



Audio Engineering Society

Convention Paper 6031

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Further Steps towards Drum Transcription of Polyphonic Music

Christian Dittmar¹, Christian Uhle²

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

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

ABSTRACT

This publication presents a new method for the detection and classification of un-pitched percussive instruments in real world musical signals. The derived information is an important pre-requisite for the creation of a musical score, i.e. automatic transcription, and for the automatic extraction of semantic meaningful meta-data, e.g. tempo and musical meter. The proposed method applies Independent Subspace Analysis using Non-Negative Independent Component Analysis and principles of Prior Subspace Analysis. An important extension of Prior Subspace Analysis is the identification of frequency subspaces of percussive instruments from the signal itself. The frequency subspaces serve as information for the detection of the percussive events and the subsequent classification of the occurring instruments. Results are reported on 40 manually transcribed test items.

1. INTRODUCTION

1.1. Motivation

Where the description of musical audio signals by means of metadata is concerned, an important branch constitutes the analysis of rhythm. Although rhythm is an essential concept for musical structure, which is contained in the voices of all sounding instruments there's little doubt that especially percussive instruments contribute to the rhythmical impression. High level description of any rhythmical content is only feasible when drum scores are available. This

information enables further categorization of musical content such as classification of genre (based on characteristic rhythmical patterns) determination of rhythmic complexity, expressivity and "groove" of a musical item. The measurement of less subjective descriptors like tempo and musical meter significantly benefits from the availability of a drum score as well. Thus, automated extraction of the drum score is an essential tool for cataloguing musical content and is able to contribute to today's music retrieval algorithms immensely.

1.2. State of the Art

The transcription of percussive un-pitched instruments represents a less challenging task than the comprehensive transcription of all played instruments in a musical piece, including harmonic sustained instruments. This is due to a number of reasons. First, no melodic information has to be detected, since with most percussive sounds pitch plays only a subordinate role. Second, percussive instruments commonly do not produce sustained notes (there are numerous exceptions, e.g. guiro, cymbal crescendos, as well as instruments residing in a grey area between un-pitched and pitched instruments, e.g. the Brazilian quica), so the duration of the notes has not to be detected in the general case. The challenge with identification of percussive instruments resides in the fact, that a great variety of sounds can be generated using a single instrument. This work focusses on the vast field of popular music, and only a limited set of percussive un-pitched instruments is presumed to be present. There are mainly two instruments classes in scope: membranophones and idiophones (as well as their electronic counterparts). The membranophones usually occupy the lower frequency regions of the audio signal. To name a few examples ordered according to their dominant frequency regions kick drum, tomtom, snare drum, timbales, conga and bongo are enumerated here. The list of examples can be continued with respect to the dominant frequency range by idiophones like woodblock, shaker, cymbal, tambourine and hi-hat. Unfortunately for the retrieval task, the instruments are not clearly separable along the frequency axis and there are many ambiguities due to different playing techniques and styles, recording situations and electronic effects, which are eventually applied to drum-sounds.

Previous work on the transcription of percussive instruments includes the doctoral thesis by Schloss [1], which addresses the transcription of pure percussive music. The developed system detects note onsets from the slope of the amplitude envelope and subsequently identifies the source of each note. The events are classified into damped and un-damped strokes, and subsequently into high and low frequency drum-sounds. The analyzed percussive instruments are membranophones, exclusively. The resulting note-list is used for the metrical analysis.

Other work relating to the detection and classification of events in musical audio signals containing only drum-sounds is described in [2], [3]. Gouyon et al. presented a system for automatic labeling of short drum kit performances, in which the instruments do not occur

simultaneously. The audio signal is segmented using a tatum grid, and each segment is represented as a vector of low-level features (e.g. spectral kurtosis, temporal centroid and zero-crossing rate). Various clustering techniques were examined to identify similar instrument sounds. Paulus et. al. described a system for the labelling of synthesized drum sequences with simultaneously occurring sounds using higher-level statistical modelling with n-grams. A manually detected tatum grid is applied for the segmentation of the drum tracks.

A number of authors have suggested systems for the detection and classification of percussive instruments in the presence of pitched instruments. McDonald proposes the use of a bank of wavelet filters to produce a spectrogram of the audio signal. The spectrogram is further processed by a bank of Meddis Inner Hair Cell models for the detection of note onsets. Note onsets are detected from the amplitude data in sub-bands of one octave width scaled with the phase congruency per sub-band. The detected events are then classified using the similarity between the sonogram data of a short excerpt following an onset and trained samples [4].

An analysis/synthesis approach to extraction of drum-sounds from polyphonic music is presented in [5]. The extraction of the two dominant percussive instruments and their occurrences is done by an iterative correlation method of matching a simple drum model with the actual drum-sounds in the analyzed signal. The extracted drum-sounds are not explicitly classified but subjected to be used as an audio-signature for the signal.

Some of the most recent work relates to the decomposition of the audio-signal using Independent Subspace Analysis (ISA). Casey et al. introduced this method for separation of sound sources from single channel mixtures. No explicit focus on percussive instruments has been emphasized, but the decomposition of a drum-loop into single sounds is featured as an illustrative example [6].

Iro Orife adopts ISA to separate and detect salient rhythmic and timing information with regard to a better understanding of rhythm, as well as computer based performance and composition [7].

In [8] ISA is employed to separate real world musical signals into percussive and harmonic sustained fragments using a decision-network based on measures describing the spectral and time-based features of a fixed number of independent components.

Further developments were conducted by Fitzgerald et al. [9] through introducing the principal of Prior Subspace Analysis, where generalized spectral profiles

for different percussive instruments are used to extract amplitude basis functions, which are then subjected to ICA to achieve statistical independence. Peak picking in such separated amplitude bases enables onset detection corresponding to the occurrence of drum instruments a priori assumed to be contained in the musical signal. The application of PSA in terms of detection and classification of drum instruments has moved from percussive to polyphonic music with promising results in [10]. This step is motivated by the assumption that the drum instruments are stationary in pitch.

2. SYSTEM OVERVIEW

2.1. Blockdiagram

An overview of the proposed system is presented in figure 1. The subsequent sections will give a more in depth account of the different stages endorsed the signal processing chain.

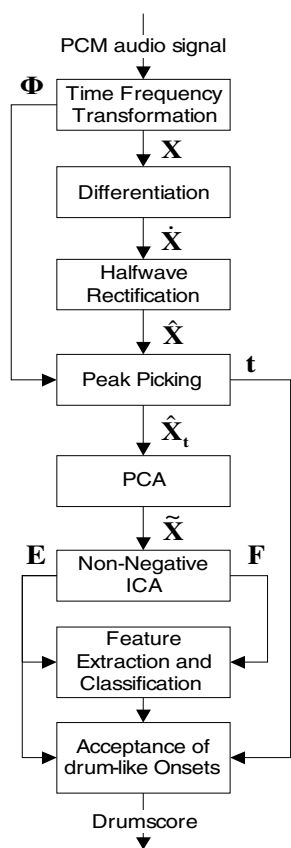


Figure 1 System Overview

2.2. Spectral Representation

The digital audio signals used for further analysis are mono files with 16bit per sample at a sampling frequency of 44.1kHz. They are submitted to pre-processing in the time domain using a software-based emulation of an acoustic effect device often referred to as Exciter. In this context, the Exciter stage emphasizes the higher frequency content of the audio signal. This is achieved by applying non-linear distortion to a high-pass filtered version of the signal and adding that distorted signal to the original. It turns out, that this is a vital issue when assessing hi-hats or similar high sounding idiophones with low intensity. Their energetic weight in respect to the whole musical signal is increased by that step, while most harmonic sustained instruments and lower sounding drum types are not affected. Another positive side effect consists in the fact that formerly MP3-encoded (and in the process low pass filtered) files can regain higher-frequency information to some extent. A spectral representation of the pre-processed time signal is computed using a Short Time Fourier Transformation (STFT). Thereby a relatively large block-size and high overlap are necessary due to two reasons. First the need for a fine spectral resolution in the lower frequency bins has to be fulfilled. Second the time resolution is increased to a required accuracy by a small hop-size between adjacent frames. From the above mentioned steps a spectrogram representation of the original signal is derived. The unwrapped phase-information Φ and the absolute spectrogram values X are taken into further consideration. The magnitude spectrogram X possesses n frequency bins and m frames. The time-variant slopes of each spectral bin are differentiated over all frames in order to decimate the influence of sustained sounds and to simplify the subsequent detection of transients. The differentiation leads to some negative values, so a half wave rectification is appended to remove this effect. This way, a non-negative difference-spectrogram \hat{X} is computed for the further processing.

2.3. Event Detection

The detection of multiple local maxima associated with transient onset events in the musical signal is conducted in a quite simple manner. At first a time tolerance is defined which separates two successive drum onsets. In this implementation 68ms have been used as a constant value that is translated to the time resolution in the spectral domain where it determines the number of frames which must at least occur between two consecutive onsets. The usage of this minimum distance

was proposed in [11] and is also supported by the consideration that a sixteenth note lasts 60ms at an upper tempo limit of 250bpm, which is quite close to the value presumed above. To derive a detection function on which the peak-picking can be executed the spectral bins of the differentiated spectrogram are simply summed up. A relatively smooth function \mathbf{e} is obtained by convolving the summed spectrogram with a suitable Hann window. To achieve the positions \mathbf{t} of the maxima a sliding window of the tolerance length is then shifted along the whole vector \mathbf{e} thus achieving the ability to detect one maximum per step. The trick is now to keep only those maxima stored which appear in the window for longer terms, because these are very likely the peaks of interest. The unwrapped phase information of the original spectrogram serves as reliability function in this context. It can be observed that a significant positive phase jump must occur near the estimated onset-time \mathbf{t} in order to avoid mistaking small ripples for onsets. The main concept of the further process is the storage of a short excerpt of the difference-spectrogram $\hat{\mathbf{X}}$ (namely one frame) at the time of the onset. From these frames the significant spectral profiles will be gathered in the next stages.

2.4. Reduction of Dimensionality

From the steps described in the preceding section the information about the time of occurrence \mathbf{t} and the spectral composition of the onsets $\hat{\mathbf{X}}_t$ is deduced. With real-world musical signals, one quite frequently encounters a high number of transient events within the duration of the musical piece. Even the simple example of a 120bpm piece shows that there could be 480 events in a 4 minute excerpt given the case that only quarter notes occur. With regard to the goal of finding only a few significant subspaces Principal Component Analysis (PCA) is applied to $\hat{\mathbf{X}}_t$. Using this well known technique it is possible to break down the whole set of collected spectra to a limited number of decorrelated principal components, thus resulting in a good representation of the original data with small reconstruction error. For this purpose an Eigenvalue Decomposition (EVD) of the dataset's covariance matrix is computed. From the set of eigenvectors the ones related to the d largest eigenvalues are chosen to provide the coefficients for a linear combination of the original vectors according to equation (1).

$$\tilde{\mathbf{X}} = \hat{\mathbf{X}}_t \cdot \mathbf{T} \quad (1)$$

Thereby, \mathbf{T} describes a transformation matrix which is actually a subset of the manifold of eigenvectors. Additionally the reciprocal values of the eigenvalues are incorporated as scaling factors yielding not only a decorrelation but also a variance normalization, which in turn implies whitening [12]. Alternatively a Singular Value Decomposition (SVD) of $\hat{\mathbf{X}}_t$ according to [6], [8] can achieve the same goal. With small modifications it is proven to be equivalent to the PCA using EVD [13]. The whitened components $\tilde{\mathbf{X}}$ are subsequently fed into the ICA-computation stage described in the next section.

2.5. Non-Negative Independent Component Analysis

Independent Component Analysis is a technique that is applied for separation of a set of linear mix signals into their original sources. A requirement for optimum performance of the algorithm is the statistical independency of the sources. Over the last years, extremely active research has been conducted in the field of ICA. One very interesting approach is the recent Non-Negative ICA [14], [15]. Where other commonly deployed algorithms like JADE-ICA [16] or FAST-ICA [17] exploit higher order statistics of the signals, Non-Negative ICA uses the very intuitive concept of optimising a cost function describing the non-negativity of the components. This cost function is related to the reconstruction error introduced by axis pair rotations of two or more variables in the positive quadrant of the joint probability density function (PDF). The assumptions for this model imply that the original source signals are positive and well grounded, which means they exhibit a non-zero PDF at zero, and they are to some extent linearly independent. The first concept is always fulfilled for the data considered in this publication, because the vectors subjected to ICA originate from the differentiated and halfwave rectified version $\hat{\mathbf{X}}$ of the amplitude-spectrogram \mathbf{X} , which does not contain any values lower than zero, but certainly some values at zero. The second constraint is taken into account when the spectra collected at onset times are regarded as the linear combinations of a small set of original source-spectra characterizing the involved instruments. This seems, of course, to be a rather coarse approximation, but it holds up well in the majority of the cases. The onset-spectra of real-world drum instruments do not exhibit invariant patterns, but are more or less subjected to changes in their spectral composition. Nevertheless, however, it may safely be

assumed that there are some characteristic properties inherent to spectral profiles of drum-sounds [9] that allow us to separate the whitened components $\tilde{\mathbf{X}}$ into their potential sources \mathbf{F} according to (2).

$$\mathbf{F} = \mathbf{A} \cdot \tilde{\mathbf{X}} \quad (2)$$

Where \mathbf{A} denotes the $d \times d$ unmixing matrix estimated by the ICA-process, which does actually separate the individual components $\tilde{\mathbf{X}}$. The sources \mathbf{F} will be named spectral profiles from here forth. Like the original spectrogram they own n frequency bins, but consist only of one frame. That means they only hold the spectral information related to the onset spectrum. To circumvent arbitrary scaling of the components introduced by PCA and ICA, a transformation matrix \mathbf{R} is constructed according to (3).

$$\mathbf{R} = \mathbf{T} \cdot \mathbf{A}^T \quad (3)$$

Normalizing \mathbf{R} with its absolute maximum value leads to weighting coefficients in a range $[-1 \dots 1]$ so that spectral profiles, which are extracted using

$$\mathbf{F} = \hat{\mathbf{X}}_t \cdot \mathbf{R} \quad (4)$$

possess values in the range of the original spectrogram. Further normalization is achieved by dividing each spectral profile by its L2-norm.

2.6. Crosstalk Profiles

As stated earlier the independence and invariance assumption for the given spectral slices suffer some weaknesses. So it is no surprise that the unmixed spectral profiles still display some dependencies. But that should not be regarded as erroneous behaviour. Tests with spectral profiles of single drum-sounds recorded under real-world conditions also yielded strong interdependence between the onset-spectra of different percussive instruments. One way to measure the degree of mutual overlapping and similarity along the frequency axis is the conduction of crosstalk measurements. As an illustrative metaphor the spectral profiles gained from the ICA-process can be regarded as transfer-function of highly frequency-selective parts in a filter-bank where overlapping pass-bands lead to crosstalk in the output of the filter-bank channels. The crosstalk measure between two spectral profiles is computed according to (5).

$$\mathbf{C}_{i,j} = \frac{\mathbf{F}_i \cdot \mathbf{F}_j^T}{\mathbf{F}_i \cdot \mathbf{F}_i^T} \quad (5)$$

for $i = 1 \dots d$, $j = 1 \dots d$ and $j \neq i$

In fact this value is related to the well known cross-correlation coefficient, but it uses a different normalization.

2.7. Extraction of amplitude bases

The preceding steps followed the main goal to compute a certain number of spectral profiles. These spectral profiles can be used to extract the spectrograms amplitude basis, from here forward referred to as amplitude envelopes according to (6).

$$\mathbf{E} = \mathbf{F} \cdot \mathbf{X} \quad (6)$$

As a second source of information the differentiated version of the amplitude envelopes can be extracted from the difference spectrogram according to (7).

$$\hat{\mathbf{E}} = \mathbf{F} \cdot \hat{\mathbf{X}} \quad (7)$$

This procedure is closely related to the principle of PSA. The main difference is that the priors used here are not some generalized class specific spectra. The second modification comprises in the fact that no further ICA-computation on the amplitude envelopes is applied. Instead, highly specialized spectral profiles very close to the spectra of the instruments really appearing in the signal are employed. Nevertheless the extracted amplitude envelopes are only in some cases nice detection functions with sharp peaks (e.g. for dance oriented music with predominant percussive rhythm tracks). Mostly they are accompanied by smaller peaks and plateaus stemming from the crosstalk effects mentioned above.

2.8. Component Classification

It is a well known problem [6] that the actual number of components is unknown for real world musical signals. ‘‘Components’’ is in this context used as general term for both the spectral profiles and the corresponding amplitude envelopes. If the number d of extracted components is too low artefacts of the suppressed component are likely to appear in some other components. If too many components are extracted the most prominent ones are likely to be split up into

several components. Unfortunately this division may even occur with the right number of components and accidentally suppress detection of the real components. Hence, special care has to be taken when considering the results.

This issue is approached by choosing a maximum number d of components in the PCA or ICA process, respectively. Afterwards, the extracted components are classified using a set of spectral-based and time-based features. The classification shall provide two sources of information. First, components should be excluded from the rest of the process which are clearly non-percussive. Second, the remaining components should be assigned to pre-defined instrument classes.

A suitable measure for the distinction of the amplitude envelopes is represented by the percussiveness, which is introduced in [8]. Here, a modified version is applied using the correlation coefficient between corresponding amplitude envelopes in $\hat{\mathbf{E}}$ and \mathbf{E} . The degree of correlation between both vectors tends to be small, when the characteristic plateaus related to harmonic sustained sounds occur in the non-differentiated amplitude envelopes \mathbf{E} . These are likely to almost disappear in the differentiated version $\hat{\mathbf{E}}$. Both vectors resemble each other far more in the case of transient amplitude envelopes originating from percussive sounds.

A spectral-based measure is constituted by the spectral dissonance, earlier described in [18], [8]. It is employed here to separate spectra of harmonic sustained sounds from the ones related to percussive sounds. In the implementation presented here, again a modified version of the computation of this measure is used, which exhibits tolerance to spectral leakage, dissonance with all harmonics and a suitable normalization. A higher degree of computational efficiency has been achieved by substituting the original dissonance function with a weighting matrix for frequency pairs.

The assignment of spectral profiles to a priori trained classes of percussive instruments is provided by a simple k-nearest neighbour classifier with spectral profiles of single instruments as training-database. The distance function is calculated from the correlation coefficient between query-profile and database-profile. To verify the classification in cases of low reliability (low correlation-coefficients) or several occurrences of the same instruments, additional features representing

detailed information on the shape of the spectral profile are extracted. These comprise global centroid, spread and skewness as measures describing the overall distribution. More advanced features are the center frequencies of the most prominent local partials, their intensity, spread and skewness.

2.9. Acceptance of drum-like onsets

Drum-like onsets are detected in the amplitude envelopes using conventional peak picking methods. Only peaks near the original times \mathbf{t} are regarded as candidates, the remaining ones are stored for further considerations. The value of the amplitude envelope's magnitude is assigned to every onset candidate at its position. If this value does not exceed a predetermined dynamic threshold then the onset is not accepted. The threshold varies over time according to the amount of energy in a larger area surrounding the onsets. Most of the crosstalk influences of harmonic sustained instruments as well as concurrent percussive instruments can be reduced in this step. Of crucial importance is the determination whether simultaneous onsets of distinct percussive instruments are indeed present or exist only due to crosstalk effects mentioned earlier. A simple solution is to accept those circumstantial instruments occurrences, whose value is relatively high in comparison to the value of the strongest instrument at the onset-time. Unfortunately, the relevance of this procedure in terms of musical sense is low.

3. RESULTS

3.1. Testdata

To quantify the abilities of the presented algorithm, drum scores of 40 excerpts from real world musical signals were extracted manually by trained listeners as a reference. Each excerpt consists of 30 seconds duration at 44.1 kHz samplingrate and 16 bits amplitude resolution. Different musical genres are contained among these examples featuring rock, pop, latin, soul and house only to name a few. They were chosen because of their distinct musical characteristics, and the intention to confront the system with a significant variety of possible percussive instruments and sounds.

3.2. Experimental Results

The drum scores automatically extracted by the proposed system were compared to the manually transcribed reference scores. The results are listed in table 1. The featured instruments represent the most frequently appearing drum types, for which the numbers are representative.

Class	Found	Missed	Added
Kick	83 %	9 %	23 %
Snare	75 %	21 %	35 %
Hi-hat	77 %	17 %	58 %
Cymbal	43 %	55 %	26 %
Shaker	60 %	35 %	93 %

Table 1 Drum Transcription Results

The detected onsets show deviations from the reference onsets. The average time difference is thereby ± 2 blocks. This value corresponds to approximately 19 ms and is below the presumed tolerance.

3.3. Discussion

Some common problems can be observed. For a small part of the test-files no satisfying separation of spectral profiles has been achieved. In those cases spectral profiles that were identified in the spectrogram by a trained human observer have not been extracted, resulting in missing instruments. That obviously happens especially when many of the components are assigned to harmonic sustained sounds. The presence of very prominent and dynamical harmonic sustained instruments (expressive singing voice, trumpet or saxophone solos) also tends to increase the number of spuriously found onsets. Even the selection of only drum-like peaks is error-prone to influences of quickly changing sustained components.

The separation of high sounding idiophones (hi-hat, cymbal, tambourine or shaker) can be particularly delicate because of their immensely overlapping spectral profiles. In contrast to lower sounding membranophones there are often no prominent partials, but a more or less broad distribution across the upper parts of the spectrum. This results in the indistinctiveness of the corresponding amplitude envelopes. So the decision if only one of those instruments is present at a certain onset time or whether

there are more of them, can not be simply deduced from intensity thresholds. That is the reason for the high numbers of erroneous shaker- and hi-hat onsets.

Unfortunately, direct comparison to the results presented in [10] is not feasible because of the disjunctive test data-bases and the wider range of percussive instruments considered in this publication.

4. CONCLUSIONS

In this paper a method for automatic detection and classification of un-pitched percussive instruments in real world music signals has been presented. The results are extremely promising when considering the extraction of significant rhythmical information rather than perfect note-to-note transcription. It can be expected that further improvements will be made in the near future with regards to the classification stage and the onset-acceptance. Furthermore, additional information has to be collected and algorithmic methods have to be invented in order to correctly assess the few exceptional situations where the ISA-model does not deliver the desired results.

5. ACKNOWLEDGEMENTS

The authors would like to thank Markus Cremer for proofreading of this paper and valuable suggestions to its clarity.

6. REFERENCES

- [1] W.A. Schloss, "On the automatic transcription of percussive music – from acoustic signal to higher-level analysis", PhD thesis, Stanford University, 1985
- [2] F. Gouyon and P. Herrera, "Exploration of Techniques for Automatic Labeling of Audio Drum Tracks' Instruments", in *Proc. of MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, 2001
- [3] J.K. Paulus, A.P. Klapuri, "Conventional And Periodic N-Grams In The Transcription Of Drum Sequences", in *Proc. of the IEEE International Conference On Multimedia and Expo*, Baltimore, USA, 2003

- [4] S. McDouglas, "Biologalesque Transcription of Percussion", in *Proc. Of the Australasian Computer Music Conference*, Canberra, 1998
- [5] A. Zils, F. Pachet, O. Delerue, F. Gouyon, "Automatic Extraction of Drum Tracks from Polyphonic Music Signals", in *Proc. of the IEEE*, 2002
- [6] M.A. Casey and A. Westner, "Separation of Mixed Audio Sources by Independent Subspace Analysis", in *Proc. of the International Computer Music Conference*, Berlin, 2000
- [7] I.F.O. Orife, "*Riddim: A rhythm analysis and decomposition tool based on independent subspace analysis*", Master thesis, Dartmouth College, Hanover, New Hampshire, 2001
- [8] C. Uhle, C. Dittmar and T. Sporer, "Extraction of Drum Tracks from polyphonic Music using Independent Subspace Analysis", in *Proc. of the Fourth International Symposium on Independent Component Analysis*, Nara, Japan, 2003
- [9] D. Fitzgerald, B. Lawlor and E. Coyle, "Prior Subspace Analysis for Drum Transcription", in *Proc. of the 114th AES Convention*, Amsterdam, 2003
- [10] D. Fitzgerald, B. Lawlor and E. Coyle, "Drum Transcription in the presence of pitched instruments using Prior Subspace Analysis", in *Proc. of the ISSC*, Limerick, Ireland, 2003
- [11] F. Gouyon, P. Herrera, P. Cano, "Pulse dependent Analyses of percussive music", in *Proc. of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002
- [12] A. Hyvärinen, J. Karhunen and E. Oja, "*Independent Component Analysis*", Wiley & Sons, 2001
- [13] A. Webb, "*Statistical Pattern Recognition*", Wiley & Sons, 2002
- [14] M. Plumbley, "Algorithms for Non-Negative Independent Component Analysis", in *IEEE Transactions on Neural Networks*, 14 (3), pp 534-543, May 2003
- [15] E. Oja, M. Plumbley, "Blind Separation of Positive Sources using Non-Negative PCA", in *Proc. of the Fourth International Symposium on Independent Component Analysis*, Nara, Japan, 2003
- [16] J.-F. Cardoso, A. Souloumiac, "Blind beamforming for non Gaussian signals", in *Proc. of the IEEE*, Vol. 140, no. 6, pp. 362-370, 1993
- [17] A. Hyvärinen, E. Oja, "A Fast and Robust Fixed-Point Algorithm for Independent Component Analysis", in *IEEE Transactions on Neural Networks*, 1999
- [18] W. Sethares, "Local Consonance and the Relationship between Timbre and Scale", in *Journal Acoust. Soc. Am.*, 94 (3), pt. 1, 1993