METHODOLOGY

A. System Architecture

Fig. 1 shows the overall system architecture. The input stage captures user speech through the INMP441 I^2S digital microphone. The ESP32-S3 forms the processing core and performs audio data processing, voice recognition, command interpretation, decision making, and IoT communication. Over Wi-Fi, the ESP32-S3 communicates with an external AI/cloud service that performs speech-to-text conversion, AI-based processing, and response generation. Recognized commands and generated responses return to the ESP32-S3, which drives two output paths: an audio path through the MAX98357A I^2S amplifier, and a visual path through the SSD1306 I^2C OLED display. In parallel, the ESP32-S3 uses its Wi-Fi connection to issue control commands to smart devices such as lights, fans, switches, and televisions, enabling voice-based control of the connected environment.

B. Hardware Implementation

The prototype hardware, shown in Fig. 2, integrates the following key components on a breadboard:

- SSD1306 OLED display — I^2C display used for visual feedback.
- Push buttons — manual trigger and mode-select inputs.

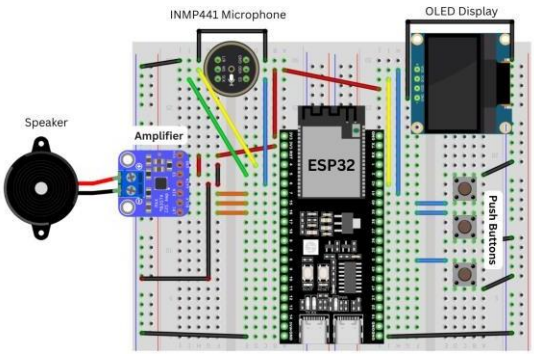
The microphone and amplifier communicate with the ESP32S3 over the I^2S bus, while the OLED display communicates over I^2C, allowing simultaneous audio capture, audio output, and visual output without protocol conflicts.

V.RESULTS

The Phase-I work reported in this paper covers system design, architecture definition, and hardware integration; full functional and quantitative evaluation is planned for Phase-II. Based on the completed design and prototype, the system is expected to demonstrate the following outcomes:

1) Real-time voice command recognition — the system is designed to understand and respond to spoken commands with minimal delay.
2) Combined audio-visual feedback — audio responses are paired with corresponding status information shown on the OLED display.
3) Wireless smart-device control — users are expected to be able to operate connected devices, including lights, fans, and switches, through voice commands.
4) Stable inter-module communication — the I^2S/I^2C/Wi-Fi links are expected to provide smooth data flow between the ESP32-S3 and connected modules.
5) Low-cost, expandable prototype — the final build targets an affordable, easy-to-expand platform suited to further smarthome and IoT development.

Quantitative measurements such as recognition accuracy and end-to-end response latency will be collected during Phase-II testing and compared against the ranges reported in the literature reviewed in Section II.



VI.CONCLUSION

This paper has presented the design of an AI-enabled voice interaction system built on the ESP32-S3 that unifies voice input, AI-assisted processing, audio-visual feedback, and smart-device control within a single low-cost, modular platform. By combining the INMP441 microphone, MAX98357A amplifier, and SSD1306 OLED display with I^2S, I^2C, and Wi-Fi connectivity, the proposed design directly addresses the cloud-dependency, limited-feedback, and cost limitations identified in the reviewed literature. The Phase-I work establishes a complete architecture and a working hardware prototype, providing a solid foundation for full functional testing and evaluation in the next phase of the project.

VII.FUTURE SCOPE

The following directions are identified for extending the proposed system beyond its current Phase-I implementation:

1) Edge AI processing — moving voice recognition on-device to reduce cloud dependency and further cut response latency.

2) Ofline operation mode — supporting core voice-command functions without requiring a continuous internet connection.

3) Expanded device compatibility — extending control to a wider range of IoT and smart-home appliances beyond the current prototype set.

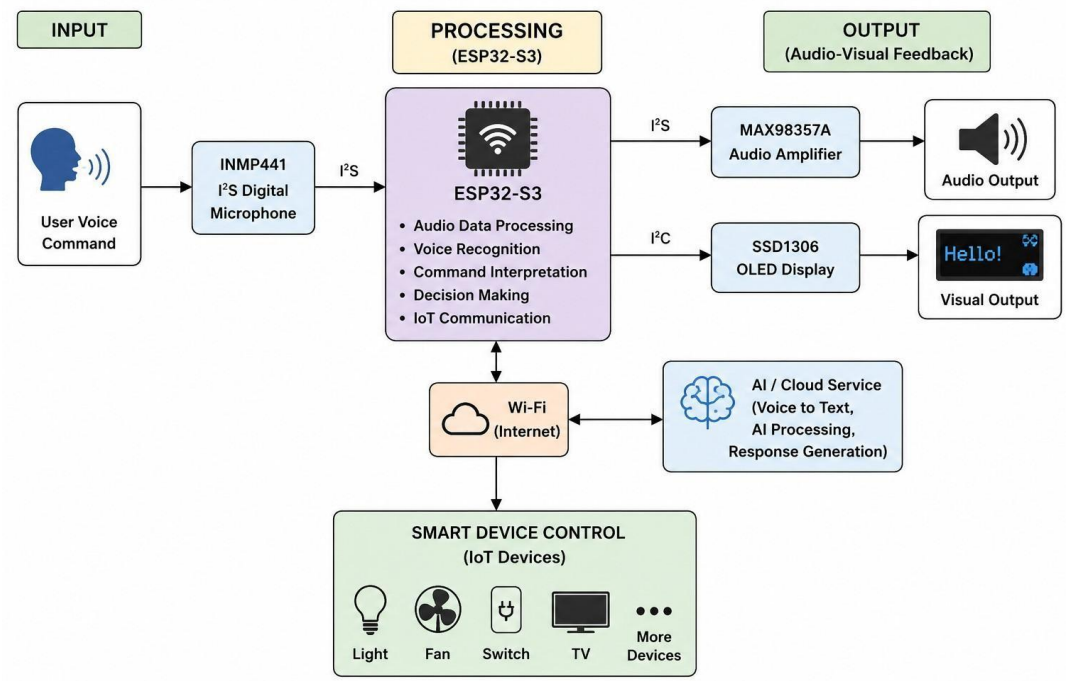


Figure 1: Proposed system architecture showing the input, processing, cloud, output, and smart-device control stages. Figure 2: Breadboard prototype wiring the microphone, amplifier, speaker, OLED display, and push-button inputs to the ESP32-S3.

4) Multilingual and natural-language support — incorporating broader language coverage and more natural conversational interaction.

5) Enhanced security and privacy — adding authentication and data-protection measures for voice data and devicecontrol commands.

6) Mobile application / dashboard integration — providing a companion interface for remote monitoring, configuration, and control.

REFERENCES

[1] J. Dai, J. Qian, Z. Tao, J. Chen, X. Zhu, H. Shao, Z. Gong, and P. Liu, "A System Integrating Speech Interaction and Vision Sensing Applying in Smart Home Scenario," in *Proc. IEEE Int. Conf. on Computer and Communications (ICCC)*, pp. 1–5.

[2] M. A. Torad, B. Bouallegue, and A. M. Ahmed, "A voice controlled smart home automation system using artificial intelligent and internet of things," *TELKOMNIKA*, vol. 20, no. 4, pp. 808–816, 2022.

[3] S. Bhaganagare, S. Chavan, S. Gavali, and V. Godase, "The vocal cortex of the smart home: A review of ESP32based control and automation architectures," *Int. J. Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 5, no. 4, 2025.

[4] Y.-W. Lin, Y.-B. Lin, Y.-F. Wu, and P.-H. Shen, "VoiceTalk: A no-code approach for creating voice-controlled smart home applications," *IEEE Open J. Computer Society*, vol. 6, 2025.

[5] A. N. Wahid, T. Hariono, and N. Yaqin, "Integration of speech recognition in IoT-based smart home automation systems," in *Proc. 3rd Multidiscipline Int. Conf.*, pp. 248–254, Nov. 2025.

[6] M. Chen, X. Zhang, M. Peng, Z. Yu, A. Papangelis, and Y. Jo, "MIST: Multimodal interactive speech-based tool-calling conversational assistants for smart homes," arXiv:2605.06897, May 2026.

[7] S. S. Lakshmi, R. G. Suresh Kumar, A. Nambi, M. Mugilan, E. Vijayakumar, and C. Satheesh, "Innovations in voice-controlled smart home technology: Exploring AI roles and future potential," *Int. J. for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 13, no. 5, pp. 2733–2737, 2025.

[8] S. S. Pusatkar, S. M. Chikate, O. A. Kale, H. N. Dhokane, and A. V. Saywan, "An intelligent voice-activated laboratory control

system powered by ESP32 and Firebase," *Int. J. Innovative Research in Technology (IJIRT)*, vol. 12, no. 11, pp. 2064–2068, 2026.

[9] W. Syamsuri, Y. Wijanarko, and I. Maja, "Smart room based on ESP32 and Google Assistant: Automation solution in electrical engineering laboratory environment," *Iota*, vol. 5, no. 2, pp. 437–443, 2025.

[10] N. Agrawal, P. Gupta, S. Barwe, C. P. Singar, and D. Agrawal, "ESP32-S3 AI voice assistant: A cloudaugmented embedded system for real-time voice interaction," *Int. J. Leading Research Publication (IJLRP)*, vol. 7, no. 5, pp. 1–9, 2026.

[11] R. Nayak, "Home automation system using Google Assistant," *Int. J. for Multidisciplinary Research (IJFMR)*, vol. 6, no. 2, pp. 1–5, 2024.

[12] A. Suryawanshi, S. Jagtap, A. Naikwade, and P. Patange, "Voice assistance using ESP32 & conversational AI," *Open Access Int. J. of Science and Engineering*, vol. 9, pp. 118–127, 2026.