BASS PLAYING STYLE DETECTION BASED ON HIGH-LEVEL FEATURES AND PATTERN SIMILARITY

Jakob Abeßer
Fraunhofer IDMT
Ilmenau, Germany
(abr@idmt.fraunhofer.de)

Paul Bräuer
Piranha Musik & IT
Berlin, Germany

Hanna Lukashevich, Gerald Schuller
Fraunhofer IDMT
Ilmenau, Germany

ABSTRACT

In this paper, we compare two approaches for automatic classification of bass playing styles, one based on high-level features and another one based on similarity measures between bass patterns. For both approaches, we compare two different strategies: classification of patterns as a whole and classification of all measures of a pattern with a subsequent accumulation of the classification results. Furthermore, we investigate the influence of potential transcription errors on the classification accuracy, which tend to occur when real audio data is analyzed. We achieve best classification accuracy values of 60.8% for the feature-based classification and 68.5% for the classification based on pattern similarity based on a taxonomy consisting of 8 different bass playing styles.

1. MOTIVATION

Melodic and harmonic structures were often studied in the field of Music Information Retrieval. In genre discrimination tasks, however, mainly timbre-related features are somewhat satisfying to the present day. The authors assume, that bass patterns and playing styles are missing complementaries. Bass provides central acoustic features of music as a social phenomenon, namely its territorial range and simultaneous bodily grasp. These qualities come in different forms, which are what defines musical genres to a large degree. Western popular music with its worldwide influence on other styles is based upon compositional principles of its classical roots, harmonically structured around the deepest note. African styles also often use tonal bass patterns as ground structure, while Asian and Latin American styles traditionally prefer percussive bass sounds. In contrast to the melody (which can easily be interpreted in "cover versions" of different styles), the bass pattern most often carries the main harmonic information as well as a central part of the rhythmic and structural information.

A more detailed stylistic characterization of the bass instrument within music recordings will inevitably improve classification results in genre and artist classification tasks. Within the field of Computational Ethnomusicology (CE) [19], the automatic detection of the playing styles of the participating instruments such as the bass constitutes a meaningful approach to unravel the fusion of different musical influences of a song. This holds true for many contemporary music genres and especially for those of a global music background.

The remainder of this paper is organized as follows. After outlining the goals and challenges in Sec. 2 and Sec. 3, we provide a brief overview over related work in Sec. 4. In Sec. 5, we introduce novel high-level features for the analysis of transcribed bass lines. Furthermore, we propose different classification strategies, which we apply and compare later in this paper. We introduce the used data set and describe the performed experiments in Sec. 6. After the results are discussed, we conclude this paper in Sec. 7.

2. GOALS

The goal of this publication is to compare different approaches for automatic playing style classification. For this purpose, we aim at comparing different classification approaches based on common statistical pattern recognition algorithms as well as on the similarity between bass patterns. In both scenarios, we want to investigate the applicability of a aggregation classification based on the subpatterns of an unknown pattern.

3. CHALLENGES

The extraction of score parameters such as note pitch and onset from real audio recordings requires reliable automatic transcription methods, which nowadays are still error-prone when it comes to analyzing multi-timbral and polyphonic audio mixtures [4, 13]. This drawback impedes a reliable extraction of high-level features that are designed to capture important rhythmic and tonal properties for a description of an instrumental track. This is one problem addressed in our experiments. Another general challenge is the translation of musical high-level terms such as syncopations, scale, or pattern periodicity into parameters that are automatically retrievable by algorithms. Information regarding micro-timing, which is by the nature of things impossible to encompass in a score [9], is left out.

4. PREVIOUS APPROACHES

Within the last years, the use of score-based high-level features became more popular for tasks such as automatic genre classification. To derive a score-based representation from real audio recordings, various automatic transcription algorithms have been proposed so far. The authors of [18], [13], and [4] presented algorithms to transcribe bass lines. Musical high-level features allow to capture different properties from musical domains such as melody, harmony, and rhythm [1, 3, 10, 11]. Bass-related audio features we used for genre classification in [18], [1], and [17].

An excellent overview over existing approaches for the analysis of expressive music performance and artist-specific playing styles is provided in [23] and [24]. In [7], different melodic and rhythmic high-level features are extracted before the performed melody is modeled with an evolutionary regression tree model. The authors of [15] also used features derived from the onset, 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. To compare different performances in terms of rhythmic and dynamic similarity, the authors of [14] proposed a numerical method based on the correlation at different timescales.

5. NOVEL APPROACH

5.1 Feature extraction

In this paper, we use 23 multi-dimensional high-level features that capture various musical properties for the tonal and rhythmic description of bass lines. The feature vector consists of 136 dimensions in total. The *basic note parameters*, which we investigate in this paper, are the absolute pitch $\boldsymbol{\Theta_P}$, the loudness $\boldsymbol{\Theta_V}$, the onset $\boldsymbol{\Theta_O^{[s]}}$and $\boldsymbol{\Theta_O^{[M]}}$, and the duration $\boldsymbol{\Theta_D^{[s]}}$and $\boldsymbol{\Theta_D^{[M]}}$of each note. The indices [s] and [M] indicate that both the onset and the duration of a note can be measured in seconds as well as in lengths of measures. All these parameters are extracted from symbolic MIDI files by using the MIDI-Toolbox for MATLAB [5].

Afterwards, further *advanced note parameters* are derived before features are extracted. From the pitch differences $\boldsymbol{\Delta\Theta_P}$ between adjacent notes in semitones, we obtain vectors containing the interval directions $\boldsymbol{\Delta\Theta_P^{(D)}}$ (being either ascending, constant, or descending), and the pitch differences in terms of functional interval types $\boldsymbol{\Delta\Theta_P^{(F)}}$. To derive the functional type of an interval, we map its size to a maximum absolute value of 12 semitones or one octave by using the modulo 12 operation in case it is larger than one octave upwards or downwards (12 semitones). Then each interval is assigned to a function interval type (prime, second, third etc.) according to well known music principles. In addition to the high-level features presented in [1], we use various additional features related to tonality and rhythm in this paper, which are explained in the following subsections.

Features related to tonality

We derive features to measure if a certain *scale* is applied in a bass pattern. Therefore, we take different binary scale templates for natural minor (which includes the major scale), harmonic minor, melodic minor, pentatonic minor (subset of natural minor which also includes the pentatonic major scale), blues minor, whole tone, whole tone half tone, arabian, minor gypsy and hungarian gypsy [21] into account. Each scale template consists of 12 values representing all semitones of an octave. The value 1 is set for all semitones that are part of the scale, the value 0 for those that are not. All notes within a given pattern, which are related to a certain scale, are accumulated by adding their normalized note loudness values $\Theta_V/\Theta_{V,max}$ with $\Theta_{V,max}$ being the maximum note loudness in a pattern. The same is done for all notes, which are not contained in the scale. The ratio of both sums is calculated over all investigated scales and over all 12 possible cyclic shifts of the scale template. This cyclic shift is performed to cope with each possible root note position. The maximum ratio value over all shifts is determined for each scale template and used as a feature value, which measures the presence of each considered scale. We obtain the relative frequencies p_i of all possible values in the vector that contains the interval directions ($\boldsymbol{\Delta\Theta_P^{(D)}}$) as well as the vector that contains the functional interval types ($\boldsymbol{\Delta\Theta_P^{(F)}}$) and use them as features to characterize the variety of different pitch transitions between adjacent notes.

Features related to rhythm

Syncopation embodies an important stylistic means in different music genres. It represents the accentuation on weak beats of a measure instead of an accentuation on a neighbored strong beat that usually would be emphasized. To detect syncopated note sequences within a bass-line, we investigate different temporal grids in terms of equidistant partitioning of single measures. For instance, for an eight-note grid, we map all notes inside a measure towards one of eight segments according to their onset position inside the measure. In a $\frac{4}{4}$ time signature, these segments correspond to all 4 quarter notes (on-beats) and their off-beats in between. If at least one note is mapped to a segment, it is associated with the value 1, otherwise with 0. For each grid, we count the presence of the following segment sequences - (1001), (0110), (0001), or (0111). These sequences correspond to sequences of alternating on-beat and off-beat accentuations that are labeled as syncopations. The ratios between the number of syncopation sequences and the number of segments are applied as features for the rhythmical grids 4, 8, 16, and 32.

We calculate the ratio $\Theta_D^{(M)}(k)/\Delta\Theta_O^{(M)}(k)$ between the duration value of the k-th note in measure lengths and the inter-onset-interval between the k-th note and its succeeding note. Then we derive the mean and the variance of this value over all notes as features. A high or low mean value indicates whether notes are played *legato* or *staccato*. The variance over all ratios captures the variation between these two types of *rhythmic articulation* within a given bass pattern. To measure if notes are mostly played on *on-beats* or *off-beats*, we investigate the distribution of notes towards the segments in the rhythmical grids as explained above for the syncopation feature. For example, the segments 1, 3, 5, and 7 are associated to on-beat positions for an eight-note grid and a $\frac{4}{4}$ time signature. Again, this ratio is calculated over all notes and mean and variance are taken as feature values. As additional rhythmic properties, we derive the frequencies of occurrence of all commonly used note lengths from half notes to 64th notes, each in its normal, dotted, and triplet version. In addition, the relative frequencies from all note-note, note-break and break-note sequences over the complete pattern are taken as features.

5.2 Classification based on statistical pattern recognition

We investigate the applicability of the well-established Support Vector Machines (SVM) using the Radial Basis Function (RBF) as kernel combined with a preceding feature selection using the Inertia Ratio Maximization using Feature Space Projection (IRMFSP) as a baseline experiment. The feature selection is applied to choose the most discriminative features and thus to reduce the dimensionality of the feature space prior to the classification. Therefore, we calculate the high-level features introduced in 5.1 for each bass pattern, which results in an 136 dimensional feature space. Details on both the SVM and the IRMFSP can be found for instance in [1].

5.3 Classification based on pattern similarity

In this paper, we apply 2 different kinds of pattern similarity measures, *pairwise similarity measures* and *similarity measures based on the Levenshtein distance*. To compute similarity values between patterns, the values of the onset vector $\boldsymbol{\Theta}_{\boldsymbol{O}}^{[M]}$ and the absolute pitch vector $\boldsymbol{\Theta_P}$ are simply converted into character strings. In the latter case, we initially subtract the minimum value of $\boldsymbol{\Theta_P}$ for each pattern separately to remain independent from pitch transpositions. This approach can of course be affected by potential outliers, which do not belong to the pattern.

5.3.1 Similarity measures based on the Levenshtein distance

The Levenshtein distance D_L offers a metric for the computation of the similarity of strings [6]. It measures the minimum number of edits in terms of insertions, deletions, and substitutions, which are necessary, to convert one string into the other. We use the Wagner-Fischer algorithm [20] to compute D_L and derive a similarity measure S_L between two strings of length l_1 and l_2 from

$$S_L = 1 - D_L/D_{L,max}\ . \tag{1}$$

The lengths l_1 and l_2 correspond to the number of notes in both patterns. $D_{L,max}$ equals the maximum value of l_1 and l_2. In the experiments, we use the rhythmic similarity measure $S_{L,R}$ and the tonal similarity measure $S_{L,T}$ derived from the Levenshtein distance between the onset $\boldsymbol{\Theta}_{\boldsymbol{O}}^{[M]}$ and the pitch $\boldsymbol{\Theta_P}$ as explained in the previous section. Furthermore, we investigate

$$S_{L,RT,Max} = \begin{cases} S_{L,R} & , S_{L,R} \geq S_{L,T} \\ S_{L,T} & , S_{L,T} > S_{L,R} \end{cases} \tag{2}$$

and

$$S_{L,RT,Mean} = \frac{1}{2}(S_{L,R} + S_{L,T}) \tag{3}$$

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

5.3.2 Pairwise similarity measures

In general, we derive a pairwise similarity measure

$$S_P = \frac{1}{2}\left(\frac{N_{n,m}}{N_n} + \frac{N_{m,n}}{N_m}\right) \tag{4}$$

$N_{n,m}$ denotes the number of notes in pattern n, for which at least one note in pattern m exists that have the same absolute pitch value (for the similarity measure $S_{P,T}$) or onset value (for the similarity measure $S_{P,R}$). $N_{m,n}$ is defined vice versa. By applying the constraint that both onset and absolute pitch need to be equal in Eq. 4, we obtain the measure $S_{P,RT}$. Furthermore, we derive the aggregated similarity measures $S_{P,RT,Max}$ and $S_{P,RT,Mean}$ analogous to Eq. 2 and Eq. 3.

6. EVALUATION

6.1 Data-set

We assembled a novel dataset from instructional bass literature [12, 21], which consists of bass patterns from the 8 genres *Swing* (SWI), *Funk* (FUN), *Blues* (BLU), *Reggae* (REG), *Salsa & Mambo* (SAL), *Rock* (ROC), *Soul & Motown* (SOU) and *Africa* (AFR), a rather general term which here signifies Sub-Saharan Popular Music Styles [16]. For each genre, 40 bass-lines of 4 measure length have been stored as symbolic audio data as MIDI files. Initial listening tests revealed that in this data set, which was assembled and categorized by professional bass players, a certain amount of stylistic overlap and misclassification between genres as for instance Blues and Swing or Soul & Motown and Funk occurs. The overlap is partly inherent to the approach of the data sets, which treat all examples of a style (e.g. Rock) as homogenous although the sets include typical patterns of several decades. In some features, early Rock patterns might resemble early Blues patterns more than they resemble late patterns of their own style [22]. Thus, the data set will be extended further and revised by educated musicologists for future experiments.

6.2 Experiments & Results

6.2.1 Experiment 1 - Feature-based classification

As described in Sec. 5.2, we performed a baseline experiment that consists of IRMFSP for chosing the best $N = 80$ features and the SVM as classifier. The parameter N has

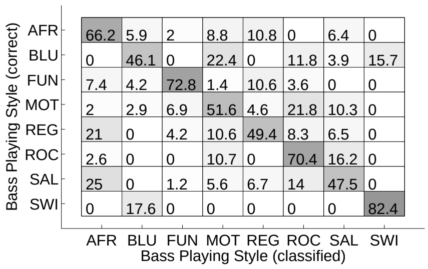

Figure 1. Exp. 1 - Confusion matrix for the feature-based pattern-wise classification (all values given in %). Mean classification accuracy is 60.8% with a standard deviation of 2.4%.

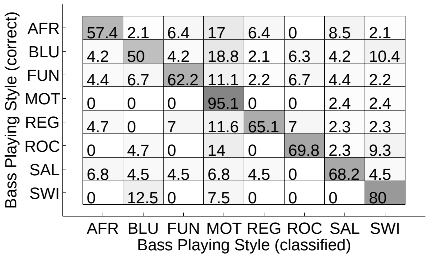

Figure 2. Exp. 2 - Confusion matrix for the best similarity-based configuration (measure-wise classification using the $S_{P,RT,Max}$ similarity measure - all values given in %). Mean classification accuracy is 68.5% with a standard deviation of 3.1%.

been determined to perform best in previous tests on the data-set. A 20-fold cross validation was applied to determine the mean and standard deviation of the classification accuracy. For a feature extraction and classification based on complete patterns, we achieved 60.8% of accuracy with a standard deviation of 2.4%. The corresponding confusion matrix is shown in Fig. 1. It can be seen, that best classification results were achieved for the styles Funk, Rock, and Swing. Strong confusions between Blues and Motown respectively Swing, Motown and Rock, Reggae and Africa as well as between Salsa and Africa can be identified. These confusions support the musicological assessment of the data-set given in Sec. 6.1. In addition, they coincide with historical relations between the styles in Africa, the Caribbean, and Latin America, as well as relations within North America as it is common musicological knowledge [8].

As a second classification strategy, we performed the feature extraction and classification based on sub-patterns. Therefore, we divided each pattern within the test set into $N = 4$ sub-patterns of one measure length. It was ensured, that no sub-patterns of patterns in the test set were used as training data. After all sub-patterns were classified, the estimated playing style for the corresponding test set pattern was derived from a majority decision over all sub-pattern classifications. In case of multiple winning classes, a random decision was applied between the winning classes. For the accumulated measure-wise classification, we achieved only 56.4% of accuracy. Thus, this approach did not improve the classification accuracy. We assume that the majority of the applied high-level features that are based on different statistical descriptors (see Sec. 5.1 for details), can not provide a appropriate characterization of the sub-patterns, which themselves only consist of 6 to 9 notes in average.

6.2.2 Experiment 2 - Pattern Similarity

This experiment is based on a leave-one-out cross-validation scheme and thus consists of $N = 320$ evaluation steps according to the 320 patterns in the data-set. Within each evaluation step, the current pattern $\mathcal{P}_k$ is used as test data while all remaining patterns $\mathcal{P}_l$ with $l \neq k$ are used as training data. We derive the class estimate $\hat{c}_k$ of $\mathcal{P}_k$ from the class label $\hat{c}$ of the best-fitting pattern $\widehat{\mathcal{P}}$ as

$$\hat{c}_k = c_{\hat{l}} \Leftrightarrow \hat{l} = \arg\max_{l} S_{k,l} \tag{5}$$

with $S_{k,m}$ representing the similarity measure between $\mathcal{P}_k$ and $\mathcal{P}_m$ in the given case. As in Sec. 6.2.1, if multiple patterns have the same (highest) similarity, we perform a random decision among these candidates. This experiment is performed for all similarity measures introduced in Sec. 6.2.2.

Exp. 2a: Pattern-wise classification. The basic approach for a pattern-based classification is to use each pattern of 4 measures length as one item to be classified.

Exp. 2b: Accumulated measure-wise classification. Bass patterns are often structured in a way, that the measure or a part of the measure, which precedes the pattern repetition, is often altered rhythmically or tonally and thus often varies greatly from the pattern. These figures separating or introducing pattern repetition are commonly referred to as *pickups* or *upbeats*, meaning that they do not vary or overlap the following pattern repetition which starts on the first beat of the new measure. A pattern-wise classification as described above thus might overemphasize the difference between the last measure because the patterns are compared over their complete length. Hence, we investigate another decision aggregation strategy in this experiment.

As described in Sec. 6.2.1, we divide each bass pattern into sub-patterns of one measure length each. Within each fold k, we classify each sub-pattern $\mathcal{SP}_{k,l}$ of the current test pattern $\mathcal{P}_k$ separately. At the same time, we ensure that only sub-patterns of the other patterns $\mathcal{P}_i$ with $i \neq k$ are used as training set for the current fold. To accumulate the classification results in each fold, we add all similarity values $\boldsymbol{S_{k,l}}$ between each sub-pattern $\mathcal{SP}_{k,l}$ towards their assigned winning pattern(s) $\boldsymbol{\mathcal{P}_{k,l,win}}$. The summation is done for each of the 6 genres separately. The genre that achieve the highest sum is considered as the winning genre.

As depicted in Fig. 3, the proposed accumulated measure-wise classification strategy led to higher classification accuracy values (blue bars) in comparison to a pattern-wise classification (red bars). This approach can be generalized and adopted to patterns of arbitrary length.

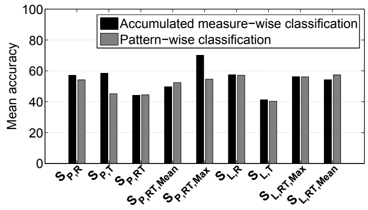

Figure 3. Mean classification accuracy results for experiment 2

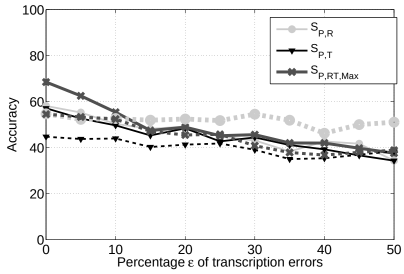

Figure 4. Exp. 3 - Mean classification accuracy vs. percentage ε of pattern variation (dotted line - pattern-wise similarity, solid line - accumulated measure-wise similarity).

The similarity measure $S_{P,RT,Max}$ clearly outperforms the other similarity measures by over 10 percent points of accuracy. The corresponding confusion matrix is shown in Fig. 2. We therefore assume that it is beneficial to use similarity information both based on pitch and onset similarity of bass patterns. For the pattern-wise classification, it can be seen that similarity measures based on tonal similarity generally achieve lower accuracy results in comparison to measures based on the rhythmic similarity. This might be explained by the frequently occurring tonal variation of patterns according to the given harmonic context such as a certain chord of a changed key in different parts of a song. The most remarkable result in confusion matrix is the very high accuracy of 95.1% for the Motown genre.

6.2.3 Experiment 3 - Influence of pattern variations

For the extraction of bass-patterns from audio recordings, two potential sources of error exist. In most music genres, the dominant bass patterns are object of small variations throughout a music piece. An automatic system might recognize the basic pattern or a variation of the basic pattern. Furthermore, automatic music transcription systems are prone to errors in terms of incorrect pitch, onset, and duration values of the notes. Both phenomena directly have a negative effect on the computed high-level features. We therefore investigate the achievable classification accuracy dependent on the percentage of notes with erroneous note parameters.

We simulate the mentioned scenarios by manipulating a random selection of ε percents of all notes from each unknown pattern and vary ε from 0% to 50%. The manipulation of a single note consists of either a modification of the onset $\mathbf{\Theta}_{\mathbf{O}}^{[M]}$ by a randomly chosen difference $-0.25 \leq \Delta\Theta_O^{(M)} \leq 0.25$ (which corresponds to a maximum shift distance of one beat for a $\frac{4}{4}$ time signature), a modification of the absolute pitch $\mathbf{\Theta}_{\mathbf{P}}$ by a randomly chosen difference $-2 \leq \Delta\Theta_P \leq 2$ (which corresponds to a maximum distance of 2 semitones), or a simple deletion of the current note from the pattern. Octave pitch errors that often appear in automatic transcription algorithms were not considered because of the mapping of each interval to a maximum size of one octave as described in Sec. 5.1. Insertions in terms of additional notes, which are not part of the pattern will be taken into account in future experiments.

As depicted in Fig. 4, the accuracy curve of the three different pair-wise similarity measures $S_{P,R}$, $S_{P,T}$ and $S_{P,RT,Max}$ falls until about 40% for a transcription error rate of 50% Interestingly, the pattern-wise classification based on $S_{P,R}$ seems to be more robust to transcription errors above 15% in comparison to the accumulated measure-wise classification even though it has a lower accuracy rate for the assumption of a perfect transcription.

6.2.4 Comparison to the related work

The comparison of the achieved results to the related work is not directly feasible. On one side, it is caused by the fact, that different data sets have been utilized. Tsunoo et al. [18] reported an accuracy of 44.8% for the GZTAN data set [1] while using only bass-line features. On the other side, the performance of only bass-line features was not every time stated. The work of Tsuchihashi et al. [17] showed an improvement of classification accuracy from 53.6% to 62.7% while applying bass-line features complimentary to other timbre and rhythmical features, but the results of genre classification with only bass features were not reported.

7. CONCLUSIONS & OUTLOOK

In this paper, different approaches for the automatic detection of playing styles from score parameters were compared. These parameters can be extracted from symbolic audio data (e.g. MIDI) or from real audio data by means of automatic transcription. For the feature-based appraoch, a best result of 60.8% of accuracy was achieved using a combination of feature selection (IRMFSP) and classifier (SVM) and a pattern-wise classification. Regarding the classification based on pattern similarity, we achieved 68.5% of accuracy using the combined similarity measure $S_{P,RT,Max}$ and a measure-wise aggregation strategy based on the classification of sub-patterns. The random baseline is 12.5%. This approach outperformed the common approach to classify the complete pattern as once.

For analyzing real-world audio recordings, further musical aspects such as micro-timing, tempo range, applied plucking & expression styles [2], as well as the interac-

[1] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transaction on Speech and Audio Processing, 10(5):293-302, 2002.

tion with other participating instruments need to be incorporated into a all-embracing style description of a specific instrument in a music recording. The results of experiment 4 emphasize the need for a well-performing transcription system for a high-level classification task such as playing style detection.

8. ACKNOWLEDGEMENTS

This work has been partly supported by the German research project *GlobalMusic2One* [2] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B). Additionally, the Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See* [3], enabling transnational cooperation between Thuringian companies and their partners from other European regions.

9. REFERENCES

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

[2] J. Abeßer, H. Lukashevich, and G. Schuller. Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[3] P. J. Ponce de Léon and J. M. Iñesta. 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, March 2007.

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

[5] Tuomas Eerola and Petri Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä, Jyväskylä, Finland, 2004.

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

[7] A. Hazan, M. Grachten, and R. Ramirez. Evolving performance models by performance similarity: Beyond note-to-note transformations. In *Proc. of the Int. Symposium*, 2006.

[8] Ellen Koskoff, editor. *The Garland Encyclopedia of World Music - The United States and Canada*. Garland Publishing, New York, 2001.

[9] Gerhard Kubik. *Zum Verstehen afrikanischer Musik*. Lit Verlag, Wien, 2004.

[10] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *Proc. of the Int. Symposium of Music Information Retrieval (ISMIR)*, 2004.

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

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

[13] 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.

[14] C. S. Sapp. Hybrid numeric/rank similarity metrics for musical performance analysis. In *Proc. of the Int. Symposium on Music Information Retrieval (ISMIR)*, pages 501–506, 2008.

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

[16] Ruth M. Stone, editor. *The Garland Encyclopedia of World Music - Africa*, volume 1. Garland Publishing, New York, 1998.

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

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

[19] George Tzanetakis, Ajay Kapur, W. Andrew Schloss, and Matthew Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, 2007.

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

[21] Paul Westwood. *Bass Bible*. AMA, 1997.

[22] Peter Wicke. *Handbuch der populären Musik: Geschichte, Stile, Praxis, Industrie*. Schott, Mainz, 2007.

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

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

[2] see http://www.globalmusic2one.net

[3] see http://www.idmt.de/eng/research_topics/songs2see.html