IDENTIFYING REPEATED PATTERNS IN MUSIC USING SPARSE CONVOLUTIVE NON-NEGATIVE MATRIX FACTORIZATION

Ron J. Weiss and Juan Pablo Bello
Music and Audio Research Lab (MARL), New York University
{ronw, jpbello}@nyu.edu

ABSTRACT

We describe an unsupervised, data-driven, method for automatically identifying repeated patterns in music by analyzing a feature matrix using a variant of sparse convolutive non-negative matrix factorization. We utilize sparsity constraints to automatically identify the number of patterns and their lengths, parameters that would normally need to be fixed in advance. The proposed analysis is applied to beat-synchronous chromagrams in order to concurrently extract repeated harmonic motifs and their locations within a song. Finally, we show how this analysis can be used for long-term structure segmentation, resulting in an algorithm that is competitive with other state-of-the-art segmentation algorithms based on hidden Markov models and self similarity matrices.

1. INTRODUCTION

Repetition has been widely-recognized to be a ubiquitous feature of music, closely related to structural units in music, such as beats, bars, motives and sections [10]. This applies both to popular music, often composed of nearly exact repetitions of a small number of sections, e.g. verse, chorus, and bridge; and to more sophisticated genres, e.g. jazz or orchestral music, where recurrences are often masked by complex transformations, including key modulations and tempo variations. The analysis of repeated patterns and their temporal organization is central to the understanding of music. However, while repetitions are apparent in symbolic representations of music, their extraction from musical audio poses a number of challenges stemming from factors such as the presence of background noise, the influence of multiple instruments and sonic textures, timing variations and other attributes of musical expression, etc.

The automatic analysis of repetition in music audio has been an important focus of attention in MIR, with applications including thumbnailing [1], retrieval [2], and, notably, long-term segmentation using methods such as self-similarity matrices and hidden Markov models [11, 8, 5]. However, with a few exceptions [7, 1], the emphasis has been on locating repetitions rather than on extracting of characteristic, repetitive patterns. Previous research on detecting motif occurrences across a collection [9] and cover-song retrieval based on short-snippets [3], illustrate the utility of extracting such patterns.

In this paper we propose a novel approach for the automatic extraction and localization of repeated patterns in music audio. The approach is based on sparse shift-invariant probabilistic latent component analysis [14] (SI-PLCA), a probabilistic variant of convolutive non-negative matrix factorization (NMF). The algorithm treats a musical recording as a concatenation of a small subset of short, repeated patterns, and is able to simultaneously estimate both the patterns and their repetitions throughout the song. The analysis naturally identifies the long-term harmonic structure within a song, while the short-term structure is encoded within the patterns themselves. Furthermore, we show how it is possible to utilize sparse prior distributions to learn the number of patterns and their respective lengths, minimizing the number of parameters that must be specified exactly in advance. Finally, we explore the application of this approach to long-term segmentation of musical pieces.

The remainder of this paper is organized as follows: Section 2 reviews the proposed analysis based on SI-PLCA and describes its relationship to NMF. Sections 3 and 4 describe prior distributions over the SI-PLCA parameters and the expectation maximization algorithm for parameter estimation. Sections 5 and 6 discuss how the proposed analysis can be used for structure segmentation and provide experimental results. Finally, we conclude in Section 7.

2. PROPOSED APPROACH

2.1 From NMF to PLCA

Conventional NMF decomposes a non-negative matrix V into the product of two non-negative matrices W and H:

$$V \approx WH \tag{1}$$

In the context of audio analysis, if V represents a time-frequency decomposition of an audio signal, each column of W can be thought of as a frequency template used repeatedly throughout V, and each row of H can be thought of as the activations of the corresponding basis in time. In this paper we focus on the analysis of beat-synchronous chromagrams [4], but the method is equally applicable to any non-negative time-frequency representation such as a magnitude spectrogram.

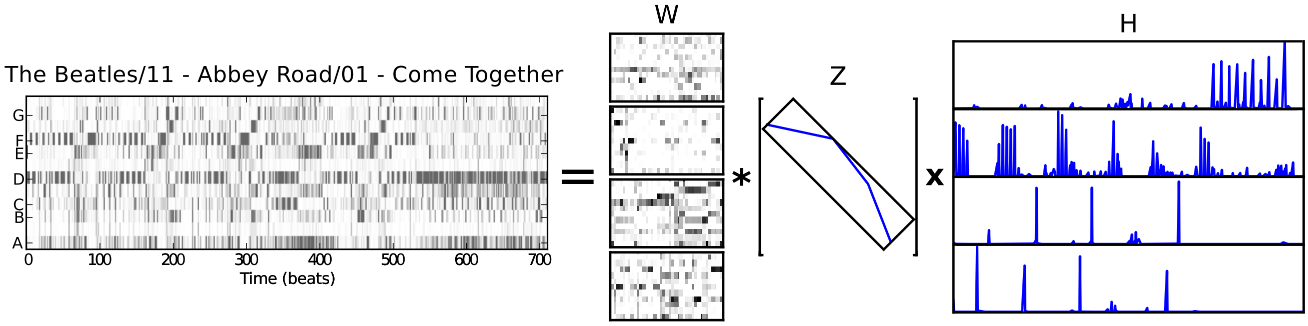

Figure 1. Demonstration of the SI-PLCA analysis of a chromagram. The decomposition was initialized with $L = 40$, and $K = 10$ with $\alpha_z = 0.98$, and no sparsity on W_k or $\mathbf{h}_k^T$. The parameter estimation algorithm pruned out most of the initial bases due to the sparse prior on $\mathbf{z}$, converging on only 4 bases.

Probabilistic Latent Component Analysis (PLCA) [14] recasts this analysis in a probabilistic framework. PLCA represents each column of W and each row of H as multinomial probability distributions and adds an additional distribution over each basis, i.e. a mixing weight. The decomposition can be rewritten in NMF terms as follows:

$$V \approx WZH = \sum_{k=0}^{K-1} \mathbf{w}_k \, z_k \, \mathbf{h}_k^T \qquad (2)$$

where $Z = \mathrm{diag}(\mathbf{z})$ is a diagonal matrix of mixing weights $\mathbf{z}$ and K is the rank of the decomposition (i.e. the number of bases in W). Contrary to standard NMF, each of V, $\mathbf{w}_k$, $\mathbf{z}$, and $\mathbf{h}_k^T$ are normalized to sum to 1 since they correspond to probability distributions.

The probabilistic foundation makes for a convenient framework for imposing constraints on the parameters $\mathbf{w}_k$, $\mathbf{h}_k^T$, and $\mathbf{z}$ through the use of prior distributions. This will be discussed in detail in Section 3.

2.2 Adding shift-invariance

A shift-invariant extension to the PLCA model which allows for *convolutive* bases is described in [14]. Unlike the single frame bases $\mathbf{w}_k$ described in Section 2.1, each SI-PLCA basis is expanded to form a fixed duration template W_k containing L frames. Therefore, the $F \times K$ matrix W becomes an $F \times L \times K$ tensor $\mathcal{W}$, and the normalized basis $\mathbf{w}_k$ becomes a normalized matrix W_k. The factors $\mathcal{W}$ and H are combined via a convolution operation instead of matrix multiplication in a process analogous to equation (2):

$$V \approx \sum_k W_k * z_k \mathbf{h}_k^T \qquad (3)$$

Figure 1 shows an example SI-PLCA decomposition of a chromagram using $K = 4$ basis patterns of length $L = 40$.

3. SPARSE PRIOR DISTRIBUTIONS

A common strategy used throughout the NMF literature is to favor sparse settings, i.e. one containing many zeros, for W or H in order to learn parsimonious, parts-based decompositions of the data. Sparse solutions can be encouraged when estimating the parameters in equation (3) by imposing constraints using an appropriate prior distribution. In the following sections we describe how this process can be used to automatically learn the number and length of the repeated patterns within a song.

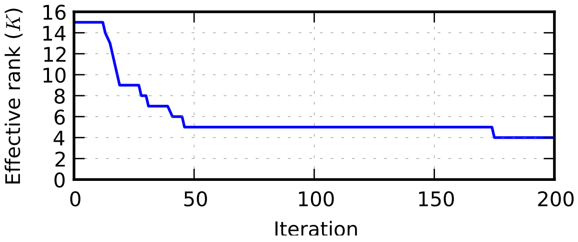

Figure 2. Typical behavior of the automatic relevance determination effect of a sparse prior on $\mathbf{z}$. The initial rank of the decomposition is set to $K = 15$, and as the estimation algorithm iterates it is pruned down to a final effective rank (the number of bases with non-zero z_k) of 4.

3.1 Learning the number of patterns K

The Dirichlet distribution is conjugate to the multinomial distributions W_k, $\mathbf{z}$, and $\mathbf{h}_k^T$, making it a natural choice for a prior. The Dirichlet prior on $\mathbf{z}$ has the following form:

$$P(\mathbf{z} \,|\, \alpha_z) \propto \prod_k z_k^{\alpha_z - 1}, \quad \alpha_z \geq 0 \qquad (4)$$

where the hyperparameter α_z is fixed across all K components. If $\alpha_z < 1$ this prior favors solutions where many components are zero, i.e. where the distributions are sparse.

If $\mathbf{z}$ is forced to be sparse, the learning algorithm will attempt to use as few bases as possible. This enables an automatic relevance determination strategy in which: (a) the algorithm is initialized to use many bases (large K), and (b) the sparse prior on $\mathbf{z}$ prunes out bases that do not contribute significantly to the reconstruction of V. Only the most relevant patterns “survive” to the end of the parameter estimation process, as is shown in the example in Figure 2. This approach is useful because it removes the need to specify the exact rank of the decomposition K in advance. The parameter estimation simply learns the underlying number of patterns needed by the data. A similar approach to automatically determining the rank of a standard NMF decomposition is described in [15].

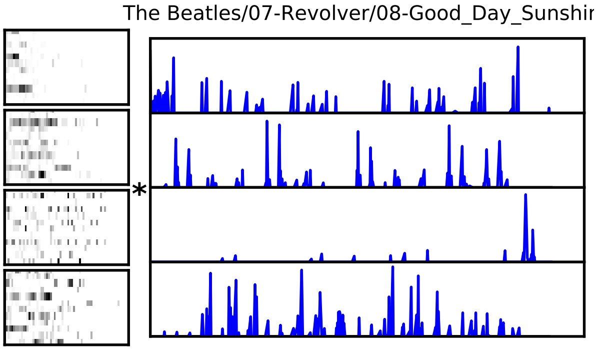

Figure 3. Demonstration of the SI-PLCA decomposition of a chromagram using $L = 60$ and sparsity in all parameters ($\alpha_z = 0.98$, $c = 16$, $m = -10^{-8}$, and $\alpha_h = 1 - 10^{-5}$).

3.2 Learning the pattern length L

The other parameter that must be specified in advance is the length L of the convolutive bases. In fact, different patterns within the same piece often have different intrinsic lengths, e.g. if the chorus uses a shorter riff than the verse or if the time signature changes. Therefore it is useful to automatically identify the length of each basis independently instead of using a fixed length across all bases.

We employ a similar strategy to that described in Section 3.1 by setting L to an upper bound on the expected pattern length and constructing a prior distribution that encourages the use of shorter bases. This is accomplished by using a Dirichlet prior on W_k with a parameter that depends on the time position τ within each basis:

$$P(W_k \,|\, \boldsymbol{\alpha}_w) \propto \prod_\tau \prod_f w_{kf\tau}^{\alpha_{w\tau}-1} \tag{5}$$

$\boldsymbol{\alpha}_w$ is constructed as a piecewise function which is uninformative for small τ and then becomes increasingly sparse:

$$\alpha_{w\tau} = \begin{cases} 1, & \tau < c \\ 1 + m\,(\tau - c), & \tau \geq c \end{cases} \tag{6}$$

This prior only effects patterns longer than c frames with a penalty that increases with the pattern length.

An example of the effect of this prior is shown in Figure 3. Most of the information in the top basis is contained within the first 12 columns, while the other bases have effective lengths between 30 and 40.

3.3 Basis/activation trade-off

It is often worthwhile to enforce sparsity on $\mathbf{h}_k^T$ using a similar approach to equation (4), with a single parameter α_h tied across all points within $\mathbf{h}_k^T$. The rationale is that if most of the activations in $\mathbf{h}_k^T$ are zero, then more of the information in V will be captured by W_k, and vice versa. A sparse $\mathbf{h}_k^T$ promotes more parsimonious patterns for W_k, at the cost of a reduced time resolution.

This is illustrated by the example in Figure 1. The second basis pattern is relatively sparse, while the corresponding row of H contains many non-zero entries. In fact, the spacing between adjacent activations in $\mathbf{h}_1^T$ is smaller than the length of the pattern; i.e. it is continually mixed with delayed versions of itself. The pattern repeats about every 8 beats, roughly corresponding to the underlying meter.

In contrast, the bottom two bases contain significantly more information while the corresponding rows of H contain only about 4 peaks. The sparsity setting α_h, in combination with $\alpha_{w\tau}$, control the trade-off between these qualitatively different solutions. A sparse H leads to more musically meaningful bases that are exactly repeated throughout the piece, while a sparse W leads to temporal patterns in H that are organized according to to the underlying rhythm.

4. PARAMETER ESTIMATION

The decomposition of equation (3) can be computed iteratively using an expectation maximization (EM) algorithm. The full derivation of the algorithm can be found in [13]. Here we extend it to incorporate the prior distributions described in Section 3. Since we are using conjugate prior distributions, this extension is straightforward to derive.

In the E-step, the posterior distribution over the hidden variables k and τ is computed for each cell in V. For notational convenience we represent this distribution as a set of matrices $\{R_{k\tau}\}$ for each setting of k and τ. Each point in the $F \times T$ matrix $R_{k\tau}$ corresponds to the probability that the corresponding point in V was generated by basis k at time delay τ. It can be computed as follows:

$$R_{k\tau} \propto \mathbf{w}_{k\tau} \otimes z_k \overset{\rightarrow\tau}{\mathbf{h}_k^T} \tag{7}$$

where $\otimes$ denotes the outer product, and $\overset{\rightarrow t}{x}$ shifts x t places to the right. The set of $R_{k\tau}$ matrices are normalized such that each point in $\sum_{k\tau} R_{k\tau}$ is one.

Given this posterior distribution, the parameters can be updated in the M-step as follows:

$$z_k \propto \sum_\tau \sum_{ft} V \cdot R_{k\tau} + \alpha_z - 1 \tag{8}$$

$$\mathbf{w}_{k\tau} \propto \sum_t V \cdot R_{k\tau} + \alpha_{w\tau} - 1 \tag{9}$$

$$\mathbf{h}_k^T \propto \sum_\tau \sum_f \overset{\leftarrow\tau}{V} \cdot \overset{\leftarrow\tau}{R}_{k\tau} + \alpha_h - 1 \tag{10}$$

where $\cdot$ denotes the element-wise matrix product and the parameters are normalized so that $\mathbf{z}$, W_k, and $\mathbf{h}_k^T$ sum to 1.

The overall EM algorithm proceeds by initializing W_k, $\mathbf{z}$, and $\mathbf{h}_k^T$ randomly, and then iterating equations (7) to (10) until convergence. This algorithm is only guaranteed to converge to a local optimum, so the quality of the factorization is somewhat dependent on initialization. In our experiments we found that initializing $\mathbf{z}$ and $\mathbf{h}_k^T$ uniformly while setting the initial W_k randomly leads to more consistent results.

5. STRUCTURE SEGMENTATION

As mentioned in the introduction, the analysis described in this paper can be applied to the task of music structure segmentation. It naturally identifies the long-term temporal

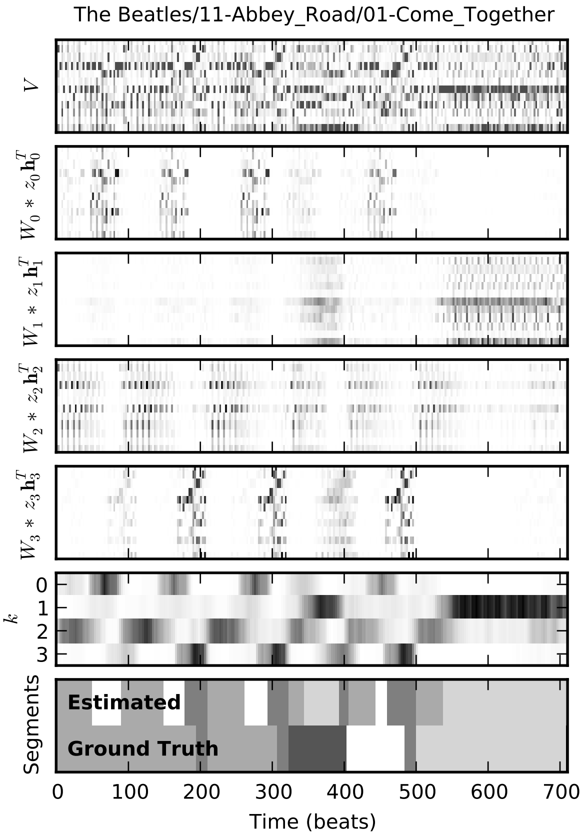

Figure 4. Song structure segmentation using the SI-PLCA decomposition shown in Figure 1. The pairwise F-measure of the estimated segmentation is 0.52.

structure within a song, encoded by H. At the same time, the short-term structure is captured within the bases $\mathcal{W}$.

We use the beat-synchronous chroma feature extraction from [4]. Each frame of V is normalized so that the maximum energy is one. Analysis of these features identifies repeated motifs in the form of chord patterns. We assume a one-to-one mapping between these chord patterns and the underlying song structure, i.e. we assume that each pattern is used within only one segment. The mapping is derived by computing the contribution of each pattern to the chroma gram by summing equation (3) across all pitch classes:

$$\ell_k(t) = \sum_f W_k * z_k \mathbf{h}_k^T \tag{11}$$

The segmentation labels are then found by smoothing the K "pattern usage" functions $\ell_k(t)$ using a rectangular window, and finding the most active pattern at each frame:

$$\ell(t) = \operatorname*{argmax}_k \; \ell_k(t) * \mathbf{1}_S \tag{12}$$

where $\mathbf{1}_S$ is a length S vector of ones. Finally, the per-frame segment labels $\ell(t)$ are post-processed to remove segments shorter than a given minimum segment length.

5.1 Examples

An example of this segmentation procedure is shown in Figure 4. The top panel shows the original chromagram of the song. The following four panels show the contribution of each pattern to the chromagram, and the bottom two panels show the smoothed $\ell_k(t)$ and the final segmentation.

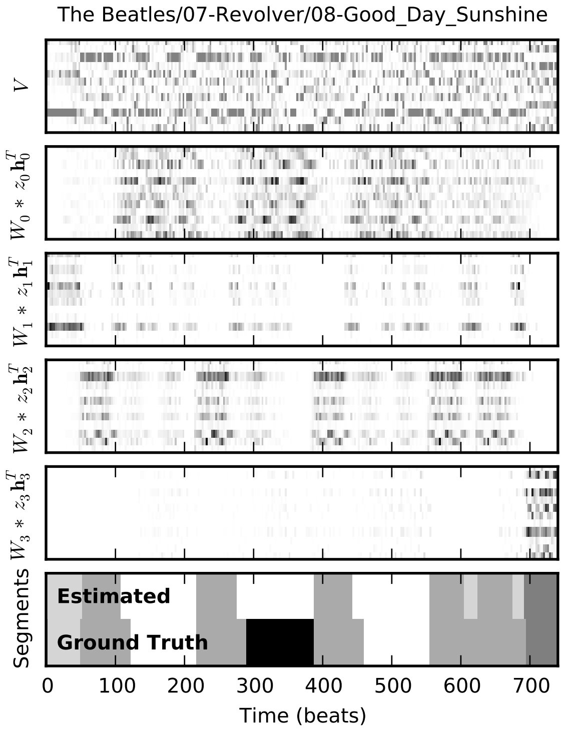

Figure 5. Song structure segmentation using the SI-PLCA decomposition shown in Figure 3 (PFM = 0.69).

There are some interesting differences between the ground truth segmentation and that derived from the proposed algorithm in Figure 4. For example, the proposed algorithm breaks the beginning of the song into repeated subsections: basis 2 (mid-gray) → basis 0 (white), while the ground truth labels this sequence as a single segment. When inspecting the actual patterns it is clear that these segments are composed of distinct chord patterns, despite serving a single musical role together ("intro/verse" as annotated in the ground truth). In fact the mid-gray and white segments are reused in different contexts throughout the song in regions with different ground-truth annotations. The analysis has no notion of musical role, so it tends to converge on solutions in which bases are reused as often as possible.

One way to address this limitation is to increase the length L of the convolutive bases (or the corresponding parameters of $\alpha_{w\tau}$), in which case the repeated sub-segments would be merged into a single long segment. This highlights an inherent trade-off in the proposed analysis between identifying simple chord patterns that are frequently repeated (short W_k, many activations in $\mathbf{h}_k^T$) as opposed to deriving long-term musical structure (longer W_k, sparser $\mathbf{h}_k^T$). This trade-off is a recognized ambiguity in the concept of musical segmentation [12].

When high-level segments are more closely correlated with the harmonic structure identified by our method, the proposed analysis leads to good segmentation. An example of this is shown in Figure 5. Note that the ground truth labels make a distinction between "verse"(white) and "verse/break" (black) which is not present in our analysis.

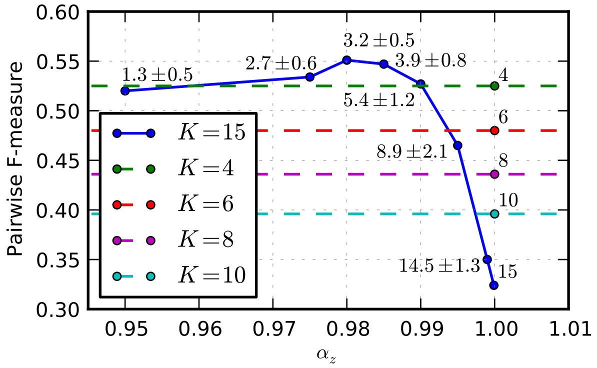

Figure 6. PFM as a function of α_z (solid line). $K = 15$, $L = 60$, and no other priors are used. The average effective rank for each setting of α_z is displayed. Also plotted is PFM for $\alpha_z = 1$ for different settings of K (dashed lines).

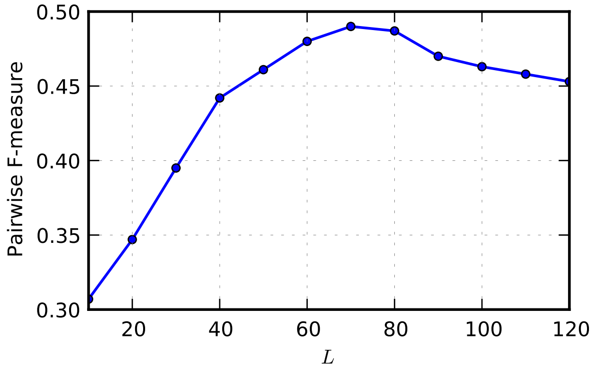

Figure 7. PFM as a function of the pattern length L. The rank is fixed at $K = 6$ and no sparse priors are used.

6. EXPERIMENTS

In this section we evaluate the proposed approach to structure segmentation. We quantify the effect of the various prior distributions described in Section 3 and compare our approach to other state-of-the-art algorithms. The test set consists of 180 songs from the recorded catalog of The Beatles, annotated into verse, chorus, refrain, etc. sections by the Centre for Digital Music. [1] Each song contains an average of about 10 segments and 5.6 unique labels.

Segmentation performance is measured using the pairwise recall rate (PRR), precision rate (PPR), and F-measure (PFM) metrics proposed in [5] which measure the frame-wise agreement between the ground truth and estimated segmentation regardless of the exact segment label. We also report the entropy-based over- and under-segmentation scores (S_o and S_u, respectively) as proposed in [6].

6.1 Number of patterns

Since our segmentation algorithm assumes a one-to-one relationship between patterns and segments, the appropriate choice of the number of patterns K is critical to obtaining good performance. We evaluate this effect by segmenting the data set with varying settings for K with $\alpha_z = 1$, and by fixing K to 15 and varying α_z. No smoothing of the resulting labels is performed ($S = 1$).

The results are shown in Figure 6. For $\alpha_z = 1$, segmentation performance decreases as K increases, peaking at $K = 4$. Performance improves when the sparse prior is applied for most settings of α_z. The average effective rank and its standard deviation both increase with decreasing α_z (increasing sparsity). The best performance is obtained for $\alpha_z = 0.98$, leading to an average effective rank of 3.2 ± 0.5. These results demonstrate the advantage of allowing the number of patterns to adapt to each song.

6.2 Pattern length

As described in Section 5.1, the length of the patterns used in the decomposition has a large qualitative effect on the segmentation. To measure this effect, we segmented the entire corpus varying L between 10 and 120 beats. No sparsity was enforced, so the pattern length remained fixed for all bases and all songs. The results are shown in Figure 7.

As predicted, segmentation performance is poor for small L since the ground truth segments are often divided into many distinct short segments. Performance improves with increasing L, until it reaches a peak at $L = 70$. When L grows larger than the average segment length in the ground truth (78 beats) the performance decreases.

Enforcing sparsity on W_k and varying c leads to similar results. However, we have found that allowing for varying pattern length has negligible effect on segmentation performance, despite often resulting in qualitatively better patterns. Following this trend, we have also found that $\alpha_h \neq 1$ has minimal effect on performance, so it is set to 1 in the remaining experiments. These results are not surprising since the segmentation is derived from the *combination* of W and H. Shifting the sparsity from one factor to another should not have significant impact on $\ell_k(t)$.

6.3 Comparison to the state-of-the-art

We compare the proposed segmentation system with other state-of-the-art approaches, including Levy and Sandler's HMM-based segmentation system [2] [5] (QMUL) and a more recent system from Mauch et al [8] based on analysis of self-similarity matrices derived from beat-synchronous chroma. As in Section 6.1, we found that QMUL has optimal PFM when the number of segments is set to 4.

We compare these to the proposed system using fixed rank $K = 4$ (SI-PLCA) and a variant using sparse $\mathbf{z}$ with $\alpha_z = 0.995$ and $K = 15$ (SI-PLCA-α_z). L was fixed at 70 for both systems, and the minimum segment length S was set to 32. Also included is a baseline random segmentation where each frame is given one of 4 randomly selected labels.

The results are shown in Table 1. The system from Mauch et al performs best, followed by SI-PLCA-α_z, SI-PLCA, and QMUL. All systems perform significantly better than the baseline. All of the segmentation systems have roughly comparable pairwise precision and S_u. The differences are primarily in the recall (and S_o) with Mauch et al

[1] http://isophonics.net/content/reference-annotations-beatles

[2] Available: http://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html

System	PFM	PPR	PRR	S_o	S_u
Mauch et al [8]	0.66	0.61	0.77	0.76	0.64
SI-PLCA-α_Z	0.60	0.58	0.68	0.61	0.56
SI-PLCA	0.58	0.60	0.59	0.56	0.59
QMUL [5]	0.54	0.58	0.53	0.50	0.57
Random	0.30	0.36	0.26	0.07	0.24

Table 1. Segmentation performance on the Beatles data set. The number of labels per song was fixed to 4 for SI-PLCA, QMUL, and Random. The average effective ranks for SI-PLCA-α_z and Mauch et al were 3.9 and 5.5, respectively.

outperforming SI-PLCA-α_z by 12% (15%), and SI-PLCA-α_z in turn outperforming QMUL by 15% (11%).

Aside from our algorithm's tendency to over-segment, the most obvious qualitative difference between Mauch et al's and the proposed system lies in more accurate boundary detection in the former system. This is partially a result of the smoothing performed in equation (12) which tends to blur out the segmentation. A more sophisticated set of heuristics for deriving segment labels from the SI-PLCA decomposition might not suffer from this problem.

7. CONCLUSION

We have described an algorithm for identifying repeated patterns in music using shift-invariant probabilistic component analysis and shown how it can be applied to music segmentation. The source code is freely available online. [3]

We demonstrate that the use of simple sparse prior distributions on the SI-PLCA parameters can be used to automatically identify the bases that are most relevant for modeling the data and discard those whose contribution is small. We also demonstrate a similar approach to estimating the optimal length of each basis. The use of these prior distributions enables a more flexible analysis and eliminates the need to specify these parameters exactly in advance.

Although this paper has focused on structure segmentation, the proposed analysis has many other potential applications. For example, basis patterns could be extracted from a collection of pieces to search for common motifs used throughout a corpus of music, e.g. retrieval of cover songs or musical variations. Similarly, Mauch et al demonstrate that chord recognition performance can be improved by pooling data from repeated sections to smooth over variations [8]. In the context of the proposed analysis this amounts to simply analyzing the bases W_k.

Other potential future work includes extracting the hierarchical structure within a piece by repeating the SI-PLCA analysis at different time scales. Finally, we mention that it is possible to extend the SI-PLCA decomposition to be key-invariant by using the 2D extension to SI-PLCA which allow for shifts in pitch class/frequency as well as time [14]. Such an extension would allow for structure segmentation that is insensitive to key modulations within a piece.

8. ACKNOWLEDGEMENTS

The authors would like to thank Matthias Mauch for sharing the implementation of the algorithm from [8]. This material is based upon work supported by the NSF (grant IIS-0844654) and by the IMLS (grant LG-06-08-0073-08).

9. REFERENCES

[1] M.A. Bartsch and G.H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. *IEEE Trans. Multimedia*, 7(1):96–104, 2005.

[2] J.P. Bello. Grouping recorded music by structural similarity. In *Proc. ISMIR*, pages 531–536, 2009.

[3] M. Casey and M. Slaney. Song Intersection by Approximate Nearest Neighbor Search. In *Proc. ISMIR*, 2006.

[4] D.P.W. Ellis and G.E. Poliner. Identifying 'cover songs' with chroma features and dynamic programming beat tracking. In *Proc. ICASSP*, pages IV–1429–1432, 2007.

[5] M. Levy and M. Sandler. Structural Segmentation of Musical Audio by Constrained Clustering. *IEEE Trans. Audio, Speech, and Language Processing*, 16(2), 2008.

[6] H. Lukashevich. Towards quantitative measures of evaluating song segmentation. In *Proc. ISMIR*, 2008.

[7] M. Marolt. A mid-level representation for melody-based retrieval in audio collections. *IEEE Trans. Multimedia*, 10(8):1617–1625, 2008.

[8] M. Mauch, K. C. Noland, and S. Dixon. Using musical structure to enhance automatic chord transcription. In *Proc. ISMIR*, pages 231–236, 2009.

[9] M. Müller, F. Kurth, and M. Clausen. Audio matching via chroma-based statistical features. In *Proc. ISMIR*, pages 288–295, 2005.

[10] A. Ockelford. *Repetition in music: theoretical and metatheoretical perspectives. Volume 13 of Royal Musical Association monographs*. Ashgate Publishing, 2005.

[11] J. Paulus and A. Klapuri. Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. *IEEE Trans. Audio, Speech, and Language Processing*, 17(6):1159–1170, 2009.

[12] G. Peeters and E. Deruty. Is Music Structure Annotation Multi-Dimensional? A Proposal for Robust Local Music Annotation. In *Proc. LSAS*, 2009.

[13] P. Smaragdis and B. Raj. Shift-Invariant Probabilistic Latent Component Analysis. Technical Report TR2007-009, MERL, December 2007.

[14] P. Smaragdis, B. Raj, and M. Shashanka. Sparse and shift-invariant feature extraction from non-negative data. In *Proc. ICASSP*, pages 2069–2072, 2008.

[15] V.Y.F. Tan and C. Févotte. Automatic Relevance Determination in Nonnegative Matrix Factorization. In *Proc. SPARS*, 2009.

[3] http://marl.smusic.nyu.edu/resources/siplca-segmentation/