Audio Engineering Society

Convention Paper

Presented at the 122nd Convention
2007 May 5–8 Vienna, Austria

Time Signature Detection by Using a Multi-Resolution Audio Similarity Matrix

Mikel Gainza[1], Eugene Coyle[2]

[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
eugene.coyle@dit.ie

ABSTRACT

A method that estimates the time signature of a piece of music is presented. The approach exploits the repetitive structure of most music, where the same musical bar is repeated in different parts of a piece. The method utilises a multi-resolution audio similarity matrix approach, which allows comparisons between longer audio segments (bars) by combining comparisons of shorter segments (fraction of a note). The time signature method only depends on musical structure, and does not depend on the presence of percussive instruments or strong musical accents.

1. INTRODUCTION

Standard staff music notation utilises symbols to indicate note durations (onset and offset times). The pitch of the notes is derived from the key signature and the position of the note symbols in the staff. In addition, the information regarding the tempo, the commencement and end of the bars, and the time signature is also included in the staff [1]. Western music describes the time signature as the ratio between two integer numbers, where the numerator indicates how many beats are in a bar and the denominator specifies the note reference. There are numerous algorithms that perform pitch detection [2, 3], onset detection [4, 5], key signature estimation [6, 7] and tempo extraction [8, 9]. However, the detection of the metrical structure or the time signature remains a relatively unexplored area. In [10], Brown obtains the meter by using the autocorrelation function under the assumption that the frequency of repetition of notes is greater on the downbeat of the musical bar. Gouyon estimates the meter (duple or triple) by tracking periodicities of low level features around beat segments [11]. Even though the title of [12] is related to music meter, the approach focuses on detecting the time signature within Greek traditional music by using an audio similarity matrix (ASM) [13, 14], which compares all possible combinations of two frames of the domain utilised to represent the audio file (e.g.: time domain, spectrogram, cepstrum…). The method described in [12] calculates

the numerator and denominator of the time signature independently. The denominator is obtained by tracking the similarities in the audio signal between instants separated by beat duration. Thus, it is assumed that successive notes will be similar. In a similar manner to [10], the time signature numerator is estimated by analysing the similarities between successive bars. However, both methods [10], [12] discard similarities between bars located at different points in the music.

In this paper, a time signature detection algorithm is presented, which estimates the number of beats in a musical bar. The method is based on the use of audio similarity matrix (ASM) [13]. The ASM exploits the repetitive nature of the structure of music, where the same musical bars, chorus or phrases frequently repeat in different parts of a musical piece. The presented approach seeks repetitions in any two possible musical bars without the requirement of the periodic repetition of any musical event or the repetition of successive musical bars. Thus, the limitations of previous approaches are overcome.

Section 2 describes the different components that comprise the time signature detector. In Section 3, a set of results that evaluate the time signature detector is introduced. Finally, a discussion of the results obtained and some future work are presented in Section 4.

2. PROPOSED APPROACH

The different parts of the time signature detection system here are described in this section. Firstly, by using prior knowledge of the tempo of the song, a spectrogram is generated with a frame length equal to a fraction of the duration of the beat of the song. Following this, the first note of the song is detected. A reference ASM is then produced by using Euclidian distance measures between the frames starting at the first note. Such fine representation allows the approach to capture the similarities between small musical events such as short notes. Then, a multi-resolution ASM approach is undertaken in order to form other audio similarity matrices representating a variety of bar length candidates. Having formed all the new ASMs within a certain range, the new ASM which provides the highest similarity between its components will correspond to the bar length. Following this, a method to detect the anacrusis of the song is also introduced, which is an anticipatory note or notes occurring before the first bar of a piece [15]. Finally, the time signature is obtained and a more accurate tempo estimation is also provided.

2.1. Spectrogram

In order to provide a more accurate input to the problem of interest here (time signature detection), the tempo is semi-automatically estimated in the same manner as [10] and [11], where the tempo and the beat locations were respectively known.

By using the tempo information, a spectrogram is generated from windowed frames of length L, which are equal to a fraction (1/32) of the duration of the beat of the song. The hop size H is equal to half of the frame length L (1/64 of the beat duration).

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

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 $k \in \{1{:}N/2\}$.

Following this, the first note of the song is detected, by obtaining the energy of the frequency ranges $E1$ = [1:3000] Hz and $E2$ = [15000:21000] Hz respectively [16]. This will disable the columns of the spectrogram that contain no useful information. If a note has been played, it is expected that $E1$ has a much higher value than $E2$. Otherwise, the energy will be spread over the frequency axis, and it will be assumed that the signal does not contain musical notes. Thus, by using a high threshold Tn, the first note played in the song will be estimated as follows:

$$\frac{E1}{E2} < Tn \quad (2)$$

2.2. Reference Audio Similarity Matrix

An Audio Similarity Matrix [13] is built 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 $m=b$ is given by:

$$ASM(a,b) = \sum_{k=1}^{N/2} \left[X(a,k) - X(b,k)\right]^2 \quad (3)$$

As an example, the spectrogram of an excerpt of a MIDI generated song is depicted of Figure 1, which is played in a 6/8 time signature. The bar lines are also depicted in white, where it can be seen that the excerpt comprises five complete bars and one incomplete bar.

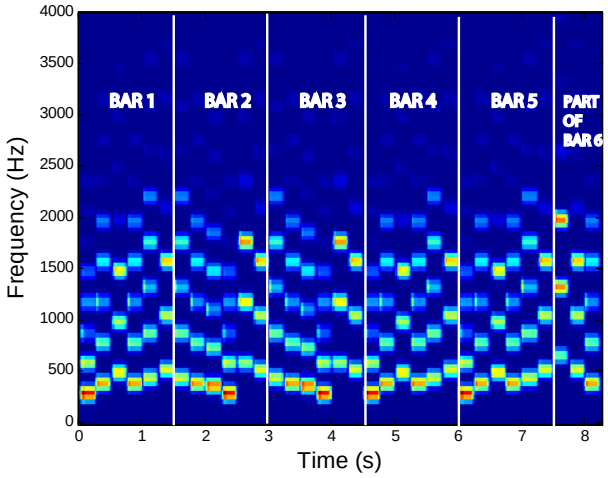

Figure 1: Spectrogram of a song played in 6/8

The ASM of Figure1's spectrogram is depicted in Figure 2, where the brightness of each matrix cell provides a measure of the similarity between two frames. Thus, a bright and a dark matrix cell represent a dissimilar and similar comparison respectively. It should be noted that the presented time signature detector is designed to work with real audio signals. However, a MIDI example has been utilised for illustration purposes, since this type of format provides steady signals using constant tempo, which generates clearer figures.

2.2.1. Multi-resolution matrices

The ellipses depicted in Figure 2 show the groups of cells in the audio similarity matrix that contain the comparisons between the frames of each possible combination of two musical bars. As an example, the group 1-2 denotes the comparison between the frames of bar 1 and 2, where the first frame of bar 1 is compared against the first frame of bar 2, the second frame of bar 1 is compared against the second frame of bar 2 and so on. From Figure 2, it can be appreciated that the group of cells denoted as 2-3, 4-5, 1-4 and 1-5 show high similarity. This indicates that bars 1, 4 and 5, and bars 2 and 3 respectively contain similar notes in their respective musical bars, as can be appreciated from a visual inspection of Figure 1. Consequently, the components of an ASM with a resolution equal to the length of the musical bars will show a high degree of similarity.

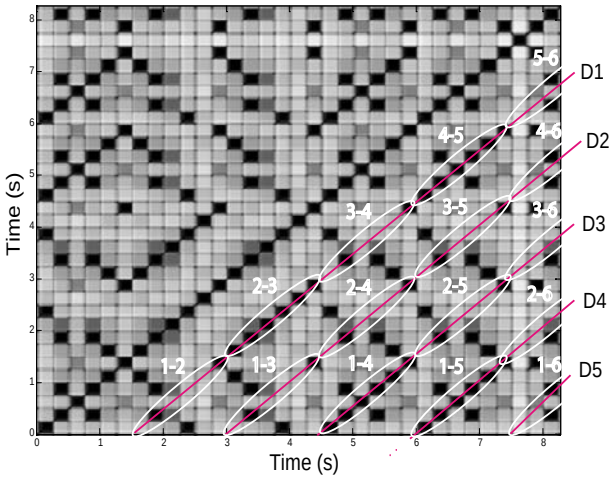

Figure 2 ASM of Figure 1´s example

The existence of any time signature within the 2/2 to 12/8 range is investigated, including complex time signatures such as 5/4, 7/8 and 11/8. Thus, the range of number of beats in a bar considered in this method is comprised in the range {2:12}. In addition, the maximum length of the bar is restricted to 3.5 s, which corresponds to a musical bar formed by 12 beats played with a bpm = 205.

In order to obtain the time signature of the piece, the method successively combines integer numbers of components of the ASM to form groups of components of length *Bar*. Considering that the length of the spectrogram frame is equal to 1/64 of the beat duration, the range of *Bar* will be within {2*64:12*64}. Thus, each of the values of *Bar* corresponds to a bar length candidate. As an example, the combination of 64*6 = 384 components will correspond to a duration of 6 beats. As it can be seen in Figure 1, this duration corresponds to the length of the musical bar of the song.

For each of the bar length candidates *Bar*, the generation of a new ASM will be simulated. This is achieved as follows; Firstly, the diagonals of one side of

the symmetric ASM (see Figure 2) which are integer multiples of *Bar* are extracted. Each of the diagonals provides information about the similarities between components of musical bar candidates separated by a different amount of bars. As an example, the diagonals depicted as D_1 and D_2 in Figure 2 provide information 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*, which we denote as *G*, and an incomplete segment, which we denote as *P*. As an example, the components inside the ellipses located at the end of the *x* axis side of Figure 2: 5-6, 4-6, 3-6, 2-6 and 1-6, correspond to the incomplete segments *P*. The remaining ellipses of Figure 2 group the components of each of the complete segments *G* (e.g: components inside the Ellipse 1-2). Then, a similarity measure of each of the complete and incomplete segments, which we denote as *SCS* and *SIS*, provides the measure of the similarity between two bars. The similarity measure is calculated as follows:

$$SCS = \sqrt{\frac{\sum_{i=1}^{Bar} G_i^{\,2}}{Bar}} \qquad (complete\,bars)$$

$$SIS = \sqrt{\frac{\sum_{i=1}^{r} P_i^{\,2}}{r}} \qquad (incomplete\,bars) \qquad (4)$$

where *Gi* and *Pi* are the i_{th} component of the complete and incomplete segments respectively, and where *r* is the length of the incomplete segment.

Each of the *SCS* and *SIS* measurs corresponds to a component of the new audio similarity matrix. The combination of these measures simulates the generation of an ASM from a spectrogram with a frame length equal to a multiple of the subdivision of the note beat. Considering Figure 2 example, the generation of a new ASM by grouping the components contained in the white ellipses will be simulated as in Figure 3. As an example, *SCS*(*1,2*) and *SIS*(*5,6*) correspond to the similarity measure between bars 1 and 2, and 5 and 6 respectively. It should be noted that only one of the symmetric sides of the ASM is considered. In addition, the main diagonal is also discarded, which does not provide any additional useful information.

					$SIS_{(5,6)}$
				$SCS_{(4,5)}$	$SIS_{(4,5)}$
			$SCS_{(3,4)}$	$SCS_{(3,5)}$	$SIS_{(3,6)}$
		$SCS_{(2,3)}$	$SCS_{(2,4)}$	$SCS_{(2,5)}$	$SIS_{(1,6)}$
	$SCS_{(1,2)}$	$SCS_{(1,3)}$	$SCS_{(1,4)}$	$SCS_{(1,5)}$	$SIS_{1,6)}$

Figure 3: New ASM of Figure 2´s example

In order to measure the similarity of each new ASM, *SM*, the following equation is utilised:

$$SM = \frac{Bar \times \sum_{i=1}^{s_c} SCS_i + r \times \sum_{i=1}^{s_i} SIS_i}{Bar \times s_c + r \times s_i} \qquad (5)$$

where s_c and s_i correspond to the number of *SCS* and *SIC* segments respectively. This equation weights the segments according to the length *r*.

Having obtained the *SM* of all the new ASMs, the bar length associated with the highest *SM* is deemed to be the bar length of the entire piece.

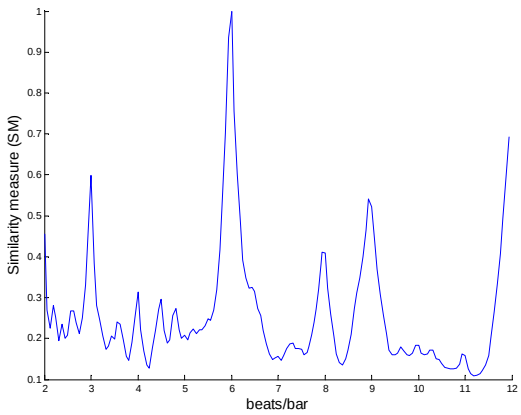

Figure 4: Beats/bar detection of Figure 1's example

The multi-resolution audio similarity matrix approach allows comparisons between longer segments (bars) by

combining shorter segments (1/32 of a note reference). The method avoids having to generate a new spectrogram and a new audio similarity matrix for each different frame length considered in the analysis. In addition, the use of short segments provides good time resolution, which is required in order to compare individual notes located in different bars.

In Figure 4, the *SM* of Figure 2´s example for all the range of bar length candidates is displayed in Figure 4, where it can be seen that the highest *SM* value corresponds to 6 beats in the bar.

2.3. Anacrusis Detection

The first note of the song displayed in Figure 2 corresponds to the first note of the first musical bar. However, this is not always the case, where other notes can be played before the first bar. In this case, the boundaries of the segmented groups from the diagonals of the ASM will not fully correspond to the start and finish of the musical bars. This problem is addressed in [17], where the location of the first beat of the first bar is obtained for dance music songs played in 4/4. The songs are successively segmented into bars by covering each possible case of groups of eight notes before the first bar. Then, an ASM is generated for each of the cases to find the ASM with more similar components.

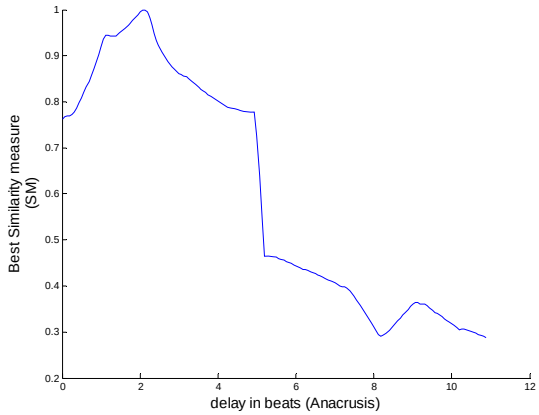

Figure 5: Anacrusis detection example

In order to detect the anacrusis of the song, a similar method to [17] is implemented by adding a sliding offset from the origin of the ASM, which is also a multiple of the subdivision of the beat duration. Thus, an anacrusis of 2 beats will correspond to an offset of 2*64=128 frames, which results in a shift from the origin of $ASM_{(1,1)}$ to $ASM_{(129,129)}$.

The anacrusis range is equal to the *Bar* range minus one full beat. Thus, for the case of a grouping of 3 beats, the maximum anacrusis value will be 2 beats.

As an example, an anacrusis of 2 eight notes is added to the example of Figure 2. The result of the detection is shown in Figure 5, where it can be seen that the most similar measure was obtained when the ASM was shifted approximately 2 beats.

2.4. Time Signature Estimation

Having obtained the number of beats *B* that provides the most similar measure *SM* for the entire beat and anacrusis range, the time signature is estimated. The time signature denominator is obtained by rounding B to the nearest integer value. Then, the denominator will be obtained as follows: if the number of beats *B* estimated is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12, the estimated time signature will be detected as 2/2, 3/4, 4/4, 5/4, 6/8, 7/8, 8/8, 9/8, 10/8, 11/8 or 12/8 respectively. Then, 2/2 and 8/8 will be estimated as 4/4 by just halving and doubling the tempo respectively.

Since the tempo does not remain constant through the entire tune, *B* will rarely be an integer number. Thus, in order to provide a more accurate average tempo, the following equation is applied:

$$newTempo = \frac{B \times tempo}{round(tempo)} \qquad (6)$$

where *tempo* is a semi-automatic tempo extraction

3. RESULTS

In order to evaluate the presented approach, a set of audio signals selected from commercial CD recordings is utilised. The songs are listed in Table 1, where a large variety of time signatures and genres are represented in the testbed. An excerpt of approximately 12 seconds was extracted in each song to obtain the time signature of the piece. From Table 1, *BPM* and *ana* correspond to the semi-automatic tempo and the anacrusis respectively.

Song num	*Song*	*Artist*	*Time Sig*	*BPM*	*ana*
1	Eleven	Primus	11/8	230	0
2	Windows To The Soul	Steve Vai	11/8	243	0
3	Watermelon In Easter Hay	Frank Zappa	9/4	55	0
4	ScatterBrain	Jeff Beck	9/8	250	0
5	Take It To The Limit	The Eagles	3/4	90	0
6	Doing It All For My Baby	Huey Lewis & The News	12/8	275	6
7	Forces… Darling	Koop	4/4-8/8	200	0
8	Sliabh	Danu	6/8	190	2
9	Money	Pink Floyd	7/8	120	0
10	Whirl	The Jesus Lizard	5/4	150	0

Table 1: TestBed content

The results can be seen in Table 2, where *newBPM, CTS and Cana* denote the new estimated tempo value, correct time signature detection and correct Anacrusis detection respectively.

Song num	*ana*	*TimeSig*	*newBPM*	*CTS*	*Cana*
1	1	11/8	228	YES	NO
2	1	11/8	242	YES	NO
3	0	2/2	53	NO	YES
4	10	11/8	248	NO	NO
5	1	3/4	90	YES	NO
6	9	12/8	276	YES	NO
7	0	8/8	208	YES	YES
8	2	6/8	200	YES	YES
9	0	7/8	121	YES	YES
10	0	5/4	153	YES	YES

Table 2: Time Signature Detection Results

In Figure 6, the similarity detection function of the song *num.* 8 “Sliabh” is depicted. The song consists on a pipe playing solo, where the tempo is not maintained constant over the song. This is apparent in Figure 6, where the most similar measure was obtained for a grouping of 5.6 beats. However, since the nearest integer is 6 beats, the time signature is correctly estimated.

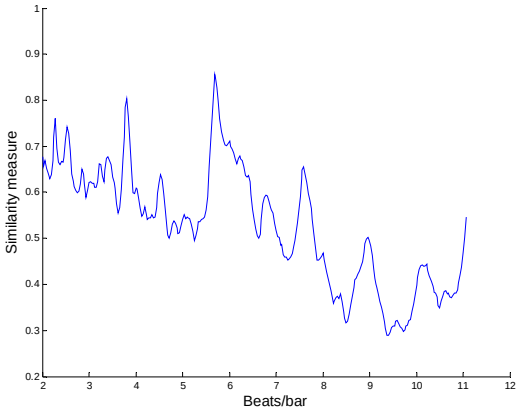

Figure 6: Beats/bar detection of “Sliabh”

Figure 7 depicts the similarity detection function of “Eleven”, which is played in the infrequent time signature 11/8. It can be seen that a very distinctive peak in the function arises at 11 beats.

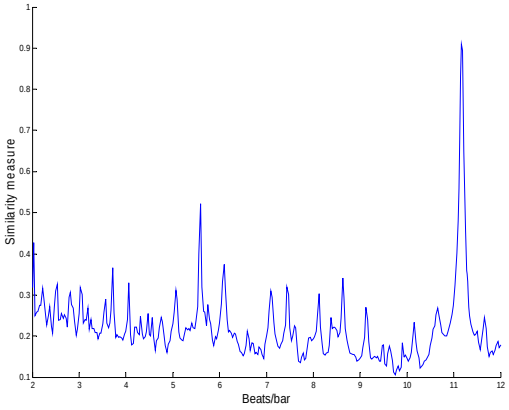

Figure 7: Beats/bar detection of “Eleven”

4. DISCUSSION AND FUTURE WORK

A system that detects the time signature of a piece of music has been presented. In addition, a method to detect the anacrusis of a song has also been introduced. The system only depends on musical structure, and does not depend on the presence of percussive instruments, strong musical accents or a particular metric structure. The system can detect simple time signatures such as

4/4 as well as complex time signatures such as 11/8. The results show the robustness of the time signature detector for a variety of time signatures, where only the song *num* 3 and 4 are detected incorrectly. It should be noted that the bar length of song *num* 3 is longer than the maximum of 3.5s allowed in the approach. However, by allowing a maximum bar length of 11s and by increasing the length of the excerpt to 1m, the correct number of beats is detected. This can be seen in Figure 8, where a clear peak in the 9 beats location arises. By applying the method to estimate the time signature described in Section 2.4, Figure 8's detection will be estimated as 9/8, since it is assumed that a bar of 9 beats will be divided into eight notes. However, a further classification based on the tempo could be incorporated to select the denominator of the time signature.

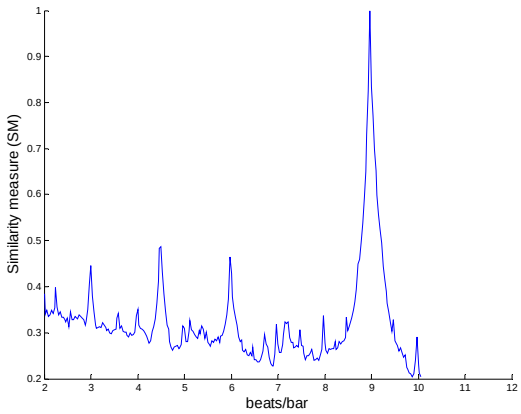

Figure 8: Beats/bar detection of "Watermelon in Easter Hay"

Only two of the excerpts of the songs were played using anacrusis. The system anticipated the correct number of notes preceding the first barline in one of the two cases. However, an anacrusis of just one beat was also detected in songs where there were no notes before the first bar. This can be due to deviations of the tempo that occur in a song, which can generate musical bars with different lengths. Consequently, improving the accuracy of the anacrusis detection should be considered as further work.

The system assumes that there is no time signature change trough the tune. A modification of the algorithm to adapt it to bar length deviations, tempo changes and time signature changes warrants future work.

5. ACKNOWLEDGEMENTS

Work supported by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project EASAIER contract IST-033902.

We would like to thank Dan Barry and David Dorran for all the relevant discussions regarding the topic of this paper and the proof-reading of the same.

6. REFERENCES

[1] Bent, I. D. and Hughes, D. W, "Notation". Grove Music Online 2006. http://www.grovemusic.com. Ed. L. Macy.

[2] Klapuri, A., *Signal Processing Methods for the Automatic Transcription of Music*. Phd Thesis, 2004.

[3] Martin, K., *Automatic transcription of simple polyphonic music: Robust front end processing.* MIT Media Laboratory. 1996

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

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

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

[7] Pauws, S. *Musical key extraction from audio*. In Proc of *International Symposium on Music Information Retrieval, Barcelona*. 2004.

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

[9] Davies, M.E.P. and M.D. Plumbley. *Causal Tempo Tracking of Audio*. In Proc *Int.l Conference on Music Information Retrieval., Barcelona, Spain*. 2004.

[10] Brown, J.C., *Determination of the meter of musical scores by autocorrelation.* Journal of the Acoustical Society of America, 1993. **4**(94): p. 1953-1957.

[11] Gouyon, F. and P. Herrera. *Determination of the meter of musical audio signals: Seeking*

recurrences in beat segment descriptors. In Proc of *AES 114 thConvention*. 2003.

[12] Pikrakis, A., I. Antonopoulos, and S. Theodoridis. *Music Meter And Tempo Tracking From Raw Polyphonic Audio*. In Proc of *5th International Conference on Music Information Retrieval-ISMIR 2004*.

[13] Foote, J. *Visualizing Music and Audio using Self-Similarity*. In Proc of *ACM Multimedia*. 1999. Orlando.

[14] Foote, J. and S. Uchihashi. *The beat spectrum: a new approach to rhythm analysis*. 2001.

[15] Dogantan, M, "Anacrusis". Grove Music Online. http://www.grovemusic.com. Ed. L. Macy. 2007

[16] Amatriain, X., et al., *Spectral Processing*, In Proc *Digital Audio Effects, DAFX*. 2002, John Wiley & Sons. Chapter 10.

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