International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 06 | June-2016 www.irjet.net p-ISSN: 2395-0072

Gesture to Speech Conversion in "HINDI" language for Hearing & Speech disabled persons in INDIA

Anjali Singh[1],Balram Timande [2]

[1]*Digital Electronics*
Department of Electronics and Telecommunication Engg.
RCET, Bhilai
Email id: anjali.singh15692@gmail.com
[2] *Associate Professor*
Department of Electronics and Telecommunication Engg
RCET, Bhilai
Email id: balramtimande@gmail.com

Abstract*: Communication to each other is a very important part of human life. Still there are some persons who are not able to speak like normal people they are known as speech impaired persons or in other words we say "dumb and deaf" persons, and for them communication is a challenge. For understanding them, we have to learn their language i.e. sign language or finger language. The proposed system of this project will help to solve this problem up to some extent. In this system gestures are converted into speech by which the communication between normal people and speech impaired person can become much easier. Purpose behind this system is two-fold. This system has different modes of operation like image segmentation, feature extraction, PCA extraction, and applying neural network. The image of the sign gestures are captured through the webcam. Then the image segmentation and feature extraction is processed. The output of feature extraction is the input to the PCA and then output of the PCA plot will given into neural network. Then the sign are converted into speech by using speech recognition system. Through this project the communication gap between speech person is become much smaller between speech impaired persons and normal person.*

***Key Words**: **Sign language, image segmentation, extraction, neural network, speech recognition**.*

1. INTRODUCTION

Communication is the activity of exchanging the information between two or more people in order to convey or get the meaningful message through a shared system of signs and semiotic rules. The main application of communication is known the motive and meaning of the information and its translation to the others.

Communication is an internal process and established between three main subject Categories: human beings, organism which comes under living and communication-enabled devices.

Communication which come under category of living beings either human being or animal usually occurs through visual, auditory. Human communication is unique for its extensive use of language. But all kind of people or persons are not able to communicate like normal people and it is main problem in today's world. The devices which are invented for them are cannot be afforded by all kinds of persons that's why it is difficult for them to communicate with other peoples.

Many Deaf people cannot speak a natural language, technically they known as: mute" [1]. When a person is mute or keep silence then we say that is a dumb person When a person is unable to make speech or any type of sounds correctly or fluently, or has problem with their voices so that problem is known as **speech disorder**. Difficulty in vocalizing the sound and words, or having utterance disorders, and stuttering are examples of speech disorders.

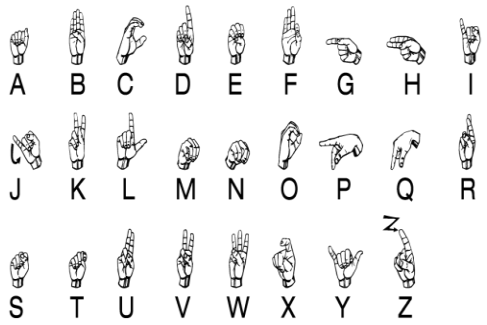

Fig.1 Indian Sign laguage Hand gesture

The gesture and sign of any word or any information is a that form of communication in which there is no use of any verbal sentences or it can be known as silent communication in which the person will not talk with the help of speech, the body gestures and visible body action helps communicate to and display picture of a particular message. We know that if the movement of a human hand or eyes or any other part of body tell some message or display some pictures of information than that movement of hand than it is type of gestures. If a person talks by using gestures of body to another person then it allows other person to transform a variety of feelings and thoughts of their mind, through the action of their hand and face, rather than to speak.

In this project basically image is captured through the camera, and then vision analysis is performed and then function in the operating system and the speech output is provided through the speech recognition system.

Fig. 2 Hand Gestures of HINDI Sign Language

1. LITERATURE SURVEY

The algorithm proposed is applied on 200 images with 40 different gestures. With the help of defined features and five bit pattern encoded sequence, it can successfully recognize the 40 different hand gestures patterns[1]. This paper proposed three main steps they are image segmentation, pre-processing, orientation detection, feature extraction and classification[1]. Present a system that will not only recognize the hand gestures automatically but also convert it into the corresponding voice output so that speaking impaired person can easily communicate with the other normal people[2]. The system G2S , has been developed using the skin colour segmentation. Pre-recorded speech tracks is made for the corresponding dataset and play at the ouput.

A method which can be used for two way communication that means gesture is converted into speech and then the speech or word converted into gesture , this system makes the communication very easy between both the hearing and speech impaired people and the normal people[3]. The Indian Sign Language(ISL) gesture are drawn out from the real time video and then map it with human-understandable speech. Second module will take natural language as input and map it with equivalent ISL animated gesture[3].

Basically this paper presnet a very innovative and enhanced method in which with the help of a represntation technique (LOTS Notation) which will map the kinematic data corresponding to each word and animate it dynamically we have improved the overall accuracy and a speed of animation[4].A novel method based dynamically animated gestures are devised by us in order to map the kinematic data for the corresponding word[4]. The word after translation into ISL will be checked in the database where in lies the notation format for each word or not. The device consist of two different parts in first there is a small portable device with a display and the gloves which have to be worn by the hand of hearing person. The device has to be held by the normal person with the LCD display person[5]. The Microsoft Kinect a motion capture devices from Microsoft is used in this project. The system is stably designed to identify 20 of human joints like (head , hands , and so on...). A series of experiments are being conducted to evaluate the system's return and limitation. In a test done for 100 spells for different signs , accuracy of 90% have been achieved[6].

2. PROPOSED METHODOLOGY

a. Image Acquisition

Image Acquisition is the main and the foremost step in this project .So in this study, the input can be an image or a sequence of images (video), or it can be the data set downloaded from the internet, the image which is taken can be from webcam of computer or can be an external camera device [1]. After collecting the images second step is to create database. There are total 46 letters in Hindi language. The whole data set of images is divided into two sets i.e., training set and test set. Each gesture represents a definite word of Hindi language [8] used in Sign language gestures. The images of hand gestures are real having different sizes and shape captured by single digital camera.

The total number of 126 images is taken to make the data set which represents the gestures of sign language. All the hand images used in this project were taken from and external camera with high resolution, all the images are taken from a single camera with simple background and stable light. Here we see some captured gesture images in Figure 3

Fig. 3 Some hand gesture images

b. Gesture Area Segmentation

To obtain the depth image of the scene (which is 320*240 px) the depth image forming principle is used [1]. Capture image is derived in to Gesture area and the background region. Gesture region are segmented by depth values. This segmentation mainly depends upon the distance between the gesture and the camera. After this the value of scene depth information is converted the gray value space. The histogram is calculated to search the threshold for segmenting gesture area from background. It is also possible that the part of the gesture is often close to camera and gesture area is smaller than background area. Pixels are obtained which change the gray value from large to small at first large values are used as threshold for segmentation.

c. Image Extraction

This is the fast and robust feature matching algorithm and is mainly used in image registration. This algorithm uses key point detecting key point description & key point matching. The gesture recognition algorithm can be applied to static gesture recognition.

Key point detection it needs to build scale space. Firstly the integral image is converted to original image and integral image is convoluted with original image in order to get the scale space of real time image.

It includes two steps firstly the critical point are determined and then corresponding key descriptors are generated. It is done in three simple steps, firstly we follow fast indexing of SURF algorithm then nearest neighbor matching ideas are chosen. Finally matching results are judged according to key point matching pairs.

d. Neural Network (NN)

The Neural Network was discovered in the years 1940 to 1970. Between these years many scientist worked in this field. Warren McCulloch and Walter Pitts are the two persons who simulate the biological neural network research. The model which is consist of more than one inputs and only one output is refer as the very first explained model of NN given by McCulloch-Pitts. The Neural Network referred as a mathematical model for transmitting processing similar to the biological neural based on the following rules [14]:

1. Neurons are the information processing units.
2. There is a communication link between the ways of neurons is known as connection link, and there is a certain weight value for each connection link.
3. The neuron produces the output only through the activation function.

The very popular neural network algorithms are feed forward back propagation algorithm which is used with the image classifier. Back-propagation algorithm comes under supervised learning. In the supervised learning the inputs and desired output is given before the training process [15].

Neural Network has two main stages in its training phase they are features extraction, and training the classifier to learn these features.

The input image can included other thing also rather than hand thus, to remove the other portion from the image; image segmentation is done for the proper hand image. In the stage of feature extraction, the NN classifier is trained to recognize the user's hand gesture. Similarly in the training stage the extraction of features is performed in the testing stage.

The structure of proposed back-propagation NN is illustrated in below Fig.11. There are total 240 neurons in the input layer and 10 neurons in the hidden layer. The number of neurons and hidden neurons were determined by "trial and error" method; and the number of neurons in the output layer is the 4 hand Gesture images. .

The Neural Network is trained to classify user's hand gesture features. The images of hand gestures in the database consist of total 126 single hand images of the user's either right

or left hands , captured by external digital camera. The random values are initialized for the weight between the input in the training phase, and the hidden and output layer. Then, the errors are computed at the output with the target values. The whole process is echoed until the error becomes less at the output. There is one activation function is used i.e., "Hyperbolic tangent sigmoid" between the input layer and the hidden layer, while another activation function i.e., "linear transfer" is used between the hidden layer and the output layer. The algorithm back propagation used is inclined descent with momentum back propagation.

Fig. 12 shows the performance of the trained neural network, determined by validation performance between the number of iterations and the mean squared error. The best validation performance was 2.6348 at epoch of 195. This result is acceptable due to the small mean-square error. However, the regression analysis illustrates that, the correlation coefficient equals to 0.98048 between the output and the target for training; which means that the both output and target are very close, which represents excellent fit .

3. EXPERIMENTAL RESULT

In this project the image data set is created by capturing the image of hand gestures of Hindi letters Sign language. All the images are of same resolution that is 320* 240 pixels and the rate is 30 frames per second. The background of all the images is uniform. As we know the colors of human skin are not same so the RGB image is converted into gray image [14] for better recognition. Figure shows the gray image of hand gesture.

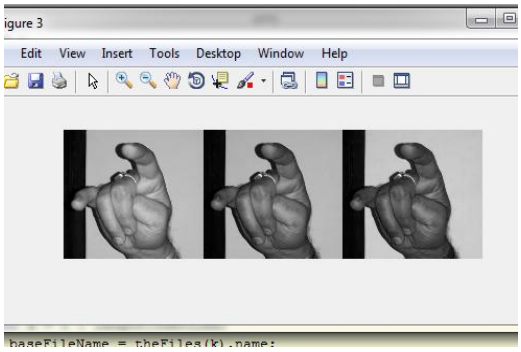

Fig. 4 Gray image of hand gestures

Now the gray image is converted into YCbCr lab space [11]. Each pixel of the image is transformed into lab space. The lab space then detects the skin color area in the image and removes the unwanted background and then it converted into binary image. Resizing of image according to the need to program is done. Because of that accuracy is increased up to the 95%.

Fig. 5 RGB to Lab Space Conversion

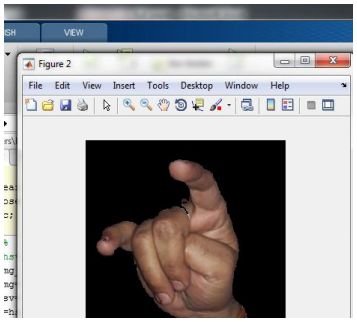

Fig. 6 Skin Detection

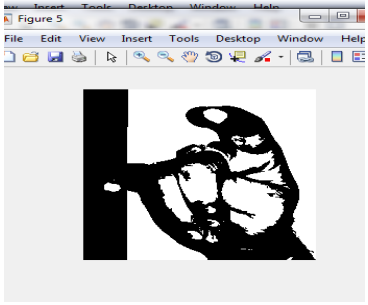

Fig.7 Binary Image

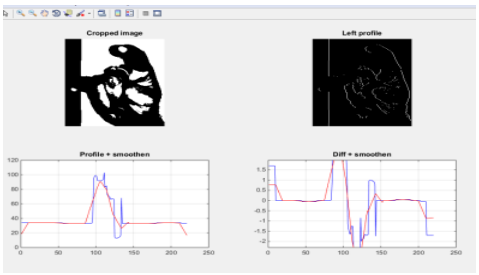

Fig.8 Cropped image

Feature extraction is done by the use of Principle Component Analysis algorithm [10]. This algorithm is the most prominent method to extract the feature of the image of the gestures. Result of PCA algorithm is shown in the below figure:

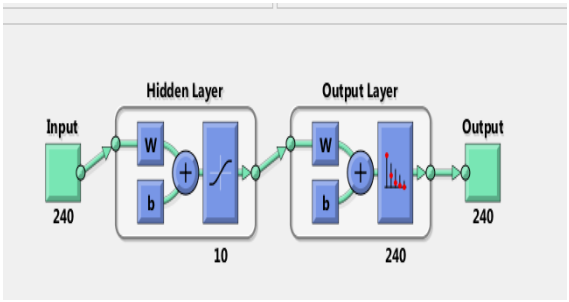

Fig.9 PCA Extraction

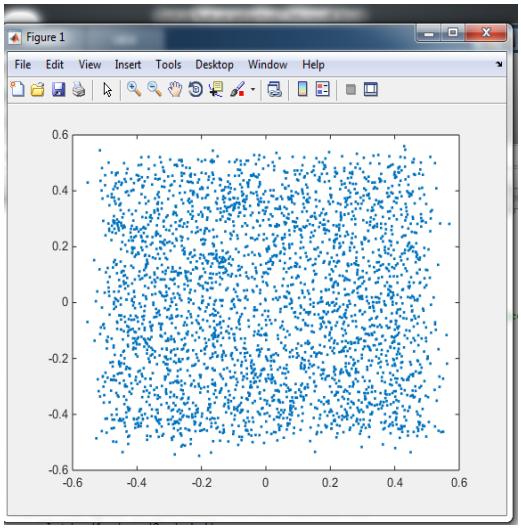

Fig.10 PCA Plot of the extracted image

The PCA output is used as the target value to train the neural network. The neural networks train the input hidden layer and the output target values and give the following output:

Fig. 11 Structure of neural network

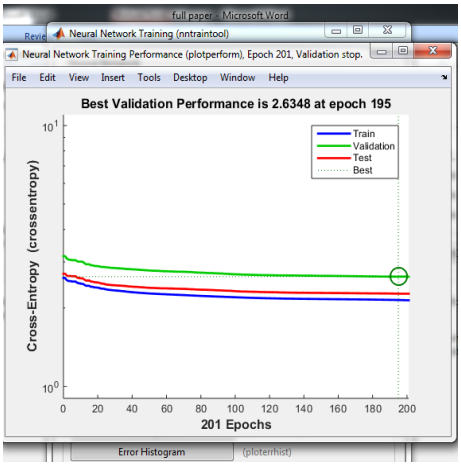

Fig. 12 NN Training performance

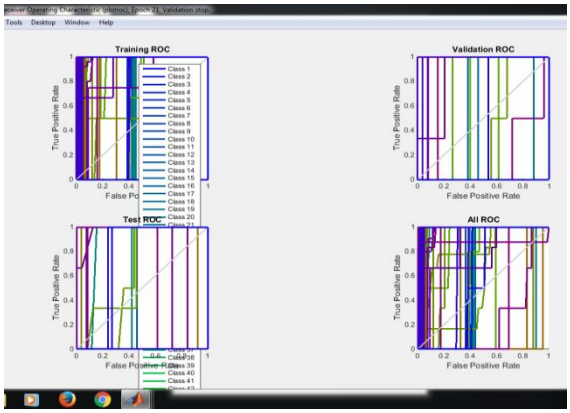

Fig. 13 Regression analysis without error

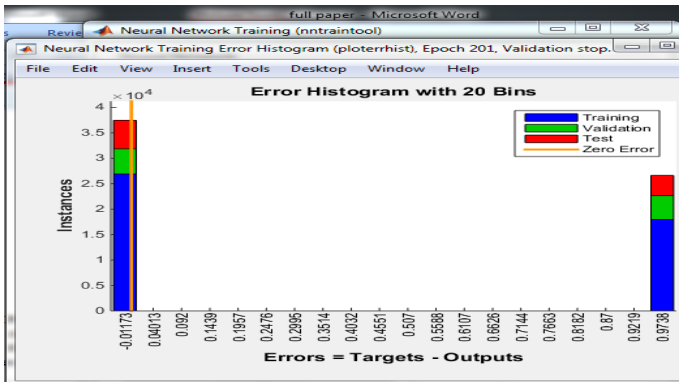

Fig. 14 Error performance

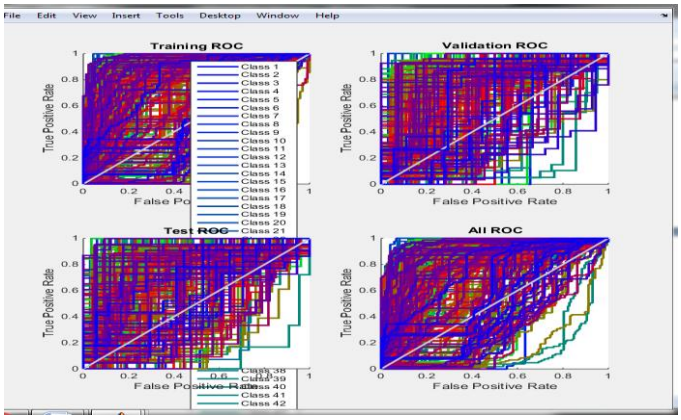

Fig. 15 Regression analysis with some error

The last process of this system is the speech recognition which is done by using a Windows text to speech system. A communication set up is done by using "actserver" command of matlab. The windows text to speech – Microsoft Speech API speech recognition system is used which gives better quality of speech output when compared to other text to speech techniques. It gives accuracy of up to 93% in speech recognition when used with matlab.

4. CONCLUSION

This system is going to be designed for better communication between dumb and deaf people and normal people. By the conversion of gesture into speech it become very much easy for us to explain our words to hearing and speech disabled persons also they can express their words to us in more efficient way by this the way of communication between hearing and speech disabled person and normal person is much developed the person who don't understand English language now they can also use this software because it is mainly those person who can understand and speak Hindi language.

REFRENCES

[1]. Hongwu Yang, Xiaochun An , Dong Oei, Yiong Liu . 2014. IEEE, Towards realizing gesture-to-speech conversion with a HMM- based bilingual speech synthesis system.

[2]. D.K. Vishwakarma and Rajiv kapoor , 2012. Simple and intelligent system to recognize the expression of speech-disabled person. IEEE , International Conference on intelligent human computer interaction.

[3]. Prof. R.RItkarkar and Dr. Anil V Nandi , 2013. Hand Gesture to speech conversion using Matlab, IEEE.

[4]. Prof. Prashant G Ahire, Kshitija B, Tilekar, Tejaswini A, Jawake, Pramod B Warale. ,2015, IEEE, International conference on computing communication control and automation.

[5]. Ruvinash J Raghavan, Kiran A Prasad, Rahul Muraleedharan, Geetha.M. 2013, Animation system for Indian sign language communication using LOTS notation.

[6]. M.Sethuraman , Abhinav Rajan, K.P Sai Saran, 2015, Beta Real Time Interface for deaf and Global Entity, IEEE ,2015 International conference on smart technology and management for computing, communication, controls, energy, and materials.

[7]. Rajaganapathy.S, Aravind.b, Keerthana.B, Sivagami.M , 2015, Conversation of Sign language to Speech with Human Gestures, ELSEVIER .

[8]. Itisha Gupta, Vinti Parmar, Shubham Tayal, February 2014, Hand Gesture Recognition for Hindi Vowels Using OpenCV, International Journal of Application or Innovation in Engineering & Management (IJAIEM).

[9]. Neha Poddar, Shrushti Rao, Shruti Sawant,Vrushali Somavanshi, Prof. Sumita Chandak, February 2015, Study of Sign Language Translation using Gesture Recognition. International Journal of Advanced Research in Computer and Communication Engineering

[10]. Shreyashi Narayan Sawant , M.S. kumbhar, 2014, Real Time Sign Language Recognition using PCA ,IEEE International Conferrence on Advanced Communication Control and Computing Technologies(ICACCCT).

[11]. Vivek Agarwal, Besma R. Abidi , Mongi A. Abidi , March 2006, An Overview of Color Constancy Algorithms. Journal of Pattern Recogniton Research.

[12]. Chen Guanghua , Zhang Xiaolong , 2016, A Method to Improve Robustness of the Gray World Algorithm, 4th International Conference on Computer, Mechatronics, Control and Electronic Engineering (ICCMCEE 2015).

[13]. Anirudh Garg, 2012, Converting American Sign Language to Voice output using RBFNN.

[14]. Hamid A. Jalab , Herman .K. Omer, Human Computer Interface Using Hand Gesture Recognition Based On Neural Network .

[15]. H, AL-JBARA.M, KIAH. H.A, JALAB.2012. Increased capacity of image based steganography using artificial neural network. AIP Conf. Proc, 2012. 20-25.

BIOGRAPHIES

1. Author

Anjali Singh[1], I am the student of M.Tech, branch Digital Electronics in RCET, Bhilai, I have completed my B.E from Ashoka Institute of Technology and Management,Rajnandgaon,Chhattisgarh in 2013. I live in Rajnandgaon.

2. Author

Prof. Balram Timande, received B.E. in Electronics from Nagpur University in 1993-94 and M.Tech. in Electronics and Tele communication Engineering from Bhilai Institute of Technology, Durg, C.G. in 2007-08.