2007-01-01

Automatic Bar Line Segmentation

Mikel Gainza
Technological University Dublin, Mikel.Gainza@tudublin.ie

Dan Barry
Technological University Dublin, dan.barry@tudublin.ie

Eugene Coyle
Technological University Dublin, Eugene.Coyle@tudublin.ie

Recommended Citation

Gainza, M., Barry, D. & Coyle, E. Automatic Bar Line Segmentation.Audio Engineering Society Convention Paper, Presented at the *123rd Convention, October, 2007,New York, NY, USA, 2007.*

Audio Engineering Society
Convention Paper
Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

Automatic Bar Line Segmentation

Mikel Gainza[1], Dan Barry[2], and Eugene Coyle[3]

[1] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
mikel.gainza@dit.ie

[2] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
dan.barry@dit.ie

[3] Audio Research Group (Dublin Institute of Technology), Kevin St, Dublin 2, Ireland
eugene.coyle@dit.ie

ABSTRACT

A method that segments the audio according to the position of the bar lines is presented. The method detects musical bars that frequently repeat in different parts of a musical piece by using an audio similarity matrix. The position of each bar line is predicted by using prior information about the position of previous bar lines as well as the estimated bar length. The bar line segmentation method does not depend on the presence of percussive instruments to calculate the bar length. In addition, the alignment of the bars allows moderate tempo deviations.

1. INTRODUCTION

Standard staff music notation utilises vertical lines to indicate the commencement and end of musical bars. The duration of the bar is governed by the time signature and tempo, which imposes the number and duration of the beats respectively that each bar is composed of.

There are numerous algorithms that perform tasks related to music transcription such as pitch detection [1, 2], onset and offset detection [3, 4], key signature estimation [5, 6] and tempo extraction [7, 8]. Recently, the detection of the time signature has also been attempted [9]. However, the detection of the position of the bar lines remains an unexplored area within music transcription research. Other applications related to musical bar segmentation include music editing operations and providing DJs automatic audio markers to perform loops. The detection of the bar line positions can also be used to estimate other hierarchical segmentation levels such as beat and music structure detection.

In this paper, an algorithm that segments the audio according to the position of the bar lines is presented.

The method is based on the system presented in [9], which estimates the time signature of a piece of music by using an audio similarity matrix (ASM) [10]. The method introduced in [9] exploits the self-similarity nature of the structure of music to estimate the time signature by detecting musical bars that frequently repeat in different parts of a musical piece. The method requires the tempo as prior knowledge in order to operate.

In the system presented in this paper, the detection of the tempo is not necessary. The approach obtains the length of the most repeating segment within a range of bar length candidates, which are derived from different tempo and time signature ranges.

Section 2 describes the different components that the bar line segmentation detection system is comprised of. In Section 3, a set of results that evaluate the bar line segmentation approach are presented. Finally, a discussion of the results obtained and some future work are presented in Section 4.

2. PROPOSED APPROACH

In Figure 1, a block diagram showing the different tasks related to the detection of the bar line positions is shown. Firstly, an audio similarity matrix is utilised in order to estimate the bar length and the anacrusis of the song. Next, the position of each bar line is predicted by using prior information about the position of previous bar lines as well as the estimated bar length. Finally, each bar line is estimated by aligning the predicted bar line position to the most prominent value in an onset detection function within a window centered at the predicted bar line.

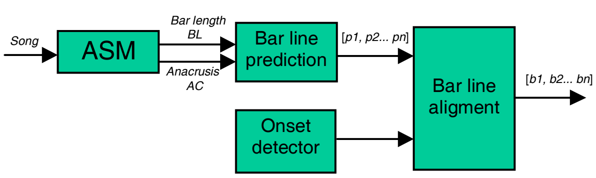

Figure 1: Bar line detection system

2.1. Bar Length Estimation

The bar line detection approach estimates the bar length and the anacrusis of the song by using a method based on the system presented in [9], which utilises prior information of the tempo of the song. In the bar line detection system, the estimation of the tempo is not necessary. Firstly, a spectrogram is generated from windowed frames of length L = 826 samples (18.7 ms), which corresponds to a fraction (1/16) of the duration of a note played at a tempo equal to 200 bpm. The hop size H is equal to half of the frame length L. Thus, it is equal to a fraction (1/32) of the reference beat duration.

$$X(m,k) = abs\left(\sum_{n=0}^{L-1} x(n+mH)w(n) * e^{-j(2\pi/N)k.n}\right) \tag{1}$$

where $w(n)$ is a Hanning window that selects an L length block from the input signal $x(n)$, and where m, N and k are the frame index, FFT length and bin number respectively. It should be noted that $k \in \{1:N/2\}$

As in [9], an Audio Similarity Matrix is generated by comparing all possible combinations of two spectrogram frames by utilising the Euclidian Distance Measure. Thus, the measure of similarity between two frames $m= a$ and $m=b$ are calculated as follows:

$$ASM(a,b) = \sum_{k=1}^{N/2} \left[X(a,k) - X(b,k)\right]^2 \tag{2}$$

As an example, the audio similarity matrix of the audio excerpt shown in Figure 2 is depicted in Figure 3.

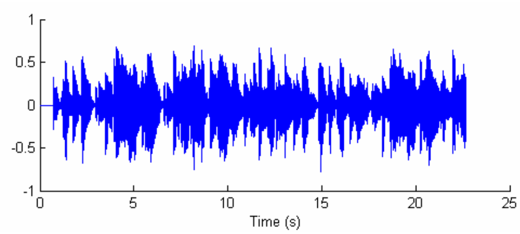

Figure 2: Excerpt of "Good Bait" by John Coltrane

The ASM based system obtains the length of the most repeating segment within a range of bar length candidates. The bar length candidates *bar* considered are within the following range:

$$bar \in \{0.6:\Delta:4\}s \tag{3}$$

where 0.6s is the shortest bar length considered, which corresponds to the bar length of a fast double meter song played at tempo equal to 200 bpm. The longest bar length candidate is equal to 4sec, which corresponds to the bar of a slow quadruple meter song played at a tempo equal to 60 bpm. The vector *bar* steps by increments of Δ, which is equal to 18.7 ms. This value corresponds to two frames of the spectrogram.

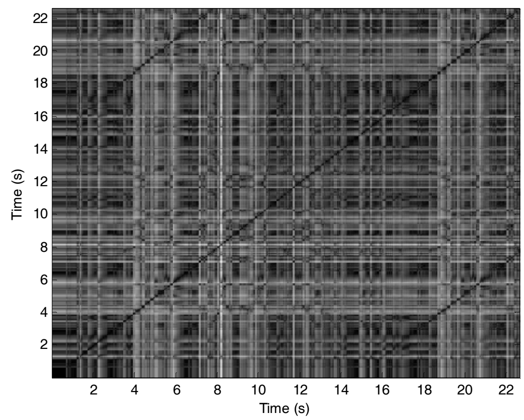

Figure 3: audio similarity matrix of Figure's 2 excerpt

The bar length is estimated by successively combining different groups of components of the ASM [9]. Each of the groups has different length, covering the entire *bar* range. The multi-resolution audio similarity matrix approach is suitable for this operation, since it allows comparisons between longer segments (bars) by combining shorter segments such as a fraction of a note.

As described in [9], for each of the bar length candidates *bar*, the generation of a new ASM is simulated. This is achieved by firstly extracting the diagonals of one side of the symmetric ASM (see Figure 3). Each of the extracted diagonals provides information about the similarities between components separated by a different amount of bars. As an example, if the i_{th} component of the bar length candidate vector is $bar(i)$ = 2s, the diagonals at 2s and 4s provide information of the similarities of components separated by one bar and two bars respectively.

Next, each of the diagonals is partitioned into non-overlapping data segments of length equal to the bar length candidate $bar(i)$. This is shown in Figure 4, where an illustrative example of a diagonal with length M segmented into n groups of length $bar(i) = b$ is depicted. The first and second segments extracted from the first diagonal, which are denoted as S_1 and S_2, will correspond to the similarity measures between the first and second bars, and the second and third bars respectively. The incomplete bar is denoted as *I*.

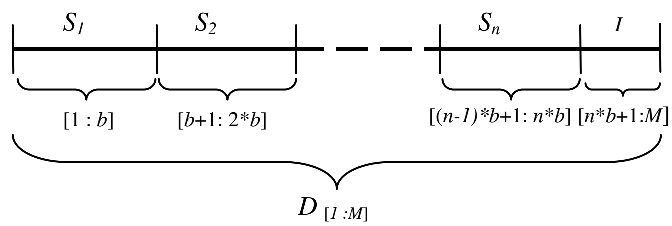

Figure 4: segmentation of a diagonal *D* with length *M* into segments of length *b*

Next, a similarity measure *SM* is obtained for each of the segments *S* and *I* of each of the diagonals extracted [9]. The extraction and segmentation of the diagonals of the ASM associated to a given bar length, combined with the further *SM* calculation of the mentioned segments simulates a new ASM. In this new matrix, each comparison between any two bars of the initial ASM will be represented by a unique cell in the new ASM. As an example, $SM(S1)_r$ corresponds to the similarity measure of the first segment of the r_{th} diagonal.

Following this, the *SM* of all the diagonal segments are combined to obtain a unique similarity measure per bar length candidate. A more detailed description of the similarity measure calculation is included in [9].

Finally, a Gaussian-like function is applied to the bar line detection function. As it can be seen in Figure 5, this function gives more weight to the values around 2s and equal weight to both edges of the bar length candidate range.

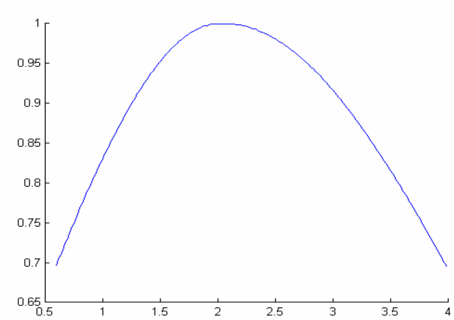

Figure 5: Gaussian-like weighting function

As an example, the weighted bar line detection function of Figure 2's example is depicted in Figure 6. The function is flipped in up/down direction. Thus, high similarity values will correspond to peaks in the detection function.

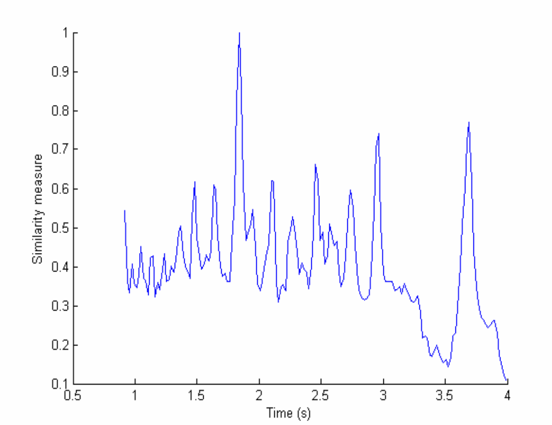

Figure 6: bar line detection of Figure 2's excerpt

2.2. Song Anacrusis detection

In Section 2.1, a method to estimate the bar length is described. However, if anacrusis is not taken into consideration, the boundaries of the segmented groups from the diagonals of the ASM will not fully align to the commencement and finish of the musical bars. This will affect the overall similarity of the new ASM. In addition, the detection of the length of the anacrusis bar represents a crucial task for estimating the bar line positions, since it provides the position of the first bar.

In [9, 11], first attempts to detect the anacrusis beats were introduced. The method used in this system firstly generates a vector of anacrusis candidates within a given segment of the recording. Following this, an anacrusis detection function is generated by calculating a similarity measure per anacrusis candidate.

2.2.1. Anacrusis candidates detection

The number of anacrusis candidates depends on the bar length candidate, where the length of the anacrusis bar should be smaller than the bar length candidate. Thus, for an i_{th} bar length candidate $bar(i) = b$, anacrusis candidates will be detected by picking peaks within a region R of an onset detection function. The region R starts at the first onset of the song, and has a length equal to the bar length candidate b. Then, by applying a moderate threshold equal to the mean of the detection function of R, peaks in that region will be considered as anacrusis candidates, which we denote as *ana*. Finally, a 100ms sliding window centered at each peak is applied, where only the most prominent peak is kept.

The complex onset detection method is utilized in this paper [3], which provides a good compromise in the detection of slow and sharp onsets. As in [12], the onset detection function, OD, is processed as follows:

$$OD(m) = HWR(OD(m) - \overline{OD}(m)) \tag{4}$$

where HWR denotes half wave rectification and $\overline{OD}$ *is* the mean of the OD within a sliding window of length equal to 1s centred at the current frame m.

As an example, Figure 7 shows the onset detection function of the first 5 seconds of Figure 2´s example. This song has an anacrusis of 1 beat, which duration is shown in Figure 7. The length of the region R is equal to the duration of the bar length candidate b, which is approximately equal to 1.83 seconds. In this case, the vector of anacrusis candidates will be formed by the peak locations, which are located at ana = [0.74, 0.88, 1.3, 1.77, 2.25] s.

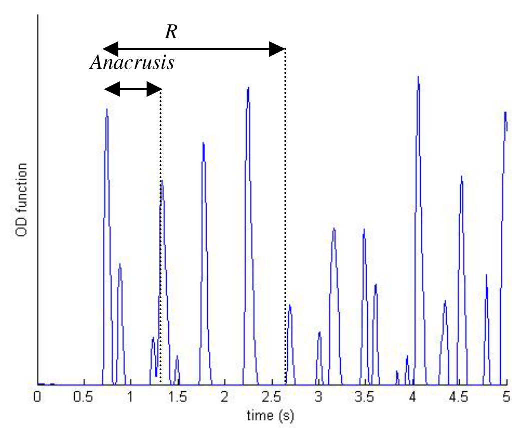

Figure 7: anacrusis candidates detection region, *R*, of Figure'2 example

2.2.2. Similarity measure of shifted ASM versions

A sliding offset from the origin of the ASM equal to each anacrusis candidate is successively applied. As an example, if the j_{th} component of *ana* is equal to $ana(j) = x$ frames, the ASM will be shifted from $ASM_{(1,1)}$ to $ASM_{(x,x)}$. Next, the same method as in section 2.1 is applied in order to obtain the anacrusis candidate that provides the best similarity measure for each bar length candidate.

The anacrusis detection function obtained by applying Figure 7's example is depicted in Figure 8. The maximum of the function is located at j=3, which corresponds to $ana(3)$ = 1.3s.

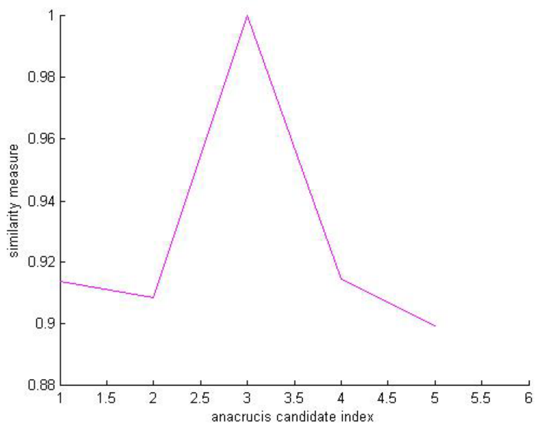

Figure 8: anacrusis detection function of figure 2's example.

The decision of incorporating anacrusis into a piece of music varies depending on the composer or performer. When anacrusis is utilized, any combination of beats to form an incomplete bar is allowed. Thus, songs with no anacrusis are more common that songs with any other combination of anacrusis beats (e.g: one, two, three, three and a half beats…). Consequently, the detection of no anacrusis, $ana(1)$ will be giving more weight as follows:

$$ana(1) = ana(1) + (ma - mi) * 0.5 \qquad (5)$$

where ma and mi are the maximum and minimum respectively of the anacrusis detection function.

2.3. Bar line prediction and alignment

The estimated bar length, which we denote as BL, represents the most repeating bar length within the audio segment analysed by the ASM. However, tempo changes can generate bars with different lengths. Consequently, the length of the bar should be dynamically updated following to each bar line prediction.

Firstly, the position and length of the first bar are initialised by using estimations of the song's anacrusis AC and the bar length BL respectively, which are provided by the ASM. Thus, $p(1) = AC$, and $BL(1) = BL$, where p denotes bar position.

In order to predict the length of the current l_{th} bar, $BLp(l)$, information of the length of the previous 6 bars is used:

$$BLp(l) = \frac{\sum_{x=l}^{M} BL(l-x)}{M} \qquad (6)$$

where M is the maximum value of [6 , l-1] and l>1.

In [8], the prediction of future beats uses information of the tempo of the recording. In the presented bar line detection method, the prediction of the position of the next bar, $pr(l+1)$, uses information of the predicted bar length as follows:

$$pr(l+1) = p(l) + BLp(l) \qquad (7)$$

Following this, the position of the next bar line $p(l+1)$ is estimated by aligning the predicted bar line position, $pr(l+1)$, to the most prominent value in an onset detection function within a 100 ms window centered at the predicted bar line position.

Then, the bar length of the current bar, $BL(l)$, is updated as follows:

$$BL(l) = p(l+1) - p(l) \qquad (8)$$

As an example, Figure 9 shows the onset detection function of a segment within figure 2´s audio signal. The current l_{th} bar line position is located at $p(l)$ = 6.78s. The predicted bar length $BLp(l)$ is equal to 1.87s. Thus, the predicted position of the next bar line will be located at $pr(l+1)$ = 8.578s. Finally, the position of the next bar is aligned to the peak in the onset detection function located at $p(l+1)$ = 8.606 s.

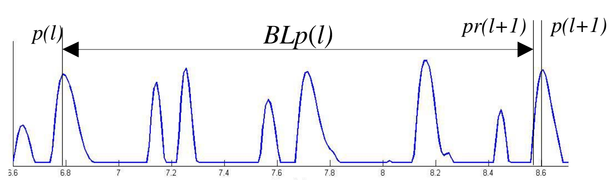

Figure 9: Bar length prediction example

3. RESULTS

In order to evaluate the presented approach, excerpts of the set of audio signals listed in Table 1 are utilised. The position of the bar lines are manually annotated and compared against the estimation of the bar line positions provided by the proposed method. The length of the anacrusis bar is also included in Table 1, where songs with no anacrusis have a length equal to 0s. The bar length is obtained by calculating the mean of the difference between consecutive manually annotated bar line positions.

Song	*Artist*	*Anacrusis bar length (s)*	*Bar length (s)*
Teenage kicks	The Undertones	0.9	1.76
Good Bait	John Coltrane	0.41	1.84
Pastor	Madredeus	0	2.86
…Meets his Maker	DJ shadow	0	3..57
Chameleon	Head Hunters	0	5.03
Sexy boy	Air	0	2.13
All mine	Portishead	0	1.82
Photo Jenny	Belle and Sebastian	0	2.85
Mami Gato	Medeski, Martin and Wood	0	3.47

Table 1: Audio signals Testbed

The results are shown in Table 2, where *AC*, *BL* denote the estimated anacrusis and bar length respectively. The correct and incorrect detections of *AC* and *BL* are denoted as YES and NO respectively. In addition, the percentage of correct and incorrect bar line positions is also provided, which is denoted as *CBL* and *IBL* respectively. The detection of anacrusis, bar lengths and bar line positions falling within a 150 ms window centred at the target locations are considered correct detections.

4. DISCUSSION AND CONCLUSIONS

In this paper, a system that automatically detect the position of the bar lines has been introduced. The accuracy of the system depends on three independent tasks: bar line detection, anacrusis detection and bar line alignment.

From Table 2, it can be seen that the detection of the bar length provides a high percentage of good results. The only song which bar length was incorrectly estimated is "Chameleon". The estimated bar length for this song is equal to half of the song's bar length. These results show the robustness of the bar line detection method, which does not depend on the presence of percussive instruments and is solely based on the repetitive nature of the majority of the music.

Song	*AC (s)*	*BL (s)*	*CBL*	*IBL*
Teenage kicks	1 YES	1.76 YES	20/20 = 100%	0
Good Bait	0.5 YES	1.83 YES	19/24 = 79.1%	5
Pastor	1.7 NO	2.85 YES	0/19 = 0%	19
…Meets his Maker	0.58 NO	3..54 YES	0/15 = 0%	15
Chameleon	0 YES	2.56 NO	17/17 = 100%	17
Sexy boy	0 YES	2.14 YES	20/21 = 95.2 %	1
All mine	0.4 NO	1.96 YES	0/22 = 0%	22
Photo Jenny	0 YES	2.83 YES	18/20 = 90%	2
Mami Gato	0 YES	3.6 YES	14/15 = 93.33%	1
TOTAL CORRECT	6 / 9 = 66.6 %	8 / 9 = 88.8%	108/ 173 = 62.4%	82

Table 2: bar line detection system results

As it can be seen in Table 2, the detection of the anacrusis is less accurate. The system incorrectly detected the use of anacrusis in songs played without the use of that technique. Due to the repetitive nature of these songs, a shift of the ASM also encountered high degree of repetition between the incorrectly aligned musical bars.

The technique utilized to align the bar line predictions to the onset detection function peaks shows a high degree of accuracy. However, the behavior of this task entirely depends on the accuracy of the bar length and anacrusis estimations. By only considering the songs were the anacrusis was correctly estimated, Table 2's results show a high percentage of good results. On the other hand, the incorrect estimation of the anacrusis results in 0% of correct bar line positions. The song "Chameleon" represents a different case, where the bar length was incorrectly estimated. Since the estimated bar length is half the song's bar length, all lthe bar lines will be correctly estimated . However, in between any two song bars, a spourious bar detection will also be estimated.

The size of the database of audio signals should be increased in order to continue the evaluation of the presented system. The development of a more robust anacrusis detector also warrants future work. The system should combine the repetitive nature of the music in conjunction with other anacrusis bar properties. As previously mentioned, the alignment of the bars allows moderate tempo deviation. However, bar length changes due to time signature or abrupt tempo changes will affect the accuracy of the results. A system that firstly segments the audio signal according to these changes should be considered as an area of future work. Thus, the system presented in this paper will be applied to each individual audio segment.

5. ACKNOWLEDGEMENTS

Work supported in part by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project EASAIER contract IST-033902. The work was also supported by Enterprise Ireland under the IMAAS project, reference CFTD/06/220.

6. REFERENCES

1. Bello, J.P. *Blackboard system and top-down processing for the transcription of simple polyphonic music*. *In Digital Audio Effects, DAFX. 2000.*

2. Smaragdis, P. and J.C. Brown. *Non-negative matrix factorization for polyphonic music transcription*. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2003.

3. Duxbury, C., et al. *Complex Domain Onset Detection For Musical SIgnals*. In 6th Int. Conference on Digital Audio Effects (DAFx-03). 2003. London, UK.

4. Gainza, M., B. Lawlor, and E. Coyle. *Onset Detection Using Comb Filters*. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2005.

5. Pauws, S. *Musical key extraction from audio*. In International Symposium on Music Information Retrieval, Barcelona. 2004.

6. Chai, W. and B. Vercoe. *Detection Of Key Change In Classical Piano Music*. *In ISMIR. 2005. London.*

7. Scheirer, E., *Tempo and Beat Analysis of Acoustic Musical Signals.* J. Acoust. Soc. Am., 1998. **103**(1): p. 588-601.

8. Davies, M.E.P. and M.D. Plumbley. *Causal Tempo Tracking of Audio*. *In Proceedings of the International Conference on Music Information Retrieval. Audiovisual Institute, Universitat Pompeu Fabra, Barcelona, Spain. 2004.*

9. Gainza, M. and E. Coyle. *Time Signature Detection by Using a Multi-Resolution Audio Similarity Matrix*. *In Audio Engieneering Society 122nd Convention. 2007. Viena.*

10. Foote, J. *Visualizing Music and Audio using Self-Similarity*. *In ACM Multimedia. 1999. Orlando.*

11. O'Keeffe, K., *Dancing Monkeys (Automated creation of step files for Dance Dance Revolution)*. MEng Thesis. 2003

12. Davies, M.E.P. and M.D. Plumbley. *Context-dependent beat tracking of musical audio*. In IEEE Transactions on Audio, Speech and Language Processing, **15**(3), 1009-1020, 2007.