Survey paper on music beat tracking

1 author:

Makarand Velankar
Cummins College of Engineering for Women

39 PUBLICATIONS   **122** CITATIONS

SEE PROFILE

International Journal of Research in
Computer and Communication Technology

ISSN (O) 2278-5841
ISSN (P) 2320-5156

Survey Paper on Music Beat Tracking

Vedshree Panchwadkar, Shravani Pande, Prof.Mr.Makarand Velankar

Cummins College of Engg, Pune , India

vedshreepd@gmail.com, shravni.pande@gmail.com, makarand_v@rediffmail.com

Abstract-Tempo in music is beats perceived by us in unit time. The tempo is measured as number of beats per minute (BPM) in a music clip. This paper includes two algorithms used to measure tempo of music file. First one is an online musical beat tracking algorithm based on Kalman filtering(KF) with an enhanced probability data association (EPDA) method is proposed. This beat tracking algorithm is built upon a linear dynamic model of beat progression, to which the Kalman filtering technique can be conveniently applied. The beat tracking performance can be seriously degraded by noisy measurements in the Kalman filtering process. Three methods are presented for noisy measurements selection. They are the local maximum (LM) method, the probabilistic data association (PDA) method and the enhanced PDA (EPDA) method. Also another algorithm called Tempo Detection Using a Hybrid Multiband Approach is used for calculating beats per minute. The model tracks the periodicities of different signal property changes that manifest within different frequency bands by using the most appropriate onset/transient detectors for each frequency band.

Index Terms—Beat tracking, Kalman filtering, probabilistic data association, music information retrieval.

I. INTRODUCTION

Rhythm is characterized by patterns of musical units that occur at different hierarchical metrical levels. The rhythmic units that occur at the primary metrical level are called beats and the rate of repetition of these beats provides the tempo of a piece of music, which is expressed in beats per minute (bpm).Therefore, beat tracking plays an important role in music transcription and musical information retrieval. Beats perceived by us are generally similar to a particular musical clip. Songs with different beat patterns present have different BPM present and it is difficult to calculate BPM automatically in such musical clips.

The beat tracking performance can be seriously degraded by two factors. First, the existence of rest notes which hide cues for beat tracking and missed-beat, which does not have an onset pulse on the expected beat's position but with a small shift results in beats without obvious onset pulses. In both cases, the lack of clear onsets make beat tracking difficult. Second, there exists variability in human performance. Even a performer attempts to keep the duration between two adjacent beats constant through the whole music piece, the actual duration tends to vary along time. These factors result in noisy measurements in the Kalman filtering process.

Three methods are presented for noisy measurements selection. They are the local maximum (LM) method, the probabilistic data association (PDA) method and the enhanced PDA (EPDA) method. The performance of the three noisy measurement selection techniques is compared. We see that the performance of EPDA outperforms that of LM and PDA significantly. In the second algorithm the audio is converted into a down sampled representation where the frames around onset times are emphasized by generating an Onset Detection Function (ODF), which tracks different signal property changes. The term Onset Detection Function (ODF) refers to a function whose peaks ideally coincide with onset times. In the context of a tempo detector, it does not necessary imply musical onset times being extracted. Next, the existing periodicities of the ODF are extracted, which results in the generation of a Periodicity Detection Function (PeDF). Finally, the PeDF is postprocessed in order to extract the periodicity that corresponds to the perceived tempo.

Our study of different methods for beat tracking can be applied to popular Hindi songs to automatically identify tempo of the song. We are compared different methods used and identified the advantages and limitations for different methods. We studied these methods and planning to verify the results for the Hindi songs. Automatic identification of tempo has varied application such as music retrieval, recommendation, DJ music, mood identification of music etc.

II. KALMAN FILTERING ALGORITHM

In Kalman Filter algorithm (Fig 1), the input is the digital music signal, from which the musical onset signal and its period are estimated. Given these estimates, the Kalman filter (KF) algorithm is used to track beat locations sequentially.
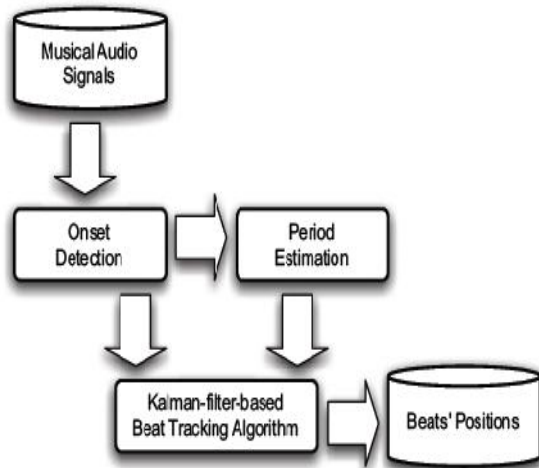

Fig.1 Kalman Filtering algorithm

A. Musical Data Pre-processing

It includes Onset Detection and Period Estimation. Musical onset signal gives the intensity change of musical contents along time. Changes can be of two types: new note arrival because of change of music pitches/harmonies and instantaneous noise-like pulses caused by percussion instruments. The cepstral distance method is used to calculate the musical onsets. The process is as follows:

First, the music contents is represented via mel-scale frequency cepstral coefficients (MFCC)[8], $c_m(n)$, for each shifting window of 20-msec with 50% overlap, where $m = 0, 1, ...,L$ 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.

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

The tempo and its inverse (*i.e.* period) are assumed to be perceptually fixed in our beat tracking system.

B. Beat tracking with kalman Iter

To apply the Kalman filter to the musical beat tracking, the first step is set up a linear dynamic system of equations 1,3 and 5.

$$x(k + 1) = (k + 1|k)x(k) + (k), \qquad (2)$$

$$y(k) = M(k)x(k) + (k), \qquad (3)$$

where k is a discrete time index, $x(k)$ is the state vector, $y(k)$ is the measurement, (k) is system noise, and (k) is measurement noise,$(k+1|k)$ is the state transition matrix, M(k) is the observation matrix.

$$x(k) = [(k), (k)]T, \qquad (4)$$

$$y(k) = (k), \qquad (5)$$

For state vector, (k) is the beat location and for measurement, (k) is the instantaneous   period, respectively. The instantaneous period, (k), is defined to be the time difference between the current and thenext beats as

$$(k) = (k + 1) - (k). \qquad (6)$$

Ideally, if there is no tempo change, period $(k + 1)$ should be the same as period (k);   namely,

$$(k + 1) = (k). \qquad (7)$$

Based on the above discussion, the state transition matrix $(k + 1|k)$ can be written as

$$(k + 1|k) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad (8)$$

and the observation matrix $M(k)$ is in form of

$$M(k) = [1\ 0], \qquad (9)$$

C. Method for noisy measurement selection

The beat tracking performance can be seriously degraded by noisy measurements in the Kalman filtering process. Following three methods are presented for noisy measurements selection.
1.Local Maximum(LM)
2.Probabilistic Data Association(PDA)
3. Enhanced PDA.

The performance of EPDA outperforms that of LM and PDA significantly. EPDA considers both information of prediction residual and music onsets intensities in a probabilistic way while the conventional method LM considers only the information of music onsets intensities. Therefore, EPDA can tackle the problem from the beats that have insignificant music onsets intensities.

Conventional method used in the Kalman method is Local Maximum. LM selects the time instance that has the maximum musical onset within a fixed window around the predicted beat location. LM fails when the beat does not have the strongest musical onset in the neighbourhood of predicted beat location. To overcome the weakness of the LM method, Probabilistic data association (PDA) is used in the Kalman filter to associate measurements with the target of interest in a confusing or disorderly state or collection.

In EPDA, we need to modify the definition of association probability because in music beat tracking, human uses not only the closeness between the measurement and the predicted beat location but also the intensity of musical onsets as cues to pick the next beat location. Hence this method is called Enhanced PDA.

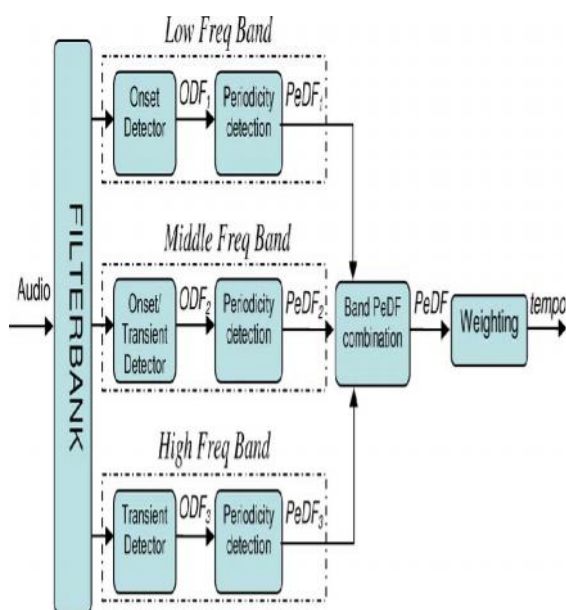## III. TEMPO DETECTION USING A HYBRID MULTIBAND APPROACH

Fig.2

The Fig 2 illustrates the different blocks that form the tempo detection system proposed here. First, a multi-band decomposition is utilized, which splits the incoming audio signal into three different frequency bands. Following this, the model attempts to use the most appropriate onset/transient detection method in each band. This is performed by exploiting the different acoustic properties of each frequency band with a different onset detector. Next, the existing band periodicities are extracted by building a *PeDF* in each band. Following this, the band *PeDF* s are combined into a single representation. Next, the combined *PeDF* is postprocessed by using a weighting function. Finally, the tempo is extracted from the weighted *PeDF*.

The algorithm is explained in following sections as Section A introduces the multiband decomposition used in the presented approach. A brief description of the onset/transient detectors is given in Section B, which includes a discussion of the suitability of the onset/transient detectors in each frequency band. Following this, the characteristics of the hybrid multiband configuration are given in Section C. Then, the periodicity detection method is described in Section D. Finally, a description of the suggested weighting method is given in Section E.

A. Multiband Decomposition

The presented multiband tempo detection system splits the audio signal into three different frequency bands. The choice of the band cut off frequencies is motivated by the different activity of certain instruments at different frequency regions. The different frequency ranges are given as follows.

• *Low-frequency band (LFB)*: frequency range: [0– 200 Hz]
Existing periodicities resulting from the presence of a bass line or percussive instruments such as a snare or a kick drum will be present in this low-frequency band.

• *Middle-frequency band (MFB)*: frequency range: [200–5000 Hz]
This band range overlaps with a large number of instrument frequency ranges. Thus, this band will contain a large amount of energy and active frequency components. The chosen band range roughly covers the fundamental frequencies of a wide range of instruments.

• *High-frequency band (HFB)*: frequency range:[above 5000 Hz]

Where corresponds to the sampling rate. The presence of percussive instruments in the recording results in transient signals spreading over the entire frequency range. Due to the low presence of non percussive instruments in this band, transients will be more localized in this band.

B. Onset/Transient Detection Function

There a large number of different onset detection functions have been used within tempo detection systems. In the presented tempo detection system, the combination of the spectral complex change onset detection method in [2], and a transient detection method presented in [3] is suggested. A brief description of the chosen onset/transient methods and its suitability to track periodicities in the above frequency bands is given as follows:

1. Spectral Complex change onset detection method (SC): This method prescribed by M. Davies [4] and S. Dixon [5] was identified as a very suitable representation for tempo extraction. The method emphasizes onsets in the *ODF* by tracking energy changes in the magnitude spectrum and unexpected deviations in the phase spectrum (e.g., a pitch change). The phase part of the complex number prediction facilitates the detection of slow onsets, such as a flute onset, and common onset energy changes occurring in the MFB. However, low-energy transients will be more difficult to track by using the *SC* in the HFB.

2. Transient detection method (TD): This method presented by Barry in [6], has not yet been utilized within a tempo detection model, which tracks the occurrence of broadband signals. This is performed by solely counting the number of bins that show an energy increase between consecutive frames larger than a threshold in dB. Due to the low number of bins that comprise the LFB, the *TD* will not be a suitable method for this band. The *TD* will track percussive occurrences in the MFB. Since the energy content of the signal does not play an important role in the *TD* method, it will also be effective in tracking transients in the HFB. Thus, even if the energies of the constituent bins of a transient signal are low, the method will effectively track a new occurrence if the transient spreads over the HFB range.

C. Hybrid Multiband Configuration

As can be derived from the description of the three frequency bands, different signal property changes manifest at different frequency bands. Consequently, the use of the most appropriate onset/transient detection method in each frequency band depending on the acoustic properties of each band should improve the performance of a tempo detection model. The advantages of both transient and complex detectors are combined together into a hybrid model.

The configuration of the suggested hybrid multiband configurations *Hyb1* and *Hyb2* is shown in Table I. In the LFB, onset energies can span over several consecutive frames. In this case, the *SC* is a more suitable method to track energy changes than the *TD* and will be used in both hybrid configurations.

In contrast, the use of *TD* in the HFB will ensure that existing broadband low energy transients will be accurately tracked. The method suitability in the MFB will change depending on the music type; singing solos or recordings with presence of slow onset instruments will benefit from the use of the *SC* (see *Hyb1* method in Table I). In contrast, the *TD* will be more appropriate to detect percussive transients within complex polyphonies (see *Hyb2* method in Table I).

As an example, the left column of Fig. 3 depicts the band *ODF*s generated using *Hyb1* method in a 10-s excerpt of Jive song "Big Time Operator" by "Big Band Batty Bernie." It can be seen that percussive transients are well localized using the *TD* in the HFB.

TABLE I: PROPOSED HYBRID MULTIBAND CONFIGURATIONS

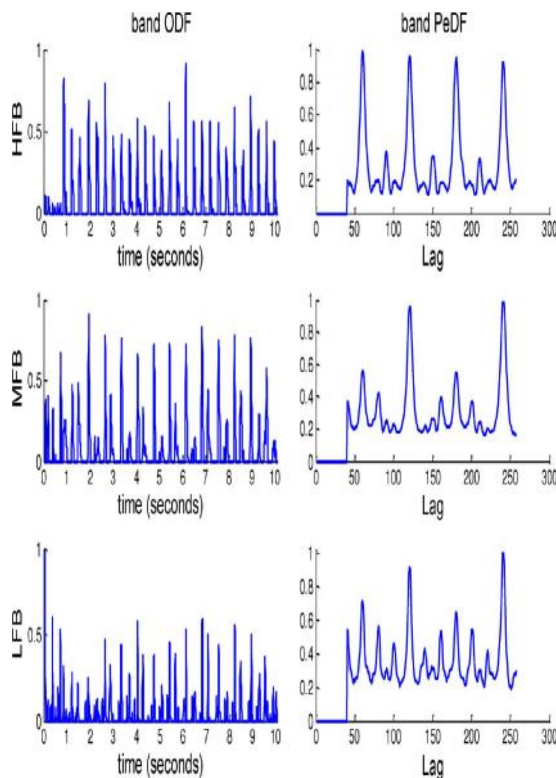| Configuration name	Low Freq Band	Middle Freq Band	High Freq Band
Hyb1	SC	SC	TD
Hyb2	SC	TD	TD

Fig.3

D. Periodicity Detection Method

As can be seen in Fig. 3, existing band periodicities are tracked by generating a *PeDF* in each band. This is performed by using the widely utilized autocorrelation function *r*

D = {minLag…maxLag}

within each band *ODF*. Existing periodicities in the lag range are tracked, where *minLag* and *maxLag* correspond to the beat period (in frames) of a tempo equal to 250 bpm and 40 bpm, respectively.

E. Weighting Method

Finally, as can be seen in Fig. 2, the combined *PeDF* is weighted in an effort to reduce the number of double and half tempo estimations. The general method weights the *PeDF* by a function that gives different weight to each beat periodicity candidate.

PeDF'(D) = PeDF(D') * W(D')

Existing approaches generate the function by using statistics derived from commonly used tempo annotations in popular music.

IV. CONCLUSION

In the tempo detection method using hybrid multiband , improved weighting method has been used   , which improves the results in all tempo detection methods. It was shown that adapting Davies *et al.* model to a multiband configuration improves the results. In addition, hybrid multiband configurations which combine the use of unique onset detectors for each frequency band were also introduced.

In the musical beat tracking algorithm based on Kalman filter, enhanced probabilistic data association (EPDA) is proposed. EPDA considers both information of prediction residual and music onsets' intensities in a probabilistic way while the conventional method LM considers only the information of music onsets' intensities.

V. FUTURE DIRECTIONS

A robust method capable of detecting the tempo in classical music is yet to be implemented, which suggests that further research in the area is still required.

The tempo detection model using hybrid multiband has difficulties to track slow and very fast tempi, which can be a result of the weighting function used. Thus the weighing function used in the proposed model requires further investigation.

The Kalman Filter algorithm is used for music clips with the constant tempo throughout. So the further improvement which can give better results for music clips with varying tempo can be thought of.

In the hybrid multiband approach, three frequency bands are used where cut-off frequencies are chosen to cover the frequency ranges of certain instrument types. Each band equally contributes to the overall periodicity estimation. So a more dynamic multiband decomposition should be considered. Thus, the reliability of the extracted periodicities in each individual band will be evaluated. This ensures that only bands in which onset detection functions provide valuable periodicities will be used.

REFERENCES

[1] M. 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.

[2] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, "Complex domain onset detection for musical signals," in *Proc. 6th Int. Conf. Digital Audio Effects (DAFx-03)*, London, U.K., 2003.

[3] Barry, D. Fitzgerald, E. Coyle, and B. Lawlor, "Drum source separation using percussive

feature detection and spectral modulation," in *Proc. Irish Signals Syst. Conf., ISSC*, Dublin, Ireland, 2005.

[4] M. Davies and M. D. Plumbley, "Comparing mid-level representations for audio based beat tracking," in *Proc. DMRN Summer Conf.*, Glasgow, U.K., 2005.

[5] F. Gouyon, S. Dixon, G. Widmer, and I. Porto, "Evaluating low-level features for beat classification and tracking," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, vol. 4, pp. 1309–1312.

[6] D. Barry, D. Fitzgerald, E. Coyle, and B. Lawlor, "Drum source separation using percussive feature detection and spectral modulation," in *Proc. Irish Signals Syst. Conf., ISSC*, Dublin, Ireland, 2005.

[7] D. P. W. Ellis, "Beat tracking by dynamic programming," *J. New Music Res., Special Iss. Beat and Tempo Extraction*, vol. 36, pp. 51–60, 2007.

[8] MFCC https://projects.developer.nokia.com/DSP/wiki/Mel_frequency_cepstral_coefficients

[9] Yu Shiu and C.-C. Jay Kuo Musical Beat Tracking via Kalman Filtering and Noisy Measurements Selection.

[10] Mikel Gainza and Eugene Coyle,Tempo Detection Using a Hybrid Multiband Approach.