Tracking the Beat: Classification of Music Genres and Synthesis of Rhythms.

Debora C. Correa, Luciano da F. Costa
Instituto de Fisica de Sao Carlos
Universidade de Sao Paulo - USP
Sao Carlos, Brazil
deboracorrea@ursa.ifsc.usp.br, luciano@ifsc.usp.br

Jose H. Saito
Departamento de Computacao
Universidade Federal de Sao Carlos - UFSCar
Sao Carlos, Brazil
saito@dc.ufscar.br

Abstract— The continuous growth of online musical databases motivates the development of fast and reliable tools for music content analysis. Music genres, widely used to organize music collections, have become particularly interesting. This paper proposes an innovative framework for automatic classification of music genres. Matrices of transition probabilities are obtained which reflect the temporal sequence of rhythm notation events. Measurements are then obtained from these matrices and used to classify the artworks. Feature extraction is performed by PCA technique (unsupervised) while the classification is obtained through a Bayesian classifier. Qualitative results computed by the kappa coefficient confirm the effectiveness of the proposed methodology. An approach for the synthesis of new rhythms with similar characteristics of respective genres and pieces is also proposed, involving random walks.

Keywords- music genres, rhythm, Markovian Model, synthesis, random walk.

I. INTRODUCTION

Online musical databases have significantly increased in number and size, motivating more reliable and faster tools for studies in music content analysis, retrieval and description. In this context, music genres constitute particularly interesting descriptors, since they reflect interplay of cultures, artists and market strategies and have been used for years to organize music collections. Even more promising, music genres summarize some common characteristics (patterns) in music pieces. Despite their ample use, music genres are not a clearly defined concept, with fuzzy and controversial boundaries and taxonomies [1,2].

The task of automatically classifying music genres can be divided into three main steps: representation, features extraction, and classification. There are many previous works concerning automatic genre classification, with different representation of music information and adopted features [3-4]. Related works that use rhythm as features in automatic genre classification can also be found in the literature [5-6].

In a previous work [7], the problem of automatic genre classification was explored in terms of rhythmic features obtained through a digraph representation. The current study complements that analysis by considering additional features. We also report a new approach for the synthesis of rhythms. We investigate the possibility to compose new genres or artwork-based rhythms based on statistics of the extracted representations. Music scores in MIDI format are used, from which several measurements are derived. Then, we attempt to identify the genres based on their rhythmic patterns. Despite the fact that there is not a clear definition of rhythm [1], it can be related with the idea of temporal regularity. The rhythm is intrinsically characteristic of musical genres, since, for example, rock and salsa (rhythmically more complex) music can be immediately distinguishable by their specific rhythmic patterns. Rhythmic patterns are derived from the occurrence of sequence of events through rhythmic notations. Matrices of probability transition are computed from the rhythms of each piece, yielding a Markov model.

Therefore, we aim at improving the classification of the genres by taking into account also measurements obtained from digraphs built from the transition matrices. Supervised classification is performed by the Bayesian classifier, which receives as input the properties of the transition matrices as well as the measurements directly obtained by the digraphs; producing as output the most likely genre. Synthesis of new rhythms is also investigated in this work, by performing random walks through the aforementioned digraphs. Such methodology is applied to four well-known genres: blues, *bossa-nova*, reggae and rock.

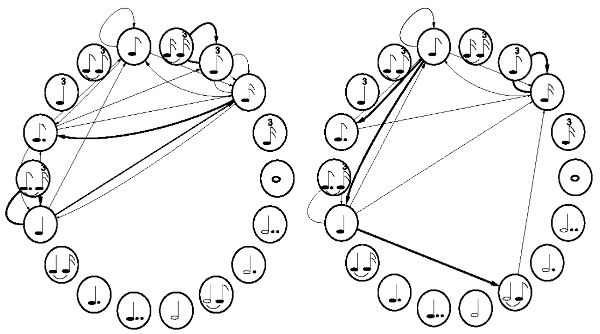
This paper is organized as follows: section II describes the proposed methodology, section III shows the results, and section IV presents the conclusions.

II. METHODOLY

A. *Representing Systems by Digraphs*

Digraphs, which are graphs with oriented edges, will be used in this work to represent the musical rhythms. A weighted digraph G can be defined by vertices (or nodes), edges (or links) and a mapping. Vertices are represented by integer numbers $i = 1, 2, ..., N$. The vertex set of the digraph G is $N(G)$ (N is the total number of vertices). Edges are set in the form (i, j) , denoting a connection from vertex i to vertex j . The mapping ω indicates that each edge (i, j) has a weight $\omega(i, j)$

associated to it. As a result, a weight matrix W can completely represent a weighted digraph [8].



(a) (b)

Figure 1. Digraphs of two music samples: (a) Looking Out the Window (Stevie R Vaughan); (b) Can't Buy Me Love (The Beatles).

Measurements which can be obtained from digraphs include: total vertex degree, strength and clustering coefficient. Total vertex degree (of vertex i) is defined as $k_i = k_i^{in} + k_i^{out}$ where k_i^{in} is the number of incoming edges; and k_i^{out} is the number of outgoing edges. The strength is defined as the total sum of weights associated to each vertex ($s_i^{in} + s_i^{out}$), that is, the sum of the weights of the respective incoming edges - in-strength $\left(s_i^{in} = \sum_j w_{ji} \right)$; and outgoing edges - out-strength $\left(s_i^{out} = \sum_j w_{ij} \right)$ of a vertex i . The clustering coefficient can be defined as $cc(i) = l_i / m_i(m_i - 1)$, where m_i is the number of neighbors of vertex i , and l_i is the number of connections between the neighbors of vertex i [8].

B. Data Description

For the purposes of this work, four music genres have been chosen: blues, *bossa-nova*, reggae and rock. They are well-known, reflecting different tendencies; and several music samples (artworks) are available in Web collections. Seventy artworks for each one of them were selected, downloaded in MIDI format. MIDI is an event-like format that works as a digital score [9]. To analyse and edit the MIDI scores, the software Sibelius was used (http://www.sibelius.com). As it is proposed to study the musical genres through their rhythms, the voice related to the percussion was extracted in each sample, since it is intrinsically appropriate to represent the rhythm of a piece. Once extracted, it is possible to analyse the elements involved in the rhythm. To perform such analysis, the free MIDI toolbox for Matlab was applied [9], in which functions can be used to study MIDI files in the Matlab computing environment.

We used the occurrence of sequences of events obtained by note duration (in beats), which are referred as rhythmic notations. A digraph is built taking into account the duration of the notes, regarding the sequence in which they occur in the sample. As the instrumentation is not been considered in this work, when two or more note events occurs in the same beat, the median duration of them is taken. The vertices of this digraph represent the rhythmic notations (quarter note, eighth note, and so on), while the edges reflect the vertices occurrence relation. For example, the presence of an edge from vertex i , represented by a sixteenth note, to a vertex j , represented by a eighth note, indicates that a sixteenth note was followed by an eighth note at least once. The thicker the edge, the larger is the strength between these two nodes, that is, more frequently these sequences occur in the piece. Fig. 1 depicts examples of these digraphs. Fig. 1 (a) shows a blues sample represented by the music *Looking Out The Window* by Stevie Ray Vaughan. Fig. 1(b) shows a rock sample, represented by the music *Can't Buy Me Love* by The Beatles. To complement the analysis, the 70 samples of each genre were used to built a unique digraph by genre: blues digraph (with all samples of blues), *bossa-nova* digraph, reggae digraph and rock digraph. They helped to visually compare the rhythms of these genres and were used in the synthesis process.

C. Feature Extraction and Analysis

Feature extraction is a fundamental step in most pattern recognition systems. Each pattern is represented by a vector in a d -dimension space, where d is the number of features or attributes. To maximize class separability, as generally desired in discrimination problems, good features need to be chosen. In particular, they should allow pattern vectors of different classes to occupy compact and distinct regions in the feature space.

As described previously, a digraph is created for each one of the 70 samples of each genre. After the exclusion of some rhythmic notations that hardly ever happens, it was verified that 18 vertices would be sufficient to represent the rhythmic notation possibilities concerning all the samples. In addition, it is important to avoid features that do not significantly contribute to the analysis, thus reducing the data dimensionality and increasing the classification performance. Therefore, all digraphs have 18 static nodes.

The features were divided into two groups. The first group corresponds to the features associated to the weight matrix W . Each element in W , w_{ij} , express the weight of the connection from vertex i to vertex j , indicating the frequency that the rhythmic notations was followed by each other in the sample. As a result, the weight matrix W of each sample has 18 rows and 18 columns and after that is reshaped by a 1 x 324 feature vector. The final feature matrix, namely here as Matrix A, has 280 rows, and 324 columns (the attributes). The second group involves three quantities directly extracted from the digraphs: total vertex degree, strength and clustering coefficient. These features were computed for each one of the vertices, and then the entropy, media, maximum value, standard deviation and variance were taken to compose the feature vector. Hence, the second feature matrix, Matrix B, has 280 rows and 15 columns.

The classification performance can be improved by normalizing the features (zero mean and unit standard deviation). Principal Components Analysis was applied for features analysis and redundancy removing [10]. This approach uses the eigenvalues and eigenvectors of the covariance matrix to apply geometric transformations to the original feature space, creating new orthogonal uncorrelated features [10].

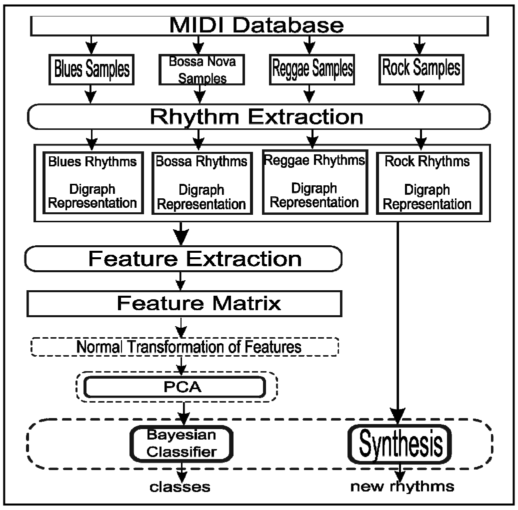


Figure 2. Block diagram of the proposed methodology.

D. *Classification and Synthesis of Rhythms*

Supervised classification is applied by the Bayesian classifier through discriminant functions. The Bayesian classifier is based on the Bayesian Decision Theory. It combines probabilities densities (likelihood) and prior probabilities (knowledge *a priori*) to assign each object to the class with maximum a posteriori probability [10].

We considered all objects from each class as the training set (in order to estimate the parameters) and all objects as the testing set. Cohen Kappa Coefficient [11] was adopted as a quantitative criterion to better evaluate the performance of the supervised classification. The kappa coefficient has good statistical properties (e.g. be asymptotically normal) and can be directly computed from the confusion matrix [12].

It is also proposed an approach, based on random walks, to compose new rhythms using the previously described digraphs. The procedure works as follows: first, an initial vertex is selected; the next vertex is randomly select among all the outgoing vertices of the initial one, but considering the probability of the edges, i.e. as being proportional to its weight. This procedure continues until a given number of steps have been performed. The random walk was applied in digraphs of artworks belonging to blues, *bossa-nova*, reggae and rock music genres (artwork-based rhythms), as well as in the four digraphs built by genre (genre-based rhythms). Figure 2 summarizes the proposed methodology.

III. RESULTS

A. *Classification Results*

Reducing data dimensionality is often a fundamental step in classification tasks. It is known that, in a high dimensional space, the classes can be easily separated, but it increases the complexity of the problem, complicating the analysis of the extracted features. For Matrix A, 15 principal components preserved 76% of the data variance, reducing the data dimensionality from 364-D to 15-D without significant loss of information. For Matrix B, the data was even more correlated, in the sense that only one principal component were sufficient to retain more than 75% of the data variance. However, it was previously verified that the number of required components to achieve suitable results depends on the classifier and on the classification task [7].

Table I presents the classification results obtained by the Bayesian Classifier (using PCA), in terms of kappa, the accuracy of the classification and the respective performance according to the value of kappa. The confusion matrices for the classification with feature Matrix A and for the classification with the combination of Matrix A and B are illustrated in Table II.

TABLE I. PCA KAPPA AND ACCURACIES

Features	*Kappa*	*Accuracy*	*Performance*
Feature Matrix A (Reduced from 324-D to 106-D[a])	0.64	73.21%	Substantial
Feature Matrix B (15-D)	0.49	61.78%	Moderate
Matrix A + Matrix B (Reduced from 339-D to 52-D[a])	0.81	85.72%	Almost Perfect

a. Quantity of components (attributes) to obtain the highest value of kappa.

TABLE II. CONFUSION MATRICES FOR THE BAYESIAN CLASSIFIER USING MATRIX A AND MATRIX A+ MATRIX B

	Matrix A				*Matrix A + Matrix B*			
	Blues	*Bossa*	*Reggae*	*Rock*	*Blues*	*Bossa*	*Reggae*	*Rock*
Blues	57	1	5	3	64	0	0	0
Bossa	2	46	5	5	0	56	0	3
Reggae	7	17	59	19	6	14	70	17
Rock	4	6	1	43	0	0	0	50

The PCA technique also helps to identify which features most contribute to the classification by verifying which elements in the first eigenvectors have higher values. For Matrix A, a detailed analysis of the obtained features can be found in [7]. The order of the six most significant features of Matrix B were: the variance and the maximum value of the clustering coefficient; the variance and the maximum value of the total vertex degree; and the standard deviation and media of the clustering coefficient.

B. *Synthesis Results*

Some synthesis results are presented in Fig. 3 and Fig. 4. We generated new rhythms considering the genre digraphs and individual samples. Fig. 3 (a) shows the rock digraph formed through all 70 artworks of rock, that is, the digraph was built from a unique weight matrix of all rock samples. Then, we applied the random walk method over this rock digraph. Fig. 3 (b) shows the resulting rock digraph obtained through the weight matrix of the composed rhythmic notations. Fig. 3 (c) presents the first fifty new rhythmic notations generated by the method. Fig. 3(d)-(f) illustrate results from the same approach, but with respect to the reggae genre. Fig. 4 (a) shows the digraph of the music *Is this Love?* by Bob Marley. Fig. 4 (b) presents the obtained digraph by the synthesis. The first fifty rhythmic notations of the original music is illustrated in Fig. 4 (c) and the first fifty rhythmic notations that form the new composed rhythmic sequence are presented in Fig. 4 (d).

IV. CONCLUSIONS

Music genres have been intensively used to organize music collections. In a general point of view, they can summarize shared characteristics of musical pieces, which is a suitable property for classification purposes. Therefore, their automatic classification has become an important topic in music research. However, music genres are not a clear concept, making the development of a definitive taxonomy a nontrivial task. In order to complement a previous study, automatic genre classification is explored in this paper taking into account the rhythms of pieces from four genres: blues, *bossa-nova*, reggae and rock. These rhythms were extracted from MIDI files and modeled as digraphs according to the respective matrix of probability transition, comprising the feature matrix together with measures obtained directly from the digraphs. PCA was applied for feature analysis and the classification was performed by the Bayesian classifier. New rhythms were composed through random walks in the digraphs.

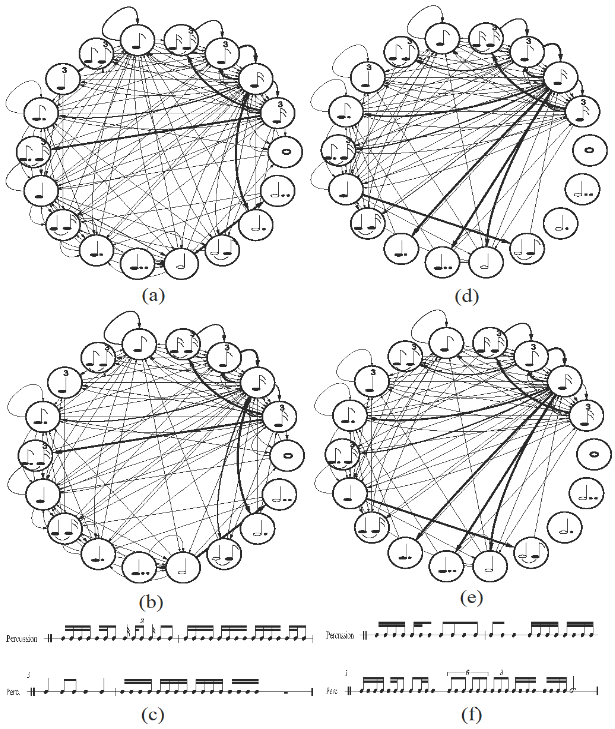


Figure 3. Synthesis results: (a) the rock and (d) reggae digraphs; (b) the respective rock and (e) reggae digraphs generated by the synthesis; (c) new composed rhythmic sequence for rock; (f) new composed rhythmic sequence for reggae.

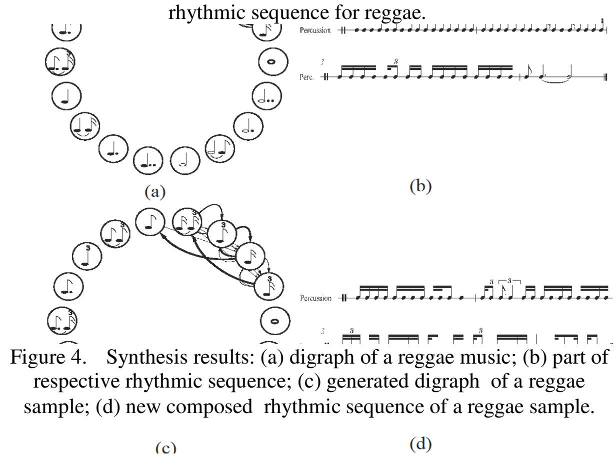


Figure 4. Synthesis results: (a) digraph of a reggae music; (b) part of respective rhythmic sequence; (c) generated digraph of a reggae sample; (d) new composed rhythmic sequence of a reggae sample.

Considering that the problem of automatic genre classification is a nontrivial task, that only one aspect of the rhythm has been studied (the occurrence of rhythmic notations), and that PCA extracts features in a unsupervised manner, the results presented in Table I and II substantiate the viability of the proposed framework. We observed that the classes are completely overlapped and that the dimensionality of the problem is high, that is, the rhythms are very complex and require many dimensions (features) to separate them.

Complementing the previous study [7], we verified that 15 attributes extracted directly from the digraphs (Matrix B) led to considerable accuracies in a lower dimensional space. In addition, we improved the classification rates by combining the two groups of measurements. According to the statistical significance of kappa, these results are different.

For the synthesis approach, we obtained particularly promising results. The quantity of runs was specified by the quantity of rhythmic notations over all samples of each genre (in the case of genre-based rhythms) or by the quantity of rhythmic notations of the specific artwork (artwork-based). Original digraphs and generated ones look similar, but with variations, which are expected when composing new rhythms. This implies that the new composed sequences of rhythms, although similar, contain different passages characteristic of their genres.

Some future works include the extraction of more measurements from the rhythms and the combination of different classifiers.

ACKNOWLEDGMENTS

We thank FAPESP for Debora C Correa (2009/50142-0) financial support and FAPESP (05/00587-5) and CNPq (301303/06-1 and 573583/2008-0) for Luciano da F. Costa financial support.

REFERENCES

[1] N. Scaringella, G. Zoia, and D. Mlynek. “Automatic genre classification of music content - a survey”. IEEE Signal Processing Magazine, pp. 133-141, 2006.

[2] F. Pachet and D. Cazaly. “A taxonomy of musical genres.” Proc. Content-Based Multimedia Information Acess (RIAO), 2000.

[3] T. Li, M. Ogihara, B. Shao, and D. Wang. “Theory and Novel Applications of Machine Learning.” Chapter: Machine Learning Approaches for Music Information Retrieval, pp. 259-278. I-Tech, Vienna, Austria, 2009.

[4] M. M. Mostafa and N. Billor. “Recognition of western style musical genres using machine learning techniques.” Expert Systems with Applications, vol. 36, pp.11378-11389, 2009.

[5] Md. Akhtaruzzaman. “Representation of musical rhythm and its classification system based on mathematical and geometrical analysis.” Proc. of the Int. Conference on Computer and Communication Engineering, pp. 466-471, 2008.

[6] F. Gouyon, S. Dixon.“A review of automatic rhythm description system.” Computer Music Journal, vol.29, pp.34-54, 2005.

[7] D. C. Correa, L. da F. Costa, and J. H. Saito. “Musical genres: beating to the rhythms of different drums.” (http://arxiv.org/abs/0911.3842) unpublished.

[8] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. “Characterization of complex networks: A survey of measurements”. Advances in Physics., vol. 56,pp.167-242, 2007.

[9] T. Eerola and P.I. Toiviainen. MIDI Toolbox: Matlab Tools for Music Research. (https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/).

[10] A. R. Webb. Statistical Pattern Recognition. John Wiley & Sons Ltd, 2002.

[11] J. Cohen. “A coefficient of agreement for nominal scales.” Educ. and Psychological Measurement, vol. 20, n.1, pp.37-46, 1960.

[12] R. G Congalton. “A review of assessing the accuracy of classifications of remotely sensed data.” Remote Sensing of Enviroment, vol. 37, pp.35-46, 1991.