Software Requirements Specification

Generating Clear speech from Speech impaired Audio

Version 1.0

Prepared by

Roseline Jerry (PES1UG21CS500)

T P Shriambhikesh (PES1UG21CS659)

Tanmay Praveen Udupa (PES1UG21CS662)

Vandana S (PES1UG21CS697)

Table of Contents

Table of Contents 2
Revision History 3
1. Introduction 4
1.1 Purpose 4
1.2 Intended Audience and Reading Suggestions 4
1.3 Product Scope 5
1.4 References 5
2. Overall Description 6
2.1 Product Perspective 6
2.2 Product Functions 6
2.3 Operating Environment 6
2.4 User Classes and Characteristics 7
2.5 Design and Implementation Constraints 7
2.6 Assumptions and Dependencies 8
3. External Interface Requirements 10
3.1 User Interfaces 10
3.2 Software Interfaces 10
3.3 Communications Interfaces 12
4. Analysis Models 13
5. System Features 14
5.1 Speech-to-Text Conversion 14
5.2 Text Correction 15
5.3 Text-to-Speech Conversion 15
5.4 User Feedback on Speech Output 16
6. Other Nonfunctional Requirements 17
Appendix A: Requirement Traceability Matrix 19

Revision History

Name	Date	Reason For Changes	Version
RSTV Capstone	14-03-24	Initial draft	1.0

1. Introduction

1.1 Purpose

The primary objective of our project is to address the communication challenges faced by individuals with speech impairments. By developing a solution that enhances impaired speech outputs, we aim to facilitate clearer and more understandable communication for these individuals. This solution seeks to empower users by providing them with tools to express themselves effectively, convey their ideas accurately, and connect with others more confidently. Through innovative software development, our project endeavors to break down barriers to communication caused by various factors such as medical conditions and developmental disorders. Ultimately, our goal is to improve the quality of life for individuals with speech impairments by enabling them to communicate more efficiently and seamlessly in both personal and professional contexts.

1.2 Intended Audience

1. **Individuals with speech impairments**

   These are the primary users who will directly benefit from the solution developed by the project. They may have various types of speech difficulties caused by medical conditions, developmental disorders, or other factors.

2. **Caregivers and family members**

   This includes individuals who provide support and assistance to those with speech impairments. They may be interested in understanding and utilizing the solution to enhance communication with their loved ones.

3. **Speech therapists and healthcare professionals**

   Professionals working in the field of speech therapy and healthcare may find the project's solution valuable in their practice. They can use it as a tool to assist their clients in improving their communication skills.

4. **Education institutions**

   Schools, universities, and special education programs may have students with speech impairments who could benefit from the project's solution. Teachers and administrators may be interested in incorporating the technology into their educational curriculum or support services.

5. **Researchers and Developers**

   Individuals involved in research or software development related to assistive technologies, accessibility, and communication disorders may find the project's findings and methodologies relevant to their work.

1.3 Product Scope

Objectives and Goals

1. Improve speech intelligibility through speech recognition, enhancement algorithms, and audio adjustments.
2. Develop an accessible and user-friendly interface with multiple input/output options to accommodate diverse user needs.
3. Enable personalization by allowing users to create speech profiles and adapt to individual speech characteristics over time.
4. Ensure seamless integration and compatibility with various operating systems, devices, and assistive technologies.
5. Implement robust data privacy, security measures, and access controls to protect user data and settings.
6. Achieve optimal performance, scalability, and provide comprehensive user support for continuous improvement.

Relationship to Corporate Goals or Business Strategies

1. **Corporate Social Responsibility:** Demonstrates commitment to inclusivity, accessibility, and improving quality of life for underserved communities.
2. **Customer Satisfaction and Loyalty**: Addresses a significant communication barrier, fostering strong customer satisfaction and brand loyalty.
3. **Strategic Partnerships and Collaborations:** Opens opportunities for partnerships with healthcare providers, advocacy groups, and organizations serving individuals with speech impairments.
4. **Corporate Values and Culture:** Aligns with core values of inclusivity, empathy, and making a positive societal impact, contributing to employee motivation and retention.

2. Overall Description

2.1 Product Perspective

The speech synthesis system will serve as a standalone application designed to enhance impaired speech outputs. It will integrate with existing communication devices and platforms, providing an additional layer of functionality for individuals with speech impairments. This system will be scalable and adaptable to different user needs and environments, ensuring seamless integration into various assistive technologies.

2.2 Product Functions

1. Speech Synthesis: The model will generate clear and natural sounding speech based on input text or user prompts.
2. Real time adaptations: It will dynamically adjust speech output based on user preferences, feedback, and environmental factors.
3. User Interactions: The system will provide a user-friendly interface for easy interaction and customization of speech settings.
4. Feedback Mechanisms: Integration of a feedback mechanism will allow users to provide input on speech quality and effectiveness

2.3 Operating Environment

1. **Hardware Platform**
    a. The model must be designed to run on a variety of hardware platforms, including: PCs, Laptops, Tablets and Smartphones.
    b. Minimum hardware requirements - a CPU with speed sufficient to run the chosen programming language and framework, adequate amount of RAM and a display screen.

2. **Operating system and version**

a. Must be compatible with Windows 10 and above
b. Must be compatible with iOS 12 and above

2.4 User Classes and Characteristics

User	Frequency of use	Technical Expertise	Security/privilege Levels	Educational Level
Individuals with speech impairments	Daily	Variable	Low to medium	Can vary- should be able to write, speak and understand language well.
Caregivers and family members	Moderate to Daily	Low to moderate	Low to medium	Can vary- should be able to write, speak and understand
Speech Therapists and Healthcare Professionals	Moderate to Daily	Moderate to high	Medium to high	High level education- expertise in the field
Developers	As needed	High	High	High level education

The software should prioritize providing a user-friendly experience for both frequent users and those who are new to the system, while also accommodating advanced users who require in-depth functionality and training capabilities.

2.5 Design and Implementation Constraints

1. **Compatibility and Integration:** The solution must be compatible with a wide range of existing communication devices and platforms commonly used by individuals with speech impairments. Integration with popular assistive technologies, communication devices, and operating systems
2. **Real time processing:** Given the real-time nature of communication, the system must

be capable of processing and synthesizing speech output with minimal latency. This constraint requires efficient algorithms and optimization techniques to ensure timely responses to user input.

3. **Resource Efficiency:** To accommodate a wide range of devices, including those with limited computational resources, the system must be resource-efficient. This involves optimizing algorithms and minimizing memory and processing requirements while maintaining high-quality speech synthesis.
4. **Customization and Adaptability:** The system should allow for user customization and adaptability to accommodate individual preferences and varying speech impairments. This constraint necessitates flexible design approaches, such as adjustable speech settings and adaptive algorithms.
5. **Privacy and Security:** Given the sensitive nature of communication, the system must adhere to strict privacy and security standards to protect user data and ensure confidentiality. This constraint requires robust encryption mechanisms, secure data storage practices, and adherence to regulatory requirements.
6. **Ethical Considerations:** The design and implementation must prioritize ethical principles, ensuring fairness, inclusivity, and respect for user autonomy. This includes considerations such as unbiased speech synthesis, transparent user consent mechanisms, and adherence to ethical guidelines in research involving human subjects

2.6 Assumptions and Dependencies

1. **Assumptions:**
    a. **Database availability:** It is assumed that the required database system will be available and properly configured for storing user account data. Any issues with the database system may impact the project.
    b. **Text Input Availability:** The system assumes that users will have access to text input methods for generating speech output. This could include typing on a keyboard, using a touch screen, or employing alternative input methods such as eye-tracking devices or switches.
    c. **Feedback Availability:** The effectiveness of the speech synthesis system relies on user feedback for continuous improvement. It is assumed that users will provide feedback on speech quality, customization preferences, and overall user experience to

inform future enhancements.

d. **Collaboration with Healthcare Professionals:** The project assumes ongoing collaboration with healthcare professionals, including speech therapists, clinicians, and researchers. This collaboration is essential for validating the effectiveness of the system, incorporating clinical insights, and ensuring alignment with best practices in speech therapy and assistive technology.

e. **Availability of Resources:** The project assumes access to necessary resources such as computing infrastructure, software development tools, and funding support. Adequate resources are essential for system development, testing, and deployment.

2. **Dependencies**

a. **Third-Party Libraries and Components:** The project may rely on third-party libraries and components for specific functionalities (e.g., security libraries, GUI frameworks, and database connectors). Any changes or issues with these dependencies can affect the project's development and functionality.

b. **Hardware and Network Infrastructure:** The software's performance and functionality depend on the hardware and network infrastructure of the host computer. The project is dependent on the availability and proper functioning of this infrastructure.

c. **Speech Synthesis Algorithms:** The project depends on the availability and effectiveness of speech synthesis algorithms for generating natural-sounding speech output. Research and development efforts will focus on optimizing these algorithms to meet the needs of individuals with speech impairments.

d. **User Engagement and Feedback:** The success of the project depends on active user engagement and feedback. Dependence on user participation for testing, evaluation, and improvement requires effective strategies for soliciting feedback, engaging with user communities, and incorporating user insights into the development process.

3. External Interface Requirements

3.1 User Interfaces

The speech enhancement solution shall provide an intuitive and user-friendly graphical user interface (GUI) that is accessible to individuals with varying levels of technical proficiency and disabilities. It shall provide clear instructions, tooltips, and contextual help, and shall incorporate intuitive icons, symbols, and visual cues. The interface shall allow users to create and manage personalized speech profiles and shall provide feedback mechanisms. The interface shall be designed with a modular and extensible architecture, allowing for future enhancements without compromising usability and accessibility.

1. **Main Screen:**

   The main screen shall allow users to select from the core functionalities of the speech enhancement solution, including speech input, speech enhancement, and output generation.

2. **Speech Input Screen:**

   The speech input screen shall enable users to provide speech input from various sources, such as microphones, audio files, or real-time audio streams.

3. **Speech Enhancement Screen:**

   The speech enhancement screen shall display the progress of the speech enhancement process and shall provide options to adjust settings like speech rate, pitch, and volume.

4. **Output Generation Screen:**

   The output generation screen shall allow users to select the desired output format, such as synthesized speech, text-to-speech conversion, or visual representations, and shall enable the user to customize the output settings according to their preferences and accessibility needs.

3.2 Software Interfaces

The software will interact with the following software components:

1. **Database:** This system will interact with a database to store and retrieve user data.

2. **Operating system:** This system will run on an operating system such as Windows or Linux.
3. **Tools:** This system will use a variety of tools to develop and deploy the software, such as a compiler, a linker, and a debugger.

4. **Libraries/Frameworks:** This system will use a variety of libraries and frameworks to provide common functionality, such as a GUI library and a database library.
5. **Integrated commercial components:** This system may use integrated commercial components to provide additional functionality.

The following data items or messages will come into the system:

1. **User account information:** This information will be retrieved from the database when the user inserts login credentials.
2. **Entering Speech:** The user will input his speech through the system's user interface.
3. **Entering Feedback:** The user will enter his feedback of the output speech through the system's user interface.

The following data items or messages will go out of the system:

1. **Output results:** The system will output the speech which can be heard.

The system will need to communicate with the following services:

1. **Database service:** The system will need to communicate with the database service to retrieve and update speech data.

The nature of the communications between the system and other software components will be synchronous. This means that the system will wait for a response from the other component before proceeding.

The following data will be shared across software components:

1. **Customer account information:** This information will be shared between the ATM simulator and the database.
2. **Transaction data:** This information will be shared between the ATM simulator and the

database. The data sharing mechanism will be implemented using a database. This means that the ATM simulator and the database will share a common pool of data.

3.3 Communications Interfaces

The ATM simulator software will require the following communications functions:

1. **Database communication:** The ATM simulator will need to communicate with a database to retrieve and update customer account information. This communication will use a database-specific protocol, such as JDBC or ODBC.
2. **Client-server communication over HTTPS:** To ensure secure data transmission between the client user and the server, HTTP or HTTPS is to be used for communication between the client and the Flask application.
3. **Electronic Forms:** HTML forms to be used to create the electronic forms for user input.

3.4 Synchronization mechanisms:

The ATM simulator will use a variety of synchronization mechanisms to ensure that data is shared correctly between software components. For example, the ATM simulator will use a database transaction to ensure that customer account information is updated correctly.

3.5 Data Transfer Rates:

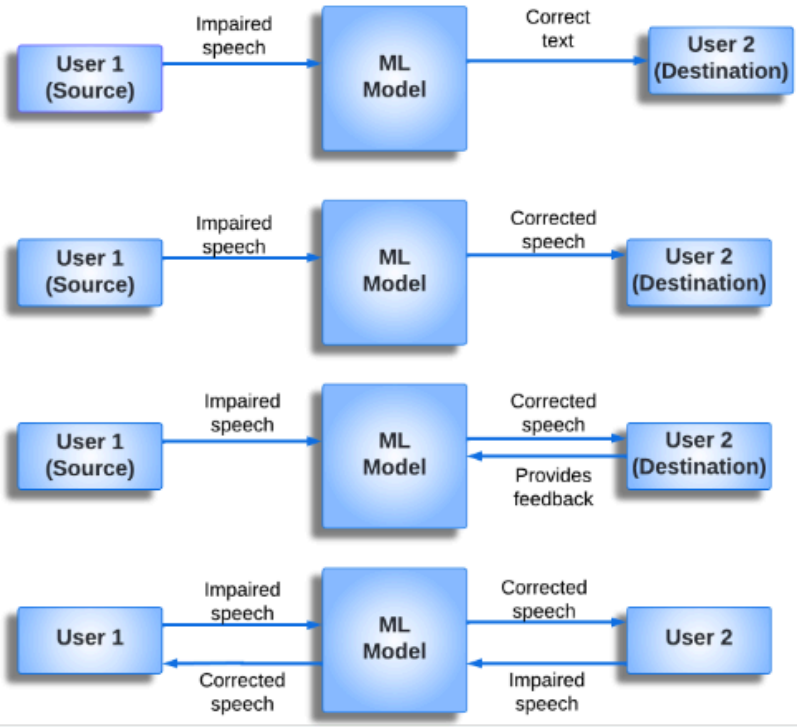
The data transfer rates will largely depend on factors such as server performance, network conditions, and database queries.

3.6 Security considerations:

1. Using strong passwords and authentication mechanisms
2. Implementing input validation to protect against cross-site scripting attacks

4. Analysis Models

4.1 High level diagram



Communication between 2 users where one has impaired speech and the speech is converted to text

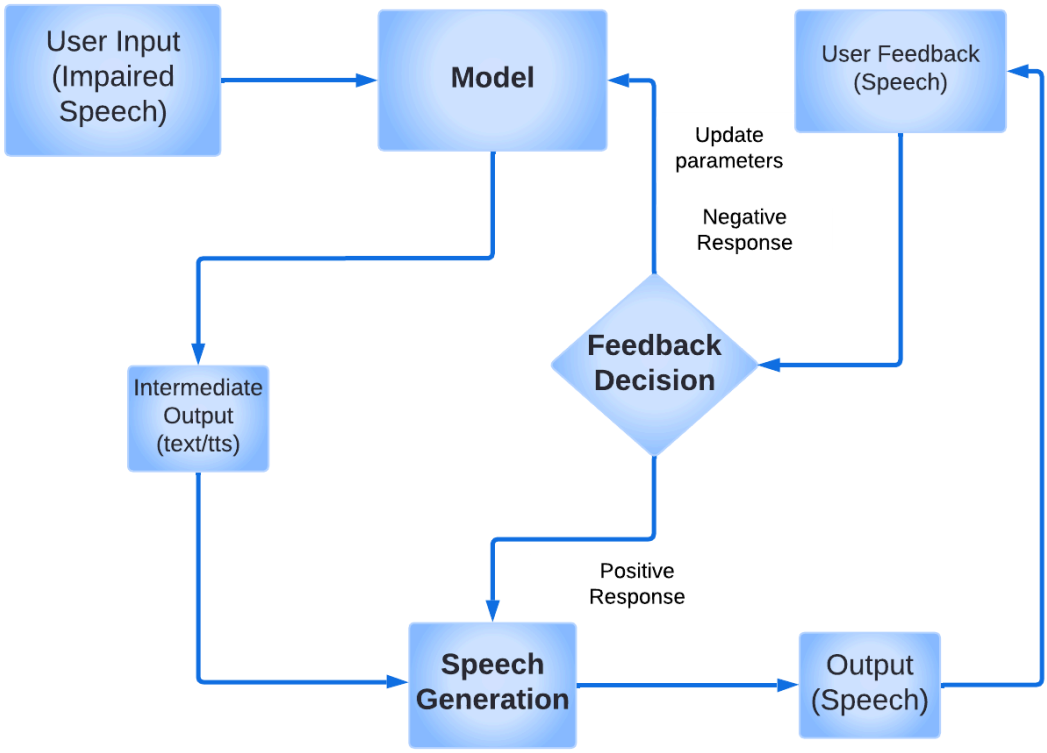
Communication between 2 users, where one has impaired speech so its converted to corrected speech

Impaired speech converted to corrected speech after which the User receiving corrected speech provides Feedback

Communication between two impaired speakers

7

4.2 Pipeline diagram



1. The impaired speech (e.g., stuttering, slurred speech) is fed into the model, which then generates a clear version of the input or translates it to another language based on its prediction.
2. The other user can then give positive or negative feedback, which is used to retrain the model or generate the speech again using the same parameters.

5. System Features

5.1 Speech-to-Text Conversion

5.1.1 Description and Priority

The ability to convert impaired or incorrect speech input from a user into text form. (High priority)

5.1.2 Stimulus/Response Sequences

Stimulus: User provides speech input (which is impaired or incorrect).

Response: The system converts the speech input into text.

5.1.3 Functional Requirements

1. Accurate speech recognition for various accents, dialects, and speech impairments.
2. Robust handling of background noise and environmental factors.
3. Ability to detect and handle multiple speakers in a conversation or audio stream.

5.2 Text Correction:

5.2.1 Description and Priority

Analyzing the converted text and correcting any errors, or impaired speech patterns. (High priority)

5.2.2 Stimulus/Response Sequences

Stimulus: The system receives the converted text from the speech-to-text component.

Response: The system applies natural language processing and correction algorithms to improve the text quality.

5.2.3 Functional Requirements

1. Natural language processing techniques to identify and correct errors, and impaired speech patterns.
2. Ability to learn and adapt to individual users' speech patterns and improve correction accuracy over time.

5.3 Text-to-Speech Conversion:

5.3.1 Description and Priority

Converting the corrected text into natural, clear speech output. (High priority)

5.3.2 Stimulus/Response Sequences

Stimulus: The system receives the corrected text from the text correction component.

Response: The system generates natural, clear speech output based on the corrected text.

5.3.3 Functional Requirements

1. High-quality speech synthesis with natural intonation and pronunciation.
2. Ability to handle complex text inputs, including abbreviations, punctuations.
3. Efficient memory and resource management for handling large or continuous text inputs.

5.4 User Feedback on Speech Output:

5.4.1 Description and Priority

Enabling users to provide feedback on the quality and effectiveness of the speech output generated by the system, and incorporating that feedback to improve the speech synthesis capabilities. (High priority)

5.4.2 Stimulus/Response Sequences

Stimulus: User provides feedback on various aspects of the speech output, such as naturalness, clarity, pronunciation, and overall quality.

Response: The system analyzes and incorporates the user feedback to enhance speech synthesis.

5.4.3 Functional Requirements

1. Incorporation of user feedback into machine learning models or algorithms to improve speech synthesis.
2. Ability to capture and analyze user feedback data to identify areas for improvement.
3. Efficient memory and resource management for handling large or continuous text inputs.

6. Other Non Functional Requirements

6.1 Performance Requirements

The system should process speech in real-time or near-real-time to ensure timely responses.

6.2 Accuracy

The generated clear speech should closely match the intended message of the disordered speech, minimizing errors and distortions.

6.3 Robustness

The system should be capable of handling various types and severities of speech disorders effectively without significant degradation in performance.

6.4 Scalability

The system should be able to handle varying loads of speech processing tasks, accommodating increasing demands without sacrificing performance.

6.5 Compatibility

It should integrate seamlessly with existing speech recognition systems or other applications where clear speech generation is needed.

6.6 Security

Measures should be in place to ensure the security and privacy of the processed speech data, especially if it contains sensitive information.

6.7 User Interface

The system should have a user-friendly interface, making it easy for users to interact with and configure settings as needed.

6.8 Accessibility

The system should be accessible to users with disabilities, such as providing alternative input methods for those who cannot produce clear speech.

6.9 Adaptability

The system should be adaptable to different languages, dialects, and speech disorders, allowing for broader usage across diverse populations.

6.10 Reliability

The system should operate reliably under varying conditions, minimizing downtime and ensuring consistent performance.

6.11 Ethical Considerations

The project should adhere to ethical guidelines, ensuring that the generated clear speech is used responsibly and ethically, without causing harm or perpetuating biases.

Appendix A: Requirement Traceability Matrix

Sl.no	Requirement	Testcase_ID	Test Result	Defect status
1				
2				
3				
4				