Generation of Musical Scores of Percussive Un-Pitched Instruments from Automatically Detected Events

Christian Uhle [1] and Christian Dittmar [2]

[1] Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany
uhle@idmt.fraunhofer.de

[2] Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany
dmr@idmt.fraunhofer.de

ABSTRACT

This publication addresses the generation of a musical score of percussive un-pitched instruments. A musical event is defined as the occurrence of a sound of a musical instrument. The presented method is restricted to events of percussive instruments without determinate pitch. Events are detected in the audio signal and classified into instrument classes, the temporal positions of the events are quantized on a tatum grid, musical meter is estimated and preparatory beats are identified. The identification of rhythmic patterns on basis of the frequency of their occurrence enables a robust identification of the tempo and gives valuable cues for the positioning of the bar lines using musical knowledge.

1. INTRODUCTION

Automatic transcription of music is a research topic of great interest with a number of different applications, e.g. the synchronization of light effects and music, tools for music education and meta-data extraction for music information retrieval. The analyzed musical material belongs to genres of popular music, but the presented method is not restricted to particular musical genres. In the following, transcription is defined as the generation of a musical score from listening to or analysis of a piece of music. The score should contain the rhythmic information (e.g. starting time and duration) and the pitch information of the notes with determinate pitch and the description of the played instruments. Although the estimation of other metrical information, namely the time signature, is not necessary for the automatic synthesis of the transcribed music, it is required for the generation of a valid musical score and for the reproduction by human performers. Therefore, an automatic transcription process can be separated into two tasks: the detection and classification of the musical events (i.e. the notes), and the creation of a musical score from the detected notes. The latter requires the estimation of the metrical structure of the music, the quantization of the temporal positions of the detected notes in the score, the recognition of preparatory beats and determination of the position of the bar lines. In this work, the creation of a musical score for percussive instruments without determinate pitch from polyphonic musical audio signals is described. The detection and classification of the events is done with an Independent Subspace Analysis method. An event is defined as the occurrence of a note of a musical instrument. The audio signal is segmented into parts of a few segments length using a distance measure between short frames of the audio signal represented by a vector of low-level features. The tatum grid and higher metric levels are estimated from the segmented parts separately. It is assumed that the metric structure does not change within one segmented part of the audio signal. The detected events are aligned with the estimated tatum grid. This process corresponds to the well-known quantization function in common MIDI sequencer software programs for music production. The bar length is estimated from the quantized event list and recurring rhythmic patterns are identified. Knowledge of the rhythmic patterns is used for the correction of the estimated tempo and the identification of the position of the bar lines using musical expertise.

As far as we are aware, no previous work exists on the generation of a musical score for percussive instruments from polyphonic audio signals. There are, however, numerous publications regarding the processing steps involved here: the detection and classification of musical events, the segmentation of the audio signal into similar regions and the estimation of the tatum grid and the metric structure of music. For an overview of research on the detection and classification it is referred to [1]. Other previous research of interest to his work is reviewed in Section 1.2. The transcription method is described in Section 2, experimental results are presented in section 3, and conclusions are drawn in Section 4.

1.1. Definition of Musical Terms

Since this publication deals with some culturally coined terms from the domain of music, a short explanation of the terms used herein is given. Two independent components of rhythmic organization exist: the grouping structure and the metrical structure [2]. A listener intuitively groups the events contained in a sound signal into distinct units such as motives, themes, and phrases. In parallel, he or she constructs a regular pattern of strong and weak beats to which he or she relates the actual musical sounds. Musical groups are heard in a hierarchical fashion, i.e. a motive appears as part of a theme, and a section containing several themes as part of a piece. The metrical structure of music is also ordered in a hierarchical manner. Meter can be defined as “the measurement of the number of pulses between more or less regularly recurring accents” [3] and exists only in the presence of pulse series. Here, a pulse is defined as one event of a series of equidistant identical stimuli. Different pulse series and meters occur on different hierarchic strata. The time spans between the pulses on the different rhythmic levels exhibit integer ratios for mono-rhythmic music. The pulse on the lowest level is the tatum [4]. The tatum may be established by the smallest time interval between two successive notes, but is in general best described by the pulse series that most highly coincidences with all note onsets. The pulse which determines the musical tempo is the beat. The integer ratio between the beat period and the tatum period is called micro-time [5]. Pulses at bar lines constitute a pulse series on a higher level. The number of beats per bar is specified by the nominator of the time signature. The denominator determines the temporal note value of the beat. The decision about the time signature is therefore crucial for the tempo estimation.

1.2. Previous Work

A large body of research is concerned with the temporal segmentation of audio signals. Different methods were proposed for the segmentation of speech and music, the segmentation into notes or phonemes and for the segmentation of musical audio signals into similar region. The latter task is of interest in our context. Foote proposed a method based on a measure of audio novelty [6]. Signal features are extracted from short successive frames in the spectral domain, and a similarity matrix is computed, comprising the distances between the feature vectors for each frame. A novelty measure is obtained by correlating the similarity matrix with a checkerboard-like kernel matrix. In [7], a segmentation method using Hidden Markov Models is presented. In a recent work [8], the audio signal is represented in a feature space by extracting low-level features of sliding windows of the signal. A matrix of similarity respectively dissimilarity measures is computed from the signal representation. A more discriminating representation is obtained by transforming the original feature vectors using *Linear Discriminant Analysis* (LDA) [8]. The segmentation is obtained by clustering the feature vectors using *Dynamic Programming*.

The estimation of the tatum is addressed in various publications. Gouyon et. al. [9] estimated the tatum period from previous detected note onsets. The most frequent *inter-onset-interval* (IOI) is detected in a histogram of IOI. The calculation of the IOI is not limited to successive onsets but rather to all pairs of onsets in a time frame. Tatum candidates are calculated as integer fractions of the most frequent IOI. The candidate is selected which best predicts the harmonic structure of the IOI histogram according to a two-way mismatch error function. The estimated tatum period is subsequently refined by calculating the error function between the comb grid derived from the tatum period and the onset times of the signal. Seppänen [10] proposed an IOI-based tatum estimation with a time-varying histogram representation in order to accommodate slight tatum changes (e.g. accelerando and ritardando). The phase of the tatum grid is obtained by minimizing the average deviation between the note onsets and the tatum grid elements.

The determination of musical meter from audio recordings is addressed in [11], [12] and [13]. Gouyon et. al. [11] investigated various low-level features and their application in an autocorrelation based processing for the determination of meter. Using a set of four descriptors, 95% of correct classifications were reached for the classification of the meter into one of the two classes “binary meter” and “ternary meter”. Klapuri [12] proposed a combined estimation of beat, meter and tatum. The approach presented there involves a decomposition of the signal energy into 36 bands using a Discrete Fourier Transform, from which the *μ*-law compressed power envelopes are calculated. The power envelopes were smoothed, differentiated and half-way rectified. The linear summation of adjacent bands yields four “registral accent signals”, which are fed into a bank of comb-filter resonators to estimate the strength of different pulse periods. A probabilistic model is applied for the interpretation of the detected periodicities. A previous experiment by Scheirer demonstrated that amplitude envelopes are a sufficient representation for rhythmic analysis [14]. In [13], the use of the autocorrelation function for the estimation of the meter from musical scores of single melodic lines is proposed.

An important prerequisite for the estimation of the musical meter from drum notes is a measure for the similarity between rhythmic pattern, since the perception of musical meter can be characterized as the detection of underlying periodicities [2], and some methods for the calculation of periodicities compare the signal to its shifted version. Rhythmic pattern can be represented by a matrix T_{ij} with $i=1\ldots n$ and $j=1\ldots m$, whereas n denotes the number of instruments and m is the number of tatum grid elements. The patterns are assumed to be quantized. The matrix T can be either a Boolean matrix, where events are marked by ones, or it may consist of the velocity or intensity values of the events. Distance measures for rhythmic patterns represented by Boolean matrices include the Hamming distance [15] and the interval vector distance [16]. The interval vector distance is obtained from the Euclidian distance between to patterns represented as vectors of IOI. In [17], rhythmic patterns are represented as a “difference of rhythm vector”, a vector of ratios of IOI between consecutive notes. The distance measure between two patterns, named rhythm error, is defined as

$$e = \left(\sum_{j=1}^{n-1} \frac{max\left(r_j, s_j\right)}{min\left(r_j, s_j\right)} \right) - \left(n-1\right) \qquad (1\ )$$

Here, r and s denote the two patterns and n is the length of each pattern.

2. METHOD

2.1. Detection and Classification of percussive events

The detection and identification of percussive events is explained in detail in [1] and therefore only a short description of the applied method is given here. Events probably assigned to onsets of percussive instruments are detected using a suitable detection function derived from differential magnitude sum as well as phase congruency information of the musical signals spectrogram. Slices of the spectrogram are cut out at the note onset times and subjected to *Non-negative Independent Component Analysis (NICA)* [18]. The extracted independent frequency bases are furthermore regarded as spectral profiles of the contained instruments and their corresponding amplitude bases are interpreted as detection function for the occurrence of the events in time. Salient peaks in the amplitude envelopes near the onset-times are accepted as percussive events. A number of spectral and time-based features is used to eliminate spectral profiles stemming from harmonic sustained sounds and to classify the percussive instruments.

2.2. Segmentation

The applied segmentation procedure for musical audio signals is based on a method initially proposed by Foote [6]. In our current implementation, the audio signal is divided into *n* adjacent frames of 30 milliseconds length each. The time signal is transformed into the frequency domain using the *Fast Fourier Transform* (FFT). A feature vector v_i, $i=1\ldots n$, is calculated for each frame, combining the *Audio Spectrum Envelope* (ASE), the *Spectral Flatness Measure* (SFM) and the *Mel-Frequency Cepstral Coefficients* (MFCC) for a number of frequency bands. Subsequently, four adjacent feature vectors are grouped by averaging the feature values per frame.

A similarity matrix *S* is constructed by calculating a distance measure between all pairs of feature vectors $D_{ij}=f(v_i,v_j)$, and arranging the distances so that the matrix element S_{ij} corresponds to the distance D_{ij}. From the variety of distance metrics, the *Cosine Distance* D_c was chosen.

$$D_c(i,j) = \frac{v_i \cdot v_j}{|v_i| \cdot |v_j|} \tag{2 }$$

From the similarity matrix, a novelty measure is computed by correlating *S* with checkerboard-like kernel matrix *K* along the diagonal.

$$K = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{3 }$$

The size of the kernel influences the width of the separated audio segments in a way that larger kernels average over short-time novelty and separate longer segments. The kernel size is enlarged by forming the *Kronecker product* of *K* with a quadratic matrix of ones and the edges are tapered with a *Gaussian* distribution. The novelty measure shows peaks at time points between segments. Because both *S* and *K* are symmetric matrices, only one half need to be computed. Furthermore, *S* is only computed for pairs of $\{v_i,v_j\}$ with $j-i \le l/2$, where *l* denotes the size of *K*. The novelty measure is smoothed using an FIR-Low-Pass-Filter and segment boundaries are detected using an *n*-point running window method.

2.3. Quantization of the detected events

The detected events are quantized on the tatum grid. The tatum grid is estimated using the note onset times of the detected events together with note onset times derived by means of a conventional note onset detection method. The generation of the tatum grid on basis of the detected percussive events only fails than short section appear without any drums and therefore an additional note onset detection is applied here.

2.3.1. Conventional note onset detection

For the detection of note onsets, amplitude envelopes of 19 frequency bands ranging from 44 Hz to 11276 Hz are extracted from the audio signal by means of a *Short-Time Fourier Transform* (STFT). A window size of 1024 samples with a hop size of 128 samples is chosen for input audio files with 44.1 kHz sampling frequency. The amplitude envelopes E_i are calculated by combining the amplitudes of the frequency bins belonging to one frequency band and smoothing the signal by means of convolution with a *Hanning-window* of 100 milliseconds length. The relative difference function D_i is computed from the amplitude envelopes E_i as

$$D_i = diff(log(E_i)) \tag{4 }$$

where the operator *diff* is the difference function, *log* is the natural logarithm and *i* is the band index. Note onsets are detected by searching for maxima in the relative difference function of each frequency band above a static and an adaptive threshold. The adaptive threshold decays from a value which is set on the detection of a preceding note onset to half of its value after 200 milliseconds. It is furthermore required that E_i drops below a threshold between two subsequent note onsets that is computed as a function of the envelope values of the two peaks corresponding to the onset times. An arbitrary intensity value is calculated for each note onset as the maximum value of the difference function at the note onset time, and is subsequently weighted by the middle ear transfer function. The phase congruency information, calculated by summation of the phase, is applied to detect note onset times more accurately [19].

2.3.2. Tatum Grid Estimation

Two alternative approaches to the estimation of the tatum grid are compared. As a first approach, the tatum grid is estimated using a *two-way mismatch procedure* (TWM). Originally proposed for the fundamental frequency estimation [20], this method has also been applied before to the estimation of the tatum grid, as described in [9]. A series of trial values for the tatum period is derived from the histogram of IOI. Various authors suggest the calculation of IOI between all note onset times within a certain range rather than between adjacent notes only. The histogram of the IOI is generated and smoothed by means of an FIR-low-pass filter. Tatum candidates are obtained by dividing the IOI corresponding to peaks in the IOI-histogram by a set of values *v={1, 2, 3, 4}*. A raw estimate of the tatum period is derived from the IOI-histogram after applying the TWM. Subsequently, the phase of the tatum grid and a more exact estimate of the tatum period are computed by means of the TWM between the note onset times and several tatum grids with periods near the previously estimated tatum period.

The second method refines and adjusts the tatum grid by computing the best match between the note onsets vector and the tatum grid utilizing the *correlation coefficient* R_{xy} between the note onset vector x and the tatum grid y.

$$R_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (5\ )$$

To follow small tempo variations, the tatum grid is estimated for adjacent frames of 2.5 seconds length each. Transitions between the tatum grids of adjacent frames are smoothed by low-pass-filtering the IOI vector of the tatum grid points and reconstructing the tatum grid from the smoothed IOI vector. Subsequently, each event is assigned to its nearest grid position. The score can then be written as a matrix T_{ik}, $i=1\ldots n$ and $j=1\ldots m$, with n denoting the number of detected instruments and m equalling the number of tatum grid elements. The intensities of the detected events can be either adopted or discarded, yielding a Boolean matrix.

2.4. Estimation of the time signature

The quantized representation of percussive events delivers valuable information for the estimation of musical meter. The periodicity on the bar level is identified in two steps: calculation of a periodicity function and estimation of the bar length.

Common periodicity functions are the *autocorrelation function* (ACF) and the *average magnitude difference function* (AMDF) as illustrated in equation (6) and (7) respectively, where x is the signal and τ denotes the lag.

$$ACF(\tau) = \sum_{i=1}^{\tau} x_i x_{i+\tau} \qquad (6\ )$$

$$AMDF(\tau) = \sum_{j=1}^{\tau}\left(x_j - x_{j+\tau}\right)^2 \qquad (7\ )$$

The AMDF has been successful applied to the estimation of the fundamental frequency for musical and speech signals [21] and to the estimation of musical meter [22].

In the general case, a periodicity function measures the similarity i.e. dissimilarity between the signal and its time shifted version. Various similarity measures are reported in the literature [13], [16], [17]. The *Hamming Distance* (HD) is known from the information theory [15] and calculates the dissimilarity between two Boolean vectors b_1 and b_2 according to equation (8).

$$HD = sum\left(b_1 \underline{\vee} b_2\right) \qquad (8\ )$$

An appropriate extension for the comparison of rhythmic patterns is different weighting of simultaneous hits and rests. The similarity B between two sections of the score T_1 and T_2 is than calculated by weighted summation of the Boolean operations

$$B = a \cdot T_1 \wedge T_2 + b \cdot \neg T_1 \wedge \neg T_2 - c \cdot T_1 \underline{\vee} T_2 \qquad (9\ )$$

where the weights a, b and c are initially set to $a=1$, $b=0.5$ and $c=0$. The similarity measure M is obtained by summation of the elements of B, as in equation (10).

$$M = \sum_{i=1}^{n}\sum_{j=1}^{m} B_{ij} \qquad (10\ )$$

This similarity measure resembles the *Hamming distance* in a sense that it considers the differences between matrix elements in a similar manner. In the following explanations, the distance measure derived from equation (10) is named *modified Hamming distance* (MHD). Additionally, the influence of distinct instruments can be controlled by means of a weighting vector v_i, $i=1\ldots n$, which can be set either using musical knowledge, e.g. assigning more importance to snare drums or to deep resonating instruments, or depending on the frequency and regularity of occurrence of the instruments.

$$M_v = \sum_{i=1}^{n} v_i \cdot \sum_{j=1}^{m} B_{ij} \qquad (11)$$

Additionally, the similarity measures for Boolean matrices can be extended by weighting B with the mean value of T_1 and T_2 in order to incorporate the intensity values. Distances respectively dissimilarities are interpreted as negative similarities in the following evaluation. The periodicity function $P=f(M, l)$ is obtained by calculating the similarity measure M between the score T and its shifted self by lag l. The time signature is estimated by comparing P to a number of metric models. The implemented metric models Q are constituted by a train of spikes at typical accent positions for different time signatures and micro-times. The best match between P and Q is obtained if the correlation coefficient assumes its maximum. In the current state of the system, 13 metric models for 7 different time signatures are implemented.

2.5. Detection of recurring patterns of the percussive events

Recurring patterns are detected in order to detect the preparatory beats and to derive a robust tempo estimate by applying musical knowledge. For the detection of drum patterns, a score T' is generated of the length of one bar b, by summation of the matrix elements T with similar metrical position,

$$T' = \sum_{k=1}^{p} T_{i,\,j+(k-1)b} \qquad (12)$$

where b denotes the estimated bar length and p the number of bars in T. In the following, T' is named the score histogram. Drum patterns are obtained from the score histogram T' by searching for score elements $T'_{i,j}$ with large histogram values. Patterns of a length of more than one bar are retrieved by means of repetition of the above described procedure for integer multiples of the measure length. The pattern length with the most hits relatively to the pattern length is chosen to derive the most representative pattern.

2.6. Generation of the musical score

The identified rhythmic patterns are interpreted using a set of rules derived from musical knowledge. In the current stage of the development, fairly simple concepts have been applied. Equidistant occurrences of single instruments are identified and evaluated with respect to the instrument class. This leads to identification of playing styles which occur frequently in popular music. One example is the very frequent use of snare drums or hand claps on the second and fourth beat in four-four time. This concept, named “backbeat”, serves as an indicator for the position of the bar lines. If the “backbeat”-pattern is present, the bar starts between snare strokes.

Another cue for the positioning of the bar lines is the occurrence of kick drum events, which is represented by means of a histogram. It is assumed that the start of a musical measure is marked by the metric position where most kick drum notes occur. Although the applied ideas are simple, they are very powerful for the analysis of popular music. The generated musical scores are subsequently represented by a matrix similar to a “piano roll” representation.

3. EXPERIMENTS AND RESULTS

An important prerequisite for a correct generation of a musical score and successful tempo and meter estimation is a correct quantization of the detected events on the tatum grid. In a first experiment, the tatum estimation methods are compared. The time signatures, tempos, micro-times and tatum periods from 161 musical excerpts of 30 seconds length each are automatically estimated and compared to manually detected values. The results in terms of correct classifications of all rhythmic features and of the tatum period are illustrated in Table 1. The correct estimation of time signature, tempo and micro-time is valuable indicator for the performance of the estimation of the tatum grid. An item was correctly classified, if all rhythmic features together (time signature, tempo and micro-time) were estimated correctly. The results are presented separately for each of the 465 segments, and for the most representative segment of each excerpt. The most representative segment is defined here as the segment with the largest temporal duration. Only segments were considered with a minimum duration of 6 seconds. In the first experiment, the modified Hamming distance has been applied and the intensity values have not been considered for the periodicity calculation. There are two common errors in meter estimation process: the confusion of the correct tempo with its double or half value and the misinterpretation of ternary four-four time as six-eight time or vice versa, resulting in a wrong estimation of tempo and micro-time (whereas the length of the musical measure is estimated correctly).

	TWM	CM
correct	88.8% / 84.0%	81.3% / 73.9%
tatum period	95.6% / 95.0%	92.3% / 91.3%

Table 1: Results of the two-way mismatch procedure (TWM) and the correlation method (CM), for the most representative segment (left) and all segments (right)

The estimation of the tatum grid is affected by timing deviations and erroneously detected events. The two-way mismatch procedure shows a more robust behaviour than the correlation method and is used in the following experiments. Although there is no large difference in the estimation of the tatum period, the results of the meter estimation indicate a better performance regarding the tatum tracking if the TWM is applied.

The influence of different similarity measures on the estimation of time signature, tempo and micro-time has been investigated with and without consideration of the intensity values of detected notes. The results are illustrated in Table 2. The presented values were obtained from the most representative segments. The best results were obtained using the modified Hamming distance without consideration of the intensity values. Wrong tempo estimates were caused by confusion with the double or half tempo of the correct tempo, or by confusion between binary and ternary micro-time, e.g. for pieces in six-eight time.

	MHD	ACF	AMDF
without intensities	88.8%	88.2%	81.3%
with intensities	83.8%	84.4%	76.9%

Table 2: Comparison of the performance of different similarity measures for the estimation of time-signature, tempo and micro-time

The influence of segmentation on the analysis was examined by comparing the meter estimation results from the analysis with and without precedent segmentation. The results in terms of correct estimated rhythmic features and tatum periods are illustrated in Table 3.

	with Segmentation	without Segmentation
correct	88.8%	75.7%
tatum period	95.6%	89.1%

Table 3: Experimental results of the investigation on the influence of precedent segmentation

In a third experiment, the scores from 40 musical excerpts of 30 seconds duration were generated and manually evaluated. Only excerpts with correct estimation of tatum period, tempo, and time signature have been considered. The evaluation has been carried out in an audio-visual manner, where errors in the generated score were detected by comparing the visualized and synthesized score to the original audio signal. The error types in this experiment are an insertion or a miss of tatum grid elements and a wrong positioning of the bar lines. Insertions or misses occurred 3 times and were mainly located at segment boundaries. The positioning of the bar lines failed in 5 cases. The error occurred if no rhythmic patterns were detected. The assumption, that the start of a measure is marked by the metric position with most occurrences of kick drum events is not appropriate in many cases.

Some problems occurred in the analysis of the musical items. Although the applied segmentation procedure does successful identify segment boundaries of the musical signal, variations of rhythmic pattern occur within segments and lead to errors in their recognition. The tempo was estimated falsely as half of the correct tempo due to misinterpreted back-beat patterns played by the kick drum, e.g. in reggae music. A confusion of the tempo value with the double or half occurred frequently in cases where no patterns were identified.

4. CONCLUSIONS AND FUTURE WORK

A system has been presented for the generation of musical scores of percussive instruments from automatically detected events. This process involves the analysis of the audio signal with respect to its rhythmic structure. A number of interesting conclusions have been obtained from experimental results. Musical knowledge can be applied if the occurring musical instruments in the signal are previously known. This increases the performance of the estimation of the rhythmic features under consideration. The precedent segmentation is advantageous for the analysis of musical items. The presented methods works well in many cases, but it fails if the drums play very expressive. The estimation of recurring patterns enables a robust analysis even if the automatic detection of the events is not reliable, e.g. if the analyzed signal features low audio quality or quiet percussive sounds. A crucial point in the analysis is the estimation of the tatum grid, especially if the model of equidistant pulses is not appropriate. The consideration of the intensity values of the detected events leads to degradation of the performance. There is some room for improvements. The training of various other patterns would probably increase the performance of the method regarding the estimation of tempo and measure length and positioning of the bar lines. An additional segmentation procedure considering the detected events would yield improvements in the recognition of the rhythmic patterns.

5. ACKNOWLEDGEMENTS

The authors thank Juergen Herre and Markus Cremer for proof-reading the paper and valuable suggestions, and Markus van Pinxteren for the improvement and implementation of the segmentation algorithm.

6. REFERENCES

[1] Dittmar, C., Uhle, C., “Further Steps towards Drum Transcription of Polyphonic Music”, to be presented at the *AES 116th Convention*, 2004.

[2] Lehrdahl, F., Jackendoff, R., *A Generative Theory of Tonal Music*, MIT Press, 1983.

[3] Cooper, G., Meyer L. B., *The Rhythmic Structure of Music*, University of Chicago Press, 1960.

[4] Bilmes J. A., *Timing is of Essence*, MSc Thesis, Massachusetts Institute of Technology, 1993.

[5] Marron, E., *Die Rhythmiklehre*, (in German), AMA Verlag, 1990.

[6] Foote, J., “Automatic Audio Segmentation Using a Measure of Audio Novelty”, in *Proc. of the IEEE Int. Conf. on Multimedia and Expo*, vol. 1, pp. 452-455, 2000.

[7] Aucouturier, J.-J., Sandler, M., “Segmentation of Musical Signals Using Hidden Markov Models”, in *Proc. of the AES 110th Convention*, 2001.

[8] Goodwin, M. M., Laroche, J., “Audio Segmentation by Feature-Space Clustering Using Linear Discriminant Analysis and Dynamic Programming”,in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003.

[9] Gouyon, F., Herrera, P., Cano, P., “Pulse-Dependent Analysis of Percussive Music”, in *Proc. of the AES 22nd Int. Conference on Virtual, Synthetic and Entertainment Audio*, 2002.

[10] Seppänen, J., “Tatum Grid Analysis of Musical Signals”, in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 101-104, 2001.

[11] Gouyon, F., Herrera, P., “Determination of the Meter of Musical Audio Signals: Seeking Recurrences in Beat Segment Descriptors”, in *Proc. of the AES 114th Convention*, 2003.

[12] Klapuri, A. P., “Musical Meter Estimation and Music Transcription”, presented at the *Cambridge Music Colloquium*, 2003.

[13] Brown, J. C., “Determination of the Meter of Musical Scores by Autocorrelation”, in *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1953-1957, 1993.

[14] Scheirer, E. D., “Tempo and Beat Analysis of Acoustic Musical Signals”, in *Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 588-601, 1998.

[15] Hamming, R. W., *Coding and Information Theory*, Prentice-Hall, 1986.

[16] Toussaint, G. T., „A Mathematical Analysis of African, Brazilian and Cuban Clave Rhythms“, in *Proc. of BRIDGES: Mathenatical Connections in Art, Music and Science,* Towson University, USA, 2002.

[17] Coyle, E. J., Shmulevich, I., “A system for machine recognition of musical patterns”, in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 1998.

[18] Plumbley, M., “Algorithms for Non-Negative Independent Component Analysis”, in *IEEE Transactions on Neural Networks*, 14 (3), pp. 534-543, 2003.

[19] McDonald, S., “Biologicalesque Transcription of Percussion”, in *Proc. of the Australasian Computer Music Conference*, 1998.

[20] Maher, R. C., Beauchamp, J. W., “Fundamental Frequency Estimation Of Musical Signals Using A Two-Way Mismatch Procedure”, in *Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2254-2263, 1994.

[21] de Cheveigné, A., Kawahara, H., “YIN, a fundamental frequency estimator for speech and music”, in *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917-1930, 2002.

[22] Paulus, J., Klapuri, A., “Measuring the Similarity of Rhythmic Patterns”, in *Proc. of the 3rd Int. Conf. on Music Information Retrieval (ISMIR)*, 2002.