CONVERTING PATH STRUCTURES INTO BLOCK STRUCTURES USING EIGENVALUE DECOMPOSITIONS OF SELF-SIMILARITY MATRICES

Harald Grohganz, Michael Clausen
Bonn University
{grohganz,clausen}@cs.uni-bonn.de

Nanzhu Jiang, Meinard Müller
International Audio Laboratories Erlangen
{nanzhu.jiang,meinard.mueller}@audiolabs-erlangen.de

ABSTRACT

In music structure analysis the two principles of repetition and homogeneity are fundamental for partitioning a given audio recording into musically meaningful structural elements. When converting the audio recording into a suitable self-similarity matrix (SSM), repetitions typically lead to path structures, whereas homogeneous regions yield block structures. In previous research, handling both structural elements at the same time has turned out to be a challenging task. In this paper, we introduce a novel procedure for converting path structures into block structures by applying an eigenvalue decomposition of the SSM in combination with suitable clustering techniques. We demonstrate the effectiveness of our conversion approach by showing that algorithms previously designed for homogeneity-based structure analysis can now be applied for repetition-based structure analysis. Thus, our conversion may open up novel ways for handling both principles within a unified structure analysis framework.

1. INTRODUCTION

The task of music structure analysis with the objective of partitioning a given audio recording into temporal segments and of grouping these segments into musically meaningful categories constitutes a central task in the field of music information retrieval [14]. Because of different structure principles including temporal order, repetition, contrast, variation, and homogeneity, finding the musical structure is a challenging and often ill-defined problem [18]. In particular, the two principles of repetition and homogeneity have been in the focus of previous research efforts [14, 15]. On the one hand, repetition-based methods target at identifying recurring patterns and, on the other hand, homogeneity-based methods try to determine passages that remain unchanged with respect to some musical property. When converting the given audio recording into a suitable feature sequence and then deriving a self-similarity matrix (SSM), repetitions typically lead to path-like structures, whereas homogeneous regions yield block-like structures. In previous research, numerous extraction and clustering techniques have been proposed that either allow for handling path structures or block structures, see, e g., [1, 5, 8, 12, 13, 14, 15]. However, dealing with both structural elements at the same time has turned out to be a challenging or even irreconcilable task.

Only few approaches that try to apply several segmentation principles at the same time exist. In [13], a unifying optimization scheme that jointly accounts for path and block structures is proposed. In [17], structural changes with regard to path and block elements are captured to derive segment boundaries. In [4, 8], approaches for homogeneity-based structure analysis are introduced, where smoothing and clustering techniques are applied for enhancing the block structure in a pre-processing step. A very interesting research direction is sketched in [16], where an audio recording is locally classified according to the properties of being repetitive or homogeneous with the goal to locally adapt the segmentation strategy.

Along these lines of research, we deal in this paper with the task of converting a repetition-based into a homogeneity-based structure analysis problem. Opposed to [4, 8], who try to enhance already latent block structure by applying smoothing and image processing techniques, we generate the block structures from path structures. As our main technical contribution, we propose a novel procedure that is based on an eigenvalue decomposition of the SSM. We show that certain path structures induce some disjoint properties of the supports (non-zero entries) of the eigenvectors. This in turn allows us to generate block-like structures when converting back the suitably clustered eigenvectors into some SSM. A typical result of our procedure is shown in Figure 1, which shows the original path structure in (b) and the resulting block structure in (e). The underlying piece of music has the musical form [1] $A_1B_1A_2B_2C_1C_2D_1D_2A_3B_3A_4B_4$. Note that in our procedure an explicit extraction of the path structure, often a fragile step used in repetition-based structure analysis, is not necessary.

The general idea of using eigenvalue decompositions of SSMs with applications to audio segmentation is not new, see e. g. [2]. However, in [2] this techniques is used for the purpose of dimensionality reduction and clustering, whereas we exploit specific properties of the eigenvectors

[1] As in [14], musical parts are denoted by the letters $A, B, C, \dots$ in the order of their first occurrence, where indices are used to indicate repetitions.

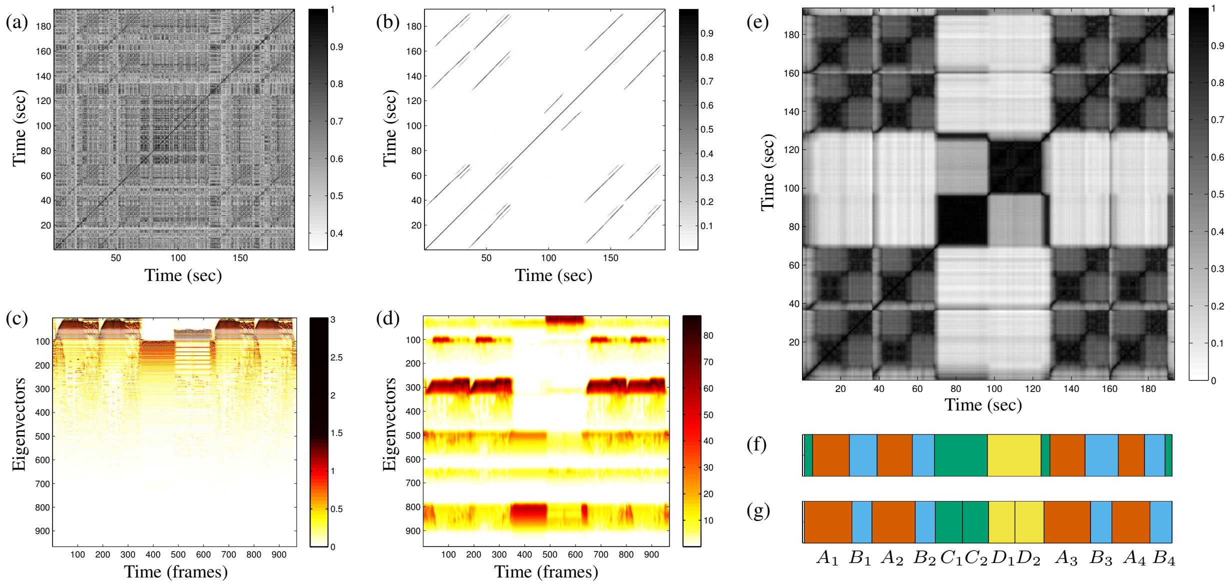

Figure 1: Illustration of the algorithmic pipeline for converting path into block structures. **(a)** Original SSM. **(b)** Path-enhanced SSM as typically used for repetition-based structure analysis. **(c)** Eigenvectors (rows) of the SSM weighted and sorted by the corresponding eigenvalues. **(d)** Eigenvectors (rows) after clustering and post-processing. **(e)** SSM obtained from (d). **(f)** Structure analysis result obtained from (e) applying some homogeneity-based clustering procedure. **(g)** Manually labeled structure (ground truth).

for converting path into block structures.

We demonstrate the effectiveness of our conversion approach by discussing a number of explicit examples and by presenting some quantitative evaluation based on two known datasets. In particular, we show that upon our conversion standard clustering procedures that are designed for homogeneity-based structure analysis can then be applied for repetition-based structure analysis. As a results, our conversion may open up novel ways for combining different music segmentation principles at an early stage of an audio structure processing pipeline.

In the remainder of this paper, we first describe the algorithmic details for our conversion procedure (Section 2), then report on our experiments (Section 3), and conclude with an outlook on future research directions (Section 4).

2. ALGORITHMIC PIPELINE

In this section, we describe our procedure for converting SSMs with path-like structures into SSMs with block-like structures, see also Figure 1 for an illustration of the overall pipeline. We first summarize a procedure for computing an SSM with path structures as typically used for repetition-based structure analysis (Section 2.1). Such matrices constitute the input of our conversion procedure. We then describe some properties of the eigenvectors obtained from such matrices (Section 2.2) and show how the eigenvectors can be used to derive SSMs with block-like structures (Section 2.3).

2.1 Computing SSMs with Path Structures

Repeating segments in music often share the same melodic and harmonic progression while showing differences in instrumentation and timbre. Therefore, in most approaches for repetition-based structure analysis, the audio signal is converted into twelve-dimensional chroma-based audio features, which closely correlate to the aspect of harmony and have become a widely used tool in processing and analyzing music data [3]. In the following, we use a chroma variant referred to as CRP (Chroma DCT-Reduced Log Pitch) features [2], which show a high degree of invariance to changes in timbre [10]. In our experiments, we adapt the feature rate according to the length of the considered audio recording, which results in feature rates (features per second) between 2 Hz and 6 Hz. Normalizing the features, we use the inner product as a similarity measure to compute a self-similarity matrix $\mathcal{S}$ by comparing the elements of the feature sequence in a pairwise fashion, see Figure 1a.

To further enhance the path structure of $\mathcal{S}$, one typical procedure is to apply some kind of smoothing filter along the direction of the main diagonal, resulting in an emphasis of diagonal information in $\mathcal{S}$ and a denoising of other structures. In our implementation, we use a smoothing variant similar to [12], which can deal with local tempo variations. Furthermore, we apply image processing and thresholding techniques to eliminate short and weak path fragments, see also [17] for similar strategies.

The resulting path-enhanced SSM, as illustrated by Figure 1b, constitutes the input of our conversion algorithm. Note that the implementation details to obtain the path-enhanced SSM are not important at this stage. Our conversion procedure is generic and works well as long as the SSM has a relatively sparse structure only showing the most relevant paths.

[2] An implementation of these features is available at `www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/`, see also [11].

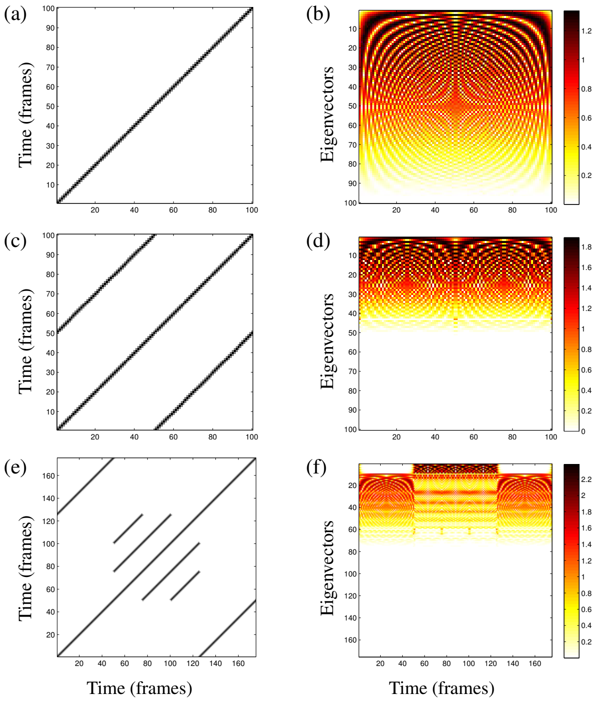

Figure 2: Various SSMs with path structures (**left**) and corresponding eigenvectors (**right**). The figures shows the eigenvectors in some transposed (row-wise) and weighted (multiplied by the corresponding eigenvalue) form. Furthermore, the eigenvectors are sorted according to decreasing eigenvalues. **(a)/(b)** SSM I_K^ϵ. **(c)/(d)** SSM reflecting the musical form A_1A_2. **(e)/(f)** SSM reflecting the musical form $A_1B_1B_2B_3A_2$.

2.2 Eigenvalue Decomposition

When comparing the elements of a feature sequence in a pairwise fashion using a symmetric similarity measure results in a symmetric SSM $\mathcal{S}$. This property may be lost by applying enhancement and image processing techniques as used in Section 2.1. However, one can restore the symmetry by considering $\frac{1}{2}(\mathcal{S}+\mathcal{S}^\top)$ instead of $\mathcal{S}$, where $\mathcal{S}^\top$ denotes the transposed matrix of $\mathcal{S}$. Doing so, we may assume in the following that $\mathcal{S}$ is a symmetric matrix of dimension $N \times N$ for some $N \in \mathbb{N}$.

Next, we apply an eigenvalue decomposition of the symmetric matrix $\mathcal{S}$ and investigate the properties of the resulting eigenvectors. Principle component analysis tells us that there exists a real-valued diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and an orthogonal matrix E such that $\mathcal{S} = EDE^\top$, where the n^{th} column e_n of E is an eigenvector of $\mathcal{S}$ with eigenvalue λ_n, i. e., $\mathcal{S}e_n = \lambda_n e_n$.

In our scenario, we assume that the matrix $\mathcal{S}$ consists of path-like structures, where a prototype of a path of length K may be modeled by the matrix [3]

$$I_K^\epsilon := \begin{pmatrix} & & & \epsilon & 1 \\ & & \epsilon & 1 & \epsilon \\ & \iddots & \iddots & \iddots & \\ \epsilon & 1 & \epsilon & & \\ 1 & \epsilon & & & \end{pmatrix} \in \{0, \epsilon, 1\}^{K \times K}.$$

[3] Opposed to existing conventions, we enumerate in this papers the rows of the matrix from bottom to top with the aim to better match the visualizations of the SSMs in the figures.

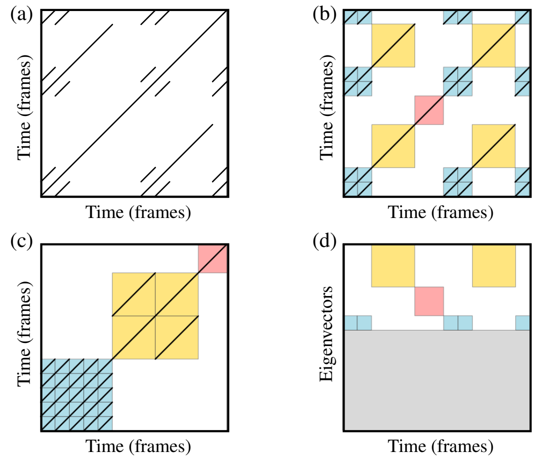

Figure 3: **(a)** SSM $\mathcal{S}$ with a path structure corresponding to the musical form $A_1A_2B_1C_1A_3A_4B_2A_5$. **(b)** SSM with highlighted block structure. **(c)** SSM in some normalized form after applying some permutation. **(d)** Illustration of the support of the suitable sorted and transposed eigenvectors.

In this matrix the non-specified entries have the value 0, see also Figure 2a. Intuitively, each path consists of a diagonal matrix with large entries on the main diagonal (represented by the value 1 in I_K^ϵ) and with decreasing entries on the diagonals above and below the main diagonal (represented by the value ϵ in I_K^ϵ). Such paths typically arise when applying path enhancement strategies based on smoothing techniques.

In [6], an explicit eigenvalue decomposition of the matrix I_K^ϵ is described. The eigenvalues are

$$\lambda_k = 1 + \epsilon \cdot 2\cos\tfrac{k\pi}{K+1}, \qquad 1 \le k \le K, \qquad (1)$$

with corresponding eigenvectors

$$e_k = \left(\sin\tfrac{k\pi}{K+1},\ \sin\tfrac{2k\pi}{K+1}\ \ldots,\ \sin\tfrac{Kk\pi}{K+1}\right)^\top. \qquad (2)$$

In particular, note that the entries of the eigenvectors are non-zero, see also Figure 2b for an illustration. This property, as we will see, becomes crucial for converting paths into blocks.

The matrix I_K^ϵ constitutes the basic building block for SSMs with more general path structures. Instead of a mathematically rigorous treatment, which is beyond the scope of this paper, we explain the general case by means of an illustrative example. Let $\mathcal{S}$ be the SSM shown in Figure 3a, which has the path structure corresponding to the musical form $A_1A_2B_1C_1A_3A_4B_2A_5$. The desired block structure is indicated in Figure 3b. By applying a suitable permutation matrix $\mathcal{P}$, it can be shown that $\mathcal{S}$ can be converted into a matrix $\mathcal{S}' := \mathcal{P}\mathcal{S}\mathcal{P}^{-1} = \mathcal{A} \oplus \mathcal{B} \oplus \mathcal{C}$, which is the direct sum of three matrices $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ corresponding to the musical parts A, B, and C (inclusive repetitions), respectively. This is illustrated by Figure 3c. As for the eigenvalue decomposition, it can be shown that the eigenvalues of $\mathcal{S}'$ are given by the eigenvalues of the matrices $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. Furthermore, the eigenvectors of $\mathcal{S}'$ are obtained by suitably extending the eigenvectors of the matrices $\mathcal{A}$,

$\mathcal{B}$, and $\mathcal{C}$ by zero entries. As a result, the supports (non-zero entries) of eigenvectors coming from different summands $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in $\mathcal{S}'$ are disjoint. This fact is illustrated by Figure 3d. Furthermore, the repetitions induce substructures in the summands $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. Each such substructure can be expressed by a suitable Kronecker product with an all-one matrix and a matrix of the form I_K^ϵ. For example, the matrix $\mathcal{B}$ corresponds to two repeating parts and can be expressed by

$$\mathbb{1}^{R \times R} \otimes I_K^\epsilon = \begin{pmatrix} I_K^\epsilon & I_K^\epsilon \\ I_K^\epsilon & I_K^\epsilon \end{pmatrix}$$

with $R = 2$ being the number of repetitions, $\mathbb{1}^{R\times R}$ being the all-one $R \times R$ matrix, and $\otimes$ being the Kronecker product. Note that the rank of $\mathcal{B}$ is K, so that in the case of $R = 2$ half of the eigenvalues are zero (therefore, the corresponding eigenvectors are not uniquely determined). Also, the permutation $\mathcal{P}$ is not known in practice. In our pipeline, we multiply the normed eigenvectors with their corresponding eigenvalue and arrange the modified eigenvectors according to their length in decreasing order. (The eigenvectors to eigenvalue 0 are irrelevant here.)

To build up some more intuition, let us consider the examples shown in Figure 2. The case of I_K^ϵ and its eigenvalue decomposition is illustrated by (a)/(b) of Figure 2, whereas the case of two repeating segments ($\mathbb{1}^{2\times 2} \otimes I_K^\epsilon$) is shown in (c)/(d). A third example corresponding to the musical form $A_1B_1B_2B_3A_2$ is shown in (e)/(f) of Figure 2. Here, the disjointness property of the supports for the eigenvalues that belong to different parts is visible. Note that the permutation matrix $\mathcal{P}$ is not known and that the eigenvectors are not ordered according to the musical parts they belong to. Furthermore, in practical applications the path structures may be noisy and distorted so that the discussed properties of the eigenvectors are not strictly fulfilled.

2.3 Deriving SSMs with Block Structures

Let $N \in \mathbb{N}$ denote the dimension of the eigenvectors, which also coincides with the number of frames. As indicated by Figure 1c, we form a matrix by defining its rows to be the transposed eigenvectors weighted and sorted by the corresponding eigenvalues. We denote this $N \times N$ matrix by $\mathcal{E}$. As discussed above, the path structure of the SSM is reflected by the support properties of the rows. In theory (assuming an ideal path structure as discussed before), two rows either have the same support (when corresponding to the same repeating musical part) or have disjoint supports (when corresponding to repeating, but different musical parts). Furthermore, the support of an eigenvector reveals all frames that belong to repeating segments of the same musical part (e. g., the frames of all A-part segments).

Motivated by this observation, we consider the columns $\mathcal{E}(n)$ of $\mathcal{E}$ as features, $n \in [1 : N]$. This yields a feature sequence $\mathcal{E}(1), \ldots, \mathcal{E}(N)$, which, in turn, can be used to define a self-similarity matrix $\mathcal{S}(\mathcal{E})$. The properties of the eigenvalues imply that two features $\mathcal{E}(i)$ and $\mathcal{E}(j)$ are similar if the frames i and j belong to repeating segments of the same musical part (or to frames of non-repeating segments), and dissimilar otherwise. As a consequence, the matrix $\mathcal{S}(\mathcal{E})$ has the desired block structure.

To make this procedure applicable for real data, we post-process the matrix $\mathcal{E}$ prior to forming the self-similarity matrix. To this end, we first replace each entry e of $\mathcal{E}$ by $\log(10|e| + 1)$ to prevent overrating the most-repeated segment. Then we apply a standard k-means clustering procedure [4] to rearrange the eigenvectors (rows of $\mathcal{E}$) so that vectors corresponding to similar structures are adjacent. Then we smooth the rearranged matrix in both directions, horizontally as well as vertically, see Figure 1d. Here, the horizontal smoothing balances out the values of the non-zero entries in the eigenvectors, whereas the vertical smoothing introduces robustness to local distortions. [5] Denoting the smoothed matrix by $\mathcal{E}'$, we compute the self-similarity matrix as above to obtain $\mathcal{S}^{\mathrm{Block}} = \mathcal{S}(\mathcal{E}')$. This matrix constitutes our final result, see Figure 1e.

3. EXPERIMENTS

To show how our conversion approach behaves on real data, we now discuss a number of explicit examples (Section 3.2) and report on some quantitative experiments (Section 3.3). Note that optimizing and investigating the specific role of the various parameters is not in the scope of this paper. Rather than numerically improving a specific structure analysis result, our main goal is to highlight the conceptual novelty of our approach. In particular, we demonstrate that procedures that are designed for homogeneity-based structure analysis (as the one described in Section 3.1) can now be applied for repetition-based structure analysis thanks to our conversion procedure.

3.1 Structure Analysis Procedure

As a typical example approach, we consider the homogeneity-based structure analysis procedure as described in [5], where a given self-similarity matrix is decomposed into a prototype matrix and an activation matrix using non-negative matrix factorization (NMF). Looking at maximizing entries in the activation matrix yields a frame-wise classification of the columns of the SSM, which in turn can be used to assign a class label to each frame. A segment is then defined as a maximal run of consecutive frames having the same class label, see [5] for more details and Figure 1f for an example.

In our experiments, we used an NMF-variant with additional sparseness constraints [7] setting the sparseness parameter to $4 \cdot \mathrm{mean}(\mathcal{S}^{\mathrm{Block}})$ and the rank parameter to 6 (assuming at most six different musical parts). The procedure was then applied to the matrix $\mathcal{S}^{\mathrm{Block}}$.

[4] In our implementation, we used 6 clusters. Our experiments showed that any number between 5 and 20 led to similar results.

[5] In our experiments, we used Gaussian smoothing using an adaptive window size vertically and 7 frames horizontally. Again these values are not crucial here.

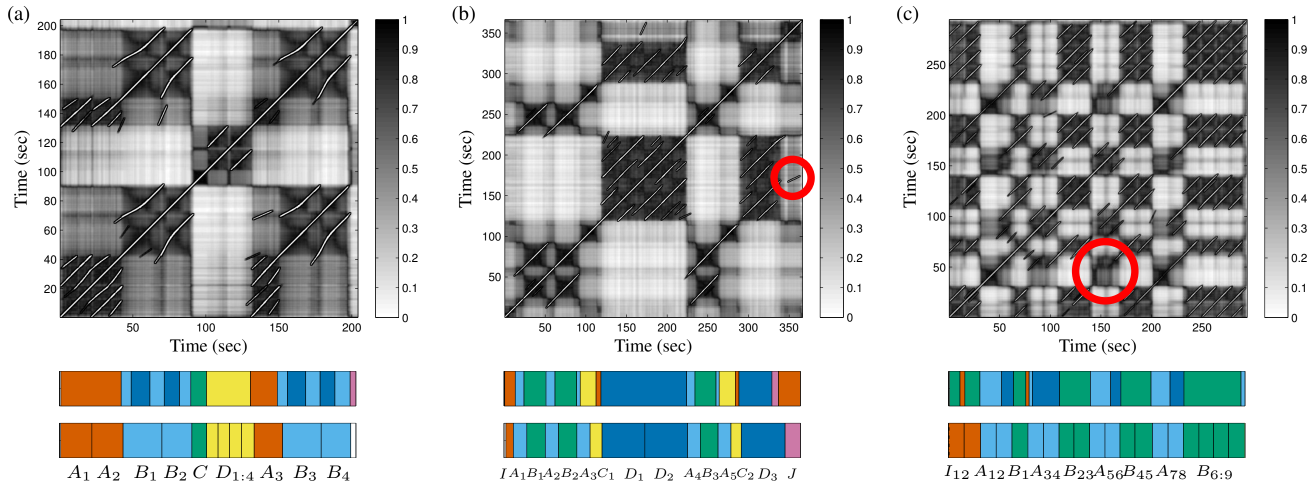

Figure 4: Results for three different audio recordings. The figure shows the computed block matrix overlaid with the path structure of the input matrix (top), the computed structure analysis results (middle) and the manually generated structure annotations (bottom). **(a)** Hungarian Dance No. 5 by Johannes Brahms. **(b)** March No. 1 from Op. 39 (Pomp and Circumstance) by Edward Elgar. **(c)** The song "The winner takes it all" by ABBA.

3.2 Qualitative Evaluation

We now discuss some specific examples to show the potential and the limitations of our conversion procedure. We start with our running example shown in Figure 1, which is a recording of the Waltz No. 2 from the Suite for Variety Orchestra by Dmitri Shostakovich. Using the path matrix $\mathcal{S}$ as shown in Figure 1b as input, our conversion procedure outputs the block matrix $\mathcal{S}^{\text{Block}}$ shown in Figure 1e. As the figure illustrates, path structures of repeating segments have been correctly converted into blocks. For example, the four repetitions of the combined AB-part are clearly visible as path structure in Figure 1b and as block structure in Figure 1e. Furthermore, Figure 1f shows the computed structure annotation obtained from $\mathcal{S}^{\text{Block}}$, whereas Figure 1g shows a manually generated structure annotation. Indeed, the homogeneity-based clustering approach applied to $\mathcal{S}^{\text{Block}}$ produced a reasonable repetition-based structure analysis result. Only subsequent repeating parts such as the two D-parts $D_1 D_2$ (which are clearly reflected by paths in Figure 1b) have not been resolved by our frame-based labeling approach. Also, note that, because of significant musical variations in harmony and melody, the two repeating C-parts are neither reflected by paths nor by blocks.

Next, we consider the three examples of Figure 4. For each example, the computed block matrix $\mathcal{S}^{\text{Block}}$ overlaid with the original path structure inputted to our conversion procedure is shown. Also the structure annotations obtained from $\mathcal{S}^{\text{Block}}$ as well as the manually generated "ground truth" annotations (for comparison) are shown. The first example shown in Figure 4a is a recording of the Hungarian Dance No. 5 by Johannes Brahms. The A-part as well as the B-part segments are well reflected in the block structure despite of some distortions and inconsistencies in the path structure. Also, tempo differences between B-part segments (B_2 and B_4 are played faster than B_1 and B_3) still led to meaningful block structures. As in the previous Shostakovich example, the subsequent repeating A-part and D-part segments were not subdivided as a result of the purely frame-based labeling procedure. Also note that even in the path representation only the repetitions $D_1 D_2$ and $D_3 D_4$ were captured, but not the finer grained subdivision (because of the chosen temporal resolution induced by the parameter setting). Finally, the B-part segments were further subdivided by our structure analysis procedure, illustrating an over-segmentation as typical for automated structure analysis methods [9].

The example shown in Figure 4b is based on a recording of the March No. 1 from Op. 39 by Edward Elgar. As before, one can say that overall the computed block structure correspond well to the inputted path structure. The erroneously extracted small path fragment indicated by the red circle has no major influence on the computed block structure as well as on the final structure. This indicates that our conversion procedure is, at least to some degree, robust to local distortions and noise. In general, our procedure tends to yield better results when the inputted path structure is sparse, thus requiring a denoising/smoothing and thresholding step to enhance the path structure as is also done in most repetition-based structure analysis approaches [1, 14].

The final example is a recording of the song "The winner takes it all" by ABBA, see Figure 4c. With this example, we want to indicate that missing path relations as marked by the red circle may be "recovered" in the block structure. Since the eigenvalue decomposition is a *global* analysis of the entire matrix $\mathcal{S}$, local deviations and missing relations are balanced out, thus enforcing some kind of transitivity on the block level.

3.3 Quantitative Evaluation

Finally, we quantitatively evaluated and compared our overall structure analysis procedure based on two well-known datasets. First, we used the Beatles dataset with

Dataset	Method	pairwise			boundary ($3s$)		
		F [%]	P [%]	R [%]	F [%]	P [%]	R [%]
BeatlesTUT	**proposed**	68.0	71.4	68.8	61.4	58.0	69.5
	[5]	60.8	61.5	64.6	N/A	N/A	N/A
	[13]	59.9	72.9	54.6	N/A	N/A	N/A
	SMGA (worst)	65.8	70.9	65.9	69.6	68.1	72.9
	SMGA (best)	71.8	65.1	80.0	75.3	73.4	79.1
Mazurka49-Rub	**proposed**	72.3	70.1	78.7	60.6	66.3	60.5
Mazurka49-Coh	**proposed**	70.0	69.3	74.2	62.7	65.3	65.9
Mazurka49-Eza	**proposed**	71.4	69.0	77.4	64.4	70.7	64.1
Mazurka2792	SMGA (worst)	68.1	75.2	65.2	65.9	70.3	65.3
	SMGA (best)	71.9	75.8	71.6	69.2	72.4	69.5

Table 1: Evaluation results for various procedures, evaluation measures, and datasets, see text for a detailed explanation.

the TUT annotations [6] described in [13]. Second, we used three complete recordings (Rubinstein 1966, Cohen, Ezaki) taken from the 2792 recordings of the Mazurka dataset [7] with manually generated structure annotations. Using standard precision (P), recall (R) and F-measure (F) for labeled pairs of frames as well as for segment boundaries (with 3 seconds tolerance), we compared our approach to [5, 13] as well as to the best performing MIREX2012 method [8] denoted by SMGA which is based on an extension of [17]. For SMGA, the results are reported for two different parameter settings corresponding to best and the worst performing setting, respectively.

Table 1 shows the results. Note that we have applied a similar NMF-based structuring algorithm as in [5], however applied to our converted matrix $\mathcal{S}^{\mathrm{Block}}$. This leads to substantial improvements compared to [5] on the Beatles dataset considering pairwise P/R/F-values. Also compared to the SMGA results, we are at least in the same range. As for the segment boundaries, however, we are worse. This is by no surprise since our approach is a purely frame-based procedure, whereas SMGA is based on a segment boundary detection step. Similar results hold for the Mazurka dataset, where SMGA has been evaluated on all 2792 recordings, which include the three versions used in our experiments. Our procedure yields for all three pianists pairwise P/R/F-values that are in the same range as the ones reported for SMGA. Again we want to emphasize that the quantitative results are not in the focus of this paper, but should only indicate the overall behavior of our conversion procedure.

4. CONCLUSIONS

In this paper, we introduced a novel method for converting SSMs with path structures into SSMs with block structure based on eigenvalue decompositions. As main technical contribution, we discussed how certain path structures translate into characteristic properties of the eigenvectors. Furthermore, as an application of our conversion, we showed how a homogeneity-based structure analysis procedure can be applied to the converted path matrix to facilitate repetition-based structure analysis. We hope that our contribution is interesting not only from a conceptual point of view, but may also open up novel ways for fusing different segmentation principles at an early stage of a structure processing pipeline. In particular, it seems promising to directly combine block-like SSMs (reflecting homogeneous musical properties) with converted path-like SSMs (reflecting repetitive musical properties), which can then be handled using the same algorithmic pipeline.

Acknowledgments: This work has been supported by the German Research Foundation (DFG CL 64/8-1, DFG MU 2682/5-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS.

[6] http://www.cs.tut.fi/sgn/arg/paulus/structure.html

[7] http://www.mazurka.org.uk

[8] http://nema.lis.illinois.edu/nema_out/mirex2012/results/struct/mrx09/

5. REFERENCES

[1] Roger B. Dannenberg and Masataka Goto. Music structure analysis from acoustic signals. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, volume 1, pages 305–331. Springer, New York, NY, USA, 2008.

[2] Shlomo Dubnov and Ted Apel. Audio segmentation by singular value clustering. Proc. ICMC, 2004.

[3] Emilia Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, UPF Barcelona, 2006.

[4] Florian Kaiser, Marina Georgia Arvanitidou, and Thomas Sikora. Audio similarity matrices enhancement in an image processing framework. Proc. CBMI, Madrid, Spain, 2011.

[5] Florian Kaiser and Thomas Sikora. Music structure discovery in popular music using non-negative matrix factorization. Proc. ISMIR, pages 429–434, Utrecht, The Netherlands, 2010.

[6] Jerry L. Kazdan. A Tridiagonal Matrix, http://hans.math.upenn.edu/~kazdan/AMCS602/tridiag-short.pdf, Retrieved 11.03.2013.

[7] Jingu Kim and Haesun Park. Toward faster nonnegative matrix factorization: A new algorithm and comparisons. Proc. ICDM, pages 353–362, Pisa, Italy, 2008.

[8] Mark Levy and Mark Sandler. Structural segmentation of musical audio by constrained clustering. IEEE TASLP, 16(2):318–326, 2008.

[9] Hanna Lukashevich. Towards quantitative measures of evaluating song segmentation. Proc. ISMIR, pages 375–380, Philadelphia, USA, 2008.

[10] Meinard Müller and Sebastian Ewert. Towards timbre-invariant audio features for harmony-based music. IEEE TASLP, 18(3):649–662, 2010.

[11] Meinard Müller and Sebastian Ewert. Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. Proc. ISMIR, pages 215–220, Miami, FL, USA, 2011.

[12] Meinard Müller and Frank Kurth. Towards structural analysis of audio recordings in the presence of musical variations. *EURASIP Journal on Advances in Signal Processing*, 2007(1), 2007.

[13] Jouni Paulus and Anssi P. Klapuri. Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. IEEE TASLP, 17(6):1159–1170, 2009.

[14] Jouni Paulus, Meinard Müller, and Anssi P. Klapuri. Audio-based music structure analysis. Proc. ISMIR, pages 625–636, Utrecht, The Netherlands, 2010.

[15] Geoffroy Peeters. Deriving musical structure from signal analysis for music audio summary generation: “sequence” and “state” approach. Proc. CMMR, Vol. 2771, LNCS, pages 143–166. Springer, 2004.

[16] Geoffroy Peeters. Music structure dicovery: Measurig the “stateness” of times. ISMIR: Late Breaking Session, 2011.

[17] Joan Serrà, Meinard Müller, Peter Grosche, and Josep Lluis Arcos. Unsupervised detection of music boundaries by time series structure features. Proc. AAAI International Conference on Artificial Intelligence, Toronto, Canada, 2012.

[18] Jordan Bennett Louis Smith, John Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. Design and creation of a large-scale database of structural annotations. Proc. ISMIR, pages 555–560, Miami, FL, USA, 2011.