BEAT HISTOGRAM FEATURES FROM NMF-BASED NOVELTY FUNCTIONS FOR MUSIC CLASSIFICATION

Athanasios Lykartsis
Technische Universität Berlin
Audio Communication Group
alykartsis@mail.tu-berlin.de

Chih-Wei Wu
Georgia Institute of Technology
Center for Music Technology
cwu307@gatech.edu

Alexander Lerch
Georgia Institute of Technology
Center for Music Technology
alexander.lerch@gatech.edu

ABSTRACT

In this paper we present novel rhythm features derived from drum tracks extracted from polyphonic music and evaluate them in a genre classification task. Musical excerpts are analyzed using an optimized, partially fixed Non-Negative Matrix Factorization (NMF) method and beat histogram features are calculated on basis of the resulting activation functions for each one out of three drum tracks extracted (Hi-Hat, Snare Drum and Bass Drum). The features are evaluated on two widely used genre datasets (GTZAN and Ballroom) using standard classification methods, concerning the achieved overall classification accuracy. Furthermore, their suitability in distinguishing between rhythmically similar genres and the performance of the features resulting from individual activation functions is discussed. Results show that the presented NMF-based beat histogram features can provide comparable performance to other classification systems, while considering strictly drum patterns.

1. INTRODUCTION

The description of musical rhythm remains an important and challenging topic in Music Information Retrieval (MIR) with applications in several areas [12, 16]. The difficulty of rhythm extraction lies in its multifaceted character, which involves periodicity and structural patterning in the signal as well as perceptual components such as musical meter [19]. An approach which has achieved some popularity over the last years is based on the creation of a periodicity representation — commonly called the beat histogram (BH) — and the subsequent extraction of features from this histogram to be used, e.g., in genre classification [4, 13, 33]. A common first processing step of all approaches is the extraction of a so-called novelty function [2] or its derivatives as the starting point for further analysis. Since a complete rhythm representation of a musical track results from the superposition of the temporal progressions of different instruments or voices [12, 16], it makes sense to include features taking into account individual temporal and spectral properties.

In western popular music (which is the focus of this paper), rhythm is most often carried from the drum section, providing the temporal grid on which other instruments can unfold their melodic or harmonic patterns. This makes the analysis of the drum track appealing for the description of rhythmic character. In order to obtain the rhythmic properties of the drum section, the extraction of temporal novelty functions per instrument is necessary. Although such methods for the extraction of specific voices or instruments have been commonly used in the area of source separation or automatic instrument transcription (the most notable being non-negative matrix factorization (NMF) [31]), their application to rhythm extraction problems is, to the best of our knowledge, sparse. We therefore propose to use a technique for source separation and drum transcription based on partially fixed NMF using the resulting activation functions as a source material for the extraction of rhythmic features based on beat histograms. This paper investigates the suitability of the proposed features in the context of rhythm-based genre classification for dance music and other styles.

The paper is structured as follows. In the second section, an overview of previous work and the goals of the current paper are presented. In section 3, the drum transcription procedure and the feature extraction are described. In the fourth section, the evaluation of the proposed features and the results are given. After discussing the results in section 5, we close by giving conclusions and suggestions for future work (sect. 6).

2. PREVIOUS WORK AND GOALS

Beat histograms have been used for a long time as rhythmic descriptions. Initially introduced in studies on beat tracking and analysis [11, 29] as a useful very low frequency periodicity representation, they were only later referred to as the *beat histogram* [33] or *periodicity histogram* [13]. The histogram is useful as an intermediate representation that can be used to extract musical parameters such as tempo as well as low-level features (e.g., statistical properties of the histogram). Traditionally, a measure of the signal amplitude envelope or its change over time is utilized as the novelty function for the extraction of a beat histogram [4, 13, 33]. However, in the field of onset detection, the proposed novelty functions take into account spectral content changes [3, 10, 15, 27]. Genre classification systems based on such representations have generally shown

promising results, although rhythm features do usually not perform as well as features from other domains such as timbre descriptors [4,28,33]. However, studies have shown that for highly rhythmical music, beat histogram features can achieve very high performance [13], a fact which has been confirmed in current studies investigating the role of using multiple novelty functions as a basis for beat histogram features [20].

Since drum tracks convey essential information about tempo, rhythm and possibly genre, they could potentially provide better representation for extracting rhythm features. To extract drum tracks from complete mixtures of music, a drum transcription system for polyphonic music would be necessary. Gillet and Richard divide systems for the drum transcription from mixtures into three categories [9]: (i) segment and classify, (ii) separate and detect, and (iii) match and adapt. Here, we focus on the second type of approaches (separate and detect). Based on the assumption that the music signal is a superposition of different sound sources, the music content could be transcribed by first decomposing the signal into source templates with corresponding activation functions, and then detecting the activities of each template. Different methods such as Independent Subspace Analysis [7], Prior Subspace Analysis [6], and Non-negative Matrix Factorization [1, 21] fall into this category. These approaches are usually easy to interpret since most of the decompositions result in spectrum-like representations. Furthermore, these approches do not require additional classes for simultaneous events, which could potentially reduce the model complexity.

In the context of NMF for music transcription, the following issues have to be taken into consideration: First, the number of sound sources and notes within a music recording is usually unknown. It is therefore difficult to determine a suitable rank r in order to obtain a clear differentiation of the decomposed components in the dictionary matrix. Second, after the unsupervised NMF decomposition process, it is difficult to identify the associated instrument of each component in the dictionary matrix W when rank is too high or too low. Third, when multiple similar entries exist in the dictionary matrix, the corresponding activation matrix could be activated at these entries simultaneously, which in turn increases the difficulty of intuitively interpreting the results.

To address the above issues, Yoo et al. proposed a co-factorization algorithm [35] to simultaneously factorize a prior drum track and a target signal, and use the basis matrix from the drum track to identify the drum components in the target signal. This method ensures that the drum components in both dictionary matrices remain percussion only over the iterations, and thus proper isolation of the harmonic components from the drum components. Since they focus on drum separation rather than drum transcription, their selection of ranks can be higher, but the approach is not directly applicable to the transcription problem because of the probable lack of interpretability of the dictionary matrix. Wu and Lerch proposed a variant of the co-factorization algorithm using partially fixed NMF (PFNMF)

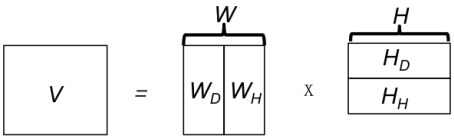

Figure 1. Illustration of the factorization process. W: dictionary matrix. H: activation matrix. Subscript D: drum components, Subscript H: harmonic components.

for drum transcription in polyphonic signals [34]. Instead of co-factorization, this method uses a pre-determined drum dictionary matrix during the decomposition process, and extracts one activation function for each of the three drums (Hi-Hat, Snare Drum, and Bass Drum).

In this paper, we apply PFNMF to transcribe drum events in polyphonic signals, and use the activation functions as the basis for the extraction of beat histogram features. The idea of using NMF with prior knowledge of targeting source within the mixture has been applied in source separation tasks [32], multi-pitch analysis [26] and drum transcription [34]. Furthermore, the use of multiple novelty functions for the extraction of beat histograms has been proposed in [20]. Here, we combine both approaches for the generation of rhythmic features which are descriptive of the percussive rhythmic content of polyphonic tracks and therefore of their general rhythmic character. We focus on two tasks: the investigation of their overall performance, in order to determine the salience of the features for genre classification; and their performance for each percussive component (drum track) separately, attempting to extract conclusions regarding the importance of drum based rhythm features and the salience of NMF activation functions.

3. METHOD

The basic concept of NMF is to approximate a matrix V with matrices W and H as $V \approx WH$ with non-negativity constraints. Given a $m \times n$ matrix V, NMF will decompose the matrix into the product of a $m \times r$ dictionary (or basis) matrix W and an $r \times n$ activation matrix H, with r being the rank of the NMF decomposition. In most audio applications, V is the spectrogram to be decomposed, W contains the magnitude spectra of the salient components, and H indicates the activation of these components with respect to time [31]. The matrices W and H are estimated through an iterative process that minimizes a distance measure between the target spectrogram V and its approximation [30].

To effectively extract drum activation functions from the polyphonic signals, PFNMF is used in this study. Figure 1 visualizes the basic concept from the work of Yoo et al.: the matrices W and H are split into the matrices W_{D} and W_{H}, and H_{D} and H_{H}, respectively. Instead of using co-factorization, PFNMF initializes the matrix W_{D} with drum components and to not modify it during the factorization process. Matrices W_{H}, H_{H}, and H_{D} are initialized with random numbers. The distance measure used in this paper is the generalized KL-divergence (or I-divergence), in which

Figure 2. Flowchart of NMF and beat histogram feature extraction and classification system.

$D_{\mathrm{KL}}(x \mid y) = x \cdot \log(x/y) + (y - x)$. The cost function as shown in (1) is minimized by applying gradient descent and multiplicative update rules, the matrices W_{H}, H_{H}, and H_{D} will be updated according to Eqs. (2)–(4).

$$J = D_{\mathrm{KL}}(V \mid W_{\mathrm{D}}H_{\mathrm{D}} + W_{\mathrm{H}}H_{\mathrm{H}}) \quad (1)$$

$$H_{\mathrm{D}} \leftarrow H_{\mathrm{D}} \frac{W_{\mathrm{D}}^T(V/(W_{\mathrm{D}}H_{\mathrm{D}} + W_{\mathrm{H}}H_{\mathrm{H}}))}{W_{\mathrm{D}}^T} \quad (2)$$

$$W_{\mathrm{H}} \leftarrow W_{\mathrm{H}} \frac{(V/(W_{\mathrm{D}}H_{\mathrm{D}} + W_{\mathrm{H}}H_{\mathrm{H}}))H_{\mathrm{H}}^T}{H_{\mathrm{H}}^T} \quad (3)$$

$$H_{\mathrm{H}} \leftarrow H_{\mathrm{H}} \frac{W_{\mathrm{H}}^T(V/(W_{\mathrm{D}}H_{\mathrm{D}} + W_{\mathrm{H}}H_{\mathrm{H}}))}{W_{\mathrm{H}}^T} \quad (4)$$

PFNMF can be summarized in following steps:

1. Construct an $m \times r_{\mathrm{D}}$ dictionary matrix W_{D}, with r_{D} being the number of drum components to be detected.
2. Given a pre-defined rank r_{H}, initialize an $m \times r_{\mathrm{H}}$ matrix W_{H}, an $r_{\mathrm{D}} \times n$ matrix H_{D} and an $r_{\mathrm{H}} \times n$ matrix H_{H}.
3. Normalize W_{D} and W_{H}.
4. Update H_{D}, W_{H}, and H_{H} using (2)–(4).
5. Calculate the cost of the current iteration using (1).
6. Repeat step 3 to step 5 until convergence.

In our current setup, the STFT of the signals is calculated using a window size and a hop size of 2048 and 512, respectively. A pre-trained dictionary matrix is constructed from the training set, consisting of isolated drum sounds. The templates are extracted for the three classes Hi-Hat (HH), Bass Drum (BD) and Snare Drum (SD) as the median spectra of all individual events of one drum class in the training set. Next, the PFNMF will be performed with rank $r_{\mathrm{H}} = 10$ on the test files. More details of the training process and the selection of rank r_{H} can be found in [34]. Finally, the activation Matrix H_D can be extracted from the audio signals through the decomposition process.

Once the activation functions of the three drum tracks have been extracted as described above, they are used as novelty functions for the calculations of beat histograms, similar to [20]. The complete procedure for the generation of a feature vector representing each track includes the following steps: For each activation function, the beat histogram is extracted through the calculation of an Autocorrelation Function (ACF) and the retaining of the area between 30 and 240 BPM. For each beat histogram, the subfeatures listed in Table 1 are extracted. The concatenation of all subfeature groups for each novelty function produces the final feature vector for an audio excerpt. Similar subfeatures as listed in Table 1 can be found in the literature, e.g., in [33] (Peak), and [4, 13] (Distribution). In total, 3 novelty functions are used for the production of as many beat histograms, from each of which 19 subfeatures are extracted, resulting in a total count of 57 features.

Distribution	Peak
Mean (ME)	Salience of Strongest Peak (A1)
Standard Deviation (SD)	Salience of 2nd Stronger Peak (A0)
Mean of Derivative (MD)	Period of Strongest Peak (P1)
SD of Derivative (SDD)	Period of 2nd Stronger Peak (P2)
Skewness (SK)	Period of Peak Centroid (P3)
Kurtosis (KU)	Ratio of A0 to A1 (RA)
Entropy (EN)	Sum (SU)
Geometrical Mean (GM)	Sum of Power (SP)
Centroid (CD)	
Flatness (FL)	
High Frequency Content (HFC)	

Table 1. Subfeatures extracted from beat histograms.

4. EVALUATION

4.1 Dataset Description

In order to evaluate the features for multiple track kinds possessing different rhythmic qualities, two datasets were considered: the Tzanetakis Dataset (**GTZAN**) [33], as an example of a dataset which is widely used, comprising 100 30 sec excerpts for each of 10 diverse musical genres; and the Ballroom Dataset [5, 13] (**Ballroom**), comprising 698 very rhythm/dance-oriented tracks of length 10 sec and therefore suitable for the evaluation of our NMF-based beat histogram features. Both datasets contain tracks with a drum section and others with only non-percussive instruments. This does not only allow to investigate if the extracted features are also suitable for music where a drum section is present and if they can generalize to other music styles, but also allows conclusions as to what genres in particular are represented satisfactory or insufficiently by the features.

4.2 Evaluation Procedure

The features were tested using the Support Vector Machine (SVM) algorithm for supervised classification. For our multiclass setting, an RBF kernel was used and the optimal parameters (C, γ) were determined through grid search. We chose the SVM classifier since it has been frequently used in similar genre classification experiments, shows generally good results (see [8]) and allows for comparability with those studies. Since the focus here lay on the features and not the classification algorithms, we refrained from using more state-of-the-art approaches such as deep learning algorithms. All experiments took place with a 10-fold cross-validation (using 90% of the data for training and 10% for testing over 10 randomly selected folds, taking the average accuracy over the folds for each dataset) and standardization (z-score) of the training and testing data. After the full

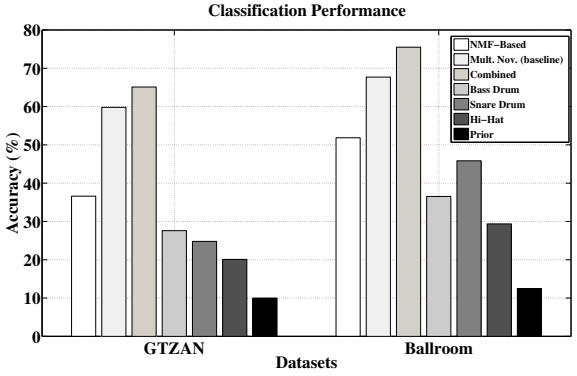

Figure 3. Classification results for both datasets.

NMF-based feature set (i.e., the features originating from all three drum activation functions) was tested, the features from each individual activation function were evaluated in turn in order to study the importance of each drum track separately. Finally, the NMF-based features are combined with other beat histogram features from a current study [20], extracted from novelty functions of amplitude (RMS), spectral shape (spectral flux, centroid, flatness and the first 13 MFCCs) and tonal components (pitch chroma coefficients and tonal power ratio) on 3 second-long frames. Those features resulted from a similar procedure as the one used here, where 30 different novelty functions were extracted and their beat histograms computed through the calculation of an ACF. A subsequent two-stage feature selection scheme (mutual information with target data [14] using the CMIM metric [25], followed by a sequential forward selection with an SVM wrapper [17]) was applied to retain the best-performing features, resulting in a total of 20 features in each case.

4.3 Results

The results are shown in Figure 3. On both datasets, the full NMF feature set (comprising features from all three drum activation functions) performs better than the individual ones (BD, SD, HH), with an attained accuracy of 36.6% and 51.9% for GTZAN and Ballroom, respectively. Those values lie considerably above the average priors of both datasets. The differences between the accuracies of the feature sets are not large (especially between the individual drum based feature sets) but are significant at the 0.05 level in all cases (based on a comparison test of the Cohen's Kappa extracted from the confusion matrices). Due to their small values (ranging from 0.2% to 0.6%), standard deviations between accuracies of the folds for each feature set are not presented in Figure 3.

The multiple novelty feature set (from [20]) outperforms the NMF-based features, reaching an accuracy of 59.8% for the GTZAN and 67.7% for the Ballroom dataset, whereas the combined set (NMF and multiple novelty) demonstrates the best performance (accuracy of 65.1% (GTZAN) and 75.5% (Ballroom)). The individual feature sets from each drum track provide performance inferior to that of the full NMF-based set, but still considerably higher than the prior. The best individual drums are the BD and SD for the GTZAN and Ballroom datasets, respectively. The worst individual percussion instrument is in both cases the HH. For the full NMF-based feature set, confusion matrices resulting from the classification can be seen in Tables 2 and 3. In general, features achieved better average performance on the Ballroom dataset than on the GTZAN. In order to evaluate the misclassifications and the performance of the individual genres, a closer observation of the confusion matrices of each dataset should be taken.

	Ch.	Ji.	Qu.	Ru.	Sa.	Ta.	Vw.	Wa.
Ch.	54	10	10	11	14	3	1	8
Ji.	17	13	5	6	10	2	2	5
Qu.	10	2	44	5	3	8	7	3
Ru.	8	3	2	53	4	7	2	19
Sa.	15	4	8	2	50	3	1	3
Ta.	2	1	6	6	2	55	7	7
Vw.	5	3	9	6	0	6	17	19
Wa.	5	0	4	16	1	4	4	76
Acc.	49	22	54	54	58	64	26	69
Pr.	15.9	8.6	11.7	14.0	12.3	12.3	9.3	15 .8

Table 2. *Confusion matrix for Ballroom dataset, average accuracy:* 51.9%. *Accuracy and Prior are given in* $\%$.

	Bl.	Cl.	Co.	Di.	Hi.	Ja.	Me.	Po.	Re.	Ro.
Bl.	15	11	16	15	4	7	9	9	11	3
Cl.	4	63	3	1	1	14	5	3	1	5
Co.	6	6	38	12	4	5	6	6	11	6
Di.	13	1	6	43	6	1	8	5	12	5
Hi.	8	4	5	4	21	8	10	20	13	7
Ja.	8	17	5	0	7	38	9	7	7	2
Me.	3	11	7	7	2	6	51	2	2	9
Po.	7	6	6	5	14	5	5	33	12	7
Re.	6	3	6	6	6	4	1	11	53	4
Ro.	6	4	10	10	17	10	11	11	10	11
Acc.	15	63	38	43	21	38	51	33	53	11
Pr.	10	10	10	10	10	10	10	10	10	10

Table 3. *Confusion matrix for GTZAN dataset, average accuracy:* 36.6%. *Accuracy and Prior are given in* $\%$.

For the **Ballroom** dataset, confusions between genres appear to be plausible based on what one would expect when extracting rhythm features only from drums tracks: genres with strongly pronounced, stable rhythm played from a drum section such as samba and *chachacha (Ch.)* are confused with each other, whereas the *waltz (Wa.)* and *tango (Ta.)* genres, having no drum section (but still a succinct rhythm) are not confused much with other genres. The latter are the two genres which also achieve the best individual performance, followed by *chachacha*, *quickstep (Qu.)*, *rumba (Ru.)* and *samba (Sa.)*. *Jive (Ji.)* and *viennese waltz (Vw.)* display the worse performance, and are confused with *chachacha* and *waltz* respectively, a result which is also expected when one considers the rhythmic proximity of those genres, whether they possess a drum section or not.

For the **GTZAN** dataset, misclassifications present a more mixed picture: On the one hand, genres which possess tracks featuring a well articulated, distinct rhythmic performed by a drum section (such as *reggae (Re.)*, *metal*

(Me.) and *disco (Di.)*) as well as the only genre without drums (*classical (Cl.)*) achieve satisfactory performance and are confused with genres which are rhythmically relatively close (*classical* with *jazz (Ja.)*, *metal* with *rock (Ro.)*, *disco* with *reggae*, and *reggae* with *pop (Po.)*). On the other hand, genres possessing tracks with a more "generic" rhythm (such as *country (Co.)* and *pop*) are confused with multiple other genres. Finally, *hiphop (Hi.)*, *blues (Bl.)* and *rock* attain the last places in individual performance and are confused with multiple other genres.

5. DISCUSSION

The results show that beat histogram features based on NMF activation functions of specific drums can be helpful in rhythm-based genre classification, as their accuracy for the used datasets is comparable to that achieved by other rhythmic feature sets used up to date (59.8% [20] and 28% [33] for the GTZAN, 67.7% [20] and 56.7% [13] for the Ballroom dataset). When taking into account that the features are solely based on drum novelty functions, their performance, especially for the Ballroom dataset, can be seen as satisfactory. It is clear, though, that for this reason, our results cannot achieve as high accuracy as other studies which use very sophisticated methods [8, 18, 22–24]. Our results are somewhat lower than the state of the art using rhythm [22, 24] or combined features [8, 23], however staying in the same range. For the sake of comparison, we report here the highest performances reached when using advanced rhythmic features: on the GTZAN dataset an accuracy of 92.4% [22] has been achieved, for the Ballroom dataset one of 96.1% [24]. The advantage of our proposed methods and features lies in the ability to pinpoint the importance of the rhythm patterns from specific drums for specific genres.

The misclassifications (reported in Tables 2 and 3) show that genres which do not feature genre-specific rhythm patterns, even if those are clearly articulated by the drum section (e.g., a 4/4 BD and SD alternating beat), tend to be confused with other similar genres (especially when drum tracks are present, such as in *rock*). Genres containing non-percussive tracks (such as *classical* and *waltz*) or very specific rhythmic patterns (*reggae*) are more easily distinguished from others. Those results indicate that the NMF-based beat histogram features indeed capture rhythmic properties related to the drum section and the regularities of their periodicities, pointing towards the suitability of those features for the extraction of drum-based rhythmic properties and the use in the discrimination of musical tracks which contain drums from ones which do not.

With regards to the feature sets, the satisfactory accuracy of the NMF-based feature set is a hint towards the appropriateness of the features for the analysis of the rhythmic character of a musical track. However, it is clear that those features, being derived only from drum tracks, cannot represent as much information as features resulting from the use of multiple novelty functions covering many aspects of the signal temporal progress. The improved performance of the combined set (NMF and multiple novelty based) is a consequence of incorporating specific, drum-related rhythm information in the feature base, showing that the NMF-based rhythm feature set can contribute information not provided by more general rhythm features and lead to significant improvement for the two evaluated datasets. The analysis of the features derived from the activation function of a specific drum track showed that mainly the snare drum and to a lesser extent the kick drum are the most important components. The tendency is strong for the Ballroom dataset, where the SD outperforms the BD, whereas for the GTZAN dataset the result is reversed but with a smaller difference. In all cases (also between the individual drum sets), the differences in accuracies between the feature sets are significant at the 5% level. Those results can be due to the very pronounced sound texture and greater power of those drums which leads to a salient activation function, as well as their role in providing the basic metric positions in most of western popular music. However, the accuracy of each subset lies below that of their combination, leading to the conclusion that the activation functions of all three percussion instruments contribute valuable information to the feature description of musical genre.

Concerning the datasets, the poorer classification performance observed for the GTZAN dataset is a sign of the more diverse character of tracks and genres in this set, containing music styles which lack a specific rhythmic character and can therefore not be distinguished effectively through beat histogram features derived from drum activation functions. Results were still better than the ones reported in [33], but their inferiority compared to the ones in other studies [13, 20] shows that when considering a multitude of different genres, solely drum based activation functions can not provide a complete rhythmic characterization. This, however, points towards the possible goal of using NMF in order to transcribe not only drums but also other instruments in order to use their activation functions as a basis for beat histogram features. The Ballroom dataset shows better performance, which was to be expected since the tracks therein are selected for belonging to different dance styles, requiring a special rhythmic pattern which is mostly conveyed by the drum section. The results are in the same range as those provided in [13] (56.7%) when using only periodicity histogram features. Furthermore, in the same study it was shown that using the tempo of the given tracks as a feature they could achieve very high results using a simple 1-NN classifier (51.7% for the "naive tempo" derived from the periodicity histogram and 82.3% for the ground-truth tempo provided with the recordings), reaching as much as 90% when combining the correct tempo with other descriptors (MFCCs) from the periodicity histogram. This shows that beat histograms (from which the tempo can be extracted) are a good tool for rhythmic analysis in datasets containing dance music such as the Ballroom.

Regarding specific genres, it is clear from the results that the NMF-based features have a twofold use: first, in representing genres which are characterized by distinct patterns in their drum sections (e.g., *reggae* or *samba*) and second, in characterizing genres which lack a drum section

at all (*waltz*, *classical*) in contrast to genres which do; the activation functions transcribed in this case are maximally different, leading to beat histogram features which can be easily discriminated by a classifier. Such a finding shows that drum-based rhythm features can be very helpful for rhythmic characterization of specific genres, which could be an argument for their further application when a specific kind of music is involved. As a general remark, it can be seen that genres possessing a stable rhythm articulated by a drum section such as *reggae* and *samba* or genres lacking drums in general (*waltz* and *classical*) perform better, whereas genres which have a very uncharacteristic rhythm (such as *rock* or *blues*) get more easily confused.

6. CONCLUSIONS

The work presented in this paper focuses on the creation of novel, NMF-based beat histogram features for rhythm-based musical genre classification and rhythmic similarity. The difference in comparison to other well-known studies for rhythm features based on beat histograms [4, 13, 24, 33] is the use of the activity functions of specific drums provided through NMF as a basis for the calculation of the beat histogram. We showed that the classification accuracy using these beat histogram features is comparable to that of other rhythm features, whereas our proposed features are better especially for characterizing tracks with specific rhythmic patterns or for distinguishing between songs with and without a drum section. It was observed that the most important percussion patterns for dance music classification were generated by the snare and the kick drum, which underlines the importance of its activation function for further tasks.

One future goal is the expansion of the use of NMF to identify more instruments or voices and use them as possible novelty functions. The goal would be to therefore capture the rhythmic patterns of every instrument, essentially joining source transcription and rhythm feature extraction into one module. Another possibility is the use of our proposed features for larger and more specific datasets, in order to further investigate their suitability for specific genres, as well as the strengths and weaknesses of the patterns extracted from individual drums in discriminating between musical genres. As an expansion of the feature selection procedure, a further idea would be to profit from the combination of NMF-based features and other acoustic features using a classifier that is capable of learning feature importance (e.g. random forest) to quantitatively investigate the importance of NMF-derived features. While NMF-based beat histogram features have been evaluated only in the context of rhythmic genre classification, we believe that they can prove useful in other tasks. Future research will focus on adjusting and using the proposed features for MIR tasks such as rhythmic similarity computation and structural analysis.

7. REFERENCES

[1] David S Alves, Jouni Paulus, and José Fonseca. Drum transcription from multichannel recordings with non-negative matrix factorization. In *Proceedings of the European Signal Processing Conference (EUSIPCO), Glasgow, Scotland, UK*, 2009.

[2] Juan P. Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.

[3] Juan P. Bello, Chris Duxbury, Mike Davies, and Mark Sandler. On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters*, 11(6):553–556, 2004.

[4] Juan José Burred and Alexander Lerch. A hierarchical approach to automatic musical genre classification. In *Proceedings of the 6th international conference on digital audio effects*, pages 8–11, 2003.

[5] Simon Dixon, Elias Pampalk, and Gerhard Widmer. Classification of dance music by periodicity patterns. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, 2003.

[6] Derry FitzGerald, Bob Lawlor, and Eugene Coyle. Drum transcription in the presence of pitched instruments using prior subspace analysis. In *Proceedings of the Irish Signals and Systems Conference (ISSC)*, 2003.

[7] Derry FitzGerald, Robert Lawlor, and Eugene Coyle. Sub-band independent subspace analysis for drum transcription. In *Proceedings of the Digital Audio Effects Conference (DAFX)*, pages 65–59, 2002.

[8] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, 2011.

[9] Olivier Gillet and Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):529–540, 2008.

[10] Masataka Goto and Yoichi Muraoka. Music understanding at the beat level – real-time beat tracking for audio signals. In *Computational auditory scene analysis*, pages 157–176, August 1995.

[11] Masataka Goto and Yoichi Muraoka. A real-time beat tracking system for audio signals. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 171–174, 1995.

[12] Fabien Gouyon and Simon Dixon. A review of automatic rhythm description systems. *Computer Music Journal*, 29(1):34–35, 2005.

[13] Fabien Gouyon, Simon Dixon, Elias Pampalk, and Gerhard Widmer. Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference*, pages 196–204, 2004.

[14] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[15] Stephen Hainsworth and Malcolm Macleod. Onset detection in musical audio signals. In *Proceedings of the International Computer Music Conference (ICMC)*, 2003.

[16] Enric Guaus i Termens. New approaches for rhythmic description of audio signals. Technical report, Music Technology Group, Universitat Pompeu Fabra, 2004.

[17] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.

[18] Chang-Hsing Lee, Jau-Ling Shih, Kun-Ming Yu, and Hwai-San Lin. Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. *Multimedia, IEEE Transactions on*, 11(4):670–682, 2009.

[19] Justin London. *Hearing in time*. Oxford University Press, 2012.

[20] Athanasios Lykartsis. Evaluation of accent-based rhythmic descriptors for genre classification of musical signals. Master's thesis, Audio Communication Group, Technische Universität Berlin, (www.ak.tu-berlin.de/menue/abschlussarbeiten/), 2014.

[21] Arnaud Moreau and Arthur Flexer. Drum transcription in polyphonic music using non-negative matrix factorisation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 353–354, 2007.

[22] Yannis Panagakis, Constantine Kotropoulos, and Gonzalo R Arce. Music genre classification using locality preserving non-negative tensor factorization and sparse representations. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR)*, 2009.

[23] Yannis Panagakis, Constantine L Kotropoulos, and Gonzalo R Arce. Music genre classification via joint sparse low-rank representation of audio features. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(12):1905–1917, 2014.

[24] Geoffroy Peeters. Spectral and temporal periodicity representations of rhythm for the automatic classification of music audio signal. *IEEE Transactions on Audio, Speech and Language Processing*, 19(5):1242–1252, 2011.

[25] Hanchuan Peng, Fulmi Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[26] Stanisław A Raczyński, Nobutaka Ono, and Shigeki Sagayama. Multipitch analysis with harmonic nonnegative matrix approximation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

[27] Axel Roebel. Onset detection in polyphonic signals by means of transient peak classification. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, 2005.

[28] Nicolas Scaringella, Giorgio Zoia, and Daniel Mlynek. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141, 2006.

[29] Eric D Scheirer. Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1):588–601, 1998.

[30] D Seung and L Lee. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[31] Paris Smaragdis and Judith C Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.*, pages 177–180, 2003.

[32] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka. Supervised and semi-supervised separation of sounds from single-channel mixtures. In *Independent Component Analysis and Signal Separation*, pages 414–421. Springer, 2007.

[33] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[34] Chih-Wei Wu and Alexander Lerch. Drum transcription using partially fixed non-negative matrix factorization. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2015.

[35] Jiho Yoo, Minje Kim, Kyeongok Kang, and Seungjin Choi. Nonnegative matrix partial co-factorization for drum source separation. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1942–1945, 2010.