Bayesian Modelling of Temporal Structure in Musical Audio

Nick Whiteley, A.Taylan Cemgil and Simon Godsill

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ, UK
{npw24, atc27, sjg}@eng.cam.ac.uk

Abstract

This paper presents a probabilistic model of temporal structure in music which allows joint inference of tempo, meter and rhythmic pattern. The framework of the model naturally quantifies these three musical concepts in terms of hidden state-variables, allowing resolution of otherwise apparent ambiguities in musical structure. At the heart of the system is a probabilistic model of a hypothetical 'bar-pointer' which maps an input signal to one cycle of a latent, periodic rhythmical pattern. The system flexibly accommodates different input signals via two observation models: a Poisson points model for use with MIDI onset data and a Gaussian process model for use with raw audio signals. The discrete state-space permits exact computation of posterior probability distributions for the quantities of interest. Results are presented for both observation models, demonstrating the ability of the system to correctly detect changes in rhythmic pattern and meter, whilst tracking tempo.

Keywords: tempo tracking, rhythm recognition, meter recognition, Bayesian inference

1. Introduction

In construction of intelligent music systems, an important perceptual task is how to infer attributes related to temporal structure. These attributes may include musicological constructs such as meter and rhythmic pattern. The recognition of these characteristics forms a sub-task of automatic music transcription - the unsupervised generation of a score, or description of an audio signal in terms of musical concepts.

For interactive performance systems, especially when an exact score is *a-priori* unknown, it is crucial to construct robust algorithms that can correctly operate under rhythmic fluctuations, ritardando/accelarando (systematic slowing down or speeding up), metric modulations, etc. For music categorisation systems, tempo and rhythmic pattern are defining features of genre. It is therefore apparent that a complete system should be able to recognise all these features.

Much work has been done on detecting the 'pulse' or foot-tapping rate of musical audio signals. Existing algorithmic approaches dealing with raw audio signals focus on the extraction of some frequency variable representing this rate [1],[2]. However these approaches do not detect more complex temporal characteristics, such as meter and rhythmic pattern, which are needed for complete transcription. Goto and Muraoka detail a multiagent hypothesis system which recognises beats in terms of the 'reliability' of hypotheses for different rhythmic patterns, given a fixed 4/4 meter [3].This system was later extended to deal with non-percussive music [4].

Cemgil and Kappen also introduce musical concepts by modelling MIDI onset events in terms of a tempo process and switches between quantised score locations [5]. Raphael independently proposed a similar system [6]. Hainsworth and Macleod build on some elements of this work inferring beats from raw audio signals using an onset detector [7], but in these works meter and rhythmic pattern are still not explicitly modelled.

Approaches to meter detection include 'Gaussification' of onsets [8], similar to the Tempogram representation of Cemgil et. al. [9], autocorrelation methods, [10],[11] and preference-rule based systems, [12],[13]. Pikrakis et al. extract meter and tempo, assuming that meter is constant [14]. Takeda et al. perform tempo and rhythm recognition from a MIDI recording by analogy with speech-recognition [15]. Klapuri et. al. define metrical structure in terms of pulse sensations on different time scales [16]. This system is successful in detecting periodicity on these time scales, but does not yield an explicit estimate of musical meter in terms of standard musical notation: 3/4, 4/4 etc.

In this paper we focus on three musical concepts: tempo, meter and rhythmic pattern. The strength of the system presented here, compared to the existing works described above, is that it explicitly models these three concepts in a consistent framework which resolves otherwise apparent ambiguities in musical structure, for example switches in rhythmic pattern or meter. Furthermore, the probabilistic approach permits robust inference of the quantities of interest, in a principled manner.

In the Bayesian paradigm one views tempo tracking and recognition of meter and rhythmic pattern as a latent state inference problem, where given a sequence of observed data

$Y_{1:K}$, we wish to identify the most probable hidden state trajectory $X_{0:K}$. Firstly we need to postulate prior distributions over $X_{0:K}$. Secondly, an observation model is defined which relates the observations to the hidden variables. The posterior distribution over hidden variables is given by Bayes' theorem:

$$p(X_{0:K}|Y_{1:K}) = \frac{p(Y_{1:K}|X_{0:K})p(X_{0:K})}{p(Y_{1:K})} \quad (1)$$

This admits a recursion for online ('causal'), potentially real-time computation of filtering distributions of the form $p(X_k|Y_{1:k})$, from which estimates of the current tempo, rhythmic pattern and meter can be made, given observations up to the present time. Defining $\alpha_k \equiv p(X_k|Y_{1:k})p(Y_{1:k})$,

$$\alpha_k = p(Y_k|X_k) \int dX_{k-1} p(X_k|X_{k-1})\alpha_{k-1} \quad (2)$$

For off-line inference, *smoothing* may be performed, which conditions on future as well as present and past observations. Intuitively, smoothing is the retrospective improvement of estimates. Defining $\beta_k \equiv p(Y_{k+1:K}|X_k)$,

$$\beta_k = \int dX_{k+1} p(X_{k+1}|X_k)p(Y_{k+1}|X_{k+1})\beta_{k+1} \quad (3)$$

A smoothing distribution for a single time index may be obtained in terms of the corresponding filtering distribution:

$$p(X_k|Y_{1:K}) \propto \alpha_k \beta_k \quad (4)$$

2. Dynamic Bar Pointer Model

We here define the 'bar pointer' as being a hypothetical, hidden object located in a space consisting of one period of a latent rhythmical pattern, i.e. one bar. The velocity of the bar pointer is defined to be proportional to tempo, measured in quarter notes per minute. Note that we therefore avoid explicitly modelling hierarchical periodicity, as described for example in [16] in terms of measure, tactus and tatum periods, but such quantities could be be extracted from the model if required.

In qualitative terms, for a given rhythmical pattern there are locations in the bar at which onsets will occur with relatively high probability. This concept is quantified and related to observed signals in section 3 below. In this section we define the prior model for the bar pointer dynamics.

We choose to define a discrete 'position' space in terms of M uniformally spaced points in the interval $[0, 1)$. Denote by $m_k \in \{1, 2, ..., M\}$ the index of the location of the bar-pointer in this space at time $k\Delta$, where Δ and $k \in \{1, 2, ..., K\}$ are respectively the discrete time period and index. Next define a discrete 'velocity' space with N elements and denote by $n_k \in \{1, 2, ..., N\}$ the index of the velocity of the bar pointer at time index k. A similar construction called a 'score position pointer' is defined in [17], but for this model exact inference is intractable and switches in rhythmic pattern and meter are not explicitly modelled.

Over time the position and velocity indices of the bar pointer evolve according to:

$$m_{k+1} = (m_k + n_k - 1) mod\,(M\theta_k) + 1 \quad (5)$$

for $1 < n_k < N$,

$$p(n_{k+1}|n_k) = \begin{cases} \frac{p_n}{2}, & n_{k+1} = n_k \pm 1 \\ 1 - p_n, & n_{k+1} = n_k \\ 0, & otherwise \end{cases} \quad (6)$$

where p_n is the probability of a change in velocity. At a boundary, $n_k = 1$ or $n_k = N$, the velocity either remains constant with probability $1 - p_n$, or transitions respectively to $n_{k+1} = 2$ or $n_{k+1} = N - 1$ with probability p_n. A modulo operation is implied in a similar context in [6], to allow calculation of note lengths in terms of quantised score locations. In the dynamic bar pointer model, this modulo operation is made explicit and exploited to allow representation of switches in meter. The meter indicator variable, θ_k, takes one value in a finite set, for example $\theta_k \in T = \{3/4, 4/4\}$, at each time index k. This facilitates modelling of switches between $3/4$ and $4/4$ meters during a single musical passage and is illustrated in figure 1. The advantage of this

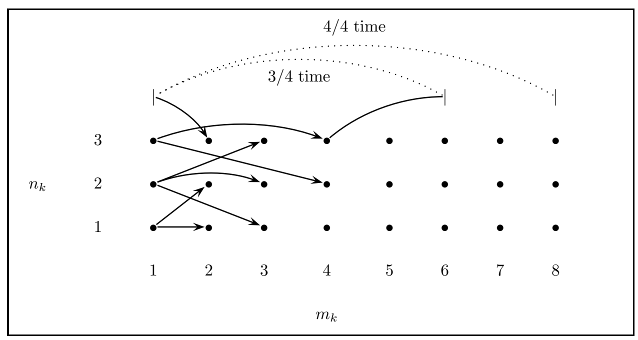

Figure 1: Toy example of the position and velocity state subspace for $M = 8$, $N = 3$. Solid lines indicate examples of possible state transitions and dotted lines indicate the effect of the modulo operation for different meters. The implication is that, for a given tempo, one bar in 3/4 meter is simply 3/4 the length of one bar in 4/4 meter. The concept generalises to other meters, eg bars in 2/4, 3/4, and 4/4 can all be represented as subsets of a bar in 5/4 meter. Note that in practice M would be chosen much larger.

approach compared to the existing resonator-based system of Scheirer, [1], is that the bar pointer continues to move whether or not onsets are observed. This explicitly models the concept that tempo is a latent process and provides robustness against rests in the music which might otherwise be wrongly interpreted as local variations in tempo. Large and Kolen formulate a phase locking resonator, but it is not given a full probabilistic treatment [18].

Switches in meter are modelled as occurring when the bar-pointer passes the end of the bar:

for $m_k < m_{k-1}$,

$$p(\theta_k|\theta_{k-1}, m_k, m_{k-1}) = \begin{cases} p_\theta, & \theta_k \neq \theta_{k-1} \\ 1 - p_\theta, & \theta_k = \theta_{k-1} \end{cases} \quad (7)$$

otherwise,

$$p(\theta_k|\theta_{k-1}, m_k, m_{k-1}) = \begin{cases} 0, & \theta_k \neq \theta_{k-1} \\ 1, & \theta_k = \theta_{k-1} \end{cases} \quad (8)$$

where p_θ is the probability of a change in meter at the end of the bar.

The last state variable is a rhythmic pattern indicator, r_k, which takes one value in a finite set, for example $r_k \in S = \{0, 1\}$, at each time index k. The elements of the set S correspond to different rhythmic patterns, described in section 3. For now we deal with the simple case in which there are only two such patterns, and switching between values of r_k is modelled as occurring at the end of the bar:

for $m_k < m_{k-1}$,

$$p(r_k|r_{k-1}, m_k, m_{k-1}, \theta_{k-1}) = \begin{cases} p_r, & r_k \neq r_{k-1} \\ 1 - p_r, & r_k = r_{k-1} \end{cases} \quad (9)$$

otherwise,

$$p(r_k|r_{k-1}, m_k, m_{k-1}, \theta_{k-1}) = \begin{cases} 0, & r_k \neq r_{k-1} \\ 1, & r_k = r_{k-1} \end{cases} \quad (10)$$

where p_r is the probability of a change in rhythmic pattern at the end of a bar.

In summary, $X_k \equiv [m_k \ n_k \ \theta_k \ r_k]^T$ specifies the state of the system at time index k. For computation, the set of all possible states may be arranged into a vector $\mathbf{x}$ and the state of the system at time k may then be represented by $X_k = \mathbf{x}(i)$, ie the ith element of this vector. Using equations 5 - 10, a transition matrix $\mathbf{A}$ may then be constructed such that:

$$\mathbf{A}(i, j) = p(X_{k+1} = \mathbf{x}(j)|X_k = \mathbf{x}(i)) \quad (11)$$

For a value of M which is high enough to give useful resolution, eg. 1000, and a suitable value of N, eg 20, this matrix is large, but extremely sparse, making exact inference viable.

3. Observation Models

3.1. Poisson Points

Denote by y_k the number of MIDI onset events observed in the kth non-overlapping frame of length Δ. All other MIDI information, (eg. key, velocity, duration) is disregarded. The number y_k is modelled as being Poisson distributed:

$$p(y_k|\lambda_k) = \frac{\lambda_k^{y_k} \exp(-\lambda_k)}{y_k!} \quad (12)$$

A gamma distribution is placed on the intensity parameter λ_k. This provides robustness against variation in the data. The shape and rate parameters of the gamma distribution, denoted by a_k and b_k respectively, are functions of the value of a rhythmic pattern function, $\mu(m_k, r_k)$. The mean of the gamma distribution is defined to be the value of this rhythmic pattern function, which quantifies knowledge about the probable locations of note onsets for a given rhythmic pattern. This formalises the onset time heuristics given in [4]. Examples of rhythmic pattern functions are given in figure 2. For brevity denote $\mu_k \equiv \mu(m_k, r_k)$.

$$p(\lambda_k|m_k, r_k) = \frac{b_k^{a_k} \exp(-b_k\lambda_k)}{\Gamma(a_k)} \lambda_k^{a_k - 1} \quad (13)$$

$$a_k = \mu_k^2 / Q_\lambda \quad (14)$$

$$b_k = \mu_k / Q_\lambda \quad (15)$$

where Q_λ is the variance of the Gamma distribution, the value of which is chosen to be constant.

Inference of the intensity parameter λ_k is not required so it is integrated out. This may be done analytically, yielding:

$$p(y_k|m_k, r_k) = \frac{b_k^{a_k} \Gamma(a_k + y_k)}{y_k! \Gamma(a_k)(b_k + 1)^{a_k + y_k}} \quad (16)$$

Figure 3 gives a graphical representation of the combination of the Poisson Points observation model and the Dynamic Bar-Pointer model.

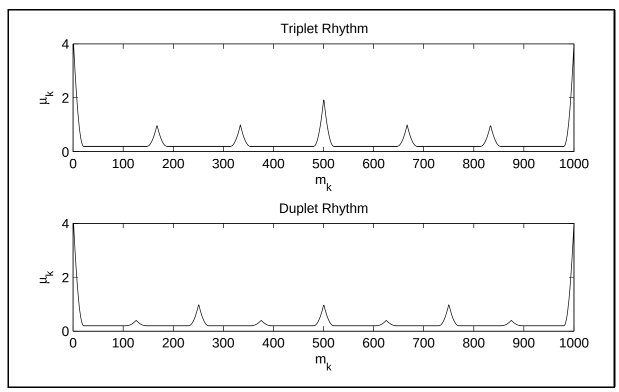

Figure 2: Two example rhythmic pattern functions for use with the Poisson points observation model, each corresponding to a different value of r_k. Top - a bar of triplets in 4/4 meter, Bottom - a bar of duplets in 4/4 meter. The locations of the peaks in the function correspond to the locations of probable onsets for a given rhythmic pattern. The heights of the peaks imply the average number of onset events at the corresponding bar locations, via a gamma distribution. The widths of the peaks model arpeggiation of chords and expressive performance. Construction in terms of splines permits flat regions between peaks, corresponding to an onset event 'noise floor'.

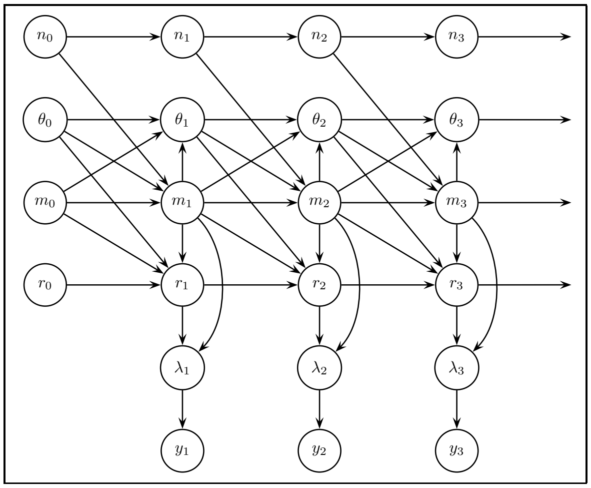

Figure 3: Graphical representation of the bar-pointer model in conjunction with the Poisson points observation model. Each node corresponds to a random variable. Directed links denote statistical dependence: each node is conditionally independent, given the values of its 'parent' nodes. The graph for the Gaussian process model is exactly equivalent.

3.2. Gaussian Process

The motivation for specifying this second observation model is to demonstrate that the dynamic bar pointer framework can be employed with raw audio as well as MIDI onset data. The Gaussian process model presented here is therefore simple and is suitable for percussive sounds only. However, it should be noted that this observation model could easily be modified to operate on other feature streams, such as those in defined in [2], taking into account changes in spectral content.

Denote by $\mathbf{z}_k$ a vector of ν samples constituting the kth non-overlapping frame of a raw audio signal. The time interval Δ is then given by $\Delta = \nu/f_s$, where f_s is the sampling rate. The samples are modelled as independent with a zero mean Gaussian distribution:

$$p(\mathbf{z}_k|\sigma_k^2) = \frac{1}{(2\pi\sigma_k^2)^{\nu/2}} \exp\left(-\frac{\mathbf{z}_k^T\mathbf{z}_k}{2\sigma_k^2}\right) \tag{17}$$

An inverse-gamma distribution is placed on the variance σ_k^2. The shape and scale parameters of this distribution, denoted by c_k and d_k respectively, are determined by the location of the bar pointer, m_k and the rhythmic pattern indicator variable r_k, again via a rhythmic pattern function, μ_k. An example of a rhythmic pattern function for use with this model is shown in figure 4.

$$p(\sigma_k^2|m_k, r_k) = \frac{d_k^{c_k}\exp(-d_k/\sigma_z^2)}{\Gamma(c_k)}\sigma_k^{-2(c_k+1)} \tag{18}$$

$$c_k = \mu_k^2/Q_s + 2 \tag{19}$$

$$d_k = \mu_k\left(\frac{\mu_k^2}{Q_s} + 1\right) \tag{20}$$

where Q_s is the variance of the inverse-gamma distribution and is chosen to be constant.

The variance of the Gaussian distribution, σ_k^2, may be integrated out analytically to yield:

$$p(\mathbf{z}_k|m_k, r_k) = \frac{d_k^{c_k}\Gamma(c_k+\nu/2)}{(2\pi)^{\nu/2}\Gamma(c_k)}\left(\frac{\mathbf{z}_k^T\mathbf{z}_k}{2} + d_k\right)^{-(c_k+\nu/2)} \tag{21}$$

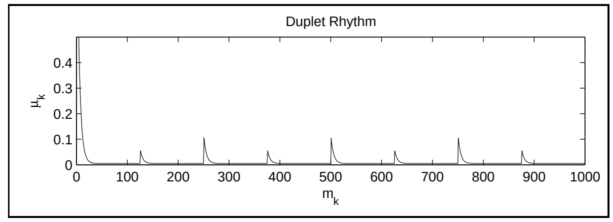

Figure 4: A duplet rhythmic pattern function for use with the Gaussian Process observation model, corresponding to one value of r_k. The function is piece-wise exponential: a simple model of energy transients for percussive onsets. The locations of the peaks correspond to temporal locations of probable onsets for a given rhythmical pattern. The heights of the peaks imply the average signal power at the corresponding bar locations, via an inverse-gamma distribution.

4. Inference Algorithm

Computation of posterior marginal filtering and smoothing distributions can be performed exactly using the forwards-backwards algorithm, see [19] and references therein for details. At each time step, the forward pass of the algorithm recursively computes a so-called 'alpha message' vector, $\boldsymbol{\alpha}_k$. Each element of this vector, $\boldsymbol{\alpha}_k(i)$, is proportional to $p(X_k = \mathbf{x}(i)|y_{1:k})$ - the corresponding element of the filtering distribution [1] . The recursion is given by:

$$\boldsymbol{\alpha}_{k+1} = \mathbf{O}_k\mathbf{A}^T\boldsymbol{\alpha}_k \tag{22}$$

$$\boldsymbol{\alpha}_0 = p(X_0) \tag{23}$$

where $\mathbf{A}$ is the state transition matrix as previously defined and $\mathbf{O}_k$ is a diagonal matrix:

$$\mathbf{O}_k(j,j) = p(y_k|X_k = \mathbf{x}(j)) \tag{24}$$

The forward pass can potentially be carried out in real time. The backward pass of the algorithm computes the beta messages, $\boldsymbol{\beta}_k$, which are then combined with the alpha messages to give the corresponding smoothing distribution:

$$\boldsymbol{\beta}_k = \mathbf{O}_{k+1}\mathbf{A}\boldsymbol{\beta}_{k+1} \tag{25}$$

$$\boldsymbol{\beta}_K = \mathbb{1} \tag{26}$$

[1] Notation for the Poisson points observation model. For the Gaussian process model, replace $y_{1:k}$ with $\mathbf{z}_{1:k}$.

$$p(X_k|y_{1:K}) \propto \boldsymbol{\alpha}_k * \boldsymbol{\beta}_k \qquad (27)$$

where $\mathbb{1}$ is a vector of ones and $*$ represents element-wise product.

5. Results

5.1. MIDI Onset Events

Figure 5 shows results for an excerpt of a MIDI performance of 'Michelle' by the Beatles, demonstrating the joint tempo-tracking and rhythm recognition capability of the system. The performance, by a professional pianist, was recorded using a Yamaha Disklavier C3 Pro Grand Piano. The Poisson points observation model was employed with the two rhythmic patterns in figure 2 and a single value of $\theta_k = 1$, ie 4/4 meter. Uniform initial prior distributions were set on m_k, n_k and r_k, with $M = 1000$ and $N = 20$. The time frame length was set to $\Delta = 0.02s$, corresponding to the range of tempi: $12 - 240$ quarter notes per minute. The probability of a change in velocity from one frame to the next was set to $p_n = 0.01$ and the probability of a change in rhythmic pattern was set to $p_r = 0.1$. The variance of the Gamma distribution was set $Q_\lambda = 10$.

This section of 'Michelle' is potentially problematic for tempo trackers because of the triplets, each of which by definition has a duration of 3/2 quarter notes. A performance of this excerpt could be wrongly interpreted as having a local change in tempo in the second bar, when really the rate of quarter notes remains constant; the bar of triplets is just a change in rhythm.

In figure 5, the strong diagonal stripes in the image of the posterior smoothing distributions for m_k correspond to the *maximum a-posteriori* (MAP) trajectory of the bar pointer. The system correctly identified the change to a triplet rhythm in the second bar and the subsequent reversion to duplet rhythm. The MAP tempo is given by the darkest stripe in the image for the velocity log-smoothing distribution - it is roughly constant throughout.

5.2. Raw Audio

A percussive pattern with a switch in meter (score given in figure 6) was performed and recorded in mono .wav format at $f_s = 11.025$kHz. The system was then run on this raw audio signal using the Gaussian Process observation model, to demonstrate joint tempo tracking and meter recognition. The single rhythmic pattern function given in figure 4 was employed in conjunction with two meter settings: $\theta_k \in \{3/4, 4/4\}$. The probability of a change in velocity was set $p_n = 0.01$ and the probability of a change in meter was set $p_\theta = 0.1$. The variance of the inverse-gamma distribution was set $Q_s = 10$. The frame length in samples was set $\nu = 256$ with $M = 1000$ and $N = 20$, corresponding to the range of tempi: $10.3 - 208$ quarter notes per minute. Uniform initial prior distributions were set on all hidden variables.

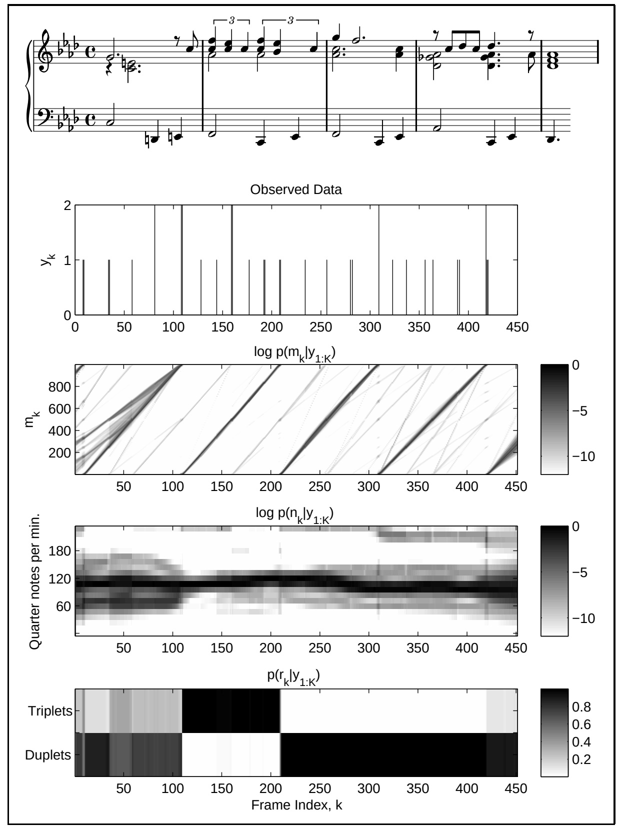

Figure 5: Results for joint tempo tracking and rhythmic pattern recognition on a MIDI performance of 'Michelle' by the Beatles. The top figure is the score which the pianist was given to play. Each image consists of smoothing marginal distributions for each frame index.

Whilst each section of this percussive pattern may appear simple at first glance, it poses two potential challenges for tempo trackers. Firstly, the lack of eighth notes in the middle two bars could incorrectly be interpreted as a reduction in tempo by a factor of two. However, the tempo measured as the rate of quarter notes remains constant. Secondly, the meter switch in the same two bars disrupts the accent pattern established in bars 1,2,5 and 6.

Figure 6 shows that the system correctly identified the metrical modulation. Note that the system is robust to the lack of eighth notes in the third and fourth bars; the MAP tempo is roughly constant throughout.

Audio files and supporting materials for further results may be found on the web at `http://www-sigproc.eng.cam.ac.uk/~npw24/ismir06/`

6. Discussion

A model of temporal characteristics of music has been presented, based around the probabilistic dynamics of a hypothetical bar pointer. Two observation models accommodate MIDI onset events or percussive raw audio, relating the observed data to the location of the bar pointer. Exact inference

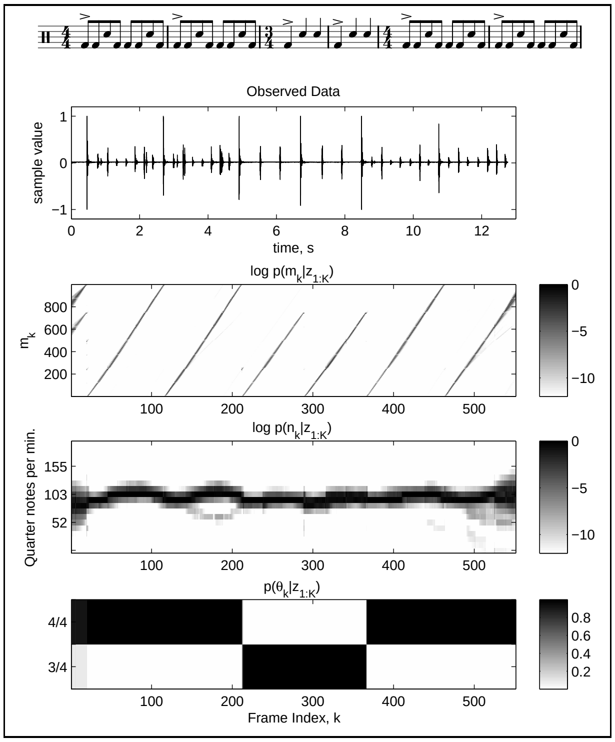

Figure 6: Results for joint tempo tracking and meter recognition from raw audio. The top most figure is the percussion score for the piece.

of tempo, rhythmic pattern and meter may be performed online (and potentially real time) from posterior filtering distributions, suitable for automatic accompaniment applications. For analysis and information retrieval purposes, off-line operation yields smoothing posterior distributions. Demonstrations of the capabilities of the system were presented for two pieces, one involving a change in rhythmic pattern and the other a switch in meter. The results show that the system robustly handles such temporal variations which might defeat simple tempo trackers.

We are currently investigating how this model could be extended to use MIDI volume information, via a marked Poisson process. Higher tempo resolution, larger numbers of meters and rhythmic patterns would require an even larger transition matrix, ultimately exceeding practical computational limits. Future work will therefore investigate the application of faster, approximate inference schemes, such as particle filtering, and the treatment of the position and velocity of the bar pointer as continuous variables.

References

[1] E. Scheirer, "Tempo and Beat Analysis of Acoustic Music Signals," in *J. Acoust. Soc. Am.*, vol 103, no. 1, pp 588-601, Jan. 1998.

[2] W. A. Sethares, R. D. Morris and J. C. Sethares. "Beat Tracking of Musical Performances Using Low-Level Audio Features." in *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 2, Mar. 2005.

[3] M. Goto and Y. Muraoka, "Music Understanding at the Beat Level - Real-Time Beat Tracking of Audio Signals," in *Proceedings of IJCAI-95 Workshop on Computational Auditory Scene Analysis*, pp. 68-75, 1995.

[4] M. Goto "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds", in *Journal of New Music Research*, vol. 30, no. 2, pp. 159-171, 2001.

[5] A. T. Cemgil and H. J. Kappen, "Monte Carlo methods for Tempo Tracking and Rhythm Quantization." in *Journal of Artificial Intelligence Research*, vol. 18, pp 45-81, 2003.

[6] C. Raphael, "Automated Rhythm Transcription" in *Proc. of the 2nd Ann. Int. Symp. on Music Info. Retrieval.*, pp. 99-107, Stephen Downie and David Bainbridge eds, 2001.

[7] S. Hainsworth and M. Macleod. "Beat tracking with particle filtering algorithms." in *Proceedings of IEEE Workshop on Applications of Signal Proc. to Audio and Acoustics*, New Paltz, New York, 2003.

[8] K. Frieler, "Beat and Meter Extraction Using Gaussified Onsets", in *Proc. of the 5th Ann. Int. Symp. on Music Info. Retrieval.*, 2004.

[9] A. T. Cemgil, H. J. Kappen, P. Desain, and H. Honing. "On tempo tracking: Tempogram Representation and Kalman filtering." in *Journal of New Music Research*, vol. 28 no. 4, pp. 259-273, 2001.

[10] J. Brown, "Determination of the meter of musical scores by autocorrelation", in *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp.1953-1957, 1993.

[11] D. Eck and N. Casagrande, "Finding Meter in Music Using an Auto-correlation Phase Matrix and Shannon Entropy", in *Proc. of the 6th Ann. Int. Symp. on Music Info. Retrieval.*, 2005.

[12] F. Lerdahl and R. Jackendoff, "A Generative Theory of Tonal Music." MIT Press, Cambridge, Massachusetts, 1983.

[13] D. Temperley and D. Sleator, "Modeling meter and harmony: A preference-rule approach." in *Computer Music Journal*, vol. 23, no. 1, pp.1027, 1999.

[14] A. Pikrakis, I. Antonopoulis and S. Theodoridis, "Music Meter and Tempo Tracking from Raw Polyphonic Audio" in *Proc. of the 5th Ann. Int. Symp. on Music Info. Retrieval.*, 2004.

[15] H. Takeda, T. Nishimoto and S. Sagayama, "Rhythm and Tempo Recognition of Music Performance from a Probabilistic Approach" in *Proc. of the 5th Ann. Int. Symp. on Music Info. Retrieval.*, 2004.

[16] A. Klapuri, A. Eronen, and J. Astola, "Analysis of the meter of acoustic musical signals" in *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14 no. 1, 2006.

[17] A. T. Cemgil, H. J. Kappen, and D. Barber, "Generative Model based Polyphonic Music Transcription" in *Proc. of IEEE WASPAA*, 2003

[18] E. W. Large and J. F. Kolen, "Resonance and the perception of musical meter," in *Connection Science*, vol. 6, no. 1, pp. 177-208, 1994.

[19] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning." *PhD Thesis, University of California, Berkeley*, 2002.