Chapter 5

2-stage NN Approaches for Tone Classification

As the base classification result is 72.21%, through the observation of the confusion matrix, we have found that the NN is biased during the process of training because of the different number of training data. Using tone classification in the real system the performance of tone classification still need to be improved. In this Chapter we proposed the method to improve performance through 2-stage NN framework.

In the Chapter, the experiment's results that we use the same training data for each tone are presented and compared with biased NN results first. Then the proposed method that uses 2-stage NN for tone classification is introduced. After that it's the experiments results and discussions. Finally the summary of the Chapter is concluded.

Table 5-1 The observation of biased and unbiased NN

First we did the experiments that using the equal training data for each tone with tone classification. In our original training data, tone0 have 1569 from 4506 tones. Tone1 have 1019 from 4506 tones. Tone2 have 861 from 4506 tones. Tone3 have 655 from 4506 tones. Tone4 have 402 from 4506 tones. To meet the least number, we choose 402 tones for each tone. Totally we got 2010 training data. The testing confusion-matrix is as follows:

Table 5-1 Tone classification
using same number of training data for each tone

Tone	0	1	2	3	4	Percent
0	264	25	33	38	24	68.8
1	43	118	22	28	38	47.4
2	9	7	219	19	1	85.9
3	22	10	29	127	5	65.8
4	8	12	1	10	108	77.7
						68.5

The confusion matrix using different number training data is:

Table 5-2 Confusion-matrix for biased NN

Tone	0	1	2	3	4	Percent
0	319	32	18	16	5	83.07
1	51	159	6	16	17	63.86
2	20	8	212	14	1	83.14
3	48	8	28	103	6	53.37
4	19	24	0	8	88	63.31
						72.21

To observe and compare these two matrixes, we can find that, although for the final performance the unbiased are less than the biased classifier, the performances for last 3 tones are increased when the NN are unbiased and the performances for the first two tones are reduced. The distribution of training data is shown in Table 5-3:

Table 5-3 Distribution of NN Training Data

Tone	Training Data	Percent
0	1569/4506	34.820%
1	1019/4506	22.614%
2	861/4506	19.108%
3	655/4506	14.536%
4	402/4506	8.921%

We can see that the number of training data for tone0 and tone1 is higher than other 3 tones. According to both of distribution information and comparison of confusion-matrix, we conclude that the base NN is biased by the training data.

At the same time, we analyzed the confused data in the second case. From the confusion-matrix, we know that there are 384 error points totally. It's not difficult to understand the NN output score should be closer for easy confusion tones. So if the first candidate of NN is incorrect then the second candidate should be most possible correct answer. Then during these 384 error points, when we choose the second candidate to be the answer, we get 207 from 384 error points correct. That's say about 54% error is because of the confusion with the second candidate.

5.2 Proposed Methods

Based on the analysis above, we propose to use two-layer NN to improve the tone classification performance. The first layer NN is the NN that we have done. And the first 2 candidates of NN output are chosen for selecting second layer NN. The second-layer NN include 10 NNs totally according to the combination of 5 tones in Thai language. The framework is shown in Figure 5-1.

In the figure, NN1 is the 5-tone classifier. NN2 is classifier of tone0 and tone1. NN3 is classifier of tone0 and tone2. Last NN is NN11 that is used to classify tone3 and tone4.

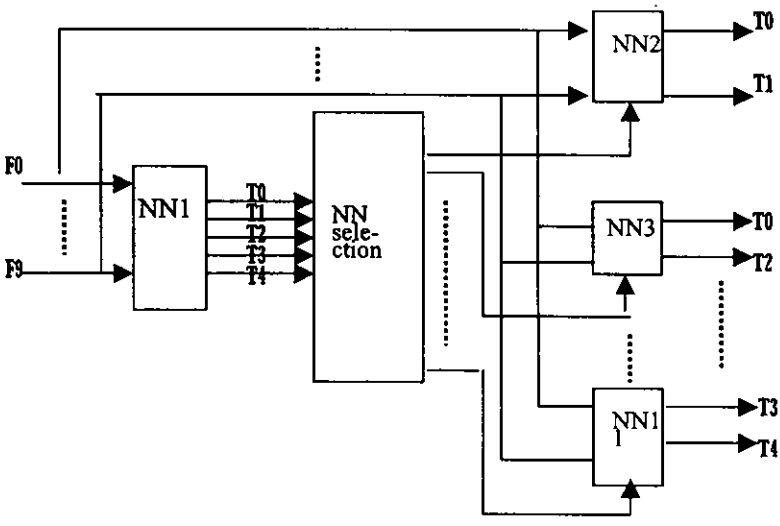

Figure 5-1 Framework 1 of 2-stage NN

5.3 Experiments and Discussions

After the experiments, we didn't find the improvements for the NN includes tone1. So we delete the 3 NNs which include tone1. Finally for this approach, we use 8 NN: first-layer one, second-layer 7. Using these 8 NNs, the output performance is increased from 881(72.21%) to 897(73.52%). From this results, we found that using the second-layer is helpful in a way. But it didn't have much improvements. The main reason that we think is because the proposed system is a series system. The error of system will be accumulated among the subsystem. The way to solve this we think is try to find a parallel system to instead.

From the confusion-matrix shown above, we found that the big confusion is happened between tone 0-1-3 (mid-low-high), 1-4 (low-rising), 2-3(high-falling). Here we propose a new method shown in the following:

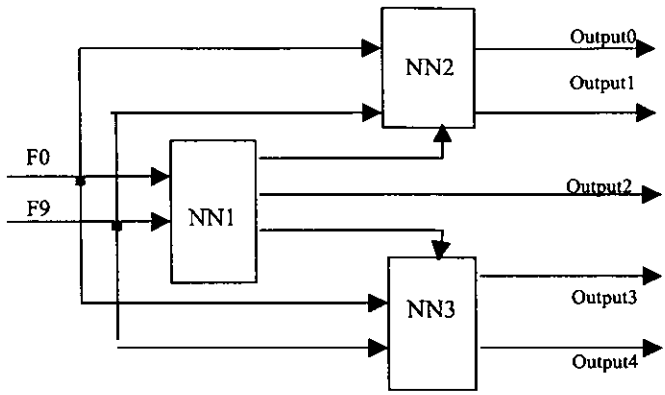

Figure. 5-2 2-stage NN framework 2 based on confusion information

Here NN1 is set according to the confusion group. NN2,NN3 are set according to the NN1 output. The detail of possible setting we have tried are introduced in the following section.

As we can see that the low tone and rising tone are easy confused, because they have the similar initial part of pitch contour. The continuous speaking of speech made the rising tone can't meet the last part of pitch contour and let it easily confuse with low tone. The same case we can find for high tone and falling tone. At the same time, the pitch contour shape of mid tone, low tone, high tone is similar. They don't have big slope. This made them another group of big-confusion. So here, we first group the tone into 0-1,4-2,3 (mid-low,rising-falling,high) three groups. Then we group them into 0,1,3-2-4 (mid,low,high-falling-rising) three groups. Finally based on the experiments results of these two grouping, we use them to be post-system of the base NN that we have had.

Expr. 1: Grouping confused tones from intial level of pitch contour.

In expr. 1, we group the tone according to confusion group 0-1,4-2,3 which is shown in the following:

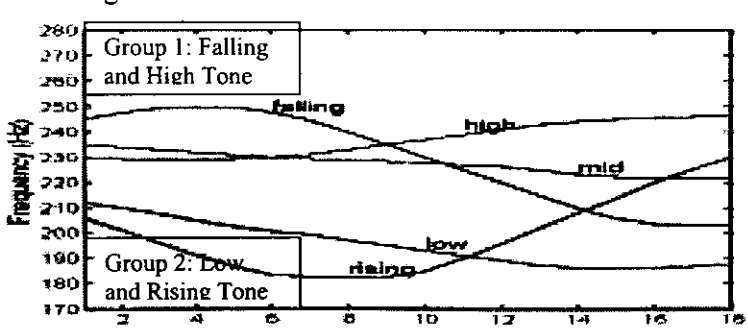

Figure 5-3 Grouping of tone according the initial level

The confusion-matrix using this grouping method is as following:

Table 5-1 Classification confusion-matrix of tone0-tone1,4-tone2,3 (tone mid-low, rising-falling, high)

Tone-Group	0	14	23
0	306	44	34
14	60	285	43
23	50	15	383

From the confusion-matrix we can find that the totally correct number is 974 that is higher than former NN about 93 tones, i.e. about 10%. But it doesn't include the confusion between tone1-4 and tone2-3 which is very easy to be confused.

After the classification among the group 0-1,4-2,3, next we use two NNs to classify tone 1,4 and tone 2,3. The classification result for tone1 and tone4 is shown in Table 5-2.

Table 5-2 Classification results for NN2 (tone low-rising)

Tone	1	4
1	165	-
4	-	8

From the result, we can see that the performance of tone 1 is higher than that of tone4. Thus we think tone4 is easier to misclassify with tone1 than the tone1 to tone4, because pitch contour of tone4 may not touch the last part for the continuous effects.

The classification result for tone2 and tone3 is shown in the following:

Table 5-3 Classification results for NN3 (tone falling-high)

Tone	2	3
2	52	-
3	-	139

Similar to above result, the performance of tone 3 is higher than that of tone2. The tone2 is easier to misclassify with tone3 than the tone3 to tone2. The reason is similar to that of tone1 and tone4.

Final result of this experiment is shown in the following table and it is compared with our base NN's results.

Table 5-4 Final Classification Results and Original Results

Tone	0	1	2	3	4	Total
Original	319	159	212	103	88	881
Expr. 1	306	165	52	139	8	670

From the table, although the total performance is not improved, we found that the performance of tone1 and tone3 are higher than original one. Then this framework will be helpful for classification of tone1 and tone3.

Expr. 2: Grouping tone based on the shape of pitch contour.

In this experiments, we group the tone into 0,1,3-2-4 (mid, low, high- falling-rising)three groups according the shape of the pitch contour. Then we use the NN framework we got above for classification of tone0,1,3(mid-low-high).

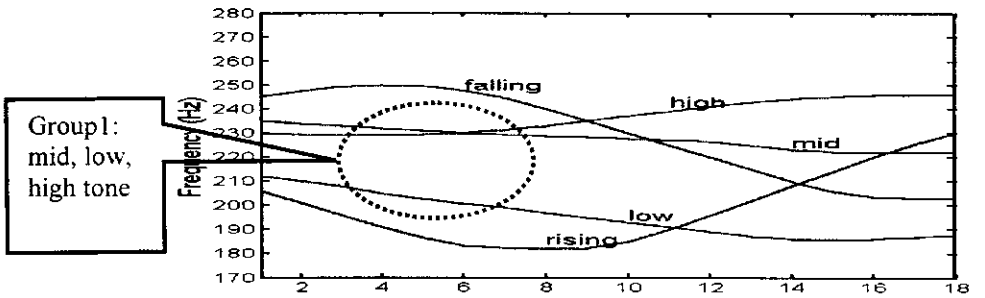

Figure 5-4 Tone Grouping based on the tone shape

The confusion-matrix for this grouping is in the following:

Table 5-5 Confusion-matrix for NN0,1,3-2-4 (tone mid, low, high- falling -rising)

Tone-Group	013	2	4
013	750	55	21
2	44	211	0
4	49	0	90

From the confusion-matrix, the total correct amount is 1051 that is higher than original about 170, i.e. 14%. But here it doesn't consider the confusion among tone0-1-3. The classification results for tone2 and tone4 are close to that of 5-tones classifier. For group tone0,1,3, we pass them into another NN for classification. The classification result is shown in Table 5-6.

Table 5-6 Classification Results for tone0-1-3 (mid-low-high)

Tone	0	1	3
0	305	-	-
1	-	156	-
3	-	-	116

Here the NN is the neural network for tone0-tone1,4-tone2,3. As we know, the NN for tone0-tone1,4-tone2,3 take advantage for classifying tone1 and tone 3. So here we use it to avoid the tone in the 013 group that is1 but similar to tone4 confused as tone0. The same is for tone3.

The final classification result is shown in Table 5-7 and compared with base performance and the results of experiment 2.

Table 5-7 Comparison of Final Classification Results

Tone	0	1	2	3	4	Total
Original	319	159	212	103	88	881
0-14-23	306	165	52	139	8	670
013-2-4	305	156	211	116	90	878

Here we set the group as tone013-tone2-tone4 (mid, low, high-falling-rising) also because the tone2 and tone4 is two uneasy confused tones. Then we first classified them from other tones and then classify other tones that are easy confused. From the table, we can find that anyway the final result is not improved through this method. The possible reason may be that the extracted tone data is not good enough for classification. That's mean if we want to improve the performance further, the processing of extracted data need to be further studied and processed.

Expr. 3: Combination of Base NN and Grouping NN.

In this experiment, we are trying to combine the base NN and the NN that we discussed above. We use the output of the base NN as the input of the NN that we discussed above.

First the tone0 (mid) outputs from the base NN are chosen to pass the NN for tone 013-2-4 to classify out tone2-4 from tone0 output. Then the rest data is pass to NN of tone0-14-23 for further classification. The classification result is shown in Table5-8. Totally results is 329 that is higher than the base NN 10 tones.

Table 5-8 Classification Results of Tone0 output

Tone	0	1	2	3	4
0	293	13	5	16	2

For the output of tone1(low), first we let them pass NN 0-14. Then we pass them into NN 01 and NN 14. The classification results are shown in table 5-9. We got 3 tones more performance.

Table 5-9 Classification Results of Tone1 Output

Tone	0	1	2	3	4
1	9	152	0	0	1

For the output of tone3 (high), they passed through NN 0-14-23. The classification results are shown in table 5-10. Here we got 3 another more from the output of tone 3.

Table 5-10 Classification Results of Tone3 Output

Tone	0	1	2	3	4
3	3	3	0	100	0

Then the final classification results are shown in table 5-11 and compared with other experiment. Finally for experiments 3, we got 14 tones more for improvement. But anyway it's still little improvement for totally 1220 tones.

Table 5-11 Comparison of Final Classification Results

Tone	0	1	2	3	4	Total
Original	319	159	212	103	88	881
Expr. 1	306	165	52	139	8	670
Expr. 2	305	156	211	116	90	878
Expr. 3	304	168	216	116	91	895

5.4 Summary

All of above approaches use the different setting of NN. We found that grouping the tone into tone0-1,4-2,3 (mid-low, rising-falling, high) is good for classifying tone1 and tone3. Grouping the tone into tone 0,1,3-2-4 (mid, low, high-falling-rising) is good for classifying tone2 and tone4. Combining these two methods into the base NN we got before, the 14-tone improvement can be achieved. But these 14 tone improvement is not much considering from 1220. Based on this, we think that the extracted tone data may be not good enough for classification. So if we want to improve the performance further, the processing of extracted data need to be further studied and processed. For example, in Expr. 1, we first classified 0-14-23 into three group. Then we use different features for tone1-4 for classification also tone2-3 in order to improve the classification performance. Which features are especially suitable for classifying tone1-4 and tone2-3 still need the further research? Basically we think the mis-classification between tone1 and tone4, tone2 and tone3 is because of the context variation. To consider the context interaction into the tone feature setting for tone1-4 and tone2-3 should be helpful. What context interaction we should extract and add into the tone feature still needs further researches.