MUSIC STRUCTURE SEGMENTATION USING SHIFT-INVARIANT PROBABILISTIC LATENT COMPONENT ANALYSIS (MIREX 2010)

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

ABSTRACT

We describe our music structure segmentation algorithm submitted to the MIREX 2010 evaluation. Our method is based on shift-invariant probabilistic component analysis, a variant of convolutive non-negative matrix factorization, applied to chroma features. Repeated harmonic patterns are identified by decomposing a chromagram into a sequence of a small number of repeated basis patterns. The patterns and their locations within a song are simultaneously estimated using an iterative expectation-maximization algorithm. The parameters of the decomposition are then used to compute the long-term structure segmentation by assuming a one-to-one mapping between the identified pattens and segment labels.

1. SEGMENTATION ALGORITHM

Our segmentation system is described in detail in [3]. In the following sections we briefly review the algorithm and describe the extensions implemented for MIREX.

The Python implementation of the algorithm is freely available under the terms of the GNU General Public License. The most recent version can be found online at `http://ronw.github.com/siplca-segmentation`

1.1 Features

The segmentation algorithm uses beat-synchronous chroma features, computed using the algorithm described in [1], and normalized so that the maximum value in each frame is 1. Example features computed from *Good Day Sunshine* by The Beatles are shown in the top left pane of Figure 1.

1.2 SI-PLCA

The beat-synchronous chromagram for a given song is decomposed using shift-invariant probabilistic latent component analysis (SI-PLCA) [2] into the convolution of k basis components, W_k, and their activations in time, $\mathbf{h}_k$. The decomposition for each point in the chromagram V can be written as follows:

$$v_{ft} \approx \hat{v}_{ft} = \sum_k \sum_\tau z_k \, w_{fk\tau} \, \overset{\rightarrow\tau}{h}_{kt} \tag{1}$$

where z_k corresponds to the mixing weight for each component and $\overset{\rightarrow\tau}{x}$ shifts x τ places to the right.

The bases W_K correspond to fixed-length chroma templates that are repeated throughout the song. The corresponding activation function $\mathbf{h}_k^T$ denotes when each component is active in time.

The number of components K is fixed to a large number (15) and unneeded components are pruned away by enforcing that the mixing weights z_k have a sparse distribution. This enables the algorithm to automatically identify the optimal number of bases needed to adequately explain the data.

For more details, including the full expectation-maximization algorithm for estimating W_k, z_k, and $\mathbf{h}_k$ from V, see [3] and [2].

1.3 Segmentation

Given the SI-PLCA decomposition described in the previous section, we derive the structure segmentation using the following likelihood function:

$$P\big(t,\, k\big) = \sum_f \sum_\tau z_k \, w_{fk\tau} \, \overset{\rightarrow\tau}{h}_{kt} \tag{2}$$

The quantity in equation (2) corresponds to the probability that the observation at time t comes from basis k. An example is shown in Figure 2. We assume that each basis corresponds to a unique segment label and compute the final segmentation from $P\big(t,\, k\big)$ by finding the optimal setting of k at each time frame. We compute this path through equation (2) using the Viterbi algorithm using a simple transition matrix designed to smooth out transitions between segments. The transition matrix is constructed to have a large weight along the diagonal to discourage spurious transitions between segments. The off diagonal components are uniform, so no preference is given for any particular state.

$$a_{ij} = \begin{cases} p & i = j \\ \frac{1}{K-1}(1-p) & i \neq j \end{cases} \tag{3}$$

p was set to 0.9 in the MIREX submission. Finally, the segmentation output by the Viterbi algorithm is processed to remove any segments shorter than 32 beats. An example segmentation is shown in Figure 2.

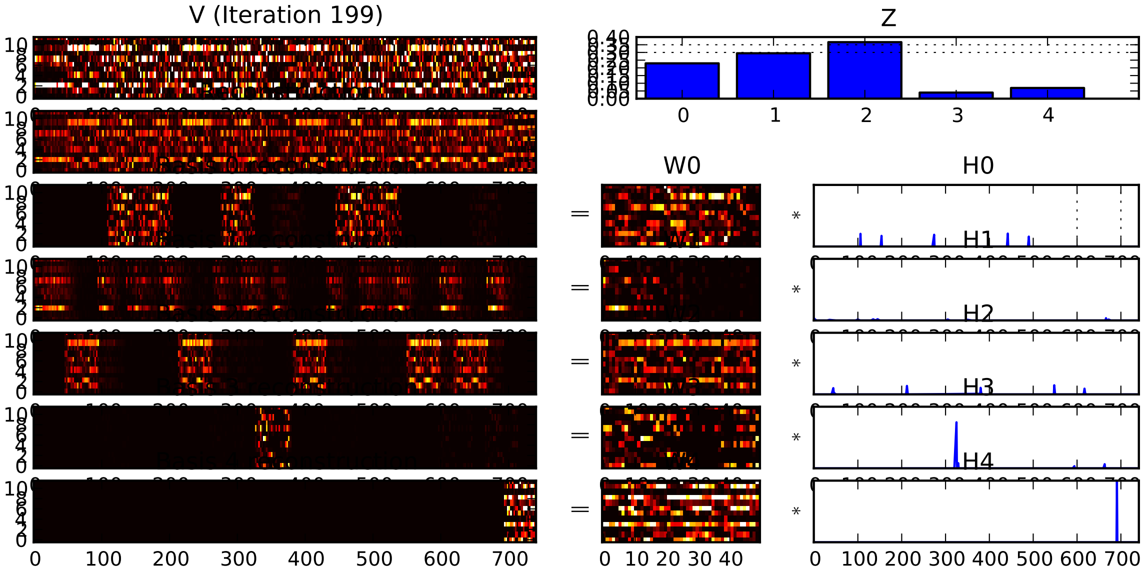

Figure 1. Example SI-PLCA decomposition. The top left pane shows the original beat-synchronous chromagram. Directly underneath is the approximation using SI-PLCA. The remaining panes in the left column contain the reconstruction using each basis alone. Finally, the parameters of the decomposition are shown in the right column.

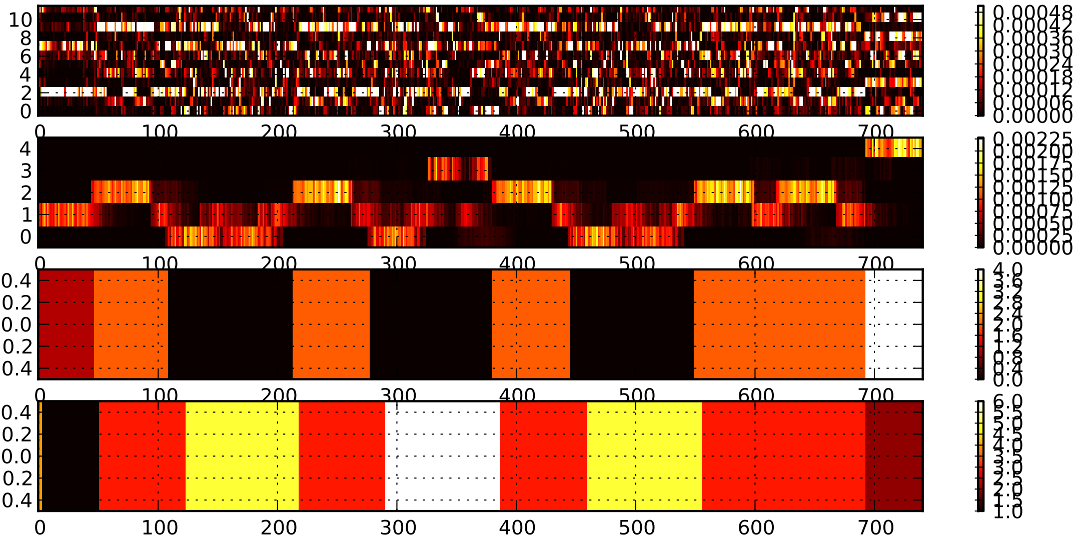

Figure 2. Structure segmentation derived from the SI-PLCA decomposition shown in Figure 1. The top pane shows the chromagram V. The middle panes show $P(t,\, k)$ from equation (2) and the resulting segmentation, respectively. The bottom pane shows the ground-truth segmentation.

2. REFERENCES

[1] D.P.W. Ellis and G.E. Poliner. Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages IV–1429–1432, 2007.

[2] P. Smaragdis, B. Raj, and M. Shashanka. Sparse and shift-invariant feature extraction from non-negative data. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2069–2072, 2008.

[3] R. J. Weiss and J. P. Bello. Identifying Repeated Patterns in Music Using Sparse Convolutive Non-Negative Matrix Factorization. In *Proc. International Conference on Music Information Retrieval (ISMIR)*, Utrecht, Netherlands, August 2010.