IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2016

Control of a Wireless Robot Using Speech Recognition in Punjabi Language

Dr. Pankaj Mohindru[1], Suraj Kumar[2], Dr. Pooja[3]

Assistant Professor, Department of Electronics, Punjabi University, Patiala, India [1, 3]

M.Tech Scholar, Department, of Electronics, Punjabi university, Patiala, India [2]

Abstract: Speech recognition and analysis have made great progress, over the last few years. Speech recognition technology make possible creating voice controlled robots. In this paper, a speech recognition system is developed by using Mel Frequency Cepstrum Coefficient (MFCC), Vector Quantization (VQ) and Artificial Neural Network (ANN) to control of a wireless robot by using regional language (Punjabi). The commands that used here are all in Punjabi language. The voice signal for both male and female are recorded in .Wav file at 16 kHz sampling rate and then modified. The features of collected voice signal are extracted by using MFCC and VQ's. Back propagation method is used to train artificial neural network. The training data include the pronunciation of six words used as the command. These commands are created from 25 people including both male and female. 750 data training is used to train artificial neural network.

Keywords: Robot, Voice Commands, Speech Recognition, MFCC, VQ's and Artificial Neural Network.

I. INTRODUCTION

Voice recognition allows human to interact with computer or robot through voice. People are used to communicate with regional or natural language. In Human speech there are certain types of unique parameters, these unique parameters are used to identify a person.

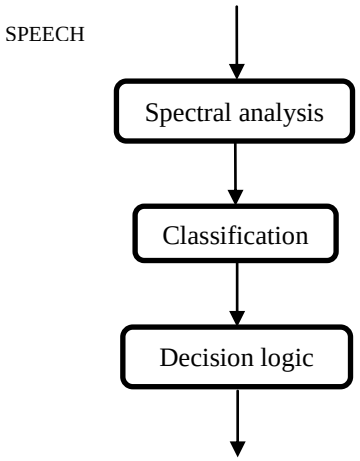
Speech recognition is a technique to recognize speaker on the basis of words spoken by the speaker. The speech recognition has mainly two task speaker identification and speaker matching.

Speaker identification also has two types' text dependent and text independent. Text dependent system means identification of person or speaker from collected database. The identification has two common task feature extraction for extracting the features of voice signal and feature matching for identifying the speaker by matching the extracted signal with stored database.

Collected voice data is extracted by using MFCC algorithm and to reduce the amount of collected voice data vector quantization is used. MFCC gives logarithmic value of features which resembles human accusatory system.

II. SPEECH RECOGNITION PROCESS

The above fig2.1 shows the structure of speech recognition process in which firstly speech signal is processed and after that features is extracted from the input speech signal and second step is to apply classification techniques on extracted features of speech.
The commonly used speech classification techniques are discussed as following;



RECOGNIZER SPEECH

A. Spectral analysis
Spectral analysis of recognition is done by various techniques. In spectral analysis process certain unique features of speech signal is extracted by using various algorithms such as LPC, MFCC, LPCC, FFT, Autocorrelation etc.

B. Vector Quantization – The output of MFCC spectral analysis is a series of vector features of time varying raw speech data [5]. If we compare the output of MFCC spectral analysis and raw speech data, we see that MFCC spectral analysis meaningfully reduce the raw speech data rate. If we denote the spectral vectors as v_l , l=1, 2……., L and each vector has k dimension vector. Consider 8 KHz sample speech signal with 16 bit amplitude. A raw speech

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2016

signal of 128000 bps is required to store the speech samples [6]. Consider for the spectral analysis vector dimension k=10 using 100 spectral vector per second. The required storage capacity is reduced to 10 to 1 if we again represent each spectral to 16 bit component. The concept of building codebook of different analysis vectors contains more code word than basic set of acoustic vectors; this is the basic idea behind vector quantization method [7].

C. Hidden Markov Model

The hidden Markov model is a set of states linked by transitions; it begins in the initial state designated. Each discrete time step, and the transition is Taken into a new state, after which it is to create a single code that came out the state [7]. Selection Transitional phase and output are both random code, the prospect judged by distributions. The HMM it can be thought of as a black box, where the sequence of output symbols generated during the observed time, the objective of the states visited the sequence over time is hidden from sight. This is why it's called Hidden Markov Model [8]. HMMs have a variety of applications. When applied to HMM speech recognition, the interpretation states acoustic models, referring to what is likely to hear the sounds during the corresponding segments of speech; while transitions provide time constraints, indicating how states may follow each other in sequence. Because speech always goes forward in time and the changes in the application of speech always go forward (or make and self- ring, and let's have the state's two arbitrary).

D. Neural network

Neural network and HMM have many similarity, both are statistical models. Hidden Markov model uses probabilities for state transition and neural network uses strength of connections and functions. Hidden Markov model works serially and neural network work parallel [1] [2] [7]. The key of neural network is set the appropriate weights of connections and the key of HMM are to finding appropriate transition and observation probabilities. Neural network and HMM is applied together in many speech processing application. Fig 2.2 shows the classification process of the neural network [8];

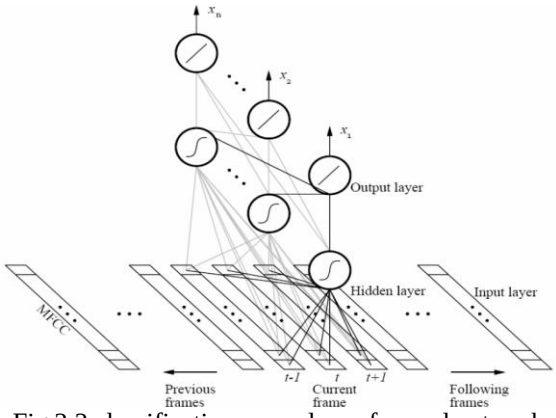


Fig 2.2 classification procedure of neural network

In this paper pattern-net neural network is used for pattern classification.

III. DATA AND RESULTS

THE RESULTS OF ALL TECHNIQUES ADOPTED FOR THIS EXPERIMENTAL WORK IS DISCUSSED AS FOLLOWING;

A. Database collection

A database of 25 persons including male and female of six voice command is recorded in Punjabi language at 16000 Hz sampling frequency in .wav file format. 'Gold wave' software is used to record the voice samples.

B.MFCC (features extraction)

Extracting of features is most important step of getting high accuracy speech recognition system. As we know that human ear percept frequency below 1000 Hz in a linear scale and in a logarithmic scale above 1 KHz. Mel frequency Cepstrum scale spacing linearly below 1000 Hz and logarithmic above 1000 Hz [4]. The speech signal has most of their energy below 1 KHz. So that MFCC is used to extract the above characteristics of human speech signal. A forward Fourier transform of spectrum of a signal is known as a cepstrum. It has many properties that make it useful in various type signal analysis. It has properties to separate the set of repeated patterns [3]. Signal presentations- speech signal is represented in different ways. The variations of amplitude in time domain are the general way to represent a signal.

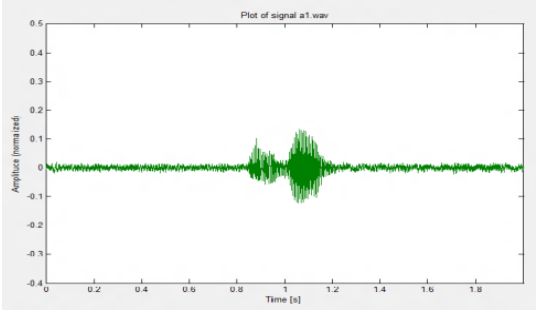


Fig 3plot of signal Agge

Filtering- The first step of MFCC is the filtering to eliminate the low frequency and high frequency noise from input signal. For this input speech signal is passed through FIR high pass filter and low pass filter. Framing and windowing-In framing filtered speech signal is converting into small overlapped frame of 20 to 40ms lengths. The speech signal is divided into frames of N samples. Typical value no. of sample in each frame N= 256 and adjacent frame M= 128, number of frame is calculated by dividing total number of samples in input speech data by 128. Windowing is the process of eliminating the discontinuity at the beginning and end of the frame in windowing each row of frame is multiplied by window function [4]. Spectrogram – Discrete Fourier transform is a simplest mathematical process to convert

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2016

signal from time domain into frequency domain. The Fast Fourier transform is used to convert each frame of N samples from time domain into frequency domain.

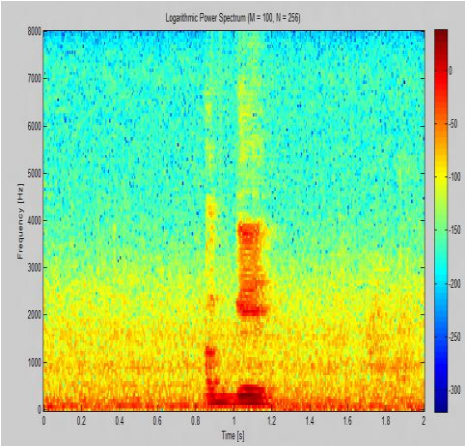


Fig .4 logarithmic power spectrum

Mel filter bank – In this previous step the modulus of Fourier transform is calculated. The magnitude of spectrum is warped accordingly into Mel scale to obtain the properties like human ear. In this process themagnitude of Fourier spectrum is segmented into number of critical bands using Mel filter bank which is typically consist of overlapping triangular band-pass filters. The purpose to apply Mel filter bank is to calculate the energy of each frame of magnitude spectrum. For this we multiply each filter bank with the magnitude spectrum [8] and then weighted sum is computed so the output of the process is according to Mel scale. After that the following equation is used to calculate Mel frequency from given frequency in hertz;

$$\text{Mel frequency} = [2595 * \log 10[1+f/700] \qquad (1)$$

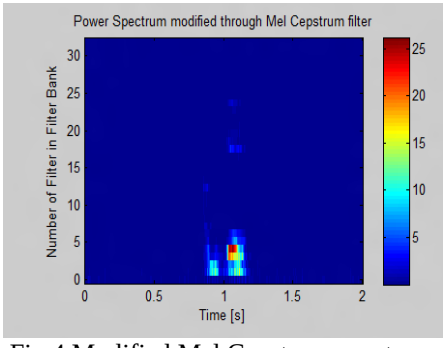


Fig.4 Modified Mel Cepstrum spectrum

Discrete cosine transform – the final step is to compute the DCT of log filter bank energies. There is two main reason to perform DCT, (1) the filter bank are all overlapping so the filter bank energies are quite correlates with each other and also the output of Mel logarithmic scale is real numbers [8]. DCT decor relates the filter bank energies and also convert log Mel spectrum into time domain. Only 12 coefficients are kept, this is because higher DCT coefficients represent fast change in the filter bank energies [7]. The set of coefficient is called acoustic vectors.

C. Neural network implementation

Pattern-net neural network is used in the experiment for word recognition. It has been tested that neural network with three hidden neural give maximum speech recognition than the network with higher number of hidden layer. Sigmoidal transfer function is used as activation function [3].

For our implementation the MATLAB neural network tool box has been used to create, train and simulates the network [10]. For every word we used 750 recorded samples. From these 70% samples used for training, while the rest 30% were used to test the network. The hidden layer consists of non-linear sigmoidal function. The amount of neurons depends on some factors like the amount of input data and output layer neurons number, the needed generalization capacity of the network and the size of the training set.

The Oja rule of thumb is applied to make a first guess on how many hidden layer neurons are required [9].

$$H = \frac{T}{5(N=M)} \qquad (3.1)$$

Where H is number of hidden layer neurons, N is the size of the output and T is the training set size.

In this experiment the input layer 12 input MFCC contained in the input layer matrix. The hidden layer contains 6 ‘tansig’ neurons. The output also has 6 linear neurons to recognize 6 words. A set of 50 samples of each word can be used as training data. The gradient is calculated for each set of training. 1000 epochs are enough to train the network.

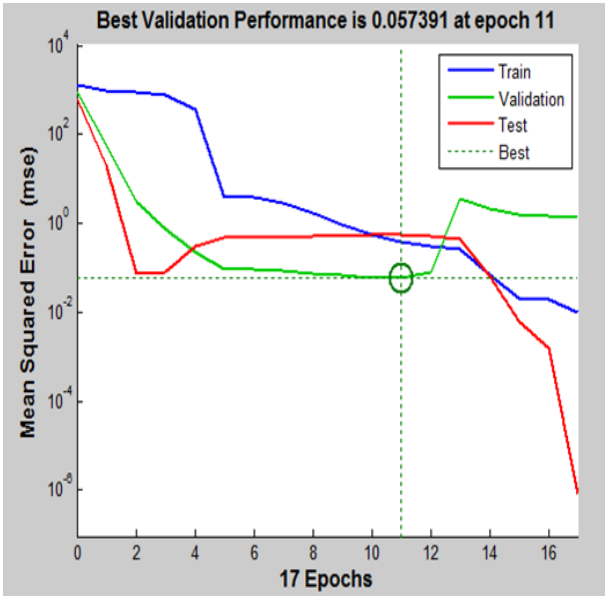


Fig.3.2 Neural network training performance graph

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2016

The performance of the network is mainly dependent on the quality of the signal processing. Fig 3.2 shows the performance graph of neural network which is show that at 11 epochs the best validation results is obtained.

The efficiencies of the isolated Punjabi words are shown in fig 3.4, and table 1, table 2. Fig 3.4 shows the overall accuracy of the speech recognition for six words. These results disclose that Pattern network is a better classifier than feed-forward network when it comes to speech recognition applications.

Its performance is best in conjugation with MFCC as the feature extraction technique. Table.1 show the accuracy of different database of voice signal and table 2 shows the accuracy of each word in percentage. From the experimental results total
96.66% accuracy of speech recognition system is obtained.

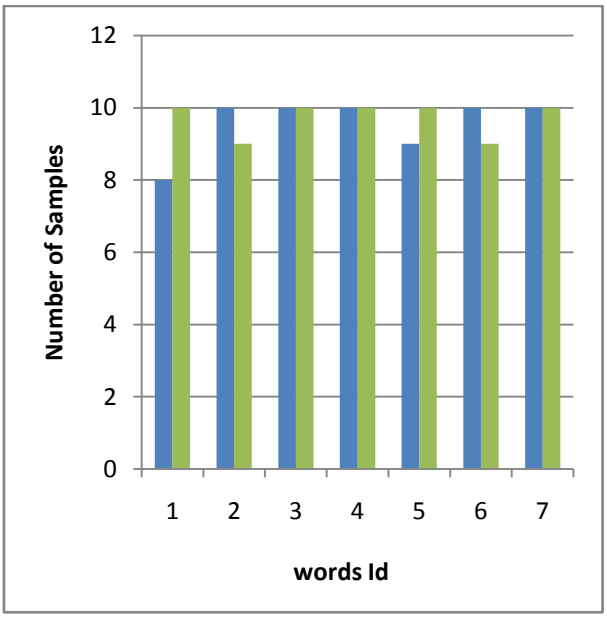


Fig 3.4 Overall recognition efficiency of Pattern network for different hidden units with MFCC feature extraction technique

Table 1.The voice command recognition accuracy Analysis in percentage

Round ID	From Database	Out of Database	Mixed Evaluation	Overall Accuracy
1	80	100	100	93.33
2	100	100	90	96.67
3	100	100	100	100
4	100	100	100	100
5	90	90	100	93.33
6	100	100	90	96.67
7	100	90	100	96.67
Total	670/7= 95.71	680/7= 97.14	680/7= 97.14	96.6%

Table 2.The voice command recognition accuracy analysis in percentage of each word

COMMAND DATA In Punjabi Language	From Database	Out of Database	Overall Accuracy
Aagge(Forward)	96.6	94.6	95.6
Kakhbe (Left)	96.6	91.6	94.1
Chal (Run)	96.1	96.3	96.7
Pishe (Backward)	95.15	95.34	95.245
Sajje (right)	97.7	94.88	95.29
Rukk (stop)	97.24	94.5	95.87
Combined	97.9	94.64	96.27
Total	97.89	95.55	96.725

D. Graphical user interface

The graphical user interface is a tool for implementation of the system where methods used in this work were integrated. C# programming language is used to create the GUI application. The GUI designed by seven basic function, named as first button ‘load’ when pressed it allows user to load the data and read the input voice sample, second function ‘plot’ when pressed it will plot the loaded signal, third function ‘linear spectrum’ when pressed gives a magnitude power spectrum, forth function ‘logarithm’ when pressed gives logarithmic power spectrum of the signal, fifth function ‘ plot different value for N’ when pressed it will plot power spectrum for different values for, sixth function ‘Mel cepstrum coefficient ’ when preesed gives plot for Mel filter bank and Modified power spectrum of signal, seven function ‘Run training ’ when pressed it will start neural network training process and give output as best matching samples from all training and testing samples.
After completion of speech recognition process the output from MATLAB GUI function is send to the receiver of robot via serial communication.

IV. CONCLUSION

In this proposed work, neural network techniques have been investigated. Pattern-net neural techniques with MFCC give features extraction gives better results than Back propagation algorithm. We have used the voice of 25 persons including male and female. It is also concluded that the performance of speech recognition is strictly depend on spectral techniques; MFCC gives better accuracy than others methods. The results from experiments showed that this proposed work is successfully implement speech recognition system that would recognize six commands words in Punjabi language. Total 96.66% accuracy of system is obtained.
The recognition accuracy of system can be increased using other recognition techniques like Neuro-fuzzy, KNN (K nearest neighbor) etc. and also multiple regional language can also be used. The number of user data can be increased.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2016

ACKNOWLEDGMENT

The authors would like to thank the concerned authority of Punjabi University Patiala, India for providing us laboratory facilities for the purpose of completing this research work.

REFERENCES

[1] Lawrence Rabiner, Biing-Hwang, and B. Yegnanarayana,"fundamental of speech recognition" 1st ED. India: Dorling Kindersley, 2009.

[2] Joe Tebelskis, "Speech Recognition using Neural Networks", Doctor of Philosophy in Computer Science,School of Computer ScienceCarnegie Mellon UniversityPittsburgh, Pennsylvania, 1995.

[3] Anjali Jain, 2O.P. Sharma,"A Vector Quantization Approach for Voice Recognition Using Mel Frequency Cepstral Coefficient (MFCC): A Review", IJECT Vol. 4, Issue Spl - 4, April - June 2013.

[4] WouterGevaert, GeorgiTsenov, ValeriMladenov, Senior Member, IEEE," Neural Networks used for Speech Recognition", journal of automatic control, university of belgrade, vol. 20:1-7, 2010.

[5] Mr. Kashyap Patel, Dr. R.K. Prasad," Speech Recognition and Verification Using MFCC & VQ", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013.

[6] Ms. Rupali S Chavan*1, Dr. Ganesh S. Sable2,"An Implementation of Text Dependent Speaker Independent Isolated Word Speech Recognition Using HMM",an Implementation of Text Dependent Speaker Independent Isolated Word Speech Recognition Using HMM",international journal of engineering sciences & research Technology,vol.2: September, 2013.

[7] Fernando I. Ablaza Jr.1, Timothy Oliver D. Danganan2, Bryan Paul L. Javier3, Kevin S. Manalang4, Denise Erica V. Montalvo5, and Engr. Leonard U. Ambata6," A Small Vocabulary Automatic Filipino Speech Profanity Suppression System Using Hybrid Hidden Markov Model / Artificial Neural Network (HMM/ANN) Keyword Spotting Framework", 7th IEEE conference on humanoid, nanotechnology, information technology communication and control, environment management, 12-14 November 2014.

[8] Hongzhi Wang, YuchaoXu and Meijing Li, "Study on the MFCC Similarity-based Voice Activity Detection(VAD) Algorithm" IEEE 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce, pp. 4391-4394, 2011.

[9] D. W. Thiang, "Limited speech recognition for controlling movement of mobile robot implemented on atmega162 microcontroller," in International Conference on Computer and Automation Engineering, 2009.

[10] MathWorks. (2013, August) Pattern recognition network. [Online]. Available: http://www.mathworks.com/help/nnet/ref/patternnet.html.

[11] J. Taheri, A. Y. Zomaya,"Artificial neural networks", In Handbook of Nature-Inspired and Innovative Computing, Springer US, pp. 147-185, 2006.