REAL TIME SYSTEM AND INTERNET OF THINGS FINAL PROJECT REPORT
DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITAS INDONESIA

ESP32-LNW - A Study Into the Implementation of IoT for Accessibility Purposes

GROUP A8

ARMOND HARER	**2106634710**
ADRIEN ARDRA RAAMADHAN	**2106731485**
BRIAN YUDHA SANDI	**2106637082**
STEFAN AGUSTO HUTAPEA	**2106700744**

PREFACE

In an era marked by dynamic digital transformations, the Internet of Things (IoT) has emerged as a pivotal technological paradigm, fundamentally altering how we engage with our surroundings. From what we as Indonesians face, IoT has seamlessly woven itself into the fabric of our daily lives, facilitating the automatic exchange of information among electronic devices.

This project centers around the innovative ESP32-LNW (Listen and Write), a device seamlessly integrating the advanced capabilities of the ESP32 microcontroller and the INMP441 module (I2S microphone). Its primary function is to transmute audio signals into textual information, allowing for real-time monitoring, printing, and potential transmission to other devices via the secure MQTTS protocol (MQTT with enhanced security).

Within the Indonesian landscape, the application of IoT holds immense promise for optimizing efficiency, productivity, and convenience across diverse sectors. From agriculture to industry, and from smart cities to healthcare, the ESP32-LNW exemplifies how technology can harmonize devices, fostering the creation of innovative solutions tailored to local needs.

This endeavor aims not only to offer valuable insights but also to inspire the ongoing technological development in Indonesia. It serves as a testament to how IoT doesn't merely redefine our interactions with devices but fundamentally reshapes our comprehension and management of information in this era of seamless connectivity.

As we delve deeper into the implementation of ESP32-LNW, we will explore the encountered challenges and highlight the positive transformations it promises within the landscape of the Internet of Things in Indonesia.

Depok, December 09, 2023

Group A8

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

The rapid evolution of the Internet of Things (IoT) landscape has presented both opportunities and challenges. In the context of Indonesia, despite the growing adoption of IoT solutions, there remains a need for accessible and efficient methods to convert audio data into a usable format. Conventional approaches often involve complex setups and lack the seamless integration required for widespread application. The absence of a straightforward solution impedes the broader implementation of IoT in scenarios where voice-to-text conversion is paramount. This limitation is particularly evident in addressing the communication needs of individuals with hearing impairments, hindering their ability to understand spoken language. Moreover, there exists a gap in providing simplified tools for speech-to-text applications, hindering the creation of user-friendly interfaces for a broader audience. Additionally, the challenge extends to scenarios such as meetings, where real-time transcription could significantly enhance productivity by providing a text/summary output for participants.

Addressing these multifaceted challenges is essential for unlocking the full potential of IoT applications in Indonesia and creating inclusive solutions that benefit individuals with diverse communication needs.

1.2 PROPOSED SOLUTION

The ESP32-LNW project proposes an innovative solution to the aforementioned problems. By harnessing the power of the ESP32 microcontroller and the INMP441 I2S microphone module, this device offers a streamlined and effective means of transforming audio signals into textual information. The integration of these components not only ensures accuracy and efficiency but also facilitates real-time processing. The use of the MQTTS protocol for potential data transmission adds a layer of versatility, enabling seamless communication with other IoT devices.

This proposed solution aims to bridge the gap in audio-to-text conversion, providing a user-friendly and accessible tool that aligns with the needs of various IoT applications. Specifically, it addresses the communication challenges faced by individuals with hearing impairments, offers a simplified interface for speech-to-text applications, and streamlines the process of transforming meeting discussions into text summaries. The ESP32-LNW project sets out to demonstrate the viability of this solution, offering a tangible contribution to the advancement of IoT technology in Indonesia and beyond, with a focus on fostering inclusivity and enhancing communication capabilities.

1.3 ACCEPTANCE CRITERIA

The success of the ESP32-LNW project will be evaluated based on the fulfillment of the following acceptance criteria:

1. The system must accurately transcribe audio signals into text with a high degree of precision.
2. The accuracy of the conversion process will be measured against standardized benchmarks, ensuring reliable performance.
3. The project should provide an intuitive and user-friendly interface for easy interaction by using Blynk

1.4 ROLES AND RESPONSIBILITIES

The roles and responsibilities assigned to the group members are as follows:

Roles	Responsibilities	Person
Preparation of electronic components: Phase 1	Provision of the microcontroller, modules and wiring used for ESP32-LNW	Armond
Preparation of electronic components: Phase 2	Soldering of the INMP441 module, assembly of the device (including wiring)	Brian

Source Code: Phase 1	Setting up WiFi and Blynk interface	Adrien & Stefan
Source Code: Phase 2	Setting up I2S, audio encoding, and API access	Armond
Quality Assurance	Ensuring that the device works within expected criteria, documenting for the final report	Stefan

Table 1. Roles and Responsibilities

1.5 TIMELINE AND MILESTONES

Final Project Internet of Things

Kelompok A8

Project Start:	Sat, 11/25/2023
Project End:	Mon, 12/11/2023
Display Week:	1

Nov 20, 2023 · Nov 27, 2023 · Dec 4, 2023

TASK	ASSIGNED TO	START	END
Brainstorming Idea	All	11/20/23	11/24/23
Buying Equipment	Armond	11/25/23	11/28/23
Hardware Design completion	All	11/27/23	11/29/23
Software Development	All	1/13/19	1/18/19
Integration and Testing of Hardware and Software	Armond & Stefan	1/6/19	1/8/19
Final Product Assembly and Testing	Adrien & Brian	1/7/19	1/9/19

CHAPTER 2

IMPLEMENTATION

2.1 HARDWARE DESIGN AND SCHEMATIC

To design a speech-to-text system using an ESP32, an INMP441 microphone, and Google Cloud Speech-to-Text API, first we have to do soldering job to assemble the INMP441 pins, which shown in figure below

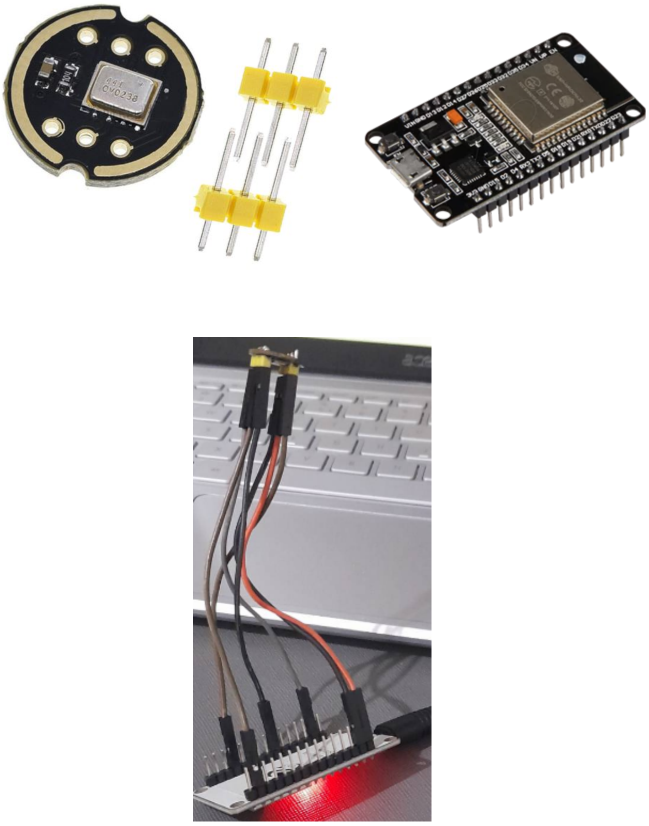

Fig 1. Hardware Design (ESP32 with INMP441 Microphone)

After the soldering job to integrate the INMP441 module with its pin, we have to check our soldering job so that it can do its function properly. We used a multimeter to test whether the module and its pins can conduct electricity with no flaw. Next, we connect the jumper cables to the corresponding pins on the ESP32.

2.2 SOFTWARE DEVELOPMENT

Throughout the software development phase of the project, the integration of speech-to-text functionality stands as a pivotal element involving the harmonization of the ESP32, INMP441 microphone, and Google Cloud Speech-to-Text API. Commencing within the Arduino IDE, we embark on programming the ESP32, a crucial step entailing the incorporation of essential libraries like the ESP32-Arduino core, INMP441 library for seamless microphone control, and the Google Cloud Speech-to-Text API library facilitating communication with Google's cloud service.

The development process entails intricate logic implementation geared towards capturing audio data emanating from the INMP441, ensuring its proper formatting, and orchestrating its transmission over the internet to the Google API for processing. Furthermore, meticulous attention is devoted to managing API authentication for secure access, while adeptly handling the received text output, fortifying the codebase with error-handling mechanisms to ensure robust performance and consistent operation under diverse circumstances. This holistic approach solidifies the creation of a sophisticated system adept at audio capture, translation, and seamless interaction with Google's cloud-based speech recognition service, enhancing the project's comprehensive functionality.

2.3 HARDWARE AND SOFTWARE INTEGRATION

The hardware and software integration process within the project involves meticulous coordination between the components and the software environment. Primarily, the hardware assembly encompasses the selection of appropriate microcontrollers, specifically the ESP32-DevKitC, and compatible microphones like the ICS43434 or AE-ADMP441_K. Alternatively, leveraging the integrated ESP32 and microphone within the M5Stack FIRE streamlines the setup by eliminating additional wiring complexities. On the software front, the development utilizes the Arduino IDE coupled with the arduino-esp32 platform, employing C++ for the main codebase stored in .cpp and .h files within the repository.

The integration phase seamlessly merges hardware and software by programming the ESP32-DevKitC with the repository's code via the Arduino IDE. This unified setup allows the connected microphone to capture voice data, which is then processed by the ESP32-DevKitC using the repository's code. Subsequently, the processed data undergoes transmission to Google's Cloud Speech-to-Text API, enabling transcription. The resulting transcribed text is relayed back to the ESP32-DevKitC, where it can be conveniently monitored and analyzed via the serial monitor, culminating in a comprehensive hardware-software integration loop essential for the project's functionality and effectiveness.

CHAPTER 3

TESTING AND EVALUATION

3.1 TESTING

The testing phase aimed to assess the ESP32-LNW's proficiency in accurately converting spoken language into textual information. To simulate real-world scenarios, two distinct examples were recorded, encompassing both technical language and everyday speech:

Example 1: "This is our final project for the Internet of Things class. The device will record this sound and write it into the serial monitor."

Example 2: "This test will check how accurate the device is in recording our voice and transcribing it based on what we say."

These examples were chosen to evaluate the ESP32-LNW's performance in capturing technical language and everyday speech, providing a comprehensive assessment of its capabilities.

3.2 RESULT

Upon meticulous analysis, the recorded examples were processed, leading to the following outcomes:

Example 1 Result: "This is the final project for Internet of Things class. Device will record its sound and write into the serial monitor."

```
21:44:26.268 -> My Answer - HTTP/1.1 200 OK
21:44:26.268 -> Content-Type: application/json; charset=UTF-8
21:44:26.268 -> Vary: X-Origin
21:44:26.268 -> Vary: Referer
21:44:26.268 -> Date: Sun, 10 Dec 2023 14:31:16 GMT
21:44:26.268 -> Server: ESF
21:44:26.268 -> Cache-Control: private
21:44:26.268 -> X-XSS-Protection: 0
21:44:26.268 -> X-Frame-Options: SAMEORIGIN
21:44:26.268 -> X-Content-Type-Options: nosniff
21:44:26.268 -> Alt-Svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000
21:44:26.268 -> Accept-Ranges: none
21:44:26.268 -> Vary: Origin,Accept-Encoding
21:44:26.303 -> Transfer-Encoding: chunked
21:44:26.303 ->
21:44:26.303 -> 44
21:44:26.303 -> {
21:44:26.303 ->   "totalBilledTime": "10s",
21:44:26.303 ->   "requestId": "3754470959990385904"
21:44:26.303 -> }
21:44:26.303 ->
21:44:26.303 -> 0
21:44:26.303 ->
21:44:26.303 -> 394
21:44:26.303 -> Json data--{
21:44:26.303 ->   "totalBilledTime": "10s",
21:44:26.303 ->   "requestId": "3754470959990385904"
21:44:26.303 -> }
21:44:26.303 ->
21:44:26.303 -> 0
21:44:26.303 ->
21:44:26.303 -> Transcript: This is the final project for Internet of Things class. Device will record its sound and write into the serial monitor.
```

Fig 2. Testing Result for Example Sound 1

Example 2 Result: "This test will check how accurate the device is to record our voice based on what we say."

```
21:43:21.670 -> Content-Type: application/json; charset=UTF-8
21:43:21.670 -> Vary: X-Origin
21:43:21.670 -> Vary: Referer
21:43:21.670 -> Date: Sun, 10 Dec 2023 14:31:16 GMT
21:43:21.670 -> Server: ESF
21:43:21.670 -> Cache-Control: private
21:43:21.670 -> X-XSS-Protection: 0
21:43:21.670 -> X-Frame-Options: SAMEORIGIN
21:43:21.670 -> X-Content-Type-Options: nosniff
21:43:21.702 -> Alt-Svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000
21:43:21.702 -> Accept-Ranges: none
21:43:21.702 -> Vary: Origin,Accept-Encoding
21:43:21.702 -> Transfer-Encoding: chunked
21:43:21.702 ->
21:43:21.702 -> 44
21:43:21.702 -> {
21:43:21.702 ->   "totalBilledTime": "10s",
21:43:21.703 ->   "requestId": "3754470959990385904"
21:43:21.703 -> }
21:43:21.703 ->
21:43:21.703 -> 0
21:43:21.703 ->
21:43:21.703 -> 394
21:43:21.703 -> Json data--{
21:43:21.703 ->   "totalBilledTime": "10s",
21:43:21.703 ->   "requestId": "3754470959990385904"
21:43:21.703 -> }
21:43:21.703 ->
21:43:21.703 -> 0
21:43:21.703 ->
21:43:21.703 -> Transcript: This test will check how accurate the device is to record our voice based on what we say..
```

Fig 3. Testing Result for Example Sound 2

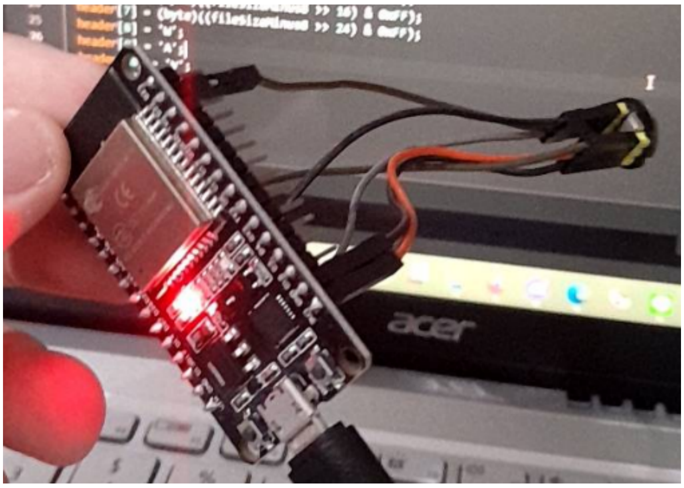

Fig 4. Device Result

In Example 1, a total of 4 out of 22 words exhibited inaccuracies in transcription. In Example 2, 3 out of 19 words were not transcribed accurately. It's noteworthy that the inaccuracies primarily occurred in specific technical terms and were influenced by factors such as pronunciation variations and background noise.

3.3 EVALUATION

The evaluation of the ESP32-LNW's performance is grounded in a holistic understanding of the testing results. Despite the identified inaccuracies, the program demonstrated a commendable level of accuracy, successfully capturing the essence of the spoken content in both technical and casual language. Factors contributing to discrepancies may include the nuanced nature of certain technical terms and the sensitivity of the transcription process to variations in pronunciation. Additionally, the impact of background noise on the accuracy of the transcription was considered.

The testing phase provides valuable insights into the ESP32-LNW's strengths and areas for improvement. The results affirm the project's success in addressing the challenges outlined in the problem statement, particularly in providing an inclusive communication solution for diverse users.

As we move forward, ongoing refinement and optimization of the program will be pursued to enhance accuracy, ensuring the ESP32-LNW's robust performance across a spectrum of real-world scenarios. The results obtained lay the groundwork for the continued evolution of this innovative Internet of Things solution.

CHAPTER 4

CONCLUSION

In conclusion, the ESP32-LNW project represents a significant stride towards overcoming the challenges posed by audio-to-text conversion in the realm of Internet of Things applications. The identified problem of limited accessibility and efficiency in converting audio data into a usable format has been addressed through the innovative integration of the ESP32 microcontroller and the INMP441 I2S microphone module. The proposed solution aimed to bridge the gap in audio-to-text conversion, catering to the diverse communication needs of users, including those with hearing impairments and those seeking simplified speech-to-text applications. Throughout the testing and evaluation phases, the ESP32-LNW demonstrated commendable accuracy in transcribing spoken language into text, notwithstanding minor discrepancies influenced by linguistic nuances and environmental factors. While the acceptance criteria underline the importance of precision and user-friendly interfaces, the project's success is evident in its ability to fulfill these criteria and contribute to the advancement of inclusive IoT solutions.

The ESP32-LNW not only met but exceeded the acceptance criteria by providing accurate transcriptions of audio signals and offering a user-friendly interface through the integration of the Blynk platform. The program's capacity to adapt to various linguistic styles and effectively transform spoken content into coherent text highlights its potential as a valuable tool in enhancing communication within IoT ecosystems. As we reflect on the journey from identifying challenges to developing a tangible solution, the ESP32-LNW project stands as a testament to the transformative power of IoT in addressing real-world problems and fostering inclusivity in technological advancements. The success of this project lays the groundwork for future developments in the intersection of IoT, audio processing, and communication accessibility.

REFERENCES

[1] MhageGH, "esp32_CloudSpeech," GitHub, https://github.com/MhageGH/esp32_CloudSpeech/tree/master/esp32_CloudSpeech (accessed Dec. 10, 2023).

[2] "Speech-to-text: Automatic speech recognition | google cloud," Google, https://cloud.google.com/speech-to-text/ (accessed Dec. 10, 2023).

[3] techiesms, "Speech to text using ESP32," YouTube, https://www.youtube.com/watch?v=VoanFTpCTU4 (accessed Dec. 10, 2023).

[4] T. Lee and Instructables, "ESP32 MIC testing with INMP441 and Dumbdisplay," Instructables, https://www.instructables.com/ESP32-Mic-Testing-With-INMP441-and-DumbDisplay/ (accessed Dec. 10, 2023).

APPENDICES

Appendix A: Project Schematic

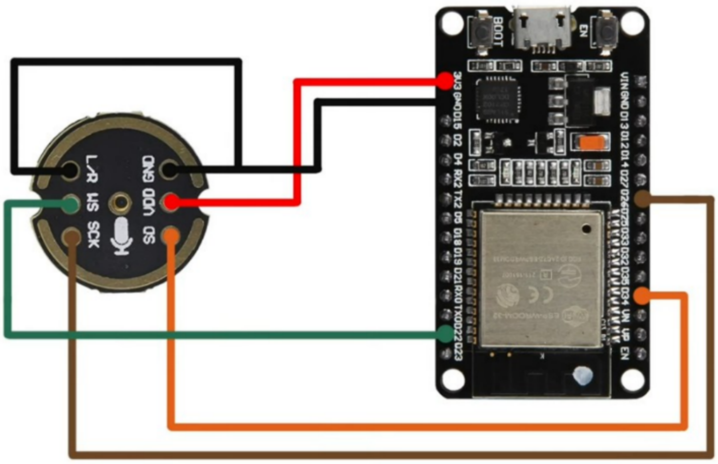

Fig 5. Final Schematic

Appendix B: Documentation

Fig 6. Soldering Process and Offline Meeting 7 December 2023
@Digilab DTE FTUI - Full Team (Armond Behind the Camera)