An Adaptive Algorithm for Music Beat Tracking

Zhanchun Gao, Yuting Liu, Yanjun Jiang
School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China
E-mail: liuyt@bupt.edu.cn

***Abstract*—On the basis of Time Domain Analysis, this paper proposes a method for music beat tracking. Through this method, music beats are detected and tracked by timestamp and intensity value. Generally, the beat-areas of music signal converge more energy than other areas, therefore, the spots of beat can be filtered out by a tracking algorithm with a dynamic threshold value. In this paper, dynamic threshold value in tracking algorithm is modified by using two sliding windows, which are Prediction Window and Detection Window. Also, a new indicator which indicates the stationarity of the signal is proposed. This factor can distinguishes the music signal with rhythmic beats from which with lone-tone and noise in time-domain. The experimental result proves the simplicity, adaptability, and robustness of this method, and it is an efficient algorithm on music beat tracking.**

Keywords—music audio analysis; music beat tracking; extraction of music feature;

I. INTRODUCTION

With the development of internet, the music data human produce everyday grew explosively over the past years. It is a significant subject that how to extract the characteristic parameters from music in both music classification and music recommendation area. As an important time-domain music feature, music rhythm carries significant weight in classifying different music types.

In paper [1], a new method is proposed which presents a multiple agent system that can track the beats of rhythmic music successfully. However, the precondition of experiment is the rhythm of music should be strong enough. Paper [2] introduces a method which uses autocorrelation phase-entropy, to analyze meter and tempo of music. The experimental result shows the method has a *97* percent of accuracy degree in music tempo induction on a data set of *100* songs. Another music extraction method is proposed in paper [3] which is based on low-pass Gaussian filter. Paper [4] presents an adaptive-whitening-based real-time algorithm for music beat tracking. This new algorithm can improve the performance when music signal strongly varying dynamically. Paper [5] and paper [6] present two music feature extracting methods based on audio content. The former is music emotion automatic detection by processing the melody and rhythm of music in MIDI files. The latter presents a feature extraction method based on frequency-domain. Paper [7] introduces the main methods on music audio analysis in recent years. As an important part of music, the note of music can detected by computer as well. A new method is presented in paper [8] which can extract the melody and rhythm feature by music note.

In terms of the recent research, there are two issues that cause the performance problems. The first issue is threshold value that set in beat tracking algorithm is not on the basis of sufficient reason. And the second issue is, the rhythm detection algorithm has good performance only with a precondition that the music rhythm is strong enough. When the music signal is stable or the music signal consists of a lot of long-tones, the performance of the algorithm will declines significantly. To solve these problems, this paper presents a new algorithm that can track beat from the music PCM data with two sliding windows which modify the dynamic threshold value in real-time. And presents a Stability Vector that indicates the stationarity of the music signal.

II. FRAMEWORK OF THE ALGORITHM

The beat tracking algorithm consists of *4* steps. The second step and the third step are the main topics in this paper.

- Preprocess the music audio signal. Filter the signal by high-pass filter in order to gain the high frequency signal which contains most information of beat.
- Calculate Stability Vector of the music signal.
- Filter out beat from music signal based on the Stability Vector.
- Remove the noise.

Minimum time-interval of two different beats is supposed to *0.1*s in this paper.

III. STABILITY VECTOR

The shape of beat-area waveform is different from the others in audio waveform. Generally, signals in beat-area have larger amplitude, so that beat-area signals can bring much more audition stimulus. Hence Stability Vector can be used to declare the stationarity of the music signal. Through this method, music signal can be predicated whether it is a soundtrack with rhythmic beats or with long-tone or noise. If the soundtrack is not stable, that means the soundtrack has rhythmic beats, and the elements in Stability Vector which stand for the beat-area are big values. Meanwhile, value range of the dynamic threshold value in beat tracking algorithm should gave a big value.

To calculate the Stability Vector, the first step is cut the PCM data to chunks C_i with length of L, $L = 0.001 \cdot samplingRate$. It means that each data chunk C_i has L sampling data. The second step is calculate the expectation for each data chunk. The third step is calculate the variance for each *20* expectations. And the last step is calculate the variance for each *5* variances. Each element of Stability Vector stands for $20 \cdot 5 \cdot L$ sampling data, which means $0.1 \cdot samplingRate$, that is *0.1*s music audio signal. Fig. 1 shows the procedure of calculation.

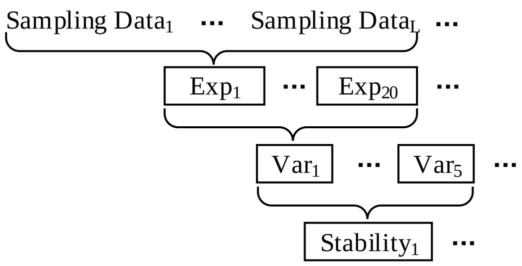

Fig. 1. Calculation of Stability Vector

Fig. 2 shows the typical long-tone music signal, Fig. 3 shows the typical music signal with rhythmic beats. Both signals are after pretreatment. Their Stability Vectors are shown in Fig. 4 and Fig. 5 respectively. As shown in the figure, music signals in beat-area have a big value in Stability Vector and vice versa. The music audio signal pieces can be classified in beat-area, gradient-area and long-tone-area by Stability Vector. The minimum of dynamic threshold value that set in beat tracking algorithm is defined in (1). Variable Ave is the average energy of the signal. ρ is the threshold value between beat-area and gradient-area. The higher ρ is, the lower tolerance for stable music signal. In conclusion, there is a negative correlation between the minimum of dynamic threshold value and Stability Vector. And how to set variable ρ is depends on how to differentiate beat-area and gradient-area.

$$MinThreshold_i = (\rho / StabilityVector_i) \cdot Ave \tag{1}$$

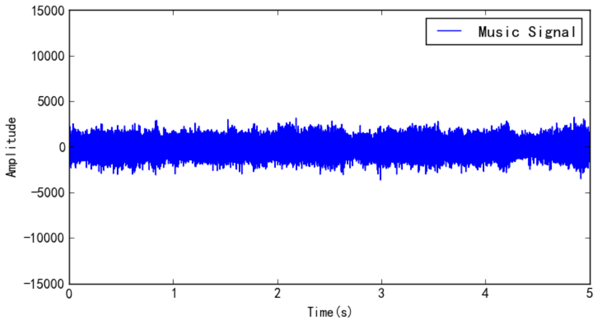

Fig. 2. Typical Long-tone Music Signal

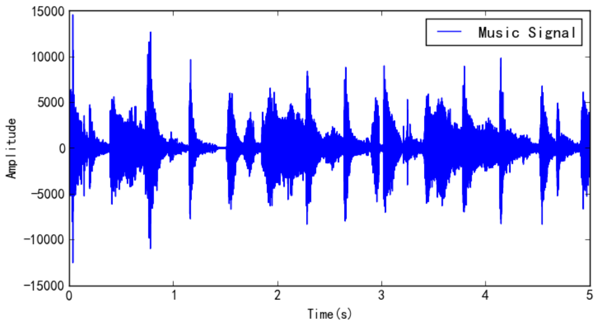

Fig. 3. Typical Music Signal with Rhythmic Beats

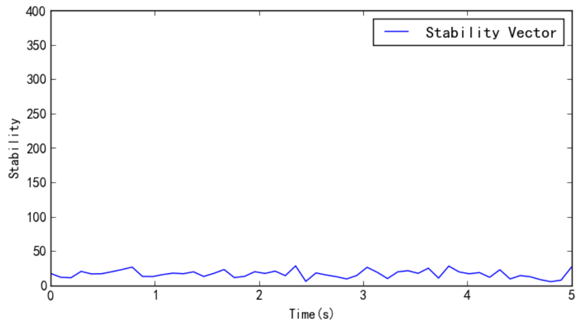

Fig. 4. Stability Vector of Long-tone Music Signal

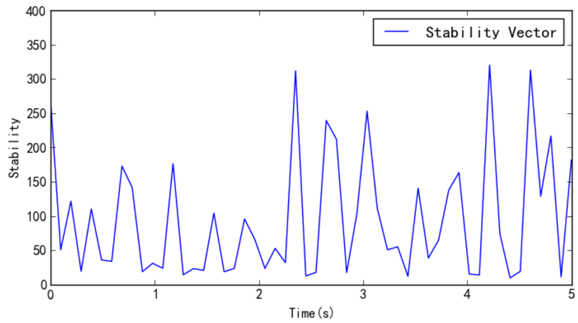

Fig. 5. Stability Vector of Music Signal with Rhythmic Beats

IV. Beat Tracking Algorithm

To filter out the beat from the music signal, the most intuitive approach is set a threshold value, then make the following decision: the signal is a beat if the value of amplitude is over the threshold and vice versa. There are two threshold value setting method so far, static method and dynamic method. Static method sets a fixed value before start algorithm. It is an easy way to tracking beats but with low accuracy. Dynamic method modifies the threshold value dynamically by the feedback of the system. Through dynamic method, beat tracking is not necessarily associated to the energy of music signal, which means the algorithm has better adaptability on most cases. Hence a new adaptive beat tracking method is proposed in this paper which modifies dynamic threshold value by adaptive factor, and the adaptive factor is associates with the feedbacks from Detection Window and Prediction Window. Dynamic threshold value is defined in (2).

$$Threshold = adaptive \cdot Ave \tag{2}$$

In this method, the PCM data flow of music are read in format of short. And the short type data are processed by beat tracking algorithm one by one. The music signal will be kept if this signal value is over the threshold value, and if the signal value is smaller than the threshold value, it will be set to zero. Consequently, a data flow which only have nonzero data in beat-area is gotten. These nonzero data can locate the time-domain position of the beats, and also with the amplitude information of the beats.

A. Detection Window

Detection window is designed to monitor the nonzero ratio in certain window length while beat tracking algorithm is processing. The nonzero ratio indicates the ratio of music signal which is over the threshold value. Constant T_d is defined in Detection Window, if nonzero ratio is higher than T_d, that means a new beat-area is detected by Detection Window or the energy of the signal is getting bigger so that plenty of signal values are higher than threshold value. And the adaptive factor will increased by Detection Window as a result. Detection Window can only increase the adaptive factor but not decrease it. Nonzero ratio is defined in (3).

$$NonzeroRatio = NonzeroCount \,/\, DWindowLength \qquad (3)$$

The length of Detection Window should not too short nor too long. If the length is too short, nonzero ratio will change rapidly. This may cause the thrashing of adaptive factor. If the length is too long, it will reduce the sensitivity of adaptive factor. In this purpose, a compromise should be reached in setting the length of Detection Window. In this paper, Detection Window length is set to *0.01·samplingRate*, T_d is set to *0.05*.

B. Prediction Window

Prediction Window is designed to predict the time interval before next beat appears.

The beats in music present a cyclical relationship. Theoretically, time intervals between beats are always the same. Hence, there exist a linear relation between beats and the timestamps of beats. The linear relation can be defined by a linear model shown in (4). x is the sequence number of the beats, f is the predicted timestamps of the beats.

$$f(x;\omega_0,\omega_1) = \omega_0 + \omega_1 \cdot x \qquad (4)$$

Constant ω_0 and ω_1 in (4) can be defined by minimizing loss function of least-square which is shown in (5).

$$L = \frac{1}{N} \cdot \sum_{n=1}^{N} L_n \cdot (t_n, f(x;\omega_0,\omega_1)) \qquad (5)$$

On the basis of mathematic deduction in [9], ω_1 can be described as (6). This formula presents the calculating method of ω_1. In (4), it is easy to know ω_0 is the value of f when x is zero.

$$\omega_1 = \frac{\overline{xt} - \bar{x}\bar{t}}{\overline{x^2} - (\bar{x})^2} \qquad (6)$$

Consequently, the timestamp of next beat can be predicted in (4). That means it is able to calculate the predicted time interval between current beat and next beat through the least square method. And this interval is the Prediction Window. The length of the window is variable, and new prediction will be made when new beat is detected in Detection Window. Additionally, an original value should be set at the beginning of the algorithm because there is no empirical data at the beginning of the algorithm for prediction.

The function of Prediction Window is decreasing the adaptive factor. The decreasing factor is defined in (7). Variable *ZeroCount* is the length of consecutive zero values that appear in Prediction Window. α is sensitive parameter of decreasing factor. The bigger α is, the faster adaptive factor decreases.

$$DecFactor = \alpha \cdot ZeroCount \,/\, PWindowLength \qquad (7)$$

The decreasing speed of adaptive factor has exponent relation to decreasing factor as shown in (8). Variable *Gradient* is a parameter that indicates how the music audio signal piece is not like a signal with long-tone or noise. Variable *Gradient* is defined in (9). γ is the threshold value between gradient-area and long-tone-area. When the Stability Vector value is under γ, it means this music audio piece is not a beat, it is noise or long-tone more likely.

$$DecSpeed = 2^{DecFactor} \cdot Gradient \qquad (8)$$

$$Gradient = StabilityVector \,/\, \gamma \qquad (9)$$

Exponent relation is defined in relation between decreasing speed and decreasing factor, so that if no beat is detected after a certain time interval that predicted. The threshold value will decrease greatly to avoid the propagation of errors. To explain that, if a linear relation is defined in decreasing speed and decreasing factor, Prediction Window of next beat will be extended when the threshold value is too high and no beat detected, this will slow down the decreasing speed as a result because the decreasing factor has a negative correlation with the length of Prediction Window. If the energy of next piece music signal is not as bigger as the one at present, the algorithm is failed that no beat will detected. Additionally, *Gradient* is another determinant factor of decreasing speed.

As the presentation of Detection Window and Prediction Window, the dynamic threshold value is modified by these two window timely and automatically. The domain of threshold value is *MinThreshold* which defined earlier to the maximum of the music signal. Fig. 6 shows a music piece with rhythmic beats, Fig. 7 shows a music piece with lone-tone, Fig. 8 and Fig. 9 shows the beats which tracked by the beat tracking algorithm of these two pieces. As the figures shown, beat tracking algorithm has convincing adaptability that the dynamic threshold value is able to filter out the beats. For the long-tone music piece, this algorithm avoids the long-tone signals and noise signals so that this method also has high accuracy.

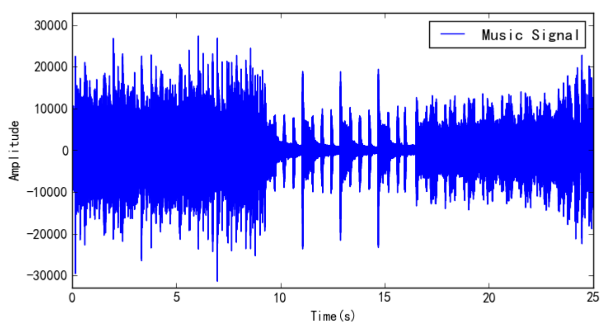

Fig. 6. Music Piece with Rhythmic Beats

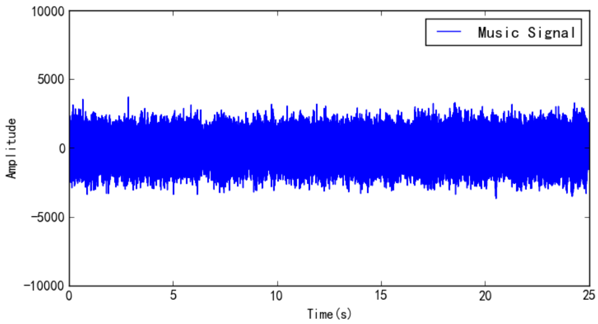

Fig. 7. Music Piece with Long-tone

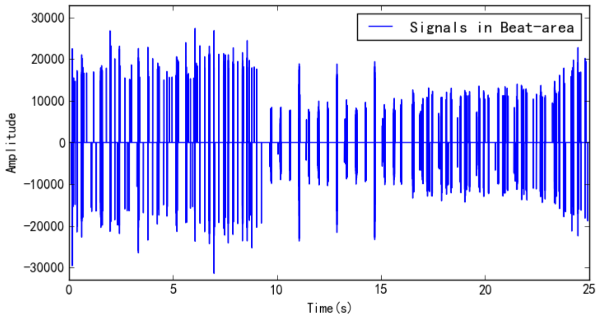

Fig. 8. Beats of Rhythmic Music that Tracked by Beat Tracking Algorithm

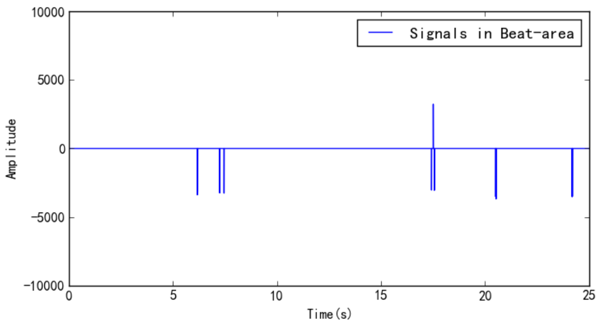

Fig. 9. Beats of Long-tone Music that Tracked by Beat Tracking Algorithm

From the experimental result, it shows that this algorithm has good astringency and rapid convergence ability. In addition, as the assumption earlier that the minimum time-interval of two different beats is *0.1*s, hence a noise reduction operation is needed afterwards because there should not exist two beats in *0.1*s. The beats of rhythmic music after noise reduction is shown in Fig. 10. From the figure, the timestamp of the beats can be exactly located in time-domain. The error range of the timestamp is in *20*ms by manual inspection. The reason causes the error is the wave length of beats are unequal. Beat tracking algorithm considers the first signal which over the threshold value as a beat, however this signal is not always the center of the beat wave. This causes the shifting of beats. The amplitude of the beats is also presents in Fig. 10, that means not only the timestamp of the beats are indicated, but also the energy of beats. It is of great significance that analyzing the music data in time-domain by these features.

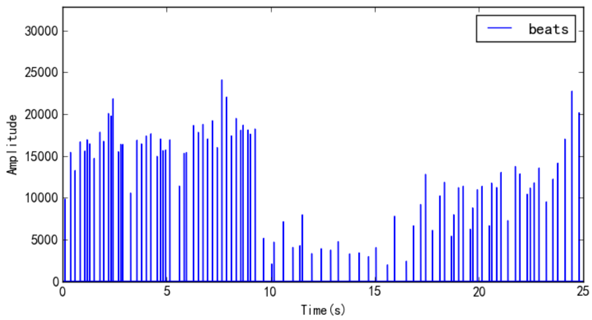

Fig. 10. Beats of Rhythmic Music After Noise Reduction

V. CONCLUSIONS

In this paper, a new algorithm on music beat tracking is presented. Through this method, a new indicator Stability Vector is proposed which indicates the stationarity of music pieces. Then a new tracking algorithm is introduced. In this algorithm, Detection Window and Prediction Window are built to modify the adaptive factor in order to adapt the threshold value for current music signal waveform automatically. By this method, error on long-tone music piece and noise are reduced effectively, and the dynamic threshold value is set reasonable. Experimental result shows that this algorithm has good adaptability and strong universality.

REFERENCES

[1] Qingyu Zeng, Xiuqing Ye, Donghui Wu, "Music Beat Induction Algorithm Based on Multi-agent," Journal of Circuits and Systems, 2005, Vol. 10, No. 1.

[2] Zhe Chen, Jieping Xu, "Content Based Music Beat Tracking," ACTA Electronica Sinica, 2009, Vol. 37, No. 12A.

[3] Jianjian Hu, Peifeng Zheng, Liping Tang, Zhuping Zhang, "Music Beat Extraction Based on Low-Pass Gaussian Filter," Journal of DongHua University(Natural Science), 2011, Vol. 37, No. 1.

[4] Yue Wang, Lei Xie, Yulian Yang, "Adaptive Whitening for Real-Time Music Beat Tracking," Application Research of Computers, 2009, Vol. 26, No. 5

[5] Qiong Peng, Cheng Zhi, "Key Issues in Music Emotion Automatic Detection by Computer," Elementary Electroacoustics, 2008, Vol. 32, No. 4.

[6] Jiming Zheng, Guohua Wei, Yu Wu, "New Effective Method on Content Based Audio Feature Extraction," Computer Engineering and Applications, 2009, 45(12): 131-133.

[7] Yibin Zhang, Jie Zhou, Zhaoqi Bian, Jun Guo, "A ReView of Content−Based Audio and Music Analysis," Chinese Journal of Computers, 2007, Vol. 30, No. 5.

[8] Xiaolan Lin, Xiaoguang Wang, hui Wang, "Key Technologies Research on Content-based Music Information Retrieval," Journal of Communication University of Chine (Science and Technology), 2010, Vol. 17, No. 4.

[9] Simon Rogers, Mark Girolami, A First Course in Machine Learning, CRC Press, Taylor & Francis Group, 2012