Voice Controlled Robot using Neural Network based Speech Recognition using Linear Predictive Coding

M Yasir Amir Khan, Haroon Rasheed, Umair Shahid

***Abstract* – Voice controlled robots, machines and systems are becoming more and more common in the current era of technology. Robots which can be operated by human verbal commands have widespread applications in the area of medicine, industry, military and as well as space exploration. This paper presents the work done in development of a voice controlled Arduino based robot. Features of the speech signal are extracted using Linear Predictive Coding, which are then processed by neural network based classifier with 10 input, 50 hidden and 5 output units. System has been implemented using MATLAB and it was found that different commands had varying amount of classification errors.**

Index Terms – Neural network, Classification, Linear Predictive Coding, MATLAB, Arduino, speech recognition

I. INTRODUCTION

Voice controlled robotic systems use verbal and non-verbal voice signals to control robotic systems. These robots have enormous applications in numerous fields ranging from medicine such as wheel chair control by disabled patients to industrial machinery control. There are two main groups in this field, one uses verbal and the other non-verbal signals. This paper focuses on verbal voice recognition to control which is in this case an Arduino based robot as shown in Figure 1.

II. LITERATURE REVIEW

Many researchers have accomplished notable contributions to this area. For instance Jeff Bilmes *et. al.* [1] proposed a system where human voice could provide control signal to manipulate real 3D robotic arm and simulated a robotic arm along with a real 3D robotic arm. Their goal was to enable individuals with motor impairment to manipulate common objects of day to day life. It was shown that using non-verbal voice communication, the robot could effectively manipulated real world objects. Their works paves the way for non-verbal voice controlled robotics and highly useful prosthetic limbs.

A type of voice recognition control technology for mobile robot system was proposed and presented in [3] where the system could robustly recognize voice of adults and children in presence of noise. Voice was captured using a wireless microphone and so as to suppress noise and reverberation, a multi-channel system consisting of a robust generalized side-lobe canceller technique was implemented. Voice activity periods were detected based end-point detection.

Another interesting work is [4], where development of a voice-controlled personal assistant robot is presented. In this work commands as human voice were fed to the robotic assistant. The speech signal commands converted to text form were communicated to the robot over a Bluetooth network. The robot could perform different movements, turns, start/stop operations and relocate an object from one place to another. Tests showed encouraging results.

Manish *et. al*. reviewed the possible application of voice controlled robot including automotive system [2].

Fig. 1 Arduino based robot used in this project.

III. METHODOLOGY

Human voice signals acquired via PC microphone have several kinds of noise and unwanted spectral components/ harmonics. Preprocessing of these signals is very important. In fact the working of system depends on correct preprocessing. Furthermore the signal needs to be normalized in terms of amplitude, because samples or actual speech signals have varying degrees of amplitude.

The preprocessed signal cannot be directly fed to the NN. For example a word on average takes around 1.5s and if sampled at a rate of 8000 samples/sec there would be 12000 samples requiring 12000 inputs to the NN. However the system could be sufficiently trained to classify with only the relevant information. Therefore the signal would be converted to frequency domain by taking discrete fourier transforms using FFT algorithm. The FFT transformed signal can then be used determines its cepstrum. MATLAB function Cepstrum coefficients C=ifft(log(abs(fft(X)))). The initial coefficients contain most of the information required

M. Yasir Amir Khan, Haroon Rasheed, Umair Shahid, Department of Electrical Engineering, Bahria University. Email: myasir@bimcs.edu.pk. Manuscript received July 25, 2016; revised on Aug 27, 2016; accepted on Sep 17, 2016.

 ISSN – 1999-4974

for the NN hence instead of 12000 samples our job should in principle be done by 20 coefficients. By using mel frequency the working of the system would drastically improve [5]. However we have used the technique of Linear Predictive Coding(LPC) which is a more widely accepted technique and faster to implement on MATLAB according to our simulations.

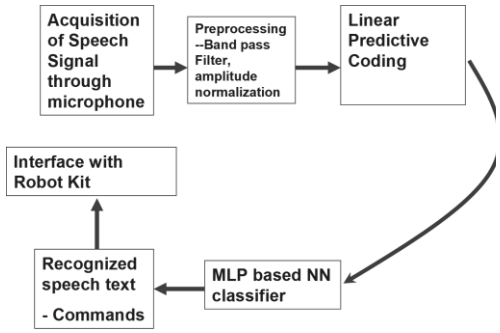

Fig. 2 Signal flow block diagram.

The LPC analyzed signal in form of features is then fed to a feed-forward multilayer perceptron with tangent-hyperbolic or log-sigmoid hidden layer and linear out put layer. The number of hidden and output layer neurons could be determined my Oja's thumb rule. The NN classifies the signal and coverts a text command which is used to control the arduino based robot interfaced with the PC using MATLAB.

The robot kit which was developed for this project comprises of arduino board and L268 dual h-bride dc motor drive. The robot is power by 7.5V dc and via USB cable connected to PC.

The signal flow block diagram of the system is shown in Figure 2. Matlab code is developed in this research work which comprises of three part or separate files. The first part is used to collect samples of commands. Twenty samples of each of five commands are acquires via PC microphone. The signal is then filter using a 50th order FIR BPF filter with 300-3750Hz frequency range. The filter signal is then processed via LPC technique and 10 features are collected. An S matrix containing features of all 100 samples 20 belonging to each word is created and stored as .mat file.

The second part of the code is used to train the 10 input 50 hidden and 5 output NN. The number of epochs can be selected in the code. The weight matrices for hidden and output layer are calculated and stored as .mat file.

The third part of the code is the main program which performs two operations. Firstly it acquires speech command signal in real time, filters it and performs the LPC. The features are then fed to NN which is previously trained and the output command is determined as one of five. Secondly it interfaces with the arduino board using arduinoIO library and sends the relevant commands.

IV. DETAILS OF DATA USED

This system is based on neural networks (NN) based classification which need extensive training. To train the system for each verbal command (word) atleast 100 samples spoken by different male/ female individuals were required and about 100 to test the system.

The system has been trained for five commands and the voice signal was acquired via microphone. The system has been implemented using MATLAB. We have acquired 20 samples of each of five commands. Figure 2 shows the histogram of the sample data. It can be seen that the distribution is overlapping making the problem very challenging.

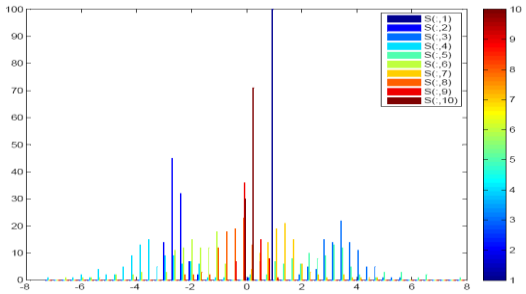

Fig. 2 Histogram of features.

The plot of the features is shown in Figure 3. It can be seen that the features show variation in their shape for different command words.

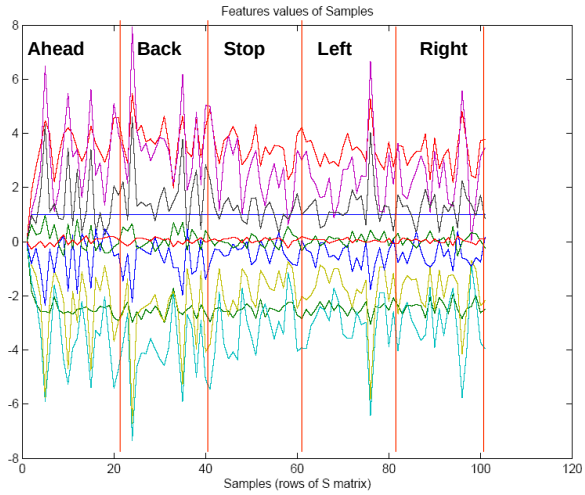

Fig. 3 Plot of features

V. EXPERIMENTAL RESULTS AND DISCUSSION

The system was tested with 100 samples with 20 samples of each of 5 commands. The NN was trained for 200 epochs and the final mean square error calculated by the program was 24. This value is fairly high and should be less than 1. The false classifications are summarized in Table 1.

It can be observed that some of the commands are more susceptible to error possibly due to more demanding features for example 'left' has error probability of 0.5 i.e 50% of the time we will get 'left' command incorrectly recognized.

ISSN – 1999-4974

Table 1. Summary of incorrect classification

Command	False Classification in 10	Probability of Error
Ahead	3	0.3
Stop	2	0.2
Back	1	0.1
Right	4	0.4
Left	5	0.5

This paper has presented a voice controlled robot system using NN based speech recognition. The system was developed using MATLAB and human voice signal was acquired using PC microphone. The robot was arduino based using h-bridge motor drive to turn the motors. System was experimentally tested and it was found that system recognized five different commands but with varying degrees of accuracy. This system can be further worked upon to increase accuracy by increasing the number of samples as only 20 samples for each command were used and also by increasing the training epochs. Future work may include increasing the number of commands.

REFERENCES

[1] House, Brandi, Jonathan Malkin, and Jeff Bilmes. "The VoiceBot: A Voice Controlled Robot Arm." (2009).

[2] Manish, N., and B. Jagadeesh Reddy. "Review of Voice Control Robot Applications." *Int. J. Adv. Eng* 1.9 (2015): 671-674.

[3] Koo, Young-mok, et al. "A study on travelling control of Mobile robot by voice commend." *Control, Automation and Systems (ICCAS), 2015 15th International Conference on*. IEEE, 2015.

[4] Mishra, Anurag, et al. "A voice-controlled personal assistant robot."*Industrial Instrumentation and Control (ICIC), 2015 International Conference on*. IEEE, 2015.

[5] Gevaert, Wouter, Georgi Tsenov, and Valeri Mladenov. "Neural networks used for speech recognition." *Journal of Automatic Control* 20.1 (2010): 1-7.

 ISSN – 1999-4974