International Journal of Emerging Technology and Innovative Engineering
Volume 1, Issue 8, August 2015 (ISSN: 2394 – 6598)

SPEECH RECOGNITION OF DEAF AND HARD OF HEARING PEOPLE BY USING NEURAL NETWORK

Ms. Pratibha Saroj
M.E. Computer Engg.
Thadomal Sahani Engineering College
Bandra West,Mumbai,India.
pratibha.saroj@gmail.com

Mrs.Shilpa Verma
Associate Professor in Computer Engg.
Thadomal Sahani Engineering College
Bandra West,Mumbai,India.
shilpaverma65@gmail.com

ABSTRACT

Speech recognition is the process of converting speech in the text format. Speech sounds are produced by modifying the flow of air from lungs to mouth by different articulators, i.e. tongue, jaw; lips etc., i.e. position of these articulators are responsible for generation of particular speech sounds. There are various types of speech impairment that can occur singly or in combination. Speech impairment may be any of several speech problems, particularly the Dysarthria and Aphasia. Dysarthria is difficult, poorly articulated speech, such as slurring. Aphasia is impaired expression or comprehension of written or spoken language. Speech recognition process contains three main stages for processing the speech which is acoustic processing, feature extraction and recognition. The acoustic processing obtains the sequence of input vector that will be used in next stages, feature extraction. For comparison purposes, Linear Prediction Coding and PLP coefficients are performed.In this dissertation, single neural network i.e. Backpropagation is used. In this the speech recognition is done for speaker independent and of isolated words spoken by deaf and hard of hearing people of 5 to 16 years of age. The recognition is done for total 17 isolated words.

Keywords

BPN, Automatic Speech Recognition, LPC, PLP, MATLAB.

1. INTRODUCTION

Speech recognition is the process of converting speech in the text format. Speech sounds are produced by modifying the flow of air from lungs to mouth by different articulators, i.e. tongue, jaw; lips etc., i.e. position of these articulators are responsible for generation of particular speech sounds [1].

There are various types of speech impairment that can occur singly or in combination. Speech impairment may be any of several speech problems, particularly the following: Dysarthria is difficult, poorly articulated speech, such as slurring. Aphasia is impaired expression or comprehension of written or spoken language. Generally, speech recognition process contains three main stages for processing the speech which is acoustic processing, feature extraction and recognition..

2. PRAPOSED SYSTEM

The use and teaching of sign-language is generally used to communicate with deaf persons. But the drawback is that everybody could not understand the sign-language and the lack of tone also makes it of very little help. The best solution that science could offer is a system which can recognize voice when an impaired person speaks which can be converted in to a text message. Project can be used in deaf people school where they can use application to communicate with them to understand there point of view more clearly instead of using the sign language.

In the proposed work the neural network will be used which consist BPN network to recognize the speech. The PLP and LPCC cepstral coefficient techniques are used to extract the feature from the speech signal. The important part of any speech recognition system is to extract features from the speech which should not change for the same words and should not change with time or be unaffected by the speaker's health. The speech signal is first divided into frames and weighted by a window, which is usually a Hamming window. After windowing, DTFT is used to convert speech frame to its frequency domain. Critical band for a given center frequency is defined to be the smallest band of frequencies around which it activates the same part of the basilar membrane of the ear. After finding the power spectrum P (w) it is warped along its frequency axis w into the Bark frequency [4].The Bark scale is used for PLP technique but not for the LPCC. Both the techniques are used differently to extract features the PLP techniques uses the twelve input features and LPCC uses the eleven input features as input to the neural network to classify the words. Using the NNtool of MATLAB the backpropagation train system is used function “trainlm” .The total dataset is of 282 from which 200 is trained an 82 is for testing purpose.

3. DESIGN CONSIDERATION

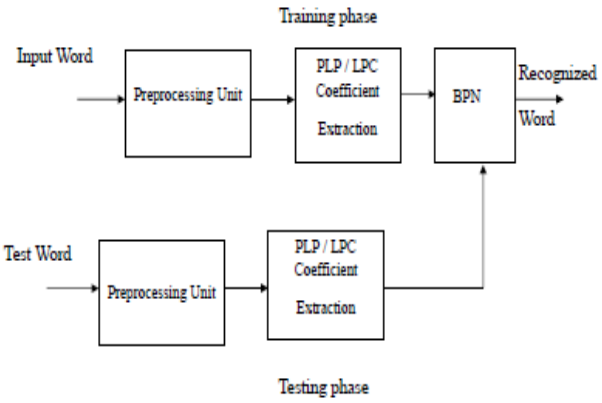

Fig1 Block Diagram of Proposed Approach

3.1 Preprocessing Unit

A Butterworth filter is one of several common infinite impulse response (IIR) filters. Other filters in this group include Bessel and Chebyshev filters. In addition, these filters are classified as feedback filters. A high pass filter is a filter which allows the high-frequency energy to pass through. It is thus used to remove low-frequency energy from a signal. A low pass filter is a filter which allows the low-frequency energy to pass through. It is thus used to remove high-frequency energy from a signal. A band pass filter may be constructed by using a high pass filter and low pass filter in series.

It covers mainly these tasks-

1. A/D conversion, Background noise filtering, Pre emphasis, Blocking and Windowing.
2. A typical representation of a speech signal is a stream of 8-bit numbers at the rate of 10,000 numbers per second clearly a large amount of data.
3. Once signal conversion is complete, background noise is filtered to keep SNR high say greater than 40 dB. The next step is the pre-emphasis.

The motivation behind it is to emphasize the important frequency components in the signal (i.e. amplify important areas of the spectrum) by spectrally flatten the signal.

3.2 Feature Extraction

In this part the details of extracting the features per frame of the speech signal are discussed. As this was not the main part of the project hence in depth analysis of the same was not delved into. The goal of feature extraction is to represent speech signal by a finite number of measures of the signal. This is because the entirety of the information in the acoustic signal is too much to process, and not all of the information is relevant for specific tasks. In present Speech Recognition systems, the approach of feature extraction has generally been to find a representation that is relatively stable for different examples of the same speech sound, despite differences in the speaker or various environmental characteristics, while keeping the part that represents the message in the speech signal relatively intact.

It should be possible to recognize speech directly from the digitized waveform. However, because of the large variability of the speech signal, it is better to perform some feature extraction that would reduce that variability. Particularly, eliminating various source of information, such as whether the sound is voiced or unvoiced and, if voiced, it eliminates the effect of the periodicity or pitch, amplitude of excitation signal and fundamental frequency etc.

The reason for computing the short-term spectrum is that the cochlea of the human ear performs a quasi-frequency analysis. The analysis in the cochlea takes place on a nonlinear frequency scale (known as the Bark scale or the mel scale). This scale is approximately linear up to about 1000 Hz and is approximately logarithmic thereafter. So, in the feature extraction, it is very common to perform a frequency warping of the frequency axis after the spectral computation.

This section is a summary of feature extraction techniques that are in use today, or that may be useful in the future, especially in the speech recognition area.

3.2.1 Perceptual Linear Prediction (PLP) Coefficients Extraction

Perceptual linear prediction, similar to LPC analysis, is based on the short-term spectrum of speech. In contrast to pure linear predictive analysis of speech, perceptual linear prediction (PLP) modifies the short-term spectrum of the speech by several psychophysically based transformations. The PLP cepstral coefficients are computed using the PLP functions defined in the analysis library.

3.2.2 Linear predictive coding (LPC)

LPC is one of the most powerful speech analysis techniques and is a useful method for encoding quality speech at a low bit rate. The basic idea behind linear predictive analysis is that a specific speech sample at the current time can be approximated as a linear combination of past speech samples. LPC is one of the most powerful signal analysis methods for linear prediction. It is predominant technique for determining the basic parameters of speech and provides precise estimation of speech parameters and computational model of speech. Speech sample can be approximated as a linear combination of past speech samples is the basic idea behind LPC. Following figure shows the steps involved in LPC feature extraction.

3.2.3 Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process

information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well [13].

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

3.2.3.1 Backpropagation Network

Backpropagation, an abbreviation for "backward propagation of errors", is a common method of training artificial neural networks used in conjunction with an optimization method such as gradient descent. The method calculates the gradient of a loss function with respects to all the weights in the network. The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function [11].

Backpropagation requires a known, desired output for each input value in order to calculate the loss function gradient. It is therefore usually considered to be a supervised learning method, although it is also used in some unsupervised networks such as auto encoders. It is a generalization of the delta rule to multi-layered feedforward networks, made possible by using the chain rule to iteratively compute gradients for each layer. Backpropagation requires that the activation function used by the artificial neurons (or "nodes") be differentiable.

The drawback of competitive learning is that some of the weight vectors that are initialized randomly may be far away from any input vector and it never gets updated. In order to avoid this, weights of both the winning neuron and the neighborhood neurons are updated.

Back propagation (BP) often gets stuck up at a local minimum mainly because of the random initialization of weights. In practice, the types of optimization algorithms that are used to optimize the weights are gradient descent, conjugate gradients.

The transfer functions mostly used for hidden layers are tan sigmoid whose output values are in-between 0 and 1 and for the output layer linear transfer function is used. In the proposed system three hidden layers are used with input and output layer.

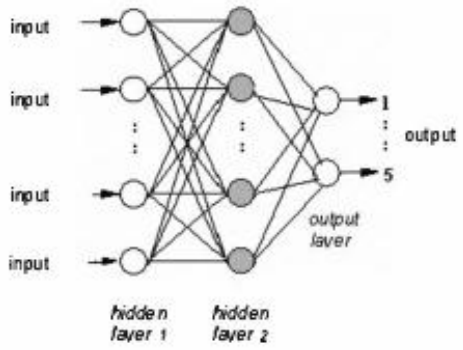

Fig 2 BPN Network

4. IMPLEMENTATION

The implementation has done of the system on platform of MATLAB 2013. First the audio signal is given as an input to the system, which is the .wav sound file of the deaf people which get converted into the feature vector format using RASTAPLP and LPC. RASTA-PLP, an acronym of Relative Spectral Transform Perceptual Linear Prediction and LPC is Linear Predictive Coefficient. PLP is a way of warping spectra to minimize the differences between speakers while preserving the important speech information. RASTA is a separate technique that applies a band-pass filter to the energy in each frequency sub band in order to smooth over short-term noise variations and to remove any constant offset resulting from static spectral coloration in the speech channel e.g. from a telephone line. Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

4.1 Speech Database

The database of deaf people speech is collected at the real time basis from two schools of deaf and dumb children. The two schools *R.S.M. SADAFULI DEAF SCHOOL SADHNA VIDYALAYA FOR DEAF* Almost 488 speech sample were collected from both the schools from 60 students for 17 isolated words each words have different numbers sample each words were spoken by different number of students whereas students were about 5 to 16 years old of age.

The table given below gives the complete idea of selected Database.

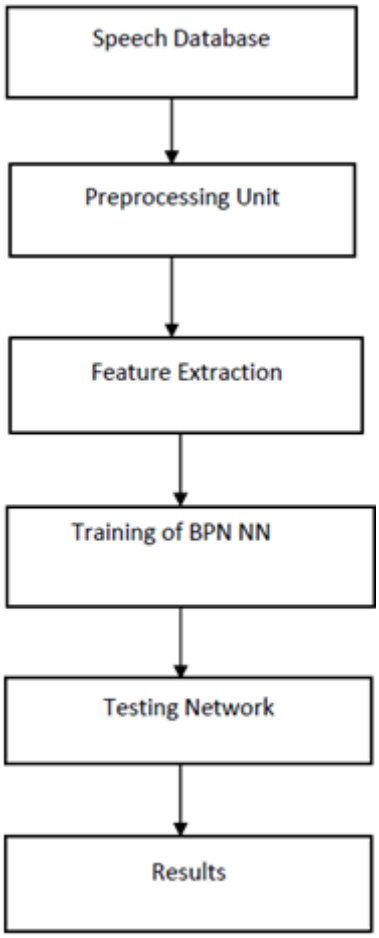

Fig 3 Implementation Flow Diagram

The total sample of 282 is used for experiment from that 200 samples are used for training purpose and 82 are used for testing purpose

Table 1. Table captions should be placed above the table

Sr.no.	Word	Number of sample
1	Bag	13
2	Ball	15
3	Bat	18
4	Book	13
5	Bye	19
6	Cat	13
7	Doll	15
8	Fan	14
9	Hello	16
10	Hi	17
11	Mummy	16
12	Phone	19
13	Namaste	13
14	Rat	9
15	Rose	12
16	School	19
17	Watch	23

4.2 Preprocessing Of Speech

Pre-processing of a signal can be said as applying any required form of processing to the signal in time domain before the feature extraction phase. Normally, in the pre-processing stage the speech signal undergoes several common processes including analog to digital (A/D) conversion, enhancement, pre-emphasis filtering and usually for SR applications silence removal or end point detection (EPD). The A/D process converts a sound pressure wave into its digital form. There are three steps in the A/D conversion process which is sampling, quantization and coding. The final product of this process is a digital version of the speech signal that can be processed by a computer [2].

4.3 Feature Extraction

This project comprises to use PLP and LPCC feature extraction techniques for both the techniques same numbers of words and samples are used. In PLP analysis, the all-pole modeling is applied to an auditory spectrum derived by (a) convolving with a critical band masking pattern, followed By (b) resampling the critical band spectrum at approximately Bark intervals, (c) pre-emphasis by a simulated fixed equal loudness curve, and finally (d) compression of the resampled and pre-emphasized spectrum through the cubic root Non-linearity, simulating the intensity-loudness power law. The low order all-pole model of such an auditory spectrum has been found to be consistent with several phenomena observed in speech perception. PLP cepstral coefficients are almost same for all the words though after normalization calculation of min, mode, phase, magnitude, skewness ,kurtosis etc. all feature were giving same values which is difficult to get

classified. Total 12 features are used. Therefor second techniques i.e. LPCC is come under account after applying LPCC, mostly it gives 12 to 14 cepstral feature from that 11 are considered. On that basis different parameters are calculated.

4.4 Training Of Network

As now the system has feature vectors for all the files and network is also ready so now the time is to train the network. For training Backpropagation algorithm is used that is 'trainlm'function of matlab. For the propose system the maximum numbers of epochs is 400 and max-fails time is 20.The system takes about 0.7msec to train one file. The training function is set as trainlm and first is to combine the entire feature vector to form an input matrix of 11x200.The input weight is 0.1 by default the learning rate is 0.01, the feedforward net is created using Matlab 2013 NNTool with all these specification and input matrix as input and target variable is set as target which is again a matrix of 17x200 dimension.

4.5 Testing Of Network

Feature vectors of the test speech are given to the trained models. The neural network model is simulated for the given input. The word is identified based on the output. First the file get selected from the test folder on the basis of file number the particular feature vector got selected and then applied to the feedforward network the resultant output send to the backpropagation network. Finally the output variable is generated of 17 values from those 17 values the value which is the highest one got selected and gets rounded off that value, if 7.9876 is the value than it will be round off to 8 and this value tells that to which class the test file belongs. The message box will display showing the word which was selected for the test, if word does not belong to any of the class it will show incorrect typed in message box.Testing of the network is done on the untrained data and their features are store in the variable called samfea which consist of 11x82 dimensional matrix. To find overall performance of testing of network is given by the test confusion matrix.

5. RESULT

It's clear that for some of the words the accuracy is very poor and for some words it is very high this because the samples for that particular words was not up to the mark the words which are same sounding like bat and bag, this kinds words are sounds exactly same from the deaf persons voice. The words are not very identical so their feature toggles between both the classes and results in wrong classification.

The system which uses feature extraction of PLP has the accuracy of 56% of correct recognition and 44% of wrong classification. The system with LPC features has quite good accuracy of almost 70% and percentage of wrong classification is 30%.The accuracy of each word is in given table.

Table 2 Accuracy of each isolated words

Sr. No.	Word	Accuracy with LPC	Accuracy With PLP
1	Bag	60%	50%
2	Ball	80%	0%
3	Bat	60%	40%
4	Book	66.7%	75%
5	Bye	50%	44%
6	Cat	100%	50%
7	Doll	100%	50%
8	Fan	66.7%	0%
9	Hello	100%	21%
10	Hi	50%	75%
11	Mummy	80%	100%
12	Phone	66.7%	80%
13	Namaste	100%	100%
14	Rat	100%	50%
15	Rose	100%	100%
16	School	55.6%	75%
17	Watch	75%	100%

According to the table the accuracy chart is given for both the techniques for each and every word which says that for some of the words the PLP is good and for some LPC.

6. CONCLUSION

Even though the deaf speech has large variation in pronunciation, recognition system has produced encouraging results. The results can be improved by varying the Back propagation neural network and it can be extended for more number of words. The implementation of this system in hardware may be useful for the deaf persons to communicate with others and their speech will be easily understood by others.

In order to increase the results the hardware configuration of the system can be improved so the number of data to be train will increase and can give better result.

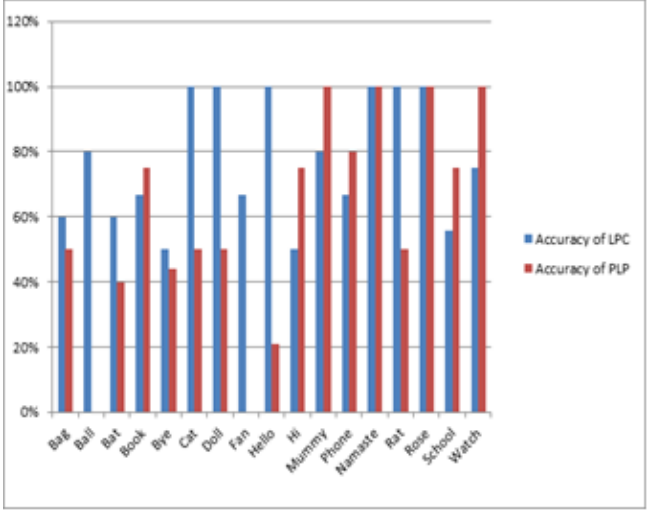

Fig 4 Accuracy Chart

7. REFERENCES

[1] C.JeyalakshmL, Dr.KrishnamurthL V, and Dr.A.Revathi,"Speech Recognition of Deaf and Hard of Hearing People Using Hybrid Neural Network,"2010 2nd International Conference on Mechanical and Electronics Engineering (ICMEE 2010).

[2] T.B. Adam and Md Salam, "Spoken English Alphabet Recognition with Mel Frequency Cepstral Coefficients and Back Propagation Neural Networks," International Journal of Computer Applications (0975 – 8887) Volume 42– No.12, March 2012.

[3] R.L.K.Venkateswarlu, R.VasanthaKumari and, A.K.V.Nagayya "Efficient Classification of Based Speech Recognition System" Evolution in Networks and Computer Communications A Special Issue from IJCA - www.ijcaonline.org.

[4] Ananthi. S, Dhanalakhsmi. P, "Survey about Speech Recognition and Its Usage for Impaired (Disabled) Persons," International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013 ISSN 2229-5518.

[5] Dr. R.L.K.Venkateswarlu, Dr. R. VasanthaKumari, and A.K.V.Nagayya"Novel Approach for Speech Recognition by Using Self-Organized Maps" International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, August 2011.

[6] E. Hosseini Aria, J. Amini, M.R.Saradjian "Back Propagation Neural Network For Classification of IRS-1D Satellite Images" Department of geomantics, Faculty of Engineering, Tehran University, Iran National Cartographic Center (NCC), Tehran, Iranector, A. Z. 1989.

[7] Navdeep Kaur, Sanjay Kumar Singh "Data Optimization in Speech Recognition Using Data Mining Concepts and ANN" International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4283 – 4286. 61.

[8] Joel-Ahmed M. Mondo, Satyanarayan Panigrahi, and Madan M. Gupta "Neural Networks Approach to Biocomposites Processing" 978-1-4577-0253-2/11/$26.00 ©2011 IEEE.

[9] Bilal Esmael1, Arghad Arnaout, Rudolf K. Fruhwirth and Gerhard Thonhauser "A Statistical Feature-Based Approach for Operations Recognition in Drilling Time Series" International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 5 (2013) pp. 454- 461 © MIR Labs, www.mirlabs.net/ijcisim/index.html.

[10] Siddhant C. Joshi, Dr. A.N.Cheeran "MATLAB Based Back-Propagation Neural Network for Automatic Speech Recognition" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 3, Issue 7, July 2014.

[11] Milind U. Nemade, Prof. Satish K. Shah "Improvement in Speech Recognition Performance using Beamforming based Speech Enhancement" International Journal of Electronics Communication and Computer Engineering Volume 3, Issue 4, ISSN.

[12] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov "Neural Networks used For Speech Recognition" International journal of automatic control, university of Belgrade, vol. 20:1-7, 2010©.

[13] Yixiong Pan, Peipei Shen, Liping Shen" Feature Extraction and Selection in Speech Emotion Recognition" Supported by national Natural Science Foundation of China under Grant No. 60873132.

[14] Bhupinder Singh, Neha Kapur, Puneet Kaur "Speech Recognition with Hidden Markov Model: A Review" International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 3, March 2012.

[15] Nidhi Desai, Prof. Kinnal Dhameliya, Prof. Vijayendra Desai" Feature Extraction and Classification Techniques for Speech Recognition: A Review"International 62 Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com ISSN 2250-2459, ISO 9001:2008 \ Certified Journal, Volume 3, Issue 12, December 2013.

[16] Zulkhairi Md. Yusof, Mohiuddin Ahmed"Malay Phoneme Classification using Perceptual Linear Prediction (PLP) Algorithm".