GENRE CLASSIFICATION USING BASS-RELATED HIGH-LEVEL FEATURES AND PLAYING STYLES

Jakob Abeßer, Hanna Lukashevich, Christian Dittmar, Gerald Schuller
Fraunhofer IDMT, {abr, lkh, dmr, shl}@idmt.fraunhofer.de

ABSTRACT

Considering its mediation role between the poles of rhythm, harmony, and melody, the bass plays a crucial role in most music genres. This paper introduces a novel set of transcription-based high-level features that characterize the bass and its interaction with other participating instruments. Furthermore, a new method to model and automatically retrieve different genre-specific bass playing styles is presented. A genre classification task is used as benchmark to compare common machine learning algorithms based on the presented high-level features with a classification algorithm solely based on detected bass playing styles.

1. INTRODUCTION

After prolonged series of publications focusing on low- and mid-level features, many works within the MIR community nowadays emphasize the importance of musical high-level features. Their application is expected to significantly increase the precision in automatic music classification and similarity search tasks that have limits using conventional modeling paradigms [2]. Various automatic transcription techniques allow the extraction of score parameters like note pitch, velocity (volume), onset time and duration from polyphonic mixtures. These parameters embody the prior foundation for a subsequent feature extraction. Due to their close relation to musicological expressions, high-level features can be easily understood by musicologists. Thus, they offer a promising opportunity to translate existing musicological knowledge into automatically retrievable properties of analyzed music.
The remainder of this paper is organized as follows. In Sec. 2, we illustrate the goals of this publication and give an overview over related work in the subsequent section. We present both novel transcription-based high-level features and a new framework to model concepts and classes for the purpose of music classification in Sec. 4. Evaluation results from different scenarios are presented and discussed in Sec. 5 and a final conclusion is given in the last section.

2. GOALS & CHALLENGES

Our goal is to design transcription-based high-level features that enable a better characterization of the bass track in different songs. Furthermore, we aim to develop a general method to translate musicological knowledge into rules on feature values that can be easily evaluated. This approach is intended to facilitate the design of an instrument-related classifier that is trained by musicological knowledge – similar to an expert system. When analyzing real audio data, the strong dependence of a well-performing transcription system still remains the biggest challenge.

3. PREVIOUS APPROACHES

Various bass transcription algorithms have been proposed so far in [13], [11], [6], and [18]. They extract the score parameters of a bass track in polyphonic audio recordings. Still, transcription errors related to pitch and onset values appear due to the high complexity of overlapping instrument spectra. These errors affect the accuracy of the deduced high-level features. As shown in [16], high-level features can be derived from different music domains like instrumentation, texture, rhythm, dynamics, pitch statistics, melody, and chords. Offering a direct access to the relevant score parameters, symbolic audio data like MIDI receives preferential treatment in many publications.
The authors of [4] applied several statistical methods to derive high-level features from note onsets, pitches, and intervals. The versatility of complexity-based descriptors based on entropy, compression, and prediction has been shown in [15]. A set of musical features derived from the bass part was introduced in [19]. The authors restricted themselves to pitch-related features and distinguished between features characterizing the pitch variability and the pitch motion. Rhythmical aspects like the swing or syncopations have been investigated in various publications as for instance in [12] and [9]. In [16], [4], [19], and [1], genre classification solely on high-level features was covered.

4. NEW APPROACHES

4.1 High-level features

High-level features allow to model and quantify musical properties that are directly observable by experienced musicologists. These are for instance the key, the time signature or measure of the harmonic consonance in a piece of music. They can be deduced from the pitch, the onset time, and the duration values of all notes.

Melody-related features

By analyzing the course of the *absolute pitch* p_A, we derive features from the incidence rate of typical pitch progressions, such as *notes with constant pitch* or *chromatic note sequences* related to the overall number of notes and the overall *pitch range* in halftones. With reference to the simultaneously sounding chords of the harmony track, we derive a feature from the ratio of *chord notes* within the bass line. Besides, we convert the absolute pitch of each note into its *functional pitch* $p_{A,F}$. It represents the interval type between each bass note and the root note of the simultaneously sounding chord. We consider all interval types from primes to sevenths ($p_{A,F} \in [1,7]$), bigger intervals are mapped into this interval range. The incidence rates of all possible values of $p_{A,F}$ are used as features that provide key-independent information about the frequency of occurrence of different interval types related to the harmony accompaniment.

The prior use of root notes, octaves, and fifths of the current chord within a bass line does not allow a conclusive differentiation between major and minor based chords by exclusively investigating the bass accompaniment. Thus, a measure of *harmonic ambiguity* is calculated proportional to the occurrence rate of primes and fifths and inversely proportional to the occurrence rate of thirds as $F_{HA} = P(p_{A,F} = 1) + P(p_{A,F} = 5) - P(p_{A,F} = 3)$.

We use a simple bar-wise distance measure combining rhythmic and melodic similarity to detect the *dominant bass pattern*. Therefore, we compute a square matrix D_τ containing the similarity between the notes in each pair of bars. We use $D_\tau(k,m) = 0.5[(1 - N_{k,m}/N_k) + (1 - N_{m,k}/N_m)]$ where N_i denotes the number of notes in bar i and $N_{i,j}$ denotes the number of notes of bar i that have a note equivalent in bar j with the same pitch (p_A) and onset [$\mathrm{mod}(\tau, 1)$]. We choose the notes of bar $n_{dom} = n$ that minimizes $\sum_i D_{i,n}$ as the dominant pattern since this bar has the lowest overall distance to the other bars. Subsequently, measures of *tonal* and *rhythmic variation* are derived from the mean distance between all bars to bar n_{dom}. For the rhythmical variation, only the aforementioned onset condition of the note equivalent is taken into account.

The interval progression of the bass line is characterized by three different representations, namely the relative pitch $p_R \in [-12, 12]$ (mapped down to a two octave range), the relative pitch mapped to functional intervals $p_{R,F} \in [-7, 7]$ (to provide a representation independent of the key-type as described above), and the interval direction $p_{R,D} \in [-1, 1]$. Subsequently, several statistical properties such as entropy and relative number of non-zero elements of the probabilities of all parameter values are extracted as features. The measures of *constant direction* F_{CD} & *dominant direction* F_{DD} furthermore quantify the temporal ratio of note passages with constant interval direction and characterize the dominant direction. Thus, they measure to what extend a melody appears to be fluent. We use $F_{CD} = N\left[p_{R,D}(i) \equiv p_{R,D}(i+1)\right]/N_{Intervals}$ and $F_{DD} = N(p_{R,I} = 1)/N_{Intervals}$.

Rhythm-related features

The *beat grid* contains the temporal positions and indices of all beats corresponding to the current time signature. After its extraction, all note onset t and duration values Δt are mapped from seconds to certain multiples of the corresponding bar lengths (resulting in τ and $\Delta\tau$). This allows a tempo-independent extraction of rhythm-related features. We applied a similar approach as described in [12] to derive the *swing ratio* related to the 8th- and the 16th-note grid.

A measure of *syncopation* related to the both aforementioned temporal grids is derived by retrieving binary patterns (like for instance “1001” in an 16th-note grid representing two notes whereas the first one is played on a downbeat and the other one on the adjacent off-beat related to the 8th-note grid).

Based on the *dominant bass pattern* and its dynamic progression, we take the number of bass notes within each bar with a velocity above 60% of the maximum occuring bass note velocity as the measure of *accent sparsity*. Percussionists often use the bass-drum to “double” the main accents of the bass line. We measure the ratio of the number of notes that both instruments played rhythmically *in unison* to the sum of all notes played by the bass and the bass drum individually.

Structure-related features

In addition, features characterizing repeating melodic and rhythmic segments are derived. Therefore, we apply a simple pattern search algorithm (*Correlative Matrix Approach* [14]) on character strings derived from the aforementioned score parameters p_A, τ, and $\Delta\tau$.

We use the statistical properties mean, median, standard deviation, minimum, and maximum from each of the pattern parameters length, incidence rate, and mean distance between similar patters as features. Overall, all single- and multidimensional high-level features result in an 154-dimensional feature vector.

4.2 Concept-based framework

To improve genre classification, we aim at modeling common bass playing styles that are typical for certain music genres. Therefore, we apply a generic framework to translate known musicological properties into explicit restrictions on feature values. The assignment of weighting factors furthermore allows to take the importance of each property into account. In the following subsections, we introduce the terms *concept*, *class*, and *property* as the major components of the framework. Afterwards we explain how *relevance values* for both properties and classes are derived to measure their significance to the investigated piece of music and close with a detailed example. Hereafter, multidimensional variables are denoted in bold print.

Concepts & classes

The term *concept* represents a general approach to categorize music. Each concept is defined by a set of *classes* as

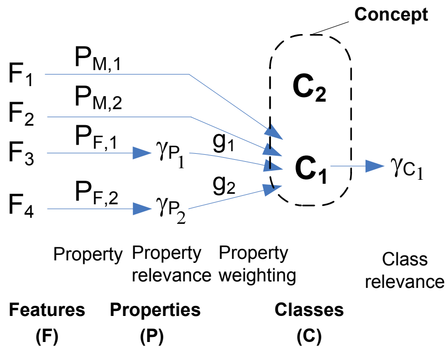

Figure 1. Concept-based framework

shown in Fig. 1. In this paper, we apply the concepts *Bass-PlayingStyle* (denoted as SB), which represents a common way of playing the bass in a specific music genre and *Genre* (denoted as G), which is a common category for musicologists. One well-known example is the bass playing style walking bass (defined as class *WalkingBass*), which is widely applied by bass players in different music genres related to *Swing*. It is covered as an example in the end of this section. The assignment between classes of both concepts is shown in Fig. 2.

Properties

Each class is defined by a number of *properties* $\boldsymbol{P}$. They translate its musicological description into explicit restrictions on the values of certain *features* $\boldsymbol{F}$.

We discern *mandatory properties* (M) and *frequent properties* (F). Mandatory properties are strictly need to be fulfilled whereas frequent properties are not mandatory for a certain class. A *weighting factor* $0 \leq g_i \leq 1$ is assigned to each frequent property. g_i is proportional to the importance of the corresponding property with regard to the current class.

Furthermore, properties are either *omnipresent* (O) or *conditional* (C). Omnipresent properties are constantly valid, whereas the validity of conditional properties depends on a certain condition. This may for instance be the presence of an instrument that a feature and thus a property is related to. Only if the condition is fulfilled, the corresponding property needs to be considered. Generally, the indices of P imply the corresponding property type. Examples are given in the end of this section. We derived the weighting factors and thresholds of all properties used in this paper from experiments with development data samples, which did not belong to the evaluation set.

Relevance values

The *property relevance value* γ_P measures to what extent a property P is fulfilled ($\gamma_P = 1$) or not ($\gamma_P = 0$). It is derived from the corresponding feature value F by using a *rating function* $r(F)$. This function depends on the type of restriction on the feature value F that is defined by P. For instance, we use $\gamma_P = r(F) = 0.5[\text{sgn}(F - V) + 1]$ to match the property $P \rightarrow F$ `isBiggerThan` V. The

A frequent use of chord tones is mandatory.
$P_{1,MO} \rightarrow F_{ChordToneRatio}$ `isBiggerThan` 0.3
2) The melodic direction is often constant within each bar. (important property - weighting factor $g_2 = 0.7$)
$P_{2,FO} \rightarrow F_{ConstantDirection}$ `isBiggerThan` 0.7
3) If quarter notes are primarily used (such as in slow and mid-tempo Jazz songs), there is a high swing factor related to the eighth note grid. (important property - weighting factor $g_3 = 0.8$)
`if` *Condition* ($F_{DominantRhythmicalGrid}$ `is` 4) $P_{3,FC} \rightarrow F_{SwingFactor,8}$ `isBiggerThan` 0.7
4) If eighth notes are primarily used (such as in up-tempo Jazz songs), there is a high swing factor related to the sixteenth note grid. (important property - weighting factor $g_4 = 0.8$)
`if` *Condition* ($F_{DominantRhythmicalGrid}$ `is` 8) $P_{4,FC} \rightarrow F_{SwingFactor,16}$ `isBiggerThan` 0.7
5) Chromatic note passages are occasionally used. (less important property - weighting factor $g_5 = 0.3$)
$P_{5,FO} \rightarrow F_{Chromatics}$ `isRelativelyHigh`

Table 1. Properties of the class *WalkingBass* (concept *BassPlayingStyle*)

rating function is designed in such a way that $0 \leq \gamma_P \leq 1$ is assured.

Subsequently, the *class relevance value* γ_C is derived for each class C from its corresponding property relevance values. γ_C quantifies to what extend a certain class is relevant for the musicological description of an analyzed piece of music.

We suggest the following algorithm to comply with the different property types. If all mandatory properties are given to be true, γ_C is calculated as a weighted sum over all frequent properties $\boldsymbol{\gamma_{P_F}}$ according to their normalized weighting factors $\widehat{\boldsymbol{g}}$ ($\sum \widehat{g}_i = 1$). Otherwise it is set to zero. This algorithm can be summarized as follows:

$$\gamma_C = \begin{cases} \sum_i \widehat{g}_i \gamma_{P_{F,i}} & \text{if } \gamma_{P_{M,j}} = 1 \, \forall \, P_{M,j} \in \boldsymbol{P_M}, \\ 0 & \text{else} \end{cases} \quad (1)$$

Example

As shown in Table 1, the class *WalkingBass* of the concept *BassPlayingStyle* is defined by 5 feature-related properties that are derived from musicological properties of this style.

5. EVALUATION

We use two data sets consisting of symbolic (MIDI) and real audio (AUDIO) each with 50 respectively 40 excerpts from each of the genres *PopRock* (POP), *Swing* (SWI), *Latin* (LAT), *Funk* (FUN), *Blues* (BLU), and *MetalHardRock* (MHR). All excerpts are derived from instrumental solo parts of the melody instruments between 20 and 35 seconds of length. Fig. 3 depicts all processing steps that precede the evaluation.

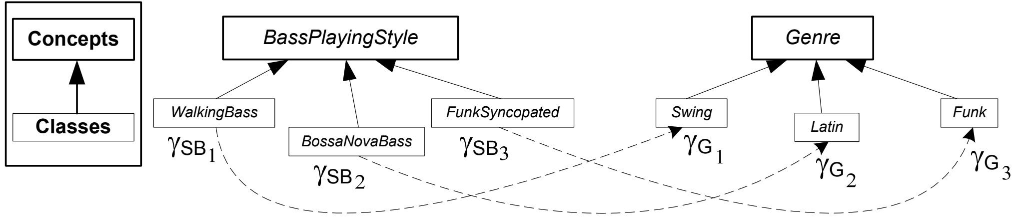

Figure 2. Assignment between the classes of the concepts *BassPlayingStyle* and *Genre*

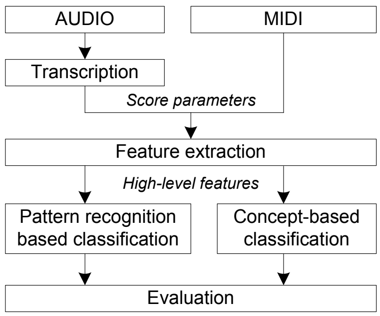

Figure 3. Processing flow-chart

5.1 Transcription & Pre-processing

We used the Transcription Toolbox [6] to extract the score parameters from real audio data. It provides algorithms to extract the bass, melody, harmony, and drum part as well as the beat grid information. As explained in Sec. 4.1, our aim is to focus on the bass and its interaction with the participating instruments. Concerning symbolic audio data, score parameters are directly extracted.

5.2 Feature Selection (FS) and Feature Space Transformation (FST)

The following feature selection and feature space transformation techniques have been utilized to reduce the dimensionality of the feature space.

Inertia Ratio Maximization using Feature Space Projection (IRMFSP).
IRMFSP was proposed in [17]. This FS algorithm is motivated by the ideas similar to Fisher's discriminant analysis. During each iteration of the algorithm, we look for the feature maximizing the ratio of between-class inertia to the total-class inertia. To avoid the next chosen feature to provide the same information on the next iteration, all features are orthogonalized to the selected one. In this evaluation we use the ISMFSP algorithms with the modifications proposed in [8].

Linear Discriminant Analysis (LDA)
LDA is one of the most often used supervised FST methods [10]. It is successfully applied as a pre-processing for audio signal classification. Original feature vectors are linearly mapped into new feature space guaranteeing a maximal linear separability by maximization of the ratio of between-class variance to the within-class variance. This mapping is conducted by multiplying the original $K \times N$ dimension feature matrix $\mathbf{X}$ with the transformation matrix $\mathbf{T}$. Reducing the dimension of the transformed feature vector from N to $D \leq N$ is achieved by considering only the first D column vectors of $\mathbf{T}$ for multiplication.

Generalized Discriminant Analysis (GDA)
Real-world classification routines often have to deal with non-linear problems, thus linear discrimination in the original feature space is often not possible. The idea of the FST technique GDA [3] is to map the features into higher dimensional (sometimes infinity dimensional) space, where the linear discrimination is possible. Dealing with a high dimensional space leads to an increase of the computation effort. To overcome this problem, the so called *kernel trick* is applied. The key idea of the kernel trick is to replace the dot product in a high-dimensional space with a kernel function in the original feature space.

5.3 Classification

We applied 4 known methods (SVM, GMM, NB, and kNN) as well as a novel concept-based approach for the purpose of classification.

Support Vector Machines
A Support Vector Machine (SVM) is a discriminative classifier, attempting to generate an optimal decision plane between feature vectors of the training classes [20]. Commonly for real-world applications, classification with linear separation planes is not possible in the original feature space. The transformation to the higher dimensional space is done using above mentioned kernel trick (we applied the RBF kernel in this paper). Transformed into a high-dimensional space, non-linear classification problems can become linearly solvable.

Gaussian Mixture Models
Gaussian Mixture Models (GMM) are commonly used generative classifiers.Single data samples of the class are interpreted as being generated from various sources and each source is modeled by a single multivariate Gaussian. The probability density function (PDF) is estimated as a weighted sum of the multivariate normal distributions. The parameters of a GMM can be estimated using the Expectation-Maximization algorithm [5].

Naive Bayes Classifier

Naive Bayes classifier (NB) is a simple probabilistic classifier. NB uses a strong assumption of feature dimensions being statistically independent and thus takes into account only means and variances over the feature dimensions for all training data of the class. Recently, applicability and efficiency of NB classifiers were discussed in detail in [21].

k-Nearest Neighbor

With k-Nearest Neighbor (kNN), the classification is based on the class assignment of the closest training examples in the feature space [7]. We used the Euclidean distance here. This type of discriminative classifier is also referred as instance based learning. The level of generalization of kNN can be tuned by adjusting the number of nearest neighbors k taken into account.

Novel approach: concept-based classifier

Using Eq. 1, we derive a class relevance value $\gamma_{SB_i} \equiv \gamma_{C_i}$ for each class of the concept *BassPlayingStyle*. We defined one common bass playing style for each of the 6 genres that were considered in the evaluation (see Sec. 5), namely *WalkingBass* (SWI), *BluesShuffle* (BLU), *FunkSyncopated* (FUN), *SteadyRiff* (MHR), *BossaNovaBass* (LAT), and *ChordRootAccompaniment* (POP). For our experiments, we used 5 different properties for each class.Using the assignment between the classes of both concepts as depicted in Fig. 2, the concept-based classifier estimates the genre $\widehat{G} = G_j$ that is assigned to the bass playing style SB_i with the highest class relevance value γ_{SB_i}. In case two or more bass playing styles related to different genres obtain the same class relevance values, the classification is considered to be correct if at least one of the candidates is related to the correct genre and false if not. As a proof of concept, we performed the evaluation experiment using the concept-based classifier on the MIDI data set.

6. RESULTS

Table 3 gives an overview over the classification scores for different FS / FST combination. For each combination, the parametrization with the best results is depicted. Further evaluation parameters such as the number of gaussians for the GMM classifiers, k for the kNN classifiers, and the number of dimensions after IRMFSP are given in brackets. We performed a 25-fold cross validation to derive mean classification scores and their standard deviations (given in brackets below) for each classifier. As shown there, best mean classification accuracies for the MIDI and AUDIO data set of 81.47% and 46.85% have been achieved applying a combined IRMFSP - GDA preprocessing for both data sets. Above all, we expect transcription errors affecting note pitch values, onset values and beat grid information to cause significantly lower classification scores for real audio data. For both data sets, the application of feature selection and feature space transformation algorithms clearly increases the accuracy values of the subsequent classifiers.

	BLU	FUN	LAT	MHR	POP	SWI
BLU	**68.0**	-	4.0	-	-	28.0
FUN	28.0	**46.0**	4.0	4.0	4.0	14.0
LAT	16.0	-	**70.0**	-	2.0	12.0
MHR	34.0	8.0	6.0	**34.0**	2.0	16.0
POP	36.0	-	20.0	2.0	**6.0**	36.0
SWI	36.0	-	22.0	-	-	**42.0**

Table 2. Confusion matrix of the concept-based classifier (MIDI data set) in %

As depicted in Table 2, the concept-based classifier achieved a mean classification accuracy of 44.3% varying in a strong way for different genres. Best results have been obtained for *Latin* (70.0%) and *Blues* (68.0%). The low results for *Pop* (6.0%) and *MetalHardRock* (34.0%) lead to the assumption, that modeling only one bass playing style per genre is not sufficient due to the high variability in the applied data set. Further steps include the evaluation based on a larger database.

7. CONCLUSIONS & FUTURE WORK

In this paper, we introduced a novel set of transcription-based high-level features related to the rhythmic, melodic, harmonic, and structural description of bass lines. Furthermore, we presented a new approach to model musical knowledge of musical styles as properties related to the values of transcription-based high-level features. The main advantage of concept-based classification approach is that significantly fewer features are necessary to model each class as in common machine learning approaches. Future steps include modeling additional genre-specific bass playing styles as well as transferring the proposed method onto other frequently used instruments like the guitar.

8. ACKNOWLEDGMENTS

This work has been partly supported by the german research project *GlobalMusic2One* [1] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B). The authors would like to thank Paul Bräuer for the fruitful discussions.

9. REFERENCES

[1] J. Abeßer, C. Dittmar, and H. Großmann. Automatic genre and artist classification by analyzing improvised solo parts from musical recordings. In *Proceedings of the Audio Mostly*, 2008.

[2] J.-J. Aucouturier, B. Defreville, and F. Pachet. The bag-of-frame approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *Journal of the Acoustical Society of America*, 122(2):881–891, 2007.

[3] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10):2385–2404, 2000.

[4] P. J. Ponce de Léon and J. M. Iñesta. Pattern recognition approach for music style identification using shallow statistical

[1] http://www.globalmusic2one.de

Dataset	FS / FST	Dim.	SVM	GMM(2)	GMM(3)	GMM(5)	GMM(10)	NB	kNN(1)	kNN(5)	kNN(10)
MIDI	-	154	69.13 (8.34)	67.45 (9.08)	66.28 (9.45)	59.52 (6.21)	60.31 (5.97)	60.03 (7.27)	66.88 (6.51)	64.17 (5.61)	62.69 (7.39)
	LDA	5	63.06 (7.20)	59.24 (5.10)	61.22 (6.69)	53.44 (7.43)	59.23 (14.43)	60.50 (7.33)	60.14 (6.97)	62.15 (7.66)	62.38 (8.53)
	GDA($\gamma = 2^{-7}$)	5	77.60 (7.65)	77.60 (7.65)	77.60 (7.65)	44.04 (8.31)	18.73 (9.54)	18.37 (6.27)	77.60 (7.65)	77.60 (7.65)	77.60 (7.65)
	IRMFSP(20)	20	73.82 (7.89)	58.70 (8.08)	65.07 (8.23)	63.75 (10.20)	64.21 (7.08)	57.99 (6.38)	**78.06** (7.31)	71.70 (6.84)	68.85 (8.72)
	IRMFSP(80) + LDA	5	72.15 (8.93)	69.87 (10.67)	69.34 (7.60)	67.95 (9.81)	65.20 (11.70)	69.65 (8.52)	69.45 (9.02)	70.48 (7.93)	69.66 (6.65)
	IRMFSP(40) + GDA($\gamma = 2^{-5}$)	5	76.99 (13.88)	19.32 (5.07)	20.15 (9.26)	13.30 (5.60)	16.09 (2.85)	18.37 (6.27)	81.10 (6.39)	**81.47** (6.20)	**81.47** (6.20)
AUDIO	-	154	41.33 (8.33)	33.45 (8.59)	34.95 (8.98)	33.73 (10.33)	33.80 (10.62)	27.24 (8.33)	36.54 (10.35)	31.42 (11.61)	32.25 (9.66)
	LDA	5	32.98 (7.39)	32.82 (7.46)	30.16 (6.88)	28.90 (7.95)	28.76 (8.33)	34.25 (7.58)	31.16 (9.00)	33.84 (8.66)	34.10 (8.28)
	GDA($\gamma = 2^{-9}$)	5	42.74 (11.53)	42.74 (11.53)	42.74 (11.53)	27.09 (15.08)	15.02 (6.90)	12.79 (6.63)	42.74 (11.53)	42.74 (11.53)	42.74 (11.53)
	IRMFSP(40)	40	43.26 (11.76)	39.19 (10.83)	38.31 (10.56)	39.83 (12.93)	36.62 (9.86)	26.69 (8.87)	45.23 (11.70)	42.04 (12.06)	37.83 (10.67)
	IRMFSP(20) + LDA	5	43.80 (10.61)	41.66 (11.16)	42.28 (10.68)	41.32 (11.50)	40.58 (13.52)	**43.90** (12.09)	35.29 (9.10)	40.69 (10.75)	41.48 (10.24)
	IRMFSP(40) + GDA($\gamma = 2^{-5}$)	5	**46.85** (9.73)	**46.85** (9.73)	**46.85** (9.73)	26.59 (13.75)	16.64 (7.10)	12.79 (6.63)	**46.85** (9.73)	**46.85** (9.73)	**46.85** (9.73)

Table 3. Mean classification accuracy [%] for the MIDI and AUDIO data set (standard deviation [%] given in brackets)

descriptors. *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews*, 37(2):248–257, March 2007.

[5] A. P. Dempster, N. M. Laird, and D. B. Rdin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[6] C. Dittmar, K. Dressler, and K. Rosenbauer. A toolbox for automatic transcription of polyphonic music. In *Proceedings of the Audio Mostly*, 2007.

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2nd edition, November 2000.

[8] S. Essid. *Classification automatique des signaux audio-fréquences : reconnaissance des instruments de musique*. PhD thesis, Université Pierre et Marie Curie, Paris, France, December 2005.

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

[10] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, September 1990.

[11] M. Goto. 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, 2004.

[12] F. Gouyon. *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, University Pompeu Fabra, 2005.

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

[14] J.-L. Hsu, C.-C. Liu, and A. L. P. Chen. Discovering nontrivial repeating patterns in music data. In *IEEE Transactions on Multimedia*, volume 3 of *IEEE Transactions on Multimedia*, pages 311–324, September 2001.

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

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

[17] G. Peeters and X. Rodet. Hierarchical gaussian tree with inertia ratio maximization for the classification of large musical instruments databases. In *Proceedings of the 6th Int. Conf. on Digital Audio Effects (DAFx)*, London, UK, 2003.

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

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

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

[21] H. Zhang. The optimality of naive bayes. In V. Barr and Z. Markov, editors, *Proceedings of the FLAIRS Conf.* AAAI Press, 2004.