**Text-To-Speech (TTS)**

**Mini Project report submitted to**

Text, logo

Description automatically generated with medium confidence

by

**Kartik Mittal**

Registration No: 215890116

Department of Computer Science and Engineering

Manipal Institute of Technology

Bengaluru

Under the supervision of

|  |  |
| --- | --- |
| **Dr. Simi VR**  Assistant professor (senior scale)  Department of Computer Science and Engineering  Manipal Institute of Technology  Bengaluru | **Dr. Deepti Gupta**  Assistant professor (senior scale)  Department of Computer Science and Engineering  Manipal Institute of Technology  Bengaluru |

**Manipal Institute of Technology**

**Bengaluru Campus-560064, Karnataka, India.**

**Certificate**

This is to certify that the **Mini Project report** entitled “**Text-To-Speech (TTS)**” submitted by **Kartik Mittal (Registration No: 215890116),** as the record of the mini project work carried out by them, is accepted as the Mini Project reportsubmission in partial fulfilment of the requirements for the award of degree of Bachelor of Technology (BTech).

|  |  |
| --- | --- |
| **Examiners Name** | **Signature with the Date** |
| ……………………………………………... | ……………………………… |
| ……………………………………………... | ……………………………… |
| ……………………………………………... | ……………………………… |

**Abstract**

A Text-to-speech synthesizer is an application that converts text into spoken word, by analysing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple web application that converts inputted text into synthesized speech and reads out to the user which can then played with different voices, speech and rate. The development of a text to speech synthesizer will be of great help to people with visual impairment and make making through large volume of text easier or [text-to-speech](https://speechify.com/text-to-speech-online/?landing_url=https%3A%2F%2Fspeechify.com%2Fblog%2Fhistory-of-text-to-speech%2F) services to listen to books, study or proofread your written work.

|  |  |  |
| --- | --- | --- |
| **Contents** | | **Page. No.** |
| **List of Figures**……………………………………………………… | | **5** |
| **1** | Introduction………………………………………………….. | **6** |
| **2** | Problem Statement…………………………………………... | **7** |
| **3** | Objective…………………………………………………….. | **8** |
| **4** | Methodology……………………………………………….... | **9** |
| **5** | Implementation Details…………………………………….... | **12** |
| **6** | Results……………………………………………………….. | **18** |
| **7** | Conclusion…………………………………………………... | **19** |
| References…………………………………………………………... | | **20** |

**List of Figures**

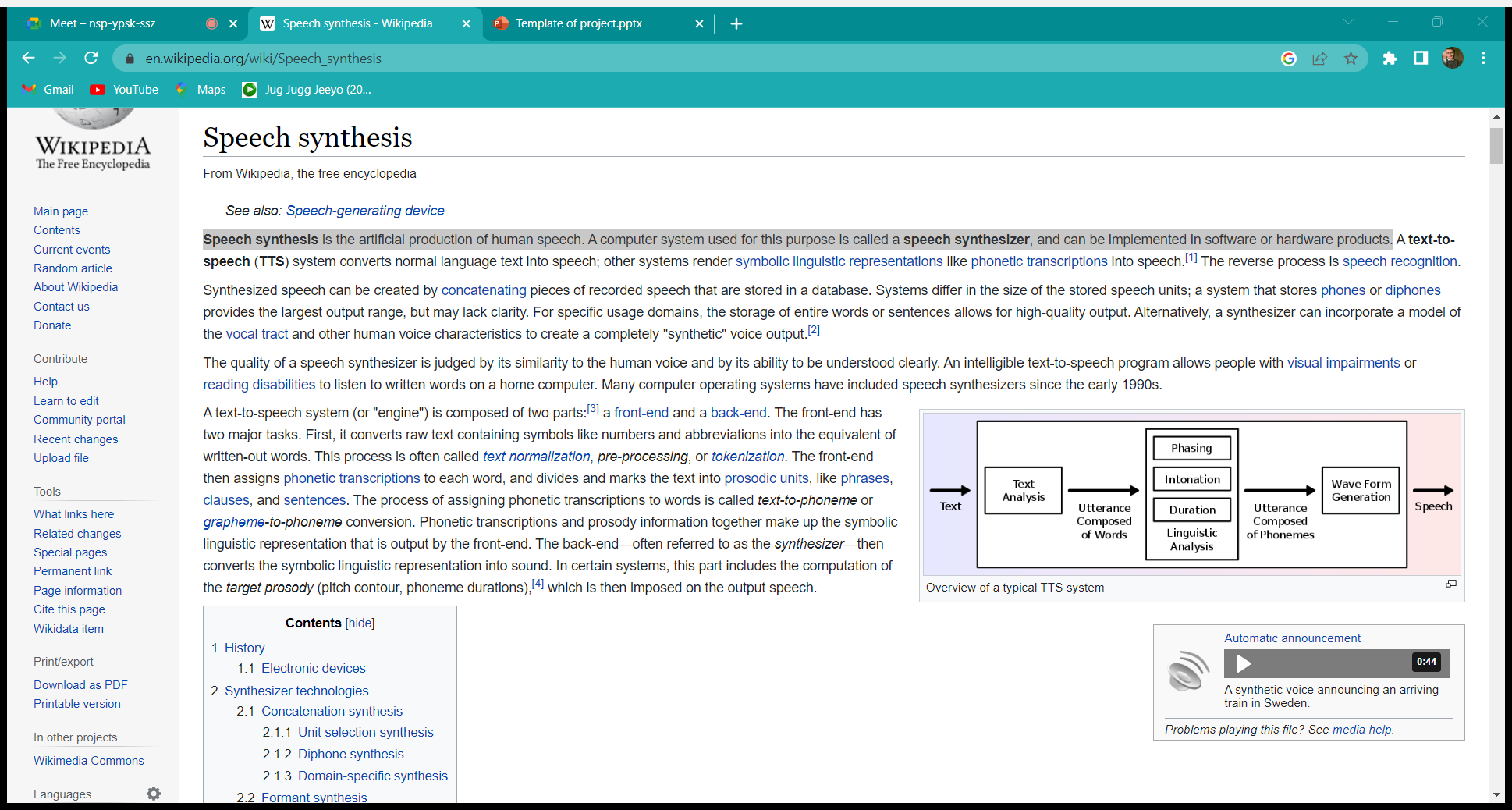
|  |  |
| --- | --- |
|  | **Page No.** |
| **Overview of a typical TTS system** 1………………………………………………………. | 6 |
| **Overview of a typical TTS system 2**………………………………………………………. | 9 |
| **Figure 1**………………………………………………… | 12 |
| **Figure 2**………………………………………………… | 12 |
| **Figure 3**………………………………………………… | 13 |
| **Figure 4**………………………………………………… | 13 |
| **Figure 5**………………………………………………… | 14 |
| **Figure 6**………………………………………………… | 14 |
| **Figure 7**………………………………………………… | 15 |
| **Figure 8**………………………………………………… | 15 |
| **Figure 9**………………………………………………… | 16 |
| **Figure 10**………………………………………………… | 16 |
| **Figure 11**………………………………………………… | 17 |
| **Figure 12**………………………………………………… | 17 |
| **Figure 13**………………………………………………… | 18 |
|  |  |
|  |  |
|  |  |

**Chapter 1**

**Introduction**

A **text-to-speech** (**TTS**) system converts normal language text into speech; other systems render [symbolic linguistic representations](https://en.wikipedia.org/wiki/Symbolic_linguistic_representation) like phonetic transcriptions into speech.[[1]](https://en.wikipedia.org/wiki/Speech_synthesis) The reverse process is [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition).

**Speech synthesis** is the artificial production of human [speech](https://en.wikipedia.org/wiki/Speech). A computer system used for this purpose is called a **speech synthesizer**, and can be implemented in [software](https://en.wikipedia.org/wiki/Software) or [hardware](https://en.wikipedia.org/wiki/Computer_hardware) products.



Overview of a typical TTS system 1

**Chapter 2**

**Problem statement**

Speech synthesis, or the artificial production of the human voice, has come a long way over the last 70 years. Whether you currently use [text-to-speech](https://speechify.com/text-to-speech-online/?landing_url=https%3A%2F%2Fspeechify.com%2Fblog%2Fhistory-of-text-to-speech%2F) services to listen to books, study or proofread your written work, there’s no doubt that text-to-speech services have made life easier for people in a variety of professions or any disabled person. Hence, a basic Text-To-Speech web app would be made so that it can help many people, using HTML and CSS for frontend and JavaScript for the working of web app using Web Speech API.

**Chapter 3**

**Objectives**

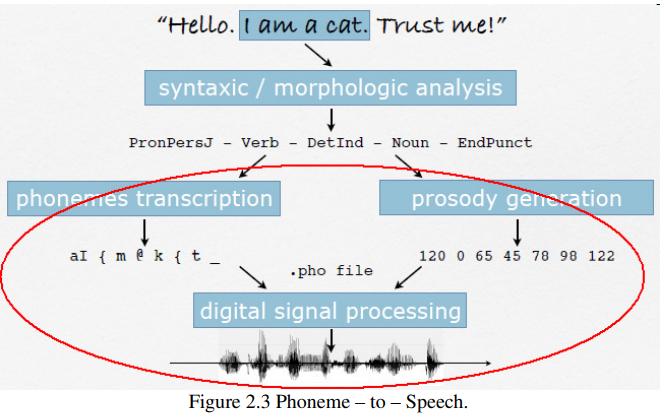
* To develop a web application using HTML/CSS/JS for implementing Text-To-Speech.

**Chapter 4**

**Methodology**

**Working of the API: Behind the scene.**

* A Text–to–Speech system (or “engine”) is composed of two main parts: Texts–to–Phoneme (Natural Language Processing, NLP) and Phoneme–to– Speech (Digital Signal Processing, DSP).
* Texts–to–Phoneme: Also called a Grapheme–to–Phoneme conversion, the process of assigning phonetic transcription to words. The text must be converted into a linguistic representation that includes the phonemes to be produced, their duration, 11 the location of phrase boundaries, and the pitch / frequency contours for each phrase.
* Phoneme–to–Speech: The Phonetic transcription and prosody information obtained in the linguistic analysis stage are converted into an acoustic waveform.



Overview of a typical TTS system 2

Files used:

1. index.html
2. style.css
3. script.js

**In the JavaScript code**,

First, I got the user text and call a function textToSpeech() with passing user text as an argument. Inside this function, using the speech synthesis property of the window object, I converted the entered text to speech. Speech Synthesis is a web speech API that controls the speech service.

After this, I got all available voices from the user device using the getVoices() method of Speech Synthesis and insert it into HTML select tag.

There are 6 properties of web speech API which we can tweak from and I used 5 of them,

1. **Text**

The text property gets and sets the text that will be synthesized when the utterance is spoken. The text can be provided as plain text.

1. **Volume**

The volume property gets and sets the volume of the utterance. It is a float that represents the volume value, between 0 (lowest) and 1 (highest). The default value is 1 if this property is unset.

1. **Rate**

The rate property gets and sets the rate of the utterance. It is a float representing the rate value which can range between 0.1 (lowest) and 10 (highest). The default value is 1 if this property is unset.

1. **Pitch**

The pitch property gets and sets the pitch of the utterance. It is a float representing the pitch value that can range between 0 (lowest) and 2 (highest). The default pitch is 1 if this property is unset.

1. **Voices**

The voice property gets and sets the voice that will be used to speak the utterance. This should be set to one of the [SpeechSynthesisVoice](https://developer.mozilla.org/en-US/docs/Web/API/SpeechSynthesisVoice) objects. If it is not set, the most suitable default voice available for the utterance’s language setting will be used.

**In HTML file,**

* An empty select menu. We will fill the empty select menu with the list of voices available using JavaScript.
* Range sliders for volume, pitch, and rate.
* A text area to type in.
* Control button for speech.

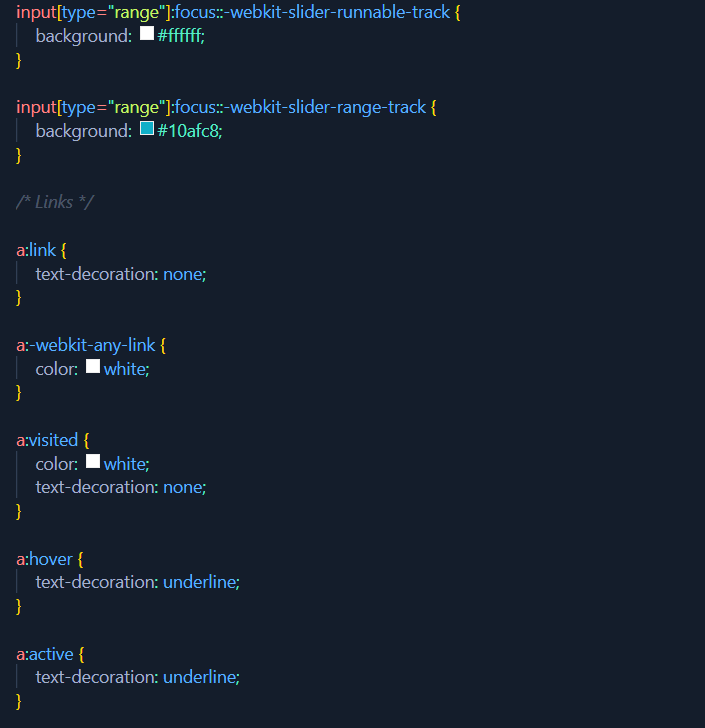
**In CSS file,**

All the design elements are implemented to improve the look and feel of the whole webpage.

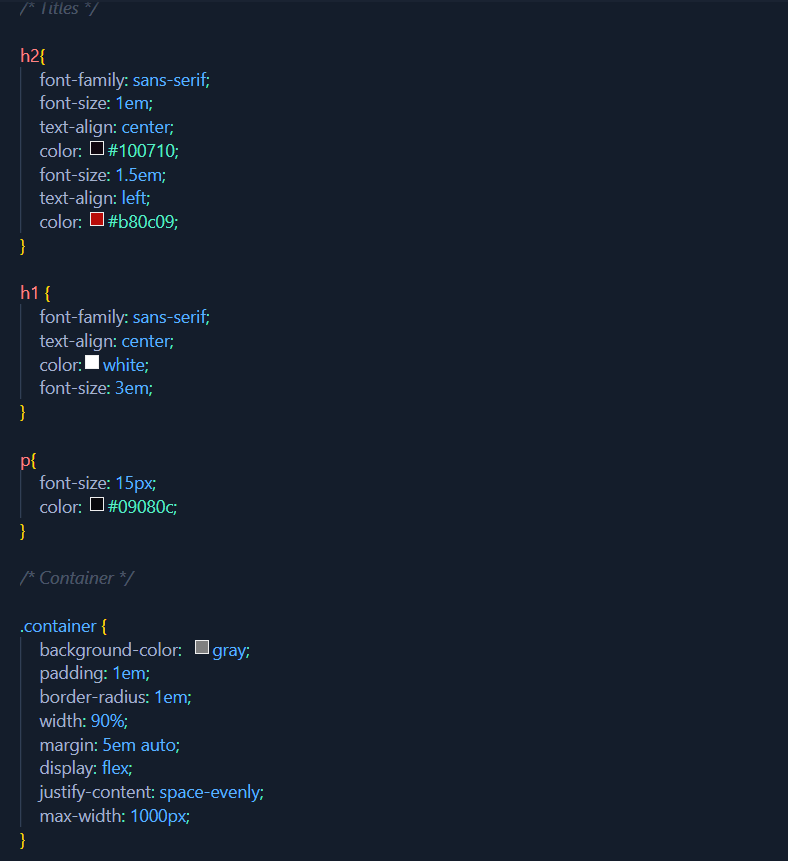
**Chapter 5**

**Implementation Details**

**CSS Code snips:**



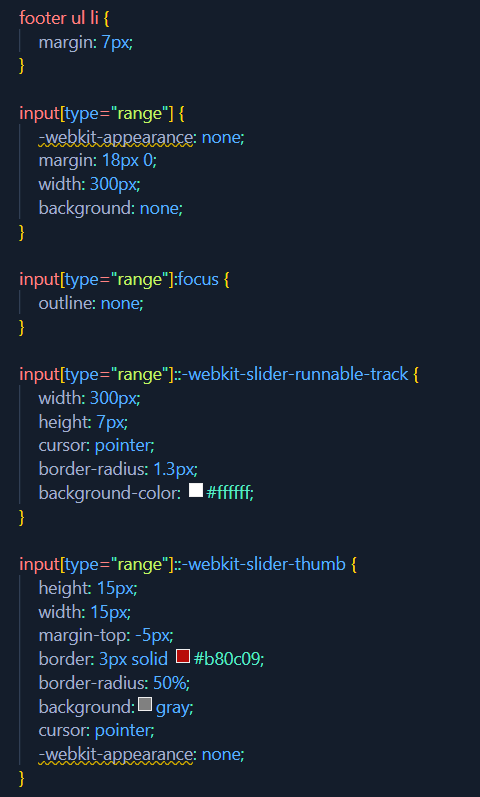
Figure



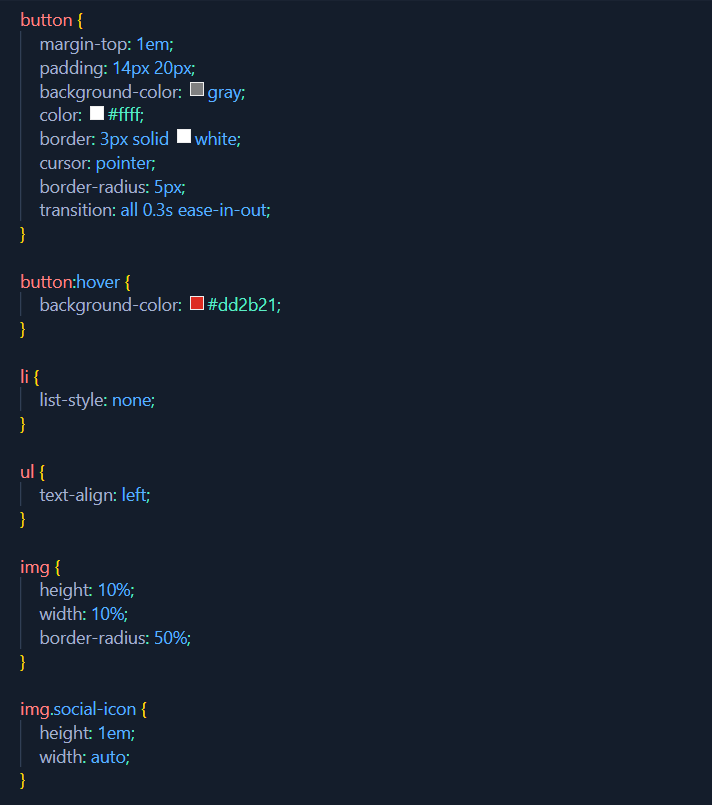
Figure



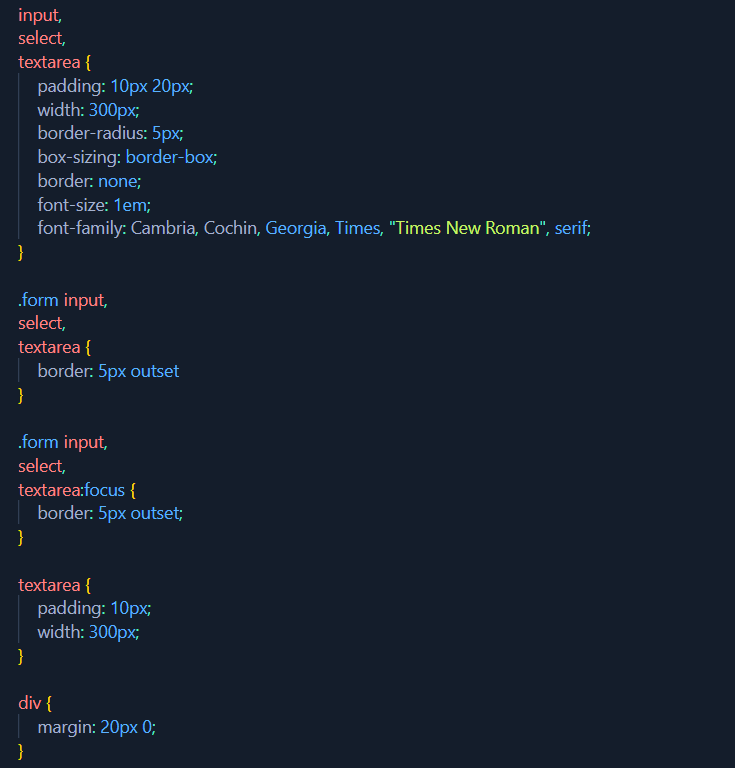
Figure



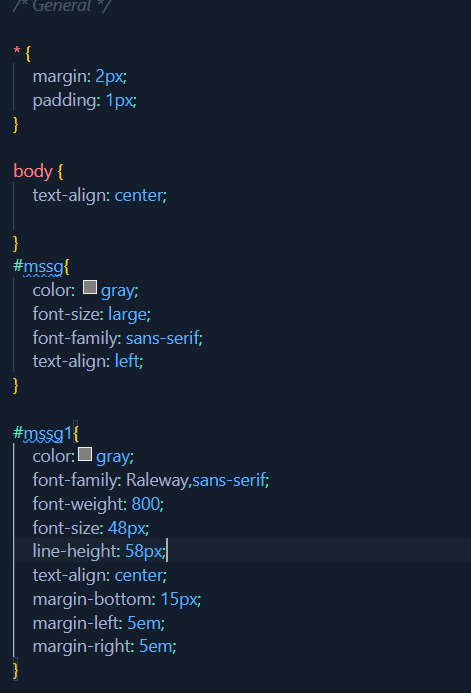
Figure



Figure



Figure



Figure

**HTML Code snips**

****

Figure

****

Figure

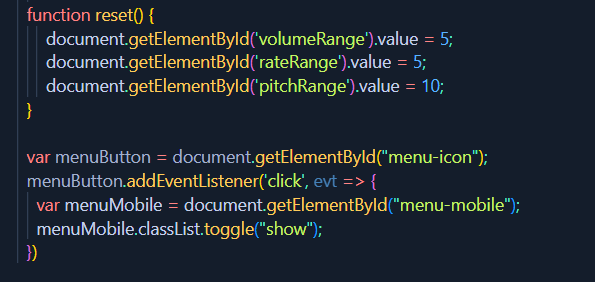
****

Figure

**JS Code snips:**



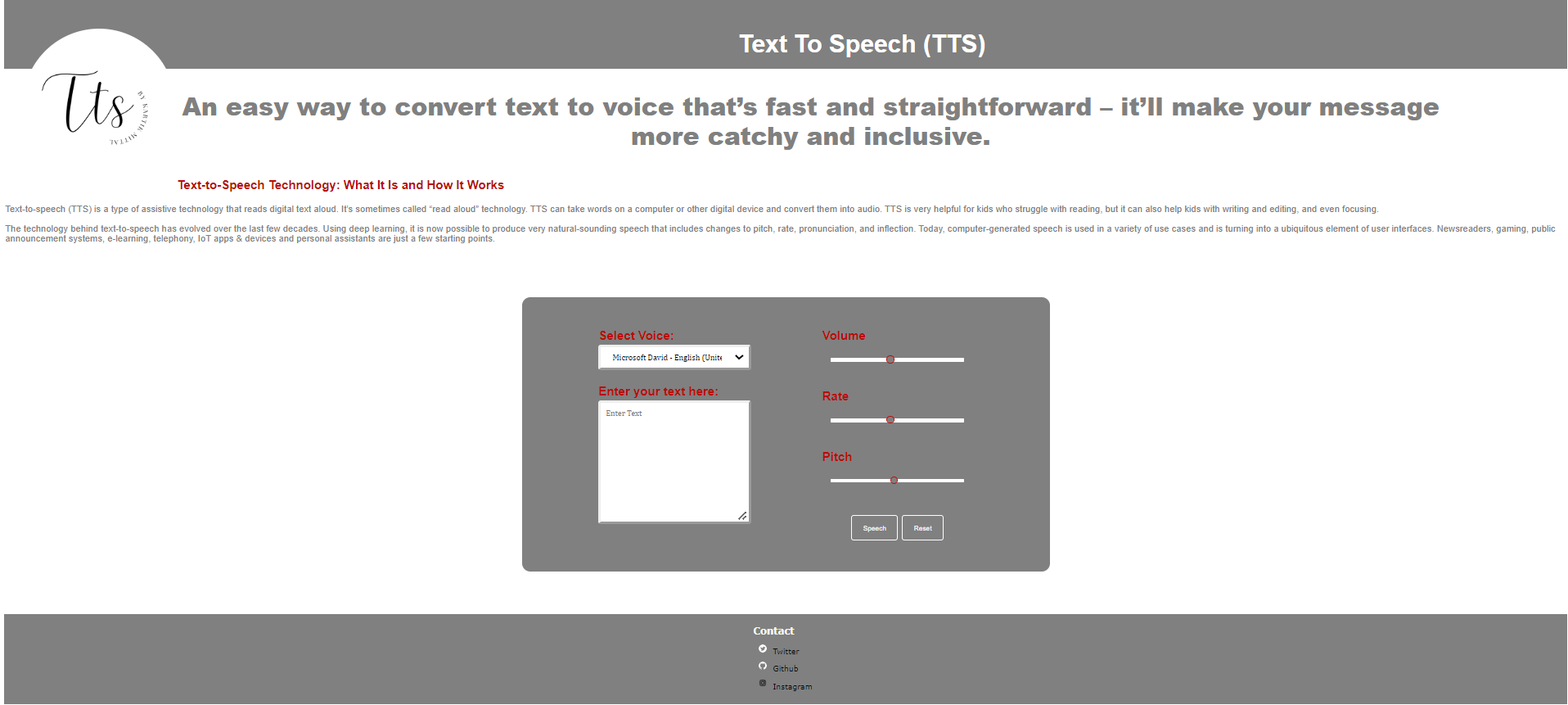
Figure



Figure

**Chapter 6**

**Results**



Figure

**Chapter 7**

**Conclusion**

Hence, we have successfully implemented Web app project of Text-To-Speech using HTML/CSS/JS by integrating Web speech API.

**References**

* <https://www.theseus.fi/bitstream/handle/10024/39427/Thesis_TextToSpeech_software_comparison_by_YingZheng.pdf;jsessionid=4E34D13EC1DD191B3FA314D123AEDCEA?sequence=1>
* [A Short History Of Text-to-Speech | Speechify](https://speechify.com/blog/history-of-text-to-speech/?landing_url=https%3A%2F%2Fspeechify.com%2Fblog%2Fhistory-of-text-to-speech%2F)