

Chapter 4: Results and Discussion

4.1 Introduction

This chapter devoted for presenting analyzing the results of the performed experiments. The results are divided into FFBP and Elman NNs results, and each network type was subdivided into single-frame and multi-frame data results. After that, the chapter discusses the major contributions of the findings to the vowel recognition area.

4.2 FFBP Results

4.2.1 Single-frame

Ten experiments were conducted for each data frame at different hidden neurons number for the three sets of data. Table 4.1 shows the highest recognition rates obtained for different single-frame data of the three sets using FFBP network. Also, the table displays the average recognition rates of the three sets for each frame according to 3-fold cross validation technique.

Table 4.1: Recognition rates (%) obtained for different single-frame data using FFBP network.

Recognition Rate (%) Signal Length (ms)	Data Set			Average
	Set 1	Set 2	Set 3	
10	78.47	79.86	78.19	78.84
15	81.94	80.56	81.81	81.44
20	82.36	82.08	81.11	81.85
25	82.78	84.31	83.47	83.52
30	83.61	82.78	81.67	82.69
35	82.78	82.64	84.86	83.43
40	83.19	83.89	84.03	83.70
45	83.33	84.31	83.61	83.75
50	83.19	83.89	84.03	83.70
55	84.17	84.17	83.61	83.98
60	83.33	84.17	83.47	83.66
65	82.50	84.17	84.17	83.61
70	82.08	82.92	85.42	83.47

Figure 4.1 depicts the highest recognition rates of each single frame using FFBP network. It can be concluded that the highest recognition rate obtained using FFBP network is 83.98%. This value was obtained from single-frame 55ms speech data. On the other hand, the lowest recognition rate is 78.84% and was obtained from single-frame 10ms.

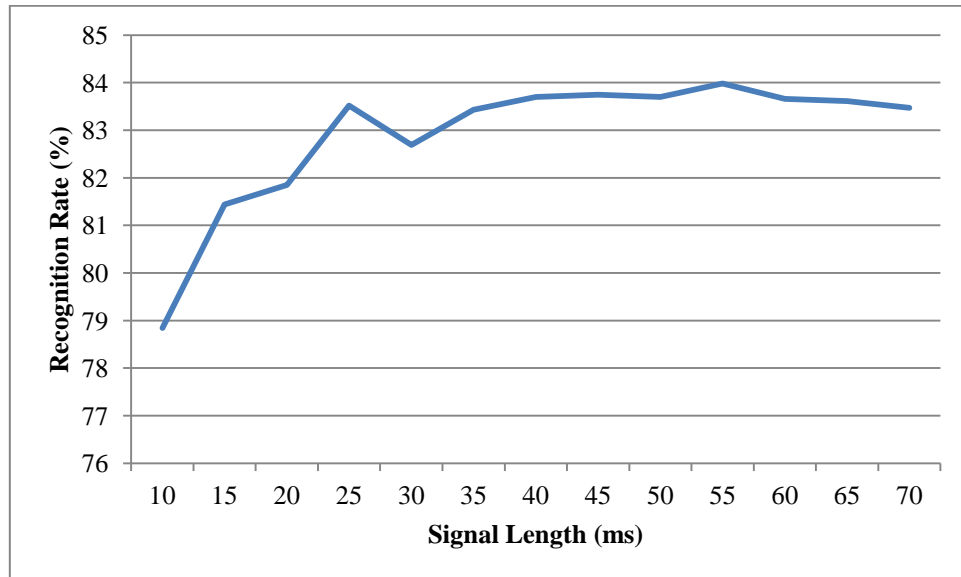


Figure 4.1: Recognition rates (%) of each single-frame speech data using FFBP network.

Different recognition rates for SF55ms speech data frame are illustrated in figure 4.2 at different hidden neuron numbers. By changing the number of neurons in the hidden layer, the accuracy of NN models was varied widely. The maximum recognition rate was achieved at 120 hidden neurons. Moreover, the lowest recognition rate was caused when using 140 hidden neurons.

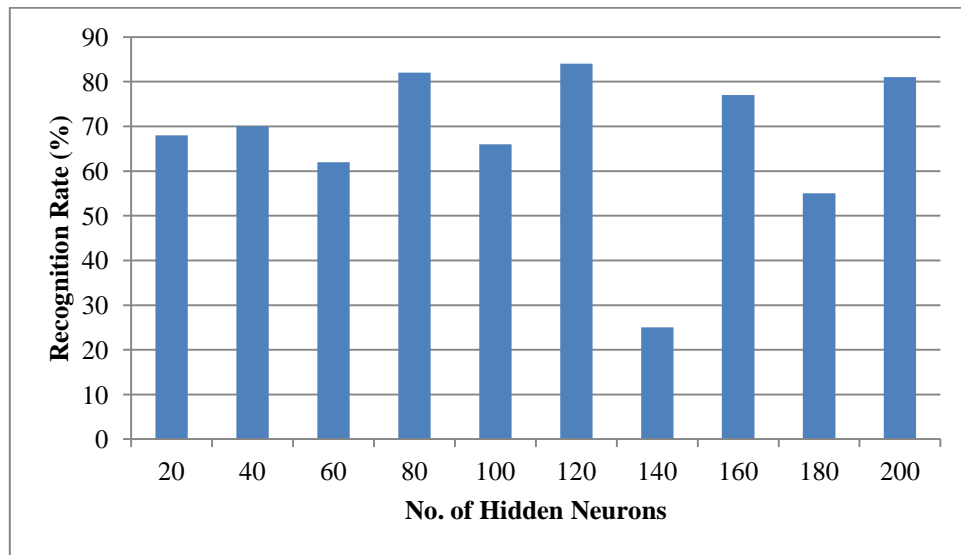


Figure 4.2: Recognition rates (%) of SF55ms speech data at different hidden neuron numbers using FFBP network.

The confusion matrix of SF55ms speech data trained using 120 hidden neurons is shown in table 4.2. The total recognition performance of the model is 84.17%. Vowels /a/ and /i/ got the highest recognition accuracy where 112 samples were correctly recognized out of 120 test samples with an accuracy of 93.33%. However, the worst vowels to be recognized were /o/ and /u/ with an accuracy of 77.50 and 63.33%, respectively. This confirms the previous studies findings where NN models confuse between /o/ and /u/ (Nong & Yunus, 2004).

Table 4.2: Confusion matrix of SF55ms speech data trained using 120 hidden neurons in FFBP.

Recognized as Vowel	/a/	/ə/	/e/	/i/	/o/	/u/	Rec. Rate (%)
/a/	112	0	6	2	0	0	93.33
/ə/	2	107	4	5	0	2	89.17
/e/	7	3	106	0	2	2	88.33
/i/	0	6	0	112	0	2	93.33
/o/	7	4	3	1	93	12	77.50
/u/	1	6	4	1	32	76	63.33
Total Recognition Rate (%)							84.17

4.2.2 Multi-frame

Table 4.3 summarizes the highest recognition rates obtained for different frame size of the three sets of data using FFBP network. Also, the table displays the average recognition rates of the three sets for each frame size according to 3-fold cross validation technique.

Table 4.3: Recognition rates (%) obtained for different frame size data using FFBP network.

Recognition Rate (%) Signal Length (ms)	Data Set			Average
	Set 1	Set 2	Set 3	
30	80.14	81.53	82.08	81.25
40	78.89	81.39	84.03	81.44
50	78.89	82.22	83.06	81.39
60	80.00	80.97	82.64	81.20
70	78.61	81.11	82.64	80.79
80	79.72	82.22	82.08	81.34
90	79.44	80.97	83.19	81.20
100	79.44	80.00	82.08	80.51

Figure 4.3 shows the recognition rates of different frame size speech signals using FFBP network. The recognition rates obtained from multi-frame data using FFBP network fluctuated gently around 81%. The highest recognition rate was 81.44% achieved by 40ms frame size. However, the lowest recognition rate was 80.51% and was obtained from data of 100ms frame size.

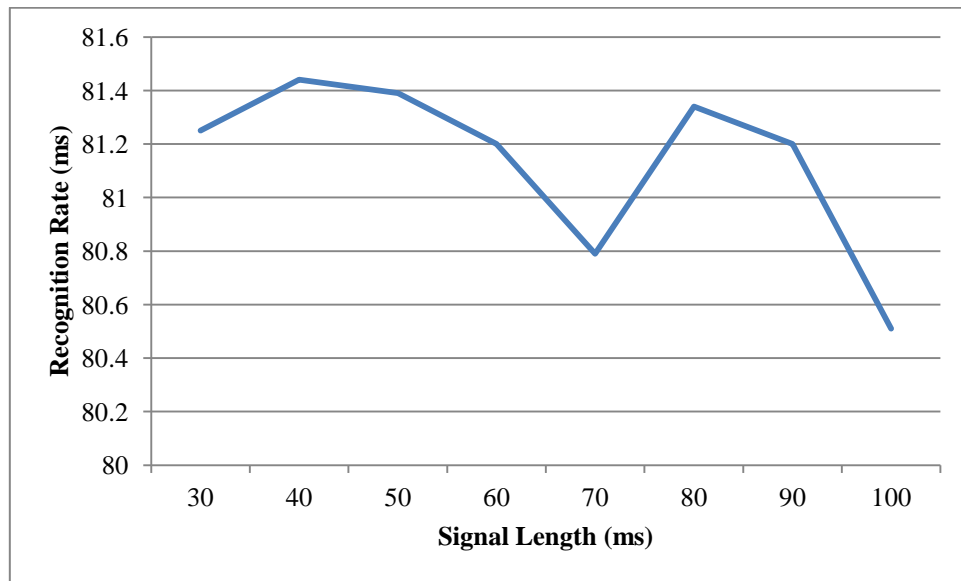


Figure 4.3: Recognition rates (%) of different frame size speech signals using FFBP network.

Single-frame speech signals provided better performance results than multi-frame speech signals; because of the efficiency difference between training algorithms implemented. LM algorithm was used for training the single-frame data; however, resilient backpropagation training function was used in training the multi-frame data due to memory requirement.

Different recognition rates of 40ms frame size data are illustrated in figure 4.4 at different hidden neuron numbers. The recognition rates range between 81 and 84% when using number of hidden neurons less than 120. However, the accuracy goes down when using more hidden neurons. This happens because of the “overfitting” phenomena where complex NN models tend to memorize the training data pattern and gives a poor performance in an external test data (Hippert, Pedreira, & Souza, 2001). The maximum recognition rate in the conducted experiments was achieved at 60 hidden neurons, while the lowest recognition rate was obtained when using 200 hidden neurons.

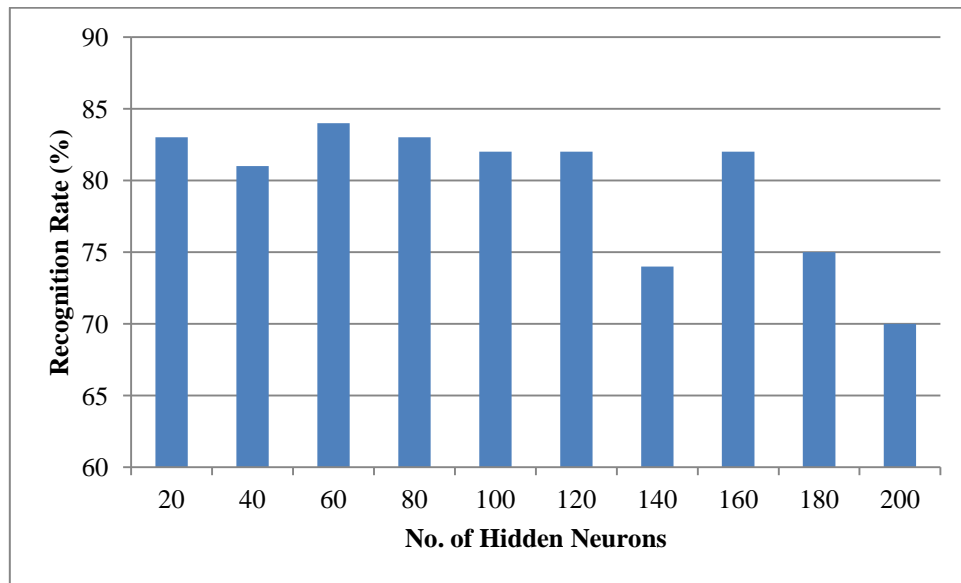


Figure 4.4: Recognition rates (%) of 40ms frame size at different hidden neuron numbers using FFBP network.

The confusion matrix of MF40ms speech data trained using 60 hidden neurons is shown in table 4.4. The total recognition performance of the model is 84.03%. Vowel /i/ got the highest recognition accuracy where 116 samples were correctly recognized out of 120 test samples with an accuracy of 96.67%. However, the vowel /o/ was confused with /u/ and got the worst recognition rate with an accuracy of 64.17%. Therefore, vowel /o/ got the lowest recognition rates for both single-frame and multi-frame signal analysis using FFBP networks.

Table 4.4: Confusion matrix of MF40ms speech data trained using 60 hidden neurons in FFBP.

Recognized as Vowel	/a/	/ə/	/e/	/i/	/o/	/u/	Rec. Rate (%)
/a/	105	1	9	0	5	0	87.50
/ə/	0	98	4	16	0	2	81.67
/e/	3	5	98	1	3	10	81.67
/i/	0	2	1	116	0	1	96.67
/o/	1	1	2	0	77	39	64.17
/u/	0	0	2	0	7	111	92.50
Total Recognition Rate (%)							84.03

4.3 Elman Results

4.3.1 Single-frame

Table 4.5 illustrates the highest recognition rates obtained for different single-frame data of the three sets using Elman network. Moreover, the table shows the average recognition rates of the three sets for each frame according to 3-fold cross validation method.

Table 4.5: Recognition rates (%) obtained for different single-frame data using Elman network.

Recognition Rate (%)	Data Set			Average
	Set 1	Set 2	Set 3	
Signal Length (ms)				
10	75.56	78.33	76.81	76.90
15	76.94	79.86	80.14	78.98
20	79.17	80.69	81.81	80.56
25	78.06	82.78	82.22	81.02
30	78.33	82.36	82.64	81.11
35	79.72	81.94	85.28	82.31
40	77.78	81.94	85.69	81.80
45	79.31	81.81	83.19	81.44
50	77.64	81.53	85.97	81.71
55	78.61	80.39	84.58	81.19
60	79.44	82.50	84.03	81.99
65	79.58	81.67	84.03	81.76
70	79.72	81.94	84.86	82.17

Figure 4.5 illustrates the highest recognition rates of each single frame using Elman network. 82.31% is the highest recognition rate obtained using Elman network and was achieved by single-frame 35ms speech data. On the other hand, the lowest recognition rate is 76.90% and was obtained from single-frame 10ms. As a result, single-frame 10ms accounts for the lowest recognition rates using FFBP and Elman networks.

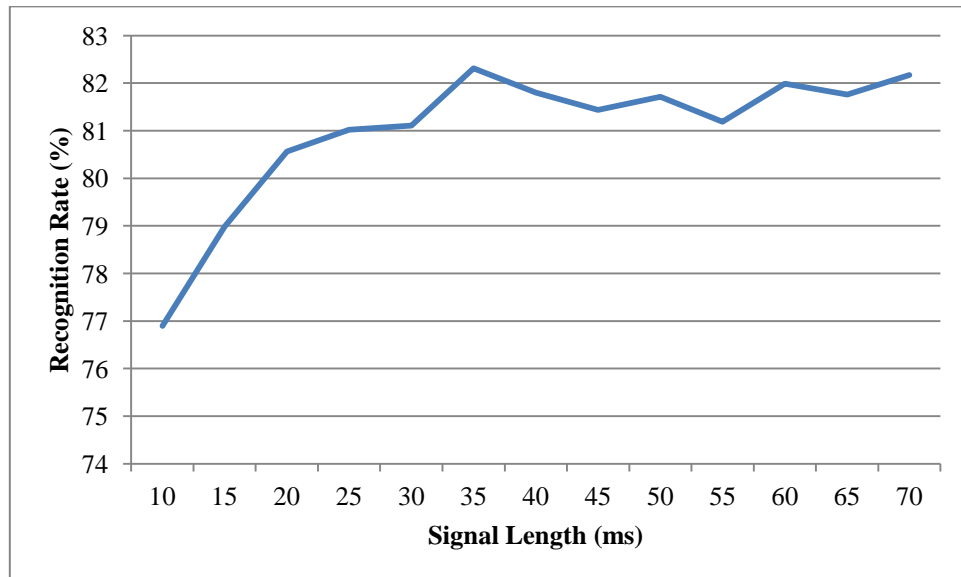


Figure 4.5: Recognition rates of each single-frame speech data using Elman network.

Different recognition rates for SF35ms speech data are illustrated in figure 4.6 at different hidden neuron numbers. The bar chart shows that the recognition rate remains at least 81% when using 100 hidden neurons or less. However, overfitting occurs when using higher number of hidden neurons making the NN model more complex. The highest recognition rate was achieved at 20 hidden neurons. Moreover, the minimum recognition rate was caused when using 140 hidden neurons. Hence, the number of 140 neurons in the hidden layer accounts for the lowest recognition rates of SF data using both FFBP and Elman networks.

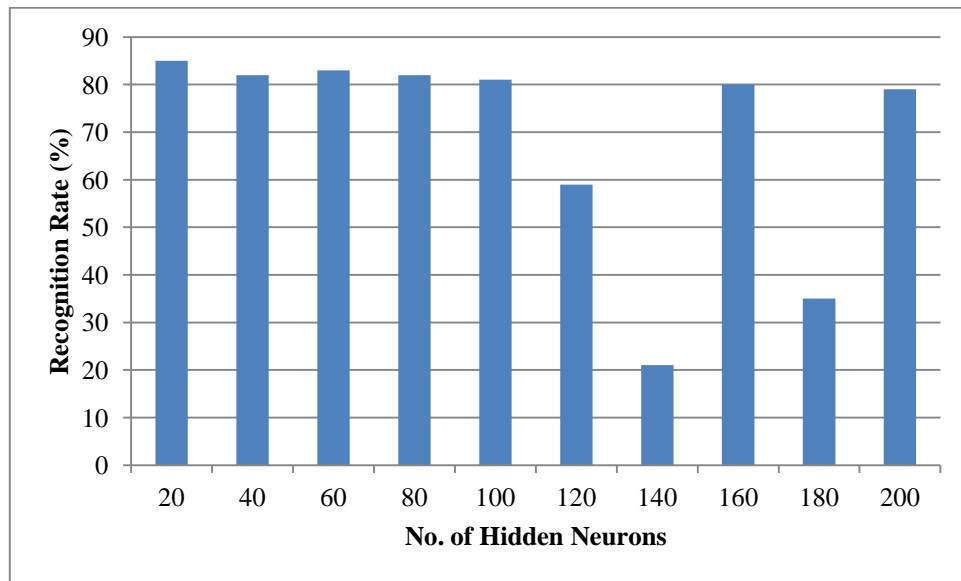


Figure 4.6: Recognition rates (%) of SF35ms speech data at different hidden neuron numbers using Elman network.

The confusion matrix of SF35ms speech data trained with 20 hidden neurons by Elman network is shown in table 4.6. The total recognition performance of the model is 85.28%. Vowel /i/ was best classified with 117 samples were correctly recognized out of 120 test samples and having an accuracy of 97.50%. However, NN models were confused between vowels /o/ and /u/, in which 38 /o/ samples out of 120 were recognized wrongly as /u/.

Table 4.6: Confusion matrix of SF35ms speech data trained using 20 hidden neurons in Elman network.

Recognized as Vowel	/a/	/ə/	/e/	/i/	/o/	/u/	Rec. Rate (%)
/a/	109	1	6	0	4	0	90.83
/ə/	0	101	6	10	3	0	84.17
/e/	2	4	107	0	1	6	89.17
/i/	0	1	0	117	0	2	97.50
/o/	1	0	4	1	76	38	63.33
/u/	0	0	7	1	8	104	86.67
Total Recognition Rate (%)							85.28

4.3.2 Multi-frame

Table 4.7 summarizes the maximum recognition rates obtained for different frame size of the three sets of data using Elman network. Moreover, the table shows the average recognition rates of the three sets for each frame size according to 3-fold cross validation technique.

Table 4.7: Recognition rates (%) obtained for different frame size data using Elman network.

Recognition Rate (%) Signal Length (ms)	Data Set			Average
	Set 1	Set 2	Set 3	
30	78.75	82.08	82.78	81.20
40	80.14	82.64	83.19	81.99
50	80.00	82.78	83.89	82.22
60	80.14	82.22	84.03	82.13
70	80.69	82.92	83.61	82.41
80	79.86	82.36	84.72	82.31
90	80.14	82.78	82.92	81.95
100	79.86	82.78	85.00	82.55

Figure 4.7 shows the recognition rates of different frame size speech signals using Elman network. All multi-frame data have recognition rates above 81%. The highest recognition rate was 82.55% and achieved by speech data of 100ms frame size.

Different recognition rates of 100ms frame size data are illustrated in figure 4.8 at different hidden neuron numbers. All the NN models achieved recognition rates above 82%. The highest performance occurred with 80 hidden neurons.

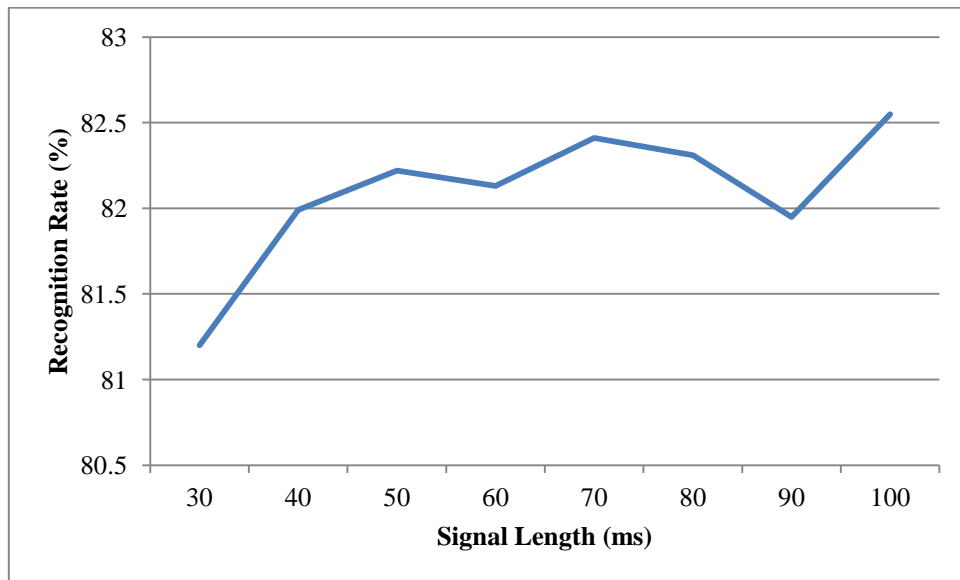


Figure 4.7: Recognition rates (%) of different frame size speech signals using Elman network.

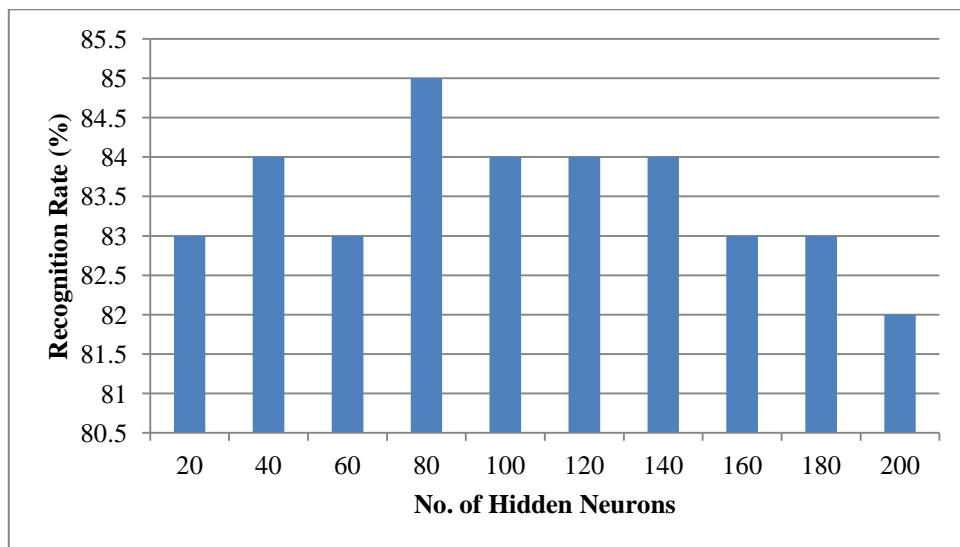


Figure 4.8: Recognition rates (%) of 100ms frame size at different hidden neuron numbers using Elman network.

The confusion matrix of MF100ms speech data trained using 80 hidden neurons is depicted in table 4.8. The total recognition performance of the model is 85.00%. The maximum recognition rate was achieved for vowel /i/ where 116 samples were correctly recognized out of 120 test samples with an accuracy of 96.67%. On the other hand, NN models were confused between vowels /o/ and /u/, in which 39 /o/ samples out of 120 were recognized wrongly as /u/. Therefore, vowel /o/ got the lowest recognition rates for both single-frame and multi-frame signal analysis using Elman networks.

Table 4.8: Confusion matrix of MF100ms speech data trained using 80 hidden neurons in Elman network.

Recognized as Vowel	/a/	/ə/	/e/	/i/	/o/	/u/	Rec. Rate (%)
/a/	105	1	10	0	4	0	87.50
/ə/	0	106	2	9	2	1	88.33
/e/	1	6	102	0	5	6	85.00
/i/	0	1	1	116	0	2	96.67
/o/	0	0	4	0	77	39	64.17
/u/	0	0	3	0	11	106	88.33
Total Recognition Rate (%)							85.00