American Journal of Applied Sciences

Article

Enhancement of Conventional Beat Tracking System Using Teager–Kaiser Energy Operator †

Matej Istvanek [1,*], Zdenek Smekal [1], Lubomir Spurny [2] and Jiri Mekyska [1]

[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 12, 61600 Brno, Czech Republic

[2] Department of Musicology, Faculty of Arts, Masaryk University, Janackovo namesti 2a, 60200 Brno, Czech Republic

Abstract: Beat detection systems are widely used in the music information retrieval (MIR) research field for the computation of tempo and beat time positions in audio signals. One of the most important parts of these systems is usually onset detection. There is an understandable tendency to employ the most accurate onset detector. However, there are options to increase the global tempo (GT) accuracy and also the detection accuracy of beat positions at the expense of less accurate onset detection. The aim of this study is to introduce an enhancement of a conventional beat detector. The enhancement is based on the Teager–Kaiser energy operator (TKEO), which pre-processes the input audio signal before the spectral flux calculation. The proposed approach is first evaluated in terms of the ability to estimate the GT and beat positions accuracy of given audio tracks compared to the same conventional system without the proposed enhancement. The accuracy of the GT and average beat differences (ABD) estimation is tested on the manually labelled reference database. Finally, this system is used for analysis of a string quartet music database. Results suggest that the presence of the TKEO lowers onset detection accuracy but also increases the GT and ABD estimation. The average deviation from the reference GT in the reference database is 9.99 BPM (11.28%), which improves the conventional methodology, where the average deviation is 18.19 BPM (17.74%). This study has a pilot character and provides some suggestions for improving the beat tracking system for music analysis.

Keywords: beat tracking system; music analysis; music information retrieval; onset detection; spectral flux; string quartet music; Teager–Kaiser energy operator

---

1. Introduction

Onset time in audio signal analysis represents the time position of a relevant sound event: usually when a music tone is created. Onset detection functions are algorithms that capture onsets (onset time positions), and thus ideally all tones in audio recordings. They can create a representation or an evolution of onset structure in given time of particular audio recording. There are also offsets of tones (indicating the end time position of a tone in a signal), e.g., see [1,2], but beat tracking systems do not need such information to work properly. The conventional beat tracking system is usually based on the calculation of repetitiveness of the dominant components in an onset function (onset curve) and its output represents a temporal framework, i.e., time instances, where a person would tap when listening to the corresponding piece of music. That is why it is important to have a robust

and computationally effective onset detector. Calculation of the beat positions and global tempo (GT) is important for musicologists and the complex music analysis. With such automated systems, tempo and agogic changes can be measured much faster than only with manual approach alone. Thus, musicologists would have to spend less time correcting calculated beat positions. Therefore, we set a new parameter—the average deviation of reference beat positions to the calculated beat positions as the average beat deviation (ABD).

Most of the onset detectors are based on energy changes in spectra: the calculation of spectral flux. For bowed string instruments there is a method called SuperFlux that can suppress vibrato in an expressive performance and reduce the amount of false-positive detections [3]. Some methods enhance the spectral flux onset detection using logarithmic spectral compression and then compute the cyclic tempogram for a tempo analysis [4]. There is also a method that calculates tempograms using Predominant Local Pulse [5]. Besides, the onset detection and beat detection could be performed in several toolboxes and libraries such as Tempogram Toolbox [6], LibROSA [7], MIR Toolbox [8], etc. [9]. The state-of-art onset detectors are usually based on deep neural networks [10,11] using spectral components and parameters as their inputs. Beat detection systems contribute from the solid onset detectors, where periodicity is identified [6,8,12–14]. As in other MIR fields, neural networks are also used.

While onset detection in percussive music is considered to be highly accurate (already at MIREX 2012 conference [15], algorithms achieved F-measure values greater than 0.95 for percussive sounds), detection of soft onsets produced by bowed string or woodwind instruments is still challenging. Although a lot of improvements in onset detection have been made, no system is truly universal for all musical instruments and all types of music.

This work aims to enhance the conventional beat tracking system and to improve the tempo analysis methodology published in [16,17] using the more sophisticated approach of tempo structure creation based on the automated beat tracking system with the Teager–Kaiser energy operator (TKEO) included. This nonlinear energy operator is used, e.g., for the improvement of onset detection in EMG signals (electromyography) [18], to decompose audio into amplitude and frequency modulation components [19], for the detection of Voice Onset Time [20], or the highly efficient technique for LOS estimation in WCDMA mobile positioning [21]. So far there is no extensive study on the use of TKEO for the analysis of musical instruments.

Since we will focus on the detection of onsets of melody instruments with low-energy attacks, we will concentrate on the onset and beat detection method based on spectral changes. We have not chosen probabilistic models, because they are usually susceptible to noisy recordings, which can be a problem in the case of old recordings.

The rest of the paper is organised as follows: Section 2 describes the onset detection function, the Teager–Kaiser energy operator, the proposed enhancement of the conventional beat tracking system and the beat detection method. It shows, how is the TKEO changing the spectra and therefore the output onset detection. Then, it introduces the reference and the string quartet database used for the GT and ABD estimation. Furthermore, a possible application is shown and the system evaluation is defined. Results are reported in Section 3 and discussed in Section 4. Finally, conclusions are given in Section 5.

2. Dataset and Methods

2.1. Onset Detection

Usually, onset detection algorithms use some pre-processing steps to reduce redundant information and to improve detection accuracy. In this study, we propose a new method of

pre-processing based on the TKEO. The TKEO ($\Psi\{s(t)\}$) is a nonlinear energy operator that can be calculated using the following formula:

$$\Psi\{s(t)\} = \left(\frac{ds(t)}{dt}\right)^2 - s(t) \cdot \frac{d^2 s(t)}{dt^2}, \tag{1}$$

i.e., we compute the square of the first derivative (which denotes the square of the rate of signal change) and then subtract the signal multiplied by the second derivative (which determines the acceleration at that point). We speed up the temporal changes of the signal module by removing the slow changes because we consider the rate of change. It is known that the faster the time changes, the higher the frequency components appear in the spectrum. By taking the first derivative into account, we increase the magnitude of higher frequencies of the spectrum [22].

In our discrete approach, we firstly downsample the input signal $x[n]$ to 22,050 Hz. Next, we apply the TKEO, i.e., we calculate the corresponding discrete non-causal form:

$$\Psi[x[n]] = x^2[n] - x[n-1] \cdot x[n+1], \tag{2}$$

which creates an energy profile of the given audio sample. In comparison to the conventional squared energy operator, the TKEO takes into account also signal's frequency [23] and it can have negative values, e.g., see Figure 1. Differences in spectra for the same audio track (clarinet recording) are shown in Figure 2. It is interesting how the dominant spectral components have changed—the clarinet has naturally strong odd harmonics, but the TKEO has changed their magnitude.

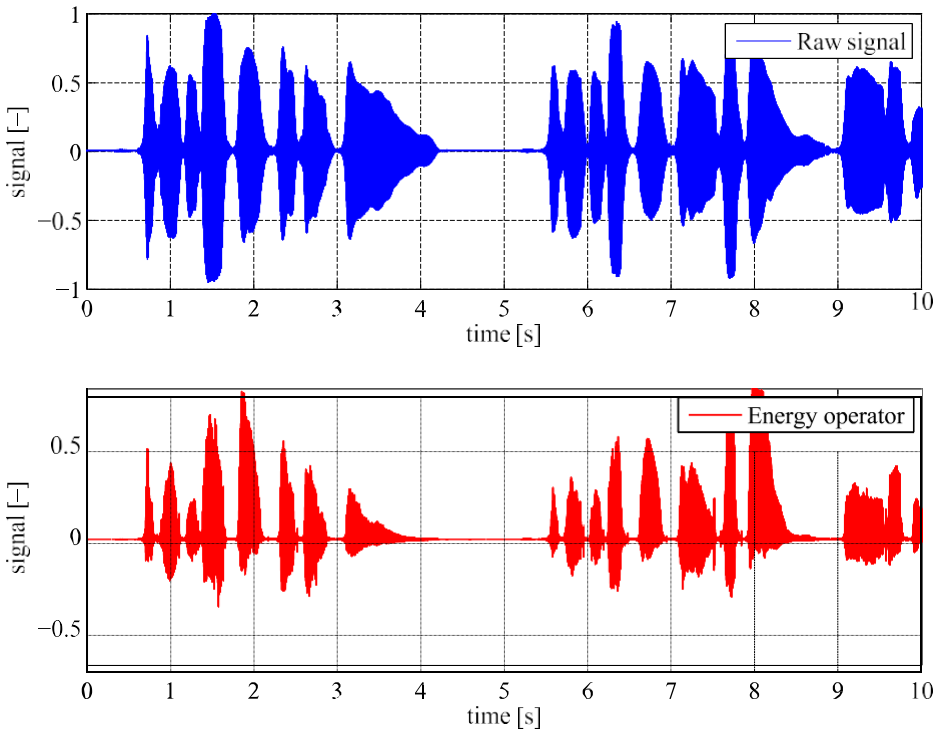

Figure 1. Original signal and the same signal after application of the TKEO pre-processing.

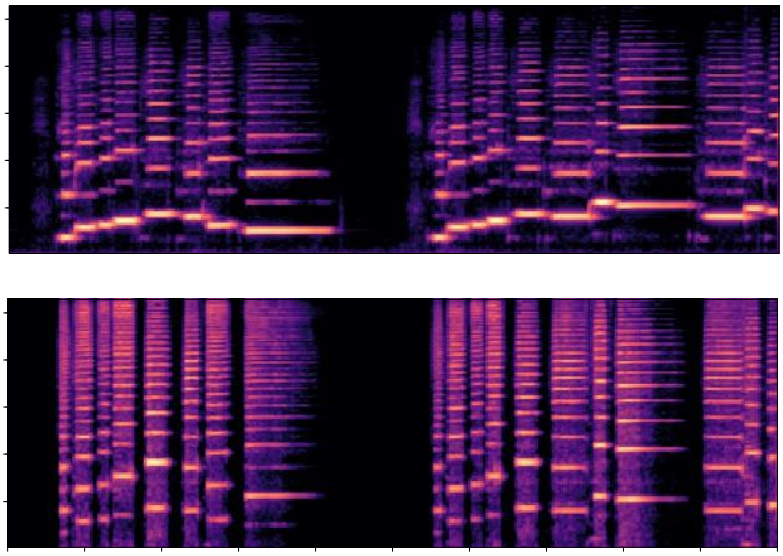

Figure 2. Spectrograms of the same clarinet recording—the second one is using a TKEO step.

In the following step, we calculate the onset envelope using the perceptual model. We use Short-Time Fourier Transform (STFT) with Hann window (hop-factor: 512 samples) and then the conversion to the perceptual model with log-power mel-frequency representation: 120 mel bands, max frequency at 10 kHz and min frequency at 27.5 Hz. We get the matrix $|X[m, k]|$, where m denotes the index of the frame and k the frequency bin or index of the mel band. These settings were inspired by SuperFlux calculation [3].

In the next step, we calculated the spectral flux. The basic version of spectral flux is defined as the l_1-norm of consecutive frames [24]:

$$SF[m] = \frac{1}{K} \sum_{k=0}^{K-1} H(|X[m+1, k]| - |X[m, k]|), \tag{3}$$

for $m = 0, 1, 2, \ldots, M-2$, where $H[x] = (x + |x|)/2$ is the half-wave rectifier, M is the number of frames, and K is half of STFT frequency bins, or number of mel bands. A half-wave rectifier is used to set negative values to zero and positive differences are summed across all frequency bands. Spectral flux gives us information, how energy in spectra changes in time. Finally, a peak-picking function is applied (default LibROSA settings) to identify time positions of onsets and therefore new tones in the audio signal.

An example of this system based on the mel-frequency representation, but without the use of TKEO, is shown in Figure 3. It represents a solo clarinet part. The onset function detected many false peaks and marked positions, where tones were not played. For comparison, Figure 4 shows the same signal, but in this case, pre-processed by the TKEO. The peak-picking function now marked all real onsets with better accuracy and without any false positive detection. The colorbar in dB (Figure 5) is presented separately because of the proper alignment of a spectrogram and onset function but is the same for all spectrograms (produced by matplotlib package) in this paper.

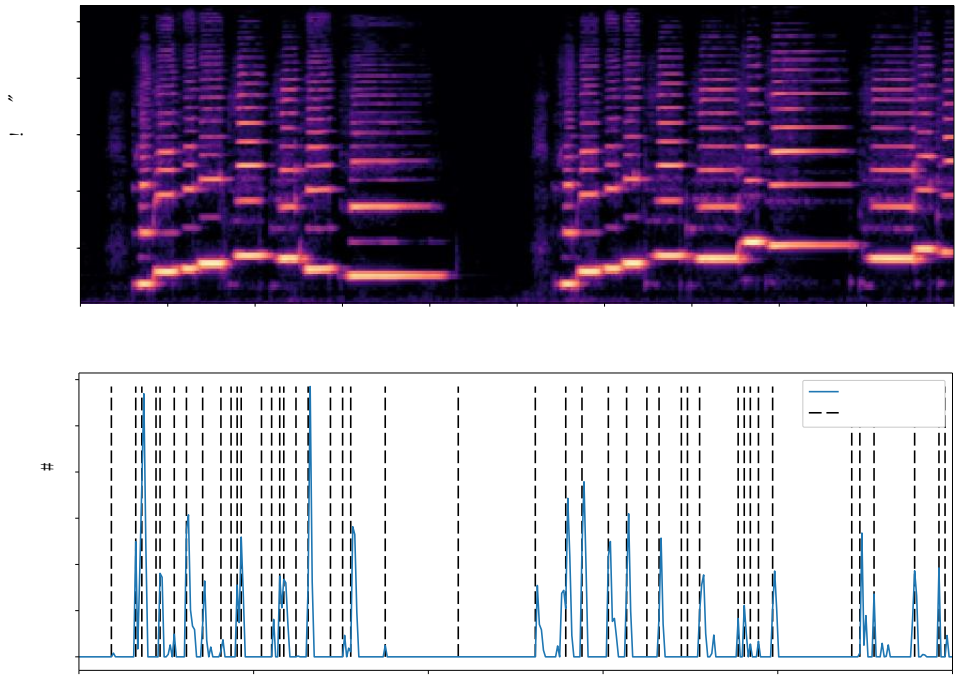

Figure 3. Spectrogram and onset detection function for a solo clarinet without the TKEO.

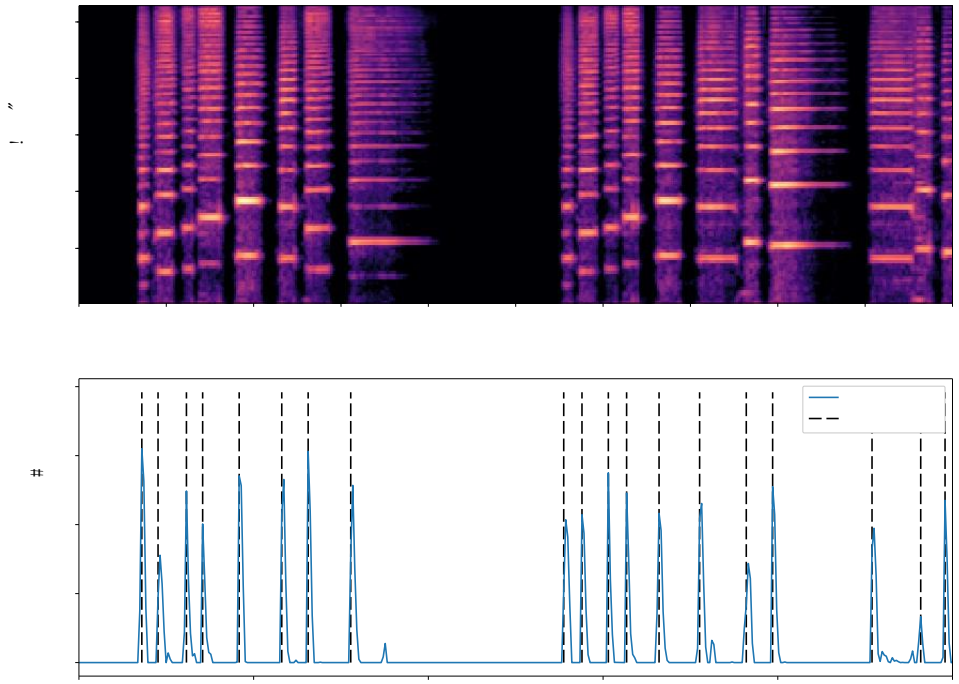

Figure 4. Spectrogram and onset detection function for a solo clarinet with the TKEO.

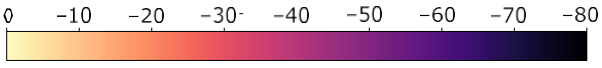

Figure 5. Colorbar.

As we can see on the second spectrogram (Figure 4), the energy in spectra changed, frequencies do not correspond properly to the original signal and new tones are sharpened and much more clear. We give this example for a good reason. Recording of a solo clarinet was the only audio track, in which the accuracy of the onset detection function was improved. Adding TKEO into this conventional detection method lowered the general detection accuracy. It decreased the number of detected false positives but also decreased the true positives. The cause of this phenomenon is explained in the following Section 2.2. We suggest that the general effect of the TKEO on onset detection function for woodwind instruments should be tested in more detail.

2.2. TKEO Influence

We applied the proposed method with the TKEO included on more recordings and observed, that in cases, where the tones are fast (e.g., violin playing thirty-second notes), or the energy difference is very low, this method does not detect every onset properly. Adding the TKEO increased the detection tolerance of fast changes in the signal. This means that the operator added additional "latency" to the signal values. It also decreased the ability of this system to capture low-energy spectral components. In general, fewer onsets were detected—only strong and more rhythmically important onsets remained. This is the advantage of the TKEO in the system. It suppresses less dominant spectral components and very fast tones even though onset detectors are usually set to do the opposite.

Figures 6 and 7 show another analysed track—a violin solo in a very fast tempo. There is a clear difference in spectrograms for the described detector and the same detection with the TKEO included. Most of the tones are quite visible in the spectrogram of the first figure. However, the system with the TKEO has its changes in the spectrum vaguer and blurry which means that onset function detected a lower number of onsets (especially between the 1st and the 4th second of this track). In this case, the conventional system detected more onsets correctly but that still does not indicate that estimation of GT would be also more accurate.

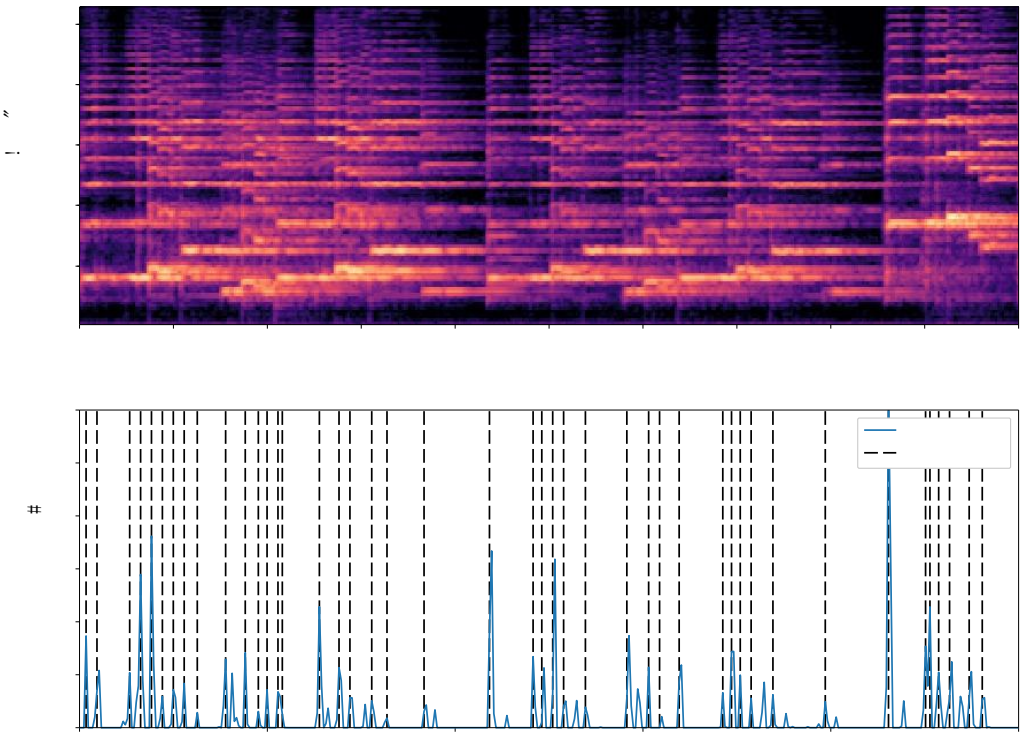

Figure 6. Spectrogram and onset detection function for a solo violin without TKEO.

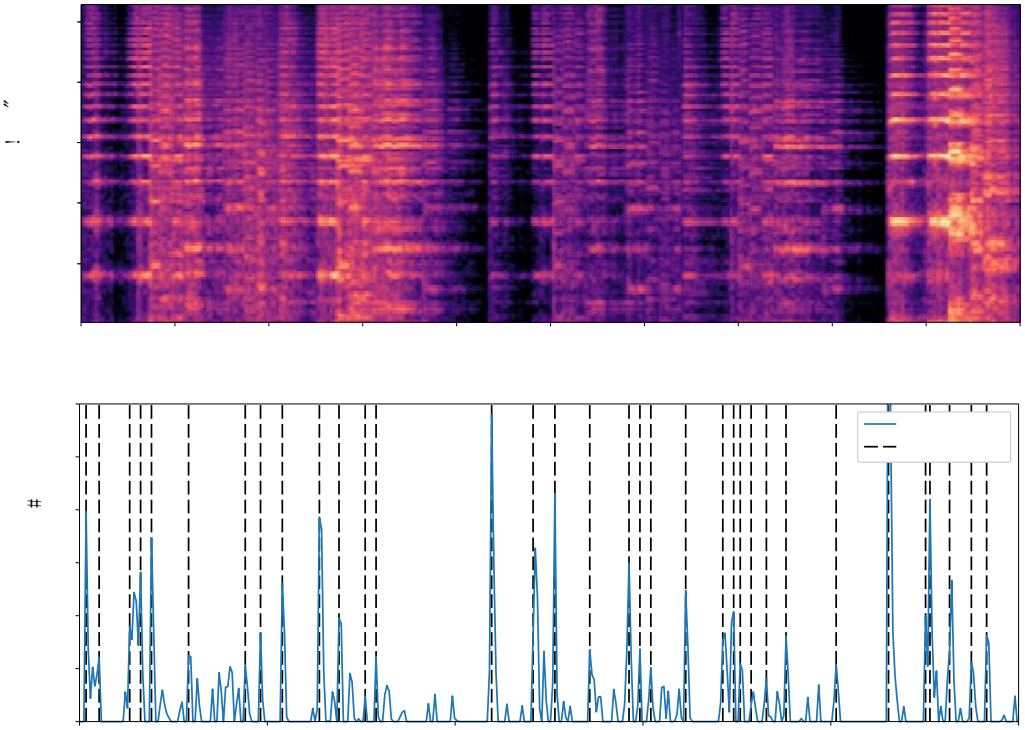

Figure 7. Spectrogram and onset detection function for a solo violin using TKEO.

2.3. Tempo Representation

To create a tempo structure of given recordings, we need a representation of tempo—how the density of onsets, or more precisely repetitiveness of significant onsets, is distributed. This can be done by several techniques, in this study we focused on the method of dynamic beat tracking system proposed in [12]. This system estimates beat positions in an onset envelope and uses them to pick the right peaks within a given interval (default tempo). The default tempo is set up before the calculation (or it is calculated automatically based on autocorrelation function with respect to the standard 120 BPM) and therefore it has to be estimated by listening to the particular audio track or estimated from the sheet music to work as we want. The calculated peak positions can deviate from the default tempo in adjustable boundaries (depends on settings, e.g., Ellis reports approximately 10% [12]). The parameter "tightness", which corresponds to the detection tolerance (from the default tempo), was set to the number 50 in all cases. At first, this looks like an inappropriate method for the varying tempo of string quartet music (second database), but with good parameterization and segmentation of particular motifs, it fits our need.

Beat detectors are based on a calculation of beats in an audio signal and therefore the metric structure from an elementary point of view. Usually, there is not enough information to consider dividing beats into bars without manual correction, but with proper segmentation, midi reference and dynamic time warping (DTW) techniques, this is possible [25]. However, one does not need such a method to calculate the GT of a given track. In this case, we only focused on the GT and ABD. Figure 8 shows how this system picks onset candidates from the onset curve and creates the beat positions by using periodicity information.

Figure 9 shows the estimated time positions of beats at the beginning of a string quartet segment. As we can see, the system is using periodicity information to calculate beat positions even at places where no onsets are detected—in this specific part, second violin and viola are playing very quietly (and no onset is detected) and then a violin solo begins. Between the 6th and the 10th second of

this track, there are strong onsets in the calculated onset curve. Their periodicity information is then used to fill the gap in the silent part of this recording, which is one of the advantages of the dynamic programming search system.

The disadvantage of such a beat tracking system is the adjustable default tempo—the algorithm searches for beat positions within a given interval, but there is no guarantee that true beat positions exist within specified limits (also concerning the tolerance parameter). The reference global tempo can be misleading if the recording is rhythmically unstable or the tempo changes significantly over time. A similar problem exists in the metric pulse. If the system detected 100 BPM as the GT and the reference is 50 BPM, it does not mean that the system is completely wrong. That is why we also calculated the ABD.

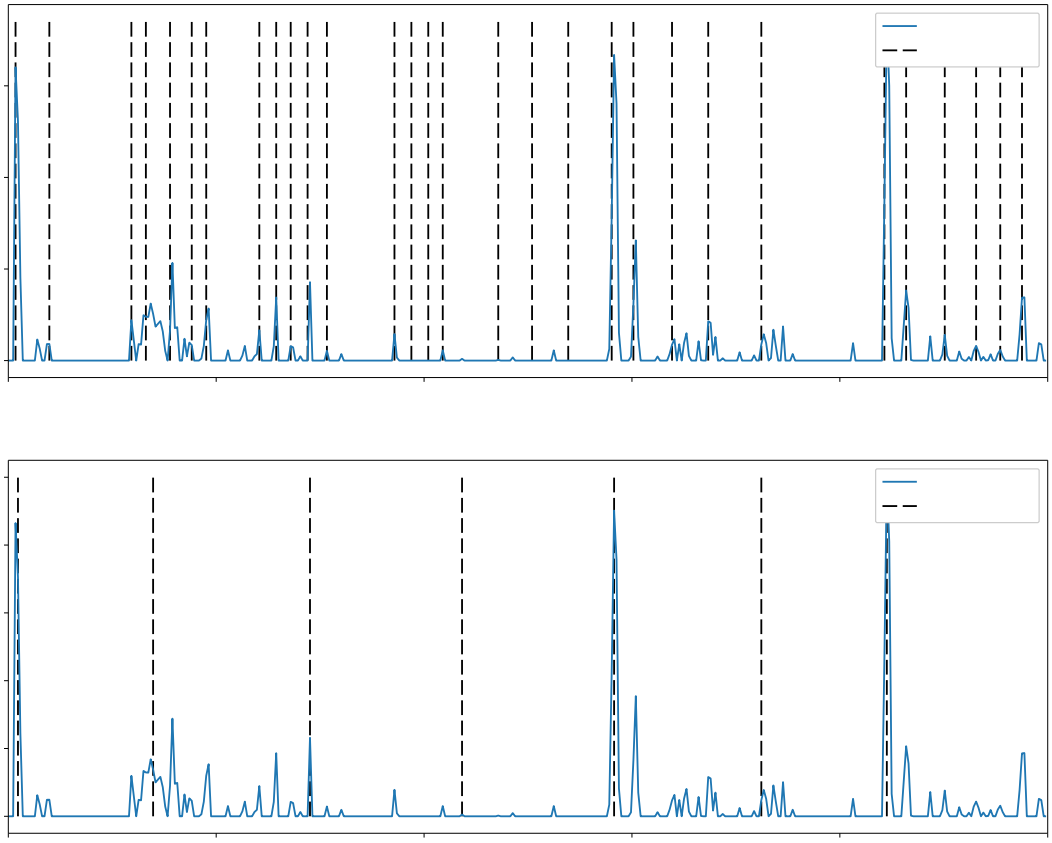

Figure 8. Comparison of the onset and beat positions.

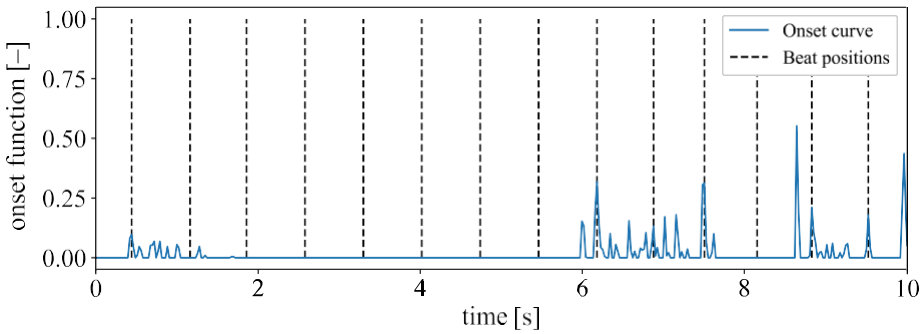

Figure 9. Estimated beat positions by the dynamic programming system.

2.4. Dataset

First of all, we tested whether the TKEO improves the estimation of the GT in general. The GT is the median of differences between the time positions of beats throughout the whole analysed track. For this purpose, we used the *SMC_MIREX* database [26], which consists of different recordings, from classical pieces to guitar solos. The recordings are sampled by 44.1 kHz. Their annotations contain manually corrected beat time positions, which will be used as a reference.

Music by string quartets is very specific because the tempo can be more or less stable but the musical ornaments, intended gaps, fermatas, or other expressive musical attributes can be present. Every musician has her/his own style of agogic performance. If we define meaningful musical parts by choosing important musical motifs, we can create segments that could be processed separately.

The second dataset consists of 33 different interpretations of *String Quartet No. 1 e minor "From My Life"*, composed by the Czech composer Bedrˇich Smetana. We also included two interpretations played by orchestra. We divided the first movement into six segments of musical motifs in the view of the musical meaning. The first movement consists of an introduction (Beg), exposition (A), coda (B), development (C), recapitulation (D) and the last coda (E). For every segment, we calculated the estimated average tempo (EAT), but without any expressive elements and information about beat positions, using a physical length of the tracks and information of rhythmic patterns in sheet music. The EAT will be used as a reference tempo for setting up the default tempo parameter in the beat tracking system. The first page of the sheet music is provided as an example in Appendix A.

2.5. Application

Beat tracking systems are used in the music analysis software for the complex tempo, timbre, dynamic or other music analysis. Example of such freeware software is Sonic Visualiser [27]. Figure 10 shows an example of tempo analysis of the string quartet music from the second tested database. The first pane is the visualisation of the audio wave, the second one is the spectrogram and the last one is a layer of manually corrected beat positions. Beat positions were calculated automatically by the beat tracking system called BeatRoot [28] (Vamp plugin) and then corrected by trained ears. The green line shows how tempo evolves in time—if the audio track is locally slowing down or the tempo increases. The method which is proposed in this paper has not been developed as a Vamp plugin for Sonic Visualiser.

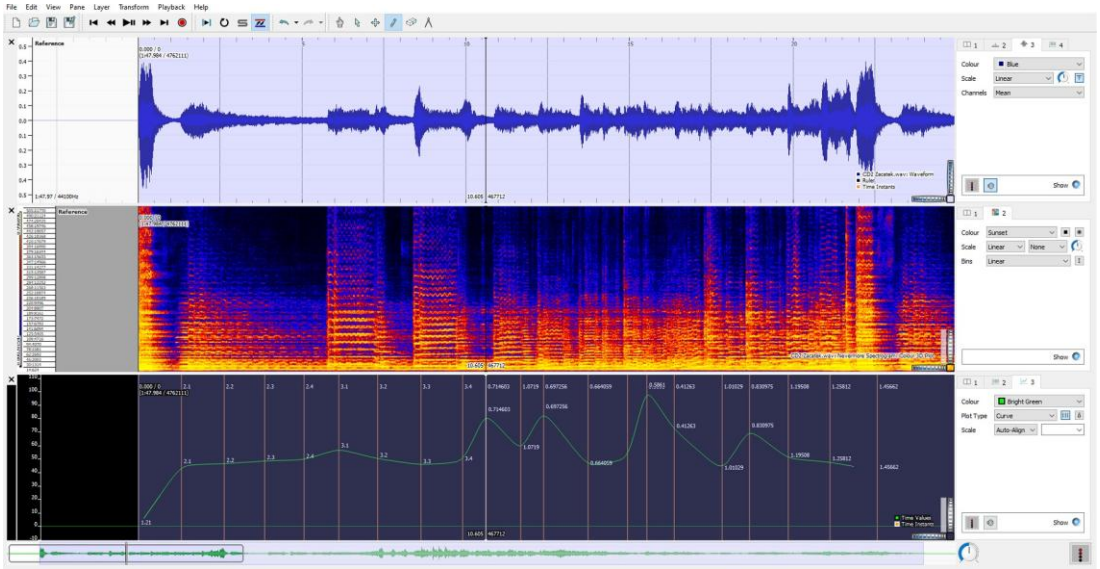

Figure 10. Possible application of the beat tracking system.

Musicologists can then draw conclusions from the measurement results. An automated beat tracking system is able to reduce the time of analysis significantly. For example, if we measure the EAT of the first motif of the second database for each recording, we get interesting results. One of the general assumptions is that presently we usually play the same piece of classical music faster than we did before. Figure 11 shows that this assumption may not be correct. There is a trend (see the slope of the linear regression line based on the sum of squares)—older recordings are on average at a faster pace. We do not have enough audio recordings to declare it as a fact, but the tendency is there. However, when we plot the EAT of the entire first movement (Figure 12), the tempo decrease is not so evident. Each black dot represents one interpretation and the blue line is a trend line. The sample from the year 1928 was an outlier and therefore we did not consider it in the regression analysis.

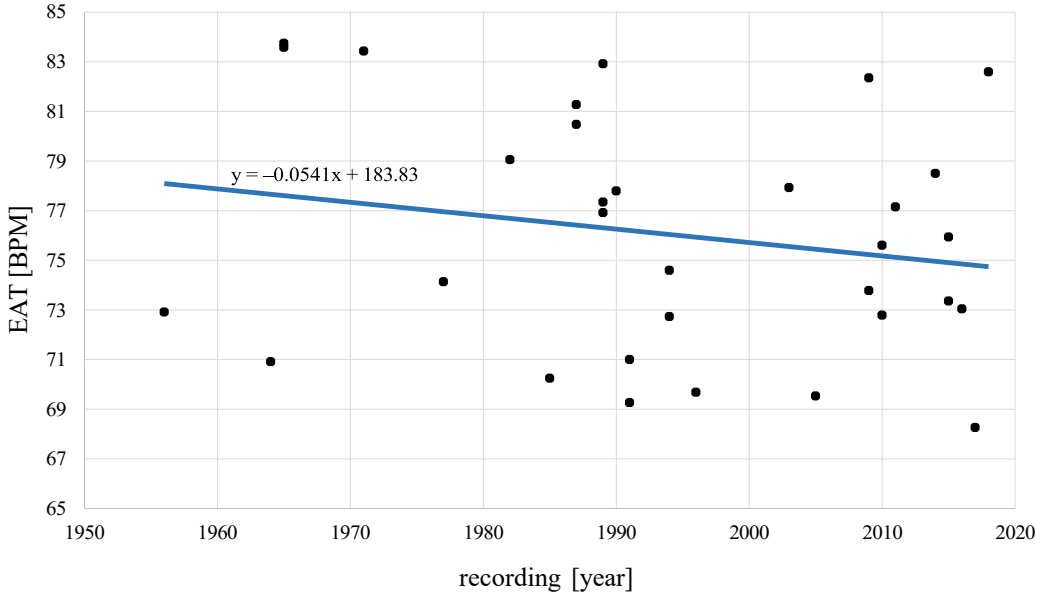

Figure 11. Results of the EAT calculation for the first motif of the string quartet database.

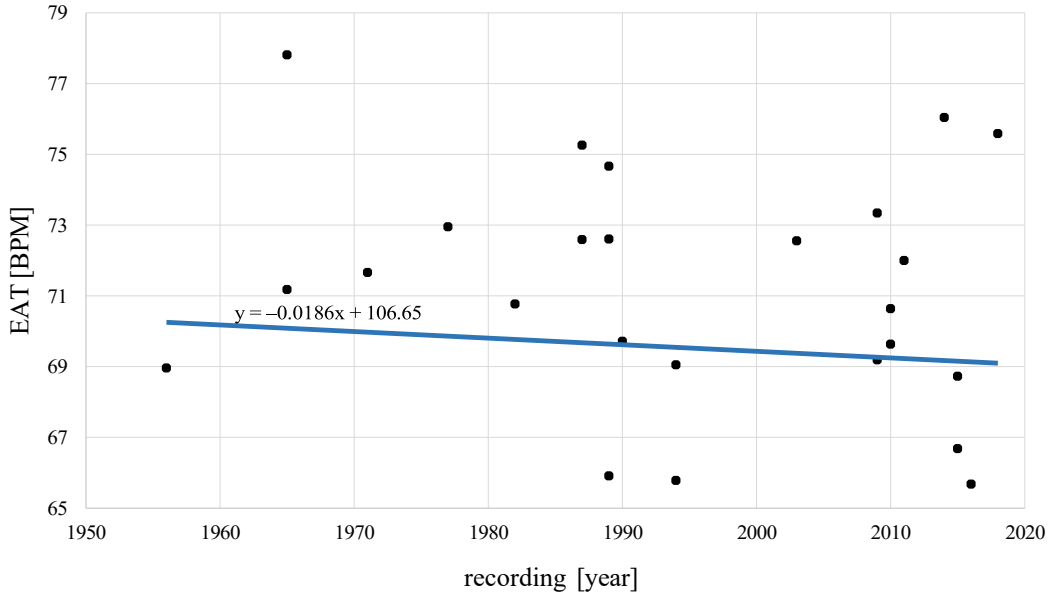

Figure 12. Results of the EAT calculation for the entire first movement of the string quartet database.

2.6. System Evaluation

During the analysis, we first used the reference dataset to determine the accuracy of the GT and ABD estimation. We computed the GT of each track by the proposed beat tracking method using both the proposed onset detection function (DS (default system)), and the same onset detection function with the TKEO (TS-system with the TKEO). Then we compared the reference values (annotation of the dataset) of each tested track with values estimated by the DS and the TS. The reference tempo was obtained as the number 60 (BPM definition) divided by the median of time differences between consecutive beat time positions. Then we calculated the median (Me) and the mean value ($\bar{x}$) of time differences of consecutive beats in all recordings and also in which the average was less than 1 s. This represents the ABD of tracks that were close to the reference tempo (some recordings achieved more than ~20 BPM difference in the GT when tested; they were excluded for the extended ABD testing).

Next, we analysed the string quartet database. First, all 33 recordings were divided into six segments with a relatively steady tempo and then all motifs were tested by the TS and the DS to estimate the GT. We computed the reference EAT of all segments of each interpretation (Table 1) by calculating the number of quarter notes (Table 2) and dividing them by the time length of each recording. The complete table is in Table A1. Finally, the EAT and the computed GT were compared. Systems were implemented using Python language (especially NumPy and LibROSA packages).

Table 1. The EAT of all motifs of the second database.

Track	Beg	A	B	C	D	E
CD01	80.61	69.37	34.41	88.56	55.60	74.50
CD02	77.80	69.03	44.14	81.84	59.52	72.43
CD03	77.93	73.19	41.60	87.09	62.36	79.14
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
CD33	76.92	63.24	42.05	74.62	56.26	68.31

All values are in BPM—Beats Per Minute.

Table 2. Calculation of quarter notes in all motifs.

Motif	Beginning	A	B	C	D	E
Bars	1–70	71–110	111–118	119–164	165–225	226–262
Quarter notes	280	160	32	184	244	148

3. Results

Table 3 presents results of the GT detection based on the first database for the first 30 analyzed tracks. The complete table is in Table A2. Average deviation from the reference tempo was 9.99 BPM (11.28%) in the case of TS and 18.19 BPM (17.74%) in the case of DS. The least accurate estimation was done on the recordings of a solo acoustic guitar. We also applied the *t*-Test (Paired Two Sample for Means) for each system (compared to the reference). P-value for the TS is 0.038 and 0.024 for the DS ($\alpha = 0.05$). Next, Table 4 presents general results of the GT testing: median, mean, standard deviation, relative standard deviation and variance for each tested system and the deviations from the reference tempo. The mean value of the reference GT was 76.78 BPM, the average computed GT 83.75 BPM for the TS and 88.97 BPM for the DS.

Table 5 shows the mean value and the median of the ABD testing for all analyzed tracks. The average difference between consecutive beat time positions of the reference and the TS was 2.30 s and 2.84 s for the DS. Table 6 shows the average of the arithmetic mean and the median of time

difference values of the recordings in which the ABD were less than 1 s. This means 11 recordings for the TS (37% of the tested database) and 9 recordings for the DS (30%). The TS detected the right metric pulse in more recordings than the DS. Average deviations from the reference beat positions were 0.39 s and 0.29 s for the TS and 0.95 s and 0.36 s for the DS respectively.

Table 2 presents the length of the first movement of each motif of the second database and the corresponding number of quartet notes. Then, the EAT was calculated. Table 1 contains results based on the EAT of all motifs of our second database—33 different interpretations of *String Quartet No. 1 e minor "From My Life"*. Finally, Table 7 shows the difference between the estimated GT and the EAT for both proposed systems. The complete table is in Table A3. The average deviation for the TS is 6.42 BPM and 6.59 BPM for the DS. Due to the nature of the results of the second dataset, no further statistical processing of the values was used.

Table 3. Reference GT and computed GT of the reference database.

Track No.	Reference (BPM)	TS (BPM)	DS (BPM)	TS Dev. (BPM)	DS Dev. (BPM)
1	48.15	47.85	47.85	0.30	0.30
2	66.99	73.83	73.83	6.84	6.84
3	68.00	95.70	95.70	27.70	27.70
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
30	63.36	63.02	63.02	0.34	0.34
Average	76.78	83.75	88.97	9.99	18.19
P-value		0.038	0.024		

TS—System with the TKEO; DS—Default system without the TKEO; Dev.—deviation from the reference global tempo; BPM—Beats Per Minute; P—*p*-value for the *t*-Test (Paired Two Sample for Means), $\alpha = 0.05$.

Table 4. Results of GT testing—the reference database.

Type	Me	$\bar{x}$	sd	rsd	var
Reference	77.11	76.78	33.01	0.43	1089.50
TS	82.05	83.75	37.30	0.45	1391.05
DS	76.07	88.97	41.05	0.46	1685.16
Dev. TS	5.31	9.99	15.75	1.58	247.93
Dev. DS	7.26	18.19	24.08	1.32	579.71

Me—median; $\bar{x}$—mean value; sd—standard deviation; rsd—relative standard deviation; var—variance; TS—System with the TKEO; DS—System without the TKEO; Dev.—deviation from the reference global tempo.

Table 5. Results of the ABD testing for all recordings.

	TS (s)	DS (s)
$\bar{x}$	2.30	2.84
Me	1.81	2.57
sd of the $\bar{x}$	1.90	2.17
sd of the Me	2.14	2.31

TS—System with the TKEO; DS—System without the TKEO; Me—median, $\bar{x}$—mean value; sd—standard deviation.

Table 6. Results of the ABD testing for recordings with the average ABD < 1 s.

	Dev. < 1 s in the Average of TS				<1 s in the Average of DS			
	TS		DS		TS		DS	
	$\bar{x}$	Me	$\bar{x}$	Me	$\bar{x}$	Me	$\bar{x}$	Me
Average	0.39	0.38	0.95	0.70	0.29	0.12	0.36	0.22

TS—System with the TKEO; DS—System without the TKEO; Dev.—deviation from the reference beat positions; Me—median; $\bar{x}$—mean value.

Figure 13 shows differences between the reference GT and calculated GT of the TS and DS of the first database. The TS generally follows the reference tempo more accurately mainly because it more often determined the correct metric pulse. The DS shows greater local deviations of the GT from the tested tracks.

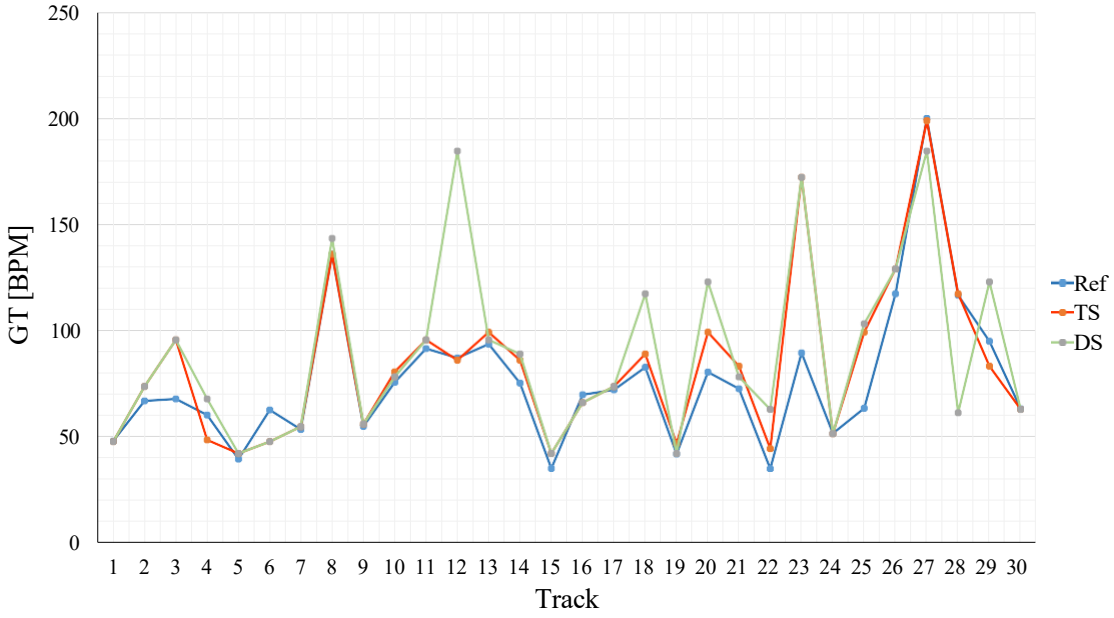

Figure 13. Visualisation of the GT computation—Ref, TS and DS estimation.

Table 7. Differences between the estimated GT and the EAT for both systems.

	TS						DS					
Track	**Beg**	**A**	**B**	**C**	**D**	**E**	**Beg**	**A**	**B**	**C**	**D**	**E**
CD01	15.09	13.98	6.61	3.73	5.92	3.80	8.49	22.92	6.61	3.73	1.82	3.80
CD02	14.49	0.81	6.53	1.51	6.74	3.57	14.49	9.27	2.84	1.51	3.50	1.40
CD03	14.36	1.41	7.15	5.20	4.94	6.99	11.17	10.16	11.13	5.20	13.64	6.99
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.		.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
CD33	3.83	6.60	9.63	0.79	5.26	3.47	3.83	6.60	9.63	0.79	11.74	5.52
Average	7.56	6.57	8.13	1.78	9.01	5.49	6.92	7.17	8.71	1.78	9.58	5.38
Result	6.42						6.59					

All values are in BPM—Beats Per Minute.

4. Discussion

Generally, the newly proposed method provided some improvements to the reference database. We analyzed 30 tracks and the results are reported in Table 4. The results suggest that the TKEO can help the proposed beat tracking system to pick better onset candidates for the beat positions and to slightly improve the GT calculation. The difference was about 8 BPM on average for all tested recordings of the first database. However, many recordings reported the same estimated GT for both methods. Then, the ABD was calculated. We used the reference database with manually corrected beat positions to determine the accuracy of both systems. We did not use F-measures, but rather average differences between consecutive beats. P values show that there is a difference between both systems. This gives us an idea of how close the beat tracking was to the reference positions. The system with the TKEO generally reported lower ABD for all settings used. The results suggest that the TKEO pre-processing improved the accuracy of the beat tracking system. This does not apply for the general onset detection function. Onset detection accuracy was reduced in most cases. The only exception was the recording of the clarinet.

As far as the string quartet database is concerned, the results were again slightly in favour of the system based on TKEO. All 33 recordings of the second database were tested. The difference between the average deviation from the EAT of the TS and the DS was only 0.17 BPM, and therefore both systems had more or less the same detection accuracy. We chose such complex music to see how the enhancement would deal with a very difficult task. The actual usefulness of the application also depends on the settings of selected parameters, not just on the TKEO pre-processing.

The idea of using TKEO in the pre-processing stage was to help the onset detection function to find more relevant onsets and therefore enhance the beat tracking system in terms of choosing better candidates for beat positions. It reduced the number of insignificant onsets detected. Onset detection accuracy has usually been reduced, but the final beat detection output may be more stable; the algorithm chooses from less and more important onsets. This is useful for analyzing tracks where we suspect a stable and non-agogic rhythm. We tested the effect of the TKEO to see how the output detection function would behave. We did not change the parameters such as tightness of the beat tracking system for each tested track; the correct setting (set for the particular piece of music) would yield better results for complex music analysis.

The limitation of this study is that the EAT in the string quartet database may be a reference value for the beat tracking system, but it is not the actual GT of a particular track since we cannot include any expressive elements in it. It does not provide any information about beat positions or local tempo changes. The same thing applies to the reference global tempo. In enhanced interpretation analysis, we need to track all beat positions in the segment and compare them to the real beat positions. However, in this case, we analysed relatively stable tracks with no abrupt tempo changes. In the future, we would like to use this system to create a database and its additional information about manually corrected beat positions of segmented string quartet music. The impact of the TKEO on audio recordings will be tested in more detail in our future work.

Cooperation between researchers and musicologists is the crucial part of such interdisciplinary projects and MIR science field. Different base knowledge and tendencies can lead to mutual misunderstandings, but both sides could benefit greatly from each other. Projects like these are the important bridge for computer scientists, MIR researchers and musicologists.

5. Conclusions

This study introduces an enhancement of the conventional beat tracking system by adding the TKEO into the pre-processing stage. It briefly describes the onset detection function and the beat tracking method with its possible application. The onset detection accuracy decreased in most analyzed tracks, but the accuracy of the GT detection and the ABD detection increased.

The influence of the TKEO was tested on different recordings and it was found, that in the case of woodwind instruments, the TKEO increased the onset detection accuracy. This phenomenon will

be studied in our future work. We would like to focus on the possible applications of the TKEO on music recordings in general. The TKEO is changing the magnitude of frequency components in a signal and acts as a filter. This could be the cause of increased onset detection accuracy, e.g., for the clarinet example.

The estimation of the GT was improved in the reference database. The average deviation from the reference GT in the reference database is 9.99 BPM (11.28%), which improves the conventional methodology, where the average deviation is 18.19 BPM (17.74%). P-values indicate that there is a clear difference between proposed systems. Both systems were also tested on the string quartet database. In this case, however, the results are not convincing. The proposed TS will be further used in the subsequent music analysis of the string quartet database. The aim is to create an automated system for capturing beat positions that are as close as possible to the actual beat positions in the recordings even for the complex music such as string quartet. In this way, it is possible to minimize the time required for manual processing and labelling. This study has a pilot character and provides some suggestions for improving the beat tracking system for music analysis.

Author Contributions: Conceptualization, M.I., Z.S. and L.S.; methodology, M.I.; software, M.I.; validation, M.I.; formal analysis, M.I., Z.S., L.S. and J.M.; investigation, M.I.; resources, M.I.; data curation, M.I.; writing—original draft preparation, M.I.; visualization, M.I.; supervision, Z.S., L.S. and J.M.; project administration, Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the grant LO1401. For the research, infrastructure of the SIX Center was used.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABD	Average Beat Deviation
BPM	Beats Per Minute
DS	Default System without the TKEO
EAT	Estimated Average Tempo
EMG	Electromyography
GT	Global Tempo
MIR	Music Information Retrieval
STFT	Short-Time Fourier Transform
TKEO	Teager–Kaiser Energy Operator
TS	System with the TKEO included

Appendix A. Sheet Music—The First Page of the First Movement

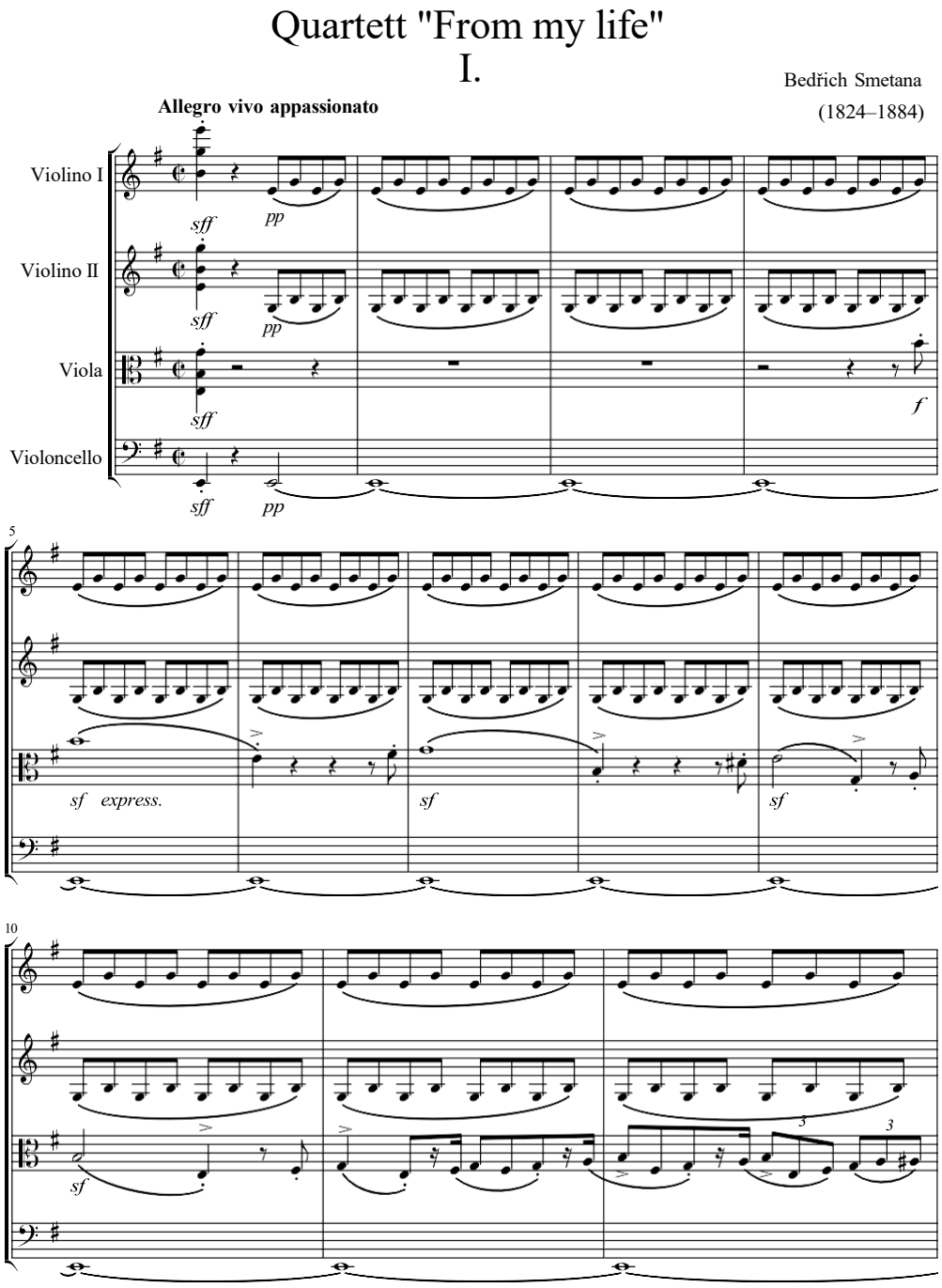

Figure A1. The beginning of the first movement.

Appendix B. Complete Tables and Results

Table A1. The EAT of all motifs of the second database.

Track	Beg	A	B	C	D	E
CD01	80.61	69.37	34.41	88.56	55.60	74.50
CD02	77.80	69.03	44.14	81.84	59.52	72.43
CD03	77.93	73.19	41.60	87.09	62.36	79.14
CD04	80.48	75.08	48.98	80.37	69.14	80.29
CD05	69.54	66.78	32.83	80.31	54.60	71.15
CD06	72.74	72.14	42.29	74.41	61.98	76.16
CD07	75.94	65.71	39.33	83.06	58.04	78.17
CD08	69.69	66.42	34.47	79.98	53.66	75.13
CD09	83.43	72.48	40.13	83.70	60.57	72.67
CD10	82.92	72.18	40.70	88.56	61.31	74.12
CD11	70.92	63.49	43.34	73.65	56.77	67.07
CD12	82.35	70.91	48.36	83.76	63.02	77.76
CD13	69.27	61.71	45.28	79.08	54.63	70.93
CD14	81.28	69.06	46.33	88.24	61.20	77.35
CD15	74.60	68.23	30.19	85.12	54.28	69.92
CD16	79.06	68.87	39.59	87.81	59.46	76.42
CD17	71.01	58.38	37.35	75.82	51.37	68.52
CD18	72.79	71.59	51.06	75.13	64.15	70.81
CD19	74.14	73.17	56.74	80.47	69.02	73.39
CD20	77.35	74.48	51.75	82.27	68.16	80.73
CD21	77.16	71.72	47.76	82.49	64.75	73.03
CD22	73.36	62.91	42.74	81.54	55.77	74.87
CD23	73.04	65.89	34.78	80.50	53.31	77.35
CD24	78.50	78.14	58.36	79.81	70.06	80.80
CD25	75.61	72.73	44.04	80.70	62.30	73.63
CD26	83.75	77.92	46.27	93.48	66.58	84.73
CD27	82.60	76.43	49.74	83.26	68.00	76.29
CD28	72.92	65.80	48.48	80.62	62.24	71.73
CD29	70.25	63.09	37.87	74.09	55.33	58.12
CD30	68.26	65.35	38.17	70.01	56.31	67.07
CD31	73.78	71.06	43.15	79.71	59.63	75.56
CD32	83.58	72.07	40.00	82.14	60.55	71.04
CD33	76.92	63.24	42.05	74.62	56.26	68.31

All values are in BPM—Beats Per Minute.

Table A2. Reference GT and computed GT of the reference database.

Track No	Reference (BPM)	TS (BPM)	DS (BPM)	TS Dev. (BPM)	DS Dev. (BPM)
1	48.15	47.85	47.85	0.30	0.30
2	66.99	73.83	73.83	6.84	6.84
3	68.00	95.70	95.70	27.70	27.70
4	60.41	48.75	68.00	11.66	7.59
5	39.71	42.36	42.36	2.65	2.65
6	62.76	47.85	47.85	14.91	14.91
7	53.67	54.98	54.98	1.31	1.31
8	136.05	136.00	143.55	0.05	7.50
9	55.15	56.17	56.17	1.02	1.02
10	75.86	80.75	78.30	4.89	2.44
11	91.63	95.70	95.70	4.07	4.07
12	87.27	86.13	184.57	1.14	97.30
13	93.75	99.38	95.70	5.63	1.95
14	75.38	86.13	89.10	10.75	13.72
15	35.34	42.36	42.36	7.02	7.02
16	70.01	66.26	66.26	3.75	3.75
17	72.20	73.83	73.83	1.63	1.63
18	82.87	89.10	117.45	6.23	34.58
19	41.99	46.98	42.36	4.99	0.37
20	80.65	99.38	123.05	18.73	42.40
21	72.73	83.35	78.30	10.62	5.57
22	35.09	44.55	63.02	9.46	27.93
23	89.71	172.27	172.27	82.56	82.56
24	51.81	51.68	51.68	0.13	0.13
25	63.56	99.38	103.36	35.82	39.80
26	117.46	129.20	129.20	11.74	11.74
27	200.00	198.77	184.57	1.23	15.43
28	116.73	117.45	61.52	0.72	55.21
29	95.09	83.35	123.05	11.74	27.96
30	63.36	63.02	63.02	0.34	0.34
Average	76.78	83.75	88.97	9.99	18.19
P-value		0.038	0.024		

TS—System with the TKEO; DS—Default system without the TKEO; Dev.—deviation from the reference global tempo; BPM—Beats Per Minute; P—*p*-value for the *t*-Test (Paired Two Sample for Means), $\alpha = 0.05$.

Table A3. Differences between the estimated GT and the EAT for both systems.

	TS						DS					
Track	**Beg**	**A**	**B**	**C**	**D**	**E**	**Beg**	**A**	**B**	**C**	**D**	**E**
CD01	15.09	13.98	6.61	3.73	5.92	3.80	8.49	22.92	6.61	3.73	1.82	3.80
CD02	14.49	0.81	6.53	1.51	6.74	3.57	14.49	9.27	2.84	1.51	3.50	1.40
CD03	14.36	1.41	7.15	5.20	4.94	6.99	11.17	10.16	11.13	5.20	13.64	6.99
CD04	0.27	0.92	34.37	0.38	17.16	3.06	0.27	3.22	31.77	0.38	9.16	3.06
CD05	4.29	3.06	4.62	0.44	10.00	7.15	8.76	11.52	10.24	3.04	21.40	4.89
CD06	3.26	6.16	8.38	1.59	4.28	4.59	3.26	6.16	5.56	1.59	14.02	4.59
CD07	7.41	8.12	8.42	3.07	9.96	2.58	7.41	6.07	14.50	3.07	15.79	7.96
CD08	11.06	9.58	7.21	3.37	10.94	5.62	11.06	9.58	6.55	0.77	10.94	5.62
CD09	8.86	5.82	3.67	2.43	15.43	3.33	8.86	5.82	3.22	2.43	15.43	5.63
CD10	12.78	6.12	15.47	0.54	12.52	4.18	9.37	3.82	10.98	0.54	3.89	6.63
CD11	7.38	4.51	16.75	2.35	7.83	2.77	5.08	0.47	16.75	2.35	0.65	0.93
CD12	1.00	6.31	7.81	3.01	3.24	2.99	1.00	7.39	5.47	0.41	1.58	5.59
CD13	9.03	14.29	10.89	1.67	5.46	2.90	6.73	19.04	8.55	1.67	15.21	2.90
CD14	11.01	11.69	8.65	0.86	8.64	3.40	11.01	6.94	8.65	4.05	0.32	8.78
CD15	11.53	12.52	8.38	1.01	5.81	10.83	8.75	12.52	7.81	1.01	19.55	19.18
CD16	13.23	7.13	12.09	1.29	10.38	2.59	13.23	9.43	5.74	1.29	8.54	1.88
CD17	7.29	13.40	5.72	2.48	4.80	5.31	9.74	0.96	9.63	0.18	2.46	9.78
CD18	5.51	6.71	10.46	0.87	1.13	21.48	3.21	6.71	0.62	0.87	4.06	0.97
CD19	1.86	10.18	4.01	0.28	11.73	4.91	1.86	7.58	13.10	0.28	9.28	4.91
CD20	3.40	8.87	0.98	1.08	3.56	0.02	3.40	6.27	0.98	1.08	0.16	11.56
CD21	6.19	4.28	2.91	0.86	11.25	7.72	8.97	6.58	10.97	0.86	0.15	7.72
CD22	7.39	3.35	11.09	1.81	14.07	11.26	12.77	1.69	9.99	1.81	5.75	3.43
CD23	16.06	3.95	10.55	2.85	2.86	3.40	16.06	10.11	5.59	2.85	24.99	3.40
CD24	0.20	5.21	1.73	0.94	10.69	8.30	2.25	2.61	0.37	0.94	10.69	2.55
CD25	2.69	8.02	0.51	2.65	16.00	2.37	2.69	10.62	4.71	2.65	9.57	2.37
CD26	8.54	6.14	11.15	2.22	16.77	7.56	5.35	6.14	11.15	2.22	14.17	4.37
CD27	3.53	6.92	5.24	0.09	15.35	0.29	3.53	6.92	4.09	0.09	12.75	4.46
CD28	5.38	10.20	2.19	2.73	7.60	9.02	0.91	1.20	3.20	2.73	7.60	11.62
CD29	8.05	1.51	8.27	4.21	10.93	0.61	8.05	1.57	19.55	4.21	7.69	0.70
CD30	7.74	2.33	7.16	0.17	2.48	22.03	0.26	5.26	6.38	2.01	9.95	6.76
CD31	4.52	2.77	6.54	1.04	10.21	0.44	4.52	2.77	14.27	1.04	14.20	0.44
CD32	12.12	3.93	3.07	1.21	13.28	2.79	12.12	8.68	6.98	1.21	15.45	7.26
CD33	3.83	6.60	9.63	0.79	5.26	3.47	3.83	6.60	9.63	0.79	11.74	5.52
Average	7.56	6.57	8.13	1.78	9.01	5.49	6.92	7.17	8.71	1.78	9.58	5.38
Result	6.42						6.59					

All values are in BPM—Beats Per Minute.

References

1. Benetos, E.; Dixon, S. Polyphonic music transcription using note onset and offset detection. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011.
2. D'Amario, S.; Daffern, H.; Bailes, F. A new method of onset and offset detection in ensemble singing. *Logop. Phoniatr. Vocol.* **2019**, *44*, 143–158. [CrossRef] [PubMed]
3. Böck, S.; Widmer, G. Maximum filter vibrato suppression for onset detection. In Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, 2–5 September 2013.
4. Grosche, P.; Müller, M. Cyclic Tempogram—A Mid-level Tempo Representation for Music Signals. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, TX, USA, 14–19 March 2010.
5. Grosche, P.; Müller, M. Extracting Predominant Local Pulse Information from Music Recordings. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 1688–1701. [CrossRef]

6. Grosche, P.; Müller, M. Tempogram Toolbox: MATLAB tempo and pulse analysis of music recordings. In Proceedings of the 12th International Conference on Music Information Retrieval, Miami, FL, USA, 24–28 October 2011.
7. LibROSA. Available online: https://librosa.github.io/librosa/ (accessed on 3 January 2020).
8. Lartillot, O.; Toiviainen, P. A Matlab Toolbox for Musical Feature Extraction From Audio. In Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Breisgau, Germany, 7–9 March 2007.
9. Zapata, J.R.; Davies, M.E.P.; Gómez, E. Multi-feature beat tracking. *IEEE Trans. Audio Speech Lang. Process.* **2014**, *22*, 816–825. [CrossRef]
10. Schlüter, J.; Böck, S. Improved musical onset detection with convolutional neural networks. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 4–9 May 2014.
11. Eyben, F.; Böck, S.; Schuller, B.; Graves, A. Universal onset detection with bidirectional long-short term memory neural networks. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, The Netherlands, 9–13 August 2010.
12. Ellis, D. Beat Tracking by Dynamic Programming. *J. New Music Res. Spec. Issue Beat Tempo Extr.* **2007**, *36*, 51–60. [CrossRef]
13. Böck, S.; Krebs, F.; Widmer, G. A Multi-Model Approach To Beat Tracking Considering Heterogeneous Music Styles. In Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, 27–31 October 2014.
14. Srinivasamurthy, A. A Data-Driven Bayesian Approach to Automatic Rhythm Analysis of Indian Art Music. Ph.D. Thesis, Pompeu Fabra University, Barcelona, Spain, October 2016.
15. 2012:MIREX Home. Available online: https://www.music-ir.org/mirex/wiki/2012:MIREX_Home (accessed on 3 January 2020).
16. Cook, N. *Beyond the Score: Music as Performance*; Oxford University Press: Oxford, UK, 2013.
17. Bowen, J.A. Tempo, duration, and flexibility: Techniques in the analysis of performance. *J. Musicol. Res.* **1996**, *16*, 111–156. [CrossRef]
18. Solnik, S.; DeVita, P.; Rider, P.; Long, B.; Hortobágyi, T. Teager-Kaiser Operator improves the accuracy of EMG onset detection independent of signal-to-noise ratio. *Acta Bioeng. Biomech.* **2008**, *10*, 65. [PubMed]
19. Kopparapu, S.K.; Pandharipande, M.; Sita, G. Music and vocal separation using multiband modulation based features. In Proceedings of the Symposium on Industrial Electronics and Applications (ISIEA), Penang, Malaysia, 3–5 October 2010.
20. Das, S.; Hansen, J. Detection of Voice Onset Time (VOT) for Unvoiced Stops (/p/, /t/, /k/) Using the Teager Energy Operator (TEO) for Automatic Detection of Accented English. In the Proceedings of the 6th Nordic Signal Processing Symposium (NORSIG), Espoo, Finland, 9–11 June 2004.
21. Hamila R.; Lakhzouri, A.; Lohan, E.S.; Renfors, M. A Highly Efficient Generalized Teager-Kaiser-Based Technique for LOS Estimation in WCDMA Mobile Positioning. *EURASIP J. Appl. Signal Process.* **2005**, *5*, 698–708. [CrossRef]
22. Kvedalen, E. Signal Processing Using the Teager Energy Operator and Other Nonlinear Operators. Master's Thesis, University of Oslo, Oslo, Norway, May 2003.
23. Dimitriadis, D.; Potamianos, A.; Maragos, P. A Comparison of the Squared Energy and Teager-Kaiser Operators for Short-Term Energy Estimation in Additive Noise. *IEEE Trans. Signal Process.* **2009**, *57*, 2569–2581. [CrossRef]
24. Müller, M. *Fundamentals of Music Processing*; Springer: Berlin, Germany, 2015; pp. 309–311.
25. Konz, V. Automated Methods for Audio-Based Music Analysis with Applications to Musicology. Ph.D. Thesis, Saarland University, Saarbrücken, Germany, 2012.
26. Holzapfel, A.; Davies, M.E.P.; Zapata, J.R.; Oliveira J.; Gouyon, F. Selective Sampling for Beat Tracking Evaluation. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 2539–2548. [CrossRef]
27. Sonic Visualiser. Available online: https://www.sonicvisualiser.org/ (accessed on 3 January 2020).
28. Dixon, S. An Interactive Beat Tracking and Visualisation System. In Proceedings of the 2001 International Computer Music Conference (ICMC 2001), Havana, Cuba, 7–22 September 2001.