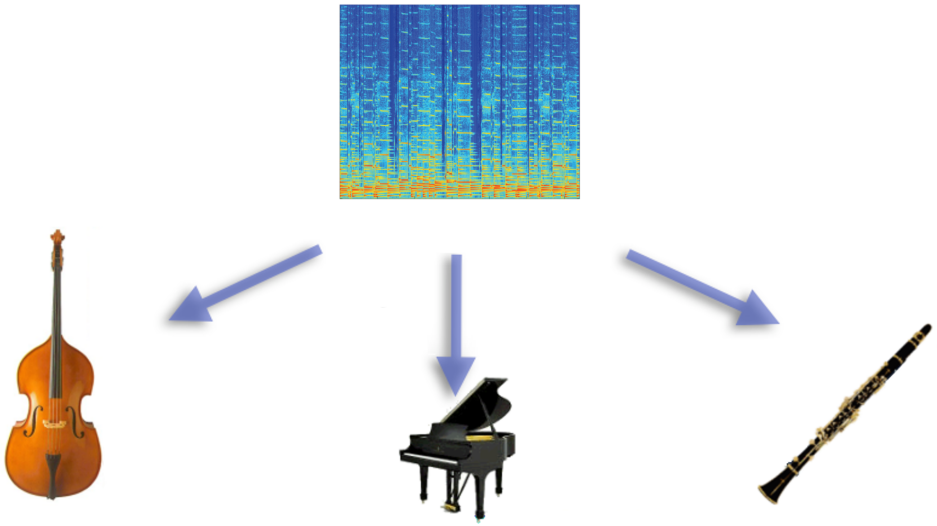
Score informed audio source separation using a parametric model of non-negative spectrogram

Romain Hennequin, Roland Badeau and Bertrand David

Telecom ParisTech
<forename>.<surname>@telecom-paristech.fr

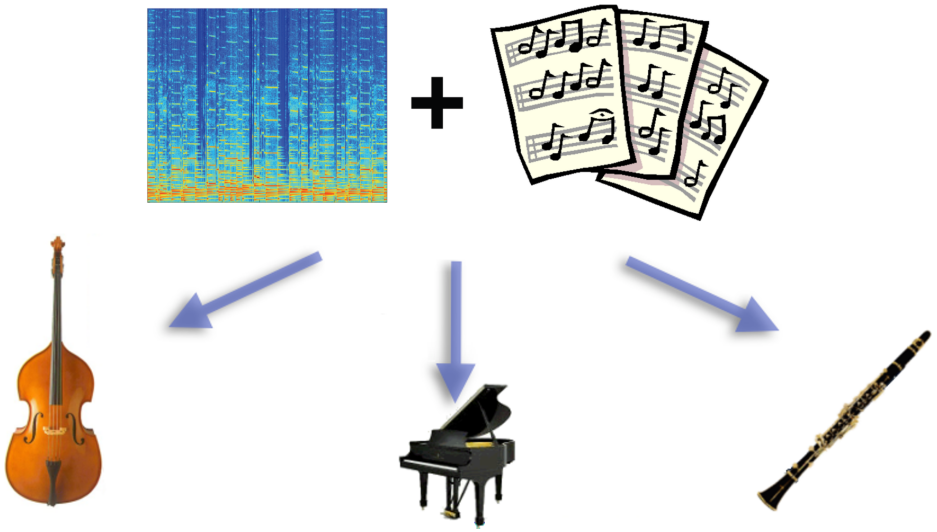
May 24, 2011

Introduction



Monaural source separation in a musical signal: separation of the signal of each instrument.

Introduction



- The information in the score of the piece is used to guide the separation.
- The score is here a MIDI file aligned on the signal (this paper does not deal with alignment).
- Only harmonic instruments are modeled.

Introduction

Overview

- Parametric spectrogram model derived from non-negative matrix factorization (NMF) to decompose the mixture spectrogram.
- A parametric time/frequency mask is computed for each instrument.
- Masks are initialized (and constrained) from the score and then finely estimated to fit the mixture spectrogram.
- Masks are used to separate the instruments (Wiener filtering).

Introduction

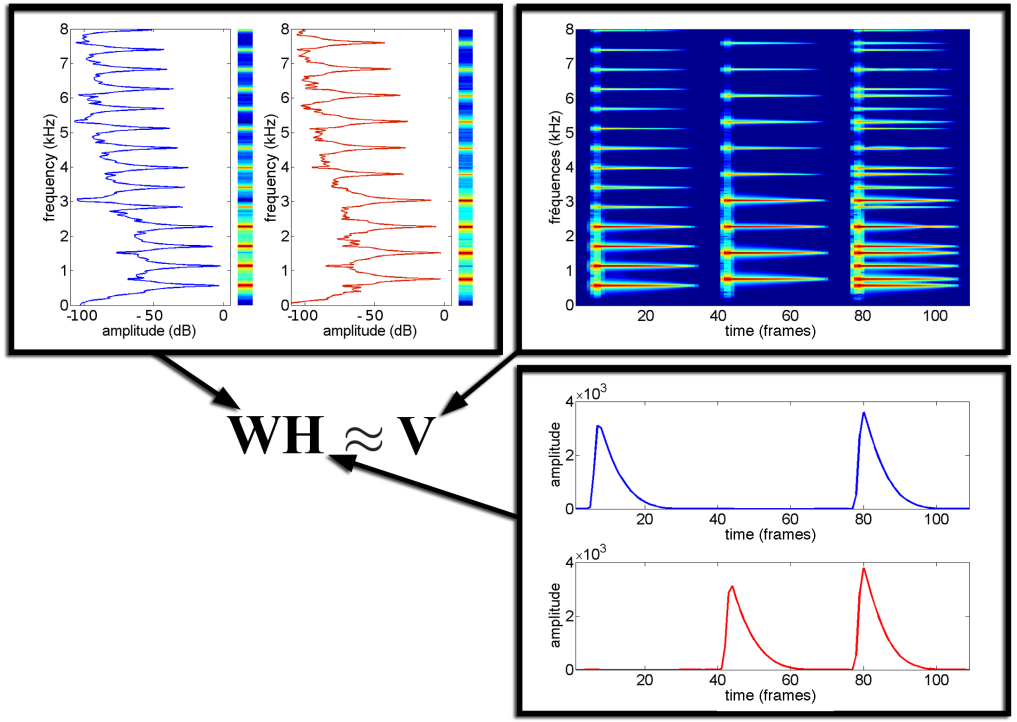
Why use the score?

- MIDI files widely available.
- Very compact description of the audio.
- Under determined blind separation remains a very difficult problem.
- Sometimes, blind separation is hopeless (separation of several voices played by the same instrument).

Outline

Outline

Principle of NMF



Low-rank approximation:

$$\mathbf{V}_{ft} \approx \hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr}\mathbf{H}_{rt} \text{ with } \mathbf{W} \geq 0, \mathbf{H} \geq 0, R \ll \min(F, T)$$

Principle of NMF

Features

- Extract redundant patterns from the data.
- Fundamental property: non-negativity constraint.
  - Atoms lie in the same space as the data.
  - Only positive combinations (no black energy).
  - Perceptive description: decomposition of musical spectrograms on a basis of notes.
- Application in automatic transcription, source separation, audio inpainting...

Principle of NMF

Limitations

- Does not permit to deal with time-frequency variations (vibrato)
- We needed a representation linked with parameters of interest (fundamental frequency)

Outline

Parametric spectrogram of a single instrument [Hennequin et al., DAFx 2010]

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.

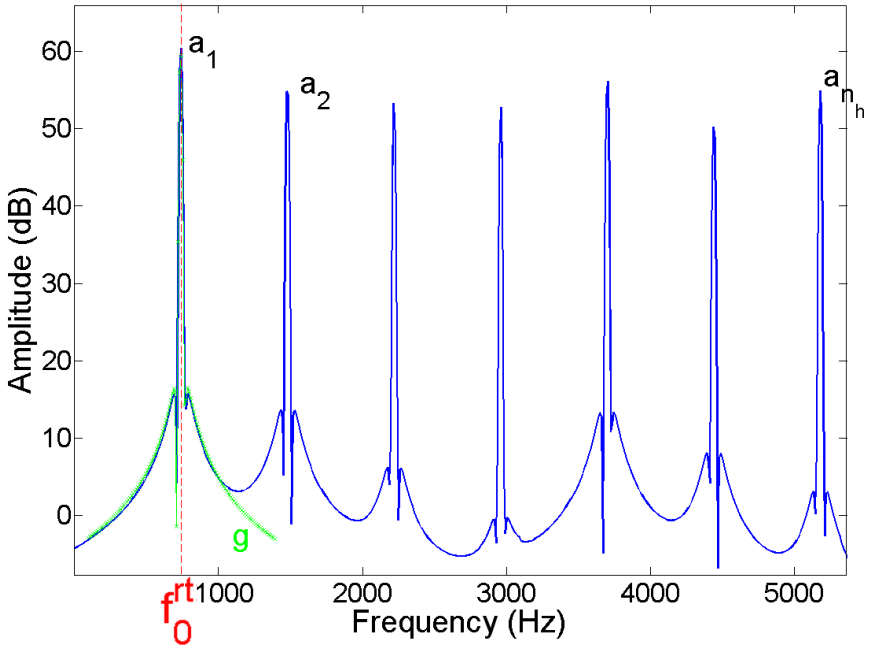
Parametric spectrogram of a single instrument

Time-varying atoms in NMF:

$$\hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr}\mathbf{H}_{rt} \quad \rightarrow \quad \hat{\mathbf{V}}_{ft} = \sum_{r=1}^{R} \mathbf{W}_{fr}^{f_0^{rt}}\mathbf{H}_{rt}$$

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

Parametric atoms



Parametric harmonic atom construction

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

Algorithm

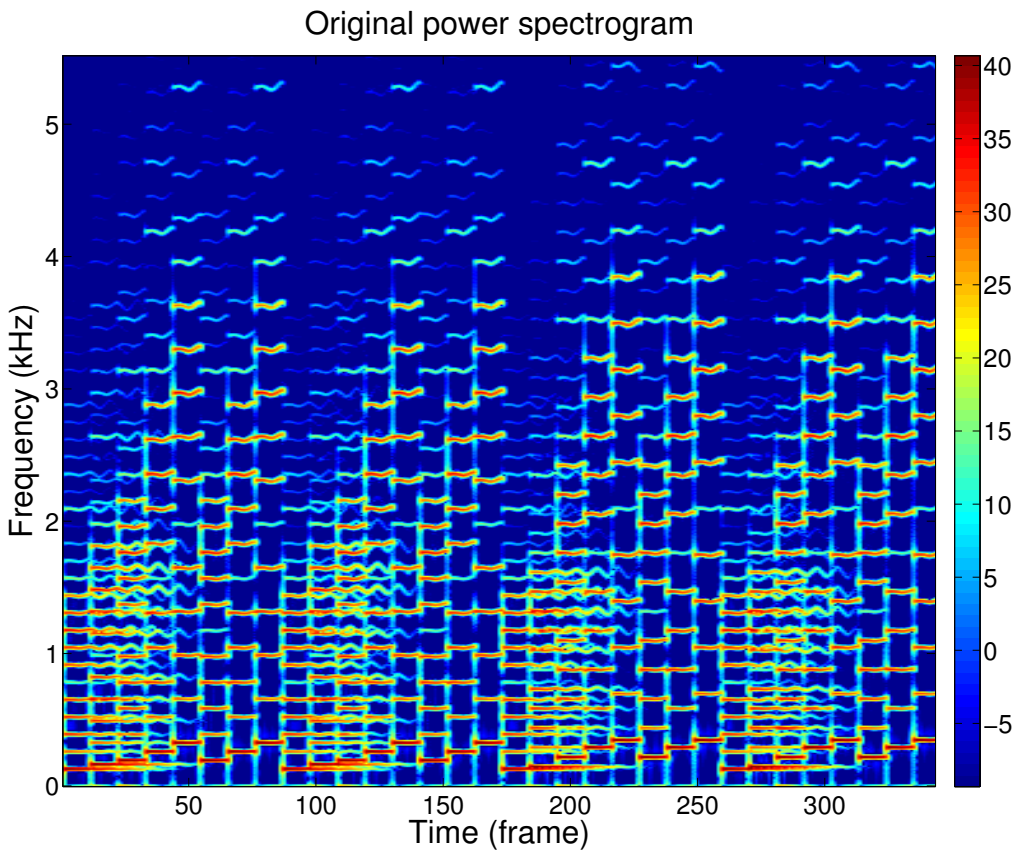
Parametric spectrogram (single instrument)

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

Minimization

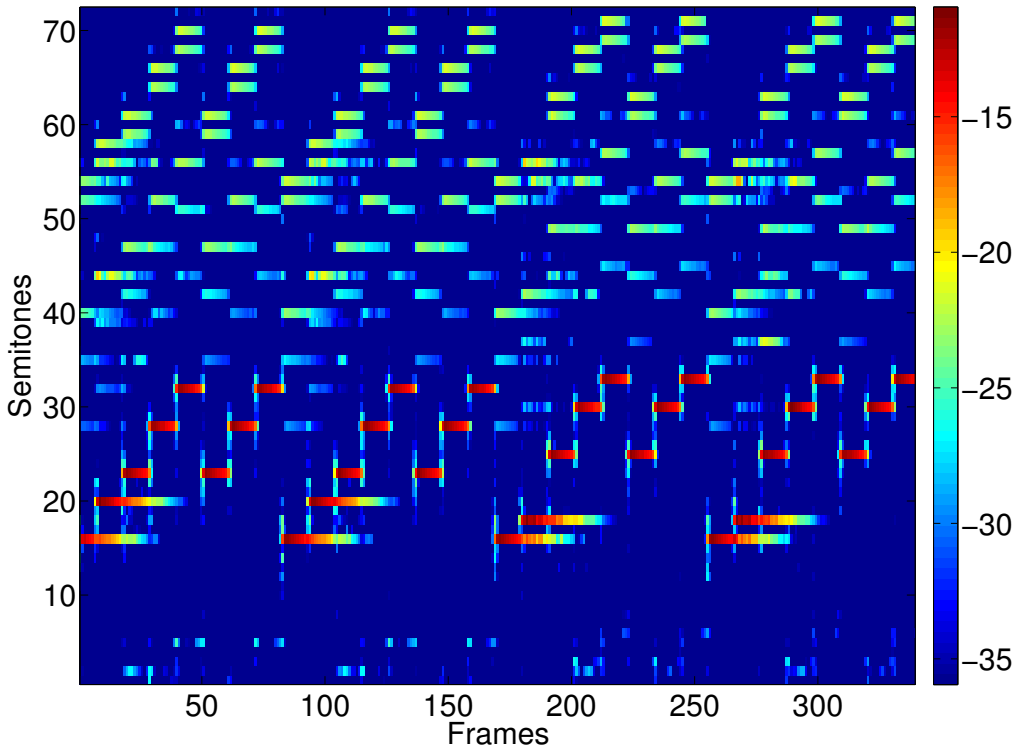
Global optimization w.r.t. f_0^{rt} is impossible (numerous local minima in cost function). $\Rightarrow$ one atom is introduced for each MIDI note. Optimization thus becomes local (fine estimate of f_0^{rt}).

Decomposition of a synthetic spectrogram



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

Obtained decomposition



Activations h_{rt} for each MIDI note.

Mixture spectrogram model

Mixture model

- The mixture is made up of K sources indexed by k. Source k is modelized with spectrogram $\hat{\mathbf{V}}^k$ following:

$$\hat{\mathbf{V}}^k_{ft} = \sum_{r=1}^{R} \underbrace{\sum_{p=1}^{n_h} a_{kp} g(f - p f_0^{krt})}_{\mathbf{W}^{f_0^{krt}}_{kfr}} h_{krt}$$

- Mixture spectrogram is then:

$$\hat{\mathbf{V}}^{\text{mix}} = \sum_{k=1}^{K} \hat{\mathbf{V}}^k$$

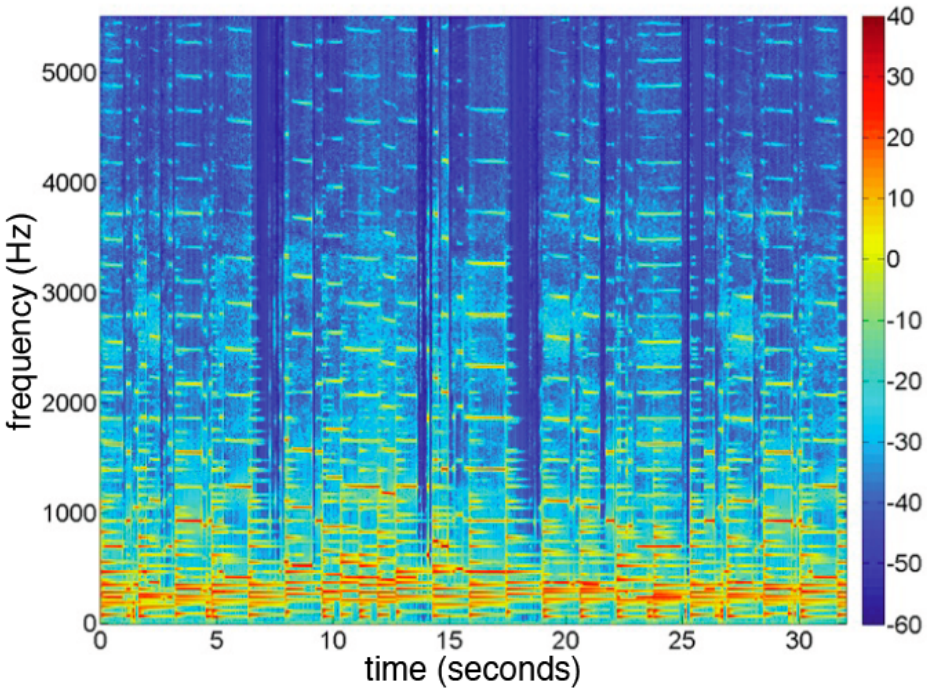
Mixture spectrogram model

Mixture model

- Parameters to be estimated for each source k:
  - Fundamental frequency of each atom r at each time t: f_0^{krt},
  - Amplitudes of harmonics: a_{kp},
  - Activations of each note r at each time t: h_{krt}.
- Decomposition obtained with a multiplicative algorithm aiming at minimizing a β-divergence between $\mathbf{V}^{\text{mix}}$ and $\hat{\mathbf{V}}^{\text{mix}}$.

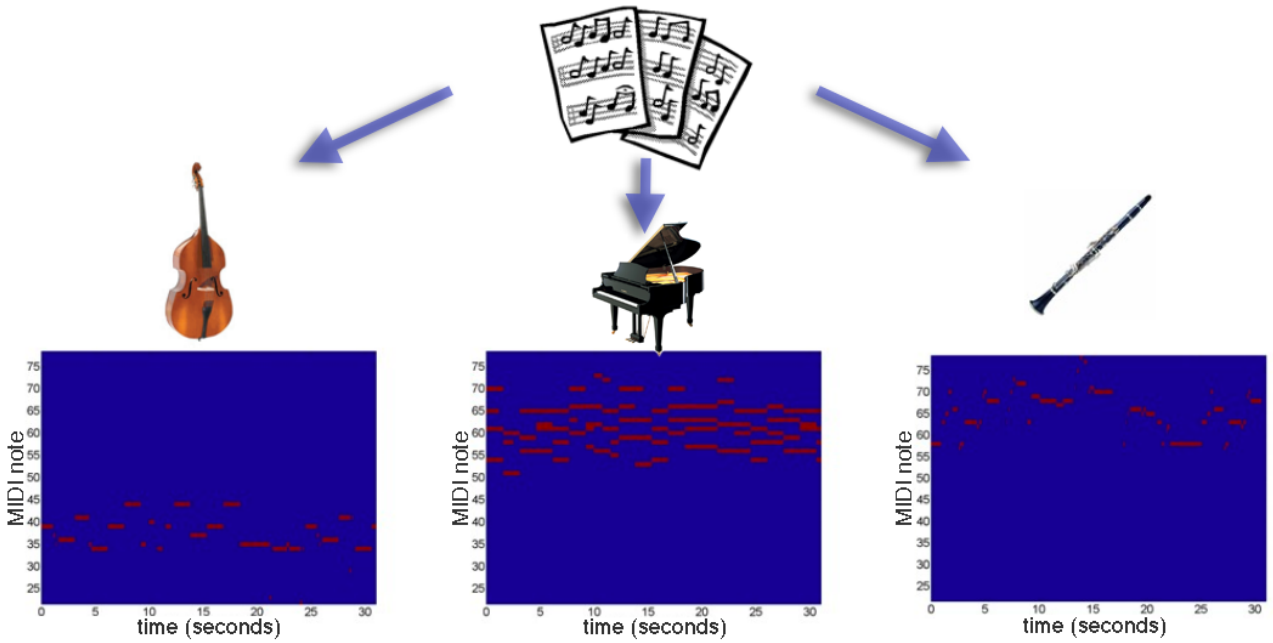
Outline

Score informed source separation



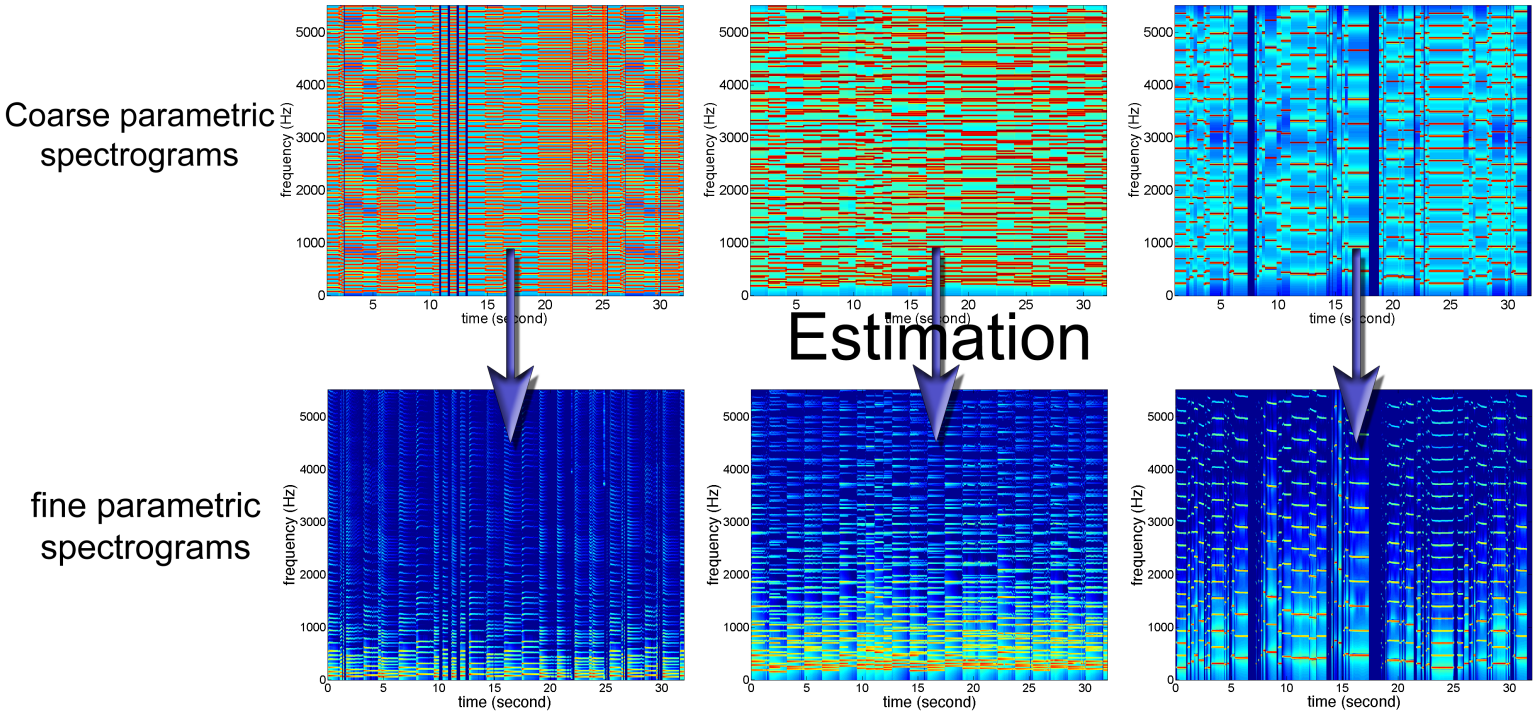
+

Score informed source separation



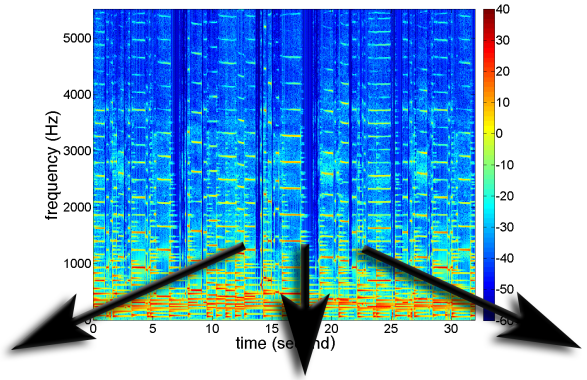
- MIDI file: notes positions and durations for each instrument.
- A piano-roll is built for each instrument.
- Piano-rolls are used to initialize and constrain activations h_{krt}.

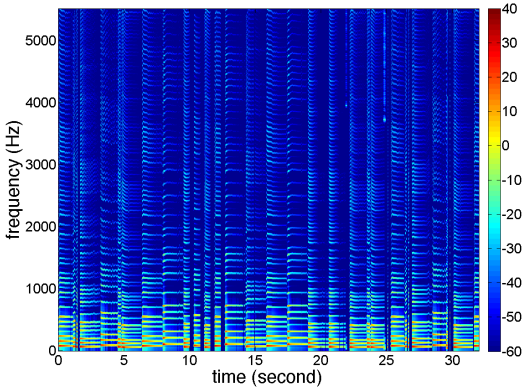
Score informed source separation

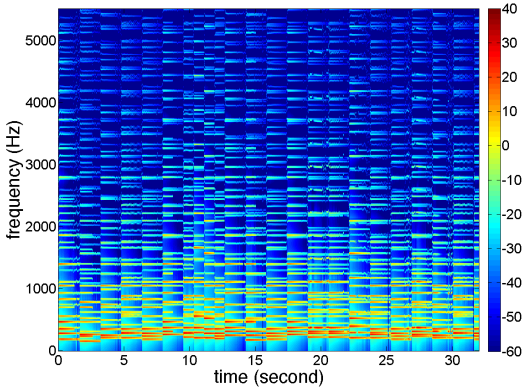


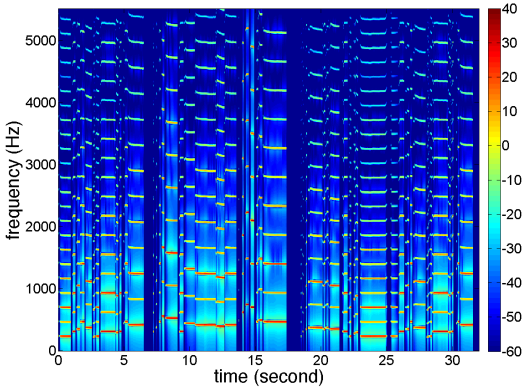
- $\hat{\mathbf{V}}_k$ are finely estimated from the actual mixture spectrogram.
- $\hat{\mathbf{V}}_k$ are used as time-frequency masks to separate the tracks (Wiener filtering).

Sound example



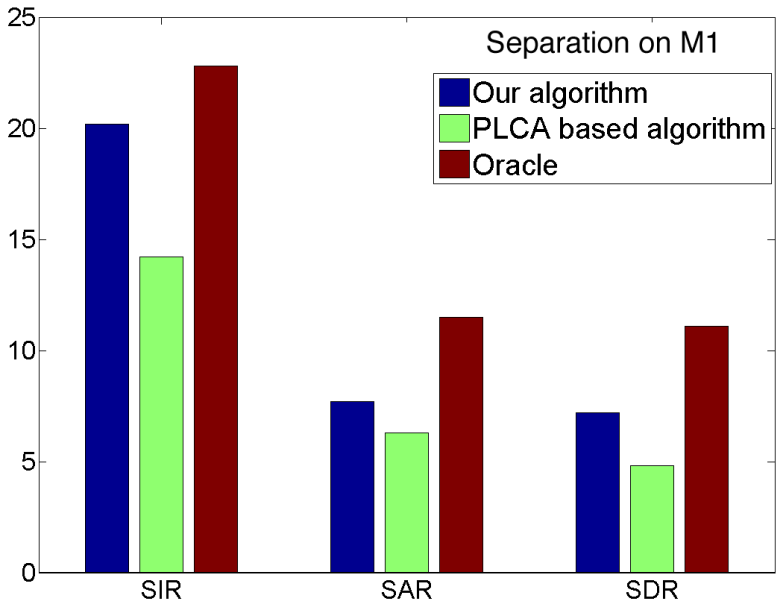


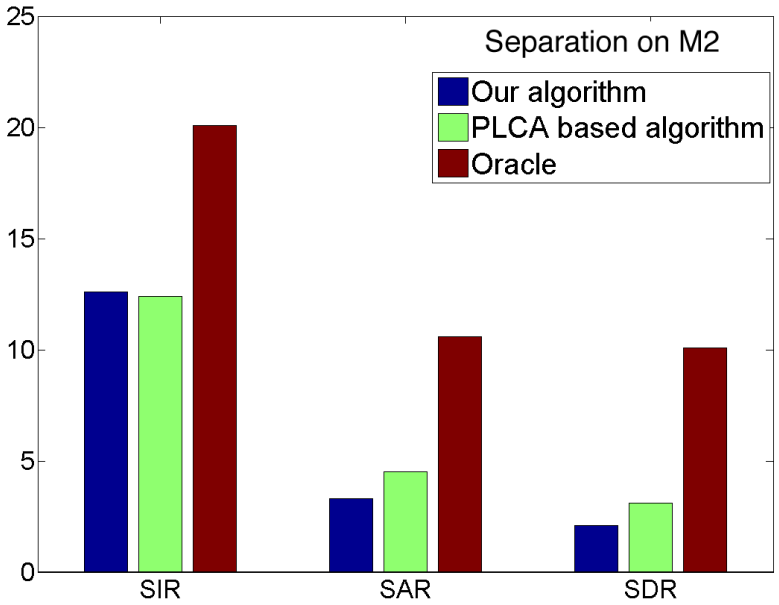




Experiment

- Comparison with a PLCA-based algorithm [Ganseman et al., ICMC 2010].
- Needs to synthesize MIDI tracks.
- Two datasets M1 and M2: same MIDI, different soundbanks.





SIR: Source to Interferences Ratio - SAR: Source to artifacts Ratio - SDR: Source to Distorsion Ratio

Conclusion

Synthesis

- Efficient method of score informed source separation.
- Parametric model: allows fine handling of the sound.

Perspectives

- Model of percussive instruments.
- Including other timbral parameters in the spectrogram model.
- Make the model more robust.
- Supervised learning of harmonic templates.

Conclusion

Questions?

Ganseman, J., Scheunders, P., Mysore, G. J., and Abel, J. S. (2010).
Source separation by score synthesis.
In *International Computer Music Conference*, New York, NY, USA.

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.