



OPEN ACCESS

Musical genres: beating to the rhythms of different drums

To cite this article: Debora C Correa *et al* 2010 *New J. Phys.* **12** 053030

View the [article online](#) for updates and enhancements.

You may also like

- [Musical Instrument Recognition using Mel-Frequency Cepstral Coefficients and Learning Vector Quantization](#)
I Maliki and Sofiyanudin
- [Research on Architecture for Long-tailed Genre Computer Intelligent Classification with Music Information Retrieval and Deep Learning](#)
Gang Sun
- [Simulation and Analysis of ANFIS \(Adaptive Neuro-Fuzzy Inference System\) for Music Genre](#)
A H Pratama, W Dwiono and M T Tamam

Musical genres: beating to the rhythms of different drums

Debora C Correa^{1,4}, Jose H Saito² and Luciano da F Costa^{1,3,4}

¹ Instituto de Fisica de Sao Carlos - Universidade de Sao Paulo,
Av. Trabalhador Sao Carlense 400, Caixa Postal 369, CEP 13560-970,
Sao Carlos, Sao Paulo, Brazil

² Departamento de Computacao—Universidade Federal de Sao Carlos,
Rodovia Washington Luis, km 235, SP-310, CEP 13565-905, Sao Carlos,
Sao Paulo, Brazil

³ National Institute of Science and Technology for Complex Systems,
24210-346 Niterói, RJ, Brazil

E-mail: deboracorreia@ursa.ifsc.usp.br and luciano.ifsc.usp.br

New Journal of Physics **12** (2010) 053030 (37pp)

Received 4 December 2009

Published 20 May 2010

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/12/5/053030

Abstract. Online music databases have increased significantly as a consequence of the rapid growth of the Internet and digital audio, requiring the development of faster and more efficient tools for music content analysis. Musical genres are widely used to organize music collections. In this paper, the problem of automatic single and multi-label music genre classification is addressed by exploring rhythm-based features obtained from a respective complex network representation. A Markov model is built in order to analyse the temporal sequence of rhythmic notation events. Feature analysis is performed by using two multivariate statistical approaches: principal components analysis (unsupervised) and linear discriminant analysis (supervised). Similarly, two classifiers are applied in order to identify the category of rhythms: parametric Bayesian classifier under the Gaussian hypothesis (supervised) and agglomerative hierarchical clustering (unsupervised). Qualitative results obtained by using the kappa coefficient and the obtained clusters corroborated the effectiveness of the proposed method.

⁴ Authors to whom any correspondence should be addressed.

Contents

| | |
|--|-----------|
| 1. Introduction | 2 |
| 2. Materials and methods | 4 |
| 2.1. System representation by complex networks | 4 |
| 2.2. Data description | 6 |
| 2.3. Feature extraction | 6 |
| 2.4. Feature analysis and redundancy removal | 8 |
| 2.5. Classification methodology | 9 |
| 3. Results and discussion | 11 |
| 3.1. Single classification results | 13 |
| 3.2. Multi-label classification results | 23 |
| 4. Concluding remarks | 25 |
| Acknowledgments | 31 |
| Appendix A. Multivariate statistical methods | 31 |
| Appendix B. Linear and quadratic discriminant functions | 32 |
| Appendix C. Agglomerative hierarchical clustering | 33 |
| Appendix D. The kappa coefficient | 34 |
| References | 35 |

1. Introduction

Musical databases have increased in number and size continuously, paving the way to large amounts of online music data, including discographies, biographies and lyrics. This happened mainly as a consequence of musical publishing being absorbed by the Internet, as well as the restoration of existing analogue archives and the advancements of web technologies. As a consequence, more and more reliable and faster tools for music content analysis, retrieval and description are required, catering for browsing, interactive access and music content-based queries. Even more promising, these tools, together with the respective online music databases, have opened new perspectives for basic investigations in the field of music.

Within this context, music genres provide particularly meaningful descriptors given that they have been extensively used for years to organize music collections. When a musical piece becomes associated with a genre, users can retrieve what they are searching in a much faster manner. It is interesting to note that these new possibilities of research in music can complement what is known about the trajectories of music genres, their history and their dynamics [1]. In an ethnographic manner, music genres are particularly important because they express the general identity of the cultural foundations in which they are incorporated [2]. Music genres are part of a complex interplay of cultures, artists and market strategies to define associations between musicians and their works, making the organization of music collections easier [3]. Therefore, musical genres are of great interest because they can summarize some shared characteristics in music pieces. As indicated in [4], music genre is probably the most common description of music content, and its classification represents an appealing topic in music information retrieval (MIR) research.

Despite their ample use, music genres are not a clearly defined concept, and their boundaries remain fuzzy [3]. As a consequence, the development of such a taxonomy is controversial and redundant, representing a challenging problem. Pachet and Cazaly [5] demonstrated that there is no general agreement on musical genre taxonomies, which can depend on cultural references. Even widely used terms such as *rock*, *jazz*, *blues* and *pop* are not clear and firmly defined. According to Scaringella *et al* [3], it is necessary to keep in mind what kind of music item is being analysed in genre classification: a song, an album or an artist. While the most natural choice would be a song, it is sometimes questionable to classify one song into only one genre. Depending on the characteristics, a song can be classified into various genres. This happens more intensively with albums and artists, since, nowadays, albums contain heterogeneous material and the majority of artists tend to cover an ample range of genres during their careers. Therefore, it is difficult to associate an album or an artist with a specific genre. Pachet and Cazaly [5] also mention that the semantic confusion existing in the taxonomies can cause redundancies that probably will not be confused by human users, but may hardly be dealt with by automatic systems, so that automatic analysis of the musical databases becomes essential. However, all these critical issues emphasize that the problem of automatic classification of musical genres is a nontrivial task. As a result, only local conclusions about genre taxonomy are considered [5].

Like other problems involving pattern recognition, the process of automatic classification of musical genres can usually be divided into the following three main steps: representation, feature extraction and the classifier design [6, 7]. Music information can be described by symbolic representation or based on acoustic signals [8]. The former is a high-level kind of representation through music scores, such as MIDI, where each note is described in terms of pitch, duration, start and end times, and strength. Acoustic signals representation is obtained by sampling the sound waveform. Once the audio signals are represented in the computer, the objective becomes to extract relevant features in order to improve the classification accuracy. In the case of music, features may belong to its main dimensions including melody, timbre, rhythm and harmony.

After extracting significant features, any classification scheme may be used. There are many previous works concerning automatic genre classification in the literature [8]–[19].

An innovative approach to automatic genre classification is proposed in the current work, in which musical features refer to the temporal aspects of the songs: the rhythm. Thus, we propose to identify the genres in terms of their rhythmic patterns. While there is no clear definition of rhythm [3], it is possible to relate it to the idea of temporal regularity. More generally speaking, rhythm can be simply understood as a specific pattern produced by notes differing in duration, pause and stress. Hence, it is simpler to obtain and manipulate rhythm than the whole melodic content. However, despite its simplicity, the rhythm is genuine and intuitively characteristic and intrinsic to musical genres, since, for example, it can be used to distinguish between rock music and rhythmically more complex music, such as salsa. In addition, the rhythm is largely independent of the instrumentation and interpretation.

A few related works that use rhythm as features in automatic genre recognition can be found in the literature. The work of Akhtaruzzaman [20], in which rhythm is analysed in terms of mathematical and geometrical properties, and then fed to a system for the classification of rhythms from different regions. Karydis [21] proposed to classify intra-classical genres with note pitch and duration features, obtained from their histograms. Reviews of existing automatic rhythm description systems are presented in [22, 23]. The authors state that despite

the consensus on some rhythm concepts, there is not a single representation of rhythm that would be applicable for different applications, such as tempo and meter induction, beat tracking, quantization of rhythm and so on. They also analysed the relevance of these descriptors by measuring their performance in genre classification experiments. It has been observed that many of these approaches lack comprehensiveness because of the relatively limited rhythm representations that have been adopted [3].

Nowadays, multi-label classification methods are increasingly required by applications such as text categorization [24], scene classification [25], protein classification [26] and music categorization in terms of emotion [27], among others. The possibility of multigenres classification is particularly promising and probably closer to the human experience.

In the current study, an objective and systematic analysis of rhythm is provided according to single- and multi-label classifications. Our main motivation is to study similar and different characteristics of rhythms in terms of the occurrence of sequences of events obtained from rhythmic notations. First, the rhythm is extracted from MIDI databases and represented as graphs or networks [28, 29]. More specifically, each type of note (regarding their duration) is represented as a node, while the sequences of notes define the links between the nodes. Matrices of probability transition are extracted from these graphs and used to build a Markov model of the respective musical piece. Since they are capable of systematically modelling the dynamics and dependences between elements and subelements [30], Markov models are frequently used in temporal pattern recognition applications, such as handwriting, speech and music [31]. Supervised and unsupervised approaches are then applied, which receive as input the properties of the transition matrices and produce as output the most likely genre. Supervised classification is performed with the Bayesian classifier. For the unsupervised approach, a taxonomy of rhythms is obtained through hierarchical clustering. The described methodology is applied to four genres: blues, bossa nova, reggae and rock, which are well-known genres representing different tendencies. A series of interesting findings are reported, including the ability of the proposed framework to correctly identify the musical genres of specific musical pieces from the respective rhythmic information. We complement the findings and results for the single classification with an approach to perform multi-label classification, assigning a piece to more than one genre when appropriate.

This paper is organized as follows: section 2 describes the methodology, including the classification methods; section 3 presents the obtained results and a discussion of them; and section 4 contains the concluding remarks and directions for future work.

2. Materials and methods

Some basic concepts of complex networks as well as the proposed methodology are presented in this section.

2.1. System representation by complex networks

A complex network is a graph exhibiting an intricate structure compared to regular and uniformly random structures. There are four main types of complex networks: weighted and unweighted digraphs and weighted and unweighted graphs. The operations of symmetry and thresholding can be used to transform a digraph into a graph and a weighted graph

Table 1. Graphs and digraphs basic concepts.

| | Graphs | Digraphs |
|----------------|--|--|
| Adjacency | Two vertices i and j are adjacent or neighbours if $a_{ij} = 1$ | Concepts of predecessors and successors. If $a_{ij} \neq 0$, i is the predecessor of j and j is the successor of i . Predecessors and successors as adjacent vertices. |
| Neighbourhood | Represented by $v(i)$, meaning the set of vertices that are neighbours to vertex i . | Also represented by $v(i)$. |
| Vertex degree | Represented by k_i , gives the number of connected edges to vertex i . It is computed as $k_i = \sum_j a_{ij} = \sum_j a_{ji}$. | There are two kinds of degrees: in-degree k_i^{in} indicating the number of incoming edges; and out-degree k_i^{out} indicating the number of outgoing edges: $k_i^{\text{in}} = \sum_j a_{ji}$ $k_i^{\text{out}} = \sum_j a_{ij}$ The total degree is defined as $k_i = k_i^{\text{in}} + k_i^{\text{out}}$ |
| Average degree | Average of k_i considering all network vertices. $\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{1}{N} \sum_{ij} a_{ij}$. | It is the same for in- and out- degrees. $\langle k^{\text{out}} \rangle = \langle k^{\text{in}} \rangle = \frac{1}{N} \sum_{ij} a_{ij}$. |

(or weighted digraph) into an unweighted one, respectively [32]. A weighted digraph (or weighted direct graph) G can be defined by the following elements:

- Vertices (or nodes). Each vertex is represented by an integer number $i = 1, 2, \dots, N$; $N(G)$ is the vertex set of digraph G , and N indicates the total number of vertices ($|N(G)|$).
- Edges (or links). Each edge has the form (i, j) indicating a connection from vertex i to vertex j . The edge set of digraph G is represented by $\epsilon(G)$, and M is the total number of edges.
- The mapping $\omega: \epsilon(G) \mapsto R$, where R is the set of weight values. Each edge (i, j) has a weight $\omega(i, j)$ associated with it. This mapping does not exist in unweighted digraphs.

Undirected graphs (weighted or unweighted) are characterized by the fact that their edges have no orientation. Therefore, an edge (i, j) in such a graph necessarily implies a connection from vertex i to vertex j and from vertex j to vertex i . A weighted digraph can be represented in terms of its weight matrices W . Each element of W , w_{ji} , associates a weight with the connection from vertex i to vertex j . Table 1 summarizes some fundamental concepts about graphs and digraphs [32].

For weighted networks, a quantity called strength of vertex i is used to express the total sum of weights associated with each node. More specifically, it corresponds to the sum of the weights of the respective incoming edges ($s_i^{\text{in}} = \sum_j w_{ji}$) (in-strength) or outgoing edges ($s_i^{\text{out}} = \sum_j w_{ij}$) (out-strength) of vertex i .

Another interesting measurement of local connectivity is the clustering coefficient. This feature reflects the cyclic structure of networks, i.e. if they have a tendency to form sets of densely connected vertices. For digraphs, one way to calculate the clustering coefficient is: let m_i be the number of neighbours of vertex i and l_i be the number of connections between the neighbours of vertex i ; the clustering coefficient is obtained as $cc(i) = l_i / m_i(m_i - 1)$.

2.2. Data description

In this work, four music genres were selected: blues, bossa nova, reggae and rock. These genres are well known and represent distinct major tendencies. Music samples belonging to these genres are available in many collections in the Internet, so it was possible to select 100 samples to represent each one of them. These samples were downloaded in MIDI format. This, event-like format contains instructions (such as notes, instruments, timbres and rhythms, among others), which are used by a synthesizer during the creation of new musical events [33]. The MIDI format can be considered a digital musical score in which the instruments are separated into voices. Each sample in the dataset belongs to a single genre, since, in general, most MIDI databases found in the Internet are still single labelled.

In order to edit and analyse the MIDI scores, we applied the software for music composition and notation called Sibelius (<http://www.sibelius.com>). For each sample, the voice related to the percussion was extracted. The percussion is inherently suitable to express the rhythm of a piece. Once the rhythm is extracted, it becomes possible to analyse all the involved elements. The MIDI Toolbox for Matlab was used in the present work [34]. This toolbox is free and contains functions to analyse and visualize MIDI files in the Matlab computing environment. When a MIDI file is read with this toolbox, a matrix representation of note events is created. The columns in this matrix refer to many types of information, such as: onset (in beats), duration (in beats), MIDI channel, MIDI pitch, velocity, onset (in seconds) and duration (in seconds). The rows refer to the individual note events, that is, each note is described in terms of its duration, pitch and so on.

Only the note duration (in beats) has been used in the current work. In fact, the durations of the notes, respecting the sequence in which they occur in the sample, are used to create a digraph. Each vertex of this digraph represents one possible rhythm notation, such as quarter note, half note, eighth note and so on. The edges reflect the subsequent pairs of notes. For example, if there is an edge from vertex i , represented by a quarter note, to a vertex j , represented by an eighth note, this means that a quarter note was followed by an eighth note at least once. The thicker the edges, the larger is the strength between these two nodes. Examples of these digraphs are shown in figure 1. Figure 1(a) depicts a blues sample represented by the music *How Blue Can You Get* by BB King. A bossa nova sample, namely the music *Fotografia* by Tom Jobim, is illustrated in 1(b). Figure 1(c) illustrates a reggae sample, represented by the music *Is This Love* by Bob Marley. Finally, figure 1(d) shows a rock sample, corresponding to the music *From Me to You* by the Beatles.

2.3. Feature extraction

Extracting features is the first step of most pattern recognition systems. Each pattern is represented by its d features or attributes in terms of a vector in a d -dimensional space. In a discrimination problem, the goal is to choose features that allow the pattern vectors belonging

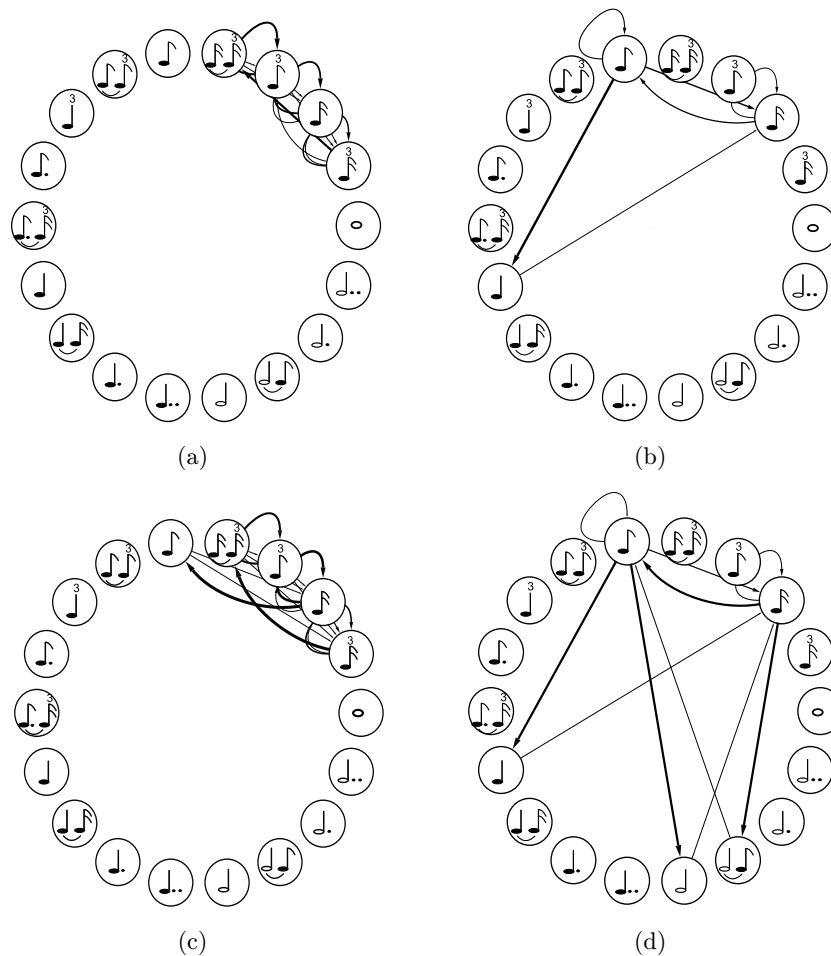


Figure 1. Digraph examples of four music samples: (a) *How Blue Can You Get* by BB King. (b) *fotografia* by Tom Jobim. (c) *Is This Love* by Bob Marley. (d) *From Me to You* by The Beatles.

to different classes to occupy compact and distinct regions in the feature space, maximizing class separability. After extracting significant features, any classification scheme may be used. In the case of music, features may belong to the main dimensions of music including melody, timbre, rhythm and harmony.

Therefore, one of the main features of this work is to extract features from the digraphs and use them to analyse the complexities of the rhythms, as well as to perform classification tasks. For each sample, a digraph is created as described in the previous section. All digraphs have 18 nodes, corresponding to the quantity of rhythm notation possibilities concerning all the samples, after excluding those that hardly ever happen. This exclusion was important in order to provide an appropriate visual analysis and to better fit the features. In fact, avoiding features that do not significantly contribute to the analysis reduces data dimensionality, improves the classification performance through a more stable representation and removes redundant or irrelevant information (in this case, minimizes the occurrence of null values in the data matrix).

The features are associated with the weight matrix W . As commented on in section 2.1, each element in W , w_{ij} , indicates the weight of the connection from vertex j to i or, in other

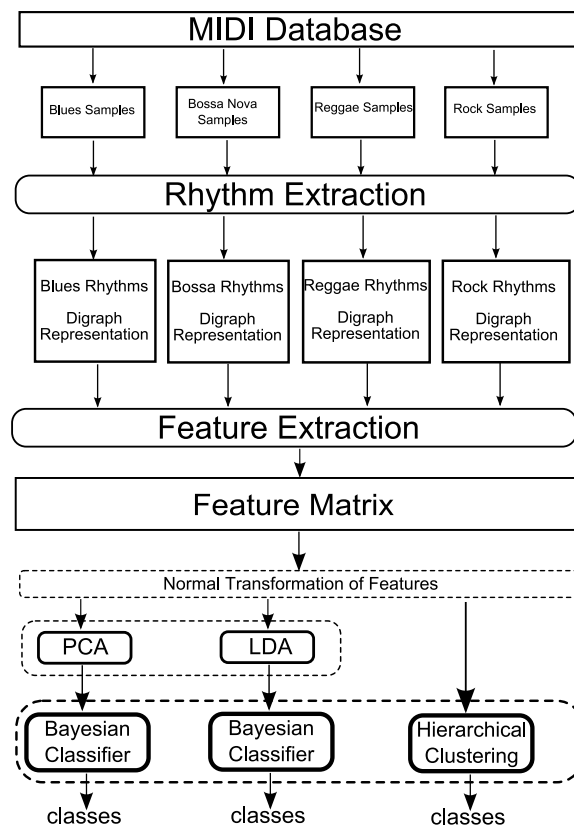


Figure 2. Block diagram of the proposed methodology.

words, they are meant to represent how often the rhythm notations follow one another in the sample. The weight matrix W has 18 rows and 18 columns. The matrix W is reshaped by a 1×324 feature vector. This is done for each one of the genre samples. However, it was observed that some samples, even belonging to different genres, generated exactly the same weight matrix. These samples were excluded. Thereby, the feature matrix has 280 rows (all non-excluded samples) and 324 columns (the attributes).

An overview of the proposed methodology is illustrated in figure 2. After extracting the features, a standardization transformation is done to guarantee that the new feature set has zero mean and unit standard deviation. This procedure can significantly improve the resulting classification. Once the normalized features are available, the structure of the extracted rhythms can be analysed by using two different approaches for features analysis: principal components analysis (PCA) and linear discriminant analysis (LDA). We also compare two types of classification methods: Bayesian classifier (supervised) and hierarchical clustering (unsupervised). PCA and LDA are described in section 2.4 and the classification methods are described in section 2.5.

2.4. Feature analysis and redundancy removal

Two techniques are widely used for feature analysis [7, 35, 36]: PCA and LDA. Basically, these approaches apply geometric transformations (rotations) to the feature space with the purpose of generating new features based on linear combinations of the original ones, aiming

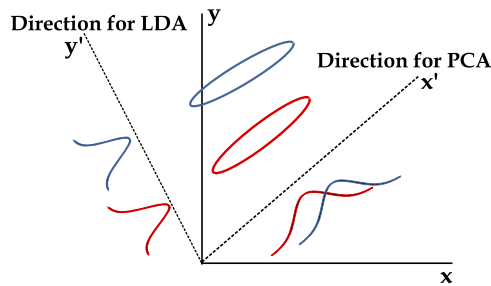


Figure 3. An illustration of PCA (optimizing representation) and LDA projections (optimizing classification), adapted from [37].

at dimensionality reduction (in the case of PCA) or to seek a projection that best separates the data (in the case of LDA). Figure 3 illustrates the basic principles underlying PCA and LDA. The direction x' (obtained with PCA) is the best one to represent the two classes with maximum overall dispersion. However, it can be observed that the densities projected along direction x' overlap one another, making these two classes inseparable. In contrast, if direction y' , obtained with LDA, is chosen, the classes can be easily separated. Therefore, it is said that the directions for representation are not always also the best choice for classification, reflecting the different objectives of PCA and LDA [37]. Appendices A.1 and A.2 give more details of these two techniques.

2.5. Classification methodology

Basically, to classify means to assign objects to classes or categories according to the properties they present. In this context, the objects are represented by attributes, or feature vectors. There are three main types of pattern classification tasks: *imposed criteria*, *supervised* classification, and *unsupervised* classification. Imposed criteria are the easiest situation in classification, once the classification criteria is clearly defined, generally by a specific practical problem. If the classes are known in advance, the classification is said to be supervised classification or by example classification, since usually examples (the training set) are available for each class. Generally, supervised classification involves two stages: *learning*, in which the features are tested in the training set; and *application*, when new entities are presented to the trained system. There are many approaches involving supervised classification. The current study applies the Bayesian classifier through discriminant functions (appendix B) in order to perform supervised classification. The Bayesian classifier is based on the Bayesian decision theory and combines class conditional probability densities (likelihood) and prior probabilities (prior knowledge) to perform classification by assigning each object to the class with the maximum *a posteriori* probability.

In unsupervised classification, the classes are not known in advance, and there is not a training set. This type of classification is usually called *clustering*, in which the objects are agglomerated according to some similarity criterion. The basic principle is to form classes or *clusters* so that the similarity between the objects in each class is maximized and the similarity between objects in different classes is minimized. There are two types of clustering: partitional (also called non-hierarchical) and hierarchical. In the former, a fixed number of clusters is

obtained as a single partition of the feature space. Hierarchical clustering procedures are more commonly used because they are usually simpler [6, 7, 36, 38]. The main difference is that instead of one definite partition, a series of partitions are taken, which is done progressively. If the hierarchical clustering is *agglomerative* (also known as bottom-up), the procedure starts with N objects as N clusters and then successively merges the clusters until all the objects are joined into a single cluster (please refer to appendix C for more details of agglomerative hierarchical clustering). *Divisive* hierarchical clustering (top-down) starts with all the objects as a single cluster and splits it into progressively finer subclusters.

Because of the difficulty of defining a clear definition of music genre taxonomy and considering the critical issues regarding the redundancies and fuzzy boundaries discussed in the introduction, we also attempt to classify the pieces using a multi-label classification approach. For the supervised classification, instead of assigning an object only to the class with the maximum *a posteriori* probability, we proceeded as follows:

1. First, we normalized the maximum *a posteriori* probabilities of each class, making their sum equal to one. Then, they were sorted in increasing order.
2. After that, given a sample, we computed the distance between its maximum *a posteriori* probabilities and each one of the vectors: one-label = [1 0 0 0], two-label = [0.5 0.5 0 0]; three-label = [0.33 0.33 0.33 0]; and four-label = [0.25 0.25 0.25 0.25]. As an example, consider sample 1 from the class blues and its maximum *a posteriori* probabilities: 0.9521156, given the class blues; 0.0108921, given the class bossa nova; 0.0162640, given the class reggae; and 0.0207283, given the class rock. When computing the difference between its maximum *a posteriori* probability vector ([0.9521156, 0.0108921, 0.0162640, 0.0207283]) and the vectors described above, certainly the minor distance will be to the one-label vector, indicating that these samples should be classified only as blues. Now consider sample 5 from the class blues and its maximum *a posteriori* probabilities: 0.4097450 for the class blues; 0.0079218 for the class bossa nova; 0.5501617 for the class reggae; and 0.0321714 for the class rock. After sorting, the maximum *a posteriori* probability vector is [0.5501617, 0.4097450, 0.0321714, 0.0079218]. The minor distance will now be to the two-label vector, indicating that this sample should be classified into two classes, namely reggae and blues.

Samples with similar maximum *a posteriori* probabilities reflect a mixture of features from more than one class. The motivation for this approach is to define a criterion in which they can be suitably labeled depending on this similarity degree. For the unsupervised case, we adopted a different approach. From the LASTFM site (<http://www.lastfm.com.br>), a known interactive web radio, we verified how the genre tags were associated with our samples, in order to set up a multi-label dataset. Actually, we entered each sample into the site and obtained its most used tags (the ones that come together with the music). If among these tags there were any of the four used in this paper (blues, bossa nova, reggae and rock), the sample would be labelled with it. For example, among the most used tags for the music *A Mess of Blues* by Elvis Presley were the tags blues (the original one) and rock. Therefore, this sample was labeled as blues and rock. This process resulted in the Venn diagram presented in figure 4.

It is interesting to note that bossa nova samples did not have any other tags (blues, reggae and rock) associated with them. Twenty blues samples had a rock tag associated with them: 14, 15, 25, 30, 31, 33, 34, 39, 50, 51, 52, 54, 58, 59, 61, 62, 63, 64 and 66. Seventeen reggae

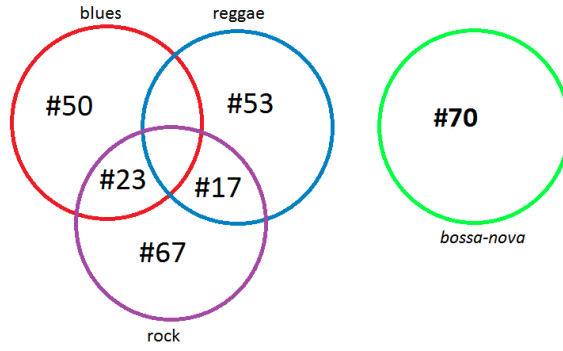


Figure 4. Venn diagram for the multi-label dataset.

samples also had a rock tag associated with them: 141, 142, 143, 159, 168, 172, 173, 174, 176, 177, 178, 195, 199, 200, 201, 202 and 203. Finally, three rock samples had a blues tag associated with them: 227, 228 and 276.

2.5.1. Performance measures for classification. To objectively evaluate the performance of the supervised classification, it is necessary to use quantitative criteria. The most used criteria are the estimated classification error and the obtained accuracy. Because of its good statistical properties, such as, for example, being asymptotically normal with well-defined expressions to estimate its variances, this study also adopted the Cohen kappa coefficient [39] as a quantitative measure to analyse the performance of the proposed method. Besides, the kappa coefficient can be directly obtained from the confusion matrix [40] (appendix D), easily computed in supervised classification problems. The confusion matrix is defined as

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1c} \\ c_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ c_{c1} & \dots & & c_{cc} \end{bmatrix}, \quad (1)$$

where each element c_{ij} represents the number of objects from class i classified as class j . Therefore, the elements in the diagonal indicate the number of correct classifications.

We computed the performance of the multi-label and single unsupervised classifications, taking into consideration the following criterion: Performance = n_i/N_i , where n_i is the number of correct classifications of class i and N_i is the number of samples of class i .

3. Results and discussion

As mentioned earlier, four musical genres were used in this study: blues, bossa nova, reggae and rock. We selected music art works from diverse artists, as presented in tables 2 and 3. Different colors were chosen to represent the genres (the color red for the genre blues, green for bossa nova, cyan for reggae and pink for rock) in order to provide a better visualization and discussion of the results.

Table 2. Blues and bossa nova art works.

| Blues art works | Bossa nova art works |
|---|---|
| 1. Albert Collins—A Good Fool | 71. Antonio Adolfo—Sa Marina |
| 2. Albert Collins—Fake ID | 72. Barquinho |
| 3. Albert King—Stormy Monday | 73. Caetano Veloso—Menino Do Rio |
| 4. Andrews Sister—Boogie Woogie Bugle Boy | 74. Caetano Veloso—Sampa |
| 5. B B King—Dont Answer The Door | 75. Celso Fonseca—Ela E Carioca |
| 6. B B King—Get Off My Back | 76. Celso Fonseca—Slow Motion bossa nova |
| 7. B B King—Good Time | 77. Chico Buarque Els Soares—Facamos |
| 8. B B King—How Blue Can You Get | 78. Chico Buarque Francis Hime—Meu Caro Amigo |
| 9. B B King—Sweet Sixteen | 79. Chico Buarque Quarteto Em Si—Roda Viva |
| 10. B B King—The Thrill Is Gone | 80. Chico Buarque—As Vitruines |
| 11. B B King—Woke Up This Morning | 81. Chico Buarque—Construcao |
| 12. Barbra Streisand—Am I Blue | 82. Chico Buarque—Desalento |
| 13. Billie Piper—Something Deep Inside | 83. Chico Buarque—Homenagem Ao Malandro |
| 14. Blues In F For Monday | 84. Chico Buarque—Mulheres De Athenas |
| 15. Bo Diddley—Bo Diddley | 85. Chico Buarque—Ole Ola |
| 16. Boy Williamson—Dont Start Me Talking | 86. Chico Buarque—Sem Fantasia |
| 17. Boy Williamson—Help Me | 87. Chico Buarque—Vai Levando |
| 18. Boy Williamson—Keep It To Yourself | 88. Dick Farney—Copacabana |
| 19. Buddy Guy—Midnight Train | 89. Elis Regina—Alo Alo Marciano |
| 20. Charlie Parker—Billies Bounce | 90. Elis Regina—Como Nossos Pais |
| 21. Count Basie—Count On The Blues | 91. Elis Regina—Na Batucada Da Vida |
| 22. Cream—Crossroads Blues | 92. Elis Regina—O Bebado E O Equilibrista |
| 23. Delmore Brothers—Blues Stay Away From Me | 93. Elis Regina—Romaria |
| 24. Elmore James—Dust My Broom | 94. Elis Regina—Velho Arvoredo |
| 25. Elvis Presley—A Mess Of Blues | 95. Emilio Santiago—Essa Fase Do Amor |
| 26. Etta James—At Last | 96. Emilio Santiago—Esta Tarde Vi Voller |
| 27. Feeling The Blues | 97. Emilio Santiago—Saigon |
| 28. Freddie King—Help Day | 98. Emilio Santiago—Ta Tudo Errado |
| 29. Freddie King—Hide Away | 99. Gal Costa—Canta Brasil |
| 30. Gary Moore—A Cold Day In Hell | 100. Gal Costa—Para Machucar Meu Coracao |
| 31. George Thorogood—Bad To The Bone | 101. Gal Costa—Pra Voce |
| 32. Howlin Wolf—Little Red Rooster | 102. Gal Costa—Um Dia De Domingo |
| 33. Janis Joplin—Piece Of My Heart | 103. Jair Rodrigues—Disparada |
| 34. Jimmie Cox—Before You Accuse Me | 104. Joao Bosco—Corsario |
| 35. Jimmy Smith—Chicken Shack | 105. Joao Bosco—De Frente Para O Crime |
| 36. John Lee Hooker—Boom Boom Boom | 106. Joao Bosco—Jade |
| 37. John Lee Hooker—Dimples | 107. Joao Bosco—Risco De Giz |
| 38. John Lee Hooker—One Bourbon One Scotch One Beer | 108. Joao Gilberto—Corcovado |
| 39. Johnny Winter—Good Morning Little School Girl | 109. Joao Gilberto—Da Cor Do Pecado |
| 40. Koko Taylor—Hey Bartender | 110. Joao Gilberto—Um Abraco No Bonfa |
| 41. Little Walter—Juke | 111. Luiz Bonfa—De Cigarro Em Cigarro |
| 42. Louis Jordan—Let Good Times Roll | 112. Luiz Bonfa—Manha De Carnaval |
| 43. Miles Davis—All Blues | 113. Marcos Valle—Preciso Aprender A Viver So |
| 44. Ray Charles—Born To The Blues | 114. Marisa Monte—Ainda Lembro |
| 45. Ray Charles—Crying Time | 115. Marisa Monte—Amor I Love You |
| 46. Ray Charles—Georgia On My Mind | 116. Marisa Monte—Ando Meio Desligado |

Table 2. Continued.

| Blues art works | Bossa nova art works |
|---|---|
| 47. Ray Charles—Hit The Road Jack | 117. Tom Jobim—Aguas De Marco |
| 48. Ray Charles—Unchain My Heart | 118. Tom Jobim—Amor Em Paz |
| 49. Robert Johnson—Dust My Broom | 119. Tom Jobim—Brigas Nunca Mais |
| 50. Stevie Ray Vaughan—Cold Shot | 120. Tom Jobim—Desafinado |
| 51. Stevie Ray Vaughan—Couldnt Stand The Weather | 121. Tom Jobim—Fotografia |
| 52. Stevie Ray Vaughan—Dirty Pool | 122. Tom Jobim—Garota De Ipanema |
| 53. Stevie Ray Vaughan—Hillbillies From Outer Space | 123. Tom Jobim—Meditacao |
| 54. Stevie Ray Vaughan—I Am Crying | 124. Tom Jobim—Samba Do Aviao |
| 55. Stevie Ray Vaughan—Lenny | 125. Tom Jobim—Se Todos Fossem Iguais A Voce |
| 56. Stevie Ray Vaughan—Little Wing | 126. Tom Jobim—Soinha De Ser Com Voce |
| 57. Stevie Ray Vaughan—Looking Out The Window | 127. Tom Jobim—Vivo Sonhando |
| 58. Stevie Ray Vaughan—Love Struck Baby | 128. Tom Jobim—Voce Abusou |
| 59. Stevie Ray Vaughan—Manic Depression | 129. Tom Jobim—Wave |
| 60. Stevie Ray Vaughan—Scuttle Buttin | 130. Toquinho—Agua Negra Da Lagoa |
| 61. Stevie Ray Vaughan—Superstition | 131. Toquinho—Ao Que Vai |
| 62. Stevie Ray Vaughan—Tell Me | 132. Toquinho—Este Seu Olhar |
| 63. Stevie Ray Vaughan—Voodoo Chile | 133. Vinicius De Moraes—Apelo |
| 64. Stevie Ray Vaughan—Wall Of Denial | 134. Vinicius De Moraes—Carta Ao Tom |
| 65. T Bone Walker—Call It Stormy Monday | 135. Vinicius De Moraes—Minha Namorada |
| 66. The Blues Brothers—Everybody Needs Somebody To Love | 136. Vinicius De Moraes—O Morro Nao Tem Vez |
| 67. The Blues Brothers—Green Onions | 137. Vinicius De Moraes—Onde Anda Voce |
| 68. The Blues Brothers—Peter Gunn Theme | 138. Vinicius De Moraes—Pela Luz Dos Olhos Teus |
| 69. The Blues Brothers—Soulman | 139. Vinicius De Moraes—Samba Em Preludio |
| 70. W C Handy—Memphis Blues | 140. Vinicius De Moraes—Tereza Da Praia |

3.1. Single classification results

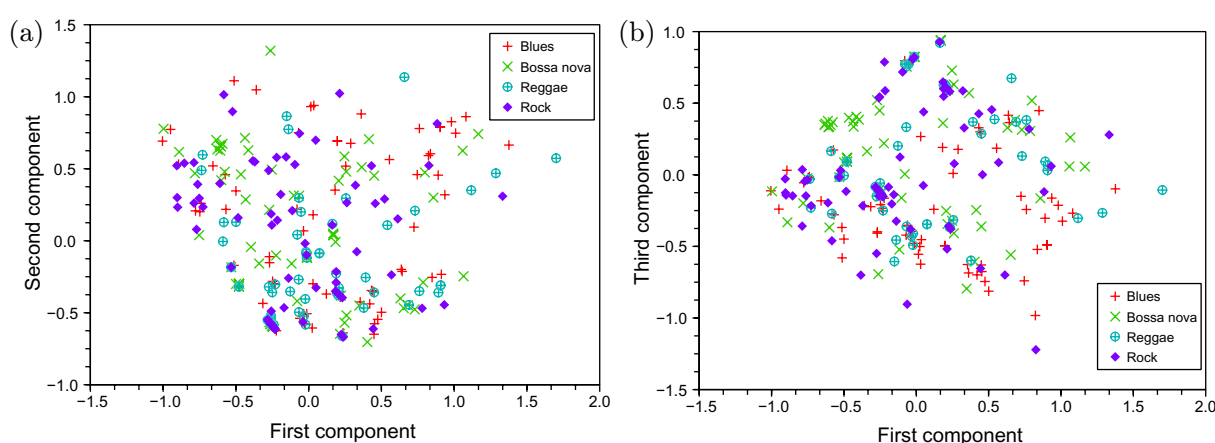
If we want to reduce data dimensionality, it is necessary to set a suitable number of principal components that will represent the new features. Not surprisingly, in a high dimensional space, the classes can be easily separated. On the other hand, high dimensionality increases complexity, making the analysis of both extracted features and classification results a difficult task. One approach to obtain the ideal number of principal components is to verify how much of the data variance is preserved. In order to do so, the l first eigenvalues (l is the quantity of principal components to be verified) are summed up and the result is divided by the sum of all the eigenvalues. If the result of this calculation is a value equal to or greater than 0.75, it is said that this number of components (or new features) preserves at least 75% of the data variance, which is often enough for classification purposes. When PCA was applied to the normalized rhythm features, as illustrated in figure 5, it was observed that 15 principal components preserved 76% of the variance of the data. That is, it is possible to reduce the data dimensionality from 364-D to 15-D without a significant loss of information. Nevertheless, as will be shown in the following, depending on the classifier and how the classification task was performed, different numbers of components were required in each situation in order to achieve suitable results. Although preserving only 36% of the variance, figure 5 shows the first three principal components, that is, the first three new features obtained with PCA. Figure 5(a) shows the first and second

Table 3. Reggae and rock art works.

| Reggae art works | Rock art works |
|--|---|
| 141. Ace Of Bass—All That She Wants | 211. Aerosmith—Kings And Queens |
| 142. Ace Of Bass—Dont Turn Around | 212. Aha—Stay On These Roads |
| 143. Ace Of Bass—Happy Nation | 213. Aha—Take On Me |
| 144. Armandinho—Pela Cor Do Teu Olho | 214. Aha—Theres Never A Forever Thing |
| 145. Armandinho—Sentimento | 215. Beatles—Cant Buy Me Love |
| 146. Big Mountain—Baby I Love Your Way | 216. Beatles—From Me To You |
| 147. Bit Mclean—Be Happy | 217. Beatles—I Want To Hold Your Hand |
| 148. Bob Marley—Africa Unite | 218. Beatles—She Loves You |
| 149. Bob Marley—Buffalo Soldier | 219. Billy Idol—Dancing With Myself |
| 150. Bob Marley—Exodus | 220. Cat Stevens—Another Saturday Night |
| 151. Bob Marley—Get Up, Stand Up | 221. Creed—My Own Prison |
| 152. Bob Marley—I Shot The Sheriff | 222. Deep Purple—Demon Eye |
| 153. Bob Marley—Iron Lion Zion | 223. Deep Purple—Hallelujah |
| 154. Bob Marley—Is This Love | 224. Deep Purple—Hush |
| 155. Bob Marley—Jammin | 225. Dire Straits—Sultan Of Swing |
| 156. Bob Marley—No Woman No Cry | 226. Dire Straits—Walk Of Life |
| 157. Bob Marley—Punky Reggae Party | 227. Duran Duran—A View To A Kill |
| 158. Bob Marley—Root, Rock, Reggae | 228. Eric Clapton—Cocaine |
| 159. Bob Marley—Satisfy my soul | 229. Europe—Carrie |
| 160. Bob Marley—Stir It Up | 230. Fleetwood Mac—Dont Stop |
| 161. Bob Marley—Three Little Birds | 231. Fleetwood Mac—Dreams |
| 162. Bob Marley—Waiting In The Van | 232. Fleetwood Mac—Gold Dust Woman |
| 163. Bob Marley—War | 233. Foo Fighters—Big Me |
| 164. Bob Marley—Wear My | 234. Foo Fighters—Break Out |
| 165. Bob Marley—Zimbabwe | 235. Foo Fighters—Walking After You |
| 166. Cidade Negra—A Cor Do Sol | 236. Men At Work—Down Under |
| 167. Cidade Negra—A Flecha E O Vulcao | 237. Men At Work—Who Can It Be Now |
| 168. Cidade Negra—A Sombra Da Maldade | 238. Metallica—Battery |
| 169. Cidade Negra—Aonde Voce Mora | 239. Metallica—Fuel |
| 170. Cidade Negra—Eu Fui Eu Fui | 240. Metallica—Hero Of The Day |
| 171. Cidade Negra—Eu Tambem Quero Beijar | 241. Metallica—Master Of Puppets |
| 172. Cidade Negra—Firmamento | 242. Metallica—My Friend of Misery |
| 173. Cidade Negra—Girassol | 243. Metallica—No Leaf Clover |
| 174. Cidade Negra—Ja Foi | 244. Metallica—One |
| 175. Cidade Negra—Mucama | 245. Metallica—Sad But True |
| 176. Cidade Negra—O Ere | 246. Pearl Jam—Alive |
| 177. Cidade Negra—Pensamento | 247. Pearl Jam—Black |
| 178. Dazaranhas—Vagabundo Confesso | 248. Pearl Jam—Jeremy |
| 179. Dont Worry | 249. Pet Shop Boys—Go West |
| 180. Flor Do Reggae | 250. Pet Shop Boys—One In A Million |
| 181. Inner Circle—Bad Boys | 251. Pink Floyd—Astronomy Domine |
| 182. Inner Circle—Sweat | 252. Pink Floyd—Have A Cigar |
| 183. Jimmy Cliff—I Can See Clearly Now | 253. Pink Floyd—Hey You |
| 184. Jimmy Cliff—Many Rivers To Cross | 254. Queen—Another One Bites The Dust |
| 185. Jimmy Cliff—Reggae Night | 255. Queen—Dont Stop Me Now |
| 186. Keep On Moving | 256. Queen—I Want It All |

Table 3. Continued.

| Reggae art works | Rock art works |
|--|---|
| 187. Manu Chao—Me Gustas Tu | 257. Queen—Play The Game |
| 188. Maskavo—Anjo Do Ceu | 258. Queen—Radio Gaga |
| 189. Maskavo—Asas | 259. Queen—Under Pressure |
| 190. Natiruts—Liberdade Pra Dentro Da Cabeca | 260. Red Hot Chili Peppers—Higher Ground |
| 191. Natiruts—Presente De Um Beija Flor | 261. Red Hot Chili Peppers—Otherside |
| 192. Natiruts—Reggae Power | 262. Red Hot Chili Peppers—Under The Bridge |
| 193. Nazarite Skank | 263. Rolling Stones—Angie |
| 194. Peter Tosh—Johnny B Goode | 264. Rolling Stones—As Tears Go By |
| 195. Shaggy—Angel | 265. Rolling Stones—Satisfaction |
| 196. Shaggy—Bombastic | 266. Rolling Stones—Street Of Love |
| 197. Shaggy—It Wasnt Me | 267. Steppenwolf—Magic Carpet Ride |
| 198. Shaggy—Strength Of A Woman | 268. Steve Winwood—Valerie |
| 199. Sublime—Badfish | 269. Steve Winwood—While You See A Chance |
| 200. Sublime—D. Js | 270. Tears For Fears—Shout |
| 201. Sublime—Santeria | 271. The Doors—Hello, I Love You |
| 202. Sublime—Wrong Way | 272. The Doors—Light My Fire |
| 203. Third World—Now That We Found Love | 273. The Doors—Love Her Madly |
| 204. Tribo De Jah—Babilonia Em Chamas | 274. U2—Elevation |
| 205. Tribo De Jah—Regueiros Guerreiros | 275. U2—Everlasting Love |
| 206. Tribo De Jah—Um So Amor | 276. U2—When Love Comes To Town |
| 207. UB 40—Bring Me Your Cup | 277. Van Halen—Dance The Night Away |
| 208. UB 40—Homely Girl | 278. Van Halen—Running With The Devil |
| 209. UB 40—To Love Somebody | 279. Van Halen—Jump |
| 210. UB 40—Red Red Wine | 280. Van Halen—Panama |

**Figure 5.** The first three new features obtained by PCA. (a) The first and second axes. (b) The first and third axes.

features and figure 5(b) shows the first and third features. It should be noted that the classes are completely overlapping, making the problem of automatic classification a nontrivial task.

In all the following supervised classification tasks, re-substitution means that all objects from each class were used as the training set (in order to estimate the parameters) and all objects

Table 4. PCA kappa coefficient and accuracy for the quadratic Bayesian classifier.

| | Kappa | Variance | Accuracy | Performance |
|-------------------|-------|----------|----------|-------------|
| Re-substitution | 0.61 | 0.0013 | 70.72% | Substantial |
| Hold-out (70–30%) | 0.33 | 0.0056 | 50% | Fair |
| Hold-out (50–50%) | 0.32 | 0.0033 | 48.57% | Fair |

Table 5. Confusion matrix for the quadratic Bayesian classifier using PCA and re-substitution.

| | Blues | Bossa nova | Reggae | Rock |
|------------|-------|------------|--------|------|
| Blues | 46 | 2 | 22 | 0 |
| Bossa nova | 1 | 45 | 23 | 1 |
| Reggae | 0 | 0 | 70 | 0 |
| Rock | 1 | 1 | 31 | 37 |

Table 6. PCA kappa and accuracy for the linear Bayesian classifier.

| | Kappa | Variance | Accuracy | Performance |
|-------------------|-------|----------|----------|-------------|
| Re-substitution | 0.64 | 0.0012 | 73.21% | Substantial |
| Hold-out (70–30%) | 0.32 | 0.0055 | 48.75% | Fair |
| Hold-out (50–50%) | 0.33 | 0.0033 | 50% | Fair |

were used as the testing set. Hold-out 70–30% means that 70% of the objects from each class were used as the training set and 30% (different ones) for testing. Finally, in hold-out 50–50%, the objects were separated into two groups: 50% for training and 50% for testing.

The kappa variance is strongly related to its accuracy, that is, how reliable are its values. The higher its variance, the lower its accuracy. If the kappa coefficient is a statistics, in general, the use of large datasets improves its accuracy by making its variance smaller. This can be observed in the results. The smaller variance occurred in the re-substitution situation, in which all samples constitute the testing set. This indicates that re-substitution provided the best kappa accuracy in the experiments. On the other hand, hold-out 70–30% provided higher kappa variance, if only 30% of the samples establish the testing set.

The results obtained by the quadratic Bayesian classifier using PCA are shown in table 4 in terms of kappa, its variance, the accuracy of the classification and the overall performance according to the value of kappa.

Table 4 also indicates that the performance was not satisfactory for the hold-out 70–30% and hold-out 50–50%. As PCA is not a supervised approach, the parameter estimation performance (of covariance matrices for instance) is strongly degraded because of the small sample size problem.

The confusion matrix for the re-substitution classification task in table 4 is illustrated in table 5. All reggae samples were classified correctly. In addition, many samples from the other classes were classified as reggae.

For comparing the two different classifiers, the results obtained by the linear Bayesian classifier using PCA are shown in table 6, again in terms of kappa, its variance, the accuracy of classification and the overall performance according to its value. The performance of the re-substitution classification task increased slightly, mainly because here a unique covariance matrix is estimated using all the samples in the dataset.

Figure 6 depicts the value of kappa depending on the quantity of principal components used in the quadratic and linear Bayesian classifiers. The last value of each graphics makes it clear that from this value onwards the classification cannot be done due to singularity problems involving the inversion of the covariance matrices (curse of dimensionality). It can be observed that this singularity threshold is different in each situation. For the quadratic classifier, this value is in the range of about 5–37 components whereas, for the linear classifier, it is in the range of about 9–106 components. The smaller quantity for the quadratic classifier can be explained by the fact that there are four covariance matrices, each one estimated from the samples for one respective class. As there are 70 samples in each class, singularity problems will occur in a smaller dimensional space (compared to the linear classifier) that uses all 280 samples to estimate one unique covariance matrix. Therefore, the ideal number of principal components allowing the highest value of kappa should be those circled in red in figure 6.

Keeping in mind that the problem of automatic genre classification is a nontrivial task, that in this study only one aspect of the rhythm has been analysed (the occurrence of rhythm notations) and that PCA is an unsupervised approach for feature extraction, the correct classifications presented in tables 4 and 6 for the re-substitution situation corroborate strongly the viability of the proposed methodology. Despite the complexity of comparing different proposed approaches to automatic genre classification discussed in the introduction, these accuracy values are very similar or even superior compared to previous works [3], [10]–[12], [15].

Similarly, figure 7 shows the three components, namely the three new features obtained with LDA. As mentioned before, the LDA approach has the restriction of obtaining only $C - 1$ nonzero eigenvalues, where C is the number of classes. Therefore, only three components are computed. If it is a supervised approach and the main goal is to maximize class separability, the four classes in figures 7(a) and (b) are clearer than those in PCA, although still involving substantial overlaps. This result corroborates that automatic classification of musical genres is not a trivial task.

Table 7 presents the results obtained by the quadratic Bayesian classifier using LDA. In contrast to PCA, the use of hold-out (70–30%) and hold-out (50–50%) provided good results, which is notable and reflects the supervised characteristic of LDA, which makes use of all the discriminant information available in the feature matrix.

Although the values of kappa and its variance obtained using LDA with re-substitution and PCA with re-substitution are similar, the two confusion matrices are slightly different from each other. However, in both cases, the misclassified art works are concentrated in one class, represented by the genre reggae. The results obtained with the LDA technique are particularly promising because they reflect the nature of the data. Although widely used, terms such as rock, reggae or pop often remain loosely defined [3]. Yet, it is worthwhile to remember that the intensity of the beat, which is a very important aspect of the rhythm, has not been considered in this work. This means that analysing rhythm only through notations, as currently proposed,

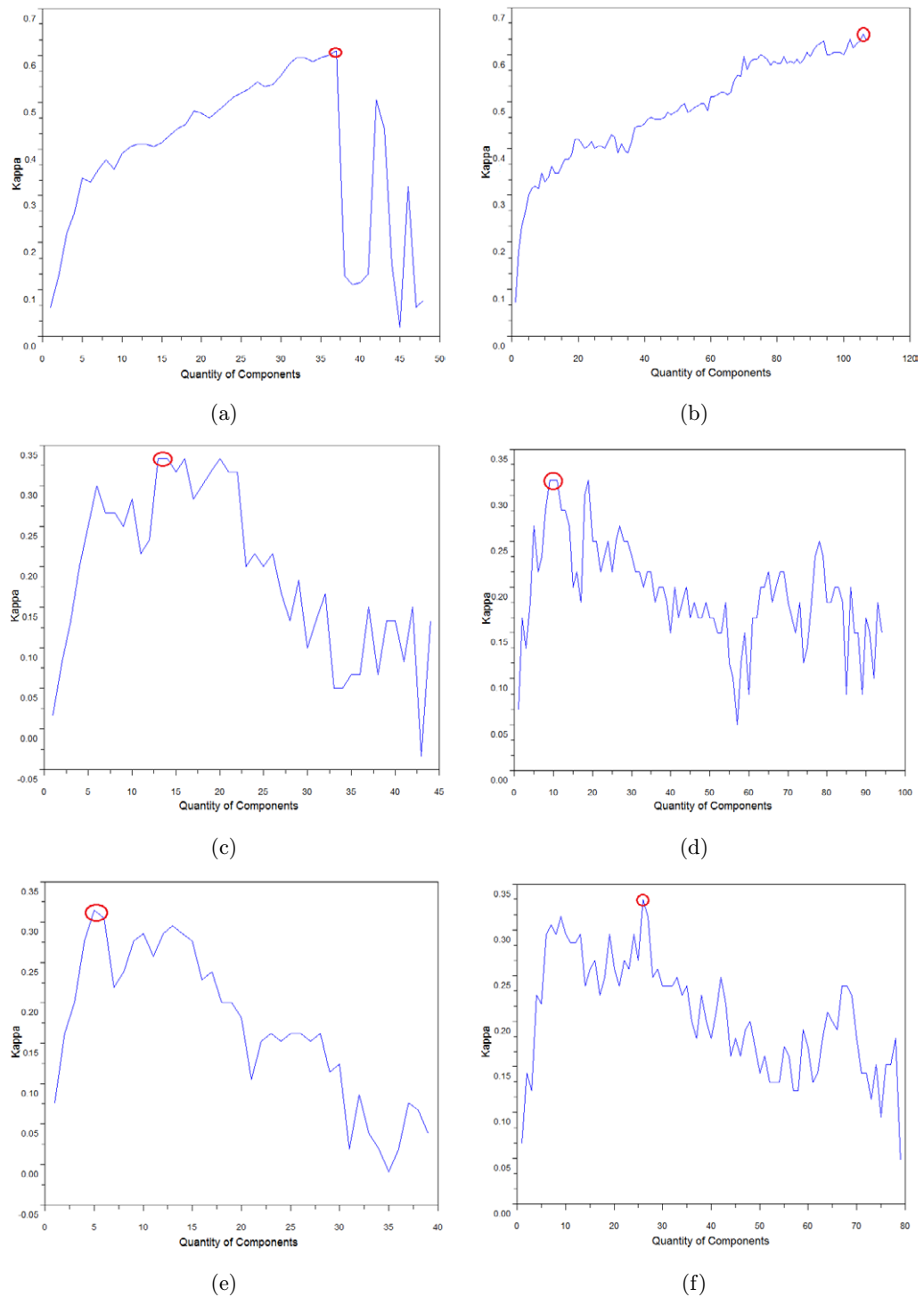


Figure 6. Values of kappa varying the number of principal components. (a) Quadratic Bayesian classifier: re-substitution. (b) Linear Bayesian classifier: re-substitution. (c) Quadratic Bayesian classifier: hold-out (70–30%). (d) Linear Bayesian classifier: hold-out (70–30%). (e) Quadratic Bayesian classifier: hold-out (50–50%). (f) Linear Bayesian classifier: hold-out (50–50%).

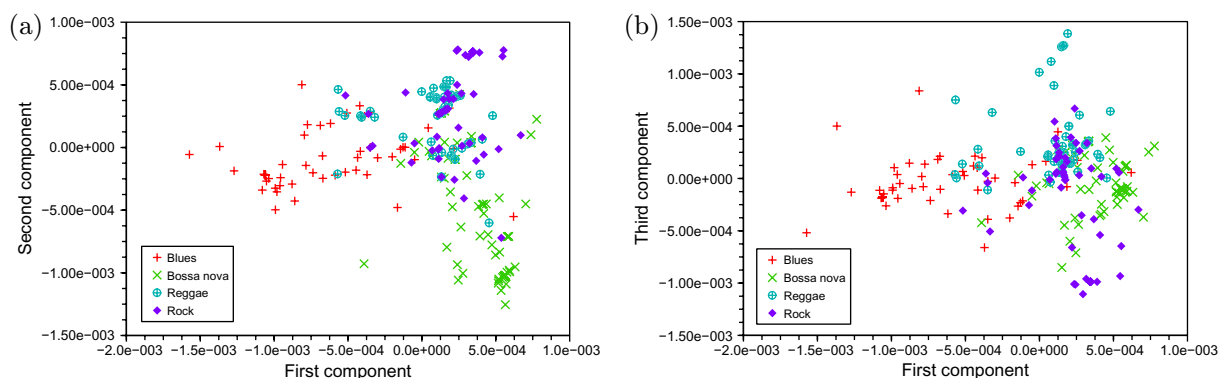


Figure 7. The first three new features obtained by LDA. (a) The first and second features. (b) The first and third features.

Table 7. LDA kappa and accuracy for the quadratic Bayesian classifier.

| | Kappa | Variance | Accuracy | Performance |
|-------------------|-------|----------|----------|-------------|
| Re-substitution | 0.62 | 0.0013 | 71.78% | Substantial |
| Hold-out (70–30%) | 0.57 | 0.0048 | 67.50% | Moderate |
| Hold-out (50–50%) | 0.6 | 0.0027 | 70% | Substantial |

Table 8. Confusion matrix for the quadratic Bayesian classifier using LDA and re-substitution.

| | Blues | Bossa nova | Reggae | Rock |
|------------|-------|------------|--------|------|
| Blues | 61 | 1 | 6 | 2 |
| Bossa nova | 2 | 45 | 20 | 3 |
| Reggae | 6 | 6 | 56 | 2 |
| Rock | 5 | 5 | 21 | 39 |

Table 9. LDA kappa and accuracy for the linear Bayesian classifier.

| | Kappa | Variance | Accuracy | Performance |
|-------------------|-------|----------|----------|-------------|
| Re-substitution | 0.63 | 0.0013 | 72.14% | Substantial |
| Hold-out (70–30%) | 0.58 | 0.0046 | 68.75% | Moderate |
| Hold-out (50–50%) | 0.66 | 0.0024 | 74.28% | Substantial |

could pose difficulties, even for human experts. These misclassified art works have similar properties described in terms of rhythm notations and, as a result, they generate similar weight matrices. Therefore, the proposed methodology, although requiring some complementation, seems to be a significant contribution toward the development of a viable alternative approach to automatic genre classification.

The results for the linear Bayesian classifier using LDA are shown in table 9. In fact, they are closely similar to those obtained by the quadratic Bayesian classifier (table 7).

As mentioned in appendix A.2, LDA also allows us to quantify the intra- and interclass dispersion of the feature matrix through functionals such as the trace and determinant computed from the scatter matrices [41]. The overall intraclass scatter matrix, denoted S_{intra} ; the intraclass scatter matrix for each class, denoted $S_{\text{intraBlues}}$, $S_{\text{intraBossaNova}}$, $S_{\text{intraReggae}}$ and $S_{\text{intraRock}}$; the interclass scatter matrix, denoted, S_{inter} ; and the overall separability index, denoted $(S_{\text{intra}}^{-1} * S_{\text{inter}})$, were computed. Their respective traces are

$$\begin{aligned}\text{trace}(S_{\text{intra}}) &= 505.55, \\ \text{trace}(S_{\text{intraBlues}}) &= 154.96, \\ \text{trace}(S_{\text{intraBossaNova}}) &= 127.41, \\ \text{trace}(S_{\text{intraReggae}}) &= 91.76, \\ \text{trace}(S_{\text{intraRock}}) &= 131.40, \\ \text{trace}(S_{\text{inter}}) &= 22.08, \\ \text{trace}(S_{\text{intra}}^{-1} * S_{\text{inter}}) &= 3.49.\end{aligned}$$

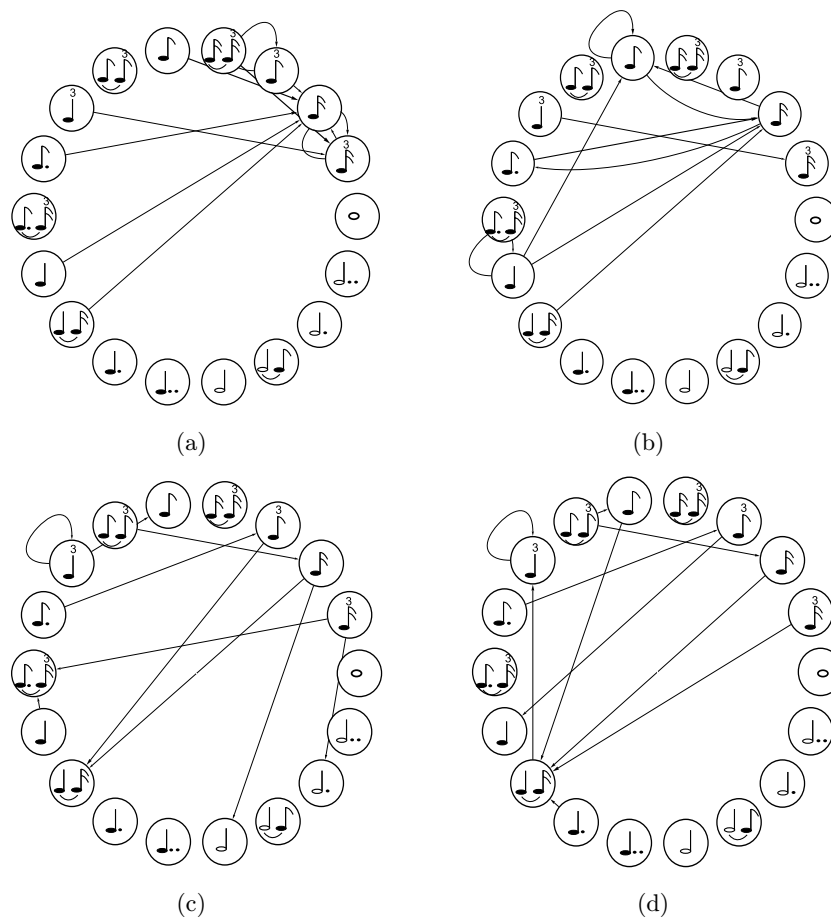
Two important observations are worth mentioning. Firstly, these traces emphasize the difficulty of this classification problem: the traces of the intraclass scatter matrices are too high, and the trace of the interclass scatter matrix together with the overall separability index, are too small. This confirms that the four classes are overlapping completely. Secondly, the smaller intra-class trace is related to the genre reggae (it is the most compact class). This may justify why, in the experiments, art works belonging to reggae were more frequently 90–100% correctly classified.

The PCA and LDA approaches help one to identify the features that contribute the most to the classification. This is an interesting analysis that can be performed by verifying the strength of each element in the first eigenvectors and then associating those elements with the original features. Within the current study, it was figured out that the first ten sequences of rhythm notations that contributed the most to separation correspond to those illustrated in figure 8. In the case of the first and second eigenvectors obtained by PCA and LDA, the ten elements with higher values were selected, and the indices of these elements were associated with the sequences in the original weight matrix. Figures 8(a) and (b) show the resulting sequences according to the first and second eigenvectors of PCA. The thickness of the edges is set by the value of the corresponding element in the eigenvector. It is interesting that these sequences are the ones that mostly frequently occur in the rhythms from all four genres studied here. That is, they correspond to the elements that play the greatest role in representing the rhythms. Therefore, it can be said that these are the *ten most representative sequences* that the first and second eigenvectors of PCA take into account. Triples of the eighth and sixteenth notes are particularly important in the genres blues and reggae. Similarly, figures 8(c) and (d) show the resulting sequences according to the first and second eigenvectors of LDA. In contrast to those obtained by PCA, these sequences are not common to all the rhythms, but they must occur with distinct frequency within each genre. Thus, they are referred to here as the *ten most discriminative sequences* if the first and second eigenvectors of LDA are taken into account.

Clustering results are discussed in the following. The number of clusters was defined as four, in order to provide a fair comparison with the supervised classification results. The idea

Table 10. Confusion matrix for the agglomerative hierarchical clustering.

| | Class 1 | Class 2 | Class 3 | Class 4 |
|---------|---------|---------|---------|---------|
| Class 1 | 25 | 11 | 12 | 22 |
| Class 2 | 32 | 16 | 14 | 8 |
| Class 3 | 46 | 8 | 6 | 10 |
| Class 4 | 30 | 8 | 19 | 13 |

**Figure 8.** The ten most significant sequences of rhythm notations according to (a) the first PCA eigenvector, (b) the second PCA eigenvector, (c) the first LDA eigenvector and (d) the second LDA eigenvector.

behind the confusion matrix in table 10 is to verify how many art works from each class were placed in each one of the four clusters. For example, it is known that from 1 to 70 art works belong to the genre blues. Then, the first line of this confusion matrix indicates that 25 blues art works were placed in cluster 1, 11 in cluster 2, 12 in cluster 3 and 22 in cluster 4. It can also be observed that in cluster 1, reggae art works are the majority (46), whereas in cluster 2, the majority are bossa nova art works (16), despite the small difference compared to the number of

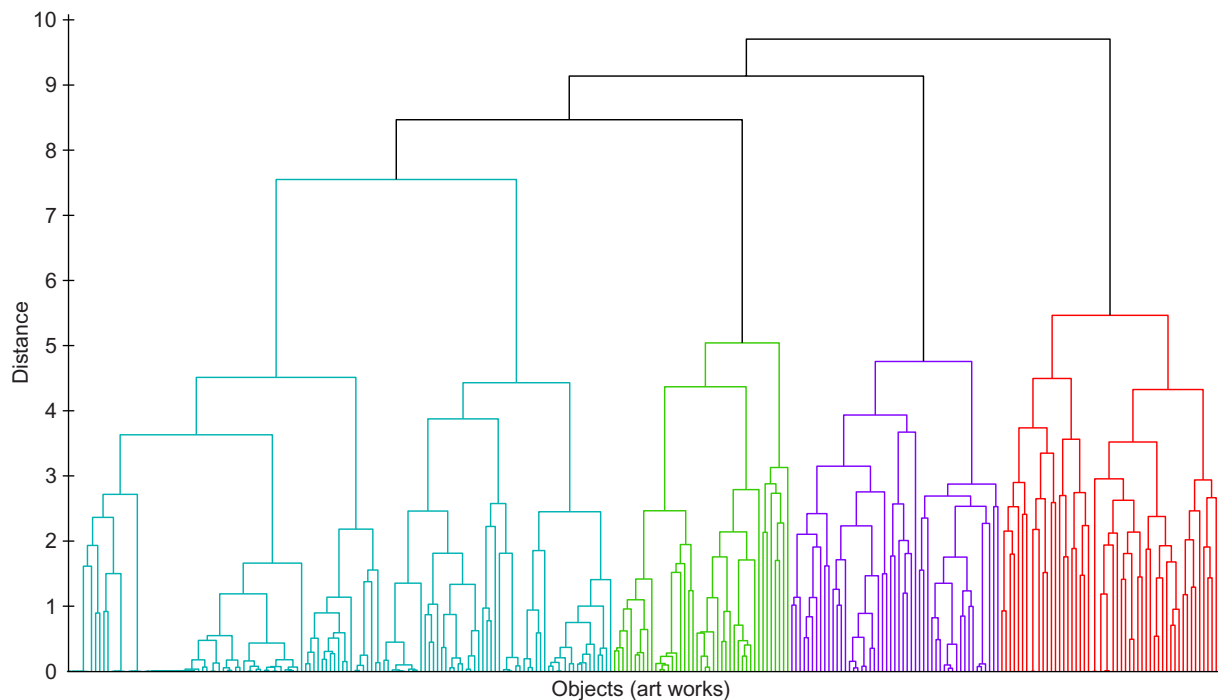


Figure 9. The dendrogram of the resulting four clusters (shown in cyan, green, pink and red).

blues art works (11); in cluster 3, the majority are rock art works; and in cluster 4, the majority are blues art works.

Comparing the confusion matrix in table 10 and the confusion matrix for the quadratic Bayesian classifier using PCA in table 5, it is interesting to notice that: in the former, cluster 1 contains considerable art works from the four genres (25 from blues, 32 from bossa nova, 46 from reggae and 30 from rock), in a total of 133 art works; in the latter, a considerable number of art works from blues (22), bossa nova (23) and rock (31) were misclassified as reggae, in a total of 146 art works belonging to this class. This means that the PCA representation was not efficient in discriminating reggae from the other genres, while cluster 1 was the one that most intermixed art works from all classes.

Figure 9 presents the dendrogram with the four identified clusters. Different colours were used for the sake of enhanced visual analysis. Cluster 1 is shown in cyan, cluster 2 in green, cluster 3 in pink and cluster 4 in red. These colours were based on the dominant class in each cluster. For example, cluster 1 is shown in cyan because reggae art works form the majority in this cluster.

The four clusters obtained are detailed in figures 10–13, in which the legends present the grouped art works from each cluster (blues art works are shown in red, bossa nova art works in green, reggae art works in cyan and rock art works in pink).

As a consequence of working in a higher dimension feature space, the agglomerative hierarchical clustering approach could better separate the data when compared to the PCA- and LDA-based approaches, which are applied over a projected version of the original measurements.

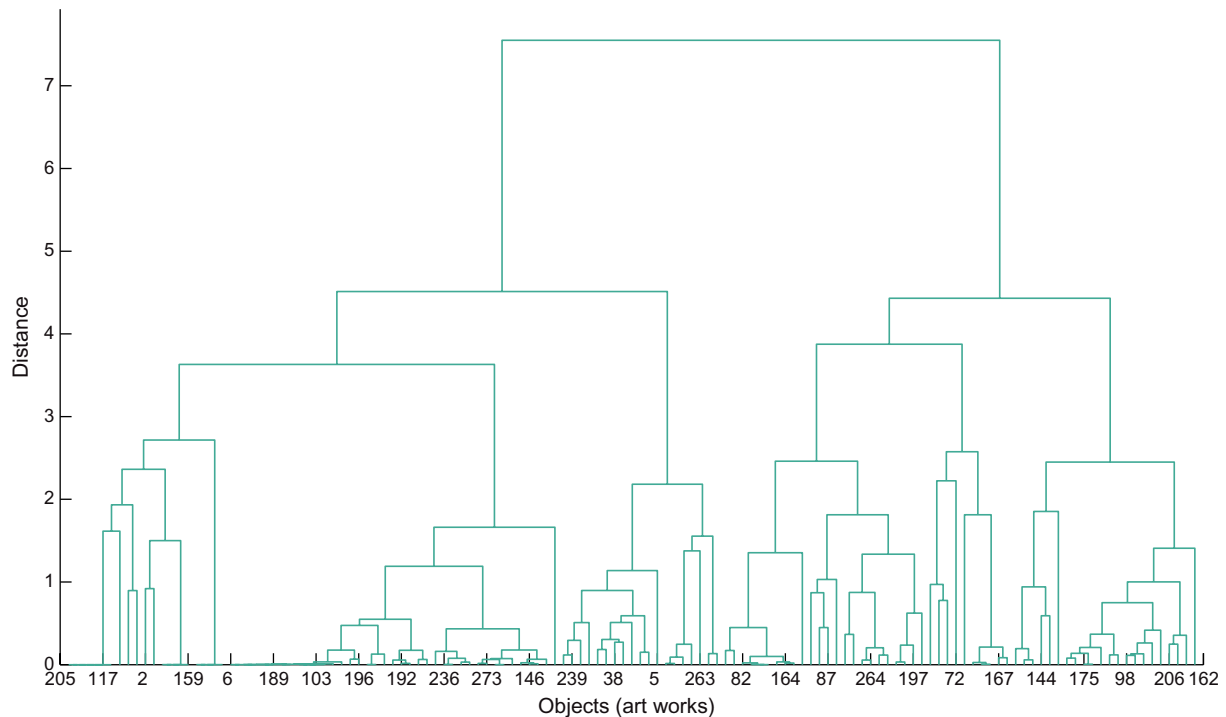


Figure 10. Detailed dendrogram for clusters in cyan. The objects were specified here, from left to right: 205 208 178 174 117 85 161 88 102 2 68 74 237 158 159 119 134 126 123 6 275 198 255 107 189 160 232 97 204 103 168 157 200 153 196 142 187 261 17 192 229 109 219 51 236 267 249 257 83 273 217 274 143 100 146 210 180 194 3 239 190 34 16 18 38 193 155 36 152 5 12 203 25 44 263 23 62 28 266 82 86 125 169 141 164 254 184 30 71 87 129 58 116 115 264 220 276 118 166 197 265 37 131 132 72 78 95 104 105 167 234 29 40 170 144 214 59 69 179 175 177 227 172 226 98 247 182 212 225 206 272 209 and 162.

We computed the performance of the unsupervised classification by using the criterion described in section 2.5.1 and assuming the dominant class in each cluster: reggae in cluster 1, bossa nova in cluster 2, rock in cluster 3 and blues in cluster 4. The performances were as follows: $46/70 = 66\%$ for reggae, $16/70 = 23\%$ for bossa nova, $19/70 = 27\%$ for rock and $22/70 = 31\%$ for blues.

3.2. Multi-label classification results

This section presents the classification that is done in a multi-label manner. The principal motivation is to complement the single classification results, considering the discussions about the redundancies of musical genres. The methodology is described in section 2.5.

Despite the fact that the best results were obtained by using the linear Bayesian classifier over 106 components of PCA (table 6) and by using the linear Bayesian classifier over three components of LDA (table 9), we chose to present the multi-label results for supervised classification by using the quadratic Bayesian classifier over the first two components of LDA

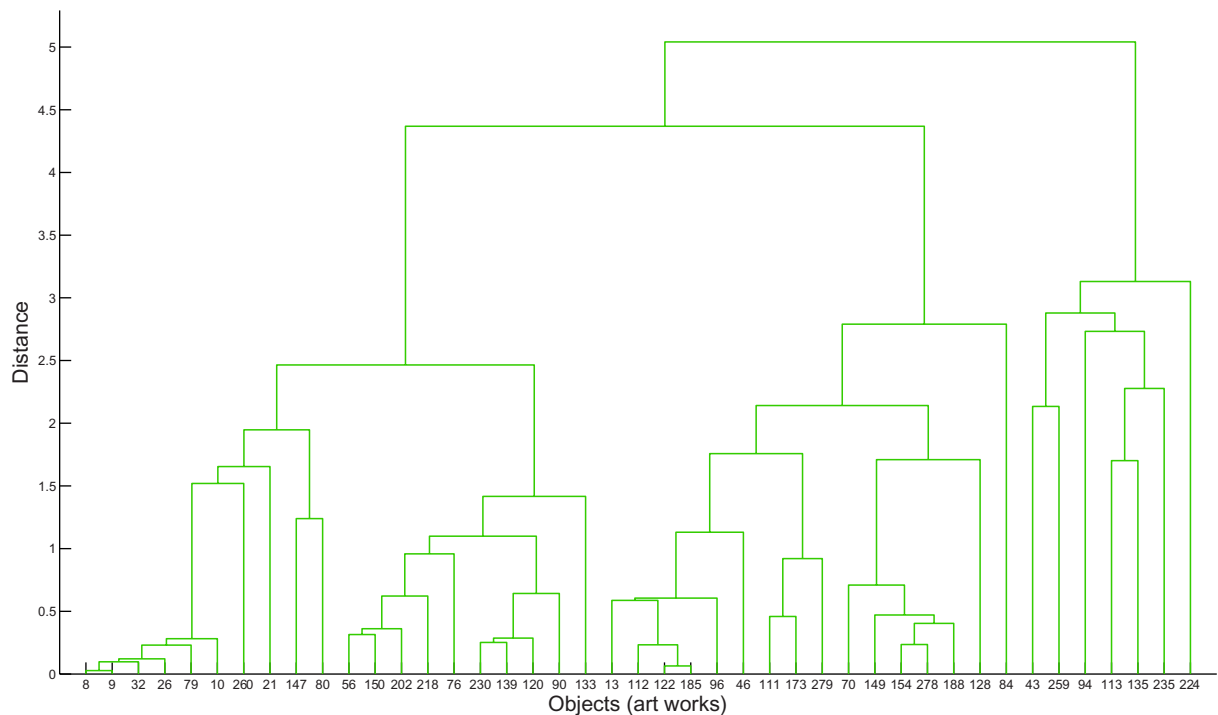


Figure 11. Detailed dendrogram for clusters in green. The objects were specified here, from left to right: 8 9 32 26 79 10 260 21 147 80 56 150 202 218 76 230 139 120 90 133 13 112 122 185 96 46 111 173 279 70 149 154 278 188 128 84 43 259 94 113 135 235 and 224.

(with $\kappa = 0.48$), since this is easier to illustrate graphically. Tables 11 and 12 present the obtained labels for each sample. In order to provide a suitable and graphical visualization of such results, figure 14 shows the contour plots of the 2D class conditional Gaussian densities together with the scatter plots of the dataset. The contour plots are obtained by defining equally spaced subintervals within the range $[p_{\min}, p_{\max}]$, where p_{\min} is the minimum probability value and p_{\max} is the maximum probability value of the distribution. In our case, illustrated in figure 14, the contour plots are given by the central isolines of such intervals (the ones that divide the interval $[p_{\min}, p_{\max}]$ in two halves). It is possible to observe that the labels of each sample are related to their position in the feature space. As an example, consider the blues sample number 30. Although it is a blues sample, its features are more similar to those of the bossa nova samples, that is, its feature vector is located closer to the centre of the reggae conditional density than to the centre of the blues density. Therefore, it is expected that this sample be classified as belonging to the bossa nova class. Samples 56 and 265 were classified as belonging to the four classes. None of the reggae samples were classified as reggae only, since reggae and rock class conditional densities are almost completely overlapping.

The performance of the multi-label unsupervised classification results was computed considering the multi-label dataset presented in section 2.5. For example, instead of considering only the 19 rock samples in the rock cluster, we now verify how many of the 17 reggae–rock samples and how many blues–rock samples there are in the rock cluster, since now they are also referred to as rock samples. The new total of rock samples is 107. The performances

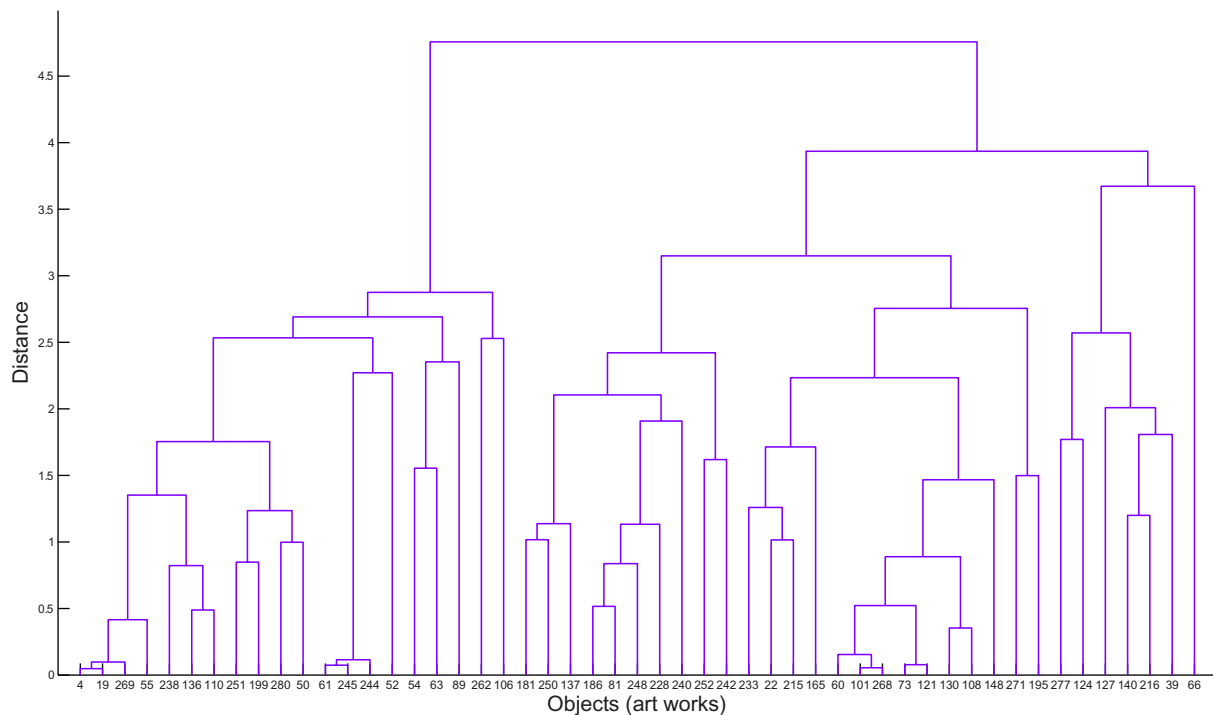


Figure 12. Detailed dendrogram for clusters in pink. The objects were specified here, from left to right: 4 19 269 55 238 136 110 251 199 280 50 61 245 244 52 54 63 89 262 106 181 250 137 186 81 248 228 240 252 242 233 22 215 165 60 101 268 73 121 130 108 148 271 195 277 124 127 140 216 39 and 66.

are: bossa nova—23%, since no blues, rock or reggae samples were also labelled as bossa nova; reggae—66%, since no blues, bossa nova or rock samples were also labelled as reggae; rock—26% $((19+2+7)/107)$, since two reggae–rock samples are in the rock cluster and seven blues–rock samples are in the rock cluster; and blues—30% $(22/73)$, since none of the three rock–blues samples are in the blues cluster.

4. Concluding remarks

Automatic music genre classification has become a fundamental topic in music research since genres have been widely used to organize and describe music collections. They also reveal general identities of different cultures. However, music genres are not a clearly defined concept so that the development of a non-controversial taxonomy represents a challenging, nontrivial task.

Generally speaking, music genres summarize common characteristics of musical pieces. This is particularly interesting when it is used as a resource for automatic classification of pieces. In the current paper, we explored genre classification taking into account the music's temporal aspects, namely the rhythm. We considered pieces of four musical genres (blues, bossa nova, reggae and rock), which were extracted from MIDI files and modelled as networks. Each node corresponded to one rhythmic notation, and the links were defined by the sequence in which they occurred in time. The idea of using static nodes (nodes with fixed positions) is particularly

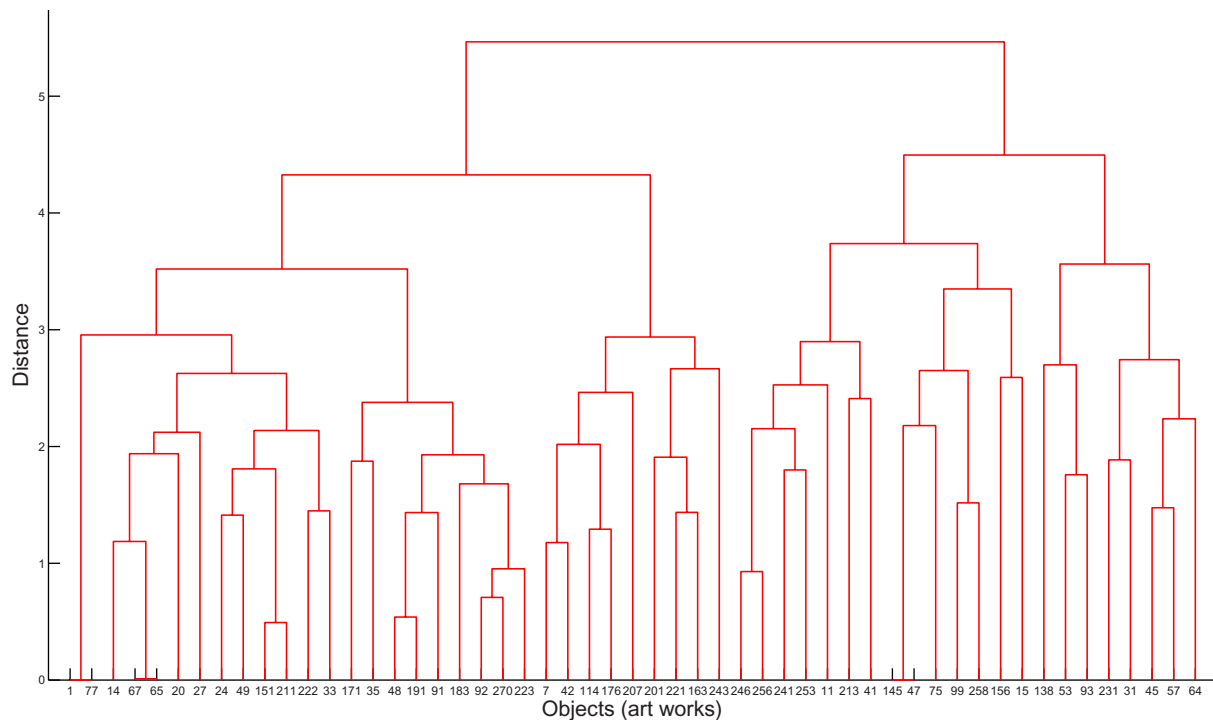


Figure 13. Detailed dendrogram for clusters in red. The objects were specified here, from left to right: 1 77 14 67 65 20 27 24 49 151 211 222 33 171 35 48 191 91 183 92 270 223 7 42 114 176 207 201 221 163 243 246 256 241 253 11 213 41 145 47 75 99 258 156 15 138 53 93 231 31 45 57 and 64.

interesting because it provides a primary visual identification of the differences and similarities between the rhythms from the four genres. A Markov model was built from the networks, and the dynamics and dependences of the rhythmic notations were estimated, comprising the feature matrix of the data. Two different approaches for features analysis were used (PCA and LDA), as well as two types of classification methods (Bayesian classifier and hierarchical clustering).

Using only the first two principal components of PCA, the different types of rhythms were not separable, although for the first and third axes we could observe some separation between three of the classes (blues, bossa nova and reggae), while only the samples of rock overlapped the other classes. However, taking into account that 15 components were necessary to preserve 76% of the data variance, it is expected that only two or three dimensions would not be sufficient to allow suitable separability. Notably, the dimensionality of the problem is high, that is, the rhythms are very complex and many dimensions (features) are necessary to separate them. This is one of the main findings of the current work. With the help of LDA analysis, another finding was reached, which supported the assumption that the problem of automatic rhythm classification is no trivial task. The projections obtained by considering the first and second, and first and third axes implied better discrimination between the four classes than that obtained by the PCA.

Unlike PCA and LDA, agglomerative hierarchical clustering works on the original dimensions of the data. The application of the methodology led to a substantially better

Table 11. Multi-label classification results over all blues and bossa nova samples.

| Blues samples | Bossa nova samples |
|---------------------------------|----------------------------------|
| 1. blues | 71. bossa nova |
| 2. reggae and rock | 72. bossa nova |
| 3. reggae and blues | 73. bossa nova |
| 4. blues | 74. bossa nova, rock and reggae |
| 5. blues | 75. bossa nova |
| 6. reggae and rock | 76. reggae, bossa nova, and rock |
| 7. blues | 77. bossa nova |
| 8. blues | 78. bossa nova |
| 9. blues | 79. bossa nova, rock and reggae |
| 10. blues | 80. rock |
| 11. blues | 81. bossa nova |
| 12. blues | 82. reggae, rock and bossa nova |
| 13. blues | 83. reggae and rock |
| 14. blues | 84. bossa nova |
| 15. blues | 85. bossa nova, reggae and rock |
| 16. blues | 86. reggae, rock and bossa nova |
| 17. reggae and rock | 87. bossa nova |
| 18. blues | 88. bossa nova |
| 19. blues | 89. reggae, rock and blues |
| 20. blues | 90. blues, reggae and rock |
| 21. blues | 91. bossa nova |
| 22. blues | 92. bossa nova |
| 23. blues | 93. bossa nova |
| 24. blues | 94. bossa nova |
| 25. blues | 95. bossa nova |
| 26. blues | 96. reggae, rock and bossa nova |
| 27. blues | 97. reggae and rock |
| 28. reggae, rock and bossa nova | 98. rock and reggae |
| 29. blues and reggae | 99. bossa nova, rock and reggae |
| 30. bossa nova | 100. reggae and rock |
| 31. blues | 101. bossa nova |
| 32. blues | 102. bossa nova |
| 33. blues | 103. reggae and rock |
| 34. blues | 104. bossa nova |
| 35. blues | 105. bossa nova |
| 36. blues and reggae | 106. reggae, rock and bossa nova |
| 37. rock, blues and bossa nova | 107. reggae and rock |
| 38. blues | 108. bossa nova |
| 39. blues | 109. reggae and rock |
| 40. blues | 110. bossa nova |
| 41. blues | 111. reggae and rock |
| 42. blues | 112. reggae, rock and bossa nova |
| 43. blues | 113. bossa nova |
| 44. blues | 114. bossa nova |

Table 11. Continued.

| Blues samples | Bossa nova samples |
|--|----------------------------------|
| 45. blues | 115. reggae, rock and bossa nova |
| 46. blues | 116. reggae, rock and blues |
| 47. blues | 117. bossa nova, reggae and rock |
| 48. blues | 118. reggae, rock and bossa nova |
| 49. blues | 119. bossa nova |
| 50. blues | 120. bossa nova |
| 51. reggae and rock | 121. bossa nova |
| 52. blues, reggae and rock | 122. reggae, rock and bossa nova |
| 53. blues | 123. bossa nova |
| 54. reggae, blues and rock | 124. bossa nova |
| 55. blues | 125. reggae, rock and bossa nova |
| 56. reggae, rock, blues and bossa nova | 126. bossa nova |
| 57. blues | 127. bossa nova |
| 58. blues and reggae | 128. bossa nova |
| 59. blues | 129. bossa nova |
| 60. blues | 130. bossa nova |
| 61. blues | 131. bossa nova |
| 62. blues | 132. bossa nova |
| 63. reggae, blues and rock | 133. bossa nova |
| 64. blues | 134. bossa nova |
| 65. blues | 135. bossa nova |
| 66. blues | 136. bossa nova |
| 67. blues | 137. bossa nova |
| 68. blues | 138. bossa nova |
| 69. rock and reggae | 139. bossa nova |
| 70. blues | 140. bossa nova |

discrimination, which provides strong evidence of the complexity of the problem studied here. The results are promising in the sense that each cluster is dominated by a different genre, showing the viability of the proposed approach.

The use of a multi-label approach was interesting and particularly appropriate, since it reflected the intrinsic nature of the dataset. Musical genre classification is a nontrivial task even for music experts, since often a song can be assigned to more than a single genre. Automatic classification is considerably more complex, because there are many samples with similar feature vectors, defining huge overlapping areas in the feature space, as observed in our study. In this context, multi-label classification plays a fundamental role, making possible generalization of the genre taxonomy, originally presented in the training set. With our proposed method, new sub-genres (for example, rock–blues) can arise from original ones. Therefore, we observed a significant improvement in the supervised classification performance. For the multi-label unsupervised approach, which does not take the data covariance structure into account, we did not observe significant changes in the classification performances. Furthermore, the labelling process that takes place in the LastFm website, through direct listener interaction, considers all music content (instruments, harmony, melody, pitch, voice and percussion, among others).

Table 12. Multi-label classification results over all reggae and rock samples.

| Reggae samples | Rock samples |
|----------------------------------|----------------------------------|
| 141. reggae, rock and bossa nova | 211. rock |
| 142. rock and reggae | 212. rock and reggae |
| 143. reggae and rock | 213. rock |
| 144. blues | 214. reggae and rock |
| 145. rock and reggae | 215. rock |
| 146. reggae and rock | 216. rock |
| 147. reggae, rock and bossa nova | 217. reggae and rock |
| 148. reggae and rock | 218. reggae, rock and bossa nova |
| 149. reggae, rock and bossa nova | 219. reggae and rock |
| 150. reggae, bossa nova and rock | 220. reggae, rock and bossa nova |
| 151. rock and reggae | 221. rock |
| 152. blues and reggae | 222. rock |
| 153. reggae and rock | 223. rock |
| 154. reggae, blues and rock | 224. rock |
| 155. blues and reggae | 225. reggae and rock |
| 156. rock and reggae | 226. rock and reggae |
| 157. reggae and rock | 227. reggae and rock |
| 158. bossa nova, rock and reggae | 228. rock |
| 159. bossa nova, rock and reggae | 229. reggae and rock |
| 160. reggae and rock | 230. reggae, bossa nova and rock |
| 161. reggae and rock | 231. rock |
| 162. bossa nova | 232. reggae and rock |
| 163. rock and reggae | 233. rock |
| 164. reggae, rock and bossa nova | 234. bossa nova |
| 165. rock and reggae | 235. rock |
| 166. reggae, rock and bossa nova | 236. reggae and rock |
| 167. bossa nova | 237. bossa nova, rock and reggae |
| 168. reggae and rock | 238. blues |
| 169. reggae, rock and bossa nova | 239. blues and reggae |
| 170. blues | 240. rock |
| 171. rock and reggae | 241. rock |
| 172. rock and reggae | 242. rock |
| 173. rock and reggae | 243. rock |
| 174. bossa nova, reggae and rock | 244. rock |
| 175. rock and reggae | 245. blues and reggae |
| 176. reggae and rock | 246. bossa nova, reggae and rock |
| 177. rock and reggae | 247. rock and reggae |
| 178. bossa nova, reggae and rock | 248. rock |
| 179. rock and reggae | 249. reggae and rock |
| 180. reggae and rock | 250. rock |
| 181. reggae, rock and bossa nova | 251. rock |
| 182. rock and reggae | 252. rock |
| 183. rock and reggae | 253. rock |
| 184. reggae and rock | 254. reggae, rock and bossa nova |

Table 12. Continued.

| Reggae samples | Rock samples |
|----------------------------------|---|
| 185. reggae, rock and bossa nova | 255. reggae and rock |
| 186. rock and reggae | 256. reggae, rock and bossa nova |
| 187. rock and reggae | 257. reggae and rock |
| 188. blues and reggae | 258. reggae and rock |
| 189. reggae and rock | 259. rock |
| 190. reggae and blues | 260. bossa nova, reggae and rock |
| 191. rock and reggae | 261. rock and reggae |
| 192. reggae and rock | 262. reggae, rock and blues |
| 193. blues and reggae | 263. reggae and rock |
| 194. rock and reggae | 264. reggae, rock and bossa nova |
| 195. rock and reggae | 265. reggae, blues, rock and bossa nova |
| 196. reggae and rock | 266. reggae, rock and bossa nova |
| 197. reggae, rock and bossa nova | 267. reggae and rock |
| 198. reggae and rock | 268. bossa nova |
| 199. rock and reggae | 269. rock |
| 200. reggae and rock | 270. rock |
| 201. rock and reggae | 271. rock |
| 202. reggae, bossa nova and rock | 272. rock and reggae |
| 203. blues | 273. reggae and rock |
| 204. reggae and rock | 274. reggae and rock |
| 205. bossa nova, reggae and rock | 275. reggae and rock |
| 206. reggae and rock | 276. reggae, rock and bossa nova |
| 207. reggae and rock | 277. bossa nova and rock |
| 208. bossa nova, reggae and rock | 278. blues |
| 209. reggae and rock | 279. reggae, rock and bossa nova |
| 210. reggae and rock | 280. rock |

Here, our focus is only on the rhythm analysis, which in practical terms reflects a drastic reduction of computational costs, since it is a substantially more compact representation.

It is clear from our study that musical genres are very complex and that they present redundancies. Sometimes it is difficult even for an expert to distinguish them. This difficulty becomes more critical when only the rhythm is taken into account.

There are several possibilities for future research implied by the reported investigation. First, it would be interesting to use more measurements extracted from rhythm, especially the intensity of the beats, as well as the distribution of instruments, which is poised to improve the classification results. Another promising area for further investigation regards the use of other classifiers, as well as the combination of results obtained from an ensemble of distinct classifiers. In addition, it would be promising to apply multi-label classification, a growing field of research in which non-disjointed samples can be associated with one or more labels [42]. Another interesting future work is related to the synthesis of rhythms. Once the rhythmic networks are available, new rhythms with similar characteristics according to the specific genre can be artificially generated.

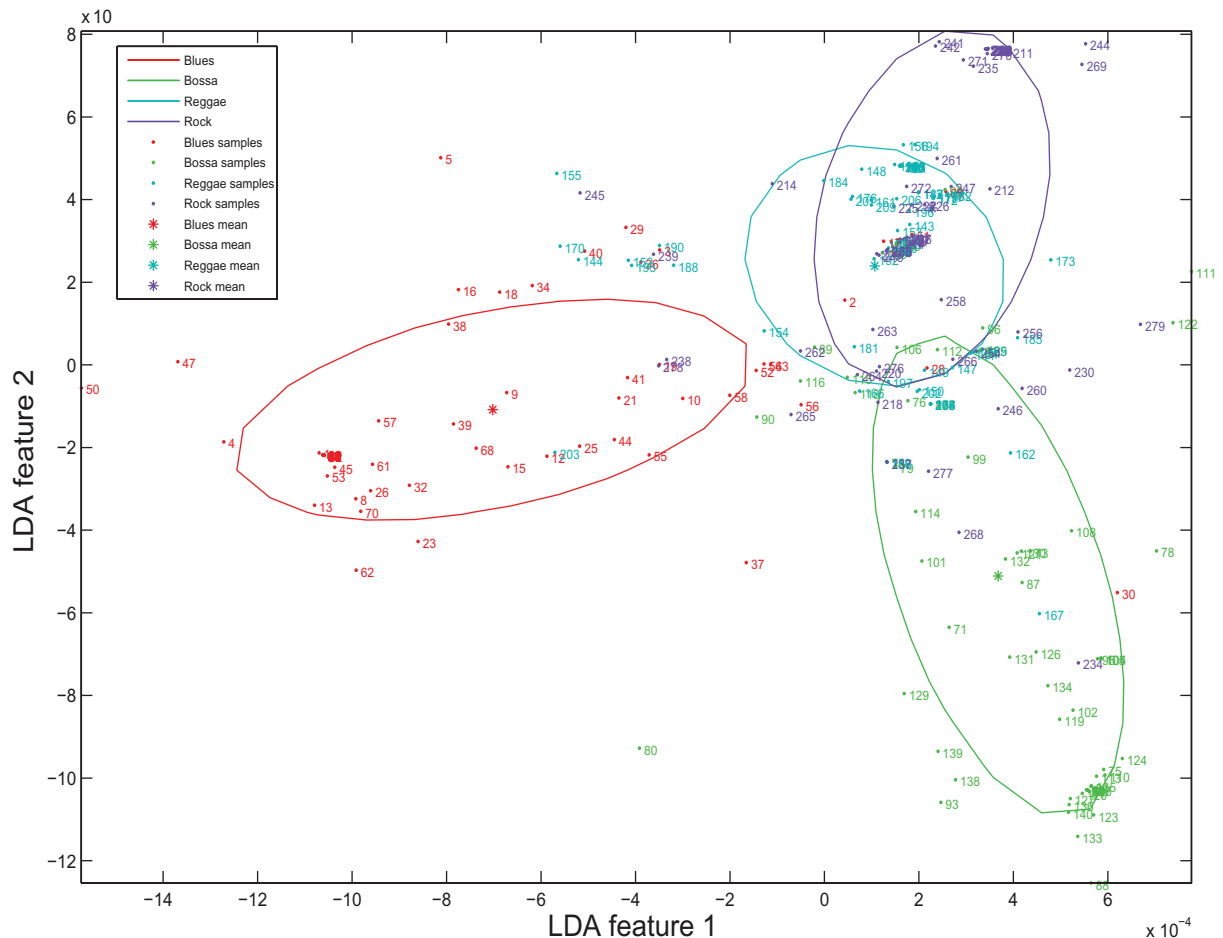


Figure 14. Contour plots of the 2D class conditional Gaussian densities and scatter plots of the dataset: all points in a given contour plot are equally likely, since they are equidistant from the mean vector of the corresponding class, according to the Mahalanobis metric.

Acknowledgments

DCC is grateful to FAPESP (2009/50142-0) for financial support and LFC is grateful to CNPq (301303/06-1 and 573583/2008-0) and FAPESP (05/00587-5) for financial support.

Appendix A. Multivariate statistical methods

A.1. PCA

PCA is a second-order unsupervised statistical technique. By second order it is meant that all the necessary information is available directly from the covariance matrix of the mixture data, so that there is no need to use the complete probability distributions. This method uses the eigenvalues and eigenvectors of the covariance matrix in order to transform the feature space, creating orthogonal uncorrelated features. From a multivariate dataset, the principal aim of PCA is to remove redundancy from the data, consequently reducing the dimensionality of

the data. Additional information about PCA and its relation to various interesting statistical and geometrical properties can be found in the pattern recognition literature, e.g. [6, 7, 36, 43].

Consider a vector \mathbf{x} with n elements representing some features or measurements of a sample. In the first step of PCA transform, this vector \mathbf{x} is centered by subtracting its mean, so that $\mathbf{x} \leftarrow \mathbf{x} - E\{\mathbf{x}\}$. Next, \mathbf{x} is linearly transformed to a different vector \mathbf{y} that contains m elements, $m < n$, removing the redundancy caused by the correlations. This is achieved by using a rotated orthogonal coordinate system in such a way that the elements in \mathbf{x} are uncorrelated in the new coordinate system. At the same time, PCA maximizes the variances of the projections of \mathbf{x} on the new coordinate axes (components). These variances of the components will differ in most applications. The axes associated with small dispersions (given by the respectively associated eigenvalues) can be discarded without losing too much information about the original data.

A.2. LDA

LDA can be considered a generalization of Fisher's linear discriminant function for the multivariate case [7, 36]. It is a supervised approach that maximizes data separability, in terms of a similarity criterion based on scatter matrices. The basic idea is that objects belonging to the same class are as similar as possible and objects belonging to distinct classes are as different as possible. In other words, LDA looks for a new, projected, feature space that maximizes interclass distance while minimizing the intraclass distance. This result can be later used for linear classification, and it is also possible to reduce dimensionality before the classification task. The scatter matrix for each class indicates the dispersion of the feature vectors within the class. The intraclass scatter matrix is defined as the sum of the scatter matrices of all classes and expresses the combined dispersion in each class. The interclass scatter matrix quantifies how disperse the classes are, in terms of the position of their centroids.

It can be shown that the maximization criterion for class separability leads to a generalized eigenvalue problem [7, 36]. Therefore, it is possible to compute the eigenvalues and eigenvectors of the matrix defined by $(S_{\text{intra}}^{-1} * S_{\text{inter}})$, where S_{intra} is the intraclass scatter matrix and S_{inter} is the interclass scatter matrix. The m eigenvectors associated with the m largest eigenvalues of this matrix can be used to project the data. However, the rank of $(S_{\text{intra}}^{-1} * S_{\text{inter}})$ is limited to $C - 1$, where C is the number of classes. As a consequence, there are $C - 1$ nonzero eigenvalues, that is, the number of new features is conditioned to the number of classes, $m \leq C - 1$. Another issue is that, for high dimensional problems, when the number of available training samples is smaller than the number of features, S_{intra} becomes singular, complicating the generalized eigenvalue solution.

More information about the LDA can be found in [7], [35]–[37].

Appendix B. Linear and quadratic discriminant functions

If normal distribution over the data is assumed, it is possible to state that

$$p(x|\omega_i) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \right\}. \quad (\text{B.1})$$

The components of the parameter vector for class j , $\vec{\theta}_j = \{\vec{\mu}_j, \Sigma_j\}$, where $\vec{\mu}_j$ and Σ_j are the mean vector and the covariance matrix of class j , respectively, can be estimated by

maximum likelihood as follows:

$$\hat{\mu}_j = \frac{1}{N} \sum_{i=1}^N \vec{x}_i, \quad (\text{B.2})$$

$$\hat{\Sigma}_j = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{\mu}_i) (\vec{x}_i - \vec{\mu}_i)^T. \quad (\text{B.3})$$

Within this context, classification can be achieved with discriminant functions, g_i , assigning an observed pattern vector \vec{x}_i to the class ω_j with the maximum discriminant function value. By using Bayes's rule, not considering the constant terms and using the estimated parameters above, a decision rule can be defined as follows: assign an object \vec{x}_i to class ω_j if $g_j > g_i$ for all $i \neq j$, where the discriminant function g_i is calculated as:

$$g_i(\vec{x}) = \log(p(\omega_i)) - \frac{1}{2} \log(|\hat{\Sigma}_i|) - \frac{1}{2} (\vec{x} - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1} (\vec{x} - \hat{\mu}_i). \quad (\text{B.4})$$

Classifying an object or pattern \vec{x} on the basis of the values of $g_i(\vec{x})$, $i = 1, \dots, C$ (C is the number of classes), with estimated parameters, defines a quadratic discriminant classifier, quadratic Bayesian classifier or quadratic Gaussian classifier [36].

The prior probability, $p(\omega_i)$, can be simply estimated by

$$p(\omega_i) = \frac{n_i}{\sum_j n_j}, \quad (\text{B.5})$$

where n_i is the number of samples of class ω_i .

In multivariate classification situations, with different covariance matrices, problems may occur in the quadratic Bayesian classifier when any of the matrices $\hat{\Sigma}_i$ is singular. This usually happens when there are not enough data to obtain efficient estimation for the covariance matrices Σ_i , $i = 1, 2, \dots, C$. An alternative to minimize this problem consists of estimating one unique covariance matrix over all classes, $\hat{\Sigma} = \hat{\Sigma}_1 = \dots = \hat{\Sigma}_C$. In this case, the discriminant function becomes linear in \vec{x} and can be simplified:

$$g_i(\vec{x}) = \log(p(\omega_i)) - \frac{1}{2} \hat{\mu}_i^T \hat{\Sigma}^{-1} \hat{\mu}_i + \vec{x}^T \hat{\Sigma}^{-1} \hat{\mu}_i, \quad (\text{B.6})$$

where $\hat{\Sigma}$ is the covariance matrix, common to all classes. The classification rule remains the same. This defines a linear discriminant classifier (also known as a linear Bayesian classifier or a linear Gaussian classifier) [36].

Appendix C. Agglomerative hierarchical clustering

Agglomerative hierarchical clustering groups progressively the N objects into C classes according to a defined parameter. The distance or similarity between the feature vectors of the objects is usually taken as such a parameter. In the first step, there is a partition with N clusters, each cluster containing one object. The next step is a different partition, with $N - 1$ clusters, the next a partition with $N - 2$ clusters, and so on. In the n th step, each object forms a unique cluster. This sequence groups objects that are more similar to one another into subclasses before objects that are less similar. It is possible to say that, in the k th step, $C = N - k + 1$.

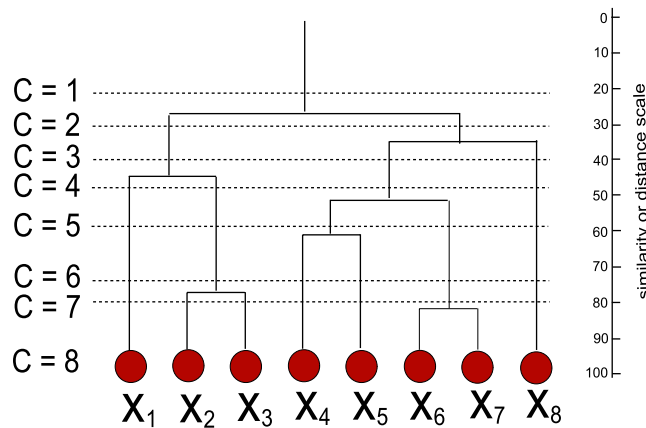


Figure C.1. Dendrogram for a simple situation with eight objects, adapted from [7].

To show how the objects are grouped, hierarchical clustering can be represented by a corresponding tree, called a *dendrogram*. Figure C.1 illustrates a dendrogram representing the results of hierarchical clustering for a problem with eight objects. The measure of similarity among clusters can be observed in the vertical axis. The different number of classes can be obtained by horizontally cutting the dendrogram at different values of similarity or distance. Hence, to perform hierarchical cluster analysis it is necessary to define three main parameters. The first regards how to quantify the similarity between every pair of objects in the dataset, that is, how to calculate the distance between the objects. Euclidean distance, which is frequently used, will be adopted in this work, but other possible distances are cityblock, chessboard, Mahalanobis and so on. The second parameter is the linkage method, which establishes how to measure the distance between two sets. The linkage method can be used to link pairs of objects that are similar and then to form the hierarchical cluster tree. Some of the most popular linkage methods are: single and complete linkage, mean linkage and Ward's linkage [6], [44]–[46]. Ward's linkage uses the intra-class dispersion as a clustering criterion. Pairs of objects are merged in such a way to guarantee the smallest increase in the intraclass dispersion. This clustering approach has sometimes been identified as corresponding to the best hierarchical method [47]–[49] and will be used in this work. Actually, it is particularly interesting to analyse the intraclass dispersion in an unsupervised classification procedure in order to identify common and different characteristics when compared with the supervised classification. The third parameter concerns the number of desired clusters, an issue that is directly related to where to cut the dendrogram into clusters, as illustrated by C in figure C.1.

Appendix D. The kappa coefficient

The kappa coefficient was first proposed by Cohen [39]. In the context of supervised classification, this coefficient determines the degree of agreement *a posteriori*. This means that it quantifies the agreement between objects previously known (ground truth) and the result obtained by the classifier. The better the classification accuracy, the higher the degree of concordance and, consequently, the higher the value of kappa. The kappa coefficient is

Table D.1. Classification performance according to kappa.

| Kappa | Classification performance |
|--------------------------|----------------------------|
| $\hat{k} \leq 0$ | Poor |
| $0 < \hat{k} \leq 0.2$ | Slight |
| $0.2 < \hat{k} \leq 0.4$ | Fair |
| $0.4 < \hat{k} \leq 0.6$ | Moderate |
| $0.6 < \hat{k} \leq 0.8$ | Substantial |
| $0.8 < \hat{k} \leq 1.0$ | Almost perfect |

computed from the confusion matrix as follows [40]:

$$\hat{k} = \frac{N \sum_{i=1}^C c_{ii} - \sum_{i=1}^C x_{i+} x_{+i}}{N^2 - \sum_{i=1}^C x_{i+} x_{+i}}, \quad (\text{D.1})$$

where x_{i+} is the sum of elements from line i , x_{+i} is the sum of elements from column i , C is the number of classes (confusion matrix is $C \times C$) and N is the total number of objects. The kappa variance can be calculated as

$$\hat{\sigma}_k^2 = \frac{1}{N} \left[\frac{\theta_1 (1 - \theta_1)}{(1 - \theta_2)^2} + \frac{2 (1 - \theta_1) (2\theta_1\theta_2 - \theta_3)}{(1 - \theta_2)^3} + \frac{(1 - \theta_1)^2 (\theta_4 - 4\theta_2^2)}{(1 - \theta_2)^4} \right], \quad (\text{D.2})$$

where

$$\begin{aligned} \theta_1 &= \frac{1}{N} \sum_{i=1}^C x_{ii}, \\ \theta_2 &= \frac{1}{N^2} \sum_{i=1}^C x_{i+} x_{+i}, \\ \theta_3 &= \frac{1}{N^2} \sum_{i=1}^C x_{ii} (x_{i+} + x_{+i}), \\ \theta_4 &= \frac{1}{N^3} \sum_{i=1}^C \sum_{j=1}^C x_{ij} (x_{j+} + x_{+i})^2. \end{aligned} \quad (\text{D.3})$$

This statistics indicates that, when $\hat{k} \leq 0$, there is no agreement, and when $\hat{k} = 1$, the agreement is total. Some authors suggest interpretations according to the value obtained by the coefficient kappa. Table D.1 shows one possible interpretation, proposed in [50].

References

- [1] Lena J C and Peterson R A 2008 Classification as culture: types and trajectories of music genres *Am. Soc. Rev.* **73** 697–718
- [2] Holt F 2007 *Genre in Popular Music* (Chicago, IL: University of Chicago Press)
- [3] Scaringella N, Zoia G and Mlynek D 2006 Automatic genre classification of music content—a survey *IEEE Signal Process. Mag.* **23** 133–41

- [4] Aucounturier J J and Pachet F 2005 Representing musical genre: a state of the art *J. New Music Res.* **32** 83–93
- [5] Pachet F and Cazaly D 2000 A taxonomy of musical genres *Proc. Content-Based Multimedia Information Access (RIAO) (Paris)* pp 1238–45
- [6] Costa L da F and Cesar R M J R 2001 *Shape Analysis and Classification Theory and Practice* (Boca Raton, FL: CRC Press)
- [7] Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* (New York: Wiley)
- [8] Cataltepe Z, Yasian Y and Sonmez A 2007 Music genre classification using midi and audio features *EURASIP J. Adv. Signal Process* pp 1–8
- [9] Li T, Ogihara M, Shao B and Wang D 2009 Machine learning approaches for music information retrieval *Theory and Novel Applications of Machine Learning* (Vienna, Austria: I-Tech) pp 259–78
- [10] Mostafa M M and Billor N 2009 Recognition of Western style musical genres using machine learning techniques *Expert Syst. Appl.* **36** 11378–89
- [11] Wang L, Huang S, Wang S, Liang J and Xu B 2008 Music genre classification based on multiple classifier fusion *4th Int. Conf. on Natural Computation* pp 580–3
- [12] Song Y and Zhang C 2008 Content-based information fusion for semi-supervised music genre classification *IEEE Trans. Multimedia* **10** 145–52
- [13] Panagakis I, Benetos E and Kotropoulos C 2008 Music genre classification: a multilinear approach *Proc. 9th Int. Conf. on Music Information Retrieval* pp 583–8
- [14] Hong J, Deng H and Yan Q 2008 Tag-based artist similarity and genre classification *Proc. IEEE Int. Symp. on Knowledge Acquisition and Modeling* pp 628–31
- [15] Holzapfel A and Stylianou Y 2007 A statistical approach to musical genre classification using non-negative matrix factorization *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* vol 2 pp II–693–6
- [16] Silla C N, Kaestner A A and Koerich A L 2007 Automatic music genre classification using ensemble of classifiers *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics* pp 1687–92
- [17] Scaringella N and Zoia G 2005 On the modeling of time information for automatic genre recognition systems in audio signals *Proc. 6th Int. Symp. on Music Information Retrieval* pp 666–71
- [18] Shao X, Xu C and Kankanhalli M 2004 Unsupervised classification of musical genre using hidden Markov model *Proc. Int. Conf. on Multimedia Explore (ICME)* pp 2023–6
- [19] Burred J J and Lerch A 2003 A hierarchical approach to automatic musical genre classification *Proc. Int. Conf. on Digital Audio Effects* pp 8–11
- [20] Akhtaruzzaman Md 2008 Representation of musical rhythm and its classification system based on mathematical and geometrical analysis *Proc. Int. Conf. on Computer and Communication Engineering* pp 466–71
- [21] Karydis I 2006 Symbolic music genre classification based on note pitch and duration *Proc. 10th East European Conf. on Advances in Databases and Information Systems* pp 329–38
- [22] Gouyon F and Dixon S 2005 A review of automatic rhythm description system *Comput. Music J.* **29** 34–54
- [23] Gouyon F, Dixon S, Pampalk E and Widmer G 2004 Evaluating rhythmic descriptors for musical genre classification *Proc. 25th Int. AES Conf.* pp 196–204
- [24] Katakis I, Tsoumakas G and Vlahavas I 2008 Multilabel text classification for automated tag suggestion *Proc. ECML/PKDD 2008 Discovery Challenge* pp 75–83
- [25] Boutell M R, Luo J, Shen X and Brown C M 2004 Learning multi-label scene classification *Pattern Recognit.* **37** 1757–71
- [26] Diplaris S, Tsoumakas G, Mitkas P and Vlahavas I 2005 Protein classification with multiple algorithms *Proc. 10th Panhellenic Conf. on Informatics (PCI 2005)* pp 448–56
- [27] Trohidis K, Tsoumakas G, Kalliris G and Vlahavas I 2008 Multilabel classification of music into emotions *Proc. 2008 Int. Conf. on Music Information Retrieval (ISMIR 2008)* pp 325–30
- [28] Albert R and Barabasi A L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47–97
- [29] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256
- [30] Booth T L 1967 *Sequential Machines and Automata Theory* (New York: Wiley)

- [31] Rohrmeier M 2006 Towards modelling harmonic movement in music: analysing properties and dynamic aspects of Pc set sequences in Bach's Chorales *PhD Thesis* Cambridge University Press
- [32] Costa L da F, Rodrigues F A, Travieso G and Boas V P R 2007 Characterization of complex networks: a survey of measurements *Adv. Phys.* **56** 167–242
- [33] Miranda E R 2001 *Composing Music With Computers* (Oxford, UK: Focal Press)
- [34] Eerola T and Toiviainen P I 2004 *MIDI Toolbox: Tools MATLAB for Music Research* (Jyväskylä, Finland: University of Jyväskylä)
- [35] Devijver P A and Kittler J 1981 *Pattern Recognition: A Statistical Approach* (Englewood Cliffs, NJ: Prentice Hall)
- [36] Webb A R 2002 *Statistical Pattern Recognition* (New York: Wiley)
- [37] Therrien C W 1989 *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics* (New York: Wiley)
- [38] Theodoridis S and Koutroumbas K 2006 *Pattern Recognition* (Amsterdam: Elsevier)
- [39] Cohen J 1960 A coefficient of agreement for nominal scales *Educ. Psychol. Meas.* **20** 37–46
- [40] Congalton R G 1991 A review assessing the accuracy of classifications of remotely sensed data *Remote Sensors Environ.* **37** 35–46
- [41] Fukunaga K 1990 *Introduction to Statistical Pattern Recognition* (New York: Academic)
- [42] Tsoumakas G and Katakis I 2007 Multi-label classification: an overview *Int. J. Data Warehousing Min.* **3** 1–13
- [43] Hyvarinen A, Karhunen J and Oja E 2001 *Independent Component Analysis* (New York: Wiley)
- [44] Jain A K and Dubes R C 1988 *Algorithms for Clustering Data* (Englewood Cliffs, NJ: Prentice Hall)
- [45] Anderberg M R 1973 *Cluster Analysis for Applications* (New York: Academic)
- [46] Romesburg H C 1990 *Cluster Analysis for Researches* (Morrisville, NC: LULU Press)
- [47] Kuiper F K and Fisher L A 1975 Monte Carlo comparison of six clustering procedures *Biometrics* **31** 777–83
- [48] Blashfield R K 1976 Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods *Psychol. Bull.* **83** 377–88
- [49] Mojena R 1975 Hierarchical grouping methods and stopping rules: an evaluation *J. Comput.* **20** 359–63
- [50] Landis J R and Koch G G 1977 The measurement of observer agreement for categorical data *Biometrics* **33** 159–74