Tempo and Time Signature Detection of a Musical Piece*

Daniel Kostrzewa[0000−0003−2781−3709] and Marek Zabialowicz

Silesian University of Technology
Department of Applied Informatics, Gliwice, Poland
daniel.kostrzewa@polsl.pl

Abstract. Tempo and time signature detection are essential tasks in the field of Music Information Retrieval. These features often affect the perception of a piece of music. Their automatic estimation unlocks many possibilities for further audio processing, as well as supporting music recommendation systems and automatic song tagging. In this article, the main focus is on building a two-phase system for extracting both features. The influence of many parameters of known methods was investigated. The results were also compared with the well-known and scientifically recognized Librosa library for music processing.

Keywords: Tempo · Time Signature · Music Information Retrieval · Audio Features · Sound Analysis · GTZAN.

1 Introduction

Music information retrieval (MIR) is a field that has become extremely popular among researchers since the beginning of the 21st century, becoming one of the most intensively developed in recent years. An increased interest due to more accessible access to music, primarily due to portable players and smartphones [18], as well as music streaming services and the competition of those for customers can be observed. In order to find similar songs, a service must acquire information about each song individually, and it must do so in a very short time.

Tempo and time signature detection are fundamental to digital audio processing and music information retrieval. It has applications for the previously mentioned streaming services and music producers who find marking bar lines helpful when editing a recording. Knowledge of tempo and meter is also needed by DJs creating live music, who need to align parallel tracks and samples, i.e., fragments of other songs, broadcasts, or recordings. Knowing a song's tempo and time signature can also help when trying to catalog it automatically. By having information about the artist, title, and genre, and additionally detecting tempo, time signature, and, for example, mood, it is possible to create a very informative database of songs. Such a database would be able to respond to a query based on the listener's tastes, current mood, or playlist destination.

* This work was supported by the Statutory Research funds of Department of Applied Informatics, Silesian University of Technology, Gliwice, Poland.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

The remainder of the paper is as follows. Section 1 describes the undertaken task, provides related works, and outlines the paper's contribution. Section 2 describes the tempo and time signature detection method in detail. The performed numerical experiments are presented in Section 3. Conclusions with the future work comprise Section 4.

1.1 Related Work

There are many papers and studies on tempo detection, but the algorithms described are very similar. Metre detection is an issue less frequently described in literature and articles [1].

The most commonly used method for tempo detection is to create a filterbank using the discrete Fourier transform (DFT). A filterbank can be created with the following, among others: a rectangular filter [4,16], with a finite impulse response (FIR) filter [8,10] or a filter with an infinite impulse response (IIR) [18]. Instead of a filter ensemble, the signal envelope [18] or the spectrogram flow function [14] can be calculated. The function calculated in this way changes more slowly than the original signal and is more suitable for processing. Compared to the filter ensemble, these methods allow for better results in music without a rhythm section.

One of the ways to detect tempo is to create a series of comb filters in which the pulsation frequency tells the beat frequency of the [4, 16] tempo, and then find the maximum sum of the products of the signal spectra from the filter ensemble and the filter [4,16]. The filter pulse frequency for which the maximum sum of energy was calculated is most likely the beat frequency of the pulses in the song, i.e., the tempo of the song. Another option for detecting tempo based on the comb filter is to use the autocorrelation function (ACF) [2, 8, 18]. ACF can also be used to detect pulses in a signal to calculate their probability of occurrence over time [10, 14].

Foote and Cooper [6] and Foote [5] in their paper presented an interesting way to represent a piece of music graphically. Using short-time Fourier transform (STFT) and spectrogram or Mel-frequency cepstral coefficients (MFCC), they create a two-dimensional matrix that represents the similarity between two excerpts of a song. This idea was used by Pikrakis et al. [15], who used the self-similarity matrix (SSM) created using MFCC to detect time signature and tempo. Gainza [7, 9] also used the SSM. Calculating a song's spectrogram creates an audio signal similarity matrix (ASM) and then a beat similarity matrix (BSM). Then, by processing the average values on the diagonals, it calculates how many rhythmic values are in a single bar, finally calculating what value is the base of the meter.

Böck et al. [3] approached the problem in a complex way, not only wanting to detect tempo, but also the position of beats over time. Improving the accuracy of detecting one value directly flows into improving the accuracy of detecting the other. To do this, they use machine learning and a temporal convolutional network (TCN). In order to extract features for learning by the algorithm, they process the spectrogram of the piece. In 2019, Davies and Böck [11] presented

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

a state-of-the-art way to tag bars in music based on TCN, improving on the previous [17] algorithm.

However, despite the significant development of artificial intelligence and deep learning, these methods still need improvement. Therefore, this paper focuses on improving the ability to detect tempo and time signature based on the beat similarity matrix.

1.2 Contribution

The contribution of the paper is two-fold. Primarily, we used algorithms known in the literature for both tempo and time signature detection to develop a two-phase detection system which is not known in the literature. Secondly, the influence of many different parameters was examined on the detection accuracy. Moreover, the tempo detection accuracy was compared to the well-known and recognized Librosa library.

2 Tempo and Time Signature Detection Method

Periodic local maxima can be observed in the graphical representation of a musical piece. These maxima are correlated with the pulse of the song and, therefore, with its tempo. Tempo detection on such a sample can be inaccurate due to overlapping sounds of multiple instruments. To compensate this effect, the signal of a piece of music is transferred to the frequency domain using a fast Fourier transform (FFT). The signal is then divided into subbands by band-pass filters. The divided signal forms a filter ensemble. Each of the spectra from the filter ensemble is transferred back to the time domain using the inverse fast Fourier transform (IFFT). However, the spectra cannot be directly transferred because the signal from which the original spectrum was created was not infinite. In addition, the second spectra of the filter set are discontinuous because they were extracted with a rectangular filter. For this reason, each spectrum must first be smoothed with a Hann time window and then transferred to the time domain. This representation contains some local maxima, but they are stretched out in time. In order to apply the comb filter, the signal has to be transformed into individual peaks. To do this, the derivative of the signal is calculated, where the value of each sample is the difference between this sample with the previous one. Figure 1 demonstrates the signal derivative of one of the filters. It clearly shows the regular beats of the song's pulse.

A comb filter signal is created for each tempo. The distance of the *step* filter pulses is calculated from the formula 1. The length of the comb filter signal is constant for all BPM values and is equal to the length of the song fragment from which the filter ensemble was created.

$$step = \frac{60}{bpm} * f_s \tag{1}$$

where: bpm - the rate expressed in beats per minute, f_s - the final sampling frequency of the signal.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

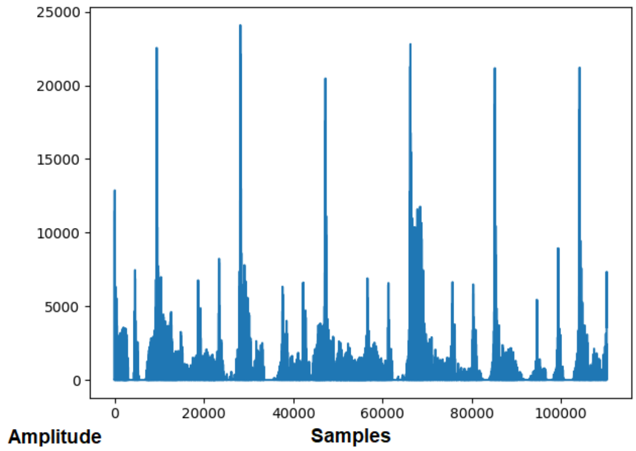

Fig. 1. Derivative of the smoothed signals from the second filter.

The filter signal for a given tempo is then transferred to the frequency domain. The calculated spectrum is successively multiplied by the spectra of the derived signals from the filter ensemble, and the products are summed. In this way, the sum of the energies of the products of the signal and filter derivative spectra is calculated. During the loop run, the maximum sum of energies is sought. The tempo corresponding to the maximum sum of energies is considered the tempo of the song.

In order to detect the time signature of a musical piece, it is necessary to know the tempo to extract the individual metrical values, divide the piece into bars and then calculate their similarity. The repetition of accents and similar bars within a song will enable effective time signature detection. The algorithm is based on the BSM method [8].

The first step is to calculate the spectrogram of a piece of music. To compute the spectrogram, the function *spectrogram* from the library *scipy.signal*[1] is used. The result is a spectrogram whose length of each segment equals a quarter note, the basic metric unit in a time signature with a base of 4. The spectrogram represents the intensity of the sound for frequencies between 0 and half the sampling frequency of the signal, which is usually 22 kHz or 11 kHz. Instruments belonging to the rhythmic group play at lower frequencies (around 6 kHz), so the frequencies of the spectrogram are truncated to this value.

The sounds in a piece of music can be divided into harmonic and rhythmic. The former are sounds from the group of melodic instruments (e.g., guitar, piano, human voice) that create the melody of a song. In contrast, rhythmic sounds are sounds from a group of rhythmic instruments that build the rhythm, providing the foundation and support for the melodic section. In a song, they deftly intertwine to form a coherent whole, but two characteristics distinguish them:

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.spectrogram.html

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

the frequency and the duration of the sound played. Nobutaka Dno et al. [13] proposed an algorithm to extract the components using the song's spectrogram and a median filter. The method is quite simple in its operation. It resembles the median filter image smoothing known from image processing methods, except that it does not use a square window but vertical and horizontal ones. Percussive components are obtained using a median filter with a vertical smoothing window. This is because the sounds of rhythmic instruments are short and can cover all frequencies. The final step of extracting the harmonic and rhythmic components is to apply binary masking and split the original spectrogram into two parts.

The next step is the calculation of the BSM bar similarity matrix. For the calculation of the distance between each frame, three distance metrics were used and tested [9]: Euclidean, cosine, and Kullback-Leibler. Using the aforementioned metrics, a bar similarity matrix is created. The distance between each frame of the spectrogram is calculated, the width of which is equal to a quarter note at the given rate. These distances are recorded into a two-dimensional BSM matrix.

To investigate the existence of similar metrical structures in a song, the diagonals of the previously obtained BSM matrix are examined. Each diagonal represents the similarity between sounds that are a different number of metric values away. This similarity can be measured by calculating the mean component of each diagonal of the BSM matrix. The function thus obtained is then inverted to build a function that shows the peaks on the diagonals where the components show maximum similarity.

Many candidates are considered in the presented approach to detect the time signatures. This includes simple and regular signatures, as well as complex and irregular ones. The resulting function gives the greatest weight to closely spaced quarter notes. The waveform of the function is presented in Figure 2. It clearly shows the maximum value for signature 5, so the time signature of the song is 5/4.

3 Quantitative experiments

3.1 Dataset

Most musical works are protected by copyright, so making them available in the public domain is challenging, even for research purposes. For this reason, most collections do not contain audio files but only meta-data about the contained works. A dataset called GTZAN [19] was used to test the effectiveness of tempo detection. It comes from Tzanetakis and Cook's paper [20]. It contains a collection of one thousand 30-second excerpts from songs. The collection also has a file where the tempo of each song is given in BPM. All songs are 16-bit audio files, sampled at 22050 Hz on one channel, saved in .wav format. This collection does not include title or artist information. Time signature information was also not included, so using our knowledge of music theory, 86 songs were drawn for which the time signature was empirically classified. Almost all the pieces are 4/4, and one is 3/4.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

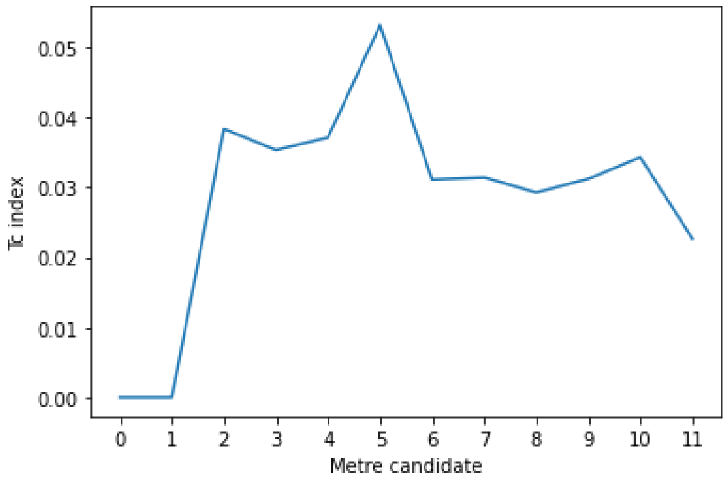

Fig. 2. An example of the time signature detection function.

The pieces included in GTZAN have little variation in time signature, so it was decided to expand the collection with an additional 14 pieces from the in-house collection. The proposal for the first ten items comes from [7]. The final collection of additional songs is presented in Table 1. 100 songs were collected to evaluate the time signature detection algorithm, with a significant prevalence of 4/4.

It has to be noted that we are aware that the skewness (imbalanced classes) of the dataset may lead to doubtful results. However, to the best of our knowledge, there is no well-known and well-prepared dataset for time signature detection. In fact, the author of this paper is working on preparing such a dataset.

3.2 Effectiveness Evaluation Criteria

Two criteria for evaluating tempo detection have been adopted in the research community – Accuracy1 and Accuracy2 [3,8]. Accuracy1 is more stringent, and with this criterion, only the case where the detected BPM value is equal to the original tempo of the song $\pm 2\%$ is considered a correct tempo detection. Accuracy2 allows for a margin of error. When the detected value is a multiple of the original tempo, such a result is also considered correct. This is because if a song is played at 70 BPM and a value of 140 BPM is detected, the pulse of the metronome set to 140 BPM will coincide with the pulse of the song, except that it will pound out eighth notes rather than quarter notes. An analogous situation occurs when a tempo of 70 is detected for a track played at 140 BPM.

Similar assumptions were made when evaluating the performance of time signature detection. With Accuracy1, only the case where an accurate measure of time signature is detected is considered correct. With Accuracy2, the arrangement of accents in the bar is ignored, and the basis of the bar measure is

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

Table 1. A collection of additional songs with different time signatures.

Title	Performer	Time signature	Tempo
Eleven	Primus	11/8	230
Windows To The Soul	Steve Vai	11/8	243
Watermelon In Easter Hay	Frank Zappa	9/4	55
ScatterBrain	Jeff Beck	9/8	250
Take It To The Limit	The Eagles	3/4	90
Doing It All For My Baby	Huey Lewis & The News	12/8	275
Forces... Darling	Koop	8/8	200
Sliabh	Danu	6/8	190
Money	Pink Floyd	7/8	120
Whirl	The Jesus Lizard	5/4	150
Take Five	Dave Brubeck	5/4	174
The Sky Is Red	Leprous	11/4	140
Hey Ya!	OutKast	11/4	158
Ageispolis	Aphex Twin	4/4	101

simplified to quarter notes. Thus, the values considered the same time signature are: 4/4, 2/4, 2/2, 8/8, 8/4, etc. For both tempo and time signature, pieces detected correctly for Accuracy2 are included in the results for Accuracy1.

In fact, the authors of this paper will focus mainly on the quantitative results of Accuracy2 in both cases (tempo and time signature detection) since the difference between the two can be difficult to discern, even for a musician or a skillful listener.

3.3 Tempo Detection

In testing the effectiveness of tempo detection, the effect of the length of the fragment of the analyzed song and the number of comb filter pulses on the results was examined. The results were also compared with those obtained using the popular Librosa library. Librosa is an open-source[2], free[3] library for audio and music analysis. It provides functions that can be used to develop a MIR system [12]. The library includes functions from digital signal processing, graphical representation of the signal, pulse, tempo analysis of the song, and extraction of various features.

An investigation of the effectiveness of tempo detection of a song began by looking at how the number of pulses of the comb filter affects the results obtained. A 25-second excerpt from each song was used to investigate this relationship, and a number of pulses ranging from 1 to 25 were used to create the signal. The results are presented in Figure 3.

The graph (Figure 3) shows that the maximum Accuracy2 performance was obtained for ten pulses. It is surprising to obtain an Accuracy2 of 50% already

[2] https://github.com/librosa/librosa

[3] https://librosa.org/doc/latest/index.html

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

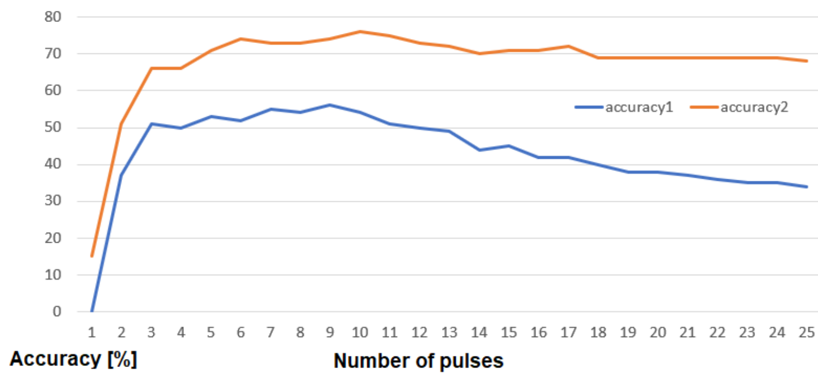

Fig. 3. The influence of the number of comb filter pulses on accuracy.

with only two pulses. With three pulses, the correct Accuracy2 result is obtained with two tracks out of three. The effectiveness increases with the number of pulses, up to a maximum Accuracy1 of 56% for nine pulses and Accuracy2 of 76% for ten pulses. Above ten pulses, the effectiveness decreases. This may be due to the fact that for comb filters with a higher number of pulses, after a while, the beats of the song begin to mismatch with the pulses of the comb filter. Accuracy2 is less susceptible to this relationship and shows a reasonably constant effectiveness of 69%. In contrast, Accuracy1, which for three and more pulses achieves a value of 50% already from 13 pulses, shows worse effectiveness than with three pulses, to finally achieve a result of only 30% with 25 pulses of the comb filter.

The next part of the study examined how the length of the analyzed song affected performance. It was observed that its length significantly influences the calculation time. In order to test this relationship, 100 test pieces were used with a fixed number of ten pulses.

The minimum tempo dictates the minimum duration of the excerpt the algorithm can detect. This value is 60 BPM, so each filter pulse is one second apart. This means that a comb filter signal with ten pulses at 60 BPM is needed to create a signal with a duration of at least 10 seconds. The results obtained are presented in Figure 4.

It can be seen from the graph (Figure 4) that, with a fixed number of ten comb filter pulses, the best effectiveness is provided by the use of a 26-second excerpt from a song. The accuracy increases steadily with a longer song fragment, but the increase is stopped at the 26-second value.

The research provided information on the optimal parameters of the proposed algorithm. In order to confront our method with other works, the results obtained were compared with the accuracy provided by the Librosa library, recognized in the scientific community. A 25-second excerpt and a comb filter with ten pulses were used to detect the tempo of each song.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

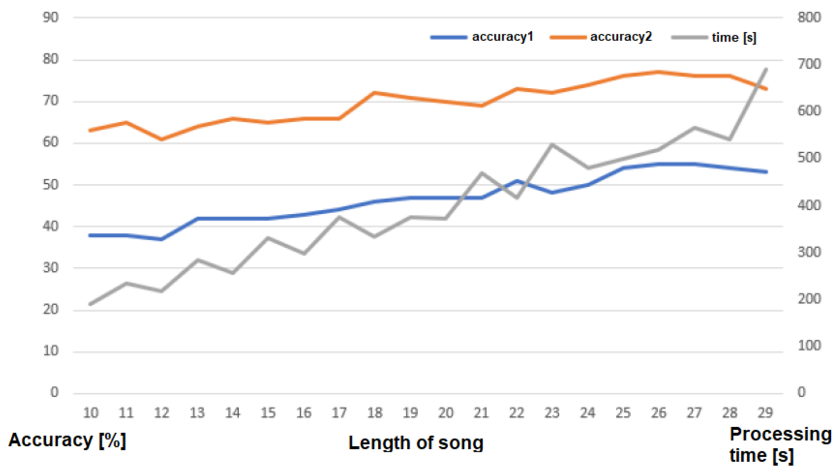

Fig. 4. The influence of the song's length on tempo detection effectiveness.

Table 2. Results of tempo detection with comb filter compared to Librosa library.

Comb filter		**Librosa**	
Accuracy1 [%]	Accuracy2 [%]	Accuracy1 [%]	Accuracy2 [%]
42,8	74,7	42,6	60

Comparing the results obtained with those obtained by using the Librosa library (Table 2), it can be seen that both methods provide the same performance for the Accuracy1 criterion. However, a significantly better result was obtained for the Accuracy2 criterion (by almost 15 percentage points). Given its popularity in the scientific community, it is surprising to see such a low detection accuracy rate from the Librosa library.

3.4 Time Signature Detection

When investigating the effectiveness of time signature detection, it was examined how the length of the excerpt from the piece being analyzed, the frequency range of the spectrogram, and the metrics of determining distance affect the results.

The research began by examining how the cut fragment's length affects a musical piece's meter detection performance. As with tempo detection, the processing time is significantly influenced. The songs from the GTZAN collection are only 30 seconds long, so they cannot be used to test how a longer excerpt impacts the effectiveness of the song's time signature analysis. The results were conducted for all 100 songs and fragment lengths ranging from 5 to 30 seconds (Figure 5).

It can be seen from the graph (Figure 5) that the best results were achieved for a 25-second fragment of the song. Therefore, it was decided that this length would be chosen in the final algorithm proposal.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

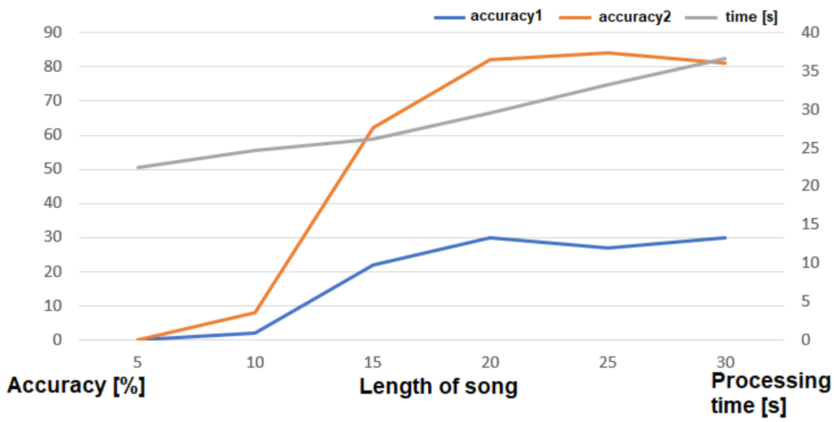

Fig. 5. The influence of fragment length on time signature detection performance.

The maximum performance for the 25-second fragment is encouraging in terms of execution time. The time required to perform the calculations shows linear complexity. Analysis of longer fragments would have taken even longer, and the efficiency of such a solution would have been negligible.

It was also examined how changing the frequency range of the spectrogram affected the results obtained and the time taken to analyze the entire collection. For this purpose, all 100 tracks were used, and, working on their 25-second excerpts, the efficiency and execution time of the computations were measured for a maximum frequency in the range from 3 to 12 kHz. The results are shown in Figure 6.

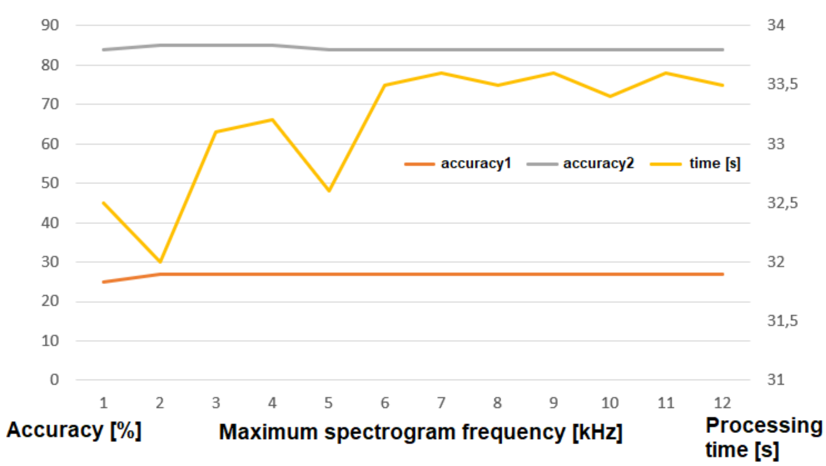

Fig. 6. The impact of reducing the frequency range of the spectrogram on the accuracy.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

Studies have shown that reducing the frequency range of the spectrogram has a negligible effect on the efficiency of time signature detection and the time required for calculation. A deterioration in efficiency of one percentage point (in terms of Accuracy2) from a frequency of 5 kHz can be observed, so in the final algorithm, the spectrogram frequency range is reduced to 4 kHz. Reducing the frequency range of the spectrogram has no significant impact on the computation time required.

Three distance measures were used in [10] when creating the BSM matrix: the Euclidean distance (EDM), the cosine distance (CD), and the Kullback-Leibler distance (K-L). It was tested what efficiency is provided by implementing each of these measures on a set of all tracks. The results are presented in Figure 7.

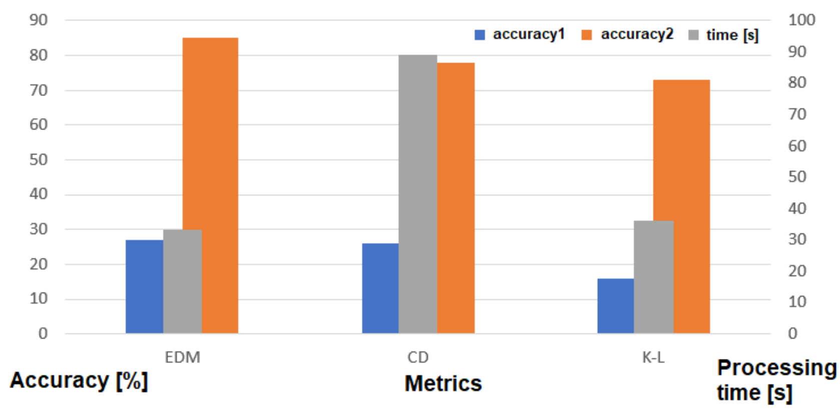

Fig. 7. The influence of the distance measure on the accuracy of time signature detection.

The results show that the Euclidean distance is the most accurate method for determining distances. Not only does the EDM show the best accuracy, but it is also supported by the shortest calculation time. As in the case of the analysis of the effect of fragment length on the accuracy, the results show that the most efficient method is also the fastest, so the proposed algorithm is correct in its performance and optimal in terms of calculation time.

The experiments delivered the most optimal and effective tempo and time signature detection algorithm. However, tests of the effectiveness of the time signature detection were carried out using the tempo information of the song taken from the test set. To test the effectiveness of the overall system and its two-stage operation, a 100-element set of songs and previously tested parameters were used to provide the maximum tempo and time signature detection performance. The results are presented in Table 3.

The presented system provides an efficiency of approximately 70% for both tempo and time signature detection in terms of Accuracy2. During a review of

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

Table 3. Results obtained in detecting the tempo and time signature of a song.

Tempo		**Time signature**	
Accuracy1 [%]	Accuracy2 [%]	Accuracy1 [%]	Accuracy2 [%]
46	70	21	74

the available literature, no work was found that proposed a system that not only detects tempo but also time signature.

4 Conclusions and Future Work

This work presents an algorithm that effectively detects a musical piece's tempo and time signature. A test set was completed to develop and test the effectiveness. The next step was to implement the adopted algorithm and study the dependencies. The study made it possible to optimize the parameters in terms of the efficiency achieved and the time needed to process the piece.

Overall, the gathered results for both tempo and time signature detection are satisfactory. Surprisingly, the proposed and optimized method for tempo detection achieved better results than the well-known and widely used Librosa library. Moreover, it has to be mentioned that time signature detection is one of the most challenging tasks in the Music Information Retrieval domain. That is the reason why only a very few papers were written on that topic.

Further work needs to start with completing a comprehensive dataset, which has not been created to date. The GTZAN collection does not have information on the song's time signature, and the song excerpts included are only 30 seconds long. However, the research conducted in this work has shown that even shorter excerpts provide better performance.

It would be ideal to empirically classify the time signature of all the pieces included in GTZAN or any other dataset. Such a procedure would take much time, but this work seems necessary to create the best possible system. It would be interesting to see if different genres of music require different parameters to detect tempo and time signature more accurately. Creating an ensemble of classifiers, where each genre had its parameters, could achieve a better result. Earlier detection of a song's genre based on, e.g., MFCC features and a neural network would provide the needed information, and such knowledge can be further used in a more comprehensive MIR system.

The proposed tempo detection method using a comb filter is acceptable but needs further research. An efficiency of 70% is a result that is too small and takes too long to process. It is possible to consider what results will be obtained from using other types of filters (e.g., triangular ones) to create an ensemble of filters.

The results obtained are satisfactory and motivate further work on this problem. The time signature detection method using the BSM matrix still has a very high potential. The methods and results presented in this work can be incorporated into a broader MIR system, which will be able to detect the genre of a

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

song based on knowledge of tempo and time signature, as well as automatically tag and recommend songs for users of music services.

References

1. Abimbola, J., Kostrzewa, D., Kasprowski, P.: Time signature detection: A survey. Sensors **21**(19), 6494 (2021)
2. Alonso, M., David, B., Richard, G.: Tempo and beat estimation of musical signals. In: ENST-GET, Département TSI (2004)
3. Böck, S., Davies, M., Knees, P.: Multi-task learning of tempo and beat: Learning one to improve the other. In: ISMIR (2019)
4. Cheng, K., Nazer, B., Uppuluri, J., Verret, R.: Beat this - a beat synchronization project. https://www.clear.rice.edu/elec301/Projects01/beat_sync/beatalgo.html
5. Foote, J.: Visualizing music and audio using self-similarity. p. 77–80. MULTIMEDIA '99, Association for Computing Machinery, New York, NY, USA (1999). https://doi.org/10.1145/319463.319472, https://doi.org/10.1145/319463.319472
6. Foote, J., Cooper, M.: Visualizing musical structure and rhythm via self-similarity (09 2001)
7. Gainza, M., Coyle, E.: Time signature detection by using a multi resolution audio similarity matrix. Journal of The Audio Engineering Society (2007)
8. Gainza, M., Coyle, E.: Tempo detection using a hybrid multiband approach. IEEE Transactions on Audio, Speech, and Language Processing **19**(1), 57–68 (2011)
9. Gainza, M.: Automatic musical meter detection. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 329–332 (2009)
10. Lie Lu, Liu, D., Hong-Jiang Zhang: Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech, and Language Processing **14**(1), 5–18 (2006)
11. MatthewDavies, E.P., Böck, S.: Temporal convolutional networks for musical audio beat tracking. In: 2019 27th European Signal Processing Conference (EUSIPCO). pp. 1–5 (2019). https://doi.org/10.23919/EUSIPCO.2019.8902578
12. McFee, B., Metsai, A., McVicar, M., Balke, S., Thomé, C., Raffel, C., Zalkow, F., Malek, A., Dana, Lee, K., et al.: librosa/librosa: 0.8.1rc2 (May 2021). https://doi.org/10.5281/zenodo.4792298
13. Ono, N., Miyamoto, K., Le Roux, J., Kameoka, H., Sagayama, S.: Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram. In: 2008 16th European Signal Processing Conference. pp. 1–4 (2008)
14. Peeters, G.: Time variable tempo detection and beat marking. In: IRCAM - Analysis/Synthesis Team (2005)
15. Pikrakis, A., Antonopoulos, I., Theodoridis, S.: Music meter and tempo tracking from raw polyphonic audio (01 2004)
16. Scheirer, E.D.: Tempo and beat analysis of acoustic musical signals. The Journal of the Acoustical Society of America **103**(1), 588–601 (1998)
17. Schlüter, J., Böck, S.: Improved musical onset detection with convolutional neural networks. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6979–6983 (2014). https://doi.org/10.1109/ICASSP.2014.6854953
18. Schuller, B., Eyben, F., Rigoll, G.: Tango or waltz?: Putting ballroom dance style into tempo detection. EURASIP Journal on Audio, Speech, and Music Processing **8**(1) (2008)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48

19. Tzanetakis, G.: Marsyas (music analysis, retrieval and synthesis for audio signals). http://marsyas.info/index.html
20. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing **10**(5), 293–302 (2002). https://doi.org/10.1109/TSA.2002.800560

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-35995-8_48