# Text To Speech - Chrome Extension

**BE-ICT Semester- VII**

Prepared at


ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

**Bhaskaracharya National Institute for Space Applications & Geoinformatics**
**Ministry of Electronics and Information Technology, Govt. of India.**

**Gandhinagar**

**Prepared By**

**Bhakti Chokshi** **Bhavya Adhiya** **Hasti Sutaria**

**Jaahanava Joshi** **Nupur Khatri**

**Guided By:**

**DR. Hitesh Chhinkaniwala**
**AIIE, Ahmedabad**

**External Guide:**

**Sidhdharth Patel**
**BISAG-N, Gandhinagar**

**Punit Lalwani**
**CISO**
**BISAG- N, Gandhinagar.**

**SUBMITTED TO**


**Gujarat Technological University**

**Adani Institute Of Infrastructure Engineering, Ahmedabad**

Bhaskaracharya National Institute for Space Applications and Geo-informatics

MeitY, Government of India

Phone: 079 - 23213081 Fax: 079 - 23213091

E-mail: info@bisag.gujarat.gov.in, website: https://bisag-n.in/

ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

# CERTIFICATE

*This is to certify that the project report compiled by* ***Ms. Bhakti Chokshi, Mr. Bhavya Adhiya, Ms. Hasti Sutaria, Mr. Jaahanava Joshi and Ms. Nupur Khatri*** *students of 7th Semester* ***BE - ICT from Adani Institute of Infrastructure Engineering, Gujarat Technological University, Ahmedabad*** *have completed their Summer internship project satisfactorily. To the best of our knowledge this is an original and bonafide work done by them. They have worked on Python based application for* ***"Text To Speech Conversion – Chrome Extension"*** *starting from June 20th, 2022 to July 20th, 2022.*

*During their tenure at this Institute, they were found to be sincere and meticulous in their work. We appreciate their enthusiasm & dedication towards the work assigned to them.*

*We wish them every success.*

**Punit Lalwani**

**CISO,**

**BISAG- N, Gandhinagar**

**T. P. Singh**

**Director General,**

**BISAG- N, Gandhinagar**

adani

Institute of Infrastructure

# Certificate of Completion

*This is to certify that the 7th Semester Summer Internship (3170001) Project entitled "**TEXT TO SPEECH CONVERSION - CHROME EXTENSION**" has been carried out by **BHAVYA ADHIYA, BHAKTI CHOKSHI, HASTI SUTARIA, JAAHANAVA JOSHI and NUPUR KHATRI** under my guidance in fulfilment of the degree of Bachelor of Engineering in INFORMATION & COMMUNICATION TECHNOLOGY ENGINEERING (7th Semester) of Adani Institute of Infrastructure Engineering - Ahmedabad during the academic year 2022-23.*

**Internal Guide**

**Dr. Hitesh Chhinkaniwala**

**Head of the Department**

**Dr. Hitesh Chhinkaniwala**

# About BISAG- N

## ABOUT THE INSTITUTE

Modern day planning for inclusive development and growth calls for transparent, efficient, effective, responsive and low cost decision making systems involving multi-disciplinary information such that it not only encourages people's participation, ensuring equitable development but also takes into account the sustainability of natural resources. The applications of space technology and Geo–informatics have contributed significantly towards the socio-economic development. Taking cognizance of the need of geo-spatial information for developmental planning and management of resources, the department of Ministry of Electronics and Information Technology, Government of India, established "Bhaskaracharya National Institute for Space Applications and Geo-informatics" (BISAG- N). BISAG- N is an ISO 9001:2008, ISO 27001:2005 and CMMI: 5 certified institute. BISAG- N which was initially set up to carryout space technology applications, has evolved into a centre of excellence, where research and innovations are combined with the requirements of users and thus acts as a value added service provider, a technology developer and as a facilitator for providing direct benefits of space technologies to the grass root level functions/functionaries.

## BISAG- N's Enduring Growth

Since its foundation, the Institute has experienced extensive growth in the sphere of Space technology and Geo-informatics. The objective with which BISAG- N was established is manifested in the extent of services it renders to almost all departments of the State. Year after year the institute has been endeavouring to increase its outreach to disseminate the use of geo-informatics up to grassroots level. In this span of nine years, BISAG- N has assumed multi-dimensional roles and achieved several milestones to become an integral part of the development process of the Gujarat State.

# BISAG- N Journey

## 2003-04

Gujarat SATCOM Network

## 2007-08

Centre for Geo-informatics Applications

## 2010-11

Academy of Geo-informatics for Sustainable Development

## 2012-13

A full-fledged Campus

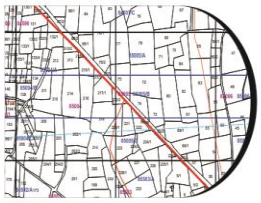
# Activities

## Satellite Communication..

for promotion and facilitation of the use of broadcast and teleconferencing networks for distant interactive training, education and extension.

## Remote Sensing..

for Inventory, Mapping, Developmental planning and Monitoring of natural & man-made resources.

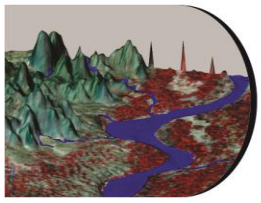


## Geographic Information System..

for conceptualization, creation and organization of multi purpose common digital database for sectoral/integrated decision support systems.

## Global Navigation Satellite System..

for Location based Services, Geo-referencing, Engineering Applications and Research.



## Photogrammetry..

for Creation of Digital Elevation Model, Terrain Characteristic, Resource planning.

## Cartography..

for thematic mapping, value added maps.

## Software Development..

for wider usage of Geo-spatial applications, Decision Support Systems (desktop as well as web based), ERP solutions.

## Education, Research and Training..

for providing Education, Research, Training & Technology Transfer to large number of students, end users & collaborators.

# Applications of Geospatial Technology for Good Governance: Institutionalization

Through the geospatial technology, the actual situation on the ground can be accessed. The real life data collected through the technology forms the strong foundation for development of effective social welfare programs benefiting directly the grass root level people. The geospatial data collected by the space borne sensors along with powerful software support through Geographic Information System (GIS), the vital spatio-temporal maps, tables, and various statistics are being generated which feed into Decision Support System (DSS).

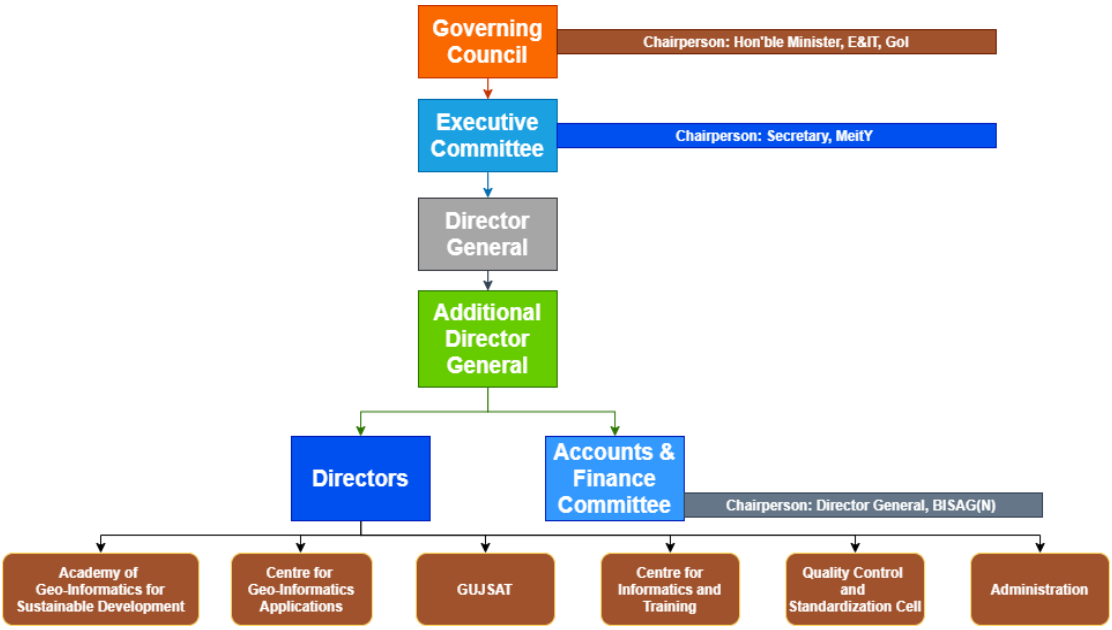
A multi-threaded approach is followed in the process of institutionalization of development of such applications. The 5 common threads which run through all the processes are: *Acceptability, Adaptability, Affordability, Availability and Assimilability*.
These are the "Watch Words" which any application developer has to meet. The "acceptability" addresses the issue that the application developed has met the wide acceptability among the users departments and the ultimate end beneficiary by way of providing all necessary data and statistics required. The "affordability" addresses the issue of the application product being cost effective. The "availability" aspect looks into aspect of easily accessible across any platform, anywhere and anytime. The applications should have inbuilt capability of easy adaptability to the changing spatio- and temporal resolutions of data, new aspects of requirements arising from time to time from users. The assimilability aspect ensures that the data from various sources / resolutions and technologies can be seamlessly integrated.

| | |
|---|---|
| **ACCEPTABILITY** | ▪ Problem definition by users<br>• Proof of Concept development without financial liability on users<br>▪ Execution through collaboration under user's ownership |
| **ADOPTABILITY** | ▪ Applications as per present systems & database<br>▪ Maximum Automation<br>▪ Minimum capacity building requirement at the user end |
| **AFFORDABILITY :** | ▪ Multipurpose geo-spatial database, common, compatible, standardized (100s of layers)<br>▪ In house developed/open source software<br>▪ Full Utilization of available assets |
| **AVAILABILITY:** | ▪ Departmental /Integrated DSS<br>▪ Desired Product delivery anytime, anywhere in the country |
| **ASSIMILABILITY** | ▪ Integration of Various technologies like RS, GIS, GPS, Web MIS, Mobile etc. |

## Organizational Setup

The Institute is responsible for providing information and technical support to different Departments and Organizations. The Governing Body and the Empowered Executive Committee govern the functioning of BISAG- N. The Institute is registered under the Societies Registration Act 1860. Considering the scope and extent of activities of BISAG- N, its organizational structure has been charted out with defined functions.

## Organizational Setup of BISAG- N




## Governing Body

For smoother, easier and faster institutionalization of Remote Sensing and GIS technology, decision makers of the state were brought together to form the Governing Body. It is the supreme executive authority of the Institute. The Governing Body comprises of ex-officio members from various Government departments and Institutes.

- Hon'ble Minister of Electronics and Information Technology..... Chairperson (Ex-Officio)
- Hon'ble Minister of State Electronics and Information Technology.....……Deputy Chairperson (Ex-Officio)
- Secretary of Government of India: Ministry of Electronics and Information Technology..... .....Executive Vice Chairperson (Ex-Officio)
- Chief Executive Officer, Niti Aayog..... ……Member (Ex-Officio)
- Chairman, Indian Space Research Organization ..... ……Member (Ex-Officio)
- Secretary to Government of India: Department of Science and Technology ..... ……Member (Ex-Officio)
- Additional Secretary to Government of India: Ministry of Electronics and Technology.. ……Member (Ex-Officio)
- Chief Secretary to Government of Gujarat..... ……Member (Ex-Officio)
- President & Chief Executive Officer, National e-Governance Division, Ministry of Electronics
- and Information Technology..... …… Member (Ex-Officio)
- Financial Advisor to Government of India: Ministry of Electronics and Information Technology…Member (Ex-Officio)
- Distinguished Professionals from the GIS field-Three (3) (To be nominated by the Chairperson)
- Director-General, Bhaskaracharya National Institute for Space Application and Geo-Informatics {BISAG(N)} ..... Member Secretary (Ex-Officio)

# Centre for Geo-informatics Applications

## Introduction

The objective of this technology group is to provide decision support to the sectoral stake holders through scientifically organized, comprehensive, multi-purpose, compatible and large scale (village level) geo-spatial databases and supporting analytical tools. These activities of this unit are executed by a well-trained team of multi-disciplinary scientists. The government has provided a modern infrastructure along with the state-of-the-art hardware and software. To study the land transformation and development over the years, a satellite digital data library of multiple sensors of last twenty years has been established and conventional data sets of departments have been co-registered with satellite data. The geo-spatial databases have been created using conventional maps, high resolution satellite 2D and 3D imagery and official datasets (attributes).The geo-spatial databases include terrain characteristics, natural and administrative systems, agriculture, water resources, city survey maps, village maps with survey numbers, water harvesting structures, water supply, irrigation, power, communications, ports, land utilization pattern, infrastructure, urbanization, environment data, forests, sanctuaries, mining areas, industries. They also include social infrastructure like the locations of schools, health centres, institutions, aganwadies, local government infrastructure etc. The geospatial database of nagar-palikas includes properties and amenities captured on city and town planning maps with 1000 GIS layers. Similar work for villages has been initiated as a pilot project.

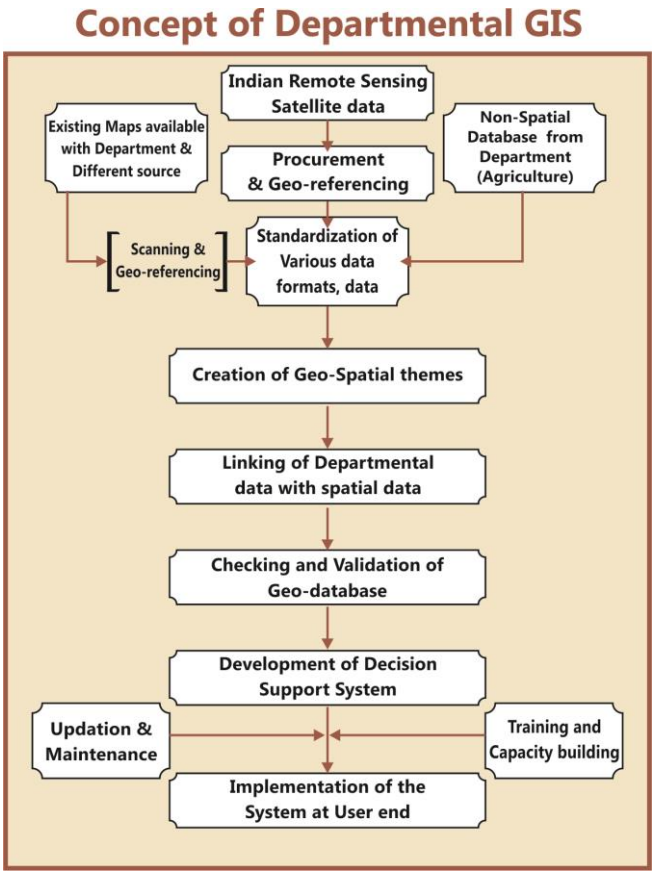
The applications of space technology and geo-informatics have been operational in almost all the development sectors of the state. Remote sensing and GIS applications have provided impetus to planning and developmental activities at grass root level as well as monitoring and management in various disciplines.

## The GIS based Applications Development

The GIS software is a powerful tool to handle, manipulate and integrate both the spatial and non-spatial data. The GIS system operates on the powerful backend data base and Sequential Query Language (SQL) to inquiry the data bases. It has the capability to handle large volume of data and process to yield values of parameters which can be input to very important government activity as Decision Support System (DSS). Its mapping capabilities help the users and specialists in generating single and multi-theme wise maps.

The GIS based applications development has been institutionalized in BISAG- N. This process can be listed as (Refer Figure for Details)

- Making the users aware of the GIS capabilities through introductory training programme and by exposing to already developed projects as success stories.
- Helping the users in defining the GIS based projects.
- Digitizing the data available with the users and encouraging them to collect any additional data as may be required.
- Generating the appropriate data bases with the full involvement of the users following the data bases standards

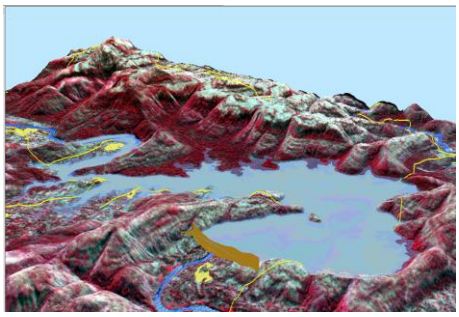



## Remote Sensing and GIS Sectoral Applications:

**Geo-informatics based Irrigation Management and Monitoring System**

- The Geo-spatial information system for Irrigation water Management and Monitoring system for command areas in Sardar Sarovar Narmada Nigam Limited (SSNL) has been developed. Satellite image-based Irrigation monitoring system has been developed in GIS. From the multi-spectral Satellite images of every month, the irrigated areas were extracted.
- The irrigated area were overlaid on the geo-referenced cadastral maps and the statistics of area irrigated has been estimated.
- The user friendly Customized Decision Support System (DSS) has been developed.

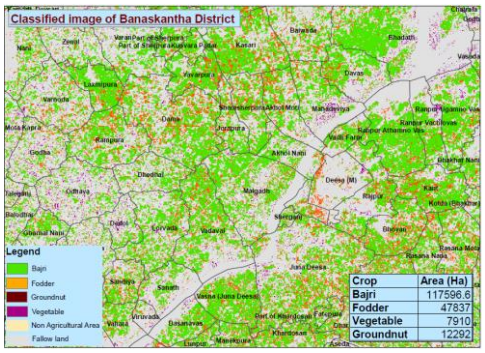
**Preparation of DPR of Par–Tapi-Narmada Link using Geo-informatics for National Water development Agency (NWDA)**



- The main objective of Par–Tapi-Narmada Link project is to divert surplus water available in west flowing rivers of south Gujarat and Maharashtra for utilization in the drought prone Saurashtra and Kachcha. On the request from NDWA, preparation of various maps for proposed DPR work was undertaken by the BISAG- N. Land use and submergence maps of proposed dams along with its statistics have been prepared by the BISAG- N. The detailed work consisted of generation of Digital Elevation Model (DEM), contour generation, Land use mapping, forest area generation of submergence extent at different levels etc.

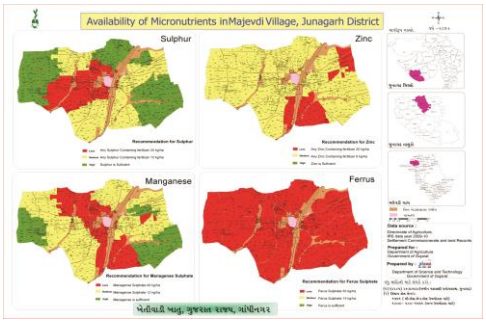
## Agriculture

**District and Village-level Crop Inventory**



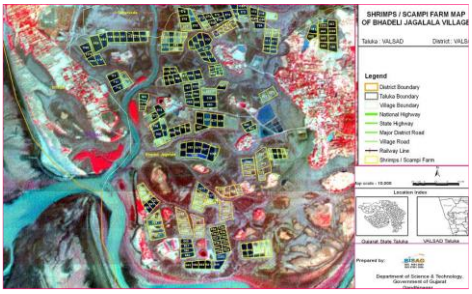

- Remote Sensing (RS) based Village-level Crop Acreage Estimation at was taken up in two villages of Anand and Mehsana districts of Gujarat state. The major objective of this study was to attempt village-level crop inventory during two crop seasons of Kharif (monsoon season) and Rabi (winter season) using single-date Indian Remote Sensing (IRS) LISS-III and LISS-IV digital data of maximum vegetative growth stage of major crops during each season.
- District-level crop acreage estimation during three cropping seasons namely Kharif, Rabi and Zaid (summer) seasons was also carried out in all the 26-districts of Gujarat State. Summer crop acreage estimation Gujarat State was carried out during 2012.

## Spatial Variability Mapping of Soil Micro-Nutrients




- The spatial variability of soil micro-nutrients like Fe, Mn, Zn and Cu in various villages of different districts, Gujarat state was mapped using geo-informatics technology. The major objectives of this study were i) to quantify the variability of Mn, Fe, Cu and Zn concentration in soil; ii) to map the pattern of micro-nutrient variability in cadastral maps, iii) suggest proper application of micro-nutrients based on status of deficiency for proper crop management and iv) preparation of village-level atlases showing spatial variability of micro-nutrients.

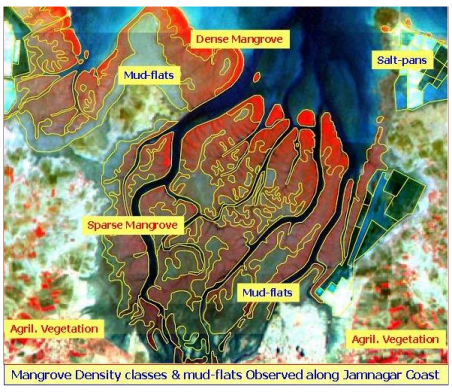
## Geo-spatial Information System for Coastal Districts of Gujarat




- The project on development of Village-level Geo-spatial Information System for Shrimp Farms in Coastal Districts of Gujarat, was taken with major objective of development of Village-level Geo-spatial Information System for Shrimp/Scampi areas using Remote Sensing (RS) and GIS. This project was sponsored by the Marine Products Export Development Authority (MPEDA), Ministry of Commerce & Industry, Government of India for scientific management of Scampi farms in the coastal districts which can help fishermen to better their livelihood and increase the economic condition on sustainable basis. The customized query shell was developed using the open source software for sharing the information amongst the officers from MPEDA and potential users. This has helped the farmers to plan their processing and marketing operations so as to achieve better remunerations.

# Environment and Forest

## Mapping and Monitoring of Mangroves in the Coastal Districts of Gujarat State



Mangrove Density classes & mud-flats Observed along Jamnagar Coast

- Gujarat Ecology Commission, with technical inputs from the Bhaskaracharya National Institute for Space Applications and Geo-informatics - N (BISAG- N) made an attempt to publish Mangrove Atlas of the Gujarat state. Mangrove atlas for 13-coastal districts with 35-coastal talukas in Gujarat, have been prepared using Indian Remote sensing satellite images. The comparison of mangrove area estimates carried out by BISAG- N and Forest Survey of India (FSI) indicates a net increase in the area under mangrove cover. The present assessment by BISAG- N, has recorded 996.3 sq. km under mangrove cover, showing a steep rise to the tune of 88.03 sq. km. In addition to the existing Mangrove cover, the present assessment also gives the availability of potential area of 1153 sq. km, where mangrove regeneration program can be taken up.

# Academy of Geo-informatics for Sustainable Development

## Introduction

- Considering the requirement of high end research and development in the areas having relevance of geo-informatics technology for sustainable development, a separate infrastructure has been established. In collaboration with different institutes in the state as well as in the country, R&D activities are being carried out in the areas of climate change, environment, disaster management, natural resources management, infrastructure development, resources planning, coastal hazard and coastal zone management studies, etc. under the guidance of eminent scientists.

- Various innovative methodologies/models developed in this academy through the research process have helped in development of various applications. There are plans to enhance R&D activities manifold during coming years.
- This unit also provides training to more than 600 students every year in the field of Geo-informatics to the students from various backgrounds like water resources, urban planning, computer Engineering, IT, Agriculture in the areas of Remote sensing, GIS and their applications.

- This Academy has been established as a separate infrastructure for advanced research and development through following schools:
  - School of Geo-informatics
  - School of Climate & Environment
  - School of Integrated Coastal Zone Management

- School of Sustainable Development Studies
- School of Natural Resources and Bio-diversity
- School of Information Management of Disasters
- School of Communication and Society

During XIIth Five year Plan advance applied research through above schools shall be the main thrust area. Already M. Tech and Ph.D. students of other Universities/ Institutes are doing research in this academy in applied sciences under various collaborative programmes.

## M. Tech. Students' Research Programme

The academy started M. Tech. students' research programme in a systematic way. It admitted 11 students from various colleges and universities in Gujarat, Rajasthan and Madhya Pradesh for period of 10 months from August 2011 to May 2012. All the students were paid stipend of Rs. 6000 per month during the tenure. The research covered the following areas:

- Cloud computing techniques
- Mobile communication
- Design of embedded systems
- Aquifer modelling
- Agricultural and Soils Remote Sensing
- Digital Image processing Techniques (Data Fusion and Image Classification).

The research resulted in various dissertations and publications in national and international journals.

• Now nine students, one from IIT, Kharagpur, three from GTU, one from M. S University, Vadodara and four from GU, are undergoing their Ph. D programme. Out of nine, two thesis have been submitted. Two students are from abroad. One each from Vietnam and Yemen. Since then (after approval of research programme from the Governing Body), 200+ papers have been published by the Academy.

# CANDIDATE'S DECLARATION

We declare that 7th semester Summer internship project report entitled "**Text to Speech – Chrome Extension**" is our own work conducted under the supervision of the external guide **Punit Lalwani & Sidhdharth Patel** from BISAG-N (Bhaskaracharya National Institute for Space Applications & Geo-informatics). We further declare that to the best of our knowledge the report for this project does not contain any part of the work which has been submitted previously for such project either in this or any other institutions without proper citation.

| Candidate 1's Signature | Candidate 2's Signature | Candidate 3's Signature |
|---|---|---|
| **Bhavya Adhiya** | **Bhakti Chokshi** | **Hasti Sutaria** |
| Student ID: M4 | Student ID: M4 | Student ID: M4 |

| Candidate 4's Signature | Candidate 5's Signature |
|---|---|
| **Jaahanava Joshi** | **Nupur Khatri** |
| Student ID: M4 | Student ID: M4 |

**Submitted To:**

Adani Institute Of Infrastructure Engineering, Ahmedabad (GTU)

# ACKNOWLEDGMENT

We are grateful to **T.P. Singh,** Director General (BISAG-N) for giving us this opportunity to work the guidance of renowned people of the field of MIS Based Portal also providing us with the required resources in the company.

We would like to express our endless thanks to our external guide **Mr. Punit Lalwani & Sidhdharth Patel** And Training Cell **Mr. Sidhdharth Patel** at Bhaskaracharya National Institute of Space Application and Geo-informatics for their sincere and dedicated guidance throughout the project development.

Also, our hearty gratitude to our Head of Department, **Dr. Hitesh Chhinkaniwala** and our internal guide **Dr. Hitesh Chhinkaniwala** for giving us encouragement and technical support on the project.

**From :**

**Bhavya Adhiya**
Student ID: M4

**Bhakti Chokshi**
Student ID: M4

**Hasti Sutaria**
Student ID: M4

**Jaahanava Joshi**
Student ID: M4

**Nupur Khatri**
Student ID: M4

# INDEX

# 1. Introduction

## 1.1 Project Details

Text to Speech Web Extension - As simple as the title is, the project is for Web users to help convert text displayed on screen to speech. With the help of Web Technologies (HTML, CSS, JavaScript), and Python, we have managed to form this idea into reality. Voice/speech synthesis is a field of computer science that deals with designing computer systems that synthesize written text. It is a technology that allows a computer to convert a written text into speech via a microphone or telephone. As an emerging technology, not all developers are familiar with speech technology. While the basic functions of both speech synthesis and speech recognition takes only minutes to understand, there are subtle and powerful capabilities provided by computerized speech that developers will want to understand and utilize. Automatic speech synthesis is one of the fastest developing fields in the framework of speech science and engineering. As the new generation of computing technology, it comes as the next major innovation in man-machine interaction, after functionality of Speech recognition (TTS), supporting Interactive Voice Response (IVR) systems.
The basic idea of text-to-speech (TTS) technology is to convert written input to spoken output by generating synthetic speech. There are several ways of performing speech synthesis:

1. Simple voice recording and playing on demand;
2. Splitting of speech into 30-50 phonemes (basic linguistic units) and their re-assembly in a fluent speech pattern;
3. The use of approximately 400 diaphones (splitting of phrases at the centre of the phonemes and not at the transition).

The most important qualities of modern speech synthesis systems are its naturalness and intelligibility. By naturalness we mean how closely the synthesized speech resembles real human speech. Intelligibility, on the other hand, describes the ease with which the speech is understood. The maximization of these two criteria is the main development goal in the TTS field.

## 1.2 Purpose

This project comes in handy when we are dealing with extensively exhausting articles, or when we are working on a project ad narration can be helpful. This will also help blind (disabled) or slightly abled people to navigate through web articles easily. This project has theoretical, practical, and methodological significance: The speech synthesizer will be very useful to any researcher who may wish to venture into the "Impact of using Computer speech program for brain enhancement and assimilation process in human beings". This text-to-speech synthesizing system will enable the semi-illiterates assess and read through electronic documents, thus bridging the digital divide. The technology will also find applicability in systems such as banking, telecommunications (Automatic system voice output), transport, Internet portals, accessing PC, emailing, administrative and public services, cultural centres and many others. The system will be very useful to computer manufacturers and software developers as they will have a speech synthesis engine in their applications.

## 1.3 Scope

It is exploding with growth at the moment. It is not only just about assisted reading, but it is becoming something bigger and more important as the days go by.
It has great scope in ML where browsers could mimic humans over speaking. TTS can add emotions to text. The study is focused on an ideal combination of a human-like behaviour with computer application to build a one-way interactive medium between the computer and the user. This application was customized using only one (1) word sentence consisting of the numeric digit 0 to 9 that could be used in operating a voice operated telephone system. Human speech is inherently a multi modal process that involves the analysis of the uttered acoustic signal and includes higher level knowledge sources such as grammar semantics and pragmatics. This project intends to focus only on the acoustic signal processing without the incorporation of a visual input.

## 1.4 Objective

Most of us have to go through tons of articles and snippets of books before we understand a topic. Reading sure is fun, but becomes very demanding after a point. What if you had someone narrate these articles at your desired speed so that you could go through them faster or listen to them while working on something else. Text-to-speech web extension can enable this and allow readers to go through the content better. The general objective of the project is to develop a Text-to-speech synthesizer for the physically impaired and the vocally disturbed individuals using English language. The specific objectives are:

1. To enable the deaf and dumb to communicate and contribute to the growth of an organization through synthesized voice.

2. To enable the blind and elderly people enjoy a User-friendly computer interface.

3. To create modern technology appreciation and awareness by computer operators.

4. To implement an isolated whole word speech synthesizer that is capable of converting text and

responding with speech

5. To validate the automatic speech synthesizer developed during the study.

## 1.5 Technology & Literature Review

- **Cloud based Text-to-Speech Study by Google. (2014)**
  Google Cloud Text-to-Speech enables developers to synthesize natural-sounding speech with 100+ voices, available in multiple languages and variants. It applies DeepMind's groundbreaking research in Wave Net and Google's powerful neural networks to deliver the highest fidelity possible. As an easy-to-use API, you can create lifelike interactions with your users, across many applications and devices.

- **Speech Processing by Google (2012)**
  Research focuses on what makes Google unique: computing scale and data. Using large scale computing resources pushes us to rethink the architecture and algorithms of speech recognition, and experiment with the kind of methods that have in the past been considered prohibitively expensive.

- The structure of the text-to-speech synthesizer can be broken down into major modules:

  • Natural Language Processing (NLP) module: It produces a phonetic transcription of the text read, together with prosody.

  • Digital Signal Processing (DSP) module: It transforms the symbolic information it receives from NLP into audible and intelligible speech. The major operations of the NLP module are as follows:

  • Text Analysis: First the text is segmented into tokens. The token-to-word conversion creates the orthographic form of the token. For the token "Mr" the orthographic form "Mister" is formed by expansion, the token "12" gets the orthographic form "twelve" and "1997" is transformed to "nineteen ninety seven".

  • Application of Pronunciation Rules: After the text analysis has been completed, pronunciation rules can be applied. Letters cannot be transformed 1:1 into phonemes because correspondence is not always parallel. In certain environments, a single letter can correspond to either no phoneme (for example, "h" in "caught") or several phoneme ("m" in "Maximum"). In addition, several letters

can correspond to a single phoneme (“ch” in “rich”). There are two strategies to determine pronunciation:

- In dictionary-based solution with morphological components, as many morphemes (words) as possible are stored in a dictionary. Full forms are generated by means of inflection, derivation and composition rules. Alternatively, a full form dictionary is used in which all possible word forms are stored. Pronunciation rules determine the pronunciation of words not found in the dictionary.

- In a rule based solution, pronunciation rules are generated from the phonological knowledge of dictionaries. Only words whose pronunciation is a complete exception are included in the dictionary. The two applications differ significantly in the size of their dictionaries. The dictionary-based solution is many times larger than the rules-based solution’s dictionary of exception. However, dictionary-based solutions can be more exact than rule-based solution if they have a large enough phonetic dictionary available.

# 2. Project Management

## 2.1 Feasibility Study

### 2.1.1 Technical Feasibility

Text-to-speech synthesis -TTS - is the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading that text. Text-to-speech synthesizer (TTS) is the technology which lets computer speak to you. The TTS system gets the text as the input and then a computer algorithm which called TTS engine analyses the text, pre-processes the text and synthesizes the speech with some mathematical models. The TTS engine usually generates sound data in an audio format as the output. We have considered 3 APIs (Application Programming Interface) for 3 different purposes. After research, we came to conclusion, feasibility of Using Google TTS API will be beneficial. TTS works on all types of Web service providers hence anyone one with such service can use it with ease.

### 2.1.2 Time Schedule Feasibility

Project was distributed between 4 weeks. Each member contributed around 6-10 hours a week with college on-going.

### 2.1.3 Operational Feasibility

With guidance of our supervisor, we were able to figure out which languages are optimal to use and easily implementable.

### 2.1.4 Implementation Feasibility

The tasks to implement TTS:

1. Create HTML Backend
2. Add JavaScript
3. Implement TTS from Google Cloud API using Promise API for using Python
4. Create Stylesheet
5. Make a manifest. json
6. HTML file for pop-up
7. Implement JS for pop-up

Requirements of Team members before implementing

1. Learn JavaScript
2. Learn how to use Google Cloud API
3. Learn how to make manifest.json

## 2.2 Project Planning

### 2.2.1 Project Development Approach and Justification

Implementation was divided in 8 steps

1. Learn the required languages and how to use API
2. Create HTML Backend
3. Add JavaScript
4. Implement TTS from Google Cloud API
5. Create Style Sheet
6. Make HTML File for pop-up
7. Make manifest.json
8. Implement JS for pop-up

The Flow was designed and updated on the basis of requirements of making a web extension.

### 2.2.2 Milestone and Deliverables

Milestones:

1. Enabling different accents
2. Enabling local language support
3. Enabling translation (from English to other commonly used languages)
4. Allowing users to customize speed
5. Allowing users to customize voice and accent

Deliverables

1. Downloadable Web Extension file
2. Report and Documentation

### 2.2.3 Roles and Responsibilities

We divided the work between 5 members over four weeks.
Week #1 - Learning JS
Week #2 - Making HTML Backend and pop-up
Week #3 - Making JSON File and access API
Week #4 - Making manifest.json file, solve errors and consolidate work

Major Responsibilities between group members were such as this -

1. Bhakti Choksi - Learning JS, researching about project work flow, accessing API, making stylesheet
2. Hasti Sutaria - Learning how to use API, researching about API, making backend
3. Nupur Khatri - Learning how to use API, making frontend and json file
4. Jahanava Joshi - Working with Backend, solving errors, consolidating code
5. Bhavya Adhiya - Learning how to use API, researching about API, making backend

Each week's progress was documented in Weekly Report which for submitted to our supervisor.

# 3. System Requirement Study

## 3.1 Study of Current System

Text-to-speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different languages, dialects, and specialized vocabularies are available through third-party publishers.

Currently, there are a number of applications, plugins, and gadgets that can read messages directly from an e-mail client and web pages from a web browser or Google Toolbar. Some specialized software can narrate RSS feeds. On one hand, online RSS narrators simplify information delivery by allowing users to listen to their favorite news sources and convert them to podcasts. On the other hand, online RSS readers are available on almost any personal computer connected to the Internet. Users can download generated audio files to portable devices, e.g., with a help of a podcast receiver, and listen to them while walking, jogging, or commuting to work.

## 3.2 Problems and weaknesses of current system

- **Text-to-Phonetic Conversion**

The first task faced by any TTS system is the conversion of input text into linguistic representation, usually called text-to-phonetic or grapheme-to-phoneme conversion. The difficulty of conversion is highly language depended and includes many problems. In some languages, such as Finnish, the conversion is quite simple because written text almost corresponds to its pronunciation. Conversion can be divided into three main phases, text preprocessing, creation of linguistic data for correct pronunciation, and the analysis of prosodic features for correct intonation, stress, and duration.

- **Pronunciation**

The second task is to find the correct pronunciation for different contexts in the text. Some words, called *homographs*, cause maybe the most difficult problems in TTS systems. Homographs are spelled the same way but they differ in meaning and usually in pronunciation (e.g., fair, lives). The word *lives* are for example pronounced differently in sentences "Three *lives* were lost" and "One *lives* to eat". Some words, e.g., *lead*, has different pronunciations when used as a verb or noun, and between two noun senses (He followed her *lead* / He covered the hull with *lead*). With these kinds of words, some semantical information is necessary to achieve correct pronunciation.

- **Prosody**

Finding correct intonation, stress, and duration from the written text is probably the most challenging problem for years to come. The intonation means how the pitch pattern or fundamental frequency changes during speech. The prosody of continuous speech depends on many separate aspects, such as the meaning of the sentence and the speaker's characteristics and emotions.

Timing at sentence level or grouping of words into phrases correctly is difficult because prosodic phrasing is not always marked in the text by punctuation, and phrasal accentuation is almost never marked (Santen et al. 1997). If there are no breath pauses in speech or if they are in the wrong places, the speech may sound very unnatural or even the meaning of the sentence may be misunderstood. For example, the input string "John says Peter is a liar" can be spoken in two different ways giving two different meanings as "John says: Peter is a liar" or "John, says Peter, is a liar". In the first sentence, Peter is a liar, and in the second one, the liar is John.

- **Chrome Extension failure**

## 3.3 User characteristics

Looking at its use cases, it is useful for individuals having high reading practices, and escaping from continuous reading. People with visual and reading impairments were the early adopters of TTS. It makes sense: TTS eases the internet experience for 1 out of 5 people who have dyslexia, low literacy readers, and others with learning disabilities by removing the stress of reading and presenting information in an optimal format.

**Students**
The teaching style, Universal Design for Learning (UDL), follows in the same vein to help all learners be successful. Teachers of all grade levels who promote UDL use a combination of auditory, visual, and kinesthetics techniques through the use of technology and adaptable lesson plans. Another useful way to use TTS in the classroom is for proofreading.

**Readers on the Go**
If you're like me, you may wish there were more hours in a day. When I want to catch up on the news, podcasts and audiobooks only take me so far.

**Multitaskers**
The life shortcuts TTS can provide are endless – from reading recipes while you cook to dictating instruction manuals when assembling furniture. The only limit to how much it can help is your own imagination!

**Mature Readers**
My grandmother has avoided upgrading from her first mobile – a flip phone circa 2004. Understandably, she wants to avoid straining her eyes from all the small text. TTS can alleviate this issue for her (and others) and introduce her to smartphones and the wonderful world of emojis (no TTS necessary).

**Visually impaired readers**
It's not simply seniors who want to avoid straining their eyes on screens. Many people have mild visual impairments or suffer from sensitivity to light.

**Foreign Language Students**
Studies show that listening to a different language aids students in learning the new dialect. TTS can help with that.

**Multilingual**
New generations raised in multilingual households may understand their (grand)parent's language, but they may not feel fluent enough to read, write or speak it. This is common in many communities, where the home language is not studied in school.

**Those with Severe Speech Impairments**
A speech-generating device (SGD), which is also known as a voice output communication aid (VOCA), can be used for those who have severe speech impairments and who would otherwise not be able to communicate verbally.

Enable the world of listening to the content you love and boost your productivity today!

## 3.4 Hardware and Software Requirements

Device used for reading include desktops, PC, I pad, Tablets, cell phones are the main hardware requirement. Software requirements include chrome usage.

## 3.5 Constraints

### 3.5.1 User Interface

Whenever text to speech chrome extension is enabled, a pop-up card is visible that has volume, rate, pitch and language adjustor. It's extremely user friendly. User needs to select the text and boom! You're all set.

Like Chrome's user interface (UI), an extension UI should be purposeful and minimal. Extensions should allow users to customize or enhance the user's browsing experience without distracting from it.

This guide explores required and optional user interface features. Use it to understand how and when to implement different UI elements within an extension.

The Action API controls the extension's action (toolbar icon). It can either open a popup or trigger some functionality when it's clicked.

Users can trigger an extension's action by expanding the extension menu and selecting the desired extension.

To make it easier to access an extension, the user may choose to pin the extension's action to the toolbar. Once pinned, the extension's action will appear to the left of the extension menu. Users can rearrange their pinned extensions by dragging and dropping their action icons to the desired order.

### 3.5.2 Communications Interface

Pop up card takes the input of volume, pitch, rate and language from the user. Now, user selects the text and the input is fed to TTS synthesizer and it converts the selected text into the speech. Language, volume, pitch and rate output values are equal to user input values.

Since content scripts run in the context of a web page and not the extension, they often need some way of communicating with the rest of the extension. For example, an RSS reader extension might use content scripts to detect the presence of an RSS feed on a page, then notify the background page in order to display a page action icon for that page.

Communication between extensions and their content scripts works by using message passing. Either side can listen for messages sent from the other end, and respond on the same channel. A message can contain any valid JSON object (null, boolean, number, string, array, or object). There is a simple API for one-time requests and a more complex API that allows you to have long-lived connections for exchanging multiple messages with a shared context. It is also possible to send a message to another extension if you know its ID, which is covered in the cross-extension messages section.

Sometimes it's useful to have a conversation that lasts longer than a single request and response. In this case, you can open a long-lived channel from your content script to an extension page, or vice versa, using runtime.connect or tabs.connect, respectively. The channel can optionally have a name, allowing you to distinguish between different types of connections.

One use case might be an automatic form fill extension. The content script could open a channel to the extension page for a particular login,

and send a message to the extension for each input element on the page to request the form data to fill in. The shared connection allows the extension to keep shared state linking the several messages coming from the content script.

When establishing a connection, each end is given a runtime.Port object which is used for sending and receiving messages through that connection.

### 3.5.3 Safety and security consideration

Third party and unauthorized extensions may perform data theft. This can risk the user's info and its circumstances can be vulnerable. Better to used reviewed and trusted chrome extensions.

Users will not install an extension if it compromises their privacy or asks for more permissions that it seems to need. Permission requests should make sense to users and be limited to the critical information necessary to implement the extension. Extensions that collect or transmit any user data must comply with the user data privacy policies .

Protect and respect extension users by including these precautions to keep their identity safe.

The APIs that an extension can access are specified in the permissions field of the manifest. The more permissions granted, the more avenues an attacker has to intercept information. Only the APIs an extension depends on should be listed, and consideration should be given to less invasive options. The less permissions an extension requests, the less permission warnings will be shown to a user. Users are more likely to install an extension with limited warnings.

Extensions should not "future proof" access to user data by requesting permissions that they do not currently need, but may implement in the future. Include new permissions with extension updates and consider making them optional.

An extension's ability to access websites and most Chrome APIs is determined by its declared permissions. Permissions should be restricted

to only what is needed for its functionality. Limiting permissions establishes an extension's capabilities and reduces possible incursion to data if the extension is compromised by an attacker. Protect extensions and their users by implementing explicit, minimal and optional permissions.

An extension's ability to access websites and most Chrome APIs is determined by its declared permissions. Permissions should be restricted to only what is needed for its functionality. Limiting permissions establishes an extension's capabilities and reduces possible incursion to data if the extension is compromised by an attacker. Protect extensions and their users by implementing explicit, minimal and optional permissions.

# 4. Requirement of Proposed System

## 4.1 Module description

The TTS system converts an arbitrary ASCII text to speech. The first step involves extracting the phonetic components of the message, and we obtain a string of symbols representing sound-units (phonemes or allophones), boundaries between words, phrases and sentences along with a set of prosody markers (indicating the speed, the intonation etc.). The second step consists of finding the match between the sequence of symbols and appropriate items stored in the phonetic inventory and binding them together to form the acoustic signal for the voice output device.

To compute the output, the system consults 1. A database containing the parameter values for the sounds within the word, 2. A knowledge base enumerating the options for synthesizing the sounds. Incorporating Expert system in the internal programs will enable the new TTS system exhibit these features: 1. The system performs at a level generally recognized as equivalent to that of a human expert 2. The system is highly domain specific. 3. The system can explain its reasoning process 4. If the information with which it is working is probabilistic or fuzzy, the system can correctly propagate uncertainties and provide a range of alternative solution with associated likelihood.

The character image is converted into text and then text into speech. The algorithm is followed.

- Firstly check the condition that if Win 32 SAPI is available in the computer or not. If it is not available then error will be generated and Win 32 SAPI library should be loaded in the computer.
- Gets the voice object from Win 32 SAPI.
- Compares the input string with Win 32 SAPI string.
- Extracts voice by firstly select the voice which are available in library.
- Choose the pace of voice.
- Initializes the wave player for convert the text into speech.

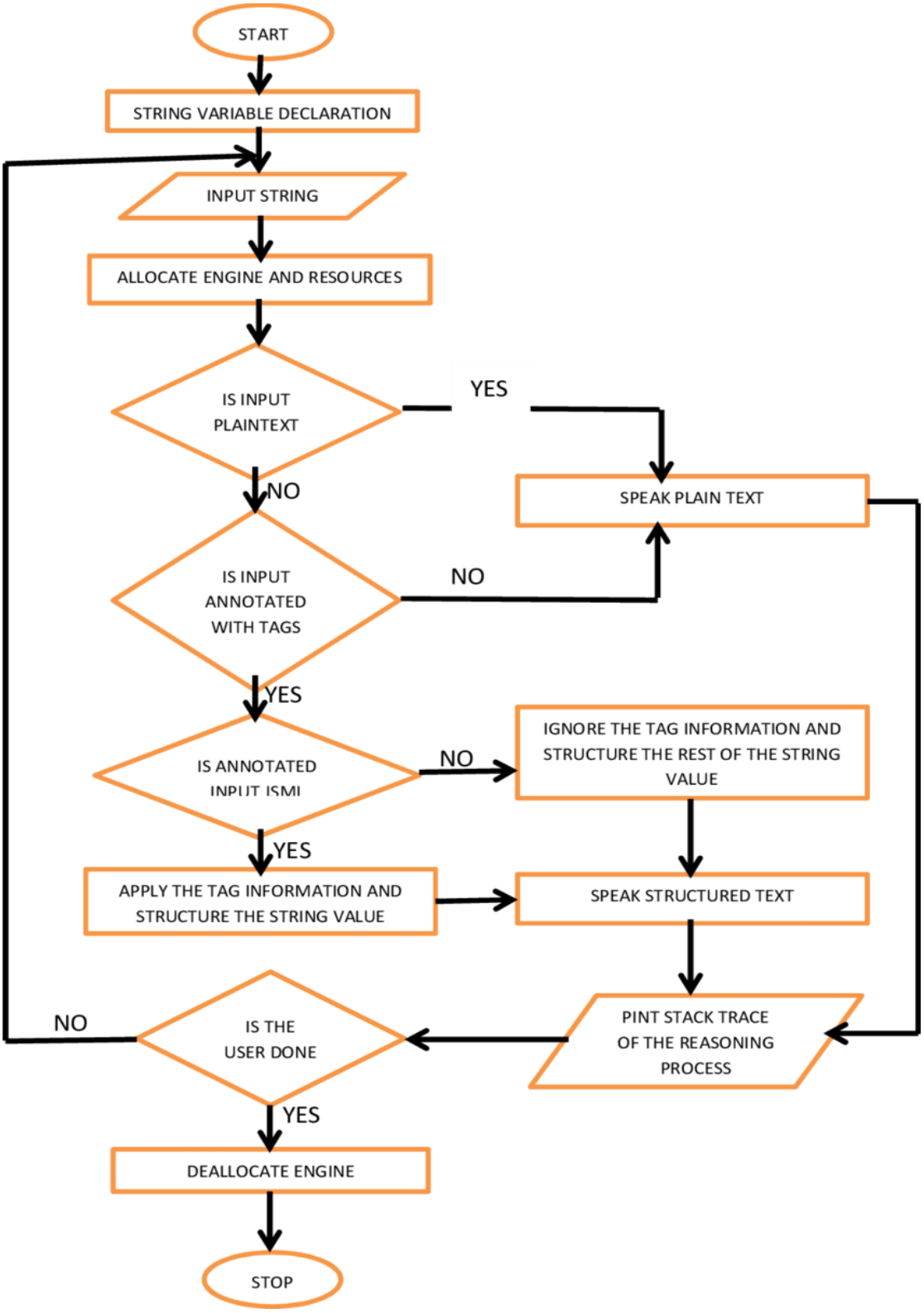
• Finally get the speech for given image. Text to speech conversion for the e-text input that directly typed in computer is also executed by the above steps.

## 4.2 Features

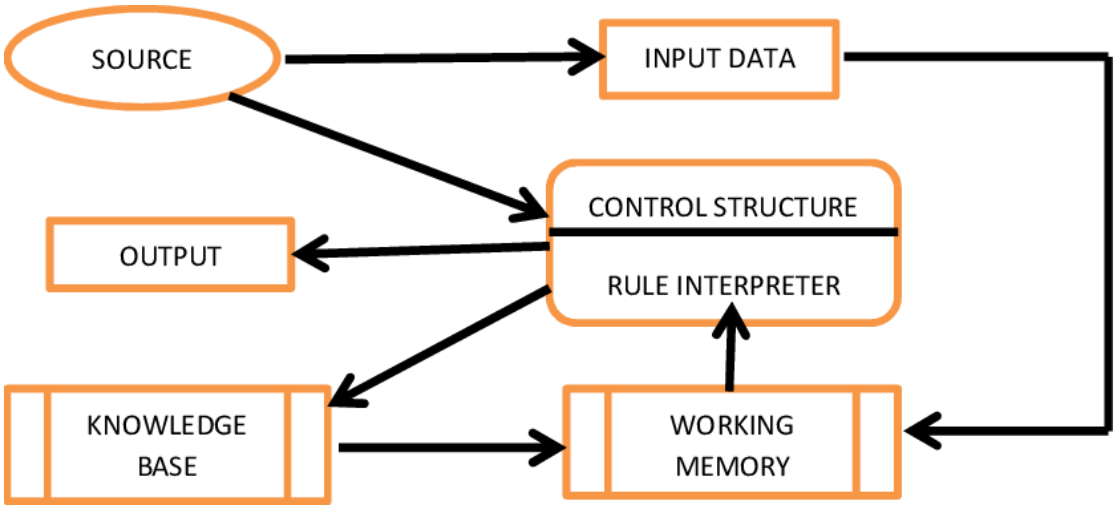
- User-friendly
- Widget with customizable options
- Has Multiple language and voice options to choose from
- Speech conversion of selected text
- Text Highlighting
- Playback Controls
- Voices sound better than the competition
- Chrome versions available
- Immersive Reader
- No registration required for usage.

# 5. System Design

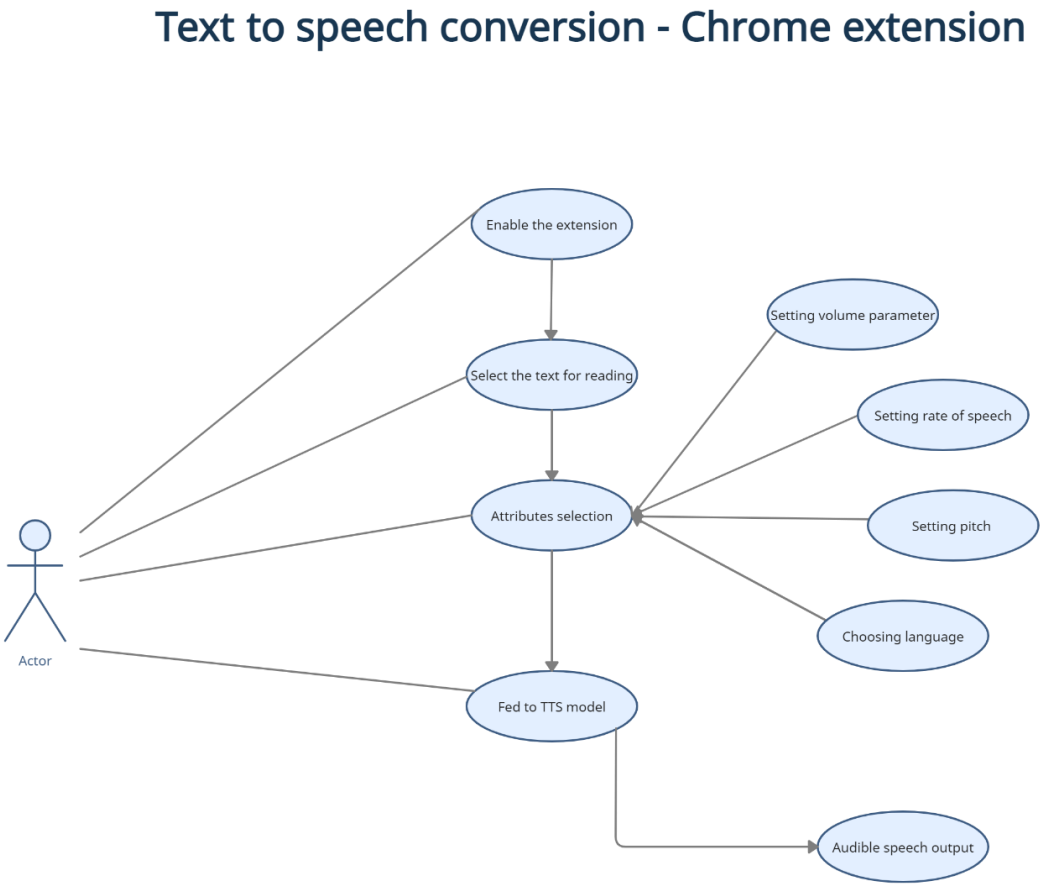
## 5.1 Flow Chart



## 5.2 Data Flow Diagram

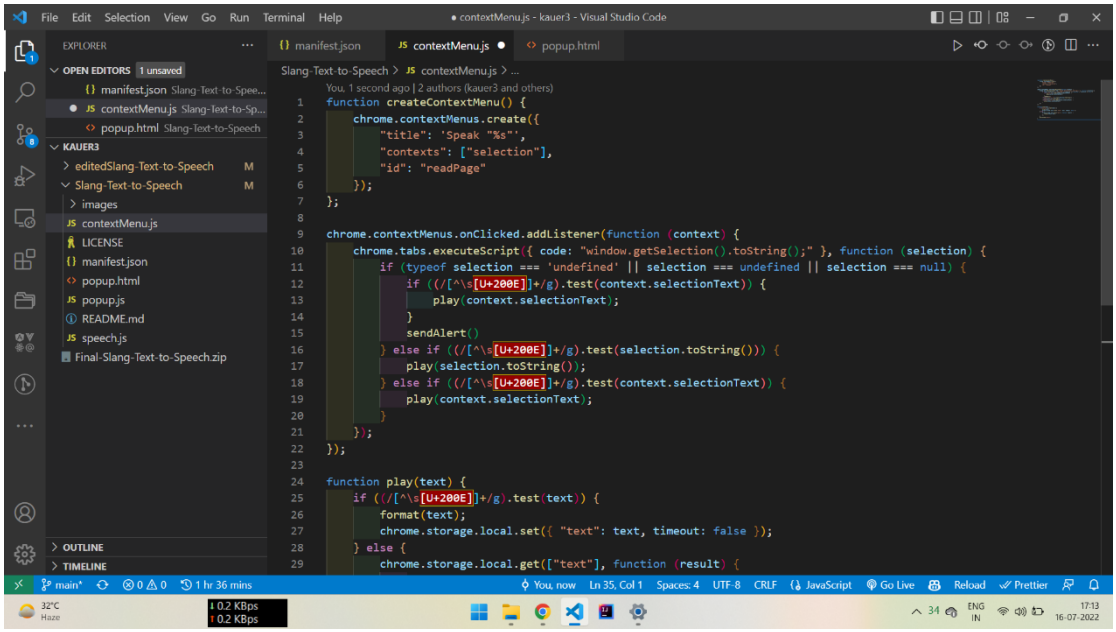


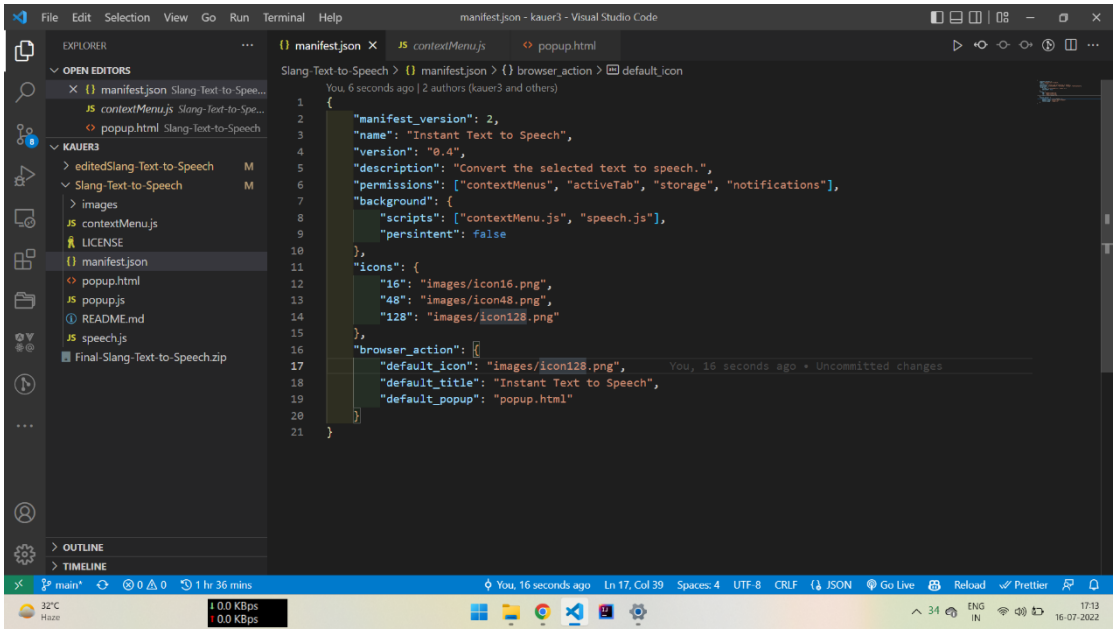

## 5.3 Use Case Diagram

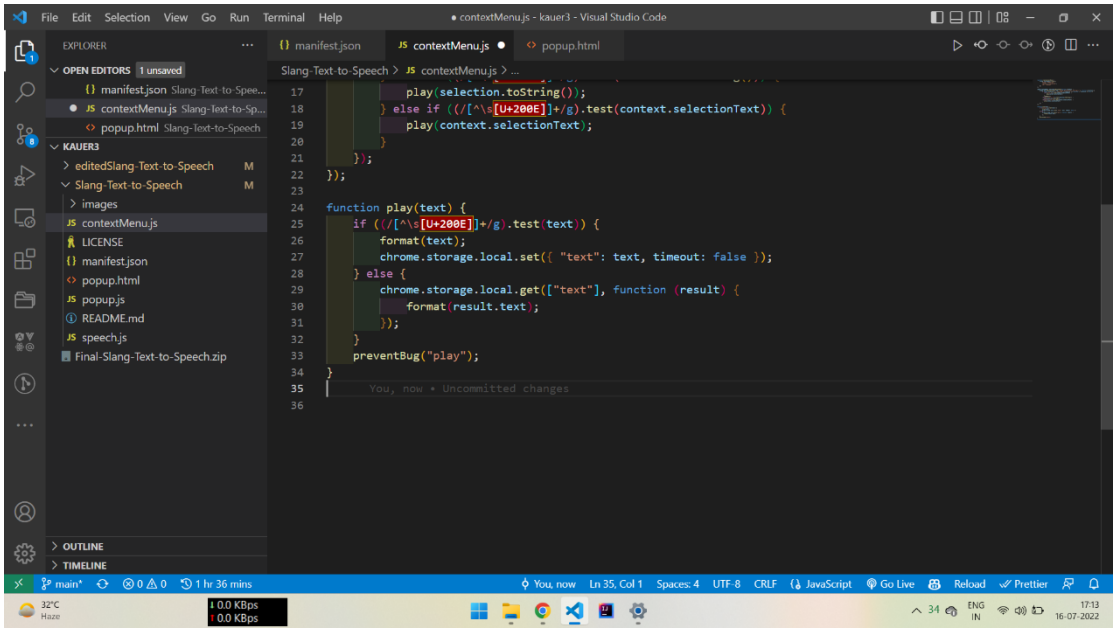


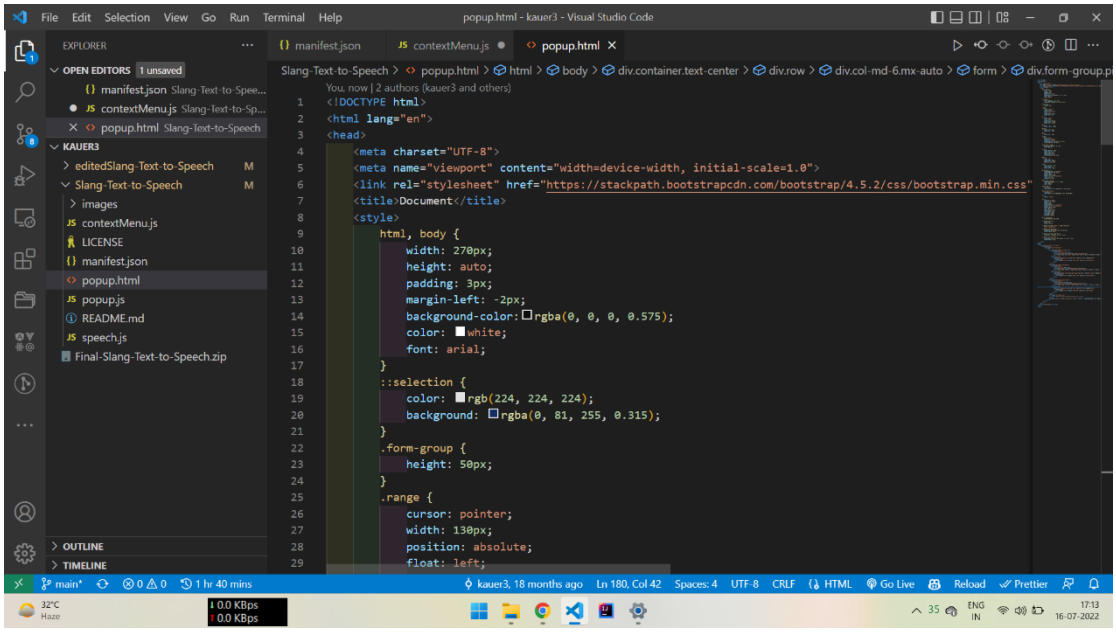
# 5.2 Snapshots

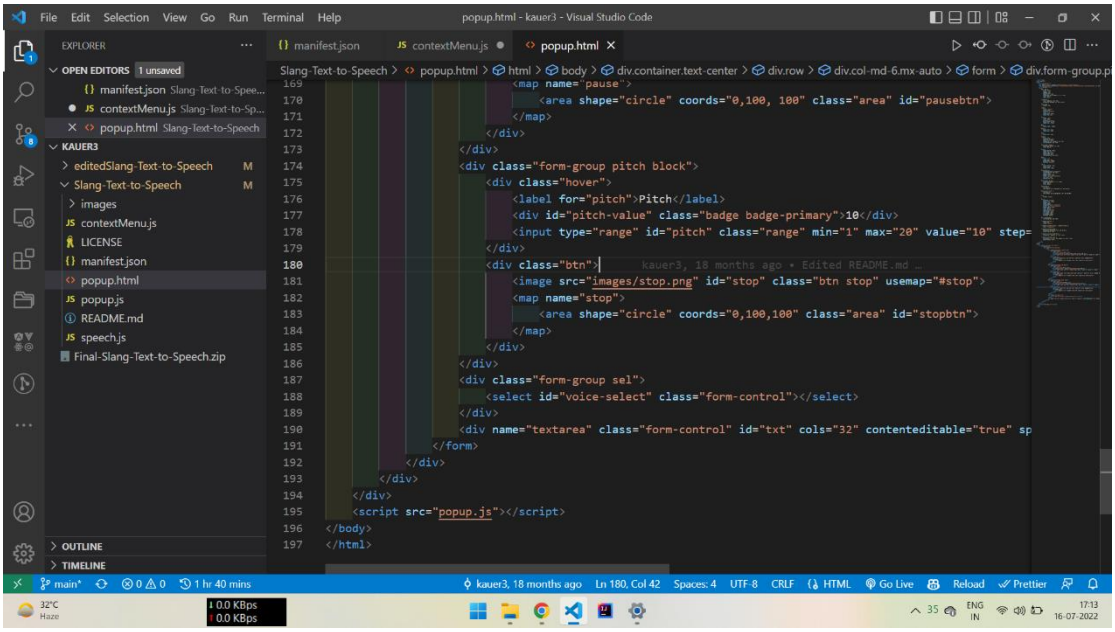
## Code Snippets




File Edit Selection View Go Run Terminal Help
• contextMenu.js - kauer3 - Visual Studio Code
EXPLORER
OPEN EDITORS 1 unsaved
manifest.json Slang-Text-to-Spee...
contextMenu.js Slang-Text-to-Sp...
popup.html Slang-Text-to-Speech
KAUER3
editedSlang-Text-to-Speech M
Slang-Text-to-Speech M
images
contextMenu.js
LICENSE
manifest.json
popup.html
popup.js
README.md
speech.js
Final-Slang-Text-to-Speech.zip
OUTLINE
TIMELINE
manifest.json
contextMenu.js
popup.html
Slang-Text-to-Speech > contextMenu.js > ...
play(selection.toString());
} else if ((/[^\s[U+200E]]+/g).test(context.selectionText)) {
play(context.selectionText);
}
});
});
function play(text) {
if ((/[^\s[U+200E]]+/g).test(text)) {
format(text);
chrome.storage.local.set({ "text": text, timeout: false });
} else {
chrome.storage.local.get(["text"], function (result) {
format(result.text);
});
}
preventBug("play");
}
You, now • Uncommitted changes
main* 0 0 1 hr 36 mins
You, now Ln 35, Col 1 Spaces: 4 UTF-8 CRLF JavaScript Go Live Reload Prettier
32°C Haze
0.0 KBps 0.0 KBps
ENG IN 17:13 16-07-2022


File Edit Selection View Go Run Terminal Help
popup.html - kauer3 - Visual Studio Code
EXPLORER
OPEN EDITORS 1 unsaved
manifest.json Slang-Text-to-Spee...
contextMenu.js Slang-Text-to-Sp...
popup.html Slang-Text-to-Speech
KAUER3
editedSlang-Text-to-Speech M
Slang-Text-to-Speech M
images
contextMenu.js
LICENSE
manifest.json
popup.html
popup.js
README.md
speech.js
Final-Slang-Text-to-Speech.zip
OUTLINE
TIMELINE
manifest.json
contextMenu.js
popup.html
Slang-Text-to-Speech > popup.html > html > body > div.container.text-center > div.row > div.col-md-6.mx-auto > form > div.form-group.p
You, now | 2 authors (kauer3 and others)
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.2/css/bootstrap.min.css"
<title>Document</title>
<style>
html, body {
width: 270px;
height: auto;
padding: 3px;
margin-left: -2px;
background-color: rgba(0, 0, 0, 0.575);
color: white;
font: arial;
}
::selection {
color: rgb(224, 224, 224);
background: rgba(0, 81, 255, 0.315);
}
.form-group {
height: 50px;
}
.range {
cursor: pointer;
width: 130px;
position: absolute;
float: left;
main* 0 0 1 hr 40 mins
kauer3, 18 months ago Ln 180, Col 42 Spaces: 4 UTF-8 CRLF HTML Go Live Reload Prettier
32°C Haze
0.0 KBps 0.0 KBps
ENG IN 17:13 16-07-2022


popup.html - kauer3 - Visual Studio Code
manifest.json
contextMenu.js
popup.html
Slang-Text-to-Speech > popup.html > html > body > div.container.text-center > div.row > div.col-md-6.mx-auto > form > div.form-group.p
<map name="pause">
<area shape="circle" coords="0,100, 100" class="area" id="pausebtn">
</map>

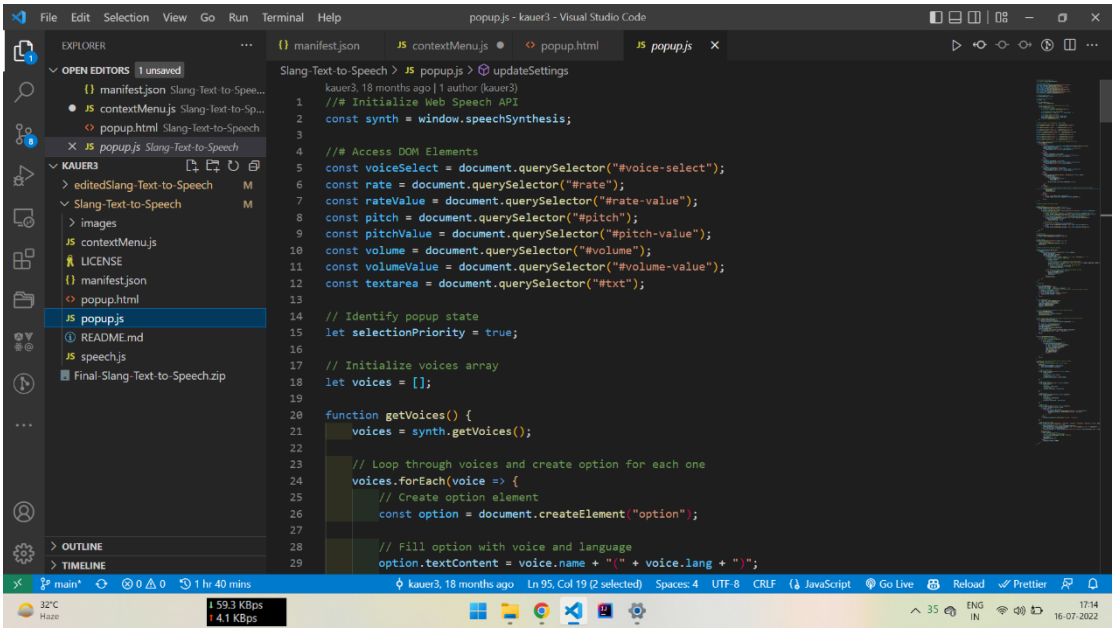
<label for="pitch">Pitch</label>
10
<input type="range" id="pitch" class="range" min="1" max="20" value="10" step=

kauer3, 18 months ago • Edited README.md
<image src="images/stop.png" id="stop" class="btn stop" usemap="#stop">
<map name="stop">
<area shape="circle" coords="0,100,100" class="area" id="stopbtn">
</map>

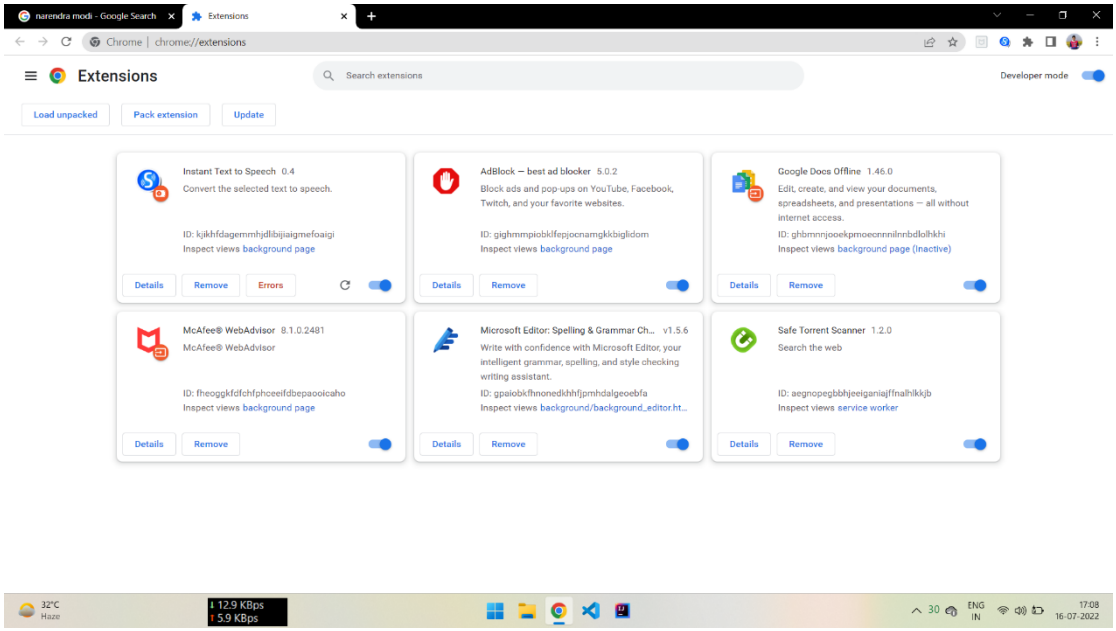
<select id="voice-select" class="form-control"></select>

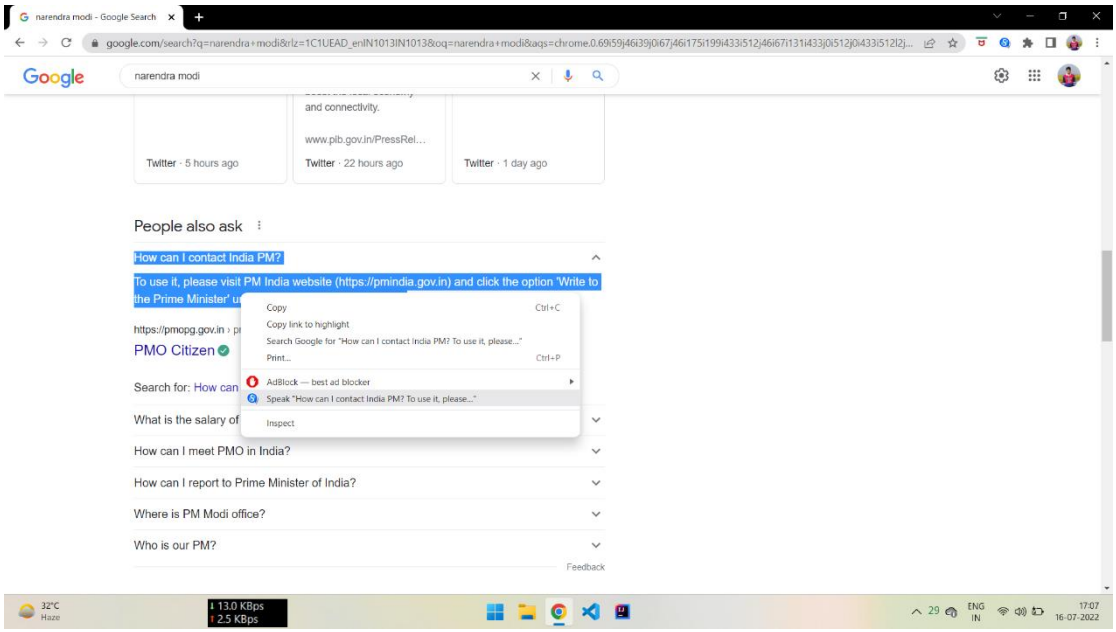
<div name="textarea" class="form-control" id="txt" cols="32" contenteditable="true" sp
</form>

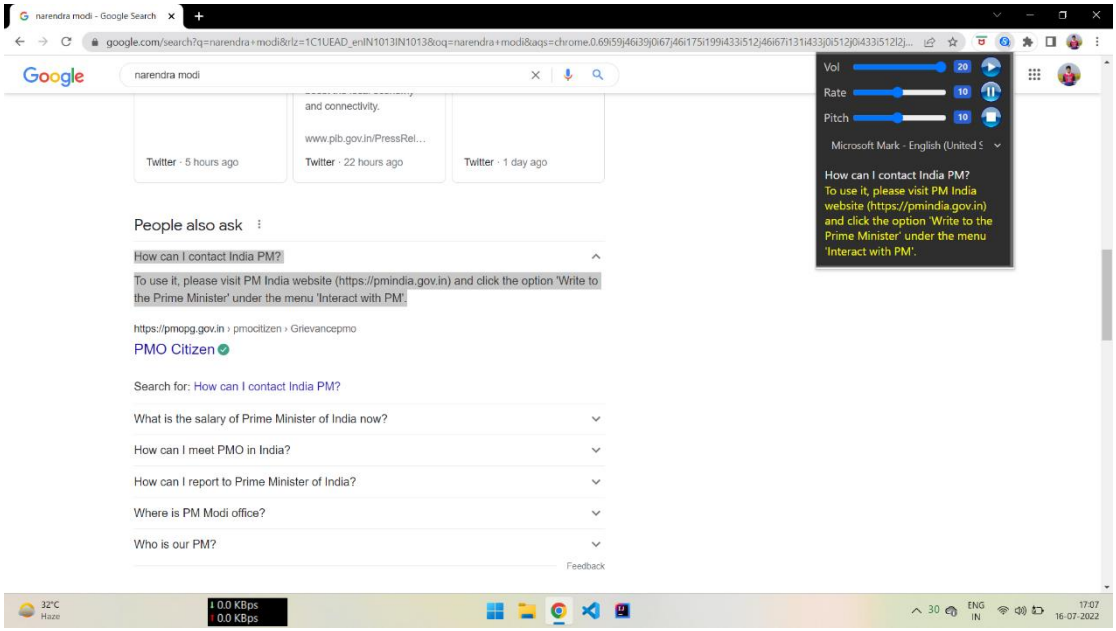
<script src="popup.js"></script>
</body>
</html>
kauer3, 18 months ago Ln 180, Col 42 Spaces: 4 UTF-8 CRLF HTML Go Live Reload Prettier

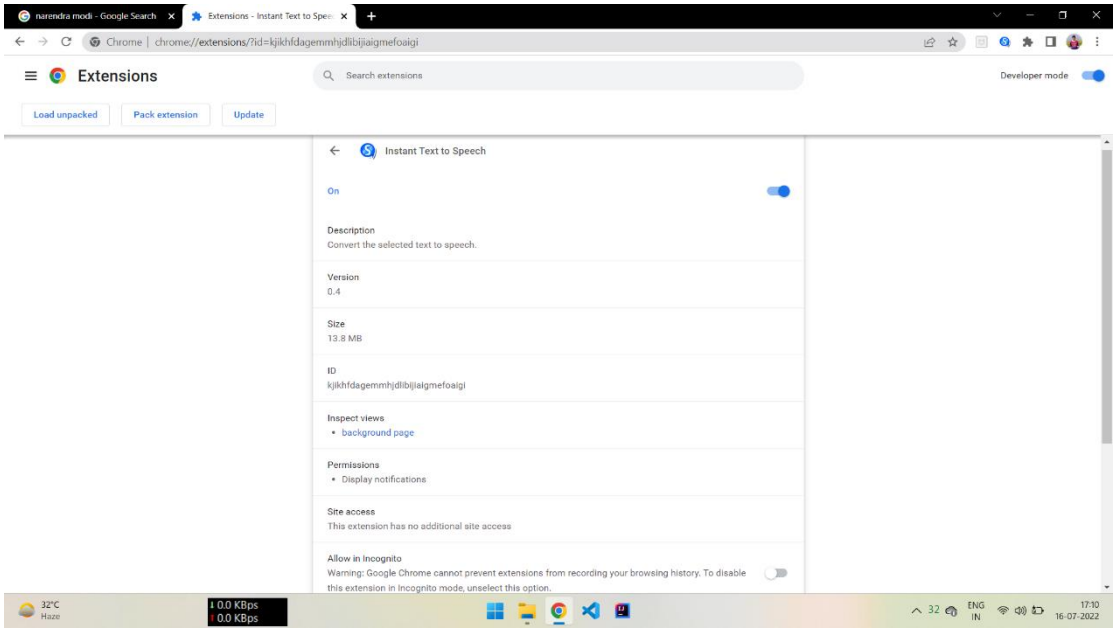

popup.js - kauer3 - Visual Studio Code
manifest.json
contextMenu.js
popup.html
popup.js
Slang-Text-to-Speech > popup.js > updateSettings
kauer3, 18 months ago | 1 author (kauer3)
//# Initialize Web Speech API
const synth = window.speechSynthesis;
//# Access DOM Elements
const voiceSelect = document.querySelector("#voice-select");
const rate = document.querySelector("#rate");
const rateValue = document.querySelector("#rate-value");
const pitch = document.querySelector("#pitch");
const pitchValue = document.querySelector("#pitch-value");
const volume = document.querySelector("#volume");
const volumeValue = document.querySelector("#volume-value");
const textarea = document.querySelector("#txt");
// Identify popup state
let selectionPriority = true;
// Initialize voices array
let voices = [];
function getVoices() {
voices = synth.getVoices();
// Loop through voices and create option for each one
voices.forEach(voice => {
// Create option element
const option = document.createElement("option");
// Fill option with voice and language
option.textContent = voice.name + "(" + voice.lang + ")";
kauer3, 18 months ago Ln 95, Col 19 (2 selected) Spaces: 4 UTF-8 CRLF JavaScript Go Live Reload Prettier

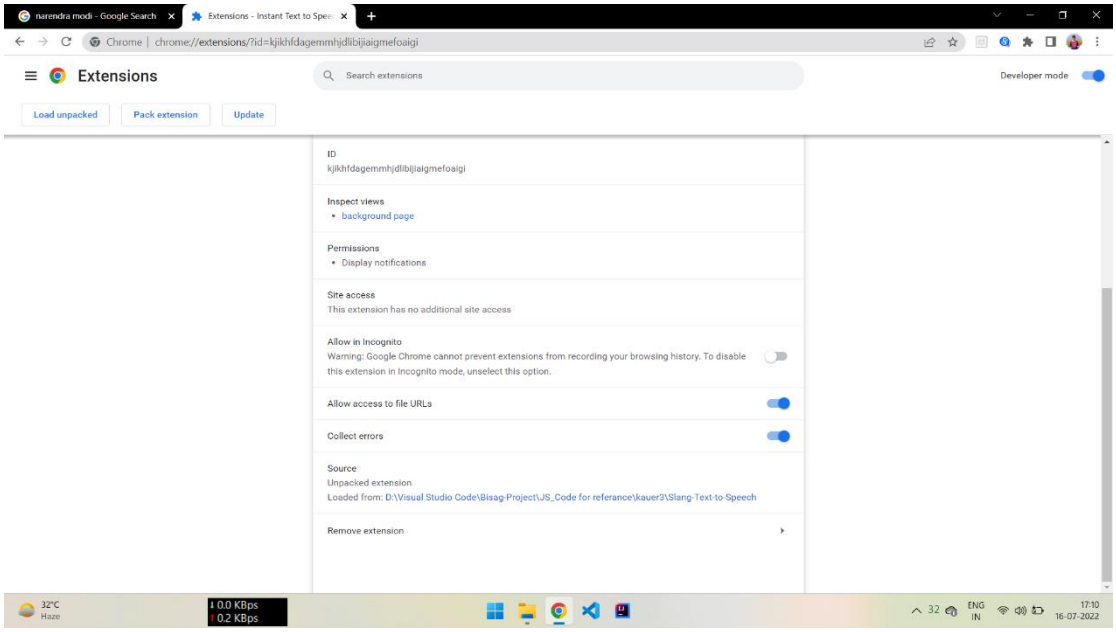
# Final output screenshots




narendra modi - Google Search
google.com/search?q=narendra+modi&rlz=1C1UEAD_enIN1013IN1013&oq=narendra+modi&aqs=chrome.0.69i59j46i39j0i67j46i175i199i433i512j46i67i131i433j0i512j0i433i512l2j...
Google
narendra modi
Twitter · 5 hours ago
and connectivity.
www.pib.gov.in/PressRel...
Twitter · 22 hours ago
Twitter · 1 day ago
Vol 20
Rate 10
Pitch 10
Microsoft Mark - English (United S
How can I contact India PM?
To use it, please visit PM India website (https://pmindia.gov.in) and click the option 'Write to the Prime Minister' under the menu 'Interact with PM'.
People also ask
How can I contact India PM?
To use it, please visit PM India website (https://pmindia.gov.in) and click the option 'Write to the Prime Minister' under the menu 'Interact with PM'.
https://pmopg.gov.in › pmocitizen › Grievancepmo
PMO Citizen
Search for: How can I contact India PM?
What is the salary of Prime Minister of India now?
How can I meet PMO in India?
How can I report to Prime Minister of India?
Where is PM Modi office?
Who is our PM?
Feedback
32°C Haze
0.0 KBps
0.0 KBps
ENG IN
17:07
16-07-2022


narendra modi - Google Search
Extensions - Instant Text to Spee
Chrome | chrome://extensions/?id=kjikhfdagemmhjdlibijiaigmefoaigi
Extensions
Search extensions
Developer mode
Load unpacked
Pack extension
Update
Instant Text to Speech
On
Description
Convert the selected text to speech.
Version
0.4
Size
13.8 MB
ID
kjikhfdagemmhjdlibijiaigmefoaigi
Inspect views
background page
Permissions
Display notifications
Site access
This extension has no additional site access
Allow in Incognito
Warning: Google Chrome cannot prevent extensions from recording your browsing history. To disable this extension in Incognito mode, unselect this option.
32°C Haze
0.0 KBps
0.0 KBps
ENG IN
17:10
16-07-2022


narendra modi - Google Search
Extensions - Instant Text to Spee
Chrome | chrome://extensions/?id=kjikhfdagemmhjdlibijiaigmefoaigi
Extensions
Search extensions
Developer mode
Load unpacked
Pack extension
Update
ID
kjikhfdagemmhjdlibijiaigmefoaigi
Inspect views
background page
Permissions
Display notifications
Site access
This extension has no additional site access
Allow in Incognito
Warning: Google Chrome cannot prevent extensions from recording your browsing history. To disable this extension in Incognito mode, unselect this option.
Allow access to file URLs
Collect errors
Source
Unpacked extension
Loaded from: D:\Visual Studio Code\Bisag-Project\JS_Code for referance\kauer3\Slang-Text-to-Speech
Remove extension
32°C
Haze
0.0 KBps
0.2 KBps
ENG
IN
17:10
16-07-2022

# 6. Implementation Planning

## 6.1 Implementation Environment: -

Our implementation environment was Visual Studio Code (version – 1.69). It comes up with excellent features such as extensions for all languages, syntax highlighting, bracket matching, box-selection, auto indentation snippets, and code beautify features. All these features helped us with easy customization and quick navigation through our code.

We have used JavaScript extension (JavaScript ES6 code snippets), Html extension (v0.2.1) and JSON reader (v2.0.2) inside the VS Code.

Because of Chrome's popularity and the rise of TTS, we will be creating a Chrome extension which converts text to speech. The extension is going to wait until the user clicks on its icon and then it will try to either find what the user has highlighted on the page that they are currently viewing or it will try to find what is copied to their clipboard. If anything is there, it will convert it to speech by first trying to use the HTML5 Speech Synthesis API and if that is not available – by calling a third-party API.

Each Chrome extension needs to have a file called manifest.json. The manifest is a JSON-formatted file which contains data crucial to the application ranging from things such as the name, description, icons, and the author of the extension to data defining the needs of the extension – which websites should the extension be able to run on (those would be the permissions which the user must grant) or what files to run when the user is browsing a particular website.

Our manifest starts by documenting the name, the description, the author, the version and the icons of the extension. You can provide numerous icons that respond to different sizes in the icons object.

Then, we have a background script called background.min.js (notice that we use minified files) defined in the background object. Background scripts are scripts which are long-running and will continue to run until the user's browser is closed or the extension is disabled.

## 6.2 Coding Standards: -

- JavaScript – ES6
- ISO/IEC 13816:2007
- ISO/IEC TS 20071-25:2017
- ISO/IEC 8825-8:2021
- ISO/IEC 15445:2000

Firstly, we create our one and only background script which hooks up an event listener that will be fired when a user clicks on the extension's icon. When this happens we call the sendMessage function which sends a message to our content script (the content script can read the DOM and find out what the user has highlighted or/and what the user has placed on their clipboard) with the help of the chrome.tabs.sendMessage(tabId, message, callback) method. We send the message to the currently opened tab – as this is what interests us and it is what we have access to – with the help of the chrome.tabs.query method whose arguments involve a callback that will be called with an argument that contains the tabs that match the query.

Now, the lengthier part is our content script. We create an object which will hold some of the data involved with the extension and then define our initialization method.

The method checks if the user has not highlighted text or if they do not have anything on their clipboard and just returns in such a case. Otherwise, it tries to produce speech with the HTML5 Speech Synthesis API. If that fails too, it finally tries to use a third-party API.

The method that checks for text does several things. It tries to get an object with the highlighted text with the help of the built-in getSelection() method and convert it to a text string with toString(). Then, if no text is highlighted it attempts to find the text on the user's clipboard. It does this by adding an input element to the page, focusing it, firing up a paste event with the help of execCommand('paste') and then saving the pasted text inside that input in a property. Then it empties the input. In either case, it returns whatever it has found.

To convert the text to speech we rely on the trySpeechSynthesizer() method. If HTML5 Speech Synthesis exists in the user's browser (window.speechSynthesis) we know that the user is able to use it and so we check if a speech is currently running (we know if it is running through the pageToSpeech.data.speechInProgress boolean).

We stop the current speech if it is in progress (as the trySpeechSynthesizer will start a new speech and we do not want two simultaneous sounds). Then, we set speechInProgress to true and whenever the speech finishes set the property to a falsy value again.

Now, I do not want to go into details as to why we are using speechUtteranceChunker but it is a bug fix related to Chrome stopping the speech synthesis while still in progress after 200-300 hundred words are uttered. Basically, it splits our text string into many smaller chunks (of 120 words in our case) and calls the Speech Synthesis API with one chunk after another.

In the end, outside of any local scope we call the addHotkeys method which will start waiting for the user to press the right hotkey and we set up a listener which will wait until a message is received from the background script. If the right message is received (speakHighlight) or the hotkey is pressed, we will initialize our text to speech conversion object.

# 7. TESTING

## 7.1 Testing Plan

### Test plan identifier

To check the accuracy of a text-to-speech convertor.

### Test items

Text to Speech Web extension.

### Features to be tested

Volume, rate, pitch, language, device friendly, user friendly, supported by most browsers.

We have to test our Text to Speech Web extension for its proper working by correctly reading the selected sentence. Each word should be pronounced correctly in the specified volume, rate, and pitch by the user. Our testing requires manual testing for the correctness of the speech.

## 7.2 Testing Strategy

- Open any web page.
- Select the sentence using the left key on the mouse.
- Then start the web extension from the extension icon.
- Check for the working & correctness of the speech at specified volume, rate, pitch & language.
- Check whether each word is spoken out at the correct place and no word in-between gets missed.
- If the extension works well and passes all the above-mentioned criteria then it passes the testing.

## 7.3 Testing Plan

With Continuous Integration (CI) and Continuous Delivery (CD) becoming standard practices in software development, automation testing has been established as a key component in testing processes. Testing tools and processes are consciously and continuously geared towards automation, with the aim of making testing faster, easier and more accurate.
However, there are certain actions that can be more difficult to automate. For example, automating the testing of browser extensions can pose difficulties.
This article attempts to explore those difficulties and provide a few solutions.

**How to test a Chrome extension with Selenium?**

**1. Download Chrome Extension**
Get the webstore URL or extension ID from Google Web Store. Now enter the URL/ID into the main field on this page: https://chrome-extension-downloader.com/
Click Download. Save the CRX file. Store the file in the same location as your script.
**Note**: The CRX might be a local file that has not yet been uploaded to the Chrome Web Store. In case the extension does not exist in the web store, install it manually by dragging the CRX file into the **Chrome://extensions** page and clicking 'Add'. Find more details on this process here.
**2. Download the Chrome Extension Source Viewer from the Google Web Store.**
**3. View source of the CRX file**
Go to the Chrome Extensions in the Google Web Store. Click the yellow icon in the browser. Go to View Source. If the CRX file is not displayed in the Web Store, the source file can be viewed manually. Just click here and upload the CRX file.
A list of all resources (javascript, images, etc.) and pages available in the CRX should be displayed.
**4. Identify the page to be tested**

To locate a specific page, extract the unique ID of the CRX in the Chrome Extension. Here's how.

- First, get the unique ID of the Chrome Extension. Right-click on it and select **Options**.
- The user will be directed to their unique ID page URL: **chrome-extension://UNIQUEID/options.html**
- Go back to the CRX resource list. Select the specific page for testing: **SPECIFICPAGE.HTML**.
  Change options.html with the specific page on the unique URL.
  Example:
  Change **chrome-extension://UNIQUEID/options.html** to **chrome-extension://UNIQUEID/SPECIFICPAGE.html**
- Copy this URL. It will be required later in the Webdriver code.

**5. Initiate Selenium script to create a new ChromeDriver**

Enter the Chrome Extension into ChromeDriver. ChromeDriver is a standalone server that implements WebDriver's wire protocol. To do this, new code needs to be added to the beginning of the script when creating the browser object.

Here's the syntax for this code in a number of languages often used in Selenium Webdriver scripts:

**Python**

```
chop = webdriver.ChromeOptions()


chop.add_extension(EXTENSION_PATH)


# create new Chrome driver object with Chrome extension


driver = webdriver.Chrome(chrome_options=chop)
```

**Javascript**

```
var chromeOptions = webdriver.Capabilities.chrome();

chromeOptions: [

binary: '/Applications/GoogleChrome.app/Contents/MacOS/Google Chrome',

args: [ ],

extensions: ['EXTENSION_LOCATION.crx']

]

var driver = new webdriver.Builder().

withCapabilities(chromeOptions).

build();
```

**6. Navigate to The ChromeDriver Website Page**

The syntax to do this is below:

```
driver.get('chrome-extension://UNIQUEID/SPECIFICPAGE.html');
```

After doing this, the tester can interact with and test the extension as they would a normal HTML webpage.

Often, an iFrame will be included in the HTML file. Switching the focus to the iFrame is easy, depicted in the examples below:

**Python**

```
driver.switch_to.frame("FRAME_NAME");
```

**JavaScript**

```
driver.switchTo().frame("FRAME_NAME");
```

Here, FRAME_NAME refers to Id, name, xpath, css_selector and other element locators.
Learn more about handling iFrames in Selenium.

## Challenges of testing Browser Extensions

Browser extensions are embedded add-ons, not regular HTML files. Since the extension is out of scope, it is not possible to simulate user actions such as clicks and scrolls, inspect web elements, etc.
Normally websites can be easily tested by automating user actions with Selenium. But in order to automate actions on a browser extension, testers have to identify where the extension's pages are located. Then, they would have to switch their scope in the web UI to interact with the extension pages as DOM elements.
The steps described above enables testers to avoid that process and interact with an extension like they do with a normal HTML webpage. Thus, one can test a Chrome Selenium plugin with automated Selenium testing. It is an uncomplicated process that just requires a few minutes of extra effort. In return, testing an important aspect of the website user experience becomes a simple task.

### 7.3.1 Defects Testing

In this testing, we aim to target the defects in our program. Defects in our Text to Speech Web Extension could be the following: -

- Missed text while reading.
- Incorrect volume, rate, pitch, or language.
- Error in code that stops proper working of extension.
- Non-compatible with any type of browser.
- All these defects could be in the following manner: -
- Checking the presence of required version compatibility.
- Running the extension once.
- Checking for any abnormality compared to our expectations.

# 8. Limitations and Future Extensions

## Limitations

- Tedious to select the text if in larger amount.
- Internet Dependency
- Monotonous hearing without any visual aids.

## Future Extension

It is expected that the new system will reduce and improve on the problems encountered in the old system. The system is expected to among other things do the following; 1. The new system has a reasoning process. 2. The new system can do text structuring and annotation. 3. The new system's speech rate can be adjusted. 4. The Pitch of the voice can be adjusted. 5. You can select between different voices and can even combine or juxtapose them if you want to create a dialogue between them 6. It has a user friendly interface so that people with less computer knowledge can easily use it 7. It must be compatible with all the vocal engines 8. It complies with SSML specification

Two methodologies were chosen for the new system: The first methodology is Object Oriented Analysis and Development Methodology (OOADM). OOADM was selected because the system has to be represented to the user in a manner that is user-friendly and understandable by the user. Also since the project is to emulate human behaviour, Expert system had to be used for mapping of Knowledge into a Knowledge base with a reasoning procedure. Expert system was used in the internal operations of the program, following the algorithm of Rule-Based computation. The technique is derived from general principles described by researchers in knowledge engineering techniques (Murray et al., 1991; 1996). The system is based on processes modelled in cognitive phonetics (Hallahan, 1996; Fagyal, 2001) which accesses several knowledge bases (e.g. Linguistic and phonetic knowledge bases, Knowledge bases about non-linguistic features, a predictive model of perceptual processes, and knowledge base about the environment).

- **Automating the task**

We don't want users to go through the trouble of selecting text that they want to listen to. Instead, we want to automate the tasks of selecting the text. So, when a user opens new web pages it directly starts reading the only contents of a given article and not the content of the header or footer provided when this extension is enabled. We wish to implement this automation task in our future extension.

- **Translation**

The next thing we want to implement in our future extension is that we want to provide an inbuilt option inside this web extension only where users can translate it to any other language which they can understand. We wish to support many local, regional languages as well.

- **Summarisation**

This is the hard part.  We want to provide an option in which users can just get the summary of the whole article or whole text. Although, this is a very complex thing. It will help users to get the whole crux of long texts into some lines of texts only.

# 9) Conclusion and Reference

## 9.1 Conclusion

Synthesizing text is a high technology advancement and artificial formation of speech given a text to be spoken. With Text-to-Speech synthesis, we can actually mediate and fill in the lacuna provided by not fully exploiting the capabilities of some handicapped individuals. It's never been so easy to use a text-to-speech program, as just one click and your computer will speak any text aloud in a clear, natural sounding voice. Therefore, there is need to use Information Technology to solve the problem for the Before the use of the new system, proper training should be given to the users. This training can come in handy with proper tutor on how to handle JSML language and how to use it to annotate text for the proper output and emphasis.

Thus, We were able to develop an extension as per our goal in which the user is able to listen the text rather than reading it. It has covered all the cases and has been tested rigorously. Our main focus was to make this user friendly and we can say that we have achieved this. The user just needs to install our software and he can start using and decide when to listen, when to pause it (whenever required) and to stop it.

## 9.2 References


- Chrome extension icon not appearing? - Stack Overflow
- json - Chrome extension icon not appearing - Stack Overflow
- chrome.browserAction - Chrome Developers
- Extension Load Error: Could not load icon 'icon.png' for browser action - iBuildMVPs
- Service Worker Debugging
- Could not load icon 'icon16.png' specified in 'icons'. Could not load manifest. - Google Search

Bhaskaracharya National Institute for Space Applications and Geo-informatics
MeitY, Government of India
Phone: 079 - 23213081 Fax: 079 - 23213091
E-mail: info@bisag.gujarat.gov.in, website: https://bisag-n.in/

BISAG-N

ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

# Report Verification Procedure

**Date:**

**Project Name:** **Text to Speech Conversion – Chrome Extension**

**Student Name & ID:**

1. **Hasti Sutaria (M4)**
2. **Bhakti Chokshi (M4)**
3. **Nupur Khatri (M4)**
4. **Bhavya Adhiya (M4)**
5. **Jaahanava Joshi (M4)**

| | Soft Copy | Hard Copy |
|---|---|---|
| **Report Format:** | ☐ | ☐ |
| **Project Index:** | ☐ | ☐ |

**Sign by Training Coordinator** **Sign by Project Guide**