TIME-DEPENDENT PARAMETRIC AND HARMONIC TEMPLATES IN NON-NEGATIVE MATRIX FACTORIZATION

13th International Conference on Digital Audio Effects

Romain Hennequin, Roland Badeau and Bertrand David

Telecom ParisTech

September 8, 2010

Introduction

- Musical spectrograms decomposition (on a basis of notes)
- Decomposition based on Non-negative Matrix Factorization (NMF)
- Spectrogram models are introduced into decomposition methods:
  - parametric harmonic atoms
  - makes it possible to model slight pitch variations
- Potential applications:
  - Multipitch estimation/transcription
  - Source separation

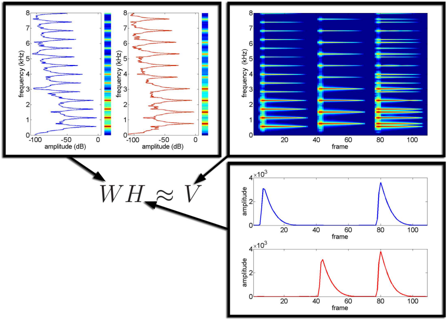
Sommaire

1. Non negative-Matrix Factorization
2. Spectrogram model
3. Decomposition of musical spectrograms

Contents

1. Non negative-Matrix Factorization
   - Principle
   - Issues
   - Proposed solution
2. Spectrogram model
3. Decomposition of musical spectrograms

Principle of NMF



Low-rank approximation:

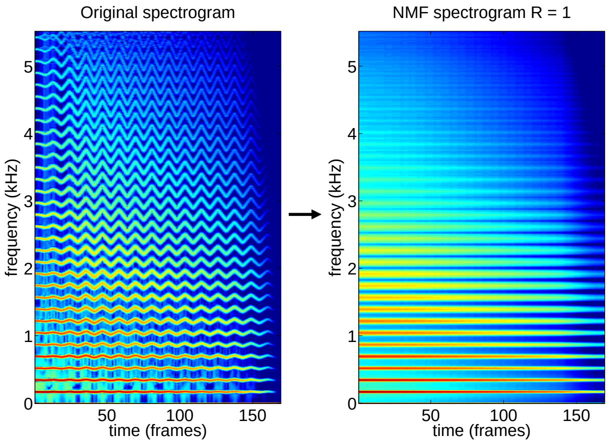
$$\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{WH} \qquad \hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr} \mathbf{H}_{rt}$$

Issues with NMF

Pitch variations

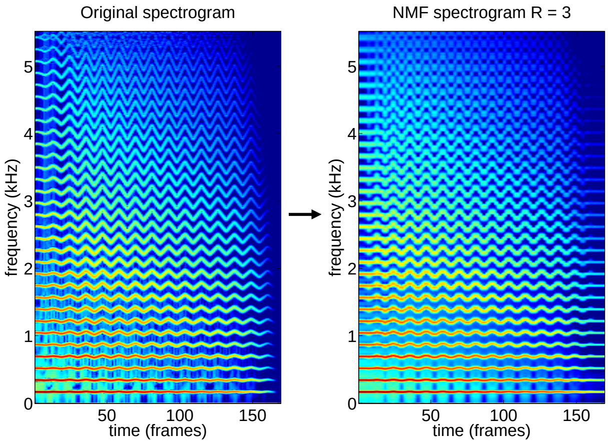
Low-rank approximation does not permit to model variations over time, such as slight pitch variations (vibrato. . .).

Issues with NMF



Note with vibrato: Decomposition with a single atom.

Issues with NMF



Note with vibrato: Decomposition with 3 atoms.

Proposed solution

What does an atom look like in a musical spectrogram?

- In a musical spectrogram most of the (non-percussive) elements are instruments notes which are generally harmonic tones.
- Parameters of interest are generally the fundamental frequency of these tones, and the shape of the amplitudes of the harmonics.
- Proposed method: parametric model of spectrogram with harmonic atoms.

Contents

1 Non negative-Matrix Factorization

2 Spectrogram model
- Parametric spectrogram
- Parametric atoms
- Algorithm

3 Decomposition of musical spectrograms

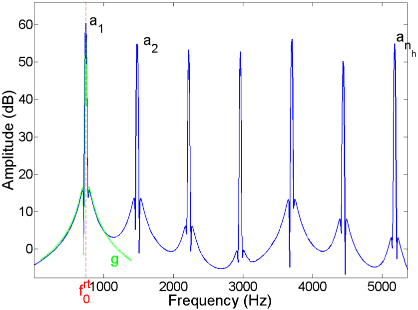
Parametric spectrogram

Time-varying atoms in NMF:

$$\hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr} \mathbf{H}_{rt} \quad \rightarrow \quad \hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr}^{\theta^{rt}} \mathbf{H}_{rt}$$

θ^{rt} is a time-varying parameter associated to each atom. In this paper, θ^{rt} is the fundamental frequency f_0^{rt} of each atom.

Parametric atoms



Parametric harmonic atom construction

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

Parametric spectrogram

Hypotheses of the model

- The harmonic part of notes is supposed to be stationary within an analysis frame.
- Interferences between harmonics are supposed to be negligible.
- Classical hypothesis of NMF about positive summation of parts.

Algorithm

Parametric spectrogram

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

Learnt parameters

A divergence between $\mathbf{V}$ and $\hat{\mathbf{V}}$ is to be minimized w.r.t.:

- f_0^{rt}: the fundamental frequency of each atom at each frame
- a_k: the amplitudes of harmonics (Atoms share the same set of amplitudes)
- h_{rt}: the activation of each atom at each frame

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

Algorithm

Minimization

- Global optimization w.r.t. f_0^{rt} is impossible (numerous local minima in $\mathcal{C}$). $\Rightarrow$ one atom is introduced for each MIDI note. Optimization thus becomes local (fine estimate of f_0^{rt}).
- Minimization achieved with multiplicative update rules.

Remark

The proposed method is no longer a rank-reduction method but still reduces the data dimension.

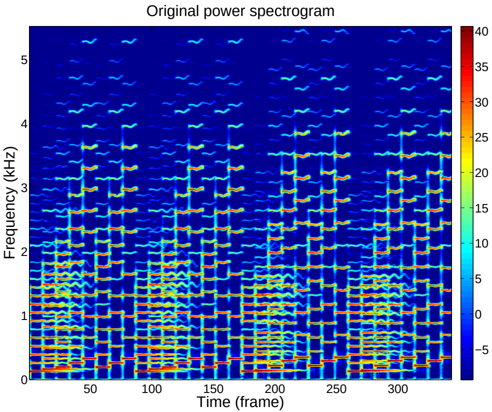
Contents

1 Non negative-Matrix Factorization

2 Spectrogram model

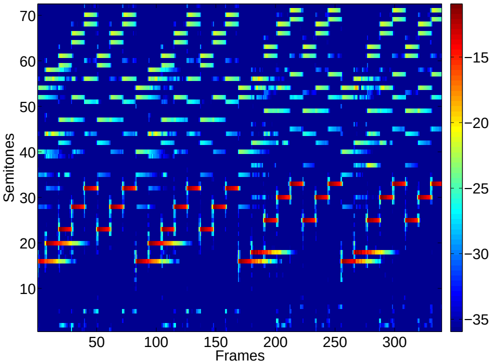
3 Decomposition of musical spectrograms
- Decomposition
- Improvement
- Estimated frequency
- Real signals

Decomposition of a synthetic spectrogram



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

Obtained decomposition



Activations for each MIDI note.

Obtained decomposition

Decomposition

- Notes appear at the right place 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

Onset

A few standard NMF atoms can be used to model onsets:

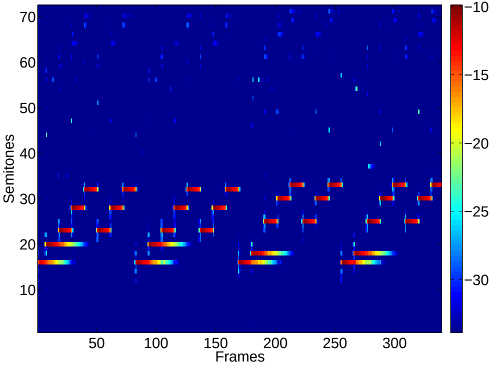
$$\hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr}^{\theta^{rt}} \mathbf{H}_{rt} + \sum_{k=1}^{K} \mathbf{A}_{fk} \mathbf{B}_{kt}$$

Octaves, twelfths...

Add constraints to the cost function:

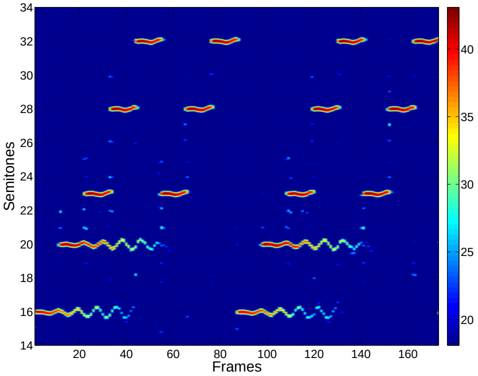
- Sparsity constraints on activations
- Decorrelation constraints (between activations of octaves...)
- Smoothness constraints on amplitudes

Obtained decomposition



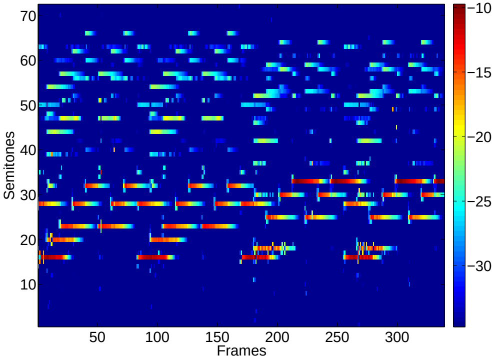
Activations for each MIDI note.

Time/frequency representation



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

Issues with real signals



Activations for each MIDI note. (Piano sound)

Issues with real signals

Issues

- The model of amplitudes of harmonics is quite rough
- Issues with onsets and octaves are more important
- Noisy components (breath. . .)
- Some instruments are not perfectly harmonic (piano. . .)

Conclusion

Summary

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

Perspectives

Improve decomposition to make it more adapted to real data:

- Better modeling of harmonic amplitudes
- Supervised learning of amplitudes
- Better onset and noise modeling

Conclusion

Any questions?