Enhancing Communication Accessibility for Deaf and Mute Individuals: An Innovative Approach for Convert Gujarati Speech into Sign Language

Dr. Abhishek Mehta[1]
Associate Professor and Post Doc Research Scholar at Parul University
Faculty of IT and Computer Science, Parul University
Parul Institute of Engineering and Technology – MCA, Parul University, Vadodara, Gujarat
Dr. Priya R. Swaminarayan[2]
Dean, Faculty of IT and Computer Science, Parul University
Principal, Parul Institute of Computer Application, Parul University, Vadodara, Gujarat

Abstract:

Deaf culture plays a critical role in uniting deaf communities worldwide through the use of various sign languages, which are visually based and diverse, with approximately 300 types identified globally. Each sign language has its unique syntax and semantics, with some languages employing one-handed gestures while others use two-handed gestures. This diversity presents challenges, particularly in terms of standardization and communication across different regions. One of the major difficulties faced by the deaf community is the lack of a universal sign language, leading to communication barriers. Moreover, specific sign languages have not yet been integrated into technologies that can bridge the gap between speech and sign language or vice versa. This gap is particularly evident in the context of Gujarati Sign Language, where no existing solutions effectively translate speech into sign language or provide a reliable communication bridge for the Gujarati deaf community. In response to this challenge, we propose a communication model that converts spoken language into Gujarati Sign Language through a series of transformations. Our model takes speech input and converts it into text, which is then mapped to HamNoSys notation, a sign language notation system. This notation is then transformed into SiGML format, which can be used to generate animated sign language via avatars. The final output is a visual representation of the original speech input in Gujarati Sign Language. This innovative approach aims to improve communication for the deaf and hard-of-hearing population in the Gujarat region, allowing for easier interaction with non-deaf individuals. By providing a tool that can seamlessly convert speech to Gujarati Sign Language, we hope to foster greater inclusion and understanding while addressing a critical communication gap within this community.

Keywords: Sign Language Conversion, Speech-to-Sign Translation, Gujarati Sign Language, Deaf Communication, Sign Language Technology

I. Introduction:

The culture of deaf individuals forms the cornerstone of their identity and sense of community, regardless of geographic location. According to the World Health Organization (WHO, 2018), approximately 466 million people globally suffer from hearing loss, with 432 million being adults and 34 million children. Roughly one-third of people over the age of 65 experience disabling hearing loss. Language and culture are inextricably linked, and for deaf people, sign language serves as their native mode of communication, complete with unique syntax, semantics, and linguistic norms. Many types of sign languages exist across various deaf communities, enabling seamless interaction among deaf individuals without the need for interpreters. Sign language is also utilized by people who can hear but are unable to speak due to physical disabilities. Communication barriers arise when hearing individuals attempt to interact with deaf people, primarily because hearing individuals are often unfamiliar with the customs and language of the deaf community. This lack of understanding can lead to the marginalization and isolation of deaf individuals. Deaf and dumb schools use sign language as the primary medium for teaching students, but the challenge lies in ensuring effective comprehension. Since sign language interpretation often requires repeated exposure for full understanding, providing students with opportunities to revisit content can enhance their learning experience. Given the popularity of sign language within the deaf and dumb community, the communication gap becomes apparent when they interact with hearing people. This paper explores a model for converting speech to sign language, specifically designed for the Gujarat region. The goal of this model is to improve communication between hearing and deaf individuals, ultimately enhancing learning and fostering more inclusive interactions between the two groups.

An estimated 1.3 billion people worldwide live with some form of vision impairment, with 188.5 million having mild impairment, 217 million experiencing moderate to severe impairment, and 36 million who are blind. Most of these individuals are over 50 years old, and India has the largest population of blind people. In terms of hearing loss, around 9.1 billion people are deaf and mute, and the World Health Organization reports that approximately 466 million people worldwide have disabling hearing loss.

Despite technological advances, these groups often do not benefit from technology in the same way as others. They face numerous obstacles, particularly in communication, a fundamental human need. Traditional communication methods like Braille and sign language can be cumbersome and often require assistance from others, reducing independence.

This paper aims to address this issue by proposing a device that uses Raspberry Pi and Google API to help visually, audibly, and vocally impaired individuals communicate more effectively and independently. This device consists of a Raspberry Pi, an integrated camera, a microphone, a speaker, and a screen, and it has three main modules designed for each group.

For those with visual impairments, the device uses a camera to capture images of text, such as books or documents, and then processes this visual information into digital text through Google Vision API. The text is then transformed into audio using text-to-speech (TTS) technology, allowing the user to hear the content.

For individuals with hearing impairments, the device records spoken audio, converts it into text, and displays it on a screen, allowing them to read the transcribed content.

For those who have vocal challenges, the device provides an on-screen keyboard where users can type their messages. The text is then converted into speech through TTS, enabling them to communicate using an artificial voice.

By integrating these features, the device aims to bridge communication gaps and promote greater independence for those with sensory and speech impairments.

II. Sign Language:

Deaf and hard-of-hearing individuals often begin learning sign language by becoming familiar with the alphabet, A-Z, through corresponding hand signs. This technique is known as finger spelling, which uses specific hand positions to represent individual letters. Finger spelling plays a critical role when there are no predefined signs for names, places, or objects, allowing sign language users to spell out these words manually. For instance, while many words have dedicated signs, others require finger spelling because they lack a standard representation. Different sign languages use distinct finger spelling systems. For example, British Sign Language (BSL) and Arabic Sign Language each have unique signs for the same alphabet letters. This variability stems from the diversity of sign languages across countries, states, and regions. In some cases, a single alphabet letter can have varying signs in different sign languages. This disparity often reflects the broader linguistic and cultural diversity among sign languages. Since sign languages are considered minority languages, their development and recognition are generally less extensive compared to spoken languages. TABLE I shows a comparison of alphabet signs from various sign languages, illustrating the differences between them. For instance, in American Sign Language (ASL), a single hand gesture represents the letter 'A,' whereas British Sign Language (BSL) requires a two-handed gesture for the same letter. Such distinctions underscore the unique nature of each sign language, reinforcing that even common elements like the alphabet can vary significantly across different regions and countries.

Figure 1: Gujarati Sign Language Alphabets

III. Literature Review:

User-centered design is essential in creating technologies and interfaces that accommodate the needs of people with disabilities. However, simply adding new features for the visually impaired may not be enough to improve usability. Many technologies, both hardware and software, aim to support visually impaired people, including devices that read printed text, Braille systems, and tools based on computer vision. Some prototypes work with smartphones and cameras to help users process visual information. For example, AudioMUD provides a multiuser virtual environment with audio cues designed specifically for blind users, allowing them to navigate and interact through text-based commands.

Another approach involves wearable technology that enhances the quality of life for visually impaired people. Systems like the one proposed by L. Gonzalez include features like facial recognition to identify individuals, obstacle detection with ultrasonic sensors, email reading through text-to-speech, medication reminders, and MP3 players for entertainment. Anusha Bhargava's work with Raspberry Pi captures images through a webcam, processes them to extract text using Optical Character Recognition (OCR), and then converts the text into speech for the user to hear.

For deaf-blind individuals, Lorenzo Monti and his team created GlovePi, which uses capacitive touch sensors with a Raspberry Pi to detect the presence, number, and facial expressions of people around the user. The system uses a peer-to-peer network and a many-to-many architecture for interaction. Another unique solution comes from Amro Mukhtar Hussain, who developed a system using infrared sensors to recognize mouth gestures, translating them into patterns corresponding to alphabetic characters.

A single device that addresses the needs of blind, deaf, and mute users is uncommon. Kumar K. and colleagues offer a setup where the visually impaired can read text through OCR, the vocally impaired can communicate through text-to-speech, and the audibly impaired can read transcribed text from speech. Rohit Rastogi's Sharon Bridge is a wearable technology that facilitates communication between people with different abilities. It features an Arduino-based glove that translates American Sign Language into audio and text, with an Arduino GSM module for wireless communication over long distances. This comprehensive approach allows for diverse communication methods, making technology more inclusive for people with different needs.

Deaf and hearing people often face communication barriers, which can be significant in environments where sign language is the primary mode of communication. My visit to a deaf-mute school in Ahmedabad highlighted how deaf-mute children communicate effortlessly using sign language, but as a hearing person unfamiliar with sign language, I struggled to interact. The need for interpreters arises when deaf and hearing individuals try to communicate. Kanwal Yousaf discusses that the deaf community includes a variety of groups, from those who are not entirely hearing or deaf, to those who belong to deaf families where sign language is the primary communication method, and those who are deaf from birth.

Sign language is widely used among deaf-mute people, but it is not as accessible to hearing individuals who lack knowledge of it. Some key facts about sign language are that it varies across regions, has unique grammar and vocabulary, and often relies on finger-spelling for unknown words. It is typically used in the present tense, and the dictionary for sign language is much smaller than those for spoken languages due to limited development in the field.

The corpus of spoken language contains about 5 million words, while sign language has fewer than 100,000 signs. When compared to spoken language, sign language databases are limited. Li, Dongxu, et al. have created a large-scale video dataset for word-level American Sign Language (ASL) with 100 signers and about 2,000 words. Challenges arise when sign language databases are scarce, given the regional differences in sign languages. Dynamic hand gesture recognition, particularly in ASL, remains a critical challenge. Camgoz, Necati Cihan, et al., presented an architecture for Continuous Sign Language Recognition (CSLR) using Connectionist Temporal Classification (CTC). Bencherif, Mohamed A., et al. introduced an Arabic Sign Language (ArSL) recognition system that uses 2D hand and body keypoints from video frames. Through a combination of 2D and 3D Convolutional Neural Networks (CNN), this system recognizes signs from recorded videos.

Object detection with CNNs offers accurate and quick results, which is why CNNs, a type of deep learning algorithm, are popular for image recognition. Sign language also faces challenges with Natural Language Processing (NLP) because of structural differences between sign language and spoken languages, along with limited annotation. Avatar generation is another hurdle, as deaf users may or may not find it acceptable. Das Chakladar, Debashis, et al., developed a sign language learning system based on 3D avatars that converts speech into English using IBM-Watson, then into Indian Sign Language (ISL) using regular expressions and script generation. Each word is transformed into a corresponding sign through a 3D avatar.

Most sign language dictionaries are created with real human videos, but synthetic animations offer more consistency in representations and require less memory. Goyal, Lalit, and Vishal

Goyal developed a synthetic animation for the Punjabi alphabet, creating HamNoSys codes for each letter. HamNoSys (Hamburg Sign Language Notation System) is a "phonetic" transcription system with about 200 symbols to describe hand location, shape, and orientation. This comprehensive approach aims to bridge communication gaps between deaf and hearing individuals.

IV. Design

The figure 2 shows the outline of the device. The raspberry Pi is the support system of the device which connects the camera, microphone, speaker and LCD display. The device works for the visually impaired as the camera clicks a picture of the document and the output is in audio format through the speaker, audibly impaired as the microphone takes the spoken words as input and displays it as text on the LCD display, and for the vocally impaired as the user types the message in the LCD and the speakers gives the output as an audio.

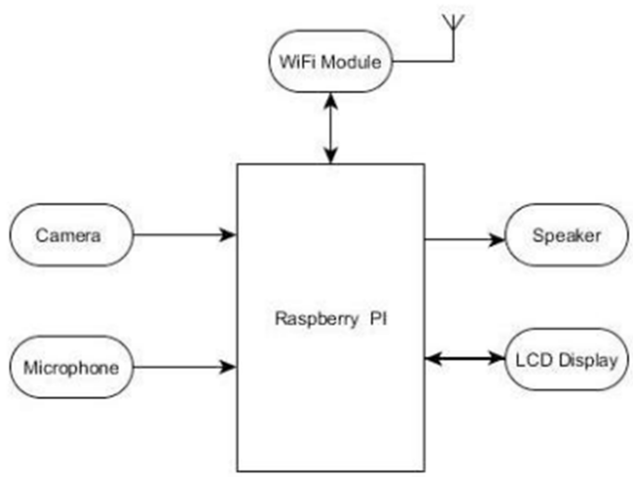

Figure 2 The system architecture with all main modules

4.1 Google Cloud Vision API

The Google Cloud Vision API (Application Programming Interfacing) encapsulates powerful machine learning models in an easy to use REST API and enables developers and users to apprehend the content of an image. It is used for classification of images into thousands of categories, detecting individual objects and faces within images, and reading printed words contained within images. Optical Character Recognition (OCR) is used to enable the user to detect text within images, along with automatic language identification. Vision API supports a huge and broad set of languages. Initially Conventional neural network (CNN) based model is used to detect localized lines of text and generates a set of bounding boxes. Script identification is done by identifying script per bounding box and there is one script per box. Text recognition is the core part of the OCR which recognizes text from image as shown in Figure 3.

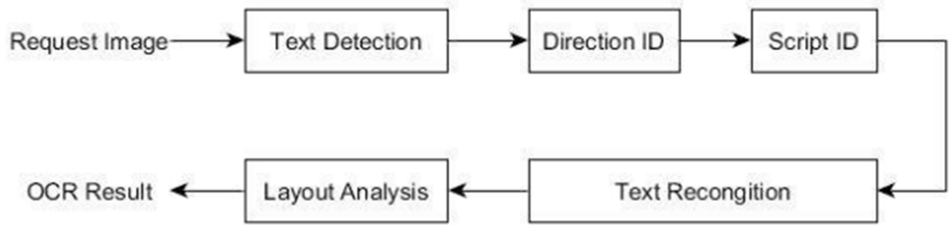

Figure 3 Architecture of Vision API converting Image to text

4.2 Tkinter

Various options for the development of graphical user interfaces are provided by python. Tkinter is the standard GUI (graphical user interface) provided as a library for python. GUI applications can be created in a faster and easier way using Tkinter, and it also provides a prevailing object-oriented interface to the Tk GUI toolkit

4.3 Speech to Text:

Google cloud Speech to text aides the developers in the conversion of audio into text as it applies robust neural network models in a convenient API. It enables voice command and control and transcribes audio. It is capable of processing real-time streaming or pre-recorded audio using Google's ML technology. The accuracy is unparalleled as the most advanced deep learning neural network algorithms are applied by Google. It streams text results, returning text as it is recognized from audio stored in a file and is capable of long-form audio as shown in Figure 4.

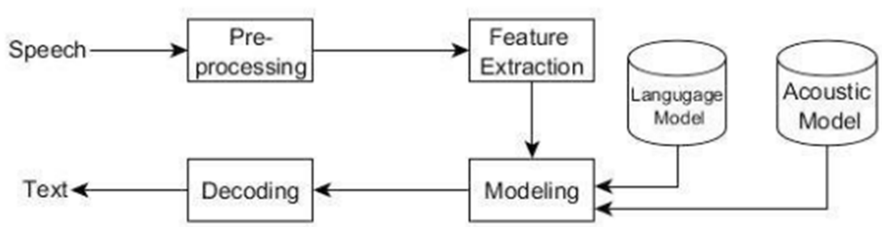

Figure 4 Google speech API converting speech to text

4.4 Text to speech

Google Text to Speech API is one of the several APIs available in python to convert text to speech as shown in Figure 5. It is commonly known as the gTTS API. It is an easy and efficient tool which converts entered text, into audio that can be saved as an mp3 file.

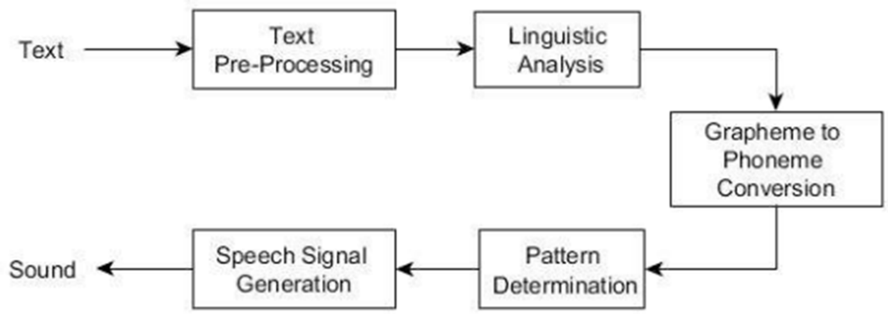

Figure 5 gTTS converting text to speech

4.5 Bitwise SSH

Bitwise SSH (Secure Shell) is one of the advanced and flexible SFTP protocol. The bitwise ssh helps us to securely connect with raspberry pi and access all the resources of raspberry pi. In addition, the user can transfer the files from localhost to raspberry pi; compile the programs; and provides a secure link for further connection.

4. 6 Raspberry Pi

Raspberry Pi, shown in Figure 6 is a low cost, credit card sized processor, which can easily perform all task we expect from a desktop. It is very easy to connect raspberry pi with computers and TVs. It also provides GPIO (General Purpose Input Output) pins to connect with other components. Because of this efficiency to intercommunicate with the cross-disciplinary domain, it has been used in a variety of projects. Raspberry pi operates in an open source environment such as Raspbian (Linux based operating system).

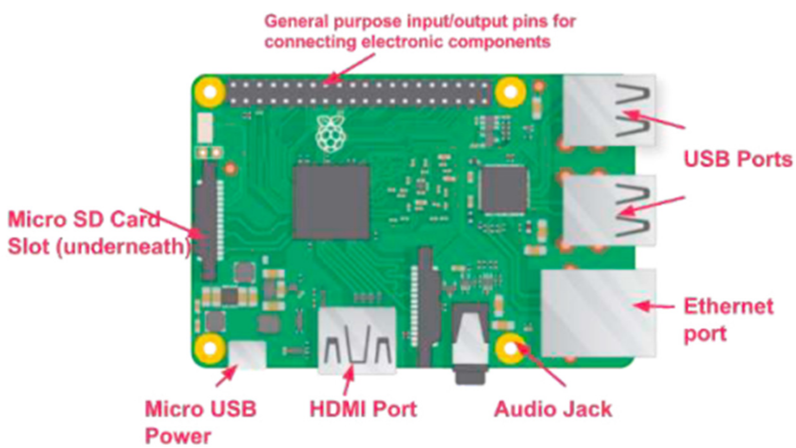

Figure 6 Raspberry Pi

Technical specification

- Broadcom Soc BCM2836 (CPU, GPU, DSP, SDRAM)
- 900 MHz quad-core ARM Cortex A7 CPU (ARMv7 instruction set)

- Broadcom VideoCore IV @ 250 MHz GPU
- 1 GB MEMORY (shared with GPU)
- 4 USB ports
- 17 GPIO(General Purpose Input Output) Peripherals plus specific functions, and HAT ID bus
- 15-pin MIPI camera interface (CSI) Video input connector
- HDMI video outputs, composite video (PAL and NTSC) via 3.5 mm jack
- I^2S audio input
- Analog audio output via 3.5 mm jack; digital via HDMI and I^2S
- MicroSD for storage
- 10/100Mbps Ethernet speed
- 800 mA power rating (4.0 W)
- 5 V power source via MicroUSB or GPIO(General Purpose Input Output) header
- 85.60mm × 56.5mm
- Weighs 45g (1.6 oz)

V. Implementation

The device has been created by formulating a unique design for assisting the differently abled people. It has been divided into three modules for enhancing the experience of the user with the device. The device consists of three modes and a three-way slider to change mode. Each mode is separately dedicated for the blind, deaf and dumb respectively in the device. The device is designed to make the user feel individualistic, self-reliant and selfsufficient. The gist of design of the device is in the figure 2. The main component of the device is the raspberry Pi.

5.1 Deaf module

The audibly impaired can virtually hear using this device as it enablesthem to read, what is being spoken. The figure 10 describes the respective procedure

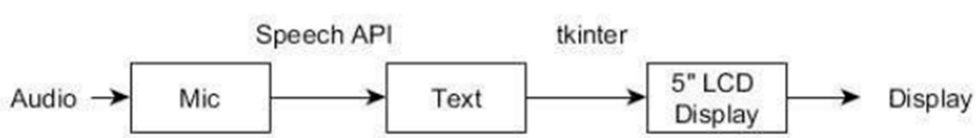

Figure 7 Working of Deaf Module

Step 1.The three-way slider is set to the deaf mode. The audio or the words being spoken to the user, who in this case might be a deaf person, are recorded as input by the USB Microphone connected to the Raspberry Pi of the device and is saved as a file in mp3 format.

Step 2.This audio file is passed to the Google Speech API which converts the audio into text for the user to understand.

Step 3.The converted text is then displayed on the 5 inch HDMI LCD screen available in the device, as a pop up window exclusively created using python tkinter for this module. This way the user understands everything that is being spoken to him quickly and efficiently. To change modes, the slider can be set accordingly.

5.2 Dumb module

This module makes the device handy for the vocally disabled as it enables them to vocalise words by typing it on the screen. The figure 11 explains the methodology.

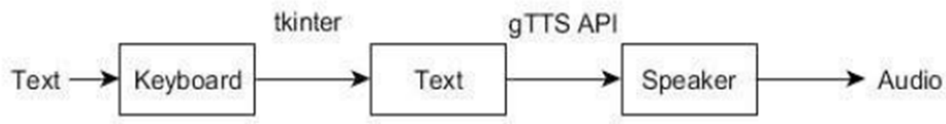

Figure 8 Working of Dumb Module

Step 1.When the three-way slider is used to set the device on the dumb mode, a pop up is displayed along with a customized keyboard which has been created using python tkinter, in it on the HDMI screen connected to the Raspberry Pi.

Step 2.The user who possibly is vocally impaired can type whatever he wants to convey using the keyboard in the screen as text

Step 3.The typed text is converted into audio format using the gTTS API and the audio file of the required text is obtained.

Step 4.The high quality speaker connected to the Raspberry Pi in the device plays this audio file thus vocalising the message given by the impaired person. Step 5.Modes can be changed in the device according to the convenience of the user.

VI. Conclusion

Through this paper, an unprecedented prototype has been created to aid the visually, vocally and audibly disabled. This project not just focuses on empowering and facilitating the differently abled, it is also compact and resource saver. The overall cost has been cut down by eliminating braille books and the energy spent in understanding them. It is a less costly solution, as all the components used in the device are cost effective and efficient. The latest and most trending technology makes this device portable, adaptable and convenient. The device proposed in this paper can be a major help in solving a few of the many challenges faced by the differently abled. To further extend the project, the device can be made more compact and wearable to make it easy for the user to use.

VII.References

[1] N. K., S. P. and S. K., Assistive Device for Blind, Deaf and Dumb People using Raspberry-pi, Imperial Journal of Interdisciplinary Research (IJIR), 3(6), 2017 [Online]. Available: https://www.onlinejournal.in/IJIRV3I6/048.pdf. [2] L. González-Delgado, L. Serpa-Andrade, K. Calle-Urgiléz, A. Guzhñay-Lucero, V. Robles-Bykbaev and M. Mena-

Salcedo, "A low-cost wearable support system for visually disabled people," 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, 2016, pp. 1-5. doi: 10.1109/ROPEC.2016.7830606

[3] Anusha Bhargava, Karthik V. Nath, Pritish Sachdeva & Monil Samel (2015), International Journal of Current Engineering and Technology, E-ISSN 2277– 4106, P-ISSN 2347– 5161

[4] J. Sanchez and T. Hassler, "AudioMUD: A Multiuser Virtual Environment for Blind People," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, 15(1), pp. 16-22, March 2007. doi: 10.1109/TNSRE.2007.891404

[5] M. Lumbreras and J. Sánchez, “Interactive 3-D sound hyperstories for blind children,” in Proc. ACM-CHI ’99, Pittsburgh, PA, 1999, pp. 318– 325.

[6] R. McCrindle and D. Symons, “Audio space invaders,” in Proc. ICDVRAT 2000, Alghero, Sadinia, Italy, Sep. 23–25, 2000, pp. 59–65.

[7] T. Westin, “Game accessibility case study: Terraformers-Real-time 3-D graphic game,” in Proc. ICDVRAT 2004, Oxford, UK, 2004, pp. 120–128.

[8] Y. H. Lee and G. Medioni, “Rgb-d camera based wearable navigation system for the visually impaired,” Computer Vision and Image Understanding, vol. 149, pp. 3–20, 2016

[9] J. Bajo, M. A. Sanchez, V. Alonso, R. Berj ´ on, J. A. Fraile, and J. M. ´ Corchado, “A distributed architecture for facilitating the integration of blind musicians in symphonic orchestras,” Expert Systems with Applications, 37(12), pp. 8508–8515, 2010.

[10] L. Monti and G. Delnevo, "On improving GlovePi: Towards a many-to-many communication among deaf-blind users," 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2018, pp. 1-5.doi: 10.1109/CCNC.2018.8319236

[11] R. Rastogi, S. Mittal and S. Agarwal, "A novel approach for communication among Blind, Deaf and Dumb people," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 605-610.

[12] A. M. Hassan, A. H. Bushra, O. A. Hamed and L. M. Ahmed, "Designing a verbal deaf talker system using mouth gestures," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, 2018, pp. 1-4. doi: 10.1109/ICCCEEE.2018.8515838