Smart Assistive Device for Speech Impaired using Silent Sound Technology

Diana John Esther
Department of Electronics and Communication Engineering
Marian Engineering College
Trivandrum, Kerala

Gayathri G.R
Department of Electronics and Communication Engineering
Marian Engineering College
Trivandrum, Kerala

Kripa Binoy
Department of Electronics and Communication Engineering
Marian Engineering College
Trivandrum, Kerala

Neha S.S
Department of Electronics and Communication Engineering
Marian Engineering College
Trivandrum, Kerala

Dhanya Mathew
Department of Electronics and Communication Engineering
Marian Engineering College
Trivandrum, Kerala

Abstract:- **One of the beautiful creation of god is human beings who have been fortunate to have senses like vision and speech. But some humans are unfortunate to have this blessing. Speech impaired people are greatly affected as they cannot communicate effectively as in the case of normal people by verbal means.Here speech impaired people communicate using sign language to communicate with others. But others who does not have the prior knowledge of what sign language is, find it very tough to understand what speech impaired person is trying to express. Smart assistive device can help speech impaired people to effectively communicate with others. This device can help people who have undergone surgery such as laryngectomy, a surgical removal of larynx which is usually done when person suffers from a critical accident or throat cancer. Using the smart assistive device it decodes the movement of lips and then convert it to text and audio. The listener on the other end can be either a normal person or a hearing impaired person. Here normal person can hear the audio of the spoken word and hearing impaired person can see the word displayed on the screen.**

I. INTRODUCTION

Communication is an important characteristic of human behaviour and embodies human beings ability to convey information, feelings, opinions to others. The main aspect of communication is that the receiver has to understand what the sender is trying to communicate. In the case of speech impaired person they communicate with others using sign language a visual means of communicating using hand signals,gestures and facial expressions. Here the person on the other end find it difficult very difficult to understand what the other person is saying if he or she does not have the prior knowledge of sign language. This project is proposed to ease the communication between speech impaired person and normal person. In this project speech impaired person's lip movements are analyzed and compared with pre-prepared database and releases output such as text and voice.

II. EXISTING SYSTEM

In the existing system it involves use of ultrasound probe, camera and a silent vocoder. Here image of vocal cords are taken from ultrasound probe and along with this the movement of lips are captured using camera. The captured images of lip and tongue movement are then given to lip reader. Lip reader then compares the input image with pre-stored images and if a match is found it generates a visual speech signal. This visual speech signal is then given to a silent vocoder. The silent vocoder consists of an HMM based visuo-phonetic decoder, audio-visual selection unit ,concatenation of the selected units, HNM based prosodic adaptation. Then the silent vocoder converts the visual speech into spoken words.
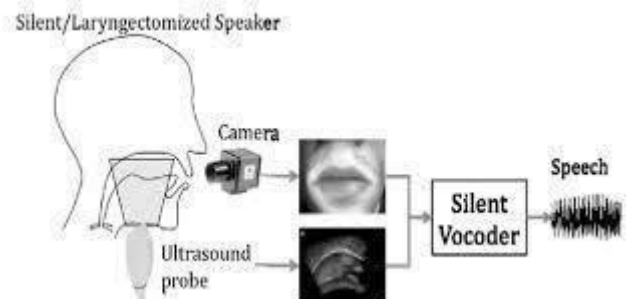
III. HARDWARE REQUIREMENTS

(i) RASPBERRY PI 4 MODEL B

It comes with a Gigabit Ethernet, along with the onboard wireless networking and Bluetooth. It has also upgraded its USB capacity, along with two USB 2 ports, two USB 3

ports are also added, which can make transfer data ten times faster.

Specifications

◆ ■ Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
◆ ■ 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (depending on model)
◆ ■ 2-lane MIPI DSI display port
◆ ■ 2-lane MIPI CSI camera port
◆ ■ 4-pole stereo audio and composite video port
◆ ■ H.265 (4kp60 decode), H264 (1080p60 decode, 1080p30 encode)
◆ ■ OpenGL ES 3.0 graphics
◆ ■ Micro-SD card slot for loading operating system and data storage
◆ ■ 5V DC via USB-C connector (minimum 3A*)
◆ ■ 5V DC via GPIO header (minimum 3A*)
◆ ■ Operating temperature: 0 – 50 degrees C ambient

(ii) RASPBERRY PI 5MP CAMERA

The Raspberry Pi Camera Board plugs directly into the CSI connector on the Raspberry Pi. It is able to deliver a crystal clear 5MP resolution image, or 1080p HD video recording at 30fps. The Raspberry Pi Camera Board features a 5MP Omnivision 5647 sensor in a fixed focus module.

The module attaches to Raspberry Pi, by way of a 15 Pin Ribbon Cable,to the dedicated 15-pin MIPI Camera Serial Interface (CSI), which was designed especially for

1V. SOFTWARE REQUIREMENTS

TENSORFLOW

TensorFlow is a machine learning software library that is both free and open-source. It can be used for a multitude of activities, but its best known for deep neural network training and inference. Tensor flow is a symbolic math library that uses dataflow and distinguishable programming to solve problems.
TensorFlow is a cross-platform programming language. It works on a wide range of systems, including GPUs and CPUs, as well as handheld and embedded platforms and even tensor processing units (TPUs)
The TensorFlow distributed execution engine abstracts away the many compatible devices and provides the TensorFlow framework with a high-performance core written in C++.

KERAS

Keras is TensorFlow 2 is high-level API: a user-friendly, highly efficient platform for resolving machine learning problems with an emphasis on modern deep learning. It provides the necessary abstractions and basics for designing and shipping high-iteration-rate machine learning solutions.
Engineers and developers can use Keras to fully exploit TensorFlow's flexibility and trans capabilities: you can execute Keras on TPUs or massive clusters of GPUs, and you can export Keras models to execute in the browser or on portable devices.
Layers and templates are Keras fundamental file formats. The Sequential model, which is a linear stack of layers, is the most basic type model. You can use the interfacing to cameras. The CSI bus is capable of extremely high data rates, and it exclusively carries pixel data to the BCM2835 processor. The board itself is tiny, at around 25mm x 20mm x
9mm, and weighs just over 3g, making it perfect for mobile or other applications where size and weight are important. The sensor itself has a native resolution of 5 megapixel, and has a fixed focus lens onboard. In terms of still images, the camera is capable of 2592 x 1944-pixel static images.

(iii) LCD DISPLAY

This 16×2 LCD packs 32 characters into an outline smaller than that of most two-line displays. An LED backlight enables optimal viewing in all lighting conditions. This unit uses the HD44780 interface found on most parallel character displays.

Keras operational API for more complex architectures, which allows you to construct arbitrary layer graphs or write models entirely from scratch through sub-classing.

IV.   DESIGN AND IMPLEMENTATION

First speakers's video is captured using a camera.From the speaker's video different frames(image) are obtained by stopping at single frequency.For example, a one sec video is converted into 30 frames/sec.From the set of frames face is localized in a video frame.Then localize the lip region as it increases the accuracy . OpenCV is used to localize the lip region OpenCV is a computer vision library, has very little support regarding neural networks/ deep learning. It can provide only computer vision/image processing utilities to your program, machine learning libraries like TensorFlow,scikit
-learn, caffe, etc.. have to be used to implement systems like LipNet . In the lip geometry based feature extraction the region of interest is the lips. Accordingly the mouth region's area is calculated. The height and width ratios of the mouth is used as features . These features are more applicable where the surrounding is more crowded area and contains lot of noise. Hence, in it only visual features are required.

In appearance based features the tongue, teeth's are considered for lip movement detection. Since the geometry based features where having the drawbacks like detecting the mouth feature reliably, lighting conditions and many more, to overcome this appearance based features are considered.
It was an alternate way for the extraction of features using the pixel data as features . Improved Local Binary Pattern (ILBP) from the three of the orthogonal planes was considered for change in the time and the space of mouth's region. The binary image of lip is also taken as the feature

taken the mouth opening which consisted of tongue and teeth's as a feature. In which the teeth area was taken as the ROI and the contour of it was taken as the feature for further processes. The lips and mouth region are the visual parts of the human speech production system; these parts hold the most visual speech information,therefore,it is imperative for any VSR system to detect/localize such regions to capture the related visual information,i.e. we cannot read lips without seeing them first.It is then compared with the lip dictionary and if it is a match text and voice are released as outputs.
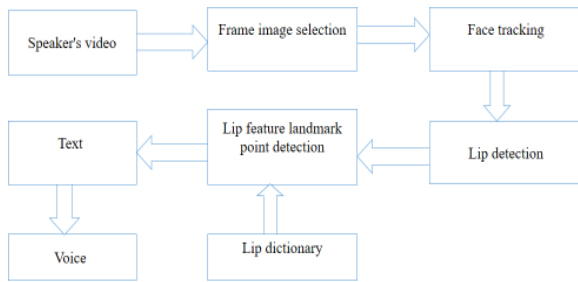
VI. CONCLUSION

By using silent sound technology it results to notice every movement of lip and turn them into sounds which could help people to make silent calls without causing any inconvenience to others.Rather than making any sounds,this would decipher the movements your lips makes and,then convert this into speech that the person on the end can hear.So substantially it reads your lips. In the proposed system speech impaired person and normal person can communicate effectively. This device could also be used for assisting laryngectomized patient.

VII. FUTURE SCOPE

In future this can be implemented as a mobile application, which brings in ease of communication and handy to carry. It will also allow people to make silent calls without bothering others.

VIII. REFERENCES

[1] Anish Kumar , Rakesh Raushan , Saurabh Aditya , Vishal Kumar Jaiswal , Mrs. Divyashree, An Innovative Communication System For Deaf, Dumb and Blind People Volume 5 Issue VI, June 2017 IC Value: 45.98 ISSN: 2321-9653

[2] S D, Lalitha & K K, Thyagharajan. (2016). A Study on Lip Localization Techniques used for Lip reading from a Video. International Journal of Applied Engineering Research. 11. 611-615.

[3] Hueber, T., Benaroya, E-L., Chollet, G., Denby, B., Dreyfus, G., Stone, M., Development of a Silent Speech Interface Driven by Ultrasound and Optical Images of the Tongue and Lips, Speech Communication (2009)

[4] J. Chung, A. Senior, O. Vinyals and A. Zisserman, "Lip Reading Sentences in the Wild," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017 pp. 3444-3453.

[5] D. E. King. Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10:1755–1758, 2009