RESEARCH ON DYNAMIC CHARACTERS OF CHINESE PITCH CONTOURS

*Zhiyong Wu** *Lianhong Cai** *Tongchun Zhou***

* Department of Computer Science and Technology, Tsinghua University,
Beijing, P. R. China, 100084
Email: wuzy@tts.cs.tsinghua.edu.cn
** Chinese Department, Beijing Normal University, Beijing, P. R. China, 100875

ABSTRACT

Chinese is a tone language. For a tone, the characters of its F0 pitch contours will be quite different in the condition of continuant speaking from the isolated speaking. The present researches about the Chinese tone are still centralized on the isolated speaking one, and about tone in fluent speech, there are some statements about the phenomenon of the two-word, three-word, four-word co-reading, but it can not represent the real characters of the tone in fluent speech very well.

Our work bases on the spoken Chinese. And with the methods of statistics and classifications, we study and analyze the characters and behaviors of the F0 pitch contours of Chinese mandarin four tones in fluent speech.

In the Paper, we put forward a mathematical model for describing the pitch contours: pitch contours time normal model (PiCTN model). In Chinese fluent speech, because of the co-reading and different surroundings, the pitch contours of one tone will be quite different from what it will be in the isolated speaking, and will even have many different characters. And about the dynamic characters of pitch contours for Chinese Putonghua four tones in fluent speech, we also give the conclusion: the F0 pitch contour of tone 1 should be high and flat, and sometimes it will be rising with the even speed, the contour of tone 2 will be rising with the accelerated speed, and the pitch contour of tone 3 will drop with the decreasing speed, and lastly the F0 contour of tone 4 should be dropping with the speed increasing.

The results of our work will be useful for the prosody modeling in Text-to-Speech (TTS) systems and the tone confirmation in the speech recognition.

KEYWORDS: Chinese Putonghua, Pitch contours, Static tone, Dynamic tone

1. INTRODUCTION

Chinese is a tone language. For a syllable, there are four lexical tones. Present studies show that the characters of different tones are determined by the F0 pitch contours, for instance, tone 1 is characterized by a high-flat F0 contour, tone 2 is characterized by a rising contour, tone 3 is characterized by a low-dip contour, and tone 4 is characterized by a falling contour from high F0.

Now, more studies about the Chinese Putonghua tones have shown that, for a tone, the characters of pitch contours will be quite different in the condition of continuant speaking from isolated speaking. One tone will have the steady-going F0 contours in the isolated speaking. While in fluent speech, because of the influence of the context, it will have one or more other different characters. Thus, someone has distinguished the Chinese tones into two different types: the static tone and the dynamic tone [1].

On the characters of the Chinese tones, the great linguistics master, Mr. Y.R.Chao had made many creative guidance research works. With his five-degree tone letters, the characters of the tone when isolated speaking can be described very well [2]. And he also made many researches on the two-word speaking and three-word speaking tone changing, and gave many rules and conclusions (co-reading rule) [3].

Now, the main applications, the prosody modeling in Text-to-Speech (TTS) systems and the tone confirmation in speech recognition, will make use of the dynamic pitch characters, especially of the tones in the fluent speech. While, on the other hand, the researches about the Chinese tones are still centralized on the static one, and about the phenomenon of the two-word, three-word, four-word co-reading, some statements have been given, but which still cannot describe the characters of the tone in the fluent speech very well. Therefore, we should take more attention and make more experiences and studies on the characters of pitch contours about Chinese tones in fluent speech, that is the dynamic tone.

Our work bases on the fluent spoken Chinese speech. And with the methods of statistics and classifications, we make the experiments and analysis about the characters and behaviors of the F0 pitch contours of the Chinese mandarin four tones. We also give some conclusions about the dynamic characters of the pitch contours for Chinese Putonghua four tones in the spoken language.

2. PITCH CONTOURS TIME NORMAL MODEL (PiCTN Model)

In order to study the characters of different tones in fluent speech, we have put forward the concept of the PiCTN Model (Pitch Contours Time Normal Model). With the model, the duration of the syllable will be normalized from 0 to 1, and the F0 contours will be kept unchanged. The model will be acceptable for the study of the pitch contours characters.

For the F0 pitch contour of one tone getting from the fluent speech, the PiCTN model data of which can be obtained by following the next steps:

(1) Time normalization

The figure 1(a) shows ten F0 pitch contours of tone 2, from which, we can see that the durations of the contours are quite different. In order to perform the succeed processing, the duration should be the same, then the time normalization of the duration should be performed. The formulation is:

$$F_0(\tau_j) = F_0(t_j / T),$$

where T is the duration of the syllable, t_j is the j th sampling time, τ is the normalized time from 0 to 1, and τ_j is the j th normalized sampling time. The figure 1(b) shows the pitch contours of which the durations have been normalized.

(2) Curve fitting and re-sampling

Usually, the sampling time is quite varied for different contours, which will not be convenient for the following processing. Actually, the sampling time should be the same. The curve fitting can be performed to the old pitch contours, and then by re-sampling, the new pitch data can be gotten.

The figure 1(c) shows the new contours from the re-sampling data. The new sampling time is: $\tau = 0,0.1,0.2,...,0.9,1.0$.

(3) Musical scale measurement of the pitch

The human is sensitive with the comparative change of the pitch, and tone shift has the character of the musical scale measurement (Z.J.Wu), so, it will be quite acceptable for the pitch data to be measured with the musical scale.

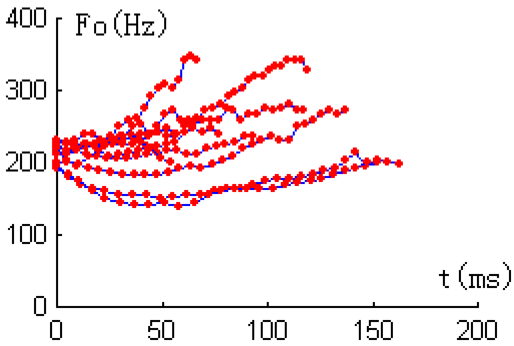
For the interval between two neighboring semitone scales of the musical scale measurement, the according ratio of the frequency will be $2^{1/12}$. If the frequency value F_s accords with the musical scale value P_s, then the relation between P_i and frequency value F_i will be:

$$P_i - P_s = 12\lg_2 \frac{F_i}{F_s}$$

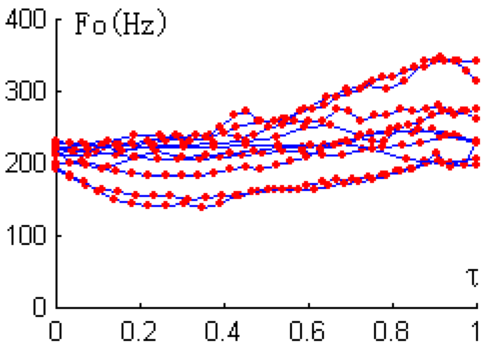
The figure 1(d) shows the new pitch contours, where $P_s = 50$, and $F_s = 320Hz$.

(4) Character analysis of the pitch contours

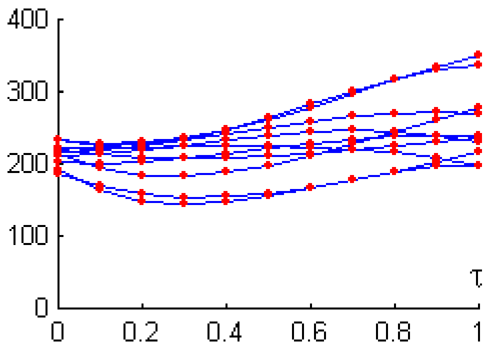
After the above processing, now the character analysis of the contours can be performed. There are many processing methods for different purposes, such as clustering, mean value getting etc. The figure 1(e) shows curve of the mean value of the above ten contours.



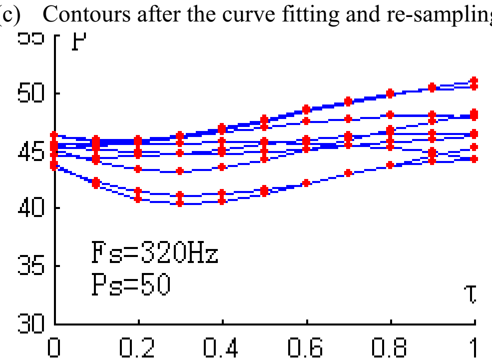
(a) Original pitch contours



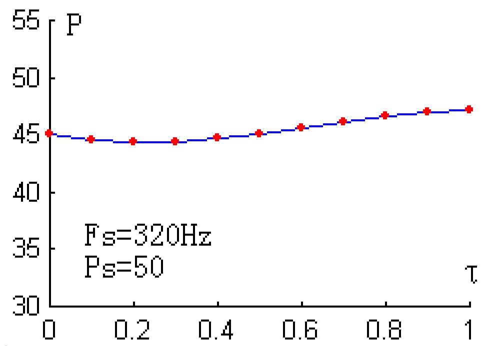
(b) Contours after the time normalization



(c) Contours after the curve fitting and re-sampling



(d) Contours in the musical scale measurement



(e) Mean value of the above ten contours

Figure 1: Pitch Contours Time Normal Model (PiCTN Model)

3. EXPERIMENT AND RESULT

To study the dynamic pitch contours characters of the tone in fluent speech, we made the following experiment, in which, we studied the pitch characters of the Chinese Putonghua four tones. And by statistics and classification, we have drawn some conclusions.

3.1. Data preparation

To accomplish the experiment or the work, we first established a speech data corpus containing about 3000 chunks (getting from the spoken sentences) and 20000 words, then we made the automatic pitch marking using the TH-SPEECH tool [4], and lastly we checked and modified the pitch contours data manually, thus we assured that our data sources were mostly accurate.

3.2. Experiment method

In the experiment, the following steps should be performed:

Firstly, get the F0 pitch contours from the speech data corpus. And then classify the contours according to different tones, thus the database of the pitch contours for each Chinese tone can be established.

Secondly, preprocess the pitch contours database. In Chinese fluent speech, there are many special phenomena, such as light-tone, lightly speaking, and so on. In these conditions, the pitch contour of the tone will be influenced by the context and will lose its own characters. So, the light-tone and the lightly speaking syllables should not be taken into account. In our experiment, such syllables have been eliminated.

Then, get the PiCTN model data for each F0 pitch contours of each tone using the above PiCTN method.

Lastly, process the pitch contours data (PiCTN model data) of each tone using the clustering method. Then the dynamic pitch contours characters of four tones can be got, by analyzing the characters, some conclusions can be drawn.

3.3. Similar measurement of pitch contours

Generally speaking, the comparability of the contours should express the curve shape difference of different contours. In figure 2, curve 1 and curve 3 should be more similar than curve 2 and curve 3.

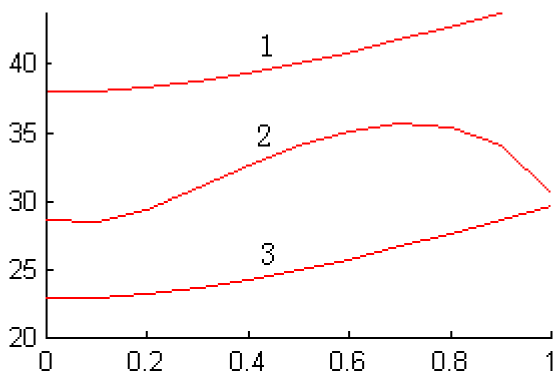


Figure 2: Similar measurement of pitch contours

In the PiCTN model, the normalized duration can be divided into 11 sampling time, that is: $t_0, t_1, \ldots, t_{10}$. And the 11 musical scale pitch data $P_j(t_i)$ can be generated for every pitch contours, where j, i represent the i th data of the j th curve.

In order to classify the curve data, the similar matrix can be gotten, and the element in the similar matrix represents the comparability of the two pitch contours samples.

For j th pitch contours, the average pitch contour can be calculated with the formulation:

$$\overline{P_j} = \frac{1}{11}\sum_{i=0}^{10} P_j(t_i),$$

The new data can be: $\hat{P}_{ji} = P_j(t_i) - \overline{P_j}$,

Then the comparability distance R_{jk} between j th contour and k th contour can be:

$$R_{jk} = \frac{2 \times \sum_{i=0}^{10}\left[\hat{P}_{ji} \times \hat{P}_{ki}\right]}{\sum_{i=0}^{10}\left[\hat{P}_{ji} \times \hat{P}_{ji}\right] + \sum_{i=0}^{10}\left[\hat{P}_{ki} \times \hat{P}_{ki}\right]}$$

Then the similar matrix can be gotten:

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{bmatrix},$$

where n is the sampling data number of the pitch contours for the current tone.

Then the pitch contours can be classified using max-tree method from the similar matrix, and the dynamic characters model of the pitch contours for the tone can be gotten using fuzzy clustering.

3.4. Dynamic model of pitch contours

By above method, and pitch contours can be classified into different types. For each type, the model of the characters can be gotten bye calculating the square root error of the contours which are in the same clustering type,

$$\overline{P}t_i = \frac{1}{n}\sum_{j=1}^{n} P_j(t_i),$$

$$\sigma = \frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{1}{11}\sum_{i=0}^{10}\left[P_j(t_i) - \overline{P}t_i\right]^2},$$

where n is the number of the pitch contours, and $\overline{P}t_i$ is the average frequency value of the i th sampling time.

4. EXPERIMENT RESULT

For the pitch contours of the Chinese tone 1, we got 3031 samples. In fuzzy clustering, we use the max-tree method and the threshold value is 0.85. And as the result, 16 types of classification have been gotten as shown in figure 3.

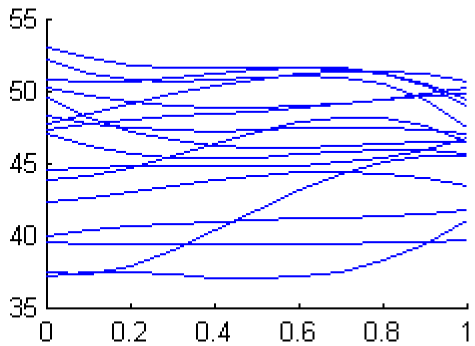


Figure 3: Experiment result. For the Chinese tone 1, 16 types of classification were gotten.

For each type, the mathematical representation is saved and can be used in the later applications.

5. CONCLUSION

In Chinese fluent speech, because of the co-reading, it will be quite different for the pitch contours character of one tone from what it will be while isolated speaking. And, because of the different context in speech, the pitch characters will have many different characters for the same tone.

Depending on the large fluent spoken speech data corpus, the F0 pitch contours for the Chinese Putonghua four tones in fluent speech can be gotten and classified into several types using the method of clustering and statistics. For each type, we can get the mathematical representation, and it will be quiet useful for the prosody modeling in Text-to-Speech (TTS) system and the tone confirmation in speech recognition.

By analyzing the experiment result, some conclusions about the dynamic characters of pitch contours for Chinese Putonghua four tones in fluent speech can be drawn. These are the conclusions: the F0 pitch contour of tone 1 should be high and flat, and sometimes it will be rising with the even speed, the contour of tone 2 will be rising with the accelerated speed, and the pitch contour of tone 3 will drop with the decreasing speed, and lastly the F0 contour of tone 4 should be dropping with the speed increasing.

6. REFERENCE

1. Guo Jinfu. Explanation and Exploration of the Chinese Tone and Intonation. Beijing: Beijing Language and Culture University Express. 1993

2. Y.R.Chao. A system of Tone Letters. Le Maiter Phonetique. 1930, 30, 24~27

3. Z.J.Wu. The Contribution of Y.R.Chao in the Study of Chinese Tone. Transaction of Tsinghua University. 1996, 11(3): 58~63

4. Tao Jianhua, Cai Lianhong. The study and development of the Speech Labeling Tool. Oriental COCOSDA99, Taiwan, 97~101