Classification of Music Genres based on Repetitive Basslines

***Abstract*—In this paper, we present a comparative study of three different classification paradigms for genre classification based on repetitive basslines. In spite of a large variety in terms of instrumentation, a bass instrument can be found in most music genres. Thus, the bass track can be analyzed to explore stylistic similarities between music genres. We present an extensive set of transcription-based high-level features related to rhythm and tonality that allows to characterize basslines on a symbolic level. Traditional classification techniques based on pattern recognition techniques and audio features are compared with rule-based classification and classification based on the similarity between basslines. We use a novel dataset that consists of typical basslines of 13 music genres from different cultural backgrounds for evaluation purpose. Finally, the genre confusion results obtained in the experiments are examined by musicologists. Our study shows that several known stylistic relationships between music genres could be verified that way by classifying typical basslines. We could achieve a highest accuracy value of 64.8 % for the genre classification solely based on repetitive basslines of a song.**

***Index Terms*—bass, music genre classification, symbolic high-level features, rule-based classification, pattern similarity**

I. INTRODUCTION

In contrast to melody lines and harmonic progressions, basslines have rarely been investigated in Music Information Retrieval (MIR). The bass track plays an important role in music genres of different historical epochs from Western classical Baroque music to contemporary genres such as heavy metal, drum and bass, or rhythm'n'blues, as well as genres from various regional traditions from Western European, American, and African countries. From military marching bands to modern dance music, the sonic qualities of basslines very often play a key role in social affiliation and hence in the perception and discrimination of musical styles [Wicke, 2007]. Typical bass playing styles have evolved in a lot of genres.

Bass frequencies provide fundamental orientation in auditory experience due to the physics of sound and the human biology. They resonate strongly in the listeners body and head and are subconsciously perceived as powerful. In music, these qualities of bass frequencies are very often used to give a a basic structure as orientation in a complex composition or performance [Bruhn et al., 2005]. The bassline carries the main rhythmic and structural information of a song and provides insight to its underlying harmonic progression. Hence, the automatic classification of extracted basslines allows for describe a music piece in terms of harmony, rhythm, and style.

In this paper, we investigate 520 basslines that represent 13 different music genres from various historical and regional backgrounds (see Tab. V). We perform a comparative study of three different classification approaches to automatically classify the genre of a song solely based on tonal and rhythmic properties of the repetitive bassline in various evaluation experiments.

This paper is organized as follows. After providing an overview of related work in Sec. II, we illustrate different symbolic audio features for the tonal and rhythmic description of a bassline in Sec. III. Then, we explain the three different classification approaches we compare in this publication in Sec. IV. In Sec. V, we introduce the applied data set, describe the evaluation experiments that we performed, and discuss their results. Finally, we discuss the results, conclude our work in Sec. VI, and give an outlook in Sec. VII.

II. RELATED WORK

In this section, we first provide an overview of music score representation. Then, we present a short summary of extraction methods for score-based audio features and classification techniques that were previously applied in related research fields.

A. Score Representations & High-Level Features

In a lot of music styles treated in this publication, compositions are commonly not written out at all. On the other hand, some of its musical predecessors are score-based, especially regarding the basslines. Therefore, it does make sense to look at basslines in terms of a score. The *score parameters* note pitch, note onset time, note duration and note loudness are commonly applied to provide a simple yet meaningful symbolic representation of each note event. While a lot of information is necessarily left out in any form of transcription, these four parameters provide a scalable and comprehensive representation of the original bassline.

Basic score parameters can be extracted from the audio data via music transcription [Hainsworth and Macleod, 2001], [Goto, 2004], [Dittmar et al., 2007], [Ryynänen and Klapuri, 2008], and [Tsunoo et al., 2009] or directly from symbolic audio data.

We refer to audio features derived from score parameters as *high-level features*. In general, these features can capture different properties of music related to melody, harmony, and rhythm [McKay and Fujinaga, 2004], [McKay and Fujinaga, 2006], [Abeßer et al., 2009], [Abeßer et al., 2010a], [de Léon and Iñesta, 2007]. Most of these features correspond to the terminology of music theory. Usually, different statistical methods were used to derive high-level features from note onsets, pitches, and intervals [Pérez-Sancho et al., 2006], [de Léon and Iñesta, 2007]. For instance, high-level features based on entropy, compression, and prediction were presented in [Madsen and Widmer, 2007]. Rhythmical aspects like swing or syncopation were investigated in various publications as for instance in [Friberg and Sundström, 2002], [Gouyon, 2005] and [Flanagan, 2008].

Since inaccurate transcription results directly impair the accuracy of these high-level features, symbolic audio formats such as MIDI or Humdrum are preferably used for feature extraction. In contrast to real audio files, symbolic audio formats offer a direct access to the relevant score parameters. In the MIDI standard, each note is considered as a single note event that is characterized by its absolute pitch, its onset time, its offset time and its loudness. We use the MIDI Toolbox for Matlab [Eerola and Toiviainen, 2004] in this work to extract the basic score parameters from MIDI files. In this publication we do not analyze real audio data but manual transcriptions of audio recordings. Our goal is to investigate, to what extend transcription-based audio features derived from the bass track are generally applicable for music genre classification.

So far, only a few publications focussed solely on the description of the bass track. Tsuchihashi et. al. [Tsuchihashi et al., 2008] confined themselves to the progression of the note pitch values and distinguished between features characterizing the pitch variability and the pitch motion. In general, high-level features extracted from different instruments were successfully applied for genre classification tasks [de Léon and Iñesta, 2007], [Tsuchihashi et al., 2008], [Abeßer et al., 2008].

Many publications approached related aspects of *expressive music performance*. They aimed at characterizing the performance of a musician in terms of rhythmic and tonal play. For example, Stamatatos et. al. used features derived from the onset values, inter-onset-interval and loudness values of note progression to quantify the performance style of piano players in terms of their timing, articulation and dynamics [Stamatatos and Widmer, 2005]. An excellent overview of existing approaches for expressive performance analysis was provided in [Widmer et al., 2003] and [Widmer and Goebl, 2004].

B. Classification Approaches

a) Classification Based on Pattern Similarity: The computation of similarity between different melodies is useful for both music retrieval and analysis. The most common application of music similarity is query by humming (QbH) [Unal et al., 2008], which allows to identify a melody that was hummed without or with minor tonal or rhythmic errors before. Usually, each note of a melody is represented by its absolute pitch and onset value in order to represent a melody as a character sequences. Therefore, different distance measures are applied to measure the similarity between melodies. Commonly applied techniques are n-grams, the edit distance (Levenshtein distance) [Müllensiefen and Frieler, 2004], the Earth Movers Dicance (EMD) and the derived Proportional Transportation Distance (PTD) [Typke et al., 2002]. Other similarity measures were derived from the perception-based Implication/Realization (I/R) model [Grachten et al., 2004] or from a graph-based representation of musical structure [Orio and Rodá, 2009].

b) Classification Based on Statistical Pattern Recognition: Statistical pattern recognition methods were widely applied for classification of musical styles or genres based on the extracted acoustical features. In order to classify the musical content based on bass-related features the following techniques were used. The Mahalanobis distance of the feature vector to the feature distribution of each genre was utilized by [Tsuchihashi et al., 2008]. The genre minimizing the Mahalanobis distance was chosen as the classification result. Support Vector Machines (SVM), Gaussian Mixture Models (GMM), Naive Bayes (NB) and k-Nearest Neighbor (kNN) classifiers were utilized in [Abeßer et al., 2009] in a combination with various feature selection and feature space transformation methods. Audio genre classification based on bassline patterns in [Tsunoo et al., 2009] was performed by means of a SVM classifier. Bass playing style detection in [Abeßer et al., 2010a] was done by applying SVM classifier with a preceding feature selection algorithm.

The related work regarding general (not only bass-related) high-level features accounts the following techniques. Automatic genre classification using large high-level musical feature sets in [McKay and Fujinaga, 2004] was performed via a novel ensemble of feedforward neural networks and kNN classifiers. The features were selected with the aid of genetic algorithms. Decision trees, decision stumps and kNN were utilized in [Gouyon, 2005] to classify audio genres based on high-level rhythmical features. A broad palette of classification paradigms including kNN, NB, multilayer perceptron (MLP) and SVM were used for symbolic genre recognition in [Pérez-Sancho et al., 2006]. Fully supervised methods NB and kNN and unsupervised Self-Organizing Maps (SOM) were applied to statistical descriptors in [de Léon and Iñesta, 2007] for music style identification.

c) Rule-based Classification & Expert Systems: In contrast to abstract genre models, which are trained using supervised classifiers such as of SVM, the modeling of different genres or styles using a list of rules is more intuitive and comprehensible for humans [Abeßer et al., 2010b], [de León et al., 2007]. These rules correspond to a set of distinct properties of a music genre. Each rule can be expressed using a simple feature-value relation. The automatic learning of these rules was previously presented for harmony progressions [Anglade et al., 2009] and melody characterization [de León et al., 2007], but also for the purpose of automatic music generation [Buzzanca, 2001], [Bresin, 2001]. The authors of [Bresin, 2001] investigated the influence of different articulation rules for the improvement of automatically generated piano performances.

III. Proposed Method - Feature Extraction

A. Score Parameters

As mentioned in the previous chapter, we focus on the processing of symbolic audio data in this publication. The absolute pitch $\boldsymbol{\theta}_{\mathrm{P}}$, the note onset time $\boldsymbol{\varphi}_{\mathrm{O}}$, and the note duration $\boldsymbol{\varphi}_{\mathrm{D}}$ provide a description of the tonal and the rhythmic context

TABLE I
SCORE PARAMETERS

Score parameter	Notation	Vector dimension	Value range
Tonality			
Absolute pitch	$\boldsymbol{\theta}_{\mathrm{P}}$	$\mathbb{Z}^N$	$[0, 127]$
(Chroma)	$\boldsymbol{\theta}_{\mathrm{P,12}}$	$\mathbb{Z}^N$	$[0, 11]$
Relative pitch	$\boldsymbol{\Delta\theta}_{\mathrm{P}}$	$\mathbb{Z}^{N-1}$	$[-127, 127]$
(Mapped to one octave)	$\boldsymbol{\Delta\theta}_{\mathrm{P,12}}$	$\mathbb{Z}^{N-1}$	$[-11, 11]$
Interval direction	$\boldsymbol{\Delta\theta}_{\mathrm{P,D}}$	$\mathbb{Z}^{N-1}$	$\{-1, 0, 1\}$
Functional intervals	$\boldsymbol{\Delta\theta}_{\mathrm{P,F}}$	$\mathbb{Z}^{N-1}$	$\{-7, ..., -2, 1, ..., 7\}$
Rhythm			
Onset *(in seconds)*	$\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{s}]}$	$\mathbb{R}^N$	$[0, \infty)$
Onset *(in measures)*	$\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$	$\mathbb{R}^N$	$[0, \infty)$
Simplified onset	$\boldsymbol{\varphi}_{\mathrm{O,1}}^{[\mathrm{m}]}$	$\mathbb{R}^N$	$[0, 1)$
Duration *(in seconds)*	$\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{s}]}$	$\mathbb{R}^N$	$(0, \infty)$
Duration *(in measures)*	$\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$	$\mathbb{R}^N$	$(0, \infty)$
Inter-onset interval *(in measures)*	$\boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$	$\mathbb{R}^{N-1}$	$(0, \infty)$
Note-duration ratio	$\boldsymbol{\varphi}_{\mathrm{DR}}$	$\mathbb{R}^{N-1}$	$(0, \infty)$

of each note. The note duration is derived by subtracting the onset time from the offset time for each note. Both the note onset and duration are commonly measured in seconds. The absolute pitch is assigned to an integer value between 0 and 127 according to the MIDI standard. In this work, we do not use the note loudness to derive features.

Tab. I provides an overview of all score parameters used in this work as well as their notation, dimensionality, and value range. Some of the score parameters have discrete values, some have continuous values. All score parameters are stored as vectors, which we denote in bold print. In the following subsections, we explain further score parameters related to rhythm and tonality that are computed before features are extracted. A example that contains an excerpt of a funk bassline and the derived score parameters is given in Tab. III.

1) Score Parameters (Tonality): We assume that a bassline consists of N notes and has a length of M measures. The *chromatic representation* $\boldsymbol{\theta}_{\mathrm{P,12}}$ of the absolute pitch $\boldsymbol{\theta}_{\mathrm{P}}$ is derived as

$$\theta_{\mathrm{P,12,k}} = \theta_{\mathrm{P,k}} \bmod 12, \qquad \boldsymbol{\theta}_{\mathrm{P,12}} \in \mathbb{Z}^N. \tag{1}$$

The chromatic values represent all absolute pitch values mapped to one octave ($\theta_{\mathrm{P,12,k}} \in [0, 11]$). By definition, $\theta_{\mathrm{P,12,k}} = 0$ corresponds to the note name C.

The relative pitch $\boldsymbol{\Delta\theta}_{\mathrm{P}}$ describes the interval size between two consecutive notes as

$$\Delta\theta_{\mathrm{P,k}} = \theta_{\mathrm{P,k+1}} - \theta_{\mathrm{P,k}}, \qquad \boldsymbol{\Delta\theta}_{\mathrm{P}} \in \mathbb{Z}^{N-1}. \tag{2}$$

We use two additional score parameters: $\boldsymbol{\Delta\theta}_{\mathrm{P,D}}$ represents the *interval direction* and $\boldsymbol{\Delta\theta}_{\mathrm{P,F}}$ provides a *functional interval representation*. The interval direction $\Delta\theta_{\mathrm{P,D,k}}$ between two notes is 1 if the interval between both notes is ascending, -1 if the interval is descending, or 0 in case both notes have the same absolute pitch value. It is computed as

$$\Delta\theta_{\mathrm{P,D,k}} = \operatorname{sgn} \Delta\theta_{\mathrm{P,k}}, \qquad \boldsymbol{\Delta\theta}_{\mathrm{P,D}} \in \mathbb{Z}^{N-1}, \Delta\theta_{\mathrm{P,D,k}} \in \{-1, 0, 1\}. \tag{3}$$

TABLE II
FUNCTIONAL REPRESENTATION OF INTERVALS

$\widehat{\Delta\theta_{\mathrm{P,k}}}$	$\Delta\theta_{\mathrm{P,F,k}}$	Interval name
-11,-10	-7	Descending seventh
-9,-8	-6	Descending sixth
-7,-6	-5	Descending fifth
-5	-4	Descending fourth
-4,-3	-3	Descending third
-2,-1	-2	Descending second
0	1	Prime
1,2	2	Ascending second
3,4	3	Ascending third
5	4	Ascending fourth
6,7	5	Ascending fifth
8,9	6	Ascending sixth
10,11	7	Ascending seventh

The functional interval representation $\boldsymbol{\Delta\theta}_{\mathrm{P,F}}$ is set according to the musical interval name that corresponds to each interval. Therefore, we map each interval to a range between one octave downwards and one octave upwards as

$$\widehat{\Delta\theta_{\mathrm{P,k}}} = \Delta\theta_{\mathrm{P,D,k}} \left(|\Delta\theta_{\mathrm{P,k}}| \bmod 12 \right) \tag{4}$$

The assignment between $\widehat{\Delta\theta_{\mathrm{P,k}}}$ and the function interval representation $\Delta\theta_{\mathrm{P,F,k}}$ is illustrated in Tab. II.

2) Score Parameters (Rhythm): The onset times and durations of note events can be expressed in both a physical and a musical time representation. We use the superscript indices *s* and *m* to indicate that both the onset φ_{O} and the duration φ_{D} of a note can be measured in seconds as well as in fractions of measure lengths. Thus, $\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$ and $\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$ provide a tempo-independent representation of the note onset times and note lengths. This allows to compute rhythmic features independent of the large range of tempo values in different music genres we investigate in this publication (see Fig. 7). Both $\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$ and $\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$ can be extracted from the MIDI files using the MIDI toolbox (see [Eerola and Toiviainen, 2004]).

The note onset values are converted into a *simplified onset representation* $\boldsymbol{\varphi}_{\mathrm{O,1}}^{[\mathrm{m}]}$ by neglecting the current measure number as

$$\varphi_{\mathrm{O,1,k}}^{[\mathrm{m}]} = \varphi_{\mathrm{O,k}}^{[\mathrm{m}]} \bmod 1, \qquad \boldsymbol{\varphi}_{\mathrm{O,1}}^{[\mathrm{m}]} \in \mathbb{R}^N. \tag{5}$$

The *inter-onset interval* (IOI) $\boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$ is derived from the onset difference between two consecutive notes, it is defined by

$$\Delta\varphi_{\mathrm{O,k}}^{[\mathrm{m}]} = \varphi_{\mathrm{O,k+1}}^{[\mathrm{m}]} - \varphi_{\mathrm{O,k}}^{[\mathrm{m}]}, \qquad \boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]} \in \mathbb{R}^{N-1}. \tag{6}$$

The *note-duration ratio* (NDR) $\boldsymbol{\varphi}_{\text{DR}}$ is computed as the ratio between the duration values of consecutive note as

$$\varphi_{\text{DR,k}} = \frac{\varphi^{[\text{m}]}_{\text{D,k+1}}}{\varphi^{[\text{m}]}_{\text{D,k}}}, \qquad (7)$$
$$\boldsymbol{\varphi}_{\text{DR}} \in \mathbb{R}^{N-1}.$$

B. Feature Extraction

In the following subsections, we explain the computation of various high-level features based on the previously extracted score parameters. Features are denoted as α with different subscripts.

Furthermore, we define two auxiliary functions $\boldsymbol{n}(\boldsymbol{x}, \boldsymbol{v})$ and $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{v})$. Consider $\boldsymbol{x}$ to be a vector that represents an arbitrary score parameter and $\boldsymbol{v}$ to be a vector with all different values taken by the elements x_i. Then n_i denotes the number of elements in $\boldsymbol{x}$ with $x_j = v_i$. Each element p_i is the (relative) frequency of its corresponding value v_i in the vector $\boldsymbol{x}$:

$$p_i = \frac{n_i}{||\boldsymbol{x}||} \qquad (8)$$

1) Tonality:

a) Basic tonal features: To capture the overall tonal range of a bassline, we compute feature as

$$\alpha_{\text{TR}} = \max_k \theta_{\text{P,k}} - \min_k \theta_{\text{P,k}}, \qquad (9)$$
$$k \in [1, N].$$

By computing the frequencies of occurrence for absolute pitch values in $\boldsymbol{\theta}_{\text{P}}$, we detect the most frequent pitch value $\theta_{\text{P,dom}}$ and the number of notes $n(\boldsymbol{\theta}_{\text{P}}, \theta_{\text{P,dom}})$ having this pitch value. If multiple pitch values appear equally often, we take the lowest of these pitch values as $\theta_{\text{P,dom}}$. We use

$$\alpha_{\text{dom}} = p(\boldsymbol{\theta}_{\text{P}}, \theta_{\text{P,dom}}) \qquad (10)$$

as a feature to measure the dominance of the most frequent pitch value.

Some basslines are based on a so-called *pedal-tone*. This term is used if the dominant pitch value is the lowest pitch value in a pattern. Furthermore, this pitch value is constantly repeated with only a few variations into higher pitches. We compute a feature as

$$\alpha_{\text{Pedal}} = \frac{\theta_{\text{P,dom}} - \theta_{\text{min}}}{\theta_{\text{max}} - \theta_{\text{min}}}. \qquad (11)$$

The use of a pedal-tone is most likely if value of α_{pedal} is very small. We do not take the duration into account here although some bass notes might be hold and thus act as a pedal-tone without being repeated.

b) Chroma-based features: Based on the chromatic pitch representation $\boldsymbol{\theta}_{\text{P,12}}$, we compute the frequencies of occurrence $\boldsymbol{p}(\boldsymbol{\theta}_{\text{P,12}}, \boldsymbol{c})$ with

$$\boldsymbol{c} \in \mathbb{Z}^{12}, \; c_i \in [0, 11] \qquad (12)$$

containing all possible chroma values between 0 and 11.

We compute the zero-order entropy H_0 over $\boldsymbol{p}(\boldsymbol{\theta}_{\text{P,12}}, \boldsymbol{c})$ as a tonal feature to measure whether only a few or a lot of chromatic values are present in a bassline. The number of different chromatic pitch values in a bassline can provide evidence about its tonal complexity.

TABLE IV
INVESTIGATED SCALES.

Scale index s	Scale name	Scale template $\boldsymbol{t}_{\text{s}}$
0	Natural minor	1 0 1 1 0 1 0 1 1 0 1 0
1	Harmonic minor	1 0 1 1 0 1 0 1 1 0 0 1
2	Melodic minor	1 0 1 1 0 1 0 1 0 1 0 1
3	Pentatonic minor	1 0 0 1 0 1 0 1 0 0 1 0
4	Blues minor	1 0 0 1 0 1 1 1 0 0 1 1
5	Whole tone	1 0 1 0 1 0 1 0 1 0 1 0
6	Whole tone half tone	1 0 1 1 0 1 1 0 1 1 0 1
7	Arabian	1 1 0 0 1 1 0 1 1 0 0 1
8	Minor gypsy	1 0 1 1 0 0 1 1 1 0 1 0
9	Hungarian gypsy	1 0 1 1 0 0 1 1 1 0 0 1

c) Scale: The choice of tonal scales is a distinctive feature of different music genres. A scale is a set of pitch values that are related by a unique interval structure. Usually, these pitch values provide the melodic material for a musical composition. We compute a measure-of-fit between the analyzed bassline and 10 different scales listed in Tab. IV to detect the most likely scale, which a bassline is based on. In this work, we do not take the local information on whether a bassline is ascending or descending in pitch into account to determine the prominent scale.

In this work, each scale is represented by a *scale template* $\boldsymbol{t}_{\text{s}} \in \mathbb{R}^{12}$. We use a simplified representation $t_{\text{s,i}} \in \{0, 1\}$ to clearly discern between outside-scale tones ($t_{\text{s,i}} = 0$) and inside-scale tones ($t_{\text{s,i}} = 1$). The first value $t_{\text{s,1}}$ represents the lowest note of a scale, which is referred to as its root note. The vector $\boldsymbol{t}_{\text{s}}$ can be rotated by r semitones using a cyclic shift operation to derive $\boldsymbol{t}_{\text{s,r}}$ with $r \in [0, 11]$. This operation does not affect the interval structure of a scale, it only changes the chromatic value of its root-note.

To compute a measure-of-fit for a scale s, we use a template-matching approach. First, we compute the frequencies of occurrence $\boldsymbol{p}(\boldsymbol{\theta}_{\text{P,12}}, \boldsymbol{c})$ of all chromatic values $\boldsymbol{c} \in [0, 11]$ in a bassline. For each scale index s and each rotation $r \in [0, 11]$, we compute

$$\gamma_{\text{s,r}} = \frac{\sum (\boldsymbol{t}_{\text{s,r}} * \boldsymbol{p}(\boldsymbol{\theta}_{\text{P,12}}, \boldsymbol{c}))}{\sum \boldsymbol{t}_{\text{s,r}}} \qquad (13)$$

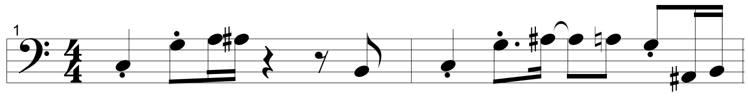
to measure the likelihood of scale s rotated by r semitones. $*$ denotes the element-wise product between 2 vectors. The summation is performed over all vector elements in each case. Finally, we compute the maximum ratio

$$\alpha_{\text{SC,s}} = \max_{\forall \text{s,r}} \gamma_{\text{s,r}} \qquad (14)$$

over all rotations $r \in [0, 11]$ to measure the likelihood of each scale $s \in [0, 9]$ for a given bassline and use all $\alpha_{\text{SC,s}}$ as features.

d) Interval Types & Progression: To characterize the interval progression, we use the frequencies of different typical note sequences in a bassline as features. Hence, we seek for

TABLE III
SCORE PARAMETERS FOR AN EXCERPT OF A FUNK BASSLINE



Note number	k	1	2	3	4	5	6	7	8	9	10	11	12
Note name		C^3	G^3	A^3	$A\sharp^3$	B^2	C^3	G^3	$A\sharp^3$	A^3	G^3	$A\sharp^2$	B^2
Absolute pitch	$\boldsymbol{\theta}_{\text{P}}$	48	55	57	58	47	48	55	58	57	55	46	47
(Chroma)	$\boldsymbol{\theta}_{\text{P,12}}$	0	7	9	10	11	0	7	10	9	7	10	11
Relative pitch	$\boldsymbol{\Delta\theta}_{\text{P}}$	7	2	1	-11	1	7	3	-1	-2	-9	1	-
(Mapped to octave)	$\boldsymbol{\Delta\theta}_{\text{P,12}}$	7	2	1	-11	1	7	3	-1	-2	-9	1	-
(Interval direction)	$\boldsymbol{\Delta\theta}_{\text{P,D}}$	1	1	1	-1	1	1	1	-1	-1	-1	1	-
(Functional intervals)	$\boldsymbol{\Delta\theta}_{\text{P,F}}$	5	2	2	-7	2	5	3	-2	-2	-6	2	-
Onset	$\boldsymbol{\varphi}_{\text{O}}^{[\text{m}]}$	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{7}{8}$	1	$1\frac{1}{4}$	$1\frac{7}{16}$	$1\frac{10}{16}$	$1\frac{3}{4}$	$1\frac{7}{8}$	$1\frac{15}{16}$
Simplified onset	$\boldsymbol{\varphi}_{\text{O,1}}^{[\text{m}]}$	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{7}{16}$	$\frac{7}{8}$	0	$\frac{1}{4}$	$\frac{7}{16}$	$\frac{10}{16}$	$\frac{3}{4}$	$\frac{7}{8}$	$\frac{15}{16}$
Duration	$\boldsymbol{\varphi}_{\text{D}}^{[\text{m}]}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{32}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
Inter-Onset-Interval	$\boldsymbol{\Delta\varphi}_{\text{O}}^{[\text{m}]}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	-
Note-duration-ratio	$\boldsymbol{\varphi}_{\text{DR}}$	$\frac{1}{2}$	1	1	2	1	$\frac{3}{4}$	2	$\frac{2}{3}$	$\frac{1}{2}$	1	1	-

sequences of consecutive notes with *constant pitch* ($\Delta\theta_{\text{P,k}} = 0$) as well as *chromatic transitions between notes and measures*. A chromatic transition between notes is characterized by two equal consecutive intervals of plus or minus one semitone:

$$\Delta\theta_{\text{P,k}} = \Delta\theta_{\text{P,k+1}},\ \Delta\theta_{\text{P,k}} \in \{-1, 1\}\,, \quad k \in [1, N-1]\,. \tag{15}$$

If the pitch difference between the last note of a measure and the first note of the consecutive measure is one semitone, we call the transition between both measures chromatic. In music genres related to Jazz, the walking bass playing style is often used. Here, chromatic transitions between measures can be found in the bassline if the last note in a measure is the leading note towards the first note of the next measure. For both types of transitions, we take the ratio between the number of note sequences and the overall number of notes as features.

The mean and the variance over the absolute interval size $|\boldsymbol{\Delta\theta}_{\text{P}}|$ are computed to characterize whether a bassline comprises mainly small or large intervals. Furthermore, we compute the number of intervals with ascending pitch $n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, 1\right)$ and the number of intervals with descending pitch $n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, -1\right)$. The feature

$$\alpha_{\text{DD}} = \frac{n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, 1\right)}{n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, 1\right) + n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, -1\right)} \tag{16}$$

measures whether an ascending or a descending interval direction is *dominant* in the bassline. In case $\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, 1\right) + n\left(\boldsymbol{\Delta\theta}_{\text{P,D}}, -1\right) = 0$, α_{DD} is set to 0.

A bassline is perceived as smooth and fluent if consecutive intervals often have the same direction. Therefore, we compute the frequencies of consecutive note transitions with the interval diraction value $\Delta\theta_{\text{P,D,k}}$ as a measure of *constant direction* α_{CD}.

$$\alpha_{\text{CD}} = \frac{1}{N-2}\sum_{k=1}^{N-2}\beta_k, \tag{17}$$

$$\beta_k = \begin{cases} 1 & \text{for } \Delta\theta_{\text{P,D,k}} = \Delta\theta_{\text{P,D,k+1}} \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

To measure the variety of applied interval types, we compute the frequencies of each possible value of $\boldsymbol{\Delta\theta}_{\text{P,F}}$ (see Tab. I).

As additional features, we compute the zero-order entropy $H(\boldsymbol{\Delta\theta}_{\text{P,F}})$ over the vector containing the functional pitch representation $\boldsymbol{\Delta\theta}_{\text{P,F}}$, the mean and the variance of the absolute size of all intervals $|\Delta\theta_{\text{P,k}}|$ as well as the absolute number of different intervals appearing in a bassline.

2) Rhythm:

a) Rhythmic Subdivisions: In this work, we investigate different rhythmic subdivisions of basslines. Therefore, each measure is divided into Q equidistant subdivisions with the indices $b \in [1, Q]$. The value of Q relates to the corresponding musical note length values such as crotchets ($Q = 4$) or eight-notes ($Q = 8$). Each beat is associated with a musical onset time

$$\varphi_{\text{Q,b}}^{[\text{m}]} = b/Q,\ b \in [1, Q]\,. \tag{19}$$

For each subdivision Q, each note k of a bassline can be quantized to an adjacent beat with the index $\hat{b}_k$ as

$$\hat{b}_k = \arg\min_{b \in [1,Q]} \left|\varphi_{\text{O,1,k}}^{[\text{m}]} - \varphi_{\text{Q,b}}^{[\text{m}]}\right|. \tag{20}$$

If the note is exactly located between two beats, it is quantized to the earlier beat. All beats with odd indices

$$b_{\text{On,i}} = 2^i - 1,\ i > 0$$

are referred to as *on-beats*, all others as *off-beats*.

b) On-beat Accentuation: Since on-beats are usually emphasized in most Western music genres, we aim to measure the accentuation of on-beats in a given bassline. We compute the relative frequency of notes $n_{\mathrm{On,Q,m}}$ that are quantized to on-beats of a given subdivision Q as

$$\alpha_{\mathrm{On,Q,m}} = \frac{n_{\mathrm{On,Q,m}}}{N} \tag{21}$$

is computed for all measures.

We use the mean and the variance over all $\alpha_{\mathrm{On,Q,m}}$ with $m \in [1, M]$ as rhythmic features. Both features are calculated for the subdivisions $Q \in \{2, 4, 8, 16, 32\}$.

These features for instance allow to distinguish between walking bass basslines in swing and basslines in reggae. Walking basslines mainly comprise of crotches that correspond to on-beat notes for $Q = 8$. The frequent use of crotchets result in high mean values over $\alpha_{\mathrm{On,Q,m}}$. In contrast, reggae basslines often contain accentuations shifted from on-beats to following off-beats and thus lead to lower mean values over $\alpha_{\mathrm{On,Q,m}}$.

c) Dominant Rhythmic Feeling: Binary and ternary rhythms are dominating in the music genres investigated in this publication. These rhythms are based on rhythmic subdivisions Q that are multiples of 2 or 3. We derive features to capture the dominant *rhythmic feeling* of the notes in a given bassline. Therefore, we investigate all rhythmic subdivisions in the matrix

$$\boldsymbol{Q}_{\mathrm{F}} = \begin{bmatrix} 2 & 4 & 8 & 16 & 32 & 64 \\ 3 & 6 & 12 & 24 & 48 & 96 \end{bmatrix}.$$

The binary subdivisions are placed in the first row and the corresponding ternary subdivision in the second row of $\boldsymbol{Q}_{\mathrm{F}}$.

For each rhythmic subdivision $Q_{\mathrm{F,i,j}}$, we compute a simplified onset representation similar to the approach previously shown in Eq. (6). In contrast, we do not use a length of one measure as reference but the inter-onset-intervals between two adjacent beats for a subdivision $Q_{\mathrm{F,i,j}}$ as

$$\varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}} = \varphi^{[\mathrm{m}]}_{\mathrm{O,k}} \bmod \frac{1}{Q_{\mathrm{F,i,j}}}, \quad k \in [1, N], \tag{22}$$

The values $\varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}}$ correspond to the onset distances between the k-th note and its adjacent preceding beat for a given subdivision $Q_{\mathrm{F,i,j}}$.

To evaluate if a note has an onset close to one of the beats of a given subdivision, we have to consider that it can be located either closely before or after a beat. Therefore we modify $\boldsymbol{\varphi}^{[\mathrm{m}]}_{\mathrm{O,Q}}$ as

$$\widehat{\varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}}} = \min\left\{\varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}}\,,\ \frac{1}{Q_{\mathrm{F,i,j}}} - \varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}}\right\} \tag{23}$$

to obtain a two-side distance measure.

To detect the dominant rhythmic feeling, we investigate the percentage of notes that can be associated to each of the subdivisions in $\boldsymbol{Q}_{\mathrm{F}}$. Thus, we use the distance criterion

$$\widehat{\varphi^{[\mathrm{m}]}_{\mathrm{O,Q,k}}} \leq 0.05 \frac{1}{Q_{\mathrm{F,i,j}}} \tag{24}$$

to decide if the the k-th note is mapped to the subdivision $Q_{\mathrm{F,i,j}}$. For each subdivision $Q_{\mathrm{F,i,j}}$, we store the percentage of notes assigned to this subdivision in a matrix

$$\boldsymbol{c}_{\mathrm{F}} \in \mathbb{R}^{2\times 6}$$

The value $c_{\mathrm{F,i,j}}$ can be thought of as a measure-of-fit for the subdivision to characterize a bassline. To investigate whether binary or ternary subdivisions are dominant, we compare the values in $\boldsymbol{c}_{\mathrm{F}}$ column-wise. We compute the number of binary subdivisions n_{B} that have a higher value $c_{\mathrm{F,i,j}}$ than their corresponding ternary subdivision as

$$n_{\mathrm{B}} = \sum_{j=1}^{6} \gamma_j,\ \gamma_j = \begin{cases} 1 & \text{for } c_{\mathrm{F,1,j}} \geq c_{\mathrm{F,2,j}} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

and the number of dominant ternary subdivisions n_{B} for $c_{\mathrm{F,1,j}} \leq c_{\mathrm{F,2,j}}$ accordingly. Finally, we obtain two features

$$\alpha_{\mathrm{BS}} = \frac{n_{\mathrm{B}}}{n_{\mathrm{B}} + n_{\mathrm{T}}} \tag{26}$$

and

$$\alpha_{\mathrm{TS}} = \frac{n_{\mathrm{T}}}{n_{\mathrm{B}} + n_{\mathrm{T}}} \tag{27}$$

to measure the dominance of binary and ternary subdivisions in a given bassline.

d) Dominant Rhythmic Subdivision: As described in the previous subsection, we can investigate the percentage of notes in a bassline that are associated to different rhythmic subdivisions. We define the *dominant rhythmic subdivision* as the subdivision whom at least 80 percent of all notes are assigned to. Therefore, we seek the lowest indices $\hat{\imath}$ and $\hat{\jmath}$ that fulfill $c_{\mathrm{F,i,j}} \geq 0.8$. We store the dominant rhythmic subdivision

$$\alpha_{\mathrm{RS}} = Q_{\mathrm{F},\hat{\imath},\hat{k}} \tag{28}$$

as feature. In case no value in $\boldsymbol{c}_{\mathrm{F}}$ is bigger than 0.8, we use

$$\alpha_{\mathrm{RS}} = \begin{cases} Q_{\mathrm{F,1,6}} & \text{, for } n_{\mathrm{B}} \geq n_{\mathrm{T}} \\ Q_{\mathrm{F,2,6}} & \text{, otherwise.} \end{cases} \tag{29}$$

e) Note Density: In order to describe the note density distribution in a particular bassline, we take the mean and variance of the number of notes per measure over all measures as features.

f) Syncopation: Syncopation is frequently used especially in Latin American music genres such as bossa nova or salsa. Instead of playing notes on on-beat positions, the accentuation is often shifted to the adjacent off-beat positions. To detect syncopated note sequences within a bassline, we again investigate multiple temporal subdivisions $Q \in \{4, 8, 16, 32\}$. For each subdivision Q, we map all notes inside a measure to closest beat inside this measure as described in Eq. (20). If at least one note is mapped to a particular beat, the beat is associated with the value 1, otherwise with 0.

In Fig. 1, we illustrate four examples of syncopated note sequences for a eight-note subdivisions ($Q = 8$). Each sequence of notes corresponds to a sequence of alternating on-beat and off-beat accentuations that musicologist refer to

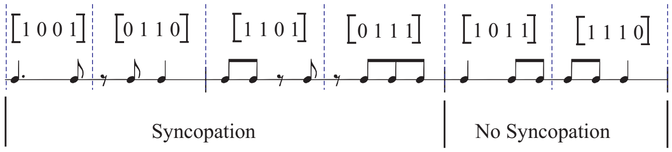


Fig. 1. Note sequences with and without syncopation for an quaver subdivision ($Q = 8$) and corresponding beat sequence shown above.

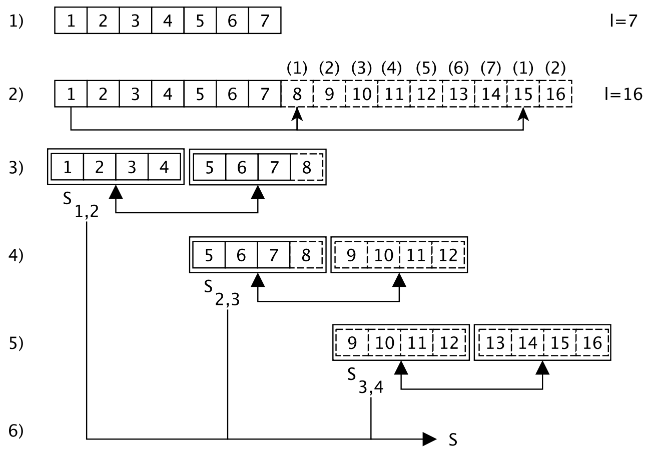


Fig. 2. Computation of the sub-pattern similarity for a pattern of a length of $l = 7$ measures and a sub-pattern length of $N_{SP} = 4$ measures. First, the pattern is elongated to a length of 16 measures by repeatedly appending the pattern to itself. Then, the similarity between adjacent sub-patterns is computed and averaged to compute the sub-pattern similarity values.

as syncopations. After quantizing the notes to a subdivision Q, we compute the number of the beat sequences - [1 0 0 1], [0 1 1 0], [1 1 0 1], and [0 1 1 1] in a bassline. The ratio between the number of syncopation sequences and the overall number of beats per measure is computed as features.

g) Sub-Pattern Similarity: The basslines used in this work have different lengths between 4 and 16 measures. They often contain one ore two measures that are repeated within the complete pattern with only slight variation. These *sub-patterns* can be characteristic for a bassline. To compute features, that capture whether sub-patterns of a bassline are similar to each other, we proceed as follows.

First, if necessary, the pattern is elongated to a length of 16 measures by simply repeatedly appending the complete pattern to itself. This approach is valid since we assume that the basslines appear as repeating patterns in real songs. Then, we compute the similarity between consecutive sub-patterns of a length of N_{SP} measures within the bassline. For this feature, we use the Levensthein distance, that will be explained in Sec. IV-C1 based on the onset values $\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$ as distance measure. Finally, all similarity values are averaged to compute the feature $\alpha_{\mathrm{Sub,N}}$ for the resolutions $N_{\mathrm{SP}} \in \{2, 4, 8\}$. Fig. 2 illustrates this procedure for a pattern of a length of $l = 7$ measures and $N_{SP} = 4$ measures.

h) Basic Feature Related to the Note Duration: In addition to the features derived from the note onset values, we obtain various characteristics regarding the note duration vector $\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$ and the inter-onset interval vector $\boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$. We calculate the mean and the variance over $\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$ as basic statistical features. To describe the variety in terms of note lengths and distances, we derive the number of different values in $\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}$ and $\boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$.

Furthermore, we compute the frequencies of occurance $\boldsymbol{p}\left(\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}, \widehat{\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}}\right)$ of 15 different note duration values

$$\widehat{\varphi_{\mathrm{D,i}}^{[\mathrm{m}]}} = \frac{a}{b},\ a \in \left\{1, \frac{2}{3}, \frac{3}{2}\right\},\ b \in \{2, 4, 8, 16, 32\} \tag{30}$$

and use all p_i as features. The values of a represent the regular note ($a = 1$), the triplet note ($a = 2/3$), and the dotted note ($a = 3/2$) and the values of b correspond to the note length of a second note ($b = 2$), a crotchet ($b = 4$), and so forth. Each note duration value $\varphi_{\mathrm{D,k}}^{[\mathrm{m}]}$ is mapped to the $\hat{\imath}$-th element of $\widehat{\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}}$ as

$$\hat{\imath} = \arg\min_i \left| \varphi_{\mathrm{D,k}}^{[\mathrm{m}]} - \widehat{\varphi_{\mathrm{D,i}}^{[\mathrm{m}]}} \right|. \tag{31}$$

i) Features Related to the Note-Duration Ratio (NDR): The NDR $\boldsymbol{\varphi}_{\mathrm{DR}}$ provides a tempo-independent representation that characterizes the progression of the note duration values in a bassline. Basslines from the music genres minimal techno, motown, hip-hop, and blues often contain note sequences of constant duration as for instance simple quaver octave patterns. Hence, the majority of the NDR values will be most likely equal to one. We compute the relative frequencies of $p(\boldsymbol{\varphi}_{\mathrm{DR}} < 1)$, $p(\boldsymbol{\varphi}_{\mathrm{DR}} = 1)$, and $p(\boldsymbol{\varphi}_{\mathrm{DR}} > 1)$ as features.

j) Rhythmic Articulation: To characterize the dominant *rhythmic articulation* which is commonly referred to by the terms *staccato* and *legato*, we use the ratio

$$\alpha_{\mathrm{SL}} = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]}(i)}{\boldsymbol{\Delta\varphi}_{\mathrm{O}}^{[\mathrm{m}]}(i)} \tag{32}$$

as a feature to measure the mean ratio between the length of a note and its distance to its successor. A high value indicates very short or no breaks between consecutive notes which refers to legato play, low values of α_{SL} point to staccato play.

Overall, we obtain a 101 dimensional feature vector representing each bassline.

IV. Proposed Method - Classification

In this publication, we compare three different approaches for the classification of the bass playing style. Fig. 3 provides an overview.

A. Classification Based on Statistical Pattern Recognition

We apply a state-of-the-art supervised classification technique, namely Support Vector Machine (SVM). SVM is a binary discriminative classifier, attempting to generate an optimal decision plane between feature vectors of the training classes [Vapnik, 1998]. Commonly for real-world applications, classification with linear separation planes is not possible in the original feature space. To overcome this problem, the so called *kernel trick* is applied. The key idea of the kernel trick is

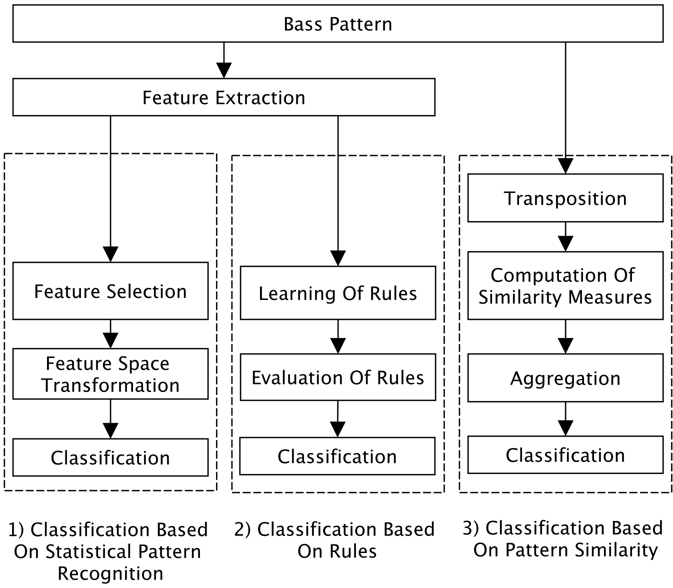


Fig. 3. Three different classification approaches.

to replace the dot product in a high-dimensional space with a kernel function in the original feature space. Transformed into a high-dimensional space, non-linear classification problems can become linearly solvable. In the experiments in this work, we use the most common type of kernel, namely Radial Basis Function (RBF).

B. *Classification Based on Rules*

The classification rules in this work are inducted using a Classification and Regression Tree (CART) recursive partitioning technique with subsequent optimal pruning strategy as proposed in [Breiman et al., 1993]. CART is a nonparametric data mining algorithm for generating decision trees, that does not make any assumption about the specific distribution of input feature vectors. Decision tree splits are made by using the features that possess the best discriminative properties in respect to the target classes. In CART each node can be split into two child nodes. This recursive partitioning is continued until the certain stopping rules are fulfilled. The optimal parameters for the stopping rules – e.g. a minimum number of items per node to be still considered for splitting – are determined experimentally. The generalization properties of the decision tree are controlled in a cross validation scenario, where the tree is pruned to a certain level in order to prevent overfitting the training data.

C. *Classification Based on Pattern Similarity*

In many genres, typical basslines have evolved that are frequently used with only some variation. We investigate two different pattern similarity measures to compare basslines with regard to their tonal and rhythmic similarity. Therefore, we use either the vector $\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$ containing the note onset values or the vector $\boldsymbol{\theta}_{\mathrm{P}}$ containing the absolute pitch values to represent a bassline $\boldsymbol{\mathcal{P}}$ as depicted in Tab. III. Hence, $\boldsymbol{\mathcal{P}}$ can be thought of as a character sequence that allows to use edit distances such as Levenshtein distance to compare two patterns. For example, the funk bassline illustrated as an example in Tab. III is represented by

$$\boldsymbol{\mathcal{P}} = [48, 55, 57, 58, \ldots] \tag{33}$$

for all tonal similarity measures and by

$$\boldsymbol{\mathcal{P}} = \left[0, \frac{1}{4}, \frac{3}{8}, \frac{7}{16}, \ldots\right] \tag{34}$$

for all rhythmic similarity measures.

Since two basslines may be very similar but notated in different keys, their absolute pitch values have to be aligned to avoid erroneous similarity values. Instead of using the lowest pitch value, we take the most frequent pitch value $\theta_{\mathrm{P,dom}}$ (see Sec. III-B1) as reference value. Again, if multiple absolute pitch appear equally often, we select the lowest candidate as the most frequent pitch value $\theta_{\mathrm{P,dom}}$.

Hence, if two arbitrary patterns $\boldsymbol{\mathcal{P}}_1$ and $\boldsymbol{\mathcal{P}}_2$ are to be compared using a pitch-related similarity measure, we modify all absolute pitch values $\theta_{\mathrm{P,2,k}}$ of $\boldsymbol{\mathcal{P}}_2$ as

$$\hat{\theta_{\mathrm{P,2,k}}} = \theta_{\mathrm{P,2,k}} + \theta_{\mathrm{P,1,dom}} - \theta_{\mathrm{P,2,dom}},\ k \in [1, N_2]\,. \tag{35}$$

We compare two similarity measures in this publication, one based on the Levenshtein distance and another one based on a pairwise pattern similarity measure.

1) Levensthein Distance: The Levensthein distance D_{L} allows to compute the similarity between character strings [Gusfield, 1997]. Therefore, the minimum number of edits in terms of insertions, deletions, and substitutions of characters is determined, which is necessary to convert one string into the other. We apply the Wagner-Fischer algorithm as described in [Wagner and Fischer, 1974] to compute D_{L}. The similarity measure S_{L} is derived as

$$S_{\mathrm{L}} = 1 - \frac{D_{\mathrm{L}}}{D_{\mathrm{L,max}}}, \tag{36}$$

where $D_{\mathrm{L,max}}$ equals the maximum value between the lengths l_1 and l_2 of both patterns:

$$D_{L,max} = \begin{cases} l_1 & , l_1 \geq l_2 \\ l_2 & , l_2 > l_1. \end{cases} \tag{37}$$

In the experiments, we use the rhythmic similarity measure $S_{\mathrm{L,R}}$ derived from the Levenshtein distance between the onset vectors $\boldsymbol{\varphi}_{\mathrm{O}}^{[\mathrm{m}]}$. A tonal similarity measure $S_{\mathrm{L,T}}$ is derived accordingly from the Levenshtein distance between absolute pitch vectors $\boldsymbol{\theta}_{\mathrm{P}}$. Furthermore, we investigate the similarity measures

$$S_{\mathrm{L,RT,Max}} = \begin{cases} S_{\mathrm{L,R}} & , S_{\mathrm{L,R}} \geq S_{\mathrm{L,T}} \\ S_{\mathrm{L,T}} & , S_{\mathrm{L,T}} > S_{\mathrm{L,R}} \end{cases} \tag{38}$$

and

$$S_{\mathrm{L,RT,Mean}} = \frac{1}{2}(S_{\mathrm{L,R}} + S_{\mathrm{L,T}}) \tag{39}$$

by using the maximum and the arithmetic mean between of $S_{\mathrm{L,R}}$ and $S_{\mathrm{L,T}}$ as aggregated similarity measures.

2) Pairwise Pattern Similarity Measure: To derive a pairwise similarity measure, we compute the number of notes $N_{1,2}$ in pattern $\boldsymbol{\mathcal{P}}_1$, for which at least one note in pattern $\boldsymbol{\mathcal{P}}_2$ exists that has the same absolute pitch value $\boldsymbol{\theta}_{\mathrm{P}}$ (for the similarity measure $S_{\mathrm{P,T}}$) or onset value $\varphi_{\mathrm{O}}^{[\mathrm{m}]}$ (for the similarity measure $S_{\mathrm{P,R}}$). We derive $N_{2,1}$ vice versa for pattern $\boldsymbol{\mathcal{P}}_2$. The *pairwise similarity measure* is computed as

$$S_{\mathrm{P}} = \frac{1}{2}\left(\frac{N_{1,2}}{N_1} + \frac{N_{2,1}}{N_2}\right) \tag{40}$$

In addition to the similarity measures $S_{P,R}$ and $S_{P,T}$, we compute three similarity measures that incorporate both the rhythmic and tonal similarity. Therefore, we use the constraint that both the onset value $\varphi_{\mathrm{O}}^{[\mathrm{m}]}$ and the absolute pitch value $\boldsymbol{\theta}_{\mathrm{P}}$ have to be equal in Eq. (40) to obtain the measure $S_{\mathrm{P,RT}}$. In addition, we compute $S_{\mathrm{P,RT,Max}}$ and $S_{\mathrm{P,RT,Mean}}$ from $S_{P,R}$ and $S_{P,T}$ analogous to Eq. (38) and Eq. (39).

3) Aggregation Strategies: In addition to the two pattern similarity approaches explained in the previous subsections, we compare two different strategies to aggregate the similarity results. We use cross-validation scenarios for the classification experiments. All folds contain the same number of basslines and the same percentage of basslines from all genres.

In each fold, we denote the set of patterns in the training set by P_{t}. According to their genre labels $g \in [1, 13]$ (see Sec. V-A), all patterns in P_{t} can be further subdivided into sub-sets $P_{\mathrm{t,g}}$ of patterns of each genre g.

a) Pattern-wise Classification: In order to classify the genre of an unknown bassline $\boldsymbol{\mathcal{P}}$, we compute a likelihood value l_{g} for each genre g. Therefore we seek the most similar pattern in each sub-set $P_{\mathrm{t,g}}$ in comparison to $\boldsymbol{\mathcal{P}}$. The corresponding similarity measures are taken as *genre likelihood* values

$$l_{\mathrm{g}} = S_{\mathrm{max,g}}. \tag{41}$$

$S_{\mathrm{max,g}}$ denotes the highest similarity detected between $\boldsymbol{\mathcal{P}}$ and all patterns in $P_{\mathrm{t,g}}$ assigned to genre g. The genre that maximizes l_{g} is the classified genre:

$$\hat{g} = \arg\max_{g \in [1,13]} l_{\mathrm{g}}. \tag{42}$$

This approach is illustrated in Fig. 4. In case two or more genres are associated with the same likelihood-value, the classified genre is randomly selected from the most likely candidates.

In the dataset we used in this publication, the basslines have different lengths between 4 and 16 measures. In order to compute the similarity between two patterns of arbitrary length, we proceed as follows. First, if necessary, all patterns are elongated by adding their first measures to achieve that all pattern lengths are powers of two. If one pattern is shorter than the other, we shift it with a step-size of two measures and compute a similarity measure between the shorter pattern and the corresponding sub-pattern of the longer pattern at the current position. Finally, we average all similarity values to compute the similarity between both patterns. This is illustrated for an example of two patterns of 4 and 7 measures length in Fig. 5.

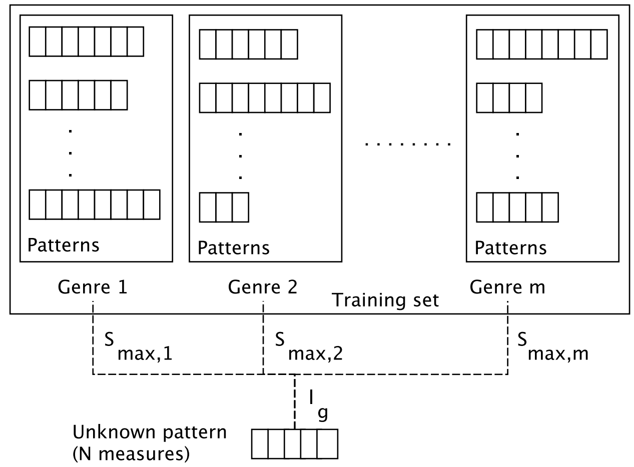


Fig. 4. Pattern-wise classification strategy. For each genre m, highest similarity $S_{\mathrm{max,m}}$ between an unknown pattern $\boldsymbol{\mathcal{P}}$ and all patterns assigned to the genre m is computed. These similarity values are used as likelihood values $\boldsymbol{l}_{\mathrm{g}}$ for the genre classification.

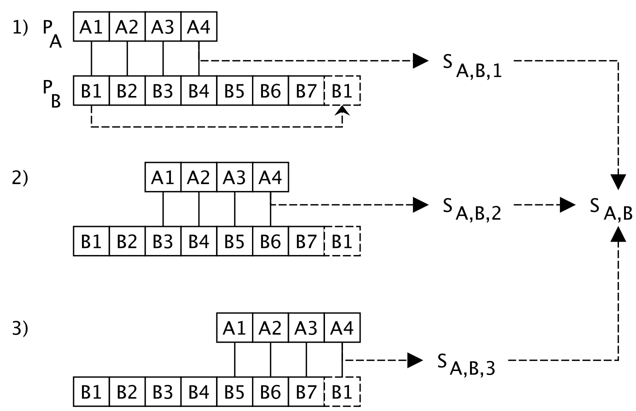


Fig. 5. Computation of pattern-wise similarity between two patterns $\boldsymbol{\mathcal{P}}_A$ and $\boldsymbol{\mathcal{P}}_B$ of arbitrary length. In this example, the patterns have the length of 4 and 7 measures. The longer pattern ($\boldsymbol{\mathcal{P}}_B$) is elongated to a length of $l(\boldsymbol{\mathcal{P}}_B) = 8$ measures by adding its first measure. To compute the similarity between $\boldsymbol{\mathcal{P}}_A$ and $\boldsymbol{\mathcal{P}}_B$, the shorter pattern ($\boldsymbol{\mathcal{P}}_A$) is shifted with a hop-size of 2 measures. For each shift, the similarity between $\boldsymbol{\mathcal{P}}_A$ and the corresponding sub-pattern of $\boldsymbol{\mathcal{P}}_B$ is computed. Finally, all similarity values are averaged.

b) Measure-wise Classification: Each pattern $\boldsymbol{\mathcal{P}} \in P_{\mathrm{t}}$ can be subdivided into sub-patterns of one measure length. The set of all sub-patterns in the training set is denoted as S_{t}. As described in the previous subsection, the set S_{t} can be subdivided into sub-sets $S_{\mathrm{t,g}}$ according to the genre labels of the corresponding sub-patterns.

To classify an unknown pattern $\boldsymbol{\mathcal{P}}$, we subdivide it into $N = |\boldsymbol{\mathcal{P}}|$ sub-patterns $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}n}$ ($n \in [1, N]$). For each sub-pattern $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}n}$, we seek the most similar sub-pattern from the training set for each genre. The genre likelihood $l_{\mathrm{n,g}}$ for each sub-pattern and each genre is computed as

$$l_{\mathrm{n,g}} = S_{\mathrm{max,n,g}}. \tag{43}$$

In order to obtain a genre likelihood value that corresponds

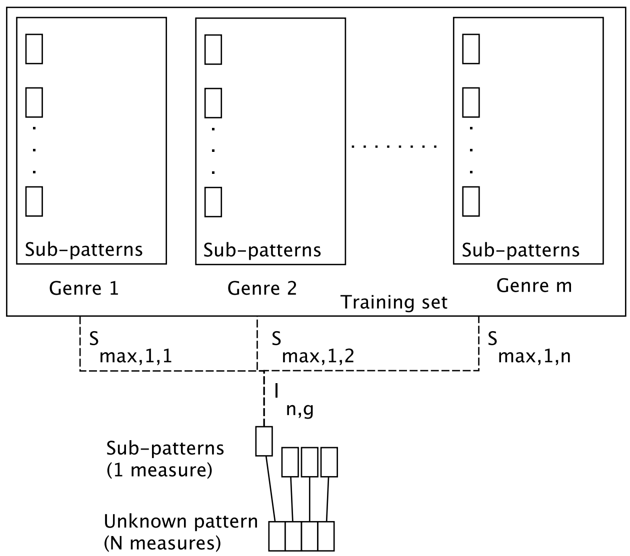


Fig. 6. Measure-wise classification strategy. Each unknown pattern $\mathcal{P}$ is divided into sub-patterns of one measure length. For the each sub-pattern $\mathcal{P}_{\mathcal{S}_n}$, the highest similarity $S_{\mathrm{max,n,m}}$ is detected for each genre m. Again, these similarity values are used as likelihood values $l_{\mathrm{n,g}}$ for the genres.

to the whole pattern $\mathcal{P}$, we average the likelihood values over all sub-patterns for each genre as

$$l_{\mathrm{g}} = \frac{1}{N}\sum_{n=1}^{N} l_{\mathrm{n,g}}. \tag{44}$$

Finally, the genre classification is performed as explained in Eq. (42) in the previous subsection. This approach is depicted in Fig. 6.

V. Evaluation

A. Data Set

The data set used in this publication is an extended version of the data set used in [Abeßer et al., 2010a]. The former collection comprised basslines from 8 different genres that have been entirely taken from instructional bass literature [Westwood, 1997] [Reznicek, 2001]. This collection has been extensively revised by musicologists. All genre-sets were modified and five new sets were added. Table V provides an overview of all genres used in this work as well as their regional origin and approximate beginning period.

Since instructional literature is not available for all of the genres, a lot of the basslines have been manually transcribed from real audio recordings by a semi-professional bass player. Musicologists selected the songs and the segments to be transcribed and checked the transcriptions afterwards. These basslines enrich the ones taken from instructional literature, which is not available for all of the genres.

The current set consists of 520 basslines with 40 basslines for each of the 13 genres listed in Tab. V. The tempo distribution over all genres is depicted in Fig. 7.

B. Experiments & Results

1) Classification Based on Pattern Similarity: We used 10-fold cross validation to compute the mean classification

TABLE V
INVESTIGATED GENRES.

g	Abbr.	Genre	Origin	First recordings
1	BLU	blues	USA	1912 (transcription)
2	BOS	bossa nova	Brasil	1958
3	FOR	forró	Brasil	1900 (transcription)
4	FUN	funk	USA	1960s
5	HIP	hip-hop	USA	1970s
6	MIN	minimal techno	USA, Germany	1994
7	MOT	motown	USA	1960
8	REG	reggae	Jamaica	1960s
9	RON	nineties rock	USA, GB	1990s
10	ROS	seventies rock	USA, GB	1970s
11	SAL	salsa & mambo	Cuba	1930s
12	SWI	swing	USA	1920s
13	ZOU	zouglou	Cote d'Ivoire	1995

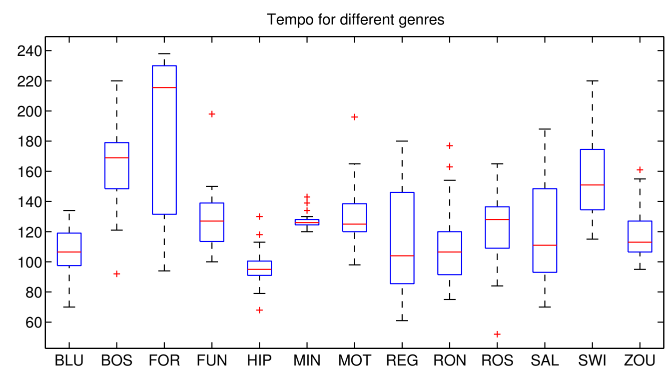


Fig. 7. Boxplots of tempo values for all genres. All tempo values are given in beats per minute (BPM). Genre abbreviations are explained in Tab. V.

accuracy for all similarity measures and both classification approaches introduced in Sec. IV-C. As illustrated in Tab. VI, the pattern-wise classification strategy clearly outperforms the measure-wise strategy for all investigated similarity measures. Apparently, sub-patterns of one measure length do not contain sufficient stylistic information and thus can not represent a genre as well as patterns of a length of four or more measures.

A second finding is that the combined similarity measures that take both the tonal and rhythmic similarity into account performed better than the similarity measures related to only one of both similarities. The highest mean accuracy of 38.9% was achieved using the pattern-wise classification and the similarity measure $S_{P,RT,Mean}$ based on the mean rhythmic and tonal pairwise-similarity between patterns.

Third, we consistently found a large variance of the classification accuracy results over all 13 genres. For instance, the upper matrix in Tab. VIII depicts all genre confusions in percent for the best configuration (similarity measure $S_{P,RT,Mean}$, pattern-wise classification). The mean accuracy values strongly vary from 97.5% for swing down to 7.5% for seventies rock.

The classification approach based on pattern similarity only provides satisfying results over 60% accuracy for the genres swing, blues, and bossa nova. We assume that for those genres, prototypic patterns exist that are often used with small rhythmic and tonal variation. The presented classification approach strongly depends on this assumption. The basslines

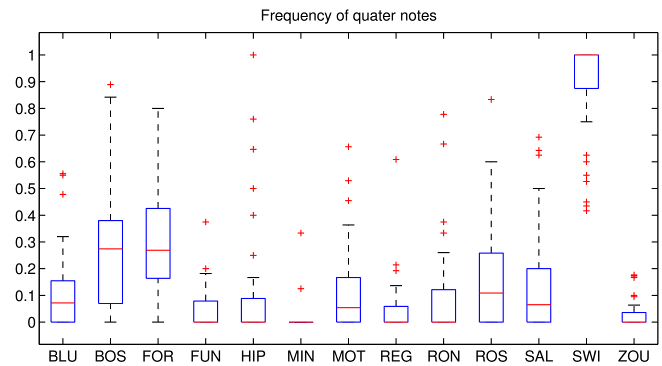


Fig. 8. Boxplots of frequency of crotchets $p(\varphi_D^{[m]} = 1/4)$ for all genres. Genre abbreviations are explained in Tab. V.

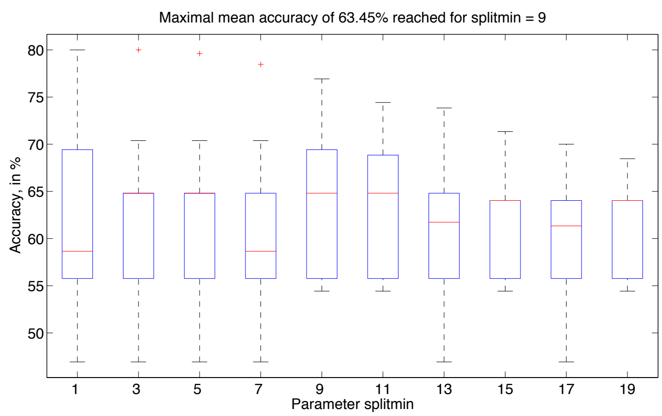


Fig. 9. Box plots of accuracy values achieved by pruned decision tree with various settings for the minimum number of items per node in CART

associated to other genres seem to be too diverse in terms of rhythmic and tonal similarity in order to be classified appropriately.

2) Classification Based on Rules: The evaluation was performed according to the following scenario. First we experimentally chose the optimal minimum number of items per node for CART. This parameter was varied from 1 to 19 in a step of two. For each parameter value the experiment was repeated 100 times. The box plots of accuracy values are presented in Fig. 9. The best results were achieved by setting minimum number of items per node to 9. The generalization properties of the decision tree were controlled in a 52-fold cross validation scenario, where the tree was pruned to an optimal level as proposed in [Breiman et al., 1993] in order to prevent overfitting to the train data.

We assume this classifier to be slightly overtrained, as the final results are achieved without clear separation of the training and test data. This scenario was chosen to allow better interpretation of one inducted set of rules instead of multiple sets of rules for various validation folds. All in all the final pruned decision tree is composed of 57 nodes and make use of 14 features. Some highly discriminative features – e.g. tempo in beats per minute (see Fig. 7) – appear multiple times in the decision nodes.

TABLE VI
CLASSIFICATION BASED ON PATTERN SIMILARITY - MEAN CLASSIFICATION ACCURACY (MN), STANDARD DEVIATION (SD), LOWEST GENRE ACCURACY (MIN), AND HIGHEST GENRE ACCURACY (MAX). ALL VALUES GIVEN IN PERCENT.

Similarity measure	Classification strategy							
	Pattern-wise				Measure-wise			
	MN	SD	MIN	MAX	MN	SD	MIN	MAX
$S_{P,R}$	33.5	22.9	12.4	93.3	16.1	19.7	0	75.3
$S_{P,T}$	21.7	8.5	7.9	35.4	7.8	12.5	0	44.5
$S_{P,RT}$	29.9	17.6	5.4	57.5	12	21	0	75.8
$S_{P,RT,Max}$	38.6	23.6	10	97.5	12.1	23.1	0	85
$S_{P,RT,Mean}$	**38.9**	23.9	7.5	97.5	14.4	25.5	0	92.5
$S_{L,R}$	33.5	22.3	10.7	93.3	16.3	18.9	1.7	73.3
$S_{L,T}$	26	15.5	3.8	51	9	12.6	0	40
$S_{L,RT,Max}$	35.8	22	13.2	93.3	15.7	17.9	1	67.8
$S_{L,RT,Mean}$	37.9	24.6	7.5	97.5	13.1	17.2	0	58.3

Classification with a pruned decision tree achieves a mean accuracy of 64.8%. Best accuracy of 92.5% is obtained for the genre swing. In comparison to the pattern similarity approach, there are no classes with the classification accuracy close to a random one. The poorest performance of 35% can be observed for Ninties Rock genre, which is often confused with blues and reggae.

3) Classification Based on Statistical Pattern Recognition: The evaluation of SVM classification was performed in a 52-fold cross validation scenario, i.e. each fold 10 items were excluded from training and used solely as test material. Additional cross validation was applied to the training data of each fold in order to estimate best kernel and regulation parameters for the model. As SVM is a binary classifier, one-against-one method with subsequent voting was utilized to enable multi-class classification.

The results of the SVM classification are presented in Table VIII. The mean accuracy over all classes comprises 55.4%. Same as for the pruned decision tree, the best result of 92.5% is achieved for genre swing. The lowest genre accuracy can be observed for motown due to the multiple confusions with blues, funk and seventies rock. More details on the confusions and their nature can be found in the next section.

VI. DISCUSSIONS & CONCLUSIONS

In this paper, we compared three different classification approaches for genre classification based on repetitive basslines. We found that the rule-based approach achieved a mean accuracy of 64.8% and thus outperformed classification based on pattern similarity and statistic pattern recognition by 25.9% and 9.4%. The random baseline accuracy for 13 classes is approx. 7.7%.

The field of genre in music is complex and ambiguous. The data set chosen for this paper is not designed to simplify this reality but is meant to reflect it. From a musicological point of view, this paper confirms and specifies known findings: low classification rates strongly correspond with what the authors would like to call *eclectic styles*, i.e. music styles, which by their own history and logic are a camouflage of older styles.

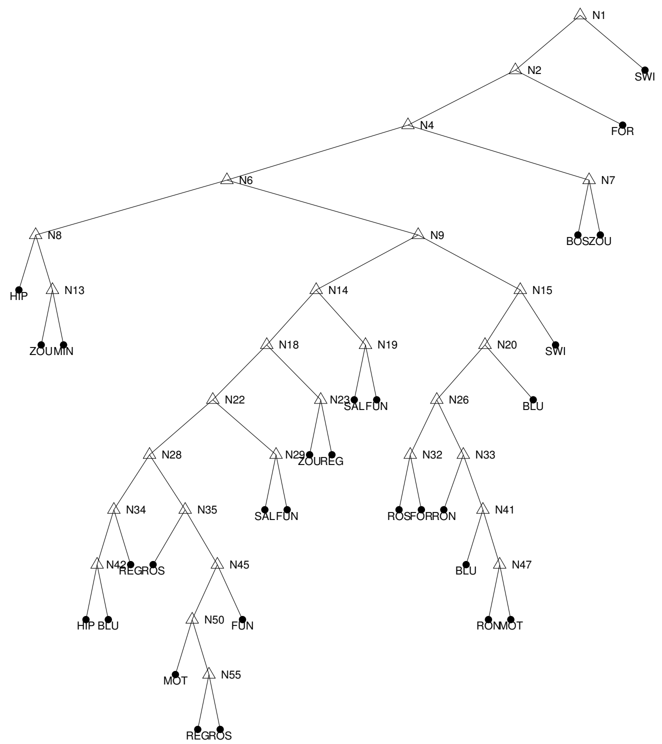


Fig. 10. Pruned decision tree, see Tab. VII for details.

TABLE VII
DECISION TREE RULES FOR FIG. 10.

Node	**IF** condition	Yes	No
N 1	p($\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]} = 1/4$) (crotchets) < 0.944	N 2	N 3
N 2	Tempo (bpm) < 202.5	N 4	N 5
N 3	genre = swing		
N 4	p($\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]} = 1/2$) (half notes) < 0.027	N 6	N 7
N 5	genre = forró		
N 6	α_{SL} (See Eq. (32)) < 0.673	N 8	N 9
N 7	Variance over number of notes per second < 1.119	N 10	N 11
N 8	Tempo (bpm) < 105.5	N 12	N 13
N 9	α_{SL} (See Eq. (32)) < 0.946	N 14	N 15
N 10	genre = bossa nova		
N 11	genre = zouglou		
N 12	genre = hip-hop		
N 13	First-order entropy over $\boldsymbol{\Delta\theta}_{\mathrm{P,D}} < -0.470$	N 16	N 17
N 14	p($\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]} = 3/8$) (dotted quavers) < 0.061	N 18	N 19
N 15	Variance over $\alpha_{\mathrm{On},32,\mathrm{m}}$(See Eq. (21)) < 0.014	N 20	N 21
N 16	genre = zouglou		
N 17	genre = minimal techno		
N 18	p(pentatonic minor scale) (see Tab IV) < 0.165	N 22	N 23
N 19	p($\boldsymbol{\Delta\theta}_{\mathrm{P,F}} = 2$) (ascending second) < 0.127	N 24	N 25
N 20	p($\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]} = 1/12$) (triplet quavers) < 0.383	N 26	N 27
N 21	genre = swing		
N 22	Mean absolute intervall size < 3.175	N 28	N 29
N 23	p($\lvert\boldsymbol{\Delta\theta}_{\mathrm{P,F}}\rvert \in \{1,2,3\}$) - p(Primes, seconds, thirds) < 0.442	N 30	N 31
N 24	genre = salsa & mambo		
N 25	genre = funk		
N 26	N_N/t_{total} (number of notes / length (s)) < 2.656	N 32	N 33
N 27	genre = blues		
N 28	Tempo (bpm) < 108.5	N 34	N 35
N 29	p(whole tone half tone scale) (see Tab IV) < 0.093	N 36	N 37
N 30	genre = zouglou		
N 31	genre = reggae		
N 32	Scale index s of most likely scale < 2.5 (See Tab. IV)	N 38	N 39
N 33	Mean over $\alpha_{\mathrm{On},2,\mathrm{m}}$(See Eq. (21)) < 0.563	N 40	N 41
N 34	p($\boldsymbol{\varphi}_{\mathrm{D}}^{[\mathrm{m}]} = 1/16$) (semi-quavers) < 0.617	N 42	N 43
N 35	p($\Delta\theta_{\mathrm{P,k}} = 0$) < 0.161 (measure of constant pitch)	N 44	N 45
N 36	genre = salsa & mambo		
N 37	genre = funk		
N 38	genre = seventies rock		
N 39	genre = forró		
N 40	genre = nineties rock		
N 41	Tempo (bpm) < 113.5	N 46	N 47
N 42	Mean number of notes per second < 2.375	N 48	N 49
N 43	genre = reggae		
N 44	genre = seventies rock		
N 45	p(chromatic note sequence) < 0.218	N 50	N 51
N 46	genre = blues		
N 47	p(minor gypsy scale) (see Tab IV) < 0.097	N 52	N 53
N 48	genre = hip-hop		
N 49	genre = blues		
N 50	Variance over $\alpha_{\mathrm{On},4,\mathrm{m}} < 0.007$	N 54	N 55
N 51	genre = funk		
N 52	genre = nineties rock		
N 53	genre = motown		
N 54	genre = motown		
N 55	p($\boldsymbol{\varphi}_{\mathrm{DR}} > 1$) (see Eq. (7)) < 0.123	N 56	N 57
N 56	genre = reggae		
N 57	genre = seventies rock		

In bossa nova (BOS) for example, Jazz and Latin styles – especially samba – are blended. However, the two styles are still recognizable. The high rate of confusions of bossa nova with other Latin styles therefore is a meaningful one.

Even stronger so in the case of hip-hop (HIP). Originally, its music is comprised of samples from tracks of other styles. This idea of a non-altering collage of elements that genuinely derive from other styles (funk, soul, jazz or rock) even persists, when the music of a hip-hop track is self made (i.e. played by a hip-hop combo instead of sampled from other recordings).

Some genres in the data set therefore are quite similar (e.g. seventies rock and nineties rock, salsa and bossa nova). The observed genre confusions in this paper as in all serious musical classification therefore must be separated into *corresponding confusions* and *non-corresponding confusions*.

As regards the non-corresponding confusions, the classification is simply mistaken. Some confusions on the other hand correspond to, i.e. reflect, ambiguities in the musical reality. For example basslines of blues music have been mistaken as rock (ROS, RON) or motown (MOT), but almost never as one of the four different Latin or Afro-Carribean music styles (SAL, ZOU, BOS, FOR). Other than mistakes or misclassification, these highly corresponding confusions are valuable information in a lot of use cases such as musicological research on genre similarity or distinguished classification of fusion styles for online shops and distributors.

Adding to the promising results, many of the genre-wise confusions are corresponding confusions (see above). Most of them are *intra-regional*. blues is confused with rock and motown. The two rock styles are confused with one another as are the urban electronic music styles of hip-hop and minimal techno. motown, as an overlap genre, also belongs to another stylistic region or cluster – self-consciously afro-american or african music styles motown, reggae, salsa and hip-hop as well as African zouglou are confused with each other significantly.

Reggae does stand in the middle of this stylistic map: corresponding to the history of this music style it is confused with

TABLE VIII
CONFUSION MATRICES - ALL VALUES GIVEN IN PERCENT.

Classification (pattern similarity - best configuration): $\mu = \mathbf{38.9}\%$

	BLU	BOS	FOR	FUN	HIP	MIN	MOT	REG	RON	ROS	SAL	SWI	ZOU
BLU	**60**	2.5	0	0	2.5	5	7.5	2.5	10	2.5	0	7.5	0
BOS	5	**67.5**	15	0	0	0	2.5	0	2.5	7.5	0	0	0
FOR	5	12.5	**40**	2.5	2.5	12.5	5	0	5	2.5	2.5	10	0
FUN	25	0	2.5	**32.5**	2.5	2.5	5	2.5	10	5	7.5	0	5
HIP	12.5	5	5	12.5	**15**	12.5	7.5	10	12.5	0	0	2.5	5
MIN	6.3	0	1.3	7.5	10	**36.3**	6.3	5	17.5	2.5	0	5	2.5
MOT	26.7	0	0	5	2.5	0	**27.5**	6.7	8.3	18.3	2.5	0	2.5
REG	12.5	0	5	2.5	7.5	2.5	17.5	**25**	5	7.5	2.5	2.5	10
RON	27.5	4.2	0	1.7	2.5	5	2.5	5	**24.2**	20	2.5	5	0
ROS	5	7.5	5	7.5	0	7.5	22.5	10	17.5	**7.5**	0	7.5	2.5
SAL	10	0	4.2	12.5	2.5	6.7	10	2.5	2.5	7.5	**36.7**	0	5
SWI	2.5	0	0	0	0	0	0	0	0	0	0	**97.5**	0
ZOU	5	0	2.5	6.3	11.3	8.8	5	5	15	0	5	0	**36.3**

Classification (pruned tree): $\mu = \mathbf{64.8}\%$

	BLU	BOS	FOR	FUN	HIP	MIN	MOT	REG	RON	ROS	SAL	SWI	ZOU
BLU	**70**	0	0	5	0	0	10	7.5	5	0	2.5	0	0
BOS	0	**67.5**	12.5	0	0	5	2.5	0	0	2.5	0	0	10
FOR	2.5	10	**77.5**	0	0	0	2.5	0	0	0	5	0	2.5
FUN	2.5	0	0	**77.5**	0	2.5	10	0	0	2.5	2.5	0	2.5
HIP	7.5	2.5	0	5	**50**	0	0	10	0	2.5	12.5	2.5	7.5
MIN	5	0	0	0	0	**60**	7.5	0	2.5	5	10	0	10
MOT	5	0	0	0	0	0	**72.5**	2.5	2.5	10	2.5	0	5
REG	10	2.5	0	2.5	2.5	0	2.5	**55**	0	0	7.5	0	17.5
RON	17.5	7.5	0	0	0	0	5	17.5	**35**	5	10	0	2.5
ROS	10	0	0	5	0	0	5	7.5	12.5	**45**	10	0	5
SAL	5	0	0	5	0	0	0	2.5	2.5	2.5	**65**	0	17.5
SWI	0	0	0	0	0	0	2.5	0	2.5	0	2.5	**92.5**	0
ZOU	2.5	0	0	2.5	2.5	0	5	2.5	0	5	5	0	**75**

Classification (SVM): $\mu = \mathbf{55.4}\%$

	BLU	BOS	FOR	FUN	HIP	MIN	MOT	REG	RON	ROS	SAL	SWI	ZOU
BLU	**60**	0	0	2.5	2.5	2.5	10	2.5	7.5	7.5	0	5	0
BOS	0	**77.5**	17.5	0	0	0	0	0	0	2.5	2.5	0	0
FOR	0	25	**57.5**	0	0	5	5	0	0	2.5	2.5	0	2.5
FUN	5	0	0	**67.5**	2.5	0	5	7.5	7.5	2.5	2.5	0	0
HIP	2.5	0	2.5	0	**45**	17.5	5	7.5	0	10	7.5	0	2.5
MIN	0	0	2.5	0	12.5	**67.5**	0	2.5	5	5	5	0	0
MOT	17.5	0	2.5	12.5	2.5	7.5	**20**	2.5	7.5	17.5	5	0	5
REG	7.5	2.5	0	7.5	10	0	7.5	**40**	10	2.5	0	0	12.5
RON	5	2.5	2.5	12.5	0	10	12.5	10	**37.5**	5	0	2.5	0
ROS	10	2.5	0	5	10	2.5	15	10	7.5	**27.5**	2.5	2.5	5
SAL	0	2.5	5	2.5	12.5	2.5	2.5	2.5	5	0	**55**	0	10
SWI	2.5	0	0	0	0	2.5	0	0	0	2.5	0	**92.5**	0
ZOU	0	0	0	0	2.5	7.5	0	7.5	0	2.5	7.5	0	**72.5**

the urban hip-hop, African zouglou and American motown. Zouglou on the other hand is further at the edge of this cluster, being confused only with salsa and reggae, which corresponds to zouglou being intertwined with Carribean styles that also influenced reggae and salsa but not so much with hip-hop, funk and soul. The confusions between seventies rock and zouglou in the statistical pattern recognition approach on the other hand simply are mistakes.

As the figures above show, SVM classification is in a lot of cases but not always worse than the pruned-tree approach. Regarding the harmonically more complex genre of bossa nova SVM approach even beats the pruned tree by 10% of accuracy. While the pruned tree approach has the best classification rates, it is worst regarding the rate of non-corresponding confusions. Remarkably, the most information of inter-stylistic musical relations could be drawn out of the pattern similarity.

VII. Future Directions

It is not surprising, that the observation of only one musical partition (i.e. the bassline) does not overrule more general approaches. However in several cases the bassline is a very successful discriminator. In combination with a corresponding rule system, the approach will produce valuable information for genre classification.

Given the ambiguity of the field of music genres, the authors are satisfied to see, that most of the observed confusions correspond to the musical reality as e.g. Latin styles are confused with each other as are urban styles or the two rock genres. Additionally the results reflect the autonomy of a style, eclectic styles like hip-hop being more prone to confusions than swing.

The limited approach of using symbolic MIDI files does not allow a detailed investigation of micro-timing deviations, which usually have a smaller order of magnitude than the smallest rhythmic unit, say a demisemiquaver. Further aspects of performance such as rhythmic micro timing, dynamic progression in note sequences and the interaction between the bass player and the drums and the harmony instrument need to be incorporated to derive a complete view on musical performance.

The automatic description of musical style by basslines must not stand alone. We assume that it can be a very useful part of a multi-layered approach that incorporates a stylistic analysis of all instrument tracks of a song. The results of the three classification approaches indicate that a combined hybrid classification framework would benefit from the strengths of the individual components. This hybrid framework could for instance be comprised of a majority-based voting including all three classifiers.

Alternatively to classification algorithms one could apply clustering methods to group and explore musical items. Constrained clustering has been developed to improve clustering methods through pairwise constraints between items in the feature space [Mercado and Lukashevich, 2010]. Based on the results of our work, these constraints could be the relations between cultural regions or music styles. Applying these constraints could help to avoid non-corresponding confusions, see Sec. VI.

Acknowledgment

Removed for blind reviewing.

References

[Abeßer et al., 2008] Abeßer, J., Dittmar, C., and Großmann, H. (2008). Automatic genre and artist classification by analyzing improvised solo parts from musical recordings. In *Proceedings of the Audio Mostly Conf., Piteå, Sweden*.

[Abeßer et al., 2010a] Abeßer, J., Lukashevich, H., Bräuer, P., and Schuller, G. (2010a). Bass playing style detection based on high-level features and pattern similarity. In *Proc. of the Int. Society of Music Information Retrieval (ISMIR), Utrecht, Netherlands*.

[Abeßer et al., 2010b] Abeßer, J., Lukashevich, H., Dittmar, C., Bräuer, P., and Krause, F. (2010b). Rule-based classification of musical genres from a global cultural background. In *Proc. of the Int. Symposium on Computer Music Modeling and Retrieval, Málaga, Spain*.

[Abeßer et al., 2009] Abeßer, J., Lukashevich, H., Dittmar, C., and Schuller, G. (2009). Genre classification using bass-related high-level features and playing styles. In *Proc. of the Int. Society of Music Information Retrieval (ISMIR Conf.), Kobe, Japan*.

[Anglade et al., 2009] Anglade, A., Ramirez, R., and Dixon, S. (2009). Genre classification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the 10th Conf. on Music Information Retrieval (ISMIR), Kobe, Japan*, pages 669–674.

[Breiman et al., 1993] Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1993). *Classification and Regression Trees*. Chapman & Hall.

[Bresin, 2001] Bresin, R. (2001). Articulation rules for automatic music performance. In *Proc. of the Int. Computer Music Conf. (ICMC), Havana, Cuba*, pages 294–297.

[Bruhn et al., 2005] Bruhn, H., Oerter, R., and Rösing, H., editors (2005). *Musik-Psychologie - Ein Handbuch*. Reinbek: Rowohlt.

[Buzzanca, 2001] Buzzanca, G. (2001). A rule-based expert system for music style recognition. In *Proceedings of the 1st Int. Conf. on Understanding and Creating Music (UCM), Caserta, Italy*.

[de Léon and Iñesta, 2007] de Léon, P. J. P. and Iñesta, J. M. (2007). Pattern recognition approach for music style identification using shallow statistical descriptors. *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews*, 37(2):248–257.

[de León et al., 2007] de León, P. J. P., Rizo, D., and Iñesta, J. M. (2007). Towards a human-friendly melody characterization by automatically induced rules. In *Proceedings of the 8th Int. Conf. on Music Information Retrieval (ISMIR), Vienna, Austria*.

[Dittmar et al., 2007] Dittmar, C., Dressler, K., and Rosenbauer, K. (2007). A toolbox for automatic transcription of polyphonic music. In *Proceedings of the Audio Mostly Conf., Ilmenau, Germany*.

[Eerola and Toiviainen, 2004] Eerola, T. and Toiviainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä, Jyväskylä, Finland.

[Flanagan, 2008] Flanagan, P. (2008). Quantifying metrical ambiguity. In *Proceedings of the Int. Conf. on Music Information Retrieval (ISMIR), Philadelphia, USA*, pages 635–640.

[Friberg and Sundström, 2002] Friberg, A. and Sundström, A. (2002). Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern. *Music Perception*, 19(3):333–349.

[Goto, 2004] Goto, M. (2004). A real-time music-scene-description system: predominant-f0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43:311–329.

[Gouyon, 2005] Gouyon, F. (2005). *A Computational Approach to Rhythm Description - Audio Features for the Computation of Rhythm Periodicity Functions and their Use in Tempo Induction and Music Content Processing*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.

[Grachten et al., 2004] Grachten, M., Arcos, J.-L., and de Mantaras, R. L. (2004). Melodic similarity: Looking for a good abstraction level. In *Proceedings of the 5th Int. Conf. in Music Information Retrieval (ISMIR), Barcelona, Spain*.

[Gusfield, 1997] Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, Cambridge, UK.

[Hainsworth and Macleod, 2001] Hainsworth, S. W. and Macleod, M. D. (2001). Automatic bass line transcription from polyphonic audio. In *Proceedings of the Int. Computer Music Conf. (ICMC), La Habana, Cuba*.

[Madsen and Widmer, 2007] Madsen, S. T. and Widmer, G. (2007). A complexity-based approach to melody track identification in midi files. In *Proceedings of the Int. Workshop on Artificial Intelligence and Music (MUSIC-AI)*.

[McKay and Fujinaga, 2004] McKay, C. and Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. In *Proceedings of the 5th Int. Conf. in Music Information Retrieval (ISMIR), Barcelona, Spain*, pages 525–530.

[McKay and Fujinaga, 2006] McKay, C. and Fujinaga, I. (2006). jSymbolic: A feature extractor for MIDI files. In *Int. Computer Music Conf. (ICMC), New Orleans, USA*, pages 302–305.

[Mercado and Lukashevich, 2010] Mercado, P. and Lukashevich, H. (2010). Feature selection in clustering with constraints: Application to active exploration of music collections. In *Proceedings of the International Conf. on Machine Learning and Applications, Washington, USA*.

[Müllensiefen and Frieler, 2004] Müllensiefen, D. and Frieler, K. (2004). Optimizing measures of melodic similarity for the exploration of a large folk song database. In *Proceedings of the 5th Int. Conf. in Music Information Retrieval (ISMIR), Barcelona, Spain*.

[Orio and Rodá, 2009] Orio, N. and Rodá, A. (2009). A measure of melodic similarity based on a graph representation of the music structure. In *Proc. of the 10th International Society for Music Information Retrieval Conference (ISMIR)*.

[Pérez-Sancho et al., 2006] Pérez-Sancho, C., de León, P. J. P., and Iñesta, J. M. (2006). A comparison of statistical approaches to symbolic genre recognition. In *Proceedings of the Int. Computer Music Conf. (ICMC), New Orleans, USA*, pages 545–550.

[Reznicek, 2001] Reznicek, H.-J. (2001). *I'm Walking - Jazz Bass*. AMA.

[Ryynänen and Klapuri, 2008] Ryynänen, M. P. and Klapuri, A. P. (2008). Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32:72–86.

[Stamatatos and Widmer, 2005] Stamatatos, E. and Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence*, 165:37–56.

[Tsuchihashi et al., 2008] Tsuchihashi, Y., Kitahara, T., and Katayose, H. (2008). Using bass-line features for content-based MIR. In *Proceedings of the Int. Conf. on Music Information Retrieval (ISMIR), Philadelphia, USA*, pages 620–625.

[Tsunoo et al., 2009] Tsunoo, E., Ono, N., and Sagayama, S. (2009). Musical bass-line pattern clustering and its application to audio genre classification. In *Proc. of the Int. Society of Music Information Retrieval (ISMIR), Kobe, Japan*.

[Typke et al., 2002] Typke, R., Giannopoulos, P., Veltkamp, R. C., Wiering, F., and van Oostrum, R. (2002). Using Transportation Distances for Measuring Melodic Similarity. In *Proceedings of the 3rd Int. Conf. on Music Information Retrieval (ISMIR), Paris, France*.

[Unal et al., 2008] Unal, E., Chew, E., Georgiou, P. G., and Narayanan, S. S. (2008). Challenging uncertainty in query by humming systems: A fingerprinting approach. *IEEE Transactions on Audi, Speech, and Language Processing*, 16:359–371.

[Vapnik, 1998] Vapnik, V. N. (1998). *Statistical learning theory*. Wiley New York.

[Wagner and Fischer, 1974] Wagner, R. A. and Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM (JACM)*, 21(1):168–173.

[Westwood, 1997] Westwood, P. (1997). *Bass Bible*. AMA.

[Wicke, 2007] Wicke, P. (2007). *Handbuch der populären Musik: Geschichte, Stile, Praxis, Industrie*. Schott, Mainz.

[Widmer et al., 2003] Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24:111–130.

[Widmer and Goebl, 2004] Widmer, G. and Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3):203–216.