# Chapter 6

## Summary and Future Works

### 6.1 Summary

As introduced, tone classification is a necessary component for speech recognition. Pitch detection is the first task for tone classification.

In Chapter 2, two pitch detection algorithms: AMDF and auto-correlation method, are implemented. The preprocessing technique: LPF and non-linear clipping, are done before pitch detection. In order to extract the tone feature from pitch contour, LMS or orthogonal polynomial algorithm, are implemented. All of them give the acceptable results.

In Chapter 3, the configurations of tone feature are discussed. As we know, the big variation lies in the pitch data. In order to reduce the effects of these variations, the further processing is necessary. The techniques related to tone-critical-segment, scaling, normalization, tone feature setting are introduced and the experiments are done on the speech data. Finally the performance of 72.21% are obtained from the best configuration that is rhythm segment, semi-tone scaling, mean normalization, 10 features. The rhythm tone critical segment is better than vowel segment. The ERB scaling is better than semi-tone scaling for tone classification. Mean normalization is better than Z-score normalization for tone classification. The 10-feature setting using five slopes and five heights of pitch contour is better than using only four coefficients of 3-order polynomial.

In Chapter 4, the method for improving the performance of tone classification using the information of final consonant is proposed and the experiments are done on the speech data. The final consonant for Thai language mainly can be grouped into 3 groups: nasal final consonant, plosive final consonant, others (approximant final consonant,fricative final consonant, lateral approximant final consonant) on the basis of the distribution of final consonant. Based on the phonetics knowledge it can be grouped into two groups: voiced final consonant and unvoiced final consonant. According to these two grouping methods, the classification experiments are done on the speech data that grouped. The best performance is improved to 77.13%. Also comparing the classification results among these groups, we found that the performance for the group that has no final consonant is obviously lower than that with final consonant. Then we conclude that the final consonant is helpful to stabilize pitch contour and then let tone

classification become easier. The performance among the tone groups with final consonant gives the different performance also. The different final consonant has different effects on the pitch contour even though it has the same tone can be concluded.

In Chapter 5, the 2-stage NN approaches for improving the tone classification performance are proposed and the experiments are done on the speech data. After analyzing the confused tone of the NN's output, we found about 54% confused tone can be deleted through choosing the second candidate other than the first candidate. It's said most of confusion lies in the first and second candidate. Then we proposed that using $2^{nd}$-layer NN framework for reducing the confusion of the first candidate and second candidate. The first two candidates of 5-tone classifier are chosen to classify again using the 2-tone classifier. But we didn't get any positive results from this method. Then we propose another method that using 2-stage NN for classification using the grouping of tone. From the confusion-matrix, it can be seen that the big confusion lies in the tone 0-1-3 (mid-low-high), tone 1-4 (low-rising), tone 2-3 (falling-high ). Based on this, the first grouping of tone is to classify tone0-1,4-2,3. Here we group tone 1,4 into one group and tone 2,3 into another one in order to ignore the confusion between them. Then we classify tone 1,4 and tone 2,3 separately using 2-tone classifier. Although the final classification result didn't improve, the classification results of tone 1 and tone 3 are increased obviously. Then we can say that this scheme is good for classifying tone 1 and tone 3. Next we grouped the tone into tone 0,1,3-2-4 in order to ignore the big confusion among tone0-1-3 first. Then tone 0,1,3 are classified using another NN. The final results are better than last proposed method, but still don't make improvements from 5-tone classifier. But the classification performances for tone 2 and tone 4 are increased. Based on all of above, we proposed that using the output of 5-tone classifier for further classification. Then we input the output of tone 0 from 5-tone classifier into the NN 0-1,4-2,3 for reducing the confused tone in tone 0 output. Here we got 10 more increased. Similarly this is done for tone 1 and tone 3. Finally we got 16 more correct. It still is not a big improvement. But it's clear that in order to improve the performance further, the main topic is to reduce the confusion between tone 0-1-3, tone 1-4 and tone 2-3. So for some future works, we can concentrate in finding the special feature setting for the classification of these easy-confused tone groups.

From all of the work described above, this thesis finally completed:

- Two pitch detection algorithms are implemented;
- The configurations of tone feature are studied and the best configuration are found based on the experiments;

- The effects of final consonant on tone classification are studied and concluded;
- The 2-stage NN approach for improving the performance of tone classification is proposed. The experiments has been done and the results are concluded;

The main contributions of this thesis work include:

- Although tone classification is not a new topic for Thai tone classification, this is the first research work of tone classification done on a newly-build Large Vocabulary Continuous Speech Database;
- Through the studying of the tone feature configurations, we got the different results from previous research that gives the new view of tone feature research;
- The effects of final consonant on tone classification are first studied in this research work. Some knowledge about this topic are concluded;
- This research work proposed several approaches for improving tone classification performance, such as: 2-stage NN approaches, biasing approaches. The further exploration of these approaches is still necessary and valuable.

## 6.2 Future Works

According to the work described above, some conclusions are already shown in the experiments. But it still has lots of knowledge and techniques need to be explored to let the tone classification used in the real speech recognition system. The main future topics can be:

- Improving Tone Classification Performance
  Through the work in this thesis, a basic tone classifier for Thai speech is already implemented. The further improvements of tone classification still need to be explored. Based on several approaches proposed in the thesis, the further studying of them is necessary. For 2-stage NN approaches, we group the tone according to the confusion information. Then how to classify these easy confused tone is one of issue that need to be studied. For biasing approaches, the choosing of biasing coefficients and the biasing method still need to be studied.

- Combining tone classification into speech recognition system

During this research work, we are using NN as the modeling technique because tone information is segment based. Currently for speech recognition the HMM is widely accepted as a better one. It's frame based. So how to let frame based framework and segment based framework incorporate with each other to get the best matching is a big issue for future work.

- the expansion of tone classification research
  This research work is a tone classification system and we use F0 as the main cues for classification. As we know, F0 is the main cue of prosodic information that not only include the tone but also a lot of other bilinguistic information. So how to expanding this research work into other region which using prosodic information also can be another issue for future work.