Digital Communicator for people with Visual and Speech Impairments

Anju K S, Apoorva H, Apoorva S Kulkarni, Chandana N
Department of Electronics& Communication Engineering
GSSS Institute of Engineering & Technology for Women, Mysuru

Abstract: **In this paper, an efficient communication method for visual and speech impaired people have been developed using several algorithms. Communication is important part of human life. Communication helps in mutual understanding. All over world, deaf and dumb people struggle in expressing their feelings to other people. Deaf and dumb people use hand gestures for their communications. These gestures are understood by only few people, hence it is necessary to convert these gestures into some other form which is understood by everyone. Thus an algorithm has been implemented to convert these gestures into speech. Another algorithm has been implemented to convert speech to text to help these people in the other way of communication.**

Keywords- ***Image processing, Image recognition, Speech processing and recognition, Linear Predictive Code(LPC), Artificial neural network, Principal component Analysis(PCA), Linear Discriminant Analysis(LDA).***

I. INTRODUCTION

1.1 OVERVIEW

A gesture may be defined as a movement, usually of hand or face that expresses an idea, sentiment or emotion. Sign language is a more organized and defined way of communication in which every word or alphabet is assigned some gesture. Sign language is mostly used by the deaf, dumb or people with any other kind of disabilities. Our aim is to design an algorithm which converts sign language to speech and speech to text so that communication is established between normal and visually or speech impaired people.

There are two processes to be implemented,

- Image to speech conversion
- Speech to text conversion

Image to text conversion involves image recognition and corresponding text generation. Image recognition is done by PCA and LDA methods. PCA is used to project images from the original image space into a gesture-subspace, where dimensionality is reduced and here intra class scattering is more and inter class scattering is less. LDA accepts high dimensional data(raw images) as input, and optimizes Fisher's criterion directly, without any feature extraction or dimensionality reduction steps. LDA is an enhancement to PCA, it constructs a discriminant subspace that minimizes the scatter between images of same class and maximizes the scatter between different class images. After image is recognised corresponding text is genertaed. Thus generated text is converted into speech using TTS(text to speech converter).

Speech to text conversion involves speech recognition and corresponding text display. Speech recognition involves speech sample training and recognition. Training of speech samples is done using LPC and neural network. Speech recognition is done by matching the input sample with the trained database. Once speech is recognised corresponding text is displayed.

1.2 OBJECTIVE

- To convert the sign language of deaf-mute people into text and voice so that it can be heard by blind people.
- The other process is to convert voice into text so that it could be read by deaf-mute people.

1.3 EXISTING SYSTEM

The existing system involves a speaking module which requires flex sensor, voice module and a microcontroller. Flex sensor and microcontroller are used in this system to capture the letters and store them. The work of flex sensor is to obtain changed position of fingers and to capture letters. To analyze these positions microcontroller is used. The microcontroller is placed on the device so that by changing the finger for conversation, flex sensor and microcontroller values will get changed, by comparing these two values, output is displayed and voice module gives the output. But this is costlier since protected layers have to be placed in order to save the circuit, battery and speaker from water. Also it's not convenient to use the module everywhere by physically challenged people.

1.4 PROPOSED SYSTEM

The process involves converting hand gestures of deaf-dumb people to speech so that normal people can understand their gesture. For the other way of communication where dumb people cannot hear what normal people speak, we are converting speech into text which is read by dumb people. For converting hand gestures into speech, we will acquire image of gesture made through webcam and process this using MATLAB. First the image is converted into text and then the text is converted to speech i.e. voice using Matlab code. In the second process of converting speech to text, speech recognition plays a major role. Matlab code is implemented to convert speech to text.

II. LITERATURE SURVEY

In the first paper, improved Linear Predictive Coding (LPC) coefficients of the frame are employed in the feature extraction method. In the proposed speech recognition system, the static LPC coefficients + dynamic LPC coefficients of the frame were employed as a basic featurE. The framework of Linear Discriminant Analysis (LDA) is used to derive an efficient and reduced-dimension speech parametric speech vector space for the speech recognition system. There are different approaches to speech recognition like Hidden Markov Model *(*HMM*)*, Dynamic Time Warping *(*DTW*)*, Vector Quantization *(*VQ*)*, etc. The second paper provides a comprehensive study of use of Artificial Neural Networks *(*ANN*)* in speech recognition. The paper focuses on the different neural network related methods that can be used for speech recognition and compares their advantages and disadvantages. The conclusion is given on the most suitable method.

The communication among human computer interaction is called human computer interface. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech to text conversion and also gives overview technique developed in each stage of classification of speech to text conversion. A comparative study of different technique is done as per stages.

Three algorithms have been proposed to convert an image to sound in the next paper,
1.Uses time-multiplexed mapping
2.Mapping of each pixel into a 3D point
3. Analyze the image with the Fast Fourier Transform

Hand gesture is recognized by capturing the image and send it to an proposed algorithm which has four steps segmentation, orientation detection, feature extraction and classification. A blind person can get information about the shape of an image through speech signal by using algorithm that uses edge detection, sound production and analysis and sound output. This algorithm has been described in sixth paper.

To convert image to text it combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS) in Raspberry pi. This device consists of two modules, image processing module and voice processing module.

Many techniques have been developed for the face recognition but in our work we just discussed two prevalent techniques PCA (Principal component analysis) and LDA (Linear Discriminant Analysis) and others in brief. These techniques mostly used in face recognition. PCA based on the eigenfaces or we can say reduce dimension by using covariance matrix and LDA based on linear Discriminant or scatter matrix. In our work we also compared the PCA and LDA.

The tenth paper presents a systematic procedure using cepstrum and linear predictive coding (LPC) based statistical approach for exploring the singers rendition for a set of song of poet Rabindra Nath Tagore sung by different renowned singers. The medium category (Mudara) of songs started with vowel `aa' has been incorporated in the data set of song. Index of perceived vocal qualities like flow phonation, depth, vowel quality and timbre have been estimated numerically based on cepstral and LPC coefficients.

III. METHODOLOGY

1. Image to Speech conersion:

In this process first text of acquired hand gesture is generated and then corresponding speech is displayed as output.
PCA and LDA methods are used for hand gesture recognition. PCA based on the eigen gestures or we can reduce dimension by using covariance matrix and LDA based on linear discriminant or scatter matrix.

PCA (Principal Component Analysis)

PCA is a dimensionality reduction technique which is used for compression and recognition problems. PCA can be done by eigen value decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering the data matrix for each attribute. The main goal of PCA is the dimensionality reduction, therefore the eigenvectors of the covariance matrix should be found in order to recognize the gestures.

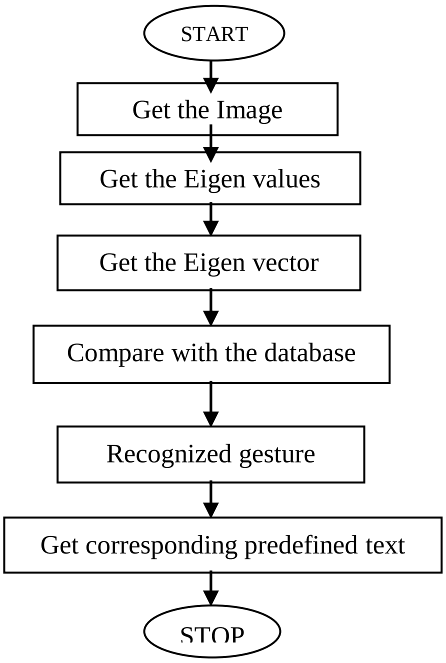


Fig 1 : Recognition Phase

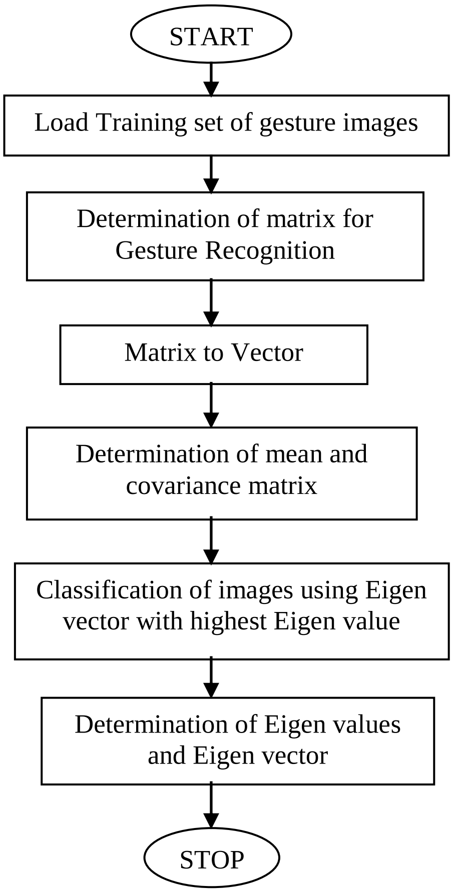

START
Load Training set of gesture images
Determination of matrix for Gesture Recognition
Matrix to Vector
Determination of mean and covariance matrix
Classification of images using Eigen vector with highest Eigen value
Determination of Eigen values and Eigen vector
STOP

Fig 2 : Learning Phase

Steps involved in PCA:

1) Create a training set and load it. Training set consist of total M images and each image is of N*N.

2) Convert the gesture images in the training set to gesture vector .we denoted it by Xi.

3) Normalize the gesture vector.

a) Calculate average gesture vector.

b) Subtract average gesture vector from each gesture vector.

Normalized gesture vector is,

$$\Phi i = Xi - \psi$$

Where ψ= Common features of the image

4) Find eigen gestures with the help of covariance matrix C.

$C = AA^T$

here A= {Φ1, Φ2,…… Φm} these are normalized gesture vectors. $A=N^2*M$

Because dimensions of M images are N*N

$C=AA^T$

$C=N^2*M \quad M*N^2$

C will become N^2*N^2 dimensions

5) Calculate eigenvectors from a covariance matrix with reduced dimensionality.

Here we use formula $C=A^TA$ (this covariance with the reduced dimensionality)

$C= A^T A$

$C=M*N^2 \quad M*N^2$

Here C will be of M*M dimensions

6) Select K best eigen gestures, such that K<M and can represent the whole training set.

7) Convert lower dimensional K eigenvectors into gesture dimensionality

Ui = AVi

Here Ui= ith vector in higher dimension space

Vi = ith vector in lower space dimension.

8) Image can be represented as a weighted sum of K eigengestures + mean or average gesture.

$$\Omega i = \begin{pmatrix} W1 \\ W2 \\ . \\ Wku \end{pmatrix}$$

Weighted vector Ωi is the eigengesture representation of the ith gesture. Weight vector for each is calculated.

9) Recognition using PCA.

LDA (LINEAR DISCRIMINANT ANALYSIS)

LDA (Linear Discriminant Analysis) is enhancement of PCA .LDA use the concept of class. Gesture images of same person is treated as of same class here. LDA is also perform dimensionality reduction. They transforms images as a vector to new space with new axes. The projection axes chosen by PCA might not provide good discrimination power. LDA tries to find projection axes, such as classes are best separated. Another name of LDA is fisher's discriminant analysis and it searches those vectors in the underlying space that are the best discriminant among classes. The goal of LDA is to maximize the between-class scatter matrix measure and minimizing the within class scatter matrix measure. LDA is a derived form of Fisher linear classifier it maximizes the ratio of the between- and within-class scatters. There are some problems in LDA one is small size problem the number of training samples is less than the sample's dimensionality so the within-class scatter matrix is singular and the Linear Discriminant Analysis (LDA) method cannot be applied directly. To remove this problem there are certain methods one of them is fishergesture it combines the both PCA and LDA to make a within class scatter matrix non singular.

There are 5 general steps for performing a LDA

1. Compute the *d*-dimensional mean vectors for the different classes from the dataset.

2. Compute the scatter matrices (between-class and within-class scatter matrix). For all samples of all classes the between class scatter matrix SB and the within-class scatter matrix SW are defined by:

$$S_B = \sum_{i=1}^{c} M_{i.}(x_i - \mu).(x_i - \mu)^T$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i).(x_k - \mu_i)^T$$

where Mi = number of training samples in class i,

c =number of distinct classes

μi is the mean vector of samples belonging to class i

Xi represents the set of samples belonging to class i with x_k being the k-th image of that class

SW = scatter of features around the mean of each gesture class

SB = scatter of features around the overall mean for all gesture classes

3. Compute the eigenvectors (e1, e2, ..., ed) and corresponding eigenvalues (λ1, λ2, ..., λd) for the scatter matrices.The goal is to maximize SB while minimizing SW, in other words, maximize the ratio det|SB|/det|Sw|.

4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $d \times k$-dimensional matrix W .

5.Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation: y = WT × x

where x =$d \times 1$-dimensional vector representing one sample

y =transformed $k \times 1$-dimensional sample in the new subspace.

In text to speech conversion the four functions are used i.e. text, pace, voice, sampling frequency. Input the text with pace, sampling frequency. The SAPI.SPvoice server is accessed. Invoke the voices from SAPI server, the input text(string) is compared with the SAPI and the speech which are matched with the input text(string) is invoked. The speech corresponding to the input text is obtained.

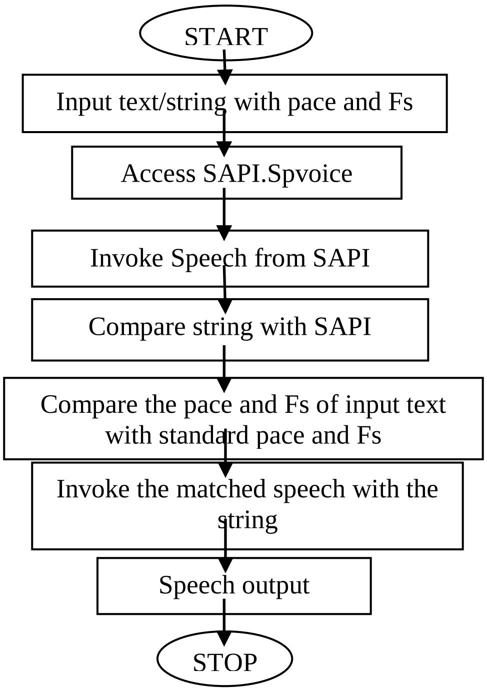

START
Input text/string with pace and Fs
Access SAPI.Spvoice
Invoke Speech from SAPI
Compare string with SAPI
Compare the pace and Fs of input text with standard pace and Fs
Invoke the matched speech with the string
Speech output
STOP

Fig 3 : Text to speech conversion

2. Speech to Text conversion:

There are two phases involved in Speech to Text conversion,

1. Training phase

2. Recognition phase

Speech samples are trained using LPC and artificial neural network. LPC is a speech compression technique that models the process of speech production. Speech compression is a method for reducing the amount of information needed to represent a speech signal. It is a digital method for encoding an analog signal in which a particular value is predicted by a linear function of the past values of the signal. The general algorithm for LPC involves an analysis or encoding part and a synthesis or decoding part. An artificial neural network(ANN) is a computer program, which attempt to emulate the biological functions of the human brain. It comprises of an input layer, one or more hidden layers and one output layer. Artificial Neurons are the basic unit of Artificial Neural Network which simulates the four basic function of biological neuron. It is a mathematical function conceived as a model of natural neuron.

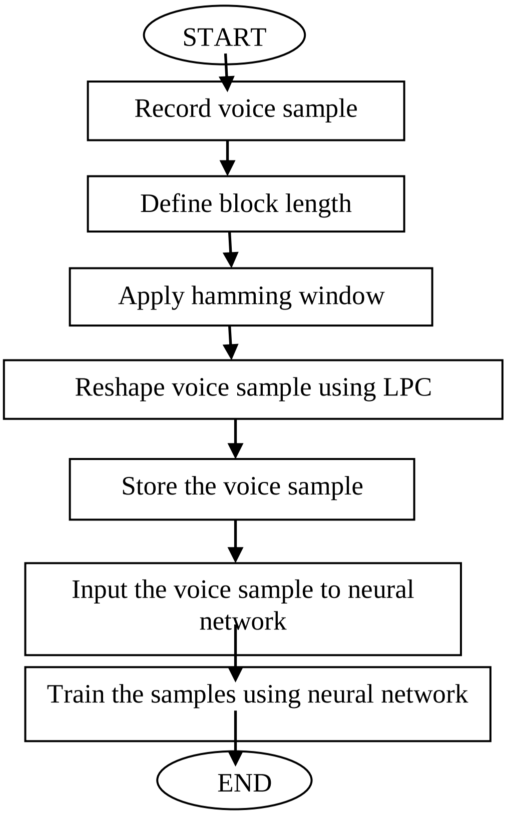

START
Record voice sample
Define block length
Apply hamming window
Reshape voice sample using LPC
Store the voice sample
Input the voice sample to neural network
Train the samples using neural network
END

Fig 4 : Training phase

In Recognition phase, first the input speech is obtained. The obtained input speech is compared with the trained database. When the match is found corresponding text is diplayed.

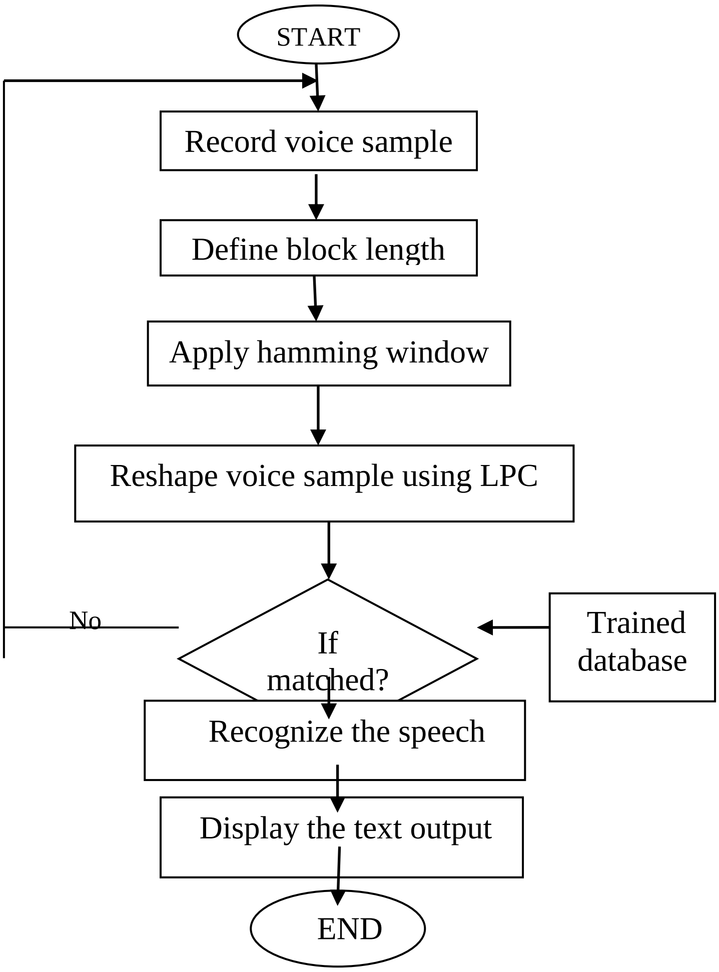


Fig 5 : Recognition phase

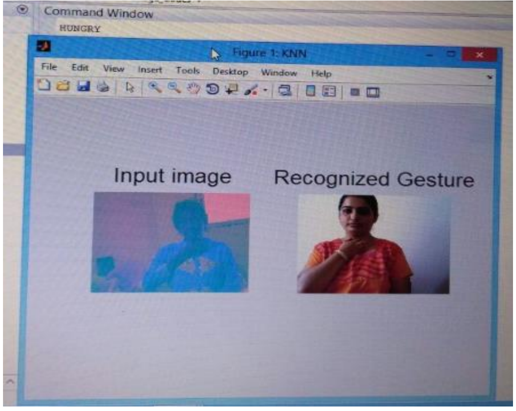
IV. RESULTS

1. Image to Speech conersion:

The database is trained for five hand gestures, they are,

1. Name
2. Understand
3. Sorry
4. Thank you
5. Hungry

The output for Image to speech conversion for the hand gesture 'Hungry' is shown below,



```
HUNGRY
```

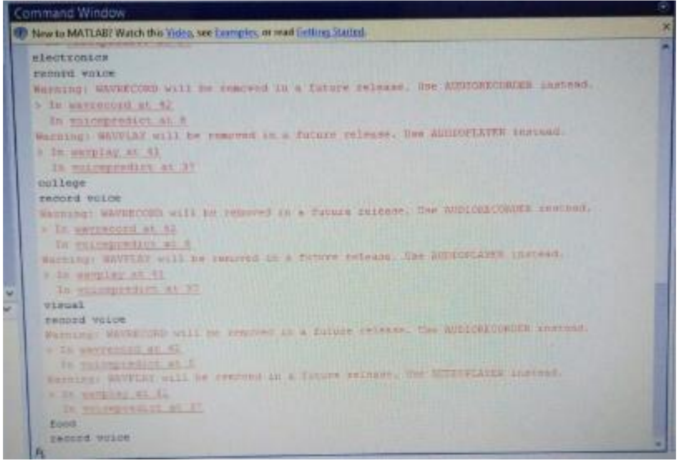
Fig 6 : Image to Text output

2. Speech to Text conversion:

The database is trained for five samples in the same order as follows,

1. Food
2. Water
3. Electronics
4. College
5. Visual

The output for all these words are shown below in the figure,


Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

```
electronics
record voice
Warning: WAVRECORD will be removed in a future release. Use AUDIORECORDER instead.
> In wavrecord at 42
  In voicepredict at 8
Warning: WAVPLAY will be removed in a future release. Use AUDIOPLAYER instead.
> In wavplay at 41
  In voicepredict at 37
college
record voice
Warning: WAVRECORD will be removed in a future release. Use AUDIORECORDER instead.
> In wavrecord at 42
  In voicepredict at 8
Warning: WAVPLAY will be removed in a future release. Use AUDIOPLAYER instead.
> In wavplay at 41
  In voicepredict at 37
visual
record voice
Warning: WAVRECORD will be removed in a future release. Use AUDIORECORDER instead.
> In wavrecord at 42
  In voicepredict at 8
Warning: WAVPLAY will be removed in a future release. Use AUDIOPLAYER instead.
> In wavplay at 41
  In voicepredict at 37
food
record voice
```

Fig 7 : Speech to text conversion output

V. ADVANTAGES AND DISADVANTAGES

Advantages:

- PCA and LDA have good acceptability, low cost.
- PCA and LDA methods have contactless acquisition.
- ANN have the ability to learn how to do task based on the data given for training, learning and initial experience.
- Computations of ANN can be carried out in parallel. ANN can be used in pattern recognition which is a powerful technique for harnessing the data and generalizing about it.
- ANN are flexible in changing environments.
- More accurate result will be obtained if database is trained for more samples in both cases.
- ANN can build informative model when conventional model fails. They can handle very complex interactions.
- LPC is generally used for speech analysis and resynthesis.
- LPC synthesis can be used to construct vocoders where musical instruments are used as excitation signal to the time-varying filter estimated from a singer's speech.

Diadvantages:

- Requires more time and samples to train the database.
- Problem solving methodology of many ANN system is not described.
- PCA and LDA methods are very sensitive to data acquisition conditions(illumination and pose).

VI. APPLICATIONS

Speech to text converter:

- Aid to Vocally Handicapped
- Source of Learning for Visually Impaired
- Games and Education
- Telecommunication and Multimedia
- Man-Machine Communication
- Voice Enabled E-mail

Image to speech converter:

- Aid to visually handicapped
- PCA method is applied to share portfolios.
- PCA method used in neuroscience to identify specific
  properties of a stimulus.

CONCLUSION

The main aim of the project is to reduce the communication gap between speech or visually imapaired people and normal people. The system is proposed to improve lifestyle of deaf-dumb people. This algorithm is also favourable for degrading the communication difference between blind and deaf-dumb people. The algorithm proposed is effective and efficient since it is been implemented using software i.e. MATLAB. The algorithm is simple and cost effective.

FUTURE SCOPE

This project can be a help (great use) for the communication between visual and speech impaired people. By increasing the efficiency and adding some more features, it can be implemented in the public places like railway station, metro station etc. This can be made user friendly by incorporating the features like sending request and queries to the admin and ensure that the expectation of the users are met.

REFERENCES

1. "Improved Linear Predictive Coding Method for Speech Recognition", Jiang Hai, Er Meng Joo,2003.
2. Speech Recognition Using Artificial Neural Network –A Review", Bhushan C. Kamble, 2016.
3. A REVIEW ON SPEECH TO TEXT CONVERSION METHODS ",Miss.Prachi Khilari, Prof. bhope V. P,2015.
4. "Algorithms and technique for image to sound conversion for helping the visually impaired people", Alexander Caan, Radu Varbanescu, Dan Popescu,2007.
5. “Hand gesture recognition for human-computer interaction” , Meenakshi panwar, Pawan singh mehra,2011.
6. “Image recognition for visually impaired people by sound” , K. Gopala Krishnan, C. M. Porkodi, K. Kanimohi,2013.
7. “Sign Language Recognition System” , Aditi Kalsh, Dr. N.S. Garewal,2013
8. "Image Text to Speech Conversion Using OCR Technique in Raspberry Pi" , K Nirmala Kumari, Meghana Reddy J ,2016.
9. "Face Recognition Using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) Techniques", Amritpal Kaur, Sarabjit Singh, Taqdir, 2010.
10. "Cepstrum and LPC based statistical approach to explore singers' rendition on a set of songs of Tagore", Indira Chatterjee, Joy Sen, Parthasarathi Bera, 2017.
11. "Improved linear predictive coding method for speech recognition", Jiang Hai, Er Meng Joo,2004.
12. "Speech recognition using artificial neural networks", C. P.Lim, S.C. Woo, A.S. Loh,2002.
13. "Speech Recognition using Artificial Neural Networks", I.A. Maaly, M. El-Obaid, 2006.