

A Hidden Markov Model Approach to Musical Beat Tracking

Yu Shiu and C.-C. Jay Kuo

Ming Hsieh Department of Electrical Engineering and Signal and Image Processing Institute
University of Southern California, Los Angeles, CA 90089-2564

E-mails: atoultaro@gmail.com, cckuo@sipi.usc.edu

Abstract—An automatic musical beat tracking algorithm using the hidden Markov model (HMM) approach is proposed. Musical onsets are used as observation. A state space called the Periodic Left-to-Right (PLTR) model is used to model the dynamics of beat progression. PLTR consists of multiple hidden states. These hidden states can be categorized into two state types: the beat state type and the non-beat state type. Their probability distributions are modeled by the Gamma distribution with different parameters. The Viterbi algorithm is adopted to find bit positions in a given music clip. It is shown by experimental results that the proposed algorithm can achieve 78.63% for MIREX data set and 96.23% Billboard Top10 data set, respectively.

Index Terms—Beat tracking, hidden Markov model (HMM), music signal processing, music information retrieval.

I. INTRODUCTION

Automatic musical beat tracking by computers can be done either on-line or off-line. On-line beat tracking algorithms [2]–[4] attempt to detect the beat location from audio waveforms on the fly. In contrast, off-line beat tracking algorithms [5]–[8] determine all beat positions of beats from a given music piece.

The hidden Markov model (HMM) has been widely used in speech recognition and other applications for several decades [1]. It can model the dynamics of a class of time series effectively such as speech signals. Temporal dependency of a time-varying signal is modeled by the state transition characterized by a first order Markov chain. Parameters of the probabilistic signal model and its dynamics are learned through a large amount of training data to capture the statistical characteristics for a particular audio input segment (*e.g.* a speech sound unit). In this work, a musical beat tracking algorithm based on HMM is proposed to determine the beat locations.

We choose a specific state space to model the dynamics of beat progression; namely, the Periodic Left-to-Right (PLTR) model. PLTR consists of N hidden states. These hidden states can be categorized into two state types: the beat state type and the non-beat state type. Only one state belongs to the beat state type while the other $N - 1$ states belong to the non-beat state type. Their probability distributions can be modeled by the Gamma distribution with different parameters. The musical onset signal gives the observation sequence. Once the framework is set, the Viterbi algorithm can be used to decode the optimal state sequence and determine beat locations.

The framework of our beat tracking system is shown in Fig. 1. The input to the system is the musical audio signal. The system consists of three main modules. First, the musical onset signal is estimated from the audio signal. Second, its period within musical onset signals is estimated. Finally, the HMM-based musical beat tracking algorithm is used to estimate all

beat locations. In this paper, we assume that the onset detection module and the period estimation module are both available, and we can focus on the HMM-based beating tracking module only.

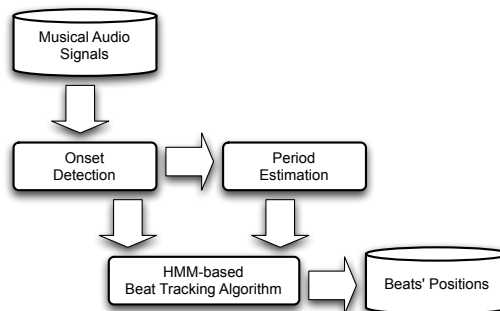


Fig. 1. An overview of the HMM-based musical beat tracking system.

II. MUSICAL DATA PRE-PROCESSING: ONSET DETECTION AND PERIOD ESTIMATION

The musical onset signal provides the intensity change of musical contents along time. It reflects two types of music content changes: instantaneous noise-like pulses caused by percussion instruments and changes of music pitches/harmonies due to the new note arrival. In our work, the cepstral distance method [1] is used to calculate musical onsets. The process is detailed below.

First, the music contents is represented via mel-scale frequency cepstral coefficients (MFCC), $c_m(n)$, for each shifting window of 20-msec with 50% overlap, where $m = 0, 1, \dots, L$ is the order of the cepstral coefficient and n is the time index. The first four low-order coefficients $c_0(n)$, $c_1(n)$, $c_2(n)$ and $c_3(n)$ are used for the computation. Then, the selected MFCCs are smoothed over p consecutive frames $c_m(n)$. In our implementation, $p = 3$ is used. Finally, we compute the change of spectral contents by examining the MFCC difference between the two adjacent smoothed cepstral coefficients $\bar{c}_m(n)$. The mel-scale cepstral distance is chosen to be the musical onset detection function at time n .

$$d(n) = \sum_{m=1}^L (\bar{c}_m(n) - \bar{c}_m(n-1))^2, \quad (1)$$

The tempo and its inverse (*i.e.* period) are assumed to be perceptually fixed in our beat tracking system. They need to be estimated before the actual task is conducted. One can estimate it by the autocorrelation function (ACF) of musical

onset signals. However, there often exists confusion between the real period and its double/half-period (or triple/one-third-period for the triplet case). We will not address the problem since our focus is the “tracking” of beats. A period is selected manually within the range of interests as an input parameter to the system.

III. PROBABILITY DISTRIBUTION OF OBSERVATIONS

Consider an HMM with N states, there are three sets of parameters as denoted by

$$\lambda = (\mathbf{A}, \mathbf{B}, \pi), \quad (2)$$

where $\mathbf{A} = \{a_{i,j}\}$ is the state transition probability distribution, $\mathbf{B} = \{b_j(k)\}$ is the probability distribution of observations on state j , and π is the initial state distribution [1]. The time axis is uniformly partitioned by a basic time unit so that $k = 1, 2, \dots$ denotes the discrete time index. The application of HMM to the music beat-tracking problem is detailed below.

A. Observations for Each State Type

The proposed beat-tracking HMM has N_0 states, which can be classified into two types; namely, the beat state type (B) and the non-beat state type (N) depending on whether a beat occurs or not inside a time unit. Observations are the musical onset intensity in each time unit.

Not all beats yield strong musical onsets. For example, even though rest notes in the end of a musical section do not have significant musical onsets, human can still perceive or believe the progression of beats. On the other hand, the occurrence of a strong musical onset does not necessarily correspond to the occurrence of a beat. Musical onsets reflect the change of music activities in the audio spectrum caused by the change of music notes and/or instruments but are not limited to beats.

To estimate the observation probability distribution of beat and non-beat state types, training data are used. Both the musical onset signal and the annotated beat location are given in the training data. Musical onset signals are collected in the vicinity of each annotated beat location to estimate the observational probability distribution of the beat state type. Likewise, the remaining musical onset signals that are away from annotated beats' time are used in estimating the observational probability distribution of the non-beat state type.

B. Observation Probability Estimation

The observations, musical onsets, are an one-dimensional signal with non-negative values. To estimate the observation probability for a given state type (denoted by $P(o(k)|j = B)$ or $P(o(k)|j = N)$ for the beat and the non-beat state types, respectively), we consider both parametric and non-parametric approaches.

The data set from MIREX 2006 beat tracking competition [9] is used to determine the observation probability conditioned on beat and non-beat. It consists of twenty 30-sec music clips with annotated beat location. Histograms of music onsets conditioned on beat/non-beat state types are shown in Fig. 2,

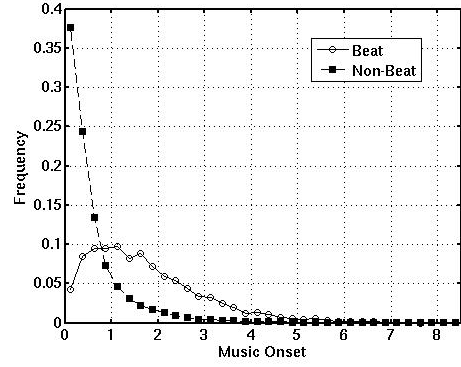


Fig. 2. The histogram of music onsets' intensities for I_B and I_N with the bin width equal to 0.25, where the x-axis is the music onset intensity while the y-axis is the frequency of occurrence.

where bin centers are uniformly located from 0.125 to 8.125 seconds with bin width 0.25 second.

We see from Fig. 2 that the onset distribution for non-beats, $P(o(k)|j = N)$, heavily concentrates on small onset values with few large onset values. The non-beat state type has a much higher probability than the beat state type at the first three bins. From the 4th bin, the musical onset probability for beats is larger than that for non-beats. Both observation probabilities $P(o(t)|B)$ for beat state type and $P(o(t)|N)$ for non-beat state type can be approximated by a parametric model of the Gamma distribution in the following form

$$f(x|k, \theta) = \frac{e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} x^{k-1},$$

where θ is a scale parameter and

$$\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt,$$

is the Gamma function of shape parameter k . To find parameters k and θ , the maximum likelihood estimation (MLE) method is used. To approximate the histograms in Fig.2, parameters are $k = 0.949$ and $\theta = 0.904$ for the beat state type and $k = 0.581$ and $\theta = 1.045$ for the non-beat state type.

IV. STATE TRANSITION MODELING

A. Periodic Left-to-Right (PLTR) Model

The left-to-right (LTR) model (or called the Bakis model) [1] is often adopted to model a time-varying signal, where each node denotes a state. In the LTR model, the allowed transition between the current state $s(k)$ and the next state $s(k+1)$ is either to proceed to the next state in the right or to stay at the current state via self-looping. No transition is allowed from a state to any of its left states. The LTR model is used to model signals that change over time in a successive manner. Since the right hand side represents the direction into which time proceeds, the left to right state transition means time progression.

In this work, a state space for HMM is proposed to model the temporal progression of beats. It is called periodic left-to-right model (PLTR) as shown in Fig. 3. In PLTR, there are totally N_0 states, where the first state belongs to the beat

state type while others belong to the non-beat state type. It models the periodic pulses from the beats. States of the same state type have the same observation probability distribution. To model the temporal progression of multiple beats, we also add a large-loop state transition from the right-utmost state back to the left-utmost state.

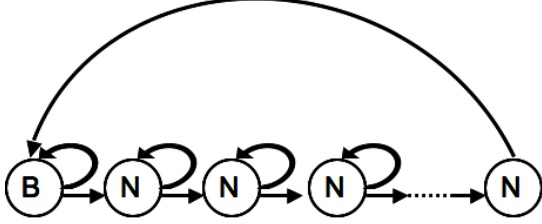


Fig. 3. The PLTR model with large-loop state transition for modeling the periodic repetition of beats.

B. State Transition Probability a_{ij}

The probability distribution $Prob(d)$ for the PLTR model with a self-loop at every state can be analyzed as

$$Prob(d) = \binom{d-1}{N_0-1} p^{d-N_0} (1-p)^{N_0}. \quad (3)$$

The expected beat period from the PLTR model of N_0 states is

$$E\{d_{N_0}\} = N_0 E\{d\} = \frac{N_0}{1-p}. \quad (4)$$

By setting $E\{d_{N_0}\} = T_0$, we can determine the value of N_0 via

$$N_0 = T_0 \cdot (1-p). \quad (5)$$

The p value can be roughly estimated from (5) by

$$p = \frac{T_0 - N_0}{T_0}. \quad (6)$$

Please note that the value of $T_0 - N_0$ is approximately one half of the window size. Since $p = \frac{1}{16}$ or $\frac{1}{32}$ is used in this work, the corresponding window size is about one-eighth of beat period ($\frac{T_0}{8}$) or one-sixteenth of beat period ($\frac{T_0}{16}$) respectively.

Finally, the initial probability distributions are assumed to be uniformly distributed at N_0 states; namely,

$$\pi_i = \frac{1}{N_0}, \quad i = 0, \dots, N_0 - 1. \quad (7)$$

V. OPTIMAL STATE SEQUENCE DECODING

Given an HMM model, λ , and a sequence of observations, $o(t)$, $t = 1, 2, \dots, T$, the Viterbi algorithm is used to decode the optimal state sequence $\hat{S} = \{\hat{s}(1), \hat{s}(2), \dots, \hat{s}(T)\}$ that has the largest posterior probability conditioned on observational sequence $o(t)$ among all possible state sequences. Mathematically, this can be written as

$$\begin{aligned} \hat{S} &= \{\hat{s}(1), \hat{s}(2), \dots, \hat{s}(T)\} \\ &= \arg \max_{s(1), s(2), \dots, s(T)} P(s(1), s(2), \dots, s(T) | o(1), o(2), \dots, o(T)) \\ &= \arg \max_{s(1), s(2), \dots, s(T)} P(s(1), s(2), \dots, s(T)) P(o(1), o(2), \dots, o(T) | s(1), s(2), \dots, s(T)). \end{aligned} \quad (8)$$

HMM assumes the first order Markov chain in the state space and that observation $o(t_0)$ depends only on state $s(t_0)$ and is independent of observations at $t \neq t_0$. Thus, we can derive the following optimization problem from Eqs. (8) as

$$\begin{aligned} \hat{S} &= \arg \max_{s(1), \dots, s(T)} P(s(1)) P(o(1) | s(1)) \cdot \\ &\quad \prod_{t=2}^T P(s(t) | s(t-1)) P(o(t) | s(t)), \end{aligned} \quad (9)$$

where $P(s_1)$ is the initial probability distribution of a certain state type. It is well known that the Viterbi algorithm can be applied to (9) to decode the best state sequence $s(1), s(2), \dots, s(T)$, which corresponds to beat locations in the beat sequence.

One example of state sequence decoding is shown in Fig. 4. The test music is the 4th music clip in the MIREX beat tracking competition data, where a female vocal singer performs in the Jazz or R&B style. The accompanying music and percussion instruments provide a clear and specialized Jazz rhythm against the singer's vocal. In some segments, beats are easy to follow due to strong musical onsets and concrete periodicity. In other segments, beats are irregular and, therefore, difficult to follow. The beats of this music clip cannot be easily followed as most pop and rock songs, in which percussion instruments are heavily used.

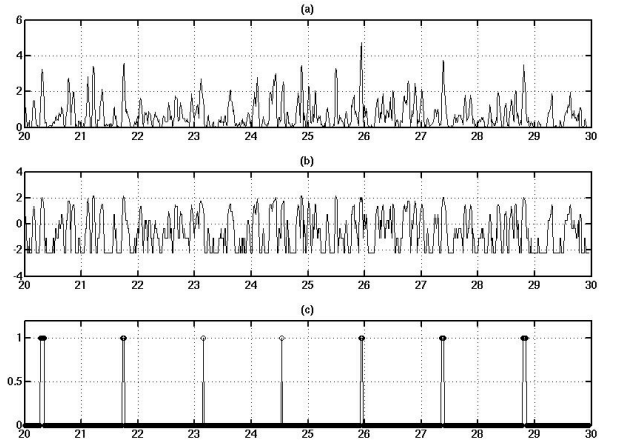


Fig. 4. An example of the decoded state sequence based on the 4th music clip in MIREX: (a) the musical onset, (b) the log-likelihood difference between the beat and the non-beat state types, (c) the decoded state sequence, where "1" represents beat state and "0" represents non-beat state, as a function of time (in the unit of second).

Results for the musical segment from 20 to 30 seconds is shown in Fig. 4. Fig. 4(a) shows the musical onset extracted from the musical audio signal. Fig. 4(b) shows the difference between the log-likelihood of the observation of the beat and the non-beat state types. Larger values in Fig. 4(b) means the observation is more likely to come from a beat state while smaller values means the observation is less likely to come from a beat state.

It is difficult for human to find beat locations via visual inspection on Fig. 4(a). In particular, musical onsets look

“congested” between 20 and 22 seconds and between 24 and 26 seconds. Even though beats usually correspond to larger musical onsets, they are buried in the duration of busy music activities due to vocal sounds and musical instruments. HMM provides a way to clean the data for us. HMM incorporates the information of periodic cycles of beat pulses. Only those periodic signals of pre-designed period T_0 can achieve a high likelihood along time. Fig. 4(c) shows the resulting state sequence, where value 1 represents a beat state and the value 0 represents a non-beat state. Beats can be extracted from the “congested” musical onsets in Fig. 4(a) since the proposed HMM applies the beat periodicity constraint.

VI. EXPERIMENTAL RESULTS

A. Experimental Data and Setup

Two data sets were used in the evaluation of HMM-based musical beat tracking. They were the MIREX 2006 beat tracking competition practice data and the Billboard Top10 songs in 80’s. The first 5 seconds of each music clip were used to calculate the beat period via the autocorrelation function. The remaining music clips were used for the performance evaluation of HMM-based musical beat tracking.

The observation probabilities of the beat and the non-beat state types were trained using the leave-one-out technique. For example, when the performance on the 1st music clip of MIREX 2006 data was tested by the parametric approach, parameters of the Gamma distribution were estimated from the remaining 19 music clips, and the resulting probability model was used to calculate the observation probabilities for the 1st music clip.

B. P-Score Performance Evaluation

The P -score and the longest correctly tracked music segment are the two metrics widely used to evaluate the performance of musical beat tracking. Since the HMM-based musical beat tracking aims at off-line beat tracking, the former metric is more appropriate. The P -score is defined as

$$P = \frac{1}{N_{max}} \sum_{n=-\infty}^{\infty} \sum_{m=-w}^w \tau_d(n)\tau_g(n-m), \quad (10)$$

where τ_d and τ_g have values 1.0 on the detected location and the ground-truth beat location, respectively,

$$N_{max} = \max(N_d, N_g), \quad (11)$$

and where N_d is detected beat number, N_g is ground-truth beat number. The window size W used throughout our experiments is 20% of the beat period.

We first evaluated the performance of several HMM settings applied to the MIREX data set. The first was the way in observation probability calculation: 1) non-parametric (histogram analysis with bin width 0.25) and 2) parametric via the Gamma distribution. The second was the value of self-loop probability $a_{ii} \triangleq p$ for all states. The performance of $p = \frac{1}{16}$ and $\frac{1}{32}$ was compared. The P -Score performance is shown in Table I. We see very similar performance under different settings.

TABLE I
PERFORMANCES OF HMM-BASED MUSICAL BEAT TRACKING WITH VARIOUS SETTINGS APPLIED TO THE MIREX DATA SET.

	Non-Parameterized	Gamma Distribution
$p = 1/16$	74.65%	77.76%
$p = 1/32$	78.63%	76.65%

Then, we evaluated the performance of the same HMM settings applied to the Billboard Top10 data set. The results are shown in Table II.

TABLE II
PERFORMANCES OF HMM-BASED MUSICAL BEAT TRACKING WITH VARIOUS SETTINGS APPLIED TO THE BILLBOARD TOP10 DATA SET.

	Non-Parameterized	Gamma Distribution
$p = 1/16$	91.67%	91.42%
$p = 1/32$	96.23%	97.53%

As compared with the results of the MIREX data set in Table II, the results of Table I are significantly better. Recall that the Billboard Top10 data set consists mainly of genres such as pop and rock while the MIREX data set has diverse genres from classical music, jazz to folk music. Pop and rock use a lot of percussion instruments. Usually, they have more regular and stronger beats than other genres such as classical and jazz music. Regular beats fit into the HMM model, which enforces the periodicity of beats. Stronger beats yield higher observation probability for the beat state type, and make themselves more easily decoded by the Viterbi algorithm.

VII. CONCLUSION

A musical beat tracking algorithm based on hidden Markov model(HMM) was proposed. A special state space is used to model the dynamics of beat progression. States belong to two state types: the beat type and the non-beat type. The Viterbi algorithm can be used to decode the optimal state sequence and, thus, indicates the time of beats through the beat state type. Experimental results are provided for the proposed HMM method.

REFERENCES

- [1] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proceedings of the IEEE*, vol. 77, No. 2, pp. 257-286, 1989.
- [2] A. T. Cemgil, B. Kappen, P. Desain and H. Honing, “On tempo tracking: tempogram representation and Kalman filtering”, *Journal of New Music Research*, 2001.
- [3] S. Hainsworth and M. Macleod, “Beat tracking with particle filtering algorithms”, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 91-94, 2003.
- [4] W. A. Sethares, R. D. Morris and J. C. Sethares, “Beat tracking of musical performances using low-level audio features”, *IEEE Trans. on Speech and Audio Processing*, vol. 13, No. 2, pp. 275-285, 2005.
- [5] S. Dixon, “Automatic extraction of tempo and beat from expressive performances”, *Journal of New Music Research*, 2001.
- [6] F. Gouyon and S. Dixon, “A review of automatic rhythm description systems”, *Computer Music Journal*, vol. 29, No. 1, pp. 34-54, 2005.
- [7] A. P. Klapuri, A.J. Eronen and J.T. Astola, ‘Analysis of the Meter of Acoustic Musical Signals’, *IEEE Trans. on Speech and Audio Processing*, vol. 14, No. 1, pp. 342-355, 2006.
- [8] E. Scheirer, “Tempo and beat analysis of acoustic musical signals”, *Journal of Acoustic Society America*, vol. 103, pp. 588-601, 1998.
- [9] “Music Information Retrieval Evaluation eXchange (MIREX) Competition”, <http://www.music-ir.org/mirexwiki/index.php>, 2006.