Appendix

Appendix 1: HTK Experiments

F0 are not used as input features and are in fact factored out from the signal when MFCC features are used (F0 information is in the higher order MFCCs --- it is usually discarded) (http://www.ling.mq.edu.au/units/slp806/slp806/prosody.html). Based on this, basically the tone classification using MFCC should not give very high performance. Because the principle cue that includes the tone information is F0 and F0 is discarded by MFCC. Although like this, the experiments still have done on HTK.

The basic steps for the experiments are:

- getting label information
- delete the duration less than 2 segments in label information
  --Because the MFCC feature can't be extracted for the speech that shorter than 2 segments.
- using label information and programming to cut the mfcfiles / wavefiles into single tone
- using hcopy to get mfc file
- edit the bat files
- run .bat files

Experiment's Results:

Some of the cut tone speech are deleted from the experiment. Because some of tone speech are too short to extract MFCC features. We use 1 state to build the HMM model to model the tone. And to get a better performance, the increased mixtures in HMM model is used.

Table A1-1 Tone Classification Results of HTK Experiments

	5 mixtures	10 mixtures	15 mixtures	20 mixtures
Percent	34.09%(254/745)	34.23%(255/745)	36.24%(270/745)	37.45%(279/745)

From the results shown above, it proved that the tone classification is essentially necessary for tonal speech recognition system.

Appendix 2: Human Testing

First the single tone is cut out from the utterance. Then they are randomly chosen for testing. Considering generally our human don't have experience to classify the single tone without the context and other related information, we give each listener 10 samples for train and every time 25 tones for classify which is randomly sampled as 5 each tones. The first round testing data is not discarded because it's quite low with respect to no former experience. Finally the best performance that we got is 72%.

This experiment isn't intent to be a statistics work for research. We chose several listeners for testing. The results are trying to give us a cue about human ability of tone classification. From the results, we found even for the human, the performance of tone classification is not high also. Thus the human understand of the tone in speech is not only based on the pitch information but also other speech information for helping tone classification.

Appendix 3: Biasing Approaches

From the speech data and the former experiments, we know that the probability of each tone happened when spoken by human is not the same. Some tone is more often used, such as tone0. Some tone is less probability happened, such as tone 4. So in order to improving the performance, we can bias the score of NN output based on this probability. The difficult here is to find the suitable coefficients for biasing that we may conclude it from some statistics linguistics data. Here we just use the data that we collect from our database in order to test this proposed method.

Table A3-1 Distribution of Tone 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%

Then we multiplied the coefficients [0.35 0.23 0.19 0.15 0.09] from the distribution of tone in this speech database with the output of NN. After that we choose the highest score output as the correct answer. But finally we didn't get any improvements of the classification. The possible reason may be the coefficients

choosing is not from scientific way or the way for biasing is too direct and simple to make the proposed method not work well.

In appendix, some other works that we have done are described. First the tone classification experiments using HTK toolkit is done. MFCC feature is not suitable for tone classification, although it's one of the most popular features used in current speech recognition system (http://www.ling.mq.edu.au/units/slp806). This is proved by our experiments. The best performance that we get is about 37.45%. Then a simple test of human tone classification ability is done on our speech data. The best performance we get is about 72%. Although this result is not gotten from statistics way, it is proved that the human tone classification ability is also not high without other information of speech and former experience. Also the human use the combination information of speech to help improve the recognition ability. Finally the approach using the biasing coefficients from the happening probability bias the NN output for improving the performance is described. The experiment result didn't give us any positive results. But some further research still can be done on this approach. The way of biasing and the choosing of biasing coefficient still need further study.

VITAE

Name Miss Li Tan
Birth Date 2 March, 1978

Education Attainment

Degree	Name of Institution	Year of Graduation
Bachelor of Engineering	Changchun University of Science and Technology, China	1999

Field of Interest: Speech Recognition

List of Publications:

1. Tan L., Karjanadecha M., "Modified Mel-frequency Cepstrum Coeffcient", *Proceeding of Information Engineering Postgraduate Workshop 2003(ICEP 2003),* pp. 127-130, Songkla, Thailand, Jan. 30-31, 2003.
2. Tan L., Karjanadecha M., "Pitch Detection Algorithm: Autocorrelation Method and AMDF", *Proceeding of the 2003 International Symposium on Communication and Information Technology (ISCIT 2003),* pp. 271-274, Songkla, Thailand, Sep. 3-5, 2003.
3. Tan L., Karnjanadecha M., Thanate K. and Pichaya T., "A Study of Thai Tone Classification", *Proceeding of Information Engineering Postgraduate Workshop 2004(ICEP 2004),* Phuket, Thailand, Jan. 22-23,2004
4. Tan L., Karnjanadecha M., Thanate K. and Pichaya T., "A Study of Tone Classification for Continuous Thai Speech Recognition", *8th International Conference on Spoken Language Processing(ICSLP2004),* Jeju Island, Korea, Oct. 4-8,2004(Submitted)