TOWARDS A MUSICAL BEAT EMPHASIS FUNCTION

Matthew E. P. Davies and Mark D. Plumbley[*]

Centre for Digital Music
Queen Mary University of London,
London E1 4NS, United Kingdom
matthew.davies@elec.qmul.ac.uk

Douglas Eck

University of Montreal
Department of Computer Science,
Montreal, Quebec H3C 3J7, Canada
eckdoug@iro.umontreal.ca

ABSTRACT

We present a new method for generating input features for musical audio beat tracking systems. To emphasise periodic structure we derive a weighted linear combination of sub-band onset detection functions driven a measure of sub-band beat strength. Results demonstrate improved performance over existing state of the art models, in particular for musical excerpts with a steady tempo.

1. INTRODUCTION

The automatic extraction of beat locations from audio signals forms a fundamental task in computer analysis of music signals. Beat tracking is strongly related to many high-level music analysis tasks including music transcription [1], automatic accompaniment [2], and structural segmentation [3]. Beat tracking systems are commonly used to perform a musically meaningful temporal segmentation of an audio input prior to higher level analysis. This use of "beat-synchronous" processing has been shown to improve performance over fixed-duration frame based analysis on high-level tasks including chord recognition [4] and music similarity [5]. Despite the growing use of beat tracking techniques in this way, the underlying problem of extracting beats is far from solved. Considerable challenges remain, including: reliable tempo tracking (especially for expressively performed music); synchronisation of beat locations to the perceived *on-beat*; and maintaining consistent beat estimates in music which is rhythmically syncopated or without prominent metrical structure.

The majority of existing techniques exploit the relationship between the start times of musical events (i.e. *note onsets*) and beat locations. The inputs to beat tracking systems either take the form of a discrete sequence of onset times [6] or a continuous mid-level representation which emphasises them [7], e.g. an onset detection function [8]. Given this input, the role of the beat tracker is then to recover a sequence of time instants (consistent with human "foot taps" in time to music) by separate [9] or simultaneous [10] estimation of beat period and phase. For a review see [11].

In this paper we address the link between input feature and beat tracking system. Our motivation is towards input features designed specifically for beat tracking and rhythm analysis. Our eventual aim is to generate a signal-dependent *beat emphasis function*.

The approach we adopt is based on the initial calculation of sub-band onset detection functions (as in [10, 7]). We hypothesise that periodic structure is often more prominent in certain sub-

bands than others. By using a measure of beat strength we derive a weighted linear combination of sub-band onset detection functions which boosts periodic structure while attenuating aperiodic information. We demonstrate that for certain types of input signal this emphasis of periodic structure can reduce the task of beat tracking to one of thresholding and peak-picking. Over a large database we show our system is able to exceed the current state of the art.

The remainder of this paper is structured as follows: in §2 we present the generation of the sub-band onset detection functions and derive periodicity-dependent linear weighting across the sub-bands; results are presented in §3 with discussion and conclusions in §4.

2. APPROACH

2.1. Sub-band onset detection functions

Within a large-scale test of input features for beat tracking, the complex spectral difference onset detection function [8] was been shown to be the most effective over a wide range of musical styles [12]. Towards our beat emphasis function we therefore adopt this as our starting point.

Given an input audio signal $x(t)$, the complex spectral difference method measures the Euclidean distance between observed and predicted short term spectral bins

$$C(k, m) = |X_k(m) - \hat{X}_k(m)|. \quad (1)$$

The standard onset detection function (DF), as used in [9], is found as the sum of Euclidean distances over all bins, k,

$$\Gamma(m) = \sum_{k=1}^{K} C(k, m) \quad (2)$$

where (as in [9]) K=512 and the temporal resolution of $\Gamma(m)$ is 11.6ms per DF sample.

In our approach we modify the process which generates the standard function by emphasising spectral bands with periodic structure. With K=512 the linear frequency spacing in (1) and (2) offers too fine a spectral resolution. To improve the likelihood of finding meaningful periodicity in spectral bands we group these narrow spectral bins into larger sub-bands. Example decompositions include having sub-bands equally spaced on the Mel scale [13] or ERB scale [7]. In our system we use a Gammatone filterbank from Slaney's Auditory Toolbox [14]. We create a matrix F whose rows are frequency responses to combine the fine spectral bands into wider sub-bands. An example filterbank is shown in Fig. 1 for 20 sub-bands. We multiply of F with C to give the
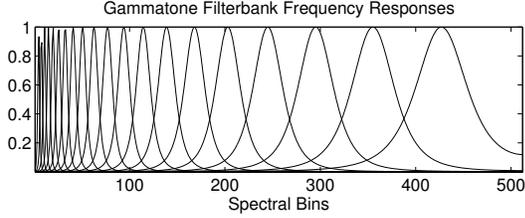
Figure 1: Frequency responses of the Gammatone filterbank for 20 sub-bands.

sub-band onset detection functions S_b,

$$S_b(m) = \sum_{b=1}^{B} F(b,k).C(k,m) \tag{3}$$

In §3 we explore the effect of varying the number of sub-bands B. To prevent any bias in the subsequent weighting of the sub-bands, we normalise each sub-band DF have to have unit variance.

2.2. Combining sub-bands

To generate the beat emphasis function we derive a weighting $w(b)$ which favours sub-bands with prominent periodic structure. To this end, we follow a technique for finding the beat period (i.e. time between beats) [9] which we now summarise.

To preserve the peaks of each sub-band DF, we calculate an adaptive "moving mean" threshold,

$$\bar{S}_b(m) = \text{mean}\{S_b(q)\} \quad m - \frac{Q}{2} \le q \le m + \frac{Q}{2} \tag{4}$$

where Q=16 DF samples. We subtract the moving mean threshold from each sub-band DF and, to prevent any negative values, half-wave rectify the output

$$\tilde{S}_b(m) = \text{HWR}(S_b(m) - \bar{S}_b(m)). \tag{5}$$

We then perform a band-wise unbiased autocorrelation

$$A_b(l) = \frac{\sum_{m=1}^{L} \tilde{S}_b(m)\tilde{S}_b(m-l)}{|l-L|} \quad l = 1, \ldots, L \tag{6}$$

where L=512 DF samples sets the maximum lag. Each autocorrelation function is then passed through a comb filtering process

$$R_b(l) = g(l) \sum_{p=1}^{4} \sum_{v=1-p}^{p-1} \frac{A_b(pl+v)}{2p-1} \tag{7}$$

where $g(l)$ is a tempo preference curve based on the Rayleigh distribution function,

$$g(l) = \frac{l}{\beta^2} \exp\left(\frac{-l^2}{2\beta^2}\right) \quad l = 1, \ldots, l_{\max} \tag{8}$$

with β=43 DF samples and $l_{\max}$=110 DF samples. The resulting sub-band comb filter outputs $R_b(l)$ are shown in Fig. 2

In [9] a single comb filter output function is calculated for one onset detection function input signal. Here, the index of the maximum value is used as an estimate of the beat period. In our approach we do not need to directly measure the periodicity (although this would clearly be possible from R_b), our interest is
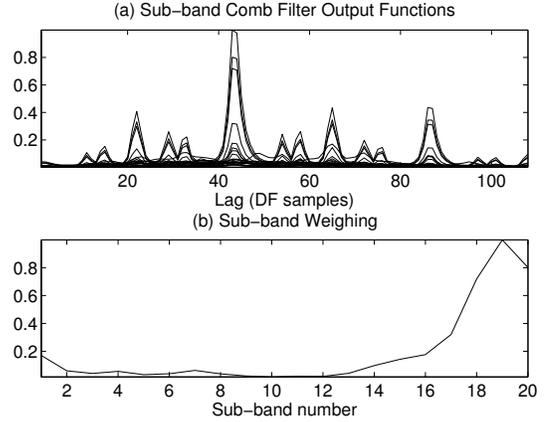

Figure 2: (a) Sub-band comb filter output functions. (b) The weighting value each sub-band, given as the peak height for the corresponding comb filter output function.

in using these comb filter functions to provide information about periodic structure. Given that each sub-band DF $S_b(m)$ is normalised to have unit variance, we can use the peak height as a measure of the strength of the periodic structure in a sub-band and use this directly to make the weighting,

$$w(b) = \max(R_b(l)). \tag{9}$$

This heuristic approach provides a simple musically meaningful way to linearly combine the sub-bands to provide a periodicity-weighted *beat emphasis function*,

$$\eta(m) = \sum_{b=1}^{B} w(b).S_b(m). \tag{10}$$

An example beat emphasis function and standard complex spectral difference onset detection function are shown in Fig. 3 for a 20 second excerpt of "Nothing compares 2U" by Sinead O'Connor. In this example the the vocals dominate the middle sub-bands, however the highest bands contain a regular pulse structure. This is emphasised to such an extent that the beat locations can be determined entirely by peak-picking without the need for further periodicity or phase estimation as would normally be required.

3. EVALUATION

In our evaluation we adopt two strategies to compare the beat emphasis function to the standard onset detection function. First, we compare the strength of each input feature at beat and non-beat positions. In the second part of the evaluation, we directly measure beat tracking performance.

We use the annotated beat tracking database from [11]. It contains 222 one minute long excerpts across six musical genres (dance, rock, jazz, folk, classical and choral), where all of the beat locations have been annotated by a musical expert.

3.1. Beat contrast measure

The primary aim of our beat emphasis function is to derive an onset detection function type representation with significant peaks at likely beat locations little energy elsewhere, as shown in Fig. 3(c).
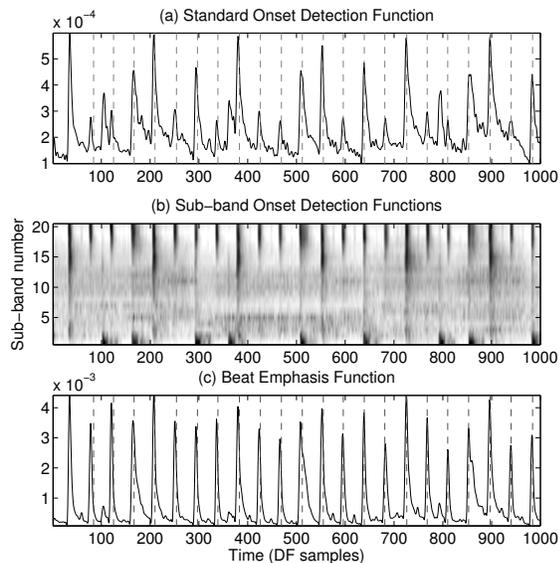
Figure 3: (a) Standard onset detection function with beat locations. (b) Sub-band onset detection functions. (c) Beat emphasis function calculated using the weighting in Fig. 2(b).

Input	Annotated	Random
Feature	Beats	Beats
BEF	1.501	1.002
ODF	1.398	1.003

Table 1: Comparison of beat contrast for the beat emphasis function (BEF) and the standard complex spectral difference onset detection function (ODF) for annotated beat positions and random beat locations.

To measure this property of increased *contrast* at beat locations we take small windows (in this instance, ± 2 DF samples) at each beat location and find the mean DF beat value, κ_{beat}. We then repeat the process finding the mean DF non-beat value over the larger windows covering all non-beat locations $\kappa_{\text{non-beat}}$. To measure beat contrast, c, we take the mean of the ratio of κ_{beat} to $\kappa_{\text{non-beat}}$ for all beats over all files n in the test database (N=222),

$$c = \frac{1}{N}\sum_{n=1}^{N}\frac{\kappa_{\text{beat}_n}}{\kappa_{\text{non-beat}_n}}. \tag{11}$$

In Table 1 we present the beat contrast measurements for the beat emphasis function and standard complex spectral difference onset detection function first using the (approx. 20,000) annotated beat locations and then with same number of random beat positions uniformly distributed across each test excerpt. For the random beat data, we find the mean beat contrast over one hundred iterations of random beats.

The results in Table 1 show that the both the beat emphasis function and standard complex spectral difference function are stronger close to beat locations than non-beat locations. When comparing the two input features the beat emphasis function shows

Beat	CML_c	AML_t
Tracker	(%)	(%)
BEF_{10}	59.7	82.5
BEF_{20}	58.7	83.2
BEF_{40}	57.8	82.0
ODF	56.1	81.0
KEA (NC)	55.7	80.0
DP (NC)	54.8	78.9

Table 2: Comparison of beat tracking performance. BEF_x refers to the hybrid beat tracker with the beat emphasis function as input and 'x' sub-bands. ODF uses the standard complex spectral difference input. KEA (NC) [7] and DP (NC) [9] are existing non-causal algorithms.

Genre	CML_c	AML_t
	(%)	(%)
Dance	78.6 (75.6)	98.0 (98.5)
Rock	83.5 (76.1)	93.1 (89.2)
Jazz	46.2 (49.2)	92.0 (90.7)
Folk	48.1 (44.2)	78.3 (68.7)
Classical	38.1 (39.7)	72.7 (74.4)
Choral	7.1 (5.6)	28.5 (27.4)

Table 3: Comparison of beat tracking for different musical genres using BEF_{20}. Performance for ODF is shown in parentheses.

greater contrast at beat positions. In the following section we investigate whether this translates to improved beat tracking performance.

3.2. Beat tracking accuracy

The beat tracking system we employ is a hybrid method based on the tempo tracking stage in [9] and the dynamic programming aspect in [13]. The two-state model for tempo tracking in [9] is replaced by a Viterbi decoding to find the best tempo path through successive comb filterbank output functions as calculated in §2 but over a single input feature. As a baseline, we provide this beat tracker with the standard onset detection function input, and then examine the effect using the beat emphasis function with different numbers of sub-bands (10, 20 & 40). We also present results from two reference systems: the two state model approach [9] which we refer to as DP, and the Klapuri et al method [7] which we refer to as KEA.

We measure beat tracking performance using the strictest and most lenient criteria from the continuity-based evaluation method [9]. The two continuity-based accuracy scores (CML_c and AML_t) reflect continuously correct tapping at the annotated metrical level (i.e. correct tempo) and the total number of correct beats allowing for tapping at twice or half the annotated tempo. Beats are considered accurate if they fall within allowance windows set at 17.5% of the inter-beat-interval. Overall beat tracking performance is summarised in Table 2 with a breakdown of accuracy by musical genre in Table 3.

4. DISCUSSION AND CONCLUSIONS

The results in Table 2 indicate that the use of the beat emphasis function as input to the beat tracking system exceeds the current state of the art as set by the DP and KEA methods. The number of sub-bands used to create the beat emphasis function does not appear critical, as performance is improved for each configuration. From a computational standpoint, it is most efficient to use fewer bands, however there is potential a trade-off with regard to masking periodic structure by using a coarser grouping of sub-bands. We intend to investigate this further.

Inspection of beat tracking performance broken down into genres in Table 3 reveals that the beat emphasis function gave most improvement to those genres characterised by a steady tempo, e.g. dance and rock. The folk category is interesting as many of the examples have a perceptually weak beat masked by strong non-rhythmic vocal lines. Our beat emphasis function is able to reduce the influence on the vocal range leading to improved tempo estimation and beat tracking.

Since our measure of beat strength is calculated once per sub-band, we should not expect it to be as effective for genres with a variable tempo (e.g. classical), where changes in tempo will lead to a flatter and therefore less informative weighting function. When the weighting is entirely flat the resultant beat emphasis function is very close in structure to the standard onset detection function.

A potential solution to calculating a beat emphasis function in a changing tempo context would be to take multiple measurements of beat strength at regular intervals across the duration of each sub-band DF. This would yield a *dynamic* weighting able to emphasise periodicity within regions of steady tempo. This approach could also be applied when analysing full length songs where it may be reasonable to assume that different sub-bands contribute the most significant periodic structure in different sections rather than one sub-band weighting per song. A section-wise weighting could be calculated given estimates of segment boundaries (e.g. from [3], or the reverse could be attempted, where the change in weighting functions could inform a segmentation algorithm.

In terms of future work, we intend to explore other techniques for weighting sub-bands, including the application of machine learning methods (e.g. [15]). In addition, we will research the use of multiple features beyond the complex spectral difference function towards a set of *style-specific* beat emphasis functions able to provide furthers increases in beat tracking performance. Finally, we wish to investigate the use of beat emphasis functions as input to other beat tracking systems.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] J. P. Bello-Correa, "Towards the automated analysis of simple polyphonic music: A knowledge-based approach," Ph.D. dissertation, Department of Electronic Engineering, Queen Mary, University of London, 2003.

[2] A. Robertson and M. D. Plumbley, "B-Keeper: A beat-tracker for live performance," in *Proceedings of the International Conference on New Interfaces for musical expression (NIME)*, New York, USA, June, 6–9 2007, pp. 234–237.

[3] M. Levy and M. Sandler, "Structural segmenation of musical audio by constrained clustering," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 156, no. 2, pp. 318–326, 2008.

[4] J. P. Bello and J. Pickens, "A robust mid-level representation for harmonic content in music signals," in *Proceedings of 6th International Conference on Music Information Retrieval*, London, United Kingdom, 2005, pp. 304–311.

[5] D. P. W. Ellis, C. Cotton, and M. Mandel, "Cross-correlation of beat-synchronous representations for music similarity," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, April 2008, pp. 57–60.

[6] S. Dixon, "Evaluation of audio beat tracking system beat-root," *Journal of New Music Research*, vol. 36, no. 1, pp. 39–51, 2007.

[7] A. P. Klapuri, A. Eronen, and J. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 342–355, 2006.

[8] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, part 2, pp. 1035–1047, 2005.

[9] M. E. P. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1009–1020, 2007.

[10] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *Journal of Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, 1998.

[11] S. Hainsworth, "Techniques for the automated analysis of musical audio," Ph.D. dissertation, Department of Engineering, Cambridge University, 2004.

[12] F. Gouyon, S. Dixon, and G. Widmer, "Evaluating low-level features for beat classification and tracking," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. IV, Hawaii, USA, April, 15–20 2007, pp. 1309–1312.

[13] D. P. W. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[14] M. Slaney, "Auditory toolox: A matlab toolbox for auditory modeling work," Interval Research Corporation, Tech. Rep., 1998. [Online]. Available: http://rvl4.ecn.purdue.edu/~malcolm/apple/tr45/AuditoryToolboxTechReport.pdf

[15] L. K. Saul and J. B. Allen, "Periodic component analysis: an eigenvalue method for representing periodic structure in speech," in *Advances in Neural Information Processing Systems (NIPS)*, 2001, pp. 807–813.