COMPUTING PREDOMINANT LOCAL PERIODICITY INFORMATION IN MUSIC RECORDINGS

Peter Grosche and Meinard Müller

Saarland University and MPI Informatik
Campus E1 4, 66123 Saarbrücken, Germany
{pgrosche,meinard}@mpi-inf.mpg.de

ABSTRACT

The periodic structure of musical events plays a crucial role in the perception of tempo as well as the sensation of note changes and onsets. In this paper, we introduce a novel function that reveals the predominant local periodicity (PLP) in music recordings. Here, our main idea is to estimate for each time position a periodicity kernel that best explains the local periodic nature of previously extracted note onset information and then to accumulate all these kernels to yield a single function. This function, which is also referred to as PLP curve, reveals musically meaningful periodicity information even for non-percussive music with soft and blurred note onsets. Such information is useful not only for stabilizing note onset detection but also for beat tracking and tempo estimation in the case of music with changing tempo.

Index Terms— onset detection, novelty curve, periodicity, phase, tempo, beat, tactus, tatum

1. INTRODUCTION

Many different methods for the detection of note onsets in music recordings have been proposed [1, 2] and applied to tasks such as music transcription, beat tracking, tempo and meter estimation, as well as music synchronization [3, 4, 5, 6]. Most of the proposed onset detectors rely on the fact that note onsets often go along with a sudden increase of the signal's energy, which particularly holds for instruments such as piano, guitar, or percussive instruments. This property allows for extracting some kind of novelty curve from a music signal, the peaks of which yield good indicators for note onset candidates [1]. Much more challenging is the detection of onsets in the case of non-percussive music, where one often has to deal with soft onsets or blurred note transitions. As a consequence, more refined methods have to be used for computing the novelty curves, e. g., by analyzing the signal's spectral content, pitch, or phase [1, 7]. In the case of weak and blurry onsets, the resulting novelty curves tend to be rather noisy exhibiting many spurious peaks. Here, the selection of the relevant peaks that correspond to true note onsets becomes a difficult or even infeasible problem.

For many of the above mentioned applications, the explicit determination of note onsets is often not required. Here, it may suffice to have weak onset indicators from which one can directly derive the desired semantically higher-level information concerning tempo or structure. In this paper, we substantiate this observation by introducing a novel function that reveals the predominant local periodicity (PLP) even for non-percussive music with soft note onsets and changing tempo. Starting with a novelty curve that possibly has a noisy and poor peak structure, we estimate for each time position a periodicity kernel that best explains the local periodic nature of the novelty curve. Since there may be a number of outliers among these kernels, one may not obtain reliable information when looking at these kernels in a one-by-one fashion. Our idea is to accumulate all these kernels to obtain a single function, which we refer to as PLP curve. As it turns out, PLP curves are robust to outliers and reveal musically meaningful periodicity information even in the case of poor onset information. The musical motivation for introducing PLP curves is that the periodic structure of musical events plays a crucial role in the sensation of note changes. In particular, weak note onsets may only be perceptible within a rhythmic context. In this sense, a PLP curve can be regarded as a periodicity enhancement of the original novelty curve, indicating musically meaningful onset positions.

This paper is organized as follows. In Sect. 2, we review the concept of novelty curves while introducing a variant used in the subsequent sections. Sect. 3 constitutes the main contribution of this paper, where we introduce the concept of PLP curves that exhibit the local periodicity of novelty curves in a robust fashion. Our experiments are described in Sect. 4 and prospects of future work are sketched in Sect. 5. Related work is discussed in the respective sections.

2. NOVELTY CURVE

The most characteristic property going along with a note onset is a sudden increase in the signal's energy. However, simultaneously occurring events in polyphonic music may lead to masking effects that even out the energy ascents and prevent any observation of an energy increase. To circumvent these masking effects, detection functions were proposed that analyze the signal in a bandwise fashion [3] to extract transients occurring in certain frequency regions of the signal. As a side-effect of a sudden energy increase, there appears an accompanying broadband noise burst in the signal's spectrum. This effect is mostly masked by the signal's energy in lower frequency regions but well detectable in the higher frequency regions [8] of the spectrum. A widely used approach to onset detection in the frequency domain is the *spectral flux* [1], where changes of pitch and timbre are detected by analyzing the signal's short-time spectra. Recently, pitch-based algorithms have been proposed [7] that allow for capturing even smooth note transitions in non-percussive music.

Combining some of these ideas, we now describe an approach for computing novelty curves that indicate note onset candidates.

The research is funded by the *Cluster of Excellence on Multimodal Computing and Interaction* at Saarland University.

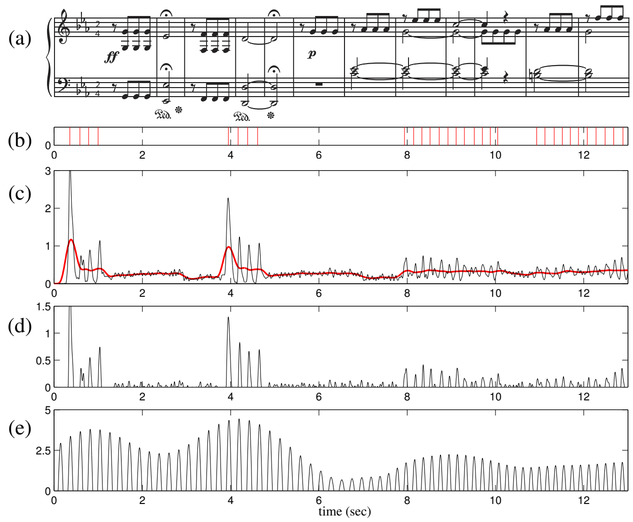

Figure 1: First 12 measures of Beethoven's Symphony No. 5 (Op. 67). **(a)** Score representation (in a piano reduced version). **(b)** Ground truth annotation of onsets (for an orchestral audio recording by Bernstein). **(c)** Novelty curve Δ (black curve) based on the compressed spectrogram Y with local average (red curve). **(d)** Novelty curve $\bar{\Delta}$. **(e)** PLP curve Γ.

Given a music recording, a short-time Fourier transform is used to obtain a spectrogram $X = (X(k,t))_{k,t}$ with $k \in [1 : K] := \{1, 2, \ldots, K\}$ and $t \in [1 : T]$. Here, K denotes the number of Fourier coefficients, T denotes the number of frames, and $X(k,t)$ denotes the k^{th} Fourier coefficient for time frame t. In our implementation, each time parameter t corresponds to 23 milliseconds of the audio. Note that the Fourier coefficients of X are linearly spaced on the frequency axis. Using suitable binning strategies, various approaches switch over to a logarithmically spaced frequency axis, e.g., by using mel-frequency bands or pitch bands, see [5, 3]. Here, we keep the linear frequency axis, since it puts greater emphasis on the high-frequency regions of the signal, thus accentuating the afore mentioned noise bursts visible as high-frequency content. Next, we apply a logarithm to the magnitude spectrogram $|X|$ of the signal yielding $Y := \log(1 + C \cdot |X|)$ for a suitable constant $C > 1$, see [3]. Such a compression step not only accounts for the logarithmic sensation of sound intensity but also allows for adjusting the dynamic range of the signal to enhance the clarity of weaker transients, especially in the high-frequency regions. In our experiments, we use the value $C = 1000$.

To obtain a novelty curve, we basically compute the discrete derivative of the compressed spectrum Y. More precisely, we sum up only positive intensity changes to emphasize onsets while discarding offsets to obtain the novelty function $\Delta : [1 : T-1] \to \mathbb{R}$:

$$\Delta(t) := \sum_{k=1}^{K} |Y(k,t+1) - Y(k,t)|_{\geq 0} \qquad (1)$$

for $t \in [1 : T-1]$, where $|x|_{\geq 0} := x$ for a non-negative real number x and $|x|_{\geq 0} := 0$ for a negative real number x. Fig. 1c shows the resulting curve for a Bernstein recording of Beethoven's Fifth Symphony. To obtain our final novelty function $\bar{\Delta}$, we subtract the local average and only keep the positive part (half-wave rectification), see Fig. 1d. In our actual implementation, we use a higher-order smoothed differentiator. Furthermore, we process the spectrum in a bandwise fashion using 5 bands [5]. The resulting 5 novelty curves are weighted and summed up to yield the final novelty function. For details, we refer to the literature.

The particular design of the novelty curve is not in the focus of this paper. Our PLP curves as introduced in Sect. 3 are designed to work even for noisy novelty curves with a poor peak structure. Of course, the overall result may be improved by employing more refined novelty curves as suggested in [7].

3. PREDOMINANT LOCAL PERIODICITY CURVE

The peaks of a novelty curve indicate note onset candidates. In order to determine the significant peaks, one generally refers to peak-picking strategies based on a combination of fixed and adaptive thresholding [1]. For soft and blurred onsets, however, novelty curves tend to be noisy and the musically meaningful peaks often become indistinguishable from spurious peaks. Instead of extracting unreliable onsets from a noisy novelty curve, we introduce a function that takes periodic properties of the novelty curve into account. This approach is closely related to the extraction of rhythmic structure [3, 5] and tempo [4] from music recordings. Most of the previous work focused on determining musical pulses on the *tactus* (the foot tapping rate or beat [3]) or *measure* level, but only few approaches exist for analyzing the signal on the finer tatum level. Here, a *tatum* or *temporal atom* refers to the fastest repetition rate of musically meaningful accents occurring in the signal. Thus, all note onsets roughly occur at tatum pulse positions. In [9], the extraction of a *tatum grid* is proposed by first detecting note onsets and then analyzing inter-onset intervals (IOI) to estimate the finest repetition rate occurring in the signal. In contrast to these approaches, our goal is to extract the *predominant local periodicity* (PLP) of accents in the music signal, which may be a pulse on the tatum, the tactus, or measure level. Furthermore, our approach does not assume constant tempo throughout the recording. Actually, our PLP curve exhibits the predominant pulse for each time position thus making local tempo information explicit.

Let $\bar{\Delta}$ be the novelty function as described in Sect. 2. To avoid boundary problems, we assume that $\bar{\Delta}$ is defined on $\mathbb{Z}$ by setting $\bar{\Delta}(t) := 0$ for $t \in \mathbb{Z} \setminus [1 : T-1]$. First, we investigate the local periodicity of the novelty curve by fitting for each time position $t \in \mathbb{Z}$ a periodicity kernel into $\bar{\Delta}$. This estimation is obtained basically by performing a harmonic analysis of $\bar{\Delta}$. More precisely, let $\Omega \subset \mathbb{R}_{>0}$ be a finite set of frequency parameters. The frequency $\omega \in \Omega$ corresponds to the period $1/\omega$. Furthermore, we fix a window function $W : \mathbb{Z} \to \mathbb{R}$ centered at $t = 0$ with support $[-N : N]$. In our experiments, we use a Hann window of size $2N + 1$. The complex Fourier coefficient $F(\omega, t)$ is obtained by

$$F(\omega, t) = \sum_{n \in \mathbb{Z}} \bar{\Delta}(n) \cdot W(n-t) \cdot \mathrm{e}^{-2\pi i \omega n}\,. \qquad (2)$$

For each $t \in [1 : T]$, we then compute the frequency parameter $\omega_t \in \Omega$ that maximizes the magnitude of $F(\omega, t)$:

$$\omega_t := \mathrm{argmax}_{\omega \in \Omega} |F(\omega, t)|. \qquad (3)$$

The corresponding phase is denoted by φ_t and can be computed by means of the following formula [6]:

$$\varphi_t := \frac{1}{2\pi} \arccos\left(\frac{\mathrm{Re}(F(\omega_t, t))}{|F(\omega_t, t)|}\right). \qquad (4)$$

Using ω_t and φ_t, the optimal periodicity kernel $\kappa_t : \mathbb{Z} \to \mathbb{R}$ for $t \in [1 : T]$ is defined as the windowed sinusoid

$$\kappa_t(n) := W(n-t)\cos(2\pi(\omega_t n - \varphi_t)) \qquad (5)$$

for $n \in \mathbb{Z}$. Fig. 2a shows various optimal periodicity kernels for our Beethoven example. Intuitively, the sinusoid κ_t best explains

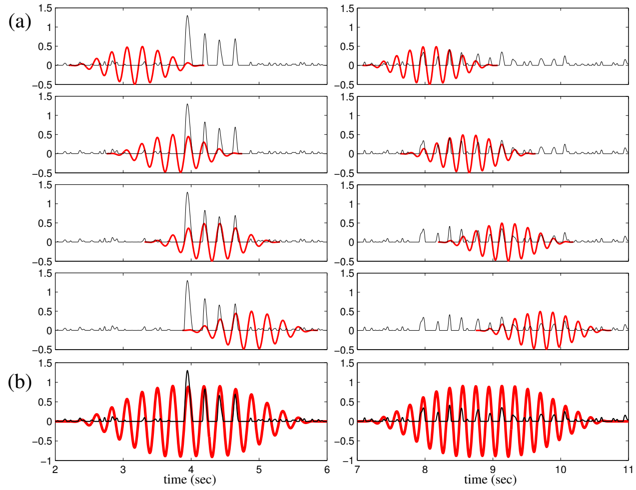

Figure 2: **(a):** Optimal periodicity kernels κ_t for various time parameters t using a kernel size of 2 seconds for two different parts of the novelty curve shown in Fig. 1d. **(b)** Accumulation of all kernels. From this, the PLP curve Γ (see Fig. 1e) is obtained by half-wave rectification.

the local periodic nature of the novelty curve at time position t with respect to the set Ω. The period $1/\omega_t$ corresponds to the predominant periodicity of the novelty curve and the phase information φ_t takes care of accurately aligning the maxima of κ_t and the peaks of the novelty curve.

The properties of the kernels κ_t depend not only on the quality of the novelty curve, but also on the window size $2N + 1$ of W and the set of frequencies Ω. Increasing the parameter N yields more robust estimates for ω_t at the cost of temporal flexibility. In our experiments, we chose a window length of 4 to 8 seconds. In the following, this duration is referred to as *kernel size*. Furthermore, by restricting Ω to certain frequency ranges one can influence the desired pulse level to be captured. We use the set $\Omega = \{k/60 \mid k \in [30 : 600]\}$, which covers the (integer) musical tempi between 30 and 600 beats per minute (BPM). Here, the upper bound of 600 BPM is motivated by the psychoacoustic fact that only events that show a temporal separation greater than 120 milliseconds (corresponding to 500 BPM) contribute to the perception of rhythm [10]. Thus, our concept covers the tatum grid as introduced in [9] in a perceptually meaningful pulse range.

The estimation of optimal periodicity kernels in regions with a strongly corrupted peak structure is still problematic. This particularly holds in the case of small kernel sizes. To make the periodicity estimation more robust, our idea is to accumulate these kernels over all time positions to form a single function instead of looking at the kernels in a one-by-one fashion. More precisely, we define a function $\Gamma : [1 : T] \to \mathbb{R}_{\geq 0}$ as follows:

$$\Gamma(n) = \textstyle\sum_{t\in[1:T]} |\kappa_t(n)|_{\geq 0} \tag{6}$$

for $n \in [1 : T]$, see Fig. 2b. The resulting function is referred to as *predominant local periodicity curve* or, for short, as *PLP curve*. As it turns out, such PLP curves are robust to outliers and reveal musically meaningful periodicity information even when starting with relatively poor onset information.

4. DISCUSSION AND EXPERIMENTS

In this section, we discuss various properties of PLP curves based on representative examples and a more quantitative evaluation on

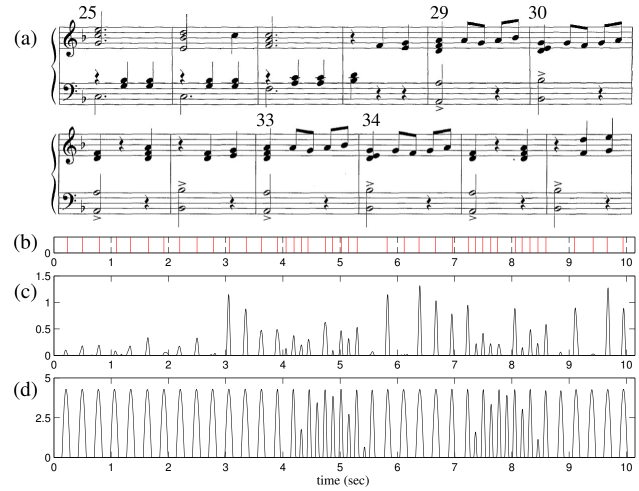

Figure 3: Excerpt of Shostakovich's 2nd Waltz from Jazz Suite No. 2. **(a)** Score representation of measures 25 to 36 (in a piano reduced version). **(b)** Annotated ground truth onsets (for an orchestral audio recording by Yablonsky). **(c)** Novelty curve $\bar{\Delta}$. **(d)** PLP curve Γ.

two different datasets in the context of onset detection.

We start with the PLP curve for our Beethoven example shown in Fig. 1e. This orchestral piece constitutes a great challenge. First, there are many significant local tempo changes (e. g., caused by fermatas). Second, besides very dominant note onsets in the fortissimo part at the beginning of the piece, there are also many soft and blurred note onsets in the piano part played by strings. Despite these properties, the maxima of Γ align well with musically relevant peaks of the novelty curve. While the detection of note onsets in the fortissimo section can be achieved with relatively simple energy-based methods, the PLP curve Γ also reveals periodically spaced peaks in the problematic noisy piano section. As it turns out, these periodically spaced peaks correspond exactly to the musically meaningful note onsets. In other words, by first estimating the Γ-encoded local rhythmic grid, one can then identify the weak onsets that lie on musically meaningful pulse positions. Furthermore, the amplitudes of the periodicity kernels have unity value and do not depend on the height of the peaks or the signal's energy. As a consequence, the PLP curve is invariant under changes in dynamics and the amplitude of Γ indicates the confidence in the periodicity estimation. Consistent kernel estimations produce constructive interferences in the accumulation resulting in high values of Γ. Contrary, outliers or inconsistencies in the kernel estimations cause destructive interferences in the accumulation resulting in lower values of Γ. In Fig. 1e, this effect is visible in the fermata sections, where no consistent onset information is available (the noisy peaks are caused by vibrato).

As second example, we consider the second Waltz from the Jazz Suite No. 2 by Shostakovich. Fig. 3a shows an excerpt (measures 25 to 36) of the piano reduced score. The audio recording is an orchestral version conducted by Yablonsky. Here, the first beats in the 3/4 Waltz are played by non-percussive instruments leading to relatively soft and blurred onsets, whereas the second and third beats are played by percussive instruments. This is also reflected by the novelty curve shown in Fig. 3c, where some of the peaks corresponding to the soft onsets are hardly visible. However, the beat period (tactus level) is perfectly disclosed by the PLP curve Γ, see Fig. 3d. Also the predominant eighth note tatum pulse in measures 29/30 and measures 33/34 is captured by Γ, which locally

Dataset		PUBLIC			PRIVATE		
Curve	KS	P	R	F	P	R	F
$\bar{\Delta}$		0.783	**0.821**	0.793	0.694	**0.732**	0.698
Γ	4	0.591	**0.933**	0.695	0.588	**0.913**	0.679
Γ	6	0.599	**0.955**	0.705	0.599	**0.907**	0.689
Γ	8	0.597	**0.944**	0.701	0.588	**0.877**	0.674

Table 1: Mean precision, recall, and F-measure values using the novelty curve $\bar{\Delta}$ and PLP curves Γ with different kernel sizes 'KS' (in seconds).

switches from the tactus to the tatum level.

In view of a more quantitative evaluation, we describe the behavior of our PLP curves by means of precision and recall values in the context of note onset detection. To this end, we use two different evaluation datasets containing audio recordings along with manually labeled reference note onsets. First, we use a publicly available[1] dataset [11], which is widely used in onset detection experiments [7]. This dataset, referred to as PUBLIC, consists of 242 seconds of audio (17 music excerpts of different genre) with 671 labeled onsets. Second, we use a dataset that particularly contains classical music with soft onsets and significant tempo changes. This dataset, referred to as PRIVATE, consists of 201 seconds of audio with 569 manually labeled onsets. For determining onsets from the original novelty curve $\bar{\Delta}$, we use a locally adaptive peak picking strategy as proposed in [1]. Similarly, we determine onsets from the corresponding PLP curve simply by picking the local maxima of Γ (for this curve, peak picking becomes a trivial task). Following the MIREX 2007 Audio Onset Detection evaluation procedure[2], each detected onset is considered a *true positive* if there is a reference onset within a tolerance bound of 50 milliseconds, otherwise a *false positive*. Furthermore, each reference onset not associated to a true positive is referred to as *false negative*. From this, we derive the standard precision (P), recall (R), and the F-measure (F) values, see [7].

Table 1 shows the resulting average P, R, and F values for the original novelty curve $\bar{\Delta}$ and for PLP curves using periodicity kernels of different sizes. Note that for the PLP curves, all peak positions of the induced local periodicity grid are taken as positives. Since musical note onsets are rhythmically correlated, one may expect that the onsets lie on PLP-defined peak positions. This expectation is confirmed by our experiments, where the PLP curves achieved higher recall rates than the original novelty curves. For example, using PLP curve Γ of a kernel size 4, the mean recall R increased from 0.821 to 0.933 for the PUBLIC set and from 0.732 to 0.913 for the PRIVATE set. This result shows that a vast majority of the relevant note onsets indeed lie on the PLP-defined pulse grid, which also indicates the robustness and accuracy of our method. Especially for the PRIVATE set, the PLP curve is still capable of capturing the local periodicity, despite of soft onsets and tempo changes. On the other side, the precision values for Γ are lower than for $\bar{\Delta}$. This is not surprising, since we consider *all* peak positions of Γ in the evaluation. Even though most note onsets are captured by the grid, not all grid positions necessarily correspond to note onsets. For example, the mean precision of Γ roughly amounts to 0.6 for both datasets, which shows that 40% of the Γ-encoded pulses do not correspond to note onsets. This is also illustrated by the Beethoven example shown in Fig. 1e, where the fermata sections are filled with periodically spaced Γ-pulses. Even though such positions are counted as false positives in our evaluation, they generally correspond to rhythmically meaningful pulses. Finally, we look at the influence of the kernel size on the PR values. Here, note that most of the excerpts in the PUBLIC dataset have a constant tempo. Therefore, using a kernel size of 6 seconds instead of 4 seconds, the kernel estimation is more robust leading to a slight increase of recall (from $R = 0.933$ to $R = 0.955$). Contrary, the PRIVATE dataset contains music with many tempo changes. Here, kernels of smaller sizes are better suited for adjusting the local periodicities according to the tempo.

[1] http://old.lam.jussieu.fr/src/Membres/Leveau/SOL/SOL

[2] http://www.music-ir.org/mirex/2007/index.php/Audio_Onset_Detection

5. CONCLUSIONS

In this paper, we have introduced a novel concept for determining the predominant local periodicity (PLP) in music recordings with possibly poor onset information and changing tempo. The PLP curves not only align to the dominant pulse level but also adapt to the local tempo. Based on the PLP-encoded local periodicity grid, one can identify even weak onsets on musically meaningful pulse positions. In this sense, PLP curves facilitate a periodicity-adaptive peak picking. First experiments show that PLP curves are also a powerful tool for beat tracking applications and local tempo estimation in particular for classical music that reveals significant variations in tempo, dynamics, and timbre. Furthermore, we plan to use PLP curves for improving the temporal accuracy of alignments in the context of music synchronization applications. Such improvements are essential for tasks such as performance analysis, where one objective is to extract expressive tempo information from music recordings.

6. REFERENCES

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

[2] N. Collins, "A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions," in *AES Convention 118*, 2005.

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

[4] G. Peeters, "Time variable tempo detection and beat marking," in *Proc. ICMC*, Barcelona, Spain, 2005.

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

[6] M. Müller, *Information Retrieval for Music and Motion*. Springer, 2007.

[7] R. Zhou, M. Mattavelli, and G. Zoia, "Music onset detection based on resonator time frequency image," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1685–1695, 2008.

[8] P. Masri and A. Bateman, "Improved modeling of attack transients in music analysis-resynthesis," in *Proc. ICMC, Hong-Kong*, 1996, pp. 100–103.

[9] J. Seppänen, "Tatum grid analysis of musical signals," in *Proc. IEEE WASPAA*, 2001, pp. 131–134.

[10] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[11] L. Daudet, G. Richard, and P. Leveau, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," in *Proc. ISMIR*, Barcelona, Spain, 2004.