DRAFT: FINDING THE METRICAL GRID IN MUSICAL SIGNALS

Jarno Seppänen

Tampere University of Technology
Signal Processing Laboratory
P.O.Box 553, FIN-33101 Tampere, Finland
jarno.seppanen@cs.tut.fi

ABSTRACT

An ad hoc algorithm for estimating the quantization step and the metrical grid based on inter-onset interval data is presented.

The metrical grid analysis is derived from and is a variant of the greatest common divisor (GCD). The algorithm consumes sound onset data and produces the metrical grid.

1. INTRODUCTION

Beat and meter induction has been conventionally based directly on sound onset data derived from either MIDI data or an acoustical signal [2], [3], [4], [5], [1]. However, as the principles between beat sensation are somewhat vague and the perceived beat can even vary between listeners, it is not easy to build a system to assign beat positions to a stream of sound onsets.

On the other hand, it is known that there are a number of shorter and longer pulse sensations present in addition to the beat. In this paper we will concentrate on the temporally shortest and perceptually lowest-level pulse, which we call the *metrical grid*. The other, temporally longer pulse sensations are all integral multiples of the shortest pulse.

We will call the duration of the shortest pulse the *quantum*.[1] The quantum is equal to the step size of the metrical grid. There has been also other names for the step size in the literature, such as the *clock* and the *tatum* [6].

Beat induction based on quantum-level representation is feasible due to two reasons: firstly, the estimation of the quantum duration is practicable since the quantum corresponds approximately to the *greatest common divisor* (GCD) of the inter-onset intervals (IOI). Secondly, the beat pulse period is equal to an integral multiple of the quantum, and thus beat induction reduces to choosing the correct integer.

In addition to beat induction, the metrical grid can be used for sound segmentation purposes in various musical signal analysis tasks.

Another very specific application field would be musically synchronized effects such as a tremolo.

2. METHOD

2.1. Quantum duration estimation

The quantum duration estimation is carried out using the inter-onset intervals (IOI) of the music as input. The IOI's are computed from the onset data detected from an acoustic music waveform. The calculation is done on a 500 ms frame-by-frame basis, and we will present the calculations done within frames here.[2]

2.1.1. Inter-onset interval computation

The quantum estimation algorithm gets a stream onset data as input and it processes the onsets in frames. However, in order to use the data, the onset stream is transformed into *inter-onset interval* (IOI) data.

The IOI's are not only computed between successive onsets but all onset pairs, whose IOI's are within a preset bound, are taken into account. The IOI computation process travels the onset stream forward one

[1] *quantization step* of temporal quantization

[2] Although the 500 ms frame duration is handled as a constant throughout the text, other values may well be substituted.

onset at a time and computes the IOI's to past onsets within the IOI bound.

2.1.2. *Greatest common divisor approximation*

The quantum corresponds approximately to the greatest common divisor (GCD) of the IOI values. In fact the quantum *is* the GCD of all the IOI's *iff* the IOI's are all integral multiples of the quantum, i.e. there is no deviation from the perfect values. Due to the fact that there always is deviation, we must use a scheme to estimate the most prominent GCD candidate from the IOI values.

The GCD of nonnegative integers a_i, $i \in I$, $I = \{1, 2, \ldots, n\}$ is defined to be the the largest positive integer which divides all a_i exactly, i.e.

$$\begin{aligned} & \gcd(a_1, a_2, \ldots, a_n) \\ = \; & \max\{r : \forall i \in I : a_i \bmod r = 0\} \end{aligned} \qquad (1)$$

When applied to the quantum estimation problem, the integers a_i would correspond to IOI values and their GCD would be the quantum.

In order to apply the GCD concept to actual IOI values, the integer input constraint has to be dropped. Due to random deviations, we also cannot require that the quantum divides all the IOI values *exactly* ($a_i \bmod r = 0$); we want to find out a quantum that divides the IOI values sufficiently well.

We shall define an error function whose local minima represent possible GCD candidates and which will go to zero if an exact GCD is found. Let q be the variable quantum duration [ms] and o_i be the inter-onset intervals (IOI) [ms] present in the music. We shall then define a least-square *remainder error function* $e(q)$ as follows:

$$e(q) = \sum_i \left[(o_i + q/2) \bmod q - q/2\right]^2 . \qquad (2)$$

The remainder error function (2) is based on the form $\sum_i (o_i \bmod q)^2$ and has been slightly modified to make the function behaviour smooth around the GCD candidates. If a GCD exists, eq. (1) will hold and the remainder error function will go to zero at the GCD. Note that the modification does not change the place of the local minimums.

Using equation (2), the GCD can be found by finding the greatest value for which the remainder error function is zero, or

$$\gcd(o_1, o_2, \ldots, o_n) = \max\{q : e(q) = 0\} \qquad (3)$$

2.1.3. *Inter-onset interval histogram*

In order for the algorithm to facilitate for quantum changes (e.g. accelerandos and ritardandos), the processing is carried out in 500 ms frames. However, the algorithm cannot base its analysis on the IOI's of onsets falling within only one 500 ms frame due to lack of information. The IOI's are therefore converted into a histogram representation and the histogram is accumulated from frame to frame. Figure 1 illustrates example histograms of the music samples in table 1, extracted in the middle of processing.

The histogram bin width is equal to the reciprocal of f_s, a histogram data 'sample rate'. f_s is used to determine the resolution of the IOI's stored in the histogram. The size of the histogram M is determined by the IOI bound mentioned in section 2.1.1.

Let $h[k]$, $k \in K$, $K = \{0, 1, \ldots, M-1\}$ represent the contents of the M-bin histogram and $h_x[k]$ represent the histogram bin centers [ms]:

$$h_x[k] = \frac{k}{f_s} \qquad (4)$$

Before accumulation in the histogram, the IOI's of each onset within a frame o_i must be discretized according to f_s. The histogram contribution of the onsets within one frame are gathered to a fill function $f[k]$, $k \in K$:

$$f[k] = n\left(\left\{i : |o_i - h_x[k]| \le \frac{1}{f_s}\right\}\right) \qquad (5)$$

After discretization the new histogram $h'[k]$ is computed by adding the fill function $f[k]$ and the past histogram $h[k]$. Weighting coefficients c_f and c_l are used for the fill function and the past histogram, respectively, to implement a leaky integrator:

$$h'[k] = c_l h[k] + c_f f[k], \qquad (6)$$

where the leak and fill coefficients are

$$c_l = (1/2)^{\frac{500\,\mathrm{ms}}{t_{1/2}}} \qquad (7)$$

$$c_f = \frac{1 - c_l}{c_l} \qquad (8)$$

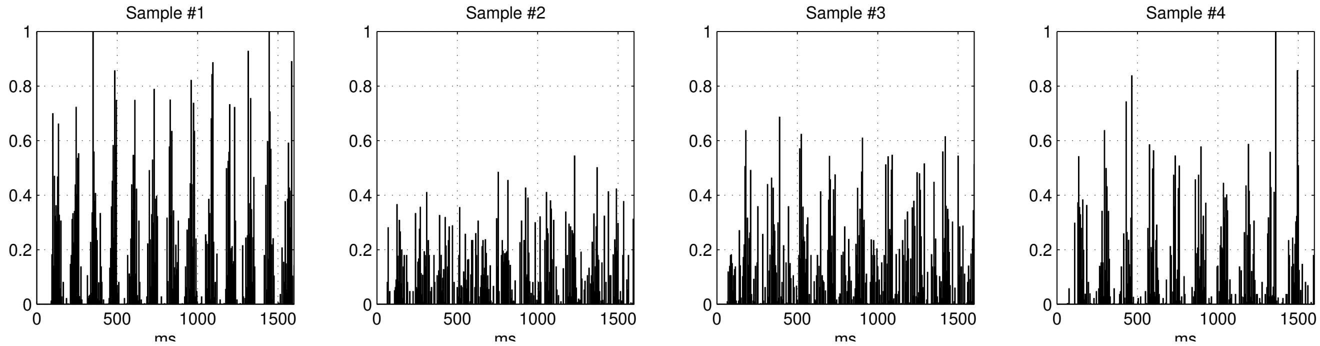

Figure 1: Accumulated histograms of the example signals at $t = 10$ s, with $f_s = 200$ Hz and $M = 321$.

and $t_{1/2}$ is the half-life [ms] of the decay of the histogram content. With this kind of weighting coefficients the histogram values don't continue to grow from frame to frame but will represent a weighted average of the past IOI data.

The remainder error function (2) cannot be used with the histogram directly and therefore we need to derive a formula for the remained error based on the histogram content. Let $d(o,q) = (o + q/2) \bmod q - q/2$. Now equation (2) becomes

$$\begin{aligned} e(q) &= \sum_i \left[d(o_i,q)\right]^2 \\ &= \underbrace{\sum_i \left[d(o_i,q)\right]^2}_{o_i=h_x[0]} + \underbrace{\sum_i \left[d(o_i,q)\right]^2}_{o_i=h_x[1]} + \cdots \\ &\quad + \underbrace{\sum_i \left[d(o_i,q)\right]^2}_{o_i=h_x[M-1]} \\ &= h[0]\left[d\left(h_x[0],q\right)\right]^2 + h[1]\left[d\left(h_x[1],q\right)\right]^2 + \cdots \\ &\quad + h[M-1]\left[d\left(h_x[M-1],q\right)\right]^2 \\ &= \sum_{k=0}^{M-1} h[k]\left[d\left(h_x[k],q\right)\right]^2 \\ &= \sum_{k=0}^{M-1} h[k]\left[(h_x[k]+q/2) \bmod q - q/2\right]^2 \end{aligned} \quad (9)$$

The form (9) could be used in computing the approximate GCD, but we will introduce a version of eq. (9) normalized with respect to histogram mass:

$$\hat{e}(q) = \frac{\sum_{k=0}^{M-1} h[k]\left[(h_x[k]+q/2) \bmod q - q/2\right]^2}{\sum_{k=0}^{M-1} h[k]} \quad (10)$$

Figure 2 illustrates the remainder error functions computed with equation (10) from the histograms drawn in figure 1.

2.1.4. Remainder error thresholding

After the computation of the remainder error function $e(q)$, the quantum value must be chosen. According to the definition of the GCD (eq. 1), the quantum would have to be the highest local minimum of the remainder error function (fig. 2).

We have decided to choose the quantum to be the highest local minimum below a certain threshold. The threshold is computed from the minimum and the median values of the remainder error function according to a coefficient α; figure 2 shows both the medians (dash-dot line) and the thresholds (dashed line) in addition to the remainder error functions (solid line).

The remainder error function threshold e_{th} is computed as follows:

$$e_{th} = \alpha \min_q e(q) + (1-\alpha)\mathrm{median}_q e(q), \quad (11)$$

where the α parameter has been set to 0.4 in the example cases in figure 2.

After computing the threshold, the highest local

Table 1: Sound samples used.

N:o	Artist	Song title	Duration [s]	Description
1	Nick Holder	Da Sambafrique	15.0	House music; instrumental; machine timing
2	The Headhunters	Tip Toe	18.6	Funk music
3	Marisa Monte	Segue o Seco	30.7	Live performance; w/ ritardando
4	The Guitar Trio	Cardeosa	15.0	3 acoustic guitars; instrumental

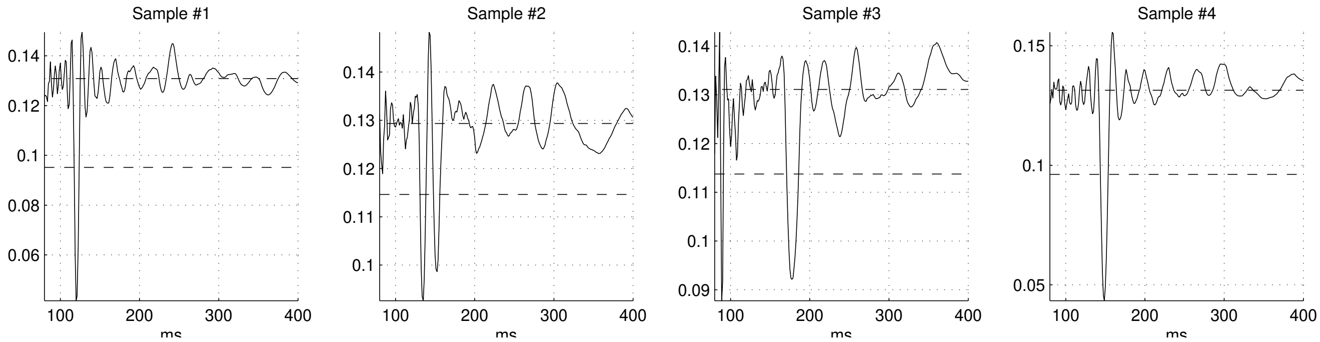

Figure 2: Remainder error functions of the example signals at $t = 10$ s. The medians and the thresholds of the remainder error functions are also shown.

minimum[3] below the threshold is picked as the quantum q:

$$q = \max_{p} \{p : e(p) < e_{th} \wedge e(p) \text{ local minimum}\} \tag{12}$$

2.2. Quantum phase estimation

The quantum phase, i.e. the exact points of the metrical grid, is decided after the quantum duration has been computed. The quantum positions are based on the quantum duration but are incrementally adjusted towards onsets. The metrical grid points within each frame are spaced by the quantum duration q with some adjustments. The adjustments are done towards onsets, since by assumption all the observed onsets are aligned on the metrical grid.

Initially, when processing starts, the quantum phase is probably out of alignment with respect to the onsets. The metrical grid will gradually become aligned with the onsets as the adjustments cumulate, controlled with a coefficient β.

The quantum phase — the points of the metrical grid — are indicated by φ [ms]. At the beginning of a frame φ is initialized to the offset of the first quantum according to quantum 'carry' variable γ:

$$\varphi = \gamma q. \tag{13}$$

The γ variable is used to carry quantum overlap information between frames.

Unless there are onsets, there is no adjustment done and the quantum positions are spaced by the quantum duration q; the unadjusted quantum position is therefore

$$\hat{\varphi}' = \varphi + q. \tag{14}$$

When there are onsets, the quantums are positioned at each adjusted quantum position φ'; the adjustment is done according to each onset:

$$\varphi' = \begin{cases} \hat{\varphi}' + \beta(o_i - \hat{\varphi}') & \text{if } \exists i : \hat{\varphi}' - \frac{q}{2} \leq o_i \leq \hat{\varphi}' + \frac{q}{2} \\ \hat{\varphi}' & \text{otherwise} \end{cases} \tag{15}$$

The equations (14) and (15) are applied repetitively, placing a grid point at every φ' and substituting $\varphi = \varphi'$ between repetitions, while $\varphi \leq 500$ ms. The metrical grid points are assigned this way.

After deciding the points of the metrical grid within the frame, the carry variable γ is updated in anticipa-

[3] In the actual implementation $e(q)$ is a discrete function and thus the local minima are straightforward to locate.

tion for the forthcoming frame:

$$\gamma = 1 - \frac{500\ \mathrm{ms} - \varphi}{q} \tag{16}$$

3. ANALYSIS

Due to the relatively long frame duration (500 ms) and the use of even longer half-life (1300 ms) in the histogram accumulation, the quantum duration output is delayed when compared with the actual onsets and the music signal. The delay can be well heard when comparing the produced metrical grid with the music signal if there are tempo changes etc. in the material. There is probably space for reducing the latency of the system by carefully tuning the timing parameters.

4. CONCLUSIONS

An ad hoc algorithm for estimating the quantization step and the metrical grid based on inter-onset interval data was presented.

4.1. Future issues

4.1.1. Beat estimation

After the quantum analysis has been performed, the beat induction problem reduces to selecting a set of points from the metrical grid.

It is known that the beat sensation is most salient within periods ranging from 400 ms to 900 ms [7]. Once the quantum duration is known, it is easy to select the set of possible integer ratios between beat duration and quantum duration.

The beat positions can be estimated by tracking how the onsets are distributed on the metrical grid. Beats can be tracked by accumulating the number of onsets on each metrical grid point, with different integer periods, and then selecting the period with most onsets on average.

4.1.2. Statistical analysis

Based on assumptions such as that the IOI deviations are Gaussian [8, p. 45] and that the IOI's are all integral multiples of the quantum, an accurate/optimal formula for the quantum could be derived.

4.1.3. Global post-processing

After the analysis based on the remainder error function there could be a global rule-based post-processing stage which would locate and eliminate erroneous quantum estimates.

5. REFERENCES

[1] Petri Toiviainen, "Modelling the perception of metre with competing subharmonic oscillators," .

[2] Judith C. Brown, "Determination of the meter of musical scores by autocorrelation," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 1953–1957, 1993.

[3] Masataka Goto and Yoichi Muraoka, "Beat tracking based on multiple-agent architecture — a real-time beat tracking system for audio signals," in *Proc. 2nd Int. Conf. Multiagent Sys.*, 1996, pp. 103–110.

[4] Leigh M. Smith, "Modelling rhythm perception by continuous time-frequency analysis," in *Proc. ICMC*, Hong Kong, 1996.

[5] Eric D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 588–601, Jan. 1998.

[6] 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, Sept. 1993.

[7] R. Parncutt, "A perceptual model of pulse salience and metrical accent in musical rhythms," *Music Perception*, vol. 11, no. 4, pp. 409–464, 1994.

[8] Eric D. Scheirer, "Extracting expressive performance information from recorded music," Master's thesis, Massachusetts Institute of Technology, Sept. 1995.