

Content-Based

39. Content-Based Methods for Knowledge Discovery in Music

Juan Pablo Bello, Peter Grosche, Meinard Müller, Ron Weiss

This chapter presents several computational approaches aimed at supporting knowledge discovery in music. Our work combines data mining, signal processing and data visualization techniques for the automatic analysis of digital music collections, with a focus on retrieving and understanding musical structure.

We discuss the extraction of midlevel feature representations that convey musically meaningful information from audio signals, and show how such representations can be used to synchronize different instances of a musical work and enable new modes of music content browsing and navigation. Moreover, we utilize these representations to identify repetitive structures and representative patterns in the signal, via self-similarity analysis and matrix decomposition techniques that can be made invariant to changes of local tempo and key. We discuss how structural information can serve to highlight relationships within music collections, and explore the use of information visualization tools to characterize the patterns of similarity and dissimilarity that underpin such relationships.

With the help of illustrative examples computed on a collection of recordings of Frédéric Chopin's Mazurkas, we aim to show how these

39.1	Music Structure Analysis	824
39.2	Feature Representation	826
39.2.1	Chroma Features	826
39.2.2	Feature Trajectories	826
39.3	Music Synchronization and Navigation	827
39.4	Self-Similarity in Music Recordings	829
39.4.1	Self-Similarity Revisited	829
39.4.2	Enhancing Self-Similarity Matrices	830
39.4.3	Structure-Based Similarity	832
39.4.4	Visualizing Structure	833
39.5	Automated Extraction of Repetitive Structures	835
39.5.1	Structure Analysis	835
	Using Matrix Factorization	835
39.5.2	Representative Patterns	835
39.5.3	Segmentation Analysis	837
39.6	Conclusions	838
	References	838

content-based methods can facilitate the development of novel modes of access, analysis and interaction with digital content that can empower the study and appreciation of music.

The rapid and sustained growth of digital music sharing and distribution is nothing less than astounding. A multitude of digital music services provide access to tens of millions of tracks, both legally and illegally. Such abundance of content, coupled with the relative ease of access and storage afforded by recent technologies, means that music is shared and listened to more than ever before in history.

Using computational methods to help users find and organize music information is a widely researched topic in industry and academia. Existing approaches can be coarsely divided into two types: in *content-based* methods the information is obtained directly

from the analysis of audio signals, scores and other representations of the music, whereas *context-based* methods are based on information surrounding the music content, such as usage patterns, tags and structured metadata. While a significant amount of research has been devoted to the former strategy – see [39.1] for an early review – the latter has been historically favored in industrial applications such as music recommendation and playlist generation. Content-based analysis is sometimes seen as providing too little *bang for the buck*, with some observers going as far as wondering whether it is at all necessary for the retrieval of music information [39.2].

Yet, we argue that there are numerous data-mining problems for which context-based analysis is insufficient, as it tends to be low in specifics and unevenly distributed across artists and styles. Consider for example the issue of tracing back the original sources of samples used in electronic or hip-hop recordings; or of identifying the many derivations of George Gershwin's *I Got Rhythm* in the jazz catalog; or of finding quotations of a given Wagner motif in 20th century modernist music; or of quantifying which movements, artists and compositions are cited most often and are therefore the most influential. These problems are motivated by the needs of sophisticated users such as media producers, Foley artists, sound designers, film and game composers, copyright lawyers, musicologists, and professional and amateur musicians, for whom music search necessarily goes beyond the passive act of music recommendation. We believe that the development of robust and scalable solutions to these problems, and many others that could be listed instead, passes through the automated analysis of the musical content.

This chapter aims to introduce the reader to computational approaches to content-based analysis of digital music recordings. More specifically, we give an overview of a number of techniques for music structure analysis, i. e., the identification of the patterns and relationships that govern the organization of sounds in

music, and discuss how the outcomes of this analysis can facilitate data mining in music. This review is intended for an audience interested in music search, organization and discovery, that is not steeped in the field of music information retrieval (MIR). Therefore the emphasis is not on technical details, which are published elsewhere in the literature, but on the presentation of examples and qualitative results that illustrate the operation and potential of the presented approaches.

The chapter is organized as follows: Section 39.1 familiarizes the reader with the basics of music structure analysis, discusses the fundamental role that repetition plays in it, and introduces the corpus of music that will be used throughout this chapter. Section 39.2 presents standard methods for music signals analysis. Section 39.3 introduces methods for temporal alignment of different representations of a given musical piece and shows how such alignments can be used to create novel user interfaces. Section 39.4 demonstrates how to characterize the patterns of repetition in music via self-similarity analysis, which has numerous applications in segmenting, organizing and visualizing music recordings. Section 39.5 introduces a powerful technique for structure analysis using matrix factorization with applications in identifying representative patterns and segmenting music signals. Finally, Sect. 39.6 presents our conclusions and outlook on the field.

39.1 Music Structure Analysis

The architectural structure of a musical piece, its *form*, can be described in terms of a concatenation of sectional units. The amount of repetition amongst these units defines the spectrum of possible forms, ranging from strophic pieces, where a single section is continuously repeated (as is the case for most lullabies), to through-composed pieces, where no section ever recurs. While repetition is not a precondition to music, it undeniably plays a central role (*the basis of music as an art form* according to [39.3]), and is closely related to notions of coherence, intelligibility and enjoyment in its perception [39.4]. Indeed, some observers estimate that more than 99% of music listening involves repetition, both internal to the work and of familiar passages [39.5].

However, the notion of repetition in structural analysis is not rigid and, depending on the composition, might include significant variations in the musical content of the *repeated* parts. This is even more true for recordings featuring changes in instrumentation, ornamentation and expressive variations of tempo and

dynamics. In other words, an exact recapitulation of the content is not required in order for a part to be considered to be repeated. As a consequence, the analysis and annotation of musical structure is, to a certain degree, ambiguous [39.6, 7].

Take for example Frédéric Chopin's Mazurka in F major, Opus 68, No. 3, a piano work to which we will refer throughout this chapter as M68-3. One way to describe the structure of this piece is as: $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{A}_3\mathcal{T}C_1C_2\mathcal{A}_4\mathcal{A}_5$. In this description, each letter denotes a pattern and each subscript the instance number of a pattern's repetition. \mathcal{T} is a special symbol denoting a transitional section. The four patterns are depicted, in score format, in Fig. 39.1.

Note that this annotation is by no means unique and implies a number of choices. First, grouping the music into a relatively small number of parts requires tolerating slight variations across repetitions. For example, pattern \mathcal{A} presents with two alternative endings, with the last bar of segments \mathcal{A}_1 and \mathcal{A}_4 differing harmonically from the last bar of segments \mathcal{A}_2 , \mathcal{A}_3

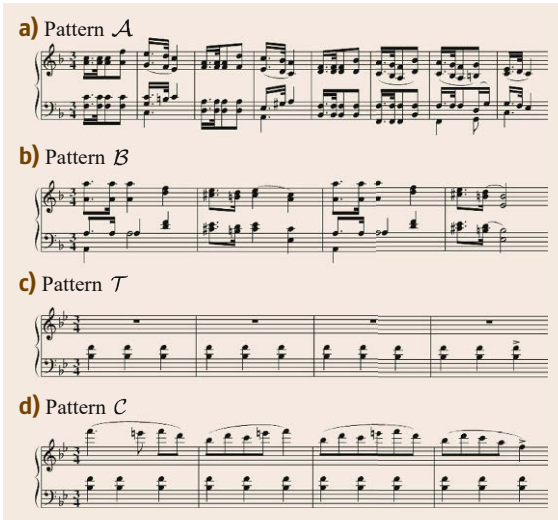


Fig. 39.1a-d Mazurka Op. 68, No. 3 (M68-3) in F major

and \mathcal{A}_5 . We could have chosen to break these occurrences into two groups, but that would ignore the high degree of overlap that otherwise exists between them. Second, the description avoids patterns consisting of a small number of repeating subpatterns. For example, \mathcal{T} and \mathcal{C} are considered to be two different parts despite their strong harmonic similarities. The alternative would be to merge $\mathcal{T}C_1C_2$ into a single pattern of highly repetitive subpatterns. In the following sections we will illustrate how these decisions relate to the infor-

mation within the music signals, and their implications for the proposed analyses.

Throughout this chapter we draw examples from the Mazurka dataset, compiled by the Center for the History and Analysis of Recorded Music in London [39.8]. The set includes 2919 recorded performances of the 49 Frédéric Chopin’s Mazurkas, resulting in an average of 58 renditions per Mazurka. These recordings, featuring 135 different pianists, cover a range of more than 100 years beginning in 1902 and ending in 2008. This makes the dataset a rich and unique resource for the analysis of style changes and expressivity in piano performance, and of the evolution of recording techniques and practices [39.9].

Our analysis is mainly focused on a subset of 298 recordings that correspond to five Mazurkas. In addition to M68-3, mentioned above, these include: Opus 17, No. 4 in A minor (M17-4); Op. 24, No. 2 in C major (M24-2); Op. 30, No. 2 in B minor (M30-2); and Op. 63, No. 3 in C^\sharp minor (M63-3). These recordings are chosen because they have been the subject of extensive musicological studies that resulted, amongst other things, in manually annotated beat positions [39.10]. Additionally, the musical structures of these Mazurkas are comparatively well defined, a fact that we have exploited to manually annotate their forms, as depicted in Fig. 39.2. Please note that the annotations were performed only once on the score representation, and then propagated to all recordings using the beat annotations mentioned above.

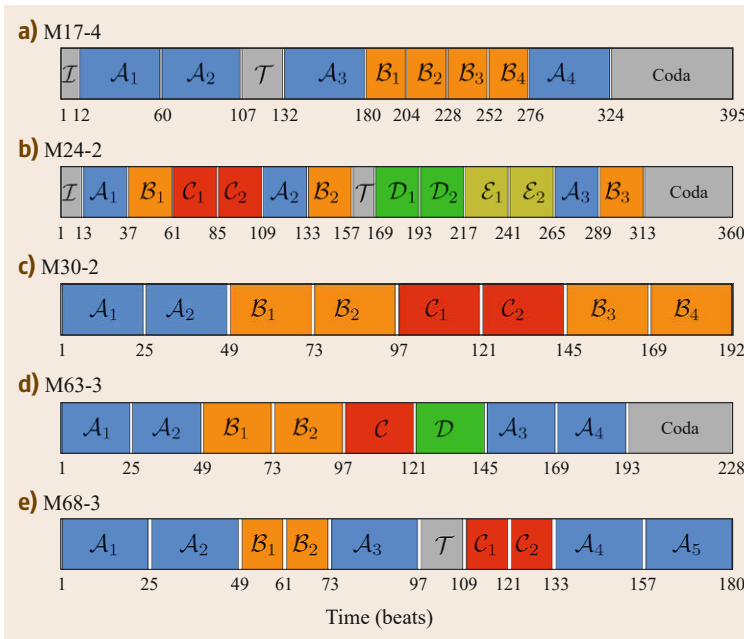


Fig. 39.2a-e Illustration of manually generated structural descriptions for five different Mazurkas

39.2 Feature Representation

All content-based music analysis begins with the extraction of a meaningful feature representation from the audio signal. This representation typically encodes information about one or more musical characteristics, e.g., harmony, melody, rhythm, or timbre as required by the specific task. There are numerous signal processing techniques that can be used to this end, such as the ubiquitous Mel-frequency Cepstral coefficients (MFCC), which are used for the analysis of timbre and texture (e.g., by [39.11, 12]). For a comprehensive list of audio features and their implementation, the reader is referred to [39.13].

In our work we use chroma features, introduced in Sect. 39.2.1, which can be used to derive information about chords. As shown in Sect. 39.2.2, these features are a powerful tool for music structure analysis, because repeating segments are often characterized by common chord progressions.

39.2.1 Chroma Features

Most musical parts are characterized by a particular melody and harmony. In order to identify repeating patterns in music recordings it is therefore useful to convert the audio signal into a feature representation that reliably captures these elements of the signal, but is not sensitive to other components of the signal such as instrumentation or timbre. In this context, *chroma features*, also referred to as *pitch class profiles* (PCPs), have turned out to be a powerful midlevel representation for describing harmonic content. They are widely used for various music signal analysis tasks, such as chord recognition [39.14], cover song identification [39.15, 16], and many others [39.17–19].

It is well known that human perception of pitch is cyclical in the sense that two pitches an integer number of octaves apart are perceived to be of the same type or class. This is the basis for the helical model of pitch perception, where pitch is separated into two dimensions: *tone height* and *chroma* [39.20]. Assuming the equal-tempered scale, and enharmonic equivalence, the chroma dimension corresponds to the twelve pitch classes used in Western music notation, denoted by $\{C, C^\sharp, D, \dots, B\}$, where different pitch spellings such as C^\sharp and D^\flat refer to the same chroma. A *pitch class* is defined to be the set of all pitches that share the same chroma. For example, the pitch class corresponding to the chroma C is the set $\{\dots, C0, C1, C2, C3, \dots\}$.

There are several methods available for the computation of chroma features from audio, usually involving the warping of the signal's short-time spectrum or its

decomposition into log-spaced subbands. This is followed by a weighted summation of energy across spectral bins corresponding to the same pitch class [39.15, 17, 19, 21]. Each chroma vector characterizes the distribution of the signal's local energy across the twelve pitch classes. Just as with the short-time Fourier transform (STFT), chroma vectors can be calculated sequentially on partially overlapped blocks of signal data, resulting in a so-called *chromagram*. The literature also proposes a number of ways in which the standard chroma feature representation can be improved by minimizing the effects of harmonic noise [39.22], by timbre and dynamic changes [39.19, 23], or by tempo variations via beat synchronization [39.15].

Figure 39.3a depicts the waveform of a 1976 recording of M68-3 performed by Sviatoslav Richter. Pattern labels are shown on the horizontal axis with boundaries marked as vertical red lines. Figure 39.3b shows the corresponding sequence of normalized chroma feature vectors, at a resolution of ten vectors per second, with the pitch classes of the chromatic scale starting in A, and ordered from bottom to top. As expected, most of the signal energy is concentrated on the classes corresponding to the F major key: F, G, A, B^\flat/A^\sharp , C, D and E. Additionally, careful inspection of the chromagram shows that the different patterns of this piece can be associated with distinct subsequences of chroma vectors.

39.2.2 Feature Trajectories

To make the latter point more evident we can use an alternative visualization of the chromagram in Fig. 39.3b. The values in a single chroma feature vector can be interpreted as a set of coordinates describing a point in the 12-dimensional space of pitch classes. Connecting those points over time results in a trajectory in feature space. Since this trajectory cannot be directly visualized in 12 dimensions, we use principal component analysis [39.24] to project this information onto its two principal components, resulting in the black line shown in Fig. 39.4. Since time is not explicitly encoded in this visualization, Fig. 39.4 also shows the two-dimensional (2-D) histograms of trajectory values, color-coded such that points contained within instances of patterns \mathcal{A} , \mathcal{B} , \mathcal{C} and \mathcal{T} are in blue, red, orange and gray respectively.

The projection clearly characterizes the main repetitions in the feature sequence as similar subtrajectories, i.e., segments of the main trajectory that are in close proximity to each other. Note that repetitions do not

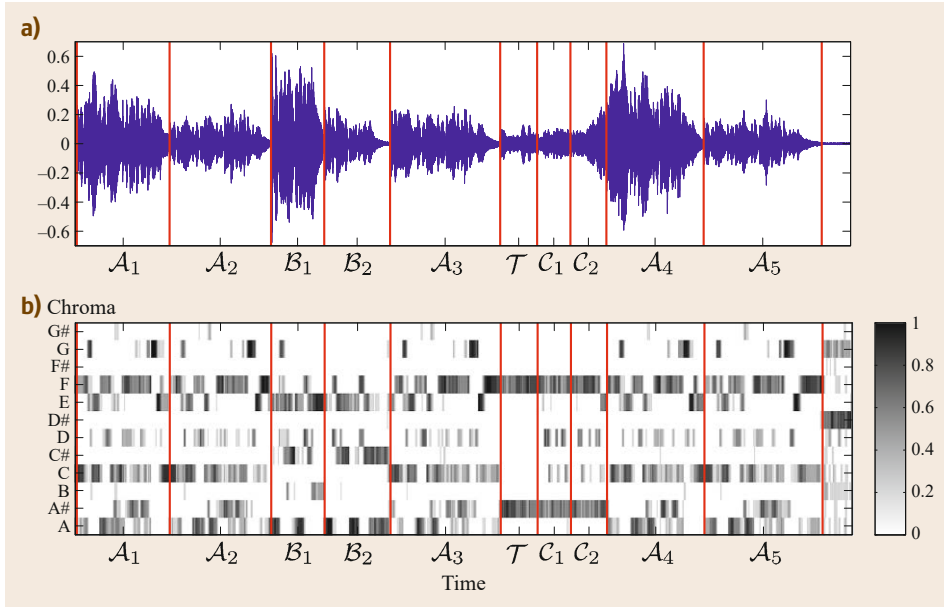


Fig. 39.3a,b Feature representations for a performance by Richter (1976) of M68-3. (a) Waveform. (b) Corresponding chromagram. Pattern boundaries are marked with vertical red lines and pattern labels are shown on the time axis

form fully overlapping subtrajectories due to expressive changes in tempo, timbre and dynamics. Despite this variability the histograms clearly show how each pattern results in distinct trajectory shapes. For example, pattern \mathcal{A} has a multimodal distribution resulting from the six different chord types that are part of its progression: F major, C major, B^b major and G major as well as D minor and A minor. In contrast, pattern \mathcal{B} , which is strongly dominated by A major chords (while also featuring some D minor and E major chords) has a much more narrow distribution. This is even more extreme for patterns \mathcal{C} and \mathcal{T} , which are highly overlapping and almost exclusively dominated by a B^b-F dyad played ostinato, as can be seen in the score representation of Fig. 39.1.

This example illustrates how chromagrams successfully capture harmonic information in the signal. In the following sections we show how these features can be used to synchronize different instances of a musical work and enable new modes of music content browsing and navigation.

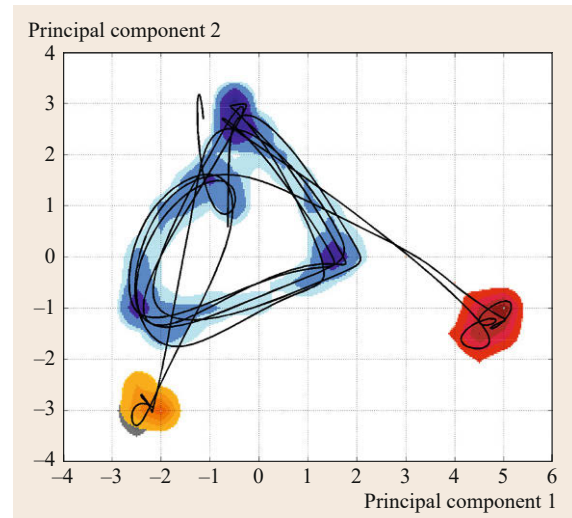


Fig. 39.4 Trajectory of 12-dimensional chroma features projected onto two dimensions using the chromagram shown in Fig. 39.3b

39.3 Music Synchronization and Navigation

Musical works can be represented in many different domains, e.g., in different audio recordings, MIDI (Musical Instrument Digital Interface) files, or as digitized sheet music. The general goal of music synchronization is to automatically align multiple information sources related to the same musical work. Here, *music synchro-*

nization denotes a procedure which, for a given position in one representation of a piece of music, determines the corresponding position within another representation [39.19, 25]. The linking information produced by the synchronization process can be used to propagate information across these representations.

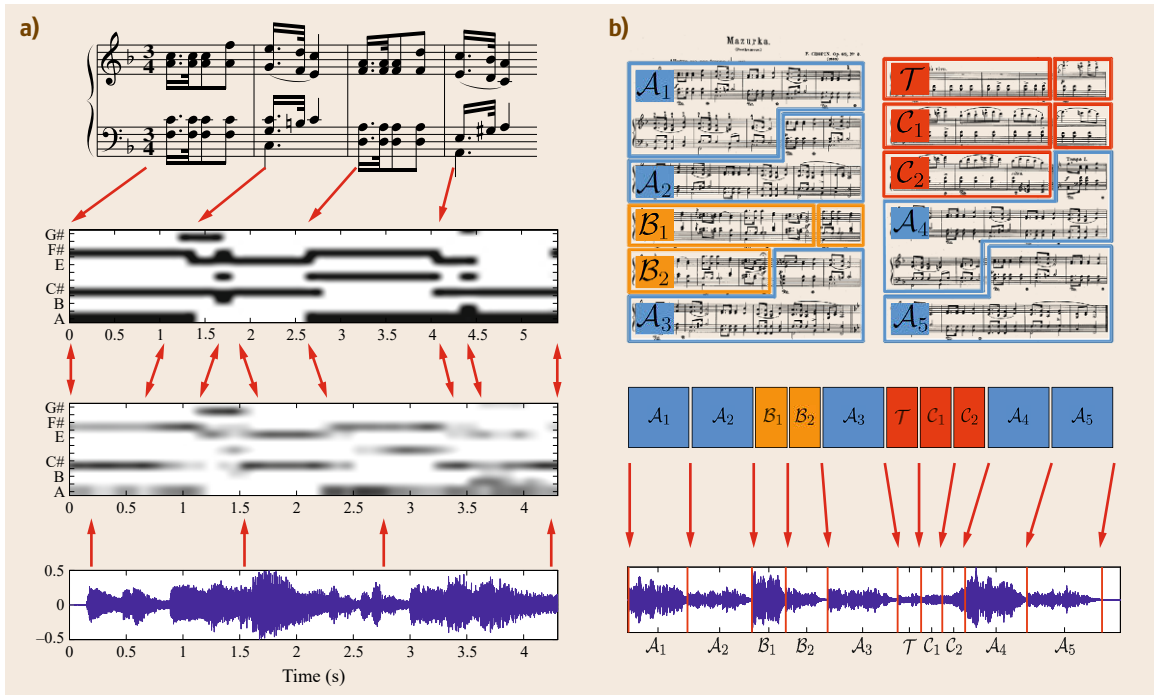


Fig. 39.5a,b Illustration of the score-audio synchronization pipeline used to transfer structure annotations from the score to the audio domain. (a) Score and audio representation with the corresponding chromagrams. The *red arrows* indicate the score-audio synchronization result. (b) The score-based annotations are transferred to the audio domain using the linking information supplied by the synchronization procedure

Consider the example in Fig. 39.5. The left column demonstrates the synchronization process between a musical score (top) and a recording of the same piece (bottom). Both the score and audio are converted to a common midlevel representation: the chromagram. While one would naturally expect the audio chromagram to be much noisier than the chromagram derived from the score, both representations are expected to contain energy in the pitch classes corresponding to the played notes, and to be organized according to the sequence of notes in the score. Therefore, standard alignment procedures based on dynamic time warping can be used to synchronize the two feature sequences. See [39.19, 25, 26] for details.

There are many potential applications of the synchronization process. For example, if the musical form of the piece has been manually annotated using the score representation – as shown in the right-hand side of Fig. 39.5 – the synchronization result can be used to transfer this, or any other annotation, to a recording of the same piece, regardless of variations of local or global tempo. This alleviates the need for the laborious process of manually annotating multiple performances of a given work. The same techniques can be applied to score following for automatic page turning or to adding

subtitles to a video recording of an opera performance, to name only some further applications.

Another application is the creation of novel user interfaces for inter- and intradocument navigation in music collections. Figure 39.6 shows an interface (from [39.27–29]) that allows the user to interact with several recordings of the same piece of music, which have been previously synchronized. The timeline of each recording is represented by a slider bar, whose indicator (a small down arrow) points to the current playback position. Note that these positions are synchronized across performances. A user may listen to a specific recording by activating the corresponding slider bars and then, at any time during playback, seamlessly switch to another recording. Additionally, the slider bars can be segmented and color coded to visualize symbolic annotations common to all recordings, such as a chord transcription or a structure segmentation as shown in the figure. Furthermore, users can jump directly to the beginning of any annotated element simply by clicking on the corresponding block, thus greatly facilitating intradocument navigation. A similar functionality was introduced in [39.30].

The interface offers three different timeline modes. In *absolute mode*, shown in the top of Fig. 39.6, the



Fig. 39.6a,b User interface for intra- and interdocument navigation of music collections. The *slider bars* correspond to four different performances of M68-3. They are segmented and color-coded according to the structural annotations in Fig. 39.2. The *down arrows* point to the current playback time, synchronized across performances. The *absolute (a)* and *relative (b)* timeline modes are shown

length of a particular slider bar is proportional to the duration of the respective recording. In *relative mode*, shown in the bottom of the figure, all timelines are linearly stretched to the same length. The third and final mode, referred to as *reference mode*, uses a single recording as a reference. All other timelines are then

temporally warped to run synchronous to the reference. In all cases, the annotations are adjusted according to the selected mode. Such functionalities open up new possibilities for viewing, comparing, interacting, and evaluating analysis results within a multiversion, and multimode, framework [39.29].

39.4 Self-Similarity in Music Recordings

Since their introduction to the music information community in [39.31], self-similarity matrices (SSM) have become one of the most widely used tools for music structure analysis. Their appeal resides in their ability to characterize patterns of recurrence in a feature sequence, which are closely related to the structure of musical pieces.

39.4.1 Self-Similarity Revisited

Given a sequence of N feature vectors, and a function to measure the pairwise similarity between them, the

self-similarity matrix \mathbf{S} is defined to be the $N \times N$ matrix of pairwise similarities between all feature vectors in the sequence. For the examples in this section we use the chroma features as introduced in Sect. 39.2.1 and the cosine similarity function, i. e., the inner product between the normalized chroma vectors.

Figure 39.7a shows the matrix \mathbf{S} for the chromagram in Fig. 39.3. Both the horizontal and vertical axes represent time, beginning at the bottom-left corner of the plot. Hand-annotated pattern boundaries are marked by vertical and horizontal red lines. Note that if the similarity function is symmetric, the matrix \mathbf{S} is also

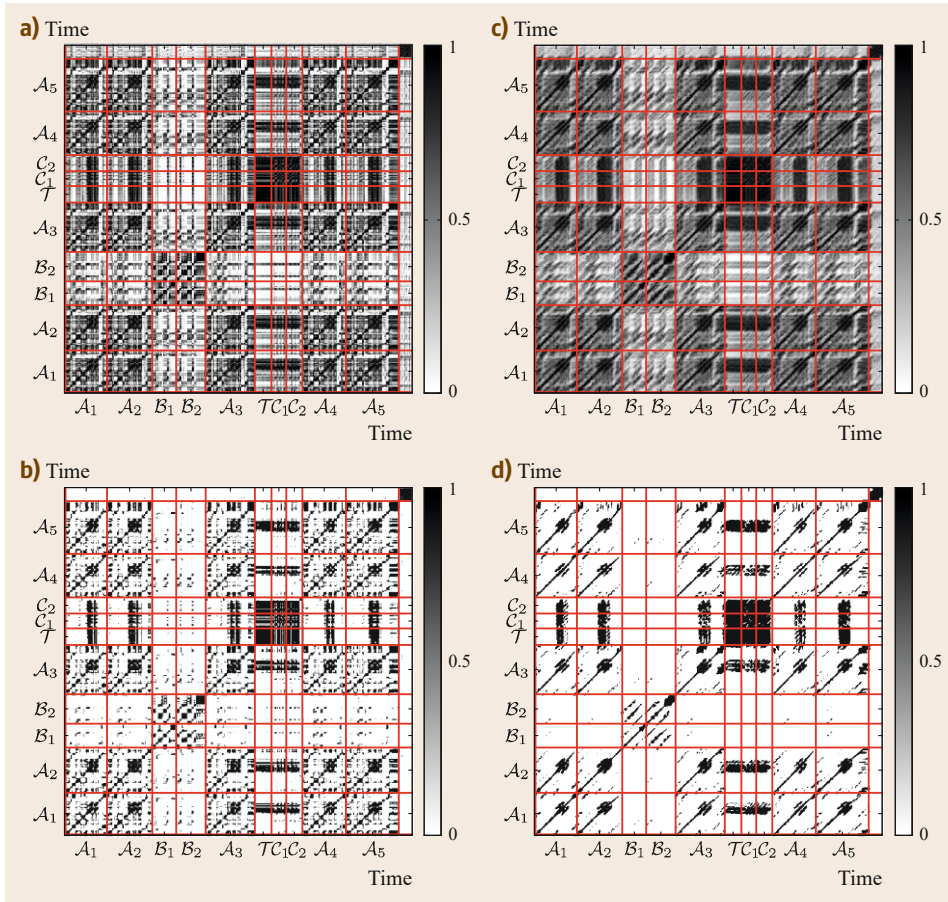


Fig. 39.7a–d Self-similarity matrices for M68-3 performed by Richter, 1976: (a) SSM, (b) thresholded SSM, (c) smoothed SSM, (d) thresholded smoothed SSM. Both vertical and horizontal axes represent time

symmetric with respect to the main diagonal, and that larger/smaller similarity values are represented respectively by darker/lighter colors in the plot. It can be readily observed that similar feature subsequences result in dark diagonal lines or stripes of high similarity. For example, all submatrices corresponding to the two instances of the \mathcal{A} part contain a dark diagonal stripe. The segments corresponding to the \mathcal{C} and \mathcal{T} parts are rather homogeneous with respect to their harmonic content. As a result, entire blocks of high similarity are formed, indicating that each feature vector is similar to every other feature vector within these segments.

This example clearly illustrates how SSMs capture the repetitive structure of music recordings and reveal the location of repeating patterns in the form of diagonal stripes of high similarity. It follows that this information can be exploited for automatic segmentation [39.32], the extraction of musical form [39.19, 33], the detection of chorus sections [39.34], or music thumbnailing [39.35]. For a comprehensive review the reader is referred to [39.36]. However, the reliable extraction of such information from the SSM is

problematic in the presence of distortions caused by variations in dynamics, timbre, note ornaments (e.g., grace notes, trills, arpeggios), modulation, articulation, or tempo. The following section describes how SSMs can be processed to minimize their sensitivity to such distortions.

39.4.2 Enhancing Self-Similarity Matrices

A number of strategies have been proposed to emphasize the diagonal structure information in SSMs. These strategies commonly include a combination of some form of filtering along the diagonals, and the thresholding of spurious, off-diagonal values. While many alternative approaches have been proposed in the literature (e.g., [39.33–35]), here we briefly discuss methods based on time-delay embedding, contextual similarity and transposition invariance.

The process of filtering SSMs along the diagonals can be framed in terms of *time-delay embedding*, a process that has been widely used for the analysis of dynamical systems [39.37]. In this process, the feature

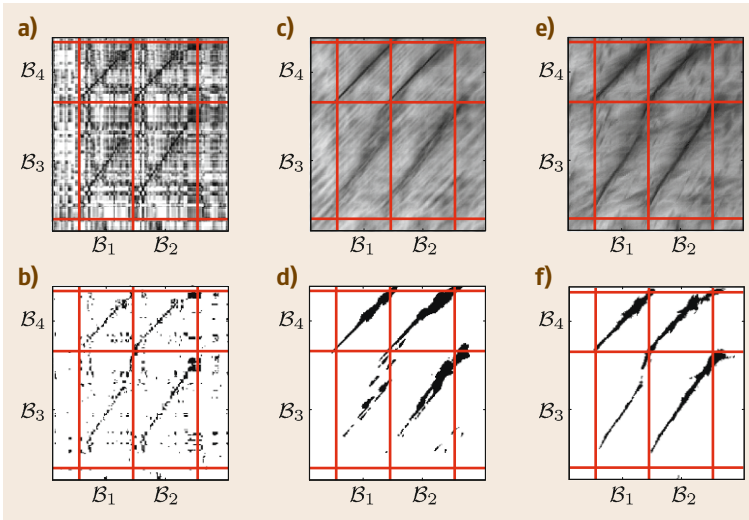


Fig. 39.8a–f SSM smoothing and thresholding in the presence of tempo variations for M30-2 performed by Jonas, 1947: (a,b) SSM, (c,d) smoothed SSM, and (e,f) smoothed SSM with contextual similarity

sequence is converted into a series of feature m -grams, each of which is composed of a stack of m feature vectors. These vectors can be taken from a contiguous window, or spaced by a fixed sample delay τ . In the context of the example in Sect. 39.2.2, setting $m > 1$ implies that pairwise similarities in the SSM computation are computed between subtrajectories instead of between individual feature vectors. Subtrajectories that are parallel to each other result in large self-similarity, and the effect of random crossings is minimized, generally resulting in a smoothed matrix. For $\tau = 1$, the time-delay embedding process is similar to textural windows [39.38] and audio shingles [39.39].

Figure 39.7a,b shows the traditional SSM (i.e., $m = \tau = 1$) computed from M68-3 before and after thresholding, respectively. Similarly, Fig. 39.7c,d shows the corresponding similarity matrices computed using $m = 15$ and $\tau = 1$. It is immediately apparent how the time-delay embedding emphasizes the diagonal line structure of the matrix and significantly minimizes the amount of off-diagonal noise that obscures the structure in the nonembedded matrix. It is worth noting that different methods can be used for thresholding, e.g., by simply ignoring distances larger than a predefined threshold (as was done in this example), by fixing the number of nearest neighbors per sequence element, or by enforcing a rate of recurrence in the thresholded SSM [39.37]. Recent studies show that the latter strategies can improve the applicability of SSMs to a variety of music analysis tasks [39.40, 41].

This form of filtering only works well if the tempo is (close to) constant across the music recording, i.e., where repeating segments have roughly the same length. Stripes corresponding to repeated patterns at the same tempo run diagonally with a slope of 1.

Music structure analysis research has most commonly operated on popular music, in which this constant-tempo assumption is often reasonable. However, in other genres such as romantic piano music, expressive performances often result in significant tempo variations. Take for example Jonas' performance of M30-2, shown in Fig. 39.8, where B_3 is played at nearly half the tempo of B_1 . This results in a diagonal stripe in the SSM with slope close to 2. Applying time-delay embedding to this data can result in broken stripes, as shown in Fig. 39.8c,d. One way to avoid this loss of information is to use a contextual similarity measure that filters the SSM along various slopes around 1 [39.42]. The resulting SSM, illustrated in Fig. 39.8e,f, has filtered out the off-diagonal noise present in the unprocessed SSM shown in Fig. 39.8a while successfully preserving diagonal structure in the presence of local tempo variations.

It is further possible to make the SSM transposition-invariant by computing the similarity between the original chromagram and its 12 cyclically shifted versions, corresponding to all possible transpositions. The invariant SSM is calculated by taking the point-wise minimum over the 12 resulting matrices [39.34, 43]. An example of the process from an excerpt of M7-5 is shown in Fig. 39.9. The excerpt contains repetitions of two patterns, \mathcal{A} and \mathcal{B} , which are actually transpositions of each other, as shown in Fig. 39.9a,b. The standard SSM, shown in Fig. 39.9c,d, differentiates between the repetitions of \mathcal{A} and \mathcal{B} , while the transposition-invariant SSM shown in Fig. 39.9e,f, successfully identifies the relationship between the two patterns.

The enhancement procedures described in this section are aimed at emphasizing the diagonal stripe structure of the SSM. In music structure analysis, the

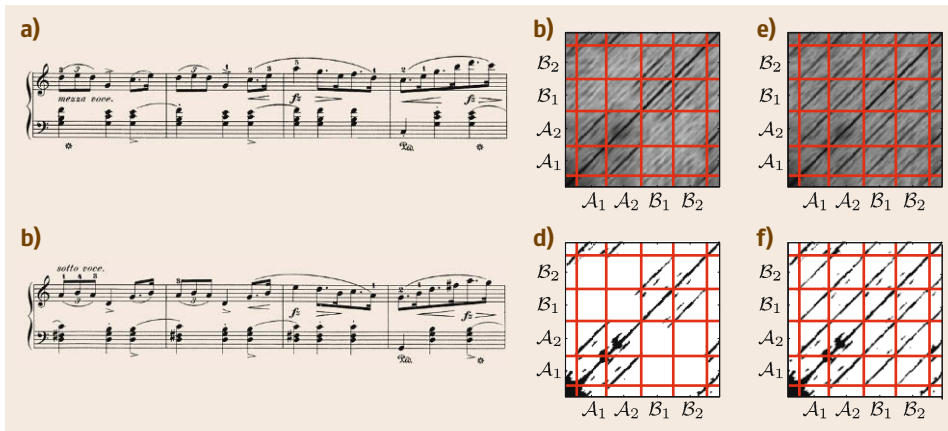


Fig. 39.9a–f Transposition invariance for M7-5 performed by Cohen, 1997: (a,b) Scores of patterns \mathcal{A} and \mathcal{B} , (c,d) SSM before and after thresholding, and (e,f) transposition-invariant SSM before and after thresholding

computation of the self-similarity matrix is usually followed by a *grouping* step, where clustering techniques and heuristic rules are used to identify the global repetitive structure from the matrix, or to select representative segments for e.g., music thumbnailing. This is a large topic that we do not aim to fully review in this chapter. For a detailed review of these solutions and discussions of their advantages and disadvantages, the reader is referred to, e.g., [39.36, 44].

39.4.3 Structure-Based Similarity

In the previous sections, we discussed the computation of SSMs as an intermediate step towards the extraction of a piece's musical form or identification of representative excerpts. In this section, we use SSMs directly as a midlevel representation of a recording's structure. Distances measured in this domain can therefore be used to characterize the *structural similarity* between different music recordings [39.45, 46].

A similar problem in bioinformatics is concerned with measuring the similarity between protein structures on the basis of SSM-like representations known as *contact maps* [39.47]. One solution to this problem makes use of the *normalized compression distance* (NCD), an approximation of the joint Kolmogorov complexity between binary objects, to measure the amount of information overlap between contact maps [39.48]. The NCD is versatile, easy to implement using standard compression algorithms, and does not require alignment between the representations. In [39.46, 49], these ideas are transferred and applied to music information retrieval. Experimental results show that these measures successfully characterize global structural similarity for large and small music collections.

Figure 39.10 depicts the entire Mazurka dataset organized by structural similarity. The plot shows a two-

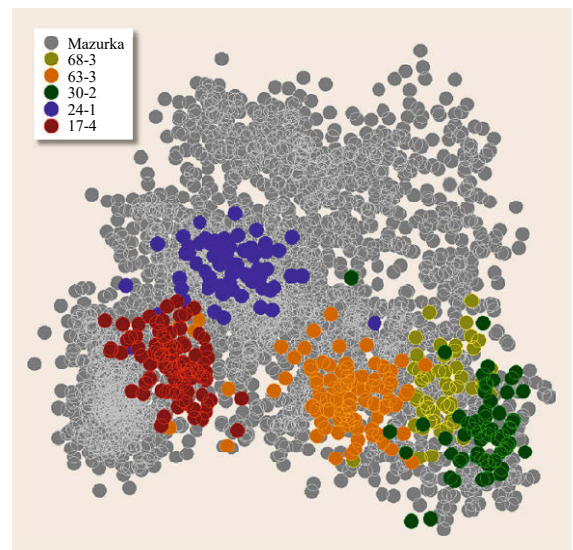


Fig. 39.10 The Mazurka dataset (*gray*) organized according to the NCD-based similarity of SSMs. *Colored points* correspond to recordings of the five Mazurkas considered in this chapter

dimensional projection, obtained using multidimensional scaling [39.50], of the matrix of pairwise NCDs between the SSMs of all 2919 recordings. Each recording is shown as a gray circle, except for those corresponding to performances of M17-4 (depicted in red), M24-1 (blue), M30-2 (green), M63-3 (orange) and M68-3 (olive green).

An informal inspection of this plot shows that most performances of a given work, which feature little or no variation in global structure, naturally form strongly knit clusters in this space. Note that the projection is deceptive in that it makes some of the clusters, e.g., the olive green and orange groups, appear closer to each other than they actually are. As

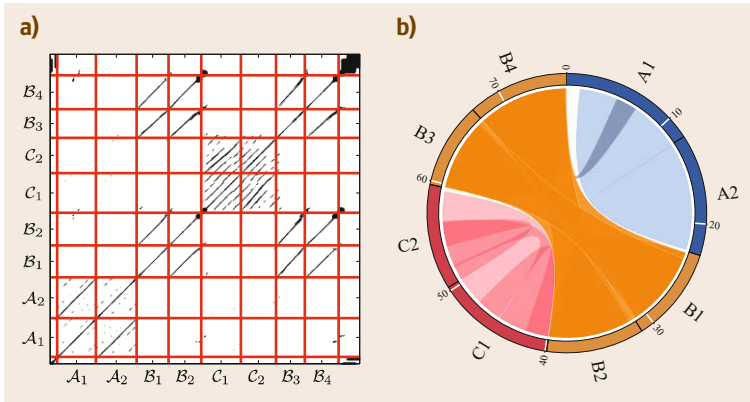


Fig. 39.11a,b Arthur Rubinstein's 1939 performance of M30-2: (a) thresholded SSM, and (b) radial convergence diagram

a result, works can be easily grouped by means of a simple nearest-neighbor search. However, the NCD is, predictably, sensitive to structural changes. This is typified by the M24-1 performance that appears the furthest from the blue cluster. This recording is part of a 1957 release of a master class by Alfred Cortot, where the pianist's spoken commentary alternates (and at times overlaps) with a partial performance of the piece including errors, nonscored pauses and repetitions. This results in a structure that diverges considerably from those of other performances. Likewise, the M30-2 (green) outlier corresponds to a 1945 performance by Vladimir Horowitz, in which the famous pianist repeats the entire second half of the piece, resulting in a unique $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{C}_1\mathcal{C}_2\mathcal{B}_3\mathcal{B}_4\mathcal{C}_3\mathcal{C}_4\mathcal{B}_5\mathcal{B}_6$ structure, see Fig. 39.2c. There are other such examples in this collection, whereby the performer took the liberty to deviate from the notated musical score by playing additional repetitions or leaving out certain parts. A structure-based similarity metric is a natural way to identify these variations.

39.4.4 Visualizing Structure

Self-similarity matrices have been used for visualizing structure in music [39.31, 51] and other domains such as programming [39.52] and computational biology [39.53]. However, SSMs are rarely used for data visualization outside of scientific research, possibly due to the nonintuitive presence of two temporal axes, which results in a complex and redundant representation.

Radial convergence diagrams [39.54] (RCDs) are an alternate vehicle for visualizing structural information derived from SSMs. These diagrams consist of a network of data points arranged on a circle. Groups of related points in the sequence are emphasized by connecting them via links or ribbons. We use the *cir-*

cos toolbox [39.55], a powerful and popular tool for information visualization based on circular graphs, to generate the RCDs shown in this section [39.54].

An example RCD and the associated SSM computed from a recording of M30-2 is shown in Fig. 39.11. To generate the RCD, the matrix is postprocessed as follows: the lower triangular part as well as the main diagonal and all points within a prespecified distance from it are ignored. The remaining repetitions (*ones* in the matrix) are stored as a list of *links* connecting pairs of elements in the feature sequence. Nearby links in time (i. e., those that form diagonal strips in the SSM) are grouped together. Small groups, either in temporal scope or in membership, are filtered out.

The remaining groups are used to construct the radial convergence diagram where time increases clockwise along the perimeter of the circle, beginning and ending at the topmost point. The outer ring shows the sequence of annotated sections as blocks, colored according to the schema in Fig. 39.2. Time markers are drawn as white lines and are placed at ten-second intervals along the ring, and section boundaries are drawn in black. Groups are represented using translucent ribbons connecting two time segments such that each ribbon edge connects the beginning of the first segment to the end of the second. If they link two instances of the same pattern, these ribbons assume the color of that pattern. Otherwise they are depicted in gray, which is also used to denote special sections such as introductions, interludes, transitions and codas.

As is shown in Fig. 39.11b, the RCD seamlessly combines information from the SSM with the piece's structural annotation, resulting in a rich and appealing visualization. The annotation in this example was manually generated on the basis of a musical score and propagated to audio recordings using the synchronization approach described in Sect. 39.3. Alternatively, the annotations can be automatically generated

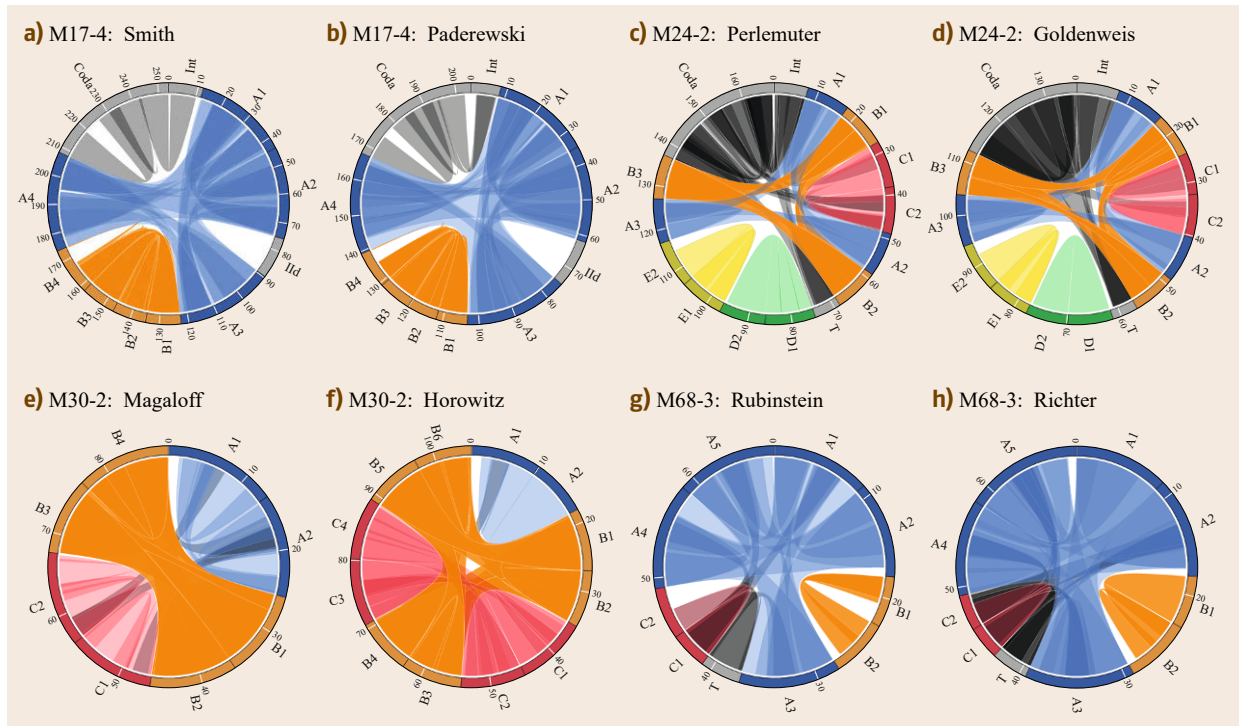


Fig. 39.12a–h Radial convergence diagrams for various Mazurka performances

by, e.g., analyzing the SSM (Sect. 39.4.2), or using the factorization approach that will be introduced in Sect. 39.5.

To illustrate the RCD’s potential for qualitative analysis, Fig. 39.12 shows diagrams for two recordings of each of the following four Mazurkas: M17-4 in Fig. 39.12a,b, M24-2 in Fig. 39.12c,d, M30-2 in Fig. 39.12e,f, and M68-3 in Fig. 39.12g,h. The presence or absence of repetitive patterns results in distinctive shapes inside the diagrams that complement the information in the block-based structural annotations. These shapes highlight, for instance, the unique (nonrepetitive) character of the interlude (bottom-right gray segment) in M17-4, the significant overlap between introduction, transition and coda (gray sections) in M24-2, and the strong predominance of pattern \mathcal{A} in the structure of M68-3.

Note that the gray ribbons linking different patterns depict relationships that are absent from the structural annotations. For example, they highlight the high similarity between instances of pattern C and the transition section in M68-3, both of which are dominated by a B^b

major chord as described previously in Sects. 39.1 and 39.2. The gray ribbons also connect the C and \mathcal{T} sections with instances of the same chord in pattern \mathcal{A} , a link that results from the piece’s only key modulation, in the transition, from F major to B^b major.

The RCD representation can also be used to qualify stylistic differences in performance practice. Take for example one of the outliers identified in the analysis in Sect. 39.4.3. Figure 39.12f shows the unique structure of Vladimir Horowitz’s performance of M30-2 when compared to performances that more closely follow the original score in Figs. 39.12e or 39.11b. Likewise, the varying-length gaps at the end of section C_2 (the final red segment) in M68-3, indicate a ritardando and an expressive pause in Rubinstein’s performance (Fig. 39.12g), which is not present in Richter’s (Fig. 39.12h).

Please note that an in-depth musicological analysis based on these diagrams is beyond the scope of this chapter. These simple observations are only intended to illustrate some of the capabilities of the RCD representation.

39.5 Automated Extraction of Repetitive Structures

Section 39.4.2 describes some general strategies for identifying repetitive structure within music recordings using self-similarity matrices. Example approaches include the clustering of diagonal elements in the SSM representation into sets of disjoint segments, and selecting the first of the most similar pair of segments, or choosing an arbitrary occurrence of the most frequent segment as a representative pattern. In this section we present an alternative approach based on matrix factorization, which jointly identifies repetitive patterns and derives a global structure segmentation from audio features.

39.5.1 Structure Analysis Using Matrix Factorization

In [39.56, 57] we propose a novel approach for the localization and extraction of repeating patterns in music audio. The idea underlying this method is that a song can be represented through repetitions of a small number of patterns in feature space. For example, recall that M68-3 has the form $\mathcal{A}_1\mathcal{A}_2\mathcal{B}_1\mathcal{B}_2\mathcal{A}_3\mathcal{T}C_1C_2\mathcal{A}_4\mathcal{A}_5$, i. e., the piece exhibits repetitions of three parts (\mathcal{A} , \mathcal{B} , and \mathcal{C}) as well as a transition \mathcal{T} that occurs only once. The only information needed to represent this piece are the feature sequences corresponding to each of the four patterns and the points in time of their occurrences.

This observation can be exploited to identify the main parts of a music recording and its overall temporal structure using an approach known as *shift-invariant probabilistic latent component analysis* (SI-PLCA). The process is illustrated in Fig. 39.13 for a performance of M68-3 by Gábor Csálog from 1996. The chromagram in Fig. 39.13a is approximated as the weighted sum of K ($K = 4$ in this example) components, each of

which corresponds to a different part. Each component is further decomposed into a short chroma *basis pattern* (shown in Fig. 39.13b) and an activation function defining the location of each repetition of that pattern in the chromagram (shown in Fig. 39.13c). The peak positions in the activation function denote occurrences of the corresponding basis pattern in the feature sequence. Given a chromagram, this decomposition into basis patterns and activations can be computed iteratively using well-known optimization techniques [39.57].

There are many advantages to this model. First, it operates in a purely data-driven, unsupervised fashion: outside of a few functional parameters, the only prior information it requires is the length of the patterns L and the number of patterns K to extract. Second, the probabilistic formulation makes it straightforward to impose sparse priors on the distribution of basis patterns and mixing weights, enabling the model to learn optimal values for L and K . In addition, the framework can be easily extended to be invariant to key transpositions using a technique similar to that described in Sect. 39.4.2. Finally, the method jointly estimates the set of patterns and their activations, thus avoiding the need for heuristics for pattern extraction, grouping and selection. The versatility of the model is demonstrated in [39.57], where it is successfully applied to riff finding, meter identification, and segmentation of popular music.

39.5.2 Representative Patterns

The SI-PLCA model can be easily extended to extract patterns jointly from all available recordings of a piece by sharing the bases across recordings but using different activation functions for each one. This

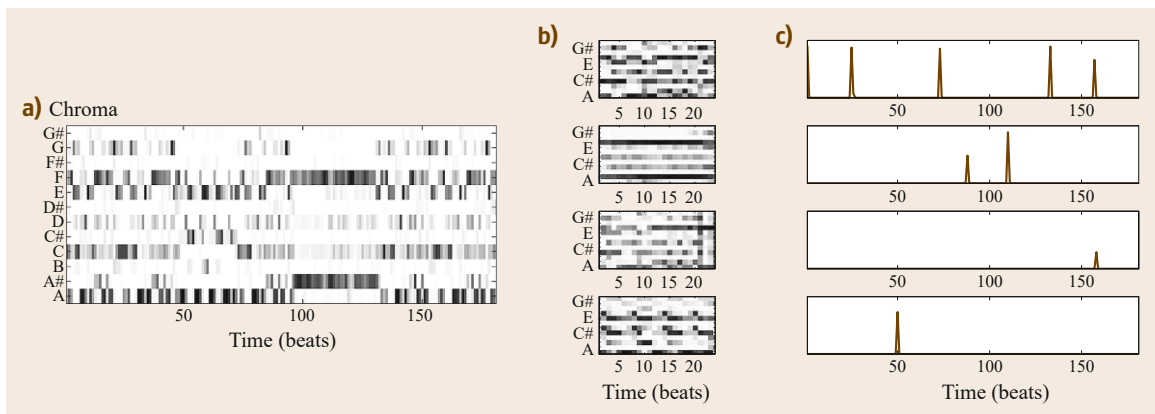


Fig. 39.13a–c Illustration of the SI-PLCA decomposition. (a) Chromagram. (b) Set of four basis patterns. (c) Respective activations in time

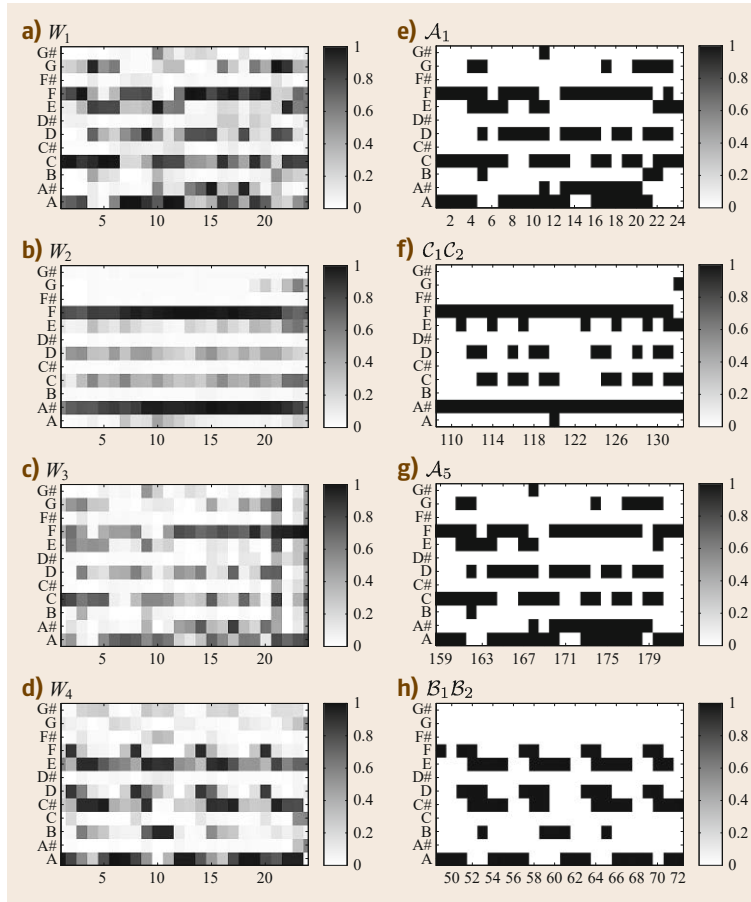


Fig. 39.14a–h Extracted patterns $W_k, k \in [1 : 4]$ for M68-3 and the corresponding MIDI-generated ground-truth patterns for the different parts. To facilitate comparison, (f) and (h) show both instances of the parts C and B respectively. The vertical axis shows time information given in beats. For the ideal patterns, the absolute beat numbers indicate the position of the pattern in the piece, see Fig. 39.2e

allows the model to identify the patterns of the piece rather than capture the idiosyncrasies of a single performance. In order to facilitate this cross-recording analysis we use beat-synchronous chromagrams, where features are averaged within beat segments [39.15]. This leads to a representation containing one feature vector per beat. We employ the manually annotated beat positions available for the Mazurka dataset to generate the chromagrams shown in this section.

Figure 39.14 shows patterns extracted for M68-3. The number of patterns in the decomposition is set to $K = 4$ and the pattern length is set to $L = 24$ beats, the length of \mathcal{A} , the longest part in the piece. The extracted patterns, denoted W_k , are shown in the left column of Fig. 39.14. The right column shows *ideal* chromagrams for the parts of M68-3, extracted from a symbolic (MIDI) representation. These sequences are paired with the basis that is most similar.

Upon close examination, it can be seen that W_1 matches every chord in \mathcal{A}_1 , aside from some ambiguity in beats 21, 23 and 24. This is due to the alternation of endings of different instances of the \mathcal{A} pattern: \mathcal{A}_1 and \mathcal{A}_4 finish with Gmaj-Cmaj-Cmaj-Cmaj chords at beats

21–24; while \mathcal{A}_2 , \mathcal{A}_3 and \mathcal{A}_5 end with a Gmin-Cmaj-Fmaj-Fmaj sequence. Correspondingly, beat 21 shows energy in pitch classes G, B, B^b/A[#] and D; while beats 23 and 24 show activity for classes A, C, E and F. Likewise, W_4 matches $\mathcal{B}_1\mathcal{B}_2$ almost perfectly, aside from a slight shift: W_4 starts in beat 2 of \mathcal{B}_1 and finishes in the first beat of \mathcal{A}_3 , which contains a transitional C major seventh chord.

The melodic components of W_2 are also shifted by a couple of beats, although the transitional C dominant seventh chord at the end is well aligned. However, most of the energy of the matrix is concentrated in the B^b fifth chord that dominates the C and \mathcal{T} patterns. As a result, W_2 matches activations of both these patterns, as can be seen in the corresponding activation vectors in Fig. 39.13c. The unintended consequence of this is that the remaining pattern W_3 , which must be assigned to some portion of the signal, absorbs a small amount of the energy from the last instance of pattern \mathcal{A} , mostly from the long F major chord at the end. W_3 is largely redundant with W_1 , and could therefore be easily discarded. This is also indicated by the low value of the corresponding activation in Fig. 39.13c.

39.5.3 Segmentation Analysis

The work in [39.56] also describes a method for segmenting an audio signal based on the contribution of each basis to the chromagram. This is done by computing a temporal localization function ℓ_k from the extracted bases and activations that represents the energy contribution of each component to the overall chromagram at each point in time. Figure 39.15 shows the localization functions for the five Mazurkas in Fig. 39.2. These results are obtained for $L = 24$ beats and the optimal setting of K for each piece.

Figure 39.15e shows ℓ_k for M68-3, our running example. Given the similarity between patterns C and T , discussed in previous sections, we have chosen to set $K = 3$, which removes the spurious component depicted in Fig. 39.14c. The resulting analysis shows a strong correspondence between the localization function and the annotated parts. The annotations for all \mathcal{A} and \mathcal{B} parts align perfectly with ℓ_2 and ℓ_1 respectively. ℓ_3 corresponds primarily to the C and T parts, and also makes a small contribution to \mathcal{A}_3 .

As shown in Fig. 39.15c, the localization functions for M30-2 correlate closely with the manual annotation, given an optimal choice of $K = 3$. Similarly for M17-4 in Fig. 39.15a, there is high correlation between ℓ_k and the annotated parts. However, there is a constant shift in the segmentation that is mainly caused by the 12-beat long introduction. Additionally, the choice of $K = 3$ results in the grouping of the introduction and interlude sections (and some of the coda) with the \mathcal{A} pattern.

The segmentation is less straightforward for M24-2 and M63-3, shown in Fig. 39.15b,d. In the case of M24-2, the choice of $K = 7$ closely matches the number of distinct parts in this piece, also visualized in the diagrams in Fig. 39.12c,d. As a result, C , D , E , the introduction and coda are slightly shifted but well segregated. However, the decomposition combines the contributions of \mathcal{A} and \mathcal{B} , which always occur in sequence, into ℓ_2 . Similarly, the ending of \mathcal{B} and the transition are merged together in ℓ_4 . For M63-3, the choice of $K = 6$ is larger than the number of parts in the annotation. This is a consequence of strong variations between the annotated \mathcal{A} parts, where only the first 12 beats are common to all instances. These variations lead to a separation of the \mathcal{A} parts into different components.

To further alleviate possible errors, we apply temporal smoothing to the localization functions, and select the maximal contributing pattern for each time position. This results in segmentation results comparable to state-of-the-art methods [39.56]. It is important to note that, while the results discussed in this section reveal a sensitivity to the setting of K , our research shows that it is

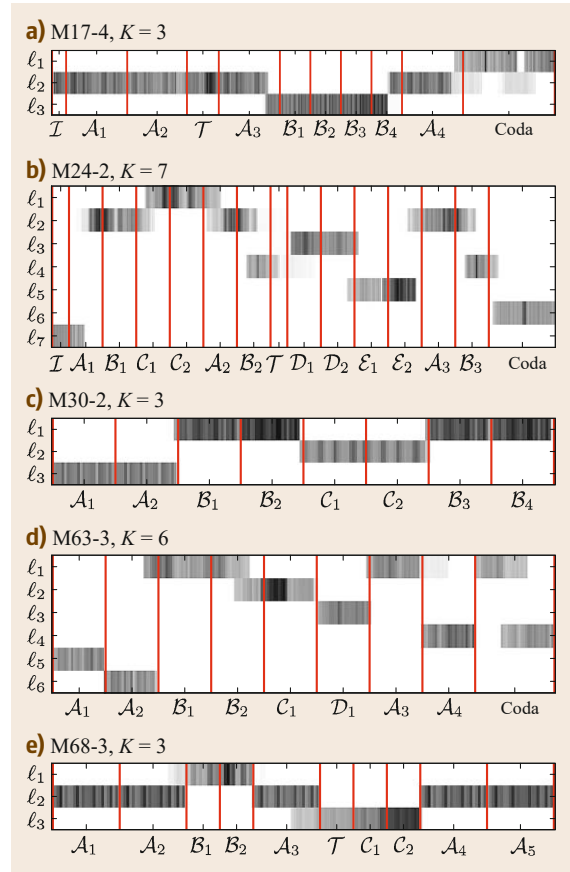


Fig. 39.15a–e Contribution ℓ_k of the basis patterns W_k to the different parts using $L = 24$ and the optimal setting of K for each of the five Mazurkas

possible to decrease this sensitivity by imposing additional sparsity constraints on the model parameters.

It is also important to observe that the above analysis assumes that all patterns are of the same length. This is of course unrealistic for most music. When analyzing popular music this is partly alleviated via beat-synchronous analysis. However, extending this strategy to the analysis of piano music of the classical and romantic canon is problematic, because the variability introduced by expressive tempo changes negatively affects the performance of beat tracking systems. See [39.58] for an in-depth discussion of this problem in the context of the Mazurka dataset. The examples in this section have avoided such complications by making use of beat-synchronous chromagrams based on manually annotated beat positions. In order to analyze music for which such annotations are not available, we could, in the future, extend the SI-PLCA model to be invariant to tempo changes by allowing the basis patterns to be stretched in time.

39.6 Conclusions

In this chapter, we have reviewed the general principles behind several computational approaches to content-based analysis of recorded music and discussed their application to knowledge discovery. For example, we demonstrated how chroma-based audio features can be used as a robust midlevel representation for capturing harmonic information, and to compute similarity matrices that reveal recurrent patterns. By including temporal contextual information, we showed how these matrices can be enhanced to reveal repetitions in the presence of local tempo variations as occurring in expressive music recordings. Furthermore, we discussed how these techniques can be applied to music synchronization, to automate audio annotation and to realize novel user interfaces for inter- and intradocument music navigation. Finally, we demonstrated how structure-based similarity measures can be used to classify music collections based on their global structural form, and how matrix factorization techniques can be used to identify the most representative building blocks of a recorded performance.

Our aim was to highlight the implications of the various approaches for the analysis and understanding of recorded music, and to demonstrate how they can empower the work of music professionals and scholars. In addition we showed how these tools facilitate novel modes of interaction with digital content that could be integrated with music distribution services as ways to enhance the listeners' understanding and appreciation of music. Throughout the text, we illustrated the potential of content-based analysis via many concrete and intuitive examples taken from a collection of recorded piano performances of Chopin's Mazurkas. However, the presented techniques and underlying principles are applicable to the analysis of a wide range of tonal music and, with appropriate modifications of the feature

representation, to percussive music or other types of time-dependent structured multimedia data.

Of course, this chapter has only covered a small part of the breadth of techniques and problems in the area of automated music processing. There are numerous challenges and open issues for future research. For instance, many of the examples shown here rely, to some extent, on the use of manually generated expert data such as beat positions or structural annotations. Despite significant research efforts, the automated generation of such annotations using purely content-based analysis techniques still poses challenging and open research problems for large parts of the existing digital music catalog. State-of-the-art algorithms often lack the robustness, accuracy and reliability needed for many music analysis applications. This is especially true for highly expressive music recordings such as the Mazurka performances described in this chapter, which exhibit subtle differences and variations in tempo, articulation, and note execution.

As the amount of digitally available music-related data grows, so does the need for efficient approaches that can scale up to the processing of millions of tracks. As automated procedures become more and more powerful they tend to gain computational complexity, making scalability an increasingly important issue. In addition to improvements in robustness and accuracy, issues related to time and memory efficiency will progressively come into the focus of future research.

Acknowledgments. This material is based upon work supported by the National Science Foundation, under grant IIS-0844654, and the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University. The authors would like to thank Craig Sapp for kindly providing access to the Mazurka dataset and beat annotations.

References

- 39.1 M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney: Content-based music information retrieval: Current directions and future challenges, *Proc. IEEE* **96**(4), 668–696 (2008)
- 39.2 M. Slaney: Web-scale multimedia analysis: Does content matter?, *Multimed. IEEE* **18**(2), 12–15 (2011)
- 39.3 H. Schenker: *Der freie Satz* (Universal, Vienna 1935)
- 39.4 A. Ockelford: *Repetition in Music: Theoretical and Metatheoretical Perspectives* (Ashgate, London 2005)
- 39.5 D. Huron: *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge 2006)
- 39.6 M.J. Bruderer, M. McKinney, A. Kohlrausch: Structural boundary perception in popular music. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR)*, Victoria (2006) pp. 198–201
- 39.7 G. Peeters, E. Deruty: Is music structure annotation multi-dimensional? A proposal for robust local music annotation. In: *Proc. 3rd Workshop Learn. Semant. Audio Signals*, Graz (2009) pp. 75–90
- 39.8 The AHRC Research Centre for the History and Analysis of Recorded Music: Website of the Mazurka Project, <http://www.mazurka.org.uk/>

- 39.9 C.S. Sapp: Comparative analysis of multiple musical performances. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 497–500
- 39.10 C.S. Sapp: Hybrid numeric/rank similarity metrics. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Philadelphia* (2008) pp. 501–506
- 39.11 E. Pampalk: *Computational Models of Music Similarity and Their Application to Music Information Retrieval*, Ph.D. Thesis (Vienna University of Technology, Vienna 2006)
- 39.12 S. Essid: *Classification Automatique des Signaux Audio-Fréquences: Reconnaissance des Instruments de Musique*, Ph.D. Thesis (Université Pierre et Marie Curie, Paris 2005)
- 39.13 G. Peeters: A large set of audio features for sound description (similarity and classification) in the CUIDADO project, http://recherche.ircam.fr/anasy/n/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf (Ircam, Analysis/Synthesis Team, Paris 2004), version 1.0
- 39.14 A. Sheh, D.P.W. Ellis: Chord segmentation and recognition using EM-trained hidden Markov models. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Baltimore* (2003)
- 39.15 D.P.W. Ellis, G.E. Poliner: Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Honolulu* (2007)
- 39.16 J. Serrà, E. Gómez, P. Herrera, X. Serra: Chroma binary similarity and local alignment applied to cover song identification, *IEEE Trans. Audio Speech Lang. Process.* **16**, 1138–1151 (2008)
- 39.17 E. Gómez: *Tonal Description of Music Audio Signals*, Ph.D. Thesis (Universitat Pompeu Fabra, Barcelona 2006)
- 39.18 M. Mauch, K. Noland, S. Dixon: Using musical structure to enhance automatic chord transcription. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Kobe* (2009) pp. 231–236
- 39.19 M. Müller: *Information Retrieval for Music and Motion* (Springer, Berlin, Heidelberg 2007)
- 39.20 R.N. Shepard: Circularity in judgments of relative pitch, *J. Acoust. Soc. Am.* **36**(12), 2346–2353 (1964)
- 39.21 T. Fujishima: Realtime chord recognition of musical sound: A system using common lisp music. In: *Proc. ICMC, Beijing* (1999) pp. 464–467
- 39.22 M. Mauch, S. Dixon: Approximate note transcription for the improved identification of difficult chords. In: *Proc. 11th Int. Soc. Music Inf. Retr. Conf. (ISMIR), Utrecht* (2010) pp. 135–140
- 39.23 M. Müller, S. Ewert: Towards timbre-invariant audio features for harmony-based music, *IEEE Trans. Audio Speech Lang. Process.* **18**(3), 649–662 (2010)
- 39.24 I.T. Jolliffe: *Principal Component Analysis* (Springer, New York 2002)
- 39.25 N. Hu, R.B. Dannenberg, G. Tzanetakis: Polyphonic audio matching and alignment for music retrieval. In: *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA), New Paltz* (2003)
- 39.26 S. Ewert, M. Müller, P. Grosche: High resolution audio synchronization using chroma onset features. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Taipei* (2009) pp. 1869–1872
- 39.27 C. Fremerey, F. Kurth, M. Müller, M. Clausen: A demonstration of the SyncPlayer system. In: *Proc. 8th Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 131–132
- 39.28 D. Damm, C. Fremerey, F. Kurth, M. Müller, M. Clausen: Multimodal presentation and browsing of music. In: *Proc. 10th Int. Conf. Multimodal Interfaces (ICMI), Chania* (2008) pp. 205–208
- 39.29 M. Müller, V. Konz, N. Jiang, Z. Zuo: A multi-perspective user interface for music signal analysis. In: *Proc. Int. Computer Music Conf. (ICMC), Huddersfield* (2011)
- 39.30 M. Goto: A chorus section detection method for musical audio signals and its application to a music listening station, *IEEE Trans. Audio Speech Lang. Process.* **14**(5), 1783–1794 (2006)
- 39.31 J. Foote: Visualizing music and audio using self-similarity. In: *Proc. ACM Int. Conf. Multimed., Orlando* (1999) pp. 77–80
- 39.32 J. Foote: Automatic audio segmentation using a measure of audio novelty. In: *Proc. IEEE Int. Conf. Multimed. Expo (ICME), New York* (2000) pp. 452–455
- 39.33 G. Peeters: Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 35–40
- 39.34 M. Goto: A chorus-section detecting method for musical audio signals. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Hong Kong* (2003) pp. 437–440
- 39.35 M.A. Bartsch, G.H. Wakefield: Audio thumbnailing of popular music using chroma-based representations, *IEEE Trans. Multimed.* **7**(1), 96–104 (2005)
- 39.36 J. Paulus, M. Müller, A. Klapuri: Audio-based music structure analysis. In: *Proc. 11th Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 625–636
- 39.37 N. Marwan, M.C. Romano, M. Thiel, J. Kurths: Recurrence plots for the analysis of complex systems, *Phys. Rep.* **438**(5/6), 237–329 (2007)
- 39.38 G. Tzanetakis, P. Cook: Musical genre classification of audio signals, *IEEE Trans. Speech Audio Process.* **10**(5), 293–302 (2002)
- 39.39 M. Slaney, M. Casey: Locality sensitive hashing for finding nearest neighbours, *IEEE Signal Process. Mag.* **25**(2), 128–131 (2008)
- 39.40 J. Serrà, X. Serra, R.G. Andrzejak: Cross recurrence quantification for cover song identification, *New J. Phys.* **11**(9), 093017 (2009)
- 39.41 T. Cho, J. Forsyth, L. Kang, J.P. Bello: Time-varying delay effects based on recurrence plots. In: *Proc. 14th Int. Conf. Digit. Audio Eff. (DAFx), Paris* (2011)
- 39.42 M. Müller, F. Kurth: Enhancing similarity matrices for music audio analysis. In: *Proc. 32nd Int. Conf. Acoust. Speech Signal Process. (ICASSP), Toulouse* (2006) pp. 437–440
- 39.43 M. Müller, M. Clausen: Transposition-invariant self-similarity matrices. In: *Proc. 8th Int. Conf. Music Inf. Retr. (ISMIR), Vienna* (2007) pp. 47–50
- 39.44 R.B. Dannenberg, M. Goto: Music structure analysis from acoustic signals. In: *Handbook of Signal*

- Processing in Acoustics*, Vol. 1, ed. by D. Havelock, S. Kuwano, M. Vorländer (Springer, New York 2008) pp. 305–331
- 39.45 T. Izumitani, K. Kashino: A robust musical audio search method based on diagonal dynamic programming matching of self-similarity matrices. In: *Proc. 9th Int. Conf. Music Inf. Retr. (ISMIR), Philadelphia* (2008) pp. 609–613
- 39.46 J.P. Bello: Measuring structural similarity in music, *IEEE Trans. Audio Speech Lang. Process.* **19**(7), 2013–2025 (2011)
- 39.47 W. Xie, N.V. Sahinidis: A Branch-and-reduce algorithm for the contact map overlap problem, *Res. Comput. Biol. (RECOMB 2006)*, *Lect. Notes Bioinform.* **3909**, 516–529 (2006)
- 39.48 N. Krasnogor, D.A. Pelta: Measuring the similarity of protein structures by means of the universal similarity metric, *Bioinformatics* **20**(7), 1015–1021 (2004)
- 39.49 J.P. Bello: Grouping recorded music by structural similarity. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Kobe* (2009)
- 39.50 I. Borg, P. Groenen: *Modern Multidimensional Scaling* (Springer, New York 1997)
- 39.51 P. Toivainen: Visualization of tonal content with self-organizing maps and self-similarity matrices, *Comput. Entertain.* **3**(4), 1–10 (2005)
- 39.52 K.W. Church, J.I. Helfman: Dotplot: A program for exploring self-similarity in millions of lines for text and code, *J. Am. Stat. Assoc., Inst. Math. Stat. Interface Found. North Am.* **2**(2), 153–174 (1993)
- 39.53 E.L.L. Sonnhammer, J.C. Wootton: Dynamic contact maps of protein structures, *J. Mol. Graph. Modell.* **16**(33), 1–5 (1998)
- 39.54 M. Lima: VC blog on *Radial Convergence*, <http://www.visualcomplexity.com/vc/blog/?p=876> (2011)
- 39.55 M.I. Krzywinski, J.E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra: Circo: An information aesthetic for comparative genomics, *Genome Res.* **19**(9), 1639–1645 (2009)
- 39.56 R.J. Weiss, J.P. Bello: Identifying repeated patterns in music using sparse convolutive non-negative matrix factorization. In: *Proc. Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 123–128
- 39.57 R.J. Weiss, J.P. Bello: Unsupervised discovery of temporal structure in music, *IEEE J. Sel. Top. Signal Process.* **5**(6), 1240–1251 (2011)
- 39.58 P. Grosche, M. Müller, C.S. Sapp: What makes beat tracking difficult? A case study on Chopin Mazurkas. In: *Proc. 11th Int. Conf. Music Inf. Retr. (ISMIR), Utrecht* (2010) pp. 649–654