Design of Sign to Speech Conversion System with Augmented Reality Assistant

[1]Nandhini V, [2]Indhumathi GK, [3]Ruth Christy J, [4]Mrs. B. Rajalakshmi
[1,2,3] UG Scholar, [4]Assistant Professor
[1,2,3,4] Sri Sairam Engineering College, Chennai, India.

Abstract: Patients with severe speech and motor impairments, who are not able to speak, require specific human-computer interfaces to communicate with the world. Existing system uses flex sensors that fit to hand. The hand movements are monitored continuously and are converted to speech signals. These hand wearing sensors may cause uncomfortable to the patients. In our system we enhance the augmented reality techniques to allow differently abled people to interact with digital world. The proposed system uses a camera that captures sign movements of the patient and are processed using C# and Python programs to identify signs shown by the patient. Then the output is displayed using augmented assisted screen as text or speech. The proposed system has proved to be efficient in terms of speed and processing time compared to formal flex sensor systems. The proposed system thus reduces complex hardware components and is compatible to handle.

***Index Terms*---Augmented Reality, Machine learning, Python, Vuforia, dumb people**

1. INTRODUCTION

Communications between deaf-dumb and a normal person have always been a challenging task. Generally dumb people use sign language for communication but they find difficulty in communicating with others who don't understand sign language. This project aims to lower this barrier in communication by developing a device that act as a speech assistant. Through this device the disabled persons can make communication, which is defined as an impossible thing for them. The existing system uses flex sensors that capture gestures since it does not require accurate coordinates like accelerometers. The flex sensors are two-terminal variable resistance sensors. It measures the amount of deflection or bending. This may use carbon elements. The programming tool such as MPLAB IDE and PICKIT are used to dump the code with PIC microcontroller. The flex sensors would give the output of a varying voltage values depending on the variation in the resistance values due to the bending of fingers. This analog output is converted to digital output by the A to D converter inside PIC and the output is displayed in LCD. The existing system also uses Arduino that produces speech output. The existing system has the following disadvantages: The soldering on flex sensor is difficult, Produce errors when used for a long time due to change in the flexibility of the sensor, too costly to build a control system for both hands as one will cost around INR 600-900. The proposed system eliminates these complex hardware components and uses more software dumped system to enable effective communication between the deaf-dumb and a normal person. The system uses camera to capture the sign shown by the disabled people and is fed through the series of image processing systems. The techniques identify the sign and then the information is displayed as text or heard through speakers. The augmented reality assistant thus uses to display the text output with specified options. There will be two options in the augmented reality designed screen such as text and speech. If the text option is touched the identified sign is displayed in the formal English language and if the speech option is touched then the person nearby can hear the information in the form of sound through speakers attached to the system. The system will have identified to provide better interaction with the deaf-dumb and normal people. Speed in sign identification is higher compared to the flex sensor systems.

2. LITERATURE REVIEW

The system with reduced dependencies of hardware devices is encouraged among the public and researchers'. The system provides the following. The system is portable to be used at anywhere at any time. The virtual keyboard is displayed in the PC. Based on the movement of head movements and eye blink the typing of letter occurs. In the proposed system the virtual screen contains two options text and speech. Only touch movements are sensed in the proposed system. Templates are included and the typed or selected word is also converted into speech. Here, we also measure the performance of our system Information Transfer Rate (ITR). Database is managed in the PC to measure the Information Transfer Rate. Based on the Experience of the user the efficiency will increase.

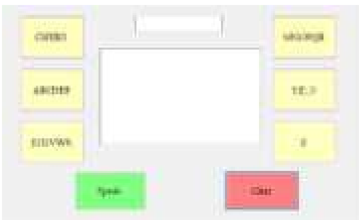

Fig. 1: virtually designed screen

The system converts the text into speech if the speech option is touched virtually. The system thus encourages the use of software supported system to do so. The system is found to have high efficiency in terms of measuring the Information Transfer Rate. Information Transfer Rate is given in bits per trial, is used as an evaluation measurement in a brain-computer interface. Thus upon continuous practice this method is observed to be efficient in communication. An observer can typically watch the screen more easily than the keyboard, and see which characters the mouse moves to. Some implementations of the on-screen keyboard may give visual feedback of the "key" clicked, e.g. by changing its color briefly. This makes it much easier for an observer to read the data from the screen. In the worst case, the implementation may leave the focus on the most recently clicked "key" until the next virtual key is clicked, thus allowing the observer time to read each character even after the mouse starts moving to the next character. The sign shown by the observer are easily monitored on continuous basis and thus clearly capture the gesture shown by the disabled people. Also when using augmented reality the seamless interaction will be good this is proved on recent survey. Thus the proposed system will have its own way of providing efficiency compared to the existing systems.

3. FLOW DIAGRAM

The flow diagram for the proposed system is shown below.

Fig. 2: Flow Diagram for the Proposed System

The above diagram uses few and simple hardware components and the system entirely depends on all the software components that process with the sign language. The gestures shown by the patient is fed to Arduino through the camera. Arduino fed the captured image to the PC system. The image is then processed through a series of programs and the information is gathered. The information is transmitted to the augmented reality screen where there will be options to choose between text or speech output.

The Arduino ATmega328 is used in the proposed system. The processing of image uses Python along with OpenCV and NumPy installations in the system. Several modules, programming files, object files and features are used for identifying and labeling the images. The relative information is then act as a source for the screen developed using augmented reality.

Vuforia is an augmented reality software development kit for mobile and PC that enables the creation of augmented reality applications. It uses computer vision technology to recognize and track planar images and simple 3D objects, in

real time. Vuforia provides Application Programming Interfaces (API) in C++, Java, Objective-C++ and the .NET languages. Vuforia has been acquired by PTC Inc. in November 2015.

VuMarks allow identifying and adding content to series of objects. This feature is used to create user interactive button for displaying the text information and to produce sound output through speakers. Speakers available for systems or additionally connected speaker can be used to display the output.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1. ARDUINO

The Arduino Uno is a microcontroller board based on the ATmega328 (datasheet). It has 14 digital input/output pins, 6 analog inputs, a 16MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button that simply connect it to computer with a USB scale or power it with an AC-to-DC adapter or battery to get started. The Uno differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it features the ATmega8U2 programmed as a USB-to-serial converter.

Fig. 3: Arduino Uno Board

"Uno" means one in Italian and is named to mark the upcoming release of Arduino 1.0. The Uno and version 1.0 will be the reference versions of arduino, moving forward. The Uno is the latest in a series of USB Arduino boards, and the reference model for the Arduino platform; for a comparison with previous versions.

4.2. C#

C# is a modern, general-purpose, object-oriented programming language developed by Microsoft and approved by European Computer Manufacture Association (ECMA), International Standard Organization (ISO). C# is designed for Common Language Infrastructure (CLI), which consists of the executable code and runtime environment that allows use of various high-level languages on different computer platforms and architectures.

4.3. Python 3.5

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding; make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python supports modules and packages, which encourages program modularity and code reuse.

4.4. VUFORIA

Vuforia can recognize and track a wide range of 3D objects as well. Object Recognition enables Objects Targets to be created by scanning physical objects. It allows you to create apps that recognize and track intricate rigid objects. The Vuforia Engine Driver Framework allows developers to provide and consume data for external systems through Vuforia Engine. This framework provides access to External Camera features. The External Camera defines all specifics required for Vuforia Engine to access external image sources.

4.5. OTHER REQUIREMENTS

WHL File – It is a package saved in wheel format (python distributions). It could compress (zip) files. Three types of WHL files

Universal Wheel - contains only python files, no compiled extensions.

Pure Python Wheel- does not support older versions Platform Wheel- python files and compiled extensions.

OpenCV – Open Source Computer Vision is a machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications. The library has more than 2500 optimized algorithms.

NumPy – Python library provides multidimensional array object, various derived objects (masked array, matrices), and sorting, selecting, statistical operations. It is nd array object. It has fixed size.

The elements in NumPy array should have same data type (same size in the memory).

```
FOR 1D ARRAY
c=[]
for i in range(len(a)):
        c.append(a[i]*b[i])
if a&b:
        for(i=0;i<rows;i++):
                c[i]=a[i]*b[i];

FOR 2D ARRAY
for(i=0;i<rows;i++):
        for(j=0;j<columns;j++):
                c[i][j]=a[i][j]*b[i][j];
```

Fig. 4: Sample codes using Python

5. METHODOLOGY

The following flow chart is used to convert the obtained image to labels.

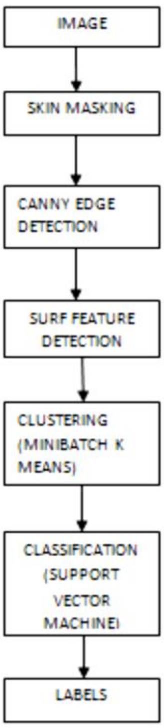

Fig. 5: Flow chart for sign recognition

The first process is skin masking. Skin detection is the process of finding skin-colored pixels and regions in an image. Skin image recognition is used in a wide range of image processing applications like face recognition, skin disease detection, gesture tracking etc., in the system preferred the skin masking is used mainly for human computer interaction.

Skin masking is followed by Canny Edge Detection. Canny Edge Detection is a popular Edge Detection algorithm. It involves stages such as noise reduction, finding intensity gradient of the image, non maximum suppression and hysteresis thresholding.

SURF- Speeded Up Robust Features, is a speeded-up version of SIFT. SURF approximates LoG with Box filter. Convolution of the box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. Also the SURF relies on determinant of Hessian matrix for both scale and location.

```
import cv2
import numpy as np
from matplotlib import pyplot as plt

img = cv2.imread('messi5.jpg',0)
edges = cv2.Canny(img,100,200)

plt.subplot(121),plt.imshow(img,cmap = 'gray')
plt.title('Original Image'), plt.xticks([]), plt.yticks([])
plt.subplot(122),plt.imshow(edges,cmap = 'gray')
plt.title('Edge Image'), plt.xticks([]), plt.yticks([])

plt.show()
```

Fig. 6: Sample codes for canny algorithm

K Mean algorithm is used for clustering of data which works on unlabeled numerical data and it will automatically and quickly group them together into 4 clusters. K-means algorithm includes three steps including initialization, cluster assignment and move the centroid.

```
randomly chose k examples as initial centroids
while true:
    create k clusters by assigning each
        example to closest centroid
    compute k new centroids by averaging
        examples in each cluster
    if centroids don't change:
        break
```

Fig. 7: K-means algorithm

Then classification of the sign shown by the patient is done by Support Vector Machine (SVM). SVM is a supervised machine learning algorithm which can be used for either classification or regression challenges. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/line). This is implemented using python programming.

The signs are classified and then is labeled using bag of words (BoW) histogram. The labeled sign is then given to augmented reality developed screen. Steps involved in creating a new button or command on the screen:

Start a new unity project

Set up a Unity Scene

Create a virtual button available in options Add the codes relevant for its operation Now the button will be ready to work with

Now the proposed system is completed.

Fig. 8: Vuforia working screen

```
using UnityEngine;

using Vuforia;

public class virtualButtonScript : MonoBehaviour, IVirtualButtonEventHandler

{         GameObject zombie;

          /// Called when the scene is loaded

          void Start() {

          zombie = GameObject.Find ("zombie");

          GameObject virtualButtonObject = GameObject.Find ("actionButton");

          virtualButtonObject.GetComponent<VirtualButtonBehaviour>
          ().RegisterEventHandler (this);

          }

          /// Called when the virtual button has just been pressed:

          public void OnButtonPressed(VirtualButtonAbstractBehaviour vb) {

          Debug.Log("button Pressed");

          zombie.GetComponent<Animation> ().Play ();

          }

          /// Called when the virtual button has just been released:

          public void OnButtonReleased(VirtualButtonAbstractBehaviour vb) {

          zombie.GetComponent<Animation> ().Stop ();

          }

}
```

Fig. 9: Sample codes for implementing buttons

Fig. 10: Final screen

VI. RESULT

Thus the designed system will produce either text output or speech output using the options provided on the virtual screen. When text option is disturbed the text corresponding to the sign shown by the patient is displayed. When speech option is disturbed by the user then speaker will produce the corresponding result. This provides effective communication for the people who don't understand sign language. They could communicate easily with dumb patients.

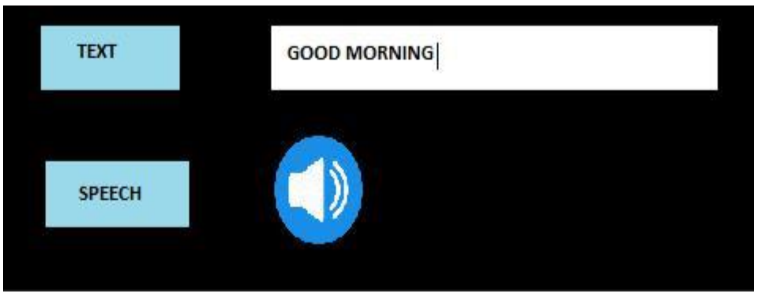

Fig. 11: Output

VII. CONCLUSION

The performance of the system includes the effective use of augmented reality technique to enable the patient to communicate with the normal people. This developed system not only benefits the dump people but also patients who are unable to speak due to various reasons such as aging. The K-means algorithm used is effective that is if the variables are huge, then most of the times computationally faster than the hierarchical clustering, if we keep k small. The proposed system thus reduces the number of hardware components used for communication. This proposed system was developed using python and Vuforia supporting programs.

VIII. FUTURE SCOPE

Further works that include development of mobile application can be developed that improves the communication between the patient and the normal people. The application thus developed will enable long distance communication. This helps when the patient is in emergency mode. Similarly designing more characters with additional options in the augmented reality developed screen is also entertained. Future scope of the proposed system may be also useful for elderly patients to communicate with their family members.

REFERENCES

[1] Virtual Signing: Capture Animation, Storage and Transmission – An Overview of the ViSiCAST Project ‖ , J>A> Bangherm, S.J. Cox, R. Elliott, J.R.W. Glauert, I.Marshel, Rankov, M.wellls(Televirtual), IEEE seminar on ―Speech and language processing for disabled and elderly people ‖ , London, April 2000.

[2] Augmented Reality Virtual Keyboard including Text to Speech Conversion ‖ T. Maheshwari, S. Menagapriya, P. Geetha, Asian Journal of Applied Science and Technology (AJAST), April 2017.

[3] Developing with Vuforia ‖ Vuforia Developer.

[4] Advantages of k-means algorithm for image processing ‖ , http://playwidtech.blogspot.com/2013/02/k-means-clustering-advantages-and.html.

[5] Hubert ceccoti “A Multimodal gaze-controlled virtual keyboard” in IEEE transaction on human-machine system 2016.

[6] Hafiz adnan habib, mufti “Real time mono vision gesture based virtual keyboard system”, IEEE transactions on consumer electronics 2006.

[7] Amer Al-rahayfeh “Eye tracking and head detection movement” IEEE journal of transaction on engineering in health and medicine in 2013.

[8] Eda akman aydin, Omer Faruk bay “ Region based brain computer interface for a home control application”, IEEE journal on Communication in 2012.

[9] “Vocal Platform for Telephone Voice Portals and Internet Based Interfaces”, m-ICTE Badajoz 2003. V. Tomico, N. Morales, E. Campos, J. Tejedor, D. Bolanos, S. Jimenez, J. Garridoy J. Colas.

[10] http://www.dylogic.com/sito/articlesDMD/PSE_3G_Gateway.html.

[11] http://www.sipfoundry.org/sipXezPhone/

[12] “Signing for the deaf using virtual humans”, J.A. Bangham, S.J. Cox, M. Lincoln, I.Marshall, M.Tutt, IEEE Seminar on “Speech and language processing for disabled and elderly people”,London, April 2000.

[13] “The HTK Book Version 3.2, S. Young, G. Evergreen, T. Hain, D.Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, Cambridge University Engineering Department, December,2002.

[14] Seo, B.K. Choi, J. Han, J.H.Park, “One handed interaction with augmented virtual objects on mobile devices. In Proceedings of the 7th ACM SIGGRAPH Internation Conference on Virtual Reality Continuum and Its Application in Industry”, 2008.

[15] M.Higuchi, T. Komuro, AR typing interface for mobile devices Vision-based hand interaction in augmented reality environment, In proceeding of the 12th International Conference on mobile and Ubiquitous Multimedia, 2013.