Bidirectional Communication Device for Deaf, Dumb and Blind

Krishnendu Nair[1], Manthan Gharat[2], Anjali Agarwal[3], Vighnesh Srinivas[4]

[1]*Professor, Dept. of I.T. Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India*
[2]*Student, Dept. of I.T. Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India*
[3]*Student, Dept. of I.T. Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India*
[4]*Student, Dept. of I.T. Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India*

Abstract - *Deaf/Dumb and Blind people, due to their special needs, face certain issues while communicating with others in their everyday life. According to WHO 39 million people are estimated as blinds and around "700,000 to 900,000" are estimated as deaf/dumb worldwide. They have a lot of difficulties communicating in their day-to-day life. Sign language is the most prominent methodology, using which hearing and speech impaired people can interact with the remainder of the planet. It is inconvenient for normal people and almost impossible for a blind to communicate this way. What we are proposing is a single device that can be used by both blind and deaf/dumb.*
This device can be used nonchalantly by everyone. The device will use a camera to record the Sign language used by the impaired and convert it to speech using certain algorithms. This device will also consist of a microphone that will take speech as input and convert it into the appropriate signs and display it in front of the eyes of deaf/dumb. Thus this device will be of great use for the blinds, deaf and dumb, and help them travel to different places.

***Key Words*: Bidirectional Conversation System, Indian Sign Language, Deaf/Dumb and Blind, Hand Signs, Object detection, Deep Learning**

1. INTRODUCTION

There has always been a barrier between us and the disabled community while communicating with each other in social interaction. To bridge this gap we came up with the following perspective to access the problem.
For communication to exist there must be a sender and a receiver and the essence of that is in the information being exchanged. But during the bi-directional conversation, the medium through which the information is conveyed by the sender and the medium through which the receiver can perceive the sent information is equally crucial
(We have considered Indian Sign Language for implementation of the system)

These are the three vital factors:

1. Speech Recognition

Speech Recognition is the process of collecting voice input from the user and the collected voice is processed to identify words and sentences in the respective language. The efficiency of speech recognition depends on the clarity of the speech recorded from the microphone which will be further processed to remove the background noise and filter out the required audio, which is converted into text. The sentences obtained from the text are further re- structured concerning Indian Sign Language Grammar rules.

2. ISL Grammer

The biggest hurdle when the problem was addressed was the grammatical complexity of spoken languages. The breakthrough for this issue was in the basics of Sign Language itself. Since grammatical complexities cannot be represented in the form of hand signs. A lot of parts of speech have been removed (e.g. articles, conjunctions, etc)and the structure of the sentence is changed which strips down the sentences of their depth to a simple format that conveys the same message. Also, ISL assigns a distinct hand sign to different words and letters and if the word is not present in its dictionary (e.g. Names) then the word is spelled letter by letter.

2. Hand Sign Detection

Hand Sign Detection is the third part of the process wherein the hand signs used by the dumb will be detected by the system and translated to text in real-time thus constructing sentences. The accuracy of Hand Sign Detection will depend on the size of the dataset, the quality of the dataset which is dependent on picture quality, lighting, etc..,. The bigger the dataset, the better the accuracy. For Hand Sign Detection, we are using TensorFlow by Google which comes with a built-in open- source object detection library that has an accuracy of 85- 99%

Training the object detection model to recognize Indian Sign Language(ISL) requires us to supply

the object detection model with a large dataset enabling the model to differentiate between different signs and objects that are not part of sign language.

2. Related Work

Various IoT technologies and techniques are used to make the lives of deaf/dumb and blind easier. This section presents a light on different technologies used :

Raspberry Pi Based Assistive Communication System for Deaf, Dumb and Blind: Suvarna Nandyal, Shireen Kausar(2019)[1]:
They are developing an efficient text-to-speech conversion technique by using the Raspberry Pi 3 processor. When the text image is captured by the camera, and stores in the cloud, A synthesizer is used to separate the text from the image. The Optical Character Recognition algorithm is then implemented to recognize the characters in the text which is then converted to speech using OpenCV libraries. For deaf people, the speech output is then converted to text with the help AMR voice app and HC-05 device. Here it offers three stairs Image to Voice (For BLIND) Text to Voice (For DUMB) Voice To Text (FOR DEAF).

Third eye for the blind using Arduino and ultrasonic sensors: M Narendran, Sarmistha Padhi, Aashita Tiwari(2018)[2]:
This proposed system consists the equipment like Arduino mini pro, ultrasonic sensor, perf-board, vibrating motor, buzzers for detecting the obstacles and letting the user know about the obstacle, Red LEDs, Switches, Jumper cable, power bank, Male and female header pins, a 3.3-volt old mobile battery which is unused or discarded, some elastic and stickers to make the device wearable as a band for wearing for the users. The wiring of the device is done in the following manner. The Ground of LED, buzzer, and vibration motor is connected to the GND of the Arduino.

Sensor-Based Assistive Devices for Visually-Impaired People: Wafa Elmannai and Khaled Elleithy(2017) [3]: Visual assistive technology is divided into three parts: Vision enhancement, Vision substitution, and Vision replacement. The sensor aids the user with the mobility task based on the determination of dimensions, range, and height of objects. The vision replacement deals with medical and technology issues. Vision replacement includes displaying information directly to the visual cortex of the brain or through an ocular nerve. In vision enhancement, the camera input is processed and then the results are visually displayed. Vision substitution comprises of the non-visual display, which can be vibration, auditory, or both based on the hearing and touch senses that can be easily controlled and felt bmy the blind user

3. Proposed System

Our proposed system works in the following way:

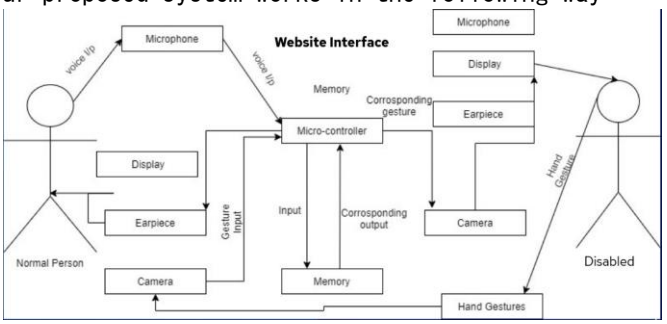

Fig -1: Conversation Between Normal and Dumb/Deaf

The above figure explains the block level representation of interactions and also the entire flow of the system.

Dumb/Deaf uses sign language to communicate with an abled person. The camera records these signs. This video is broken down into individual frames. The image in this frame is then interpreted using the model that is trained and then is interpreted concerning the data set. This is then converted to a text format. Using the text to voice module, we can then translate the acquired text to a voice format which is then sent to the speaker of the system. The abled person will provide a voice input using the microphone which is then sent to the back-end which then processes the audio input and converts it into sign language. The converted sign language output is then displayed on the screen on the system. These steps are carried out multiple times during a conversation.

3.1 Algorithm

Natural Language Processing
Natural Language Processing or NLP is a field of Artificial Intelligence that helps to comprehend and procure meaning from human languages. It is a discipline that focuses on the interaction between computers and human language and is scaling to lots of industries. It automatically manipulates the natural language of humans like text and

speech. NLP uses various algorithms to identify and extract various NLP rules which are converted in a form that is understood by the computer.

a) Bag of Words: It is a model which is frequently used that allows you to count all the words present in a text. The bag of words creates an occurrence matrix that is used for the sentence or document, ignoring grammar and word order. These word frequencies or occurrences are then used as features for training a classifier

b)Tokenization: It is the method of segmenting running text into sentences and words. In essence, it' s the task of cutting a text into pieces called tokens, and at an equivalent time discard certain characters, like punctuation. at the same time removing characters, such as punctuation.

c)Lemmatization: It breaks down words to their dictionary form known as the lemma for which it requires detailed dictionaries during which the algorithm can check out and link words to their corresponding lemmas. for instance, the words "calling", "calls" and "called" are all sorts of the word "call", so "call" is that the lemma of all the previous words.

Convolutional Neural Networks:

A convolution neural network is a category of deep learning model usually used for classifying images and detecting faces and objects by learning the weights and biases and the various aspects of images with the help of datasets. Even though CNN requires a training dataset, it is far less than compared to other classification algorithms. Also, CNN can learn from its real-time application and increase its accuracy.

Contours

Contours are defined as a curve joining all the continuous points (along the boundary), having the same color or intensity. Contours are of great help when it comes to analysing shapes and detecting and recognizing objects. To enhance accuracy, binary images should be used.

Convex Hull

Even though the convex hull might seem similar to contours, it isn' t. It checks the curve for convexity defects and corrects it. The curves which bulge out from inside are called convexity defects. The convex hull contains arguments such as points, hull, clockwise, and return Points.

3.2 Implementation

The system has two phases depending on the following: Phase1:
Audio as an input
In this section, the sender speaks something and the audio has to be decoded by the receiver. If the receiver is Deaf/Dumb then this becomes a barrier in communication. To deal with this we are taking the audio from the sender and converting it into text which can be processed to show the representation of the same in hand signs.
Phase 2:
Hand Sign as an input
In this section the sender makes hand signs and the signs have to be interpreted by the receiver. If the receiver is profound with Sign Language then this again becomes a barrier in communication.

To deal with this we are detecting the Hand Signs in real- time from the sender and converting them into a medium that is understood by the receiver. Consider a scenario where a normal person is trying to have a conversation with a deaf and dumb person. The normal person will "speak" his message. For example, if the person says "The sword is sharp". This will be taken as an audio input by the system.

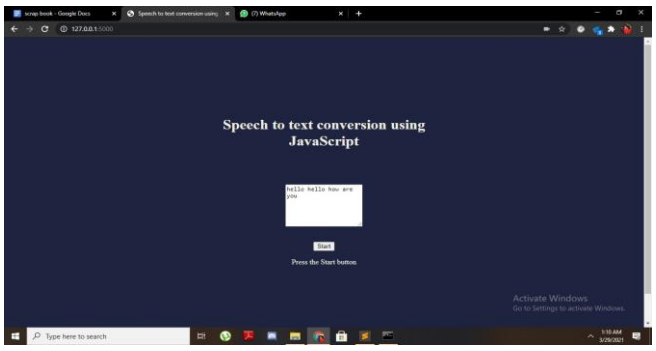

Fig -2:Web interface to collect audio

The figure mentioned above show the audio collected from the user which is converted to text This input will then be converted into text by the speech- to-text module.
Once the sentence is received, it is processed and each word from it will be tokenized and labeled with its respective parts of speech.
After being tokenized it' ll remove the redundant POS. It will also find words that have plural forms by their lemma_tags and convert those to their root form, hence converting them into a

singularity. After all the operations are applied finally the sentence is converted into 'Remaining words in exact order + verbs'.
After processing, we get "sword sharp"
The generated sentence is looked up in the database of words. If the word is found a video is created by concatenating all the videos from the database in the sentence resulting in one final video and if a word is not found then it's broken down into letters and those are concatenated to the video which conveys the sign to the deaf and dumb person.

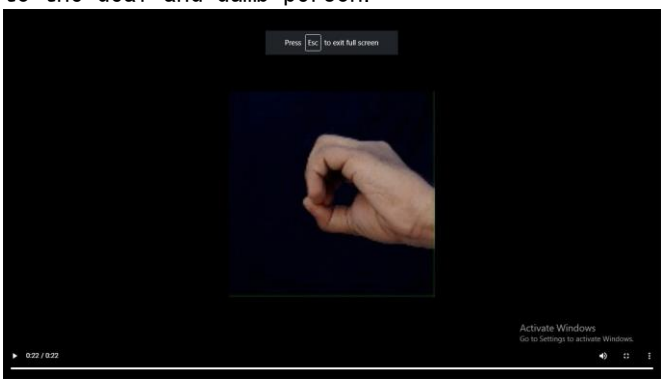

Fig -3:Snapshot of the video output

In the second part of the conversation, the disabled person will use sign language to communicate. Here we are going with Indian Sign Language(ISL) so whatever sign language translation is done, it is translated to ISL. Each sign is captured by the camera, detected, and interpreted in real-time by our object detection system, and the corresponding letter, word, or sentence is conveyed as per the perceived medium of the receiver.
We have achieved the Hand detection Problem using two approaches
1.Using Contours and Convex Hull

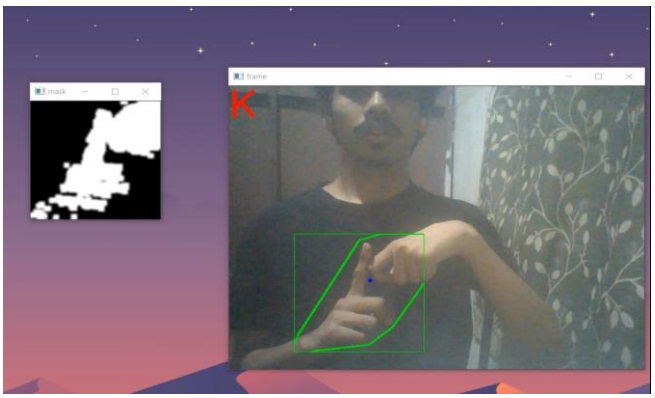

Fig -4:Hand Sign Detection using contours

In this approach, we are detecting the hand signs on the basis of the area created by the sign and the convex hull defects of the HSV mask passed through various filters. The thought process behind this is that every sign assigned is distinct, the pair of the area and convex hull defects should be different.

Cons:

Since it depends on the HSV range of skin color, if the background has similar color the results are severely affected

There are signs which have similar areas which lead to false negatives

Not reliable for real-time application

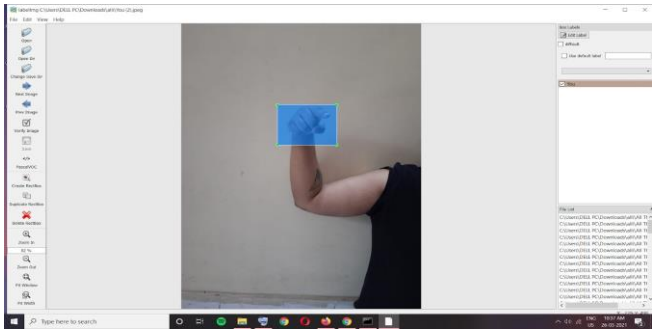

Fig -5: Data Set Labelling

The hand signs are labeled and serve as the training dataset for the model which takes the frames of live feed from the camera to detect the and signs present in it with an accuracy ranging from 85- 99%

Unlike contours, this method is not affected by any background colors or surroundings as well as the hand need not be in a particularly marked frame rectangle, the sign is detected wherever it is made on the frame.

Due to which it is highly reliable in real-time application

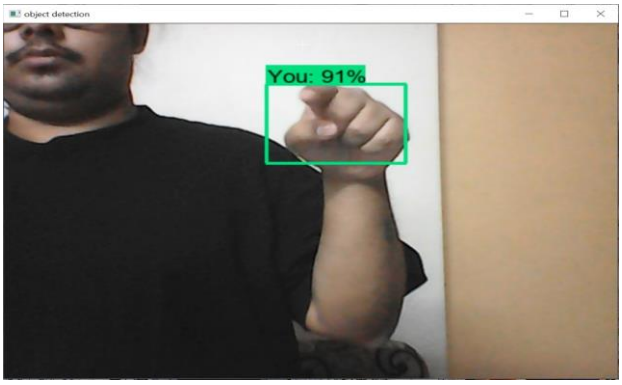

Fig -6:Hand Sign Detected using Object Detection

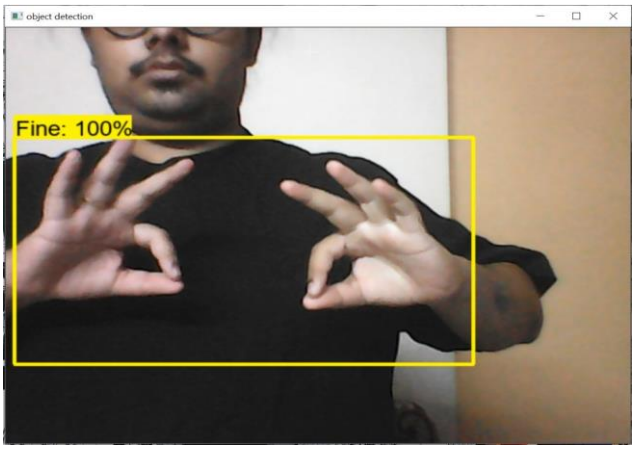

Fig -7:Hand Sign Detected using Object Detection

4. CONCLUSION

The proposed system is a bidirectional communication device for disabled people. To construct our "Knowledge Base", We created our own dataset of different words, sentences, alphabets, and numbers as well as consulted Sign Language instructors. These data forms and sources are used in a combination with an expert system, generating features to further improve our knowledge base.

The future scope of the system is to not restrict it to only Indian sign language it can be expanded over various other sign languages such as ASL and incorporated with them as well as translate between different spoken languages.

REFERENCES

1. Raspberry Pi Based Assistive Communication System for Deaf, Dumb and Blind: Suvarna Nandyal, Shireen Kausar(2019).
2. Third eye for the blind using Arduino and ultrasonic sensors: M Narendran, Sarmistha Padhi, Aashita Tiwari(2018).
3. Sensor-Based Assistive Devices for Visually-Impaired People: Wafa Elmannai and Khaled Elleithy(2017).
4. Automatic Translation of English Text to Indian Sign Language Synthetic Animations: Lalit Goyal, Assistant Professor, DAV College, Jalandhar, Dr. Viahal Goyal, Associate Professor, Punjabi University, Patiala(2016.
5. A Large Scale Dataset for Indian Sign Language Recognition: Sridhar, Advaith; Ganesan, Rohith Gandhi; Kumar, Pratyush; Khapra, Mitesh(2020).
6. https://www.talkinghands.co.in/