ISSN No. 0976-5697
Volume 8, No. 5, May – June 2017
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Voice Controlled Smart Electric-Powered wheelchair based on Artificial Neural Network

Mohammed H. Jabardi
Information Technology Research and Development Center,
University of Kufa, Najaf, Iraq

Abstract: Electric-Powered Wheelchair (EPW) used by people with disabilities, who cannot move or with a limited ability to move. The traditional technique that used to control the powered wheelchair can be a challenging and stressful operation. Therefore, there is a demand to find alternative ways to perform this process. In that, the voice recognition is a natural form of pointing that can use to replace the joystick whilst still allowing for similar control. Neural network one of the effective ways to recognize the speech commands, because of its ability to grant a trainable powered wheelchair controller with a convenient way, ease of use and attractiveness. In this situation, the neural networks are providing a trainable system for each person irrespectively of his disability.

Smart Electric-Powered Wheelchair (SEPW) is an electric-powered wheelchair based on Artificial Neural Networks (ANN) Technique can recognize the voice commands and replace the traditional control of powered wheelchairs with intelligent technology. So, a neural network for the feed-forward Multi-layer perceptron has been trained to recognize isolated spoken words (commands), such as ‘right, ‘left,' ‘stop' and so on. Initially, the design of the system's network started with one hidden layer using five neurons (according to [1] formula). During the training process, the number of hidden neurons has been increased to keep the mean square error (MSE) of the system minimum as possible. Furthermore, the system structure is designed based on a Multi Layers Perceptron (MLP) neural system for five output neurons, which represent the powered wheelchair's commands.

Keywords: Electric-Powered Wheelchair; Smart Electric-Powered Wheelchair; voice recognition; Artificial Neural Networks; disabilities**;** mean square error; feed-forward Multi-layer

I- INTRODUCTION

According to the Centers for Disease Control and Prevention (CDC),there are, 53000,000 persons suffers from some kind of disability in the USA, Approximately ,2000,000 persons of them used wheelchairs for independent mobility and daily tasks[2]. The objective of this work is to design and develop a new technology for reliable and usable electric wheelchairs controlled by speech recognition. Mobility is restricted or absent at the extreme of paralyzed and multiple sclerosis patients, in particular for Patients who suffer from the disease of the white matter tissue. This issue leads to demand for entirely new technology for independent mobility to help these people. One of the intelligent solutions to such problems is designing powered wheelchair control using voice commands. Initially, the system receives the speech signal, then after some signal, processing it converts this signal to the particular motion of the wheelchair. With this regulated movement of the wheelchair, most of the lost mobility can be retained.

Consequently, many input methods have implemented to perform this task such as bio-signals and non-bio-signals . The bio-signals completely depends on the brain signal that can be recording by EMG, EEG and EEG devices, this type of technology is too sophisticated and depend on power of brain signals [3,4].

In non-bio-signal completely depends on human action such as use of head movement, tongue control, joystick control ,eyes movement and voice controlled [5].

In this approach, several algorithms based on neural networks have been presented to deal with speech recognition through training the net, so neural network is considered powerful technique and accuracy [6].Neural networks are able to deal with these challenges and present some advantages over than conventional techniques.

A. Speech Recognition

The speech is the core mode of communication between the people, so they learn the required skills naturally at the childhood stage or at different stages of their lives. The main problem with the voice recognition is the difference in characteristics of the same word specked by varies people or even by the same speaker with various conditions. This variation can result because of several reasons: acoustic variability where the same command can pronounce in a variety of ways, speaking variability where the speaker can talk slowly or quickly and Phonetic variability where people from a different background (from different regions) can be expressed with various accents for the same command.

For the human, it comes gradually and simply without any complexity because of the abilities of the human’s brain. On the other hand, the speech recognition process through the machine is a very complex issue (because of this variation in the speech signal), and the real-time speech recognition is a fundamental problem for any speech based system.

The efficiency of any speech controlled system can vary depending on the following parameters[7]

1. Vocabulary size and language constraints:-

In general, the growth of the words size in speech recognition operation leads an increase of error rates, if we take a small set of words, such as numbers from zero to nine, we get a very small percentage of errors[8] compared to a large size of Vocabulary [9][10]. On the other hand side, is not always easy to distinguish small vocabulary due to containing the vowel character.

2. Speaker dependence and independence:-

A dependent system is a system that uses an only particular person, while the independent system is a more reliable system which can use by any speaker. However, the independence system hard to meet because generally, the characteristic of the system is tuning to accurate the speaker(s). According to J. Tebelskis the speaker dependent systems has small error rate than independent systems may reach 3 to 5 times[11] As in-between the two systems, there are multi-speaker systems that can be used by a small group of people, by training the system with their voices.

3. Adverse conditions:-

These refer to external conditions such as the environmental noise , sound distortions, changing the microphone and so on. Therefore try to reduce the effect of these conditions, the training should be implemented with samples that recorded with different conditions. Moreover, an increasing number of samples contributes to a degree to solve this problem. However, these solutions still limited and these problems are associated with the voice recognition process by all techniques. So, some hardware solutions can deal with this issue better than the software such as individual microphones.

B. Artificial Neural Networks

An Artificial Neural Networks (ANN) is a computer system based on functions and design of biological neural networks .ANN system mimic the way of accessing the information in the human brain. Generally, most ANNs are composed of three types of layers: an Input layer, one or more hidden layers and output layer, each layer be composed of a number of neural units and each neural unit connected to other units in the previous and forward layers by edges or line called weight. The learning procedure is done through updating the value of weights according to a particular function ,such as sigmoid function. Connections between neural units carry out an activation signal of varying strong that may be transferred to another node or not. The idea of the artificial neural systems started in 1943 by McCulloch and Pitts when they outline the first formal model of the neuron's computing [1] This idea has been developed to propose a learning scheme to update the neuron's connection by so-called Hebbian learning rule which implemented by Donald Hebb[12] Later in 1954, Minsky built first neurocomputers which adopted connections automatically[13]. The development of the artificial neural network has continued to discover successful extensions of neural networks knowledge in 1986.

After this year, many types of research, journals, conferences, and programs have investigated this field. Nowadays, it has become an attractive field of study because it offers high efficient systems. In that, Zurada argues that the artificial neural network "is still in its early stages of development" and he expects more and more from this nets [1].

The features of ANN are Trainability; these networks acquire their ability by training and learning with any patterns. Nonlinearity, the neural networks are able to deal with the nonlinear system, especially speech recognition systems (because of these at most nonlinear systems).

COMPONENTS OF A SPEECH RECOGNITION SYSTEM

The operation starts from receiving the voice signal (by microphone or recorder) and finishes at command's execution, as shown in figure 1.

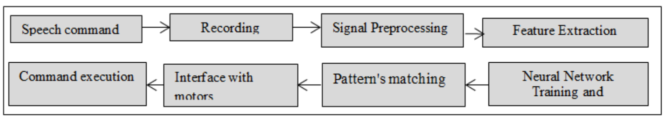

Figure 1 speech recognition steps

The primary operations of the proposed system are normalized and extract features of the speech signal, classify inputs by a neural network (after design and train it), and finally execute the command according to the output, which means the start action (move forward, backward, left or right and stop) .Figure 2 shows the block diagram of the proposed system and figure 3 shows the system architecture.

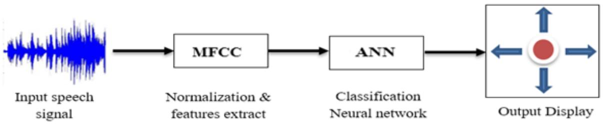

Figure 2 the block diagram of proposed system

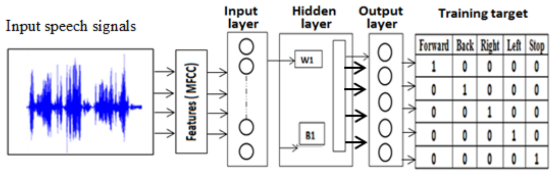

Figure 3 System Architecture

II. METHODOLOGY

A. Feature extraction and normalization

1.Feature extraction

The primary purpose of this operation is to analyze or to code the input signal. The digitized signal is not enough to be input to the neural network because the elements of this signal are too many and they are not very expressive. Feature extraction means to perform a series of steps or operations on (something) in order to obtain feature vector from raw signal. So, this process is essential to any automatic speech recognition system. The most common methods for feature extraction are

Linear Predictive Coding (LPC),Mel-Frequency Discrete Wavelet Coefficient (MFDWC) and Mel-Frequency Cepstral Coefficients (MFCCs) .

1.1MFCC's Algorithm.

One of the ways to extract the features is Mel Frequency Cepstral Coefficients (MFCCs). This firstly introduced by Davis and Mermelste in the 1980's[14] and has been most significant at all [7]. MFCC is the most common method that used to debrief features of spectral in speech recognition. MFCC based on frequency range, while the other techniques based on domain features, there for it supposed more precise form the others techniques[26][27]. The figure (4) shows the steps of extract these coefficients (features) using the MFCC method.

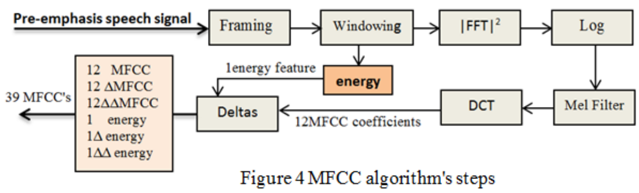

Figure 4 MFCC algorithm's steps

During this operation, the speech signal will process according to below steps:-

1- **Framing** the first step in mfcc algoritm by divided the signal into number of blocks(frame),typically into 20-30 ms, standard is 25 ms.Suppose we have sample rate R, K ms, so the number of frams equal a[N]=KxL .For example if we have ample rate R=2o KHz with K=20 ms, the frame length equal to RxK=20x20=400 samples.

2- **Windowing**, because of the spectrum varies rapidly across time, which means that the speech is a non-stable signal. Therefore we cannot extract features from the speech signals. As an alternative, by choosing a small specific region (window) of speech, we can obtain spectral features, with considered that signals in that window are stationary, so windowing is used to prevent sudden jump at the ends of the signal.

In order to process the windowing, we need to determine three parameters: what is the windows' shape, how far the distance between two consecutive windows, and what the window's range in milliseconds). There are many types of windows: rectangular ,Hamming, Blackman, Gauss, and the triangular window. In the rectangular window, the signal at its boundaries can suddenly cuts, which cause the problem during applying Fourier analysis. Therefore, Hamming window used to overcome the problem in the rectangular window by shrinks the signal's value at boundaries to zero.

The windowing function applied for each frame with the following form

$w[n] = 0.54 - 0.46 \cos(2\pi n/N - 1)$ …………… (1)

N represent window length, where $0 < n < N-1$

3- **Discrete Fourier Transformation**: Because of the different tunes in speech signals associated with different energy, required to apply Discrete Fourier Transform (DFT) to calculate the magnitude frequency of the frame. The DFT equations are:

$$S_i(K) = \sum_{n=1}^{N} S_i(n) h(n) e^{-j2\pi kn/N} \quad \text{…………(2)}$$

where $1 \leqslant k \leqslant K$

Where h(n) is N sample and K is the length of the DFT. The best algorithm for calculating the DFT is Fast Fourier Transformation FFT.The power spectral estimate for the speech frame is given by:

$$P_i(k) = \frac{1}{N} |S_i(k)| \quad \text{………. (3)}$$

4- Calculate the Mel-spaced filter bank ,typically 40 (26 is standard) triangular filters applied to the spectral power estimate from step 2 to extract frequency

$$L_p(m,k) = \{\sum_{k=0}^{N-1} M(m,k) * |X_p(k)|\} \text{………. (4)}$$

where m = 1, 2,…….,F and p = 1, 2 ……, P.

The resulting MFCC features are 39 (12 Cepstral Coefficients ,12 delta Cepstral Coefficients ,12 double delta, 1 energy coefficients, 1 delta energy coefficient and 1 double delta energy coefficient) numbers for each frame) are called Mel-Frequency Cepstral Coefficients. IN summary, the MFCC operation results in a new vector which will be the input to the next stage, in the other words, at this level we prepare the input of the next stage which is the neural network .

Figure 5 shows the MFCC operation for a sample of the word "left" which is one of the 1200 samples as an input signal

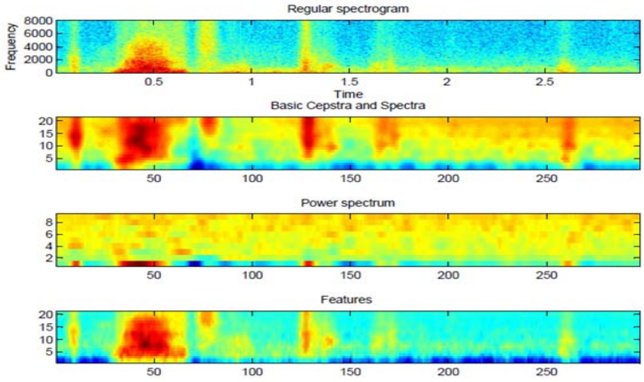

Figure 5 Mel-frequency cepstral coefficients for "LEFT" sound wave

1.2 Normalization

The results of the MFCC operation are 20 bits with different values. So, the input layer contains 20 nodes in addition to the bias (-1).
Theoretically, the network performance should not have an effect of the range of inputs value, since the learning operation done by updating the weights and adjusting the bias of the hidden units. However, inputs normalization in the specific range makes the network learn more efficient.
In order to make a fair comparison between our inputs representations, we first normalized all of them to the uniform boundaries, [-1, 1] as shown in the figure 6.

-15.5862	-15.1840	-13.4664
-1.7647	-1.8142	-0.7801
-8.2611	-10.8527	-11.1257
-5.0904	-6.5110	-6.9271
-2.9156	-3.8321	-3.6893
-2.6344	-3.3208	-3.4420
1.2170	0.5041	1.8378
-3.7787	-4.3306	-3.9424
-5.1036	-5.9846	-5.3257
-4.2741	-4.7355	-3.7031
-3.5303	-3.8638	-4.4253
-0.9835	-1.8915	-1.3072
-1.0974	-1.3545	-2.1016

a) Before the normalization

-0.6005	-0.5636	-0.4061
0.3471	0.3382	0.5255
-0.5044	-0.9528	-1.0000
-0.1502	-0.4464	-0.5332
-0.0285	-0.2576	-0.2219
-0.2924	-0.4259	-0.4495
0.6577	0.4837	0.8093
-0.6712	-0.8321	-0.7190
-0.4417	-0.6788	-0.5015
-0.3632	-0.5147	-0.1757
-0.0303	-0.1391	-0.3221
0.2293	-0.0862	0.1168
0.0471	-0.0286	-0.2485

b) after the normalization

Figure 6 part of the input matrix

B. Classification (The neural network)

1.Learning algorithm (Back-propagation)

Back-propagation learning algorithm is the most significant algorithm in neural networks [21][22][23][24][25]. the Back-propagation algorithm spread of errors inversely to the back during the training process where the errors in the output layer determine the error of hidden layer which is used a bias to adjust the weights of the link between output layer and hidden layer and continue the treatment and recycling of outputs for a number of iterations until the arrival of the solution optimized for by achieving the lowest rate of error[6][28]. For a given training set of input and output pairs [x(k) , d(k)] where k= 1,2,3,…………,N, the procedure for updating the wieghts of the network explain as follow.

Procedure in Back-propagation algorithm

1- Given P are training pairs {Z1, d1, Z2, d2,……., Zp, dp}, Where Zi is (i × 1), di is (K × 1), and i = 1, 2, . . . P.

2- X[j×1] an input matrix of data(Digits features)

3- W[k × j] , $\overline{W}$ [J×1] W and $\overline{W}$ weights are initialized to small random values;
4-Learning rate λ
5-The transfer function:- $f(y) = \frac{1-e^{-y}}{1+e^{-y}}$
Steps of propagation algorithm
Step 1: If n > 0, Exam chosen.
Weights W and $\overline{W}$ are set to small random values

Step 2: Starting training.
Input is presented, and the output layer computed Z= Zp, d=dp

$\overline{W}j = f(\overline{W}j' * Z)$ ……. (5) for i = 1, 2. . ., i

$O_K = f(W_k' * y)$ ……. (6) for k = 1, 2 . . . , K
Step 3: Error value(E):
$E = 0.5\,(d(k) - o(k))^2 + E$ ……… (7)
Step 4: Error signal vectors ESO (Error Signal Output layer), and ESH (Error Signal Hidden layers) are calculated as follow:
$SO(k) = 0.5(d(k) - O(k))(1 - O(k)^2)$ …….. (8)
for k =1, 2 . . . , K.
SH y (j) = SO (k) x W (k, j) x f(y) ……… (9)
for j = 1, 2 . . . , J
Step 5: Adjusting weights of the output layer.
$Wkj = Wkj + \lambda \times SO(k) \times yj$ ……… (10)
for k = 1, 2. . . K and j = 1, 2 . . . , J.
Step 6: Adjusting weights of hidden layer.
$\overline{W}ji = \overline{W}ji + \lambda \times SH_y \times Zi$ …….. (11)
for j= 1, 2, . . . , J and i = 1, 2, ... ,I .

Step 7: The procedure of training has done.

If (E < Emax) terminate the training session. The output are weights (W and $\overline{W}$) and E.

If (E > Emax) then let p=1, E = 0, and go to Step 2 for a new training cycle.

End

IV. Experiment result

Designing any neural network depends on two main issues network architecture and training procedure. Artificial Neaural Network formed from input layer, hidden layers and output layer. Network's architecture issues such as, how many number of hidden layers, how many number of units per layer and what the appropriate activation functions. As any speech based system, the proposed system will deal with these factors and this section explains all specifications of the network's architecture, training procedure, and the result.

A. The hidden layer:-

Our network has one hidden layer, because:

1.Single hidden layer with enough hidden neurons can perform a function instead of using multiple hidden layers without effected on the network's accuracy [19].

2.The training time increases substantially for networks with multiple hidden layers .

The number of hidden units has a high impact on the performance of an MLP. With more hidden units, the network will have more complex decision surfaces, and hence the better classification accuracy it can attain [3]. There are many proposed a technique for many researchers used to find the numbers of hidden units, In 1993, Arai [17] suggested that 2n hidden units ,while Tamura and Tateishi [18] proposed in three layers neural network that (n-1) hidden units are enough where n present number of inputs . In 2003, Huang presented two different formula to calculate the hidden units

($\sqrt{(m+2)n} + 2\sqrt{(n/(m+2)}$) hidden units for a single hidden layer and ($m\sqrt{n/(m+2)}$)hidden units for double hidden layer , In 2012 [6] improve that's no need to trial-and-error rule, and we can use n+1 hidden units in single hidden layer network. Some techniques can use to reach the actual appropriate number such as a genetic algorithm or try and error. In our system, we depend on the try and error to know the required neurons. In fact, we started from five neurons then we increased them gradually until 100, and we find that 30 neurons give us the best response and the lowest error. Therefore, the hidden layer will include 30 neurons as a final decision.

The figure (7) demonstrates how to change the number of the neurons in the hidden layer can affect the response and the value of the error. We chose only five cases to keep the graph clear to the reader.

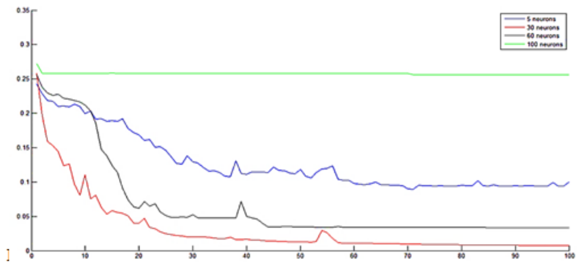

Figure 7 The root mean square errors with a different number of neurons in the hidden layer

B. The Output Layer

The output layer of neural network is related with the input layer, because in any mathematical operation, the outputs basically depends on processing of inputs, subsequently the number of output units directly related to work of the neural networks. Generally, any type of EPW should be move for four direction and stopping, so our neural network deal with these five tasks by command's classifying(five speech commands),which means, one output neural for each speech command. In our proposed system we have five commands (stop, left, right, forward, backward).So we will expect that the network has five output neurons as shown in table 1.

Table 1 Coding of output neurons

Forward	Backward	Right	Left	Stop
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

During the classification process, all output neurons will be all (0) except one (1) which represents the correct class. So, any result includes more than one positive output will be ignored and classify as an unrecognized pattern

Therefore, we need to set the target of our samples. We organize our samples to be in sequence forward, back, left, right, stop, forward, back… and so on.

A network of the system will be built from three layers (Input layer, Hidden layer, Output layer).

C. Training procedure

Training procedure means training the network on the related data to make the network capability to distinguish speech command to perform the proper action. Many parameters can effects on training operation, such as best learning rate, number of hidden unite ,the bias (1 or -1), number of the cycles, and type of algorithm that used to update the weights. This part presents our research on training procedures, including learning rate schedules, data presentation, and update procedure.

Learning Rate Schedules.

The learning rate selection is crucial issue during the training process. If the learning rate is too small, the network will move toward optimal solution very slowly, while the network will swing back and forth when the learning rate is too high.

If two different networks are trained with the same learning rate schedule, it will be unfair to compare their results to a limited number of cycles, because the learning rate schedule may have been optimal for one of the networks but not for the other. We eventually spend a long time to understand the effect of learning rate schedules on network performance. We began by applying different values of learning rates, in order to reach the best result .

Figure 8 shows the learning curves (for different values of learning rates in the Range 0.05 to 3. We see that a learning rate of 0.05 is too small while 2.5 is too large. In the mean time, a learning rate of 0.15 gave the best result at the beginning, but 0.1 has been proved better later. These curves illustrate how the response can influence with changing the learning rate within the range with same initial weights.

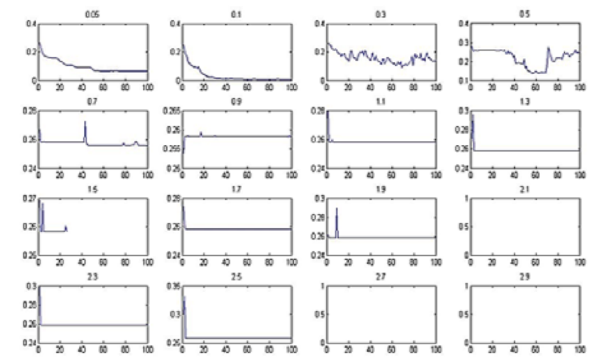

Figure 8 The root mean square errors with different learning rates

In our system, we took 1200 samples for 12 different people; they divided into three groups:-

Table 2 Percentage of dataset grouping

Types	Total number	The percent
Training set	360 samples	30%
Validation set	360 samples	30%
Test set	480 samples	40%
Total	**1200**	

Each group is divided into three parts:-

1- Samples for one person

2-Samples for a group of persons.

3-People from outside of the training scope (they are not same people who train the network).

D. The training and validation process

The neural network training is not restricted nonlinear minimization problem in which weights of a net are continuousely updated to reduce the total squared error between the actual and desired output values for all output neuron over all input patterns.

The figure (9) shows the training and validation operations and illustrates that 250 cycles are enough to reach to the final weights that can offer high accuracy. We will check the accuracy during test stage.

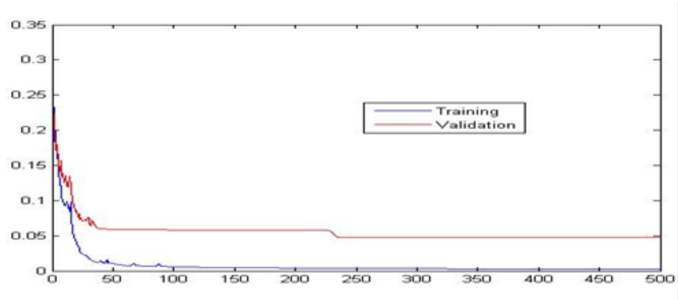

figure (9) shows the training and validation operations

E. Results

A test set is a set of samples that was not used in any way whatsoever during the training process (which described in the previous part). The error of the test set is the primary result to be presented for any learning problem used. We trained and tested the system with three different cases.

First case: - Training with samples for specific person, then test the system by another sample to the same person

Table 3 The results of the first case

Commands	Total number	Correct	Unrecognized	Wrong
Forward	20	17 85%	2 10%	1 5%
Back	20	20 100%	0 0%	0 0%
Right	20	18 90%	1 5%	1 5%
Left	20	20 100%	0 0%	0 0%
Stop	20	18 90%	1 5%	1 5%
Total	100	93	4	3
Percentage		93%	4%	3%

Second case: - Training with samples for group of people, then test the system with different samples to same people

Table 4 The results of the second case

Commands	Total number	Correct		Unrecognized		Wrong	
Forward	40	33	82.5%	4	10%	3	7.5%
Back	40	36	90%	1	3%	3	7%
Right	40	33	82.5%	4	10%	3	7.5%
Left	40	31	77.5%	5	12.5%	4	10%
Stop	40	39	97%	0	0%	1	3%
Total	**200**	**172**		**18**		**14**	
Percentage		**86%**		**7%**		**7%**	

Third case: - Training with samples for specific person but test it by another one

Table 5 The results of the third case

Commands	Total number	Correct		Unrecognized		Wrong	
Forward	20	15	75%	4	20%	1	5%
Back	20	16	80%	2	10%	2	10%
Right	20	14	70%	4	20%	2	10%
Left	20	15	75%	3	15%	2	10%
Stop	20	16	80%	2	10%	2	10%
Total	100	76		15		9	
Percentage		76%		15%		9%	

Initially, the dependent and independent system is well described in the first section in this paper.Among of all, as shown in the figure 10, the first system that trained by one person and designed to use by the same user offers the highest accuracy (93%) to follow by the second one who trained and used by small group of people and provide 86%. However, the third network can be classified as a weak net with lower accuracy (only 76%).

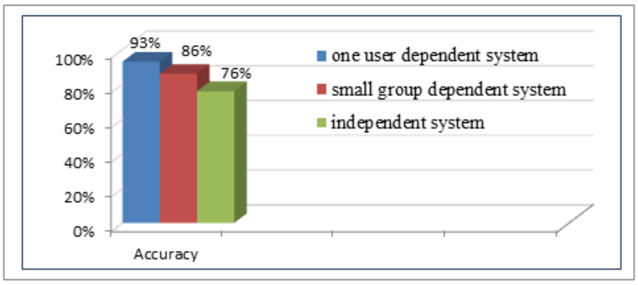

Figure 10 Results of dependent and independent

These results prove that our technique for speech recognition is good enough to provide the disabled people with reliable and convenient wheelchair.

V. Conclusion and future works

In this paper, using neural networks to control wheelchair through speech recognition technology has been applied. Inaddition, an automatic speech recognition system has been designed using MATLAB programming. By the fully automated training and recognition process, an accuracy of more than 86% is achieved for a group of people and 76% with unknown pattern (spoken by unknown speakers). This can help many disabled people by providing them with a trainable system for each person irrespectively of his disability.

The results show that the performance of a network gets better by increasing the training samples with people who benefit from the system. Furthermore, the network performed optimally in speaker dependent and limited one language (English). The definitive goal of this study is to establish an accurate speech recognition system, which is capable of dealing with distinct speech inputs, rather than just normal speech. Upon preliminary attempts, this network seems to be robust enough to satisfy that task.As a future direction, we will attempt to

1- Increase the accuracy of the system, in particular for the unknown patterns (which are spoken by anonymous speakers) to be similar to the accuracy of the known people. This can be done by increasing the samples and take more samples from different people with various conditions (environment conditions). This can support the system to be or more reliable and trusted.

2-International System, we will increase the usability of the system by supporting different languages, which means more people can use it.

3-Multifunctional system, try to support system with Bluetooth , sensors and WIFI technology in order to give the system the capabilities of connection with others device likes Mobile and control devices to serve different tasks.

VI. REFRENCES

[1] Zurada, J.M. 1992, Introduction to Artificial Neural Systems, West Publishing Company, St. Paul.

[2] Mohammed Hayyan ,"A Study on Smart Wheelchair Systems", INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY AND SCIENCES (IJETS(,Vol.4 (1) December 2015

[3] Murali Krishnan, Muralindran Mariappan," EEG-Based Brain-Machine Interface (BMI) for Controlling Mobile Robots: The Trend of Prior Studies", IJCSEE, Vol. 3, pp.159-165, Issue 2, 2015.

[4] Vijay Khare, Jayashree Santhosh, Sneh Anand, Manvir Bhatia," Brain Computer Interface Based Real Time Control of Wheelchair Using Electroencephalogram", International Journal of Soft Computing and Engineering (IJSCE), Vol. 1, Issue 5, pp. 41-45, November 2011

[5] Jobin Jose, "Development of EOG Based Human Machine Interface Control System for Motorized Wheelchair", Master of Technology, Department of Biotechnology and Medical Engineering National Institute of Technology Rourkela, Odisha, 769 008, India, June 2013

[6] Mohammed H. Jabardi. & Harleen Kaur.(2014). Artificial Neural Network Classification for Handwritten Digits Recognition, International Journal of Advanced Research in Computer Science .

[7] Joe Tebelskis,1995, Speech Recognition using Neural Networks, PhD thesis, Carnegie Mellon University, Pittsburgh. Journal of Rehabilitation research and Development, July/august 2005

[8] Doddington, G. (1989). Phonetically Sensitive Discriminates for Improved Speech Recognition. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing.

[9] Itakura, F. (1975). Minimum Prediction Residual Principle Applied to Speech Recognition.IEEE Trans. on Acoustics, Speech, and Signal Processing, 23(1):67-72, February. Reprinted in Waibel and Lee (1990).

[10] Kimura, S. (1990). 100,000-Word Recognition Using Acoustic-Segment Networks. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing.

[11] Hebb, D. O. (1949). The Organization of Behavior: A Neuropsychological Theory. New York: Wiley and Sons. ISBN

9780471367277

[12] Minsky, Marvin Lee (1954). Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain Model Problem (PhD thesis). Princeton University

[13] Davis, S. Mermelstein, P. (1980) Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366

[14] Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. Journal Mathematics of Control, Signals, and Systems, vol. 2, no. 4

[15] X. Huang, A. Acero, and H. Hon. Spoken Language Processing: A guide to theory, algorithm, and system development. Prentice Hall, 2001.

[16] M. Arai, “Bounds on the number of hidden units in binary-valued three-layer neural networks,” Neural Networks, vol. 6, no. 6, pp. 855–860, 1993

[17] S. Tamura and M. Tateishi, “Capabilities of a four-layered feedforward neural network: four layers versus three,” IEEE Transactions on Neural Networks, vol. 8, no. 2, pp. 251–255, 1997

[18] G. B. Huang, “Learning capability and storage capacity of two-hidden-layer feedforward networks,” IEEE Transactions on Neural Networks, vol. 14, no. 2, pp. 274–281, 2003

[19] D. Hunter, H. Yu, M. S. Pukish III, J. Kolbusz, and B. M. Wilamowski, “Selection of proper neural network sizes and architectures: a comparative study,” IEEE and Management Science,MIT,april 1985.

[20] Bryson, A.E., and Ho, Yu-Chi, Applied Optimal Control, Blaisdell, New York, 1969.

[21] Werbos, Paul J., Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Scienc Mathematics, Harvard University, November, 1974

[22] Lecun, Y. (1988). A theoretical framework for back-propagation. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), Proceedings of the 1988 Connectionist

[23] Parker, David B.,"Learning-Logic" , Technical Report TR-47, Center for Computational Research in Economics

[24] Rumelhart, David E., Hinton, Geoffrey E., and William, Ronald J., "Learning representations by back-propagating errors",Nature,1986

[25] Lei Xie, Zhi-Qiang Liu, “A Comparative Study of Audio Features For Audio to Visual Conversion in MPEG-4 Compliant Facial Animation,” Proc. of ICMLC, Dalian, 13-16 Aug-2006.

[26] Alfie Tan Kok Leong, “A Music Identification System Based on Audio Content Similarity,” Thesis of Bachelor of Engineering, Division of Electrical Engineering, The School of Information Technology and Electrical Engineering, The University of Queensland, Queensland, Oct-2003.

[27] Namrata Dave "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition", International journal for advance research in engineering and technology, Volume 1, Issue VI, July 2013

