International Journal of Hybrid Information Technology
Vol. 9, No.10 (2016), pp. 161-166
http://dx.doi.org/10.14257/ijhit.2016.9.10.16

An Improved Tone Model for SVM Based Tone Recognition

Qian Liu[1,2], Jinxiang Wang[1] and Mingjiang Wang[3]

[1.] Microelectronics Center,Harbin Institute of Technology, Harbin, P. R. China
[2.] Department of Electronic Science and Technology, Harbin University of Science and Technology, Harbin, P. R. China
[2.] Shenzhen Graduate School, Harbin University of Science and Technology, Harbin, P. R. China
jxwang@hit.edu.cn, liuqian0428@126.com

Abstract

Tonal information is important for Chinese Mandarin speech recognition and understanding. Tone information is carried on the shape of fundamental frequencies. When constructing a tone recognition system based on support vector machine, the fundamental frequencies should be normalized by curve fitting method. The fitting coefficients are input data of support vector machine, but it cannot precisely represent the shape of fundamental frequencies. An improved tone model is proposed, which added some other information of fundamental frequencies, such as the maximum and minimum value, the position of maximum and minimum values, the fundamental frequency values of first and last point. The experimental result shows that the proposed model is more accurate with the shape of fundamental frequencies and the accuracy rate of tone recognition is increased by using the improved tone model.

Keywords: *tone model, support vector machine, tone recognition, speech recognition*

1. Introduction

Chinese Mandarin is known as a tonal language. Tone information is very important for Mandarin speech recognition. It is proved that the accuracy rate of speech recognition is increased by introducing tone information to the matching stage[1]. Tone information is carried on fundamental frequencies. The four tones can be classified by its fundamental frequency shapes. Literature [2] uses second order polynomial to fit the fundamental frequencies, and uses the three coefficients to classify the four tones. The classification process employs eight thresholds, which are defined by experimental results. Literature [3] uses normalized fundamental frequency, logarithm energy, the first order difference of fundamental frequency and logarithm energy, and the second order difference of fundamental frequency and logarithm energy as feature parameter. The tone recognition result is derived from decision tree. After machine learning theory is successful used in speech recognition system [4], literature [7] and [8] adopts HMM and ANN for tone recognition respectively. Literature [9] and [10] employ SVM to tone recognition process.

For SVM based tone recognition system [11], the feature parameter, which is derived from fundamental frequency [8], logarithm energy [8] and MFCC [10], should be normalized at the same length as the input of SVM. The normalization method in literature [8] is curve fitting, and the tone model is built by fitting coefficients, which are the input for SVM. The fitting coefficients are computed by least square theory, which generates some deviations between the shape of fundamental frequencies and the fitting curve. It means that the fitting coefficients cannot describe the shape of fundamental frequency precisely. An improved tone model, which is more precise with the shape of fundamental frequency, is proposed for SVM based tone recognition.

2. Curve Fitting Based Normalization

Tone information is carried on the shape of fundamental frequency. For SVM based tone recognition, fundamental frequency, logarithm energy, first order difference of fundamental frequency and first order difference of logarithm energy are used as feature vectors to build tone model.

As a static classifier, the dimension of tone model in SVM should be consistent [12]. So, it is necessary to make the fundamental frequency and logarithm energy with different length into tone model with same length. The feature vectors are normalized by curve fitting method, and the tone model is generated by composing the fitting coefficients. The least square method based on the Legendre Polynomials basis functions is used for curve fitting. The top 6 orders Legendre Polynomials are as follows:

$$P_0(x) = 1 \tag{1}$$

$$P_1(x) = x \tag{2}$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1) \tag{3}$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x) \tag{4}$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \tag{5}$$

$$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x) \tag{6}$$

The top 4 orders Legendre Polynomials are used only, which means that the feature vectors of each frame turns to a tone model with sixteen fitting coefficients.

3. Improved Tone Model for SVM based Tone Recognition

Figure 1 gives four curve fitting result of Chinese syllable”ci4”,"ren2","jiu3"and"wan4". The fitting curve of syllable”ci4” is the best; the fitting curve of syllable “ren2" and “jiu3" are the worst. The deviation is introduced by least square method. By adding some other information in tone model, such as the maximum and minimum value, the values of first and last point, the position of maximum and minimum values, *et al*, can make the tone model is more similar to the feature parameters.

An improved tone model is built by adding difference of first and last fundamental frequency point, maximum and minimum value of fundamental frequencies, the position of maximum and minimum fundamental frequencies, and difference of maximum and minimum fundamental frequencies to the sixteen curve fitting coefficient.

For high and smooth tone, its fundamental frequency curve should be approximately a straight line. Its first and last fundamental frequency point should almost the same, and so did of the difference of maximum and minimum fundamental frequencies. For rising tone, its fundamental frequency curve should be approximately a rising line. The difference of maximum and minimum fundamental frequencies and the difference of first and last fundamental frequency point are relatively large, and at the same time, its maximum and minimum fundamental frequencies should near the first and last fundamental frequency point. For falling and rising tone, its fundamental frequency curve should be decline first and then rising. Its minimum fundamental frequency should near the middle of fundamental frequencies. For falling tone, its fundamental frequency curve should be approximately a falling line. The significant difference of rising tone and falling tone is that the difference of first and last fundamental frequency point. For rising tone, its value is below 0; for falling tone, its value is above 0. Therefore, the new 22 dimensional feature parameter dimension can describe the fundamental frequency sequence more precisely.

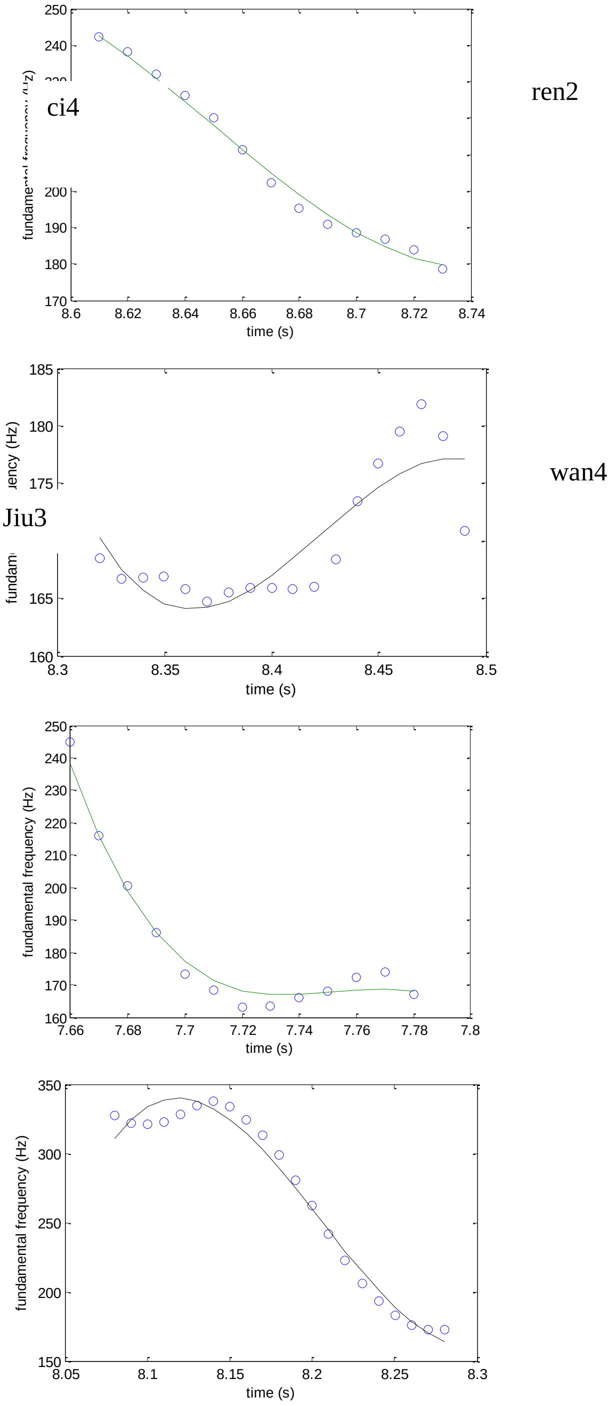

Figure 1. Comparation of Fundamental Frequency Sequence and Curve Fit Result

4. Experimental Results

Randomly choose 35.36s speech samples from standard Mandarin speech library to complete the experiment which helps to determine the parameters in SVM training period.

The parameters are p1 in Radial Basis Function and penalty factor C. The sampling rate of the speech signal is 16 kHz, and is quantized in 16 bit. The ratio of speech segment is 72%. The length of the analysis window is 25ms, and the frame shift is 10 ms.

Table 1 gives the reference value of p1 and C. Six tone recognition process are implemented with different values of p1 and C. The optimal value of p1 and C then determined by comparing the six tone recognition error rate and its optimal classification interval.

Table 1. Reference Value Table of P1 and C

parameters	Reference value					
	First group	Second group	Third group	Forth group	Fifth group	Sixth group
p1	1	10	100	1	1	10
C	10	10	10	1	100	100

The tone recognition error rate with different values of p1 and C is shown in Figure 2. Except for the second group, the error rates of other groups are 0, which reaches the best situation of 100% accuracy rate. The optimal parameters for tone recognition need to be determined according to the margin between two hyper planes. According to the classification theory of SVM, when the training samples are fixed, the margin between the two hyper planes should with its maximum value. Figure 3 gives the margin of hyper plane with different kernel parameters. By analyzing Figure 3, the optimal values of p1 and C are the third group.

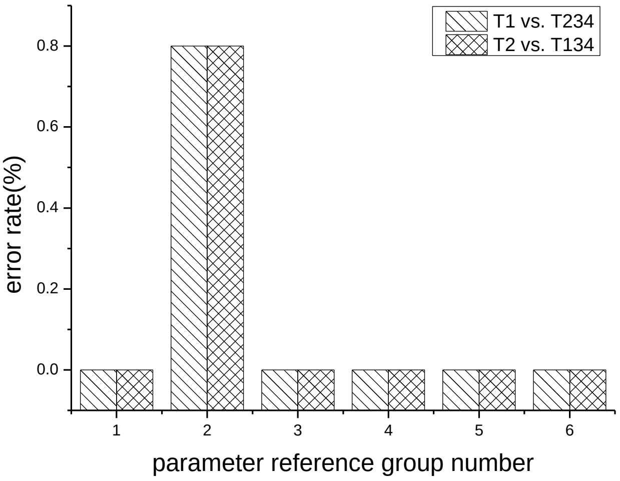

Figure 2. Margin of Hyper Plane with Different Kernel Parameters

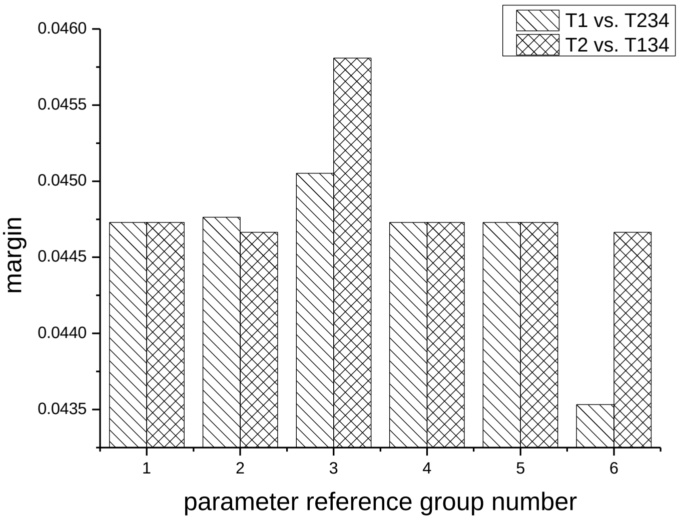

Figure 3. Margin of Hyper Plane with Different Kernel Parameters

Randomly choose 288.9s speech samples from the Chinese news speech database to analysis the performance of tone recognition system. The ratio of the speech sample was 25.37%. 201.94s speech samples are chosen as training data, and the other 86.96s speech samples are as testing data. The sampling rate of the speech signal is 16 kHz, which is quantized of 16 bits. Its analysis window is 25 ms, and the frame is shifted every 10 milliseconds.

The fundamental frequency and logarithm energy of each frame are extracted from speech file, and then use them to build both tone model and improved tone model. In the tone recognition program, four “one vs. all” classifiers were constructed: tone 1 vs. tone 2, tone 3 and tone 4; tone 2 vs. tone 1, tone 3, and tone 4; tone 3 vs. tone 1, tone 2, and tone 4; tone 4 vs. tone 1, tone 2, and tone 3. The class label is appended according to the classifier it was attached to. For example, the label of syllable “mai” with rising tone should be 1 for the classifier of tone 2 vs. tone 1, tone 3, and tone4; it should be -1 for the classifier of tone 1 vs. tone 2, tone 3, and tone4.

Figure 4 shows the tone recognition error rate under different tone models. The error rate of four tones all decreased by using the improved tone model. It can be seen that the improved tone model is better than the typical one. For falling and rising tone, its error rate decreased most significantly. It is because that its fundamental frequency shape is similar to a second order polynomial, the typical tone model cannot fit it precisely, but the improved tone model can. The improved tone model can increase the tone recognition accuracy rate.

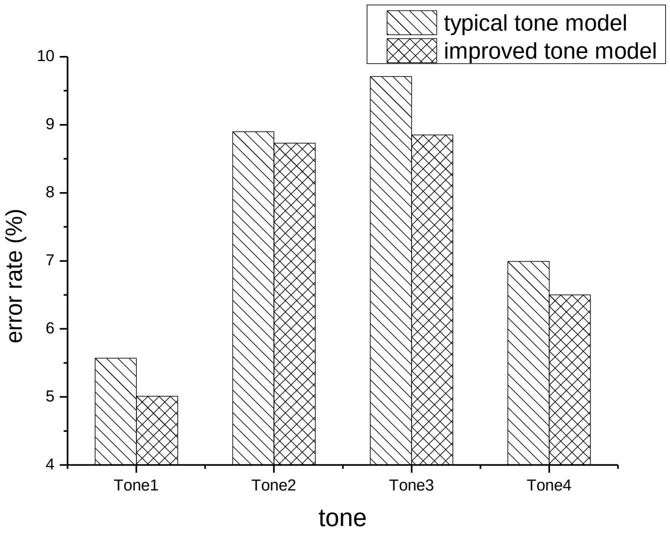

Figure 4. Error Rate of Tone Recognition under Different Model

5. Conclusions

Tone information is important for Mandarin speech recognition. The typical tone model for SVM based tone recognition system is built by curve fitting method. The tone model is consisted by fitting coefficients. An improved tone model is proposed which added some other information of fundamental frequency to make the tone model more precisely. The experimental results show that the tone recognition error rate by using improved tone model is smaller than that by using typical tone model.

References

[1] J. Chaiwongsai and Y. Miyanaga, “Improved tone model for low complexity tone recognition”, Proceedings of the SICE Annual Conference, **(2014)**, pp.1124-1129.

[2] D. Liu, C. Xu and S. Sun, “A Chinese four tone detection algorithm”, The third National Human-computer Speech Communication Conference, **(1994)**, pp.121-124.

[3] C. Yang, D. Yong and Z. Hong, “Decision tree based mandarin tone model and its application to speech recognition”, IEEE International Conference on Acoustics, Speech and Signal Processing, no. 3, **(2000)**, pp.1759-1762.

[4] M. Espi, S. Miyabe and T. Nishimoto, “Using spectral fluctuation of speeach in multi-feature HMM-based voice activity detection”, Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, **(2011)**, pp. 2613-2616.

[5] K. Hirose, H. Hu and X. Wang, “Tone recognition of continuous speech of standard Chinese using neural network and tone nucleus model”, 9th International Conference on Spoken Language Processing, no. 5, **(2006)**, pp. 2394-2397.

[6] S. Wang, Z. Tang and Y. Zhao, “Tone recognition of continuous Mandarin speech based on binary-class SVMs”, 1st International Conference on Information Science and Engineering, **(2009)**, pp.710-713.

[7] L. Zhao, C. Zou and Z. Wu, “Tone recognition of Chinese continuous speech based on continuous distributed HMM”, Signal Processing, vol. 16, no. 1, **(2000)**, pp.20-23.

[8] L. Tang, J. Yin and Z. Su, “Mandarin tone recognition system based on two level BP model”, Computer Engineering and Application, no. 25, **(2004)**, pp.96-99.

[9] M. Gu, Y. Xia and Y. Yang, “Tone recognition based on support vector machine”, Acoustic Technology, vol. 26, no. 6, **(2007)**, pp.1186-1190.

[10] H. Xiao and C. Cai, “Speaker-independent tone recognition based on support vector machine”, Computer Engineering and Applications, vol. 45, no. 9, **(2009)**, pp.174-176.

[11] G. Zheng-Yan, Z. Yu-Shuang and W. Mu-Kun, “Speaker Recognition Using KernelK-mean Clustering and SVM”, Journal HARBIN UNIV.SCI.＆ TECH., vol. 13, **(2008)**, pp.40-46

[12] V. N. Vapnik, “The Nature of Statistical Learning Theory”, New York,pp. Springer-Verlag, **(1995)**, pp. 90~156