

Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	21 October 2022
Team ID	PNT2022TMID50460
Project Name	Project – Real -Time Communication System Powered By AI For Specially Abled
Maximum Marks	4 Marks

Technical Architecture:

Obtain Output text in the form of speech

Output text

Character : A
Word :
Sentence : HI I

Recognized hand gesture

Gesture Recognition Hand Tracking System

Extracted image

Feature extraction

Preprocessed image

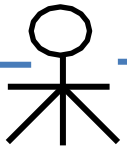
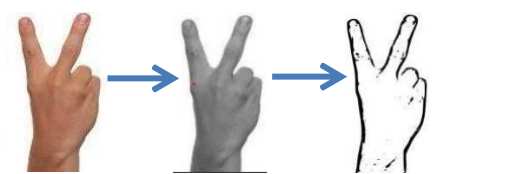
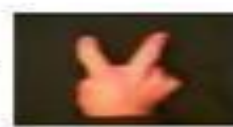
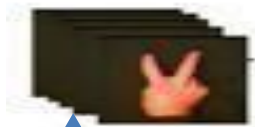
Image pre-processing

Hand movement

web cam

captured image

input image



Background is popped and original sign is extracted

User interface

Capture the sign within ROI

End-user

Original

Gray-scale

Gaussian blur filter to get outline

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application i.e. Desktop usage and clicking the concerned app.	HTML, CSS, JavaScript and Angular JS
2.	Application Logic-1	<ul style="list-style-type: none">❑ Camera detects the sign shown by the user.❑ Captures the sign within ROI.	Adaboost face detector is used to differentiate between faces and hand as both involves similar skin-colour.
3.	Application Logic-2	<ul style="list-style-type: none">❑ Background is popped and original sign is extracted.	By default, Original image captured is converted into Gray-scale image.
4.	Application Logic-3	<ul style="list-style-type: none">❑ Extract the edges of the gray-scale image.	Apply Gaussian-blur filter and threshold to the frame taken with Open CV to get the processed image after feature-extraction.
5.	Application Logic-4	<ul style="list-style-type: none">❑ Convert the output text into speech	The Final text obtained is converted to speech using the speech assistant implemented , which in turn produce sound from speaker.
6.	Database	<ul style="list-style-type: none">❑ Binary Large Object(BLOB) is the data type used to store the images in the dataset.❑ <code>/etc/mysql/my.cnf</code> is the default configuration / directories for MYSQL that is used.	MySQL database is used.
7.	File Storage	<ul style="list-style-type: none">❑ Create a BLOB column for the image files, whether they be JPEG, PNG, PSD or whatever, and then load the images into the table/column, created for them.	Local File system is used for storing the images.
8.	Machine Learning Model	Allows the user to feed a computer algorithm an immense amount of data and have the computer analyze and make data-driven recommendations and decisions based on only the input data	Supervised and Unsupervised learning model etc.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
------	-----------------	-------------	------------

1.	Open-Source Frameworks	<ul style="list-style-type: none"> • Palm detector operates on full images and outputs an oriented bounding box. • Hand landmark takes the cropped image defined by the palm detector and returns 3D hand key points. • Gesture recognizer then classifies the previously computed key point configuration into a discrete set of gestures 	Media Pipe Framework is used. Within this framework, the pipeline is built as a directed graph of modular components.
2.	Scalable Architecture	<ul style="list-style-type: none"> • It's a Three –Tier Architecture comprises the following technology, Convolutional neural network can be scaled in three dimensions: <i>depth, width, resolution</i>. • Depth of the network corresponds to the number of layers in a network. • Width is associated with the number of neurons in a layer. • Resolution is the image resolution that is being passed to CNN. Increasing the depth, by stacking more convolutional layers, allows the network to learn more complex features. 	Convolution Neural Networks is used.
3.	Availability	Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non verbal form of communication. Hand gesture recognition is a process of understanding and classifying meaningful movements by the human hands	CNN, Media Pipe, Gaussian blur filter, Machine learning models along with Speech assistant is used.