Speaker Verification System Using Artificial Neural Network : MLP

Ali Mohammed Ben Ramadan
Misurata university –Misurata Libya
Science Faculty
henaish@yahoo.com

Abstract— In a world where authentication and privacy are taking a lot of our daily efforts, it is becoming more important for us to prove our identity to different systems every day so that we can access required and useful services by speaker verification system.

Speaker verification (Is the speaker who we think he or she is ?) .In addition , speaker verification can be closed-set (The speaker is always one of a closed set used for training.) or open-set (speaker from outside the training set may be examined .).
Also, each variant may be implement as text-dependent (The speaker must utter one of a closed set of words.) or text-independent (The speaker may utter any type of speech).

In this project we explore the ability of a multilayer perceptron (MLP) to perform text-dependent speaker verification .

Our networks are trained on sets of acoustical parameters extracted form samples obtained from a closed set of speakers uttering a set of known words .Our primary feature extraction tools are Mel Frequency Cepstral Coefficient (MFCC).

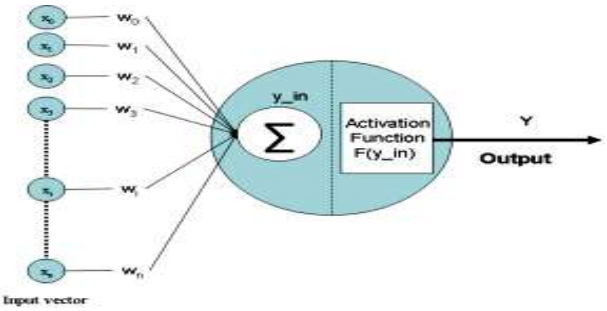
All software for this project was created using Matlab Version 6.5 , and neural network processing was carried out using the Matlab toolbox .

Keywords— ***Speaker verification, multilayer perceptron (MLP)***

I. Introduction

1.1 Artificial Neural networks
Artificial neural networks is a system of information processing has advantages of certain performance in a manner that simulates neural networks vital is a Wizard, a large distributor in parallel, and component units of the treatment is simple, search stores the practical knowledge to make it available to the user and that by adjusting the weights.
Neural networks were used as examples of mathematical way of thinking based on human and how to deal with nerves of information, the following steps show the components of the neural network and how to address the industrial to the information.

• Information is processed in the processing elements called nodes "Nodes"
• the signals transmitted between the nodes via communication lines.
• For each line of contact a certain weight "Weight", which hits with signals entering the node.
• each node applies a function to activate "Activation Function", which are usually non-linear income for her as an output signal in Figure (1).

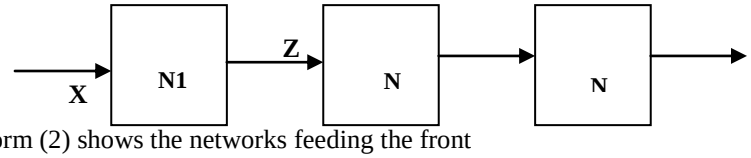


Figure(1) shows the neuron industrial

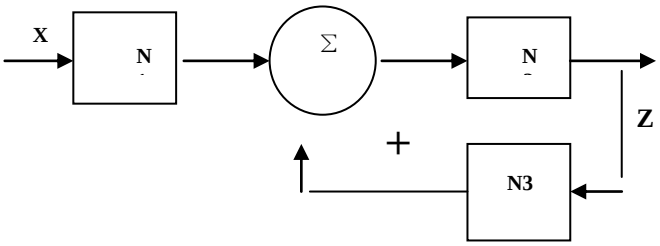
2.1 Rating Artificial Neural Networks
Be the nodes in Artificial Neural Networks arranged in layers, so that is not required to have an equal number of nodes in each layer, and arrange the nodes in the layer and the pattern of contact with other classes is called the structure of neural network (Networks Architecture).
And can be classified as Artificial Neural Networks Bloatmad the main components, namely:

• the pattern of contact between the nodes is called the "Architecture"
• modify the way the weights between the nodes (the way of education) Learning "Algorithm"
• The type of activation function "Activation Function Type"

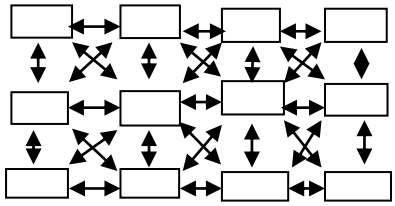
Classification of neural networks in terms of industrial structure:
Can be classified as neural networks in terms of industrial structure as follows: -
Feed Forward Network
In this type of network output nodes are calculated depending on the Mdkhaladthawalta intervention of the external structure or hold other classes, and the advantage of this type of neural networks Bastaqraria system as shown in Figure(2)):



Form (2) shows the networks feeding the front
Feed Back Networks
This product is the same kind in the former structure, and is characterized by the presence of a background of feeding the output to the input as shown in Figure (3):



Form (3) shows the networks feeding back
Cellular Networks
Arranged in the form of contract where the matrices so that the nodes are connected to each other directly and take the matrices in which rectangular or triangular or any other form as shown in Figure (4):



Form (.4) illustrates the cellular networks
Classification Artificial Neural Networks in terms of the number of layers
2. Single Layer Networks
This consists of networks of layer income receives the signal from the outside world and layer output that we get them on the response of the network, and put one set of weights on the lines of communication between the holding of two layers of input and output, but this type of networks compared to networks with multiple classes can not solve many of the complex problems, where they can be exposed to the problem can not be solved at all using a single-class network even if trained for a long time.

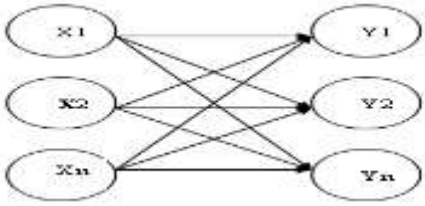
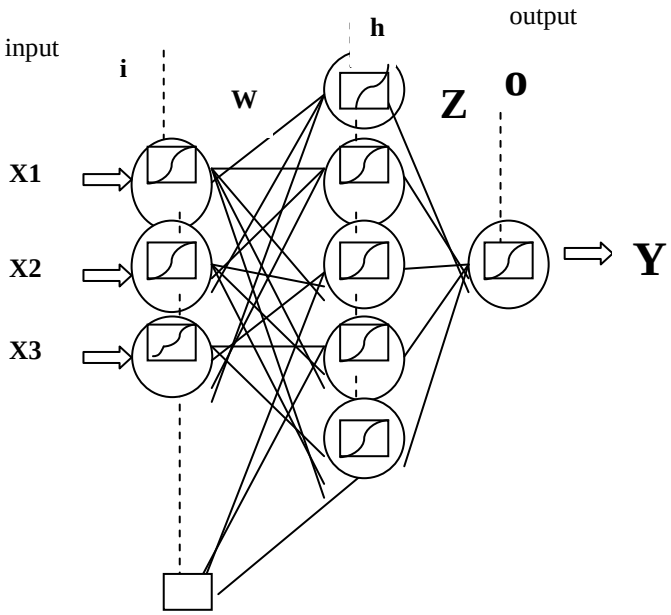
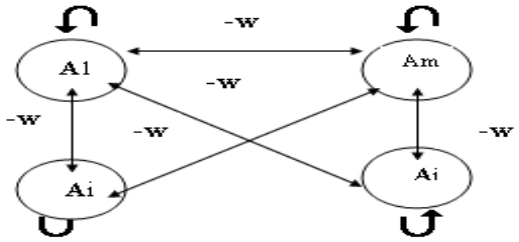


Figure (5. A) shows the single-layer networks
2.1 Multi Layer Networks
These networks consist of one layer or more nodes which are called the hidden layer and the layer placed between the income and output layer.
You may compared to multi-layered networks of networks and a single class to solve many complex problems, but the training of these networks take longer, but can be training for these networks more successful than others.
Layer hidden layer income
Output layer
Input output



Form (5B) shows the multi-layered networks
2.1Networks Competitive Layer
Constitute a large part of the Artificial Neural Networks, where nodes are connected in these networks through competitive links as described as follows:



shows the layered competitive networks igure (5. C)
3.Approaches to learning of Neural Networks
Learn the network by giving them a set of examples, which should be selected carefully, because that is the speed of learning. And range of examples of this category is called training.
And methods of education are divided into two neural network training by category of exposure to the network. Are: -
Supervised Learning
In this way, the training class that displays on the network a couple of vectors, vector input is a valuable input to the network, and the output vector which is a value that must be graduating network.
Example
Input: (0 1 0 1 0 0 0 1)
Output (0 1 1)
Unsupervised learning
In this way, the training class is an input vector only without displaying the output on the network.
Common Activation function
The process involves the basic algorithms learning: collecting

input signals multiplied with weights and then apply the continued activation ((activation)) and must use the accessories to activate certain algorithms of education. And enables the ability of neural networks multilayer compared with the networks and a single layer in use for the consequences of activated non-linear with the possibility of take advantage of the consequences of activation of linear networks used in a single layer, with the following set of aftershocks activation common algorithms used in learning.

3.1Linear Function

He uses a linear networks in a single layer and transfer income is used to form the network output, a suitable reference value of the variable and constant, and the following figure illustrates this.

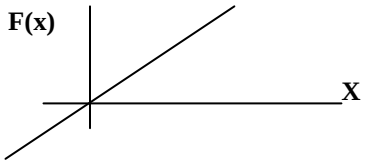


Figure (5D) shows the linear relationship

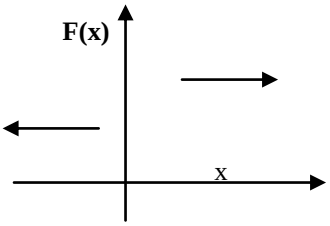
Sports and shape to him: -

F (x) = x for all x

(Binary step function (with threshold)

And it is collected weights of the signals input and comparing the value of certain limit or threshold ((θ) Threshold) If the sum of weights of the signals over all the (Threshold (θ) (where the signal output are (1), and if it was less the result is (0) and the following figure illustrates this.

Figure (5B) shows bilateral, continue with step (with a threshold θ)



Sports and shape him

$$f(x) = \begin{cases} 0 \text{ if } x > 0 \\ 1 \text{ if } x \le 0 \end{cases} \qquad (1)$$

3.2 Binary Sigmoid

It went very useful and there is of it two, disciples and relative accessories hyperbola disciples are commonly used, and has special advantages have to be used in algorithms to train the neural network with spread back "Back Propagation", and there is a relationship link between the disciples and their derivatives at a certain point ease the burden on the network During the training phase in order to address the accounts, and located the value of the disciples within the domain [1 0] and is often used Ktwaba activation for neural networks which have a value of output desired is a binary value or a variable within the domain [1 0], and the following figure illustrates this Council:

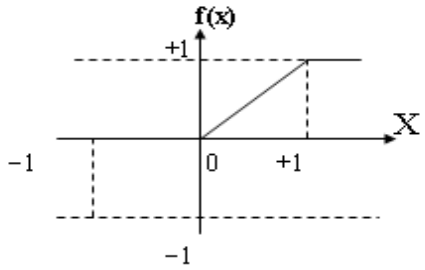


Figure (5E) shows the continued revitalization of bilateral Siegmoad

Sports and shape him:

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

Where σ is called the regression coefficient

3.3 Sigmoid Bipolar

The most common area for use is [+1, -1], which uses a bipolar child Siegmoad The following figure illustrates the order of Para meter slope 1 = σ

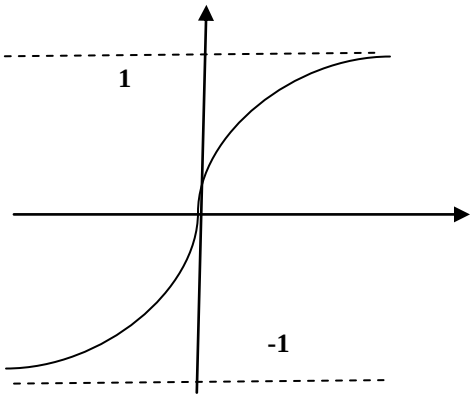


Figure (5.F) shows the continued revitalization of Alsiegmoad bipolar

And shape of the sport for this:

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

4. Advantages of Neural Network

Among the most important features of neural networks, the following points:

-1 Circular
Neural networks to memorize the training data only, and because it can generalize from training data to the training data to new examples so that they can not Altenbabsloc cases are not teaching the network in accordance to the behavior of the general gained from her education at the initial cases, and this is necessary to distinguish the sound.
-2 Parallel
The property of parallelism of the most important features of neural networks, to that with the big problems we need a large amount of information to be processed in a short time and therefore this feature will allow addressing Sriahjadda especially for words and other data.
-3 Permittivity
Where that knowledge is distributed in neural networks to all parts of the network so the loss of part of the network have no affect on the general behavior of the neural network.
4.1 Applications of Artificial Neural Networks
That the type of problem that needs to be resolved by the neural network determines the network in a manner and method of training.
Neural networks are fed through the input variables, and produce some of the variables of the output. Badzlk can be used when you have some of the known information and want to infer some information is known such as:
1 - excellence in optical instruments to read the codes.
2 - calculate and determine the personal bank loan applications.
3 - distinguish faces, speech, and fingerprints in security systems.
4 - the movement of automatic control and treatment.
5 - in the diagnosis of heart attack.
6 - the ability to recognize distorted images.
7 - in data mining.
8 - used purely in the engines.
9 - The ability to recognize the validity of signatures.
10 - signal processing to get rid of noise in the communication.
4.2 Standard Back propagation
Is a gradual regression method in order to find the smallest value of the square error for the total output value calculated by the network.
I have been using many of the training methods in the back diffusion networks (multi-layers, or forward deployment training networks in accordance with the rules of the back diffusion). And these networks can be used to solve many problems in multiple domains.
Includes the training of the network deployment in a back three stages: -
1 - the front of the stage of nutrition training samples Income
2 - the stage of the account and spread the rear of the error on the Al-Kharj
3 - phase synthesis of weights
The following figure shows the network deployment back with one hidden layer between input and output layers

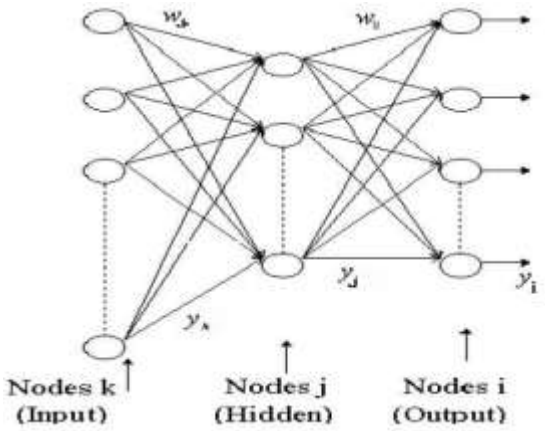


Figure (6) shows the network spread back
4.3 Symbols used in the algorithm
Xi-income units layer
Zj units of hidden layers
vij weights between the hidden layer and a layer of income
wjk weights between the hidden layer and output layer
Yk and output layer units
Vojgwor transactions bias
 K correction factor weights of the second layer between the hidden layer and output layer
tk represents the value of the target you want to access
α coefficient of education
(Y-ink) f differential activation function when the value of the network output
Learning Algorithm
Algorithm can be summarized in the education networks, the back diffusion index the following points: -
1 - to impose the values of initial random weights.
* Stage deployment front
2 - as long as the condition to stop the said point 8 Not achieved, repetitive steps 3 to 8 for the purpose of education at a time.
3 - Each unit income (xi, i = 1,2, ------, n) receives input signal xi and sends this signal to all units in the top layer and the hidden units (zij, j = 1,2, - ----, p) collects the input signal weights.

$$Z_{-in_j} = V_{0j} + \sum_{i=1}^{n} x_i * V_{ij} \quad \text{----(4.3)}$$

4 - The application of activation output signal to find her:
Zj = f (Z_inj)
It then sends this signal to all units in the layer above it (Units of output).
5 - Each unit of output (yk, k = 1,2, ------, m) collects input signals multiplied by the weights:

$$Y_{-in_k} = W_{0k} + \sum_{j=1}^{p} Z_j * W_{jk} \quad \text{----(5.3)}$$

After that the activation function is applied to calculate the value of the output signal according to the equation:
Yk = f (y_ink) -------- (6.3)
6 - Each unit output (Yk, K = 1, --------, m) compared to its

output (output stage deployment front) with the aim to find a value error, and based on this error factor is the distribution of errors which uses the weights of the equation : -

$$\sigma_k = (t_k - Y_k) * f'(Y_{-in_k}) \text{ ----(7.3)}$$

The correction factor is calculated by weight which is used to weight Wjk challenged for the second layer weights of the equation: -

$$\Delta W_{jk} = \alpha \sigma_k Z_j \text{ ----(8.3)}$$

It also calculates the correction factor weights which is used to Haddat bias weights of the equation: -

$$\Delta W_{0k} = \alpha \sigma_k \text{ ----(9.3)}$$

* Phase-Aligned congenita

7 - Each hidden unit Z j, j = 1, ------- p)) collect income from the units in the layer above it according to the equation:

$$\sigma_{-In_j} = \sum_{k=1}^{m} \sigma_k W_{jk} \text{ ----(10.3)}$$

And hit her in the differential coefficient of error is calculated in each unit, according to the following equation:

$$\sigma_j = \sigma_{-In_j} * f'(Z_{-In_j}) \text{ ----(11.3)}$$

The correction factor is calculated weights of the first layer which is used to weight challenged Vij of the equation:

$$\Delta V_{ij} = \alpha \sigma_j X_i \text{ ----(12.3)}$$

And calculates the correction factor weights which is used to bias challenged weights of the equation:

$$\Delta V_{0j} = \alpha \sigma_j$$

Each unit output challenged the weights according to the relationship:

Wjk (new) = Wjk (old) + Δ Wjk

Each unit also challenged the hidden weights according to the relationship:

Vij (new) = Vij (old) + Δ Wij

8 - to stop the test condition according to the error value returned is greater than permittivity steps 3 to 8 by using weights that were updated, but if the error is less than allowed will be to stop the process of education.

The test algorithm

After completion of the learning process the network is the testing process by forward deployment in accordance with the following steps: -

1 - read the primary weights (obtained from the learning algorithm).

2 - is made of the steps (3-5) per unit of factor income.

3 - Each unit income (Xi, i = 1, -------, n) receives an input signal Xi and sends this signal to all units in the next class.

4 - Each unit in this layer ((compile the input signal multiplied by the weights

$$Z_{-in_j} = V_{0j} + \sum_{i=1}^{n} X_i * V_{ij}$$

The application of activated function to find its output signal:

Zj = f (Z_inj)

And send a reference to all the units in the layer, followed by (output units).

5 - The grandmother of all units of output ((k = 1, ----, m Tdja input signals multiplied by the weights:

$$Y_{-in_k} = W_{0k} + \sum_{j=1}^{p} Z_j * W_{jk}$$

Function activated and applied to calculate the output signal.

Yk = f (y_ink)

It is noted that the basis of mathematical algorithm to spread the rear is a method known as "jumbo regression function Gradient Descent", which have an important role in most of the teaching methods based on the sum of squared errors, and contains these algorithms on a constant (coefficient of education) is always greater than zero to determine the size of the jump at update the weights.

The process of selection of primary weights will affect whether the network will amount to less error in "Global Error" and it also affects the speed of convergence of the function.

The process of updating the weights between two nodes depends on the derivative of a function of the node activation function of the upper and the lower activation of the node, and for this reason was to avoid primary test weights that are likely to make any of the activation functions or their derivatives equal to zero.

Kman weights elementary selected should not be too large so as not to reach the input signals primary to the area where the derivative of a function activation with a small value (the saturation), In contrast, if the weights primary very small, the income of the units output and layers of hidden will be roughly comparable to a very to zero, which causes very slow in the education network. Among the most famous methods used to create the weights and bias:

$$\beta = \frac{0.5}{Ya\min} \text{ ----------(13.3)}$$

$$\beta = 0.7\sqrt[n]{p} \text{ ----------(14.3)}$$

Where β: the weight gradient between the value of the first and the next and so on.

P: the number of nodes in the hidden layers

n and famin: the number of contract income at each stage.

The amendment of the weights of the initial simplex (by making them random values between(0.5; -0.5), or (-1.1]), or how appropriate the model gives a very high speed of Education, has been the use of equation (14.3) to determine the initial weights in this project.

5. The Education Network

When the proliferation network run back in the starting of education try to access the correct response for the models of education and the response is good for models Aljdladh income.

It is essential that we continue in the education of the network until it reaches the square error of less value to the public is allowed, it is advisable to use two sets of data during the

education, data collection for the Knowing, the data set to test education.
Each group is separate and unrelated to others, either modify the weights to be based on models of education, and between now and then calculates the error by using models of education, if this error is decreasing, we continue in the education process, but if the amount of error is increasing, the network will have begun in loss of control over learning, in which case you must stop the learning process of the network and the networks that have more than one hidden layer and one is coefficient of error frequently, Vtba units hidden to the weights of units of income will be repeated each hidden layer will be the other repeatedly in the back diffusion and thus be more accurate calculations, to get better results..
reasons for not learning network
Kaditm network design and training, but ultimately fail! And causes of failure in fact many, including:
• coefficient of education is not appropriate.
• the training group were not selected carefully.
• The quality of the network to the application that disproportion trained him, and this requires the selection of another network according to the desired application.
• weights that starts by the network is appropriate.
• base conversion is not appropriate.
• the number of processing units is appropriate.

5.1. Noise Removal

Noise is a collection of random signals and unwanted in their presence because they generate errors.
To get rid of the noise using filter Filter is that controls the traffic signals and to design an appropriate candidate must know the frequency noise is it higher than the sound frequency or less If the top can be removed by Miaraf candidate pass the low-frequency low pass filter and, if less can use a filter Tmreraltrddat senior high pass filter.
It uses low-pass filter frequency after an initial phase to get the audio signal, in order to delete the vehicle high frequency overlapped with the reference compounds, which are noisy.

5.2 Silence Removal

It is natural to have periods of silence to speak before the start of speech signal when recording audio. The discovery of the starting point of the audio signal used to obtain the region of importance in the sound where it is to remove the silence that precedes the word which is followed by silence, by deleting the silent or Allassotah each region of the word.
Therefore, set the start and end of the sample used to obtain the region of interest in the sound where silence is removed before the word and figure (8.2) illustrates the audio signal before you delete the silence them and shape) (7shows after deleting the reference sound of silence from them.

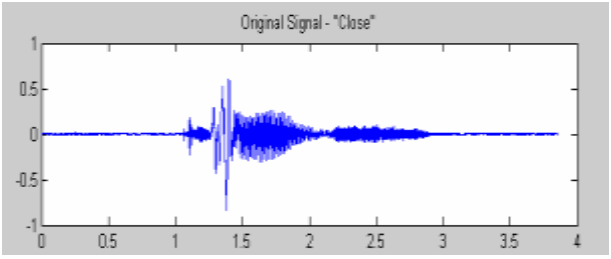


Figure (7) illustrates the audio signal before you delete the silence them.

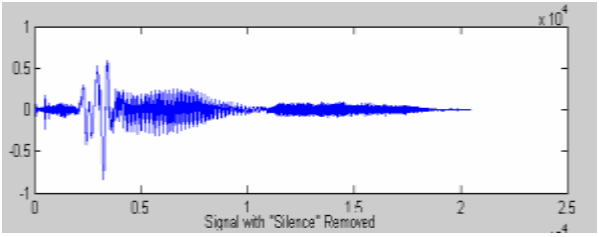


Figure (8) shows after deleting the reference sound of silence.

5.3 Features Extraction

The phase extraction characteristics are very important stage because they significantly affect the efficiency of the verification system and where to draw landmarks to indicate the sound so as to obtain a set of properties it is assumed that these characteristics vary from person to person, and of course all Mazdat those properties which are extracted sound whenever the recognition process easier and less error rate, then the shortfall after the extract is then inserted so that transactions are applied to the artificial neural network. I have used MFCC in this project is to extract the characteristics of a sound conversion Msofh points in the time domain to the frequency of this method is the most famous techniques in the extraction properties of the vote verification systems to speak.

6 Results

The following tables show the results obtained from the education program and testing (when teaching the neural network and artificial) for different values of the input (number of people, the number of forms for each person, coefficient of education, the number of hidden layers and the number of nodes in the hidden layer), and the results were as follows:

6.1 Education Network

The first test:

Table (1) show the eduction and the efficiency of network performance when the number of people2 the number offorms 2 thenumber of hidden layers 2 ' sama alive error (1e-10);

No:ofcouress	No; ofnodes in thehidden layer	No; of courses at (0.1)for the coefficient	No; of courses of education at (0.5)	Time of education (0.1)	Minutes at atime ofeduction (0.5)	Results
9000	200	941	2984	2.56	2.91	The LEARNED
2500	120	322	361	2.12	2.50	The LEARNED
1500	250	993	4199	3.93	4.98	The LEARNED
3000	180	826	1473	2.41	2.59	The LEARNED
1500	150	610	1297	2.00	2.33	The LEARNED
2000	230	985	3333	3.88	3.00	The LEARNED
6000	100	137	222	1.98	1.99	The LEARNED

Table (1) shows the first Education Network

The second test:

Table (2) shows the education and the efficiency and network performance when the number of people 2, the number of forms 4, the number of hidden layers 2, Sama alive error (10-1e).

No; of courses	No; of nodes in the hidden layer	No; of courses at (0.1)fo thecoefficient	No; of courses of education at (0.5) for the coefficient	Time of education (0.1)	Minutes at a time ofeducation (0.5)	Results
8000	250	9522	3993	6.85	7.98	The LEARNED
1600	120	224	1454	4.92	5.98	The LEARNED
4900	200	1724	3452	5.53	6.94	The LEARNED
3000	180	550	2232	5.12	6.41	The LEARNED
1500	100	102	1362	4.71	5.98	The LEARNED
7000	230	1925	3828	6.23	7.88	The LEARNED
1700	150	487	1658	4.99	5.97	The LEARNED

Table (2) shows the second network education

Third test:

Table (3) shows the education and the efficiency and performance of the network when you have: the number of people 5, the number of forms 6, the number of hidden layers 2, Sama alive error (10-1e).

No;of courses	No; nodes in the hidden layer	No; of courses at (0.1)	No:of courses at (0.5) for the coefficient	Time of aducation (0,1)	Minutes of a time of education	Resuits
402900	200	402900	402900	17.43	18.04	NO LEARNED
126500	120	126500	126500	12.52	15.44	NO LEARNED
889000	250	889000	889000	22.00	23.01	NO LEARNED
300000	180	300000	300000	17.02	17.51	NO LEARNED
137100	150	137100	137100	14.59	16.49	NO LEARNED
701200	230	701200	701200	18.03	20.18	NO LEARNED
105500	100	105500	105500	10.42	15.03	NO LEARNED

Table (3) The third shows the education network

The network: Test

After the process has to teach the neural network models of different voices, you will be in this part of the network performance test to identify the models sounds that are not teaching them the network, and the efficiency of the test is determined using the following equation:

Test the efficiency = (number of votes that have been identified / number of votes entered) * 100.

The first network:

No; of people	No; of models	No; of votes	No; of votes entered	efficiency test Identified
2	2	1	1	%100

Table (4) shows the results of the first test

The number of people the number of models the number of votes entered the number of votes Identified efficiency test

Second network:

The number of people the number of models the number of votes entered the number of votes Identified efficiency test

No; of people	No; of models	No; of votes	No; of votes entered	efficiency test Identified
2	4	3	3	%100

Table (5.) Shows the results of the second test

The third network:

The number of people the number of models the number of votes entered the number of votes Identified efficiency test

No; of people	No; of models	No; of votes	No; of votes entered	efficiency test Identified
6	24	28	3	%10

Table (6) shows the results of the third test

7. Conclusions

Through the design of the system and see the results we have obtained from the implementation of this system has been reached to a set of conclusions may be summarized as follows:

1. The design and implementation of a system to verify the speaker is not easy at all, but if Aldhirash and research in verification systems is extensive and well and provide enough time, there may be a kind of simplicity in design and construction of this type of system.

-2 To write the program language of MATLAB for all phases of the system make the record sound stage and treatment is somewhat easy.

-3 At increasing the number of nodes in the hidden layer, the accuracy of discrimination is increasing, but that Istmrany a certain extent after which the process of education in the very slow decrease with precision.

-4 When the coefficient of education small, this leads to the slow process of education, the greater the coefficient of education, this leads to instability of the system.

REFERENCES

1-Joe Tebelskis ,"Speech Recognition Using Neural Networks", School of Computer Science,Carnegie mellon University ,pittsborgh, Pennsylvania .

2-Anthony Zaknich ,"Principles of Artifical Neural Networks: for Scientists, Engineers and other ",Fremantle , western Austalia .

3-Ronald S.Burns ," Advanced Control Engineering " Department of Mecanical and Marine .

4- Automatic Person Verification Ashy Speech and Face Information ,School of microelectronic Ery,Grffith University, 2002.

5-Sanjoy Sharma ,"Digital Signal Processing ",Macrch,2007.

6-Ling Feg,"Speaker Recognition ",Technical University of Denmark ,2004.

7- Gerik Alexander,Von Graeventiz,"About Speaker Recognition Technology" , Bonn,Germany.

8-todor dimitrov Ganchev , "speaker Recognition", University of Patras,Greece,November 2005.

9-Scott chin,Kelvin Lav,Lindsey Lev, "A speaker Verification System ", Department of Electrical and Computer Engineering ,University of Victoria,2002.

10-Antti Eronen,"AUTOMATIC MUSICAL INSTRUMENT RECOGNITION",Department of Information Technology,2001