BEAT AND METER EXTRACTION USING GAUSSIFIED ONSETS

Klaus Frieler
University of Hamburg
Department of Systematic Musicology
kgf@omniversum.de

ABSTRACT

Rhythm, beat and meter are key concepts of music in general. Many efforts had been made in the last years to automatically extract beat and meter from a piece of music given either in audio or symbolical representation (see e.g. [11] for an overview). In this paper we propose a new method for extracting beat, meter and phase information from a list of unquantized onset times. The procedure relies on a novel method called 'Gaussification' and adopts correlation techniques combined with findings from music psychology for parameter settings.

1. INTRODUCTION

The search for methods and algorithms for extracting beat and meter information from music has several motivations. First of all, one might want to explain rhythm perception or production in a cognitive model. Most of classical western, modern popular and folk music can be described as organized around a regularly sequence of beats, this is of utmost importance for understanding the cognitive and productive dimensions of music in general. Second, meter and tempo information are important meta data, which could be useful in many applications of music information retrieval. Third, for some tasks related to production or reproduction such information could also be helpful, e.g., for a DJ who wants to mix different tracks in a temporal coherent way or for a hip-hop producer, who wants to adjust music samples to a song or vice versa.

In this paper we describe a new method, which takes a list of onset times as input, which might come from MIDI-data or from some kind of onset detection system for audio data. The list of onsets is turned into a integrable function, the so-called Gaussification, and the autocorrelation of this Gaussification is calculated. From the peaks of the autocorrelation function time base (smallest unit), beat (tactus) and meter are inferred with the help of findings from music psychology. Then the best fitting meter and phase are estimated using cross-correlation of prototypical meters, which resembles a kind of matching algorithm. We evaluated the system with MIDI-based data, either quantized with added temporal noise or played by an amateur keyboard player, showing promising results, especially in the processing of temporal instabilities.

2. MATHEMATICAL FRAMEWORK

The concept of Gaussification was developed in the context of extending autocorrelation methods from quantized rhythms to unquantized ones ([1], [4]). The idea behind is that any produced or perceived onset can be viewed as an imperfect rendition (or perception) of a point on a perfect temporal grid. A similar idea was used by Toiviainen & Snyder [11], who assumed a normal distribution of measured tappings time for analysis purposes. However, the method presented here was developed independently, and the aims are quite different. Though a normal distribution is a natural choice it is not the only possible one, and the Gaussification fit into the more general concept of functionalisation.

Definition 1 (Functionalisation) *Let $\mathcal{R} = \{t_i\}_{1 \le i \le N}$ be a set of time points (a* ***rhythm****) and $\{\psi_i\}_{1 \le i \le N}$ a set of (real) coefficients Moreover, let f be a L^2-integrable function:*

$$\int_{-\infty}^{\infty} f(t)dt < \infty$$

Then

$$f_{\mathcal{R}}(t) = \sum_{i=1}^{N} \psi_i f(t - t_i) \tag{1}$$

is called a ***functionalisation*** *of $\mathcal{R}$.*

We denote by $g(t; \mu, \sigma)$ the gaussian kernel, i.e.,

$$g(t; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \tag{2}$$

Then

$$\begin{aligned} G_{\mathcal{R}}(t) &= \sum_{i=1}^{N} \psi_i g(t; t_i, \sigma) && (3) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{i=1}^{N} \psi_i e^{-\frac{(t-t_i)^2}{2\sigma^2}} && (4) \end{aligned}$$

is called a ***Gaussification*** *of $\mathcal{R}$.*

A Gaussification is basically a linear combination of gaussians centered at the points of $\mathcal{R}$. The advantage of a functionalisation is that the transformation of a discrete set into a integrable (or even continous and differentiable) function, so that correlation and similar techniques are applicable. An additional advantage of Gaussification is that the various correlation functions can be easily integrated out. One has

Proposition 1 *Let $T_\tau : G(t) \to G(t+\tau)$ be the time translation operator. Then the time-shifted scalarproduct of two gaussfications G_1, G_2 is the cross-correlation function $C_{G_1G_2}$:*

$$\begin{aligned} C_{G_1G_2}(\tau) &:= \langle G_1, T_\tau G_2 \rangle \qquad (5) \\ &= \frac{1}{2\sigma\sqrt{\pi}} \sum_{i,j=1}^{N_1,N_2} \psi_i \phi_j g(\tau; \Delta t_{ij}^{12}, \sqrt{2}\sigma) \end{aligned}$$

with $\Delta t_{ij}^{12} = t_i^{(1)} - t_j^{(2)}$.
The autocorrelation function $A_G(\tau)$ of a Gaussification G is given by:

$$\begin{aligned} A_G(\tau) &:= \langle G, T_\tau G \rangle \qquad (6) \\ &= \frac{1}{2\sigma\sqrt{\pi}} \sum_{i,j=1}^{N} \psi_i \psi_j g(\tau; \Delta t_{ij}^{12}, \sqrt{2}\sigma) \end{aligned}$$

The next thing we need is the notion of a temporal grid.

Definition 2 (Temporal grid) *Let $\Delta T > 0$ be a real positive constant, the timebase. Then the set*

$$G_{\Delta T} = \{n\Delta T, n \in \mathbb{N}_0\} \qquad (7)$$

*is called a **temporal grid**. For $k < p \in \mathbb{N}_0$ the (k,p)-subgrid of G is the set*

$$G_{\Delta T}(k,p) = \{(k+np)\Delta T, n \in \mathbb{N}\} \qquad (8)$$

with phase k and period p. The value

$$\Theta = \frac{1}{m\Delta T} \qquad (9)$$

*is called the **tempo** of the (sub)grid. Any subset $\mathcal{R} \subset G_{\Delta T}$ of a temporal grid is called a **regular rhythm**.*

It is now convenient to define the notion of a metrical hierarchy.

Definition 3 (Metrical hierarchy) *Let $G_{\Delta T}$ be a temporal grid, $M = \{m_i, 1 < m_1 < m_2 < \cdots < m_N\}$ a set of ordered natural numbers and $k < m_1$ a fixed phase. The subgrid $G_{\Delta T}(k, p_n) \equiv \Gamma(k,n)$ with $p_n = \prod_{i=1}^n m_i, n \leq N$ is called a subgrid of level n and phase k.*

A (regular) metrical hierarchy is then the collection of all subgrids of level $n \leq N$:

$$\mathcal{M}(k; m_1, \ldots, m_N) = \{\Gamma(k,n), 1 \leq n \leq N\} \qquad (10)$$

We are now able to state some classic problems of rhythm research.

Problem 1 (Quantization) *Let $\mathcal{R} = \{t_i\}_{1\leq i \leq N}$ be a given rhythm (w.l.o.g. $t_1 = 0$) and $\epsilon > 0$. The task of quantization is to find a time constant ΔT and a set of quantization numbers $\{n_i \in \mathbb{N}_0\}$ such, that*

$$|t_i - n_i \Delta T| < \epsilon, \forall i \qquad (11)$$

*The mapping $Q(t_i) = n_i\Delta T$ is called a **quantization** of $\mathcal{R}$.*

It is evident that a solution does not necessarily exist and is not unique. For any ΔT and any natural number k, $\Delta T_k = \frac{\Delta T}{k}$ gives a another solution. Therefore the requirement of minimal quantization, i.e. $\sum n_i = min$ should be added. Many algorithms can be found in the literature for solving the quantization problem (see [11] for an overview) and the related problems of beat and meter extraction, which can be stated as follows.

Problem 2 (Beat and meter extraction) *Let $\mathcal{R}$ be the measured onsets of a rhythm rendition. Furthermore, assume that a subject was asked to tap regularly to the rhythm, and the tapping times were measured, giving a rhythm $T(\mathcal{R})$. The task of beat extraction is to deduce a quantization of $T(\mathcal{R})$ from $\mathcal{R}$. If the subject was furthermore asked to mark a 'one', i.e. a grouping of beats, measured into another rhythm $\mathcal{M}(\mathcal{R})$ the task of meter extraction is to deduce a quantization of $\mathcal{M}$ and to find its relative position to the extracted beat.*

We will present a new approach with the aid of Gaussification. For musically reasonable applications more constraints have to be added, which naturally come from music psychological research.

3. PSYCHOLOGY OF RHYTHM

Much research, empirical and theoretical, has been done in the field of rhythm, though a general accepted definition of rhythm is still lacking. Likewise there are many different terms and definitions for the basic building blocks, like tempo, beat, pulse, tactus, meter etc. We will only assemble some well-known and widely accepted empirical facts from the literature, which serve as an input for our model. In addition we will restrict ourselves to examples from westen music which will be considered to have a significant level of beat induction capability, and can be described with the usual western concepts of an underlying isochronous beat and a regular meter.

A review of the literature on musical rhythm speaks for the fact, that there is a hierachy of time scales for musical rhythm related to physiological processes. (For a summary of the facts presented here see e.g [10] or [7] and references therein). Though music comprises a rather wide range of possible tempos, which range roughly from 60-300 bpm (200 ms - 1s), there is no general scale invariance. The limitations on either side are caused from perceptual and motorical constraints. The fusion threshold, ie, the minmal time span at which two events can be perceived as distinct lies around 5-30 ms, and order relation between events can established above 30 - 50 ms. The maximal frequency of a limb motion is reported to be around 6-12 Hz ($\Delta T =$ 80-160 ms), and the maximum time span between two consecutive events to be perceived as coherent, the so-called subjective present, is around $2 - 3$ s. Furthermore, subjects asked to tap an isochronous beat at a rate of their choice tend to tap around 120 bpm ($\Delta T = 500$ ms), the so-called spontaneous tempo ([3], [7], [12]). Likewise, the preferred tempo, i.e. the tempo where subjects feel most comfortably while tapping along to music lies around within a similar range, and is often used synonymously to spontaneous tempo.

With this facts in mind, we will now formulate an algorithm for solving the quantization task and the beat and meter extraction problem.

4. METRICAL HIERARCHY ALGORITHM

Input to our algorithm is the rhythm $\mathcal{R} = \{t_i\}_{1 \leq i \leq N}$ as measured from a musical rendition. For testing purposes we used MIDI files of single melodies from western popular music. Without loss of generality we set $t_1 = 0\ ms$.

1. Prepare a Gaussification $G(\mathcal{R})$ with coefficeints coming from temporal accent rules.
2. Calculate the autocorrelation function A_G.
3. Determine set of maxima and maxima points M of A_G
4. Find beat T_B and timebase ΔT from $M(A_G)$
5. Get a list of possible meters p_i with best phases φ_i and weights w_i with cross-correlation.

4.1. Gaussification with accents rules

The calculation of a Gaussification from a list of onsets was already describe above. We chose a value of $\sigma =$ 25 ms for all further investigations. The crucial point is the setting of the coefficients ψ_i. We will consider the values of a Gaussification as accent values, so the question is how to assign (perceptual) meaningful accents to each onset. it is known from music psychology that there are a lot of sources for perceived accents, ranging from loudness, pure temporal information along pitch clues to involved harmonical (and therefore highly cultural dependent) clues. Since we are dealing with purely temporal information, only temporal accent rules will be considered. Interestingly enough, much of the temporal accent rules ([7], [8], [9]) are not causal, which seems to be evidence for some kind of temporal integration in the human brain. For sake of simplicity we implemented only some of the simplest accent rules, related to inter-onset interval (IOI) ratios.

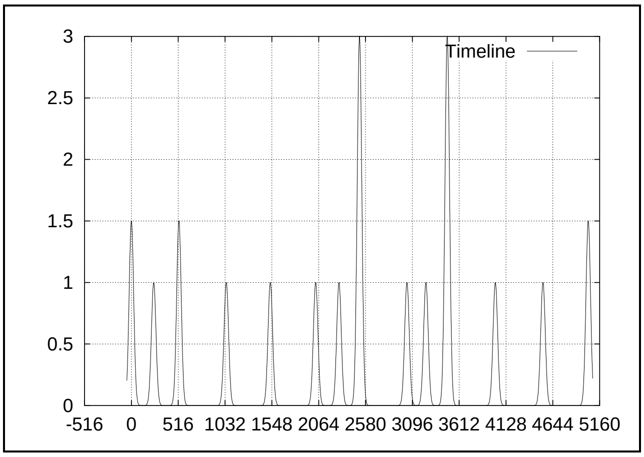

Figure 1. Example: Gaussification of the beginning of the Luxembourgian folk song 'Plauderei an der Linde', at 120 bpm with temporal noise added ($\sigma = 50\ ms$).

Let $a_{MAJ} > a_{MIN} > 1$ be two free accent parameters for major and minor accents respectively. Furthermore, we write $\Delta t_i = t_i - t_{i-1}$ for IOIs. Then the accent algorithm is given by

1. INITIALIZE
   Set $\psi_i = 1, \psi_1 = a_{MIN}, \psi_N = a_{MIN}$
2. MINOR ACCENT
   If $(\Delta t_{i+1} - 2\sigma)/\Delta t_i > 1$ then $\psi_i = a_{MIN}$
3. MAJOR ACCENT
   If $(\Delta t_{i+1} + \sigma)/\Delta t_i > 2$ then $\psi_i = a_{MAJ}$

The second rule assigns a minor accent.to every event, which following IOI is significantly longer then the preceding IOI. The third rule assigns a major accent to an event, if the following IOI is around two times as long as the preceding IOI. It seems that accent rules, even simple one like these, are inevitable for musically reasonable results. After some informal testing we used values of $a_{MAJ} = 3$ and $a_{MIN} = 2$ throughout.

4.2. Calculation of A_G and its maximum points

The calculation of the autocorrelation function is done according to equation 6. Afterwards the maxima are searched and stored for further use. We denote the set of maxima and corresponding maximum points with

$$M(A_G) = \{(t_i, y_i), y_i = A_G(t_i) = max,\ 0 \leq i < N\}$$

4.3. Determination of beat and time-base

4.3.1. Determination of the beat

It is a widely observed fact that the 'beat'-level in a musical performance is the most stable one. First, we weight the autocorrelation with a tempo preference function, and then choose the point of the highest peaks to be the beat

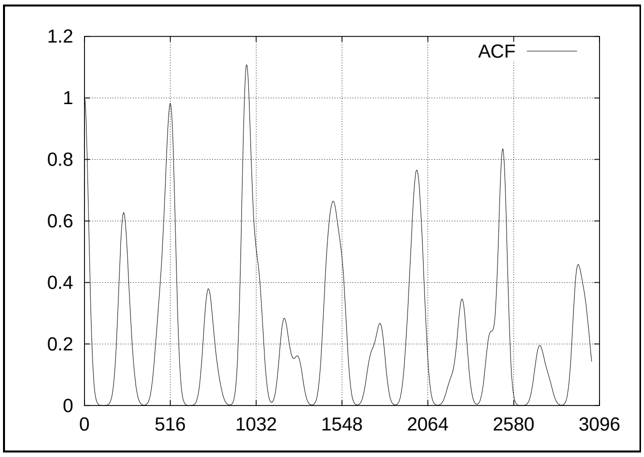

Figure 2. Example: Autocorrelation of the beginning of 'Plauderei an der Linde'. One clearly sees the peaks at the timebase of 246 ms, at the beat level of 516 ms and at the notated meter 2/4 (975 ms)

T_B. The tempo preference function can be modelled fairly well by a resonance curve with critical damping as in [12]. Parncutt [7] also uses a similar curve, derived from a fit to tapping data , which he calls pulse-period salience. Because the exact shape of the tempo preference curve is not important, we used the Parncutt function, which has a more intuitive form:

$$w(t) = e^{-\beta \log_2^2 t/t_s}, \quad (12)$$

where t_s denotes the spontaneous tempo, which is a free model parameter that was set by us to 500 ms throughout, and β being a damping factor, which is another free parameter ranging from $\beta = 1$ to $\beta = 2$. (See Fig. 3).

The set of beat candidates can now be defined as

$$T_b = \{t_i \in M(A_G), w(t_i)y_i = \max\} \quad (13)$$

But another constraint has to be applied on T_B to achieve musical meaningful results, coming from the corresponding timebase. The timebase is defined as the smallest (ideal) time unit in a musical piece [1] , and must be a integer subdivision of the beat. But subdivisions of the beat are usually only multiples of 2 ('binary feel') or 3 ('ternary feel'), or no subdivision at all. So, the final definition of the beat is:

$$T_B = \min_i \{t_i \in M'(A_G), w(t_i)y_i = \max\}, \quad (14)$$

with

$$M'(A_G) = \{t_i, \left[\frac{t_i}{\Delta T(t_i)}\right] = 2^n 3^m, n, m \in \mathbb{N}_0\}, \quad (15)$$

where the symbol $[\cdot]$ denotes the nearest integer (rounding) operation, and we take the minimal candidate in the extremely rare case of more than one possibility.

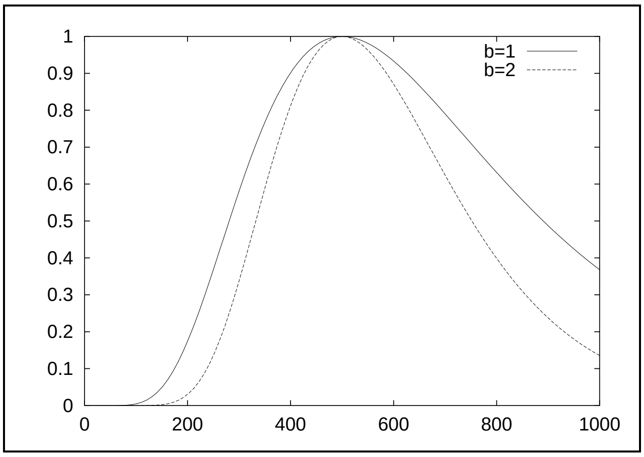

Figure 3. Tempo preference function with different dampings

4.3.2. Determination of the timebase

For a given beat candidate T_B the timebase ΔT can be derived from $M(A_G)$ with the following algorithm.

Consider the set of differences

$$\Delta M = (\Delta t_1, \Delta t_2, \ldots, \Delta t_N)$$

of the points from $M(A_G)$, with the properties $\Delta t_i \leq T_B$ and $\Delta t_i \geq 3\sigma$. The second property rules out 'unmusical' timebases, which might be caused by computational artifacts or grace notes. Then the timebase $\Delta T \in \Delta M$, is defined by

$$\left|\frac{T_B}{\Delta T} - \left[\frac{T_B}{\Delta T}\right]\right| = \min, \left[\frac{T_B}{\Delta T}\right] = 2^n 3^m, n, m \in \mathbb{N}_0 \quad (16)$$

If there is no such a timebase for a beat candidate, the candidate is ruled out. If for all beat candidates no appropiate timebase can be found, the algorithm stops.

4.4. Determination of meters and phases

Given the beat, the next level in a metrical hierarchy is the meter. It is defined as a subgrid of the beat grid. Although it can be presumed that the total duration of a (regular) meter should not exceed the subjective present of around $2 - 3\ s$, there are no clear measurements as, e.g., for the preferred tempo. Likewise, meter is much more ambiguous than the beat level, as e.g. the decision between 2/4 or 4/4 meter is often merely a matter of convention (or notation).

So the strategy used for meter determination is more heuristic, resulting in a list of possible meters with a weight, which can be interpreted as a relative probability of perceiving this meter, and which can be tested empirical. The problem of determining the correct phase is the most difficult one. One might conjecture that the interplay of possible but different phases for a given meter, or even of different meters, is a musical desirable effect, which might account for notions like groove or swing.

[1] sometimes called pulse

Meter period	Relative Accents
2	{2,0}
3	{2,0,0}
4	{2,0,1,0}
5	{2,0,0,1,0}
5	{2,0,1,0,0}
5	{2,0,0,0,0}
6	{2,0,1,0,1,0}
6	{2,0,0,1,0,0}
7	{2,0,1,0,2,0,0}
7	{2,0,0,2,0,1,0}
7	{2,0,0,2,0,2,0}

Table 1. List of prototypical accent structures

Nevertheless, our strategy is straightforward and is basically a pattern matching process with the help of cross-correlation of gaussifications. For the most common musical meters in western music prototypical accent patterns ([6]) are gaussificated on the base of the determined beat T_B, and then the cross-correlation with the rhythm is calculated over one period of the meter. The maximum value of this cross-correlation is defined as the match between the accent pattern and the rhythm, and along this way we also acquired the best phase for this meter. The matching value is then multiplied with the corresponding value of the autocorrelation function, this is the final weight for the meter.

The prototypical accent patterns we used can be found in Tab. 1. For some meters several variants are given, because they can be viewed as compound meters.

So from an accent pattern Ψ_p for a meter with period p and beat T_B we get the following Gaussification:

$$G_p(t; \Psi_p, T_B) = \sum_{k=1}^{N_{max}} \psi_i g(t; kT_B, \sigma), \qquad (17)$$

with N_{max} such, that $N_{max}T_B \geq t_N$

The match m_p is the maximum of the cross-correlation

$$m_p = \max_{0 \leq \tau < pT_B} C_{G_p G}(\tau) \qquad (18)$$

and the best phase φ is the corresponding time-lag. The weight w_p is the value

$$w_p = A_g(pT_B)m_p$$

5. EXAMPLES

In Fig. 1 the Gaussification of a folk song from Luxembourg ('Plauderei an der Linde') is shown. The input was quantized but distorted with random temporal noise of magnitude $\sigma =$ 50 ms. The original rhythm was notated in 2/4 meter with a two eight-note upbeat. The grid shown in the picture is based on the estimated beat $T_B = 516$ ms. Fig. 2 displays the corresponding autocorrelation function.

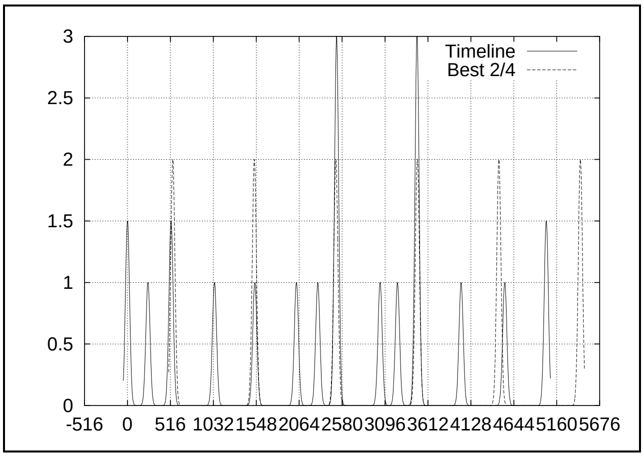

Figure 4. Best 2/4 Meter for 'Plauderei an der Linde'. One can see how the algorithm picks the best balancing phase.

Meter	Phase	Match	Weight
2	545 ms	1.39953	1.55218
3	540 ms	0.882693	0.587578
4	545 ms	1.04741	0.803957

Table 2. Phases, match and total weights for 'Plauderei an der Linde'

The important peaks are clearly identifiable. In Fig. 4 the best 2/4 meter is shown along with the original Gaussification. The cross-correlation algorithm searches for a good interpolating phase. The correponding cross-correlation function can be seen in Fig. 5 The weights, matches and best phases for this example are listed in Tab. 2

We also tested a MIDI rendition of the German popular song 'Mit 66 Jahren' by Udo Jürgens (Fig. 6) played by an amateur keyboard player. The autocorrelation can be seen in Fig. 7. Though the highest peak of the autocorrelation is around 303 ms, the algorithm chooses the value of 618 ms ($\sim$ 97 bpm) for the beat, cause of influence ot

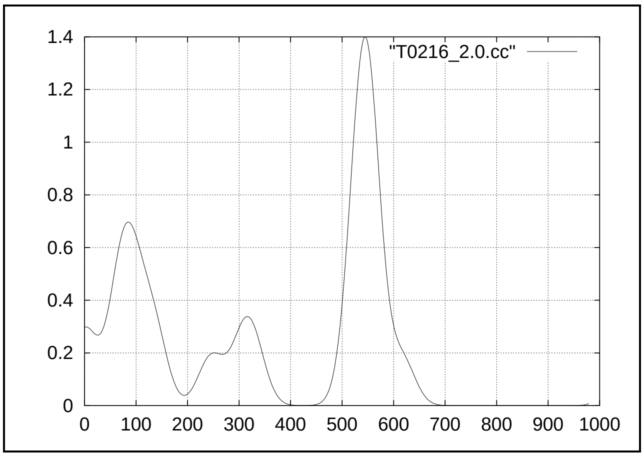

Figure 5. Cross-correlation function for 2/4 meter for 'Plauderei an der Linde'.

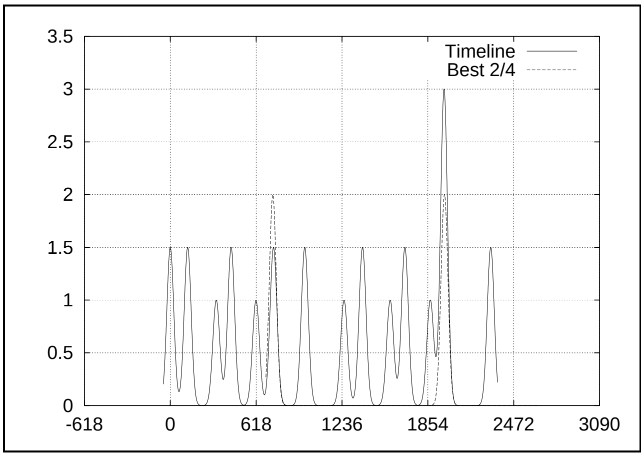

Figure 6. Gaussification of 'Mit 66 Jahren' and best 2/4 meter.

the tempo preference curve. The timebase is chosen to be 103 ms, indicating thet the player adopted a ternary feel to the piece, which is reasonable, because the original song has kind of a blues shuffle feel. The best meter is 2/4 (or 4/4 for the half beat), but the best phase is 738 ms. Compared to the original score, which is notated in 4/4, the calculated meter is phase-shifted by half a measure.

6. SUMMARY AND OUTLOOK

We presented a new algorithm for determining a metrical hierarchy from a list of onsets.

The first results are promising. For simple rhythm like they can be found in (western) folksongs, the algorithm works stable giving acceptable results compared to the score.

For more complicated or syncopated rhythm, as well as for ecological obtained data the results are promising, but not perfect in many cases, especially for meter extraction. However, it is the question, whether human listener are able to determine beat, meter and phase from those rhythms in a 'correct' way, if presented without the musical context and with no other accents present. This will be tested in the near future.

The algorithm can be expanded in a number of ways. The extension to polyphonic rhythms should be straightforward and might even stabilize the results. Furthermore, a window mechanism could be implemented, which is necessary for larger pieces and to account for tempo changes as accelerations or decelerations.

7. REFERENCES

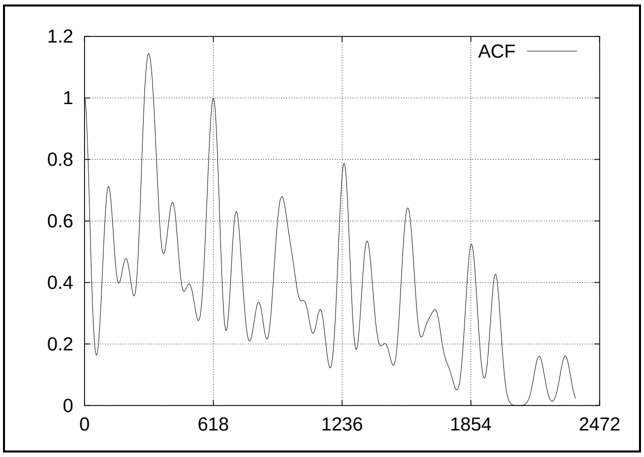

Figure 7. Autocorrelation of 'Mit 66 Jahren'

[1] Brown, J. "Determination of the meter of musical scores by autocorrelation", *J.Acoustic.Soc. Am*, 94(4), 1953-1957, 1993

[2] Eck, D. *Meter through synchrony:Processing rhythmical patterns with relaxation oscillators*. Unpublished doctotal dissertation, Indiana University, Bloomington, 2000.

[3] Fraisse, P. "Rhythm and tempo", in D.Deutsch (Ed.), *Psychology of music*, New York: Academic Press, 1982

[4] Frieler, K. *Mathematical music analysis*. Doctotal dissertation (in preparation), University of Hamburg , Hamburg.

[5] Large, E., & Kolen, J.F. "Resonance and the perception of musical meter", *Connection Science*, 6(1), 177-208, 1994

[6] Lerdahl, F & Jackendoff, R. *A generative theory of tonal music*. MIT Press,Cambridge, MA, 1983.

[7] Parncutt, R. "A perceptual model of pulse salience and metrical accents in musical rhythms", *Music Perception*, 11, 409-464, 1994

[8] Povel, D.J., & Essens, P. "Perception of temporal patterns", *Music Perception*, 2, 411-440, 1985

[9] Povel, D.J., & Okkermann, H. "Accents in equitone sequences", *Perception and Psychophysics*, 30, 565-572, 1981

[10] Seifert, U., Olk, F., & Schneider, A. "On rhythm perception: theoretical Issues, empirical findings", *J. of New Music Research*, 24, 164-195, 1995

[11] Toiviainen, P. & Snyder, J. S. "Tapping to Bach: Resonance-based modeling of pulse", *Music Perception*, 21(1), 43-80, 2003

[12] van Noorden, L. & Moelants, D. "Resonance in the the perception of musical pulse", *Journal of New Music Research*, 28, 43-66, 1999