

CYCLIC TEMPOGRAM – A MID-LEVEL TEMPO REPRESENTATION FOR MUSICSIGNALS

Peter Grosche, Meinard Müller*

Saarland University and MPI Informatik
Campus E1 4, 66123 Saarbrücken, Germany

Frank Kurth

Fraunhofer-FKIE, Abteilung KOM
53343 Wachtberg-Werthhoven, Germany

ABSTRACT

The extraction of local tempo and beat information from audio recordings constitutes a challenging task, particularly for music that reveals significant tempo variations. Furthermore, the existence of various pulse levels such as measure, tactus, and tatum often makes the determination of absolute tempo problematic. In this paper, we present a robust mid-level representation that encodes local tempo information. Similar to the well-known concept of cyclic chroma features, where pitches differing by octaves are identified, we introduce the concept of cyclic tempograms, where tempi differing by a power of two are identified. Furthermore, we describe how to derive cyclic tempograms from music signals using two different methods for periodicity analysis and finally sketch some applications to tempo-based audio segmentation.

Index Terms— tempo, tempogram, chroma, music signals, audio segmentation

1. INTRODUCTION

For the processing of music signals, various feature representations have been proposed that account for musical dimensions such as pitch, harmony, rhythm, or tempo. For example, exploiting the fact that most Western music is based on the equal-tempered scale, music signals can be decomposed into logarithmically spaced frequency bands that correspond to the pitch scale [1]. From the resulting pitch representation one obtains the well-known *chroma representation* by pooling pitches that correspond to the same chroma. Here, the chromas correspond to the twelve pitch spelling attributes C, C[♯], D, . . . , B of the equal-tempered scale [2]. Chroma features closely correlate to harmonic aspects and have turned out to be a powerful mid-level representation for various music analysis and retrieval tasks such as chord transcription, audio matching, or music segmentation.

Besides pitch and harmony, the aspects of rhythm and tempo constitute further important dimensions of music. In this context, various representations have been proposed that reveal local tempo and beat information [4, 5, 6, 7]. To accomplish this task, most approaches proceed in two steps.

First, note onset candidates are extracted from the music signals by exploiting the fact that note onsets typically cause sudden changes of the signal’s energy and spectrum. Based on this property so-called *novelty curves* are derived, the peaks of which yield good indicators for note onset candidates [8, 9]. Then, using autocorrelation [5], Fourier [6, 7], or comb filter methods [4], the novelty curves are analyzed for locally periodic patterns, from which the local tempo and beat information can be estimated.

In the case that the music recording reveals significant tempo changes, the detection of locally periodic patterns becomes a challenging problem. Furthermore, there are various pulse levels that contribute to the human perception of tempo such as the measure, tactus, and tatum levels [10]. As an analogy, these different levels may be compared to the existence of harmonics in the pitch context. Inspired by the concept of chroma features, in this paper, we introduce the concept of *cyclic tempograms*, where the idea is to form tempo equivalence classes by identifying tempi that differ by a power of two. This concept has first been suggested in [11] and will be formalized and expanded in this paper. The resulting cyclic tempo features constitute a robust mid-level representation that reveals local tempo characteristics of music signals while being invariant to changes in the pulse level. Being the tempo-based counterpart of the harmony-based chromagrams, cyclic tempograms are suitable for music analysis and retrieval tasks, where harmony-based criteria are not relevant.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the novel concept of cyclic tempograms. Then, in Sect. 3, we discuss two methods on how to extract these features from music signals. Finally, in Sect. 4, we sketch various applications and discuss future work.

2. CYCLIC TEMPOGRAM

We now introduce the concept of cyclic tempograms in an abstract form using, for the sake of clarity, a continuous formulation. In the next section, we then discuss how this concept can be realized in practice by using discrete parameters. Similar to a spectrogram, we define a *tempogram* to be a time-tempo representation for a given time-dependent signal. Mathematically, a tempogram is mapping $\mathcal{T} : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ depending on a time parameter $t \in \mathbb{R}$ measured in sec-

*The authors are funded by the *Cluster of Excellence on Multimodal Computing and Interaction (MMCI)* at Saarland University.

onds and a tempo parameter $\tau \in \mathbb{R}_{>0}$ measured in beats per minute (BPM). Intuitively, the value $\mathcal{T}(t, \tau)$ indicates to which extend a pulse of tempo τ is present at time t . For example, let us suppose that a music signal has a dominant tempo of 120 BPM around position $t = 20$ seconds. Then the resulting tempogram \mathcal{T} should have a large value $\mathcal{T}(t, \tau)$ for $\tau = 120$ and $t = 20$. Because of the above mentioned ambiguity concerning the pulse levels, one typically also has large values of \mathcal{T} for integer multiples $\tau, 2\tau, 3\tau, \dots$ (referred to as *harmonics* of τ) and integer fractions $\tau, \tau/2, \tau/3, \dots$ (referred to as *subharmonics* of τ), see Fig. 1.

To reduce the impact of such kind of tempo confusions, we apply a similar strategy as in the computation of chroma features [2]. Recall that two pitches having fundamental frequencies f_1 and f_2 are considered as octave equivalent, if they are related by $f_1 = 2^k f_2$ for some $k \in \mathbb{Z}$. Similarly, we say that two tempi τ_1 and τ_2 are *octave equivalent*, if they are related by $\tau_1 = 2^k \tau_2$ for some $k \in \mathbb{Z}$. Then, for a given tempo parameter τ , the resulting tempo equivalence class is denoted by $[\tau]$. For example, for $\tau = 120$ one has $[\tau] = \{\dots, 30, 60, 120, 240, 480 \dots\}$. Now, the *cyclic tempogram* \mathcal{C} induced by \mathcal{T} is defined by

$$\mathcal{C}(t, [\tau]) := \sum_{\lambda \in [\tau]} \mathcal{T}(t, \lambda). \quad (1)$$

Note that the tempo equivalence classes topologically correspond to a circle. Fixing a reference tempo ρ (e. g., $\rho = 60$ BPM), the cyclic tempogram can be represented by a mapping $\mathcal{C}_\rho : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ defined by

$$\mathcal{C}_\rho(t, s) := \mathcal{C}(t, [s \cdot \rho]), \quad (2)$$

for $t \in \mathbb{R}$ and $s \in \mathbb{R}_{>0}$. Note that $\mathcal{C}_\rho(t, s) = \mathcal{C}_\rho(t, 2^k s)$ for $k \in \mathbb{Z}$ and \mathcal{C}_ρ is completely determined by its values $s \in [1, 2)$. As illustration, Fig. 1 shows various tempograms for a click track of increasing tempo. Fig. 1b shows a tempogram with harmonics and Fig. 1c the resulting cyclic tempogram. As in the pitch context, the tempo class $[3\tau]$ is referred to as the *tempo dominant*. In Fig. 1c, the tempo dominant is visible as the increasing line in the middle. Similarly, Fig. 1d shows a tempogram with subharmonics and Fig. 1e the resulting cyclic tempogram. Here, the tempo class $[\tau/3]$ is referred to as the *tempo subdominant* see the increasing line in the middle of Fig. 1e.

3. COMPUTING CYCLIC TEMPOGRAMS

We now describe how discrete cyclic tempograms can be computed in practice for digital music signals. In a first step, a novelty curve is extracted (Sect. 3.1). In a second step, local periodic patterns are derived from the novelty curve. Here, we discuss two different methods that yield tempograms with harmonics (Fourier tempogram, Sect. 3.2) and with subharmonics (autocorrelation tempogram, Sect. 3.3), respectively. Finally, we show how cyclic versions are obtained from these tempograms (Sect. 3.4).

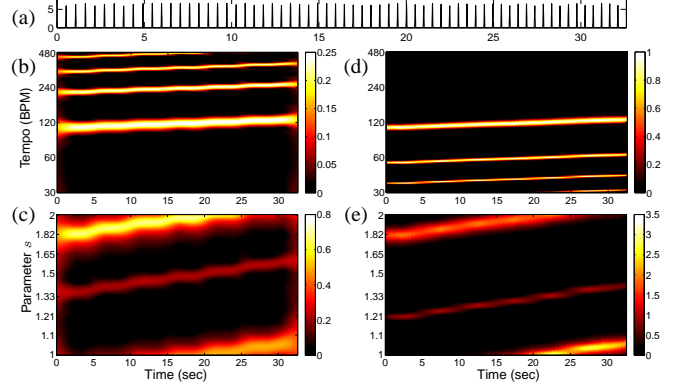


Fig. 1. (a) Novelty curve of click track of increasing tempo (110 to 130 BPM). (b) Fourier tempogram (showing harmonics). (c) Cyclic tempogram \mathcal{C}_{60} induced by (b). (d) Autocorrelation tempogram (showing subharmonics). (e) Cyclic tempogram \mathcal{C}_{60} induced by (d).

3.1. Novelty Curve

Most methods for deriving novelty curves from audio signals are based on the fact that note onsets typically cause a sudden change of the signal's energy and spectrum, see [8, 9]. Here, we exemplarily discuss one such basic approach. Given a music signal, a short-time Fourier transform is used to obtain a spectrogram $X = (X(t, k))_{t, k}$, $k \in [1 : K] := \{1, 2, \dots, K\}$, $t \in \mathbb{Z}$, where K is the number of Fourier coefficients and $X(k, t)$ is the k^{th} Fourier coefficient for time frame t . In our implementation, each time parameter t corresponds to a step size of $r = 0.023$ seconds. Next, we apply a logarithm to the magnitude spectrogram $|X|$ yielding $Y := \log(1 + C \cdot |X|)$ for a constant $C > 1$, see [10]. 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 curve $\Delta : \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$:

$$\Delta(t) := \sum_{k=1}^K |Y(t+1, k) - Y(t, k)|_{\geq 0} \quad (3)$$

for $t \in \mathbb{Z}$, where $|x|_{\geq 0} := x$ for a non-negative and $|x|_{\geq 0} := 0$ for a negative real number x . To obtain our final novelty curve, again denoted by Δ , we subtract the local average and only keep the positive part, see Fig. 2a for an example. For refinements and variants of this procedure, we refer to [8, 9].

3.2. Fourier Tempogram

In the next step, we analyze the novelty curve Δ with respect to local periodic patterns using a short-time Fourier transform similar to [6, 7]. To this end, we fix a window function $W : \mathbb{Z} \rightarrow \mathbb{R}$ centered at $t = 0$. In our experiments, we use a Hann window of a width corresponding to six seconds. Then, for a frequency parameter $\omega \in \mathbb{R}_{\geq 0}$, the complex Fourier coefficient $\mathcal{F}(t, \omega)$ is defined by

$$\mathcal{F}(t, \omega) = \sum_{n \in \mathbb{Z}} \Delta(n) \cdot W(n - t) \cdot e^{-2\pi i \omega n}. \quad (4)$$

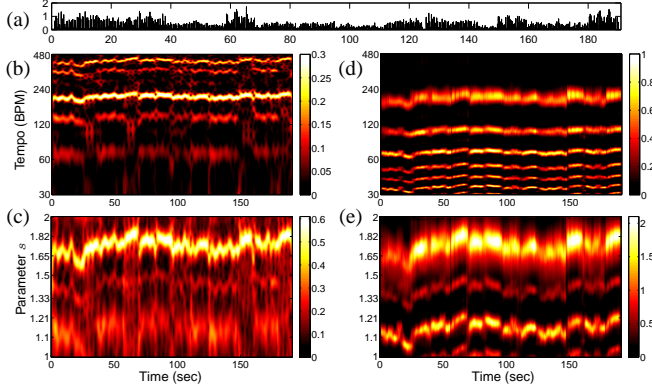


Fig. 2. (a) Novelty curve Δ of an audio recording of a Waltz by Shostakovich. (b) Fourier tempogram \mathcal{T}^F . (c) Cyclic tempogram \mathcal{C}_{60}^F . (d) Autocorrelation tempogram \mathcal{T}^A . (e) Cyclic tempogram \mathcal{C}_{60}^A .

In the musical context, we rather think of tempo measured in beats per minutes (BPM) than of frequency measured in Hertz (Hz). Therefore, we use a tempo parameter τ satisfying the equation $\tau = 60 \cdot \omega$. Furthermore, we compute the tempi only for a finite set $\Theta \subset \mathbb{R}_{>0}$. In our implementation, we cover four tempo octaves ranging from $\tau = 30$ to $\tau = 480$. Furthermore, we sample this interval in a logarithmic fashion covering each tempo octave by M samples, where the integer M determines the tempo resolution. Then, the discrete *Fourier tempogram* $\mathcal{T}^F : \mathbb{Z} \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ is defined by

$$\mathcal{T}^F(t, \tau) = |\mathcal{F}(t, \tau/60)|. \quad (5)$$

As an example, Fig. 2b shows the tempogram \mathcal{T}^F of a recording of a Waltz by Shostakovich. In \mathcal{T}^F , the tempo on the beat level (roughly $\tau = 216$ BPM) and the second harmonics of this tempo are dominant. However, the tempo on the measure level of the three-quarter Waltz (roughly 72 BPM, third subharmonics of $\tau = 216$) is hardly noticeable. Actually, since the novelty curve Δ locally behaves like a track of positive clicks, it is not hard to see that Fourier analysis responds to harmonics but suppresses subharmonics, see also [6].

3.3. Autocorrelation Tempogram

In the context of tempo estimation, also autocorrelation-based methods are widely used to estimate local periodicities [5]. Since these methods, as it turns out, respond to subharmonics while suppressing harmonics, they ideally complement Fourier-based methods, see [6]. To obtain a discrete *autocorrelation tempogram*, we proceed as follows. Again, we fix a window function $W : \mathbb{Z} \rightarrow \mathbb{R}$ centered at $t = 0$ with support $[-N : N]$, $N \in \mathbb{N}$. This time, we use a box window of a width corresponding to six seconds. The local autocorrelation is then computed by comparing the windowed novelty curve with time shifted copies of itself. More precisely, we use the

unbiased local autocorrelation

$$\mathcal{A}(t, \ell) = \frac{\sum_{n \in \mathbb{Z}} \Delta(n) \Delta(n + \ell) \cdot W(n - t)}{2N + 1 - \ell}, \quad (6)$$

for time $t \in \mathbb{Z}$ and time lag $\ell \in [0 : N]$. Recall from Sect. 3.1 that each time parameter $t \in \mathbb{Z}$ corresponds to r seconds (in our implementation we used $r = 0.023$). Then, the lag ℓ corresponds to the tempo $\tau = 60/(r \cdot \ell)$ BPM. We therefore define the *autocorrelation tempogram* \mathcal{T}^A by

$$\mathcal{T}^A(t, \tau) = \mathcal{A}(t, \ell). \quad (7)$$

for each tempo $\tau = 60/(r \cdot \ell)$, $\ell \in [1 : N]$. Finally, using standard resampling and interpolation techniques applied to the tempo domain, we derive an autocorrelation tempogram $\mathcal{T}^A : \mathbb{Z} \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ that is defined on the same tempo set Θ as the Fourier tempogram \mathcal{T}^F , see Sect. 3.2. The tempogram \mathcal{T}^A for our Shostakovich example is shown in Fig. 2d, which clearly indicates the subharmonics.

3.4. Cyclic Tempogram

Recall that the tempo parameter set Θ comprises four tempo octaves ranging from $\tau = 30$ to $\tau = 480$, where each octave is covered by M logarithmically spaced samples. Therefore, one obtains a discrete cyclic tempogram \mathcal{C}^F (resp. \mathcal{C}^A) from the tempogram \mathcal{T}^F (resp. \mathcal{T}^A) simply by adding up the corresponding values of the four octaves as described in Eq. (1). Using a reference tempo of $\rho = 60$ BPM, we obtain the cyclic tempogram \mathcal{C}_{60}^F (resp. \mathcal{C}_{60}^A). Note that these discrete cyclic tempograms are M -dimensional, where the cyclic tempo axis is sampled at M positions. As an illustration, Fig. 2c (resp. Fig. 2e) shows the discrete cyclic tempogram \mathcal{C}_{60}^F (resp. \mathcal{C}_{60}^A), where we used a tempo resolution of $M = 120$. Note that the subharmonic tempo at measure level corresponding to roughly 72 BPM ($s = 1.2$) is clearly visible in \mathcal{C}_{60}^A , but not in \mathcal{C}_{60}^F .

4. APPLICATIONS AND FUTURE WORK

As mentioned before, the cyclic tempograms are the tempo-based counterparts of the harmony-based chromagrams. Compared to usual tempograms, the cyclic versions are more robust to tempo ambiguities that are caused by the various pulse levels. Furthermore, one can simulate changes in tempo simply by cyclically shifting a cyclic tempogram. Note that this is similar to the property of chromagrams, which can be cyclically shifted to simulate modulations in pitch. As one further advantage, even low-dimensional versions of discrete cyclic tempograms still bear valuable local tempo information of the underlying musical signal.

To illustrate the potential of our concept, we sketch how cyclic tempograms can be used for automated music segmentation, which is a central task in the field of music information retrieval [1, 3, 12]. Actually, there are many different

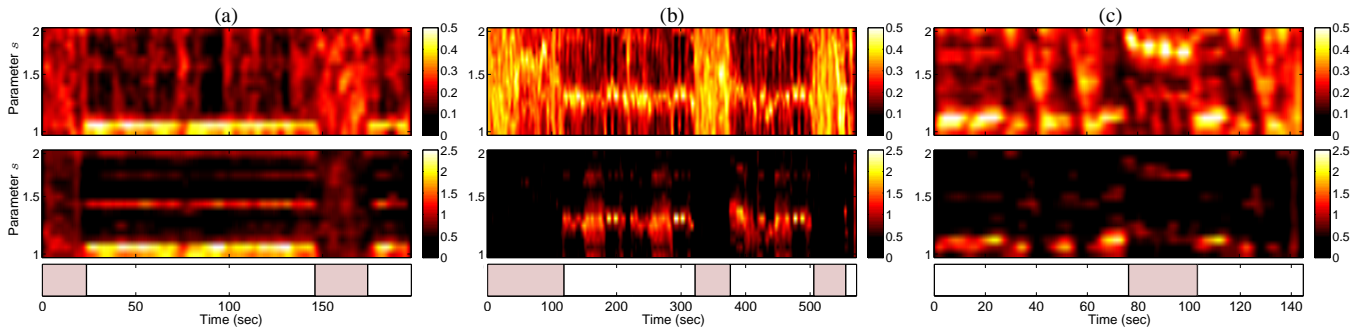


Fig. 3. Cyclic tempograms C_{60}^F (top) and C_{60}^A (middle) with $M = 15$ as well as tempo-based segmentations (bottom) for three different pieces. (a) *In The Year 2525* by Zager and Evans. (b) *Piano Sonata Op. 13 (Pathétique)* by Beethoven. (c) *Hungarian Dance No. 5* by Brahms.

strategies for segmenting music signals such as novelty-based, repetition-based, and homogeneity-based strategies. In the latter, the idea is to partition the music signal into segments that are homogenous with regard to a specific musical property [12]. In this context, timbre-related audio features such as MFCCs or spectral envelopes are frequently used, resulting in timbre-based segmentations. Similarly, using chroma-based audio features results in harmony-based segmentations. We now indicate, how our cyclic tempograms can be applied to obtain tempo-based segmentations (using a simple two-class clustering procedure for illustration). In the following examples, we use low-dimensional versions of C_{60}^A and C_{60}^F based on $M = 15$ different tempo classes. In our first example, we consider the song *In The Year 2525* by Zager and Evans. This song starts with a slow intro and contains a slow interlude of the same tempo. The remaining parts (basically eight repetitions of the chorus section) are played in a different, much faster tempo. As can be seen in Fig. 3a, both cyclic tempograms, C_{60}^F and C_{60}^A , allow for separating the slow from the fast parts. As second example, we consider a recording of the first movement of Beethoven’s *Piano Sonata Op. 13 (Pathétique)*. After a dramatic *Grave* introduction, the piece continues with *Allegro di molto e con brio*. However, it returns twice to *Grave*—at the beginning of the development section as well as in the coda. Using a purely tempo-based segmentation, the occurrences of the three *Grave* sections can be recovered, see Fig. 3b. Here, in particular the autocorrelation tempogram C_{60}^A yields a clear discrimination. Finally, as a third example, we consider a piano version of Brahms’ *Hungarian Dance No. 5*, a piece with many abrupt changes in tempo. This property is well reflected by the cyclic tempograms shown in Fig. 3c. In particular, the Fourier tempogram C_{60}^F separates well the slow middle part from the other, much faster parts.

In these three example, the cyclic tempograms yield musically meaningful segmentations purely based on a low dimensional representation of tempo. Actually, these segments can not be recovered using MFCCs or chroma features, since the homogeneity assumption does not hold with regard to timbre or harmony. In practice, various strategies based on different musical dimensions are needed to cope with the richness

and diversity of music [3]. For the future, we integrate our concept of cyclic tempo features into a hierarchical segmentation and structure extraction framework. Furthermore, having low-dimensional tempo features (in the order of the 12-20 dimensions of chroma and MFCCs), makes it possible to employ index-based range and nearest neighbor searches, which is important in view of efficient music retrieval. Finally, we closer investigate the musical meaning of tempo harmonics and subharmonics as well as the characteristics and combinability of the various tempograms.

5. REFERENCES

- [1] Meinard Müller, *Information Retrieval for Music and Motion*, Springer Verlag, 2007.
- [2] Mark A. Bartsch, Gregory H. Wakefield, “Audio thumbnailing of popular music using chroma-based representations,” *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 96–104, 2005.
- [3] Kristoffer Jensen, “Multiple scale music segmentation using rhythm, timbre, and harmony,” *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 159–159, 2007.
- [4] Eric D. Scheirer, “Tempo and beat analysis of acoustical musical signals,” *J. Acous. Soc. of America*, vol. 103, no. 1, pp. 588–601, 1998.
- [5] Daniel P. W. Ellis, “Beat tracking by dynamic programming,” *J. New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.
- [6] Geoffroy Peeters, “Time variable tempo detection and beat marking,” in *Proc. ICMC*, Barcelona, Spain, 2005.
- [7] Peter Grosche, Meinard Müller, “Computing predominant local periodicity information in music recordings,” in *Proc. IEEE WASPAA*, New Paltz, New York, USA, 2009.
- [8] Juan P. Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, Mark B. Sandler, “A tutorial on onset detection in music signals,” *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.
- [9] Ruohua Zhou, Marco Mattavelli, Giorgio Zoia, “Music onset detection based on resonator time frequency image,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1685–1695, 2008.
- [10] Anssi P. Klapuri, Antti J. Eronen, Jaakko Astola, “Analysis of the meter of acoustic musical signals,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 342–355, 2006.
- [11] Frank Kurth, Thorsten Gehrmann, Meinard Müller, “The cyclic beat spectrum: Tempo-related audio features for time-scale invariant audio identification,” in *Proc. ISMIR*, Victoria, Canada, pp. 35–40, 2006.
- [12] Mark Levy, Mark Sandler, “Structural segmentation of musical audio by constrained clustering,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 318–326, 2008.