# Software Requirements Specification
for
Bengali Sign Language Interpreter

Prepared by:

Faiza Feroz (298)
Neelima Monjusha Preeti (297)
Fariha Akter Promi (293)
Sushmoy Barua Rahul (337)
Nazma Sarker (294)

Jahangirnagar University, Dhaka
Software Engineering

Remote, 06 March 2021

Contents

Chapter 1

Introduction

This document is a Software Requirement Specification of Bengali Sign Language Interpreter application using CNN and TensorFlow.Through this document we are going to provide all specifications ,classes and functionalities as well as what our resultant application will provide .This document will provide specification of external interface which will be with end users , performance features , attributes , that is if our application is portable , maintainability ,design constraints on implementation , input-output specifications , implementation language as well as output expectations.

1.1 Purpose

The aim of this document is to specify the features, requirements of the final product , specifications and the interface of BSL using TensorFlow. It will explain the scenario of the desired project and necessary steps in order to succeed in the task. To do this throughout the document, overall description of the project, the definition of the problem that this project presents a solution and definitions and abbreviations that are relevant to the project will be provided. The preparation of this SRS will help consider all of the requirements before design begins, and reduce later redesign, recording, and retesting. If there will be any change in the functional requirements or design constraints part, these changes will be stated by giving reference to this SRS in the following documents.

1.2 Intended Audience

Project manager, developers and testers will have the access throughout the whole document. Sales and marketing team will only have the access of Introduction and Overall Description part.

1.3 Intended Use

Project manager will overview the document and ensure that all the functional requirements meet user need properly and user can user the intended product smoothly, PM will also discuss about functional requirements with developers and testers to ensure that all possible criteria and dependencies are cleared. Developers will follow the requirements and develop the project while testers will write test cases based on the requirements. Then after the development process, tester will test and make sure all those criteria are fulfilled.

1.4 Product Scope

The project which is going to be presented in this document is called Bengali Sign Language Interpreter. This application is planned to be used by hearing and speech impaired people in order to ease their life and express themselves to the rest of the world, and also government offices that should serve all of its citizens equally,private companies that want to reach and serve hearing and speech impaired people as well, organisations which aims to help hearing and speech-disordered people. The application will work as the person whose gestures to be recognized (called "user" here after) stand in front of the device(here web application). For system to specify the special points on the body, the user should stand in a certain position in order to track his/her body.After tracking the user's body, system is ready to use.It will provide an output of what it sees in front of its camera in the form of the hand movements mainly on the body. The project will use the output come from using tensorFlow and will match them with the pre-defined gestures.In order to define gestures, we will use tensorFlow again, and store the output of each gesture to a database.After this when a user performs a gesture, the program will try to determine and match it with a pre-defined gesture.If successful , it will give the meaning of the gesture in text format or audio version as user chooses.Again if user types text or inputs audio by selecting an avatar output will be converted into Bengali sign language and avatar will show the output as sign language. In this project it is aimed to interpret Bengali Sign Language and translate it into a text/audio in Bengali only. Other languages will neither be considered as input language nor as output language. Obviously, our program will have some limitations like gestures based on finger movements, so the program will not cover abstract words of BSL. For the beginning the program will recognize only 10 gestures pre-defined. Yet, it will be possible to define more gestures once it is proved that the program works well enough. The program will work on PC environment mainly.

1.5 Risk Definition

If model is not trained properly with database could predict wrong words which will create miscommunication that will create fatality in all communication sectors between speech impaired people and able/normal people. Database should contain clear words and vocabulary . Without clarity words could be misinterpreted.

Chapter 2

Overall Description

2.1 User Classes and Characteristics

This application will be used by 2 categories of users. The users are speech and hearing-impaired people and normal human beings.
The user, speech and hearing-impaired person, is required to know Bengali sign language. Normal user must know Bengali language. Both users have to know basic functions of a smartphone specifically operating camera and microphone.

2.2 User Needs

The user, speech and hearing-impaired person, wants to communicate with a normal person without any help of a human interpreter and vice-versa.
The speech and hearing impaired user wants to convert an audio recording to sign gestures. The user may also want to learn Bengali sign language or look up the meaning of a sign gesture really quickly.

2.3 Operating Environment

1. The system will be a hybrid app that will be available in both android and iOS.
2. We will use TensorFlow 2.0 which requires python version 3.5-3.7 supported in a 64-bit system
3. Windows operating system requires more than windows 7 and for Linux Ubuntu 16.04 or later

2.4 Constraints & Assumptions

1. A Standard Language: Every area in Bangladesh has a unique use of language. Each region in Bangladesh has its own dialect. The process of using sound, pronunciation, expression differs from one to another. This will be a big challenge for us to overcome this complication, but for now, we are willing to focus on a standard version.
2. Front-line User: Our system is equally compatible for a normal person, but we are targeting speech & hearing impaired peoples as our primary users.
3. Completion time: As this will be the first attempt to make a complete interpreting system, this will require quite a bit of time to complete this system.
4. Lack of sufficient light: As per the system it is built on recognizing the hand gestures of the user. So sufficient lighting is a must to have the correct interpretation.

5. Background Noise: To communicate using audio messages, it is mandatory to have a quiet background.
6. Dependency on the third-party tool: As per our feature of having an association with online meeting platforms like(Zoom, Google Meet) we are dependent on their input form for interpretation and connectivity issues for collaborating with us.
7. Distance from device to object The user should stand between 1 to 1.5 meters away from the interface The program will work on the PC environment mainly.

Chapter 3

Requirements

3.1 Functional Requirements

3.1.1 Sign to Audio:

This feature will convert the recognized gesture into the textual meaning of the gesture and then convert it into audio and output the converted audio to the user.

3.1.1.1 Description and Priority

System will recognize the appropriate movement of the hands and will search its database to match the movement with the pre-defined gestures. After matching system will add the meaning of the sign and parse it into audio format and output it in audio format.

3.1.1.2 Stimulus

Data Flow:

Normal Flow of Events:

1. User selects a voice personality.
2. select sign input.
3. select open camera option.
4. select output as audio.
5. User performs the movement.
6. Gesture is recognized and has a match.
7. Gesture is converted to a text file.
8. Text is parsed and converted to audio.
9. Audio is added to the file and outputs to user.

Alternative flow of the events:

1. Gesture could not be recognized due to background noise or standing out of device's range.
2. User gets a feedback explaining the correct position or existence of noise in the background.
3. Connection lost . Try again Later.

3.1.1.3. Functional Requirements:

REQ-1: Performed movement must be a pre-defined movement in order to get the meaning of it as return value.
REQ-2: User should be positioned 6 (1.8 m) to 8(2.4 m) feet away from the system.
REQ-3: Background should be external noise free.
REQ-4: Internet connection must be there .

3.1.2. Audio to Sign:

This feature will convert audio into text and then into sign gesture and display the output gesture to the user.

3.1.2.1 Description and Priority

System will parse the audio and after converting into text,search database for pre-defined gesture .After matching system will add sign of the text and show the output.

3.1.2.2 Stimulus

Data Flow:

Normal Flow of Events:

1. User selects a avatar.
2. User presses on microphone option as long as he/she speaks.
3. Release microphone and speech is stored locally.
4. Audio is converted to text.
5. Dependency parser for analysing grammatical structure of the sentence and establishing relationship between words.
6. Text has a match and recognizes with pre-defined gesture.
4. Audio is converted to text.
7. The gesture is added to the file and displayed to the user.

Alternative flow of the events:

1. Converted text does not have a match with any pre-defined gesture in the database.
2. User gets a feedback to pronounce the sentence/word clearly or correct grammatical structure of sentence/-word.
3. User gets a feedback to reduce noise from background .

3.1.2.3. Functional Requirements:

REQ-1: Converted text must be in the database to match with a pre-defined gesture.
REQ-2: Text included in the database should be clear and concise and without ambiguity .
REQ-3: Noise free background and as close as possible to microphone for getting a match with converted text from database. .

REQ-4: Internet connection must be there.

3.1.3 Sign to Text:

This feature will convert the recognized gesture into the textual meaning of the gesture and display the converted text to the user.

3.1.2.1 Description and Priority

System will recognize the appropriate movement of the hands and will search its database to match the movement with the pre-defined gestures. After matching system will add the meaning of the sign and show the output.

3.1.3.2 Stimulus

Data Flow:

Normal Flow of Events:

1. User selects a avatar.
2. select sign input.
3. select open camera option.
4. select output as text.
5. User performs the movement.
6. Gesture is recognized and has a match.
7. The text is added to the file and displayed to the user

Alternative flow of the events:

1. Gesture could not be recognized due to background noise or standing out of device's range.
2. User gets a feedback explaining the correct position or existence of noise in the background.
2. Connection lost . Try again Later.

3.1.3.3. Functional Requirements:

REQ-1: Performed movement must be a pre-defined movement in order to get the meaning of it as return value.
REQ-2: User should be positioned 6 (1.8 m) to 8(2.4 m) feet away from the system.
REQ-3: Internet connection must be there .

3.1.4. Text to sign:

This feature will convert text into the sign gesture and display the output gesture to the user.

3.1.4.1 Description and Priority

System will parse the text and search database for pre-defined gesture .After matching system will add sign of the text and show the output.

3.1.4.2 Stimulus

Data Flow:

Normal Flow of Events:

1. User selects a avatar.
2. Writes text in text box.
3. select open camera option.
4. Text has a match and recognizes with pre-defined gesture .
5. The gesture is added to the file and displayed to the user.

Alternative flow of the events:

1. Text doesn't have a match with any pre-defined gesture in the database.
2. User gets a feedback explaining to change word or type synonyms of that sentence or word to match output.

3.1.4.3. Functional Requirements:

REQ-1: Inputted text must be in the database to match with a pre-defined gesture.
REQ-2: Text included in the database should be clear and concise and without ambiguity .
REQ-3: Internet connection must be there.

3.1.5. Choose Avatar:

This feature will choose an avatar for the user to display the sign gestures.

3.1.4.1 Description and priority:

System will provide a few avatars e.g male, female, kid and after selecting the avatar the sign outputs will be shown according to the avatar.

3.1.5.2 Stimulus:

Data flow:

Normal flow of events:
1. User clicks the option for choosing an avatar.
2. User selects an avatar according to preference.

Alternative flow of the events:
There's no alternative flow for this functionality.

3.1.5.3 Functional Requirements

REQ-1: Internet connection must be there.

3.1.6. Voice personality:

This feature will choose a voice personality for the user to output audio, converted from sign gestures.

3.1.4.1 Description and priority:
System will provide some voice personalities. After the user chooses their preferred voice personality, the audio output will be played with this voice.

3.1.6.2 Stimulus:

Data flow:

Normal flow of events:
1. User clicks option for choosing voice personality
2. User selects a voice personality according to preference

Alternative flow of the events:
There's no alternative flow for this functionality.

3.1.6.3 Functional Requirements

REQ-1: Internet connection must be there.

3.1.7. Recording to sign:

This feature will upload any audio recording of max 2 minutes with mp3 or wav file format to the cloud server of the application and show the sign output of the audio through avatar.

3.1.7.1 Description and priority:
System will detect the audio from the recorded file and recognize the matched sign from database by searching. Then it will present the the associate sign one by one through the avatar.

3.1.7.2 Stimulus:

Data flow:

Normal flow of events:
1. User clicks the option for uploading any audio record.
2. User selects option for "Watch Sign" when the output it ready to show.
3. User selects an avatar according to preference to watch the output.

Alternative flow of the events:
1. If any sign match for any word is not found then a toast will show with a message that "Sign not found for audio."
2. If the recording is not uploaded to the cloud then an error message will show "Error occurred with uploading file, please try again."

3.1.7.3. Functional Requirements:

REQ-1: Internet connection must be there.
REQ-2: Audio to sign feature must be done to initiate this feature.

3.1.8. Reload:

This feature will show the sign output again which was just shown on display.

3.1.7.1 Description and priority:

System will store every sign output as cache and delete the second last sign every time after the last sign encounters. Then it will show the stored sign using the avatar.

3.1.7.2 Stimulus:

Data flow:

Normal flow of events:
1. User clicks the option for reload on the display.
2. User sees the previous sign through the selected avatar.

Alternative flow of the events:
1. If the system can't store the sign then an error message will show "Failed to reload. Please input again."
2. If the sign output isn't produced then an error message will show "Couldn't find match. Please try again."

3.1.7.3. Functional Requirements:

REQ-1: Sign output must be produced.
REQ-2: Showing though avatar must be functional.

3.1.9. Association:

This feature will collaborate with third-party communication tools like (Zoom, Google Meet) by which interpretation can be done at online conferences and meet-ups.

3.1.9.1 Description and priority:

The system will take audio as input from the third-party communication tools and will convert it into the sign and represent it through avatar to the end-user(speech & hearing impaired person) and vice versa when an impaired person is on speaker mode

3.1.9.2 Stimulus:

Data flow:

Normal flow of events:
1. Confirm access of interpreter in meeting software for first time.
2. Select start interpreter for capturing audio
3. Input audio .
5. User will be asked to select an avatar or stay with the default one.
6. Audio will be parsed into list of text and checked for validation.
7. User selects open camera option

8. And accordingly selects output as sign option.
9. Valid texts will be visually demonstrated by the selected avatar.

Alternative flow of the events:
1. Connection error will be popped up in case system fails to connect.
2. User will get option of 'Remember me' to avoid repeated confirmations.

3.1.9.3. Functional Requirements:

REQ-1: Inputted text must be in the database to match with a pre-defined dataset.
REQ-2: User must have internet connectivity

3.1.10. Sign Dictionary:

This feature will show a visual representation of the text given by the user by recognizing text.

3.1.10.1 Description and priority:

The system will take input as a valid text and compare it with the internal dataset of sign language. If the match is found a pictorial display will be shown in the case of a single alphabet or number. In the case of a sentence, a visual demonstration will guide the user.

3.1.10.2 Stimulus:

Data flow:

Normal flow of events:
1. The user selects a preferred avatar or continues with default one.
2. The user selects the dictionary option from menu-bar.
3. Then the user gets redirected to the dictionary interface that asks for input from the user.
4. The system accepts the user input and checks the validity.
5. Successful validation brings a pictorial or visual demonstration of that input in sign language.

Alternative flow of the events:
1. User input doesn't match with dictionary database, shows a pop up of "no match found"
2. User gets a list of similar words as a suggestion of the following input.

3.1.10.3. Functional Requirements:

REQ-1: Inputted text must be in the database to match with a pre-defined dataset.
REQ-2: User must have internet connection.

3.2 Non Functional Requirements

3.2.1 Performance Requirements

1. The system must translate sign to text/audio and vice-versa with at least 98% accuracy.
2. The system should complete the required conversion in real-time.

3.2.2 Design Requirements

For the time being, we will use TensorFlow to control the input stream and python programming language in the Visual Studio will be used with several additional libraries and CNN for neural network implementation. Software architecture will be based on real time continuous gesture recognition methods. However these are not strictly decided constraints, there will be some changes in training model and SDK or programming language etc. Any alteration that we made in the design constraints will be mentioned in the following documents.

3.2.3 Error Handling

1. Invalid input will provide a pop up of "Please enter valid input" to the user
2. "Can't recognize hand" will show up if user tries to use the system in absence of light
3. Absence of required functionality (Ex: trying to get sign output without selecting avatar, using audio before allowing the system to access microphone etc) will be shown in asterisk
4. Databases will use sharding to be redundant to prevent loss of data.
5. Backups of the databases should be done accordingly with the expansion of signs .

3.2.4 Interface requirement

1. User interface requirement Front-end software: React Native Back-end software : Node js , Django Databases software : Nosql, AWS/Azure

2. Hardware interface requirement
i. This will be a hybrid app, the software will be compatible with android, mac and windows operating system, so compatible devices are Android smartphones, Iphone, windows tablets and pc.
ii. Mobile camera mobile microphone with speaker, desktop external camera and microphone with speaker will be required to use this software.

3. External interface requirement: Here the first external requirement is a cloud database. The database will accept sign and validate with the predefined signs. Second requirement is specific associated third party software like Zoom, Google Meet.

3.2.5 Safety Requirements

1. System must have secure database.
2. Only developer can modify or edit,delete database.User don't have access to modify them.
3. The conversion which took place with the user,can't decoded by outsider third party.

Appendices

Appendix A

Glossary

AWS - Amazon Web Service

BSL - Bangla Sign Language

CNN - Convolutional Neural Network

Dataset - A collection of data with a defined structure.

Hybrid App - A hybrid app is a software application that combines elements of both native apps and web applications. Hybrid apps are essentially web apps that have been put in a native app shell. Once they are downloaded from an app store and installed locally, the shell is able to connect to whatever capabilities the mobile platform provides through a browser that's embedded in the app. The browser and its plug-ins run on the back end and are invisible to the end user.

iOS - iPhone Operating System.

PM - Project Manager

Sharding - Sharding is a method of splitting and storing a single logical dataset in multiple databases.

SRS - Software Requirement Specification

TensorFlow - It is an open source artificial intelligence library, using data flow graphs to build models.