**ECE 445**

**Spring 2025**

**Senior Design Design Document**

**AI-based Meeting Transcription Device**

Team 45

Ziyang Huang, Gao Gao, Chang Liu

TA: Jiankun Yang

Content

1. **Introduction**
 1.1 Problem and Solution
  1.1.1 Problem
  1.1.2 Solution
 1.2 Visual Aid
 1.3 High-Level Requirements List
2. **Design**
 2.1 Block Diagram
 2.2 Subsystem Overview
  2.2.1 Speech Processing Unit
  2.2.2 Central Controller
  2.2.3 Display Module
  2.2.4 Power Manager
 2.3 Subsystem Requirements & Verification
 2.4 Tolerance Analysis
3. **Cost and Schedule**
 3.1 Cost Analysis
 3.2 Schedule
4. **Ethics and Safety**
 4.1 Ethics
 4.2 Safety
5. **References**
 5.1 Safety & Ethics Documentation
 5.2 Hardware & Software Documentation

**1. Introduction**

* 1. **Problem and solution**

**1.1.1 Problem**

In many academic, professional, and social settings, clear and accurate communication is essential. However, individuals who rely on speech-to-text technology—such as students taking lecture notes, professionals transcribing meetings, and those with hearing impairments—often face challenges when real-time transcription is not readily available. Current solutions for live transcription, such as Zoom’s automatic captions or mobile speech-to-text applications, are largely platform-dependent and require a stable internet connection. This limits their effectiveness in offline environments, such as in-person meetings, classrooms, and research discussions, where accurate note-taking is crucial.

For individuals with hearing impairments, the lack of reliable transcription tools creates significant accessibility barriers. Many venues, including classrooms and workplaces, do not provide real-time captioning, leaving individuals dependent on third-party applications that may not be readily available. Furthermore, for professionals and students who need detailed, distraction-free transcripts, mobile-based solutions can be unreliable due to background noise, latency, or limited processing power.

While commercial AI-powered transcription services exist, they often require cloud connectivity and recurring subscription costs, making them unsuitable for offline or private use cases. Additionally, reliance on cloud services raises data privacy concerns, as sensitive conversations and confidential meetings may be processed externally. Given these challenges, there is a need for an independent, portable transcription device that can function reliably without an internet connection while offering real-time, high-accuracy transcription in a variety of settings.

**1.1.2 Solution**

To address these challenges, we propose the AI-based Meeting Transcription Device, a standalone, portable system that captures, transcribes, and displays spoken text in real time without requiring an internet connection. Unlike existing solutions that rely on cloud-based APIs, our device will leverage on-device AI processing to ensure low-latency transcription, enhanced privacy, and greater accessibility.

Our system consists of the following key components:

1. A microphone module to capture audio input from the speaker.
2. A speech processing unit (Raspberry Pi 5) running the VOSK speech-to-text model, which converts spoken words into text.
3. An STM32 microcontroller, which serves as the central controller, handling user interactions, formatting text output, and managing storage operations.
4. An LCD screen to display real-time transcriptions, ensuring the user can follow the conversation seamlessly.
5. External memory (SD card) for storing past transcriptions, allowing users to review meeting notes.
6. A power system (battery with efficient power management) to provide a portable and reliable power source for extended operation.

By combining real-time speech-to-text processing with a standalone embedded system, our device offers a versatile, privacy-conscious, and highly accessible solution for professionals, students, and individuals with hearing impairments. This system will provide a reliable alternative to cloud-based transcription services, allowing users to operate in any environment without connectivity constraints while maintaining full control over their data

**1.2 Visual Aid**



Figure 1: Visual aid for the AI-based Meeting Transcription Device.

**1.3 High-level Requirements List**

The device must operate for at least 3 hours on a full charge under normal usage conditions, ensuring uninterrupted transcription throughout an average-length meeting or lecture.

The Speech Processing Unit must transcribe spoken sentences with a maximum delay of 2 seconds, ensuring that the device processes and displays each sentence before the next is spoken.

The transcribed text must be displayed clearly and accurately on the LCD screen, with no more than 5% error rate in normal speaking conditions.

The system must store at least 50 transcribed sentences in local memory (SD card), ensuring users can review and retrieve past transcriptions when needed.

**2. Design**

**2.1 Block Diagram**



Figure 2: Block Diagram of AI-based Meeting Transcription Device.

**2.2 Subsystem Overview**

**2.2.1 Speech Processing Unit**

The Speech Processing Unit is responsible for the compute-intensive workload of the transcription system. It consists of a Raspberry Pi 5, a microphone module, and an SD card for storage. The Raspberry Pi 5 is used to run the VOSK Speech-to-Text Model from Alpha Cephei, which processes the audio input and converts it into text.

**Function in the System:** This unit receives real-time audio from the microphone, applies noise reduction processing, and runs the VOSK model to produce textual output. The processed text is then sent to the Central Controller via UART communication.

**Justification:** The Raspberry Pi 5 was chosen due to its quad-core processor and sufficient RAM to efficiently run the AI model while keeping power consumption reasonable.

**Interfaces:**

Microphone Module → Raspberry Pi 5 (Audio Input)

Raspberry Pi 5 → STM32 (UART Communication for Text Data Transmission)

**2.2.2 Central Controller**

The Central Controller acts as the main processing unit responsible for managing data flow, user interactions, and communication between other components. It consists of the STM32F4 microcontroller and flash memory.

**Function in the System:** The STM32F4 receives transcribed text data from the Speech Processing Unit, formats it, and sends it to the Display Module for visualization. It also manages storage operations for saving recent transcriptions in flash memory.

**Justification:** The STM32F4 was selected due to its high-speed SPI communication, low power consumption, and adequate flash storage options.

**Interfaces:**

STM32 ↔ Speech Processing Unit (UART for receiving transcribed text)

STM32 ↔ Display Module (SPI for text rendering)

STM32 ↔ Flash Memory (SPI for saving recent transcriptions)

**2.2.3 Display Module**

The Display Module is responsible for showing real-time transcriptions to the user. It consists of a 2.8-inch TFT LCD screen (ILI9341), which communicates with the STM32 microcontroller via SPI.

**Function in the System:** Displays real-time text output from the central controller.

**Justification:** The ILI9341 LCD was chosen for its high refresh rate, SPI compatibility, and low power consumption.

**Interfaces:**

STM32 ↔ LCD (SPI Communication for text rendering)

**2.2.4 Power Manager**

The Power Manager ensures a stable power supply for all components. It consists of a Li-ion rechargeable battery, a 5V boost converter, and a charger module.

**Function in the System:** Provides regulated power to all subsystems.

Justification: The Li-ion battery offers high energy density and rechargeability, making it ideal for portable applications.

**Interfaces:**

Battery → Voltage Regulator → STM32 (3.3V)

Battery → Boost Converter → Speech Processing Unit (5V)

**2.3 Subsystem Requirements & Verification**

|  |  |  |
| --- | --- | --- |
| **Subsystem**  | **Requirements** | **Verification** |
| **Speech Processing Unit** | Must process speech at a rate of ≥4 words per second to ensure real-time transcription. | Play a recorded lecture at 4 words per second and verify that the system completes transcription before the next sentence begins. |
|  | Must correctly process and output transcribed text to the STM32. | Feed predefined speech samples into the microphone and check if the processed text matches expected output. |
| **Central Controller**  | Must handle SPI communication with the display at ≥1Mb/s for smooth text updates. | Use an oscilloscope to measure SPI communication speed between STM32 and LCD. |
|  | Must store at least 50 transcribed sentences in flash memory for later retrieval. | Transcribe a simulated lecture and verify that at least 50 sentences are correctly stored and accessible. |
| **Display Module** | Must correctly render real-time text output without noticeable lag. | Display test sentences at real-time speed and visually confirm smooth rendering. |
| **Power Management** | The boost converter must output a stable 5V ±0.1V for the Raspberry Pi. | Measure output voltage with a digital multimeter under various loads. |
|  | The battery must support ≥3 hours of operation on a full charge. | Fully charge the device, run continuous transcription, and measure uptime. |

**2.4 Tolerance Analysis**

**2.3.1 Speech Processing Unit**

* According to Alpha Cephei, the VOSK can be run on Raspberry Pi, so we need not worry too much about the board’s performance.
* Assume the lecturer speaks 4 words per second (the normal speaking speed should be 2 – 3 words per second). Each letter consumes 7 bits of resource during transmission, and we assume each word has 8 letters on average (which should normally be 6). Hence, the transmission speed requirement is:

The speed is relatively low compared to the peak transmission capacity of any ports on the board. Therefore, we need not worry too much about it.

* Based on the official data sheet, the typical bare-board active current consumption of Raspberry Pi 5 is 800mA. Hence, the energy capacity of a battery that can support the device to work for 3 hours:

**2.3.2 Central Controller**

* According to the official datasheet of STM32F4, the run mode current is about 45 mA. Therefore, the battery capacity required to let the chip run for 3 hours is:
* Since there is no further change of text data in the central controller, the estimated communication rate required for a regular meeting is also 224 b/s, significantly lower than the max data rate of SPI of 4 Mb/s. Therefore, there should be no issue during the SPI communication between the controller and the display module.
* Assume a lecturer delivers 50 sentences in 3 minutes, and each sentence has 10 words. Based on the mentioned assumption in the previous subsystem, we can estimate the storage capacity required for the flash memory as follows:

A standard flash memory usually has a size of 128 Mb; therefore, the time required to fulfill the flash memory is as follows:

The result is large enough to save the entire meeting’s transcription in the flash memory. Therefore, we should not worry that the device cannot save the latest 50 sentences in the actual situation.

**2.3.3 Display Module**

* Assume the average current consumption of the screen is about 5 mA (standard current consumption should be around 3 mA). Therefore, the battery capacity required to support it to run 3 hours is:

**2.3.3 Power Manager**

* According to previous calculations, the estimated battery capacity required to power the entire system for 3 hours is:

To ensure everything runs safely, the estimated battery capacity should be at least 3000 mAh.

1. **Cost and Schedule**
	1. **Cost Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Component**  | **Quantity** | **Unit Cost ($)** | **Total Cost ($)** |
| Raspberry Pi 5 | 1 | 120 | 120 |
| STM32F401RE | 1 | 5.35 | 5.35 |
| Microphone Module | 1 | 0.67 | 0.67 |
| LCD | 1 | 16.9 | 16.9 |
| SD Card | 1 | 21.99 | 21.99 |
| Power Bank | 1 | 67.99 | 67.99 |
| LDV1117-33 | 3 | 0.64 | 1.92 |
| INA128UA/2K5 | 1 | 8.17 | 8.17 |
| PCM1808PWR | 1 | 1.47 | 1.47 |
| CAP 10UF | 30 | 0.078 | 2.34 |
| CAP 0.1UF | 30 | 0.006 | 0.18 |
| CAP 1UF | 30 | 0.011 | 0.33 |
| CAP 3.3UF | 30 | 0.131 | 3.93 |
| HEADER VERT 2.54MM | 30 | 0.124 | 3.72 |
| RES 2.2K | 30 | 0.011 | 0.33 |
| RES 10K | 30 | 0.025 | 0.75 |
| RES 1K | 30 | 0.012 | 0.36 |
|  |  |  |  |
| Larbor | 3 | 5000 (40\*50\*2.5) | 15000 |
| Total |  |  | 15256.4 |

* 1. **Schedule**

|  |  |
| --- | --- |
| **Week**  | **Assignment - Person** |
| Feb 24 - Mar 2 | Initial Setup - Gao/Ziyang/ChangParts ordering - GaoPCB Design - Gao/Ziyang/Chang |
| Mar 3 - Mar 9 | PCB Design - Gao/Ziyang/ChangDesign Document - ZiyangBreadboard - Gao/Ziyang/Chang |
| Mar 10 - Mar 16 | Breadboard - Gao/Ziyang/Chang |
| Mar 24 - Mar 30 | Sound Processing Unit:PCB Design - Gao/ChangSoftware Development - Ziyang/Chang |
| Mar 31 - April 6 | Monitor Module:PCB Design - Gao/ZiyangSoftware Development - Gao/Chang |
| April 7 - April 13 | Power System:PCB Design - Gao/ChangSoftware Development - Ziyang/Chang |
| April 14 -April 20 | STM32:PCB Design - Gao/ZiyangSoftware Development - Ziyang/Chang |
| April 21- April - 27 | Report - Gao/Ziyang/ChangPresentation Preparation - Gao/Ziyang/Chang |
| April 28 - May 4 | Report - Gao/Ziyang/ChangPresentation Preparation - Gao/Ziyang/Chang |
| May 5 - May 11  | Final Presentation - Gao/Ziyang/Chang |

1. **Ethics and Safety**

**4.1 Ethics**

Our AI-based Meeting Transcription Device raises several ethical considerations, especially related to the collection and use of audio data. In keeping with the principles outlined by the IEEE Code of Ethics and the ACM Code of Ethics, we identify key issues and approaches to preventing unethical outcomes:

**4.1.1 Privacy and Consent (ACM 1.6 & IEEE I.1)**

Because the device is designed to capture and transcribe spoken communications, there is a risk that conversations could be recorded without the knowledge or permission of participants. To address this, we will require user interaction to inform meeting participants that transcription is active.

**4.1.2 Data Security and Confidentiality (ACM 1.6 & 1.7)**

The device could be misused for unauthorized surveillance or eavesdropping. As highlighted by professional codes of ethics, engineers have a responsibility to anticipate how a product might be misused and take steps to limit harmful outcomes. We will use design features (e.g., audible alert whenever the recording is active) to discourage covert use.

**4.2 Safety**

Safety concerns for our device primarily revolve around two areas: (1) electrical and mechanical safety of the hardware (notably the lithium-ion battery and its charging circuitry), and (2) compliance with relevant regulations governing electronic devices.

**4.2.1 Battery and Electrical Safety**

The device relies on a lithium-ion battery and a boost converter to power the Raspberry Pi 5 and other components. Lithium-ion batteries can pose fire or explosion risks if they are damaged, improperly charged, or short-circuited.

**4.2.2 Safe Handling and Disposal**

Users will be instructed on safe handling of the device, including proper storage, transport, and disposal of the lithium-ion battery according to environmental regulations.

**4.2.3 Regulatory Compliance**

Even though our device primarily functions as a stand-alone transcription, we must still review relevant federal regulations (e.g., FCC Part 15 in the United States) for any emissions from the microcontroller or other components.

**5. References**

**5.1 Safety & Ethics Documentation**

1. IEEE Code of Ethics, IEEE, [Online]. Available: <https://www.ieee.org/about/corporate/governance/p7-8.html>.

2. ACM Code of Ethics and Professional Conduct, ACM, [Online]. Available: <https://www.acm.org/code-of-ethics>.

**5.2 Hardware & Software Documentation**

3. Raspberry Pi Documentation, Raspberry Pi Foundation, [Online]. Available: <https://github.com/raspberrypi/documentation/blob/develop/documentation/asciidoc/computers/raspberry-pi/power-supplies.adoc>.

4. STM32F4 Power Modes & Efficiency, STMicroelectronics, Application Note AN4365, [Online]. Available:<https://www.st.com/resource/en/application_note/an4365-using-stm32f4-mcu-power-modes-with-best-dynamic-efficiency-stmicroelectronics.pdf>.

5. ILI9341 Display Controller Datasheet, Adafruit, [Online]. Available: <https://cdn-shop.adafruit.com/datasheets/ILI9341.pdf>.

6. VOSK Speech Recognition Toolkit, AlphaCephei, [Online]. Available: <https://alphacephei.com/vosk/>.