

Software Requirements Specification

Speech Emotion Recognition

**Abdullah Özder - 202011410, Furkan Duran - 201811027, Elif Aybüke Coşkun - 201811018,
Şima Kayısı - 201811043, İhsan Bardakçı - 201717007**

Table of Contents

1. INTRODUCTION	2
1.1 Purpose	2
1.2 Scope of Project	2
1.3 Glossary	3
1.4 References	3
1.5 Overview of the Document	4
2. OVERALL DESCRIPTION	4
2.1 Product Perspective	4
2.1.1. Development Methodology	4
2.2 User Characteristic	4
2.2.1. Participants	4
2.2.2. Admin	5
3. REQUIREMENTS SPECIFICATION	5
3.1 External Interface Requirements	5
3.1.1. User interfaces	5
3.1.2. Hardware interfaces	7
3.1.3. Software interfaces	7
3.1.4. Communications interfaces	7
3.2 Functional Requirements	7
3.2.1. Home Page Use Case	7
3.2.2. Log In Page Use Case	8
3.2.3. Speech Page Use Case	9
3.2.4. Text Page Use Case	10
3.2.5. Result Page Use Case	11
3.3 Performance Requirement	12
3.4 Software System attributes	13
3.4.1. Availability	13
3.4.2. Ease of Use	13
3.4.3. Reliability	13
3.4.4. Maintainability	13
3.4.5. Security	13
3.4.6. Scalability	13
3.5 Safety Requirement	14
4. REFERENCES	14

1. INTRODUCTION

1.1 Purpose

The purpose of this document is to introduce and detail the system called Speech Emotion Recognition. This system makes an emotion analysis from voice data and produces a result among the options "happy", "angry", "sad", "neutral". It is planned to be used both in a system that normal people can use in their daily lives and in a system where priority is given to the emotional state of the customer, especially for those working in the call center. This document also mentions the details and requirements of the project.

1.2 Scope of Project

Speech is the main and most effective way of human communication. There is no transfer of emotion just by speaking the words in a straight tone. There is a transfer of emotion in every person's speech, even if he is not aware of it. Detecting this emotional state is also a difficult task, because when a person says "awesome" they may be making an allusion to it. But it is a fact that there are physiological changes in people according to their emotional state. There are also devices that determine mood by transferring data from things such as heart rate, blood pressure, blood pressure to a device. But without the need for a device, it is possible to determine the emotion from facial expressions and voice.

The Speech Emotion Recognition system, on the other hand, aims to perform an emotion analysis only through voice without the need for a machine. Through this system, this system can be used in many areas in daily life. For example, in job interviews, it is aimed to understand the emotional state of a candidate from his voice. It is important in the process of recruiting people, in the process of assessing human emotions, in matters such as detecting lies and other issues like this.

Another area that can be used is education. It can be understood which student is nervous, happy or angry from the voices of the students during the online education. In this way, more or less comments can be made about the efficiency level of the course.

Another area is customer service. Customer service is one of the most important factors that increase or decrease the value of a company on the consumers' side. For this reason, it is aimed to further improve and reach the best levels of communication between customers and businesses around the world. With the increasing e-commerce sector, many people are connected to customer service. Most of the people who connect to customer service are connected to customer service because they have problems. And while waiting in line to connect, many customers get nervous or angry. The Speech Emotion Recognition system is also a candidate to be one of the ideal solutions for this scenario.

It aims to connect angry customers to customer service by measuring the emotional state of the customers and making a ranking. Another of the potential uses of the Speech Emotion Recognition system.

In the Speech Emotion Recognition system, there are 2 actors, the admin and the user whose voice will be analyzed. In order for the system to be used and for the voice analysis to start, the voice data must be processed by the system. The owner of this voice data, usually the user, is aware that his voice is being processed and must give the necessary data privacy permission.

The Speech Emotion Recognition system will produce a result after the required permission and voice processing is complete. The options for these results are "Angry", "Sad", "Neutral" and "Happy".

1.3 Glossary

Term	Definition
SER	The main project is an abbreviation of the name Speech Emotion Recognition.

1.4 References

[1] Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia*, 51(1), 98-105,2013

1.5 Overview of the Document

The second part of the document includes titles and details such as External Interface Requirements, Functional Requirements, Software System attributes, Safety Requirement.

The Requirement Specification chapter is written for software developers and details of the functionality of the system are described in technical terms.

The remainder of the document covers the details of the Speech Emotion Recognition system in a technical way.

2. OVERALL DESCRIPTION

2.1 Product Perspective

Speech Emotion Recognition is a project that is designed to be used in daily life, education, job interviews and many other areas, and produces a result by analyzing emotion from voice. The basic dynamic of the project is to take a voice file as input and process it with certain methods and produce a result. 4 different results can be produced according to the content of this produced voice file. With this project, it is to do emotion analysis only with sound, without the need for any tool or machine.

2.1.1. Development Methodology

2.1.2. Product Features

The software described in this SRS document will be used to help people understand their mood from speech or text that users write. Since it provides the continuation of human and speech life, it is a software that can adapt to any desired area.

2.2 User Characteristic

2.2.1. Participants

2.2.1.1. Student

2.2.1.2. Customer

2.2.1.3. Worker

2.2.2. Admin

- 2.2.2.1. Company
- 2.2.2.2. Teacher
- 2.2.2.3. Employer

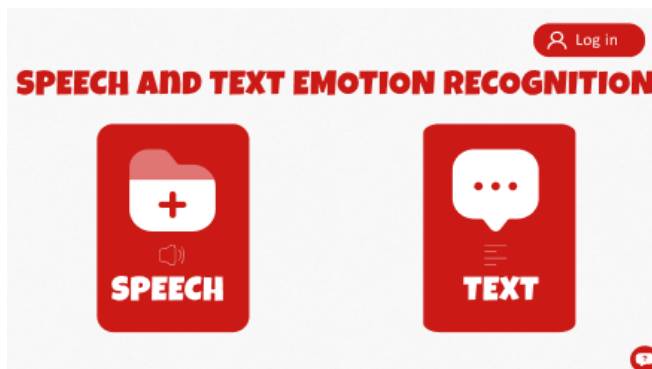
3. REQUIREMENTS SPECIFICATION

3.1 External Interface Requirements

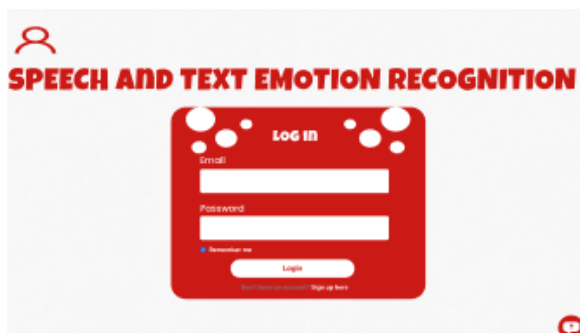
3.1.1. User interfaces

Our software works actively on all platforms with Python installed. Here's what the user can do with interface:

- Files can be added externally (voice or text)



- Can provide contact details.



-Can only manually write the text he wants to find his emotion.

The screenshot shows a web interface titled "SPEECH AND TEXT EMOTION RECOGNITION". In the top right corner, there is a user profile icon and the text "user_name". The main heading is "TEXT" in large red letters. Below it is a large red rectangular box with a white inner area for text input. A small placeholder text "Enter speech/text here" is visible. At the bottom of the red box is a red button labeled "SUBMIT". In the bottom right corner, there is a red speech bubble icon.

-After people upload the audio files they have previously recorded or after the text they have written, they enter this page and select the speech or text page they have applied, and the result is evaluated. As a result of the evaluation, there is a screen that concludes which of the 5 basic emotions is.

The screenshot shows a web interface titled "SPEECH AND TEXT EMOTION RECOGNITION". In the top right corner, there is a user profile icon and the text "user_name". The main heading is "RESULT" in large red letters. On the left, there are two radio buttons: "Speech" (selected) and "Text". Below them is a red button labeled "SELECT". Below the radio buttons is a red button labeled "EVALUATE". To the right of the "SELECT" button is a red box titled "RESULT TABLE" containing the following data:

Emotion	Percentage
Happy 😊	70%
Neutral 😐	10%
Sad 😞	5%
Angry 😡	10%
Fear 😨	5%

Below the "RESULT TABLE" is a yellow box titled "-RESULT-" containing the word "Happy" in blue and a yellow smiley face emoji 😊. In the bottom right corner, there is a red speech bubble icon.

-This page contains information about using the system. When the user has a question about the system, it allows them to take advantage of this page and continue the process.

The screenshot shows a web interface titled "SPEECH AND TEXT EMOTION RECOGNITION". In the top right corner, there is a user profile icon and the text "Log in". The main content area contains three numbered questions:

- 1- WHAT IS SPEECH EMOTION RECOGNITION SYSTEM?
.....
- 2- HOW CAN I USE THIS SYSTEM?
....
- 3-.....

In the bottom right corner, there is a red speech bubble icon.

3.1.2. Hardware interfaces

Computer must have one USB port for video recording. It also requires one microphone input for voice recording.

3.1.3. Software interfaces

-Computer must have the libraries associated with Python. Example: Librosa.

-**Visual Studio Code:** To implement the project, we chose HTML, CSS, and Javascript languages that provide more interactive support.

- **Database**

3.1.4. Communications interfaces

Internet connection required to run this software.

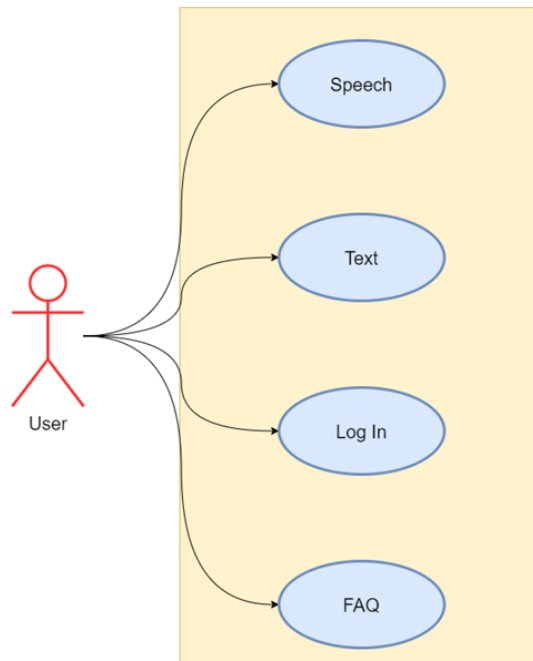
3.2 Functional Requirements

3.2.1. Home Page Use Case

Use Case:

- Speech
- Text
- Log In As
- FAQ

Use Case Diagram:



Description:

The home page schema has basic operations on what the "User" role can access. The user encounters this menu when user first logs into the system. User can use "Speech", "Text", "Log In" and "FAQ" operations.

Initial Step by Step Description:

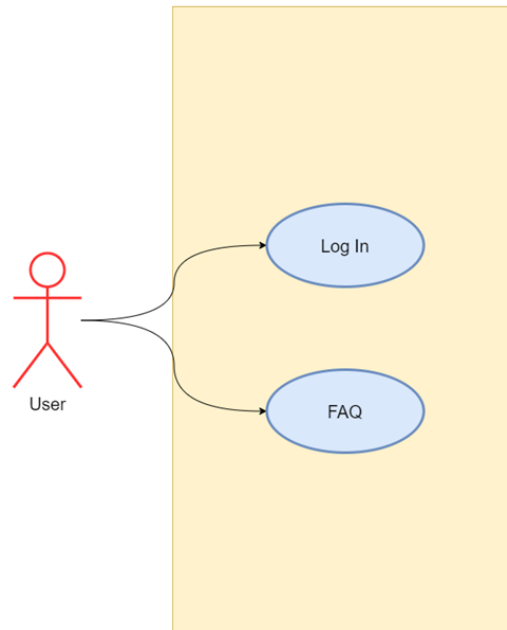
1. The user must enter the home screen.
2. The user should choose the theme for emotion recognition. If it is speech emotion recognition, "Speech", if it is to recognize emotion from text, click the "Text" button.
3. If the user wants to log in, user must click the "Log In" button and log in. If validation is not obtained, the log in must be repeated.
4. The user can use the "FAQ" button for questions that come to mind. This screen contains frequently asked questions.

3.2.2. Log In Page Use Case

Use Case:

- Log In
- FAQ

Use Case Diagram:



Description:

The login page only focuses on login. The user can log in by filling in the "User Name" and "Password" sections on this page. If verification is not achieved, the entry is repeated.

Initial Step by Step Description:

1. The user comes to the login screen.
2. User fills in "User Name" and "Password".
3. The user clicks the "Log In" button and logs into the system.

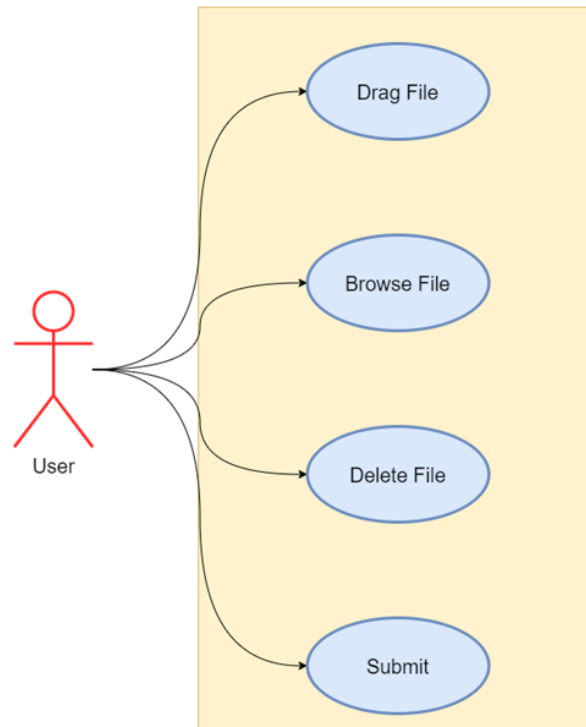
3.2.3. Speech Page Use Case

Use Case:

- Drag File
- Browse File
- Delete File

- Submit

Use Case Diagram:



Description:

User will use this screen for emotion recognition from audio file. From this screen, the user can upload and delete audio files and submit the audio for emotion recognition. From this screen, the user can use "Drag File", "Browse File", "Delete File" and "Submit" operations.

Initial Step by Step Description:

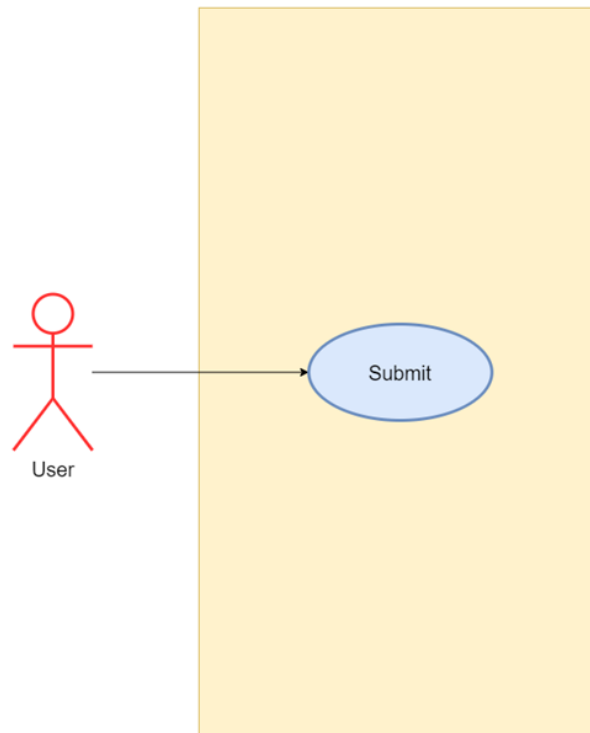
1. The user must first click the "Browse File" button to select the audio file from his computer. If user does not want to select an audio file from his computer, user should drag and drop the audio file into the "Drag File" section.
2. If the user uploaded a wrong file, they can click the "Delete File" button to delete the selected file and upload a new file.
3. After uploading the audio file, the user should click the "Submit" button to see the results.

3.2.4. Text Page Use Case

Use Case:

- Submit

Use Case Diagram:



Description:

User will use this screen for emotion recognition from text file. The user can write text from this screen and submit the text for emotion recognition. The user can use the "Submit" operation from this screen.

Initial Step by Step Description:

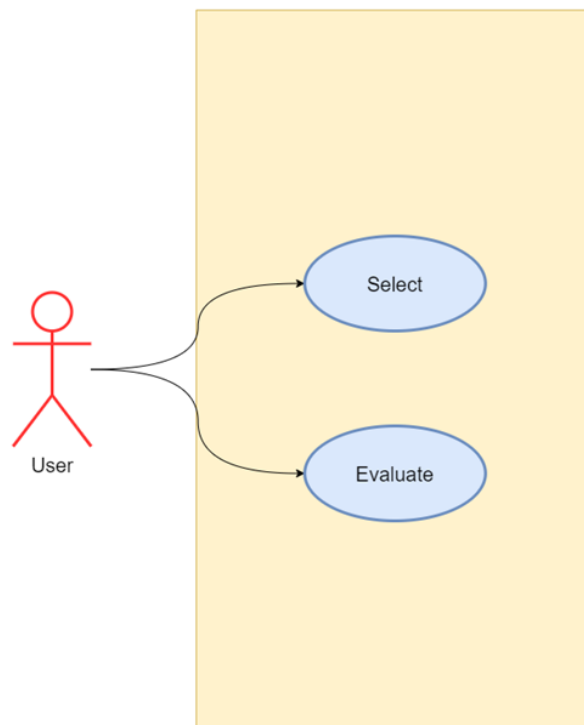
1. The user must first fill in the text part with the text for emotion recognition.
2. After typing the user text, he should click the "Submit" button to see the results.

3.2.5. Result Page Use Case

Use Case:

- Select
- Evaluate

Use Case Diagram:



Description:

Finally, the user uses this screen to see the results of the uploaded text or audio file.

Initial Step by Step Description:

1. The user must first select "Text" if he has uploaded a text file, or "Speech" if user has uploaded an audio file, and then click the "Select" button.
2. After making the selection, the user can click the "Evaluate" button to see the results of the uploaded audio or text file.

3.3 Performance Requirement

Since the Speech Emotion Recognition System does not require high-performance hardware, a strong internet connection will be sufficient. In addition, a personal computer suitable for daily use will be sufficient.

3.4 Software System attributes

3.4.1. Availability

- The system will work on Windows-based and Android-based operating systems.

3.4.2. Ease of Use

- This system is being created as a user-oriented project. It should be simple for the user to use. This simplicity should be the same for the first user and should be the same for the frequent user. The interface we have created will be simple, user-friendly and understandable as a result.

3.4.3. Reliability

- The reliability of the System depends on the sound quality or the accuracy of the text to be analyzed.
- The data provided by the user will be used to compare with the result and to measure the reliability.
- If enough data are collected, the most recent machine learning algorithms should be able to reliably utilize the user's data.

3.4.4. Maintainability

- The administrator will review and update the training and testing files throughout the period of three months.
- The maintenance period is not a matter because the reliable version is always run on the server which allows users to access past summarization.

3.4.5. Security

- Since no data is acquired and saved from the run time, there is no adaptability requirement.

3.4.6. Scalability

- Since only one participant uses the system at a time, there is no scalability requirement.

3.5 Safety Requirement

Users' data will be retained in the database so it can be utilized for system development, upgrades, and maintenance, which are scheduled to be carried out every three months. The system will be strengthened and its stability will be increased using the data. A pre-approval text will be declared at the stage of system membership, declaring that the data will only be used for system improvement, in order for these data to be utilized as the consent of persons.

4. REFERENCES

- [1] H. Lei, F. Ganjeizadeh, P. Jayachandran and P. Ozcan, "A statistical analysis of the effects of Scrum and Kanban on software development projects", 2016.
- [2] "HOW-TOS - What are the minimum system requirements?", *Vive.com*, 2016. [Online]. Available: https://www.vive.com/eu/support/category_howto/839576.html. [Accessed: 18-Nov- 2016].