International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 5,May 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

AIDING SYSTEM FOR DEAF AND DUMB PERSONS BASED ON SPEECH AND IMAGE PROCESSING ALGORITHM

Anju V.Sathyan[1], Bindu.V[2]
Student, ECE,Sree Chitra Thirunal College of Engineering,Trivandrum,India[1]
Assistant Professor, Dept of ECE, Sree Chitra Thirunal College of Engineering,Trivandrum,India[2]

Abstract-Communication plays an important role in our everyday life. Simply communication means exchanging ideas meaningfully. The capability of hearing and speaking has a major role in day to day human-human interaction. So hearing and speech impaired peoples find its difficulty for communicating with others. The only way for those peoples is sign language. But sign languages are only effective means of communication among disabled persons. Sign language medium is not an effective mode of communication for the disabled persons with normal one. This proposed system enhances the quality of communication of disabled persons. System consists of a two way communication. It aims at providing an interface between the voice and the visual sign language and vice versa. This is achieved using Natural Voice processing and Digital Image Processing algorithms. This system that enables impaired people to further connect with their society and aids them in overcoming communication obstacles created by the society's incapability of understanding and expressing sign language.

Keywords: Deaf and Dump; image processing; speech processing;MFCC; DTW

I. INTRODUCTION

Hearing and speech impaired persons suffers difficulty to communicate with normal one and vice versa. So these mute people are isolated from the non-impaired people's community. Sign language is only way for communicating the mute with others. But sign language not serves an effective way for communicating the mute with normal persons. So the only way for enhance the communication between mute people and normal people is recognition of sign language and converting it to the corresponding voice. For further enhancing the communication of normal persons with mute is done by recognizing the speech signal and converting it to the corresponding sign language. So the proposed aiding system consists of conversion of sign language into corresponding voice and conversion of speech signals into sign language which is understandable for the mute persons. American Sign Language is used for simulating the results. For the simulation results ASL numbers 1-9 static signs are used.

II. DATABASE

Recognition for hand gestures creation of a train data set is needed. Different hand gestures of size 400*300 of .jpeg color images are created. The dataset consists of static signs of asl1-9 numbers. Similarly a database for speech signals is also created.

III. SYSTEM DESCRIPTION

The proposed system consists of mainly two process i.e. recognition of hand gestures and conversion to sound signals and recognizing the sound signals and converting to the corresponding sign language gestures.

A. Sign gesture to sound conversion

For the conversion of sign gesture to voice conversion image processing methods are used. First capturing the image using 16 megapixel web camera. The RGB color images are then cropped and converting to gray scale image. Then filtering the image to remove noise and converting to binary image. Trace the hand gesture region and finding the region of interest. Then feature extraction is done .Finally feature matching is done for gesture recognition and corresponding voice is produced.

B. Sound to sign gesture conversion

For the conversion of voice to corresponding sign recognition, input is speech signal .The input speech signal is recorded using micro- phone. For speech feature extraction Mel Frequency Cepstral Coefficients with Cubic Log compression is used[5].For feature matching Dynamic Time Warping algorithm is used and corresponding sign gesture is produced at the output.

IV. SIGN GESTURE TO SOUND CONVERSION

The block diagram for sign gesture to voice conversion is shown in Fig .4.1.

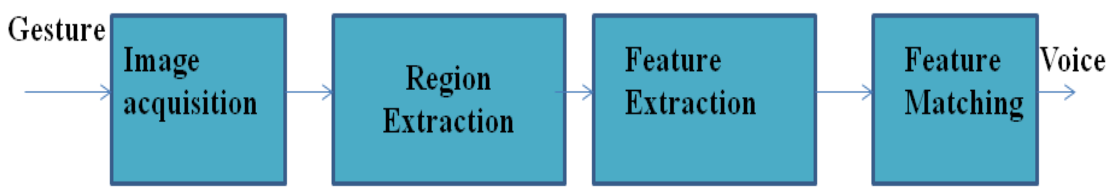


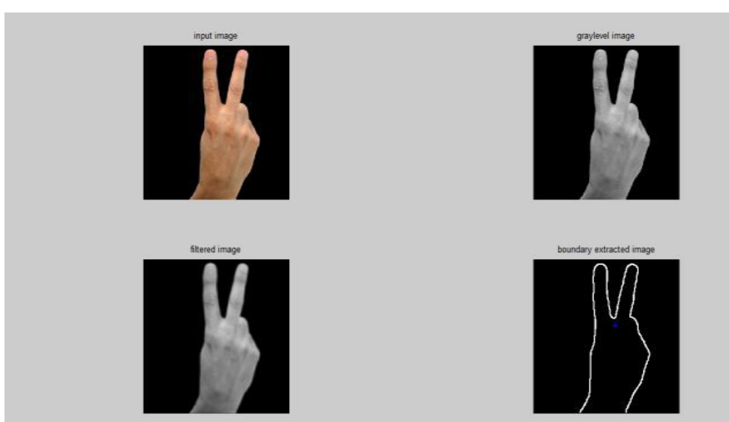
Fig .4.1 Block diagram of Gesture to sound conversion

A. Image acquisition

In this step the color RGB image is converted to gray scale image using matlab inbuilt command and removing the noise from the image using Gaussian filter. Converting the filtered gray scale image to binary using matlab inbuilt command.

B. Region Extraction

The second step involved in gesture recognition is region extraction of signs. 8- component connectivity of binary image is calculated and matlab command imcrop is used for finding the region of interest(ROI).Canny's algorithm is used for edge detection of region extracted binary image. The main advantages of canny's edge detection algorithm are better detection of edges and improved signal to noise ratio.



International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 5,May 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

Fig .4.2 Region Extraction of sign gesture

C. Feature Extraction

Feature extraction is one of the important steps of gesture recognition. For feature Extraction different features like binary area, centroid of the region extracted binary image is calculated. By using the matlab function bwarea binary area of different sign gestures and centroid is calculated using function region props. Feature vector formed by centroid [2] is given by (C_x, C_y)

$$(\boldsymbol{Cx}, \boldsymbol{Cy}) = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{xi} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{yi}$$

The feature vectors of database images calculated and store as a .mat file and the input sign gesture is calculated at run of the program.

D. Feature Matching

Feature Matching is the final step of sign gesture recognition. Here Distance Euclidean method is used for gesture recognition. Euclidean distance between input image and database image is calculated using

$$\boldsymbol{D} = \sqrt{(\boldsymbol{x}_{\boldsymbol{i}+\mathbf{1}} - \boldsymbol{x}_{\mathbf{1}})^{\mathbf{2}} + (\boldsymbol{y}_{\boldsymbol{i}+\mathbf{1}} + \boldsymbol{y}_{\mathbf{1}})^{\mathbf{2}}}$$

V. SOUND TO SIGN GESTURE CONVERSION

The block diagram for sound to sign gesture conversion is shown in Fig.5.1

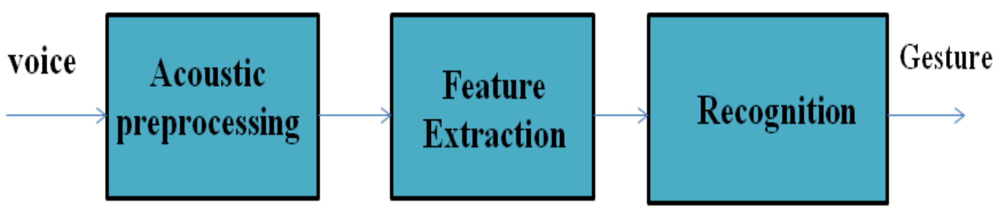


Fig.5.1. Block diagram of sound to gesture conversion

A. Acoustic preprocessing

In this step the input voice signal recorded using microphone is used for further preprocessing. The recorded signals are converted to .wav file. For processing with matlab platform .wav sound files are needed. Acoustic preprocessing involving removal of noise from the speech signals.

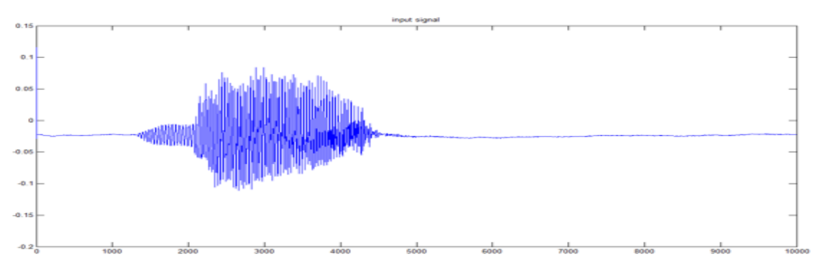


Fig.5.2. Input Speech Signal

B. Feature Extraction

This is one of the important step used for speech recognition. The feature extraction stage involving converting the speech signals into acoustic feature vectors. The main methods of speech feature extraction Mel frequency cepstral coefficients are used. But here instead of logarithmic compression cubic log compression MFCC [5] is used. The block diagram for steps involved in speech feature extraction is shown in Fig.5.2.

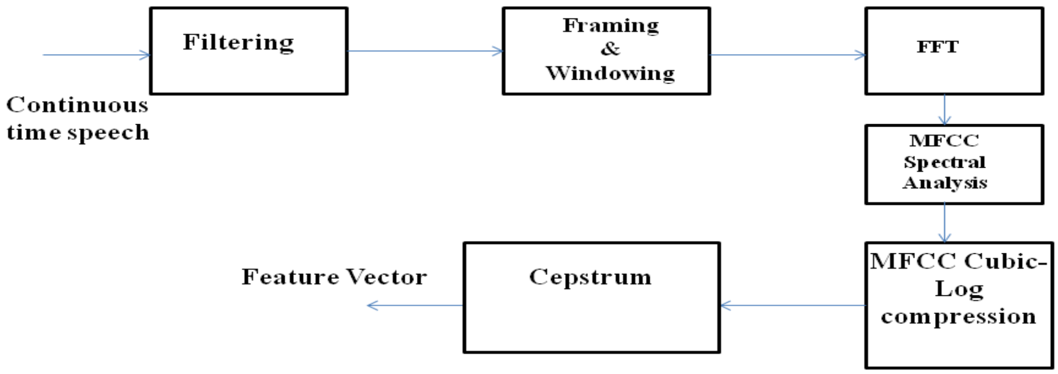


Fig.5.3.Block diagram for steps involving feature extraction.

a. Filtering

Speech waveform has high dynamic range. In order to reduce this filtering is done. Band stop filter [7] is used for filtering. A band stop filter consisting of placing a low pass filter parallel with a high pass filter.

b. Framing and windowing

Here the filtered speech signal is divided into frames of N samples with adjacent frames are separated with M where(M<N).The first frame consists of first N samples .The second frame begins M samples after the first frame and overlaps it by N-M samples and so on. The process is continued for till the input signal is accounted for. Windowing is applied for each frame in order to reduce the signal distortions at starting and ending of each frame. Hamming window is used.

$$H(k) = 0.54 - 0.46cos\left({}^{2\pi k}/_{N-1}\right)$$

Where $0 \leq k \leq N-1$

c.FFT

FFT is used to convert each frame of N samples in time domain to frequency domain. The inbuilt Matlab function is used.

d. MFCC with cubic log compression

In this step the frequency axis is initially warped to the Mel scale which is below 2 KHz and logarithmically above this point. Triangular filter are equally spaced in the Mel scale are applied on the warped spectrum and then cubic log compression function is applied on the spectrum. Finally DCT is applied to obtain the cepstrum of the signal. The whole process can be explained step by step as follows.

Step1:

The input signal is filtered using band stop filter.

Step 2:

The preprocessed speech signal is transformed into frequency domain by using Fast Fourier transform method.

$$X(m) = \sum_{n=0}^{N-1} x(n) * w(n) * e^{-j2\pi mn/N}$$

Where N is the frame size and w(n) is the hamming window.

Step 3:

Calculate the energy spectrum

$$\mathrm{Y(m)}=|X(m)|^2$$

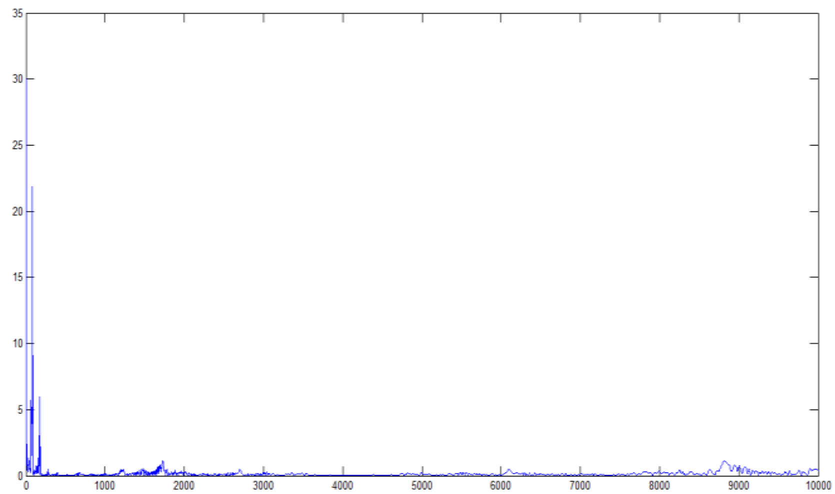


Fig .5.4. Mel Warping

Step 4:

Calculate the energy in each Mel window

$$s[k] = \sum_{m=0}^{N/2-1} w_k(m) * Y(m)$$

Where $1 \leq k \leq M$, M is Mel windows number which is generally 20 or 24.W_k (m) is the triangular weighted window function.

Step 5: Find cubic log cosine transform

$$c[n] = \sum_{k=1}^{M} log^3(s[k]) * cos[n * (k + 1/2)\pi]/M$$

Where $1 \leq n \leq L$, L is the desired order of M.

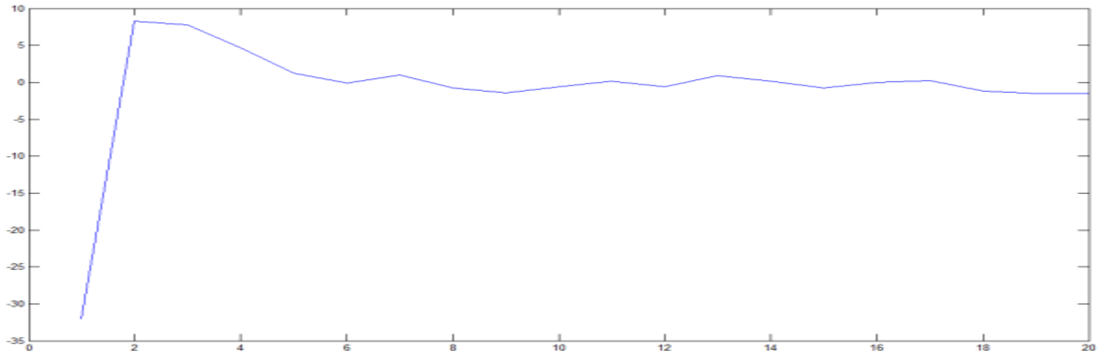


Fig.5.5 Cepstrum

C. Dynamic time warping algorithm

Dynamic time warping algorithm is used for speech recognition. In this type of speech recognition technique the test data is converted to templates. The DTW algorithm finds an optimal match between two sequences of feature vector. The recognition consists of matching the incoming speech with the stored templates. The template with the lowest distance measure from the input is recognized the speech signal. The distance measure between two feature vector is calculated using Euclidean distance metric. Therefore the distance measure of feature vector X of signal 1 and feature vector y of signal 2 is given as

$$d(x, y) = \sqrt{\left(\sum_i x_i - y_i\right)^2}$$

To obtain a global distance between two signals a time alignment must be done. Finally recognizing the input speech signal and producing the corresponding sign gesture.

VI. CONCLUSION

This system provides an aiding for the deaf and dump peoples by recognizing the sign gestures and produce corresponding sound signal and also recognizing the sound signal and producing the sign gesture as output. It provides an interface between the voice and the visual sign language and vice versa. This system that enables impaired people to further connect with their society and aids them in overcoming communication obstacles created by the society's incapability of understanding and expressing sign language.

REFERENCES

[1] Prashanth Suresh, Niraj Vasudevan, Nilesh Ananthanarayanan *"Computer-aided Interpreter for Hearing and Speech Impaired"* Fourth International Conference on Computational Intelligence, Communication Systems and Networks, 2012 .

[2] Jayashree R. Pansare1, Shravan H. Gawande2, Maya Ingle. *"Real-Time Static Hand Gesture Recognition for American Sign Language (ASL) in Complex Background"* Journal of Signal and Information Processing, 2012, 3, 364-367.

[3] J.Rekha, J. Bhattacharya, S. Majumder. *"Hand Gesture Recognition for Sign Language: A New Hybrid Approach",* In proceeding International Conference on Image Processing, Computer Vision and Pattern Recognition, IPCV ,2011

[4] Pankaj Pathak, *"Speech Recognition Technology: Applications & Future",* International journal on Advance Research on computer science, volume-1, Nov-Dec 2010.

[5] M. Ramya Devi, *"A Novel Approach for Speech Feature Extraction by Cubic-Log Compression in MFCC"* Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22.

[6] Q. Y. Zhang, F. Chen and X. W. Liu, *"Hand Gesture Detection and Segmentation Based on Difference Back-ground Image with Complex Background,"* Proceedings of the 2008 International Conference on Embedded Soft-ware and Systems, Sichuan, 29-31 July 2008, pp. 338-343

[7] J.-S. Lee, Y.J. Lee, E.H. Lee and S.-H. Hong, *"Hand Region Extraction and Gesture Recognition from Video Stream with Complex Background through Entropy Analysis*," Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, 1-5 September 2004, pp. 1513-1516.

[8] Hiroshi Matsumoto and Masanori Moroto *" Evaluation of Mel-LPC cepstrum in a large vocabulary continuous speech Recognition ".* IEEE Transactions on Audio, Speech, and Language Processing, 2001

[9] S.K., Podder, "Segment-based Stochastic Modelings for Speech Recognition". PhD Thesis. Department of Electrical and Electronic Engineering, Ehime University, Matsuyama 790-77,Japan, 1997.

