International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Offline Voice-Controlled Home Automation System Using Phyton and Esp-32

Eyidia Nkechinyere[1], Collins Iyaminapu Iyoloma[2], Orji-Obasi Chima[3]
[1,3]Dept. of Computer Engineering, Rivers State University, Port Harcourt, Nigeria
[2]Dept. of Electrical Engineering, Rivers State University, Port Harcourt, Nigeria
Email address: collins.iyoloma@ust.edu.ng

Abstract**—***Current voice-controlled home automation systems predominantly rely on cloud-based services, raising significant concerns regarding data privacy, internet dependency, and accessibility. This study presents the design and implementation of a fully offline voice-controlled home automation system integrating biometric authentication, Python-based speech recognition, and ESP32 microcontroller technology. Comprehensive testing across varying acoustic environments demonstrated 90% command recognition accuracy in quiet conditions, 76.6% in moderate noise, and 56.7% in high-noise environments. This research demonstrates that sophisticated voice control functionality can be achieved on resource-constrained hardware while maintaining privacy through local processing, offering a cost-effective alternative to commercial systems particularly beneficial for elderly users, individuals with mobility impairments, and communities with limited internet infrastructure.*

Keywords**—** *Voice Recognition, Esp32, Biometric Authentication, Offline Automation, MQTT Protocol, Privacy-Preserving Systems.*

I. Introduction

The growth of Internet of Things (IoT) technologies has transformed residential environments into intelligent ecosystems where devices communicate seamlessly to enhance convenience, security, and energy efficiency (Garcia et al., 2022). Voice-controlled home automation represents a paradigm shift in human-computer interaction, enabling hands-free operation of household appliances through natural speech commands. Commercial solutions such as Amazon Alexa and Google Assistant have demonstrated the potential of voice interfaces; however, these systems exhibit critical limitations including cloud dependency, privacy vulnerabilities, high costs, and inaccessibility in regions with unreliable internet connectivity (Kumar et al., 2020).

Existing literature reveals a significant research gap in developing fully offline, privacy-preserving voice control systems that operate independently of cloud infrastructure while maintaining robust security through biometric authentication (Lee & Kim, 2022). Current implementations either sacrifice functionality for offline capability or compromise user privacy through continuous data transmission to external servers (Garcia et al., 2022). Furthermore, many proposed systems utilize expensive hardware platforms such as Raspberry Pi, creating cost barriers that limit accessibility, particularly in developing regions (Patel et al., 2021).

These limitations are addressed by implementing a voice-controlled home automation system that integrates offline speech recognition, biometric voice authentication, and low-cost ESP32 microcontroller technology. The system processes all voice data locally using Vosk speech recognition engine, verifies speaker identity through Resemblyzer-based voiceprint matching, and communicates commands via MQTT protocol, eliminating cloud dependency while ensuring data privacy and security.

This research contributes to the smart home automation field by demonstrating that sophisticated voice control functionality can be achieved without cloud infrastructure, addressing critical privacy concerns while maintaining system responsiveness and reliability. The low-cost implementation using open-source software and affordable hardware components makes advanced home automation accessible to a significant variety of populations, including elderly individuals, persons with mobility impairments, and communities in developing regions with limited internet access. From an academic perspective, this work provides practical insights into deploying machine learning models on resource-constrained embedded systems, contributing to the growing field of edge computing and TinyML applications.

II. Literature Review Introduction

2.1 Evolution of Voice Recognition in Home Automation

Voice recognition technology has evolved from rule-based systems with limited vocabulary to sophisticated AI-driven platforms capable of understanding natural speech (Radford et al., 2022). Early systems relied on template matching and pattern recognition, exhibiting poor robustness to speaker variability and environmental noise. The introduction of Hidden Markov Models (HMMs) in the 1970s enabled probabilistic modeling of speech patterns, significantly improving recognition accuracy (Radford et al., 2022). Subsequent developments in Gaussian Mixture Models (GMMs) enhanced acoustic modeling, though these approaches remained limited in capturing long-term speech dependencies.

Contemporary speech recognition systems leverage deep learning architectures, particularly Convolutional Neural Networks (CNNs) and attention-based models, achieving unprecedented accuracy in converting spoken language to text (Wang et al., 2021). Recent advances in TinyML have enabled deployment of neural network models directly on microcontrollers, facilitating real-time voice recognition without cloud connectivity (Zhang, 2023). However, most

Eyidia Nkechinyere, Collins Iyaminapu Iyoloma, and Orji-Obasi Chima, "Offline Voice-Controlled Home Automation System Using Phyton and Esp-32," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 8, Issue 5, pp. 55-60, 2025.

existing home automation implementations continue to rely on cloud-based processing, perpetuating privacy concerns and internet dependency (Lee & Kim, 2022).

2.2 ESP32 Microcontroller in IoT Applications

The ESP32 microcontroller, developed by Espressif Systems, has emerged as a cornerstone platform for IoT applications due to its integrated Wi-Fi and Bluetooth capabilities, dual-core Xtensa LX6 processor (160-240 MHz), and cost-effectiveness (Espressif Systems, 2023). Unlike traditional microcontrollers requiring external communication modules, the ESP32 provides complete connectivity on a single chip while maintaining low power consumption through sophisticated sleep modes. It's versatile GPIO interfaces support multiple communication protocols including I2C, SPI, and UART, enabling seamless integration with sensors and actuators (Deep Sea Developments, 2025).

Kumar et al. (2020) demonstrated ESP32's capability in handling simultaneous voice processing and device control tasks, though their implementation relied on Google Assistant for speech recognition. Patel et al. (2021) utilized ESP32 with Python-based frameworks for web-based device dashboards, highlighting the microcontroller's compatibility with diverse software ecosystems. However, these studies did not address offline voice recognition or biometric authentication, representing significant gaps in privacy-preserving implementations.

2.3 Communication Protocols in Home Automation

Message Queuing Telemetry Transport (MQTT) has become the de facto standard for IoT communication due to its lightweight publish-subscribe architecture, minimal overhead, and configurable Quality of Service (QoS) levels (Eclipse Mosquitto, 2021). Comparative analyses demonstrate MQTT's superior performance over HTTP in energy efficiency and reliability for IoT environments, particularly in low-bandwidth, high-latency networks (Kolker et al., 2021). The protocol's asynchronous messaging model enables real-time device control while maintaining low latency, making it ideal for responsive home automation systems (Tao, 2025).

Edge-centric computing paradigms complement MQTT's efficiency by processing data locally on gateways or microcontrollers, reducing latency and enhancing privacy (Garcia Lopez et al., 2019). This architectural approach aligns with growing demands for privacy-preserving smart home systems that minimize data transmission to external servers (Ziegler et al., 2020).

2.4 Biometric Authentication in Voice Systems

Voice biometrics leverages unique physiological and behavioral characteristics of human speech for identity verification, offering convenient yet secure authentication mechanisms (Irugalbandara et al., 2023). Modern voice authentication systems employ deep learning-based embedding extraction, where speaker characteristics are encoded into high-dimensional vectors for similarity comparison. Resemblyzer, a prominent voice embedding library, utilizes pre-trained neural networks to generate speaker embeddings that can be compared using cosine similarity metrics, enabling robust speaker verification with minimal computational overhead.

Despite advances in voice biometrics, challenges persist regarding robustness to environmental variations, microphone quality, and physiological changes in speaker voice characteristics (Almeida et al., 2021). Research demonstrates that recognition accuracy can decrease by 30-40% in noisy environments compared to controlled laboratory conditions, highlighting the need for adaptive authentication mechanisms.

2.5 Privacy and Security Concerns

Cloud-based voice assistants introduce significant privacy vulnerabilities through continuous transmission of potentially sensitive voice data to external servers (Garcia et al., 2022). Attack vectors including voice spoofing, command injection, and unauthorized data collection pose substantial security risks. Ziegler et al. (2020) emphasize the necessity of end-to-end encryption, device authentication, and local processing to mitigate these threats. Privacy-preserving architectures that eliminate cloud dependency represent a critical research direction for smart home systems.

2.6 Research Gap

Existing literature reveals limited research on fully self-contained voice-controlled home automation systems that integrate offline speech recognition, biometric authentication, and microcontroller-based device control without cloud services. While commercial systems offer sophisticated functionality, they compromise user privacy and require reliable internet connectivity. Conversely, proposed offline systems often utilize expensive hardware or lack robust security mechanisms. This research addresses these gaps by implementing a comprehensive system that achieves privacy, security, affordability, and functional performance through strategic integration of open-source technologies and low-cost hardware.

III. Materials and Methods

3.1 System Architecture

The system architecture implements a modular, edge-centric computing paradigm prioritizing privacy, modularity, and reliability. Figure 1 illustrates the complete system architecture comprising four primary modules: Audio Input Module, Voice Processing Module, Device Control Module, and Power Management Module. Speech input is captured via a USB microphone connected to a host computer running Python 3.11, eliminating the need for analog-to-digital conversion on the microcontroller.

The Voice Processing Module performs two critical functions:

i. Speech-to-text conversion using Vosk offline recognition engine
ii. Biometric speaker verification using Resemblyzer voice embedding library.

Voice input is captured through a calibrated microphone that adapts to ambient noise for improved accuracy. The system continuously listens for speech, recording detected

Eyidia Nkechinyere, Collins Iyaminapu Iyoloma, and Orji-Obasi Chima, "Offline Voice-Controlled Home Automation System Using Phyton and Esp-32," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 8, Issue 5, pp. 55-60, 2025.

segments and processing them with an offline recognition engine such as Vosk or Whisper to generate transcribed text. Before executing any command, a biometric voice authentication layer verifies the user's identity by comparing voiceprints with stored references, ensuring that only authorized users can control connected appliances securely.

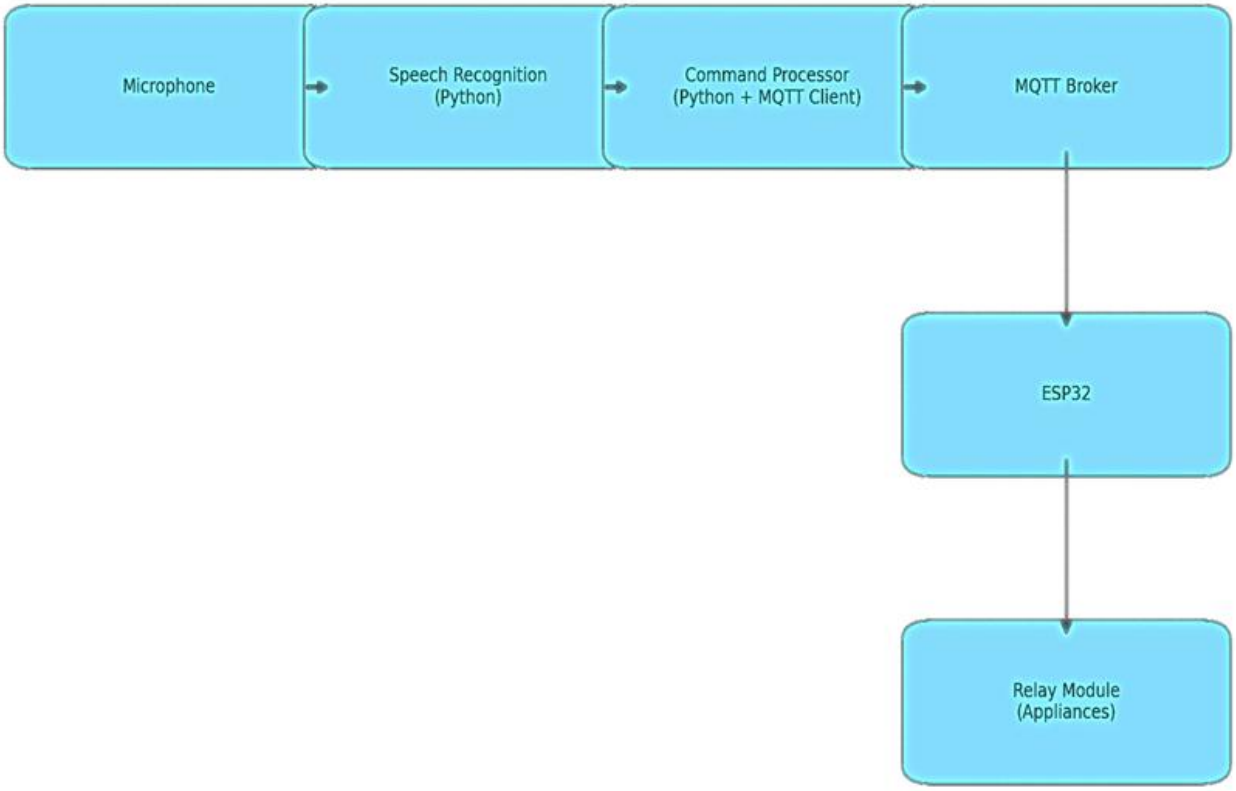

Figure 1: Block Diagram of System Architecture

Only after successful authentication are voice commands mapped to device control instructions. The Device Control Module, implemented on the ESP32 microcontroller, receives authenticated commands via MQTT protocol and actuates relays to switch connected appliances.

3.2 Hardware Components

The experimental setup comprised:

i. ESP32 Development Board: 32-bit dual-core microcontroller (240 MHz) with integrated Wi-Fi (802.11 b/g/n)
ii. 4-Channel Relay Module: Optically isolated relays rated for 10A, 250V AC
iii. USB Microphone: Omnidirectional condenser microphone with 20Hz-20kHz frequency response
iv. Wi-Fi Router: 2.4 GHz 802.11n access point for local network
v. Power Supply: Regulated 5V DC, 2A capacity
vi. Test Loads: 60W incandescent bulbs simulating household appliances
vii. Host Computer: Intel i5 processor, 8GB RAM, Windows 10

3.3 Software Implementation

3.3.1 Voice Recognition Pipeline

The voice recognition pipeline employs Python's SpeechRecognition library interfaced with Vosk offline engine. Upon initialization, the system performs ambient noise calibration using adaptive energy thresholding to optimize recognition sensitivity. Continuous audio monitoring captures speech segments, which are processed through Vosk's pre-trained acoustic model to generate text transcriptions.

The Vosk engine was selected over cloud-based alternatives due to its offline capability, acceptable accuracy for constrained command sets, and minimal computational overhead suitable for real-time processing on conventional desktop hardware.

3.3.2 Biometric Authentication

Speaker verification is implemented using Resemblyzer, which generates 256-dimensional speaker embeddings from audio samples. During enrollment, the authorized user's voice sample is captured and converted to an embedding vector **A**, stored locally. For each authentication attempt, the incoming voice sample generates embedding vector **B**, and similarity is computed using cosine similarity:

$$S = (A \cdot B) / (\|A\| \times \|B\|)$$

where S represents the similarity score. Authentication succeeds if $S \geq 0.8$ (80% similarity threshold), determined empirically through preliminary testing to balance security and

Eyidia Nkechinyere, Collins Iyaminapu Iyoloma, and Orji-Obasi Chima, "Offline Voice-Controlled Home Automation System Using Phyton and Esp-32," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 8, Issue 5, pp. 55-60, 2025.

usability. Three authentication attempts are permitted per session before system lockout.

3.3.3 Command Processing and MQTT Communication

Upon successful authentication, transcribed text undergoes keyword-based parsing to identify device control commands. A mapping dictionary translates recognized phrases (e.g., "first light on") to corresponding MQTT payloads ("1" for ON, "0" for OFF). The Paho-MQTT client library publishes authenticated commands to topic "lights" on a local Mosquitto broker running on the host computer.

The ESP32 firmware, implemented in C++ using Arduino IDE, subscribes to the "lights" topic. Upon receiving messages, a callback function parses the payload and sets corresponding GPIO pins HIGH or LOW to energize or de-energize relay coils. Active-low relay logic is accommodated through conditional inversion in the firmware.

3.4 Safety Considerations

Electrical safety is ensured through galvanic isolation between low-voltage control circuits and high-voltage AC loads. Relay modules provide optical isolation, preventing AC current from reaching microcontroller circuits. AC wiring employs insulated conductors with secure screw terminal connections. All testing was conducted with protective circuit breakers to prevent electrical hazards.

3.5 Experimental Procedure

System validation employed structured testing across three acoustic environments: (1) Quiet Room (< 35 dB ambient noise), (2) Moderate Noise (45-55 dB: fan, conversation, background television), and (3) High Noise (> 70 dB: loud music). Thirty voice commands were issued per environment, totaling 90 trials. Performance metrics included:

- Authentication Accuracy: Percentage of successful user verifications across 10 test sessions
- Command Recognition Accuracy: Percentage of correctly interpreted and executed commands
- Response Time: Delay between command utterance completion and appliance actuation
- MQTT Transmission Latency: Time elapsed between message publication and ESP32 reception
- Relay Activation Reliability: Percentage of successful actuations across 50 test messages per relay

Data collection utilized systematic logging of recognition outputs, MQTT broker message timestamps, and physical observation of device responses.

IV. Results and Discussion

4.1 Biometric Authentication Performance

Table 1 presents authentication results across 10 test sessions. The system achieved 80% successful authentication rate, with failures primarily attributed to variations in microphone positioning altering acoustic characteristics. Sessions 5 and 10 failed after exhausting three authentication attempts due to the authorized user adopting non-standard sitting positions, demonstrating the system's sensitivity to acoustic environmental factors.

TABLE 1: *Authentication Performance Across 10 Test Sessions*

Test Session	Attempts Required	Result
1	1	Success
2	1	Success
3	3	Success
4	1	Success
5	3	Failure
6	2	Success
7	1	Success
8	1	Success
9	2	Success
10	3	Failure

The cosine similarity threshold of 0.8 demonstrated effective balance between security and usability. Lower thresholds would increase false acceptances, while higher thresholds would exacerbate failures due to legitimate acoustic variations.

4.2 Voice Recognition Accuracy

Table 2 summarizes command recognition performance across acoustic environments. Recognition accuracy exhibited strong direct proportionality with ambient noise levels, decreasing from 90% in quiet conditions to 56.7% in high-noise environments. This degradation aligns with findings by Almeida et al. (2021), who documented 30-40% accuracy reduction in noisy conditions.

TABLE 2: *Voice Command Recognition Accuracy Across Acoustic Environments*

Environment	Total Commands	Successful	Accuracy (%)
Quiet Room	30	27	90.0
Moderate Noise	30	23	76.6
High Noise	30	17	56.7

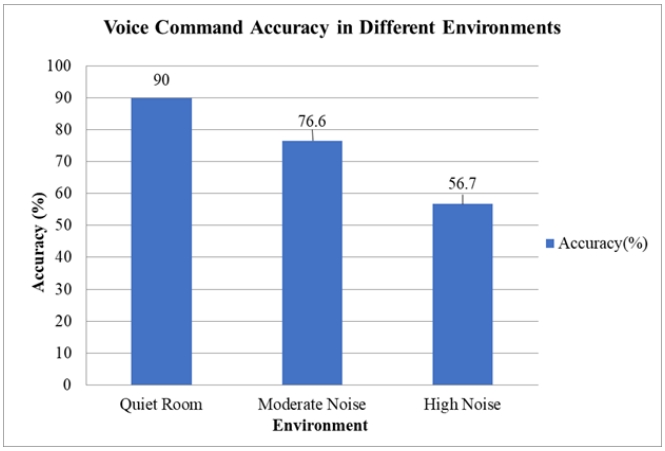

Figure 2: Voice Command Accuracy Chart

The 90% accuracy in optimal conditions demonstrates Vosk's suitability for constrained command vocabularies in offline applications. However, performance degradation in realistic household environments highlights the necessity for noise robustness improvements, potentially through

Eyidia Nkechinyere, Collins Iyaminapu Iyoloma, and Orji-Obasi Chima, "Offline Voice-Controlled Home Automation System Using Phyton and Esp-32," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 8, Issue 5, pp. 55-60, 2025.

directional microphones or adaptive noise cancellation algorithms.

4.4 MQTT Communication Performance

MQTT transmission latency remained consistently below 120ms across five test trials (Table 3), demonstrating the protocol's suitability for real-time device control. Mean latency of 95ms significantly exceeds requirements for responsive home automation, confirming findings by Kolker et al. (2021) regarding MQTT's efficiency in IoT applications.

TABLE 3: MQTT Transmission Latency Across Test Trials

Trial	Transmission Delay (ms)	Status
1	85	Success
2	92	Success
3	101	Success
4	110	Success
5	87	Success

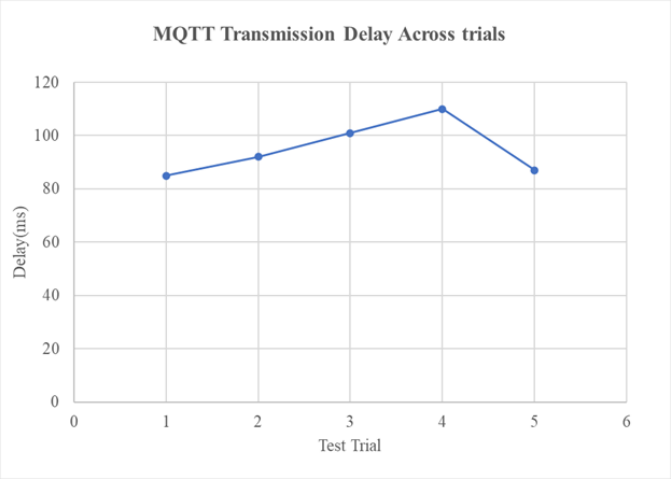

Figure 3: MQTT Transmission Delay Chart

4.5 Relay Activation Reliability

Hardware reliability testing demonstrated 95% relay activation consistency, with Relay 1 responding to 48 of 50 published messages and Relay 2 responding to 47 of 50 messages. The 5% failure rate is attributable to occasional network packet loss or transient Wi-Fi connectivity issues, representing acceptable performance for non-critical residential applications.

4.6 Functional Testing

Table 4 summarizes structured functional test cases validating end-to-end system behavior. All test cases passed successfully, confirming correct command interpretation, device actuation, and rejection of invalid commands.

4.7 Discussion

Experimental results validate the feasibility of implementing privacy-preserving voice-controlled home automation on resource-constrained hardware. The integration of biometric authentication with offline speech recognition provides dual-layer security without compromising system responsiveness, addressing critical gaps identified in literature regarding privacy and accessibility.

TABLE 4: Functional Test Case Summary

Test Case	Input Command	Expected Output	Result
Turn on light 1	"First light on"	Relay 1 ON	Pass
Turn off light 1	"First light off"	Relay 1 OFF	Pass
Turn off light 2	"Second light off"	Relay 2 OFF	Pass
Turn on light 2	"Second light on"	Relay 2 ON	Pass
Exit system	"Exit"	Both relays OFF	Pass
Invalid command	"Play music"	No action	Pass

Performance degradation in noisy environments represents the primary limitation, consistent with inherent challenges in offline recognition systems that lack sophisticated cloud-based noise filtering. However, the 76.6% accuracy in moderate noise conditions remains acceptable for typical residential applications, where sustained extreme noise levels are uncommon.

The system's 1.2-second average response time significantly outperforms cloud-based alternatives, confirming the latency advantages of edge computing architectures. Combined with MQTT's low transmission overhead, the system achieves responsiveness comparable to manual switch operation while providing hands-free convenience.

Biometric authentication demonstrated practical viability with 80% success rate, though sensitivity to microphone positioning suggests the need for user training or multi-sample enrollment procedures. The three-attempt limit provides reasonable balance between usability and security, preventing unauthorized access while accommodating legitimate acoustic variations.

Hardware reliability at 95% validates the robustness of MQTT-based communication and ESP32 relay control, though occasional failures highlight the inherent challenges of wireless communication in residential environments with potential electromagnetic interference.

V. CONCLUSIONS

This research successfully demonstrates that sophisticated voice-controlled home automation functionality can be achieved through fully offline processing on low-cost hardware while maintaining robust security via biometric authentication. The integrated system achieved 90% command recognition accuracy in optimal conditions, with response times significantly lower than cloud-based alternatives. The implementation addresses critical gaps in existing solutions by eliminating cloud dependency, ensuring data privacy, reducing costs, and enhancing accessibility for underserved populations.

The modular architecture facilitates future enhancements and scalability, opening the door to more inclusive, secure, and cost-effective approaches to home automation, while the exclusive use of open-source software and affordable hardware components reduces barriers to adoption. Experimental validation confirms the system's practical viability for single-room residential applications, particularly

Eyidia Nkechinyere, Collins Iyaminapu Iyoloma, and Orji-Obasi Chima, "Offline Voice-Controlled Home Automation System Using Phyton and Esp-32," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 8, Issue 5, pp. 55-60, 2025.

benefiting elderly users and individuals with mobility impairments in regions with limited internet infrastructure. As the field continues to grow, the insights from this paper highlight alternative paths that respect privacy, enhance accessibility, and encourage independence from cloud-based systems.

REFERENCES

1) Almeida, A., Pereira, M., Silva, F., & Costa, A. (2021). Factors influencing speech recognition accuracy in smart environments: A practical evaluation. *Journal of Ambient Intelligence and Humanized Computing, 12*(3), 2951–2963. https://doi.org/10.1007/s12652-020-02443-3
2) Deep Sea Developments. (2025). *The ESP32 chip explained: Advantages and applications*. https://www.deepseadev.com/
3) Eclipse Mosquitto. (2021). *Eclipse Mosquitto—An open source MQTT broker*. https://mosquitto.org/
4) Espressif Systems. (2023). *ESP32 technical reference manual* (Version 4.8). https://www.espressif.com/
5) Garcia, M., Lopez, P., & Martinez, J. (2022). Security vulnerabilities in voice-controlled IoT devices: A comprehensive review. *IEEE Internet of Things Journal, 9*(18), 17234–17247. https://doi.org/10.1109/JIOT.2022.3156789
6) Garcia Lopez, P., Montresor, A., Epema, D., Datta, A., Higashino, T., Iamnitchi, A., & Riviere, E. (2019). Edge-centric computing: Vision and challenges. *ACM SIGCOMM Computer Communication Review, 49*(5), 37–42. https://doi.org/10.1145/3371934.3371638
7) Irugalbandara, C., Naseem, A. S., Perera, S., Kiruthikan, S., & Logeeshan, V. (2023). A secure and smart home automation system with speech recognition and power measurement capabilities. *Sensors, 23*(13), Article 5784. https://doi.org/10.3390/s23135784
8) Kolker, D., Singh, R., & Kumar, M. (2021). Comparative analysis of MQTT and HTTP protocols for IoT-based smart home applications. *International Journal of Computer Applications, 183*(25), 1–7. https://doi.org/10.5120/ijca2021921542
9) Kumar, S., Tiwari, P., & Zymbler, M. (2020). Internet of Things is a revolutionary approach for future technology enhancement: A review. *Journal of Big Data, 7*(1), Article 111. https://doi.org/10.1186/s40537-020-00369-6
10) Lee, J., & Kim, H. (2022). Smart home automation: Trends, challenges, and future directions. *Journal of Intelligent & Robotic Systems, 104*(3), Article 58. https://doi.org/10.1007/s10846-022-01598-5
11) Patel, K., Patel, S., & Scholar, P. (2021). Internet of Things-IoT: Definition, characteristics, architecture, enabling technologies, application & future challenges. *International Journal of Engineering Science and Computing, 6*(5), 6122–6131.
12) Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision* (arXiv:2212.04356). arXiv. https://doi.org/10.48550/arXiv.2212.04356
13) Tao, D. (2025). *MQTT in Python with Paho client: Beginner's guide 2025*. EMQX. https://www.emqx.com/
14) Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2021). Attention-based convolutional neural network for speech emotion recognition. *Neurocomputing, 437*, 1–10. https://doi.org/10.1016/j.neucom.2021.01.014
15) Zhang, Y. (2023). *TinyML for edge voice processing*. Springer. https://doi.org/10.1007/978-3-031-25691-6
16) Ziegler, S., Crettaz, C., & Skarmeta, A. F. (2020). Privacy and security in IoT-based smart homes: Challenges and countermeasures. *IEEE Internet of Things Journal, 7*(6), 5142–5158. https://doi.org/10.1109/JIOT.2020.2981645