ECE 445

Spring 2025

Senior Design Project Proposal

AI-based Meeting Transcription Device

Team 45

Ziyang Huang, Gao Gao, Chang Liu

TA: Jiankun Yang

1. Introduction

1.1 Problem

During the pandemic, we found Zoom's live transcription very useful, as it helped the audience catch up quickly with the lecturer. In many professional and academic settings, real-time transcription of spoken communication is essential for note-taking. Additionally, individuals with hearing impairments face challenges in following spoken conversations, especially in environments where captions are unavailable.

Existing solutions, such as Zoom's live transcription or mobile speech-to-text apps, require an internet connection and are often tied to specific platforms. To address this, we propose a standalone, portable transcription device that can capture, transcribe, and display spoken text in real time. The device will be helpful since it provides a distraction-free way to record and review conversations without relying on a smartphone or laptop.

1.2 Solution

Our AI-based Meeting Transcription Device will be a portable, battery-powered device that records with a microphone, converts speech into real-time text, and displays it on an LCD screen. The system consists of the following key components:

1. A microphone module to capture audio input.
2. A speech processing unit (Raspberry Pi 5) running the VOSK speech-to-text model to transcribe the captured speech.
3. An STM32 microcontroller, which serves as the central controller for managing user interactions, processing text display, and storing transcriptions.
4. An LCD screen to display transcriptions in real-time.
5. External memory (SD card or NOR flash) for saving transcribed conversations.
6. A power system (battery with efficient power management) to enable portability.

1.3 Visual Aid

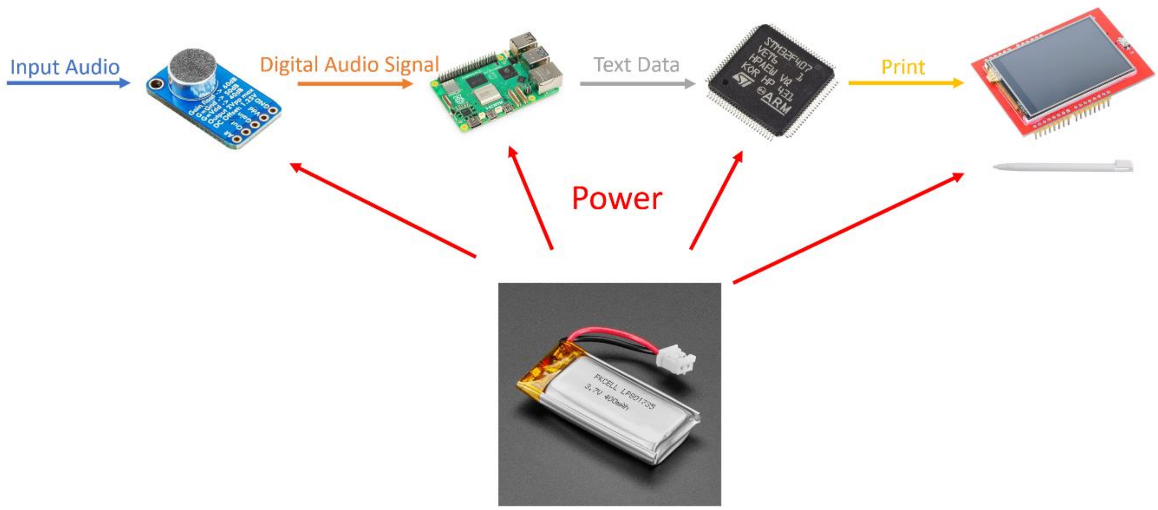

Figure 1: Visual aid for the AI-based Meeting Transcription Device.

1.4 High-level Requirements List

- The battery life should last at least 2-3 hours after charging under normal usage conditions. This requirement ensures that the transcription device can continue working until the end of a meeting.
- The transcribed text should be accurately, clearly, and completely displayed on the screen. Meanwhile, users can access transcribed text in local storage, such as an SD card. This requirement confirms the usability of the device.
- The Speech Processing Unit should finish with one sentence before the lecturer delivers the following sentences. This ensures the device's functionality.

2. Design

2.1 Block Diagram

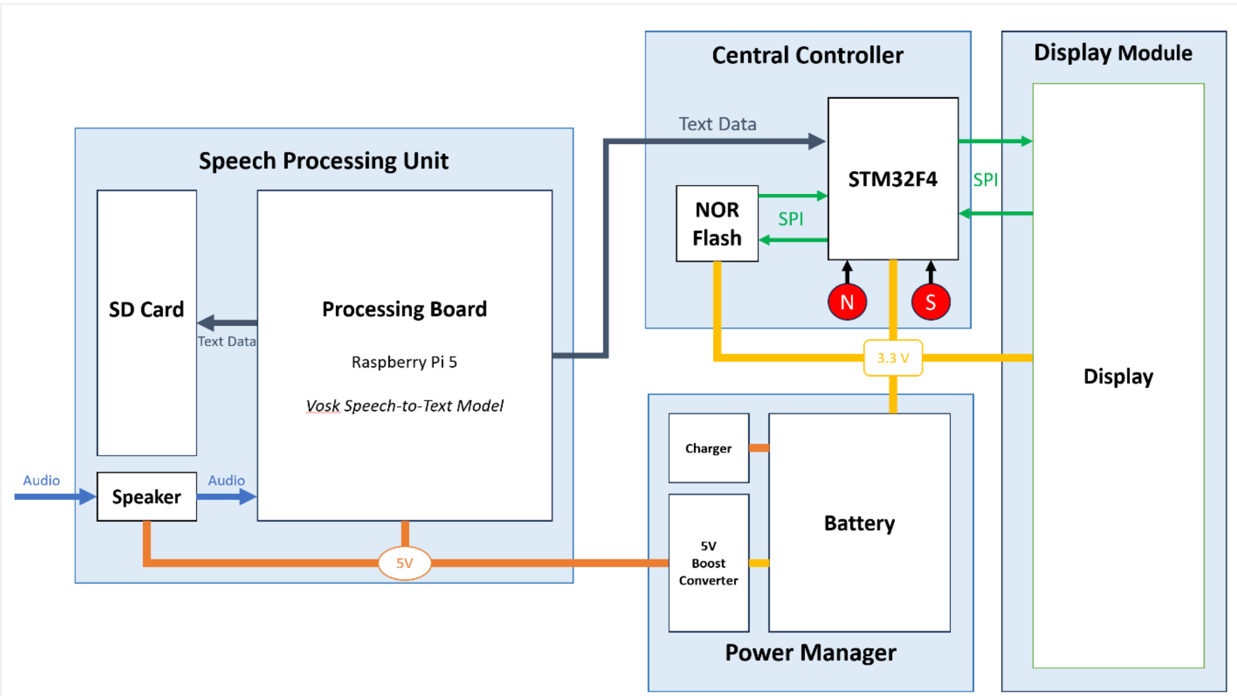

Figure 2: Block Diagram of AI-based Meeting Transcription Device.

2.2 Subsystem Overview

2.2.1 Speech Processing Unit

The speech processing unit handles the compute-intensive work under our design. It comprises a processing board (Raspberry Pi 5), speaker, and storage component (SD card). The Raspberry Pi 5 will be built to a platform that runs the VOSK Speech-to-Text Model from Alpha Cephei. The speech processing unit receives audio from the lecturer and converts it to text; then, it transmits text data to the central controller for the remaining procedures. The unit required 5V to power on, so we designed a voltage booster inside the power manager to satisfy its voltage requirement.

2.2.2 Central Controller

The central controller is the core subsystem that utilizes communication within the device. It consists of STM32F4, flash memory, and buttons. The STM32F4 is primarily used to deliver text data to the display module within the proper time and format. Flash memory optimizes the system's performance when dealing with sparse or dense inputs. Two buttons provide users with a mode-select option: they can decide to save the entire meeting's transcription or navigate to specific recorded transcription.

2.2.3 Display Module

The module is basically a screen component that communicates with the central controller via SPI. Its communication rate should be fast enough to keep up with the lecturer's speed.

2.2.4 Power Manager

The power manager provides reliable and portable power for all components. The subsystem comprises a Li-ion rechargeable battery, a 5V boost converter, and a charger. It delivers energy to the other components and does not communicate data with the different subsystems.

2.3 Subsystem Requirements

2.3.1 Speech Processing Unit

- The processing rate should be as fast or faster than the lecturer's speaking speed to ensure no missed sentences.
- The unit should properly feed input audio signal to the AI model and correctly deliver transcribed text data to the central controller.

2.3.2 Central Controller

- Its SPI communication rate with the display module should be large enough to display text evenly and adequately on the screen (we initially proposed the rate be 1 Mb/s).
- Its SPI communication rate with the flash memory should be fast enough to ensure the recent 50 transcribed sentences are entirely and accurately saved.
- The controller can accurately receive a transcribed text from the speech processing subsystem. We initially proposed that the rate should be one sentence per second.

- The button should correctly control the mode selection. When the save is on, the subsystem starts to save the last 50 transcribed sentences. When the navigation is on, users can read through the saved sentence while the device saves transcription.

2.3.3 Display Module

- The display module should accurately display received text data on the screen.
- The display module should follow commands from the controller. When the user requests to navigate recent transcriptions and the save mode is on, the screen should accurately display the saved transcriptions.

2.3.4 Power Manager

- The subsystem should provide stable and sufficient 3.3V +/- 0.1V to the central controller and display module and 5V +/- 0.1V to the speech processing unit.
- The 5V boost converter should raise the output voltage to 5V.

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:

$$Speed = 4\ \frac{word}{s} \cdot 8\ \frac{letter}{word} \cdot 7\ \frac{bit}{letter} = 224\ b/s$$

  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:

$$E = It = 800mA \cdot 3hr = 2400mAh$$

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:

$$E = It = 45mA \cdot 3hr = 135mAh$$

- 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:

$$Storage = 50\ sentence \cdot 10\ \frac{word}{sentence} \cdot 8\frac{letter}{word} \cdot 7\ \frac{bit}{letter} = \ 2800\ bits$$

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

$$Time = 128\ Mb \div\ 2800\ b\ \ \cdot 3\ \text{min} = 13714\ min$$

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:

$$E = It = 3mA \cdot 3hr = 9mAh$$

2.3.3 Power Manager

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

$$E = 9 + 135 + 2400 = 2544mAh$$

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

3. Ethics and Safety

3.1 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:

3.1.1 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.

3.1.2 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.

3.2 Safety

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

3.2.1 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.

3.2.2 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.

3.2.3 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.

4. Reference

4.1 Safety/Ethics Documentation

- https://www.ieee.org/about/corporate/governance/p7-8.html
- https://www.acm.org/code-of-ethics

4.2 Hardware/Software Documentation

- **Raspberry Pi:** https://github.com/raspberrypi/documentation/blob/develop/documentation/asciidoc/computers/raspberry-pi/power-supplies.adoc
- **STM32F4:** https://www.st.com/resource/en/application_note/an4365-using-stm32f4-mcu-power-modes-with-best-dynamic-efficiency-stmicroelectronics.pdf
- **ILI9341:** https://cdn-shop.adafruit.com/datasheets/ILI9341.pdf
- **VOSK:** https://alphacephei.com/vosk/