VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

ARTIFICIAL INTELLIGENCE

Slot: G1

Project Title: Speech to Indian Sign Language Translator

Group Members:

Rohit Raut (19BIT0403)

Kanishk Bansal (20BIT0256)

Megha Saxena (20BIT0366)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology and Engineering

May-June, 2022

DECLARATION BY THE CANDIDATE

We here by declare that the project report entitled **"Speech to Indian Sign Language Translator"** submitted by us to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Artificial Intelligence (ITE2010)** is a record of bonafide project work carried out by us under the guidance of **Prof. Chiranji Lal Chowdhary.** We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other course.

Place : Vellore Signature

Date : 21-04-2022

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology & Engineering [SITE]

CERTIFICATE

This is to certify that the project report entitled **"Speech to Indian Sign Language Translator"** submitted by **Rohit Raut (19BIT0403), Kanishk Bansal (20BIT0256) and Megha Saxena (20BIT0366)** to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Artificial Intelligence (ITE2010)** is a record of bonafide work carried out by them under my guidance.

Prof. Chiranji Lal Chowdhary

GUIDE

Asso. Professor, SITE

1. Abstract:

Be it communication, playing computer games, attending seminars or video conferences, deaf people always miss out the fun that a normal person does. Communication is the most important difficulty they face with normal people and also every normal person does not know the sign language. The aim of our project is to develop a communication system for the deaf people. It converts the audio message into the sign language. This system takes audio as input, converts this audio recording message into text and displays the relevant Indian Sign Language images or GIFs which are predefined. By using this system, the communication between normal and deaf people gets easier.

2. Introduction:

It is said that Sign language is the mother language of deaf people. This includes the combination of hand movements, arms or body and facial expressions. There are 135 types of sign languages all over the world. Some of them are American Sign Language (ASL), Indian Sign Language (ISL), British Sign Language (BSL), Australian Sign Language (Auslan) and many more. We are using Indian Sign Language in this project. This system allows the deaf community to enjoy all sort of things that normal people do from daily interaction to accessing the information.

3. Motivation:

According to an independent census, there are about 700,000 – 800,000 mute people in the world out of which thousands live in India. Quite often, these people also suffer from hearing impairments. Therefore, the only means of communication for them is the Sign Language which is not very popular. This makes life harder for the people who suffer fromthis disability.

Our aim is to integrate our knowledge of Computer Sciences with this issue and create aprogram that will take the audio of the speaker as input and show the hand signs of the Indian language according to the audio. This will help make communication easier, and itwill do away with the need of us learning the sign language.

4. Related Works:

[1] As per Amit Kumar Shinde on his study of sign language to text and vice versa in Marathi Sign language recognition is one of the most important research and it is the most natural and common way of communication for the people with hearing problems. A hand gesture recognition system can help deaf persons to communicate with normal people in the absence of an interpreter. The system works both in offline mode and through web camera.

[2] Neha Poddar, Shrushti Rao, Shruti Sawant, Vrushali Somavanshi, Prof. Sumita Chandak in their paper discussed about the prevalence of deafness in India is fairly significant as it is the second most common cause of disability. A portable interpreting device which convert higher mathematics sign language into corresponding text and voice can be very useful for the deaf people and solve many difficulties.

[3] The glove based deaf-mute communication interpreter introduced by Anbarasi Rajamohan, Hemavathy R., Dhanalakshmi is a great research. The glove comprises of five flex sensors, tactile sensors and accelerometer. The controller matches the gesture with pre-stored outputs. The evaluation of interpreter was carried out for ten letters _A,, _B,, _C,, _D,, _F,, _I,, _L,, _O,, _M,, _N,, _T,, _S,, _W,,.

[4] As per the Neha V. Tavari A. V. Deorankar Dr. P. N. Chatur in his report discuss that many physically impaired people rely on sign language translators to express their thoughts and to be in touch with rest of the world. The project introduces the image of the hand which is captured using a web camera. The image acquired is processed and features are extracted. Features are used as input to a classification algorithm for recognition. The recognized gesture is used to generate speech or text. In this system, flex sensor gives unstable analog output and also it requires many circuits and is thus very expensive.

5. Architecture:

5.1 Functional and Non-Functional Requirements

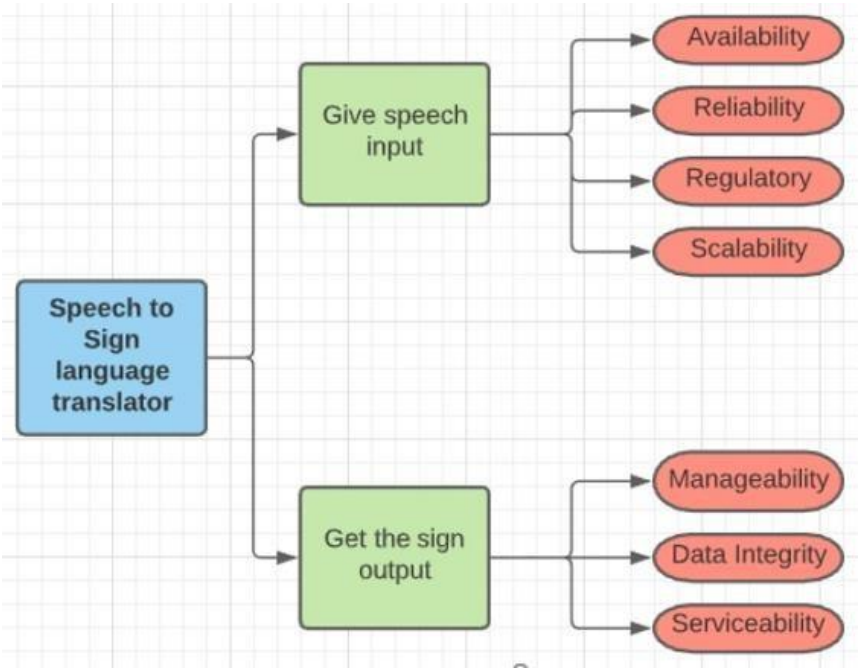

Figure 1. Functional and Non-functional requirements

5.2 Block Diagrams

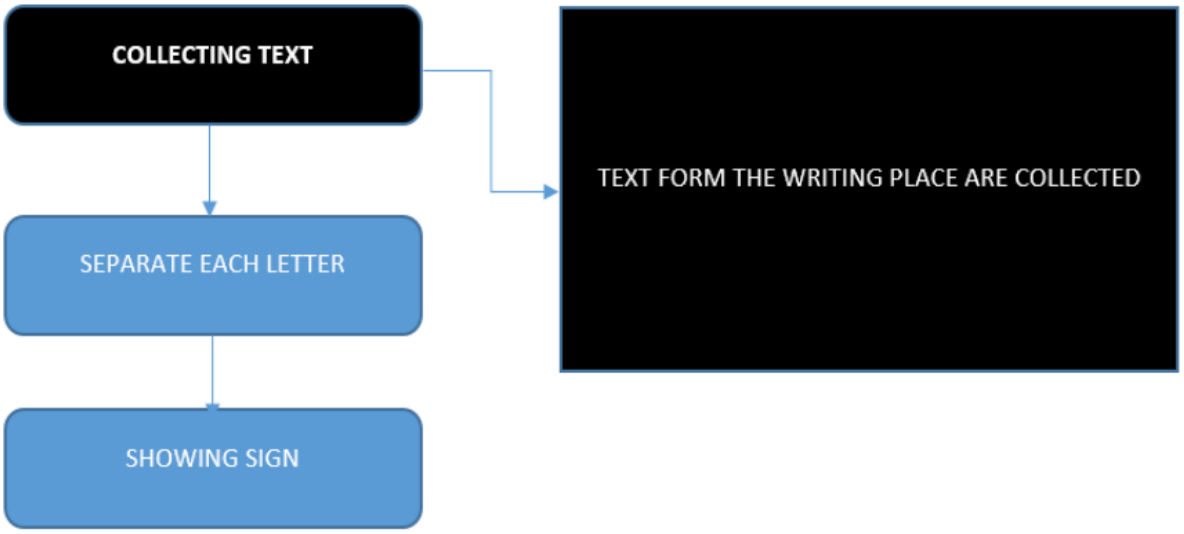

Figure 2. Text Collection

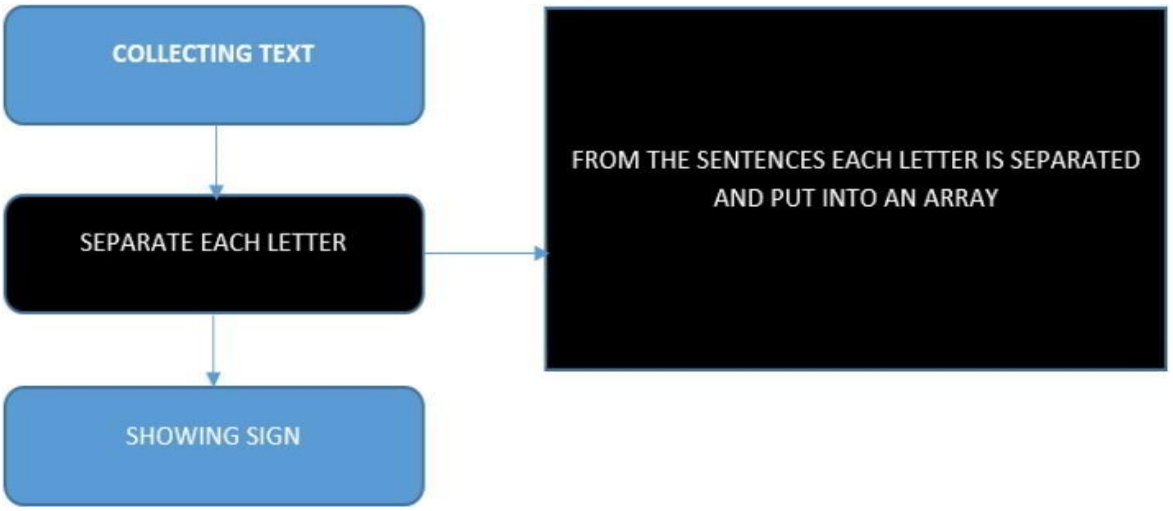

Figure 3. Text Separation

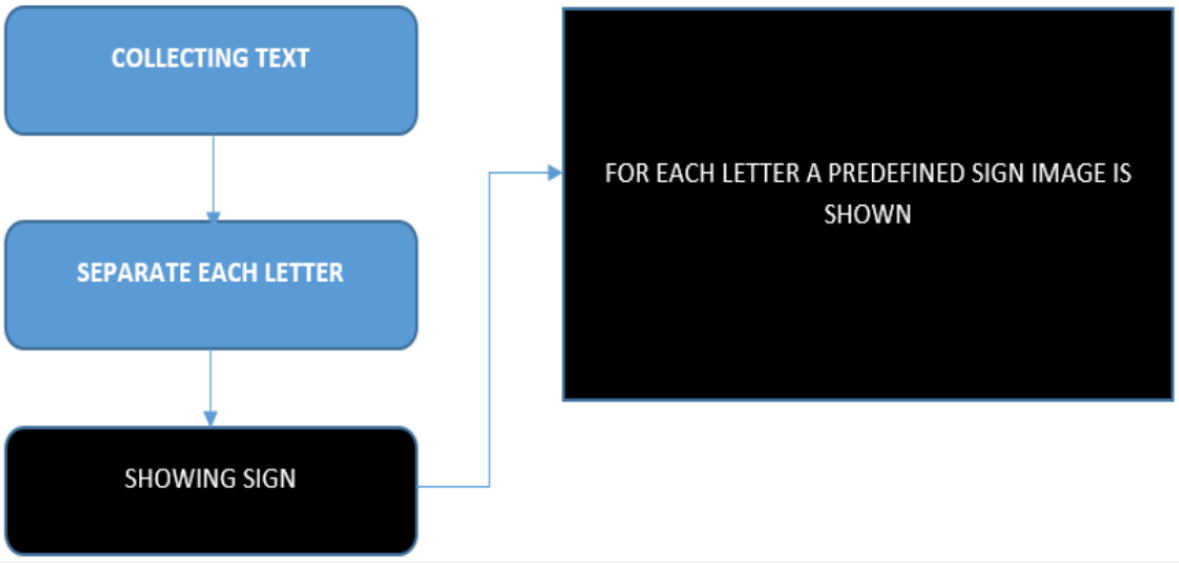

Figure 4. Text conversion to Sign Language

5.3 Audio to Text Conversion

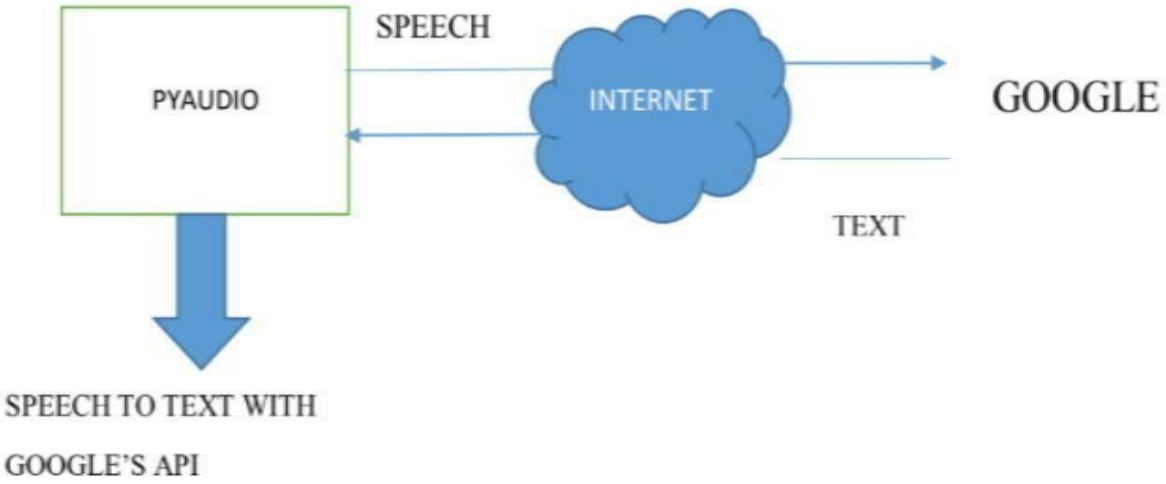

Figure 5. Google Speech-to-Text (Converts audio to text by applying neural network models in an easy-to-use API)

5.4 Pre-Processing of Text

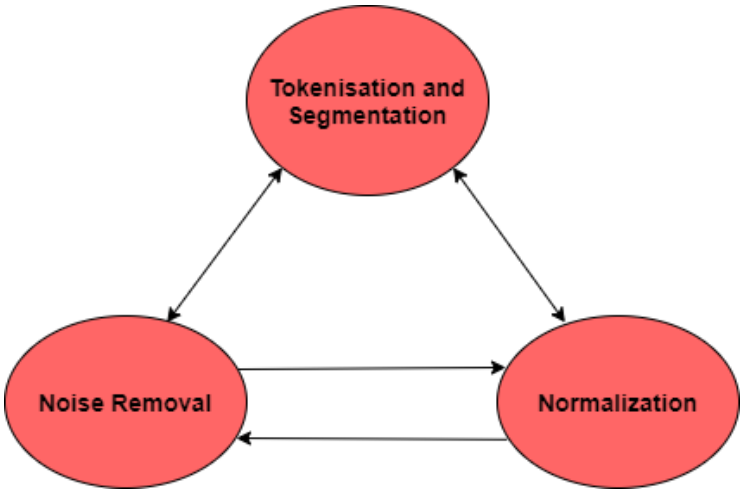

Figure 6. Text Preprocessing

6. Methodology:

6.1 Tools employed:

An application which takes in live speech or audio recording as input, converts it into textand displays the relevant Indian Sign Language images or GIFs.

- Front-end using EasyGui.
- Speech as input through microphone using PyAudio.
- Speech recognition using Google Speech API.
- Text Preprocessing using NLP.
- Dictionary based machine translation is done.
- Implementation in Visual Studio

6.2 Work Breakdown Structure:

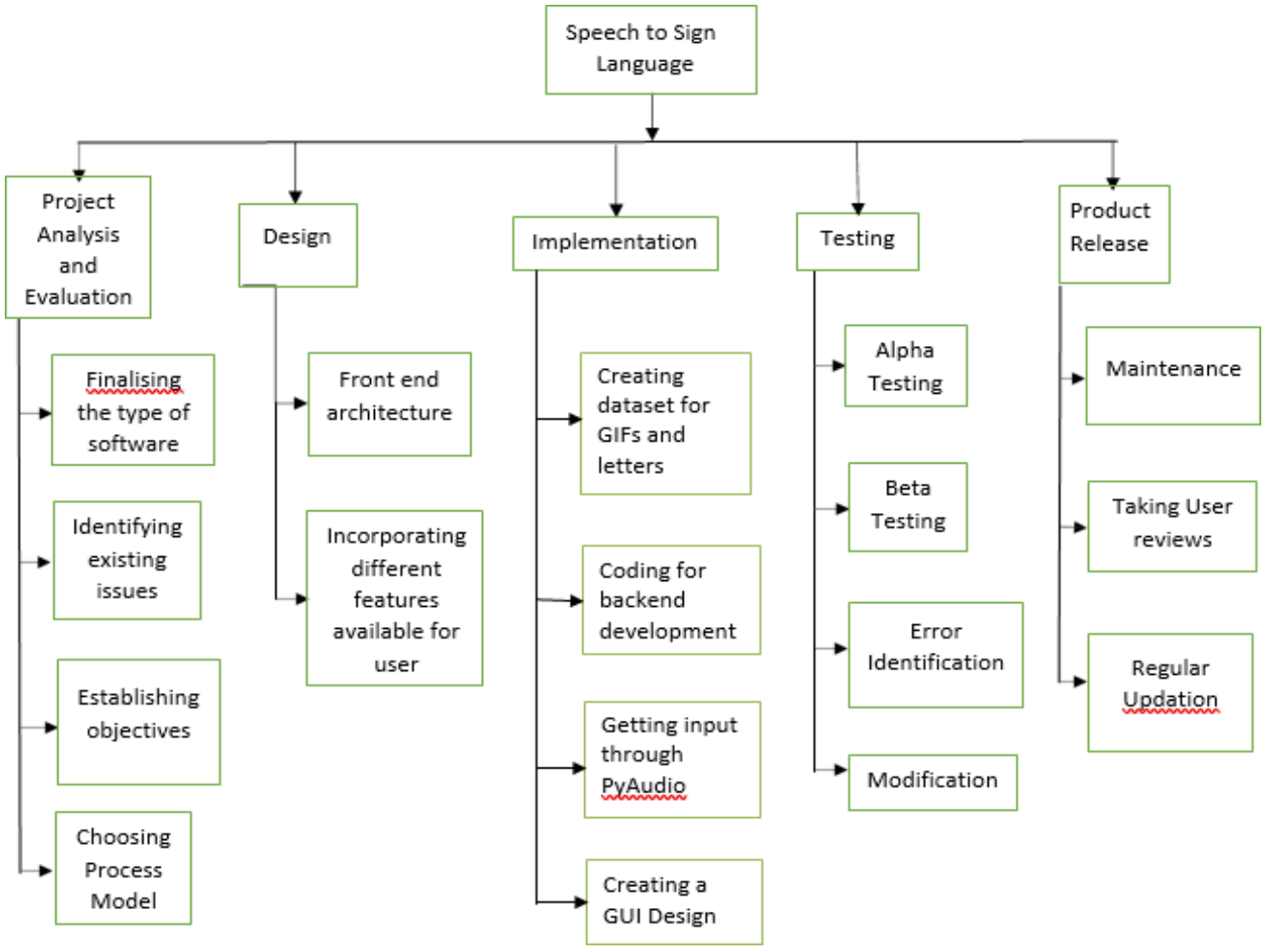

Figure 7. Work Breakdown Structure

6.3 Stakeholders and USE Case Diagram:

1. Project Managers
2. Project Sponsors
3. Coders
4. Executives
5. Programmers
6. Software Developers

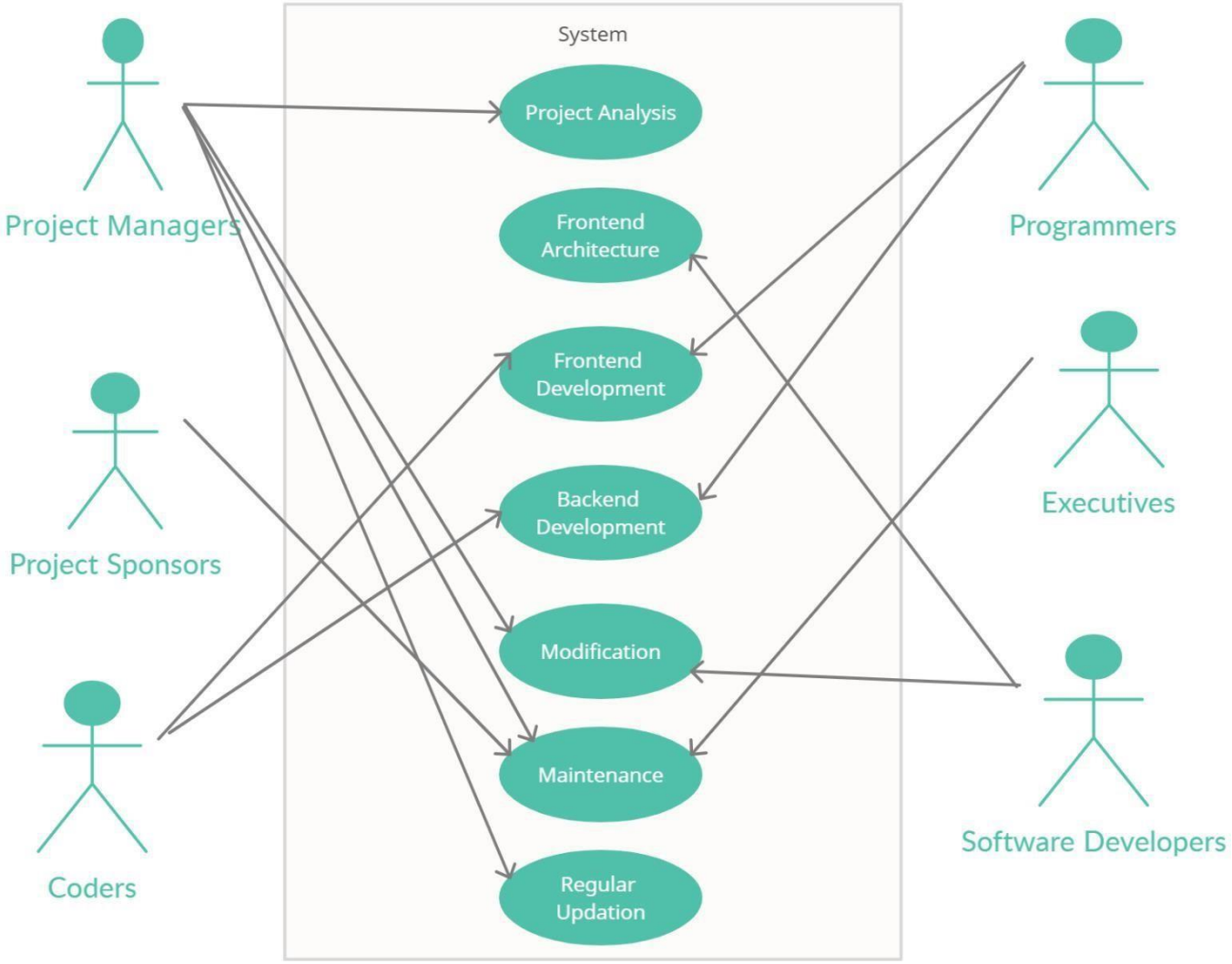

Figure 8. USE Case Diagram

6.4 Predefined Gestures:

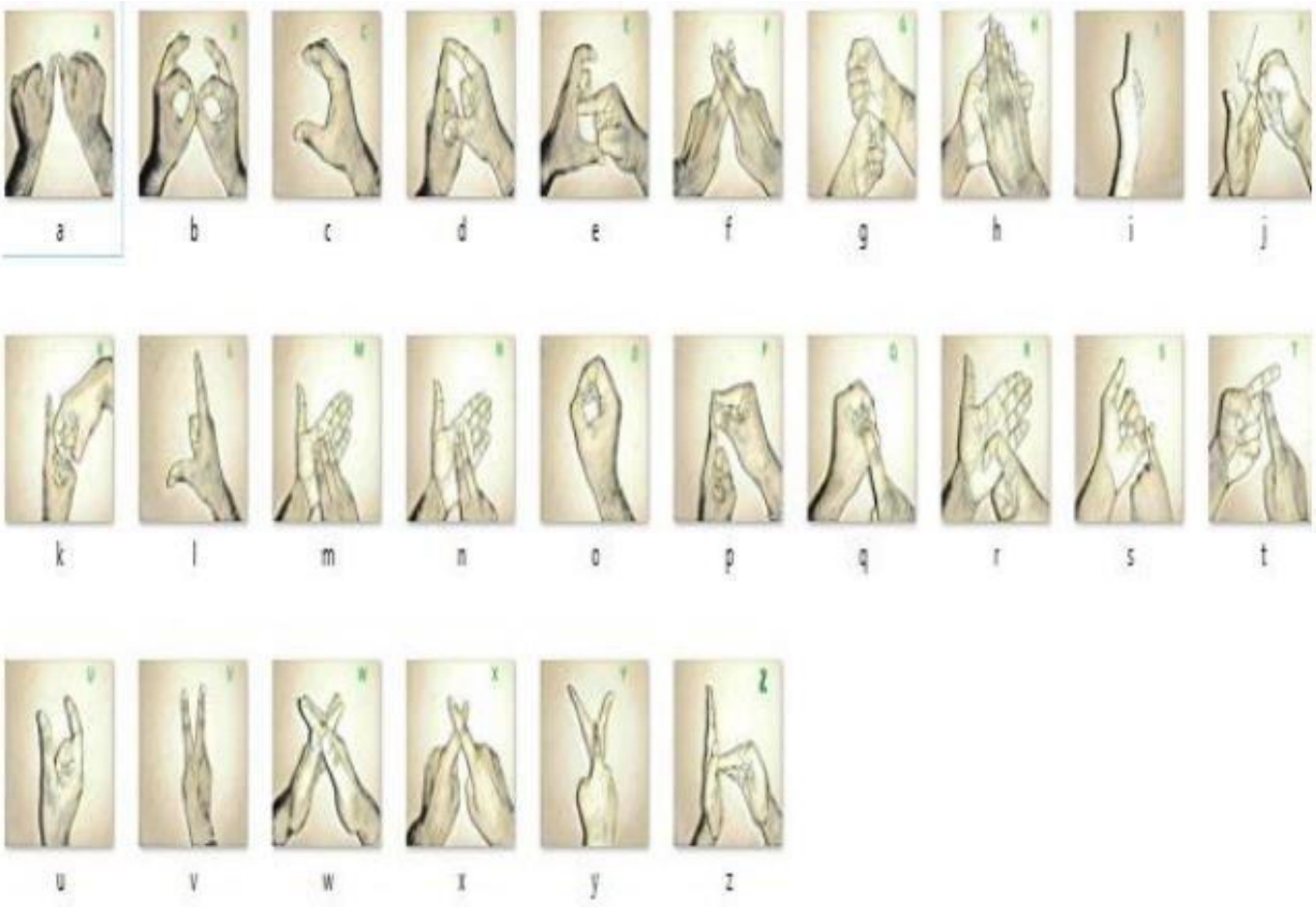

Figure 9. Hand Signs for each English alphabet

7. Modules:

7.1 Audio to Text Conversion:

1. Audio input is taken using python PyAudio module.
2. Conversion of audio to text using microphone
3. Dependency parser is used for analyzing grammar of the sentence and obtaining relationship between words.

7.2 Text to Sign Language:

1. Speech recognition using Google Speech API.
2. Text Preprocessing using NLP.
3. Dictionary based Machine Translation.
4. ISL Generator: ISL of input sentence using ISL grammar rules.
5. Generation of Sign language with signing Avatar.

8. Implementation:

8.1 Code:

```
import speech_recognition as sr

import numpy as np

import matplotlib.pyplot as plt

import cv2

from easygui import *

import os

from PIL import Image, ImageTk

from itertools import count import

tkinter as tk

import string

#import selecting

# obtain audio from the microphonedef

func():

    r = sr.Recognizer()

    isl_gif=['any questions', 'are you angry', 'are you busy', 'are you hungry', 'are you sick','be careful',

        'can we meet tomorrow', 'did you book tickets', 'did you finish homework', 'do yougo to office', 'do you have money',

        'do you want something to drink', 'do you want tea or coffee', 'do you watch TV','dont worry', 'flower is beautiful',

        'good afternoon', 'good evening', 'good morning', 'good night', 'good question','had your lunch', 'happy journey',

        'hello what is your name', 'how many people are there in your family', 'i am aclerk', 'i am bore doing nothing',

         'i am fine', 'i am sorry', 'i am thinking', 'i am tired', 'i dont understand anything', 'igo to a theatre', 'i love to shop',

        'i had to say something but i forgot', 'i have headache', 'i like pink colour', 'i live innagpur', 'lets go for lunch', 'my mother is a homemaker',

        'my name is john', 'nice to meet you', 'no smoking please', 'open the door','please call me later',

        'please clean the room', 'please give me your pen', 'please use dustbin dontthrow garbage', 'please wait for sometime', 'shall I help you',
```

'shall we go together tommorow', 'sign language interpreter', 'sit down', 'standup', 'take care', 'there was traffic jam', 'wait I am thinking',

'what are you doing', 'what is the problem', 'what is todays date', 'what is yourfather do', 'what is your job',

'what is your mobile number', 'what is your name', 'whats up', 'when is yourinterview', 'when we will go', 'where do you stay',

'where is the bathroom', 'where is the police station', 'you are wrong','address','agra','ahemdabad', 'all', 'april', 'assam', 'august', 'australia', 'badoda','banana', 'banaras', 'banglore',

'bihar','bihar','bridge','cat', 'chandigarh', 'chennai', 'christmas', 'church', 'clinic', 'coconut', 'crocodile','dasara',

'deaf', 'december', 'deer', 'delhi', 'dollar', 'duck', 'febuary', 'friday', 'fruits', 'glass', 'grapes','gujrat', 'hello',

'hindu', 'hyderabad', 'india', 'january', 'jesus', 'job', 'july', 'july', 'karnataka', 'kerala', 'krishna','litre', 'mango',

'may', 'mile', 'monday', 'mumbai', 'museum', 'muslim', 'nagpur', 'october', 'orange', 'pakistan','pass', 'police station',

'post office', 'pune', 'punjab', 'rajasthan', 'ram', 'restaurant', 'saturday', 'september', 'shop','sleep', 'southafrica',

'story', 'sunday', 'tamil nadu', 'temperature', 'temple', 'thursday', 'toilet', 'tomato', 'town','tuesday', 'usa', 'village',

'voice', 'wednesday', 'weight','please wait for sometime','what is your mobile number','whatare you doing','are you busy']

```
arr=['a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p','q','r', 's','t','u','v','w','x','y','z']
with sr.Microphone() as source:#
    image        = "signlang.png"
    # msg="HEARING IMPAIRMENT ASSISTANT"
    # choices = ["Live Voice","All Done!"]
    # reply     = buttonbox(msg,image=image,choices=choices)
    r.adjust_for_ambient_noise(source)
    i=0
    while True:
        print("I am Listening") audio
        = r.listen(source)
```

```
                # recognize speech using Sphinxtry:
                    a=r.recognize_google(audio)a =
                    a.lower()
                    print('You Said: ' + a.lower())

                    for c in string.punctuation:a=
                        a.replace(c,"")

                    if(a.lower()=='goodbye' or a.lower()=='good bye' or a.lower()=='bye'):
                            print("oops!Time To say good bye")
                            break

                    elif(a.lower() in isl_gif):


                        class ImageLabel(tk.Label):
                                """a label that displays images, and plays them if they are
gifs"""
                                def load(self, im):
                                    if isinstance(im, str):
                                        im = Image.open(im)
                                    self.loc = 0
                                    self.frames = []

                                    try:
                                        for i in count(1):
                                            self.frames.append(ImageTk.PhotoImage(im.copy()))
                                            im.seek(i)
                                    except EOFError:
                                        pass

                                    try:
```

```
                    self.delay = im.info['duration']
                except:
                    self.delay = 100

                if len(self.frames) == 1:
                    self.config(image=self.frames[0])
                else:
                    self.next_frame()

            def unload(self):
                self.config(image=None)
                self.frames = None

            def next_frame(self):if
                self.frames:
                    self.loc += 1
                    self.loc %= len(self.frames)
                    self.config(image=self.frames[self.loc])
                    self.after(self.delay, self.next_frame)
        root = tk.Tk()
        lbl = ImageLabel(root)
        lbl.pack()
        lbl.load(r'ISL_Gifs/{0}.gif'.format(a.lower()))
        root.mainloop()
    else:
        for i in range(len(a)):
                    if(a[i] in arr):

                        ImageAddress = 'letters/'+a[i]+'.jpg' ImageItself =
                        Image.open(ImageAddress) ImageNumpyFormat =
                        np.asarray(ImageItself)
                        plt.imshow(ImageNumpyFormat)
```

```
                                        plt.draw()
                                        plt.pause(0.8)
                                    else:
                                        continue


                    except:
                        print(" ")
                    plt.close()
while 1:
  image     = "signlang.png"
  msg="HEARING IMPAIRMENT ASSISTANT"
  choices = ["Live Voice","All Done!"]
  reply    = buttonbox(msg,image=image,choices=choices)if reply
  ==choices[0]:
      func()
  if reply == choices[1]:
      quit()
```

**Separate dataset consisting of all GIFs has been manually created for this project. Alongwith this, another dataset with predefined gestures corresponding to every alphabet has also been made.

**A picture from external sources was loaded into the software for GUI Design.

8.2 GUI Design:

Figure 10. GUI Design for front end by EasyGui

8.3 Speech input:

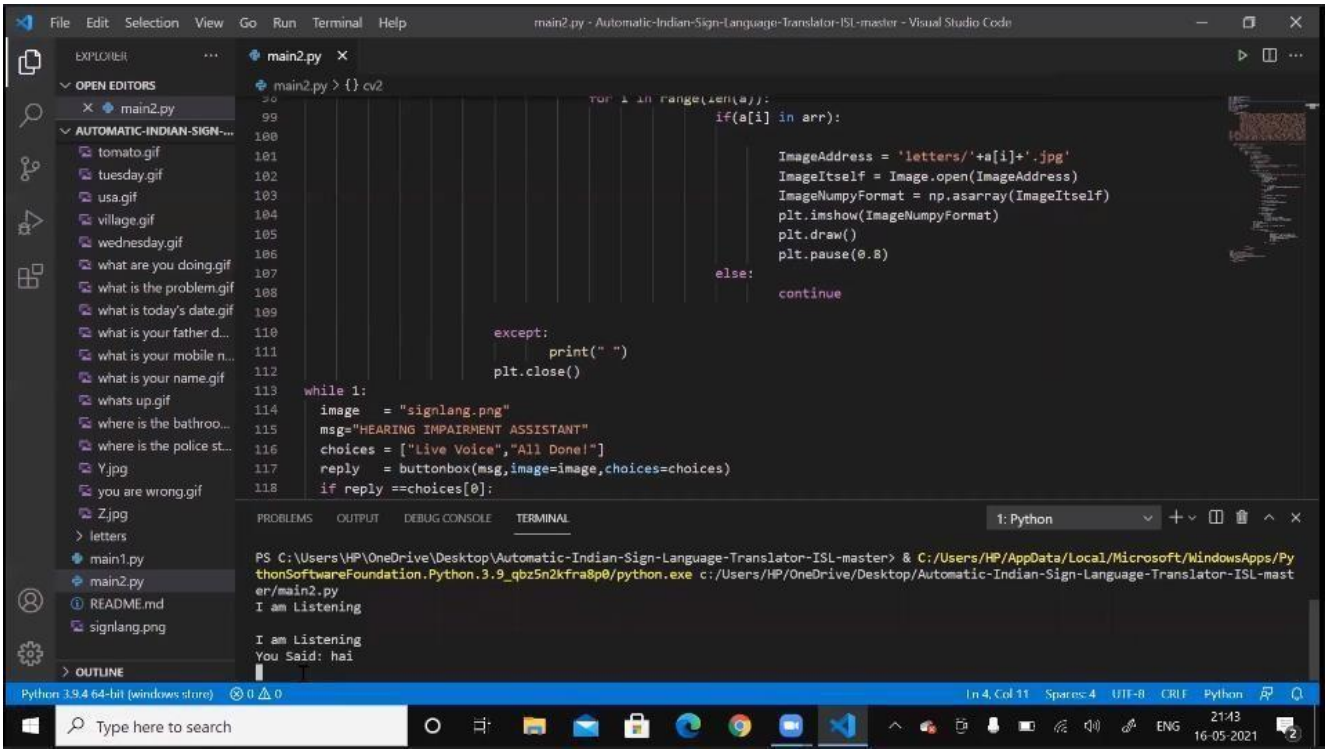

Figure 11. Speech Input taken through microphone using PyAudio package

8.4 Text Pre-processing:

Machine can understand binary language (i.e., 0 and 1) only. To make the machine understand human language, NLP was introduced. Natural Language Processing is the ability of the machine where it processes the text that was said and structures it. NLP creates an algorithm that translates text into word by labelling them based on the position and function of the words in the sentences. Human language is converted meaningfully into a numerical form. This allows computers to understand the nuances implicitly encoded into our language. Text preprocessing consists of three things - Tokenization, Normalization and Noiseremoval.

NLP helps the machine in understanding human language through the following steps:

1. We give audio as input to the machine.
2. The machine records that audio input.
3. Then machine translates the audio into text and displays it on the screen.
4. The NLP system parses the text into components; understand the context of the conversation and the intention of the person.
5. The machine decides which command to be executed, based on the results of NLP.

9. Results:

9.1 Letter-by-letter output:

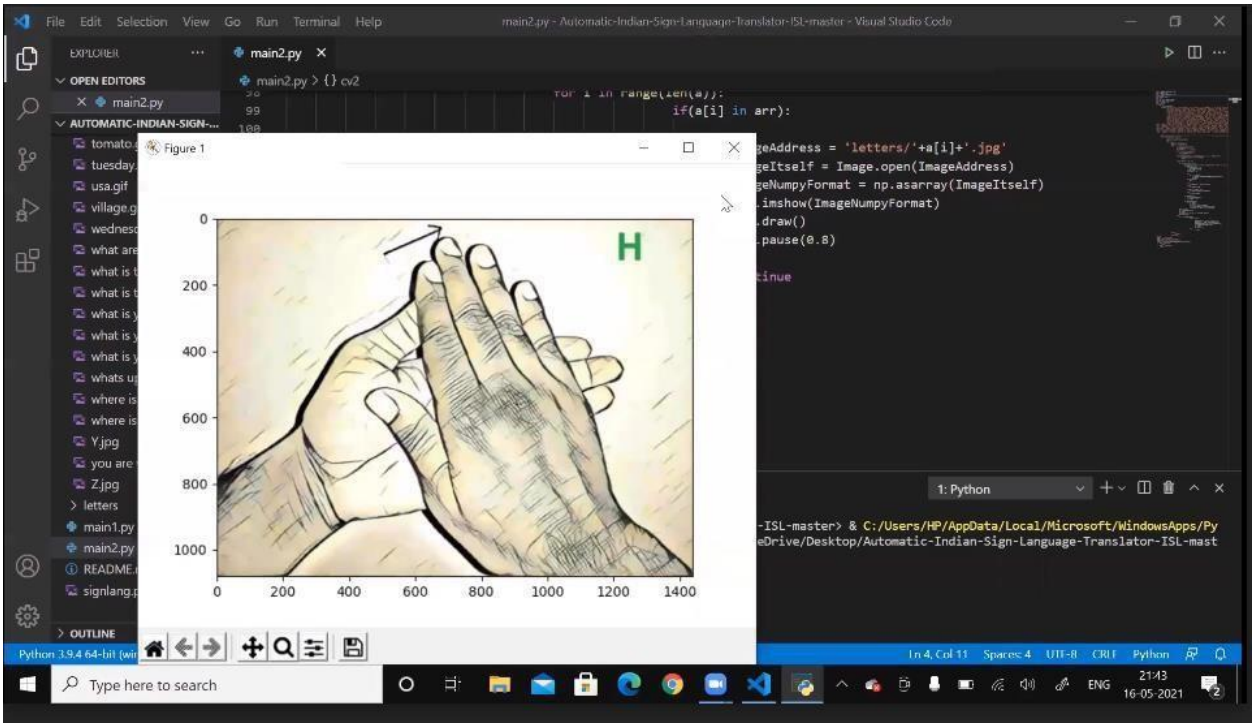

Figure 12. Word by word output (Letter "H")

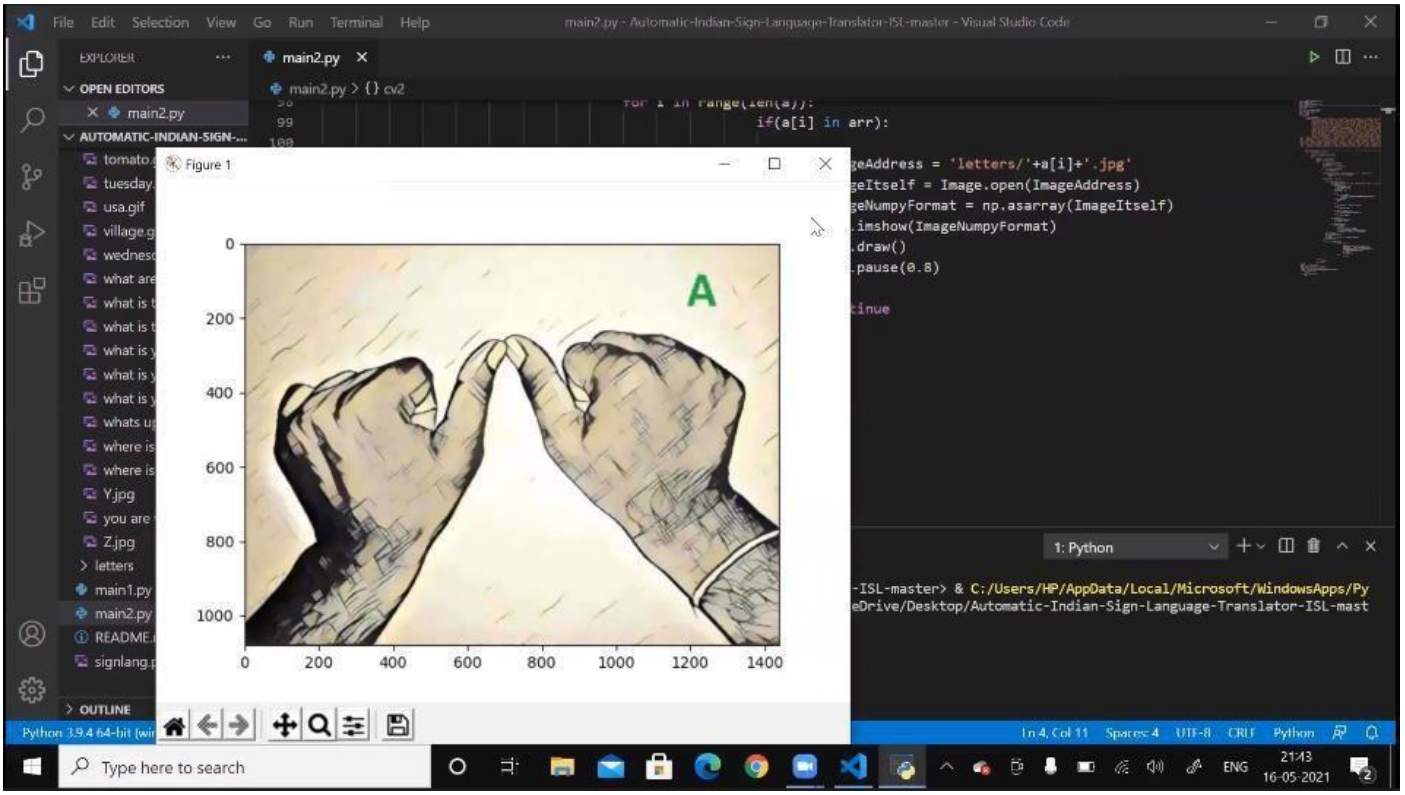

Figure 13. Word by word output (Letter “A”)

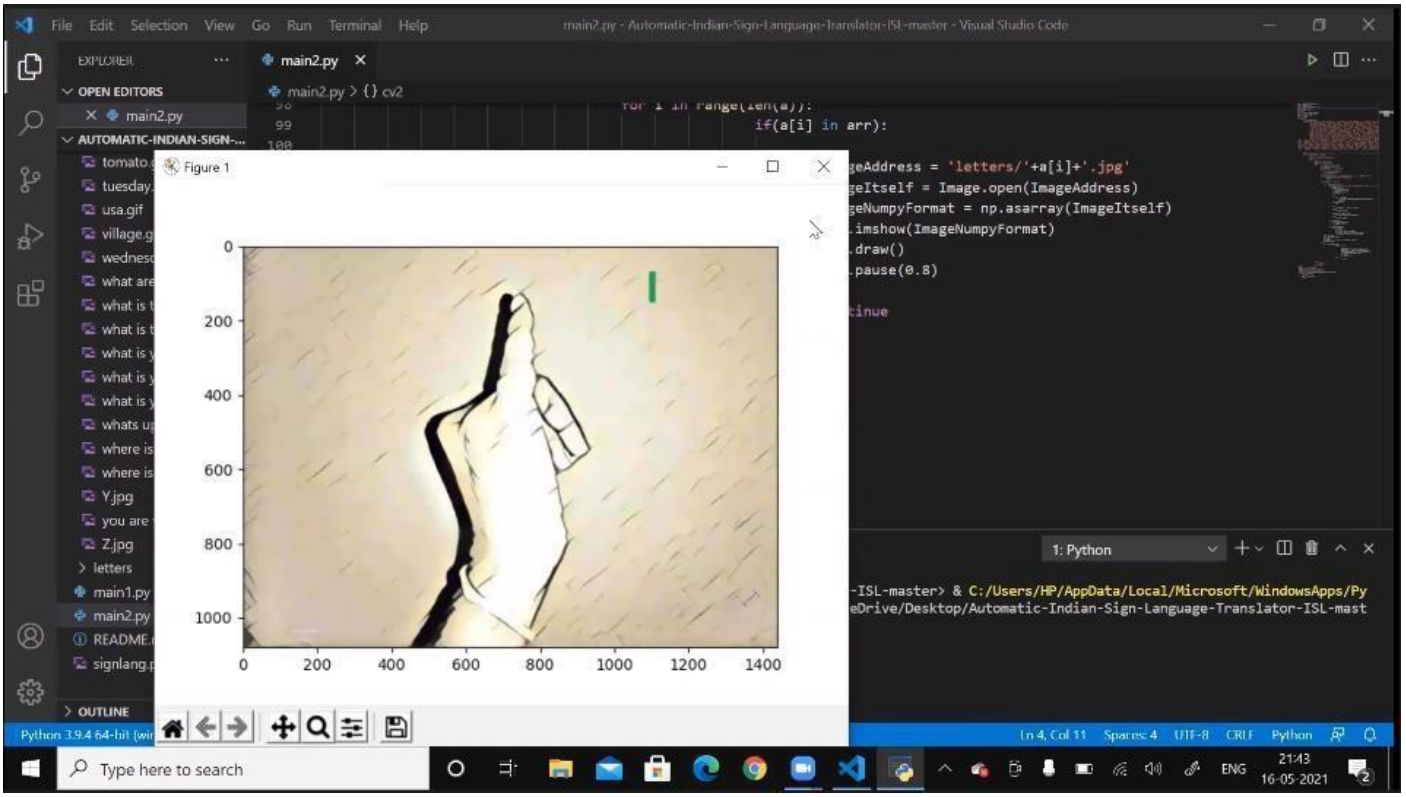

Figure 14. Word by word output (Letter “I”)

9.2 GIF Output:

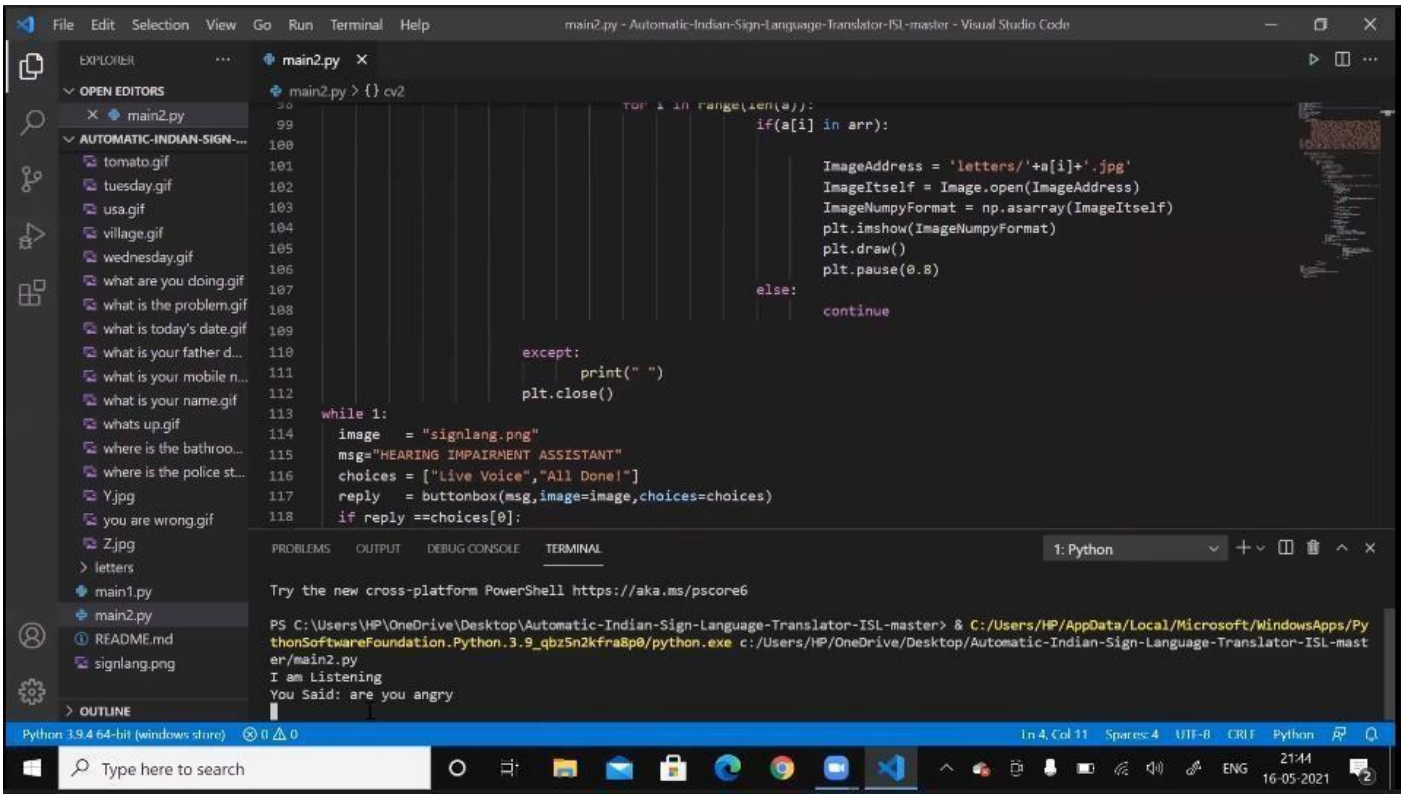

Figure 15. Speech input of an existing phrase

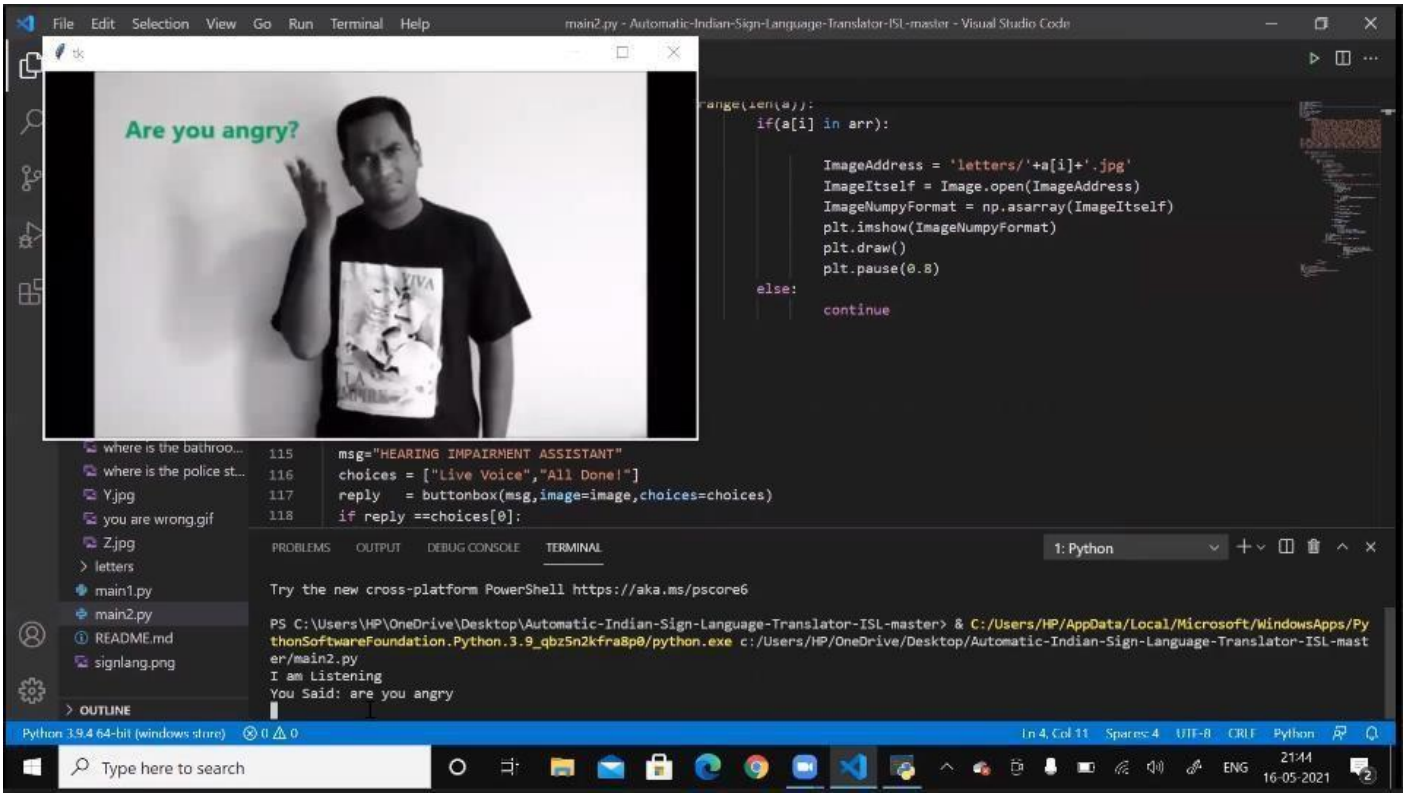

Figure 16. GIF Output

9.3 Test cases for Homophones:

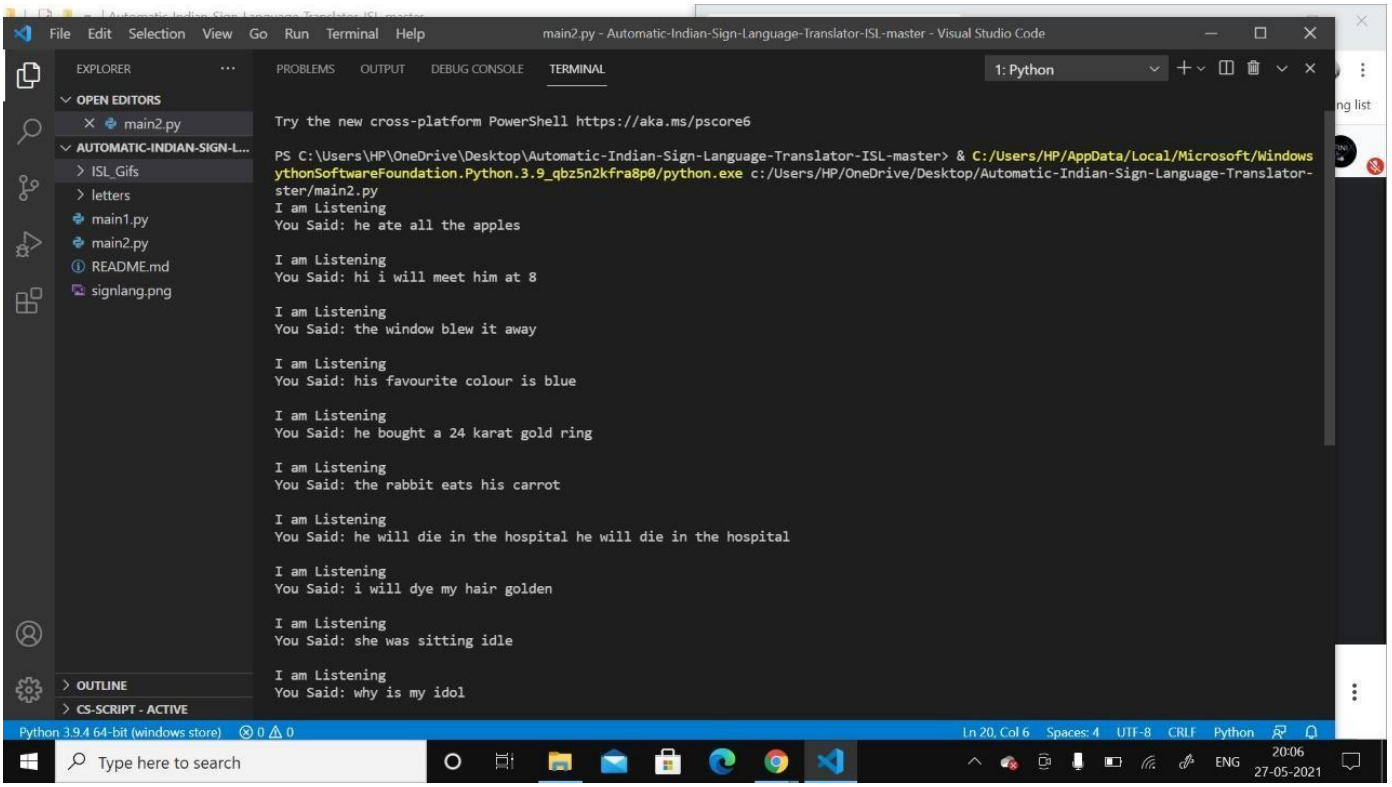

Figure 17. Test cases for homophones

Test Case ID	Test Objective	Test Data	Expected Result	Actual Result	Pass/Fail
1001	Checking speech recognition for homophones	Sentence 1	ate	ate	Pass
1002	Checking speech recognition for homophones	Sentence 2	8	8	Pass
1003	Checking speech recognition for homophones	Sentence 3	blew	blew	Pass
1004	Checking speech recognition for homophones	Sentence 4	blue	blue	Pass
1005	Checking speech recognition for homophones	Sentence 5	karat	karat	Pass
1006	Checking speech recognition for homophones	Sentence 6	carrot	carrot	Pass
1007	Checking speech recognition for homophones	Sentence 7	die	die	Pass
1008	Checking speech recognition for homophones	Sentence 8	dye	dye	Pass
1009	Checking speech recognition for homophones	Sentence 9	idle	idle	Pass
1010	Checking speech recognition for homophones	Sentence 10	idol	idol	Pass

We see a 100% accuracy when it comes to predicting the spelling of words that sound the same but have different meanings and are spelt differently. This is because Google Speech API uses n-gram model to predict words while processing. For e.g., in the sentence “He ate all his apples”, it doesn’t make sense to use “8” instead of “ate”. Therefore, because of n-gram model, Google Speech API predicts the correct spelling by checking the words preceding it.

The link of the paper discussing the use of n-gram models in Google Speech Recognition: 43819.pdf (googleusercontent.com)

9. Conclusion

The project has been successfully implemented. We have created an application that has proven to be a big help to people suffering from hearing disabilities. The report talks at length about the necessity of such a project, as well as its societal impacts. Our project uses PyAudio to take audio input and converts it to text using text-preprocessing. The final text is then ran through the database of the GIFs. If a match is found, the output in the form of corresponding GIF is given. If not, then the text is broken letter-by-letter and the hand signal of each letter is then plotted on a graph as the final output. The time lag between 2 successive plots is 0.8 seconds.

10. Utility

Form link:

https://forms.gle/mS9sDSNmovapiHRE7

Total response: 126

Results:

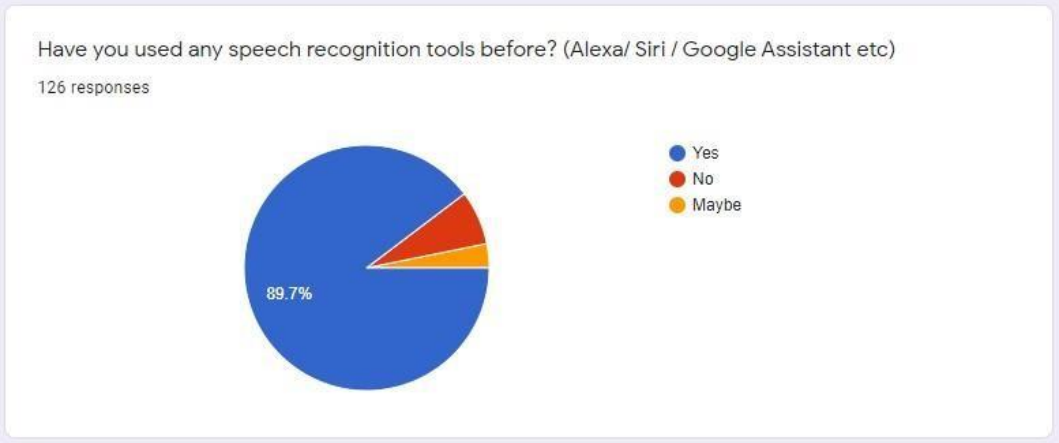

Figure 18

Insight: Most of the people have used a speech recognition tool before

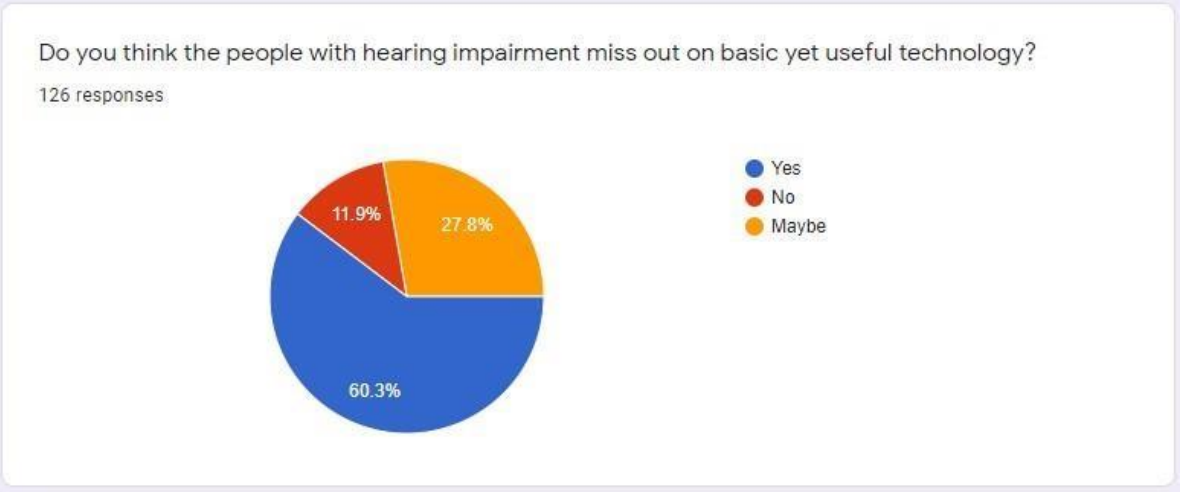

Figure 19

Insight: 81.1% of the people believe that people with hearing impairment miss out on basic but useful technology in some way or the other

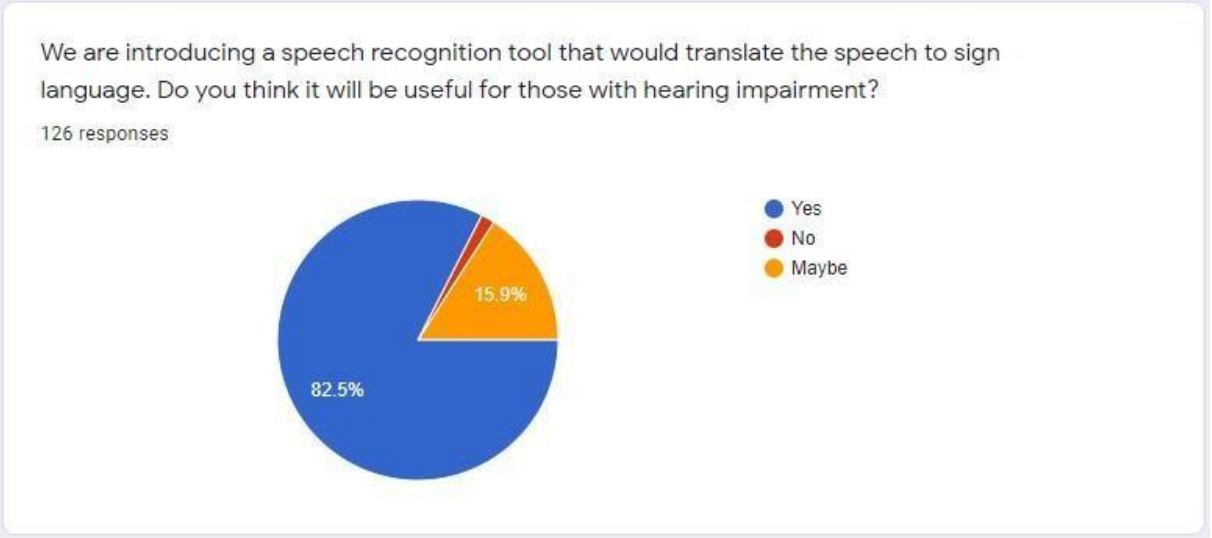

Figure 20

Insight: 98.4% of the people think that our tool will help people with hearing disability by making their lives easier through effective communication.

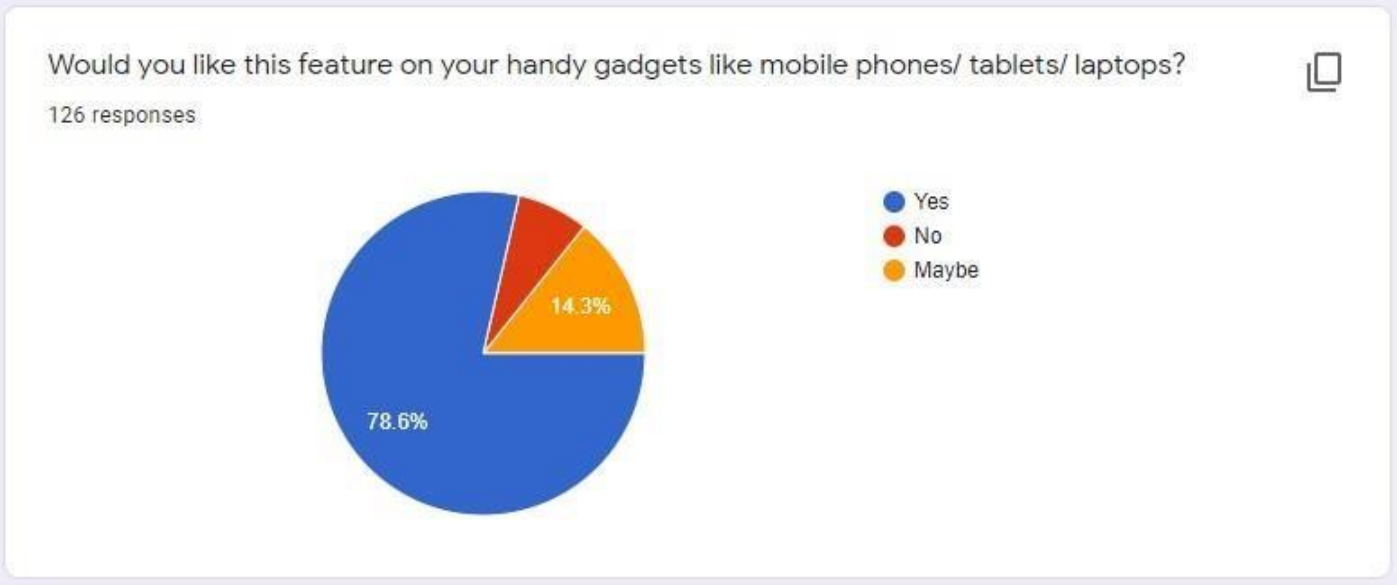

Figure 21

Insight: People largely believe that having such an app in phones will be helpful. It is the need of the hour so that people with hearing disabilities can cope up and live their lives as normal people with no communication barriers do.

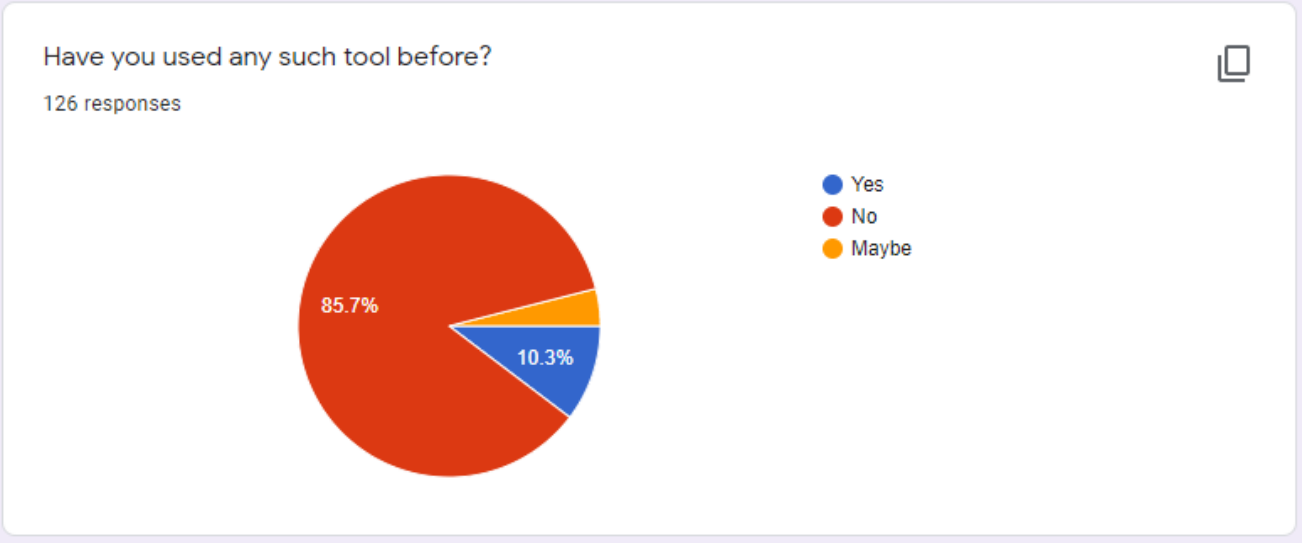

Figure 22

Insight: Mostly, people have never used such a tool which explains how necessary and relevant it is. It is our aim to make this tool available to everyone including normal people as well as those with hearing disability to bridge the communication gap between them.

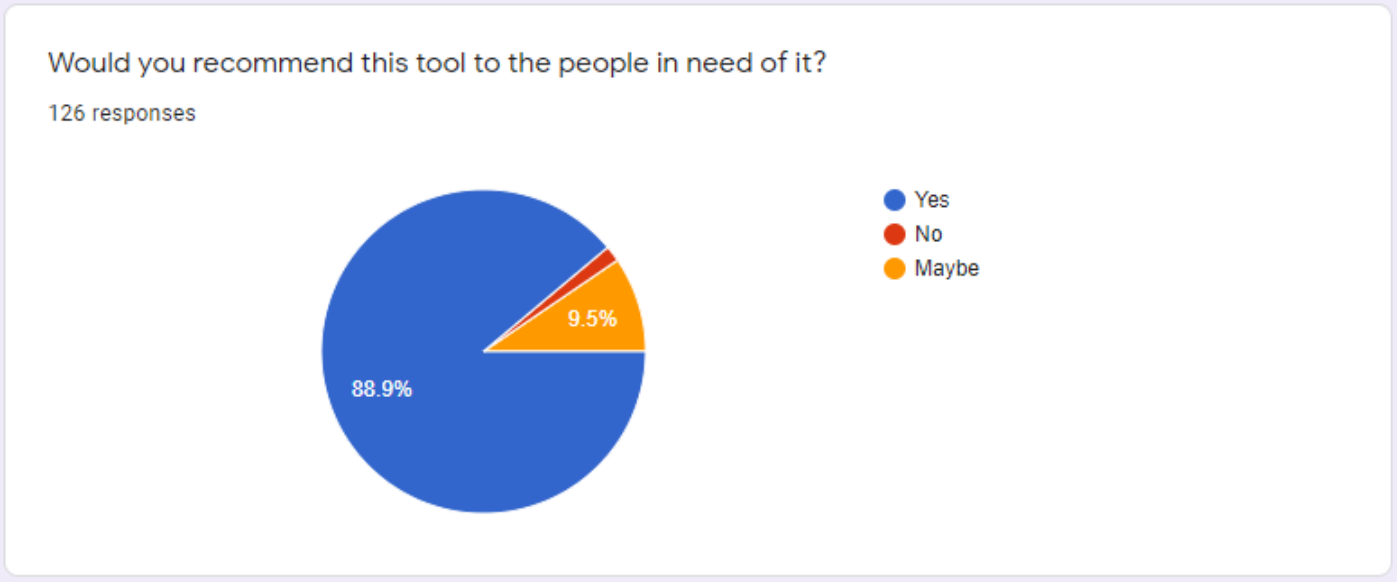

Figure 23

Insight: People largely recommend this tool to those with hearing issues, which shows it a general notion that a tool like this could be extremely important in the lives of deaf people.

9. Future Works:

Sign language translator is extremely useful in various areas. In schools, colleges, hospitals, universities, airports, courts, etc., anywhere anyone can use this system for understanding the sign language to communicate. It makes communication between a normal hearing person and a person with hearing disability easier.

The future work is to develop an application that can be used by news channel. A small screen at corner of the screen could display sign-language visuals so that a deaf person is also able to listen to the news. At present, only DD news is using this kind of presentation technique, but they show a person manually translating the newsreader and conveying the information through hand signals. Therefore, by using this application, we can convey news faster and without error.

We also look forward to expanding the project by including facial expressions into the system. This would make the application more user-friendly.