2016 Intelligent Sensing Summer School
Matrix Decomposition Methods for Audio Analysis

Emmanouil Benetos

Machine Listening Lab
Centre for Digital Music
Queen Mary University of London
http://machine-listening.eecs.qmul.ac.uk/

September 2016

Outline

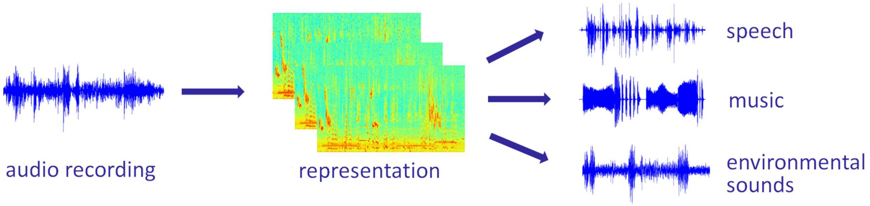
Outline

1. Intro: Machine Listening
2. Matrix/Spectrogram Factorization
3. Application: Music Signal Analysis
4. Application: Sound Scene Analysis
5. Discussion

Machine Listening

Machine listening: extracting meaningful information from audio signals.

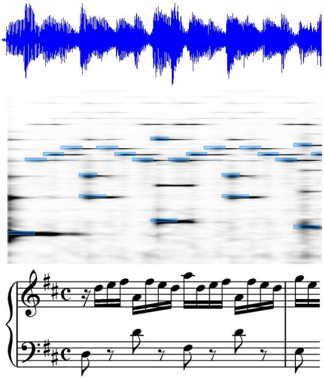
- Sounds: speech, music, environmental/everyday sounds
- Disciplines: signal processing, machine learning, acoustics, perception



Machine Listening for Music Signals

Core problems:

- Tonality: multi-pitch detection, chord/key estimation
- Rhythm: onset detection, beat tracking, meter induction
- Source (instrument) separation/identification



Applications:

- Music information retrieval
- Interactive music systems
- Computer music
- Musicology

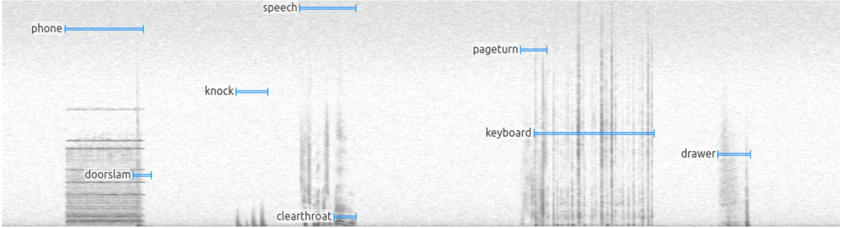
Machine Listening for Everyday Sounds

Core problems:

- Sound event detection
- Sound scene recognition
- Source separation

Applications:

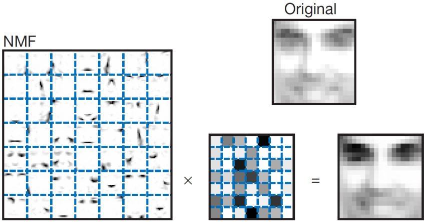
- Audio archiving
- Security/surveillance
- Smart homes/cities
- Acoustic ecology



Outline

Matrix/spectrogram factorization (1)

- Non-negative Matrix Factorization (NMF) - (Lee and Seung, 1999)
- Unsupervised algorithm for factorizing a matrix into a low rank decomposition
- Constraint: non-negative data
- This allows a parts-based representation
- Applications: detection, dimensionality reduction, clustering, classification, denoising, prediction...



Matrix/spectrogram factorization (2)

NMF Model:

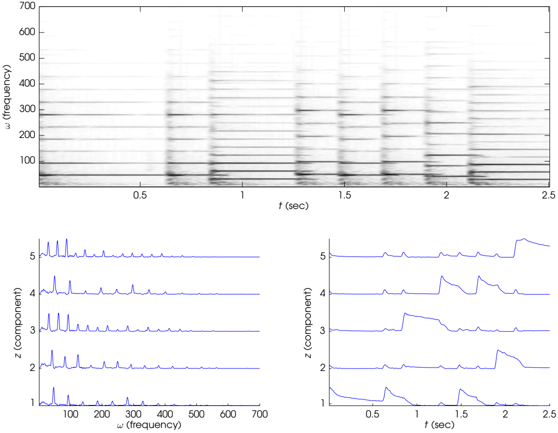
- Given a non-negative matrix V find non-negative matrix factors W and H such that:

$$V \approx WH$$

  where:
  - $V \in \mathbb{R}^{n \times m}$
  - $W \in \mathbb{R}^{n \times r}$
  - $H \in \mathbb{R}^{r \times m}$
- The rank r of the factorization is chosen as $(n+m)r < nm$, so that data is compressed
- Various algorithms (e.g. EM, ALS) and cost functions (e.g. KL, IS) have been proposed in the literature

Matrix/spectrogram factorization (3)

NMF can be applied to audio (magnitude) spectrograms:



Matrix/spectrogram factorization (4)

Probabilistic Latent Semantic Analysis (PLSA)

- In 1999, Thomas Hofmann proposed a technique for text processing and retrieval, called Probabilistic Latent Semantic Indexing (PLSI)
- ...which is also called structure Probabilistic Latent Semantic Analysis (PLSA)
- ...and also Probabilistic Latent Component Analysis (PLCA)!
- In fact, PLSA/PLSI/PLCA are the probabilistic counterparts of NMF using a specific cost function (KL divergence)
- This interpretation offers a framework that is easy to generalise and extend

Matrix/spectrogram factorization (5)

PLCA model

$$V_{\omega,t} \approx P(\omega, t) = P(t) \sum_z P(\omega|z)P(z|t)$$

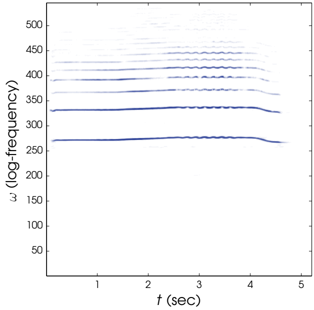
$V_{\omega,t}$: input spectrogram, $P(t)$ frame energy, $P(\omega|z)$: basis spectra, $P(z|t)$: component activations.

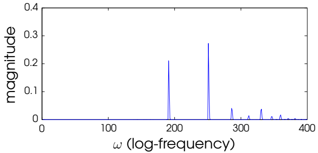
- PLCA can decompose a spectrogram into a series of spectral bases that correspond to 'sound events' and a series of event activations
- Applications in machine listening: sound event detection, multi-pitch detection, source separation/identification

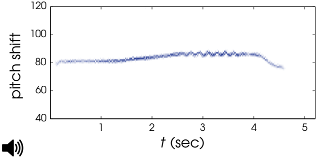
- Convolutive models: extracting shifted structures from non-negative data
- Shift-invariant PLCA (across 1 dimension):

$$V_{\omega,t} \approx P(\omega,t) = \sum_z P(z) \sum_f P(\omega - f|z)P(f,t|z)$$

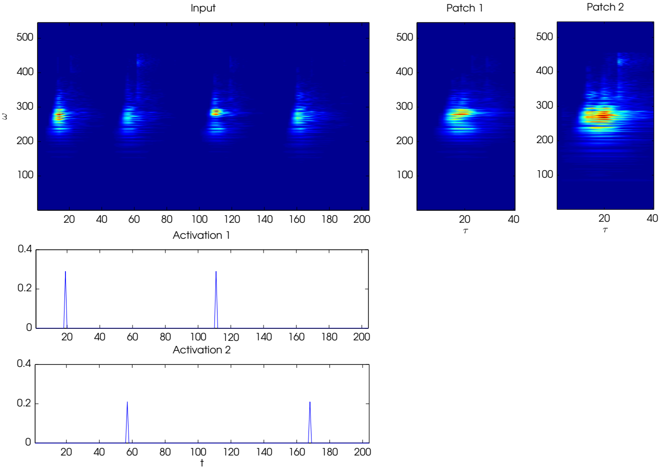
$V_{\omega,t}$: log-frequency spectrogram, $P(z)$: component prior, $P(\omega|z)$: basis spectra, $P(f,z|t)$: pitch impulse distribution.







SIPLCA example (across 2 dimensions):
detecting footstep sounds

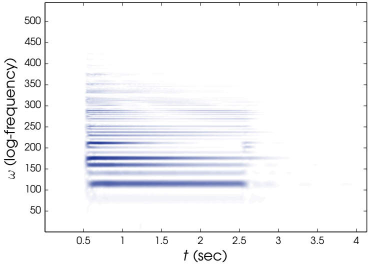


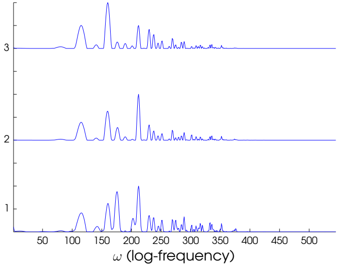
Outline

Application: Music Signal Analysis (1)

Goal:

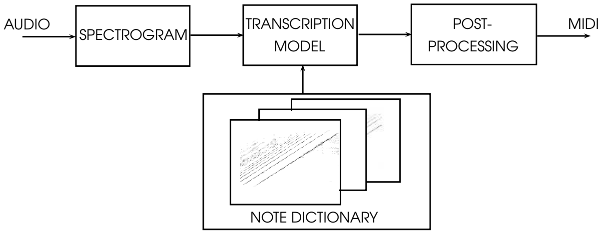
- Create a multiple-instrument music transcription system
- Express frequency modulations & tuning changes through shift-invariance
- Express each note as a temporal succession of *sound state* spectral templates





Application: Music Signal Analysis (2)

- Proposed model: Hidden Markov Model-constrained Shift-Invariant PLCA
- Fixed (pre-extracted) dictionary of spectral templates per instrument, pitch, sound state



E. Benetos and S. Dixon, "Multiple-instrument polyphonic music transcription using a temporally constrained shift-invariant model", *Journal of the Acoustical Society of America*, 133(3):1727-1741, 2013

Application: Music Signal Analysis (3)

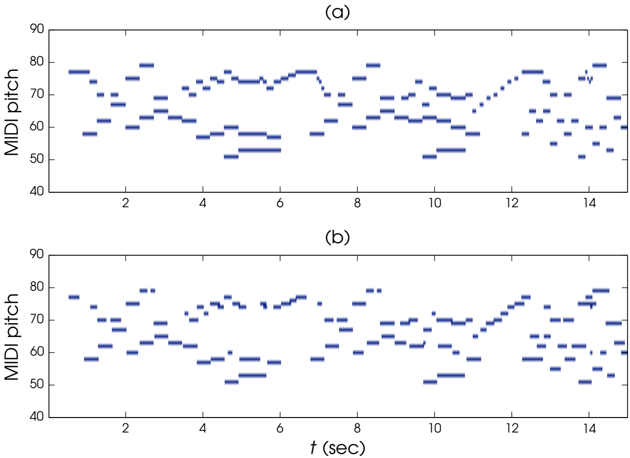


Figure : (a) The ground-truth piano-roll of Mozart's Piano Sonata K.333, 3rd mvt. (b) The transcription output piano-roll. Original recording: Synthesized transcription:

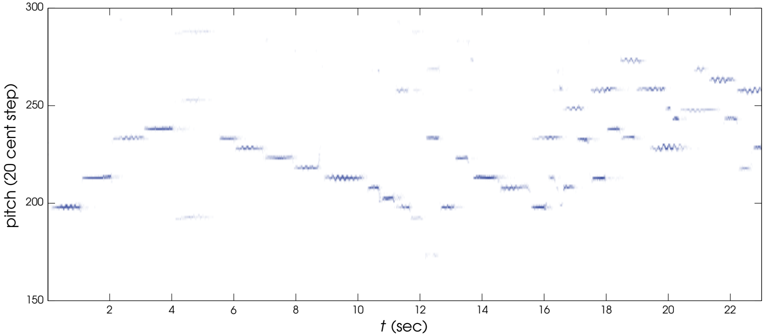
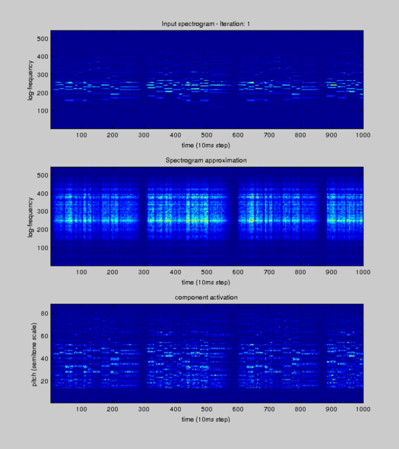


Figure : A time-pitch representation in 20 cent resolution for a recording of J.S. Bach's Musical Offering, BWV 1079.

NMF convergence (video)



Application: Music Signal Analysis (6)

Speed:

- Convergence is observed with 10-15 iterations
- Runtime: 1-2.5 $\times$ real-time (CPU)
- Faster matrix multiplications with GPU processing (0.3 $\times$ real-time)

Evaluation:

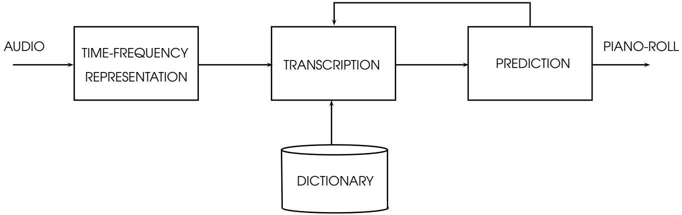
- Proposed method ranked first in public evaluations for multiple-F0 estimation and note tracking (MIREX 2013, 2015)

Code:

- `http://www.eecs.qmul.ac.uk/~emmanouilb/code.html`

Priors in PLCA / Music Language Models

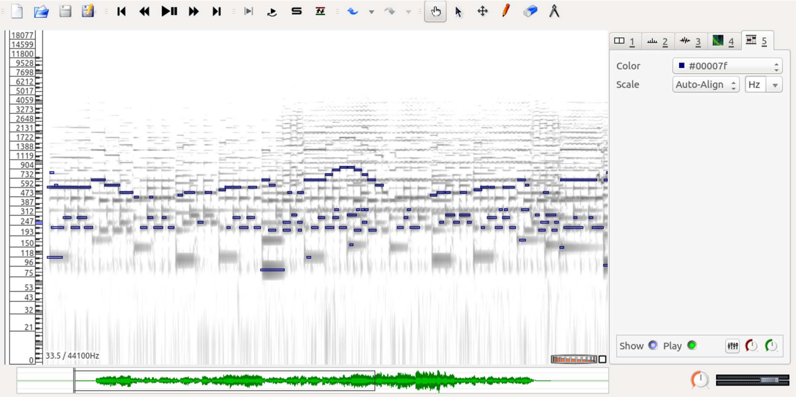
- Combine PLCA acoustic model with Recurrent Neural Network music language model (MLM) for prediction
- MLM is incorporated into PLCA using Dirichlet priors



S. Sigtia, E. Benetos, S. Cherla, T. Weyde, A. d'Avila Garcez, and S. Dixon, "An RNN-based music language model for improving automatic music transcription", in Proc. ISMIR, 2014.

Application: Music Signal Analysis (8)

- Silvet note transcription plugin:
  `https://code.soundsoftware.ac.uk/projects/silvet/`



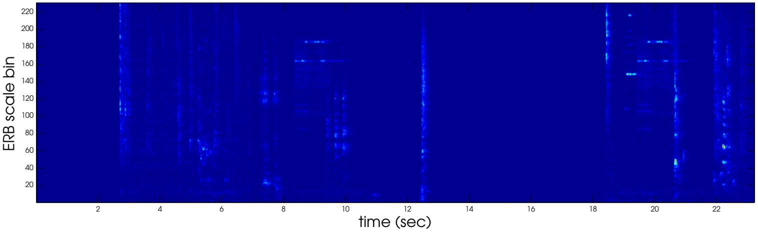
Outline

Application: Sound Scene Analysis (1)

Motivation

- Adapt PLCA model for detecting overlapping sound events
- Each event class contains several exemplars; each exemplar consists of a sequence of sound states
- Appearance of sound states controlled by event-wise HMMs

- Input representation: (auditory-motivated) Equivalent Rectangular Bandwidth filterbank



Application: Sound Scene Analysis (2)

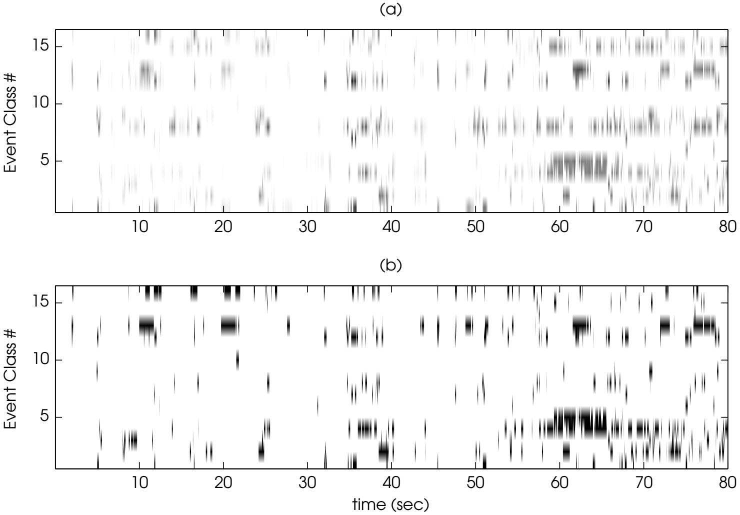
Proposed Model

$$V_{\omega,t} \approx P(\omega,t) = P(t) \sum_{q,c,s} P(\omega|q,c,s)P(s|t)P(c|s,t)P(q|s,t)$$

- $V_{\omega,t}$: ERB spectrogram
- $P(t)$: spectrogram energy (known quantity)
- $P(\omega|q,c,s)$: spectral template for event class s, exemplar c, and sound state q (pre-extracted, fixed)
- $P(s|t)$: event activation over time
- $P(c|s,t)$: exemplar contribution for each event class, over time
- $P(q|s,t)$: sound state contribution for each event class, over time

- $P(q|s,t)$ controlled by an event-wise HMM

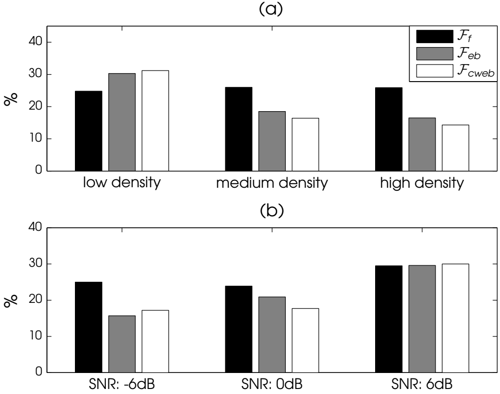
Application: Sound Scene Analysis (3)



Top: $P(s|t)$. Bottom: post-processed binary event-roll.

Application: Sound Scene Analysis (4)

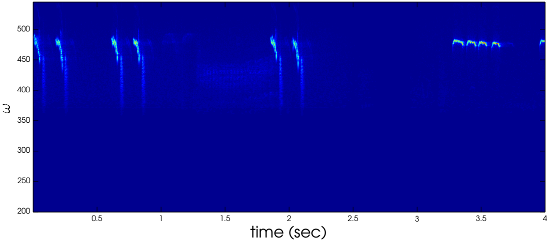
- Dataset: DCASE 2013 OS (office sounds in different density/noise conditions)



E. Benetos, G. Lafay, M. Lagrange, and M. D. Plumbley, "Detection of overlapping acoustic events using a temporally-constrained probabilistic model", ICASSP, 2016.
https://code.soundsoftware.ac.uk/projects/sound-event-detection-plca

Application: Sound Scene Analysis (5)

- Goal: identifying multiple bird species from continuous recordings



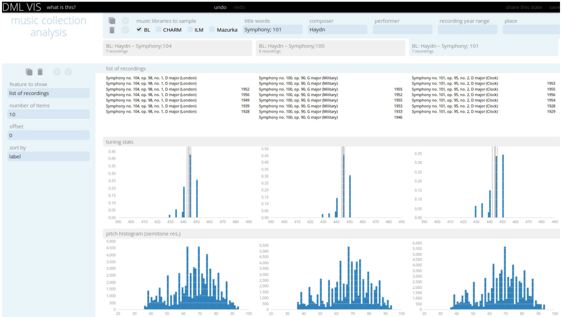
- Approach: PLCA with pre-extracted templates per species
- Participated in ICML 2013 Bird Challenge on identifying 35 bird species; system in the top 25% of submissions

E. Benetos, "Acoustic identification of bird species using probabilistic latent component analysis", in ICML Workshop on Machine Learning for Bioacoustics, pp. 77-78, 2013.

Outline

Conclusions

- Matrix factorization: powerful tool for audio analysis
- Interpretable, extensible, computationally efficient
- Additional uses: sound scene recognition, drum transcription, instrument recognition...
- Used in several commercial & public tools

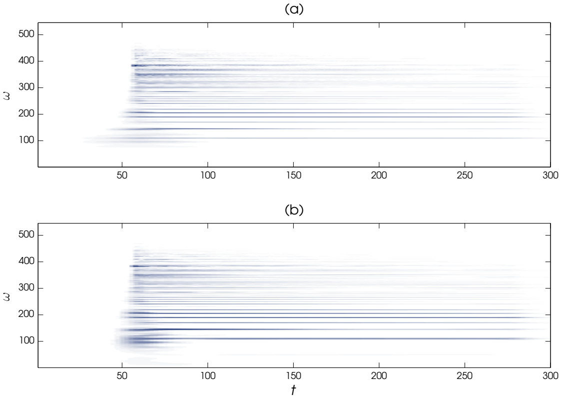


http://dml.city.ac.uk/vis/

Discussion (2)

Future Directions

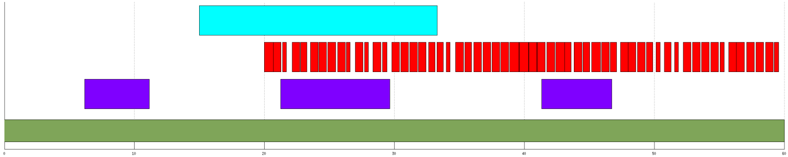
- Input time-frequency representations
- Beyond the Short-time Fourier Transform: CQT, VQT, auditory spectra...



Discussion (3)

Future Directions (cont'd)

- Acoustic language models



- Sound event taxonomy
- Source/acoustic environment adaptation

Many thanks to

- Chris Cannam
- Simon Dixon
- Artur d'Avila Garcez
- Dimitrios Giannoulis
- André Holzapfel
- Anssi Klapuri
- Grégoire Lafay
- Mathieu Lagrange
- Mark Plumbley
- Siddharth Sigtia
- Dan Stowell
- Dan Tidhar
- Tillman Weyde
- Daniel Wolff

SUPPORT:

ROYAL ACADEMY OF ENGINEERING

Arts & Humanities Research Council