

BEAT CRITIC: BEAT TRACKING OCTAVE ERROR IDENTIFICATION BY METRICAL PROFILE ANALYSIS

Leigh M. Smith

IRCAM

leigh.smith@ircam.fr

ABSTRACT

Computational models of beat tracking of musical audio have been well explored, however, such systems often make “octave errors”, identifying the beat period at double or half the beat rate than that actually recorded in the music. A method is described to detect if octave errors have occurred in beat tracking. Following an initial beat tracking estimation, a feature vector of metrical profile separated by spectral subbands is computed. A measure of subbeat quaver (1/8th note) alternation is used to compare half time and double time measures against the initial beat track estimation and indicate a likely octave error. This error estimate can then be used to re-estimate the beat rate. The performance of the approach is evaluated against the RWC database, showing successful identification of octave errors for an existing beat tracker. Using the octave error detector together with the existing beat tracking model improved beat tracking by reducing octave errors to 43% of the previous error rate.

1. STRUCTURAL LEVELS IN BEAT PERCEPTION

The psychological and computational representation of listeners experience of musical time is of great application to music information retrieval. Correctly identifying the beat rate (*tactus*) facilitates further understanding of the importance of other elements in musical signals, such as the relative importance of tonal features.

Considerable research has proposed theories of an hierarchical structuring of musical time [12–14, 18, 20, 27], with the favouring of particular temporal levels. The *tactus* has been shown to be influenced by temporal preference levels [10], proposed as a resonance or inertia to variation [25]. At the metrical level¹, [21] argue that pre-established mental frameworks (“schemas”) for musical meter are used during listening. They found a significant difference in performance between musicians and non-music-

¹ A periodic repetition of perceived accentuation, notated in music as $\frac{4}{4}$, $\frac{3}{4}$ etc.

ians, arguing that musicians hold more resilient representations of meter, which favours hierarchical subdivision of the measure, than the non-musicians.

The fastest pulse has been used in ethnomusicology [16, 24] or reciprocally, the *tatum* in cognitive musicology [1] as a descriptive mechanism for characterising rhythmic structure. While it is not assumed to be a model of perception used by listeners and performers [16], the *tatum* is used to form a rhythmic grid of equally spaced intervals. It therefore represents the limit of hierarchical temporal organisation in complex rhythmic structures.

2. ERRORS IN BEAT TRACKING

Beat tracking or foot-tapping has a long history [7, 19], spurred on by the demands of music information retrieval [8, 15, 22, 23]. Common methods of beat tracking involve extraction of a mid-level representation, or onset detection function [23], typically derived from the spectral flux, thereby avoiding the requirement of identifying each individual onset. A number of methods have been proposed to then determine a time varying frequency analysis of the onset detection function, including comb filterbanks [6, 15, 23], autocorrelation [2, 9], dynamic time warping [8], Bayesian estimation [3], combined frequency and time lag analysis [22], coupled oscillators [17] and wavelet analysis [4].

Despite reporting very good results, there are areas for improvement to these approaches. A common task faced by many of these approaches is selecting the appropriate structural level from several viable candidates. It is a common occurrence to select a beat rate which is twice as fast as the actual performed rate, termed an “octave error”. For many of these systems, a reselection of the correct structural level from the candidates would be possible if the octave error could be detected.

The concept of fastest pulse can be used as an indicator of the highest structural level and therefore a datum. This appears in terms of the fastest alternation of events. Checking for quaver (1/8 note) alternation indicates if there is evidence of the fastest pulse appearing at the expected structural level, given the assumed *tactus* level. This paper proposes a method to evaluate the beat tracking and identify octave errors using an analysis of metrical profiles. This forms a combined feature vector of metrical profile over separate spectral subbands, described in Section 3. The behaviour of the metrical profile is analysed in

terms of quaver alternation to identify beat tracking which has performed an octave error. This approach is evaluated against an annotated dataset for beat tracking and tempo estimation as described in Section 4. The results of evaluation against datasets of recorded music are reported in Section 5.

3. METHOD

To identify the fastest pulse or tatum requires identifying the higher level rhythmic structural levels. To do so, the beat period (tactus) and metrical period (duration of the bar) is computed from the audio signal of the musical example using a beat-tracker, in this case as developed by Peeters [22]. From the nominated beat times, a metrical profile is computed.

3.1 Metrical Profile

The metrical profile, indicating the relative occurrence of events in each metrical position within the measure, has been demonstrated by [21] to represent metrical structure and matches closely with listeners judgements of metrical well-formedness. The metrical profile is computed from the likelihood of an onset at each tatum (shortest temporal interval) within a measure. The likelihood of onsets are determined from the presence of onset detection function (ODF) energy e described in [22]. The probability of an onset o_t at each tatum location t is

$$o_t = \begin{cases} \frac{\bar{e}_t}{\bar{e} + \gamma \sigma_e + \epsilon}, & o_t < 1 \\ 1 & o_t > 1 \end{cases} \quad (1)$$

where \bar{e}_t is the mean energy of the ODF over the region of the tatum t , \bar{e} and σ_e are the mean and standard deviation of the entire ODF energy respectively, ϵ is a small value to guard against zero \bar{e} , and γ is a free parameter determining the maximum number of standard deviations above the mean to assure an onset has occurred. By informal testing, $\gamma = 2$. The onset likelihoods are then used to create an histogram m_t , for $t = 1, \dots, n$, of the relative amplitude and occurrence at each tatum, by averaging each o_t across all M measures

$$m_t = \frac{\sum_{\mu=0}^M o_{t+n\mu}}{M}. \quad (2)$$

To normalise for varying tempo across each piece and between pieces, the duration of each measure is derived from the beat-tracker [22]. Using the beat locations identified by the beat-tracker, each beat duration is uniformly subdivided into 1/64th notes (hemi-demi-semiquavers), that is $0 < t < 64$ for a measure of a semibreve (whole note) duration. Such a high subdivision attempts to categorise swing timing occurring within the measure and to provide sufficient resolution for accurate comparisons of metrical structure. Using the tatum duration set to equal subdivisions of each beat duration does not capture expressive timing occurring within that time period. However, the error produced from this is minimal since the expressive timing which modifies each beat and measure period is respected.

Channel c	Low band ω_c (Hz)	High band ω'_c (Hz)
1	60	106
2	106	186
3	186	327
4	327	575
5	575	1012
6	1012	1781
7	1781	3133
8	3133	5512

Table 1. Sub-band channel frequency ranges used to calculate local spectrum onset detection functions in Equation 3.

The effect of this error is to blur the peak of each tatum onset. The metrical profile is then downsampled (by local averaging of 4 tatums) to semiquavers (1/16 notes).

3.2 Spectral Sub-band Profiles

Listeners categorise sounds using their individual spectral character, and the identification of their reoccurrence aids rhythmic organisation. To distinguish the possibly competing timing of different instruments and in order to match categorization used by listeners, metrical profiles are separated by spectral energy. This is produced by computing spectral sub-bands of the half wave rectified spectral energy. The sub-bands are computed by summing over non-overlapping frequencies:

$$F_{c,t} = \sum_{b=b_c}^{b'_c} e_{HWR}(\omega_b, t), \quad (3)$$

where $F_{c,t}$ is the spectral flux for the sub-band channel $c = 1, \dots, C$ at time t , over the spectral bands $b = [\omega_b, \omega'_b]$ of the half-wave rectified spectral energy $e_{HWR}(\omega_b, t)$ at frequency band ω_b computed as described by [22]. The sub-band channels used are listed in Table 1 for $C = 8$. These form logarithmically spaced spectral bands that approximate different time keeping functions in many forms of music. A set of subband metrical profiles is then m_{tc} for $t = 1, 2, \dots, n, c = 1, \dots, C$.

3.3 Quaver Alternation

With the metrical profile reduced to semiquavers, a measure of the regularity of variation at the supposed quaver period can be calculated. Since the tatums at strong metrical locations are expected to vary strongly regardless of metrical level, only the variation for the *sub-beats* falling at metrically weaker locations is used. For example, in a $\frac{4}{4}$ measure, $n = 16$, metrically strong semiquavers are $r = \{1, 5, 9, 13\}$. The subbeat vector of length S is defined as $s = r \nearrow t$. Using the same example meter, $s = \{2, 3, 4, 6, 7, 8, 10, 11, 12, 14, 15, 16\}$.

The average quaver alternation q for a rhythm is the normalised first order difference of subbeat profiles m'_s

$$q = \frac{\sum_{c=1}^C \sum_{i \in s} |m'_{ic}|}{SC \max(m_s)}. \quad (4)$$

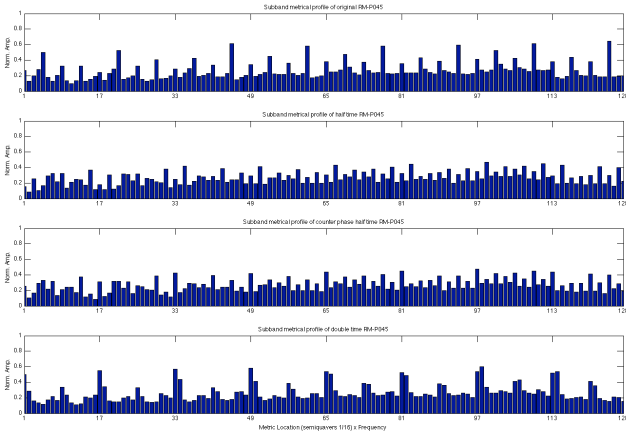


Figure 1. Metrical profiles of an example from the RWC dataset which was beat tracked with octave error. The top plot displays a metrical profile of 16 semiquavers per measure for each of the spectral subbands ($c = 1, \dots, 8$). The second, third and fourth plots displays the subband metrical profiles created for half time, half time counterphase and double time interpretations respectively.

A low quaver alternation measure indicates that variation between adjacent sub-beat semiquavers is low. This is most likely either in the case that there is little activity in the music, or the structural level chosen as the quaver is incorrect, i.e an octave error has occurred. To identify the case of an octave error, the quaver alternation of the metrical profile of a track is compared to metrical profiles of the same track formed from half and double the number of beats. The half tempo profile \hat{q} is formed from simply skipping every second beat identified by the beat tracker. A similar counter-phase half tempo profile \check{q} is formed by also skipping the initial beat. The double time profile \ddot{q} is formed from sampling at onsets o_t linearly bisecting each original inter-beat interval.

Comparisons between metrical profiles of an example rhythm is shown in Figure 1. The metrical pattern is displayed on the top plot, with $n = 16$ tatum per measure, the $C = 8$ subband profiles arranged adjacent in increasing frequency band. On the lower plots, the patterns created by assuming half tempo, half tempo counterphase, and double tempo are displayed. It can be seen that the alternation which occurs on the half tempo and half tempo counter-phase plots is more regular than the original metrical pattern or the double time pattern. This indicates that for this example, an octave error has occurred.

A measure of octave error e is computed by comparing the ratio of the half tempo quaver alternation to original quaver alternation and the ratio of double tempo to original quaver alternation,

$$e = \frac{\hat{q} + \check{q}}{2q} + \frac{\ddot{q}}{q}. \quad (5)$$

Equation 5 represents the degree that the alternation at the half or double time exceeds the original quaver alternation. Values of $\frac{\hat{q} + \check{q}}{2q} > 1$ or $\frac{\ddot{q}}{q} > 1$ indicates there is an octave error from either the double or half quaver alternation being greater, but in practice, the threshold $e > e'$ needs to be higher. The threshold was determined exper-

imentally as half a standard deviation above \bar{e} as derived from the RWC dataset at $e' = 3.34$.

3.4 Reestimation of Tempo

The beat tracking for each piece which was nominated by the algorithm as being an octave error is then recomputed with the prior tempo estimate set to half the tempo first computed. In the case of the Viterbi decoding of the beat tracker used [22], this prior tempo estimate weights the likely path of meter and tempo selection towards the half rate. In this case, even if the prior tempo is set at half, it is not guaranteed to be chosen as half the rate, if the original tempo is a more likely path which outweighs the new reestimation. This makes the beat tracker robust to false positive classifications from the beat critic.

4. EVALUATION

Two evaluation strategies for octave errors are possible: 1) evaluation of beat tracking, where the phase of the beat tracking is correct, but the beat frequency is twice the true rate and 2) evaluation of tempo alone, where the beat frequency is twice the true rate and the phase of the beat tracking is not assessed. These two evaluations meet different needs, the former if beat tracking accuracy is required, the latter if a correct median tempo measure is sufficient.

To evaluate the discrimination of the algorithm, the commonly used RWC dataset was used [11]. This dataset consists of 328 tracks in 5 sets (Classical, Jazz, Popular, “Genre” and “Royalty Free”) annotated for beat times. A subset of 284 tracks was produced by eliminating pieces whose annotations were incorrect or incomplete in the RWC dataset.²

Since the algorithm evaluates metrical profiles, this requires meter changes to be accurately identified by the beat tracker, which currently lacks that capability. Therefore pieces with changing meters are expected to reduce the performance of the algorithm. However since this would have reduced the dataset further, and added beats or time signature changes are common in many genres of music, the dataset was used with these potential noise sources.

To evaluate octave error detection independent of the quality of the beat tracking, pieces which were incorrectly beat tracked were eliminated from the test set. This was defined as a beat tracking F-score below 0.5 using a temporal window of each annotated beat position within 15% of each inter-beat interval [5, 26]. A ground truth set of octave error examples was produced by comparing the ratio of the beat tracking recall R to precision P measures, with:

$$\hat{e} = \lfloor R/P + 0.5 \rfloor, \quad (6)$$

where $\hat{e} = 2$ indicates an octave error. These ground truth candidates were then manually auditioned to verify that they were truly octave errors.

This produced a resulting dataset of 195 pieces, termed “Good”, with 46 pieces identified as actually being beat tracked at double time (an octave error). This formed the

² For several of the Jazz examples and the Genre examples, only the minim (half note) level was annotated.

Dataset	C.	True	S.	Prec.	Rec.	F
Good	30	46	55	0.545	0.652	0.594
Full	29	46	82	0.354	0.630	0.453

Table 2. Results of octave error detection by metrical profile analysis (beat critic). “C.” indicates the number of tracks correctly identified as an octave error, “True” as the ground truth number of octave errors manually identified. “S.” indicates the number of tracks selected as being an octave error. “Prec.,” “Rec.” and “F” indicates the precision, recall and F-score measures respectively.

Dataset	Meth.	Size	Pre-Reest.		Post-Reest.		%
			OE	NE	OE	NE	
Good	BT	195	46		20		43
Good	BPM	195	44	10	24	12	54
Full	BT	284	63		37		58
Full	BPM	284	57	42	38	46	66

Table 3. Number of tracks with beat tracking octave errors (OE) before (Pre) and after (Post) reestimation using the beat critic. The column labelled “%” indicates the reduction in octave errors. NE columns indicates non-octave errors.

ground truth to evaluate the octave error identification algorithm. From these, standard precision, recall and F-score measures can be computed [26]. The entire set of 284 pieces (termed “Full”) was also used to evaluate performance when beat tracking does not perform optimally.

To determine the improvement the beat critic makes to beat tracking, pieces which were determined to be beat tracked with octave error were recomputed with half the prior tempo. This would occur for false as well as true positives. The beat tracker would then use the new weighting towards the half tempo, but could produce the same result as the original beat tracking if the Viterbi decoding still biased towards the original tempo estimate [22].

The Good and Full datasets were also assessed for their fidelity to the annotated median tempo measurement τ of each track. This was computed as $\tau = 60/\tilde{i}$, where \tilde{i} is the median inter-beat interval in seconds. A beat tracked tempo which was within 3% of the annotated tempo was deemed a successful tempo estimation.

5. RESULTS

The results of evaluating the beat critic with the Good and Full RWC datasets appear in Table 2. On the “Good” dataset, while the critic is able to identify 65% of the pieces with octave errors (the recall), it produces a sizeable number of false positives (the precision) which reduces the F-score. As to be expected, with the “Full” dataset, the performance is worse. The substantially higher number of false positives for this dataset indicate that the octave error measure is sensitive to beat tracking error. As the algorithm is defined, the measure of sub-beat alternation is

probably too reliant on the expectation that the beat is correctly tracked.

Despite the relatively low scoring results, Table 3 indicates the success of the beat critic when used to reestimate the beat tracker. The column “Meth.” describes the method of evaluation, either “BT” for beat tracking, comparing each beat location against annotated beats, or “BPM”, comparing estimated tempo against annotated tempo. “Size” describes the number of tracks in the dataset. “OE” indicates the number of tracks that were beat tracked that are evaluated to have been an octave error. “Pre” and “Post” indicates the number of tracks before and after reestimating using the beat critic to bias prior tempo of the beat tracker. “NE” indicates the number of tracks that were not beat tracked correctly but were not octave errors. While it is possible to identify non-octave errors with BPM evaluation within a perceptually meaningful tolerance (3%, see Section 4), this can not be defined properly when the measure of beat tracking is calculated in terms of precision, recall and F-score.

In the case of the BT evaluation, the number of octave errors were reduced to 43% and 58% of the former number of errors for the Good and Full datasets respectively. This indicates that the Viterbi decoding of the beat tracker has benefitted from reestimation and is reasonably robust to the false positives identified as octave errors. The tempo evaluation also showed similar improvements, reducing octave errors to 54% and 66% (Good and Full). The slight increase in non-octave errors after reestimation indicates cases where the false positives have lead to mistracking. Depending on the application, this may be an unacceptable deterioration in performance despite an increase in the overall number of correctly tracked pieces.

6. CONCLUSIONS

A method for the detection of octave errors in beat tracking has been proposed and evaluated. The approach was evaluated with an audio dataset that represents a variety of genres of music. This approach, while currently applied to only one beat tracker, depends only on the presence of a mid-level representation, and the determination of beat and meter periods, commonly produced by many beat trackers. It is applicable to beat trackers which benefit from reestimation or convergence in the selection of the beat tracking frequency.

While the performance of the beat critic is well below perfection, when applied to a beat tracker, it has been shown to improve overall performance, reducing the number of octave errors, at the cost of a slight increase in mistracking. The beat critic’s applicability and usefulness is ultimately dependent on the cost of false positives.

A number of improvements are possible. The use of a threshold for the octave error classification is simplistic and possibly difficult to set accurately. A machine learning classifier promises to perform better in this task. However, the best features to be used are not yet clear, preliminary experiments with the quaver alternation measures q , \dot{q} , \ddot{q} and \ddot{q} indicate that these are insufficient features to dis-

criminate the octave error classification. The alternative, using the entire profiles, or reductions thereof, as features produces too high a dimensionality for accurate learning. Another issue is the relative computational cost of such an approach, when the current threshold approach is computationally low. In principle the approach could be used to identify beat tracking at half the correct rate, although such beat tracking errors did not occur using the dataset and therefore have not been evaluated.

The beat critic exploits knowledge of rhythmic behaviour as represented in musicologically based models of metrical profiles to compare temporal levels. The comparison of the relative activity of levels is used to identify octave errors. By examining the behaviour of events in the time domain, the goal has been to circumvent limitations in the temporal resolution of frequency based analysis in the identification of beat levels.

7. ACKNOWLEDGEMENTS

This research was supported by the French project Oseo “Quaero”. Thanks are due to Geoffroy Peeters for provision of the beat-tracker and onset detection code.

8. REFERENCES

- [1] Jeffrey A. Bilmes. Timing is of the essence: Perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm. Master’s thesis, Massachusetts Institute of Technology, September 1993.
- [2] Judith C. Brown. Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4):1953–7, 1993.
- [3] Ali Taylan Cemgil and Bert Kappen. Monte Carlo methods for tempo tracking and rhythm quantization. *Journal of Artificial Intelligence Research*, 18:45–81, 2003.
- [4] Martin Coath, Susan Denham, Leigh M. Smith, Henkjan Honing, Amaury Hazan, Piotr Holonowicz, and Hendrik Purwins. Model cortical responses for the detection of perceptual onsets and beat tracking in singing. *Connection Science*, 21(2):193–205, 2009.
- [5] Matthew E. P. Davies and Mark D. Plumbley. A spectral difference approach to downbeat extraction in musical audio. In *EUSIPCO*, 2006.
- [6] Matthew E. P. Davies and Mark D. Plumbley. Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1009–20, 2007.
- [7] Peter Desain and Henkjan Honing. Foot-tapping: A brief introduction to beat induction. In *Proceedings of the International Computer Music Conference*, pages 78–9. International Computer Music Association, 1994.
- [8] Simon Dixon. Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research*, 36(1):39–50, 2007.
- [9] Douglas Eck. Beat induction with an autocorrelation phase matrix. In M. Baroni, A. R. Addessi, R. Caterina, and M. Costa, editors, *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC)*, page 931, Bologna, Italy, 2006. SMPC and ESCOM.
- [10] Paul Fraise. Rhythm and tempo. In Diana Deutsch, editor, *The Psychology of Music*, pages 149–80. Academic Press, New York, 1st edition, 1982.
- [11] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC music database: Popular, Classical, and Jazz music databases. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 287–288, October 2002.
- [12] Mari Riess Jones. Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review*, 83(5):323–55, 1976.
- [13] Mari Riess Jones. Musical time. In *Oxford Handbook of Music Psychology*, pages 81–92. Oxford University Press, 2009.
- [14] Mari Riess Jones and Marilyn Boltz. Dynamic attending and responses to time. *Psychological Review*, 96(3):459–91, 1989.
- [15] Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech and Language Processing*, 14(1):342–55, 2006.
- [16] James Koetting. What do we know about African rhythm? *Ethnomusicology*, 30(1):58–63, 1986.
- [17] Edward W. Large and John F. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6(2+3):177–208, 1994.
- [18] Justin London. *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press, 2004.
- [19] H. Christopher Longuet-Higgins and Christopher S. Lee. The perception of musical rhythms. *Perception*, 11:115–28, 1982.
- [20] James G. Martin. Rhythmic (hierarchical) versus serial structure in speech and other behaviour. *Psychological Review*, 79(6):487–509, 1972.
- [21] Caroline Palmer and Carol L. Krumhansl. Mental representations for musical meter. *Journal of Experimental Psychology - Human Perception and Performance*, 16(4):728–41, 1990.
- [22] Geoffroy Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Advances in Signal Processing*, (67215):14 pages, 2007. doi:10.1155/2007/67215.

- [23] Eric D. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1):588–601, 1998.
- [24] Uwe Seifert, Fabian Olk, and Albrecht Schneider. On rhythm perception: Theoretical issues, empirical findings. *Journal of New Music Research*, 24(2):164–95, 1995.
- [25] Leon van Noorden and Dirk Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66, 1999.
- [26] C. V. van Rijsbergen. *Information Retrieval*. Butterworth, London; Boston, 2nd edition, 1979.
- [27] Maury Yeston. *The Stratification of Musical Rhythm*. Yale University Press, New Haven, 1976. 155p.