Reliability-Informed Beat Tracking of Musical Signals

Norberto Degara, *Student Member, IEEE*, Enrique Argones Rúa, *Member, IEEE*, Antonio Pena, *Member, IEEE*, Soledad Torres-Guijarro, Matthew E. P. Davies, *Member, IEEE*, and Mark D. Plumbley, *Member, IEEE*

***Abstract*—A new probabilistic framework for beat tracking of musical audio is presented. The method estimates the time between consecutive beat events and exploits both beat and non-beat information by explicitly modeling non-beat states. In addition to the beat times, a measure of the expected accuracy of the estimated beats is provided. The quality of the observations used for beat tracking is measured and the reliability of the beats is automatically calculated. A k-nearest neighbor regression algorithm is proposed to predict the accuracy of the beat estimates. The performance of the beat tracking system is statistically evaluated using a database of 222 musical signals of various genres. We show that modeling non-beat states leads to a significant increase in performance. In addition, a large experiment where the parameters of the model are automatically learned has been completed. Results show that simple approximations for the parameters of the model can be used. Furthermore, the performance of the system is compared with existing algorithms. Finally, a new perspective for beat tracking evaluation is presented. We show how reliability information can be successfully used to increase the mean performance of the proposed algorithm and discuss how far automatic beat tracking is from human tapping.**

***Index Terms*—Beat-tracking, beat quality, beat-tracking reliability, k-nearest neighbor (k-NN) regression, music signal processing.**

I. Introduction

THE task of beat tracking consists in automatically detecting the moments of musical emphasis in an audio signal. This task is the equivalent to the human act of tapping music with a foot so it is not surprising that the beat rate is often described as the foot-tapping rate. In the following, we use the term *beat* to describe the individual temporal events that define this metrical level and *beat period* to denote the regular time between events. As in [1], the term *beat phase* is used to indicate the location of a beat with respect to the previous beat. The beat is the most salient of the underlying periodicities of a musical signal. It is the basic time unit of music and it determines the temporal structure of an audio signal, making beat tracking a very important task in *music information retrieval* (MIR) research [2]. Thus, beat estimation enables the beat synchronous analysis of musical audio [3] and it is of interest in multiple applications including, structural segmentation of audio [4], interactive musical accompaniment [5], cover-song detection [6], music similarity [7], chord estimation [8], and music transcription [9].

The automatic extraction of beats from musical signals is a challenging process due to both musical and physical reasons. *Musical properties* such as the rhythmic complexity of a performance have a large impact on beat tracking accuracy as discussed in [10]. In [11], critical passages that are prone to beat tracking errors are identified and the erroneous beats are classified. Thus, beats that do not correspond to any note event, boundary beats, ornamental beats, weak bass beats or constant harmony beats make beat tracking difficult. In addition, there are *physical properties* that impact beat tracking accuracy such as the poor condition of a recording or the presence of high reverberation. To face the difficulties of estimating beat times in audio signals multiple strategies have been proposed.

A. Related Work

A brief description of some of the existing approaches to beat tracking is presented in this section. For more details, good reviews of tempo induction and beat tracking algorithms can be found in [1] and [12].

A multi-agent approach has been proposed by Dixon in [13]. This approach extracts a sequence of onset events and derives a number of beat period candidates from an analysis of the inter-onset-interval distribution of the sequence of onsets. As in Goto *et al.* [14], a number of competing agents evaluate multiple beat hypotheses to determine the best sequence of beat times. Laroche [15] uses a least-square estimation of the local tempo followed by a dynamic programming stage used to obtain the beat locations. Similarly, Ellis [16] first identifies the beat period and then finds the beat phases by using a dynamic programming algorithm, and Stark *et al.* [3] implement a real-time beat tracking based on this approach.

Other approaches formulate the beat tracking problem using a probabilistic framework. Based on the symbolic data model of Cemgil *et al.* [17], Hainsworth [18] explores the use of particle filtering where the beat locations are modeled as a periodic sequence driven by a time-varying tempo process. Davies

Manuscript received December 23, 2010; revised February 28, 2011; accepted June 20, 2011. Date of publication June 27, 2011; date of current version November 11, 2011. This work was supported by the Galician Regional Government (2007/000023-0, 2009/062), the Spanish Government (TEC2009-14414-C03-03), the European Regional Development Fund, and EPSRC Grants EP/G007144/1 and EP/E045235/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bryan Pardo.

N. Degara, E. A. Rúa, A. Pena, and S. Torres-Guijarro are with the Signal Theory and Communications Department, University of Vigo, Vigo 36310, Spain (e-mail: ndegara@gts.uvigo.es; eargones@gts.uvigo.es; apena@gts.uvigo.es; marisol@gts.uvigo.es).

M. E. P. Davies and M. D. Plumbley are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: matthew.davies@eecs.qmul.ac.uk; mark.plumbley@eecs.qmul.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2011.2160854

et al. [19] propose a two-state model for beat tracking. A general state tracks the beat period and a context-dependent state is used to enforce continuity within a tempo hypothesis. A hidden Markov model (HMM) is proposed by Klapuri *et al.* [20] to simultaneously estimate the tatum, tactus, and measure metrical levels. Beat phases are independently estimated using an additional HMM whose hidden state models beat time instants.

More recently, Peeters [21] introduced a probabilistic framework formulated as an inverse Viterbi problem. Instead of decoding the sequence of beats along time, the system proposed by Peeters decodes beat times over beat-numbers. Following the idea of Laroche [15], a beat template is used to model tempo-related expectations on an onset signal. Thus, instead of estimating the beat observation likelihood using a single onset observation, the system calculates the observation likelihood through a cross-correlation of the onset signal and the estimated beat template. This template needs to be learned from a dataset and results depend on the musical genre.

B. Motivation

Despite the number of beat tracking strategies, there are still some issues that need to be addressed. Previous probabilistic approaches model the likelihood of a beat at a particular time either using a single observation, as for example in [20] and [16], or using a correlation template, as in [15] and [21]. However, the observations at non-beat time instants provide extra information that can potentially be exploited for beat tracking.

In addition, earlier work has concentrated on comparing the mean performance of different beat tracking methods such as in [1], [20] and [19]. The risk of focusing the analysis of the performance on average values overlooks the reasons beat trackers fail to correctly estimate beats. As discussed by Grosche *et al.* [11], beat tracking accuracy is determined by the musical and physical properties of a performance. However, the specific limitations of a particular beat tracking algorithm also have to be taken into account. Understanding these limitations is essential to improving the performance of beat tracking methods. Doing so could lead to the eventual automatic prediction of the behavior of beat tracking algorithms and the ability to combine them according to their expected performance.

C. Proposed Model

The aim of this paper is to present a reliability-informed beat tracking method for musical signals. To integrate musical-knowledge and signal observations, a probabilistic framework that models the time between consecutive beat events and exploits both beat and non-beat signal observations is proposed. This differs from [20] that models beat time instants and only uses beat information. Simple approximations for the parameters of this probabilistic model are also provided using musical knowledge. Contrary to the current trend in beat tracking which exclusively estimate beat locations, the specific limitations of the proposed probabilistic model are identified and a measure of the expected accuracy of the estimated beats is also provided. The idea of automatically measuring the expected performance of a beat tracking algorithm is general and can potentially be extended to any other system by identifying its own limitations.

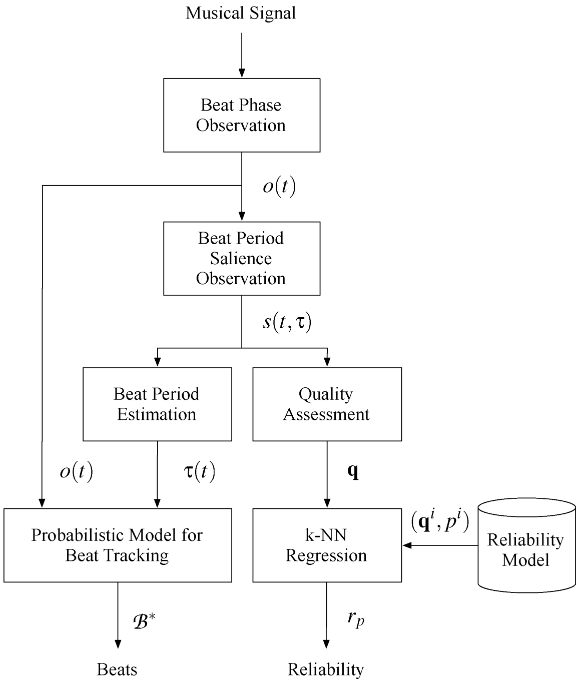


Fig. 1. Block diagram of the reliability-informed beat tracking system. It returns the set of estimated beat times $\mathcal{B}^*$ and a measure of the reliability of the estimates r_p.

The system analyzes the input musical signal and extracts a beat phase and a beat period salience observation signal from which the beat period is calculated. Then, the beat tracking probabilistic model takes as input parameters the phase observation signal and the beat period estimation, returning the set of beat time estimates. Finally, the quality of the beat period salience observation signal is assessed and a k-nearest neighbor algorithm is used to measure the reliability of the beat estimates. Fig. 1 shows an overview of the proposed beat tracking system.

The remainder of this paper is structured as follows. Section II describes the different elements of the reliability-informed beat tracking system shown in Fig. 1. Then, Section III describes the database and the evaluation measures used to compare the proposed algorithm with state-of-the-art beat tracking methods. Section IV presents the experimental results where we evaluate the importance of the different elements of the beat tracking model, discuss the use of a learning algorithm for the automatic estimation of the parameters of the model, compare the proposed method with existing systems and discuss the benefits of using reliability information. Finally, the main conclusions and future work are summarized in Section V.

II. Beat Tracking System

This section describes the different elements of the reliability-informed beat tracking method illustrated in Fig. 1. The proposed beat tracker is publicly available under the GNU Public License.[1] Section II-A presents the feature extraction process. Then, Section II-B introduces the method used for beat period estimation. The proposed probabilistic beat tracking model is described in Section II-C. Finally, the quality analysis is presented in Section II-D and the calculation of the reliability measure in Section II-E.

[1] http://www.gts.uvigo.es/~ndegara/Publications.html

A. Feature Extraction

In beat tracking, an *onset detection function* is commonly used as a midlevel representation that reveals the location of transients in the original audio signal. This detection function is designed to show local maxima at likely event locations [22]. Many methods exist to emphasize musical events and performance often depends on the features used for beat tracking [23]. The *complex spectral difference* method [24] shows good behavior for a wide range of audio signals and has been successfully used in other beat tracking systems [19]. It works in the complex domain, emphasizing onsets due to a change in the spectral energy and/or a deviation in the expected phase. Although the proposed probabilistic framework can accept any onset signal, the complex spectral difference has been selected as the reference method used to discuss results.

In the following, the complex domain onset signal at time t is denoted as $o(t)$. As in [19], the time-resolution for $o(t)$ is 11.6 ms. As shown in the block diagram of Fig. 1, the onset signal $o(t)$ constitutes the phase observation used to determine the beat locations $\mathcal{B}^*$ and extract the beat period salience signal $s(t, \tau)$.

The periodicity of the phase observation signal $o(t)$ is analyzed to determine the beat period salience of the musical signal. For that, the shift-invariant comb filterbank approach described in [19] is adopted. The method can be summarized as follows. First, the signal $o(t)$ is segmented into frames of 6 s in length and an overlap of 75%, equivalent to a resolution of 1.5 s. The length of the analysis window is long enough to correctly estimate the beat period and the resolution short enough to track changes. Then, the signal is normalized using an adaptive mean threshold and half-wave rectified. The autocorrelation of the resulting signal is calculated to discard phase-related information and emphasize potential periodicities. Finally, the autocorrelation is processed by a shift-invariant comb filterbank weighted by a beat period preference curve. The beat period salience information is assumed to stay constant for the 1.5 s that define its original time resolution, then the same time index t can be effectively used for $o(t)$ and $s(t, \tau)$. For a more detailed description of $s(t, \tau)$ see the derivation of the beat period salience signal in [19].

Fig. 2 presents examples of the observation signals $o(t)$ and $s(t, \tau)$. Fig. 2(a) shows the phase observation signal (i.e., the onset detection function) $o(t)$ and the annotated beat time instants. In general, the phase observation signal $o(t)$ will present large values at beat locations and small values at non-beat time instants. Fig. 2(b) shows the beat period salience signal $s(t, \tau)$ for $t = 0$ and the annotated beat period of the input music signal. The signal $s(t, \tau)$ is a measure of the salience of each beat period candidate τ at time t. The beat period τ can take any value in $\{1, \ldots, 128\}$, in time frame units. Thus, the maximum beat period allowed is 1.5 s given the fixed time-resolution of 11.6 ms. This feature is used to track the tempo and to assess the quality of the beat period estimate as shown in Fig. 1.

B. Beat Period Tracking

The proposed beat tracking system estimates the beat period and phases independently. Like the beat phase observation $o(t)$, the beat period estimate $\tau(t)$ is an additional parameter to the beat tracking model shown in Fig. 1. To extract the sequence of periods $\tau(t)$ from the beat period salience observation signal $s(t, \tau)$, an offline version of the hidden Markov model method presented in [3] is used. The system assumes the beat period to be a slowly varying process and the transition probabilities are modeled using a Gaussian distribution of fixed standard deviation. For a complete description of the beat period tracking method see [3].

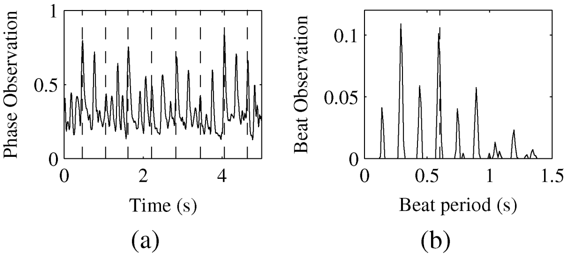


Fig. 2. Example of the extracted observation signals. (a) Phase observation signal $o(t)$ (continuous line) and beat annotations (dotted line). (b) Beat period salience signal $s(t, \tau)$ for $t = 0$ (continuous line) and annotated beat period (dotted line).

C. Probabilistic Model for Beat Tracking

Music is highly structured in terms of the temporal ordering of musical events defining a context that can be used to determine beat events. In particular, beats are regularly spaced in time with small deviations from the beat period. To integrate this contextual knowledge with signal observations and then estimate beat phases, a hidden Markov model (HMM) is used [25]. This probabilistic framework has been shown to be useful for modeling temporal dependencies. Examples of using a HMM to model the temporal nature of music can be found in [20], [26] and [21].

The proposed beat tracking system defines a first-order HMM where a hidden variable ϕ represents the phase state and measures the elapsed time, in frames, since the last beat event. The total number of states N_{τ_t} is determined by the estimated beat period $\tau(t)$, denoted in the following as τ_t. The possible states for ϕ are $\{0, 1, \ldots, N_{\tau_t} - 1\}$ (see Section II-C3 for details). Thus, state $\phi = n$ indicates that there have been n frames since the last beat event and state $\phi = 0$ denotes the beat state. A state at time frame t is denoted as ϕ_t and a particular state sequence $(\phi_1, \phi_2, \ldots, \phi_T)$ as $\phi_{1:T}$.

The temporal structure of the beat sequence is encoded in the state transition probabilities $a_{ij} = \mathrm{P}(\phi_t = j \mid \phi_{t-1} = i)$. Then, as the phase state variable ϕ_{t-1} measures the elapsed time since the last visit to the beat state 0 at time $t - 1$, the allowed transitions are from $\phi_{t-1} = n$ to $\phi_t = n + 1$ or to the beat state $\phi_t = 0$. The observable variable for the phase states, o_t, is the phase observation signal $o(t)$ and $o_t = o(t)$ in the following. The phase observation o_t is assumed to be independent of any other state given the current state, and then the state-conditional observation probability is $P(o_t \mid \phi_t)$.

The first-order HMM model introduced above is summarized in Fig. 3(a). The hidden variable ϕ_t is shown with circles and the observation o_t variable with boxes. Links represent the conditional dependencies between the state and observation variables. Additionally, transitions between hidden states are shown in Fig. 3(b) where the states are represented by circles and transitions by links. There are only two possible transitions from

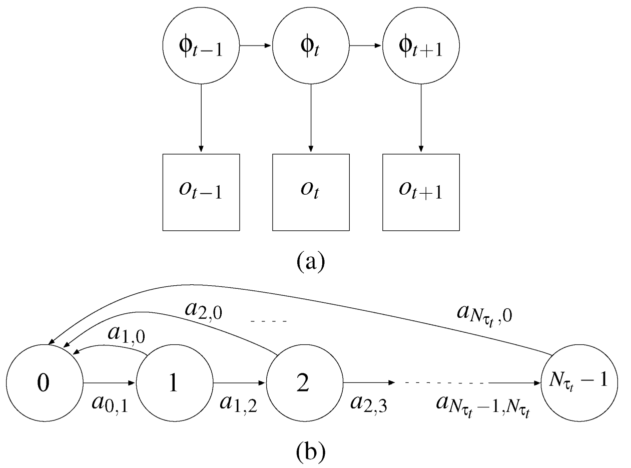


Fig. 3. Hidden Markov model for beat tracking. (a) Hidden state and observation variables conditional dependencie. (b) State transition diagram for the hidden state ϕ_t.

a particular state which considerably reduces the search space. Unlike other works where only the beat "strength" is considered [20], [16], [21], we specifically model non-beat states and account for non-beat observations.

1) Estimation Goal: The goal of the proposed probabilistic model is to estimate the sequence of beats which best explains the phase observations, o_t. To do so, the most likely sequence of hidden states $\phi^*_{1:T}$ that led to the set of observations $o_{1:T}$ is estimated as

$$\phi^*_{1:T} = \underset{\phi_{1:T}}{\operatorname{argmax}} \mathrm{P}(\phi_{1:T} \,|\, o_{1:T}) \tag{1}$$

where T denotes the number of frames of the input audio signal. This optimization problem can be easily solved using the well-known Viterbi algorithm [27]. Once the optimal sequence of hidden states $\phi^*_{1:T}$ has been decoded, we are ready to obtain the set of beat times $\mathcal{B}^*$. We do this by selecting the time instants where the sequence $\phi^*_{1:T}$ visited the beat state. Thus,

$$\mathcal{B}^* = \{t : \phi^*_t = 0\}. \tag{2}$$

Considering the model assumptions presented in Fig. 3(a), the posterior probability of (1) can be written as

$$\mathrm{P}(\phi_{1:T} \,|\, o_{1:T}) \propto \mathrm{P}(\phi_1) \prod_{t=2}^{T} \mathrm{P}(o_t \,|\, \phi_t)\mathrm{P}(\phi_t|\phi_{t-1}) \tag{3}$$

where $\mathrm{P}(\phi_1)$ is the initial state distribution, $\mathrm{P}(\phi_t|\phi_{t-1})$ the transition probabilities and $\mathrm{P}(o_t \,|\, \phi_t)$ the observation likelihoods. These probabilities constitute the parameters of the proposed beat tracking model and reasonable estimates are provided below.

2) Estimation of the Observation Likelihoods: The observation likelihoods $\mathrm{P}(o_t \,|\, \phi_t)$ need to be estimated for the N_τ states of the model. A common approach to determine the parameters of the HMM is to model the observation distributions with a Gaussian mixture model (GMM) and automatically learn these distributions using a Baum–Welch algorithm [25]. However, this approach is computationally very demanding and it requires a large number of training samples. To avoid this situation, reasonable estimates for the state-conditional distributions can be obtained.

Recall that the phase observation signal o_t is designed to show large values at event locations. As a result, it is reasonable for the beat state observation likelihood $P(o_t \,|\, \phi_t = 0)$ to be assumed proportional to the observation

$$\mathrm{P}(o_t \,|\, \phi_t = 0) \propto o_t. \tag{4}$$

Similarly, reasonable estimates for the non-beat state observation likelihoods $\{\mathrm{P}(o_t \,|\, \phi_t = n) : n \neq 0\}$ can be obtained. Although the observation likelihoods of states submultiples of the beat period will probably show a different distribution, the observation model is simplified by assuming that all non-beat states $\{\phi_t : \phi_t \neq 0\}$ are identically distributed. This state-tying approach is equivalent to the data model simplification introduced in [28]. We could try to find a suitable distribution for each of the non-beat states; however, state-conditional distributions show significant variability from genre to genre as discussed in [15]. Again, it is acceptable to assume that the phase observation signal o_t will show small values at non-beat locations. Then, a reasonable estimate for the non-beat state observation likelihood functions $\{\mathrm{P}(o_t|\phi_t = n) : n \neq 0\}$ is

$$\mathrm{P}(o_t \,|\, \phi_t = n) \propto 1 - o_t. \tag{5}$$

These estimates are equivalent to using a first-order polynomial to model the state-conditional distributions.

Section IV-B discusses the goodness of these observation likelihood estimates, comparing this simple model with a trained approach where the model parameters are automatically learned.

3) Estimation of the Initial and Transition Probabilities: The initial probability $\mathrm{P}(\phi_1)$ models the time instant when the first beat is expected to be. We do not make any assumption over the location of the first beat, therefore a discrete uniform distribution for $\mathrm{P}(\phi_1)$ is chosen.

The transition probabilities $\mathrm{P}(\phi_t \,|\, \phi_{t-1})$ encode the temporal structure of the sequence of beats. Specifically, beats are expected to be regularly spaced in time with small deviations from the beat period τ_t. The probability density function of the time between consecutive beats at any time instant, Δ, is modeled to be proportional to a Gaussian distribution centered at the beat period τ_t

$$\mathrm{P}(\Delta = n) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n-\tau_t)^2}{2\sigma^2}\right) \tag{6}$$

where the standard deviation σ models the tolerance to tempo deviations that occur in musical performances [12] and timing deviations caused by the temporal resolution of the onset signal [29]. The Gaussian distribution is normalized to sum to unity in order to be a valid probability distribution. As in [21], a value of 0.02 s is chosen for the standard deviation and then $\sigma = 1.72$ frames.[2]

The number of states of the HMM will be determined by the largest time between beats allowed. Assuming a maximum time

[2] Recall that the time-resolution is 11.6 ms, Section II-A.

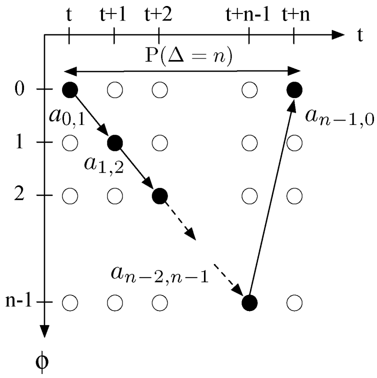


Fig. 4. Relation between state transition probabilities a_{ij} and the distribution of the time between beats $\mathrm{P}(\Delta)$. States are represented by circles and transitions by links.

between beats of $\tau_t + 3\sigma$, we account for 99% of the support of the Gaussian distribution in (6). This value must agree with the maximum time between beats measured by the hidden state variable ϕ, which is $N_{\tau_t} - 1$. Therefore, the total number of states of the proposed beat tracking model is given by

$$N_{\tau_t} = \tau_t + 3\sigma + 1. \tag{7}$$

As shown in Fig. 4, if there are $\Delta = n$ frames between two consecutive beats, the state transition probabilities $a_{ij} = \mathrm{P}(\phi_t = j|\phi_{t-1} = i)$ and the distribution of the time between beats $\mathrm{P}(\Delta)$ in (6) can be related as

$$a_{n-1,0} = \frac{\mathrm{P}(\Delta = n)}{\prod_{k=0}^{n-2} a_{k,k+1}} \tag{8}$$

$$a_{n-1,n} = 1 - a_{n-1,0} \tag{9}$$

with $n \in \{1, \ldots, N_{\tau_t}\}$. Note that (9) reflects that the only possible transitions allowed by our model are the transitions from state $\phi_{t-1} = n$ to the following non-beat state $\phi_t = n + 1$ or to the beat state $\phi_t = 0$.

In summary, the estimates of the observation likelihoods $\mathrm{P}(o_t \,|\, \phi_t)$, the initial probabilities $\mathrm{P}(\phi_1)$ and the transition probabilities $\mathrm{P}(\phi_t \,|\, \phi_{t-1})$ define the proposed beat tracking model and the sequence of beats, $\mathcal{B}^*$, can be obtained using a Viterbi algorithm as described in Section II-C1.

D. Beat Tracking Quality Assessment

Beat tracking accuracy is determined by the musical and physical properties of a performance [10], [11] but also by the specific limitations of the beat tracking algorithm. In particular, the behavior of the probabilistic framework proposed here relies on the correctness of the beat period estimation. In some cases, the quality of the beat period salience observations used for period estimation can be poor. For example, the time–frequency analysis may not be appropriate to the characteristics of the musical signal or the signal does not show any clear periodicity.

In order to characterize the quality of the feature signals used for beat period estimation, three measures are calculated. First, a peak-to-average ratio, q_{par}, that relates the maximum amplitude of the beat period salience observation signal with its root-mean-square value is computed as

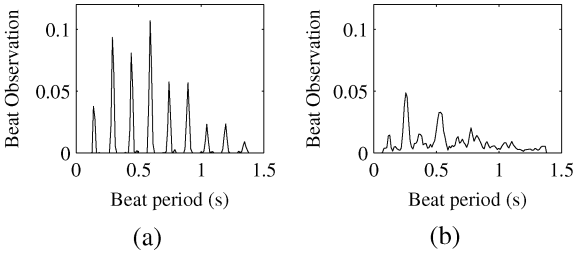


Fig. 5. Time average of the beat period salience observation $s(t, \tau)$ showing (a) a clear rhythmic structure and (b) noisy beat period salience observation.

$$q_{\mathrm{par}} = \frac{\max_\tau |\bar{s}(\tau)|}{\sqrt{\frac{1}{\tau_{\max}} \sum_{\tau=1}^{\tau_{\max}} \bar{s}(\tau)^2}} \tag{10}$$

where $\tau_{\max}$ is the maximum beat period (in time frames). The signal $\bar{s}(\tau)$ denotes the time average of the beat period salience observation $s(t, \tau)$ used for tempo estimation

$$\bar{s}(\tau) = \frac{1}{T} \sum_{t=1}^{T} s(t, \tau). \tag{11}$$

The second quality value $q_{\max}$ measures the maximum of the beat period salience observation time average and it is simply calculated as

$$q_{\max} = \max_\tau |\bar{s}(\tau)|. \tag{12}$$

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

$$q_{\mathrm{kur}} = \min_t k_{s(t,\tau)} \tag{13}$$

where $k_{s(t,\tau)}$ is the sample kurtosis of $s(t, \tau)$ in the variable τ. This quality measure q_{kur} measures how outlier-prone the beat period salience observation sample distribution is.

The vector of quality measures is defined as $\mathbf{q} = [q_{\mathrm{par}}\ q_{\max}\ q_{\mathrm{kur}}]$. Large values of these quality measures are expected for beat period salience observations $s(t, \tau)$ that reflect a clear periodic structure. As an example, Fig. 5 shows the time average of the beat period salience observation signal $s(t, \tau)$ used for tempo estimation in two audio excerpts of the database described in Section III-A. While a clear periodic structure can be seen in Fig. 5(a), the beat period salience observation shown in Fig. 5(b) is noisy and therefore we do not expect to obtain a good beat period estimate.

E. Reliability Estimation

Based on the quality measure vector $\mathbf{q}$, a quantity that reflects the reliability of the set of beat estimates $\mathcal{B}^*$ obtained by the beat tracking algorithm is calculated. This reliability measure, denoted as r_p, is determined by using a k-Nearest Neighbor (k-NN) regression algorithm [30]. The reliability value of the tracking algorithm for a given musical signal is assigned to be the average of the values of its k nearest neighbors which are calculated using the Euclidean distance. Informal tests show that the Euclidean distance provides slightly better accuracy results than other metric spaces, including Mahalanobis and standardized Euclidean distances.

Let p represent a measure of performance of the beat estimates. The performance measure p can be any of the evaluation criteria discussed in [31] and introduced in Section III-A, e.g., the AMLc criterion. Let $\mathcal{I} = \{1, \ldots, I\}$ be a set of training audio signals, $\{\mathbf{q}^i : i \in \mathcal{I}\}$ the set of quality vectors and $\{p^i : i \in \mathcal{I}\}$ the set of performance measures for each of the training samples. Given a new audio signal with quality $\mathbf{q}$, the distance to the quality measures of the training set is calculated as

$$d^i = ||\mathbf{q} - \mathbf{q}^i||_2 \tag{14}$$

where $||\cdot||_2$ denotes the euclidean norm. Then, the set of indexes of the k nearest neighbors can be easily calculated by sorting the set of distances $\{d^i : i \in \mathcal{I}\}$ and it is denoted as $\mathcal{K}$. Finally, the reliability under the performance criteria p of its beat estimates $\mathcal{B}^*$ is calculated as the mean performance of its k nearest neighbors as

$$r_p = \frac{1}{k} \sum_{j \in \mathcal{K}} p^j. \tag{15}$$

In summary, the system learns the relationship between the quality measures $\{\mathbf{q}^i : i \in \mathcal{I}\}$ and the beat tracking performance $\{p^i : i \in \mathcal{I}\}$ and predicts the performance of a new audio signal r_p based on the measured quality $\mathbf{q}$ using a k-NN. Therefore, the beat tracking reliability measure r_p can be interpreted as the expected performance accuracy in terms of the evaluation criteria p. Although these quality measures are specifically designed to address the limitations of the proposed beat tracking algorithm, the reliability analysis presented here defines a general framework that can be potentially applied to any beat tracking method. First the limitations of the new beat tracker have to be identified, then a suitable set of quality measures should be defined and finally a regression method like the one presented here can be used to predict the accuracy of the new beat estimates.

The proposed reliability-informed beat tracking algorithm includes both a set of beat estimates $\mathcal{B}^*$ and a measure of the reliability of those beat estimates r_p. Thus, the user of the beat tracker is additionally informed with the reliability of the beat estimates provided by the automatic beat tracking algorithm. As shown in Section IV-D, we will be able to successfully predict the performance of the beat tracking algorithm and, introducing an innovative evaluation framework, show how the performance of the proposed beat tracker can be increased by identifying and removing musical excerpts where the beat tracker has very low confidence.

III. Experimental Setup

This section describes the database and the performance measures used to evaluate the proposed beat tracking system. In addition, we detail the systems used for comparison and describe how the methods are compared.

A. Database and Evaluation

For the evaluation of the proposed beat tracking method, the database described in [18] and studied in [32], [19], and [3] is used. The database has been designed for beat tracking evaluation and consists of 222 musical audio files, divided into six categories: Dance (40), Rock/Pop (68), Jazz (40), Folk (22), Classical (30), and Choral (22). The database includes a reasonable number of styles, tempos and time signatures. Audio files are around 60 seconds in length with time-variable tempo. The files were annotated by a trained musician, recording a human clapping signal and using the claps as beat locations. Difficult examples were manually corrected by moving beat locations interactively.

Evaluating a beat tracking system is not trivial. A manually annotated beat is an estimate of the actual beat location and an exact match between the estimated beat position given by an algorithm and the annotated beat is unlikely. In addition, there is an ambiguity associated to the metrical level annotation since human tapping responses to the same musical excerpt can be very different [33]. The most common situations are the *anti-phase* tapping (a set of annotations on the on-beat and the other set on the off-beat) and the *half* and *double* tapping rate (the rate of an annotation set is half or twice the other set). Therefore, many methods have been proposed to evaluate the performance of beat trackers: the well-known F-measure [24], the mean Gaussian error accuracy presented by Cemgil *et al.* [34], the cross-correlation based P-score [1], the binary accuracy measure of Goto *et al.* [35], the information gain measure presented in [36] and the continuity-based evaluation methods [18], [20]. A detailed description and comparison of the different evaluation methods can be found in [31].

To evaluate the performance of the proposed beat tracking algorithm, the continuity-based measures have been chosen. This allows us to analyze both the ambiguity associated to the annotated metrical level and continuity in the beat estimates. These accuracy measures consider regions of continuously correct beat estimates relative to the length of the audio signal analyzed. Continuity is enforced by defining a tolerance window of 17.5% relative to the current inter-annotation-interval [31]. Also, to allow initializations, events within the first five seconds of the input audio signal are discarded. The continuity-based criteria used for performance evaluation are the following:

- CMLc (Correct Metrical Level with continuity required) which gives information about the longest segment of continuously correct beat tracking;
- CMLt (Correct Metrical Level with no continuity required) which accounts for the total number of correct beats at the correct metrical level;
- AMLc (Allowed Metrical Level with continuity required) the same as CMLc but it accounts for ambiguity in the metrical level;
- AMLt (Allowed Metrical Level with no continuity required) the same as CMLt but it accounts for ambiguity in the metrical level.

For the AML measures, the annotations are resampled to allow tapping at half and double the correct metrical level and tapping at the off-beat. As in the MIREX beat tracking evaluation task

[2], we use the beat tracking evaluation toolbox[3] presented in [31].

In [20], the impact of beat estimation errors is analyzed from a human perspective. It was found that continuity is very important and that metrical ambiguity is not very disturbing. Therefore, it seems that a relevant evaluation criterion is the AMLc measure. In our discussion, we will pay special attention to this performance criterion.

B. Reference Systems

The performance of the proposed model is compared with four beat tracking algorithms: the publicly available beat tracking algorithms of Dixon [13] and Ellis [16], the context-dependent beat tracker of Davies *et al.* [19] and the probabilistic beat tracker of Klapuri *et al.* [20]. To informally analyze the behavior of our automatic system with respect to a human tapper, the human tap times from [19] are also included. These taps were recorded by a human tapper with some musical experience using a computer keyboard but, contrary to the ground truth annotations, no manual correction was applied.

To compare the different systems, the mean values of the performance measures across all the audio files of the test database are presented. For a more detailed analysis, box plots showing the median and 25th and 75th percentiles are also presented. Following [37], statistical significant difference on the mean values is also checked. We use an analysis of variance test (ANOVA) [38] and a multiple comparison procedure [39] when comparing with the reference systems. A multiple comparison procedure is useful to compare the mean of several groups and determine which pairs of means are significantly different. A pairwise comparison could lead to spurious statistical difference appearances due to the large number of pairs to be compared. To overcome this situation, multiple comparison methods provide an upper bound on the probability that any comparison will be incorrectly declared significant. A significance level of 5% is chosen to declare the difference statistically meaningful. This value is commonly used in hypothesis testing.

IV. Results

In this section, the performance of the proposed and publicly available[4] beat tracking system is analyzed. We evaluate the relevance of the different elements of the phase observation model and analyze the convenience of using a Baum–Welch algorithm [25] to automatically determine the parameters of the HMM model. Also, the performance of the proposed beat tracker is compared with state-of-the-art systems. Finally, a new perspective on beat tracking evaluation is presented. The beat reliability measure is used to discuss how to improve the performance of the algorithm and how far automatic beat tracking is from a human performance example.

A. Phase Observation Model Relevance

TABLE I
Beat Tracking Performance Analysis (%) for Different Model Simplifications

	Evaluation measure			
Model assumptions	CMLc	CMLt	AMLc	AMLt
1. Non-beat states disabled	**56.5**	62.9	71.5	80.6
2. Beat state disabled	55.0	60.8	70.3	79.5
Proposed model	56.1	**62.9**	**71.9**	**81.5**

Table I shows the results of the proposed probabilistic system under different model assumptions. The proposed model is the one described in Section II-C, which exploits both the beat and the non-beat state information. The relevance of the different elements of the model is evaluated by selecting the information the model uses. The first model assumption disables the non-beat state information by setting the phase observation likelihood for the non-beat states to a non-informative uniform distribution, $P(o_t \,|\, \tau_t = n) = 1$ for $n \neq 0$. Thus, this model looks for the sequence of time instants where the phase observation likelihood $P(o_t \,|\, \tau_t = 0)$ in (4) is large. This assumption slightly degrades the performance of the proposed model in AML. The assumption used in this experiment is then analogous to the probabilistic beat tracking approach of Klapuri *et al.* [20] and Peeters [21]. These methods decode, respectively, the time instants where a beat occurs looking at the beat "strength" at that time instant. The second assumption, instead disables the beat information state by setting the phase observation likelihood for the beat state to a flat distribution, $P(o_t \,|\, \tau_t = 0) = 1$. In this case, the model looks for a sequence of time instants where the phase observation between beats is low as given in (5). The system is still competitive. This is interesting considering that the approach does not use the observations at beat time instants and only accounts for the observations between beat times to be low. Although we only find statistically significant differences in AMLt when comparing the proposed model with the first model assumption, these experiments suggest that both the beat and non-beat state observations can be exploited for beat tracking.

The model proposed in this paper is somewhat related to the beat tracking algorithms presented by Peeters [21] and Laroche [15]. In these works, only the beat information is considered and a beat-template is used to estimate the beat likelihood from the observations. In short, this beat template reflects that large observation values are expected at multiples of the beat period. However, our system also exploits non-beat information by explicitly modeling non-beat states.

B. Training the Beat Model

An alternative to determining the parameters of the beat tracking model is to automatically learn the transition probabilities and observation likelihoods from a set of training samples. In order to evaluate the convenience of the simplifications introduced in Section II-C2, a learning experiment has been conducted. For each audio file, the parameters of the HMM are determined using the Baum–Welch algorithm [25] where the phase observations constitute the training samples. The observation-likelihood distributions are modeled with a GMM.

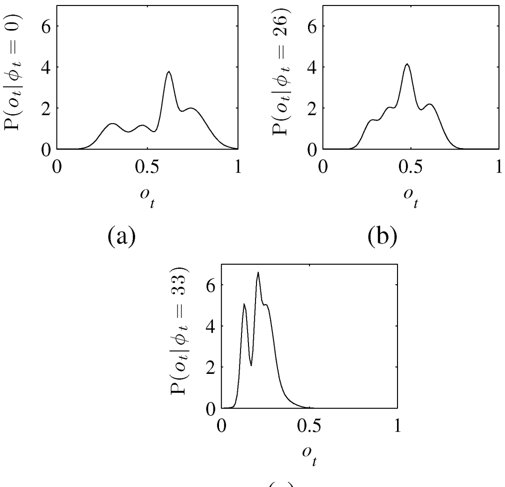
Table II shows the performance of the beat tracking algorithm for a different number of mixtures in the GMM and the proposed model in the last row. The performance increases with

[3] http://www.elec.qmul.ac.uk/digitalmusic/downloads/beateval/

[4] http://www.gts.uvigo.es/~ndegara/Publications.html

TABLE II
BEAT TRACKING MEAN PERFORMANCE ANALYSIS (%) FOR DIFFERENT NUMBER OF MIXTURES IN THE GMM

	Evaluation measure			
Mixtures	CMLc	CMLt	AMLc	AMLt
1	53.2	61.3	67.0	78.0
2	54.1	61.6	69.0	79.0
4	54.6	62.2	70.0	80.4
8	55.5	62.0	71.2	80.7
16	45.2	57.7	59.9	76.0
Proposed model	**56.1**	**62.9**	**71.9**	**81.5**



(a) (b)

(c)

Fig. 6. Observation likelihood $\mathrm{P}(o_t \mid \phi_t)$ estimates using a GMM with four mixtures for (a) the beat state 0, (b) half the beat period, 26 samples, and (c) a state not related with the beat period, in this example 33. The audio excerpt is the first file of the database and the annotated beat period is 52 samples.

the number of mixtures and for 16 mixtures performance decreases, likely due to overfitting problems. The best AMLc and AMLt mean performance values are obtained with a GMM with eight mixtures but these values are still smaller than the corresponding performance values of the proposed model. This result supports the validity of the observation likelihood simplification introduced in Section II-C2. Although it is found that the mean accuracies are not significantly different, it seems reasonable to choose this simplification because it is much less demanding in computational terms and its generalization ability is demonstrated in terms of performance.

To analyze the learned observation likelihood distributions, Fig. 6 shows estimates of the observation likelihood $P(o_t \mid \tau_t)$ using a GMM with four mixtures for: 1) the beat state, 2) half the beat period, and 3) a state not related with the beat period. The annotated beat period of the audio example is 52 samples. As shown by the distribution of the beat state and half the beat period state, large observations are more likely for beat period related states. On the contrary, smaller observation values are obtained for states that are not related with the beat period. This agrees with the rhythmic nature of music since events are more likely to happen at beat-period related instants.

As shown in Table II, modeling each state *individually* does not lead to better performance results. A HMM is a generative model and the Baum–Welch algorithm learns the parameters that best explain the observations. Thus, this learning approach does not imply that a beat-tracking performance measure is maximized. In fact, the state-tying model simplification introduced in Section II-C2 assumes that all non-beat states $\{\phi_t : \phi_t \neq 0\}$ are identically distributed and, as shown in Table II, its mean performance is higher than any of the GMM models. Therefore, it is reasonable to choose this simpler model where all non-beat states are tied together.

TABLE III
BEAT TRACKING MEAN PERFORMANCE (%) OF THE DIFFERENT METHODS

	Evaluation measure			
Method	CMLc	CMLt	AMLc	AMLt
Proposed model	**56.1**	**62.9**	**71.9**	**81.5**
Klapuri et al. [20]	55.6	62.0	69.7	79.3
Davies et al. [19]	54.7	60.9	67.1	76.3
Ellis [16]	45.6	51.0	67.8	76.6
Dixon [13]	36.8	47.9	52.0	72.3
Human tapper	52.6	80.5	57.2	87.0

C. Comparison to Other Systems

We turn now to compare the performance of the proposed beat tracker with a human tapper and the systems introduced in Section III-B. The same complex domain detection function was used for the proposed reference model, Davies *et al.* [19] and Ellis [16] methods. The Klapuri *et al.* [20] algorithm uses a more elaborate sub-band based detection function and a joint estimation of the beat, tatum, and measure pulse periods. Dixon's method [13] uses the spectral flux detection function described in [24].

In Table III, the mean accuracy of the different beat tracking algorithms is compared. The original implementations of the reference systems have been used to evaluate their performance on the selected database. The performance of the human tapper introduced in [19] is also included.

Relatively low performance of the automatic beat trackers is observed when continuity at the correct metrical level is required (CMLc). The reason for this low performance is that beat estimates must agree with the metrical level chosen by the human annotator. Interestingly, the human tapper performs worse in terms of CMLc than the two best performing automatic approach but this difference is not statistically significant. When correct metrical level is required but not continuity (CMLt), the human tapper performs statistically better than automatic trackers. These results suggest that the human tapper agreed more often with the annotator in terms of the metrical level, but was prone to isolated tapping errors which adversely affected the performance scores where temporal continuity of beats was enforced.

When comparing accuracy results with allowed metrical levels (AML), we find statistically significant differences between the performance of automatic trackers and human tappers. On the one hand, the AMLc performance of the human tapper, 57.2%, is substantially lower than the proposed beat tracker, 71.9%. On the other hand, if continuity is not required, the human tapper outperforms any of the automatic approaches and these differences are statistically significant.

To analyze the influence of annotations, it is interesting to compare CML and AML criteria. Larger values for the AML measures are found. This suggests that, unlike the human tapper, the automatic tempo induction methods fail to accurately estimate the metrical level chosen by the annotator. Therefore, low performance shown in terms of CML is imposed by the tempo induction method that informs the beat tracker and not by the beat tracking algorithm itself. We expect that improvements in tempo induction should lead to improvements under the CML criteria.

Finally, we compare the proposed beat tracking algorithm with the reference systems. As shown in Table III, the proposed method outperforms the reference methods in the mean value for all of the evaluation criteria. However, not all the differences are statistically significant. We find statistically significant differences between the proposed algorithm and the following reference methods for the evaluation criteria specified next:

- CMLc, Ellis and Dixon methods;
- CMLt, Ellis and Dixon methods;
- AMLc, Davies *et al.* and Dixon methods;
- AMLt, Davies *et al.*, Ellis and Dixon methods.

We do not find statistically significant differences between the proposed beat tracker and Klapuri *et al.* system. Both methods define a probabilistic framework based on a hidden Markov model and the number of states is equivalent in both systems since it is determined by the length of the beat period. Whereas the number of transitions from each state is two in our system (from one state to the next state or the beat state), the number of transitions per state in Klapuri *et al.* method is potentially equal to the number of states.

For a more detailed analysis of the results, box plots for the AML performance measures are also presented in Fig. 7(a) and (b). The central mark is the median, the edges of the box are the 25th and 75th percentiles and the lines extend to the most extreme data points not considered outliers. The 25th AMLt percentile is 79.5% for the proposed algorithm (*Prop.*) and 63% for Klapuri *et al.* (*Klap.*) approach. This means that the AMLt performance of the proposed algorithm is above 79.5% in 75% of the input files. On the contrary, the 75th percentile of Klapuri *et al.* (*Klap.*) is 99.0%, slightly larger than the 75th percentile of the proposed system which is 98.2%. It can be also observed that the interquartile range (the difference between the 75th and 25th percentiles) for the proposed system are the smallest in both figures, suggesting a more robust behavior of the proposed probabilistic model.

D. Reliability Analysis

The risk of focusing our analysis in performance *averages* is to neglect the reason a beat tracker is not able to correctly estimate the beat positions for a particular audio signal. As shown in Section II-D, the observations used to extract the beat period can be very noisy because either the signal analysis is not appropriate to the characteristics of the signal or the signal does not show any clear periodicity. This can potentially lead to a wrong tempo estimation and, as a result, to an erroneous determination of the beat positions.

But, is it possible to automatically predict a poor behavior of the proposed beat tracking algorithm? To answer this question we analyze the relation between the quality measures introduced in Section II-D and the performance of the proposed beat tracking algorithm on the test database. Fig. 8(a) and (b) show an scatter plot of the AMLc and AMLt measures. Each circle represents a test audio signal and the color of each marker is based on the values of the performance criteria, low performance is mapped to black and high performance to white. The circles are displayed at the locations specified by two of the quality measures introduced in Section II-D: the kurtosis, q_{kur}, and the maximum, $q_{\max}$. Looking at these figures, it is clear that there is a strong correlation between these quality measures and the performance of the algorithm for both the AMLc and AMLt performance measures. In fact, low accuracy results can be expected when the beat period salience observation quality measures are low. Any other pair combination of the quality measures would show a similar correlation between quality and performance.

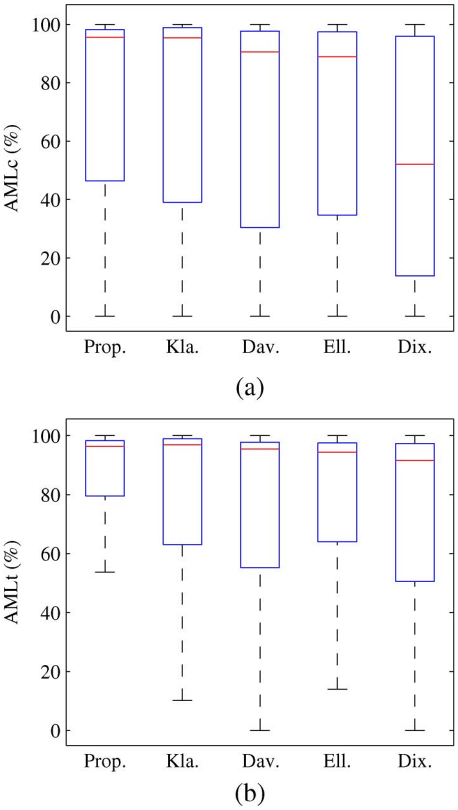


Fig. 7. Box plot for the AML (%) measures. (a) AMLc. (b) AMLt. Each column represents an algorithm: the proposed algorithm (*Prop.*), Klapuri *et al.* [20] (*Klap.*), Davies *et al.* [19] (*Dav.*), Ellis [16] (*Ell.*), and Dixon [13] (*Dix.*). The central mark is the median, the edges of the box are the 25th and 75th percentiles and the lines extend to the most extreme data points not considered outliers.

The system presented in this paper learns the relationship between the quality measures and the expected beat tracking performance r_p using a k-NN as defined in (15). As in [40], evaluation is done using a leave-one-out strategy: the reliability of a musical signal is estimated using all the other musical signals in the test set as training samples. Informal tests show that $k = 3$ nearest neighbors are enough to correctly estimate the reliability measure r_p.

This reliability-informed approach opens a new perspective in beat tracking. Just as humans often have some insight into how difficult it is to tap along to an audio signal, the beat tracking reliability measure r_p represents the expected performance accuracy (in terms of the evaluation criteria p, for example AMLc) of the beat tracking algorithm on the input musical signal. Therefore, the output of our reliability-informed

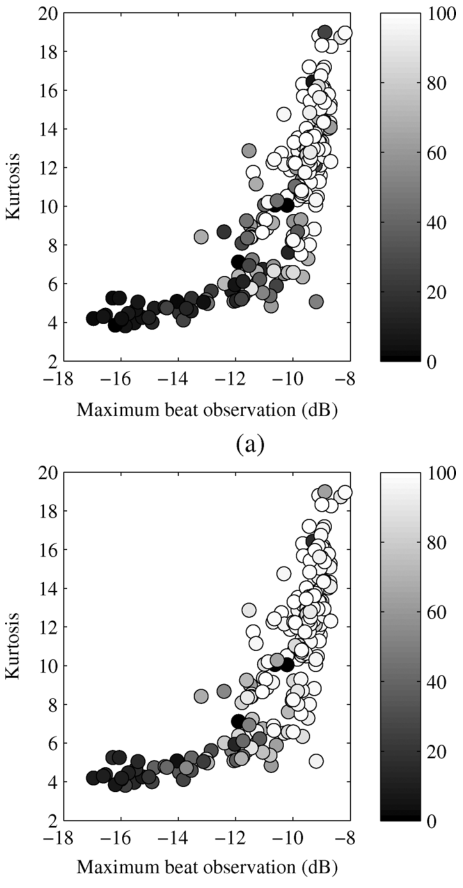


Fig. 8. Scatter plot of the performance measure versus the kurtosis ratio and the maximum beat period salience observation. (a) AMLc criterion. (b) AMLt criterion. Each circle represents a test audio signal and the color of each marker are based on the values of the performance criteria, low performance is mapped to the black color and large performance to white.

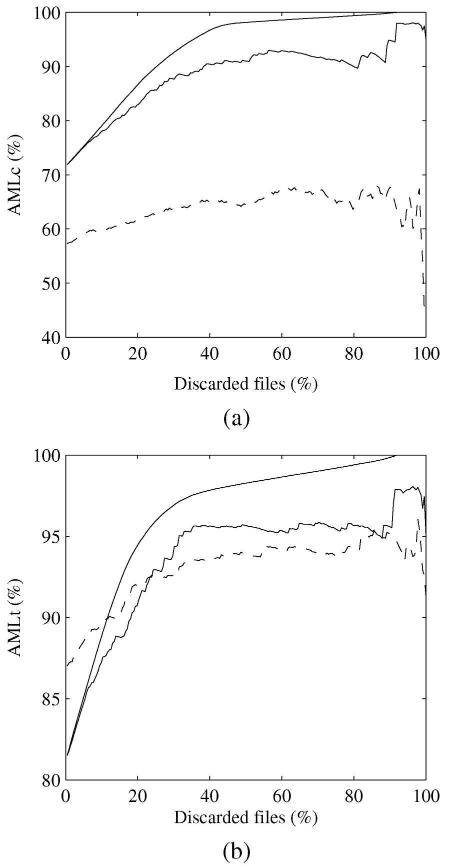


Fig. 9. Mean performance versus discarded number of files. (a) AMLc criterion. (b) AMLt criterion. The black solid-line represents the mean performance of the proposed quality-based algorithm and the dashed-line the performance of the human tapper on the files selected according to the reliability measure. The gray solid-line represents the "oracle" performance which discards files according to the actual score.

beat tracking algorithm includes both a set of beat estimates and a measure of the reliability of those beat estimates. The user of the beat tracker is then informed with the reliability of the beat estimates provided by the automatic beat tracking algorithm. Therefore, if we sought to annotate the beats of a musical signal with the assistance of a beat tracking algorithm the reliability could be used to decide whether to trust the set of beat estimates or to enter the tap times manually.

Reliability information can be successfully exploited to increase the mean accuracy of the proposed beat tracker if some files are discarded. Instead of analyzing the performance on the whole test set as in MIREX [2], a target on the number of files allowed to be discarded can be defined. Using the reliability information to identify the musical excerpts where the beat tracker has very low confidence in its beat output, we can re-evaluate overall performance of the beat tracker systematically discarding these "poorly tracked" files, weakest first. In this way we can automatically determine a sub-set of the evaluation database and, in effect, improve the performance of our beat tracker.

Fig. 9(a) and (b) shows the mean AMLc and AMLt performance for different target values of files to be discarded. The black solid-line represents the mean performance of the proposed algorithm and the dashed-line the performance of the human tapper on the selected files. The gray solid-line is the "oracle" mean performance which discards files according to the actual performance of the proposed beat tracking algorithm and not the reliability measure. Obviously, we cannot use the actual score value to automatically discard files but it gives insight on the accuracy of the reliability measure. As can be seen in the figure, the difference in mean performance of the proposed beat tracking system is smaller than 5% compared to that of the "oracle" when up to 40% of the files are discarded. For larger numbers of discarded files, the difference is larger, but we still have a fair approximation of the "oracle" performance.

As can be seen in Fig. 9(a) and (b), the mean performance of the proposed algorithm and the human tapper agree with the mean results presented in Table III when we do not discard any files. However, as we discard audio files according to the reliability measures[5], r_{AMLc} and r_{AMLt}, the mean performance of the proposed algorithm significantly increases, both in AMLc and AMLt. For example, the mean AMLc increases from 71.9% to 85.8% and the mean AMLt from 81.5% to 92.9% when discarding 25% of the input files. This indicates that the reliability measure introduced in Section II-D is a good indicator of the goodness of the beat estimates provided by the beat tracker.

Finally, it is also interesting to compare the accuracy of the quality-based beat tracking approach with the human tapper. On

[5] Note that p is replaced by the name of the performance criteria AMLc and AMLt in (15).

the one hand, Fig. 9(a) compares the proposed beat tracker and the human tapper in terms of the rate of files to be decoded using the AMLc criterion. As can be seen, the performance of the proposed beat tracker is initially superior to the human tapper in terms of AMLc and the difference increases even more when using the reliability measure information. On the other hand, using the AMLt criterion in Fig. 9(b), we see that the human performs better than the automatic approach when all the files have to be decoded. By automatically selecting files according to the expected performance of our proposed beat tracker, we can informally demonstrate that it outperforms a typical human tapper when allowed to choose a subset of 80% (or fewer of) the input files. This is far from a rigorous comparison between human tapping and computational beat tracking as the human taps used were entered in real-time and were left unaltered whereas the presented beat tracking algorithm is non-causal. However we can use this result to demonstrate that, by removing automatically the files where the beat tracker fails catastrophically, we can observe a distinct improvement in performance.

V. Conclusion

In this paper, a reliability-informed beat tracking method that analyzes musical signals has been presented. To integrate musical-knowledge and signal observations, a probabilistic framework that exploits both beat and non-beat information is proposed. The influence of the different elements of the proposed probabilistic model has been evaluated and results show that a significant increment in AMLt performance is obtained by including non-beat information. In addition, reasonable estimates for the parameters of the model are proposed. To validate the accuracy of these estimates, a large learning experiment where the parameters of the model were determined using a Baum–Welch algorithm has been completed. Results show no significant differences between the trained approach and the proposed simplification.

The proposed beat tracking system has been compared with four reference systems. The method outperforms all the reference systems in the mean value under all the evaluation criteria used. We find significant differences in three of the four references systems when comparing AML criteria. A more detailed analysis of the distribution of the performance scores shows that the proposed system achieves the highest 25th percentile value. Also, the interquartile range of our probabilistic framework is the smallest, suggesting a more robust behavior.

We also studied if we are able to predict a poor performance of the system, finding a strong correlation between the observation quality measures and the performance of the beat tracker. In addition, a k-nearest neighbor regression algorithm to automatically measure the reliability of the beat estimates is proposed. This differs from current beat tracking systems which exclusively estimate beat locations and do not account for the specific limitations of the algorithm. We show that we can successfully exploit reliability information by discarding those files where an unacceptable performance of the algorithm is expected. In this way, mean accuracy significantly increases, increasing from 71.9% to 85.8% in AMLc and from 81.5% to 92.9% in AMLt when discarding 25% of the input files. We informally demonstrated that the beat tracking system can outperform a typical human tapper (using AMLt) by exploiting the proposed reliability measure; in effect, allowing the beat tracker to pick a subset of the evaluation database itself.

The conclusions extracted from our reliability-informed analysis result in a number of ideas for future work. We plan to explore the combination of different beat tracking algorithms. Files discarded for having a low reliability measure could be handled by a different beat tracking algorithm with a higher predicted reliability so as the final performance of the global system is higher. Similarly, multiple input features could be combined or fused together in order to obtain a better representation of the rhythmic structure of the musical signal to be analyzed. Future work will also concentrate on exploiting users' inputs such as a human estimate of the actual tempo of an audio signal, the genre of the signal to be tapped or the estimated difficulty of the example. Finally, we plan to explore the joint estimation of the beat phases and periods using a probabilistic framework.

References

[1] M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri, "Evaluation of audio beat tracking and music tempo extraction algorithms," *J. New Music Res.*, vol. 36, no. 1, pp. 1–16, 2007.

[2] The Music Information Retrieval Evaluation eXchange (MIREX), Beat Tracking Evaluation Task, Jun. 2010, [Online]. Available: http://www.music-ir.org/mirex/wiki/2010:Audio_Beat_Tracking

[3] A. M. Stark, M. E. P. Davies, and M. D. Plumbley, "Real-time beat-synchronous analysis of musical audio," in *Proc. 12th Int. Conf. Digital Audio Effects (DAFx-09)*, Como, Italy, Sep. 2009.

[4] M. Levy and M. Sandler, "Structural segmentation of musical audio by constrained clustering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 2, pp. 318–326, Feb. 2008.

[5] A. Robertson and M. D. Plumbley, "B-Keeper: A beat-tracker for live performance," in *Proc. Int. Conf. New Interfaces for Musical Expression (NIME)*, New York, Jun. 6–9, 2007, pp. 234–237.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] S. Dixon, "An empirical comparison of tempo trackers," in *Proc. 8th Brazilian Symp. Comput. Music*, Fortaleza, Brazil, 31 Jul.–3 Aug. 2001, pp. 832–840.

[11] P. Grosche, M. Müller, and C. S. Sapp, "What makes beat tracking difficult? A case study on chopin mazurkas," in *Proc. 11th Int. Conf. Music Information Retrieval (ISMIR)*, Utrecht, The Netherlands, Aug. 2010.

[12] F. Gouyon and S. Dixon, "A review of automatic rhythm description systems," *Comput. Music J.* , vol. 29, no. 1, pp. 34–54, 2005.

[13] S. Dixon, "Evaluation of audio beat tracking system BeatRoot," *J. New Music Res.*, vol. 36, no. 1, pp. 39–51, 2007.

[14] 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.

[15] J. Laroche, "Efficient tempo and beat tracking in audio recordings," *J. Audio Eng. Soc.*, vol. 51, no. 4, pp. 226–233, 2003.

[16] D. P. W. Ellis, "Beat tracking by dynamic programming," *J. New Music Res.*, vol. 36, pp. 51–60, 2007.

[17] A. T. Cemgil and H. J. Kappen, "Monte Carlo methods for tempo tracking and rhythm quantization," *J. Artif. Intell. Res.*, vol. 18, pp. 45–81, 2003.

[18] S. W. Hainsworth, "Techniques for the automated analysis of musical audio," Ph.D. dissertation, Univ. of Cambridge, Cambridge, U.K., Sep. 2004.

[19] 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.

[20] A. P. Klapuri, A. J. Eronen, and J. T. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 342–355, Jan. 2006.

[21] G. Peeters, "Beat-tracking using a probabilistic framework and linear discriminant analysis," in *Proc. 12th Int. Conf. Digital Audio Effects (DAFx-09)*, 2009.

[22] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, Sep. 2005.

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

[24] S. Dixon, "Onset detection revisited," in *6th Int. Conf. Digital Audio Effects (DAFx-06)*, Montreal, QC, Canada, Sep. 18–20, 2006, pp. 133–137.

[25] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[26] M. P. Ryynänen and A. P. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music," *Comput. Music J.*, vol. 32, no. 3, pp. 72–86, 2008.

[27] J. Forney and G. D. , "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[28] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 360–370, Apr. 1999.

[29] L. Daudet, G. Richard, and P. Leveau, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," in *Proc. 5th Int. Conf. Music Inf. Retrieval (ISMIR)*, Barcelona, Spain, Oct. 10–14, 2004.

[30] R. O. Duda, P. E. Hart, and D. G. Stork, "Nonparametric techniques," in *Pattern Classification*. New York: Wiley-Interscience, 2000.

[31] M. E. P. Davies, N. Degara, and M. D. Plumbley, "Evaluation methods for musical audio beat tracking algorithms," Centre for Digital Music, Queen Mary Univ., Tech. Rep. C4DM-TR-09-06, 2009.

[32] S. W. Hainsworth and M. D. Macleod, "Particle filtering applied to musical tempo tracking," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 2385–2395, 2004.

[33] D. Moelants and M. McKinney, "Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous," in *Proc. 8th Int. Conf. Music Percept. Cognit. (ICMPC8)*, 2004.

[34] A. T. Cemgil, H. J. Kappen, P. Desain, and H. Honing, "On tempo tracking: Tempogram representation and Kalman filtering," *J. New Music Res.*, vol. 28, no. 4, pp. 259–273, 2001.

[35] M. Goto and Y. Muraoka, "Issues in evaluating beat tracking systems," in *Working Notes of the IJCAI-97 Workshop Iss. AI and Music—Eval. Assess.*, Aug. 1997, pp. 9–16.

[36] M. E. P. Davies, N. Degara, and M. D. Plumbley, "Measuring the performance of beat tracking algorithms using a beat error histogram," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 157–160, 2011.

[37] A. Flexer, "Statistical evaluation of music information retrieval experiments," *J. New Music Res.*, vol. 35, no. 2, pp. 113–120, Jun. 2006.

[38] R. V. Hogg and J. Ledolter, *Engineering Statistics*. New York: MacMillan, 1987.

[39] H. Y. and A. C. Tamhane, *Multiple Comparison Procedures*. New York: Wiley, 1987.

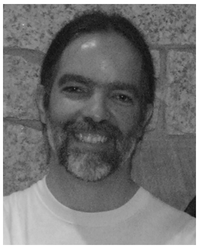
[40] A. Eronen and A. Klapuri, "Music tempo estimation with k-NN regression," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 50–57, Jan. 2010.

Norberto Degara (S'10) received the telecommunications engineer degree from the University of Vigo, Vigo, Spain, in 2001 and the M.S. degree in electrical engineering from the University of Texas (UT) at Austin, in 2007. He is currently pursuing the Ph.D. degree at the University of Vigo.

From 2004 to 2005, he worked as a Research Engineer at LMS International, Belgium, as a Marie Curie Fellow. He was recipient of a Pedro Barrié de la Maza Foundation fellowship for continuation of studies at UT Austin. In 2009, he visited the Centre for Digital Music, Queen Mary University of London, London, U.K. His research focuses on audio and music signal processing, including onset detection, beat tracking, and rhythm analysis.

Enrique Argones Rúa (M'10) received the M.Sc. and Ph.D. degrees (Honors) in telecommunications engineering from the University of Vigo, Vigo, Spain, in 2003 and 2008, respectively.

He is actually working on several projects related to biometrics. His research interests include face verification, video processing, online signature verification, biometric cryptosystems, and other pattern recognition problems.



Antonio Pena (M'93) received the M.S. and Ph.D. degrees in telecommunication engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 1990 and 1994, respectively.

He has been an Associate Professor with the Universidad de Vigo, Vigo, Spain, since 1995. His research activity has been closely related to real-time implementations of MPEG audio coders for broadcast equipment from 1992 to 2001. Nowadays, research on sound signal analysis for applications on acoustics, including sound source separation and subjective evaluation, and the coordination of the trademark "Sonitum," providing both acoustic consulting and teaching, are his main activities.

Soledad Torres-Guijarro received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Universidad Politcnica de Madrid (UPM), Madrid, Spain, in 1992 and 1996, respectively.

She actually enjoys a professor position at the Universidad de Vigo, Vigo, Spain. Previously, she enjoyed a Parga Pondal researcher position in the R&D Department of the Metrology Laboratory of Galicia and a Ramón y Cajal researcher position in the UPM and Universidad de Vigo. She has also enjoyed several teaching positions in the following universities: Valladolid, Carlos III de Madrid, and Europea de Madrid. She has taken part in 21 research projects with public founding, three of them with European founding, and nine research projects with private founding. She is coauthor of ten papers, one book chapter, and one patent. Her research interests include digital signal processing applied to speech, audio, and acoustics.

Matthew E. P. Davies (M'07) received the B.Eng. degree in computer systems with electronics from King's College London, London, U.K., in 2001 and the Ph.D. degree in electronic engineering from Queen Mary University of London, London, U.K., in 2007.

From 2007 until 2011, he was a Postdoctoral Researcher in the Centre for Digital Music, Queen Mary University of London. He has recently joined the Sound and Music Computing Group at INESC in Porto, Portugal. His research interests include beat tracking and rhythm analysis, evaluation methods, music therapy, and sparse representations.

Mark D. Plumbley (S'88–M'90) received the B.A. (Honors) degree in electrical sciences and the Ph.D. degree in neural networks from the University of Cambridge, Cambridge, U.K., in 1984 and 1991, respectively.

From 1991 to 2001, he was a Lecturer at King's College London, London, U.K. He moved to Queen Mary University of London in 2002, where he is now an EPSRC Leadership Fellow and Director of the Centre for Digital Music. His research focuses on the automatic analysis of music and other audio sounds, including automatic music transcription, beat tracking, and audio source separation, and with interest in the use of techniques such as independent component analysis (ICA) and sparse representations.

Prof. Plumbley chairs the ICA Steering Committee, and is a member of the IEEE SPS TC on Audio and Acoustic Signal Processing.