Decomposition of musical spectrograms informed by spectral synthesis models

—

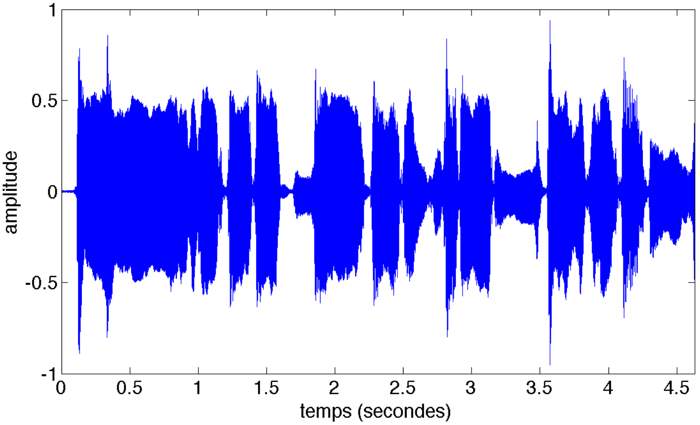
Modeling of time variations in sound elements

Romain Hennequin

Telecom ParisTech

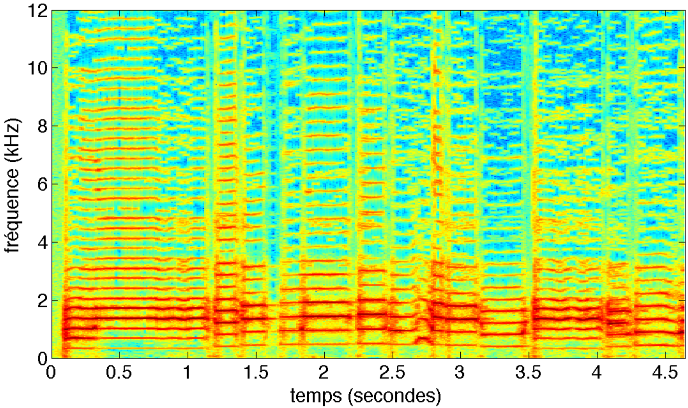
21 november 2011

Introduction



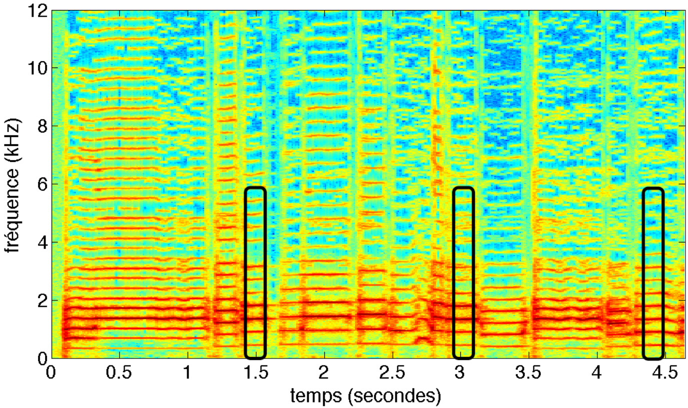
Waveform

Introduction



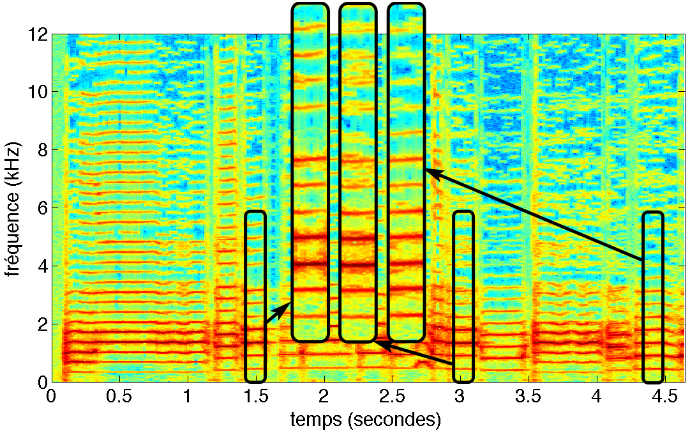
spectrogram

Introduction



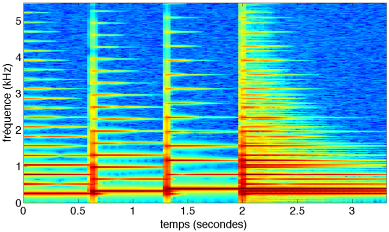
spectrogram

Introduction

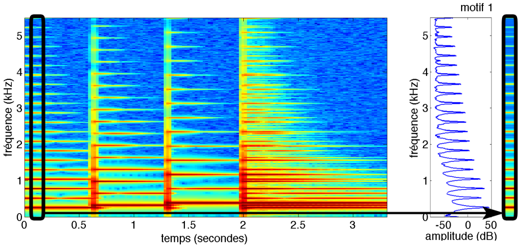


spectrogram

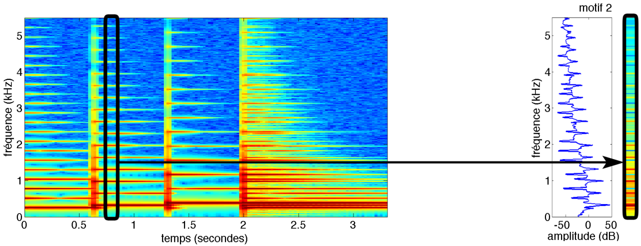
Redundancy extraction: Non-negative Matrix Factorization (NMF)



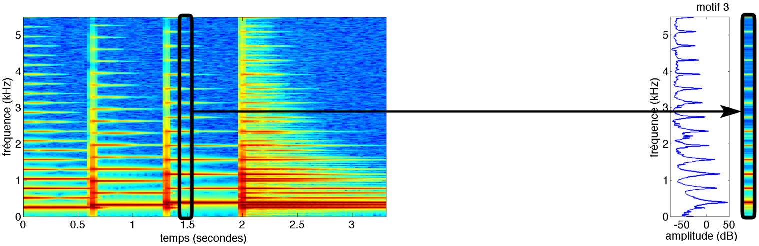
Redundancy extraction: Non-negative Matrix Factorization (NMF)



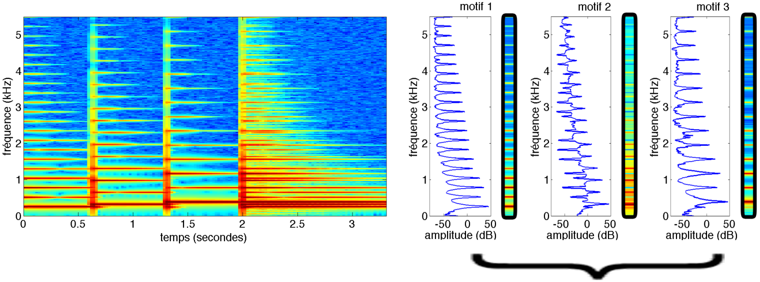
Redundancy extraction: Non-negative Matrix Factorization (NMF)



Redundancy extraction: Non-negative Matrix Factorization (NMF)

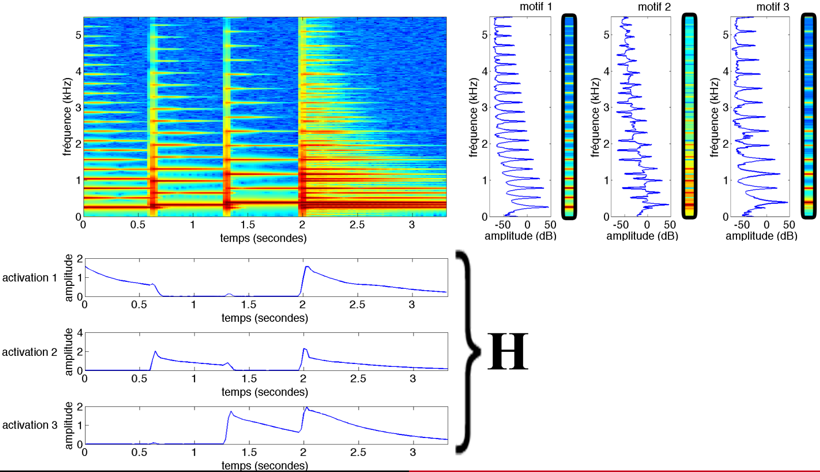


Redundancy extraction: Non-negative Matrix Factorization (NMF)

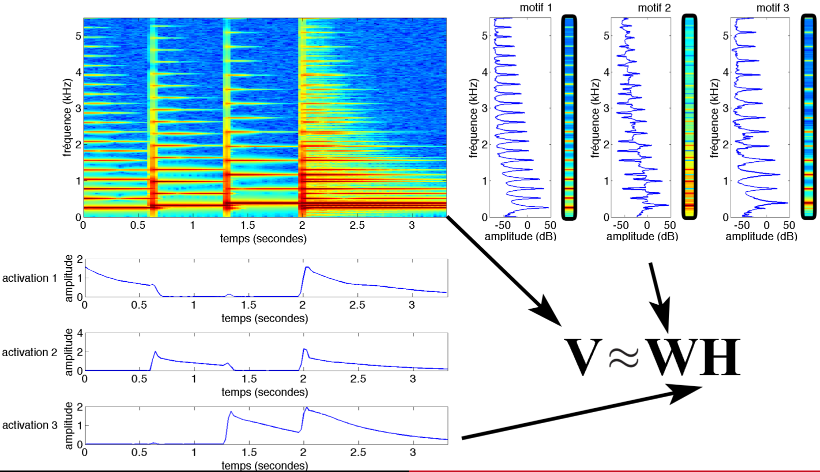


$\mathbf{W}$

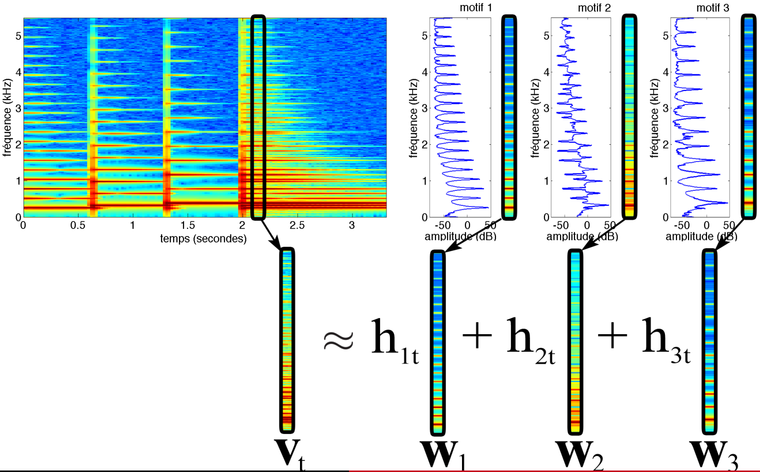
Redundancy extraction: Non-negative Matrix Factorization (NMF)



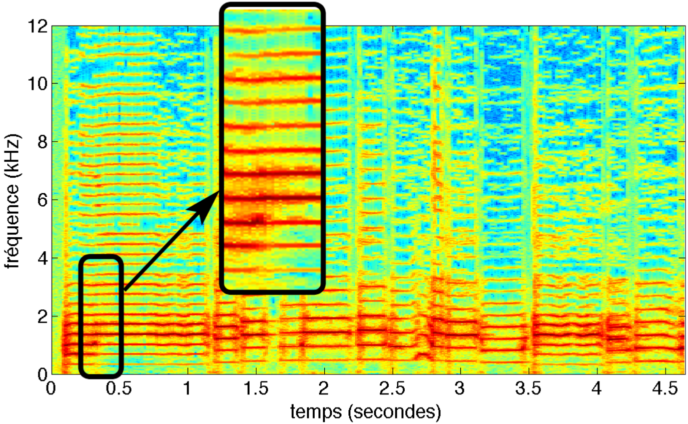
Redundancy extraction: Non-negative Matrix Factorization (NMF)



Redundancy extraction: Non-negative Matrix Factorization (NMF)

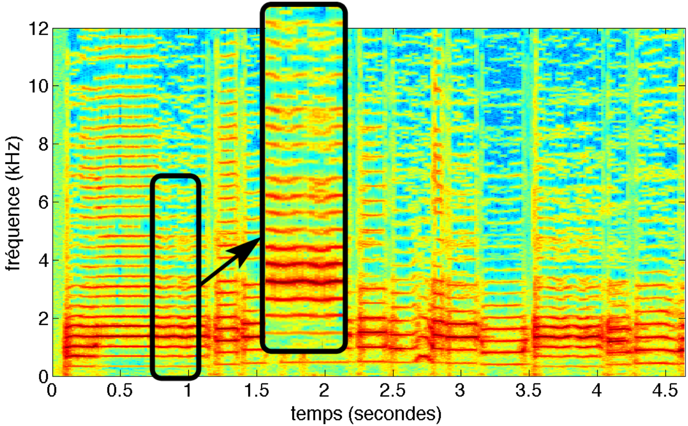


Introduction



Spectral envelope variations

Introduction



Variations de fréquence fondamentale

Introduction

Framework and goal

- Automatic decomposition of musical spectrograms in a "perceptual way": extracted elements should have a perceptual meaning.
- Decomposition based on NMF: extraction of simple redundancies.
- Spectrogram models based on sound synthesis technics are introduced into decomposition methods in order to model time variations.

Introduction

Possible applications:

- Automatic transcription (score)
- Source separation
- Selective transformation of sounds
- ...

Contents

Contents

Principle of NMF

Low rank approximation:

$\mathbf{V}$ being a non-negative matrix (amplitude or power spectrogram), NMF approximates $\mathbf{V}$ in the following way:

$$\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$$

$$v_{ft} \approx \hat{v}_{ft} = \sum_{r=1}^{R} w_{fr} h_{rt}$$

- $\mathbf{W}$ matrix $F \times R$ and $\mathbf{H}$ matrix $R \times T$.
- Non-negativity constraints: $\mathbf{W}, \mathbf{H} \geq 0$.
- Rank reduction: $R \ll \min(F, T)$

Algorithm

Cost function

- Generally a scalar divergence between $\mathbf{V}$ and $\hat{\mathbf{V}} = \mathbf{WH}$:

$$\mathcal{C}(\mathbf{W}, \mathbf{H}) = D(\mathbf{V} || \mathbf{WH}) = \sum_{f,t} d([\mathbf{V}]_{ft} | [\mathbf{WH}]_{ft})$$

- Common divergences in spectrogram factorization:
  - Kullback-Liebler divergence
  - Itakura-Saito divergence
  - Generalized divergence: β-divergence, Bregman divergence...

Algorithm

Multiplicative algorithm

- Alternated optimization with respect to (wrt) $\mathbf{W}$ and $\mathbf{H}$.
- Decomposition of the gradient as a difference of two positive terms:

$$\nabla_{\mathbf{W}}\mathcal{C}_{\mathbf{V}}(\mathbf{W},\mathbf{H}) = \mathbf{P}_{\mathbf{W}} - \mathbf{M}_{\mathbf{W}} \text{ where } \mathbf{P}_{\mathbf{W}} > 0 \text{ and } \mathbf{M}_{\mathbf{W}} > 0$$
$$\nabla_{\mathbf{H}}\mathcal{C}_{\mathbf{V}}(\mathbf{W},\mathbf{H}) = \mathbf{P}_{\mathbf{H}} - \mathbf{M}_{\mathbf{H}} \text{ where } \mathbf{P}_{\mathbf{H}} > 0 \text{ and } \mathbf{M}_{\mathbf{H}} > 0$$

- Update rules:

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{\mathbf{M}_{\mathbf{W}}}{\mathbf{P}_{\mathbf{W}}}$$
$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{M}_{\mathbf{H}}}{\mathbf{P}_{\mathbf{H}}}$$

Algorithm

Properties of multiplicative algorithms

- Ensure the non-negativity of the parameters
- Local descent direction
- Zeros of the gradient are fixed point of the update rules

More meticulous framework: Majoration/Minimization algorithms.

NMF properties

Features

- Extraction of redundant patterns.
- Fundamental property: non-negativity constraint.
  - Atoms lie in the same space as the data
  - Only non-negative combinations (no black energy).
  - Perceptive description: decomposition of musical spectrograms on a basis of notes.
- Numerous applications in audio: automatic transcription [Smaragdis and Brown, 2003], source separation [Cichocki et al., 2006], audio inpainting [Le Roux et al., 2008] . . .

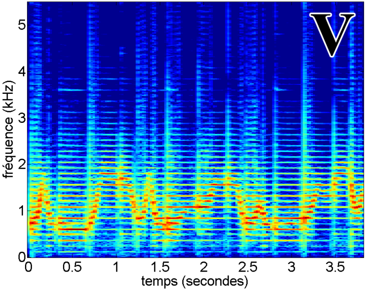
Issues with NMF

Time variations

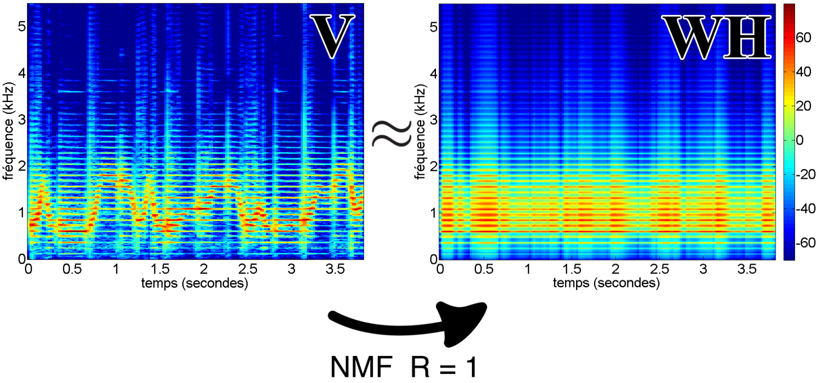
A low-rank approximation does not permit to model efficiently variations over time:

- spectral envelope variations
- pitch variations (vibrato, prosody. . .).

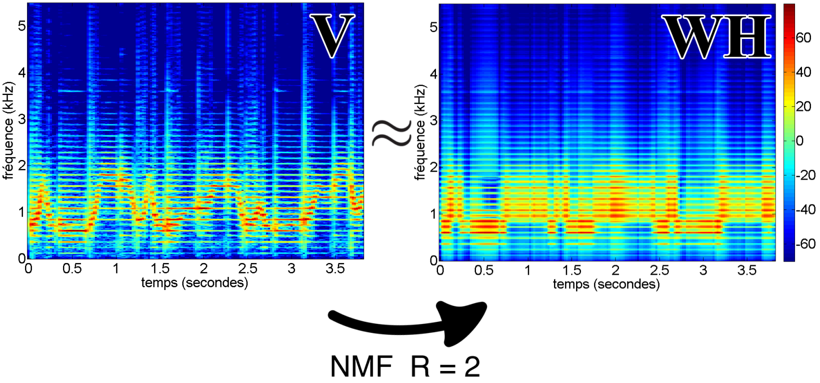
Issues with NMF: spectral envelope variations



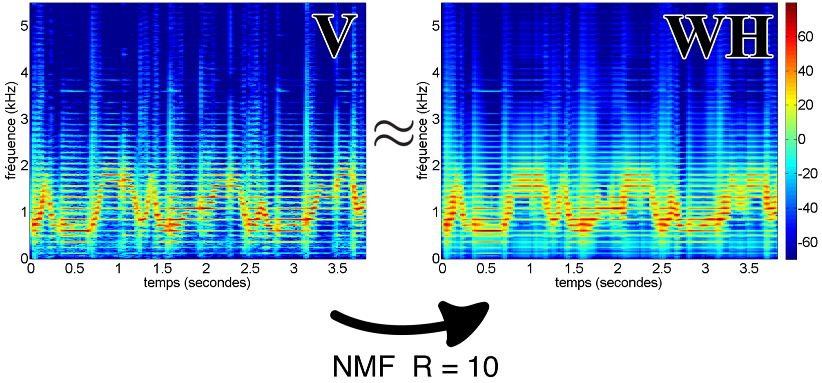
Issues with NMF: spectral envelope variations



Issues with NMF: spectral envelope variations



Issues with NMF: spectral envelope variations

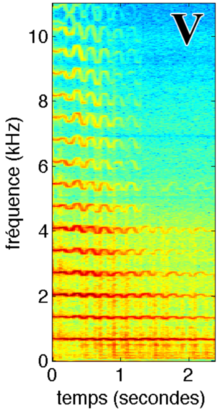


Issues with NMF: spectral envelope variations.

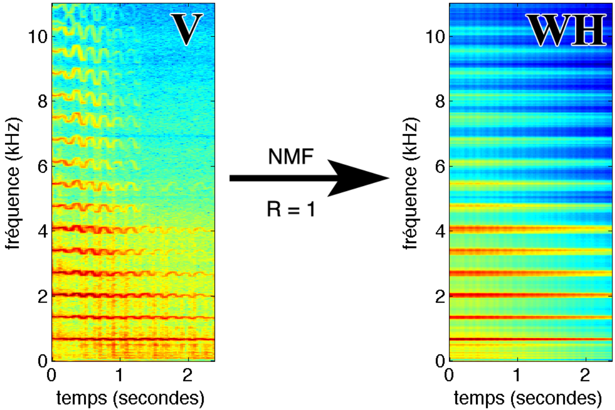
Issues

- Spectral variations of each note is discarded.
- Inefficient for sounds with strong spectral variations.

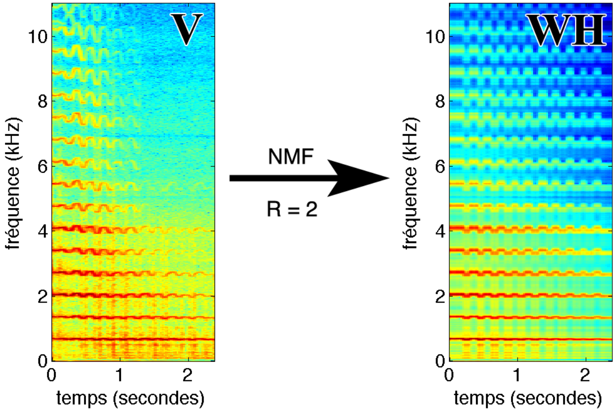
Issues with NMF: fundamental frequency variations



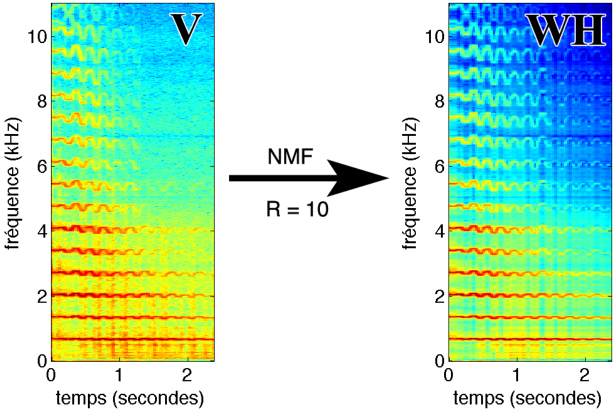
Issues with NMF: fundamental frequency variations



Issues with NMF: fundamental frequency variations



Issues with NMF: fundamental frequency variations



Solutions

Proposed solutions

- Extraction of underlying redundancies.
- Introduction into NMF of generative spectrogram models based on simple sound synthesis technics.

Solutions

Time variations and sound synthesis

- Spectrale envelope variations: introduction of source/filter synthesis in NMF.
- Fundamental frequency variations:
  - Additive synthesis: additive synthesis of a parametric harmonic atoms.
  - Wavetable synthesis: scale-invariant decomposition.

Remark

Proposed decompositions are no longer rank-reduction methods but still reduce the data dimension.

Contents

Time/frequency activations [Hennequin et al., 2011a]

Principle

- Model inspired by source/filter synthesis.
- Temporal activations are replaced by time-varying filters:

$$v_{ft} \approx \sum_{r=1}^{R} w_{fr} h_{rt} \quad \rightarrow \quad v_{ft} \approx \sum_{r=1}^{R} w_{fr} \boxed{h_{rt}(f)}$$

- Limitation of the number of parameters: $h_{rt}(f)$ should be parametric and smooth (with respect to f).

ARMA modeling

Filters parametrization

$h_{rt}(f)$ is the frequency response of an ARMA filter:

$$h_{rt}^{\text{ARMA}}(f) = \sigma_{rt}^2 \frac{\left|\sum_{q=0}^{Q} b_{rt}^q e^{-i2\pi\nu_f q}\right|^2}{\left|\sum_{p=0}^{P} a_{rt}^p e^{-i2\pi\nu_f p}\right|^2}$$

- b_{rt}^q: MA coefficients.
- a_{rt}^p: AR coefficients.
- σ_{rt}^2: global gain.

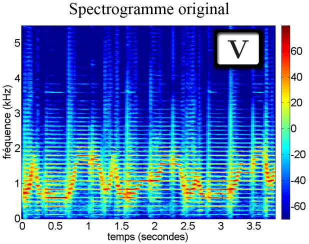
(ν_f: normalized frequency.)

Source/filter decomposition

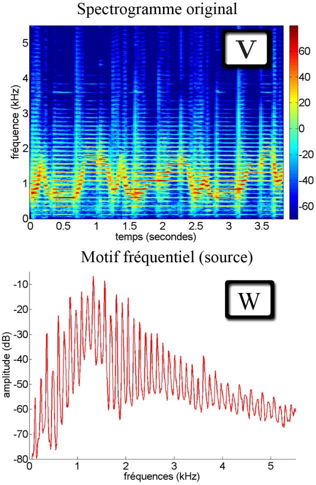
Decomposition

- Decomposition: $v_{ft} \approx \hat{v}_{ft} = \sum_{r=1}^{R} w_{fr} h_{rt}^{\text{ARMA}}(f)$
- Decomposition obtained with a multiplicative algorithm similar to those used in NMF.

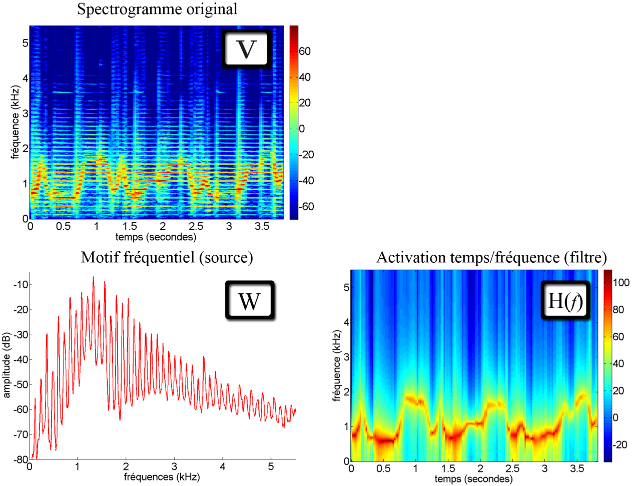
Decomposition of the jew harp sound, 2nd order AR filter



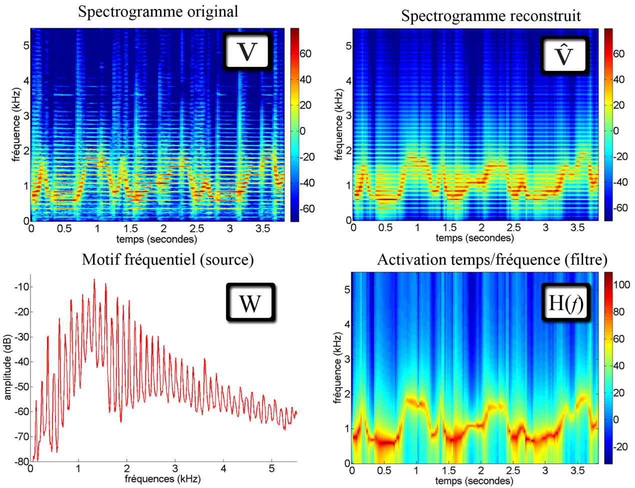
Decomposition of the jew harp sound, 2nd order AR filter



Decomposition of the jew harp sound, 2nd order AR filter



Decomposition of the jew harp sound, 2nd order AR filter



Conclusion

Source/filter paradigm:

Each frame of the spectrogram is a combination of filtered spectral patterns:

- w_{fr}: spectral pattern of the source r.
- $h_{rt}(f)$: time-varying filter of the source r at time t.

Efficient decomposition:

- A single atom for a single sound element.
- Spectral envelope variation (resonance) finely modeled.

Contents

Harmonic atoms [Hennequin et al., 2010]

Model inspired by sinusoidal additive synthesis

- Most of (non-percussive) elements of a musical spectrogram are instrument tones which correspond to harmonic patterns.
- Parameters of interest are generally the fundamental frequency of these tones, and the shape of the amplitudes of the harmonics.
- One more time the goal is to extract what is actually redundant (in this case: the amplitude of the harmonics) from what varies over time (the fundamental frequency).
- Proposed method: parametric model of spectrogram with harmonic atoms.

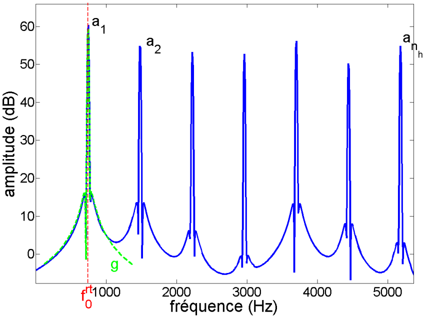
Parametric spectrogram

Time varying atoms

$$\hat{v}_{ft} = \sum_{r=1}^{R} w_{fr} h_{rt} \qquad \rightarrow \qquad \hat{v}_{ft} = \sum_{r=1}^{R} w_{fr}^{f_0^{rt}} h_{rt}$$

f_0^{rt} is the time-varying fundamental frequency of each atom.

Parametric atoms



Harmonic atom synthesis

$$w_{fr}^{f_0^{rt}} = \sum_{k=1}^{n_h(f_0^{rt})} a_k g(f - k f_0^{rt})$$

Algorithm

Parametric spectrogram

$$\hat{v}_{ft} = \sum_{r=1}^{R} \underbrace{\sum_{k=1}^{n_h} a_k g(f - k f_0^{rt})}_{w_{fr}^{f_0^{rt}}} h_{rt}$$

Learnt parameters:

Optimization with respect to the following parameters:

- f_0^{rt}: fundamental frequency of each atom at each time
- a_k: amplitudes of the harmonics
- h_{rt}: activations of each atom at each time

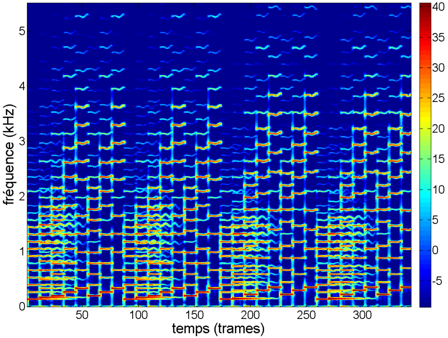
Cost function: $\mathcal{C}(f_0^{rt}, a_k, h_{rt}) = D(\mathbf{V}|\hat{\mathbf{V}})$

Algorithm

Minimization

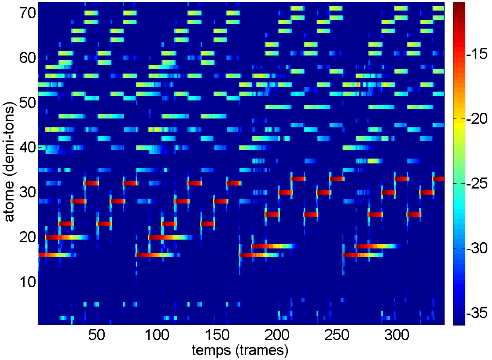
- Global optimization with respect to f_0^{rt} is impossible (numerous local minima in $\mathcal{C}$).
  $\Rightarrow$ Introduction of an atom for each note of the chromatic scale (*i.e.* for each MIDI note)
  $\Rightarrow$ Local optimization wrt f_0^{rt} (fine estimation of f_0^{rt}).
- Minimization achieved with multiplicative update rules.

Decomposition of a synthetic spectrogram



Spectrogram of the first bars of Bach's first prelude played by a synthesizer.

Obtained decomposition



Activations for each note of the chromatic scale (MIDI note).

Obtained decomposition

Decomposition

- Notes appear at the right time with decreasing amplitudes.
- Numerous atoms activated at onset time.
- Notes activated at octave, twelfth and double octave of the right note (note with many common partials).

Improvement

Onsets

Standard NMF atoms are added to the decomposition:

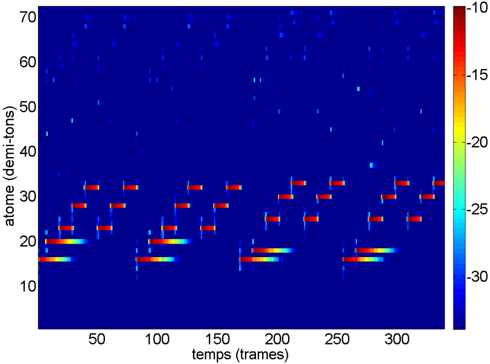
$$\hat{v}_{ft} = \sum_{r=1}^{R} w_{fr}^{f_0^{rt}} h_{rt} + \sum_{k=1}^{K} w'_{fk} h'_{kt}$$

Octave, twelfth...

Penalization is added to the cost function:

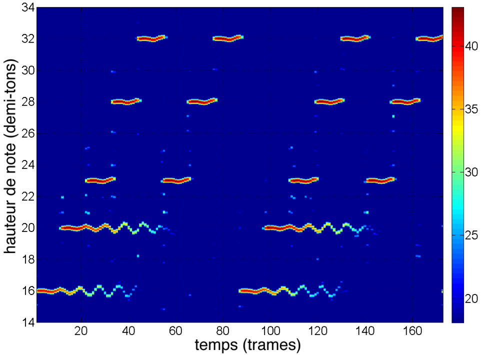
- Decorrelation constraint (on octave activations...)
- Smoothness constraint (on the amplitudes of the harmonics)
- ...

Improvement



Decomposition with right penalizations and a few standard NMF atoms.

Representation with estimated fundamental frequencies



Activations centered on estimated frequency for each MIDI note: vibrato appears.

Conclusion

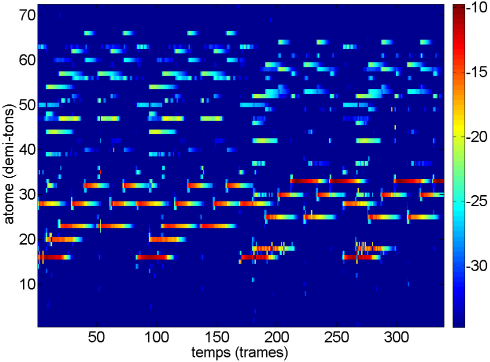
Summary

- New way of decomposing musical spectrograms with slight pitch variations in constituting elements.
- Parametric thus flexible model.

Issues

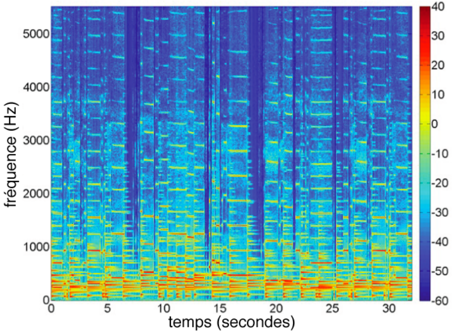
- The shape of the atoms is very constrained.
- Robustness issues to decompose real signals.

Real signal



Same piece but played by a piano.

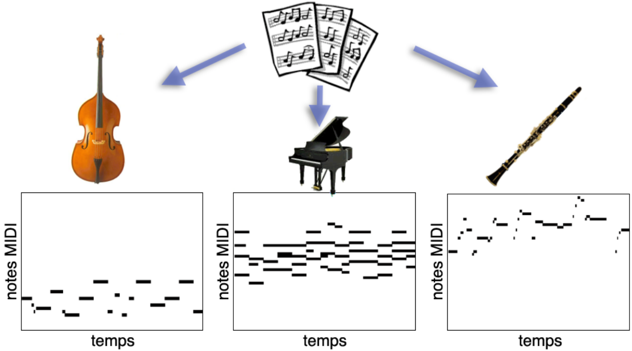
Score informed source separation [Hennequin et al., 2011c]



+

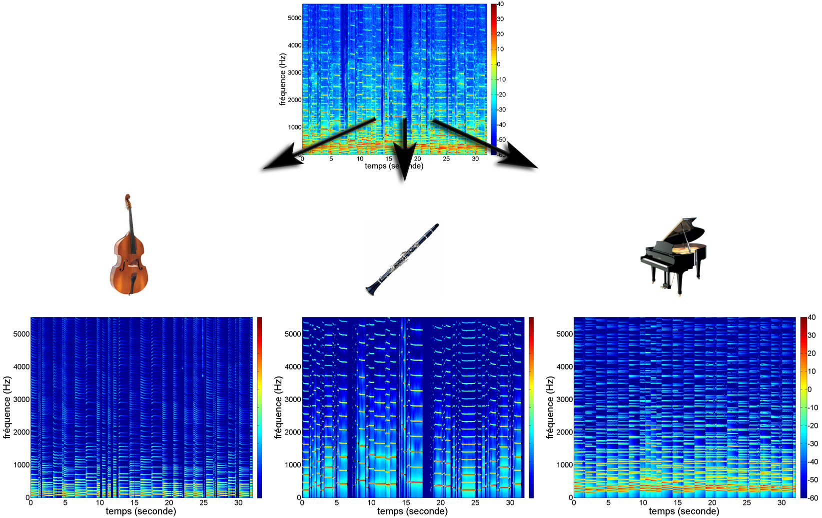
Mixture signal + Aligned score (MIDI)

Score informed source separation



The score is used to initialize (and thus constrain) the decomposition.

Score informed source separation



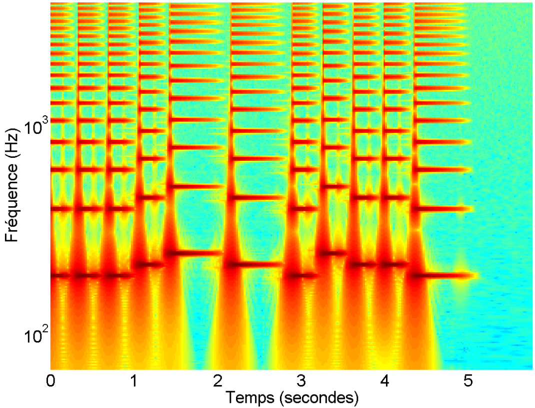
Contents

Scale-invariant decomposition

Model inspired by wavetable synthesis:

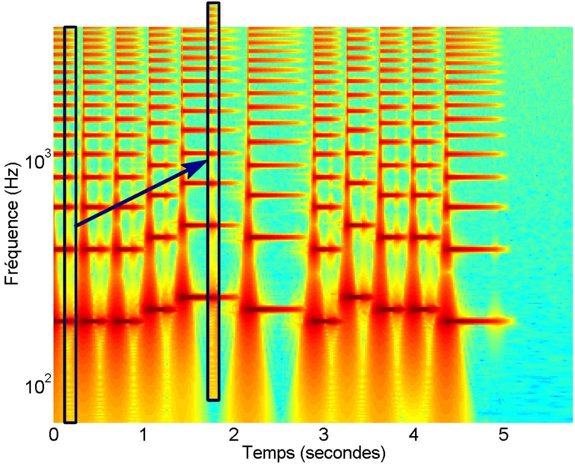
- A single atom to model all the notes of an instrument.
- Transformation of the atom to rebuild all the range of the instrument.

Shift-invariant decomposition



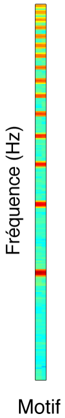
Constant-Q spectrogram of the french lullaby *Au Clair De La Lune* (played by a synthesizer).

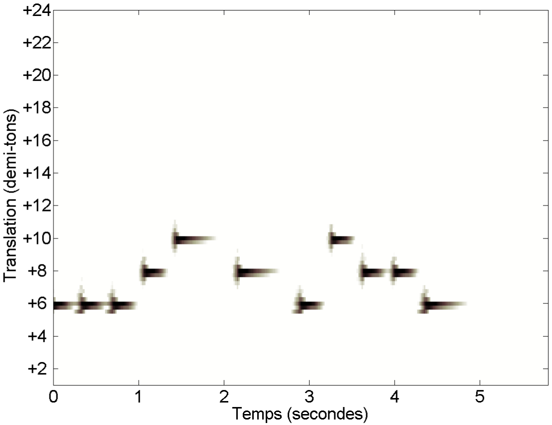
Shift-invariant decomposition



Shift-invariance in a constant-Q spectrogram.

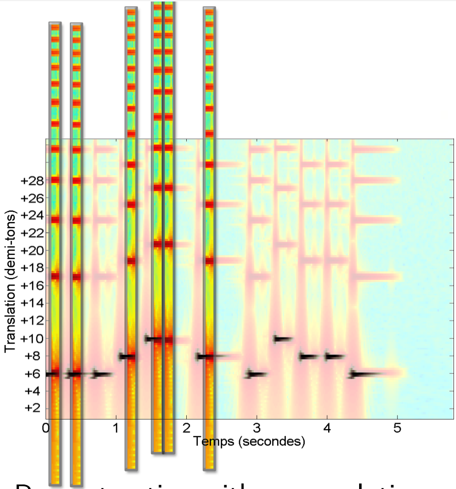
Shift-invariant decomposition





Decomposition: a single pattern and a map of the shifts (impulse distribution).

Shift-invariant decomposition



Reconstruction with a convolution

Scale-invariant decomposition [Hennequin et al., 2011b]

Goal

- Adapt shift-invariant decomposition to decompose "standard" spectrograms (with a linear frequency resolution).
- Simple and straight reconstruction of the separated components with Wiener filtering.
- The resolution of the decomposition (resolution of the homothety) is not linked to the frequency resolution of the spectrogram (in opposition to shift-invariant decomposition).

Scale-invariant decompositino [Hennequin et al., 2011b]

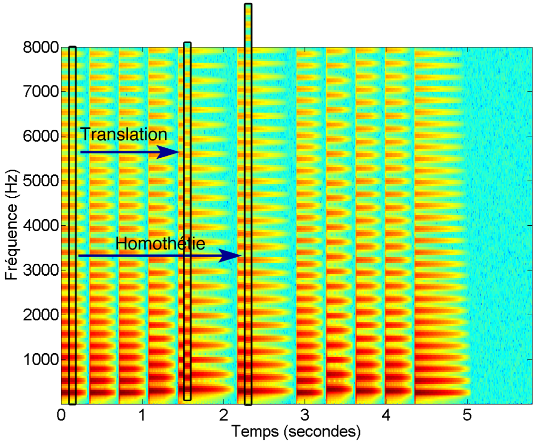
Principle

- Shift replaced by a scaling (homothety).
- New issues:
  - Partials can be compressed or dilated.
  - Non-integer scaling necessitates a continuous model.

Probabilist Latent Component Analysis (PLCA)

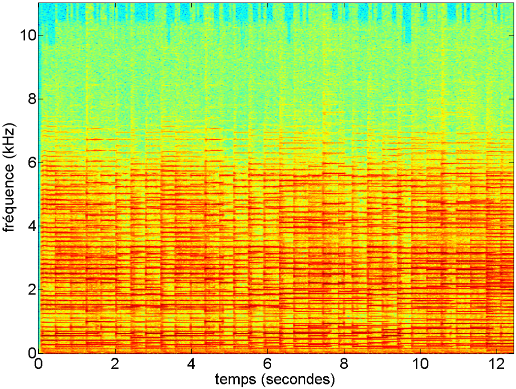
- Probabilistic variant of NMF.
- Spectrograms considered as histograms generated from a structured drawing of two random variables t (time) and f (frequency).
- Natural framework for a continuous model.

Scale-invariant decomposition



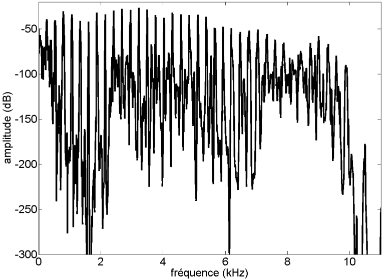
Scale-invariance in STFT spectrograms

Example

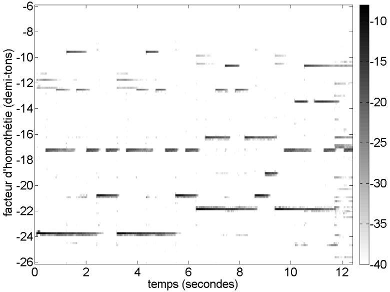


Spectrogram of the introduction of *Because* by the Beatles

Example

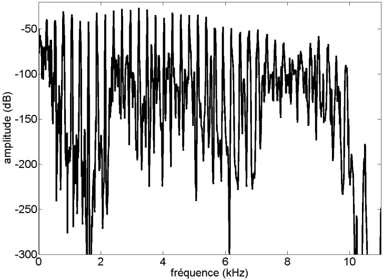


Pattern

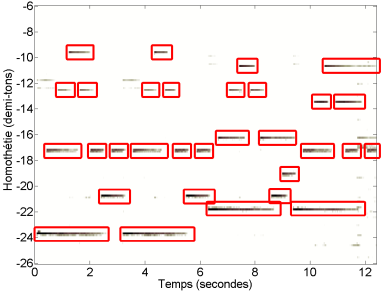


Impulse distribution

Example

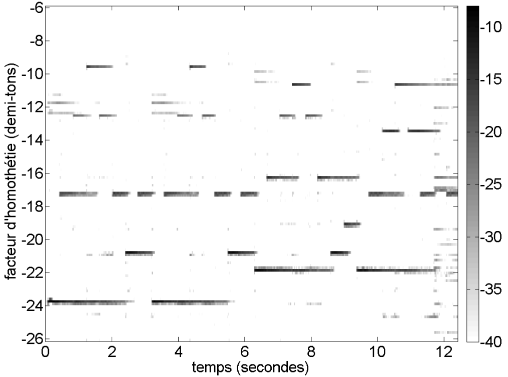


Pattern



Impulse distribution

Modification of isolated notes in a polyphonic mixture

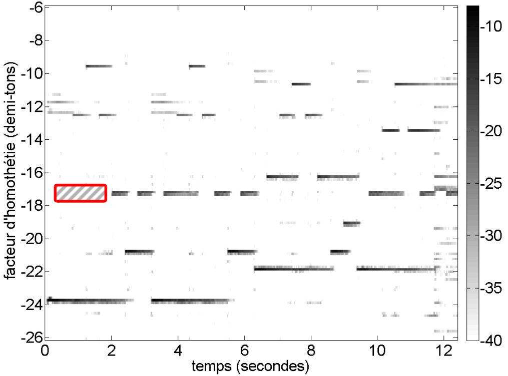


Impulse distribution of the introduction of *Because*

Modification of isolated notes in a polyphonic mixture

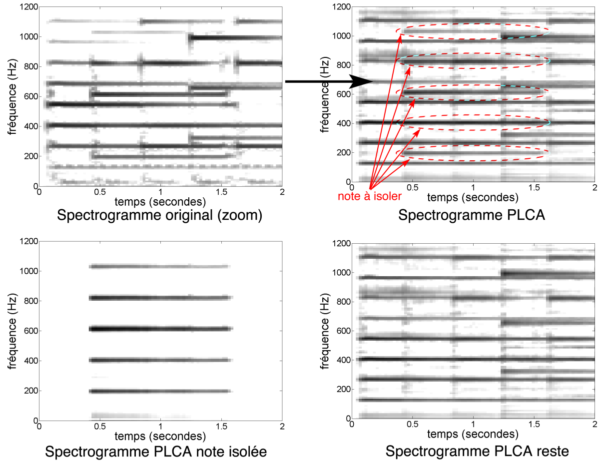
Isolated note

Modification of isolated notes in a polyphonic mixture

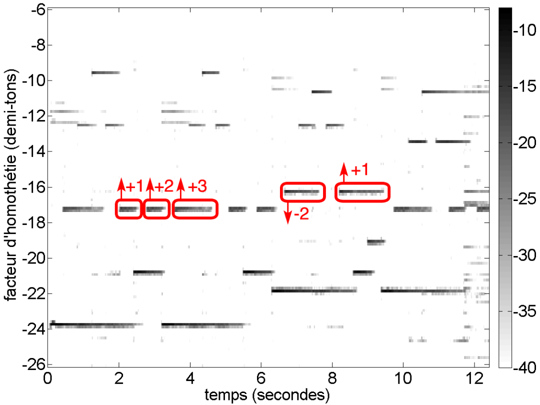


Rest

Modification of isolated notes in a polyphonic mixture



Modification of isolated notes in a polyphonic mixture



Conclusion

Scale-invariant decomposition

- Decomposition inspired by wavetable synthesis.
- More robust than the model with paramtric atoms: atoms are free and then more flexible.

Conclusion

Summarization

- Introduction of generative models of spectrograms in NMF.
- Models inspired by simple sound synthesis methods:
  - Source/filter synthesis
  - Additive synthesis
  - Wavetable synthesis
- New non-negative decomposition to model:
  - Spectral envelope variations.
  - Fundamental frequency variations.

Conclusion

Future work

- Percussive sound and onset models
- Structuring of temporal variations
- Model order estimation
- How to evaluate a decomposition/representation (with no application)

Conclusion

Thank you for your attention.

Questions?

Cichocki, A., Zdunek, R., and ichi Amari, S. (2006).
New algorithms for nonnegative matrix factorization in applications to blind source separation.
In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 621 – 625, Toulouse, France.

Hennequin, R., Badeau, R., and David, B. (2010).
Time-dependent parametric and harmonic templates in non-negative matrix factorization.
In *International Conference On Digital Audio Effects*, pages 246–253, Graz, Austria.

Hennequin, R., Badeau, R., and David, B. (2011a).
NMF with time-frequency activations to model nonstationary audio events.
IEEE Transactions on audio, speech, and language processing, 19(4):744–753.

Hennequin, R., Badeau, R., and David, B. (2011b).
Scale-invariant probabilistic latent component analysis.
Technical report, Telecom-ParisTech.

Hennequin, R., David, B., and Badeau, R. (2011c).
Score informed audio source separation using a parametric model of non-negative spectrogram.
In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic.

Le Roux, J., Kameoka, H., Ono, N., de Cheveigné, A., and Sagayama, S. (2008).
Computational auditory induction by missing-data non-negative matrix factorization.
In *ITRW on Statistical and Perceptual Audio Processing*, Brisbane, Australia.

Smaragdis, P. and Brown, J. C. (2003).
Non-negative matrix factorization for polyphonic music transcription.

In *Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 177 – 180, New Paltz, NY, USA.