3rd Multidiscipline International Conference 2025
November 13 2025, Page. 248 - 254
ISSN 2809-6142

Integration of Speech Recognition in IOT-Based Smart Home Automation Systems

Ahmad Nur Wahid, Tholib Hariono, Nurul Yaqin
Information Systems,Universitas KH. A. Wahab Hasbullah
*Email: anurwahd8866@gmail.com

ABSTRACT

This research presents the design and development of an Internet of Things-based Smart Home system that allows users to control household electronic devices using voice commands. The system utilizes an ESP32-S3 microcontroller integrated with an INMP441 digital microphone and Deepgram speech recognition service through a cloud-based intermediary server. The microphone captures user speech and sends digital audio data to the ESP32-S3 via an I2S interface, which then sends it to the server for real-time conversion into text. The recognized text is sent back to the ESP32-S3 to execute specific commands that activate or deactivate electronic devices through a four-channel relay module. Experimental results show that the system can recognize Indonesian voice commands with an average accuracy of 80 to 90 percent, with a response time of approximately one second and effective performance within a distance of three meters at a noise level of 40 to 55 decibels. This system provides a convenient, efficient, and modern way for users to control household appliances without physical contact or additional applications. It also offers significant benefits for the elderly or people with disabilities. Overall, the proposed design demonstrates a practical implementation of Internet of Things-based voice-controlled Smart Home technology, which improves convenience, accessibility, and energy efficiency in everyday home automation.

Keywords: *Deepgram, ESP32-S3, Internet of Things, Smart Home, Voice Control.*

INTRODUCTION

In the modern era, technology continues to develop rapidly and impacts nearly every aspect of daily life. One of the most significant technological developments today is the Internet of Things (IoT).(Maya, 2021). IoT allows various electronic devices to be connected to the internet so they can communicate, exchange data, and be controlled remotely.(Nizam et al., 2022)The concept of IoT is also becoming increasingly important because it has been widely applied in various aspects of modern life.(Susanto et al., 2022)This concept has given birth to many innovations, one of which is the Smart Home system.(Restu Mukti et al., 2022). Smart home systems have also been proven to increase occupant comfort with device automation that is responsive to user needs.(Kurniawan et al., 2025).

Smart Home technology allows users to automatically control household appliances such as lights, fans, and other electronic devices, providing convenience, energy efficiency, and comfort in daily life. Furthermore, controlling electronic devices via the internet has been proven to simplify daily operations.(Efendi, 2018)However, many existing Smart Home systems still require users to operate them via a smartphone app or a physical switch.(Anam, 2020)This makes the system less practical and user-friendly, especially for the elderly or people with disabilities who may have difficulty using a mobile interface. Therefore, it is necessary to develop simpler and more natural control methods that can be easily used by all types of users. One of the best alternatives is voice-based control.(Karim et al., 2023)With voice commands, users can control devices without having to touch a switch or manually open an app, making interaction with home devices faster, more natural, and more accessible for all ages.

In developing Internet of Things (IoT)-based smart home systems with voice commands, various studies have been conducted using various methods and approaches. Some systems implement control of household devices through instant messaging applications or text-based GSM networks. These systems offer the convenience of remote control, but still have limitations in terms of speed and efficiency, and do not yet support natural voice-based interaction for users with physical disabilities.(Lalu Delsi Samsumar et al., 2023).

Additionally, research has developed a smartphone app-based smart home system without voice recognition support. This system relies solely on the app interface as a control medium, requiring users to press buttons on their phones to operate the device. While this method can work well in stable environments, its accuracy is not affected by voice; its effectiveness is highly dependent on the phone itself. This makes the system less practical, especially for users who have difficulty accessing touchscreens.(Mukin & P, 2023).

In general, previous research shows that most systems are still limited in terms of accuracy, error, and real-time survivability. Therefore, this study aims to address these limitations by developing an IoT-based Smart Home system capable of recognizing voice commands in real time through the integration of cloud-based speech recognition services.(Prakosa et al., 2024).

In this study, a Smart Home system was designed using an ESP32-S3 microcontroller connected to an INMP441 digital microphone.(Andrianto & Intan Saputra, 2020)The microphone captures the user's voice and sends the audio data to the cloud over the internet. Deepgram's speech recognition service is used to convert the voice input into text in real time. The recognized text is then sent back to the ESP32-S3 to control electronic devices via a four-channel relay module.(Professor, 2024)This design eliminates the need for local storage or complex configuration, making the system more efficient and easier to implement.

With this approach, the research is expected to improve ease of access and effectiveness of use for various groups, especially the elderly and people with disabilities who experience difficulties operating devices with buttons or touchscreens. Furthermore, this system has the potential to become the basis for further development in IoT-based smart home technology, particularly in providing more intuitive, energy-efficient, and easily accessible home automation solutions.(Akbar et al., 2017)Thus, this research not only offers technical solutions but also contributes to improving the quality of life and accessibility of technology for all levels of society.

METHOD

This research employed the Research and Development (R&D) method, with stages that refer to a systematic system development process, from problem identification to product testing and refinement (Luthfi et al., 2023). This method was chosen because it is suitable for producing a prototype IoT-based Smart Home system that utilizes voice commands as the primary control. The use of experimental and prototype testing methods has also been widely used in previous research related to IoT devices and microcontroller-based systems.(Salsabila, 2022).

1. Potential and Problems

   At this stage, researchers identified user needs for a smart home system that could be controlled via voice commands. The main problem identified was the limitations of home automation systems, which still require physical interaction via switches or smartphone apps, making them impractical, especially for the elderly or users with disabilities. This situation suggests the potential for developing a more natural and accessible voice-based control system.
2. Data collection

   This phase involved reviewing literature related to smart home technology, IoT, artificial intelligence-based voice recognition services, and supporting devices such as the ESP32-S3 and INMP441 microphone. Additionally, simple interviews with potential users were conducted to determine their feature needs and preferences for using a voice control system at home.
3. Product Design

   In the design phase, researchers designed a Smart Home system architecture that included a hardware block diagram, voice processing flow, ESP32-S3 integration with the Deepgram API speech recognition service through the Replit intermediary server, and a relay module as the main actuator. This design was created so that the system could work in real time, be responsive, and remain easy to use by a wide range of users.

   The system model is built using a central ESP32-S3 module connected to an INMP441 digital microphone as a voice input receiver and a 4-channel relay module to control electronic devices. Indicator LEDs are used to display connection status and device conditions. However, this system has not been designed to support multiple microphones in different rooms. This means that voice commands can only be given effectively from the central microphone location connected to the ESP32-S3.
4. Product Manufacturing

   The product manufacturing stage is carried out by assembling a Smart Home system prototype based on a previously prepared design. The hardware is assembled using the ESP32-S3 as the main module, an INMP441 digital microphone as a sound capture, and a 4-channel relay module as an actuator to control electronic devices such as lights, fans, TVs, and speakers. Each relay channel is equipped with an indicator LED that functions to display the ON/OFF condition of the device, while an additional LED is used to indicate the status of the WebSocket connection. All components are assembled on a work board so that they can function stably and structured.

On the software side, the ESP32-S3 is programmed using the Arduino IDE to manage the WiFi connection, read audio input from the INMP441 via the I2S protocol, convert the signal to 16-bit PCM, and process commands received via WebSocket. Integration with the Deepgram API is implemented through the Replit intermediary server, so that audio data can be sent in real-time to be transcribed into command text.

As part of the implementation, the system workflow was designed and tested to ensure each step was running correctly. The process begins when the user issues a voice command, and the microphone captures the audio and sends it to the ESP32-S3. The audio data is passed to Deepgram for transcription, where it is processed by Replit, and then sent back to the ESP32-S3 as text. The microcontroller matches the text to the programmed instructions and executes the action through a relay module. Indicator LEDs display changes in device status and the system connection status. Once a command is executed, the system automatically returns to a state ready to receive the next command.

5. Product Trial

   Once the prototype was completed, a series of tests were conducted to assess the accuracy, response, and stability of the system.

   a. Black-Box Testing

      Testing involves issuing voice commands such as " *nyalakan kipas* " or " *matikan lampu* " and checking whether the device responds appropriately without reviewing internal processes. This testing ensures the system's core functions are working as designed.

   b. Speech Recognition Accuracy Testing

      Each command is tested repeatedly to obtain a recognition success rate. Accuracy is calculated using the formula:

$$\text{Akurasi } (\%) = \frac{\text{Jumlah pengenalan benar}}{\text{Jumlah total percobaan}} \times 100$$

      The test results show that the system is able to recognize most commands with a high success rate.

   c. Microphone Distance Testing

      Testing was conducted at distances of 0.5 to 4 meters from the INMP441 microphone. The system demonstrated optimal performance at a maximum distance of approximately 3 meters, with accuracy decreasing beyond that limit.

   d. Noise Tolerance Testing

      Testing was conducted in an environment with noise levels of 30–60 dB. The system remained responsive up to 45 dB, but accuracy began to decline significantly when noise levels exceeded 50 dB.

   e. Command Response Testing against Devices

      This test verifies the system's ability to activate and deactivate devices via a relay. Response time is measured from the time the command is spoken to the time the device executes the action. The average response time recorded was approximately 1 second.

6. Product Revision

   Based on initial testing results, adjustments were made to several aspects, such as microphone placement, optimization of the ESP32-S3's processing logic, and refinement of cloud service integration. These revisions aim to improve speech recognition accuracy and system response consistency.

7. Usage Trial

   After the revision process, the system was tested in a home environment to simulate real-world usage conditions. This testing aimed to assess connection stability, usability, and system effectiveness when used by general users of various ages.

8. Final Product

   The final stage produces a prototype IoT-based Smart Home system with voice command control that's ready to use. The product comes with technical documentation and a user guide for easy user operation.

RESULT AND DISCUSSION

A Smart Home system prototype was successfully designed and implemented using an ESP32-S3 microcontroller, an INMP441 digital microphone, and a four-channel relay module. The system is capable of recording audio input through the microphone, sending it to the Deepgram speech recognition service via a cloud-based intermediary server, and then executing commands based on the recognized text. Figure 1 shows the overall architecture and prototype of the developed Smart Home system.

Figure 1. Prototype of IoT-Based Smart Home System with Voice Control

1. System Design Results

   An IoT-based smart home system with voice commands was successfully built using an ESP32-S3 microcontroller, an INMP441 digital microphone, and a 4-channel relay module. Integration with the Deepgram API enables the system to perform real-time voice-to-text conversion through the Replit intermediary server. Implementation results show that all hardware components operate stably and are well-connected.

   Indicator LEDs on each relay channel successfully display the device's status (ON/OFF), while additional LEDs indicate the WebSocket connection status. These features help users monitor system status directly without needing to open additional applications.

   This Smart Home system is designed to be easy to use by various user groups without requiring special technical skills. The system can be operated by children, adults, the elderly, and even users with physical disabilities except for those with hearing impairments, making it an inclusive solution and applicable to various usage conditions. In addition, the system is able to recognize voice commands at various speech rates, whether fast, medium, or slow, as long as the pronunciation is clear and the articulation is correct. This capability makes interactions more natural and practical, so users can control electronic devices at home without obstacles and without the need for physical switches or additional applications.
2. System Test Results

   Testing was conducted to evaluate the accuracy, response time, connection stability, and overall reliability of the system. The test results demonstrated that the system performed as designed and met the research objectives.

   a. Command Recognition Accuracy

   The system also comes with a list of pre-programmed commands so users can give instructions naturally. Here's a list of voice command keywords the system can recognize:

Table 1. List of voice commands

No	Voice Commands	Functions / Actions Performed
1	Nyalakan kipas	Activate relay 3 to turn on the fan
2	Nyalakan tv	Activate relay 2 to turn on the TV
3	Nyalakan lampu	Activate relay 1 to turn on the light
4	Nyalakan speaker	Activate relay 4 to turn on the speaker
5	Hidupkan kipas	Activate relay 3 to turn on the fan
6	Hidupkan tv	Activate relay 2 to turn on the TV

7	Hidupkan lampu	Activate relay 1 to turn on the light
8	Hidupkan speaker	Activate relay 4 to turn on the speaker
9	Matikan kipas	Turn off relay 3 to turn off the fan
10	Matikan tv	Turn off relay 2 to turn off the TV
11	Matikan lampu	Turn off relay 1 to turn off the light
12	Matikan speaker	Turn off relay 4 to turn off the speaker

Each command was tested 15 times. The test results showed an accuracy rate of ≥ 90% in ambient conditions between 40–50 dB and at a distance of approximately 1.5 meters. This demonstrates that the Deepgram API integration is able to consistently identify commands as long as they are clearly spoken and undisturbed by excessive noise.

b. Microphone Distance Testing

Distance testing was conducted at distances of 0.5, 1, 2, 3, and 4 meters. This testing was conducted at noise levels between 40–50 dB, by giving the commands "turn on the light" and "turn off the light" at several different distances. The results showed that speech recognition worked optimally up to a distance of 3 meters, 20 times, in accordance with the system design plan. Above this distance, accuracy began to decline due to the decrease in sound intensity.

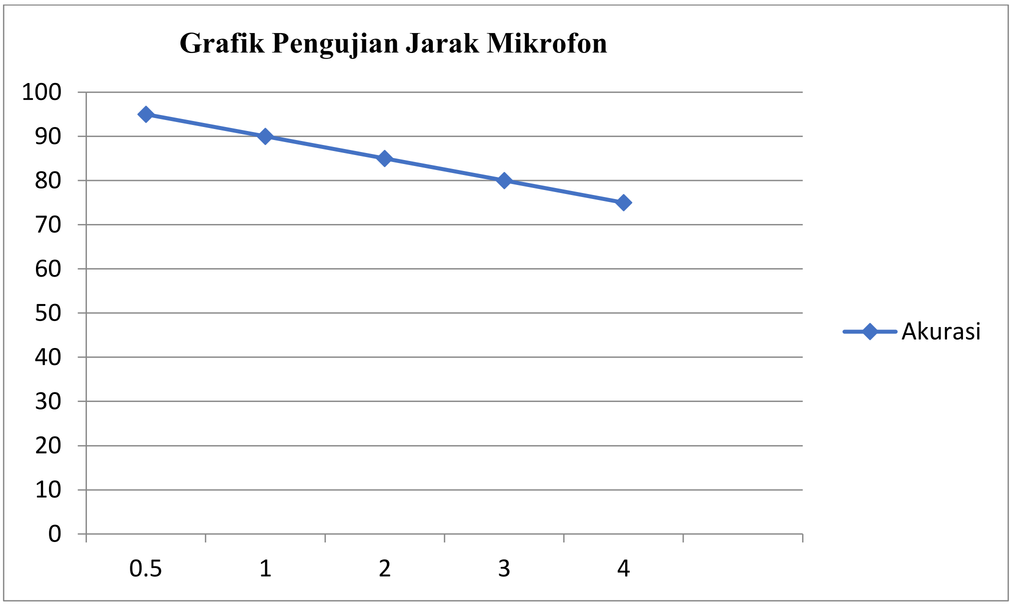

Figure 2. Microphone Distance Testing Chart

c. Noise Tolerance Testing

Testing was carried out at noise intensities:

Table 2. Noise testing

no	Noise level (db)	Number of tests (times)	Correct number (times)	Accuracy (%)
1	30 - 40	20	20	100
2	41 - 50	20	19	95.0
3	51 - 60	20	18	90.0
4	61 - 70	20	11	55.0

This test was conducted at a distance of 1 meter using the commands "*nyalakan lampu*" and "*matikan lampu*" 20 times. Thus, the system remained stable in environments with a maximum noise level of 45 dB.

d. Device Response Testing

Response time is measured from the time the user utters a command until the electronic device is turned on/off. This test was conducted at a noise level between 40-50 dB, at a distance of 0.5 m. The test results show that the system is able to respond to voice commands quickly, with an average response time of about one second from the time the command is uttered until the device is activated. The relay module works consistently without experiencing significant delays, and the entire communication process via WebSocket is stable throughout the test, thus ensuring that the device can be controlled in real-time and without interruption. This indicates that the system is capable of working in real-time as the research objective.

Discussion

The results of the study show that the developed IoT-based Smart Home system with voice commands is able to work accurately, responsively, and stably in a normal usage environment. The speech recognition accuracy rate reaching more than 90% in various test conditions confirms that the integration between ESP32-S3, INMP441 microphone, Deepgram service, and Replit intermediary server provides consistent performance. These results support the theory that modern speech recognition technology is able to improve the efficiency of human-device interaction through a more natural communication model and does not rely on physical input, as stated by(Lalu Delsi Samsumar et al., 2023)that GSM or text message based control systems still have limitations in terms of speed and flexibility compared to voice based systems.

When compared with research(Mukin & P, 2023)Using a smartphone app as a control medium, this study's results show an improvement in user experience because the system eliminates the need for interaction with visual devices or app menus. The study also found that app usage remains highly dependent on device and environmental conditions, and its accuracy decreases as noise levels increase. Conversely, the system developed in this study maintains performance down to noise levels of around 45 dB, making it more adaptable for use in domestic environments.

Distance testing also showed that the system remained accurate up to three meters from the microphone, consistent with the theory that I2S-based digital audio capture devices have good sensitivity in rooms with low echo levels. This finding aligns with research.(Devitra & Purbaningtiyas, 2022)which proves that the relay module can function optimally when paired with a microcontroller capable of real-time communication, thus strengthening the concept that voice control can be used as the main interface in home automation.

Furthermore, the discussion in this study shows that the system is not only technically successful but also has significant social implications. With the ability to understand various speech rates and be usable by children, seniors, and people with disabilities, this system supports the concept of inclusive technology. This enriches modern Smart Home theory, which emphasizes accessibility and ease of use as key components of a successful home automation system. This research reinforces the idea that IoT technology aims not only to improve energy efficiency and comfort but also to enhance users' quality of life by providing more humane and accessible interactions.

Theoretically, the results of this study integrate the concepts of the Internet of Things, artificial intelligence in speech recognition, and the design of a microcontroller-based interactive system. This contribution can be the basis for the development of new models of voice-based Smart Homes that are more independent, more responsive, and easier to implement. With the increasing maturity of cloud-based STT (Speech-to-Text) technology, this research opens up opportunities for system modifications in the next stage, such as multi-room integration with multiple microphones or the development of special features for users with special needs.

The practical implications of this research are quite broad, particularly in the context of implementing Smart Homes in modern homes, public facilities, or environments requiring simple yet effective automation. The system has been proven to operate in real time with a response time of approximately one second, making it highly suitable for use as an alternative to conventional home controls. Furthermore, this system can form the basis for developing commercial voice-based Smart Home products at a more affordable cost than existing commercial devices.

CONCLUSION

This research has successfully designed and built an Internet of Things (IoT)-based Smart Home system controlled via voice commands using the integration of ESP32-S3, INMP441 microphone, Deepgram API service, and Replit intermediary server. The developed system is able to recognize voice commands in real-time with a high level of accuracy, fast response time, and stable performance under various usage conditions. These findings are in line with the research objectives explained in the introduction, namely to present a home control system that is more practical, responsive, and easy to use by various user groups.

The system's ability to recognize commands at various speech rates, along with its ease of operation without requiring technical expertise, demonstrates its inclusiveness and applicability to children, adults, the elderly, and users with physical disabilities. These results also reinforce previous theory and research on the effectiveness of voice-based control as a more natural and efficient solution for home automation.

Although the system has functioned well, this research still has limitations, particularly the microphone's sensitivity to high environmental noise and a maximum coverage distance of approximately three meters. Furthermore, the system only supports a single microphone point, requiring commands to be issued from a single room. Based on this, further research can be directed at developing a multi-microphone system to support multi-room use, improving the noise filtration algorithm to improve performance in noisy conditions, and offline speech recognition capabilities.

Thus, this system not only provides a practical contribution in increasing the comfort and efficiency of using electronic devices at home, but also opens up opportunities for the development of smarter, more inclusive, and affordable Smart Home technology in the future.

REFERENCES

Akbar, R. I., Purnama, D. G., & Salsabila, A. (2017). *Pengembangan Model SmartHome berbasis IoT*. 1–8.

Anam, K. (2020). Smart Home Pengendali Lampu Rumah Berbasis SMS Gateway dan Arduino Menggunakan Smartphone Android. *Jurnal Ilmiah Informatika*, *5*(2), 122–132. https://doi.org/10.35316/jimi.v5i2.945

Andrianto, H., & Intan Saputra, G. (2020). Smart Home System Berbasis IoT dan SMS Smart Home System Based on IoT and SMS. *Telka*, *6*(1), 40–48.

Devitra, A., & Purbaningtiyas, R. (2022). Prototype Smart Home System Menggunakan Voice Control Pada Perangkat Iot. *Just IT : Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, *13*(1), 53–59.

Efendi, Y. (2018). Internet Of Things (Iot) Sistem Pengendalian Lampu Menggunakan Raspberry Pi Berbasis Mobile. *Jurnal Ilmiah Ilmu Komputer*, *4*(2), 21–27. https://doi.org/10.35329/jiik.v4i2.41

Karim, D. A., Anwar, N., Informatika, T., Ilmu, F., Universitas, K., & Unggul, E. (2023). *SMART HOME BERBASIS IOT MENGGUNAKAN ARDUINO UNO*. *7*(3), 47–53.

Kurniawan, I., Andriyanto, S., & Dwisaputra, I. (2025). *Aplikasi Smarthome Berbasis IoT Menggunakan Android Studio*. *03*(2).

Lalu Delsi Samsumar, Zaenudin Zaenudin, Ardiyallah Akbar, Emi Suryadi, & Beni Ary Hidayatullah. (2023). Sistem Monitoring dan Kontrol Rumah Pintar Berbasis Internet Of Things Untuk Peningkatan Efisiensi Energi. *Jurnal Teknik Informatika Dan Teknologi Informasi*, *3*(2), 33–50. https://doi.org/10.55606/jutiti.v3i2.3541

Maya, I. R. (2021). Smart Home Technology Based of Internet of Things As an Activity Support in the Home Interior Design in the Digital Era. *Idealog: Ide Dan Dialog Desain Indonesia*, *6*(1), 77. https://doi.org/10.25124/idealog.v6i1.3832

Mukin, Y. D., & P, N. (2023). Simulasi Jaringan Smart Home dengan Sistem Berbasis IoT. *Jurnal Komunikasi, Sains Dan Teknologi*, *2*(1), 159–168. https://doi.org/10.61098/jkst.v2i1.34

Nizam, M. N., Haris Yuana, & Zunita Wulansari. (2022). Mikrokontroler Esp 32 Sebagai Alat Monitoring Pintu Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 767–772. https://doi.org/10.36040/jati.v6i2.5713

Prakosa, S., Nugraha, A., Sunuharjo, L., & Atiq, M. ’. (2024). Komunikasi Arduino I2C, SPI dan UART. *Jurnal Sains Dan Teknologi Informasi*, *2*(4), 80–85.

Proffesor, A. (2024). *Implementation of Real-Time Audio-to-Text Conversion and Processing for Seamless Transformation in Classroom Environment*. *9*(9).

Restu Mukti, A., Mukmin, C., Randa Kasih, E., Palembang Jalan Jenderal Ahmad Yani No, D., Ulu, S. I., & Selatan, S. (2022). Perancangan Smart Home Menggunakan Konsep Internet of Things (IOT) Berbasis Microcontroller. *Jurnal JUPITER*, *14*(2), 516–522.

Salsabila, D. N. (2022). Analisis Sound Level Meter Berbasis Arduino dengan Implementasi Kalman Filter. In *Jurnal Fakultas Sains dan Teknologi*.

Susanto, F., Prasiani, N. K., & Darmawan, P. (2022). Implementasi Internet of Things Dalam Kehidupan Sehari-Hari. *Jurnal Imagine*, *2*(1), 35–40. https://doi.org/10.35886/imagine.v2i1.329