iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33829

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

Insight Reader using Deep Learning

Aarthi R[1], Anto Gifta Hebzibha A[2], Sabastin S[3]
[1, 2, 3]*Computer Science and Engineering, Loyola-ICAM College of Engineering and Technology, Chennai, India.*

Abstract: This paper aims to supply solution for the communication problems among the visually impaired, hearing/speech impaired or a standard person. The visually impaired can hear only voice while most of the hearing/speech impaired knows only the sign language. It is difficult for a visually impaired person to find out the sign language and also a standard one who doesn't know or understand sign language cannot communicate with them likewise. To unravel this problem we have built an application which will convert text to voice sign language, voice to text or sign language, sign language to text or voice. Thus our application helps them to speak with one another using their favourable languages.

I. INTRODUCTION

The main objective of this paper is to assist everyone to speak with each other regardless of their physical challenges. If a visually impaired person uses this application, the input which might be within the sort of text or sign language can be converted into voice message. If the speech/hearing impaired person uses this application, then the message may be converted into sign language. In both the cases, the input is received from the other person by clicking on the button which represents the kind of message i.e) text, voice, sign language. If the input that we are getting is voice then the microphone is switched on and therefore voice is being recorded. If the input is sign language then a picture is uploaded which suggests any of the 26 English alphabets or the numbers from 0 to 9. These images are set default and chosen by the user. Thus the people can give the input based on their needs and find it converted into other format of messages. Fig. 1 depicts the flow of the application.

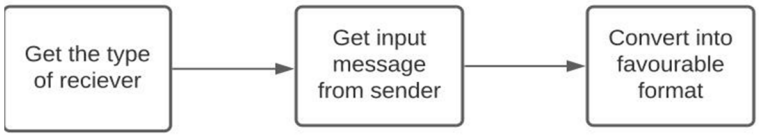


Fig. 1

II. PROPOSED SYSTEM

The proposed system is a python application which converts

A. Text to Speech,
B. Text to sign language,
C. Speech to Text,
D. Speech to sign language,
E. Sign language to Text and
F. Sign language to Speech.

There are totally six modules among which 4 are the basics. The fundamental modules are text to speech, text to sign language, speech to text and sign language to text. The other two modules like speech to sign language can be implemented by combining the speech to text and text to sign language modules and the sign language to speech can be done by combining the sign language to text and text to speech modules. Thus a python application is developed to make communication easier for everybody irrespective of their physical challenges.

Fig.2 shows the fundamental system architecture diagram.

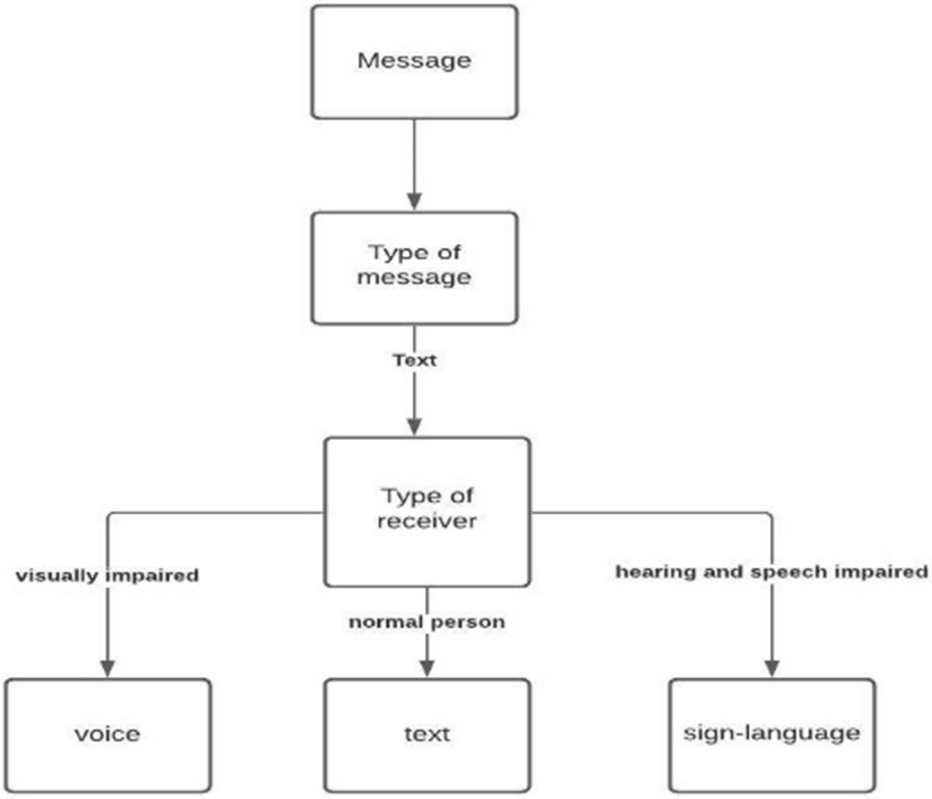


Fig.2

III. MODULES

There are 4 basic modules

A. Text to Speech

To convert the text to speech, a python library called “pyttsx3” is employed. Pyttsx3 could be a text to speech conversion library. It works offline with both python 2 and python 3. The application first invokes the pyttsx3.init() function to get the reference of the pyttsx3 library. And then the reference object is used to call the say(message) function which converts the message string into speech. Then the runAndwait function is used. This function waits for the engine to convert text to speech, and do not allow the engine to close. After the process is over stop() function is used shutdown the engine.

B. Speech to Text

In this module pyaudio is used to record the audio. pyaudio provides python bindings for portaudio which is audio library. After the audio is recorded, python library called speech_recognition is used to convert audio to text.

C. Text to Sign Language

Every 26 alphabets of the English language and the numbers from 0 to 9 are assigned an image which depicts their sign language. When user use the text and clicks on “convert” button, the respective sign for every letter is retrieved from the database and is showed to the receiver.

D. Sign Language to Text

To convert sign language to text, Convolutional Neural Network (CNN) is used. The system is trained with default images of 26 alphabets of the English language and the numbers from 0 to 9. When a hearing/speech impaired wants to send a message, the upload this trained images to the application which then the uses CNN to identify the message and convert it into text.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

IV. ALGORITHM

The algorithm used to recognize sign language is "Convolutional Neural Network". CNN is a class of deep neural network which is used for image classification. It consists of multiple layers of artificial neurons. The basic building block of Neutral Network is Neurons. They get one or more inputs, perform some computations on the input and gives the output.

Fig 3.1 shows the functionality of the neurons.

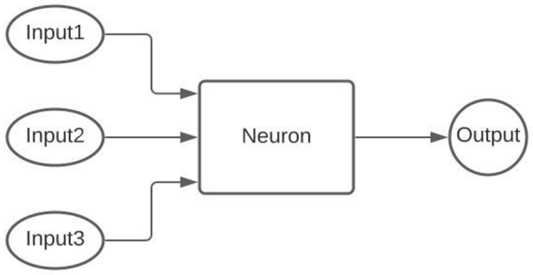


Fig 3.1

Thus the neurons are mathematical functions that gets multiple inputs and convert them into an output. There are 3 layers in CNN.

A. Convolution Layer

It gets the image as the input and detects the features of the image. This is done by performing the operation of the convolution between the image and a filter which is a 2D array of a particular size. The filter is slider over the image to get the dot product. The output of this layer is called the feature map which contains the various features of the image like the corners and edges.

B. Pooling Layer

This layer reduces the computational costs by reducing the size of the feature map and decreasing the connection between the various layers. There are various types of pooling, average pooling. These pooling operations are like filters applied to the feature map. The size of the filter is smaller than the size of the feature map so as to reduce the size of the feature map.

C. Fully Connected Layer

In this layer, the neurons of each layer is connected to all the neurons of the previous layer i.e) pooling layer. All the mathematical functions are in this layer. The classification of images takes place in this layer. This layer is placed before the output layer. The features of the image from the previous layer is the input to this layer and the mathematical functions are used in this to classify the images. Fig 3.2 represents the different layer of CNN.

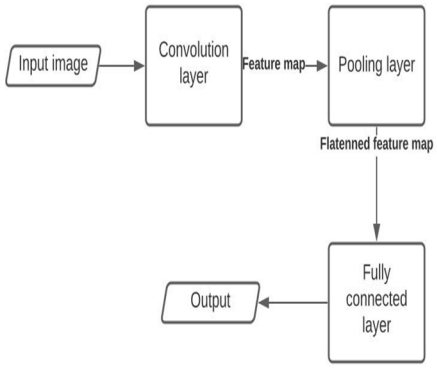


Fig 3.2

V. CONCLUSION

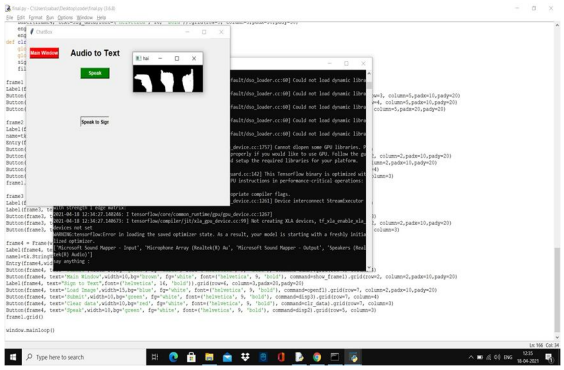
The application is built in such a way to help visually impaired, hearing/speech impaired and a normal person to communicate with each other. pyaudio and pysttx3 are used for converting text to speech and speech to text. For sign language recognition, CNN is used which classifies the images.

A. Future Work

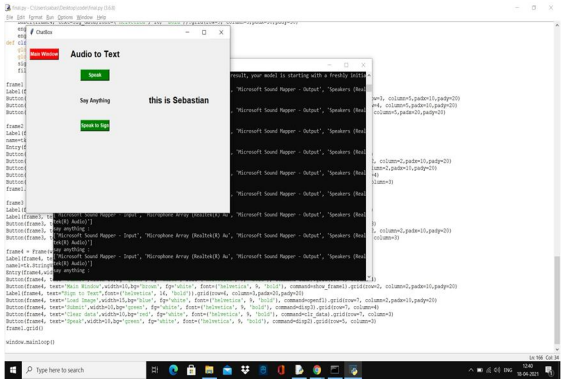
In future, this application can be built as a web application for a live chat so that the sender and receiver need not be in the same physical location. For hearing/speech impaired person, sign language can be directly read from the camera instead of using default images. For visually impaired person this application can be made as voice controlled.

VI. IMPLEMENTATION

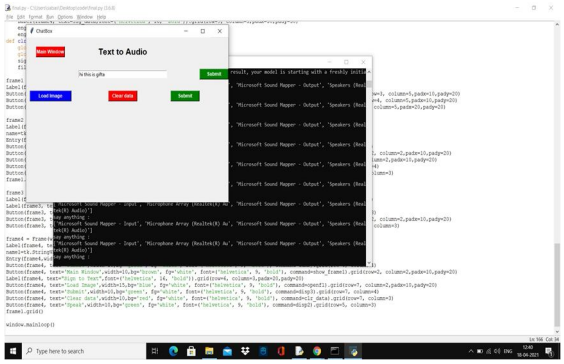
This is the main window which is meant for converting a text to voice, voice to text, text to sign language and voice to sign language.



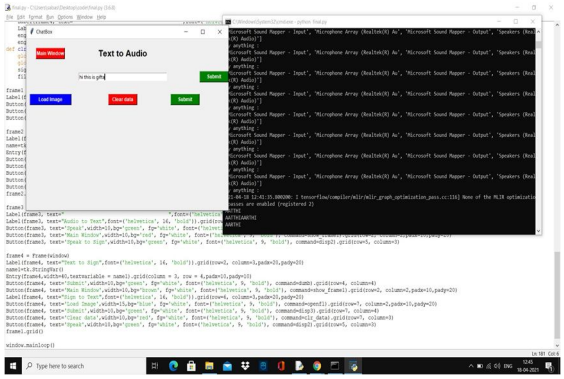
It is used for converting Audio to Sign language.



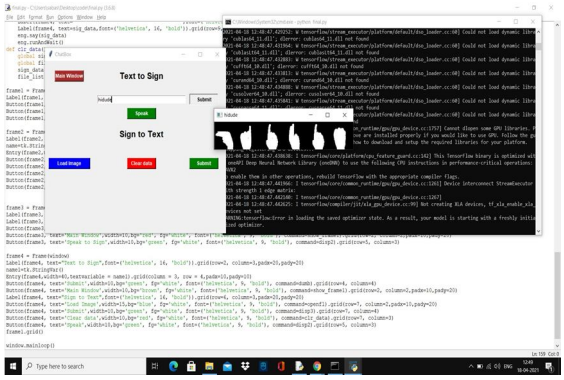
It is used for converting Audio to Text.



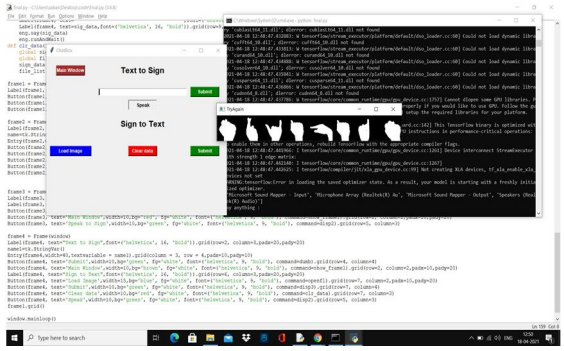
This window is used to convert Text to Audio.



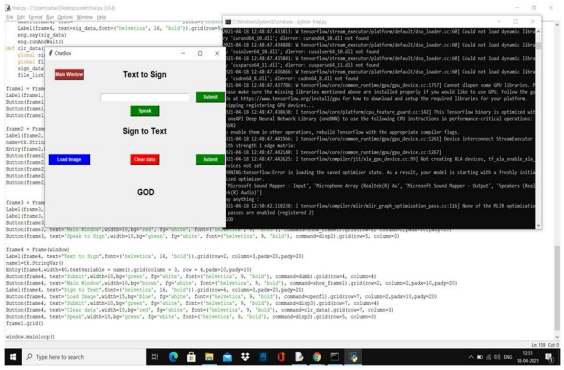
This is the window where the audio is played.



It is the window which is meant for converting Text to Sign language.



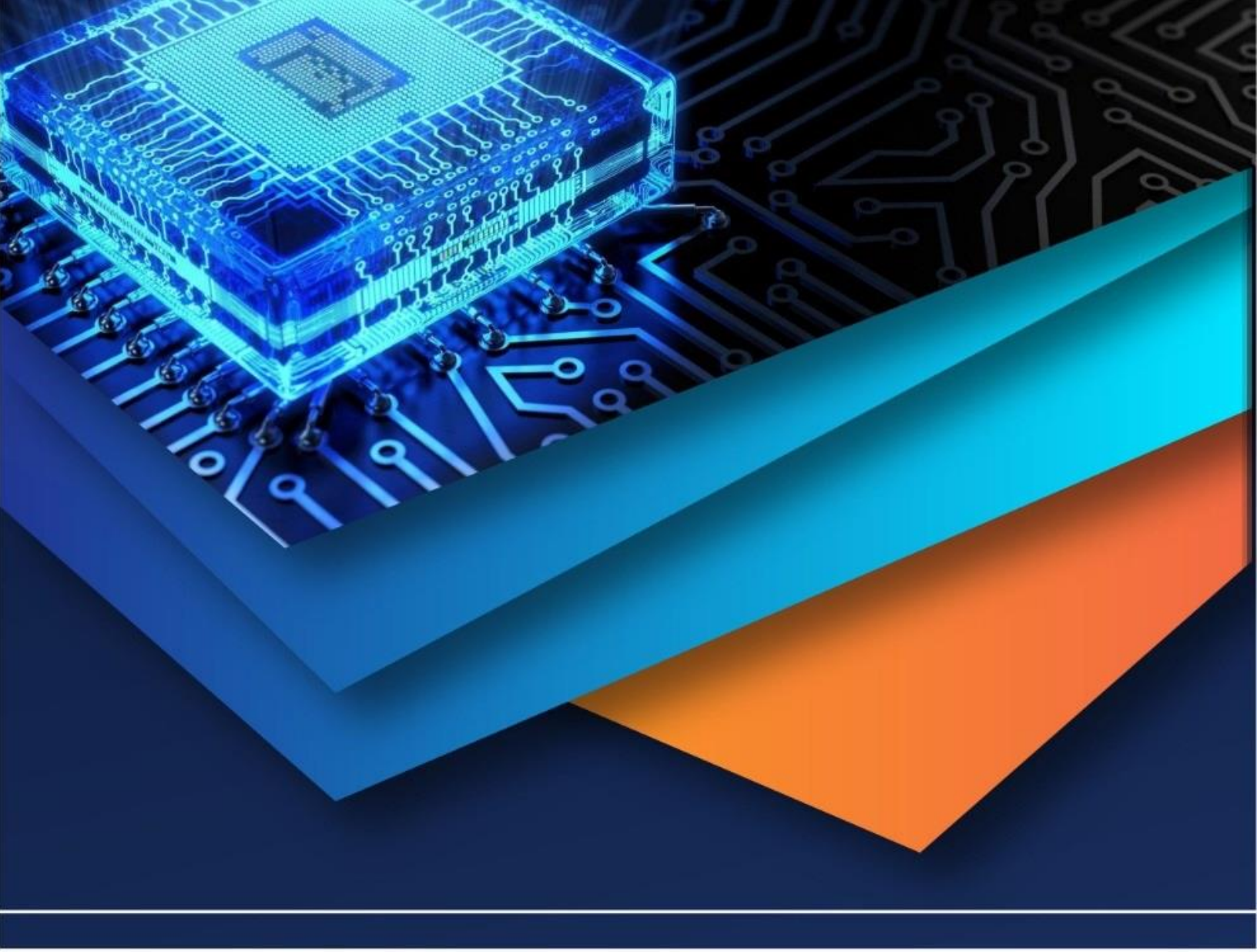
This is used for converting Audio to Sign language when Speak button is enabled.



This is used for converting the sign language to text by loading the image one by one.

REFERENCES

[1] Shadman Shahiar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar, Celia Shahnaz, Shaikh Anowarul Fattah. “Real-time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning”, 2018.
[2] Manasa Srinivasa H S, Laxmi Tyapi, Suresha H S. “Online Hand Gesture Recognition by OpenCV”, May 2015.
[3] Brandon Garcia, Sigberto Alarcon, Viesca. “Real-time American Sign Language Recognition with Convolutional Neural Network”.
[4] Karan Chauhan, Shrwan Ram. “Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensor Flow and Keras”, 2018.
[5] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr.Dayananda. “Sign Language to Text and Speech Recoginition in Real Time Using Convolutional Neural Network”, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY