# PROJECT REQUIREMENTS

## FeelSpeak : Generating Emotional Speech with Deep Learning

***Submitted by:***


| | |
|---|---|
| **M H SOHAN** | **PES1UG20CS235** |
| **RAHUL ROSHAN G** | **PES1UG20CS320** |
| **ROHIT ROSHAN** | **PES1UG20CS355** |
| **S M SUTHARSAN RAJ** | **PES1UG20CS362** |

**Prof. V R Badri Prasad**
Associate Professor
PES University


**January-May 2023**


**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**
FACULTY OF ENGINEERING
**PES UNIVERSITY**
(Established under Karnataka Act No. 16

# PROJECT REQUIREMENTS

## TABLE OF CONTENTS

# PROJECT REQUIREMENTS

## 1. Introduction

The purpose of this document is to define the requirements for the development of a system that can generate emotional speech from text. The system will utilize deep learning models for text preprocessing, emotion classification, and prosody modeling. The system will take a user-inputted text and output speech with appropriate emotional expression. The system will be designed to be integrated into existing virtual assistant applications or other speech-based applications that require emotional expression in speech.

### 1.1. Project Scope

The purpose of this project is to develop a system that can generate emotional speech from given text inputs. The system will use deep learning techniques to understand the context and emotions conveyed in the text, and then generate corresponding emotional speech in a natural-sounding way.

The benefits of this system include providing a new tool for individuals who have difficulty expressing their emotions verbally, as well as enhancing the emotional expressiveness of various applications such as virtual assistants, chatbots, and video games.

The objectives of the project are to develop a machine learning model that can accurately identify and classify the emotions conveyed in text, to develop a text-to-speech synthesis system that can generate emotional speech with natural-sounding intonation, and to integrate these two components into a working prototype.

The scope of the project is limited to the English language and a set of predetermined emotions such as happiness, sadness, anger, and neutral. The system will be designed to work on a desktop or laptop computer with a standard microphone and speakers, and will not require any specialized hardware or software. The accuracy of emotion classification and the quality of generated speech will be the primary focus of this project.

## 2. Product Perspective

The " FeelSpeak : Generating Emotional Speech " system is a software product that is intended to be used as a tool for generating emotional speech from text. The system is designed to be integrated into existing text-to-speech systems or used as a standalone tool for generating speech. The product is designed to provide users with a more engaging and emotionally expressive experience when interacting with speech-enabled systems, such as virtual assistants or customer service chatbots, text editors, etc…

# PROJECT REQUIREMENTS

The product is being developed specialized techniques of natural language processing and speech technology. The system utilizes advanced machine learning algorithms to analyze text and generate emotional speech in real-time. The system is being developed with the goal of improving the user experience and increasing the effectiveness of speech-enabled systems across a variety of industries, including healthcare, customer service, and entertainment.

## 2.1. Product Features

The major features of the product include:

1. Text-to-Speech Conversion: The system can convert written text into emotional speech using various emotions such as happiness, sadness, anger, etc.
2. Emotion Selection: The user can select the emotion they want the speech to convey using a dropdown menu or other interface options.
3. Voice Selection: The user can select the voice they want the speech to be spoken in using a dropdown menu or other interface options.
4. Audio File Export: The system allows the user to export the generated speech as an audio file in popular formats such as MP3, WAV, etc.
5. Text Input: The user can input the text they want to be converted into emotional speech using a text box or other interface options.
6. Emotion Detection: The system can detect the emotion in the input text and suggest appropriate emotions for the speech output.
7. Multi-language Support: The system supports multiple languages for input text and speech output.
8. User Management: The system allows for user management features such as user registration, login, and profile management in the text editor.
9. Security and Privacy: The system implements security and privacy features to ensure user data and interactions are protected.
10. API Integration: The system can integrate with third-party APIs for additional features and functionalities.

## 2.2. Operating Environment

The operating environment for the system includes the following components:

- Hardware platform: The system can be run on any standard computer hardware with a processor, memory, and storage space capable of running the required software.

- Operating system: The system will be developed and tested on Windows and Linux operating systems. It will be compatible with Windows 7 or later, and Linux distributions such as Ubuntu, Fedora, and CentOS.

- Software components: The system requires the following software components to be installed:

  Python 3.6 or later, TensorFlow 2.0 or later, NumPy, Pandas, NLTK, Flask

In addition, the system requires an internet connection to access cloud-based speech synthesis services.

### 2.3. General Constraints, Assumptions and Dependencies

The following are the general constraints, assumptions, and dependencies for the system:

- Hardware Limitations: The system requires a computer with a minimum configuration of a quad-core processor, 8GB RAM, and 1GB free disk space.

- Operating System: The system will be developed and tested on Windows 10 (64-bit) operating system.

- Robust Dataset : This project requires a dataset which has labeled text with emotion and a corresponding speech dataset with emotion. Also, an unlabeled text dataset with emotion is required for validations.

- Assumptions: The system assumes that the input text is in the English language and contains no grammatical errors.

- Interface Dependencies: The system will use external libraries for speech synthesis and natural language processing.

- Safety and Security Considerations: The system should follow all the safety and security considerations related to data privacy and security.

### 2.4. Risks

# PROJECT REQUIREMENTS

As with any software development project, there are risks involved with the development of the proposed system. The following risks have been identified:

- Technical Risks: There is a risk that the system may not be able to generate accurate emotional speech from text due to the complexity of the natural language processing algorithms involved. This risk can be mitigated by thorough testing and refining of the algorithms used.

- Schedule Risks: There is a risk of delays in the project schedule due to unexpected technical issues or changes in project requirements. This risk can be mitigated by setting realistic timelines, regularly monitoring progress, and having contingency plans in place.

- Resource Risks: There is a risk of inadequate resources (such as hardware, software, or personnel) that may impact the project's success. This risk can be mitigated by careful planning, as well as having backup resources in case of emergencies.

- Security Risks: There is a risk that the system may be vulnerable to security threats, such as data breaches or hacking attempts. This risk can be mitigated by implementing strong security measures, such as encryption and access controls, and regularly monitoring the system for any suspicious activity.

## 3. Functional Requirements

Sure, here are some functional requirements based on the given parameters:

• Validity tests on inputs:

The system shall validate the input text to ensure that it is in the English language.
The system shall check the length of the input text to ensure that it is within the acceptable limits for the model. The system shall verify that the input text does not contain any profanity or offensive language.

• Sequence of operations:

- The system shall first preprocess the input text to remove any unnecessary characters, punctuations, or digits.
- The system shall then tokenize the preprocessed text into words and phrases.
- The system shall apply the emotional embedding algorithm to each token to generate emotional features.
- The system shall use the emotional features to generate emotional speech signals using the TTS system.

• Error handling and recovery:

The system shall generate an error message if the input text is not in English.
The system shall provide an error message if the input text is too long.
The system shall provide an error message if the emotional embedding algorithm fails to generate emotional features for a token. The system shall provide a fallback option to use a default emotion or neutral emotion if emotional speech signals cannot be generated for a given input.

• Consequences of parameters:

The system shall adjust the emotional intensity of the speech signal based on the specified emotional parameters (e.g., happy, sad, angry, etc.).
The system shall vary the speaking rate of the generated speech signal based on the emotional parameters.

• Relationship of outputs to inputs:

The system shall generate emotional speech signals that match the emotional content of the input text.

The system shall provide an option to output the emotional speech signal in different audio formats.

## 4. External Interface Requirements

### 4.1. User Interfaces

The system will have a user interface that allows users to input the text for which they want to generate emotional speech. The user interface will consist of a text box where users can enter the text and a button to initiate the speech generation process. The user interface will follow standard GUI design principles, including consistent layouts, color schemes, and font sizes.

The system will display the generated speech output to the user in the form of an audio file. The audio file will be played back using the user's default audio player, and the system will not provide any playback controls or options. The audio file will be available for download to the user's device.

In case of errors, the system will display appropriate error messages to the user in a separate pop-up window. The error messages will provide detailed information about the error, including possible solutions or workarounds. The system will also provide a help button that users can click to access a user manual or other resources.

### 4.2. Hardware Requirements

The hardware requirements for the "Generation of emotional speech from text" system are as follows:

Computer or mobile device with at least 8GB of RAM and a dual-core processor.
Operating System: Windows 10 or later, MacOS, or Linux.
Sound card and microphone for recording audio.
Speakers or headphones for audio playback.
Internet connection for downloading and installing required software packages.

The system should be compatible with a range of hardware devices, including laptops, desktops, and mobile devices. The system should support standard audio input/output devices and protocols, such as 3.5mm audio jacks, USB, and Bluetooth. The system should also be capable of using the internet connection to access and use cloud-based resources if required.

### 4.3. Software Requirements

Since the project is about developing a software system, there are no software requirements beyond the system itself. However, it can be assumed that the software requirements for this system include:

• Name and Description: The software system for generating emotional speech from text
• Version / Release Number: Initial release version 1.0
• Databases: No specific database requirements have been mentioned
• Operating Systems: The system should be compatible with major operating systems, including Windows, Linux, and macOS.
• Tools and libraries: The system may require various tools and libraries to generate speech from text, such as text-to-speech (TTS) engines, natural language processing (NLP) libraries, and audio processing libraries.
• Source (if any): The source code for the software system may be provided to the client or maintained by the development team, depending on the agreement between the parties involved.

### 4.4. Communication Interfaces

Audio Input/Output Interface: The system must have an interface to capture audio input from the user, and also generate emotional speech output to be played through the speakers or headphones. The interface must support standard audio formats, such as WAV or MP3.

Network Interface: If the system is designed to work in a client-server architecture, it should have an interface to communicate with the server over a network. This interface must support standard network protocols, such as TCP/IP or HTTP.

Text Input/Output Interface: The system must have an interface to capture text input from the user and generate emotional speech output based on the input text. The interface must support standard text formats, such as plain text or XML.

External API Interface: The system may need to integrate with external APIs, such as speech recognition or natural language processing APIs. The interface must be compatible with the API protocols and specifications.

## 5. Non-Functional Requirements

### 5.1. Performance Requirement

The system shall generate emotional speech from text within 2 seconds of receiving the input.
The system shall be able to handle a minimum of 1000 concurrent users without a decrease in performance.
The system shall be available for use 24/7 with an uptime of at least 99%.
The system shall have a maximum error rate of 1% in generating emotional speech from text.
The system shall be able to process text inputs of up to 1000 characters in length.
In terms of quality attributes, some possible non-functional requirements could include:

The system shall be reliable and have a mean time between failures (MTBF) of at least 500 hours.
The system shall be robust and able to handle unexpected inputs and errors without crashing or corrupting data.
The system shall have a user-friendly interface with clear and concise error messages.
The system shall be secure, with all user data and interactions encrypted and stored in a secure database.
The system shall be scalable and able to handle increased usage and functionality without significant changes to the underlying architecture.

### 5.2. Safety Requirements

As the product involves generation of emotional speech from text, there are no safety requirements to be addressed in the product.

### 5.3. Security Requirements

- Authentication

  The system must require users to provide valid login credentials before allowing access to the system. Passwords must meet complexity requirements and be stored securely.

- Authorization

  The system must enforce role-based access control to ensure that users can only access functionality and data that is relevant to their role.

- Data Privacy

The system must comply with applicable data privacy regulations and protect sensitive data from unauthorized access. Any personal or sensitive data collected by the system must be encrypted in transit and at rest.

- Security Auditing

  The system must maintain an audit trail of all user actions, including login attempts, data access, and any changes made to the system configuration. The audit trail must be tamper-evident and securely stored.

- Incident Response

  The system must have procedures in place to detect and respond to security incidents. This includes notifying system administrators and users of any suspected breaches, and taking appropriate steps to contain and mitigate the impact of any security incidents.

## 6. Other Requirements

- Ethical Considerations
  The system must comply with ethical considerations in generating emotional speech. The system should not be used to generate hate speech or any other form of speech that may cause harm to others. It should also respect the privacy and consent of the user.

- Data Protection Requirements
  The system must comply with data protection requirements. It should not collect or store any personal data of the user without their explicit consent. The system should also provide an option for users to delete their data.

- Accessibility Requirements
  The system must comply with accessibility requirements. It should be designed to be accessible to users with disabilities, such as providing options for text-to-speech or audio descriptions. The system should also be user-friendly and easy to use for all users.

- Compatibility Requirements
  The system must be compatible with different devices and operating systems to ensure accessibility to a wide range of users. It should be able to function seamlessly across different platforms, including desktop and mobile devices.

## Appendix A: Definitions, Acronyms and Abbreviations

Definitions:

Emotional speech: Speech that conveys an emotion, such as happiness, sadness, anger, or fear.
Text-to-speech (TTS) synthesis: The process of generating speech from written text.
Natural language processing (NLP): A branch of artificial intelligence that focuses on making computers understand human language.
Pitch: The perceived highness or lowness of a sound, determined by the frequency of sound waves.
Prosody: The rhythm, stress, and intonation of speech that convey meaning beyond the words themselves.
Acronyms:

SRS: Software Requirements Specification
API: Application Programming Interface
GUI: Graphical User Interface
DSP: Digital Signal Processing
ML: Machine Learning
NN: Neural Network
Abbreviations:

Hz: Hertz (unit of frequency)
dB: Decibel (unit of sound intensity)
LSTM: Long Short-Term Memory (a type of neural network)
GPU: Graphics Processing Unit

**Appendix B: References**

[Provide the list of the documents or web addresses to which the Requirement Specification refers. It may include user interface style guides, standards, system requirements specification and use cases. The reference documents shall describe the title, version number, dates, authors and publishers, whatever is applicable.]