

Chapter 5

Neural network Analysis

5.1 Neural Network Approach for Vowel Recognition

Vowels are of different percentages of power in different frequency bands. The percentages of power in each frequency band are found in the spectrogram analysis method. The spectrogram for each vowel for a number of normal speakers, are used to extract vowel characteristics. Using those percentages, templates for each vowel can be designed.

The same ratios of power contents in 16 different frequency bands can be used for the neural network application.

5.2 Selection of a suitable Neural Network

A neural network with a single-layered, feed forward structure and error gradient-based training algorithm, is known as the 'Perceptron'[21].

A multi-layer perceptron (MLP) represents a generalization of the perceptron. MLPs have been applied successfully to resolve many complex real-world problems, which consist of non-linear decision boundaries[22]. The MLP network is trained in a supervised manner with the error back-propagation algorithm, which is based on the error correction learning rule.

Multi-layered neural networks usually consist of many neuron units, each operating on a sigmoid activation function. These units are arranged into several layers. The input layer receives the signal from the outside, and sends its output to the first hidden layer. The output layer receives the signal from the last hidden layer and produces the final output. There may be several hidden layers[33]. Layers with neurons whose outputs are not directly accessible are called hidden layers.

The sigmoid function is derivative and the back-propagation learning algorithm can be derived and is applicable.

Error back-propagation algorithm may be viewed as a generalization of an adaptive filtering algorithm: the delta rule[22].

5.3 Designing a Multi-Layered Neural Network for Vowel Recognition

A 3 layer MLP is used to train the selected 5 vowel sounds. Power ratios of 16 frequency bands are used as inputs. Therefore the number of input nodes is taken as 16. Number of output layer neurons are 5 because there are 5 outputs which represent the 5 vowel sounds. A hidden layer is also used, which collectively makes the network a 3 layered MLP.

Initially 8 neurons were used for the hidden layer and the training procedure was not successful. The network was not converging.

If the number of input nodes is n , the number of hidden layer neurons can be taken as $(2n+1)$ according to the Kolmogorov theorem[22]. Therefore the hidden layer neurons were increased to 33.

This is a fully connected network shown in Figure 5.1.

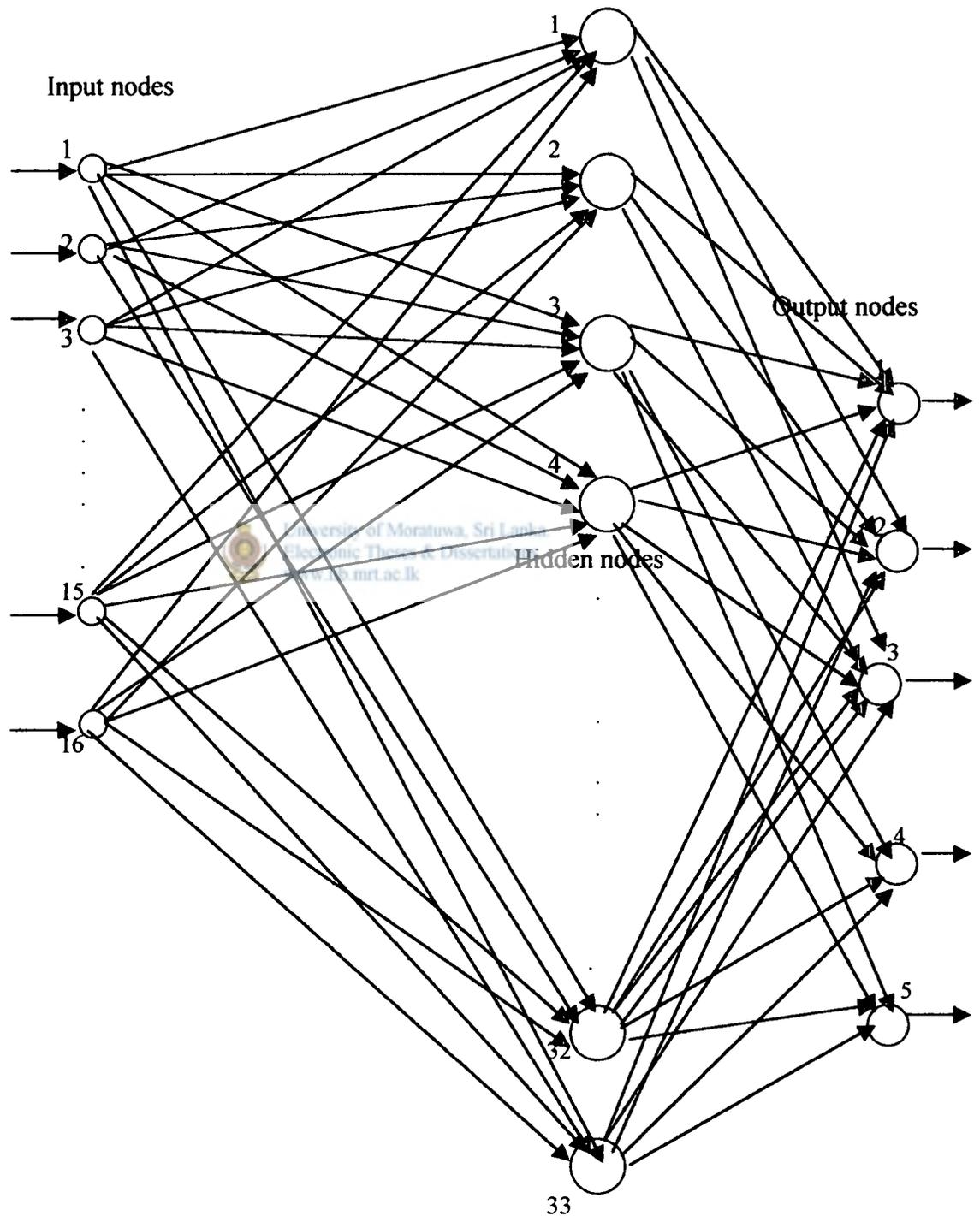


Figure 5.1. Structure of the 3 layer Feed Forward Neural Network

Inputs, outputs and the weights of the network can be shown as in Figure 5.2 where **X** and **O** are input and output respectively. **S** is the input to the hidden layer and **Y** is the output from the hidden layer. **U** is the synaptic input to the output layer. **V** and **W** are weight values of the network.

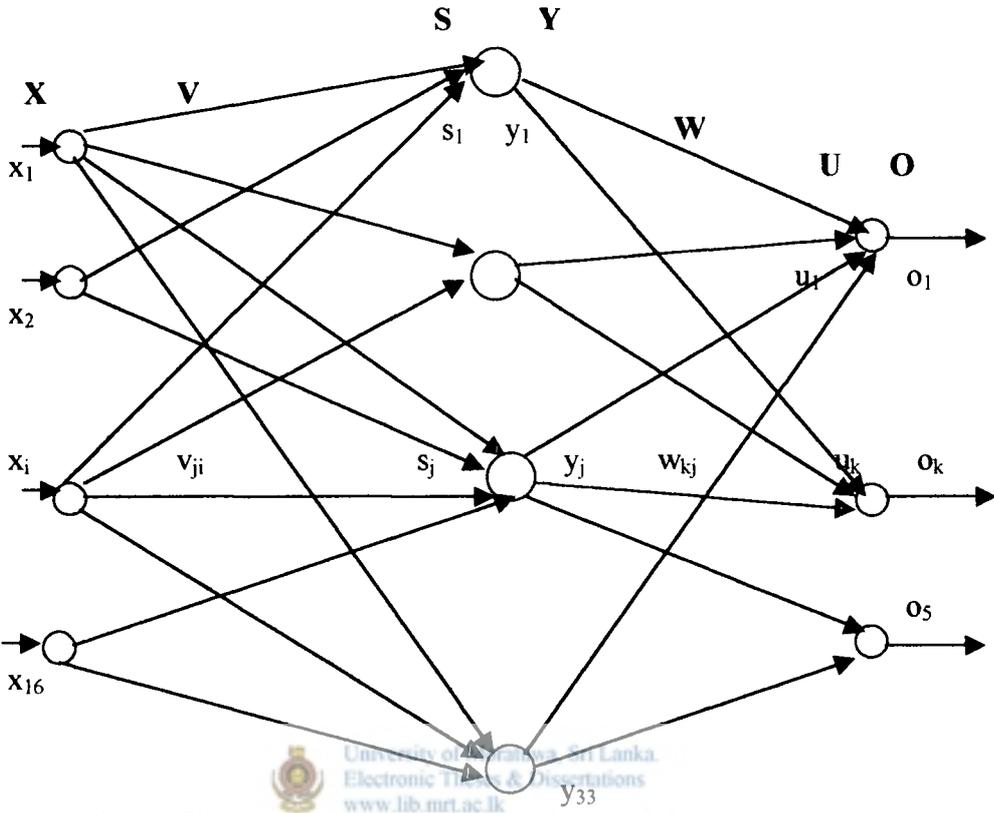


Figure 5.2. Inputs, outputs and weights of the Network

The output of the j^{th} neuron in the hidden layer and the activation of the k^{th} neuron in the output layer of the network are denoted by y_j and o_k , respectively. The weight w_{kj} connects the output of the j^{th} neuron in the hidden layer with the input to the k^{th} neuron in the output layer. The weight v_{ji} connects the i^{th} input node with the input to the j^{th} neuron of the hidden layer.

The total synaptic input to the output layer $U=(u_1 \ u_2 \ \dots \ u_k)^T$ can be expressed using the vector notation,

$$U = WY$$

where

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ u_5 \end{pmatrix} \quad W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{133} \\ w_{21} & w_{22} & \dots & w_{233} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ w_{51} & w_{52} & \dots & w_{533} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_{33} \end{pmatrix}$$



When the activation function of the k^{th} neuron at the output layer is o_k

$$O = \begin{pmatrix} o_1 \\ o_2 \\ \cdot \\ \cdot \\ o_5 \end{pmatrix} = \begin{pmatrix} f(u_1) \\ f(u_2) \\ \cdot \\ \cdot \\ f(u_5) \end{pmatrix}$$

The desired output can be expressed as

$$D = \begin{pmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ d_5 \end{pmatrix}$$

Input to the hidden layer can be given in vector notation as

$$S = VX$$

where

$$S = \begin{pmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ s_{33} \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{16} \\ v_{21} & v_{22} & \dots & v_{26} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{331} & v_{332} & \dots & v_{3316} \end{pmatrix} X = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{16} \end{pmatrix}$$

5.4 Selection of Sigmoidal as Activation Function

The total synaptic input will be transformed to the output activation via a nonlinear activation function (transfer function). Threshold function, ramp function and sigmoid function are often employed in neural networks.

Sigmoidal is the most pervasive and biologically plausible activation function[22]. It is also the most frequently used activation function. It is nondecreasing and can limit the output to a range of (0,1) [23].

There are two types of sigmoid functions. For unipolar sigmoid function, the output is

$$g(u) = a / (1 + \exp(-bu))$$

where a and b are real constants.

For bipolar sigmoid function, the output is

$$g(u) = a \frac{1 - \exp(-bu)}{1 + \exp(-bu)}$$

In a sigmoid function 'a' denotes the gain or the amplitude of the transfer function and 'b' denotes the slope or steepness parameter of the transfer function.

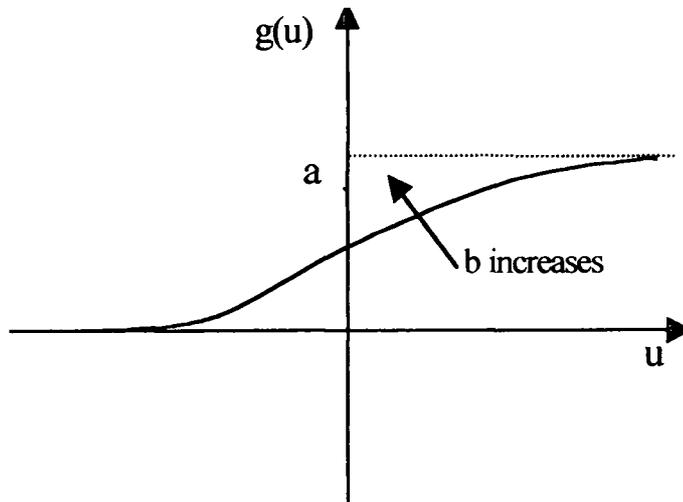


Figure 5.3. Unipolar Sigmoidal Function

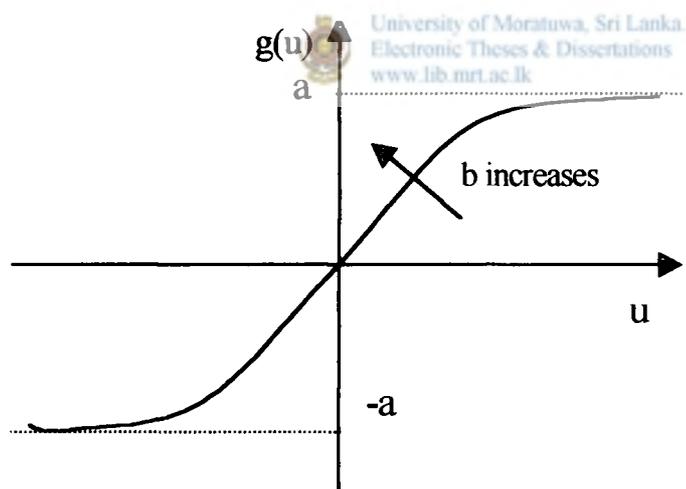


Figure 5.4. Bipolar Sigmoidal Function

Unipolar sigmoid is shown in Figure 5.3 and bipolar sigmoid is shown in Figure 5.4. When 'a' increases the amplitude of the transfer function increases. Slope of the curve increases with the increase of 'b'. The sigmoid function is differentiable.

If the output activation is given by $O = a / (1 + \exp(-bu))$

(a unipolar sigmoid)

Derivative of O can be proved as , Der O = $bO[1 - (O/a)]$

where Der O = $\partial O / \partial u$.

If the output activation is given by $O = a (1 - \exp(-bu)) / (1 + \exp(-bu))$
 (a bipolar sigmoid)
 Derivative of O can be proved as , $\text{Der } O = ab[1 - (O/a)^2]$.

Using generalized delta learning rule weight updating formula for the output layer can be written in matrix form:

$$W^{\text{new}} = W^{\text{old}} + \alpha \delta_o Y^T$$

where α is the learning factor
 and

$\delta_o = (\delta_{o1} \ \delta_{o2} \ \dots \ \delta_{o5})^T$; o represents the output layer.

δ_{ok} is the error signal term produced by the k^{th} neuron of the output layer.

$\delta_{ok} = (d_k - o_k)$. $\text{Der } o_k$, for $k = 1, 2, \dots, 5$

For the hidden layer, weight adjustment can be given as

$$V^{\text{new}} = V^{\text{old}} + \alpha \delta_y X^T$$

where

$\delta_y = (\delta_{y1} \ \delta_{y2} \ \dots \ \delta_{y33})^T$; y represents the hidden layer.

δ_{yj} is the error signal term produced by the j^{th} neuron of the hidden layer.

$\delta_{yj} = \text{Der } S. \sum_{k=1}^5 \delta_{ok} w_{kj}$, for $j = 1, 2, \dots, 33$



5.5 Training procedure of the network

The following flow chart shows the back-propagation learning procedure.

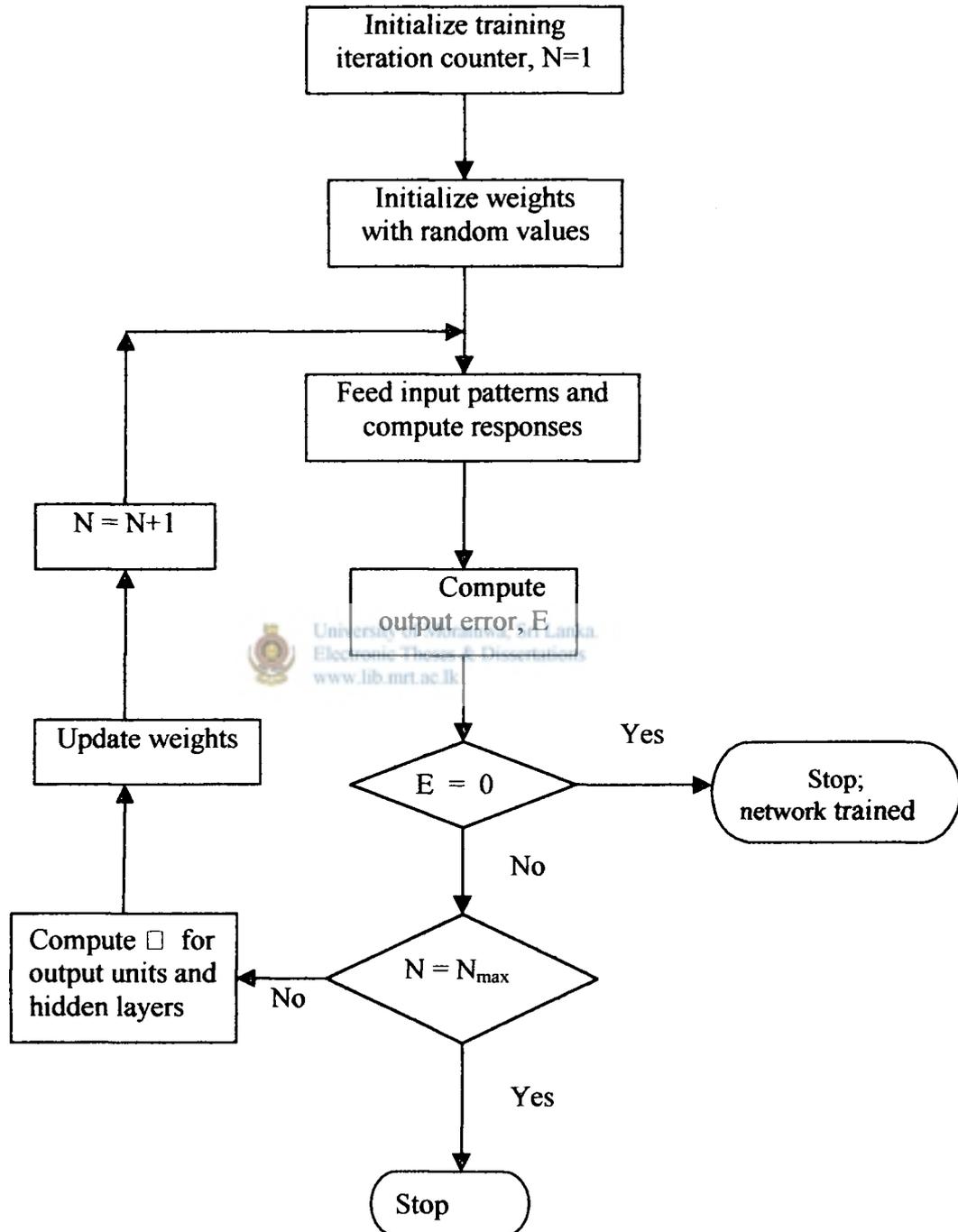


Figure 5.5. Flow Chart for learning algorithm

A total of 210 normal speaker samples, consisting 42 samples of each vowel were used for training the network.

The desired outputs were taken as,

$$\begin{aligned} O_1 &= [1\ 0\ 0\ 0\ 0]^T, \\ O_2 &= [0\ 1\ 0\ 0\ 0]^T, \\ O_3 &= [0\ 0\ 1\ 0\ 0]^T, \\ O_4 &= [0\ 0\ 0\ 1\ 0]^T, \\ O_5 &= [0\ 0\ 0\ 0\ 1]^T \end{aligned}$$

for /a:/, /æ/, /u:/, /o/ and /eI/ sounds respectively.

Taking a=1, b=1 and unipolar sigmoid for hidden layer and output layer, trained the network using random values as initial weights (*Appendix [H(i)]*). The learning factor used was 0.6. The network was not converging even though the number of iterations was greater than 100,000 times.

Keeping the same a and b values, used bipolar sigmoid for each layer. Again the results were not satisfactory because the weight values were far away from convergence.

The network showed its convergence when the learning factor was reduced to a value less than 0.1. In order to achieve convergence it was required to reduce the training parameter gradually.



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Having understood the behavior and the usage of training factor, trained the network for different values of b for hidden and output layers. Unipolar and bipolar activation functions and the combinations of the two were used to find the best performance of the neural network.

For each case the weight values after convergence (*Appendix [H(ii)]*), were used to test the program for vowel recognition.

The training program and the testing program were developed using MATLAB (*Appendix [I]*).

5.6 Testing the neural network

The same input sample values used for training were used for testing the network.

Results were calculated by applying the following formula.

$$\text{No. of correct decisions} = \left(\begin{array}{c} \text{Correct pronunciation and correct test results} \\ + \\ \text{Incorrect pronunciation and incorrect test results} \end{array} \right)$$

For different values of b and different types of activation functions the results for 5 vowel sounds can be summarized as shown in Table 5.1.

Hidden layer			Output layer			Test results				
Activation function	a	b	Activation function	a	b	/a:/	/ae/	/u:/	/o/	/eI/
unipolar	1.0	0.5	unipolar	1.0	1.0	87.50%	100%	97.37%	62.50%	47.22%
unipolar	1.0	0.8	unipolar	1.0	1.1	72.50%	95.55%	60.52%	40.63%	36.11%
unipolar	1.0	0.9	unipolar	1.0	1.2	-	-	-	-	41.25%
unipolar	1.0	1.0	unipolar	1.0	0.5	90.00%	96.55%	100%	59.38%	36.11%
unipolar	1.0	1.0	unipolar	1.0	1.0	90.00%	96.55%	94.74%	75.00%	41.25%
unipolar	1.0	1.0	unipolar	1.0	1.1	-	-	-	-	30.56%
unipolar	1.0	1.0	unipolar	1.0	1.2	90.00%	96.55%	86.84%	75.00%	52.78%
unipolar	1.0	1.0	unipolar	1.0	1.3	-	-	-	-	38.89%
unipolar	1.0	1.1	unipolar	1.0	1.2	-	-	-	-	25.00%
unipolar	1.0	1.2	unipolar	1.0	1.2	82.50%	100%	52.63%	56.25%	41.67%
unipolar	1.0	0.5	bipolar	1.0	0.5	65.00%	93.10%	15.79%	34.38%	83.33%
unipolar	1.0	0.5	bipolar	1.0	0.7	87.50%	96.55%	23.68%	52.50%	77.78%
unipolar	1.0	0.5	bipolar	1.0	1.0	82.50%	96.55%	2.63%	25.00%	100%
bipolar	1.0	0.5	bipolar	1.0	0.3	87.50%	96.55%	42.11%	71.88%	66.67%
bipolar	1.0	0.5	bipolar	1.0	0.5	90.00%	96.55%	60.52%	81.25%	38.89%
bipolar	1.0	0.5	bipolar	1.0	0.8	87.50%	96.55%	42.11%	71.88%	69.44%
bipolar	1.0	0.5	bipolar	1.0	1.0	87.50%	96.55%	71.05%	84.38%	44.44%
bipolar	1.0	1.0	bipolar	1.0	0.5	90.00%	96.55%	47.37%	90.63%	22.22%
bipolar	1.0	1.0	bipolar	1.0	1.0	90.00%	96.55%	55.26%	84.38%	25.00%

Hidden layer			Output layer			Test results				
Activation function	a	b	Activation function	a	b	/a:/	/ae/	/u:/	/o/	/eI/
bipolar	1.0	1.0	unipolar	1.0	1.1	90.00%	96.55%	92.11%	75.00%	55.56%
bipolar	1.0	1.0	unipolar	1.0	1.2	97.50%	96.55%	76.32%	78.13%	61.11%
bipolar	1.0	1.0	unipolar	1.0	1.3	90.00%	96.55%	92.11%	78.13%	66.67%
bipolar	1.0	1.0	unipolar	1.0	1.4	92.50%	100%	92.11%	75.00%	72.22%
bipolar	1.0	1.0	unipolar	1.0	1.5	87.50%	100%	92.11%	75.00%	75.00%
bipolar	1.0	1.0	unipolar	1.0	1.6	87.50%	100%	92.11%	75.00%	77.78%
bipolar	1.0	1.0	unipolar	1.0	1.7	87.50%	100%	94.74%	75.00%	88.89%
bipolar	1.0	1.0	unipolar	1.0	1.8	87.50%	100%	94.74%	78.13%	91.67%
bipolar	1.0	1.0	unipolar	1.0	1.8 5	87.50%	100%	94.74%	78.13%	94.44%
bipolar	1.0	1.0	unipolar	1.0	1.9	85.00%	100%	94.74%	78.13%	94.44%
bipolar	1.0	1.0	unipolar	1.0	1.9 5	85.00%	100%	94.74%	78.13%	94.44%
bipolar	1.0	1.0	unipolar	1.0	1.9 8	85.00%	100%	94.74%	78.13%	94.44%
bipolar	1.0	1.0	unipolar	1.0	2.0	82.50%	100%	94.74%	78.13%	91.67%
bipolar	1.0	1.0	unipolar	1.0	2.0 5	90.00%	100%	94.74%	75.00%	94.44%
bipolar	1.0	1.0	unipolar	1.0	2.1	90.00%	100%	94.74%	75.00%	94.44%
bipolar	1.0	1.0	unipolar	1.0	2.2	90.00%	100%	94.74%	75.00%	94.44%
bipolar	1.0	1.0	unipolar	1.0	2.5	90.00%	100%	94.74%	78.13%	94.44%
bipolar	1.0	1.0	unipolar	1.0	3.0	90.00%	100%	94.74%	78.13%	97.22%
bipolar	1.0	1.0	unipolar	1.0	5.0	90.00%	100%	94.74%	78.13%	97.22%

Table 5.1. Summary of Test Results

According to the results in Table 5.1 it is evident that the hidden layer requires a bipolar sigmoid while output layer requires a unipolar sigmoid.

Test results were the best for a bipolar function with $a=1.0$ and $b=1.0$ for the hidden layer and for a unipolar function with $a=1.0$ and $b=5.0$ for the output layer.

Using above values test results for hearing impaired samples are shown in Table 5.2.

Vowel sound	No. of Pronunciation OK and Neural network test 'Passed'	No. of Pronunciation Bad and Neural network test 'Failed'	No. of Correct Decisions	% of Correct Decisions
/a:/	9/45	20/45	29/45	64.44%
/æ/	8/44	29/44	37/44	84.09%
/u:/	3/43	21/43	24/43	55.81%
/o/	7/43	19/43	26/43	60.47%
/eI/	7/43	30/43	37/43	86.05%

Table 5.2

Using bipolar sigmoid activation function with $a=1.0$ and $b=1.0$ for the hidden layer and unipolar sigmoid activation function with $a=1.0$ and $b=1.85$ for the output layer, results for hearing impaired samples are shown in Table 5.3.

Vowel sound	No. of Pronunciation OK and Neural network test 'Passed'	No. of Pronunciation Bad and Neural network test 'Failed'	No. of Correct Decisions	% of Correct Decisions
/a:/	9/45	20/45	29/45	64.44%
/æ/	8/44	30/44	38/44	86.36%
/u:/	3/43	21/43	24/43	55.81%
/o/	7/43	18/43	25/43	58.14%
/eI/	6/43	30/43	36/43	83.72%

Table 5.3

Using unipolar sigmoid activation function with $a=1.0$ and $b=1.0$ for the hidden layer and unipolar sigmoid activation function with $a=1.0$ and $b=1.2$ for the output layer, results for hearing impaired samples are shown in Table 5.4.

Vowel sound	No. of Pronunciation OK and Neural network test 'Passed'	No. of Pronunciation Bad and Neural network test 'Failed'	No. of Correct Decisions	% of Correct Decisions
/a:/	9/45	19/45	28/45	62.22%
/æ/	7/44	30/44	37/44	84.09%
/u:/	3/43	24/43	27/43	62.79%
/o/	8/43	17/43	25/43	58.14%
/eI/	5/43	30/43	35/43	81.39%

Table 5.4

Using bipolar sigmoid activation function with $a=1.0$ and $b=0.5$ for the hidden layer and bipolar sigmoid activation function with $a=1.0$ and $b=1.0$ for the output layer, results for hearing impaired samples are shown in Table 5.5.

Vowel sound	No. of Pronunciation OK and Neural network test 'Passed'	No. of Pronunciation Bad and Neural network test 'Failed'	No. of Correct Decisions	% of Correct Decisions
/a:/	9/45	19/45	28/45	62.22%
/æ/	7/44	28/44	35/44	79.55%
/u:/	3/43	31/43	34/43	79.07%
/o/	8/43	14/43	22/43	51.16%
/eI/	3/43	30/43	33/43	76.74%

Table 5.5

When considering the results in Table 5.1,

/a:/ has the highest % when $a=1.0$ and $b=1.0$ (bipolar)-----hidden layer
and $a=1.0$ and $b=1.2$ (unipolar)-----output layer (97.50%)

/æ/ has the highest % when $a=1.0$ and $b=0.5$ (unipolar)-----hidden layer
and $a=1.0$ and $b=1.0$ (unipolar)-----output layer (100%)

/u:/ has the highest % when $a=1.0$ and $b=1.0$ (unipolar)-----hidden layer
and $a=1.0$ and $b=0.5$ (unipolar)-----output layer (100%)

/o/ has the highest % when $a=1.0$ and $b=1.0$ (bipolar)-----hidden layer
and $a=1.0$ and $b=0.5$ (bipolar)-----output layer (90.63%)

/eI/ has the highest % when $a=1.0$ and $b=0.5$ (unipolar)-----hidden layer
and $a=1.0$ and $b=1.0$ (bipolar)-----output layer (100%)

Using above combinations for each vowel, percentages for hearing impaired speaker samples were found. Results are shown in Table 5.6.

Vowel sound	No. of Pronunciation OK and Neural network test 'Passed'	No. of Pronunciation Bad and Neural network test 'Failed'	No. of Correct Decisions	% of Correct Decisions
/a:/	11/45	19/45	30/45	66.67%
/æ/	7/44	29/44	36/44	81.82%
/u:/	3/43	20/43	23/43	53.49%
/o/	8/43	13/43	21/43	48.84%
/eI/	8/43	16/43	24/43	55.81%

Table 5.6



According to the above results, test results were the best for a bipolar function with $a=1.0$ and $b=1.0$ for the hidden layer and for a unipolar function with $a=1.0$ and $b=5.0$ for the output layer.

A summary of results for percentage of correct decisions using neural network method is shown in Table 5.7.

Sound	Speaker Type	No.of Pronunciation OK and Algorithm 'Passed'	No.of Pronunciation Bad and Algorithm 'Failed'	No.of Correct Decisions	% of Correct Decisions
/a:/	Normal	36/40	0/40	36/40	90.00
	Hearing Impaired	9/45	20/45	29/45	64.44
/æ/	Normal	29/29	0/29	28/29	100
	Hearing Impaired	8/44	29/44	37/44	84.09
/u:/	Normal	36/38	0/38	36/38	94.74
	Hearing Impaired	3/43	21/43	24/43	55.81
/o/	Normal	25/32	0/32	25/32	78.13
	Hearing Impaired	7/43	19/43	26/43	60.47
/eI/	Normal	35/36	0/36	35/36	97.22
	Hearing Impaired	7/43	30/43	37/43	86.05

Table 5.7. Summary of results for neural method