Expert Systems With Applications 60 (2016) 190–210

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Short Survey

A survey on symbolic data-based music genre classification

Débora C. Corrêa*, Francisco Ap. Rodrigues

Institute of Mathematics and Computer Science, University of São Paulo, 4000 Av. Trabalhador Sao-Carlense, São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 19 November 2014
Revised 5 April 2016
Accepted 6 April 2016
Available online 8 April 2016

Keywords:
Musical genres
Music information retrieval
Symbolic music data
Music descriptors
Classification algorithms

ABSTRACT

Music is present in everyday life and used for a wide range of objectives. Musical databases have considerably increased in number and size over the past years, therefore, the development of accurate tools for music information retrieval (MIR) has become an important topic in computer science. The increasing theoretical advances in machine learning algorithms together with the abundance of recordings available in digital audio formats, the growing quality and accessibility of on-line symbolic music data, and availability of tools and toolboxes for the extraction of musical properties have motivated many studies on machine learning and MIR. Relevant problems in MIR include classification of songs into genres, which enables the summarization of common features (or patterns) shared by different songs. The automatic classification of music genres plays a fundamental role in the context of music indexing and retrieval, so that websites and device music engines can manage and label music content. Most studies have dealt with such an issue by extracting music characteristics from the audio content, and some have provided overviews of audio features and classification algorithms for music genre classification. However, precise high-level musical information can be extracted from symbolic data (e.g. digital music scores), known to be closely related to the way humans perceive music. A number of approaches use such musical information to process, retrieve and classify music content. This manuscript provides an overview of the most important approaches that deal with music genre classification and consider the symbolic representation of music data. Current issues inherent to such a music format, as well the main algorithms adopted for the modeling of the music feature space are presented.

1. Introduction

Online musical databases and user-interactive applications have considerably increased since the last decade, therefore, the development of fast and effective automatic tools for the classification and retrieval of musical content has become an essential issue (Casey et al., 2008). The principal aim of such tools is the extraction of suitable and compact music information for the representation and organization of music data.

An inherent way of organizing music collections is forming groups of songs in which music facets share similarities. Music similarity can be addressed in different ways, depending on the application domain. Music information retrieval (MIR) is the research area that aims at the development of computational approaches for the retrieval of useful information from music and the classification of pieces according to the categories described by their music content. Each category is expected to comprise songs with common music features (or patterns), whereas songs from different categories should present distinct patterns. In MIR research, the most common ways of labeling categories are relating them regarding genre/style, emotion/mood and composer/artist (Fu, Lu, Ting, & Zhang, 2011).

Musical genre is particularly important, as it facilitates and enhances search for music, reveals the interplay of cultures and importance of cultural features, and is used in the organization of music collections (McKay & Fujinaga, 2006b; Scaringella, Zoia, & Mlynek, 2006). Research has revealed users generally prefer to browse music by genres than by other popular alternatives, as artist similarity or recommendation (Lee & Downie, 2004). Moreover, studies have suggested the genre associated with a particular piece can influence the listener's taste (North & Hargreaves, 1997). Regarding music similarity, musical genres are particular important in the science of MIR. Despite their broad use, however, music genres are not a well defined concept, as they express controversial taxonomies and their boundaries remain fuzzy (Aucouturier & Pachet, 2003). Their manual annotation also imposes inconsistencies due to its subjective manner and is certainly not feasible for a large amount of data.

On the other hand, the automatic classification of music genres is a hard task, due to the lack of reliable ground truth data

* Corresponding author.
E-mail addresses: debcris.cor@gmail.com (D.C. Corrêa), francisco@icmc.usp.br (F.Ap. Rodrigues).

http://dx.doi.org/10.1016/j.eswa.2016.04.008
0957-4174/

crucial for an adequate application of pattern recognition methods. Moreover, the fuzzy definition of music genres may yield a multi-categorization of individual songs. The lack of consensus among human annotators regarding the categories in which a song should be classified is another drawback for automatic classification (Kotsifakos, Kotsifakos, Papapetrou, & Athitsos, 2013; McKay & Fujinaga, 2006b).

The automatic classification of music genre by different music descriptors and classification schemes has been the focus of several studies. The classification problem is usually addressed from five different perspectives in the MIR community (Silla & Freitas, 2009), namely (i) audio content-based, (ii) symbolic content-based, (iii) lyrics-based, (iv) community meta-data based, and (v) hybrid approaches. Audio content-based approaches explore music features in digital audio signals, whereas in symbolic content-based approaches, the features are extracted from symbolic data formats, like MIDI and KERN, in which musical events are presented in a higher level of abstraction. Content-based music retrieval (CBMR) has drawn interest over the past decades and become an emerging topic in MIR studies. Lyrics-based approaches take advantage of text mining techniques to perform classification according to musical information of the lyrics content. Along similar lines, community meta-data based systems explore web-mining methods to capture relevant online musical information from songs. In a most recent perspective, hybrid-based systems use music features from the combination of two or more previous approaches.

Most methods that deal with music genre classification use the music content itself, instead of music meta-data and lyrics (see, for example, Fu et al., 2011; Scaringella et al., 2006). In fact, most studies on the automatic classification of music systematically consider music features based on music content (Fu et al., 2011). Although the recognition of similar music features may be direct and intuitive for listeners, their automatic identification may be a non-trivial task. Determining if two songs belong to the same genre category usually requires the extraction of suitable music descriptors and use of machine learning or pattern recognition techniques for the inference of music similarity. Musical descriptors can cover diverse sources, as audio, meta-data, user ratings and social tasks. Although closely related to the form humans perceive music similarly, user ratings and social tags are conditioned to the availability and reliability of unbiased user entries (Bobdanov, Serrá, Wack, Herrera, & Serra, 2011).

In the MIR community, the representation of music content is usually divided into two different perspectives, namely (i) audio-recorded or (ii) symbolic content. The former is the acoustic representation of music, obtained by the sampling of a sound waveform. Audio descriptors are extracted mainly with the help of Fourier analysis and signal processing techniques. This category represents the majority of studies on music genre classification and includes music descriptors, as MFCCs (Mel-frequency Cepstral coefficients), spectral shape features and temporal and energy features. Symbolic representation, on the other hand, generally uses music scores formats, as MusicXML or KERN, and music notation protocols, as MIDI, to capture music events. Symbolic data offer high-level music representation, since it presents music in terms of instructions or directions that is supposed to be followed by a performer. For instance, note events are described in terms of pitch, duration, strength and so on. Common music descriptors of this category include key, rhythm, tempo, meter, pitch and instrumentation (e.g. McKay & Fujinaga, 2004; 2005a). Furthermore, such music representation has also the advantage of requiring little space, which facilitates storage and communication.

Both representations have advantages and drawbacks. Whereas in audio all information is mixed together, symbolic data cannot store voice or inhibit scalability. Nevertheless, it is encouraging to think music classification could benefit from both representations. Music features from symbolic data can be directly and systematically analyzed without the interference of audio noises, which impose irrelevant properties on the music. Symbolic data can also drive transcription techniques applied to the audio, offering a support for the extraction of high-level features (cf. Section 2), or even complementing low-level features extracted from audio recordings (cf. Section 4). For a feasible task of music classification in real taxonomies, accurate transcription systems can be applied to audio files and seek for the same music features easily handled with symbolic formats. Music classification performed by humans is generally accomplished by high-level information, which is more accurately provided by symbolic formats. Such an aspect emphasizes the usefulness of studies on music genre classification with high-level features, whereas studies on low-level features should be pursued in parallel.

Notwithstanding, most studies adopt audio as the music representation. The comparison of their performance is a hard task, since different genre taxonomies are adopted. Over the past years, the availability of audio databases has increased, therefore, such a comparison has become more practicable. For the genre classification of audio signals, examples of available datasets include the dataset provided by ISMIR 2004 conference[1] and the databases of Homburg, Mierswa, Möller, Morik, and Wurst (2005) and Tzanetakis and Cook (2002). Symbolic music genre databases for academic purposes were made available by Goto, Nishimura, Hashiguchi, and Oka (2002), McKay (2004), and others (a complete list is provided in Appendix A).

Fu et al. (2011) conducted a review of audio-based music classification and annotation techniques, including a discussion on recent and principal studies in such a context and their description in terms of music features, classification schemes and performance accuracies. The survey included five main tasks, namely genre and mood classification, artist identification, instrument recognition and music annotation. Strum (2013) reviewed music genre recognition systems regarding the evaluation metrics adopted and showed a more systematic analysis of the system behavior should be taken into account in the evaluation of systems for music genre classification.

To the best of our knowledge, no systematic study has provided an overview of the frameworks developed for music genre classification based on symbolic data. This manuscript addresses a review of works that consider symbolic data through a summary of features and techniques used for symbolic-based music genre classification that highlights the music features, taxonomies, classification schemes and respective accuracies. Appendix A provides a list of databases and specifications on their availability for research purposes. The review can be of great value for the summarization of the most suitable features and approaches regarding symbolic genre classification and motivate audio transcription techniques toward such features. It makes sense to suppose, in the near feature, automatic transcription systems will enable audio recordings to be converted into symbolic formats and extraction of high-level features. Therefore, a summary of studies on high-level features could also be of great interest.

The remainder of the manuscript is organized as follows: Section 2 specifies the principal features adopted for the music classification tasks; Section 3 presents the learning algorithms and classification methods; Section 4 reports some hybrid approaches that use both audio and symbolic data for music genre classification; finally, Section 5 provides the main conclusions.

[1] http://ismir2004.ismir.net/genre_contest/index.htm.

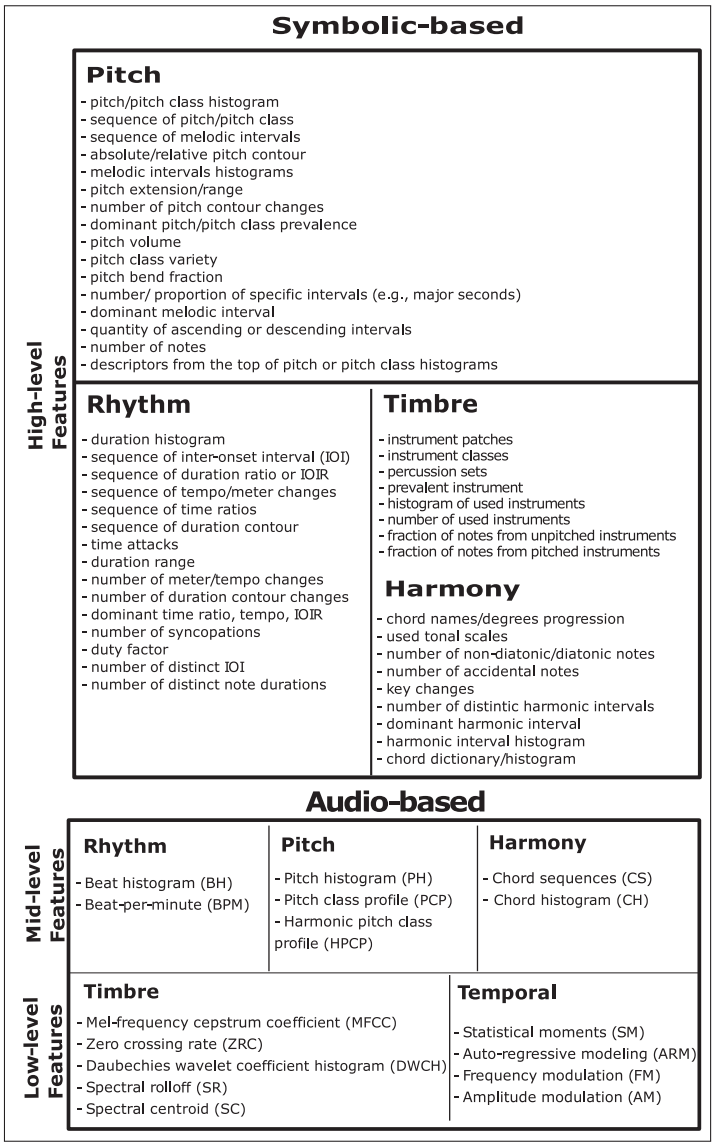


Fig. 1. Characterization of audio and symbolic-based features. The taxonomy of audio features is based on Fu et al. (2011).

2. Symbolic features

In the context of human perception of music, music categories usually reflect groups of songs that share a same interpretation of a feeling. Therefore, some specific music features are expected to be responsible for the classification of music as peaceful, sentimental or aggressive. Particularly, music genre offers an important description of emotions (Shan & Kuo, 2003; Shan, Kuo, & Chen, 2002) and are used for music classification. However, the determination of the features that better express a specific genre is a challenging aspect in music genre classification systems. Feature extraction is a crucial step for the classification of music content, and the adopted music features should reflect the significant and discriminative characteristics of different types of genres. The choice of suitable features is fundamental for successful classification tasks, whereas the performance of the classifier, although important, usually reflects the feature selectivity (Karydis, 2006).

The music descriptors used for music classification have covered the several dimensions of music, such as timbre, pitch, harmony and rhythm. Many authors have attempted to group them according to music dimensions, degree of abstraction or human music feeling (such as low-level, mid-level, high level or perceptual features) and degree of locality (such as short-term or long-term features) (Fu et al., 2011). As proposed by Fu et al. (2011), in the categorization of audio features, the standard taxonomies are applied hierarchically (Fig. 1).

In the literature on symbolic music classification, a standard taxonomy usually divides music descriptors into three groups, namely pitch, timbre and rhythm (Scaringella et al., 2006). While they are closely related to the mid-level descriptors for audio classification, they are mainly referred to as high-level descriptors, as they are extracted from music data that take into account a higher-level abstraction of music (the notes) instead of audio samples. The access to high-level music information is straightforward, since the symbolic format usually provides separated tracks or voices for each instrument and for the melody and vocals. Timbre features are generally addressed with the use of instruments patches, directly available in symbolic data. Mid-level and high-level music features are mainly related to the inherent characteristics of music as perceived and enjoyed by listeners. The most commonly used symbolic features for music classification are categorized in Fig. 1. Appendix C provides a list of the main studies that use distinct

types of music descriptors and categorized according to the temporal characteristics of music features.

Pitch/melodic and harmonic features are sometimes characterized in conjunction (Scaringella et al., 2006). Here, we considered pitch/melodic features when the model basically captured pitch, pitch class information, or sequential interval between pitches (e.g., pitch range, pitch contour, pitch histograms, melodic intervals). Harmonic features include chords, key, and harmonic intervals. In this case, the relationship between pitches and harmonic functions of chords progressions depends mainly on the ongoing tonality.

Music descriptors can also be divided according to their temporal characteristics. Event-based features model the music information in function of the sequential way they occur in time (sequential structure of the song), i. e., they capture the temporal patterns in music according to the time-order of music events and are also referred to as sequence-based features. Pitch contour information and sequence of melodic intervals are examples of pitch event-based features. Global features, on the other hand, ignore sequence information and try to summarize the entire song by single values. Global properties are often represented by a feature vector that can be viewed in a feature space. Pitch range, number of silences and distinct intervals are examples of global features.

Statistical functionals can be applied to all event-based features. They summarize the main aspects of a feature, as, for instance, its dispersion or central tendency. Measures commonly used are mean, standard deviation and minimum and maximum values of the features. For pitch, examples of statistical functionals are mean, minimum, maximum and standard deviations of melodic intervals, pitch volume and MIDI number. For rhythm, examples include the mean time between attacks, the mean metronome time, and the mean, minimum, maximum and standard deviations of rests and durations. Harmonic descriptors use, for instance, the mean and standard deviations of accidental notes and the minimum, maximum, mean and standard deviations of harmonic intervals.

The next sections summarize the main features adopted in symbolic music genre classification tasks.

2.1. Pitch-related features

Pitch, an essential aspect of music, is referred to in the literature as one of the most predominant musical characteristics (Karydis, Nanopoulos, & Manolopoulos, 2006). It is basically a perceptual attribute that enables sounds to be organized on a frequency-related scale (Haynes and Cooke). More specifically, it considers what the ear determines as the most fundamental frequency of a tone Haynes and Cooke, although it has a subjective character and contemplates the presence of harmonics (i.e. signal frequencies that are multiples of the fundamental frequency).

In studies on symbolic music data, pitches are usually represented as alphabetic letters with an octave indication (e.g., C4 for note C in the fourth octave—middle C), or MIDI numbers (e.g., 60 for C4, and 72 for C5) that range from 1 to 128. Most pitch-related features can be used in two different counterparts, i. e., folded, with octave normalization, and unfolded, without octave normalization. In the later case, the octave information is not taken into account, so that all notes are placed into a single octave, e.g., from C4 to B4. Such an abstraction is also called pitch class representation and the values range from 1 to 12. Folded pitch indication, on the other hand, considers the octave information, so that C notes that are one octave apart are considered two different notes, and, therefore, follow two distinct representations.

Perhaps the easiest way of capturing pitch information is through pitch histograms (PH). PH is a common pitch feature simple to be obtained and has showed considerably good performance for music genre classification tasks (Karydis, 2006; Tzanetakis, Ermolinskyi, & Cook, 2002; 2003). Basically, it models the distribution of pitches by indicating their frequency of occurrence in a song or a specific track. Each bin is indexed by the pitch name (or pitch number, e.g., MIDI note number) and its value indicates the frequency corresponding pitch has occurred in the melody. PH can be built in folded and unfolded fashions. Folded PHs are also established as pitch-class histograms, where adjacent bins indicate adjacent notes on the chromatic scale (C, C#, D,..., B4). In this case, a note is mapped to its folded version index according to the mod12 calculation (Tzanetakis et al., 2002).

Alternatively, folded PHs can be obtained in a circle of arrangement of fifths, i. e., adjacent bins in the histogram indicate notes separated by a fifth interval (rather than by a semitone) (Tzanetakis et al., 2002). The circle of fifths is closely related to the tonal dynamics of music and early results suggest features extracted from such a histogram yield more accurate classifications (Karydis, 2006; Tzanetakis et al., 2002). As the octave information is present in unfolded PH, C notes that are one octave apart are considered two different bins. Such possibilities enable the verification whether the capture of the range of a piece or octave independence could lead to different discriminative information for music classification.

Music descriptors can also be extracted on the top of the PH, which yields a lower dimensional feature space. For example, based on the folded and unfolded versions of pitch histograms, Karydis (2006) extracted four one-dimensional features, namely PITCH-FOLD, i.e., the bin index of the maximum peak of the folded histogram, AMPL-Fold, i.e., the amplitude (value) of the maximum peak of the folded histogram, PITCH-Unfold, i.e., the period of the maximum peak of the unfolded histogram, and DIST-Fold, which expresses the interval (in bins) between the two highest peaks of the folded histogram. Statistical moments, e.g. mean, standard deviation, entropy and uniformity can also be derived from pitch histograms and used for classification (Valverde-Rebaza et al., 2014).

Karydis et al. (2006) applied the notion of repeating patterns (RPs) and described them as important parts (sequence of notes) closely connected to phrases, motives or themes in a piece. According to the authors, an extended temporal dependency between pitches can increase the classification accuracies. They also adopted the string-join approach, denoted as RP-trees (Hsu, Liu, & Chen, 2001) to detect the non-trivial repeating patterns in a piece. Once all non-trivial patterns have been extracted, a co-occurrence matrix C, where $C(i, j)$ indicates the frequency at which pitches i and j co-appear in the same RP within a window of length w predefined by the author, is computed for each music piece. Several statistical features, such as variance, correlation, entropy, energy, and contrast are computed from each co-occurrence matrix. Then, a feature selection algorithm is adopted for forming the final feature set. The combination of the selected features and the three attributes from the duration histograms was applied for an intra-class music genre classification and reached a higher accuracy level for the same database in comparison to pitch histograms (Karydis, 2006).

Pitch sequence has also been explored in the literature from different perspectives. However, the main purpose is the capture of the temporal order of notes, in which time can be represented, for example, as beats or onset notes. In most studies, if two or more pitch events overlap in time, the highest one is taken. An absolute pitch sequence simply catches the temporal sequence of pitches, as MIDI numbers or pitch symbols and considers the relation between pitches and numbers exactly as they occur in an octave. A pitch class sequence maps all pitches in a single octave. However, the most common perspective is the use of melodic intervals, which express the temporal sequence in terms of distance (in semitones) between two adjacent pitches (see, for instance Hillewaere, Manderick, & Conklin, 2012;

Kotsifakos et al., 2013; Şimşekli, 2010). Relative melodic intervals use a reference note to determine the intervals. For example, if C4 is the reference note represented by number 0, C#4 is represented by +1, D4 is represented by +2, G4 is represented by +7 and B3 is represented by −1. The reference note can be determined according to the tonal center (Cruz-Alcázar, Vidal-Ruiz, & Pérez-Cortés, 2003), or the most frequent note in the song (Cilibrasi, Vitányi, & de Wolf, 2004b).

Melodic intervals are invariant under transpositions. They are usually expressed as symbols (Lin, Ning-Han-Liu, Wu, & Chen, 2004) that denote a type of interval, or numbers (e.g., 4 is an ascending major third, 3 is an ascending minor third, and −2 is a descending major second). For event-based features, a song is generally converted into a temporal sequence of intervals, according to its temporal pitch sequence. Similarly to pitch sequences, the melodic intervals of a song can be represented in a histogram, where bins indicate the relative frequency at which melodic intervals occurred within an octave (or across octaves) (Şimşekli, 2010).

A pitch contour (also referred to as pitch direction) is a simpler version of melodic intervals, since only the ascending or descending behavior of the interval sequence is captured. The reference pitch for the obtaining of a pitch contour can be the middle C or relative to some specific pitch (for example, the average pitch). Pitch sequence, pitch contour or melodic interval sequences are often used in conjunction with language models (e.g., *n*-grams) Conklin (2006); Hillewaere et al. (2012); Pérez-Sancho, Iñesta, and Calera-Rubio (2005) or probabilistic models (e.g., HMM, Markov chains) (Chai & Vercoe, 2001; Corrêa, Saito, & Costa, 2010).

In most studies, the information on pitch is extracted from the melody track. An exception is the work of Şimşekli (2010), in which the discriminative power of the pitches from the bass lines of different genres was analyzed. Bass is a low-pitched instrument and the author was motivated to use the bass line as the connection between the melody and the rhythm. While experiments were conducted on MIDI data, the availability of accurate systems for bass line transcription (such as in Ryynänen & Klapuri, 2007) make the study possible on polyphonic audio. Şimşekli (2010) used the temporal order in which the notes are played to construct normalized melodic interval histograms as key independent pitch features. The value ascribed to the bin represents the difference in semitones that two adjacent notes were played in the bass line, which implicit modeled the dependence of adjacent pitches. The magnitude of a bin indicates the relative frequency at which the respective interval occurred in a piece. The author also tested several distance metrics to analyze if the melodic interval histograms from distinct genres could be distinguishable.

Examples of global features of pitch are (i) pitch extension/range (usually expressed by the highest pitch minus the lowest one), (ii) minimum, maximum, average, and standard deviations of pitch and melodic intervals, (iii) number of pitch contour changes, (iv) pitch or pitch class prevalence, (v) dominant pitch class prevalence, (vi) pitch class variety, (vii) pitch bend fraction, (viii) relative frequency of specific intervals (as major seconds or major thirds), (ix) average pitch volume, (x) most used melodic interval, (xi) relative frequency of descending or ascending intervals, and (xii) number of notes. Fig. 2 shows some event-based and global pitch features for a melodic sequence extracted from J. S. Bach's Invention number 08.

2.2. Rhythmic-related features

Rhythm is another common music feature used for music classification and similarity (Fu et al., 2011; Gouyon & Dixon, 2005). It provides a perception of temporal regularity and can be basically understood as a specific pattern produced by notes that differ in duration, pause (silence) and intensity. In a more informal way, rhythm is related to the "danceability" of a song. Duration is the most explored aspect in rhythm in the context of symbolic genre classification. Intuitively, it reflects the number of notes played in a measure from the reference tempo (smaller durations enable more notes within a measure and lead to a faster rhythm).

A direct way of capturing rhythmic information is the use of duration histograms (DH), which are related to the beat histograms for audio analyses (Tzanetakis & Cook, 2002; Tzanetakis, Ermolinskyi, & Cook, 2003). Karydis (2006) proposes a duration histogram of 25 bins, where each bin represents the eight standard notes values, their dotted and double dotted versions and the breve duration. A DH is indexed by the duration value and the bin value reflects the frequency of occurrence of each of the 25 note values in a musical piece. Three features are extracted on top of the DH, namely duration of highest frequency, number of its occurrences and distance (regarding relative temporal duration) between the two most frequently durations (Karydis, 2006; Karydis et al., 2006).

Inter-onset interval (IOI) is a rhythmic descriptor similar to the note duration, however, it considers the presence of rests in a music piece. It describes the time interval (for example, in MIDI ticks) between the onset times of successive notes. In fact, the information of duration and IOI can be directly taken from symbolic data. Duration ratios and inter-onset time ratios (IOIR) commonly account for differences in speed or tempo in music sequences (Ruppin & Yeshurun, 2006). Two musical sequences can be considered equivalent if they show a fixed duration ratio (for example, two quarter notes and a half note compared to two eight notes and a quarter note). In the literature, IOI is expressed as

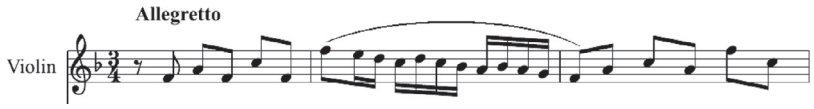
$$IOIR_i = \frac{Onset_{i+2} - Onset_{i+1}}{Onset_{i+1} - Onset_i} \quad (\text{for } i = 1, 2, \ldots, n-2) \tag{1}$$

where $Onset_i$ represents the beginning time of note i. The equation provides the IOIR of a sequence. Examples of its use are reported in Pérez-Sancho et al. (2005), de León, Iñesta, and Pérez-Sancho (2006, chapter 3), Pérez-Sancho, de León, and Iñesta (2006).

The duration sequence can also be represented by symbols (Lin et al., 2004) that express a relative duration between two consecutive notes. Another possibility is the representation of duration histograms in terms of IOIR and variations, since such features are time invariant.

Probabilistic models, as Markov chains (Gamerman & Lopes, 2006) can capture dependencies of note duration events. Markov chains determinate the probability of future events given the occurrence of one or more past events. The chain order is determined by the number of past events considered in the analysis. A first-order Markov chain takes into account only the predecessor of an event, while a second-order Markov chain considers two past events. In general, a Markov chain of order n is represented by a probability transition matrix of $n+1$ dimensions, which provides the probability of occurrence of an event, given the previous n states or events. The transition probability of pairs or triples of note durations was modeled as Markov chains in Corrêa et al. (2010), Corrêa, Costa, and Levada (2011a). First-order Markov chains tend to define subsequent relationships between more significant notes, whereas second-order Markov chains compute the probability of a sequence of three consecutive notes, capturing additional information on the dynamics and dependencies of notes values since they contemplate an extended music context. The authors suggest the rhythmic patterns established by Markov chains are sensitive to the genre discrimination, since some sequences of notes are common to all genres and others are distinct and characteristics of each one.

Other examples of event-based rhythmic features include (i) sequence of tempo or meter changes, (ii) sequence of time ratios, and (iii) duration contour change, which follows the same idea of pitch



(a) absolute pitch sequence

65 69 65 72 65 77 76 74 72 74 72 70 69 70 69 67 65 69 72 69 77 72

(b) pitch class sequence

6 10 6 1 6 6 5 3 1 3 1 12 10 12 10 8 6 10 1 10 6 1

(c) sequence of melodic intervals

+4 -4 +7 -7 +12 -1 -2 -2 +2 -2 -2 -1 +1 -1 -2 -2 +4 +3 -3 +8 -5

(d) pitch contour

+1 -1 +1 -1 +1 -1 -1 -1 +1 -1 -1 -1 +1 -1 -1 -1 +1 +1 -1 +1 -1

(e) pitch range (highest - lowest)

77 - 65 = 12

(f) minimum, maximum, average and standard deviation of pitch

65, 77, 70.45, 3.75

(g) minimum, maximum, average and standard deviation of melodic intervals

-7, 12, 0.33, 4.61

(h) number of pitch contour changes

13

(i) proportion of major thirds

0.14

(j) proportion of ascending interval

0.43

(k) pitch class prevalence

C - 0.23 C#(Db) - 0 D - 0.09 D#(Eb) - 0 E - 0.05 F - 0.27 F#(Gb) - 0
G - 0.05 G#(Ab) - 0 A - 0.25 A#(Bb) - 0 B - 0.09

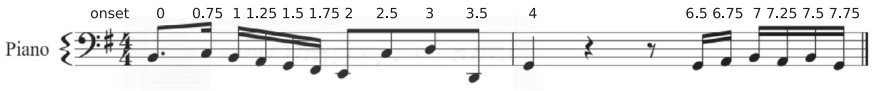
Fig. 2. Music descriptors of pitch. (a), (b), (c), (d): examples of event-based features. (e), (f), (g), (h), (i), (j), (k): examples of global features.

contour, e.g., +1 might indicate the subsequent duration is higher than the previous one, and −1 for the opposite. Examples of global features for rhythm classification are (i) minimum, maximum, average, and standard deviations of note durations and rests, (ii) duration range (usually expressed as the ratio between the maximum and minimum values), (iii) number of meter/tempo changes, (iv) number of duration contour changes, (v) mean metronome time, (vi) dominant tempo or time ratio, (vii) number of syncopations, (viii) duty factor, (ix) average time between attacks, (x) number of distinct IOI and (xi) number of distinct note durations.

Fig. 3 shows some event-based and global rhythmic features for the melodic sequence extracted from J. S. Bach's Invention number 07.

2.3. Harmonic-related features

In symbolic music genre classification, event-based harmonic features are often associated with the temporal progression of chords. A chord is a combination of at least three notes played simultaneously. The basic formation of a chord, usually known as a triad, consists of the root note (the one that establishes the chord name, for example, C), the third and the fifth. There are basically four types of triads, namely major, minor, augmented and diminished. Seventh chords include the seventh note (that forms the seventh interval from the chord's root) in the chord formation and basically exist in the following forms: major, minor, dominant,



(a) duration sequence

0.75 0.25 0.25 0.25 0.25 0.25 0.5 0.5 0.5 0.5 1 0.25 0.25 0.25 0.25 0.25 0.25

(b) IOI sequence

0.75 0.25 0.25 0.25 0.25 0.25 0.5 0.5 0.5 0.5 2.5 0.25 0.25 0.25 0.25 0.25

(c) duration ratio sequence

1/3 1 1 1 1 2 1 1 1 2 1/4 1 1 1 1 1

(d) duration contour sequence

-1 0 0 0 0 +1 0 0 0 +1 -1 0 0 0 0 0

(e) duration range (ratio between maximum and minimum durations)

$0.75/0.25 = 3$

(f) minimum, maximum, average and standard deviation of duration

0.25, 0.75, 0.3823, 0.2186

(g) number of duration contour changes

6

Fig. 3. Music descriptors of rhythm. (a), (b), (c), (d): examples of event-based features. (e), (f), (g): examples of global features. (timing resolution: one tick per quarter note).

half-diminished, diminished, augmented major and minor major (minor third and major seventh).

The chords perform different tonal functions, depending on their positions on the scale of the underlying tonality. Positions are commonly represented as degrees. For example, chord Mi minor (Em) can be the first degree of the Em key, or the third degree of the C major key, or the sixth degree of the G major key, and so on. The representation of chords as names does not capture their position on a scale. Roman numerals indicate the chord location related to the scale for the association of positions and functions. In the example above, chord Em would be represented as I, III and IV, respectively. The degrees can also include the chord extension and the seventh information. All such possibilities have been explored in symbolic music genre classification regarding harmonic features.

Fig. 4 illustrates different chords representations on C major tonality. The use of *n*-grams is common for the modeling of chord progressions (Pérez-Sancho, Rizo, & Iñesta, 2008a; 2009; Shan & Kuo, 2003).

A scale is a set of subsequent intervals that exhibit a specific structure. Different genres may use distinctive tonal scales in the music content. Abeßer, Bräuer, Lukashevich, and Schuller (2010), Abeßer, Lukashevich, and Bräuer (2012) explored tonal scales in the context of genre discrimination.

Examples of global features of harmony include (i) number of non-diatonic or diatonic intervals, (ii) number of accidental notes, (iii) average and standard deviations of accidental notes, (iv) number of key changes, (v) number of distinct harmonic intervals, (vi) dominant harmonic interval, (vii) maximum, minimum, average and standard deviations of harmonic intervals, (viii) harmonic intervals histogram, (ix) chord dictionary and (x) chord histogram.

2.4. Timbre-related features

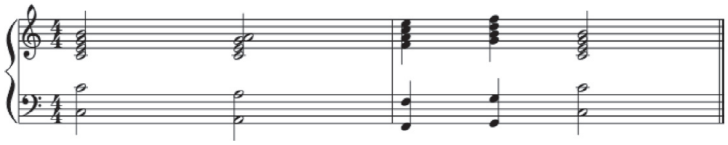
The notion of timbre in symbolic data is generally expressed by the analysis of the instrument patches, instrument classes and percussion sets. This information is provided in MIDI files by means of "voices" or patches. The number of possible instrument patches and percussion instruments used in a song is restricted to 128[2] when the General Standard MIDI (GSM) is used.

The 128 instrument patches of the GSM have been used in music classification frameworks to account for the timbre aspect. The representation is usually addressed by Boolean features, where a vector $\vec{x}$, with components $x_i \in \{0, 1\}$, indicates the presence or absence of a patch i in the instruments (Basili, Serafini, & Stellato, 2004; Pérez-García, Iñesta, & Rizo, 2009). Another possibility of representing instrument patches is the grouping of drum instruments into categories, according to their characteristics (Hübler & Hoffmann, 2013).

The General MIDI standard map assigns percussion instruments to channel 10. Each MIDI note number corresponds to a unique percussion instrument. Studies on instrument features for music genre classification deal with such a percussion mapping in several ways. For example, Basili et al. (2004) used the eight different drumsets associated with channel 10 and represented them as Boolean features. Pérez-García et al. (2009) used the number of MIDI patches in a song and incorporated the percussion sets by grouping the percussion instruments into three categories, namely (i) instruments usually present in drumkits, (ii) Latin percussion instruments and (iii) other percussion instruments.

McKay (2004) specified the following 20 music features related to instrumentation: pitched instruments present in the song,

[2] http://en.wikipedia.org/wiki/General_MIDI.



(a) chord names

C7M Am7 F7M G7 C7M

(b) chord names as triads

C Am F G C

(c) degree names

I7M VI7 IV7M V7 I7M

(d) degree names as triads

I VIm IV VI

(e) degree names without seventh and triad information

I VI IV V I

(f) bigrams with simplified degrees

I - VI VI - IV IV - V V -I

(g) trigrams with simplified degrees

I - VI - IV VI - IV - V IV - V - I

Fig. 4. Examples of representations of chord progressions.

unpitched instruments present in the song, note prevalence of pitched instruments, note prevalence of unpitched instruments, time prevalence of pitched instruments, variability in note prevalence of pitched instruments, variability in note prevalence of unpitched instruments, number of pitched instruments, number of unpitched instruments, percussion prevalence, string keyboard fraction, acoustic guitar fraction, electric guitar fraction, violin fraction, saxophone fraction, bass fraction, woodwinds fraction, orchestral strings fraction, string ensemble fraction, and electric instrument fraction. McKay and Fujinaga (2005a) observed the time prevalence of pitched instruments was the most significant instrumentation feature. In the same study, instrumentation showed the most important group of features among pitch, rhythm, dynamics, texture, and melody.

Other possible timbre-related features are (i) histogram of instruments used, (ii) total number of instruments played, (iii) fraction of note-ons belonging to unpitched instruments and (iv) fraction of note-ons belonging to pitched instruments (McKay, 2003).

2.5. Combined features

The combination of pitch and rhythm is predominant in the literature on symbolic music genre classification and can be configured in different ways. Karydis (2006) combined four features from pitch histograms (Section 2.1) and three features from duration histograms (Section 2.2) for forming a seven-dimensional feature vector. Since the PH can exist in folded and unfolded versions, the author provided both versions of the combination. The seven-dimensional feature vector is also tested in a weighted fashion. The contribution of the two groups of features was predicted according to a weighting scheme that applies a weight $0 \leq w \leq 1$ to pitch features and $1 - w$ to rhythm features.

Anan, Hatano, Bannai, and Takeda (2011) used the main melody to extract a sequence of notes and rests. In this case, a note expressed both pitch and duration, while a rest was expressed only by a duration value. The duration reference was determined by the sixteenth note. The note events were converted to string data and three types of note string were computed: pitch string, rhythm string and note string. For the pitch string, each note or rest is represented by the corresponding letter in an alphabet of 13 possibilities (octave independence). The number of letters of each note or rest depends on the relative duration value in sixteenths. For example, the C quarter note is represented by C C C C, while the D eighth note is represented by D D. The idea of dividing each note or rest into sixteen notes (or rests) is kept for the rhythm string, where the size of the alphabet is four and the symbols used are N, when a note starts, n, for a non-starting note, R, when a rest is starting, and r, for a non-starting rest. Finally, the note string reflects the combination of pitch and rhythm strings and the size of the alphabet is 26. For example, G quarter note is represented by G g g g, while a rest of the eighth note is represented by R r. For instance, the second measure of the melodic sequence in Fig. 3 might have the following string representation: pitch string = G G G G G A B A B G; rhythm string = N n n n R r r r r r N N N N N N; and note string = G g g g R r r r r r G A B A B G. Similar representations of notes can be found in Chai and Vercoe (2001); Lin et al. (2004).

Ruppin and Yeshurun (2006) extracted invariant pitch and rhythmic features by defining music transformations, such as transposition, augmentation/diminution, sequential modulation

and crab form and represented music as a function of time, $f(t)$, that reflects n-dimensional pitch and duration vectors. To ensure transposition and time invariance, the derivatives of the pitches and durations $\frac{d}{df}f(t)$ are computed for the determination of relative pitch and time, although the authors suggest note duration changes are better captured with the use of time ratios, for instance, the duration logarithm. Only the pitch direction is preserved for the Crab transformation, while the absolute pitch direction is discarded (a process referred to by the authors as the second derivative).

Apart from those combinations of pitch and rhythm, some studies on symbolic music classification adopt a set of diverse music properties to cover additional aspects of music, as volume, dynamics or musical texture. Dannenberg, Thom, and Watson (1997) used a set of 13 features based on pitch, duration, duty factor and volume. McKay and Fujinaga (2004) extracted 109 high-level features from MIDI files related to dynamics, instrumentation, pitch, melody, rhythm and chords. Posteriorly, the authors extended the set of features to comprise 111 (McKay & Fujinaga, 2005b) and 160 (McKay & Fujinaga, 2006a) high-level music features. Examples of features are fraction of melodic intervals comprising a tritone, average time between attacks, prevalence of most common vertical interval and average changes in loudness. A detailed description of the features can be found in McKay (2004). Those 111 features were also used by DeCoro, Barutcuoglu, and Fiebrink (2007) for the testing of a Bayesian aggregation classification.

Additional general features include number of notes, track duration, number of pauses, average number of notes per beat and average duration of melodic arcs. Some studies have also included features related to normality using the Dágostino statistics, such as pitch distribution normality, note duration distribution normality, interval distribution normality, IOI distribution normality, among others (de León & Iñesta, 2004a; 2007; de León et al., 2006; de León, Iñesta, & Rizo, 2008).

3. Classification tasks

In this section we review and summarize the principal techniques used for symbolic music genre classification. The approaches are separated into two perspectives, namely supervised classification (Section 3.1) and clustering (Section 3.2).

3.1. Music genre classification as a supervised classification problem

The process of automatic classification of music can be divided into feature extraction and classifier design (Costa & César, 2001; Duda, Hart, & Stork, 2001). The first step consists in the summarization of music examples by music descriptors (generally referred to as feature vectors), which can be viewed as data-points in the feature space. The labels of a subset of music examples (training set) are used in the testing of music features for a designed classification rule. The main objective is to find a classification rule that predicts the labels of unseen music examples (testing set) with the minimum prediction error (Duda et al., 2001). For an interesting study on the comparison of machine learning techniques, the reader is invited to read the survey of Amancio and others (Amancio et al., 2014).

The choice of learning algorithms and decision making processes tends to depend on the type of feature used (event-based or global). For instance, statistical and information-theoretical tools are more often used for modeling event or sequence-based features, due to their inherent characteristics of capturing temporal aspects in music. Moreover, this type of feature requires an approach that combines local decisions or probability scores for the obtaining of a final classification decision. Common choices of such methods are n-grams, hidden Markov models (HMM), Markov chains and the Naive Bayes classifier. Global features, on the other hand, aim at summarizing meaningful information into single values, and a simple decision making process is sufficient.

Common choices of classifiers for symbolic music content are k-nearest neighbor (k-NN), support vector machines (SVM), and Bayesian classifier. k-NN and SVM are also the most common classifiers for the genre classification of audio signals (Fu et al., 2011). Artificial neural networks (ANN), self-organizing maps (SOM), HMM, logistic regression and associative rules have also been adopted for the symbolic music genre classification. Basili et al. (2004) tested the performance of voting feature intervals (VFI), Quinlan algorithm, projective adaptive resonance theory (PART), Nearest neighbor with generalization (NNge), and repeated incremental pruning to produce error reductions (RIPPER).

k-NN is a simple algorithm that uses the labels of the (k) nearest samples (feature vectors) provided by the training set to predict the label of a testing sample. Euclidean distance is commonly adopted as a distance metric, however, other distance measures can be applicable (see, for example, Kotsifakos et al., 2013; Şimşekli, 2010). SVM (Boser, Guyon, & Vapnik, 1992; Cristianini & Shawe-Taylor, 2000) is a binary classifier based on the maximum-margin principle, which has solid bases in the statistical learning theory by Vapnik (1995), Vapnik and Chervonenkis (1971). Basically, it uses labeled samples from a training set to find the optimal hyperplane that best separates the samples from the two classes. The hyperplane is calculated so that the distance between the nearest samples from each class and the hyperplane is maximized. Such samples are referred to as supporting vectors and represent the closest samples to the hyperplane with labels most likely to be overlapped.

The Bayesian classifier uses the Bayesian Decision Theory (Duda et al., 2001), in which class conditional probability densities (likelihood) and prior probabilities (prior knowledge) are estimated from a training set. Classification is achieved through the assignment of each sample to the class of maximum a posteriori probability:

$$y_{MAP} = \underset{i}{\operatorname{argmax}}\, P(y_i|x), \tag{2}$$

where

$$P(y_i|x) = \frac{P(y_i)P(x|y_i)}{P(x)}. \tag{3}$$

The method in Eq. (2) is known as MAP (Maximum A Posteriori) estimate; y specifies the class labels and x is a given sample. If the conditional densities are properly estimated (suitable number of examples), the Bayes classification rule is considered optimal (Duda et al., 2001). As in a real problem a suitable number of examples might be difficult to be accomplished, hypotheses on the data are usually assumed, which configures different classifiers. The most common solution seems to be the use of Gaussian naive Bayes classifier, which assumes features in each class are independent and normally distributed, therefore, the product rule can be applied for the estimation of the conditional probabilities (Duda et al., 2001).

Table 1 shows a representative list of methods for symbolic music genre classification. Since symbolic databases for performance comparisons are less usual than audio databases, the number of genres (# C) and songs (# S) is specified. The features and classifiers adopted are also displayed for each reference. When possible, we tried to cover different features for the same datasets for a simple comparison of classification accuracies (percentage of songs correctly classified). All results are expressed in terms of classification accuracy (Ac), as reported in the original paper.

Table 1 shows the most important studies on symbolic genre classification. The results reported can assist in the identification of interesting features and methods for the classification task. Section 3.3 addresses discussions on the topic.

Table 1
Features and classifiers used in symbolic music genre classification.

Reference	# C	# S	Features	Algorithm	Ac (%)
Dannenberg et al. (1997)	8	200	Average and standard deviations of MIDI key number, duration, pitch, duty factor and volume + number of notes, pitch bend messages and volume change messages (13 features)	Naive Bayes	90
Cruz-Alcázar and Vidal-Ruiz (1998)	3	100	Melodic intervals + duration	Grammatical Inference	95
Chai and Vercoe (2001)	3	491	Pitch intervals	HMM	63
			Pitch contours		59
Shan and Kuo (2003)	2	190	Chord degrees progression	Associative rules	84.2
Basili et al. (2004)	6	300	Relative frequency of pitch intervals	Naive Bayes	36[3]
			Note extension		26[3]
			Changes in meter/time		41[3]
			Instrument patches		72[3]
			Instrument classes		61[3]
			[All features]		65[3]
Cruz-Alcázar et al. (2003)	3	100	Melodic intervals + duration	*n*-grams	97
				Grammatical Inference	92.3
McKay (2003)	**3**	100	20 global features of pitch, rhythm and timbre	ANNs	85
	9				65
Li and Sleep (2004a)	4	749	Statistics from pitch, duration and a bi-gram model	SVM + hierarchical	75.7
McKay and Fujinaga (2004)	**3**	950	109 features of chords, dynamics, instrumentation, musical texture, pitch and rhythm	(hier) *k*-NN and ANNs	98
	9				90
	3			(flat) *k*-NN and ANNs	96
	9				86
	9			(hier) *k*-NN and ANNs	81
	38				57
Li and Sleep (2004b)	4	771	Pitch intervals	*k*-NN	92.4
			Absolute pitch		88.1
McKay and Fujinaga (2005b)	9	225	111 features of chords, dynamics, instrumentation, pitch, musical texture and rhythm	(flat) *k*-NN and ANNs	84.4
				(hier) *k*-NN and ANNs	90
	38	950		(flat) *k*-NN and ANNs	46.1
				(hier) *k*-NN and ANNs	64.3
Pérez-Sancho et al. (2005)	3	300	*n*-grams of pitch intervals and IORs	Naive Bayes	93.7
Karydis (2006)	5	250	Features extracted from pitch histograms (unfolded)	*k*-NN	56
			Features extracted from pitch histograms (folded)		55
			Features extracted from duration histograms		58
			Features from pitch histograms and duration histograms		66
			Weighted scheme: features from pitch histograms and duration histograms		70
Karydis et al. (2006)	5	250	Weighted scheme: pitch statistical measurements from the **entire piece** + features from duration histograms	*k*-NN	75
			Weighted scheme: pitch statistical measurements from **repeated patterns** + features from duration histograms		92
Li et al. (2006)	6	14000	Pitch and duration sequences	Fact. Language Model	72.5
Ruppin and Yeshurun (2006)	3	50	Pitch intervals + time ratios	*k*-NN	80
			Pitch contours + time ratios		85
Juhász (2006)	6	9043	Pitch-time contours (pitch as integer number + duration)	SOM	90
Conklin (2006)	2	337	Segmented viewpoints of pitch	*n*-grams + Bayesian fwk	98
DeCoro et al. (2007)	38	950	111 features of instrumentation, rhythm, dynamics, and chords	SVM + Bayesian Agg	60.1
Kofod and Ortiz-Arroyo (2008)	9	225	1024 features of chords, dynamics, pitch, rhythm, musical texture and instrumentation	Hidden Naive Bayes	90
	38	950			64
Abeßer et al. (2009)	6	300	154 bass-line features of pitch, rhythm, harmony and structure	K-NN	81.5
Hillewaere et al. (2009)	6	3367	62 global features from McKay and Fujinaga (2004)	Logistic regression	67.8
			150 global features (122 of pitch and rhythm)	SVM	69.7
			Pitch intervals + duration ratio	*n*-grams	72.7
Abeßer et al. (2010)	8	320	136 bass-line features of pitch, rhythm, and harmony	SVM	60.8
Şimşekli (2010)	3	225	pitch interval histograms	*k*-NN	100
	9				86.6

(*continued on next page*)

Table 1 (*continued*)

Reference	# C	# S	Features	Algorithm	Ac (%)
Corrêa et al. (2010)	4	280	first-order Markov chains of durations	Quadratic Bayesian	70
Corrêa, Costa, and Saito (2011b)			second-order Markov chains of durations		87.5
Cuthbert et al. (2011)	2	1943	174 global pitch and rhythmic features	Naive Bayes	91
Abeßer et al. (2012)	13	520	101 bass-line features of pitch, rhythm and harmony	Regression Tree	64.8
Lin and Chen (2012)	5	2187	Melodic patterns as a sequence of pitch intervals + duration ratios	Pattern clustering	70
Hillewaere et al. (2012)	9	2198	Melodic intervals	*n*-grams	66.1
			IOI		76.1
			22 global features of duration	*k*-NN	69
Khoo et al. (2012)	5	312	*solfege*, interval and duration features	Feature Density Map	72.1
Velarde et al. (2013a)	26	360	Sequence of notes represented as MIDI note numbers	K-NN	88
			Wavelet representation of note sequences		85.5
Velarde, Weyde, and Meredith (2013b)			Wavelet representation of note sequences from melodic segments		85.5
van Kranenburg et al. (2013)	26	360	Melodic interval sequence	K-NN	92
			Global pitch features		74
			Duration ratio sequence		74
			Global rhythm features		55
			Combination of global pitch and rhythm features		82
			Combination of local features		99
		4470	Melodic interval sequence		60
			Global pitch features		29
			Duration ratio sequence		33
			Global rhythm features		13
			Combination of global pitch and rhythm features		34
			Combination of local features		73
Conklin (2013)	6	3367	viewpoints of melodic intervals and IOIs	Ensemble of Bayesian classifiers	72.4
			Viewpoints of melodic interval, absolute pitch, duration and onset		76.1
			Viewpoints of melodic interval, IOI, pitch class, pitch contour, duration contour and duration ratio		79.2
Hübler and Hoffmann (2013)	4	504	Drum patterns	Transducers Models	88.8
Kotsifakos et al. (2013)	4	100	Pitch interval + IOIR	*k*-NN	40
Valverde-Rebaza et al. (2014)	4	919	Chord sequences	network-only link-based	0.96[4]
Hedges et al. (2014)	9	434	Chord progressions	viewpoints of chords, pitches and durations	57.6

[3] Precision.
[4] AUC measure—area under the receiver operating characteristic curve.

Some studies have adopted feature extraction, feature selection or feature space transformation techniques before the classification methods. Such techniques include Principal Component Analysis (PCA) (Corrêa et al., 2010; de León & Iñesta, 2004a; Ellis & Arroyo, 2004; Hillewaere, Manderick, & Conklin, 2009), Linear Discriminant Analysis (LDA) (Abeßer, Lukashevich, Dittmar, & Schuller, 2009; Corrêa et al., 2010), Generalized Discriminant Analysis (GDA)) (Abeßer et al., 2009), Self-Organizing Maps (de León & Iñesta, 2003; 2004b; de León, Pérez-Sancho, & Iñesta, 2004), Inertia Ratio Maximization with Feature Space Projection (IRMFSP) (Abeßer et al., 2010; Abeßer et al., 2009), a hybrid feature selection algorithm based on Bayesian Expectation Maximization (Karydis et al., 2006), Average Mutual Information (AMI) (de León et al., 2006; Pérez-García, Pérez-Sancho, & Iñesta, 2010; Pérez-Sancho et al., 2006; Pérez-Sancho et al., 2005; Pérez-Sancho et al., 2008a; 2009), genetic algorithms (McKay, 2004; McKay & Fujinaga, 2005a; 2005b), and CfsSubset—filtering type and correlation based feature selection algorithm (Kofod & Ortiz-Arroyo, 2008).

Studies in a same database are less common for symbolic music genre classification, since efforts for making symbolic datasets available are recent. Tables 2 and 3 show two sets of studies that used the same dataset for an initial comparison of music descriptors regarding classification performance. The corpus for the results of Table 2 comprises 45 classical and 65 jazz pieces. For the final dataset, the authors split the MIDI sequences into segments of $w = 8$ bars (de León & Iñesta, 2002; 2003; 2004b), which yielded 522 classical samples and 430 jazz samples. They adopted integer values $w \in [1, 100]$ or a $w - g$ parametrization to determine the length of the segment (de León & Iñesta, 2004a; 2007; de León et al., 2004; Pérez-Sancho et al., 2006).

Table 3 shows the accuracies for a corpus of 856 songs. The root genres are popular (283), jazz (378), and academic (235). Popular songs were further divided into blues (84), pop (100), and Celtic (99). Jazz songs were separated into pre-bop (178), bop (94), and bossa nova (66), while academic songs were categorized into baroque (56), classical (50), and romanticism (129). The classification tasks involved the discrimination of the root and leaf genres.

3.2. Music genre classification as a clustering problem

The classifiers presented in the previous section are considered predictive models, i.e., given a set of labeled samples (training set), an estimator for the prediction of the label of an unknown sample can be constructed. They are also referred to as supervised learning. In a different perspective, descriptive models (also referred to as unsupervised learning or clustering) aim at inferring relevant information from the data. Such models are generally based on similarity measures.

The Euclidean and cosine distances are usually adopted for numerical data, although the choice of the appropriated metric

Table 2
Accuracies for the same database: 65 jazz and 45 classical pieces (split into short segments).

Reference	Features	Algorithm	Ac (%)
de León and Iñesta (2002)	21 melodic and harmonic statistical descriptors	ANN - SOM	(m) 77.2
de León and Iñesta (2003)	7 melodic and rhythmic statistical descriptors	ANN - SOM	84.2
de León and Iñesta (2004a)	12 melodic, rhythmic and harmonic statistical descriptors	Bayesian	89.5
de León et al. (2004)	10 melodic and rhythmic statistical descriptors	*k*-NN	83.3
de León and Iñesta (2004b)	7 melodic and rhythmic statistical descriptors	*k*-NN	88.7
Pérez-Sancho et al. (2005)	*n*-grams of pitch intervals and IORs	Naive Bayes	88.4
de León et al. (2006)	28 melodic, rhythmic and harmonic statistical descriptors	*k*-NN	97.3
	n-grams of pitch interval and IOR	Naive Bayes	92.3
Pérez-Sancho et al. (2006)	28 melodic, rhythmic and harmonic statistical descriptors	*k*-NN	95.2
de León and Iñesta (2007),de León et al. (2008)	28 melodic, rhythmic and harmonic statistical descriptors (segments)	*k*-NN	89.2
	28 melodic, rhythmic and harmonic statistical descriptors (melody)	*k*-NN	93

Table 3
Accuracies for the same database: 856 songs for taxonomies of three and nine classes.

Reference	Features	Algorithm	Ac (%)
	Three classes		
Pérez-Sancho et al. (2008a)	chords vocabulary	*n*-grams	86 ± 3
	pitch intervals + duration ratios		80 ± 6
	chords vocabulary + pitch intervals + duration ratios	classifiers ensemble + hier.	90.1
Anglade et al. (2009a)	chords names + extensions	grammars + decision tree	66.5
Anglade, Ramirez, and Dixon (2009b)	chords degrees + extensions		80.8
Pérez-García et al. (2009)	instrument + percussion patches	Naive Bayes	93 ± 2
Pérez-Sancho et al. (2009)	chord degrees + extensions	*n*-grams	87 ± 4
	chord names + extensions		86 ± 4
	chord degrees (no extensions)		83 ± 4
	chord names (no extensions)		82 ± 5
	chord degrees + extensions	Naive Bayes	86 ± 4
	chord names + extensions		83 ± 3
	chord degrees (no extensions)		72 ± 5
	chord names (no extensions)		74 ± 6
Pérez-García et al. (2010)	chord vocabulary + instrument patches + percussion sets	Naive Bayes	95 ± 2
	Nine classes		
Pérez-Sancho et al. (2008a)	chords vocabulary	Naive Bayes	62 ± 6
	melodic words (intervals + duration ratios)		55 ± 12
	chords vocabulary + pitch intervals + duration ratios	classifiers ensemble + hier.	63.4
Pérez-Sancho et al. (2009)	chord degrees + extensions	Naive Bayes	62 ± 3
	chord names + extensions		64 ± 4
	chord degrees (no extensions)		48 ± 7
	chord names (no extensions)		50 ± 9
	chord degrees + extensions	*n*-grams	41 ± 8
	chord names + extensions		40 ± 10
	chord degrees (no extensions)		50 ± 11
	chord names (no extensions)		50 ± 9
Pérez-García et al. (2009)	instrument and percussion patches	Naive Bayes	68 ± 5
	chord vocabulary + instrument patches + percussion sets		79 ± 3

depends on the analysis of the problem (Gan, Ma, & Wu, 2007). For string data, common measures of distances are normalized compression distance (NCD) and the edit-distance (Levenshtein distance) (Li, Chen, Li, Ma, & Vitányi, 2004; Müllensiefen & Frieler, 2004). Basically, they compute the minimal number of edit operations required for the transformation of a string into another. The motivation is the estimation of how much information two songs share.

Corrêa et al. (2010), Corrêa et al. (2011a) tackled the classification of music genres from two clustering perspectives using a hierarchical clustering algorithm and a community detection in complex networks. The features correspond to rhythmic patterns determined by the temporal dependency of the musical notes present in the percussion. Each song is represented by a vector of conditional probabilities between pairs and triples of notes computed by the use of first- and second-order Markov chains. For the complex network modeling, each vertex represented a song, while the edges between vertexes indicated the distance between two songs by means of their respective feature vectors.

Anan et al. (2011) addressed the problem of music genre classification by combining the (dis)similarity-based learning approach proposed by Wang, Sugiyama, Yang, Hatano, and Feng (2009) and the 1-norm soft margin optimization formulation. They compared the results with those of support vector machines and string kernels and showed the approach proposed outperformed SVMs. For SVMs, they used *n*-gram and mismatch kernels. Moreover, the length of the Longest Common Subsequence (LCS) and the edit distance were adopted as dissimilarity measures.

3.3. Discussion

The selection of suitable features is an essential step in the development of a classification system. In music genre, this process is especially important because of the redundancy of genre taxonomies. Good features are expected to be appropriate for the discrimination of music genres and connection of the feature space to genre labels with the minimum prediction error. Tables 1–4 show the principal approaches for symbolic music genre classification and represent the main contributions of our research. Based on the results reported by those tables and on the information collected during the research, we provide in the following discussions for a better comprehension of music features and their discriminative

Table 4
Similarity metrics used in symbolic music genre classification.

Reference	# C	# S	Features	Metric	Ac (%)
Lin et al. (2004)	7	500	Repeated patterns of pitch intervals	Edit distance and dynamical programming	49.2[5]
			repeated patterns of relative duration		40.2[5]
Cilibrasi et al. (2004b), Cilibrasi et al. (2004a)	3	36	Melodic interval with modal note as a reference	Hierarchical clustering	0.8 [6]
Abeßer et al. (2010)	8	320	136 bass-line features of pitch, rhythm, and harmony	Aggregation similarity measures	68.5
Corrêa et al. (2010)	4	280	First-order Markov chains of durations	Cosine distance + hierarchical clustering	36.5
Corrêa et al. (2011a)			second-order Markov chains of durations	Cosine dist. + community detection	44
Anan et al. (2011)	2	238	Pitch sequence	Dissimilarity-based learning: LPBoost + edit distance	86.1
			Pitch sequence - transposed to C		86.1
			Rhythmic sequence		92.9
			Note sequence		87.4
Hillewaere, Manderick, and Conklin (2014)	26	360	Mel. interval sequence	Edit distance	94.4
			Duration ration sequence		80.3
	6	3367	Mel. interval sequence		49.5
			Duration ration sequence		47.5
	9	2198	Mel. interval sequence		50
			Duration ration sequence		63.2

[5] Precision.
[6] Score.

power for symbolic music genre classification. However, such results strongly depend on both the algorithm and the database.

Melody seems to be the predominant musical dimension taken into account in the studies analyzed so far. Previous research suggests it makes music memorable and enables people to distinguish one song from another (Selfridge-Field, 1998; Snyder, 2009). Moreover, the chorus (repeating pattern of a melody) is generally repeated several times and is the most memorable part of the melody. The study of Shan et al. (2002) indicated that the overall accuracy for pairs of genres is not significantly influenced if only the chorus is used, rather than the whole melody, which is an indication features extracted from the chorus would be enough to represent music genres. Ruppin and Yeshurun (2006) also reported the importance of repetition as an essential aspect for music interpretation of music.

The studies consider different approaches to obtain a monophonic melody from MIDI files. The most common approach is the merging of all tracks, except the percussion ones, and their combination in a single track. If two or more note events occur simultaneously, only the note of highest pitch is preserved. This method is known to provide good results (Uitdenbogerd & Zobel, 1998). In de León et al. (2008), the melody track is selected by a random forest classifier according to the music content information of the tracks. Other possibilities include a manual specification of the principal melody track, use of measures, as entropy for inference on the melody, or use of transcription systems. Shan et al. (2002) modified the all-mono algorithm (which removes all simultaneous notes except for the highest one) to obtain the melody track of polyphonic MIDI files. Studies that do not use the melody track for the classification of music genres consider the extraction of music features from other instruments as the bass lines (Abeßer et al., 2010; Abeßer et al., 2012; Abeßer et al., 2009; Şimşekli, 2010), analysis of percussion tracks (Corrêa et al., 2011a; Corrêa et al., 2010), and analysis of many tracks in MIDI files (Basili et al., 2004; Dannenberg et al., 1997; DeCoro et al., 2007; Hillewaere et al., 2009; 2012; McKay & Fujinaga, 2004; 2005a; 2005b; Valverde-Rebaza et al., 2014).

Tables 1–4 suggest music descriptors from pitch and rhythm are predominant in the classification of music genres from symbolic data. This aspect was also considered in different reviews (for instance, Fu et al., 2011; Scaringella et al., 2006). The tables also indicate in general, rhythmic descriptors overcome the pitch ones for Western genres. An exception is the study of McKay and Fujinaga (2010), in which pitch features show a higher relative importance (27.8%) in comparison to rhythmic features (19.5%). Rhythm is a genuine and inherent characteristic of Western music genres largely explored in MIR tasks (Fu et al., 2011; Scaringella et al., 2006). However, for folk songs that are melodies from oral traditional cultural material, rhythmic-related features are usually outperformed by pitch-related features. This finding is interesting and was not reported in the literature before. However, more systematic experiments are required for an accurate confirmation of such a statement.

Regarding pitch descriptors, results indicate the representation of melodic intervals is prevalent and contribute to the extraction of more discriminative information among genres than using absolute pitches or pitch contours. When absolute pitches are taken into account, a preprocessing step is usually adopted, which consists of the transposition of all melodies to a same tonality, although this approach seems do not guarantee an increase in the classification performance (Anan et al., 2011). The results of folding and unfolding are quite similar (Karydis, 2006) and the identification of repeated patterns in a melody seems to be an important step for a better classification performance (Karydis et al., 2006; Lin & Chen, 2012; Lin et al., 2004). Duration ratios, variations in IOIR, and descriptors derived from duration seem to be suitable music features for rhythm. All such sceneries are predicted in the literature for the classification of audio signals and were confirmed by the collected results on symbolic data.

Descriptors from instruments are significant for symbolic music genre classification, as high classification accuracies are achieved when instrument patches are taken into account. The results in Tables 1 and 3 show features related to instrumentation performed better than rhythmic descriptors or chord sequences, which is in agreement with a systematic study on the importance of instrument identification (McKay & Fujinaga, 2005a).

Harmonic features are less explored in symbolic data and covered in terms of chord progressions. The studies analyzed so far have shown modeling chords as degrees (representing their functional position on the scale) is a good choice that yields better performance in comparison, for example, to chords names (Hedges, Roy, & Pachet, 2014). Moreover, the inclusion of chord extensions

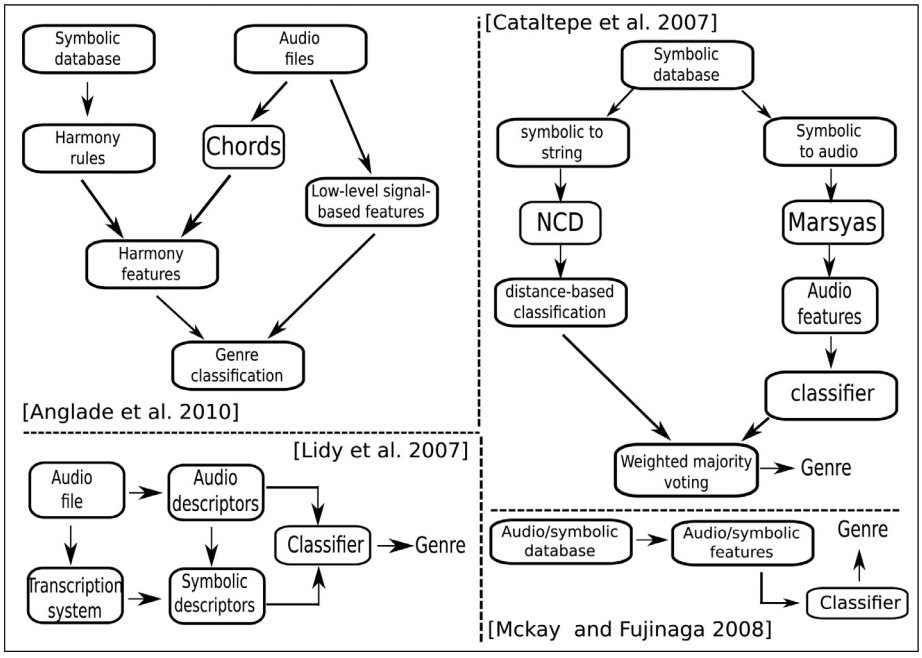


Fig. 5. Block diagrams of different approaches that combine audio and symbolic features for music genre classification.

seems to be discriminative information in the classification (Pérez-Sancho, Rizo, & Iñesta, 2009).

Most studies reported in this survey adopted a supervised learning model and the most common classifiers for symbolic genre classification are SVM, K-NN, Naive Bayes and Neural Networks, which seems to provide a feature space that capture with success relevant music properties for the identification of distinct music genres. Suitable descriptors associated with genres are extracted from the song content and used as input for a learning algorithm that must learn a function and predict new instances with good performance. The direct comparison of the classifiers performance is subjected to many conditions that are not accessible using only the classification accuracies reported in the studies.

Clustering techniques have been little explored for tasks involving symbolic or audio-based music genre classification. According to McKay and Fujinaga (2004), Scaringella et al. (2006), the notion of genre can disappear if the formed groups are based only on similarity patterns. These groups may not reflect a taxonomy of genres closely related to the ones used by humans, since such a taxonomy is inconsistent and comprehends aspects beyond music similarity. However, as groups obtained by clustering methods are formed according to an objective similarity function, subjective aspects tend to be minimized. For symbolic data, clustering strategies in Table 4 include the edit distance combined or not with other methods (for instance, dynamical programming), hierarchical clustering and community detection in complex networks. Results from Table 4 indicate better accuracies were achieved for relatively small datasets, while for a larger dataset, the average accuracy was approximately 50%. For the classification of audio data, studies that adopt an unsupervised perspective for the discrimination of musical genres use k-means, SOM and HMM methods, and the clustering performance seems to be comparable to that of supervised techniques (Mostafa & Billor, 2009; Shao, Xu, & Kankanhalli, 2004).

Event-based features tend to perform better than global descriptors, since they capture the sequential information from music events, as verified in the literature (for instance, Hillewaere et al., 2009; van Kranenburg, Volk, & Wiering, 2013) and confirmed by the reports in Tables 1 and 3. Hillewaere et al. (2009) showed classification accuracies of four well-known global feature sets with the use of standard learning algorithms were outperformed by event-based models of pitch and rhythm for folk song classification. However, while the extraction of event-based features is trivial for monophonic melodies, its application for polyphonic music may reveal a non-trivial task. The authors suggested a future investigation to analyze whether the superiority of event-models holds for polyphonic files. Results of another study with the use of folk songs indicated that event-based models are both more accurate and scalable for the task of music classification (van Kranenburg et al., 2013). Although statistical descriptors of music content have been proved useful for music genre classification de León, the capture of temporal dynamics has showed to increase the system's performance. In fact, the identification of temporal patterns is crucial for the modeling of dependencies of music events among different genres (Hillewaere et al., 2009; van Kranenburg et al., 2013).

Finally, results for larger datasets (more than 2000 songs) have shown a 75% average accuracy. Exceptions are the results reported by Juhász (2006), in which a 90% accuracy was achieved in a database of 9043 songs and six different genres. However, even those databases do not represent realistic scenarios (with million songs). Such indicative strong suggest multi-modal feature models (for instance, combining features from audio, symbolic scores, cultural data and meta-data) as promising alternatives to enhance and make the classification of music genre a more accurate and scalable task McKay and Fujinaga (2008).

4. Hybrid approaches

The studies addressed in the previous section use symbolic data for music genre classification. Some relatively recent approaches combine audio and symbolic descriptors for enhancing the classification performance by applying transcription systems to audio files (Lidy, Rauber, Pertusa, & Iñesta, 2007), transforming symbolic data into audio data (Cataltepe, Yaslan, & Sonmez, 2007), or using two matching databases, one containing symbolic scores and another containing similar songs in digital audio signals (McKay et al., 2010; McKay & Fujinaga, 2008; 2010). Concerning the later

Table 5
Features and classifiers used in the symbolic music genre classification with audio and symbolic data.

Reference	Data	# C	# S	Features	Ac (%)
Tzanetakis et al. (2002, 2003)	MIDI	5	500	Four features extracted from pitch histograms	50 ± 7
Tzanetakis et al. (2002, 2003)	Audio from MIDI	5	500	Four features extracted from pitch histograms	43 ± 7
Tzanetakis et al. (2002, 2003)	Audio	5	500	Four features extracted from pitch histograms	40 ± 6
Lidy et al. (2007)	Audio	10	1000	179 audio features from statistical spectrum descriptors and onset	74.5
Lidy et al. (2007)	MIDI-from audio	10	1000	37 high-level features	41.3
Lidy et al. (2007)	Audio + MIDI-from audio	10	1000	179 audio features derived from statistical spectrum descriptors and onset + 37 high-level features derived from transcribed notes	76.1
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio	5	250	28 features of frequency and time domains	82.8
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	MIDI	5	250	111 features related to texture, rhythm, instrumentation, dynamics, pitch statistics and melody	86.4
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio + MIDI	5	250	28 audio features + 111 symbolic features	92.4
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio + MIDI + metadata	5	250	28 audio features + 111 symbolic features + cultural features	96.8
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio	10	250	28 features of frequency and time domains	67.6
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	MIDI	10	250	111 features of texture, rhythm, instrumentation, dynamics, pitch statistics and melody	66.4
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio + MIDI	10	250	28 audio features + 111 symbolic features	75.6
McKay and Fujinaga (2008), McKay and Fujinaga (2010)	Audio + MIDI + metadata	10	250	28 audio features + 111 high-level features + cultural features	78.8
Cataltepe et al. (2007)	Audio-from MIDI	3	225	Audio features related to timbre, rhythm, and pitch	86 ± 1
Cataltepe et al. (2007)	MIDI	3	225	109 high-level features	75 ± 1
Cataltepe et al. (2007)	MIDI + audio-from MIDI	3	225	Audio features + symbolic features	93 ± 1
Cataltepe et al. (2007)	Audio-from MIDI	9	225	Audio features related to timbre, rhythm, and pitch	63 ± 1
Cataltepe et al. (2007)	MIDI	9	225	109 high-level features	42 ± 1
Cataltepe et al. (2007)	MIDI + audio-from MIDI	9	225	Audio features + symbolic features	62 ± 1
Abeßer et al. (2008)	MIDI	6	–	148 high-level features related to melody, harmony rhythm, structure and interaction	84
Abeßer et al. (2008)	Transcripted audio	6	–	148 high-level features related to melody, harmony rhythm, structure and interaction	63.4

possibility, Mckay and others provided interesting comparisons regarding the discrimination performance of different types of features (lyrical, cultural, audio and symbolic) and their combination.

Fig. 5 shows block diagrams of four studies and their respective references. Lidy et al. (2007) analyzed the contribution of audio and symbolic descriptors for music genre discrimination in three different datasets. They observed symbolic features performed better than some audio descriptors (e.g. onset features and rhythmic histograms), but worse than others (e.g. rhythmic patterns and statistical spectrum descriptors). However, the combination of rhythmic and symbolic features improved the classification accuracies. On the other hand, Cataltepe et al. (2007) used pitch intervals, duration ratios and the normalized compression distance (NCD) to compare the classification performance of songs in MIDI files, together with a number of rhythmic, pitch and timbral audio descriptors in Midi-to-audio converted files. The authors evaluated the results using the three-root and nine-leaf dataset provided by McKay (2004). With the symbolic features, the average root accuracy was 0.75, while with audio features it was 0.86. The combination of both sets of features improved the average root accuracy to 0.93.

Other studies have used symbolic and audio descriptors for comparisons of their classification performance (Abeßer, Dittmar, & Grossmann, 2008; Abeßer et al., 2009; Tzanetakis et al., 2002; 2003). Völkel, Abeßer, Dittmar, and Broßmann (2010) constructed a knowledge base of reference rhythmic patterns of percussion instruments from MIDI files to find rhythmic similarities and dissimilarities among several music genres. Pérez-Sancho, Rizo, Kersten, and Ramirez (2008b) used chord progressions from a set of symbolic data as the ground-truth for a chord transcription system to classify digital audio music.

Anglade, Benetos, Mauch, and Dixon (2010) analyzed the combination of low-level features with a harmony-based classifier (based on chord progressions) for the classification of musical genres. The performance of the harmony-based classifier was verified in a symbolic dataset and in audio data synthesized from MIDI. When pitch, rhythmic and timbral descriptors were combined with the past harmony-based classifiers, mean accuracies of 91.13% and 95.3% were achieved for GZTAN and ISMIR04 datasets, respectively. With the use of only audio descriptors (without harmony), the mean accuracies were 88.66% and 93.77%, respectively. The results were obtained by the SVM classifier.

Table 5 shows results of the aforementioned studies. As the number and type of features change for each situation, a fair comparison is not possible. However, as shown in Table 5, i) in all cases, the combination of audio and symbolic features led to a more accurate classification in comparison to their separated use, although the difference is not substantially high; ii) the use of audio features derived from transcribed MIDI files provided the same results obtained directly from audio; and iii) the use of high-level features derived from transcripted audio decreased the systems' performance. According to Tzanetakis et al. (2002), music descriptors extracted from audio or from audio derived from MIDI were less discriminative than the same descriptors obtained from the original MIDI files. The authors suggested MIDI data are accurate by definition and the disparity between MIDI and audio-from MIDI would be a consequence of errors introduced by pitch detection algorithms.

5. Concluding remarks

Music genre is a crucial aspect by which humans organize their music collections. While the classification of some music into one or more genres is quite simple for humans (Gjerdingen & Perrott, 2008), its automatic classification is still a challenging task. We have presented a review of the most important studies on musical genre classification of symbolic data and investigated the most common music features and classification techniques used, addressing their performance and feasibility. We have also collected available databases and toolboxes for future studies on the subject.

The survey has revealed pitch and rhythm as the principal musical aspects explored in symbolic data for music genre classification that lead to good classification accuracies. For audio-based systems, pitch and rhythm seem not to provided good enough results when considered separately, but their use have increased classification performance when they are adopted as to complement low-level features (Fu et al., 2011). High-level information is precise and straightforward to be obtained on symbolic data, and our analysis revealed pitch and rhythm information to be discriminative among different music genres.

Simple modeling of pitch and rhythm, e.g., pitch contours and changes in meter/time, lead to worse classification accuracies in the reported studies. This can be due to an effect of reducing important variations of pitch and duration dependencies. The representation of pitch and rhythm as music features that lead to better performances takes into account tonality transpositions (as, for example, with the use of melodic intervals) and relative duration values (as, for example, with the use IOIs). Regarding the majority of Western genres as addressed in this overview, results obtained with rhythmic features were in most cases better than the ones obtained with pitch descriptors for a same database. Similar evidences are found in previous overviews for audio-based music genre classification.

Harmonic descriptors are usually addressed through the use of chord progressions and the representation adopted may result in substantial different classification performances. Results on a same database showed the use of chord degrees led to higher classification accuracies than chord names and the use of extension can increase performance substantially. The importance of chord sequences has been also observed for audio-based classification.

Our results have also corroborated the findings of prior researchers who have investigated the performance of global features in comparison to event-based ones. When examining the results over a same dataset, we have observed that event-based properties tend to show higher classification accuracies than global features or statistical moments. The temporal dynamics of music pitch, duration and harmony seems to be discriminative and important for the characterization of music genres. These findings indicate that, as for audio-based classification, temporal features show an important class of descriptors for symbolic data.

Instrumental features are also important for genre classification on symbolic data, which is in agreement with audio-based studies, indicating that MFCC (Mel-frequency Cepstrum Coefficient) descriptors have produced quite good classification results on audio data (Fu et al., 2011). Finally, the combination of features from distinct music aspects provides better classification results, as in the case of audio-based classification.

The comparison of the performance of different machine learning algorithms is a hard task, since they have proper ways of working in the feature space, comprehending many parameters to be configured which are most of time not revealed in the papers. Results from Tables 1 to 4 on a same set of features and music collection confirmed the feature set used for classification is of crucial importance, while the learning algorithms have marginal contribution to the final accuracies. In classification tasks, the spatial organization of data is of great importance, since it provides relevant information about the convergence of classifiers, as well as their generalization performance for examples that are not present in the training stage.

Despite our efforts to present important aspects concerning the state-of-the-art approaches in symbolic music genre classification, this study has limitations and requires more research. First, most datasets addressed here are considerably small, in the sense that realistic sizes for music collections are much larger. Second, the datasets has the predominance of Western genres that may bring a biased analysis of the results. Third, contrariwise to the case of audio datasets, few classification systems are tested on previous adopted symbolic datasets, which makes the suitable comparison of results a difficult task.

The insights presented here can be used in future studies with larger and more realistic datasets that include a large variety of musical genres. It would be interesting to perform on such datasets a more systematic study on the accessibility and efficiency of most important features for symbolic music genre classification. Moreover, in view of the lack of symbolic music database for validation purposes, we expect that our list of free available symbolic databases will help future researchers to validate their classification results in common data collections, as it generally occurs for audio databases.

Analyses of the most promising features in a frame-based level are another interesting aspect to be addressed. Many audio classification systems split the audio signal into local frames and the size of such frames is usually well established depending on the audio feature under analysis. Music features are extracted frame by frame and further combined for classification (Fu et al., 2011). Although such a process offers many advantages, it has not been well explored for symbolic data. Different studies use distinct parameters for extracting features in sliding windows with or without overlapping, and a future study could verify how performance may be influenced for distinct window sizes.

The discussion on the most discriminative up-to-date symbolic features is appealing from many perspectives. It can guide audio transcription systems toward the development of methods that capture such features in audio files and study their combination with the state-of-the-art audio features. Second, we expected this survey motivates researchers in music classification of symbolic data to explore new possibilities of music descriptors, for instance, by exploring new metaphors to model dependencies and dynamics properties of melodic intervals and IOIR descriptors, as they have proved useful for characterizing music genres. We consider the development of more efficient strategies to capture their dynamism in a song as promising indication for future research, since it will certainly improve the discrimination power of such descriptors. Moreover, the combination of audio and symbolic features can be more consistently addressed. The studies reported in Section 4 show a strong evidence the combination of different sources of features can enhance the classification performance.

Finally, future studies may also include the review of other classification tasks of symbolic data, e.g., music annotation, artist classification and mood/emotion classification.

Acknowledgments

Débora C. Corrêa is indebted to FAPESP (2012/17961-0) for the financial support provided to this research. Francisco A. Rodrigues acknowledges CNPq (grant 305940/2010-4), Fapesp (grants 2011/50761-2 and 2013/26416-9) and NAP eScience-PRP-USP for their financial support.

Fig. A.6. Genre taxonomy of the Bodhidharma database (McKay, 2004).

Appendix A. Databases

This section provides the datasets and genre taxonomies used in the main references of this study. Symbolic databases for research purposes are less common than audio datasets. During the preparation of the survey, we detected some databases used are freely available for downloading. Table A.6 shows a comprehensive list of the datasets adopted, their respective taxonomies, data format and an indication if they are available to purchase (footnote indication). We expect this report increases the accessibility of symbolic datasets and support future research in this area.

To date, the most diverse and largest database is provided by McKay (2004) and used in their Bodhidharma system (McKay & Fujinaga, 2005b). The taxonomy consists of nine root labels, eight intermediate labels, and 38 leaf labels (see Fig. A.6) and encapsulates ambiguities found in realistic taxonomies.

Table A.6 shows the size of the datasets varies from 50 to 3367, with an average of 633 songs. The number of classes varies from 2 to 38, with an average of eight genres. Classical and folk are the two most used genres, probably due to the high variability of scores at the WWW. The last two entries account for datasets that contain different types of musical data and are used in the hybrid approaches (Section 4) whose main objectives are the comparison or combination the performance of musical descriptors from distinct types of musical data.

Appendix B. Toolboxes, frameworks and feature sets for symbolic music data

Below are frameworks, toolboxes and features sets available at the World Wide Web (WWW) and intended to provide music descriptors and meta-data information from MIDI files or other types of symbolic scores. The reader may find the respective web page in the corresponding references.

- Jesser (1991): 40 pitch and duration statistics;
- de León and Iñesta (2004a): 28 global features;
- MATLAB MIDI Toolbox (Eerola & Toiviainen, 2004; 2006):;
- ACE - Autonomous Classification Engine (ACE, 2005; McKay, Fiebrink, McEnnis, Li, & Fujinaga, 2005): meta-learning software package that selects, optimizes and applies machine learning techniques to music research;
- McKay's software package jSymbolic (Mackay, 2013; McKay & Fujinaga, 2006a): 111 global features developed for the classification of orchestrated MIDI files.
- meta-MIDI Toolbox (Pérez-García et al., 2009; 2012): an open source tool for the extraction of descriptive, structural and technical meta-data from standard MIDI files.
- Fantastic – Feature ANalysis Technology Accessing Statistics (In a Corpus) (Fan, 2009; Müllensiefen, 2009): developed by Müllensiefen and comprising 92 global features.
- `music21` Toolkit (Cuthbert, Ariza, & Friedland, 2011,Cuthbert, Ariza, & Friedland): A python-based and open source toolkit for analyses, searches and transformations of symbolic music data. It integrates features from other toolkits, as jSymbolic and enables the development of new features.

Appendix C. Main references on distinct types of features

Below are the main references regarding distinct aspects of features. Some of them used a combination of different types of features, including additional aspects, as dynamics, musical texture, volume, among others.

- **Pitch-repeated:** Abeßer et al. (2010); Abeßer et al. (2012); Abeßer et al. (2009); Anan et al. (2011); Basili et al. (2004); Chai and Vercoe (2001); Cilibrasi, Vitányi, and de Wolf (2004a, 2004b); Conklin (2006, 2009, 2013); Cruz-Alcázar and Vidal-Ruiz (1998, 2003); Cruz-Alcázar et al. (2003); Cuthbert et al. (2011); de León and Iñesta (2002, 2004a, 2003, 2004b, 2007); de León et al. (2006); de León et al. (2008); de León et al. (2004); Hedges et al. (2014); Hillewaere et al. (2009, 2012, 2014); Juhász (2006); Karydis (2006); Karydis et al. (2006); Khoo, Man, and Cao (2012); Kotsifakos et al. (2013); Li and Sleep (2004a, 2004b); Li, Ji, and Bilmes (2006); Lin and Chen (2012); Lin et al. (2004); Pérez-Sancho et al. (2006); Pérez-Sancho et al. (2005); Pérez-Sancho et al. (2008a); Ruppin and Yeshurun (2006); Şimşekli (2010); Valverde-Rebaza et al. (2014); van Kranenburg et al. (2013); Velarde, Weyde, and Meredith (2013a, 2013b). Total: 45.
- **Rhythm-related:** Abeßer et al. (2010); Abeßer et al. (2012); Abeßer et al. (2009); Anan et al. (2011); Basili et al. (2004); Chai and Vercoe (2001); Conklin (2006, 2009, 2013); Corrêa et al. (2010); Cruz-Alcázar and Vidal-Ruiz (1998, 2003); Cruz-Alcázar et al. (2003); Cuthbert et al. (2011); de León and Iñesta (2002, 2004a, 2003, 2004b, 2007); de León et al. (2006); de León et al. (2008); de León et al. (2004); Ellis and Arroyo (2004); Hedges et al. (2014); Hillewaere et al. (2009, 2012, 2014); Juhász (2006); Karydis (2006); Karydis et al. (2006); Khoo et al. (2012); Kotsifakos et al. (2013); Li and Sleep (2004a); Li et al. (2006); Lin and Chen (2012); Lin et al. (2004); Pérez-Sancho et al. (2006); Pérez-Sancho et al. (2005); Pérez-Sancho et al. (2008a); Ruppin and Yeshurun (2006); Valverde-Rebaza et al. (2014); van Kranenburg et al. (2013). Total: 41.
- **Timbre-related**: Basili et al. (2004); Hübler and Hoffmann (2013); Pérez-García et al. (2009); Pérez-García et al. (2010). Total: 4.
- **Harmonic-related:** Abeßer et al. (2010); Abeßer et al. (2012); Abeßer et al. (2009); Anglade, Ramirez, and Dixon (2009a, 2009b); de León and Iñesta (2002, 2004a, 2003, 2004b, 2007); de León et al. (2006); de León et al. (2008); de León et al. (2004); Hedges et al. (2014); Pérez-García et al. (2010); Pérez-

Table A.6
Databases for symbolic music genre classification ("# S" is the number of songs and "# C" is the number of classes).

Reference	# S # C and Classes	Format
Dannenberg et al. (1997)	200-8: Performance styles: "frantic", "lyrical", "pointillistic", "syncopated", "high", "low", "quote", "blues"	MIDI
Chai and Vercoe (2001)	491-3: Irish, German and Austrian folk music[7]	KERN, EsAC
Shan and Kuo (2003)	190-4: New age, Beatles, Chinese and Japanese music	MIDI
Cruz-Alcázar and Vidal-Ruiz (1998)	300-3: Gregorian Chant, J. S. Bach sacred music, Scott Joplin's Ragtime	-
Basili et al. (2004)	300-6 Blues, classical, disco, jazz, pop, rock	MIDI
Li and Sleep (2004b)	749-4: Classical, Chinese music and jazz	MIDI
Cilibrasi et al. (2004b)	36-3: Classical, jazz and rock	MIDI
Lin et al. (2004)	500-7: Blues, country, dance, jazz, Latin, pop and rock	MIDI
McKay and Fujinaga (2004)	950-9-38: Refer to Fig. A.6[8]	MIDI
Karydis (2006)	250-5: Ballads, chorales, fugues, mazurkas and sonatas[9]	KERN
Li et al. (2006)	14000-6: Music from England, Ireland, Scotland, France, Scandinavia, and S. E. Europe[10]	ABC
Pérez-Sancho et al. (2005)	300-3: Gregorian, Baroque and ragtime	MIDI
Ruppin and Yeshurun (2006)	50-3: Classical, pop, Japanese	MIDI
Juhász (2006)	9043-6: Music from Slovak, French, Sicilian, Bulgarian, and Appalachian English	-
Conklin (2006)	337-2: Nova Scotia folk songs and Bach Chorale[11]	-
de León and Iñesta (2002)	110-2: Classical and jazz	MIDI
de León et al. (2008)	3114-3: Jazz, classical and popular music	MIDI
Pérez-Sancho et al. (2008a)	856-3-9: Popular (pop, blues and celtic), jazz (pre-bop, bop and bossa nova) and academic (baroque, classicism and romanticism)[12]	MIDI and BIAB
Abeßer et al. (2009)	300-6: Pop/rock, swing, Latin, funk, blues, metal	MIDI
Hillewaere et al. (2009), Hillewaere et al. (2014)	3367-6: Folk songs divided into six geographic regions: England, France, Ireland, Scotland, S. E. Europe and Scandinavia[13]	MIDI
Abeßer et al. (2010)	320-8: Reggae, swing, salsa & mambo, funk, blues, rock, soul & motown and Africa	MIDI
Corrêa et al. (2010)	280-4: Pop/rock, Brazilian music, reggae and blues	MIDI
Şimşekli (2010)	225-3-9: Jazz (bebop, swing, bossa nova), rhythm & blues (blues rock, funk, rock'n roll), and rock (hard rock, metal, alternate rock[14])	MIDI
Anan et al. (2011)	238-2: Japanese pop songs, Enka songs	MIDI
Abeßer et al. (2009)	300-6: Pop/rock, swing, Latin, funk, blues, metal hard-rock	MIDI
Hillewaere et al. (2012), Hillewaere et al. (2014)	2198-9: Dancing styles: bourrée, hornpipe, jig, march, polska, reel, schottische, strathspey, waltz[15]	MIDI
Lin and Chen (2012)	2187-5: Jazz, lyric rock, classical and others	MIDI
Abeßer et al. (2012)	520-13: Blues, bossa nova, forró, funk, hip-hop, techno, motown, reggae, nineties rock, seventies rock, salsa & mambo, swing, zouglou	MIDI
Khoo et al. (2012)	312-5: Chinese folk songs from five regions: Jiangsu, Dongbei, Guangdong, Shanxi and Sichuan[16]	KERN
Kotsifakos et al. (2013)	100-4: Blues, rock, classical and pop	MIDI
Hübler and Hoffmann (2013)	504-4: Rumba, samba, tango and waltz[17]	MIDI
Velarde et al. (2013a), van Kranenburg et al. (2013), Hillewaere et al. (2014)	360-26: Families of Dutch folk songs[18]	MIDI
Conklin (2013)	1630-7: Geographic region: *Alava, Gipuzkoa, Navarra, Lapurdi, Bizkaia, Nafarroa Beherea, Zuberoa*[19]	MIDI
Conklin (2013)	951-3: *Danza*, *religiosa* and *amorosa*[20]	MIDI
Valverde-Rebaza et al. (2014)	919-4: Classical, Brazilian Backcountry, Pop/Rock and Jazz[21] MIDI	
Hedges et al. (2014)	434-9: jazz subgenres	MIDI
Hybrid datasets		
Tzanetakis et al. (2003)	500-5:Electronica, classical, jazz, Irish folk, and rock	MIDI, MP3

(*continued on next page*)

Table A.6 (*continued*)

Reference	# S # C and Classes	Format
McKay et al. (2010)	250-5-10: Modern blues and traditional blues; baroque and romantic; bebop and swing; hardcore rap and pop rap; and alternative rock and metal[22]	MIDI, MP3, lyrics, meta

[7] http://essen.themefinder.org/
[8] http://www.music.mcgill.ca/~cmckay/protected/Bodhidharma_MIDI.zip
[9] The Humdrum website: http://kern.humdrum.net.
[10] http://essen.themefinder.org/
[11] www.ccarh.org
[12] http://grfia.dlsi.ua.es/cm/projects/prosemus/database.php#9GDB
[13] http://essen.themefinder.org/
[14] http://www.music.mcgill.ca/~cmckay/protected/Bodhidharma_MIDI.zip
[15] http://trillian.mit.edu/~jc/cgi/abc/tunefind
[16] http://humdrum.ccarh.org/#data
[17] www.midi.de
[18] http://www.liederenbank.nl/index.php?lan=en
[19] http://www.euskomedia.org/, http://www.eresbil.com/
[20] http://www.euskomedia.org/ ,http://www.eresbil.com/
[21] http://www.icmc.usp.br/asoriano/download.html
[22] http://jmir.sourceforge.net/Codaich.html

Table C.7
Main references for symbolic music genre classification that use event or global-based features.

	References
Event-based features	Abeßer et al. (2010); Abeßer et al. (2012); Abeßer et al. (2009); Anan et al. (2011); Anglade et al. (2009a, 2009b); Chai and Vercoe (2001); Cilibrasi et al. (2004a, 2004b); Conklin (2006, 2009, 2013); Corrêa et al. (2010); Cruz-Alcázar and Vidal-Ruiz (1998, 2003); Cruz-Alcázar et al. (2003); Ellis and Arroyo (2004); Hedges et al. (2014); Hillewaere et al. (2009, 2012, 2014); Hübler and Hoffmann (2013); Juhász (2006); Karydis et al. (2006); Khoo et al. (2012); Kotsifakos et al. (2013); Li and Sleep (2004a, 2004b); Li et al. (2006); Lin and Chen (2012); Lin et al. (2004); Pérez-Sancho et al. (2006); Pérez-Sancho et al. (2005); Pérez-Sancho et al. (2008a, 2009); Ruppin and Yeshurun (2006); Shan and Kuo (2003); Shan et al. (2002); Şimşekli (2010); Valverde-Rebaza et al. (2014); van Kranenburg et al. (2013); Velarde et al. (2013a); 2013b). Total: 42.
Global features	Basili et al. (2004); Cuthbert et al. (2011); Dannenberg et al. (1997); DeCoro et al. (2007); de León and Iñesta (2002, 2003, 2004a, 2004b, 2007); de León et al. (2006); de León et al. (2008); de León et al. (2004); Hillewaere et al. (2009, 2012); Karydis (2006); Kofod and Ortiz-Arroyo (2008); McKay (2003); McKay and Fujinaga (2004, 2005a, 2005b); Pérez-García et al. (2009); Pérez-García et al. (2010); Valverde-Rebaza et al. (2014); van Kranenburg et al. (2013). Total: 24.

Sancho et al. (2006); Pérez-Sancho et al. (2008a, 2009); Shan and Kuo (2003); Shan et al. (2002); Valverde-Rebaza et al. (2014). Total: 21.

- **Combined:** Dannenberg et al. (1997); DeCoro et al. (2007); Hillewaere et al. (2009, 2012); Kofod and Ortiz-Arroyo (2008); McKay (2003); McKay and Fujinaga (2004, 2005a, 2005b). Total: 9.

Table C.7 divides the studies according to the characteristic for music descriptors: event-based and global features.

References

Abeßer, J., Bräuer, P., Lukashevich, H., & Schuller, G. (2010). Bass playing style detection based on high-level features and pattern similarity. In *Proceedings of the 11th international symposium on music information retrieval, Utrecht, Netherlands* (pp. 93–98).

Abeßer, J., Dittmar, C., & Grossmann, H. (2008). Automatic genre and artist classification by analyzing improvised solo parts from musical recordings. In *Proceedings of the audio mostly conference, Pitea, Sweden* (pp. 127–131).

Abeßer, J., Lukashevich, H., & Bräuer, P. (2012). Classification of music genres based on repetitive basslines. *Journal of New Music Research, 41*(3), 239–257.

Abeßer, J., Lukashevich, H., Dittmar, C., & Schuller, G. (2009). Genre classification using bass-related high-level features and playing styles. In *Proceedings of the 10th international symposium on music information retrieval, Kobe, Japan* (pp. 453–458).

ACE (2005). Ace 2.0. http://jmir.sourceforge.net/index_ACE.html.

Amancio, D. R., Comin, C. H., Casanova, D., Travieso, G., Bruno, O. M., Rodrigues, F. A., & Costa, L. d. F. (2014). A systematic comparison of supervised classifiers. *PLoS ONE, 9*(4), e94137.

Anan, Y., Hatano, K., Bannai, H., & Takeda, M. (2011). Music genre classification using similarity functions. In *Proceedings of the 12th international symposium on music information retrieval* (pp. 693–698). Miami, FL: University of Miami.

Anglade, A., Benetos, E., Mauch, M., & Dixon, S. (2010). Improving music genre classification using automatically induced harmony rules. *Journal of New Music Research, 39*(4), 349–361.

Anglade, A., Ramirez, R., & Dixon, S. (2009). First-order logic classification models of musical genres based on harmony. In *Proceedings of the 2009 sound and music computing, Porto, Potugal* (pp. 309–314).

Anglade, A., Ramirez, R., & Dixon, S. (2009). Genre clasification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the 10th international symposium on music information retrieval, Kobe, Japan* (pp. 669–774).

Aucouturier, J.-J., & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research, 32*(1), 83–93.

Basili, R., Serafini, A., & Stellato, A. (2004). Classification of musical genre: a machine learning approach. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain*.

Bobdanov, D., Serrá, J., Wack, N., Herrera, P., & Serra, X. (2011). Unifying low-level and high-level music similarity measures. *IEEE Transactions on Multimedia, 13*(4), 687–701.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152).

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based information retrieval: Current directions and future challenges. *Proceedings of the IEEE, 96*(4), 668–696.

Cataltepe, Z., Yaslan, Y., & Sonmez, A. (2007). Music genre classification using MIDI and audio features. *EURASIP Journal on Advances in Signal Processing, 2007*(36409), 1–8.

Chai, W., & Vercoe, B. (2001). Folk music classification using hidden markov models. In *Proceedings of the 2001 international conference on artificial intelligence* (pp. 1–6).

Cilibrasi, R., Vitányi, P., & de Wolf, R. (2004). Algorithmic clustering of music. In *Proceedings of the fourth international conference on web delivering of music (WEDELMUSIC04), Barcelona, Spain* (pp. 1–8).

Cilibrasi, R., Vitányi, P., & de Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal, 28*(4), 49–67.

Conklin, D. (2006). Melodic analysis with segment classes. *Machine Learning, 65*, 349–360.

Conklin, D. (2009). Melody classification using patterns. In *Proceedings of the international workshop on machine learning and music, Bled, Slovenia* (pp. 37–41).

Conklin, D. (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research, 42*(1), 19–26.

Corrêa, D. C., Costa, L. d. F., & Levada, A. L. M. (2011). Finding community structure in music genre networks. In *Proceedings of the 12th international symposium on music information retrieval* (pp. 447–452). Miami, FL: University of Miami.

Corrêa, D. C., Costa, L. d. F., & Saito, J. H. (2011). Using digraphs and a second-order Markovian model for rhythm classification. In L. F. Costa, A. Evsukoff, G. Mangioni, & R. Menezes (Eds.), *Complex networks*. In *Communications in computer and information science.: Vol. 116* (pp. 85–95). Berlin, Heidelberg: Springer.

Corrêa, D. C., Saito, J. H., & Costa, L. d. F. (2010). Musical genres: beating to the rhythms of different drums.. *New Journal of Physics, 12*(053030), 1–37.

Costa, L. d. F., & César, R. M., Jr. (2001). *Shape analysis and classification*. FL, USA: CRC Press.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, UK: Cambridge University Press.

Cruz-Alcázar, P. P., & Vidal-Ruiz, E. (1998). Learning regular grammars to model musical style: Comparing different coding schemes. In V. Honavar, & G. Slutzki (Eds.), *Grammatical inference*. In *Lecture notes on computer science: Vol. 1433* (pp. 211–222). Berlin, Heidelberg: Springer-Verlag.

Cruz-Alcázar, P. P., & Vidal-Ruiz, E. (2003). Modeling musical style using grammatical inference techniques: a tool for classifying and generating melodies. In *Proceedings of the third international conference WEB delivering of music, Leeds, UK* (pp. 1–8).

Cruz-Alcázar, P. P., Vidal-Ruiz, E., & Pérez-Cortés, J. C. (2003). Musical style identification using grammatical inference: the encoding problem. In A. Sanfeliu, & J. Ruiz-Schulcloper (Eds.), *13th IberoAmerican congress on pattern recognition*. In *Lecture notes on computer science.: Vol. 2905* (pp. 375–382). Berlin, Heidelberg: Springer-Verlag.

Cuthbert, M. S., Ariza, C., & Friedland, L. (a). music21: a toolkit for computer-aided musicology. web.mit.edu/music21/.

Cuthbert, M. S., Ariza, C., & Friedland, L. (2011). Feature extraction and machine learning on symbolic music using the `music21` toolkit. In *Proceedings of the 12th international symposium on music information retrieval* (pp. 387–392). Miami, FL: University of Miami.

Dannenberg, R. B., Thom, B., & Watson, D. (1997). A machine learning approach to musical style recognition. In *Proceedings of the international computer music conference, Thessaloniki, Greece* (pp. 344–347).

DeCoro, C., Barutcuoglu, Z., & Fiebrink, R. (2007). Bayesian agrregation for hierarchical genre classification. In *Proceedings of the 8th international symposium on music information retrieval, Vienna, Austria* (pp. 77–80).

de León, P. J., & Iñesta, J. M. (2002). Musical style identification using self-organizing maps. In *Proceedings of the second international conference on web delivering of music, DC, USA* (pp. 82–89).

de León, P. J., & Iñesta, J. M. (2004). Statistical description models for melody analysis and characterization. In *Proceedings of the international computer music conference, Miami, USA* (pp. 1–8).

de León, P. J. P. (c). A statistical pattern recognition approach to symbolic music classification (Ph.D. thesis). Universidad de Alicante.

de León, P. J. P., & Iñesta, J. M. (2003). Feature-driven recognition of music sytles. In F. J. Perales, A. J. C. Campilho, N. P. de la Blanca, & A. Sanfeliu (Eds.), *Pattern recognition and image analysis (ibPRIA 2003)*. In *Lecture notes on computer science: Vol. 2652* (pp. 773–781). Berlin, Heidelberg: Springer.

de León, P. J. P., & Iñesta, J. M. (2004). Musical style classification from symbolic data: a two-styles case study. In U. K. Wiil (Ed.), *Computer music modeling and retrieval*. In *Lecture notes on computer science: Vol. 2771* (pp. 167–178). Berlin, Heidelberg: Springer.

de León, P. J. P., & Iñesta, J. M. (2007). Pattern recognition approach for music style identification using shallow statistical descriptors. *IEEE Transactions on Systems, Man and Cybernetics, 37*(2), 248–257.

de León, P. J. P., Iñesta, J. M., & Pérez-Sancho, C. (2006). Classifier ensembles for genre recognition. In F. Pla, P. Radeva, & J. Vitriá (Eds.), *Pattern recognition: Progress, directions and applications* (pp. 41–53). Centre de Visió per Computador - Universitat Autónoma de Barcelona.

de León, P. J. P., Iñesta, J. M., & Rizo, D. (2008). Mining digital music score collections: melody extraction and genre recognition. In P.-Y. Yin (Ed.), *Pattern recognition techniques, technology and applications* (pp. 559–590). Vienna, Austria: I-Tech.

de León, P. J. P., Pérez-Sancho, C., & Iñesta, J. M. (2004). A shallow description framework for musical style recognition. In A. Fred, T. M. Caelli, R. P. W. Duin, A. C. Campilho, & D. de Ridder (Eds.), *Structural, syntactic and statistical pattern recognition*. In *Lecture notes on computer science: Vol. 3138* (pp. 876–884). Berlin, Heidelberg: Springer.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: John Wiley & Sons Inc.

Eerola, T., & Toiviainen, P. (2004). Mir in matlab: The midi toolbox. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 22–27).

Eerola, T., & Toiviainen, P. (2006). Midi toolbox. https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/.

Ellis, D. P. W., & Arroyo, J. (2004). Eigenrhythms:drum pattern basis sets for classification and generation. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 1–4).

Fan (2009). Modelling melodic memory and the perception of melodic similarity. http://doc.gold.ac.uk/isms/mmm.

Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia, 13*(2), 303–319.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd). Chapman and Hall/CRC.

Gan, G., Ma, C., & Wu, J. (2007). *Data clustering: Theory, algorithms, and applications*. Alexandria, Virginia: SIAM: Society for Industrial and Applied Mathematics.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research, 37*(2), 93–100.

Goto, M., Nishimura, T., Hashiguchi, H., & Oka, R. (2002). Rwc music database: Popular, classical, and jazz music databases. In *Proceedings of the 3rd international symposium on music information retrieval, Paris, France* (pp. 1–2).

Gouyon, F., & Dixon, S. (2005). A review of automatic rhythm description systems. *Computer Music Journal, 29*(1), 34–54.

Haynes, B., & Cooke, P. (b). Grove music online. http://www.oxfordmusiconline.com/subscriber/book/omo_gmo.

Hedges, T., Roy, P., & Pachet, F. (2014). Predicting the composer and style of jazz chord progressions. *Journal of New Music Research, 43*(3), 276–290.

Hillewaere, R., Manderick, B., & Conklin, D. (2009). Global feature versus event models for folk song classification. In *Proceedings of the 10th international symposium on music information retrieval, Kobe, Japan* (pp. 729–733).

Hillewaere, R., Manderick, B., & Conklin, D. (2012). String methods for folk tune genre classification. In *Proceedings of the 13th international symposium on music information retrieval, Porto, Portugal* (pp. 217–222).

Hillewaere, R., Manderick, B., & Conklin, D. (2014). Alignment methods for folk tune classification. In *Data analysis, machine learning and knowledge discovery* (pp. 369–377). Springer.

Homburg, H., Mierswa, I., Möller, B., Morik, K., & Wurst, M. (2005). A benchmark dataset for audio classification and clustering. In *Proceedings of the 6th international symposium on music information retrieval* (pp. 528–531). London, UK: University of London.

Hsu, J.-L., Liu, C.-C., & Chen, C.-y. (2001). Discovering non-trivial repeating patterns in music data. *IEEE Transactions on Multimedia, 3*(3), 311–325.

Hübler, S., & Hoffmann, R. (2013). Modelling drum patterns with weighted finite-state transducers. In *Proceedings of the 38th IEEE international conference on acoustics, speech, and signal processing* (pp. 719–723). Vancouver, Canada: Institute of Electrical and Electronics Engineers, Inc.

Jesser, B. (1991). *Interaktive Melodieanalyse*. Bern, Switzerland: Peter Lang.

Juhász, Z. (2006). A systematic comparison of different european folk music traditions using self-organizing maps. *Journal of New Music Research, 35*(2), 95–112.

Karydis, I. (2006). Symbolic music genre classification based on note pitch and duration. In Y. Manolopoulos, J. Pokorný, & T. K. Sellis (Eds.), *Advances in databases and information systems (ADBIS 2006)*. In *Lecture notes on computer science: Vol. 4152* (pp. 329–338). Berlin, Heidelberg: Springer-Verlag.

Karydis, I., Nanopoulos, A., & Manolopoulos, Y. (2006). Symbolic musical genre classification based on repeating patterns. In *Proceedings of the 1st ACM workshop on audio and music computing multimedia (AMCMM'06), California, USA* (pp. 53–57).

Khoo, S., Man, Z., & Cao, Z. (2012). Automatic han chinese folk song classification using the musical feature density map. In *Proceedings of the 6th IEEE international conference on signal processing and communication systems* (pp. 1–9). Gold Coast, QLD: Institute of Electrical and Electronics Engineers (IEEE).

Kofod, C., & Ortiz-Arroyo, D. (2008). Exploring the design space of symbolic music genre classification using data mining techniques. In *Proceedings of the IEEE international conference on computational intelligence for modelling control and automation* (pp. 43–48). Viena, Austria: IEEE Computer Society.

Kotsifakos, A., Kotsifakos, E. E., Papapetrou, P., & Athitsos, V. (2013). Genre classification of symbolic music with SMBGT. In *Proceedings of the 6th international conference on PErvasive technologies related to assistive environments, Island of Rhodes, Greece* (pp. 1–7).

Lee, J. H., & Downie, J. S. (2004). Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 1–4).

Li, M., Chen, X., Li, X., Ma, B., & Vitányi, P. M. B. (2004). The similarity metric. *IEEE Transactions on Information Theory, 50*(12), 3250–3264.

Li, M., & Sleep, R. (2004). Improving melody classification by discriminant feature extraction and fusion. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 238–242).

Li, M., & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In *Proceedings of the 2004 sound and music computing, Paris, France* (pp. 1–4).

Li, X., Ji, G., & Bilmes, J. (2006). A factored language model of quantized pitch and duration. In *Proceedings of the international computer music conference, San Francisco, CA* (pp. 556–563).

Lidy, T., Rauber, A., Pertusa, A., & Iñesta, J. M. (2007). Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In *Proceedings of the 8th international symposium on music information retrieval, Vienna, Austria* (pp. 1–6).

Lin, B.-S., & Chen, T.-C. (2012). Genre classification for musical documents based on extracted melodic patterns and clustering. In *Proceedings of the 17th conference on technologies and applications of artificial intelligence, Tainan, Taiwan* (pp. 39–43).

Lin, C.-R., Ning-Han-Liu, Wu, Y.-H., & Chen, A. L. P. (2004). Music classification using significant repeating patterns. In Y. L. a. J. Li, K.-Y. Whang, & D. Lee (Eds.), *Database systems for advanced applications(DASFAA 2004)*. In *Lecture notes on computer science: Vol. 2973* (pp. 506–518). Berlin, Heidelberg: Springer.

Mackay, C. (2013). jMIR 2.0. http://jmir.sourceforge.net/.
McKay, C. (2003). Using neural networks for musical genre classification. http://music.mcgill.caa/cmckay/projects.html.
McKay, C. (2004). Automatic classiffication of MIDI recordings. Master's thesis.
McKay, C., Burgoyne, J. A., Hockman, J., Smith, J. B. L., Vigliensoni, G., & Fujinaga, I. (2010). Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features. In *Proceedings of the 11th international symposium on music information retrieval, Utrecht, Netherlands* (pp. 213–214).
McKay, C., Fiebrink, R., McEnnis, D., Li, B., & Fujinaga, I. (2005). Ace: A framework for optimizing music classification. In *Proceedings of the 6th international symposium on music information retrieval, Barcelona, Spain* (pp. 1–6). London, UK: Queen Mary University of London.
McKay, C., & Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 1–6).
McKay, C., & Fujinaga, I. (2005). Automatic music classification and the importance of instrument identification. In *Proceedings of the conference on interdisciplinary musicology, Montreal,Canada* (pp. 1–4).
McKay, C., & Fujinaga, I. (2005). The Bodhidharma system and the results of the MIREX 2005 symbolic genre classification contest. In *Proceedings of the 6th international symposium on music information retrieval, Victoria, Canada* (pp. 1–4).
McKay, C., & Fujinaga, I. (2006). jsymbolic: A feature extrator for MIDI files. In *Proceedings of the international computer music conference, New Orleans, USA* (pp. 1–4).
McKay, C., & Fujinaga, I. (2006). Musical genre classification: Is it worth pursuing and how can it be improved? In *Proceedings of the 7th international symposium on music information retrieval, Victoria, Canada* (pp. 101–107).
McKay, C., & Fujinaga, I. (2008). Combining features extracted from audio, symbolic and cultural sources. In *Proceedings of the 8th international symposium on music information retrieval, Philadelphia, USA* (pp. 597–602).
McKay, C., & Fujinaga, I. (2010). Improving automatic music classification performance by extracting features from different types of data. In *Proceedings of the 11th ACM international conference on multimedia information retrieval, Philadelphia, USA* (pp. 257–266).
Mostafa, M. M., & Billor, N. (2009). Recognition of western style musical genres using machine learning techniques. *Expert Systems with Applications, 36*(8), 11378–11389.
Müllensiefen, D. (2009). Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus): Technical report v1.5. *Technical report*. Goldsmiths University of London.
Müllensiefen, D., & Frieler, K. (2004). Optimizing measures of melodic similarity for the exploration of a large folk song database. In *Proceedings of the 5th international symposium on music information retrieval, Barcelona, Spain* (pp. 274–280).
North, A. C., & Hargreaves, D. J. (1997). Liking for musical styles. *Musicae Scientiae, 1*(1), 109–128.
Pérez-García, T., Iñesta, J. M., & Rizo, D. (2009). metamidi: A tool for automatic metadata extraction from MIDI files. In *Proceedings of the 2009 workshop on exploring musical information spaces (WEMIS), Corfu, Greece* (pp. 36–40).
Pérez-García, T., Iñesta, J. M., & Rizo, D. (2012). Resources. http://grfia.dlsi.ua.es/gen.php?id=resources.
Pérez-García, T., Pérez-Sancho, C., & Iñesta, J. M. (2010). Harmonic and instrumental information fusion for musical genre classification. In *Proceedings of the 3rd international workshop on machine learning and music (MML 2010) ACM, Firenze, Italy* (pp. 49–52).
Pérez-Sancho, C., de León, P. J. P., & Iñesta, J. M. (2006). A comparison of statistical approaches to symbolic genre recognition. In *Proceedings of the international computer music conference, New Orleans, USA* (pp. 1–4).
Pérez-Sancho, C., Iñesta, J. M., & Calera-Rubio, J. (2005). Style recognition through statistical event models. *Journal of New Music Research, 34*(4), 331–339.
Pérez-Sancho, C., Rizo, D., & Iñesta, J. M. (2008a). Stochastic text models for music categorization. In N. d. V. Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos, & M. Loog (Eds.), *Structural, syntactic, and statistical pattern recognition*. In *Lecture notes on computer science: Vol. 5342* (pp. 55–64). Berlin, Heidelberg: Springer.
Pérez-Sancho, C., Rizo, D., & Iñesta, J. M. (2009). Genre classification using chords and stochastic language models. *Connection Science, 21*(2 & 3), 145–159.
Pérez-Sancho, C., Rizo, D., Kersten, S., & Ramirez, R. (2008). Genre classification of music by tonal harmony. In *Proceedings of the international workshop on machine learning and music, Helsinki, Finland* (pp. 1–2).
Ruppin, A., & Yeshurun, H. (2006). Midi music genre classification by invariant features. In *Proceedings of the 7th international symposium on music information retrieval, Victoria, Canada* (pp. 397–399).
Ryynänen, M., & Klapuri, A. (2007). Automatic bass line transcription from streaming polyphonic audio. In *Proceedings of the international conference on acoustics, speech, and signal processing, Hawaii, USA* (pp. 1437–1440).
Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content - a survey. *IEEE Signal Processing Magazine, 23*(2), 133–141.
Selfridge-Field, E. (1998). Conceptual and representational issues in melodic comparison. In W. B. Hewlett, & E. Selfridge-Field (Eds.), *Melodic similarity, concepts, procedures, and applications: Vol. 11* (pp. 3–64). MIT Press.
Shan, M.-K., & Kuo, F.-F. (2003). Music style mining and classification by melody. *IEICE Transactions on Information and Systems, 3*, 1–6.
Shan, M.-K., Kuo, F.-F., & Chen, M.-F. (2002). Music style mining and classification by melody. In *Proceedings of the 2002 IEEE international conference on multimedia & expo, Lausanne, Switzerland* (pp. 97–100).
Shao, X., Xu, C., & Kankanhalli, M. S. (2004). Unsupervised classification of music genre using hidden markov model. In *IEEE international conference on multimedia and expo (ICME), Taipei, Taiwan* (pp. 2023–2026).
Silla, C. N., Jr., & Freitas, A. A. (2009). Novel top-down approaches for hierarchical classification and their application to automatic music genre classification. In *Proceedings of the IEEE international conference on systems, man and cybernetics, San Antonio, Texas* (pp. 3499–3504).
Şimşekli, U. (2010). Automatic music genre classification using bass lines. In *Proceedings of the 11th international symposium on music information retrieval, Utrecht, Netherlands* (pp. 4137–4140).
Snyder, B. (2009). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology: Vol. 1* (pp. 107–117). Oxford University Press.
Strum, B. L. (2013). Classification accuracy is not enough: On the evaluation of music genre recognition systems. *Journal of Intelligent Information Systems, 41*(3), 371–406.
Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302.
Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2002). Pitch histograms in audio and symbolic music information retrieval. In *Proceedings of the 3rd international symposium on music information retrieval, Paris, France* (pp. 1–8).
Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2003). Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research, 32*(2), 143–152.
Uitdenbogerd, A. L., & Zobel, J. (1998). Manipulation of music for melody matching. In *Proceedings of the sixth ACM international conference on multimedia*. In *MULTIMEDIA '98* (pp. 235–240). New York, NY, USA: ACM.
Valverde-Rebaza, J., Soriano, A., Berton, L., de Oliveira, F., Cristina, M., & De Andrade Lopes, A. (2014). Music genre classification using traditional and relational approaches. In *2014 Brazilian conference on intelligent systems (BRACIS)* (pp. 259–264). IEEE.
van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research, 42*(1), 1–18.
Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, USA: Springer-Verlag.
Vapnik, V. N., & Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications, 16*(2), 283–305.
Velarde, G., Weyde, T., & Meredith, D. (2013). An approach to melodic segmentation and classification based on filtering with the haar wavelet. *Journal of New Music Research, 42*(4), 325–345.
Velarde, G., Weyde, T., & Meredith, D. (2013). Wavelet-filtering of symbolic music representations for folk tune segmentation and classification. In *Proceedings of the international workshop for folk music analysis* (pp. 1–7). Amsterdam, Netherlands: Meertens Institute, Department of Information and Computing Sciences, Utrecht University.
Völkel, T., Abeßer, J., Dittmar, C., & Broßmann, H. (2010). Automatic genre classification of Latin American music using characteristic rhythmic patterns. In *Proceedings of the 5th audio mostly conference, Pitea, Sweden* (pp. 16–22).
Wang, L., Sugiyama, M., Yang, C., Hatano, K., & Feng, J. (2009). Theory and algorithm for learning with dissimilarity functions. *Neural Computation, 21*(5), 1459–1484.