BS Project Report

**Final Report**

**Text-to-Speech & Speech-to-Text System**

Farah Malaeb

A report submitted in part fulfilment of the degree of

**BSc in Computer Science**

**Supervisor:** Dr. Antoine Melki

Department of Computer Science

Faculty of Arts and Sciences

University of Balamand

November 30, 2022

Table of Contents

1. Abstract
2. System Background
3. System Specifications
4. Problems Encountered
5. Capstone Project
6. Future Works
7. References
8. Abstract

Problem:

* Some people have difficulties in reading, and some can’t read
* Some people have problems in expressing themselves, and in their communication skills

Solution:

* Text to speech helps people with difficulties to communicate with others using the text to speech and the translator
* Speech to text system helps people to express themselves without typing or writing, helps in communication, and helps in improving speaking skills

Technologies used:

* Good internet connection
* Visual Studio Code
* Programming language used is JavaScript

Contribution:

* Web-based application that helps people with reading difficulties
* Web-based application that helps people in expressing themselves and improving their reading and communication skills

Problems Encountered:

* Time management
* Problems during the code development

Future Work:

* Introducing AI and data analytics to this app in order to overcome a fundamental need of the last generation solutions, that will result in good quality output.
* Programming AI in speech to text systems to understand different dialects and accents
* Programming AI in speech to text systems to identify in homophones

1. System Background

* **Project Group Member(s)**

Farah Malaeb, A1910600

* **Project Abstract**
* The project is a text-to-speech & speech-to-text system
* Project contains a translator that helps the user in text translation if needed
* Web-based app that transforms any text entered, into spoken equivalent and converts speech to text
* **Introduction**

Text-to-speech & speech-to-text system is an app that transforms written text in a given language into its spoken equivalent and vice versa

* Text-to-speech converts an input sentence which is usually a sequence of letters or symbols into machine-generated speech.
* Speech-to-text is a speech recognition software that enables the recognition and translation od spoken language into text through computational linguistics
* **Motivation and Scope**
* Text to speech system helps people that have difficulties in reading.
* This app may be developed into more advanced way to provide solutions in home-systems, vehicle-systems, etc.…
* **Related Work**

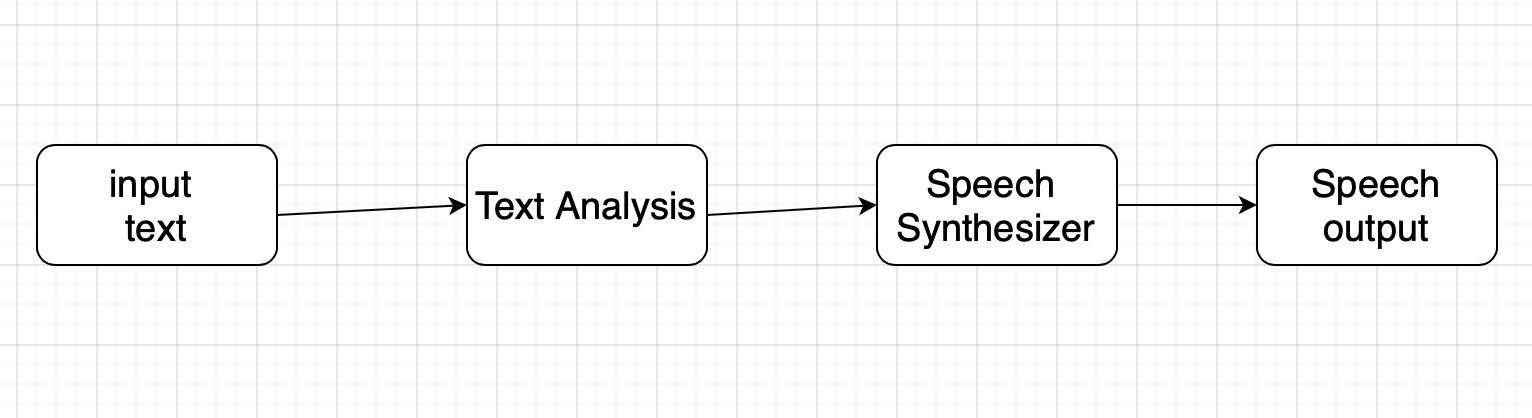
Text-to-speech:

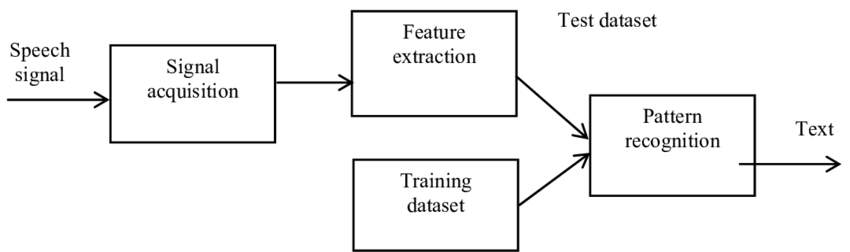
* Murf is a text-based voice-over maker, where the user can type his/her script or upload voice recording, and the tool converts it into hyper-realistic AI voices.
* Synthesys is a popular text-to-speech software to produce professional audio and video content for marketing

Speech-to-text:

* Alexa
* Google Assistant and Siri

* **System Architecture**



****

* **Goals and Objectives**
* The user must be able to access the app
* Converts an input sentence which is usually a sequence of letters or symbols into machine-generated speech.
* Help users with difficulties in reading
* Speech to text system helps people to express themselves without typing or writing, helps in communication, and helps in improving speaking skills
* Translator that helps the user
* Google translator API to help user translate text on the website
* **Gantt Chart**



* **Future Work**
* Introducing AI and data analytics to this app in order to overcome a fundamental need of the last generation solutions, that will result in good quality output.
* **Tools and Technologies**
* Good internet connection
* The app will be developed using Visual Studio Code
* Programming language used is JavaScript

1. System Specifications

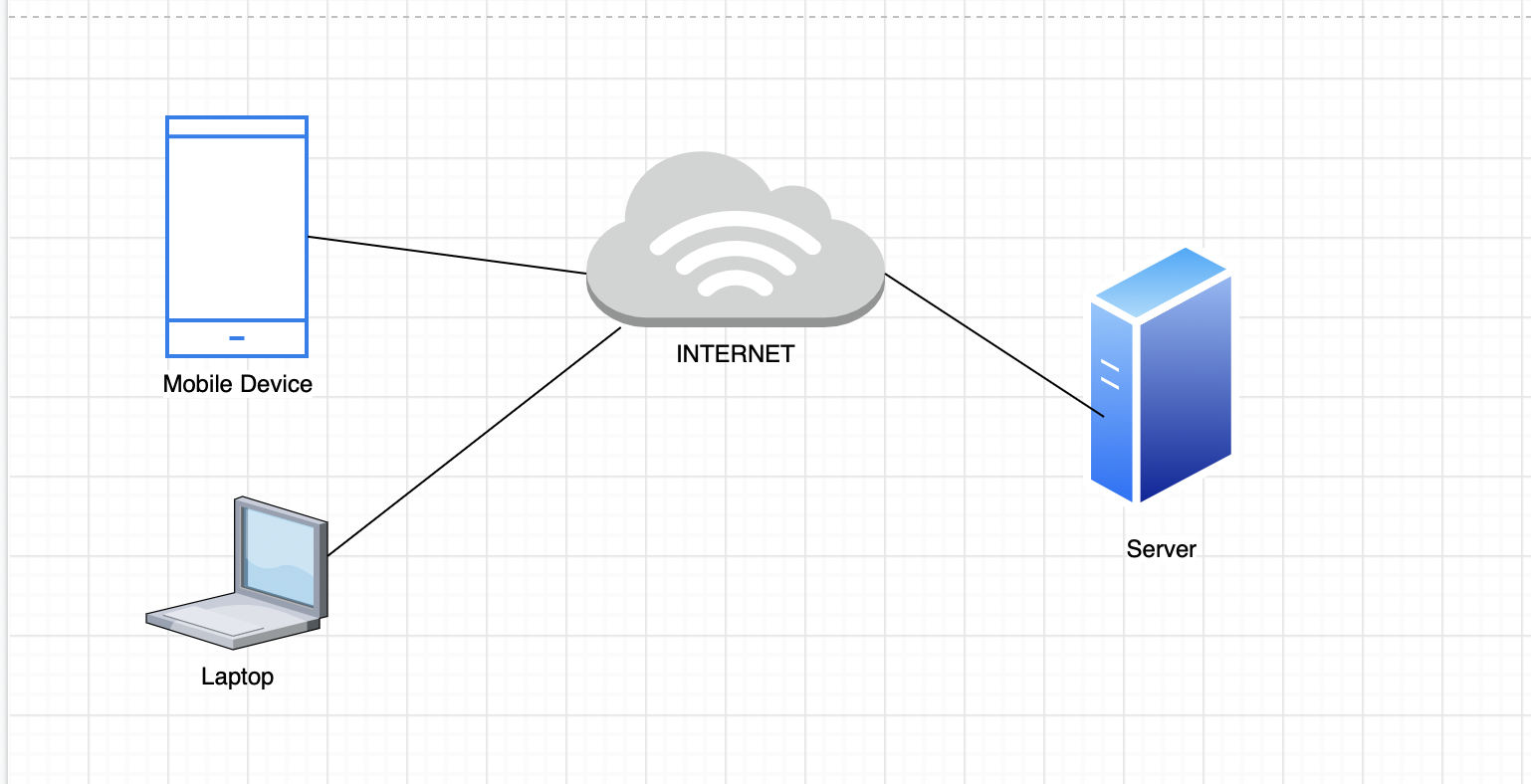
# SYSTEM ARCHITECTURE

* 1. **System Hardware Architecture**

List of components:

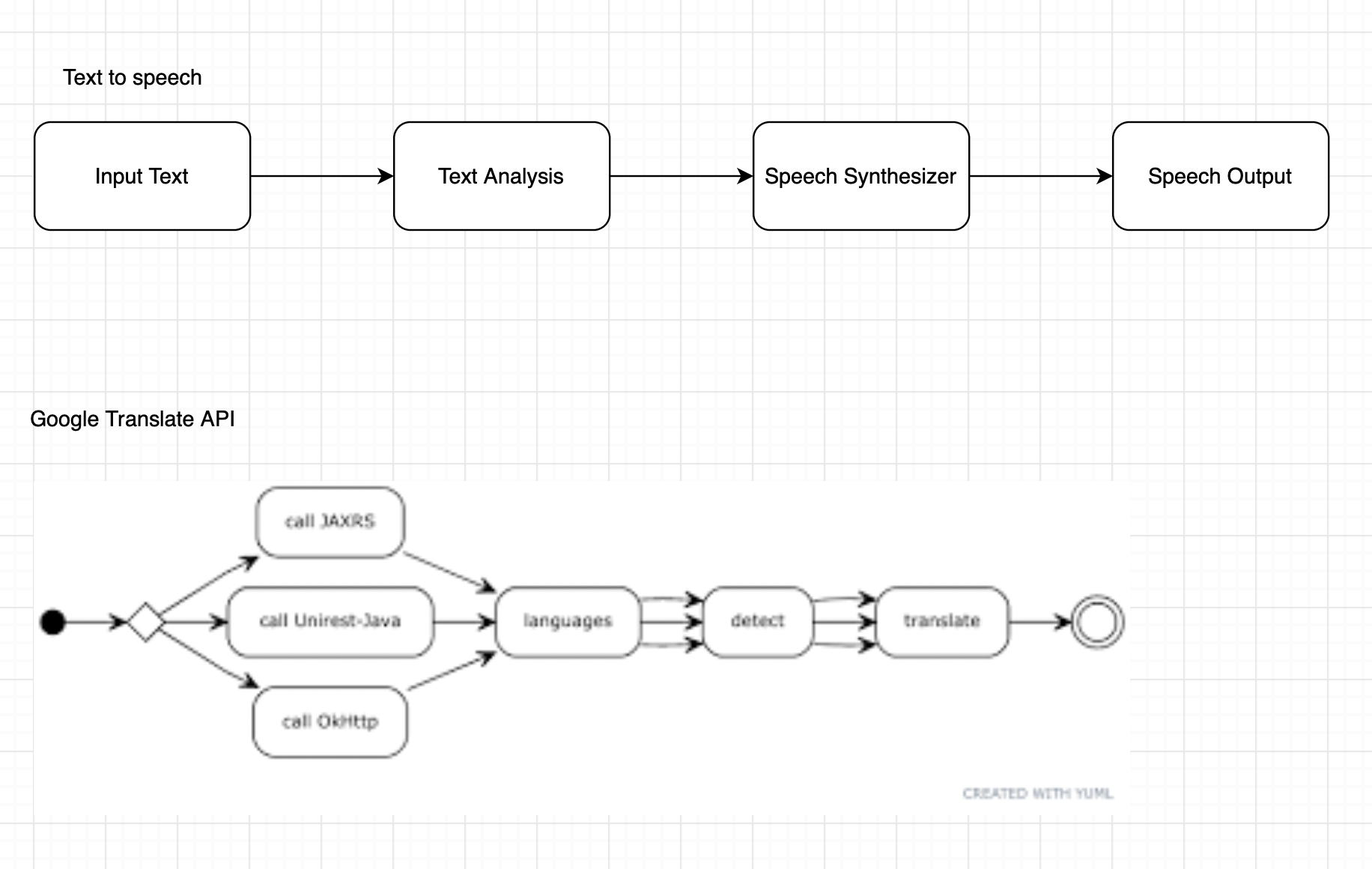
* Mobile Device or Laptop: The user will use the mobile or laptop to use the text-to-speech and speech-to-text system
* Internet Connectivity: To connect the users to the servers.
* Application server: To host the application

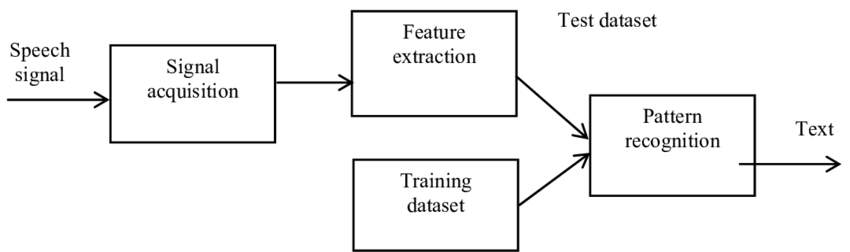
Diagram showing the connectivity of the components:



## System Software Architecture

Structured Organized Diagram



****

Use case scenarios

|  |  |  |
| --- | --- | --- |
| Scenario Name: | Text to speech | |
| Short Description: | transforms any text entered, into spoken equivalent | |
| Trigger: | Start by entering the text to be transformed to speech | |
| Type: | External | |
| Major Inputs | | |
| Description | | Source |
| enters the text | | User keyboard |
| Major Outputs | | |
| Description | | Destination |
| Speech of the text is returned to the user | | Speaker |
| Major Steps Performed | | Information for steps |
| User enter a text | | The text spoken equivalent is  returned |

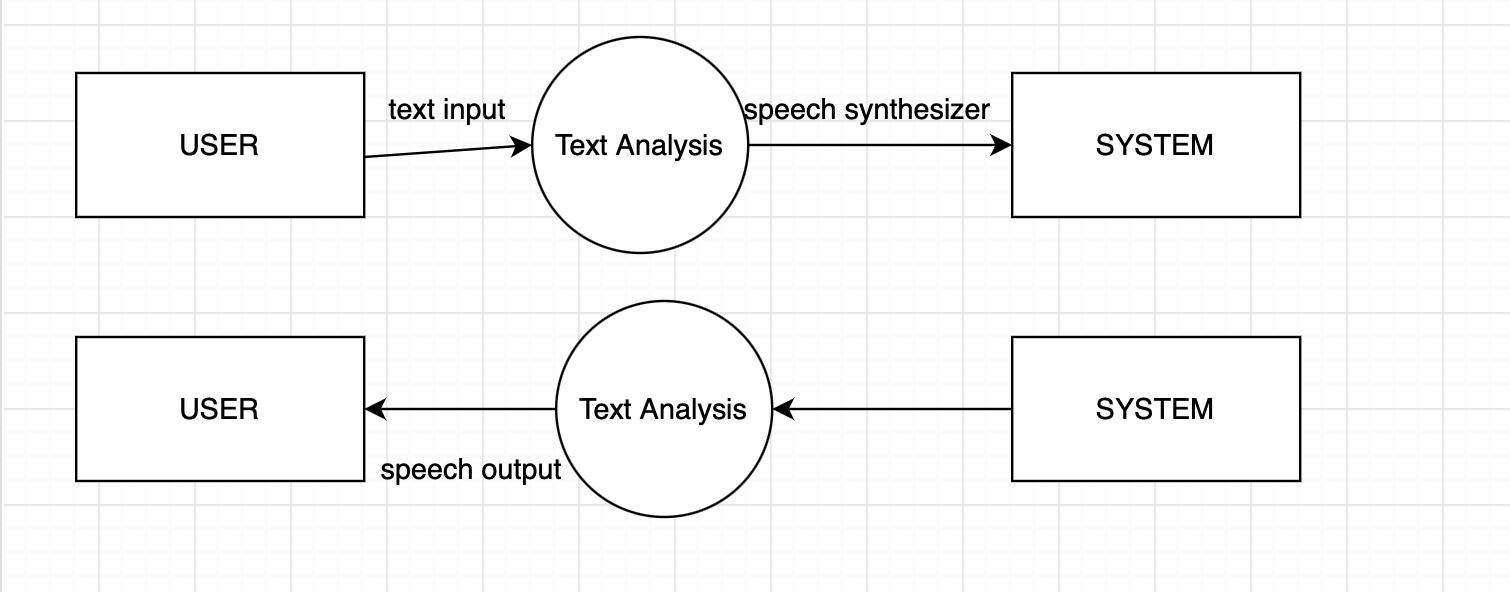
|  |  |  |
| --- | --- | --- |
| Scenario Name: | Speech to text | |
| Short Description: | enables the recognition and translation od spoken language into text | |
| Trigger: | Start by allowing the mic and speaking to convert to text | |
| Type: | External | |
| Major Inputs | | |
| Description | | Source |
| Speaks | | User microphone |
| Major Outputs | | |
| Description | | Destination |
| text of the speech is returned to the user | | screen |
| Major Steps Performed | | Information for steps |
| User speaks | | The speech spoken is returned as a text |

|  |  |  |
| --- | --- | --- |
| Scenario Name: | Translator | |
| Short Description: | Translates text from any language | |
| Trigger: | Start by typing the text the user wants to translates after selecting the languages he/she wants | |
| Type: | External | |
| Major Inputs | | |
| Description | | Source |
| Enter text to be translated | | User keyboard |
| Major Outputs | | |
| Description | | Destination |
| Text translated is returned to the user | | screen |
| Major Steps Performed | | Information for steps |
| User enters text | | The text is returned translated as another text |

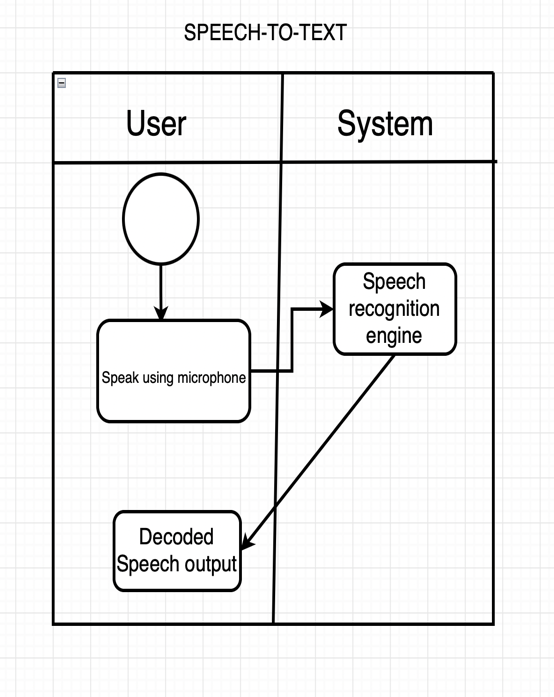
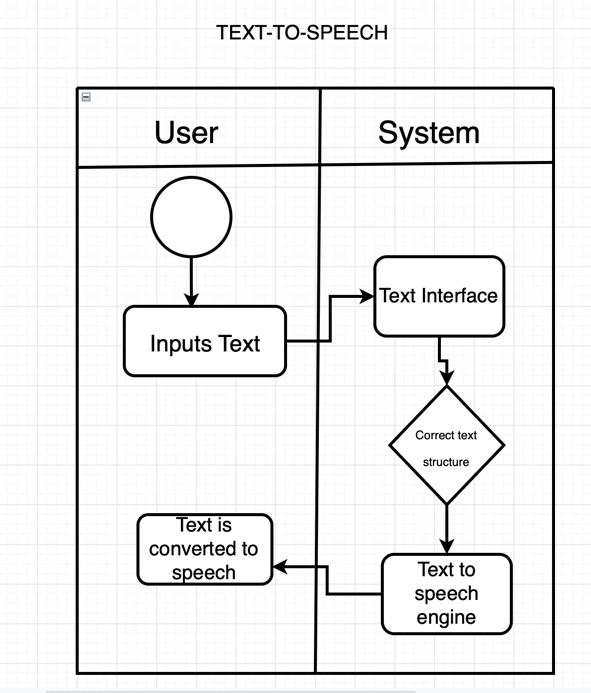
|  |  |  |
| --- | --- | --- |
| Scenario Name: | Text from image extractor | |
| Short Description: | Extracts texts from images | |
| Trigger: | Start by uploading any image of type .png | |
| Type: | External | |
| Major Inputs | | |
| Description | | Source |
| Enter image to extract text from it | | User keyboard and files contains images |
| Major Outputs | | |
| Description | | Destination |
| Text extracted from image and displayed on the screen | | screen |
| Major Steps Performed | | Information for steps |
| User uploads image | | The text is returned after extracted from the image |

## Internal Communications Architecture

DFD Level 0:



Activity Diagram:



# HUMAN-MACHINE INTERFACE

3.1. Inputs

* The user will input the text he/she wants to convert to speech
* The text is translated into aural information by speech synthesizer
* The user can use Google Translate API to translate any text before converting into speech
* The user will use the microphone to convert his speech to text
* The user will enter any text in the translator to be translated
* The user will upload any image to extract a text from it

3.2. Outputs

* The text is heard after it’s translated
* The speech is displayed on the screen
* The text is translated to the needed language
* The text will be extracted from any image and displayed on the screen

1. Problems Encountered

* Problems in time management
* Problems in code development
* The project was supposed to be a mobile app using python, it was changed to a web-based app using JavaScript due to problems in downloading python extensions in vscode

1. Capstone Project

In this project I used what we have learned in the Advances in computer science course and the web programming course in order to complete the project.

|  |  |
| --- | --- |
| Competency | Course |
| Website Design | Web Programming(CSIS228) |
| Creating functionalities in website | Web Programming and Advances in Computer Science(CSIS228 & CSIS279) |

1. Future Works

* More advanced web-app
* A mobile app instead of web-based app
* Adding a speech emotion recognition which recognize the emotion of the user based on his tone and pitch
* Adding more languages to the text to speech system since in this project it only supports english language
* Adding a feature in my app that detect bad words or words related to racism and prevent it from typing it or converting it into speech

1. References

Mark, M., About MarkMy name is Mark and I am a freelance web designer, &amp; says, M. A. (2022, August 7). Best text-to-speech software 2022 - top 19. courselounge.