International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Musical Audio Beat Tracking using Hidden Markov Model

Lata Dhulekar[1], S. K. Shah[2]

[1]Smt.Kashibai Navale College of Engineering, vadgaon (Bk), Pune, India

[2]Smt. Kashibai Navale College of Engineering, E&TC Department, vadgaon (Bk), Pune, India

Abstract: ***The proposed method describes the beat tracking system of musical audio signals. Here, the beat tracking deals with the explicit modeling of non-beat states and estimates the time between consecutive beat events. The expected accuracy of the estimated beats is also provided additionally. To predict the accuracy of the beat estimate a k-nearest neighbor regression algorithm is used. A database of 222 musical signals of various genres is used to statistically evaluate the performance for beat tracking system. A new perspective for beat detection is presented which can be used to enhance the performance of the proposed algorithm and relates it with the human tapping.***

Keywords: Beat-tracking, onset detection, k-nearest neighbor (k-NN) regression, Hidden Markov Model, Musical Audio Processing.

1. Introduction

Beat tracking is considered as one of the most challenging subject in the music-audio signal processing and it is defined as locating the times in an audio signal where beats are perceived or notated in the corresponding score. The act of tapping one's foot in time to music is computationally equivalent to beat tracking. It is a problem of both periodicity detection and location of the periods inside a signal. The basic unit of music is beat which describes the individual temporal events that define the metrical level. Beat provides the temporal structure of an audio signal which is useful in music information retrieval (MIR) research. The various application of beat tracking includes cover-song detection [2], music similarity [3], chord estimation [4], and music transcription [5].

The paper is organized as - Section 2 provides literature survey on various methods for beat tracking of musical audio signals. Section 3 briefly explains beat tracking system. Experimentation of the proposed system is described in Section 4. Results and discussions for the given system are presented in Section 5, followed by conclusion and future scope in Section 6 and Section 7 respectively.

2. Literature Survey

This paper concerns the beat tracking problem, the early approach for the same are described below:

- A multi-agent approach proposed by Dixon [6] *Beatroot*, extracted sequence of note onset times from an audio signal or from a symbolic representation is processed. From clustering inter-onset-intervals it derives tempo hypotheses which are used to form multiple beat agents with varying tempo and phase. To track beats in expressively performed music Dixon's algorithm is designed.
- Agent based Goto's approach [7], analysis for tracking of beats (at the ¼ note level) is additionally extended to the ½ and whole note levels. Across 7 parallel subbands spectral models are used to extract share and bass drum events in an onset analysis. To infer the beats and higher metrical level structure chord changes and predefined rhythmic pattern templates are used. In real-time, provided the input signal has a steady tempo and a 4/4 time signature i.e. four beats per bar, Goto's system operates accurately.
- The particle filtering is used in [9] proposed by Hainsworth where the note onsets are extracted from subband analysis using two distinct modes. The first mode is for finding transient events and another is designed to detect harmonic change. This particle filtering statistical framework for beat extraction which are modeled as a quasi- periodic sequence driven by time-varying tempo and phase processes. Computational complexity is being the main limitation of this technique.
- Klapuri proposed a Hidden Markov Model (HMM) in [11] to adopt a more robust registral accent signal across four parallel analysis bands as the input to their system. In this method comb filters banks are used within a probabilistic framework to simultaneously track 3 metrical levels. The first metrical level *tantum* is the lowest or fastest metrical level. The second level *tactus* is the tapping rate of human foot and the third metrical level *measure* denotes the grouping of beats into bars. The performance results between for Klapuri tested over a large annotated database.
- A probabilistic framework formulated as an inverse viterbi problem introduced in [12] proposed by Peeters. In this approach Peeters choose to decode the sequence of beats along time over beat-numbers. As the beat template is used to model tempo-related expectations on an onset signal, the system calculates the observation likelihood through a cross- correlation of the onset signal and the estimated beat template.
- A dataset needed to learn this template and musical genre decides the results. Some issues that need to be addressed in spite of different beat at a particular time either using a single observation shown in [8] and [11] or using correlation template in [12] modeled in early approaches. The extra information provided by non-beat observation, which needed to be exploited to track beats.

3. Beat Tracking

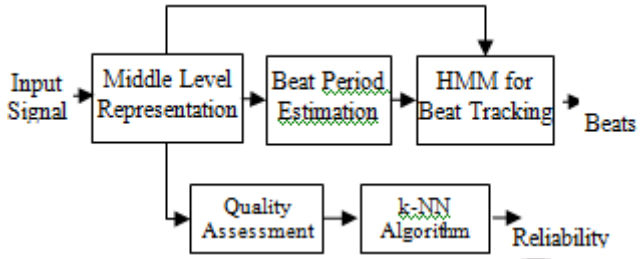

Figure 1: Overview of proposed beat tracking system

A probabilistic framework that models the time between consecutive beat events and exploits both beat and non-beat signal observations is proposed to integrate musical-knowledge and signal observations. The expected performance of a beat tracking algorithm is general and can be potentially be extended to any other system by identifying its own limitations.

The overview of the proposed best tracking system is shown in Figure 1. that analyses the input musical signal and extracts a beat phase and a best period salience observation is then used to calculate beat period. Then, beat period estimation and phase observation signal together taken as input parameters by probabilistic model i.e. HMM. Finally, to measure reliability of the beat estimates k-NN regression algorithm is used.

4. Experimentation

The difference elements of beat tracking method illustrated in Figure.1 described in this section.

4.1 Feature Extraction

The location of transients in the original audio signal revealed by an onset detection function i.e. mid-level representation in beat tracking. The reference method selected is the complex spectral difference [13] emphasizing onsets due to change of spectral energy. The shift-invariant comb filterbank approach is adopted described in [10] for analyzing periodicity of phase observation signal.

4.2 Beat Period Tracking

In beat tracking system the beat period and phases are estimated independently. Assume transition probabilities are modeled using Gaussian distribution and beat period as slowly varying process for beat period tracking.

4.3 HMM For Beat Tracking

Beat events can be determined by defining a context as music is highly structured in terms of the temporal ordering of musical events. Beats are regularly spaced in time with small deviations from the beat period. A hidden Markov model (HMM) is used to integrate this contextual knowledge with signal observations and then estimate beat phases.

A first-order HMM is defined in the proposed beat tracking system where a hidden variable ϕ represents the phase state and measures the elapsed time, in frames, since the last beat event. The estimated beat period τ_t is used to determine the total number of states $N\tau_t$. The states for ϕ are {0, 1, 2,….., $N\tau_t$-1}. A particular state sequence (ϕ_1, ϕ_2,……,ϕ_T) denoted as $\phi_{1:T}$ and a state at time frame t is denoted as ϕ_t.

The state transition probabilities $a_{ij} = P(\phi_{t=j} \mid \phi_{t-1} = i)$ encodes the temporal structure of beat signal and then phase state variable ϕ_{t-1} measures the elapsed time due to the last visit to the beat state 0 at time t-1. The phase observation signal o(t) is the observation variable for the phase state o_t i.e. $o_t = o(t)$ in the following. Assume o_t to be independent of any other state given the current state and then the state-conditional observation probability is $P(o_t \mid \phi_t)$.

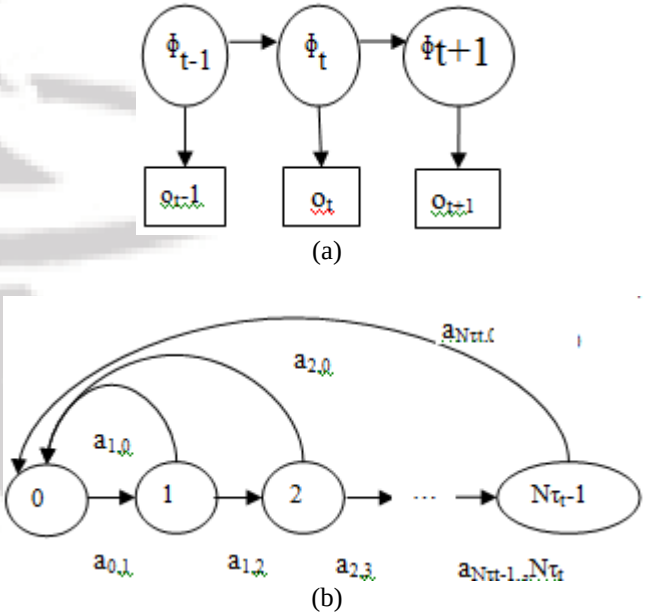

Figure 2: HMM for beat tracking. (a) hidden state and observation variables conditional dependencies. (b) State transition diagram for the hidden state ϕ_t [1].

Figure 2 (a) shows first order HMM model introduced above. The circles show the hidden variable ϕ_t and boxes shows the observation of o_t variable. Links between the state and observation variables represents the conditional dependencies. In Figure 2(b) transition between the hidden states are shown where the state are represented by circles and transitions by links.

1. Estimation Goal:

The aim of the proposed system is estimation of the sequence of beats which best explains the phase observation, o_t. The set of observation $o_{1:T}$ is obtained from hidden states $\phi^*_{1:T}$ and estimated as [1]

$$\Phi^*_{1:T} = \text{argmax } P(\phi_{1:T} | o_{1:T}) \quad (1)$$

Where T denotes the number of frames of the input musical signal. To obtain the set of beat times B^*, select the time instants where the sequence $\phi^*_{1:T}$ visited the beat state. Thus,

$$B^* = \{t : \phi^*_t = 0\} \quad (2)$$

As assumed in Fig. 2(a), the posterior probability of (1) can be written as

$$P(\phi_{1:T}|o_{1:T}) \alpha P(\phi_1) \Pi_{t=2}^{T} P(o_t | \phi_t) P(\phi_t| \phi_{t-1}) \quad (3)$$

Where $P(\phi_1)$ is the initial state distribution, $P(\phi_t | \phi_{t-1})$ is the transition probabilities and $P(o_t|\phi_t)$ is the observation likelihoods.

2. Estimation of observation likelihood:

The $N\tau$ states of the model are estimated from the observation likelihoods $P(o_t | \phi_t)$. The phase observation signal o_t is designed to show large values at event locations as discussed earlier. Assume beat state observation likelihood $P(o_t | \phi_t=0)$ –

$$P(o_t | \phi_t =0) \alpha o_t \quad (4)$$

To obtain reasonable estimates for non-beat state observation likelihood functions $\{P(o_t | \phi_t = n) : n \neq 0\}$ similar assumptions are used. Again, assume that the phase observation signal o_t will show small values at the non-beat state observation likelihoods $\{P(o_t | \phi_t = n) : n \neq 0\}$ is

$$P(o_t | \phi_t = n) \alpha 1-o_t \quad (5)$$

These are equivalent as used in first-order polynomial model the state-conditional distributions.

3. Estimation of the initial and transition probabilities:

The time instant when the first beat is expected to be modeled in the initial probability $P(\phi_1)$. A discrete uniform distribution for $P(\phi_1)$ is chosen , as no assumptions are made over the location of the first beat. Modeling the probability density function of time between consecutive beats at any time instant, Δ, to be proportional to a Gaussian distribution centered at the beat period τ_t as –

$$P(\Delta = n) \alpha (1/\sqrt{2\pi\sigma^2})\exp(-(n-\tau_t)^2 / 2\sigma^2) \quad (6)$$

Where σ denotes the standard deviation. A value of 0.02 is chosen for the standard deviation [5] then σ = 1.72 frames. The largest time between beats allowed determines the number of HMM. Assume maximum time between beats τ_t +3σ. Hence, the total number of states given by

$$N\tau_t = \tau_t +3\sigma+1 \quad (7)$$

If Δ = n frames between two consecutive beats then the state transition probabilities $a_{ij} = P(\phi_{t=j} | \phi_{t-1} = i)$ and the time distribution between beats P(Δ) related as

$$a_{n-1,0} = \frac{P(\Delta = n)}{\prod_{k=0}^{n-1} a_{k,k+1}} \quad (8)$$

$$a_{n-1,n} = 1-a_{n-1,0} \quad (9)$$

with $n \in \{1,\ldots, N\tau_t\}$.

4.4 Beat Tracking Quality Assessment

The correctness of the beat period estimation decides the behavior of the probability frame work proposed. For beat period estimation, three measures are calculated in order to characterize the quality of the feature signals. A peak-to-average ratio, q_{par}, is the first measure that relates the maximum amplitude of the beat period salience observation signal with its root-mean-square value computed as

$$q_{par} = \{\max_\tau |\hat{s}(\tau)|\} / \{\sqrt{(1/\tau_{max}) \sum_{\tau=1}^{\tau_{max}} \hat{s}(\tau)^2}\} \quad (10)$$

where, τ_{max} denoted as the maximum beat period (in the frames), $\hat{s}(\tau)$ denotes the time average of the beat period salience observation $s(t,\tau)$ used for tempo estimation.

$$\hat{s}(\tau) = (1/T)\left(\sum_{t=1}^{T} s(t,\tau)\right) \quad (11)$$

The second value, q_{pax}, maximum quality value measures the maximum of the beat period salience observation time average and calculated as

$$q_{pax} = \max(\tau) |\hat{s}(\tau)| \quad (12)$$

At last, the third quality measure q_{kur} calculates the minimum value of kurtosis of $s(t,\tau)$ along time as

$$q_{kur} = \min(t) \{k_s(t,\tau)\} \quad (13)$$

where, $k_s(t,\tau)$ is the sample kurtosis of $s(t,\tau)$ in the variable τ.

Defining $q = [q_{par}\, q_{max}\, q_{kur}]$ as the vector of quality measures. For beat period salience observations $s(t,\tau)$ that reflects the clear periodic structure, large values of these quality measures are expected.

4.5 Reliability Estimation

A quality that reflects the reliability of the set of beat estimates B^* obtained by the beat tracking algorithm is calculated based on the quality measure vector q. k-NN regression algorithm is used to determine this reliability measure denoted as r_p.

Let measure of performance of the beat estimate represented by p. Assume I= {1, ….., I} be a set of training audio signals, $\{q^i : i \in I\}$ the set of quality vectors and $\{p^i : i \in I\}$ the set of performance measures for each of the training samples. The distance to the quality measures of the training set is calculated, given a new audio signal with quality q, as

$$d^i = \| \mathbf{q} - \mathbf{q^i} \|_2 \quad (14)$$

Where, $\| . \|_2$ is the Euclidean norm. Then, K denotes the set of indexes of the K-NN. Finally, reliability is calculated as the performance of the K nearest neighbors under the performance criteria p of its beat estimates B^*, as

$$r_p = \frac{1}{K} \sum_{i \in K} p^i \quad (15)$$

Therefore, the expected performance accuracy in terms of evaluation criteria p can be interpreted as the beat tracking reliability measure r_p.

4.6 Database

The database for proposed beat tracking method consists of 222 musical audio files. This database divided into six categories Dance(40), Rock/Pop(68), Jazz(40), classical(30), and choral(22). A reasonable number of styles, tempos and time signatures are included in the database. Around 60 seconds in length the audio files are with time-variable tempo.

5. Results and Discussion

The onset detection function - a continuous signal which exhibits peaks at likely onset locations is chosen for the mid-level representation. This transformation of the audio signal is more suitable for identifying beat locations. The spectral difference between short term analysis frames is measured to calculate the onset detection function. The method used for this representation is the complex spectral difference method. An example of input audio musical signal and its onset detection function (middle level representation) are shown in the Fig.3 and Fig.4.

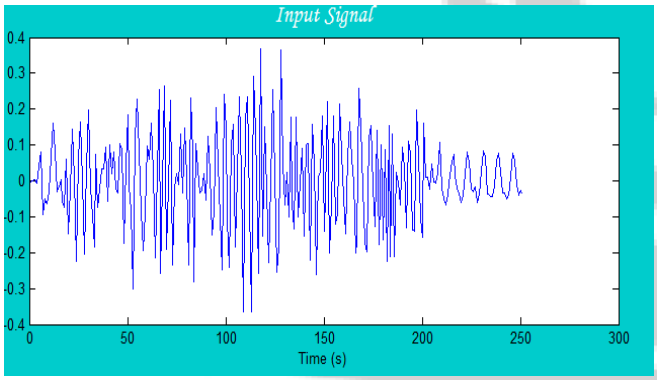

Figure 3: An example of the input audio signal.

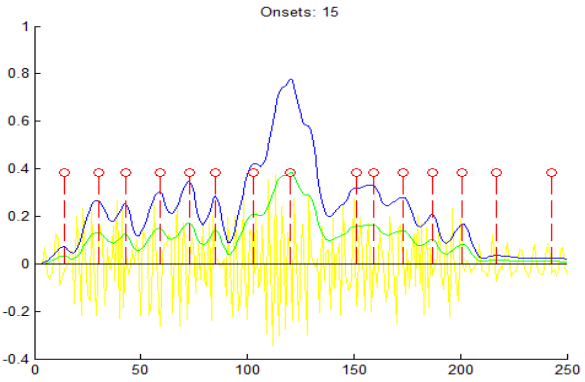

Figure 4: Onset detection function of input signal.

From the mid-level representation beat period is extracted by using comb filter resonators. An autocorrelation function discards phase-related information for that detection function is processed to emphasize the strongest and discard the least significant peaks.

To achieve an autocorrelation function, first the onset detection function is half wave rectified as shown in Fig.5(a). Furthermore, this function is then passed through comb filter-type structure to infer the meter of musical scores and symbolic performances observing periodicity integer multiples of the beat level with strong peak at the measure. This comb filter output is shown in the fig. 5(b).

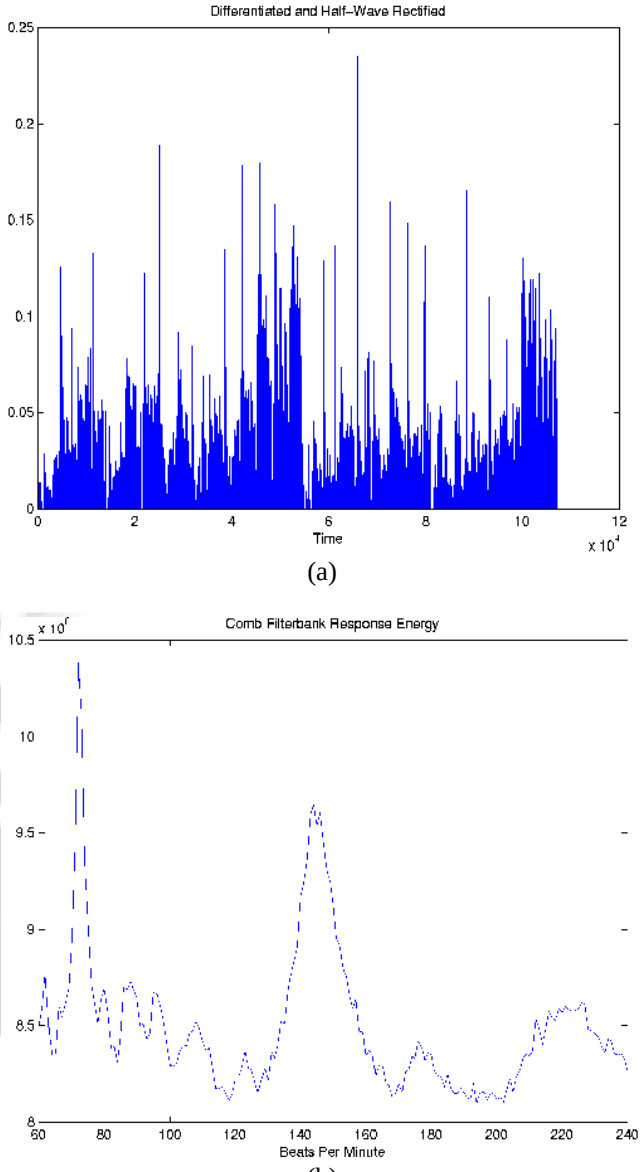

Figure 5: Autocorrelation function of the detection function. (a) Half wave rectified signal. (b) Comb Filter bank output.

The maximum beat observation of the input audio signal is shown in the Fig. 6. When looking at the beats of "flute" musical audio under the timing measures, it is found that overall beats are spread along the time axes and does not follow any periodicity. The onsets are found to be 15 in number for the same signal. Therefore if accurate beat tracking to be the result of agreement with the onset locations is considered then the onsets needed to be continuously consistent over one extended period.

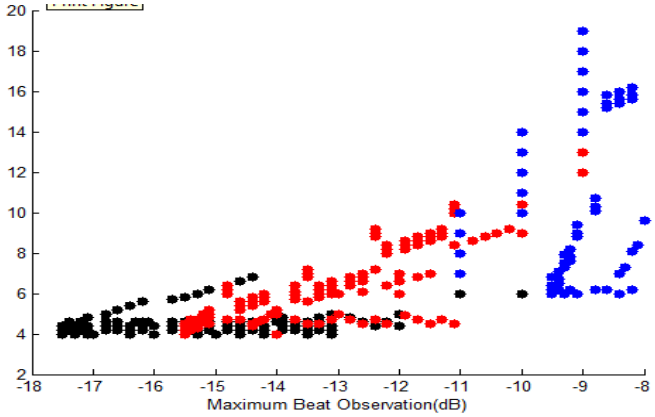

Figure 6: Beat observation of the input audio.

6. Conclusions

The proposed probabilistic framework to track beat is the smallest and suggesting a more robust behavior due the exploitation of both beat and non-beat information. Additionally it automatically measures the reliability of beat tracking system using k-NN regression algorithm while existing other algorithms exclusively estimates beat locations and do not account for the specific limitations of the algorithm.

References

[1] Norberto Degara, , Enrique ArgonesRúa,, Antonio Pena,, Soledad Torres-Guijarro, Matthew E. P. Davies, and Mark D. Plumbley, "Reliability-Informed Beat Tracking of Musical Signals," IEEE Trans. Audio, Speech, And Language Processing,, Vol. 20, No. 1, pp. 290-301, Jan. 2012.
[2] S. Ravuri and D. Ellis, "Cover song detection: From high scores to general classification," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Mar. 2010, pp. 65–68.
[3] D. Ellis, C. Cotton, and M. Mandel, "Cross-correlation of beat-synchronous representations for music similarity," in Proc. IEEE Int. Conf.Acoust., Speech, Signal Process. (ICASSP), Apr. 2008, pp. 57–60.
[4] M. Mauch and S. Dixon, "Simultaneous estimation of chords and musical context from audio," IEEE Trans. Audio, Speech, Lang. Process., vol. 18, no. 6, pp. 1280–1289, Aug. 2010.
[5] J. P. Bello-Correa, "Towards the automated analysis of simple polyphonic music: A knowledge-based approach," Ph.D. dissertation, Dept. of Electron. Eng. , Univ. of London, Queen Mary, U.K., Jan. 2003.
[6] S. Dixon, "Evaluation of audio beat tracking system BeatRoot," J. New Music Res., vol. 36, no. 1, pp. 39–51, 2007.
[7] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," J. New Music Res., vol. 30, no. 2, pp.159–171, 2001.
[8] D. P. W. Ellis, "Beat tracking by dynamic programming," J. New Music Res., vol. 36, pp. 51–60, 2007.
[9] S. W. Hainsworth, "Techniques for the automated analysis of musical audio," Ph.D. dissertation, Univ. of Cambridge, Cambridge, U.K., Sep.2004.
[10] M. E. P. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 3, pp. 1009–1020, Mar. 2007.
[11] P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter ofacoustic musical signals," IEEE Trans. Audio, Speech, Lang. Process.,vol. 14, no. 1, pp. 342–355, Jan. 2006.
[12] G. Peeters, "Beat-tracking using a probabilistic framework and linear discriminant analysis," in Proc. 12th Int. Conf. Digital Audio Effects (DAFx-09), 2009.
[13] S. Dixon, "Onset detection revisited," in 6th Int. Conf. Digital Audio Effects (DAFx-06), Montreal, QC, Canada, Sep. 18–20, 2006, pp.133–137.
[14] R. O. Duda, P. E. Hart, and D. G. Stork, "Nonparametric techniques," In Pattern Classification. New York: Wiley-Interscience, 2000.

Author Profile

Lata Dhulekar received the B.E. degree in Electronics and Tele-communication Engineering from Babasaheb Naik College of Engg., Pusad, Maharashtra, India. Now, she is pursuing M.E. degree in Signal Processing from Smt. Kashibai Navale College of Engg., Pune, Maharashtra, India.