Musical Beat Tracking via Kalman Filtering and Noisy Measurements Selection

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*— We study the problem of automatic musical beat tracking from acoustic data, *i.e.*, finding locations of beats of a music piece by computers on-the-fly, in this work. An online musical beat tracking algorithm based on Kalman filtering (KF) with an enhanced probability data association (EPDA) method is proposed. The 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. We see that the performance of EPDA outperforms that of LM and PDA significantly.**

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

I. Introduction

Beat tracking plays an important role in music transcription and musical information retrieval. This research is concerned with real-time musical beat tracking from acoustic data. Automatic musical beat tracking can be done either on-line or off-line. Although beat tracking has been extensively studied, only several methods apply to on-line audio processing, *e.g.*, [1], [2], [3].

Scheirer [2] used a comb filter to estimate the tempo and the beat location with an open-loop approach. New estimates do not take prediction residuals in the past into account. Beat tracking methods in [1] and [3] adopt the particle filtering technique. The particle filter is more general than the Kalman filter since it makes no assumption on the linearity of the tracking system and the Gaussianarity of underlying signals. However, its complexity is significantly higher and it does not address the problem of incorrect measurements as discussed in this work.

The beat tracking performance can be seriously degraded by two factors. First, the existence of rest notes and missed-beat syncopation results in beats without obvious onset pulses. Rest notes hide cues for beat tracking. Missed-beat syncopation has similar characteristics in that it does not have an onset pulse on the expected beat's position but with a small shift. 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 duratioin 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.

II. Beat Tracking with Kalman Filters

The system of the proposed beat tracking algorithm is shown in 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.

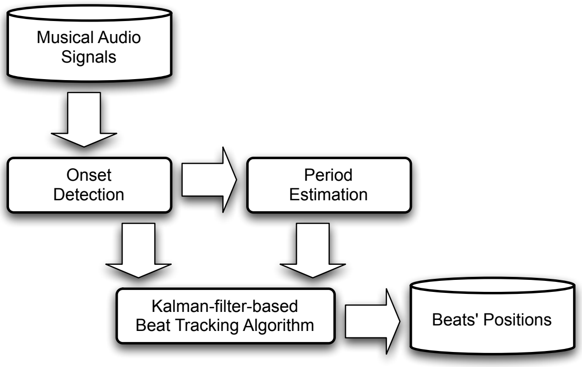

Fig. 1. Overview of the proposed musical beat tracking system.

A. 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 [4] 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 $m = 0, 1, ..., L$ 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.

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

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 the 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 the input parameter.

B. Beat Tracking with Kalman filter

To apply the Kalman filter to the musical beat tracking, the first step is set up a linear dynamic system of equations of the following form [1], [3], [5]:

$$\begin{aligned} x(k+1) &= \Phi(k+1|k)x(k) + \mu(k), & (2)\\ y(k) &= M(k)x(k) + \upsilon(k), & (3) \end{aligned}$$

where k is a discrete time index, $x(k)$ is the state vector, $y(k)$ is the measurement, $\mu(k)$ is system noise, and $\upsilon(k)$ is measurement noise. The state vector and the measurement as

$$\begin{aligned} x(k) &= [\tau(k), \Delta(k)]^T, & (4)\\ y(k) &= \tau(k), & (5) \end{aligned}$$

where $\tau(k)$ and $\Delta(k)$ are the beat location and the instantaneous period, respectively. The instantaneous period, $\Delta(k)$, is defined to be the time difference between the current and the next beats as

$$\Delta(k) = \tau(k+1) - \tau(k). \tag{6}$$

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

$$\Delta(k+1) = \Delta(k). \tag{7}$$

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

$$\Phi(k+1|k) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{8}$$

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

$$M(k) = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{9}$$

III. Methods for Noisy Measurements Selection

A. LM (Local Maximum) Method

Measurement selection in the conventional Kalman filter is called the Local Maximum (LM) method as shown in Fig. 2. Simply speaking, LM selects the time instance that has the maximum musical onset within a fixed window around the predicted beat location $\hat{\tau}(k+1|k)$ as measurement $y(k+1)$. Mathematically, this can be written as

$$y(k+1) = \tau(k+1) = \underset{|m-\hat{\tau}(k+1|k)|<w/2}{\arg\max} \; d(m), \tag{10}$$

where $d(m)$ is the onset signal, w is the window width, and

$$\hat{\tau}(k+1|k) = \hat{\tau}(k|k) + \hat{\Delta}(k|k). \tag{11}$$

is the predicted beat location.

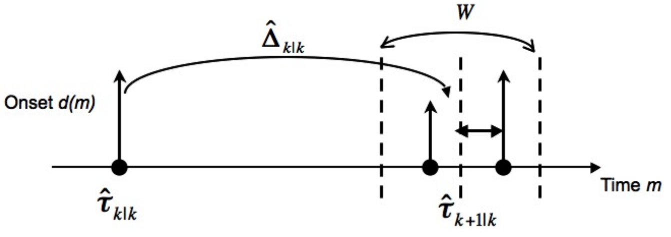

Fig. 2. Measurement selection in the conventional Kalman filter with the local maximum (LM) method.

LM fails when the beat does not have the strongest musical onset in the neighborhood of predicted beat location $\hat{\tau}(k+1|k)$. The time instant having strongest musical onset will be selected by LM as the beat's location. In particular, when there is not musical onset pulse for the beat from rest notes and there are pulses due to non-beat notes or percussion sounds nearby, LM often fails and causes the Kalman filter to lose proper tracking performance.

B. Probabilistic Data Association (PDA) Method

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 clutter environment [6]. PDA helps the Kalman filter to maintain the track by avoiding choosing a single measurement from several candidates as the LM method does. Instead, it considers all candidate measurements and their associations with the current track under a Bayesian framework.

The measurement validation aims to remove measurements that are very unlikely to be a correct measurement for PDA. A validation region is engulfed by a multi-dimentional probabilistic threshold in the measurement's space. PDA will consider only measurements within the validation region so that the computation load in PDA can be significantly reduced by removing non-validated measurements.

The validation region is derived from the covariance matrix of predicted measurement $\hat{y}(k+1|k)$:

$$\begin{aligned} S(k+1) &\triangleq E[\tilde{y}(k+1|k)\tilde{y}^{'}(k+1|k)|Y^k] \\ &= M(k+1)P(k+1|k)M'(k+1) + \upsilon(k), \end{aligned} \tag{12}$$

where

$$\begin{aligned} P(k+1|k) &= E[\tilde{x}(k+1|k)\tilde{x}^{'}(k+1|k)|Y^k] \\ &= \Phi(k+1|k)P(k|k)\Phi'(k+1|k) + Q(k), \end{aligned} \tag{13}$$

and Y^k consists of every set of validated measurements $Y(j)$ from time index $j = 1$ to k. It can be defined in the measurement space via [6]

$$\tilde{V}_\gamma(k+1) \triangleq \{y : [y - \hat{y}(k+1|k)]' S^{-1}(k+1)[y - \hat{y}(k+1|k)] \leq \gamma\}, \tag{14}$$

where $S(k+1)$ is the covariance matrix of the predicted measurement $\hat{y}(k+1|k)$. It is shown in [6] that the weighted norm of prediction residual is Chi-square distributed with the degree of freedom equal to the dimension of the measurement. For musical beat tracking, the measurement dimension is 1 because musical onset is used as the measurement. Then, the probability for the region to include true measurements is 99.7% when $\gamma = 9$ is chosen in (14) while it is 95.4% when $\gamma = 4$ is chosen [6].

Eqs. (9), (12) and (14) can be used to derive the validation region for the proposed musical beat tracking algorithm based on Kalman filtering. It is equal to

$$\tilde{V}_{\gamma}(k+1) \triangleq \{y : \frac{[y - \hat{y}(k+1|k)]^2}{p_{11} + \upsilon} \leq \gamma\}, \tag{15}$$

where p_{11} is the variance of beat location $\tau(k)$.

The PDA method uses a weighted average of estimates from candidate measurements within the validation region to replace $\hat{x}(k+1|k+1)$. The weight is determined by the probability for measurement $y_i(k+1)$ to be a correct measurement. Mathematically, PDA decomposes the estimate into a linear combination of estimates of all measurements within the validation region as

$$\begin{aligned} &\hat{x}(k+1|k+1) = E[x(k+1)|Y^{k+1}] \\ &= \sum_{i=0}^{m_{k+1}} E[x(k+1)|\theta_i(k+1), Y^{k+1}] Pr\{\theta_i(k+1)|Y^{k+1}\} \\ &= \sum_{i=0}^{m_{k+1}} \hat{x}_i(k+1|k+1)\beta_i(k+1), \end{aligned} \tag{16}$$

where

$$\beta_i(k+1) \triangleq Pr\{\theta_i(k+1)|Y^{k+1}\}, \quad i = 0, 1, ..., m_{k+1}. \tag{17}$$

The weight $\beta_i(k+1)$, also known as the "association probability", is related to the distance between candidate measurement $y_i(k+1)$ and prediction $x(k+1|k)$ as defined in Eq. (17). That is, the smaller the distance, the larger $\beta_i(k+1)$ is. Intuitively, if measurement $y_i(k+1)$ is closer to prediction $x(k+1|k)$, its contribution will weigh more.

C. Enhanced PDA (EPDA) Method

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. Thus, we need to modify the definition of association probability $\beta_i(k+1)$. The resulting method is called the enhanced PDA (EPDA) method.

In [7], the intensity of the observed signal is introduced to the association probability calculation via

$$\begin{aligned} &\beta_i(k+1) \triangleq Pr\{\theta_i(k+1)|I_Y(k+1), Y^{k+1}\} \\ &\propto Pr\{I_Y(k+1)|\theta_i(k+1), Y^{k+1}\} Pr\{\theta_i(k+1)|Y^{k+1}\}. \end{aligned} \tag{18}$$

where $I_Y(.)$ is the distribution function of measurement intensity, (*i.e.*, musical onset intensity). Note that the term $Pr\{\theta_i(k+1)|Y^{k+1}\}$ in the right-hand-side of Eq. (18) is $\beta_i(k+1)$ in Eq.(17), which considers only the prediction residual. As shown in Eq. (18), the modified $\beta_i(k+1)$ is the product of two terms. One is contributed by musical onset's intensity and the other by the prediction residual.

The term contributed by musical onset's intensity in Eq. (18) can be further decomposed [7] as

$$P\{I_Y(k+1)|\theta_i(k+1), Y^{k+1}\} = \frac{I_i(y_i)}{I_0(y_i)} \prod_{j=1}^{m_{k+1}} I_0(y_j), \tag{19}$$

for $i = 1, 2, ..., m_{k+1}$, where $I_i(y_i)$ is the probability distribution of correct measurement y_i and $I_0(y_i)$ is the probability distribution of y_i as an incorrect measurement. We can re-write (19) as

$$P\{I_Y(k+1)|\theta_i(k+1), Y^{k+1}\} = I_i(y_i) \prod_{j=1, j \neq i}^{m_{k+1}} I_0(y_j). \tag{20}$$

It is difficult to compute $I_i(y_i)$ with $1 \leq i \leq m_{k+1}$ in (20) efficiently and accurately for two reasons. First, the number of candidate measurements, m_{k+1}, is determined dynamically by validated region $\tilde{V}(k+1)(\gamma)$ at each time step k. Second, for particular i and $k+1$, there are not enough samples in estimating the probability distribution $I_i(k+1)$ accurately. To address this issue, the probability distribution of $I_i(k+1)$ is replaced by a general probability distribution of the intensity of measurement $y(.)$ for all i and k when $y(.)$ is a "correct" measurement. That is, we have

$$I_i(k+1) \simeq I_B, \quad i = 1, 2, ..., m_{k+1}.$$

where I_B is the probability distribution of musical onset intensities when their corresponding measurements are truly beats. Thus, I_B can be found by collecting intensities of musical onsets that correspond to beats.

On the other hand, $I_0(y_i)$ is the probability distribution of the intensity of candidate measurements y_i which do not correspond to a beat. Again, it is difficult to get the accurate distribution all specific i and $k+1$. Instead, we replace it by a general probability distribution

$$I_0(k+1) \simeq I_N \tag{21}$$

where I_N is probability distribution of musical onset intensities when their corresponding measurements are not beats.

The intensity of musical onsets are one-dimensional signals with non-negative values. To estimate the its probability distribution I_B and I_N, we consider non-parametric approach. Histograms of music onsets' intensities for I_B and I_N are shown in Fig. 3, where bin centers are uniformly located from 0.125 to 8.125 seconds with bin width 0.25 second.

We see from Fig. 3 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.

IV. Experimental Results

A. Experimental Data and Setup

Two data sets were used in our experiments. The first data set was MIREX 2006 beat tracking competition practice data

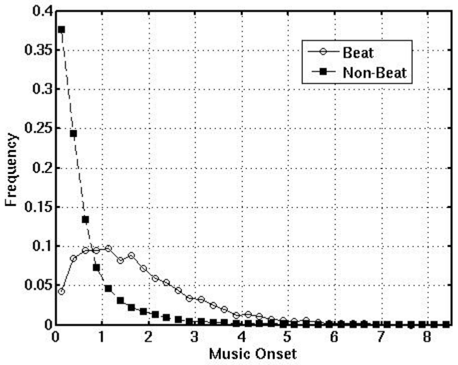

Fig. 3. The histogram of music onsets' intensities for I_B and I_N with the bin width equal to 0.25, where the x-axis is the music onset intensity while the y-axis is the frequency of occurence.

[8]. It consisted of twenty 30-sec music clips of diverse genres. A single period that is agreed by most of 40 listeners was used for tracking in our experiment. The second data set was 20 Billboard Top-10 songs in 80's. The genre includes pop and rock. For each song, a 30-sec music clip was segmented from the original song. The first 5 seconds were used to initialize the state vector while the remaining 25 seconds were used to evaluate the beat tracking performance. The sampling rate Musical onsets is 100 Hz.

B. Performance Metric

A metric that is similar to the P-score in MIREX 2006 [8] was used to evaluate the musical beat tracking performance. It evaluates the correct rate of detected beats while considering the impact of false-alarms and is defined as

$$P = \frac{1}{N_{max}} \sum_{n=-\infty}^{\infty} \sum_{m=-w}^{w} \tau_d(n)\tau_g(n-m), \qquad (22)$$

where τ_d is the unit pulse in the detected beat location, τ_g is the unit pulse in the ground-truth beat location, $2w$ is the tolerable window size and

$$N_{max} = max(N_d, N_g), \qquad (23)$$

and where N_d and N_g are the numbers of the detected and the ground-truth beats, respectively. τ_d and τ_g are indication function taking only values 0 and 1. The window size $2w$ was chosen to be 20% of the beat duration throughout the experiment. Note that the value of P lies between 0 and 1.

C. Performance of P-Score for MIREX and Billboard Data Set

For the MIREX data set, the average performance of all music clips are shown the first row of Table I. We compare the performance of the Kalman-filter-based (KF-based) beat tracking algorithm with LM, PDA and EPDA measurment selection methods. EPDA improves the performance greatly over LM by an average of 13.25%. In contrast, PDA has a performance similar to LM.

TABLE I

P-SCORE COMPARISON OF THE KF-BASED BEAT TRACKING ALGORITHM WITH LM, PDA AND EPDA.

	LM	PDA	EPDA
MIREX	74.08%	72.67%	87.33%
Billboard	77.80%	87.81%	94.68%

The P-Score performance for the Billboard Top-10 data set is is shown in the second row of Table I. For the Billboard Top-10 data set, the P-Scores of LM, PDA and EPDA all improve but with a different degree. The improvement of PDA is most significant while EPDA still offers the best performance. Actually, 19 out of 20 songs with EPDA can achieve a P score higher than 94%. In contrast, LM achieves similar performance on MIREX and Billboard Top-10 dataset. There is only 3.72% difference between two data sets using LM. As described earlier, MIREX has rather diverse genres while songs in the Billboard Top-10 data set are more homogeneous. Generally speaking, the Billboard data set has more regular beats throughout each music clip than the MIREX data set.

V. CONCLUSION

In this paper, a musical beat tracking algorithm based on Kalman filter and 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. Therefore, EPDA can tackle the problem from the beats that have insignificant music onsets' intensities. The experimental results show that EPDA improves the performance of P-score over LM for both MIREX 2006 dataset and Billboard Top-10 dataset.

REFERENCES

[1] 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.

[2] E. Scheirer, "Tempo and beat analysis of acoustic musical signals", *Journal of Acoustic Society America*, vol. 103, pp. 588-601, 1998.

[3] 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 Proccessing*, vol. 13, No. 2, pp. 275-285, 2005.

[4] L. Rabiner, "A tutorial on hidden Markov models and selected applications inspeech recognition", *Proceedings of the IEEE*, vol. 77, No. 2, pp. 257-286, 1989.

[5] A. T. Cemgil, B. Kappen, P. Desain and H. Honing, "On tempo tracking: tempogram representation and Kalman filtering", *Journal of New Music Research*, 2001.

[6] Y. Bar-Shalom and T. E. Fortmann, "Tracking and Data Association", *Academic Press*, Orlando, Florida, 2001.

[7] C.-M. Huang, D. Liu and L.-C. Fu, "Visual tracking in cluttered environments using the visual probabilistic data association filter", *IEEE Transaction on Robotics*, vol. 22, No. 6, 2006.

[8] "Music information retrieval evaluation exchange (MIREX) Competition", *http://www.music-ir.org/mirexwiki/index.php*, 2006.

[9] M. Goto and Y. Muraoka, "Issues in evaluating beat tracking systems", *Proc. IJCAI-97 Workshop on Issues in AI and Music*, pp.9-16, 1997.

[10] A. P. Klapuri, A. J. Eronen and J. T. Astola, "Analysis of the meter of acoustic musical signals", *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, No. 1, pp.342-355, 2006.