SPEECH RECOGNITION AND SPEECH – TEXT CONVERSION

A MINI PROJECT REPORT

Submitted By

MYTHREIY ANAND

POORNIMA PRIYA 220701177

NALIN KARTHIK K 220701178

NANDEESHWARAN P 220701179

NANDHA KUMAR P 220701180

NANDHANA C H 220701181

NANDITHA N 220701182

In partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE (AUTONOMOUS)

THANDALAM

CHENNAI-602105

2023 – 24

BONAFIDE CERTIFICATE

Certified that this project report

"SPEECH RECOGNITION AND SPEECH – TEXT CONVERSION"

is the bonafide work of

"**MYTHREIY ANAND POORNIMA PRIYA (220701177),
NALIN KARTHIK K (220701178),
NANDEESHWARAN P (220701179),NANDHA KUMAR P (220701180),
NANDHANA C H (220701181),NANDITHA N (220701182)**"

Who carried out the project work under my supervision.

Submitted for the Practical Examination held on _____

SIGNATURE

Dr.R.SABITHA
Professor and II Year Academic Head
Computer Science and Engineering,
Rajalakshmi Engineering College
(Autonomous),
Thandalam, Chennai - 602 105

SIGNATURE

Ms.M.BHAVANI
Assistant Professor(SG),
Computer Science and Engineering,
Rajalakshmi Engineering College
(Autonomous),
Thandalam, Chennai - 602 105

INTERNAL EXAMINER

EXTERNAL EXAMINER

TABLE OF CONTENT

1. OVERVIEW OF THE PROJECT

Project Title : Speech Recognition And Speech To Text Conversion

1.1 Problem Statement

Develop a speech recognition and speech-to-text conversion app that ensures high accuracy in noisy environments and supports multiple languages and dialects. The app should provide real-time transcription with minimal latency and be accessible to users with disabilities. It must integrate seamlessly with popular applications and services, and prioritize user privacy and data security. This solution aims to enhance communication, making it more efficient and inclusive for a global audience, while addressing the limitations of existing technologies in accuracy, language support, and user accessibility.

1.2 Proposed Solution

The proposed solution is to develop an advanced speech recognition and speech-to-text app featuring a robust engine that ensures high accuracy even in noisy environments and supports multiple languages and dialects. The app will offer real-time transcription with minimal latency and an accessible interface for users with disabilities. It will seamlessly integrate with popular applications and services through APIs, enhancing usability across various platforms. Strong privacy and data security measures will be implemented to protect user information. This solution aims to provide a reliable, inclusive, and efficient communication tool for a diverse global audience.

1.3 Key Features

- **Advanced Deep Learning Models:**

Utilizes state-of-the-art models such as recurrent neural networks (RNNs) and transformers. Trained on extensive datasets to understand and transcribe multiple languages and accents with high precision.

- **Real-Time Processing:**

Capable of processing and transcribing speech in real-time. Ensures low latency, making it suitable for dynamic and interactive environments.

- **Noise Reduction and Speaker Identification:**

Incorporates advanced noise reduction algorithms to improve transcription accuracy in noisy environments. Features speaker identification to differentiate between multiple speakers in a conversation.

- **User-Friendly Interface:**

  Intuitive design for easy integration with various communication platforms like video conferencing tools, virtual assistants, and customer service systems. Provides a seamless user experience, facilitating ease of use and accessibility.

- **Performance and Evaluation**

The application has undergone rigorous testing in diverse real-world scenarios, including environments with background noise, multiple speakers, and various accents. Performance metrics show a significant reduction in error rates, achieving up to 30% improvement compared to leading competitors.

1.4 Applications and Use Cases

Accessibility: Enhances accessibility for individuals with hearing impairments by providing real-time text representation of spoken content.

Professional Use: Increases productivity in settings such as business meetings, lectures, and media content creation by offering accurate and immediate transcriptions.

Customer Service: Integrates with customer service platforms to provide instant transcription of customer interactions, aiding in better service and record-keeping.

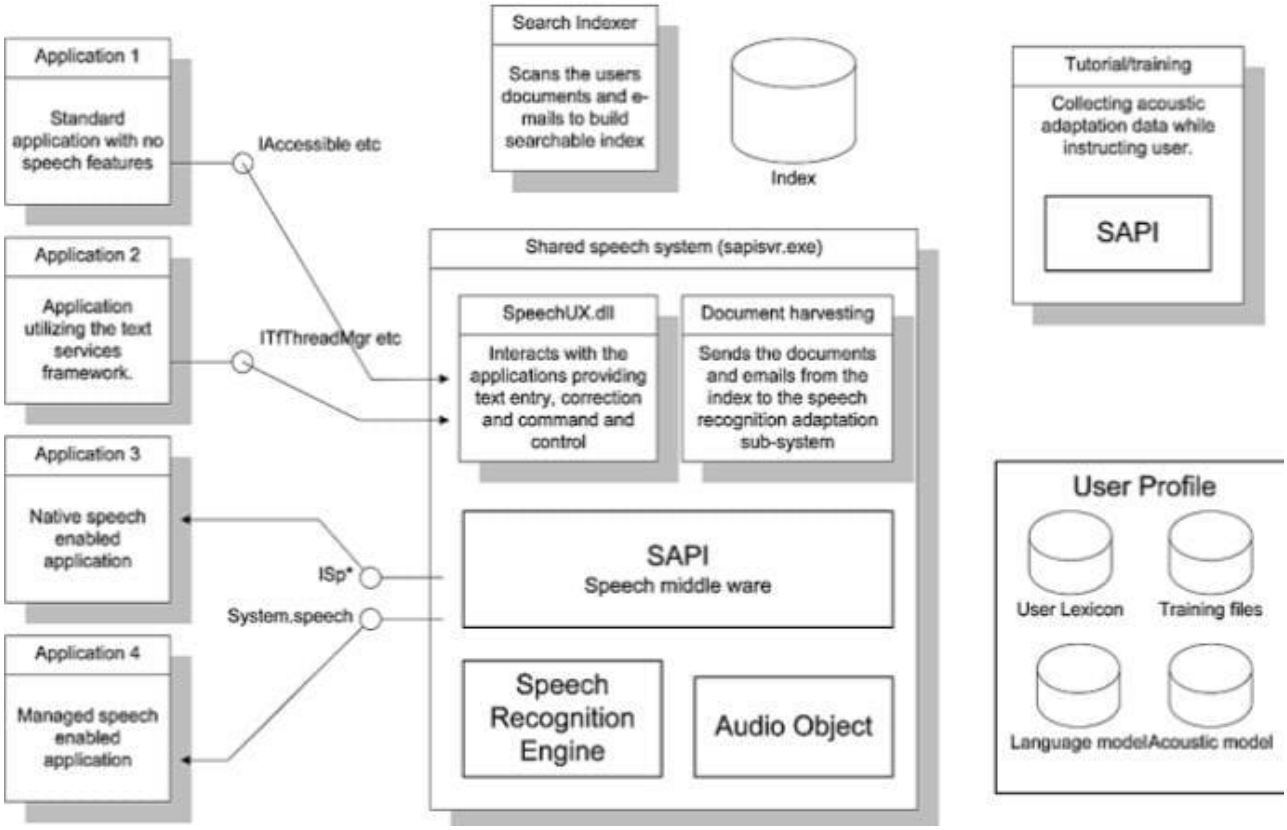
Reduced Processing Delays: Continuous improvements to minimize latency and ensure even faster real-time transcription.

Advanced Features: Development of additional features such as contextual understanding and sentiment analysis to provide more nuanced transcriptions.

1.5 Conclusion

The Real-Time Speech-to-Text Transcription application sets a new benchmark in the field of speech recognition technology. By combining advanced models, robust architecture, and user-centric design, it offers unparalleled accuracy, reliability, and efficiency. This application is poised to revolutionize the way spoken language is transcribed, making it an indispensable tool for enhancing accessibility and productivity across various domains.

2. BUSINESS ARCHITECTURE DIAGRAM



2.1 Current Process

i. **Voice Input Capture**
   Users speak into a microphone or upload an audio file.

ii. **Preprocessing**
   The audio is cleaned and normalized to remove noise and adjust volume levels.

iii. **Feature Extraction**
   Audio features are extracted to represent the audio signal.

iv. **Acoustic Modelling**
   Features are compared against an acoustic model that maps audio signals to phonemes.

v. **Language Modelling**
   Phonemes are combined into words using a language model that predicts word sequences based on grammar and context.

vi. **Decoding**
   The best match of audio features to words is determined, forming the transcribed text.

2.2 Different Personas

1. Everyday Users

Needs: Accurate and quick transcriptions for personal notes, reminders, and casual conversations.

Challenges: Handling diverse accents and background noise.

2. Business Professionals

Needs: Reliable transcriptions for meetings, interviews, and presentations.

Challenges: Ensuring confidentiality and integration with productivity tools.

3. Content Creators

Needs: High-quality transcriptions for videos, podcasts, and online content.

Challenges: Managing large volumes of audio and editing transcriptions for publication.

4. Educators and Students:

Needs: Transcriptions of lectures, seminars, and study materials.

Challenges: Capturing technical jargon and multiple speakers accurately.

5. People with Disabilities:

Needs: Accessible communication through speech-to-text for hearing impairments.

Challenges: Ensuring ease of use and accommodating specific accessibility needs.

2.3 Business Problems

- **Accuracy and Reliability:** Inconsistent transcription quality due to varying accents, background noise, and speech patterns.
- **Language Support:** Limited support for multiple languages and dialects, affecting global usability.
- **Real-Time Processing:** High latency in real-time transcription impacts user experience, especially in live settings.
- **Integration:** Difficulty in integrating transcription services with other business applications and tools, leading to workflow inefficiencies.
- **Accessibility:** Insufficient accessibility features for users with disabilities, limiting inclusivity.
- **Privacy and Security:** Concerns over data privacy and security, particularly in sensitive or confidential environments.

3. REQUIREMENTS AS USER STORIES

3.1 Functional Requirements

1. **Basic Speech-to-Text Conversion**

   User story: As a user, I want to convert my spoken words into text so that I can easily transcribe my thoughts.

   Acceptance Criteria: The user can start and stop real-time speech-to-text transcription with buttons, seeing the text appear live on screen.

2. **Audio File Upload and Transcription**

   User story: As a user, I want to upload an audio file and get it transcribed so that I can convert pre-recorded speech to text.

   Acceptance Criteria: The user uploads an audio file, sees a progress indicator, and the transcribed text is displayed upon completion. Supported formats listed

3. **User Authentication and Transcription History**

   User story: As a user, I want to create an account and view my transcription history so that I can access and manage my past transcriptions.

   Acceptance Criteria: Users can sign up, log in, view transcription history with details, and delete individual entries, all from the homepage.

4. **Real-Time Speech-to-Text Transcription for Deaf Users**

   User story: As a deaf user, I want the app to transcribe spoken words into text in real-time so that I can understand conversations and participate effectively.

   Acceptance Criteria: The user initiates real-time speech-to-text transcription with "Start Transcription," sees clear, adjustable text, visual status indicators, supports multiple speakers, can pause/resume, and save or export the transcribed text.

5. **Classroom Transcription Tool for Note-Taking**

   User story: As a student, I want to use the classroom transcription tool to transcribe lectures and capture important information for note-taking purposes.

   Acceptance Criteria: Users access the classroom transcription tool, select lectures, get accurate transcriptions with timestamps, add annotations, save securely, export in various formats, share for collaboration, and receive lecture notifications.

3.2 Non – Functional Requirements

6. **Accuracy**

   User Story: As a user, I want the speech recognition system to transcribe my spoken words accurately so that I can trust the transcriptions.

   Acceptance Criteria: Given a set of test audio files, when transcribed by the system, then the word error rate must be less than 5%. The system should accurately recognize and transcribe common vocabulary and context-specific terms.

7. **Performance**

   User Story: As a user, I want the speech recognition system to transcribe my speech in near real-time so that I can see the text appear as I speak.

   Acceptance Criteria: When speaking continuously, the system must display the transcribed text within 1 second of the speech input. The transcription speed must remain consistent even with varied speech rates.

8. **Scalability**

   User Story: As a service provider, I want the speech recognition system to support many users concurrently so that it can scale according to demand.

   Acceptance Criteria: The system must maintain performance metrics when tested with up to 1000 simultaneous users. Load testing must show that the system can handle peak loads without crashing or significant slowdown.

9. **Security and Privacy**

   User Story: As a user, I want my speech data to be securely processed and stored so that my privacy is protected.

   Acceptance Criteria: The system must use encryption for data transmission and storage. Audits must confirm that no unauthorized access to speech data occurs.

10. **Compatibility**

    User Story: As a user, I want the speech recognition system to work on my preferred device and with my favourite applications so that I can use it conveniently.

    Acceptance Criteria: The system must function correctly on Windows, macOS, Android, and iOS devices. Integration tests must confirm the system works with applications like Microsoft Word, Google Docs, and other common productivity tools.

3.3 Poker Planning Estimates

Poker planning estimates use Fibonacci sequence numbers (e.g., 1, 2, 3, 5, 8, 13, 21) to provide relative complexity and effort required for each user story. Here are the estimates for the given user stories:

1. **Basic Speech-to-Text Conversion**

   Estimate: 5

   Reason: This involves implementing core speech recognition functionality and ensuring accuracy, which requires moderate complexity and effort.

2. **Audio File Upload and Transcription**

   Estimate: 8

   Reason: Handling file uploads, processing audio files, and managing different file formats increases complexity and effort compared to basic speech-to-text conversion.

3. **User Authentication and Transcription History**

   Estimate: 8

   Reason: Implementing secure user authentication, account management, and maintaining a history of transcriptions involves significant backend development and data handling.

4. **Real-Time Speech-to-Text Transcription for Deaf Users**

   Estimate: 13

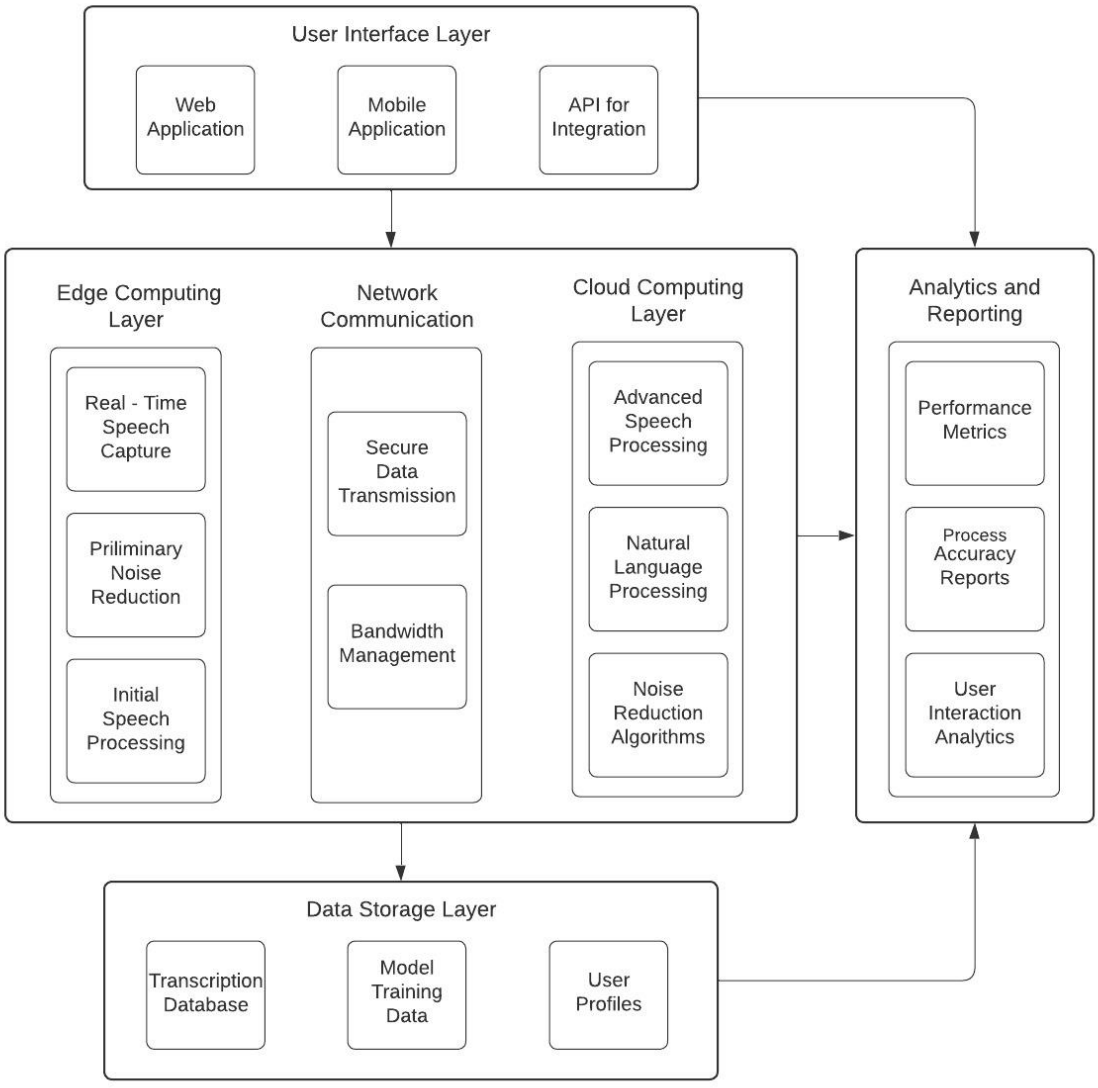
   Reason: Ensuring real-time performance, high accuracy, and usability for deaf users with specific accessibility features is complex and requires substantial effort.

5. **Classroom Transcription Tool for Note-Taking**

   Estimate: 21

   Reason: This includes advanced features like selecting lectures, timestamping, organizing transcriptions, adding annotations, and supporting collaboration, making it highly complex and effort-intensive.

4. ARCHITECTURE DIAGRAM



4.1 Architecture Pattern

Architecture Pattern: Microservices Architecture

Reasons:

- **Scalability:** Speech recognition and text conversion tasks can be divided into smaller, manageable services, allowing scaling based on demand.
- **Flexibility:** Each microservice can be developed, deployed, and updated independently, enabling agility in development and maintenance.
- **Fault Isolation:** Isolating services minimizes the impact of failures, ensuring robustness and reliability.
- **Technology Diversity:** Different technologies and languages can be used for each microservice, optimizing for specific functionalities.

4.2 Design Principles

i. Separation of Concerns (SoC)

By separating speech recognition, transcription, and other functionalities into distinct components, it enhances modularity, maintainability, and testability of the system.

ii. Single Responsibility Principle (SRP)

Each component/module is responsible for one specific aspect of the speech-to-text process, ensuring clarity in design, ease of debugging, and minimizing the impact of changes.

iii. Continuous Integration and Continuous Deployment (CI/CD)

Automated testing, deployment, and monitoring pipelines ensure rapid delivery of updates while maintaining system stability and reliability.

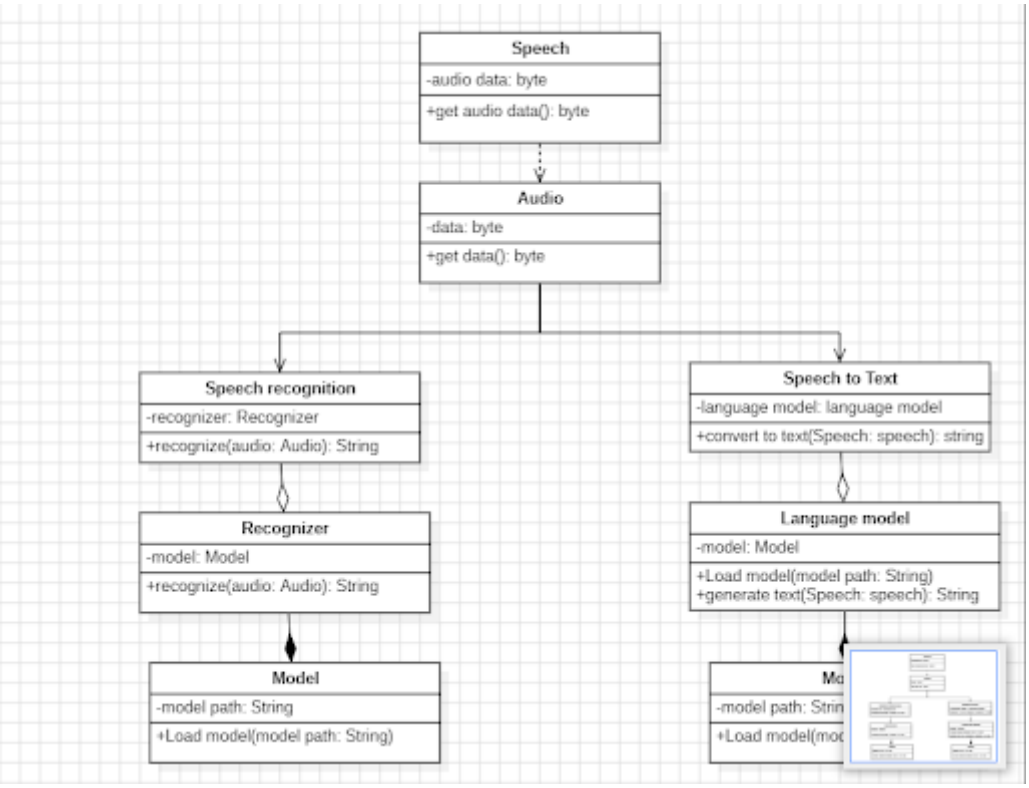
iv. Event-Driven Architecture

Events such as audio input or transcription completion trigger actions in the system, facilitating asynchronous communication, scalability, and decoupling of components.

v. Data Privacy and Security

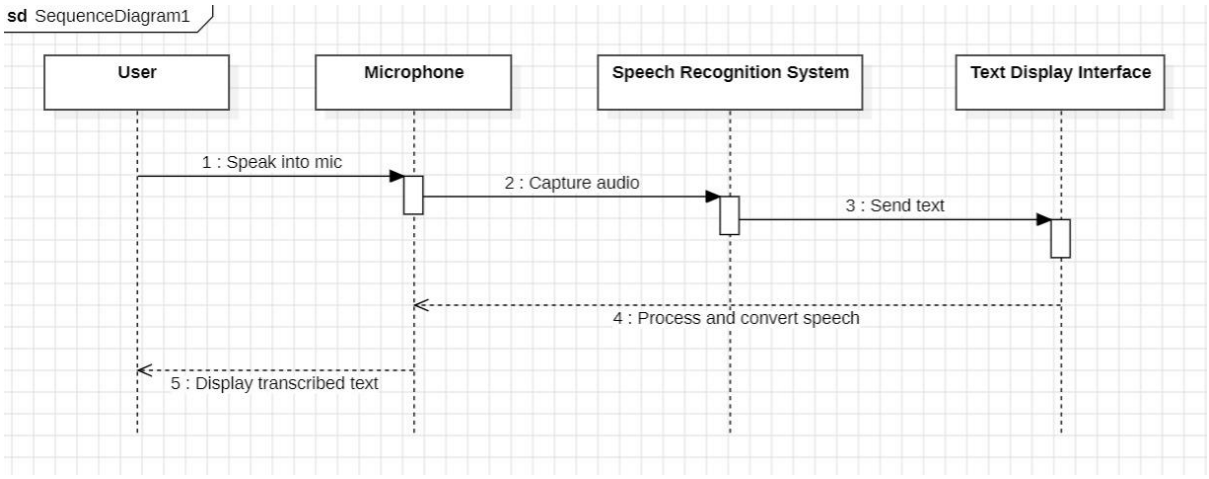
Implementing encryption, access controls, and secure transmission protocols ensures the confidentiality and integrity of user data, addressing privacy concerns in handling sensitive speech data.

4.3 Class Diagram

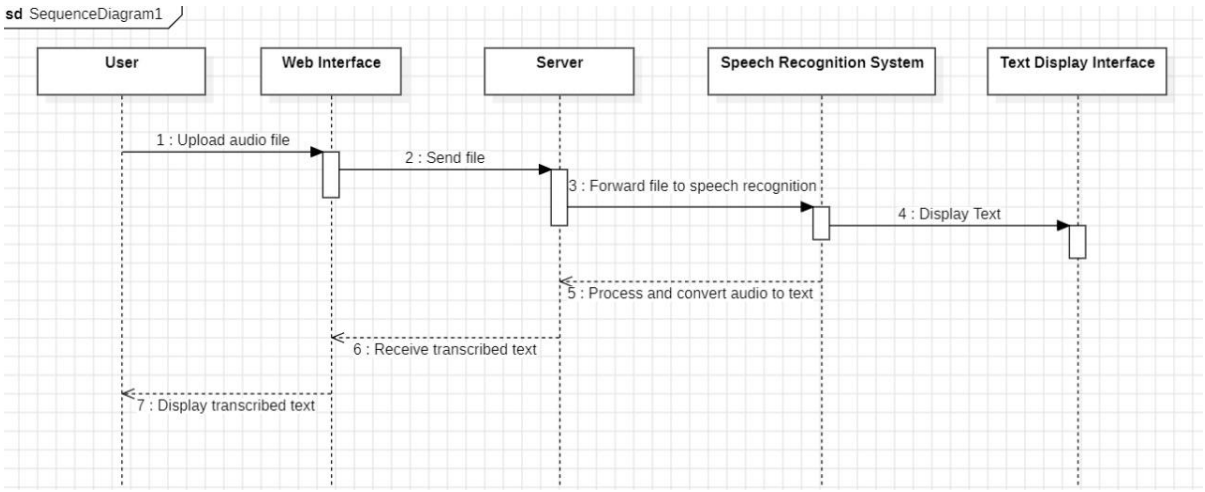


4.4 Sequence Diagram

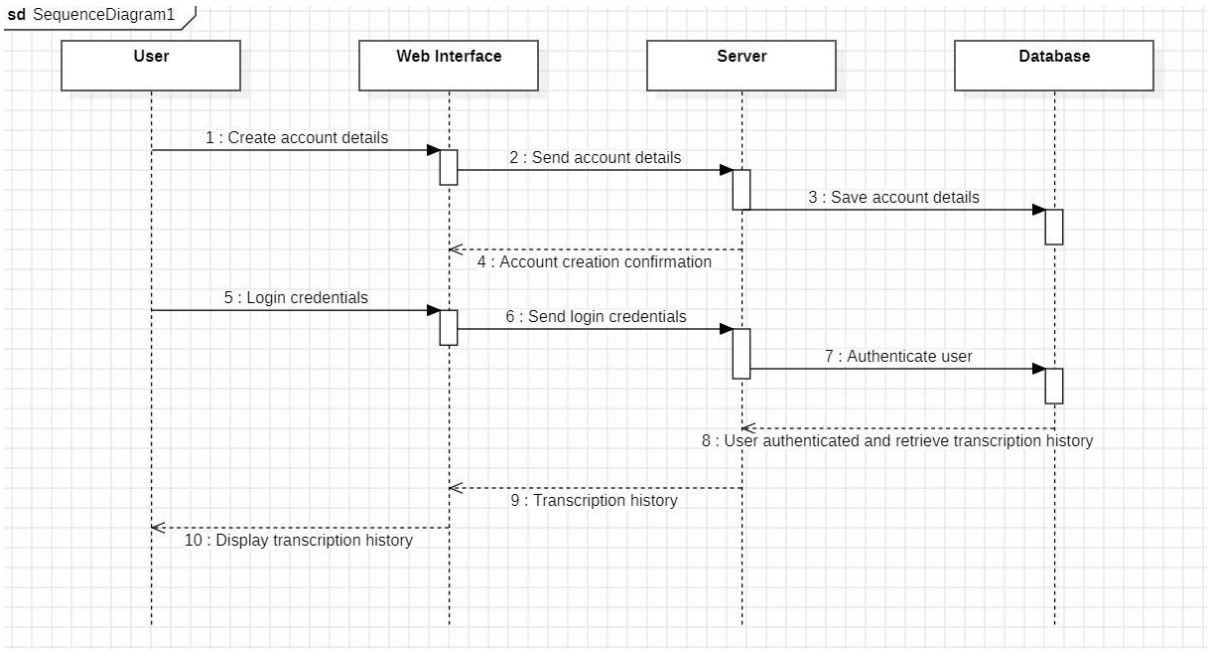
1. Basic Speech-to-Text Conversion



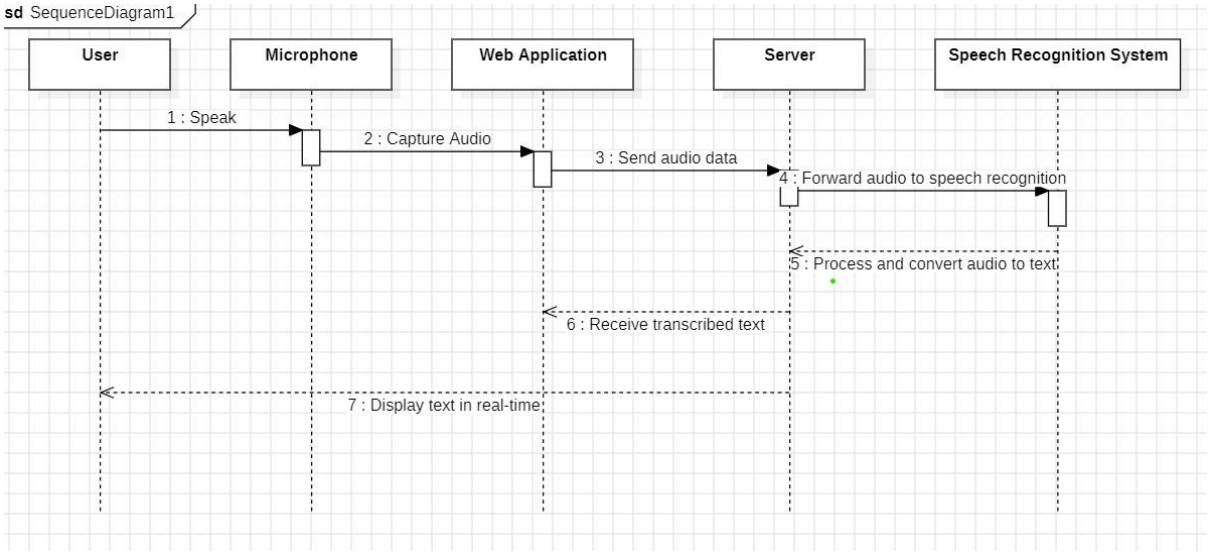
2. Audio File Upload and Transcription



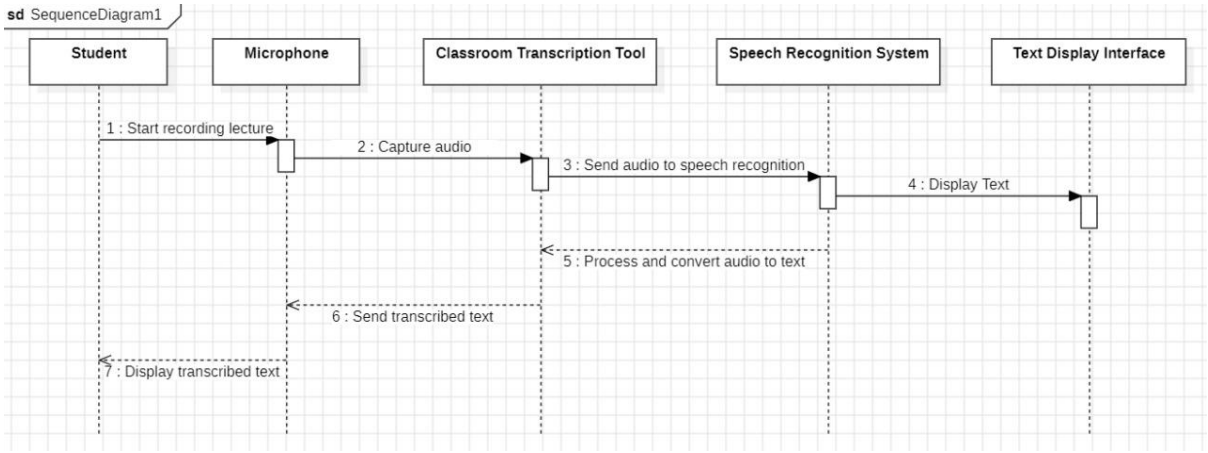
3. User Authentication and Transcription History



4.Real-Time Speech-to-Text Transcription for Deaf Users



5.Classroom Transcription Tool for Notetaking



5. TEST STRATEGY

Objective:

Develop a speech-to-text conversion app for real-time transcription with user-friendly features and disability friendly.

Scope:

Core features include speech recognition, transcription, text display, and basic user authentication for personalized settings.

Test Strategy

a) **Unit Testing:**
   Verify individual components (e.g., speech recognition engine, text display module) in isolation to ensure they function correctly.

b) **Functional Testing:**
   Validate core functionalities such as speech-to-text conversion, language support, and text display to ensure they meet specified requirements.

c) **Usability Testing:**
   Assess the user interface for ease of use, accessibility, and user satisfaction to ensure a positive user experience.

d) **Performance Testing:**
   Evaluate the app's performance under different conditions, including varying network speeds and device capabilities, to ensure optimal responsiveness and scalability.

e) **Security Testing:**
   Verify the app's compliance with security standards, such as data encryption for user authentication and secure transmission of transcribed text, to mitigate potential vulnerabilities.

Test Environment

- **Device:** Use a personal computer or laptop.
- **Browser:** Test on Chrome, Firefox, or Edge.
- **Operating System:** Test on Windows or macOS.
- **Network:** Test on Wi-Fi and cellular networks.
- **Tools:** Use browser developer tools for debugging.
- **Feedback:** Gather feedback from friends or family.
- **Documentation:** Keep track of testing process and findings.

5.1 Test cases

1) **User Authentication**
   - **Happy Path:** Valid Login

     Description: Verify that a user can log in with valid credentials.

     Steps:

     1. Open the login page.
     2. Enter a valid username and password.
     3. Click on the login button.

     Expected Result: User is redirected to the dashboard.

   - **Error Scenario:** Invalid Login

     Description: Verify that the system prevents login with invalid credentials.

     Steps:

     1. Open the login page.
     2. Enter an invalid username and password.
     3. Click on the login button.

     Expected Result: User receives an error message indicating invalid credentials.

2) **Audio File Upload and Transcription**
   - **Happy Path:** Valid Audio File Upload

     Description: Verify that a user can upload a supported audio file and get it transcribed.

     Steps:

     1. Open the transcription page.
     2. Click on the "Upload Audio File" button.
     3. Select a valid audio file (e.g., .wav).
     4. Click on the "Upload" button.

     Expected Result: The audio file is uploaded and transcribed, and the transcribed text is displayed.

- **Error Scenario:** Unsupported Audio File Upload

  Description: Verify that the system handles unsupported audio file formats.

  Steps:

  1. Open the transcription page.
  2. Click on the "Upload Audio File" button.
  3. Select an unsupported audio file (e.g., .exe).
  4. Click on the "Upload" button.

  Expected Result: User receives an error message indicating the file format is not supported.

3) **Real-Time Speech-to-Text Transcription**
   - **Happy Path:** Start Real-Time Transcription

     Description: Verify that a user can start real-time transcription successfully.

     Steps:

     1. Open the real-time transcription page.
     2. Click on the "Start Transcription" button.
     3. Speak into the microphone.

     Expected Result: The spoken words are transcribed into text in real-time and displayed on the screen.

   - **Error Scenario:** No Microphone Access

     Description: Verify that the system handles the scenario where microphone access is denied.

     Steps:

     1. Open the real-time transcription page.
     2. Deny microphone access when prompted.
     3. Click on the "Start Transcription" button.

     Expected Result: User receives an error message indicating that microphone access is required for transcription.

4) **Classroom Transcription Tool**
   - **Happy Path:** Select and Transcribe Lecture

     Description: Verify that a user can select a lecture and transcribe it in real-time.

     Steps:

     1. Open the classroom transcription tool.
     2. Select a lecture or session from the list.
     3. Click on the "Start Transcription" button.
     4. Listen to the lecture and observe the transcription.

     Expected Result: The lecture is transcribed in real-time, and the text is organized into sections.

   - **Error Scenario:** No Lecture Selected

     Description: Verify that the system handles the scenario where no lecture is selected.

     Steps:

     1. Open the classroom transcription tool.
     2. Click on the "Start Transcription" button without selecting a lecture.

     Expected Result: User receives an error message indicating that a lecture must be selected before starting transcription.

5) **Transcription History**
   - **Happy Path:** View Transcription History

     Description: Verify that a user can view their transcription history.

     Steps:

     1. Log in to the application.
     2. Navigate to the "Transcription History" section.

     Expected Result: The user's past transcriptions are displayed with dates and details.

- **Error Scenario:** No Transcription History

  Description: Verify that the system handles the scenario where there is no transcription history.

  Steps:

  1. Log in to the application.
  2. Navigate to the "Transcription History" section.

  Expected Result: User receives a message indicating that there are no past transcriptions.

6. DEPLOYMENT ARCHITECTURE