

# Musical Instrument Separation Applied to Music Genre Classification

Aldona Rosner<sup>1</sup>(✉) and Bozena Kostek<sup>2</sup>

<sup>1</sup> Institute of Informatics, Silesian University of Technology,  
Akademicka 16, 44-100 Gliwice, Poland

aldona.rosner@polsl.pl

<sup>2</sup> Audio Acoustics Laboratory, Faculty of Electronics, Telecommunications  
and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland

bokostek@audioakustyka.org

**Abstract.** This paper outlines first issues related to music genre classification and a short description of algorithms used for musical instrument separation. Also, the paper presents proposed optimization of the feature vectors used for music genre recognition. Then, the ability of decision algorithms to properly recognize music genres is discussed based on two databases. In addition, results are cited for another database with regard to the efficiency of the feature vector.

**Keywords:** Music Information Retrieval · Musical instrument separation · Music genre classification · Decision systems

## 1 Introduction

The subject of music genre classification, though visible not only in the Music Information Retrieval (MIR) research but also in commercial applications, still needs attention in terms of effectiveness and quality of classification [6–9, 12, 21]. This is especially important when applied to big music databases available for public [15]. One of the most popular query criteria are: artist, genre, mood, tempo or specific title, when looking for a specific audio track. Lately, genre has become one of the most popular choices, but not always this information is stored in a specific music track. That's why the subject of a more deep content exploring, i.e. taking into consideration sound source separation in the context of music recognition should be addressed, as separation of individual auditory sources, apart from instrument recognition and automatic transcription systems, may be very useful in genre classification. The instrument separation approach seems to be very useful in improving the genre recognition process, since the separation of individual auditory sources and instrument recognition systems get high applicability during the last few years. Therefore, one of the aims of this paper is to propose a set of parameters, which after the audio track separation preprocessing may enable to describe musical content of a piece of music in a more efficient way for the purpose of better distinguishing between selected musical genres. Such an operation may bring the efficiency of genre classification in the databases containing thousands of music tracks similar to earlier experiments carried out on small music databases.

This paper overviews related research, then presents databases and parameters used in experiments. In addition main principles of the algorithm used for musical instrument separation are presented. Then, proposed optimization of the feature vectors is shortly discussed. The following Section contains music genre classification results based on two decision systems, separation of music tracks and concerning two databases. A summary is also included providing the most important conclusions.

## 1.1 Related Research

Because of the extraordinary increase of the available multimedia data, it is necessary to make the recognition process automatic. Although the division of music into genres is subjective and arbitrary, there are perceptual criteria related to the texture, instrumentation and rhythmic structure of music that can be used to characterize a particular genre. Humans can accurately predict a musical genre based on 25 s, which confirms that we can judge the genre using only the musical surface, without constructing any higher level of theoretical descriptions.

The experiments presented by Tzanetakis and Cook (2002) [17] resulted in 61 % of effectiveness for 10 music genres, what is comparable to results aimed by human musical genre classification. The authors proposed three music feature sets for representing timbral texture, rhythmic content and pitch content. Other experiments were made by Kirss (2007) [10] on 250 musical excerpts from five different electronic music genres such as deep house techno, uplifting trance, drum and bass and ambient. By using SVM (Support Vector Machine) 96.4 % of accuracy was reached. It shows that it is possible to obtain a high evaluation efficiency on small datasets and small number of genres, which can also be classified as subgenres. The ISMIS 2011 contest left us with the final results [9] of almost 88 % of accuracy. The tests were concluded in the ISMIS database, which consists of 1310 songs of six genres. It is worth to mention that most misclassification was between Rock and Metal, and Classical and Jazz genres.

The importance of feature extraction in the topic of musical genre classification is confirmed by many researchers. In recent years, an extensive research has been conducted on the subject of audio sound separation, and resulted in interesting ideas and solutions. Uhle *et al.* (2003) [18] designed a system for drum beat separation based on Independent Component Analysis. In contrast, Smaragdīs and Brown (2003) [16] applied Non-Negative Matrix Factorization (NMF) to create a system for transcription of polyphonic music that showed remarkable results on piano music. Helen and Virtanen (2005) [3] used NMF, combined with a feature extraction and classification process, and achieved promising results in drum beat separation from popular music. Similar techniques were used by other researchers [1, 13, 19] in percussive-harmonic signal separation or in instrument separation [4].

It is noteworthy that musical social systems, in addition to the area of computer games, have become one of the most profitable financial ventures in recent years. That is why the development of such kind of applications makes it necessary to improve classification of music genre and other most searched criteria, which is still far from being satisfying. Despite major achievements in the field of MIR, there are still some challenges in this area, to name a few: the problem of scalability, large size of the data,

different standards (formats) of storing the music information (also media, e.g. a soundtrack), methods of transmitting multimedia data, varying degrees of compression, synchronizing the playback of various media elements, and others.

## 2 Genre Classification Experiments

### 2.1 Separation of Music Tracks

Separation of music tracks is based on the NMF method [11]. This method performs well in blind separation of drums and melodic parts of audio recordings. NMF performs a decomposition of the magnitude spectrogram  $\mathbf{V}$  ( $V \approx W \bullet H$ ) obtained by Short-Time Fourier Transform (STFT), with spectral observations in columns, into two non-negative matrices  $\mathbf{W}$  and  $\mathbf{H}$  (where  $W \in R_{\geq 0}^{n \times m}, H \in R_{\geq 0}^{n \times m}$  and constant  $r \in N$ ). Matrix  $\mathbf{W}$  resembles characteristic spectra of the audio events occurring in the signal (such as notes played by an instrument), and matrix  $\mathbf{H}$  measures their time-varying gains. Columns of  $\mathbf{W}$  are not required to be orthogonal as is in principal component method. Specifically, an approach based on an iterative algorithm for computing two factors based on the Kullback-Leibler divergence of  $\mathbf{V}$  given  $\mathbf{W}$  and  $\mathbf{H}$  is used in our experiments. This means that the factorization process is achieved by iterative algorithms minimizing cost-functions, which interprets the matrices  $\mathbf{V}$  and  $(\mathbf{W}, \mathbf{H})$  as probability distributions [14, 20].

Then, to each NMF component (column of  $\mathbf{W}$  and corresponding row of  $\mathbf{H}$ ) we apply a pre-trained SVM classifier to distinguish between percussive and non-percussive components. The task of this pre-trained SVM classification which bases on features such as harmonicity of the spectrum and periodicity of the gains is to distinguish between percussive and non-percussive signals bases. By selecting the columns of  $\mathbf{W}$  that are classified as percussive and multiplying them with their estimated gains in  $\mathbf{H}$ , we obtain an estimate of the contribution of percussive instruments to each time-frequency bin in  $\mathbf{V}$ . Thus, we can construct a soft mask that is applied to  $\mathbf{V}$  to obtain an estimated spectrogram of the drum part, which is transferred back to the time domain through the inverse STFT using the OLA (overlap-add) operation between the short-time sections in the inverting process. It should be reminded that the redundancy within overlapping segments and the averaging of the redundant samples averages out the effect of the window analysis. More details on the drum separation procedure can be found in the introductory paper by Weninger et al. (2011) [19].

### 2.2 Databases

Music Information Retrieval systems enable to search music basing on metadata: set of parameters which describe the specific track, track title, author, album, year, genre, etc. Another important topic of MIR is related to music databases available for research and experiments. Two music databases were employed in the experiments shown in this paper, namely ISMIS and SYNAT [8] databases.

The ISMIS music database was prepared for a data mining contest associated with the 19th Internat. Symp.on Methodologies for Intelligent Systems [9]. It consists of over 1300 music tracks of 6 music genres: classical, jazz, blues, pop, rock and heavy metal (see Table 1). For each of 60 performers there are 15-20 music tracks, which are then partitioned into 20-s segments and parameterized. Music tracks are prepared as stereo signals (44.1 kHz, 16 bit, .wav format).

**Table 1.** List of cardinality of the music genres for ISMIS database

Genre	No. of music excerpts
Classical	320
Jazz	362
Blues	216
Rock	124
Heavy Metal	104
Pop	184
Overall	1310

The SYNAT database consists of over 50.000 music tracks of 30-second long excerpts of songs in mp3 format, retrieved from the Internet by a music robot. ID3 tags of music excerpts were automatically assigned to songs by the music robot. The tags were saved in a fully automatic way without human control [5]. SYNAT contains 22 genres: Alternative Rock, Blues, Broadway and Vocalists, Children’s Music, Christian and Gospel, Classic Rock, Classical, Country, Dance and DJ, Folk, Hard Rock and Metal, International, Jazz, Latin Music, Miscellaneous, New Age, Opera and Vocal, Pop, Rap and Hip-Hop, Rock, R and B, and Soundtracks. However, for the experiments carried out within this study over 8.000 music excerpts from the SYNAT database representing 13 music genres were used. The cardinality of the music excerpts for the original and separated signals, in relation to specific music genre, is presented in Table 2. From the original audio signal harmonic (H), drum (D), piano (P), trumpet (T) and saxophone (S) signals were retrieved using different options of the Non-Negative Matrix Factorization separation method (cost function; extended KL-divergence method, window sizes: 20, 30, 40 ms, window function (Hann), window overlap (0.5), number of components (5, 10, 20, 30).

Marsyas (GTZAN) [2] is a commonly used database in MIR. It consists of 1000 songs, representing 10 genres (100 songs per each genre). They are as follows: Pop, Rock, Country, Rap and Hip-Hop, Classical, Jazz, Dance and Dj, Blues, Hard Rock and Metal, Reggae. For the purpose of showing the usability of the original feature vector used, preliminary results obtained for this database are recalled here.

### 2.3 Basic Parametrization

The so-called ‘basic’ parameters were adapted from previous studies [9] to be able to compare previous results. The list of parameters for ISMIS and SYNAT databases is

**Table 2.** Cardinality of the classes for original and separated signals, based on 8244 elements received from SYNAT database, representing 13 music genres

Genre	No. of music excerpts	H	%	D	%	P	%	T	%	S	%
Alternative Rock	207	207	100.0	204	98.55	137	66.18	48	23.19	10	4.83
Blues	264	264	100.0	257	97.35	242	91.67	134	50.76	6	2.27
Classical	953	953	100.0	408	42.81	920	96.54	272	28.54	208	21.83
Country	1039	1039	100.0	1034	99.52	837	80.56	261	25.12	51	4.91
Dance & DJ	259	258	99.61	257	99.23	185	71.43	0	0.00	19	7.34
Hard Rock & Metal	602	601	99.83	597	99.17	101	16.78	153	25.42	5	0.83
Jazz	568	567	99.82	527	92.78	537	94.54	211	37.15	33	5.81
Latin Music	444	444	100.0	444	100.00	285	64.19	256	57.66	10	2.25
New Age	567	560	98.77	327	57.67	489	86.24	120	21.16	211	37.21
Pop	796	796	100.0	798	100.25	539	67.71	236	29.65	66	8.29
R&B	610	610	100.0	610	100.00	482	79.02	253	41.48	14	2.30
Rap&Hip-Hop	1012	1011	99.90	1010	99.80	715	70.65	582	57.51	27	2.67
Rock	923	922	99.89	914	99.02	509	55.15	216	23.40	13	1.41
TOTAL	8244	8232	99.85	7387	89.60	5978	72.51	2742	33.26	673	8.16

given in Table 3. Most of the parameters are based on MPEG 7 standard, others are Mel Frequency Cepstral Coefficients (MFCC) as well as some dedicated time-domain-related descriptors.

Before the Feature Vectors (FVs) were used in the classification experiments, two normalization methods were employed for data pre-processing, i.e. Min-Max normalization, Zero-Mean normalization and both methods used jointly. However, the normalization of training and test datasets are performed in that way that the mean and standard deviation values are calculated only for training dataset, and only the current value is retained from training and test datasets (used respectively for normalization of training and test datasets). Also in the classification process separability of data was checked based on Best First, Greedy, Ranker and PCA methods, and in the main experiments reduced FVs were employed.

**Table 3.** List of parameters for ISMIS and SYNAT databases

#	ID	Audio Feature Description	Comment
1	TC	Temporal Centroid	
2	SC, SC_V	Spectral Centroid – mean and variance	
34	ASE 1-34	Audio Spectrum Envelope (ASE)- average values in 34 frequency bands	for the SYNAT database only 29 subbands occur
1	ASE_M	Mean ASE (for all frequency bands)	

(Continued)

**Table 3.** (Continued)

#	ID	Audio Feature Description	Comment
34	ASEV 1-34	ASE variance in 34 frequency bands	for the SYNAT database only 29 subbands occur
1	ASE_MV	Mean ASE variance (for all frequency bands)	
2	ASC, ASC_V	Audio Spectrum Centroid (ASC) – average and its variance	
2	ASS, ASS_V	Audio Spectrum Spread (ASS) – average and its variance	
24	SFM 1-24	Spectral Flatness Measure (SFM) – average values for 24 frequency bands	for the SYNAT database only 20 subbands occur
1	SFM_M	Mean SFM (for all frequency bands)	
24	SFMV 1-24	SFM variance (for 24 frequency bands)	for the SYNAT database only 20 subbands occur
1	SFM_MV	Mean SFM variance (for all frequency bands)	
20	MFCC 1-20	Mel Function Cepstral Coefficients (MFCC) – first 20 (mean values)	
20	MFCCV 1-20	MFCC Variance – first 20 values	
3	THR_[1,2,3] RMS_TOT	No. of samples higher than single/double/triple RMS value	Dedicated parameters (24) in time domain based on the analysis of the distribution of the envelope in relation to the RMS value
6	THR_[1,2,3] RMS_10FR_ [MEAN,VAR]	Mean/Variance of THR_[1,2,3] RMS_TOT for 10 frames	
1	PEAK_RMS_TOT	A ratio of peak to RMS (Root Mean Square)	
2	PEAK_RMS10FR_ [MEAN,VAR]	Mean/variance of PEAK_RMS_TOT for 10 frames	
1	ZCD	No. of transition by the level Zero	
2	ZCD_10FR_ [MEAN,VAR]	Mean/Variance values of ZCD for 10 frames	
3	[1,2,3]RMS_TCD	Number of transitions by single/double/triple levels of RMS	

(Continued)

**Table 3.** (Continued)

#	ID	Audio Feature Description	Comment
6	[1,2,3]RMS_TCD_10FR_[MEAN, VAR]	Mean/Variance value of [1,2,3]RMS_TCD for 10 time frames	
	TOTAL	191 (for ISMIS) and 173 (for SYNAT)	

### 3 Experiments

#### 3.1 Algorithms Used in Music Genre Classification

The most popular methods for music genre classification are: Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Decision Trees, Rough Sets and Minimum-distance methods, to which a very popular  $k$ -Nearest Neighbor ( $k$ -NN) method belongs. Since preliminary experiments carried out by the authors showed that SVM (including co-training applied for SVM) classification returns best accuracy of classification, that's why the results obtained while employing this method are to be show [13]. Also, despite computational expensiveness of the  $k$ NN algorithm, it is a good method for solving multi-class classification problem and is also commonly used in MIR area, what makes it possible to compare the experimental results. For these reasons results of this method are also presented. The core experiment, as well as normalization methods, were prepared in Java programming language using Eclipse environment. Weka library for Java was used for data managing: selecting attributes by their name, selecting instances (FVs of specific music excerpts) and in classification process. Selecting the best attributes (Best First, Greedy, Ranker and PCA methods) was done in Weka graphical interface.

#### 3.2 Feature Vector Optimization - Adding New Parameters

In the separation process several specific instrument/path (as mentioned earlier, i.e. harmonic, drum, piano, trumpet, saxophone) were retrieved aimed at FV optimization. For this purpose new FVs were created containing the originally extracted parameters and those based on separated music tracks. In that way an original signal (O) with added harmonic (H) components formed OH feature vector, etc. Several mixtures were tried, and a summary of results is shown in Table 4. Table 4 presents the results of overall correctness of classification for 16 mixtures of signals for full set of FVs (p\_173 per each signal) for the SYNAT database, as presented above.

The best results were obtained for the OH signal, however the differences between the specific mixed signals are not high. It may also be observed that the mixture of two signals is the most promising resulting in the highest correctness, however a combination of three signals (OHD) was also retained in experiments. In Table 5 results (Precision (Prec) and True Positive (TP)) for the  $k$ NN algorithm and optimized FV (using the PCA method only 59 parameters were retained) are shown for SYNAT database.

**Table 4.** Results of the overall correctness of classification for 16 mixtures of signals for a full set of FVs (p<sub>173</sub> per each signal)

signal mixture/	OH	ORIG	OHS	OD	OP	OHD	OHP	OHDS
correctness	72.37	72.19	71.92	71.57	71.42	71.40	71.35	71.32
signal mixture/	OS	OHDP	OHT	OHDPTS	OHDTS	OHDT	OHDPSP	OT
correctness	71.31	71.30	71.29	71.24	70.94	70.91	70.66	70.43

Table 6 presents the results for SVM algorithm. OH signal gave much better results ( $\sim 4\%$  better) in comparison to Original signal for such genres as alternative rock, blues, jazz, new age, pop (true positive rate) and Latin music (precision). That confirms that separating harmonic path for genres where harmonic plays significant role is useful. On the basis of the results some other conclusions may be derived, such as: for the OH signal an improvement of level of confusion between Pop and New Age, Pop and Latin Music and Pop and Rap and Hip-Hop is also observed and is equal  $\sim 2.77\%$  (in case of New Age) and  $\sim 0.88\%$  in case of two others. In case of using Co-training for the SVM algorithm results gained approx. 1% of correctness.

**Table 5.** *k*NN-based classification results

<i>k</i> NN	OHD		OH		OD		DH		ORIG	
	TP	Prec	TP	Prec	TP	Prec	TP	Prec	TP	Prec
Alternative Rock	13.04	43.46	14.49	38.68	8.21	24.05	11.11	42.00	14.49	44.76
Blues	7.20	33.01	12.50	48.08	4.55	29.86	4.92	30.35	9.47	39.56
Classical	60.13	82.14	95.80	82.18	39.56	73.87	40.29	69.36	95.17	84.08
Country	90.95	45.85	92.78	47.38	88.64	46.56	87.68	44.13	92.20	50.96
DanceDJ	58.30	95.38	43.25	94.84	59.86	90.28	46.74	89.42	62.17	87.10
Hard Rock/Metal	82.89	88.91	84.39	85.93	82.22	83.19	77.24	87.86	87.87	84.18
Jazz	59.68	65.64	64.96	71.50	54.23	64.82	53.34	61.08	64.79	70.27
Latin Music	38.29	79.84	43.92	79.55	39.64	82.41	30.63	77.74	47.97	81.15
NewAge	44.09	92.40	74.60	84.95	40.74	93.99	36.86	90.16	76.54	87.84
Pop	25.75	40.76	22.10	45.01	25.25	42.58	22.11	40.48	24.87	49.82
Rap HipHop	87.75	64.34	86.66	88.16	86.17	47.11	84.09	45.99	90.71	86.36
RB	61.32	59.63	60.16	64.34	56.73	64.67	57.71	57.03	63.45	69.23
Rock	63.16	51.76	64.90	58.27	62.62	53.56	59.48	52.23	68.26	61.05
Ov. correctn.	60.93		67.52		56.88		54.50		69.84	

Table 7 presents the summary results of the overall CCI (Correctly Classified Instances) for the small (ISMIS) and big (SYNAT) databases conducted with the Co-SVM method, as the one which gives the highest correctness of classification, involving the original FV (191 parameters for ISMIS) and reduced ones (VoP p<sub>52</sub>, p<sub>59</sub>, p<sub>60</sub>).

Even though the results of CCI for the bigger database (SYNAT) are  $\sim 10\%$  lower than for the smaller one (ISMIS), it is still a very good result considering the fact, that



**Table 6.** Results for the SVM algorithm

SVM	OHD		OH		OD		DH		ORIG	
Genre	TP	Prec	TP	Prec	TP	Prec	TP	Prec	TP	Prec
Alternative Rock	43.48	33.48	41.55	37.30	41.06	33.89	35.27	31.75	33.33	37.14
Blues	41.67	36.10	40.53	38.31	42.80	41.01	37.12	35.08	38.26	45.24
Classical	90.66	87.72	91.08	88.77	89.72	87.82	88.35	85.04	90.87	87.50
Country	77.00	74.01	79.88	73.96	75.55	72.67	73.63	69.42	77.00	72.26
DanceDJ	74.16	67.69	74.18	69.74	72.20	66.08	69.53	63.30	74.13	72.71
Hard Rock/Metal	84.55	83.92	84.22	83.96	84.88	81.86	80.73	80.20	88.03	83.13
Jazz	67.61	68.03	73.24	70.15	68.31	68.24	65.67	66.52	70.43	66.45
LatinMusic	67.79	71.76	69.59	75.27	68.02	71.60	66.22	72.54	69.37	71.96
New Age	77.07	82.57	77.60	81.21	75.13	76.98	71.08	69.56	72.31	68.67
Pop	39.32	39.58	40.83	42.76	40.83	41.65	39.19	40.02	36.68	41.96
Rap HipHop	87.55	90.06	86.56	88.13	88.04	88.35	87.45	88.23	89.92	87.92
RB	60.18	66.83	63.28	69.10	62.14	70.42	57.71	64.24	62.95	69.97
Rock	59.49	63.36	62.41	64.73	60.02	63.91	55.04	59.36	64.46	64.55
Ov. correctn.	70.38		71.79		70.37		67.58		71.05	

**Table 7.** Comparison of results of overall classification for small and big database obtained for the original and mixed signals involving Co-training method

classifier	Co-SVM				Co-SVM		
type of FV	Original				OD		
database	ISMIS		SYNAT		ISMIS		SYNAT
Vector	FV191	VoP p_60	VoP p_59	FV191	VoP p_52	VoP p_59	
no. of classes	4		13		4		13
no. of elements	465		8244		465		8244
CCI [%]	<b>80.86</b>	<b>79.78</b>	<b>71.49</b>	<b>81.08</b>	<b>81.72</b>	<b>71.29</b>	

there are as many as 13 genres to be classified instead of only few as in the case of ISMIS database. Moreover some of those 13 genres are similar and very often confused not only by the machine learning methods but also by human listeners, which makes it considerably more difficult to be classified correctly.

Finally, eight music genres common for SYNAT and GTZAN databases were compared in the context of the FV robustness. In the experiments the *k*NN algorithm (with the Euclidean function) and the ‘basic’ FV (containing 173 parameters) were used and this experiment resulted in the following classification (approx.): Classical 88 %, Blues 70 %, Country 59 %, DanceDj 52 %, HardRock 78 %, Jazz 68 %, Pop 71 %, Rap 58 %, Rock 51 %. The results obtained for the GZTAN database compared to the SYNAT database are comparable. The reason for the a bit lower classification efficiency for these particular genres may be a greater variety of songs in the GZTAN database. Also, the database contains also recordings with reduced quality, which may have an adverse effect on the effectiveness of the overall parameterization.

## 4 Summary

In this paper a new strategy to music genre classification was proposed. The main principle is separating music tracks at the pre-processing phase and extending vector of parameters by descriptors related to a given musical instrument components that are characteristic for the specific musical genre. This allows for more efficient automatic musical genre classification. It was also shown that extending the original signal even with one or two parameters (such as OP, OT and OS signals) influences the results for classification for specific genres, what confirms the importance of specific parameters in relation to music genre. It has also to be noted that the overall classification gave better correctness for any type of mixed signals in comparison to the original signal.

**Acknowledgments.** The study was partially supported by the EU from the European Social Fund (UDA-POKL.04.01.01-00-106/09) and the project PBS1/B3/16/2012 financed by the Polish National Centre for R&D.

## References

1. Canadas-Quesada, F.J., Vera-Candeas, P., Ruiz-Reyes, N., Carabias-Orti, J., Cabanas-Molero, P.: Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints. *EURASIP J. Audio, Speech Music Proc.* **2014**(26), 1–17 (2014)
2. GTZAN database. [http://marsyasweb.appspot.com/download/data\\_sets/](http://marsyasweb.appspot.com/download/data_sets/). Accessed May 2015
3. Helen, M., Virtanen, T.: Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine. In: 13th EUSIPCO'2005, Antalya (2005)
4. Heittola, T., Klapuri, A., Virtanen, T.: Musical instrument recognition in polyphonic audio using source-filter model for sound separation. In: 10th ISMIR'2009, Kobe (2009)
5. Hoffmann, P., Kostek, B.: Music genre recognition in the rough set-based environment. In: Kryszkiewicz, M., Bandyopadhyay, S., Rybinski, H., Pal, S.K. (eds.) *PREMI 2015*. LNCS, vol. 9124, pp. 377–386. Springer, Heidelberg (2015)
6. ISMIR website [accessed 2015 February]. <http://ismir2014.ismir.net>
7. Kostek, B.: Music information retrieval in music repositories. In: Suraj, Z., Skowron, A. (eds.) *Rough Sets and Intelligent Systems Refer*, pp. 463–489. Springer, Berlin (2013)
8. Hoffmann, P., Kostek, B.: Music data processing and mining in large databases for active media. In: Ślęzak, D., Schaefer, G., Vuong, S.T., Kim, Y.-S. (eds.) *AMT 2014*. LNCS, vol. 8610, pp. 85–95. Springer, Heidelberg (2014)
9. Kostek, B., Kupryjanow, A., Zwan, P., Jiang, W., Raś, Z.W., Wojnarski, M., Swietlicka, J.: Report of the ISMIS 2011 contest: music information retrieval. In: Kryszkiewicz, M., Rybinski, H., Skowron, A., Raś, Z.W. (eds.) *ISMIS 2011*. LNCS, vol. 6804, pp. 715–724. Springer, Heidelberg (2011)
10. Kirss, P.: Audio based classification of electronic music. University of Jyväskylä (2007). [https://jyx.jyu.fi/dspace/bitstream/handle/123456789/13592/urn\\_nbn\\_fi\\_jyu-2007601.pdf?sequence=1](https://jyx.jyu.fi/dspace/bitstream/handle/123456789/13592/urn_nbn_fi_jyu-2007601.pdf?sequence=1). Accessed May 2015
11. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999)

12. Ras, Z., Wieczorkowska, A.: *Advances in Music Information Retrieval*. Springer Publishing, Berlin, Heidelberg (2010)
13. Rosner, A., Schuller, B., Kostek, B.: Classification of music genres based on music separation into harmonic and drum components. *Arch. Acoust.* **39**(4), 629–638 (2014)
14. Rosner, A.: *Multi-instrumental automatic recognition of musical genres*. Ph.D. Dissertation, Silesian University of Technology, Institute of Informatics, Gliwice, Poland (2015)
15. Schnitzer, D., Flexer, A., Widmer, G.: A fast audio similarity retrieval method for millions of music tracks. *Multimedia Tools and Applications* **58**(1), 23–40 (2012)
16. Smaragdis, P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. In: *Proceedings of WASPAA*, pp. 177–180 (2003)
17. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **10**(5), 293–302 (2002)
18. Uhle, C., Dittmar, C., Sporer, T.: Extraction of drum tracks from polyphonic music using independent subspace analysis. In: *Proceedings 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA)*, Nara, pp. 843–848 (2003)
19. Weninger, F., Woelmer, M., Schuller, B.: Automatic assessment of singer traits in popular music: gender, age, height and race. In: *Proceedings of 12th ISMIR*, pp. 37–42 (2011)
20. Weninger, F., Schuller, B.: Optimization and parallelization of monaural source separation algorithms in the openBliSSART toolkit. *J. Sig. Process. Syst.* **69**(3), 267–277 (2012)
21. Wieczorkowska, A., Kubera, E., Kubik-Komar, A.: Analysis of recognition of a musical instrument in sound mixes using support vector machines. *Fundamenta Informaticae* **107**(1), 85–104 (2011)