Identifying Repeated Patterns in Music Using Sparse Convolutive Non-Negative Matrix Factorization

ISMIR 2010

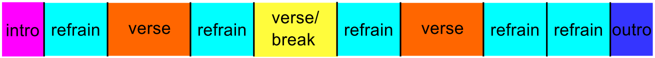
Ron Weiss Juan Bello
{ronw,jpbello}@nyu.edu

Music and Audio Research Lab
New York University

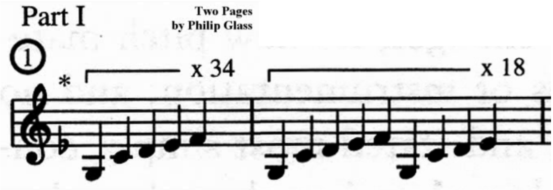
August 10, 2010

Repetitive patterns in music

- Repetition is ubiquitous is music
  - long-term verse-chorus structure

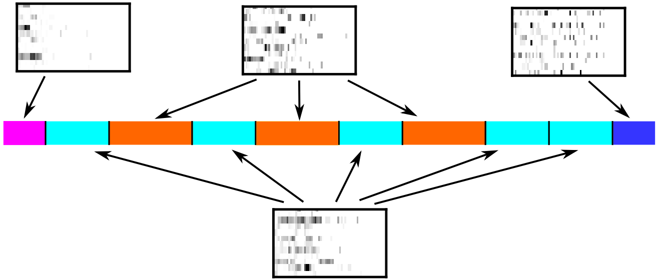


  - repeated motifs



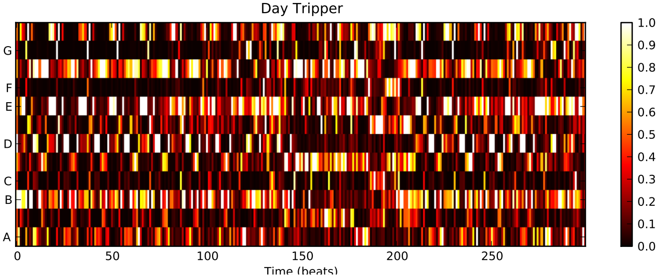
- Can we identify this structure directly from audio?
  - What about the repeated units?

Proposed approach



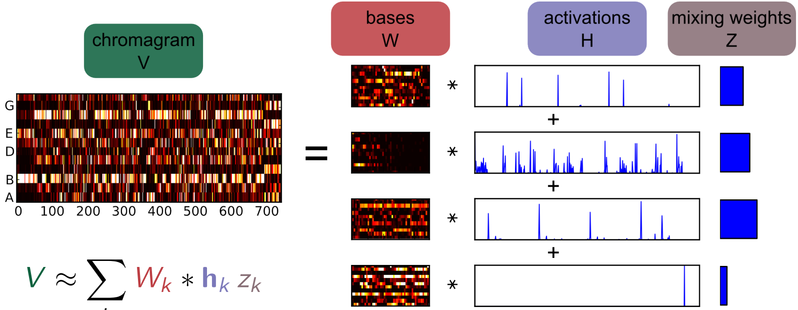
- Treat song as concatenation of short, repeated template patterns
- Inspired by source separation / text topic modeling
  - Convolutive Non-negative Matrix Factorization (NMF)

Beat-synchronous chroma features [Ellis and Poliner, 2007]



- Summarize energy at each pitch class during each beat
- Normalize frame energy to ignore dynamics

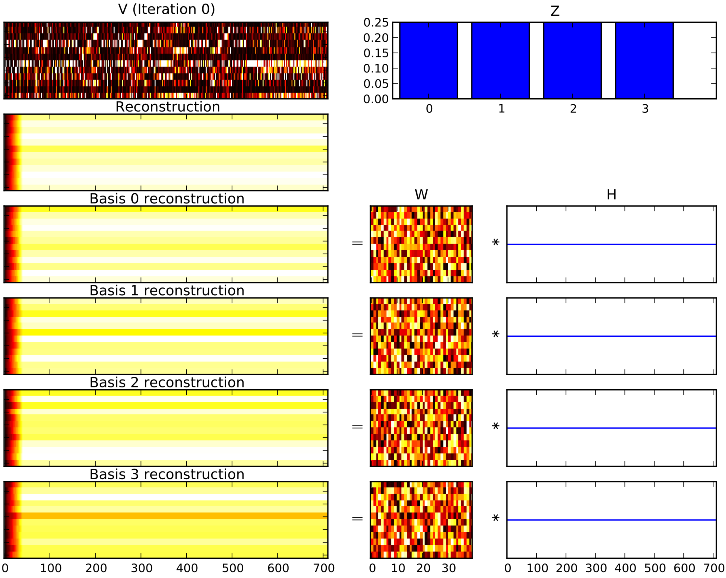
- Shift-invariant Probabilistic Latent Component Analysis
  i.e. probabilistic convolutive NMF



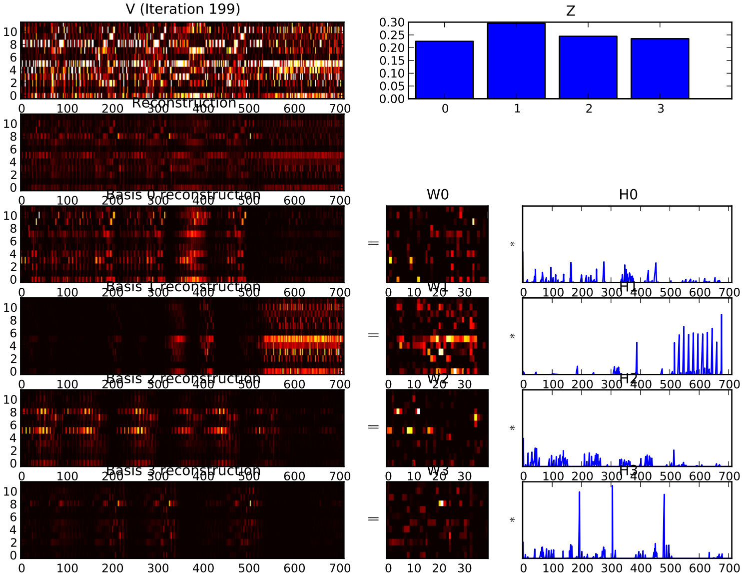
$$V \approx \sum_k W_k * \mathbf{h}_k \, z_k$$

- Decompose matrix V into weighted (by Z) sum of latent components
  - each component is convolution of basis W with activations H
- Short-term structure in W, long-term structure in H
- Must specify number, length of patterns
- Iterative EM learning algorithm

Learning algorithm example – Initialization

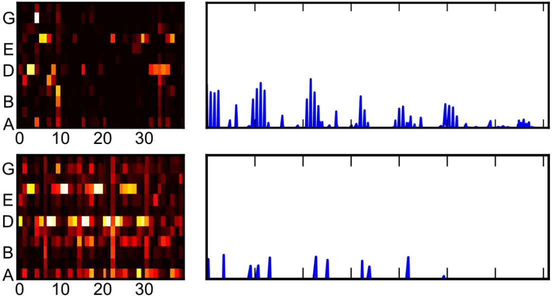


Learning algorithm example – Converged



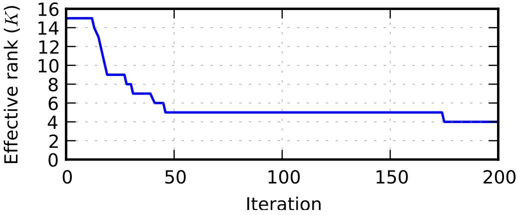
Sparsity

- Encourage sparse (mostly zero) parameters using prior distributions
- Use entropic prior over activations H [Smaragdis et al., 2008]
    - low entropy $\implies$ less uniform
- Leads to more meaningful patterns
    - but reduces temporal information in activations
    - sparse H $\implies$ dense W

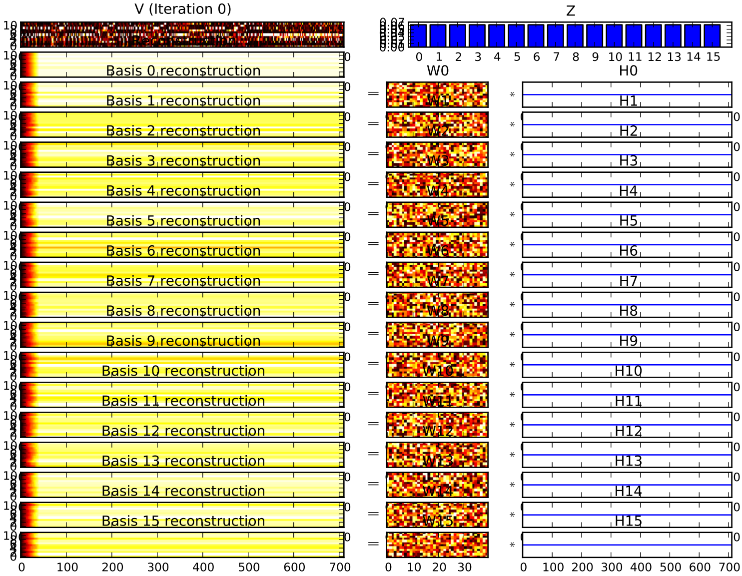


Automatic relevance determination [Tan and Févotte, 2009]

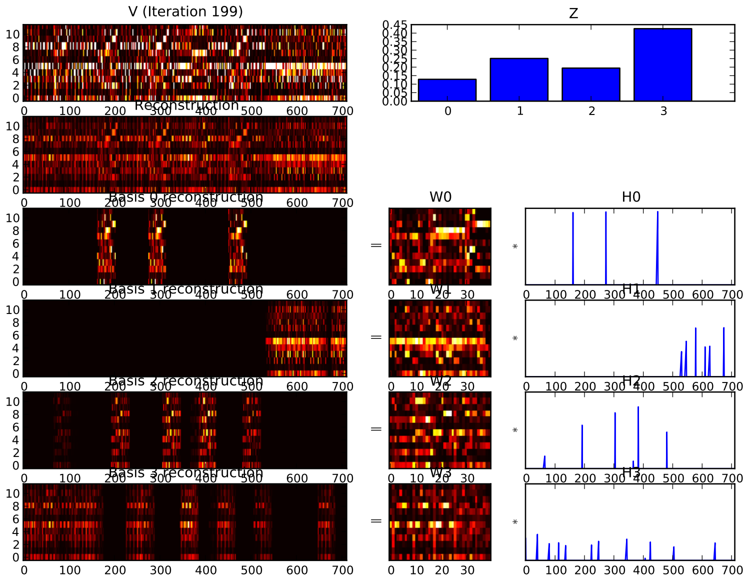
- Avoid having to specify number of patterns in advance
  - Initialize decomposition with large number of patterns
  - Sparse Dirichlet distribution over mixing weights Z
  - Discard unused patterns



Sparse learning example – Initialization

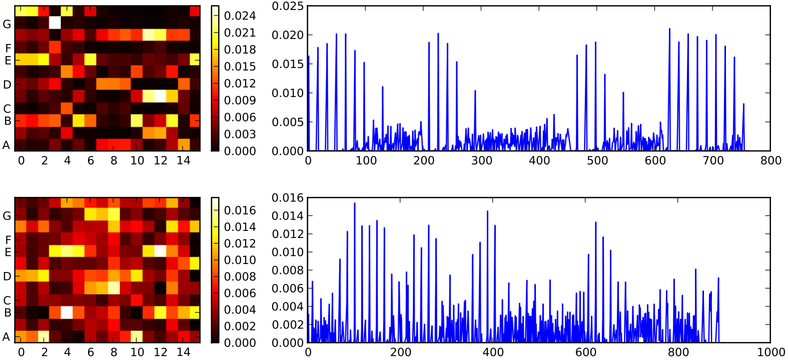


Sparse learning example – Converged

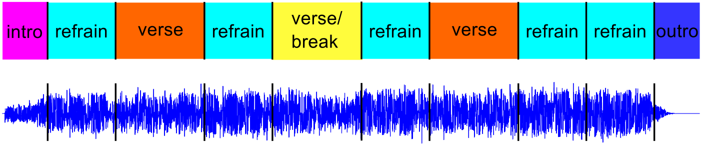


Applications: Riff identification / Thumbnailing

- Reconstruct song using a single pattern
  - Sparse activations
  - Riff length known in advance (for now)
  - Thumbnail corresponds to largest activation in H

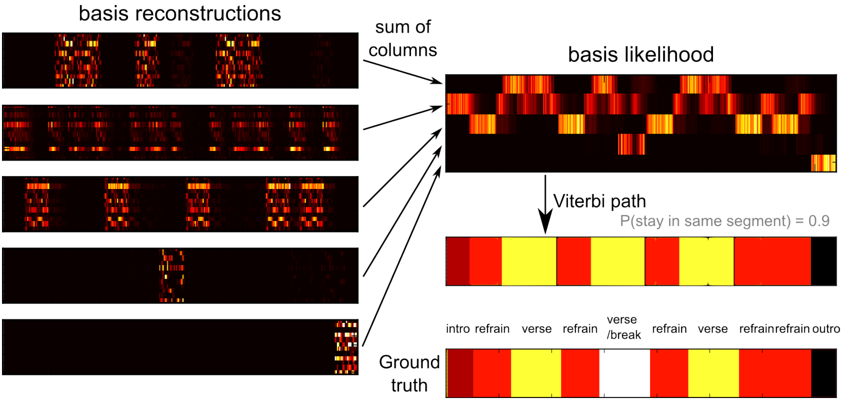


Applications: Structure segmentation



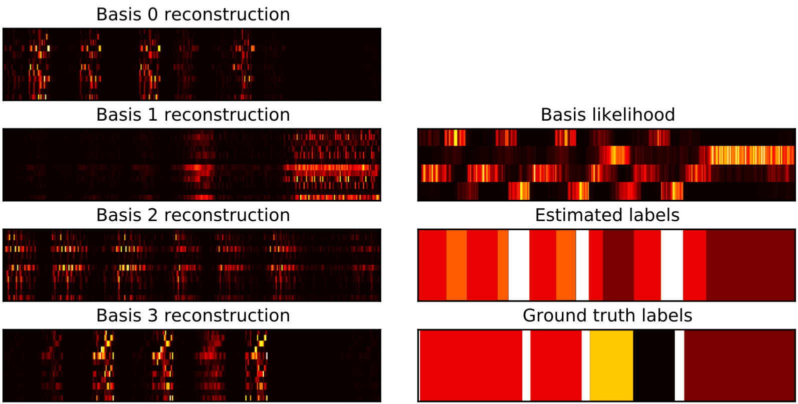
- Identify long-term song structure (verse, chorus, bridge, etc.)
- Assume one-to-one mapping between chroma patterns and segments
- Use SI-PLCA decomposition with longer patterns
  - no prior on activations

Structure segmentation example



Estimated	intro	refrain	verse	refrain	verse	refrain	verse	refrain	refrain	outro
Ground truth	intro	refrain	verse	refrain	vs/break	refrain	verse	refrain	refrain	outro

Structure segmentation example 2



segments tend to be broken into multiple motifs

Est verse1 verse2 verse1 verse2 refrain. verse1 verse2 refrain. verse1 outro. verse1 refrain. verse1 out

GT verse verse refrain. verse refrain. $\frac{1}{2}$ verse inst. $\frac{1}{2}$ verse refrain. outro

Experiments

- Evaluate on 180 songs from *The Beatles* catalog

System	f-meas	prec	recall	over-seg	under-seg
[Mauch et al., 2009]	0.66	0.61	0.77	0.76	0.64
SI-PLCA (sparse Z)	0.60	0.58	0.68	0.61	0.56
SI-PLCA (rank=4)	0.58	0.60	0.59	0.56	0.59
[Levy and Sandler, 2008]	0.54	0.58	0.53	0.50	0.57
Random	0.30	0.36	0.26	0.07	0.24

- Compare to systems based on self-similarity and HMM clustering
  - middle of the pack performance
  - sparse Z gives $\sim 10\%$ improvement in recall over fixed rank
- Needs better post-processing?

Summary

- Novel algorithm for identifying repeated harmonic patterns in music
- Use sparsity to minimize number of fixed parameters, control structure
- Applications to thumbnailing and structure segmentation
- Future work
  - Adaptive model of pattern length, better downbeat alignment
  - 2D convolution to compensate for key changes
  - Time-warp invariance (beat-tracking errors, fixed hop size)

Open source Python/Matlab implementation available:
http://ronw.github.com/siplca-segmentation

References

Ellis, D. and Poliner, G. (2007).
Identifying 'cover songs' with chroma features and dynamic programming beat tracking.
In *Proc. ICASSP*, pages IV–1429–1432.

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

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

Smaragdis, P. and Raj, B. (2007).
Shift-Invariant Probabilistic Latent Component Analysis.
Technical Report TR2007-009, MERL.

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

Tan, V. and Févotte, C. (2009).
Automatic Relevance Determination in Nonnegative Matrix Factorization.
In *Proc. SPARS*.