Voice Controlled Robot in Industrial Automation

Sunil J.Pancha[1], P.B.Borole[2]
[1]MTech (Electronics), Electrical Engineering Department,
[2]Prof., Electrical Engineering Department,
Veermata Jeejabai Technological Institute,
Mumbai, India

Abstract- The purpose of this paper is a systematic approach to design and implement a Real-Time voice activated programmable multipurpose robot which can be used to assist an operator on industrial automation platform a generic pick and place application and mobility. A user can program its functions simply by pointing to object in work area that aims to explore ways to command an industrial robot using the human voice. This feature can be interesting with several industrial, laboratory and clean room applications where a close cooperation between robots and human is desirable.

Index Terms— MFCC, Speech Recognition, Voice Controlled Robot, Robotics in Industrial automation and control ,Artificial Intelligence in Robotics. Artificial Intelligence in Industrial automation.

I. INTRODUCTION

In the last few decades, large enterprises in high volume markets have managed to remain competitive and maintain qualified jobs by increasing their productivity through, among others, the incremental adoption and use of advanced ICT and Robotics technologies. In the 70s-80s robots have been introduced for the automation of a wide spectrum of tasks such as assembly of cars, white goods, electronic devices, machining of metal and plastic parts and handling of work pieces and objects of all kinds. Robotics as has thus soon become a synonym for competitive manufacturing and a key contributing technology for strengthening the economic base of the world. So far there are several methods have been developed to automate and control robots, talking to machines is a thing normally associated with science and fiction movies and less with current industrial manufacturing systems.

In fact most of the research papers about speech recognition start with something related with artificial intelligence or a robot used in movie and cartoons etc.where machines talks like humans, understand the complex human speech without problems.Neverthless, industrial manufacturing systems would benefits very much from speech recognition for human-machine interface (HMI) even if technology is not so advanced [12]. Gains in terms of autonomy, efficiency and agility seem evident. The modern word requires better products using faster and cheaper procedures. This means autonomy, having a system that requires less operator interventions to operate normally, better human-machine interfaces and cooperation between humans and machines sharing the same working platform as real coworkers. The final objective is to achieving some cases semi-autonomous systems. In this paper it is aimed to control a robot with speech commands. The robot is able to recognize to move correctly and perform its mechanical functions for object pick and place.e.g.to give direction to robot first voice command is sent to the computer or cell phone (HMI) using a microphone the HMI recognizes the command by speech recognition system. And then HMI converts the voice command to direction command that predefined and recognized by robot. When robot gets the direction command, it moves according to spoken command.

II. METHODOLOGY

The voice data is taken from the microphone. This data is stored in an array. This array is passed on to a function which extracts words from the array. These words are sent to a function which extracts frequency as a function of time; this is the frequency vector of the spoken word. This vector is compared with the reference vectors. The comparison is done using the standard inner product of two vectors; one of the reference vectors would match. The command corresponding to this reference vector is fed to electronic circuit mounted on robot base would then interpret actuators to perform mechanical action.

III. SPEECH RECOGNISATION

The speech signals are captured coming from the microphone attached to the HMI. The Software running on HMI processes the signals to recognize the voice commands ‘Forward’, ‘Stop’, ‘Left’, ’Right’ , ‘Reverse’, ‘Pick’ and ’Place’. The software also provides a facility to train itself for the above commands. For training we use Artificial Neural Networks. The software is written using MATLAB. To identify words ,various methods are used for identifications of words such as Linear preductive coding(LPC) and its residual Mel Frequency Cepstral coefficients(MFCC), First, a spoken word is recognized and then a set of parameters called ‘Cepstral Co-efficient’ was calculated from the voice data samples belonging to that word based on MFCC. Then those co-efficient are processed by the trained neural network to decide whether that word is any of the above specific commands. If the word is identified as one of those commands, then a relevant signal is sent to the robot platform via RS232.The speech processing involves following steps

A. *Word Capturing*

The signal coming from the microphone is processed only when speaker speaks something. The program waits until the sample value exceeds some threshold value. When the program is triggered by a significant sample. A number of samples are captured to process after that to determine the actual boundaries of the word spoken 'Edge Detection' is performed.

B. *The MFCC Processor*

A block diagram of the structure of an MFCC processor is given in Figure 1.The speech signal is recorded at the sample rate of 22050Hz.This sampling frequency is chosen to minimize the effect of aliasing in the analog to digital conversion. The Figure 1 shows The Structure of an MFCC Processor

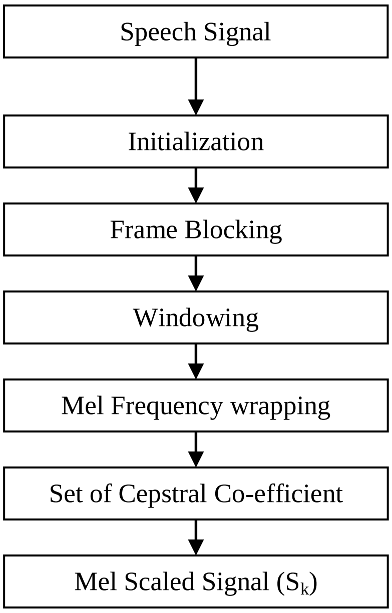

Fig.1. The Structure of an MFCC Processor

The control algorithm of an MFCC Processor has been described below.

a) Initialization

This operation removes Noise (DC components and low frequency) components from speech signal. FIR filtering was applied to the signal in the time domain using the MATLAB function 'filter'.

b) Frame Blocking

In this step the continuous speech signal is segmented into frames of N samples with 25% overlapping frames.

c) Windowing

This step is applied for spectral analysis of speech signal. Each individual frame is windowed to minimize the signal discontinuities at the border of each frame. The concept here is to minimize the spectral distortion by using window to tapper the signal to zero at the border of each frame. The set of samples for each frame is multiplied by the time domain version of 'Hamming Window' with size equal to the frame length, then the result of windowing is the signal.

$$\mathrm{W(n)}=0.54 - 0.46\times\left(\cos\left(2\times\frac{n}{N-1}\right)\right) \quad (1)$$

Where $N \le n \le N-1$

d) Fast Fourier Transform (FFT)

FFT was applied on each frame to obtain the spectral information from time domain signal. The generalized equation for DFT is given bellow,

$$\mathrm{x(n)}=\sum_{\mathrm{k=0}}^{\mathrm{N-1}} \mathrm{x(k)}\, e^{\frac{-j2\pi kn}{N}} \quad (2)$$

Where n= 0, 1, 2…., N-1

e) Mel-Frequency Wrapping

The spectrum obtained from the FFT output is "Mel Frequency Wrapped". The Mel frequency scale is linear frequency spacing bellow 1000Hz and logarithmic spacing above 1000Hz.As a reference point, the pitch of a 1KHz tone, 40db above the perpetual earring threshold is defined as 1000Mels. For a given frequency, the Mel scale is being calculated by the following equation.

$$Mel(f)=2595\times log_{10}(1+\frac{f}{700}) \quad (3)$$

The major work done in this process is to convert the frequency spectrum to Mel Spectrum. For each tone with an actual frequency "f" measured in Hz, a subjective pitch is measured on a scale called the "Mel" Scale**.**

f) Cepstrum

In this step the log Mel spectrum is converted back to time. The result is called the Mel frequency Cepstral coefficients (MFCC).The Cepstral representation of speech provides a good representation of the local spectral properties of the signal for the given frame analysis. Finally the discrete cosine transform (DCT) was applied to the signal in order to obtain MFCC coefficients

$$c[\mathrm{n}]=\sum_{k=1}^{k}\log(\mathrm{S_k})\left[\cos\left(k-\frac{1}{2}\right)\frac{n\pi}{k}\right] \quad (4)$$

Where n=1, 2, 3… k

Where Sk the Mel Scaled signal got after wrapping Cn is the Cepstral coefficient

C. *feature Extraction from MFCC*

For the short period of time the characteristics of the speech signal are fairly stationery, therefore the short-time spectral analysis is the most common way to characterize the speech signal[].The input speech signal is segmented into number of frames. Windowing operation is performed to capture the static property of the signal. Hamming window with 20ms size and 25% overlapping has been used here. Then fast Fourier transform is applied to produce the spectral characteristics of the speech signal. for the given frequency the Mel frequency was calculated by equation 3.Finally the log Mel spectrum is converted back to real time domain in order to get frequency Cepstral Coefficients(MFCC).

IV. Feature Matching

The state-of-the-art feature matching techniques used in speech reorganization include Dynamic Time warping **(DTW)**, Hidden Markov Modeling **(HMM)** and Vector Quantization **(VQ),**K-Nearest Neighbor**(K-NN)** algorithm, Artificial Neural Network**(ANN)**.In some cases hidden Markov models(HHM's) and Neural Networks(NN's) are used together in speech recognition. In this paper artificial neural network is used.

A. Artificial Neural Network

To recognize the speech the robot must be intelligent. Hence artificial neural network is used to make robot intelligent through the learning process. Artificial neural networks (ANN's) are systems consisting of interconnected computational nodes working somewhat similarly to human neurons. Neural networks can be used e.g.to approximate functions or classify data into similar classes e.g. phonemes, sub-phoneme units, syllables or words in the speech recognition domains. The ability to learn by adapting strengths of interconnections is a fundamental property of artificial neural networks. There are countless different structures for a neural network few are discussed here,

a) Feedforward Networks

Feedforward neural network is an artificial neural network where connections between the units do not form a directed cycle[11]. This is different from recurrent neural networks. The Feedforward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. In a feed forward network information always moves one direction; it never goes backward

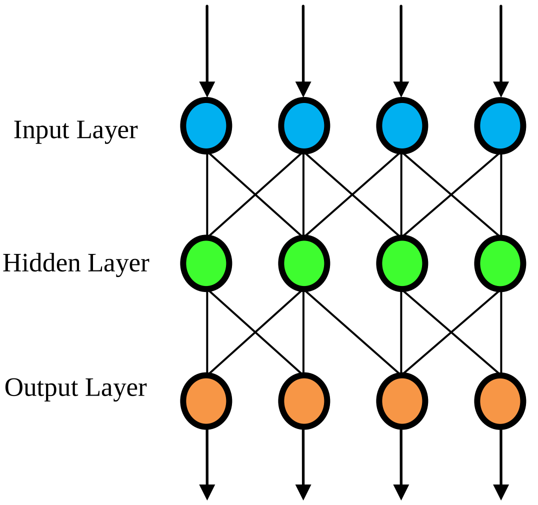

Fig. 2. A feedforward neural network

A Feed forward network has connections only forward in time, a neuron in layer i can only send data to neutrons j j>i .Only adjacent layers can be connected to each other as in multilayer perceptrons, or there can be forward "shortcuts" between layers that are next to each other. A time delay neural network (TDNN) uses Feed forward connections with weighted delays, which makes it possible to use contextual information (i.e. previous values of e.g. acoustic speech vectors) in classification of data. This adds complexity to the network and requires a more complex learning rule such as backpropagation through time, but enhances accuracy.

b) Perceptrons and multi-layer perceptrons

A perceptrons is a simple neuron model that has a set of inputs, a weight for each input and a (often nonlinear) activation function that the neuron performs to the weighted sum of inputs (plus possible bias) before sending the value to its output . The perceptrons model is shown in Figure 3, where x is an input vector, w is a weight vector, w0 is the bias and the activation function is a step function.

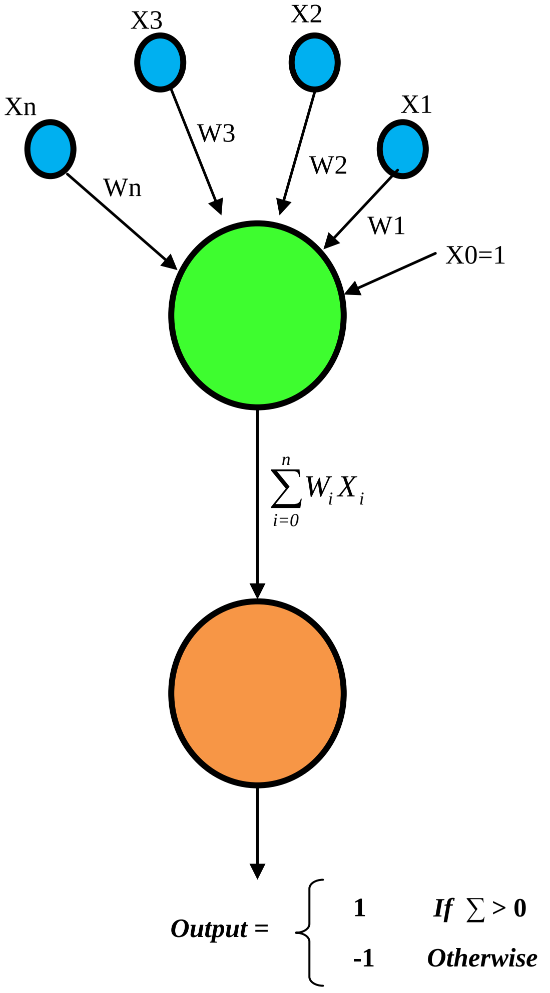

Fig. 3. The perceptrons model

A multi-layer perceptrons (MLP) consists of at least two layers of perceptrons: it has an input layer, one or more hidden layers and output layer. The hidden layers act as a feature extractor and use a nonlinear function such as sigmoid or a radial-basis function to generate (often complex) functions of input. The outputs of all the neurons in the hidden layer serve as input to all of the neurons on the next layer. The output layer acts as a logical net that chooses an index to send to the output on the basis of inputs it receives from the hidden layer, so that the classification error is minimized.

B. *Feature extraction with MLP's*

If a multi-layer perceptrons is fed the acoustical information as an input and it is trained to give the same data out in its outputs, the hidden layer(s) of the network will learn a representation of the data. If the amount if hidden neurons is smaller than the size of the input data vector and the accuracy of the input-output-function is good, the network has managed to extract the essential information from the data and can reproduce the data accurately enough from this information. The internal state of the hidden layer can thus be seen to contain the features needed to classify sounds into classes.

Input layer transfer function

$$Output(y) = mx \tag{5}$$

Hidden layers & output layers transfer function is given by,

$$Output = \frac{1}{1+e^{-\lambda x}} \tag{6}$$

The neural network has been designed in MATLAB computing environment and it process data in real time and the program give the outputs according to recognize the voice commands. In this first of all calculated set of Cepstral coefficients for that spoken word is fed to input layer of artificial Neural Network (ANN) algorithm and calculate the output. During the training process the set of voice commands samples are fed to ANN where output is to be produced. After each cycle the network calculates the actual network output and compare with the stored voice data if error exists it is backpropagate(Fig.4) to minimize it in next cycle of execution.

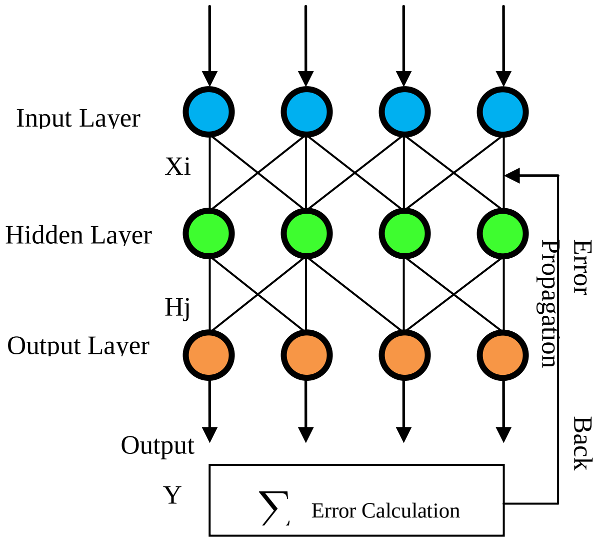

Fig. 4. Error Back Propagation

V. Human Machine Interface

The main objective of the HMI is to provide user friendly interface to an operator. The HMI has been designed such that it should be able to facilitate all phases of voice recognition algorithm including monitoring and control of robot activities from the control panel (PC).The control panel of HMI consists of, Mode selection panel

Training Mode:-Use Cepstral coefficients to build the neural network in order to get neural network coefficients.

Tuning Mode:- Tuning of known voice samples to calculate the threshold levels.

Running Mode:- Each voice input is sampled real time and use the ANN to get the output.

Communication Setup:- In this area chose com port of the computer which is to be connected to robot platform via communication link.

Connected:-This option is selected if communication between HMI and Robot has been established.

Not connected:-This option is selected if communication between HMI and Robot is failed.

Control Panel:-In this area the control and Monitoring of robot activity is done fig.5 shows HMI panel for voice controlled robot, it has been designed by using MATLAB software. If respective button is pressed robot moves forward, reverse left, right, stop, pick the objects and place. Exit and activates its arm for pick and place.

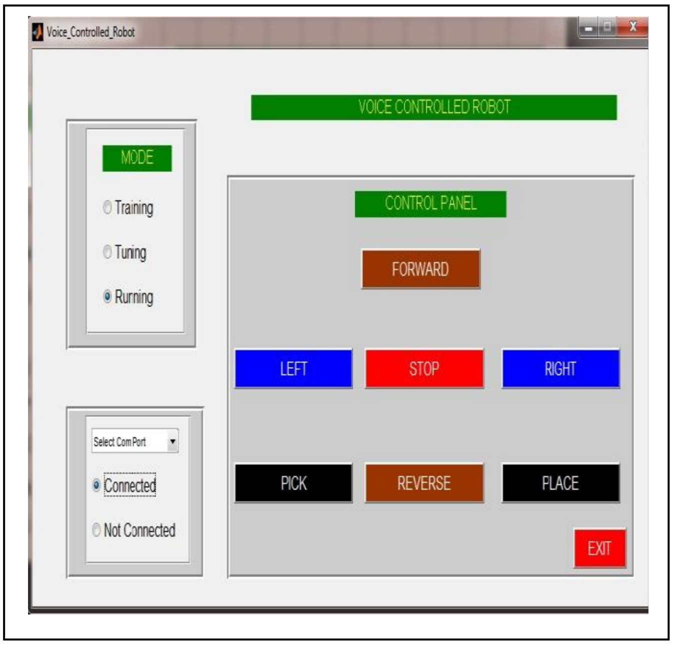

Fig. 5. Human Machine Interface Control Panel

VI. Hardware

The voice controlled robot is a vehicle which moves on the four wheels, it has control hardware includes voice recognition module (HM2007), microcontroller and decoder circuitry module, Motor driver Circuit, mechanical assembly, buffer and computer as HMI. The most challenging part of the entire system is designing and interfacing various stages together and converting analog voice signal into digital. The frequency and pitch of words be stored in memory. These stored words will be used for matching with the spoken words. When match is found, the system computes and outputs the address of the stored words. Hence this address is to be decoded and according to the address sensed, the robot vehicle will perform the required task. The use of RF module makes the robotic vehicle wireless. The address was decoded using microcontroller and then fed to RF module. This together with driver circuit and receivers end made complete intelligent system. The hardware is shown in the figures bellow.

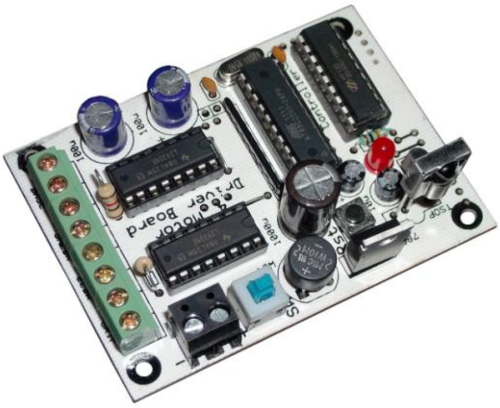

Fig. 6. Robot Driver Circuit

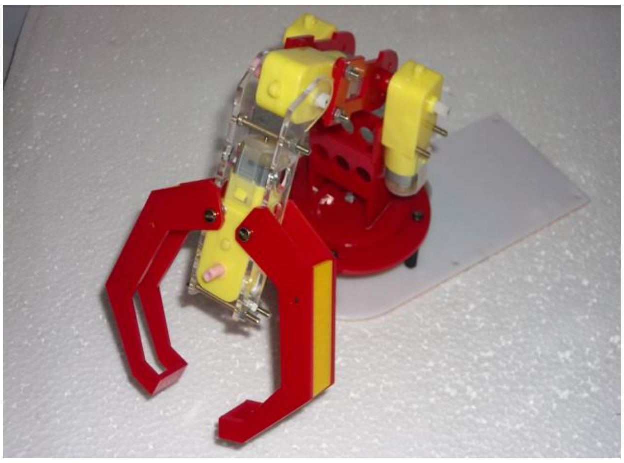

Fig. 7. Robot Arm

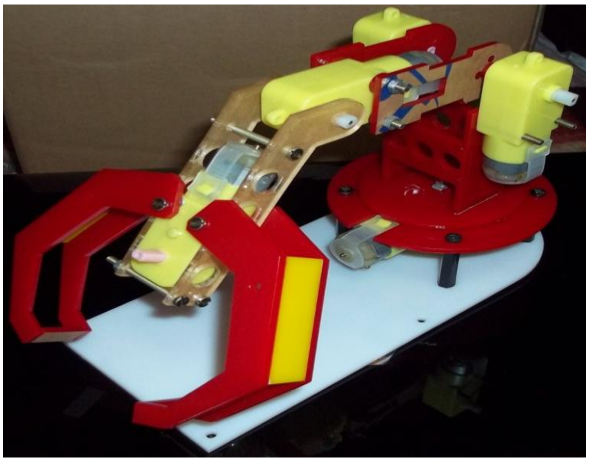

Fig. 8. Robot Arm Gripper Open

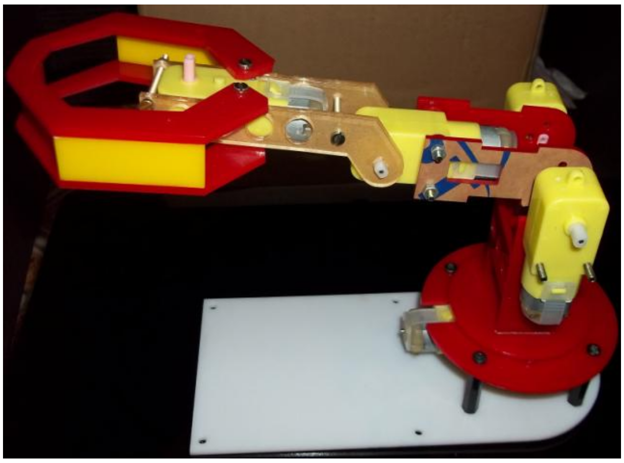

Fig. 9. Robot Arm Reach Horizontal

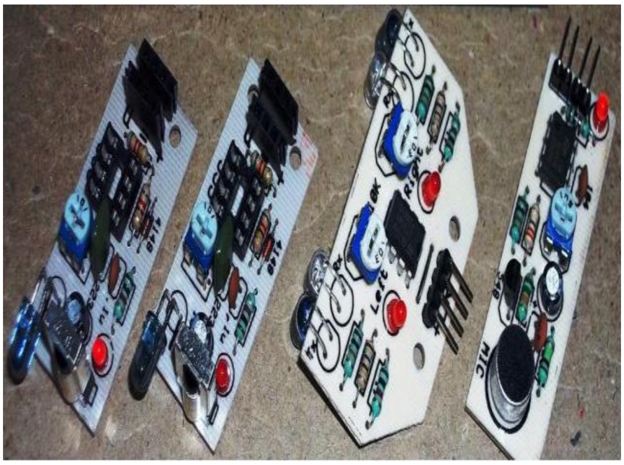

Fig. 11. Sensors and Mounts

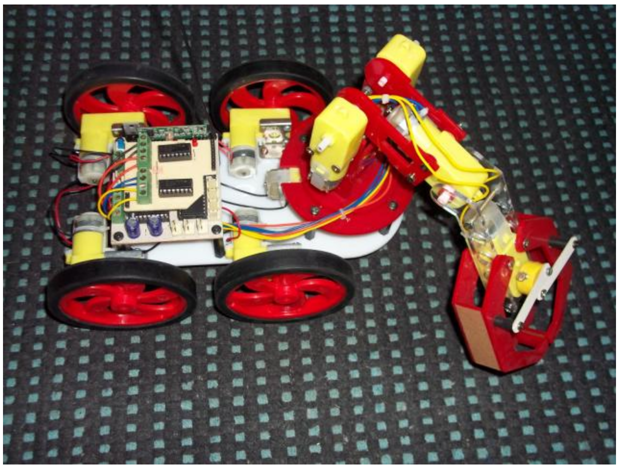

Fig. 12. Experimental Platform for Voice Controlled Robot

VII. Experiment results

In this project, a robot is controlled with the speech commands. The features of commands are extracted with MFCC algorithm. The commands are recognized using Artificial Neural Network (ANN).The recognized command converted to the form in which the robot can recognize. The final form of the commands is sent to the robot and robot moves accordingly. The system is tested with different command sets and for both current user and different users. Generally the system recognizes the commands with 90% to 100% success ratios for current users and 75% to 85% for other users.

VIII. Conclusion

So it is concluded that the voice recognition software had accuracy around 75% in correctly identifying the voice command, but it is highly sensitive to surrounding noises. The sound coming from motors also has a significant effect on accuracy.

Acknowledgment

We acknowledge support of Center of Excellence (CoE) in embedded systems, Complex and Non-linear Dynamical Systems.

References

[1] Kan, Phak Len Eh, Allen, Tim, and Quigley, F,: A GMM-Based Speaker IdentificationSystem on FPGA. In: 6th international symposium on Reconfigurable Computing:Architectures, Tools and Applications. Bangkok, Thailand march 2010, LNCS (2010)

[2] Toh, A.M., Togneri, R., Northolt, S.: Spectral entropy as speech features for speechrecognition. In: The Proceedings of PEECS, Perth, pp. 22–25 (2005)

[3] Lindasalwa, M., Begam, M., Elamvazuthi, I.:Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques.In: Jour. of Computing, vol. 2, Issu 3, pp. 138-143 (2010)

[4] Dutta, Tridibesh: Dynamic Time Warping Based Approach to Text Dependent SpeakerIdentification Using Spectrograms. In: Congress on Image and Signal Processing, Vol. 2,pp. 354-360 (2008)

[5] P. Nauth, “Speech and Image Recognition for Intelligent Robot Control with Self Generating Will”, IWSSIP 2010 - 17th International Conference on Systems,Signals and Image Processing

[6] R. Jain and S. K. Saxena, “Voice Automated Mobile Robot”, International Journal of Computer Applications (0975 – 8887) vol.16, no.2, February 2011

[7] Z. Zhang, L. Zicheng, M. Sinclair, A. Acero, L. Deng, J. Droppo,X. Huang and Y. Zheng, “Multi-sensory microphones for robust speech detection, enhancement and recognition,” Proceedings of the 2004 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP `04), Vol. 3, pp. iii-781-4, May2004.

[8] Zarycki, and J. Levin, “Human-machine interface for telerobotic operation: mapping of tongue movements based on aural flow monitoring,”Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Vol. 1, pp. 859-865, September - October 2004.

[9] S. Theodoridis and K. Koutroumbas, Pattern Recognition, Second Edition, Academic Press, San Diego, California, 2003.

[10] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceedings of the IEEE, Vol.77, No.2, pp.257- 286,February 1989.[Rabiner, 1993] L. R. Rabiner and B-H. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.

[11] www.Wikipedia.comhttp://en.wikipedia.org/wiki/Feedforward_neural_network,

[12] J.Norberto Pires,” Robot-by-voice: experiments on commanding an industrial robot .”,Industrial robot,An International journal,Emrald group publishing Ltd.volume 32,Number 6,2005

[13] http://home.iitk.ac.in/~amit/courses/768/00/gatram/ “Control of a Robot by Voice Input - IITK - Indian Institute ”

[14] http://www.ifp.illinois.edu/~minhdo/teaching/speaker_recognition/speaker_recognition.doc” Digital Signal Processing Mini-Project”

[15] http://info2myfriends.blog.com/files/2010/11/speech-recognisation-using-DSP.doc