Decomposition of musical spectrograms informed by spectral synthesis models

—

Modeling of time variations in sound elements

Romain Hennequin

Telecom ParisTech

21 november 2011

Decomposition of musical spectrograms informed by spectral synthesis models
—
Modeling of time variations in sound elements

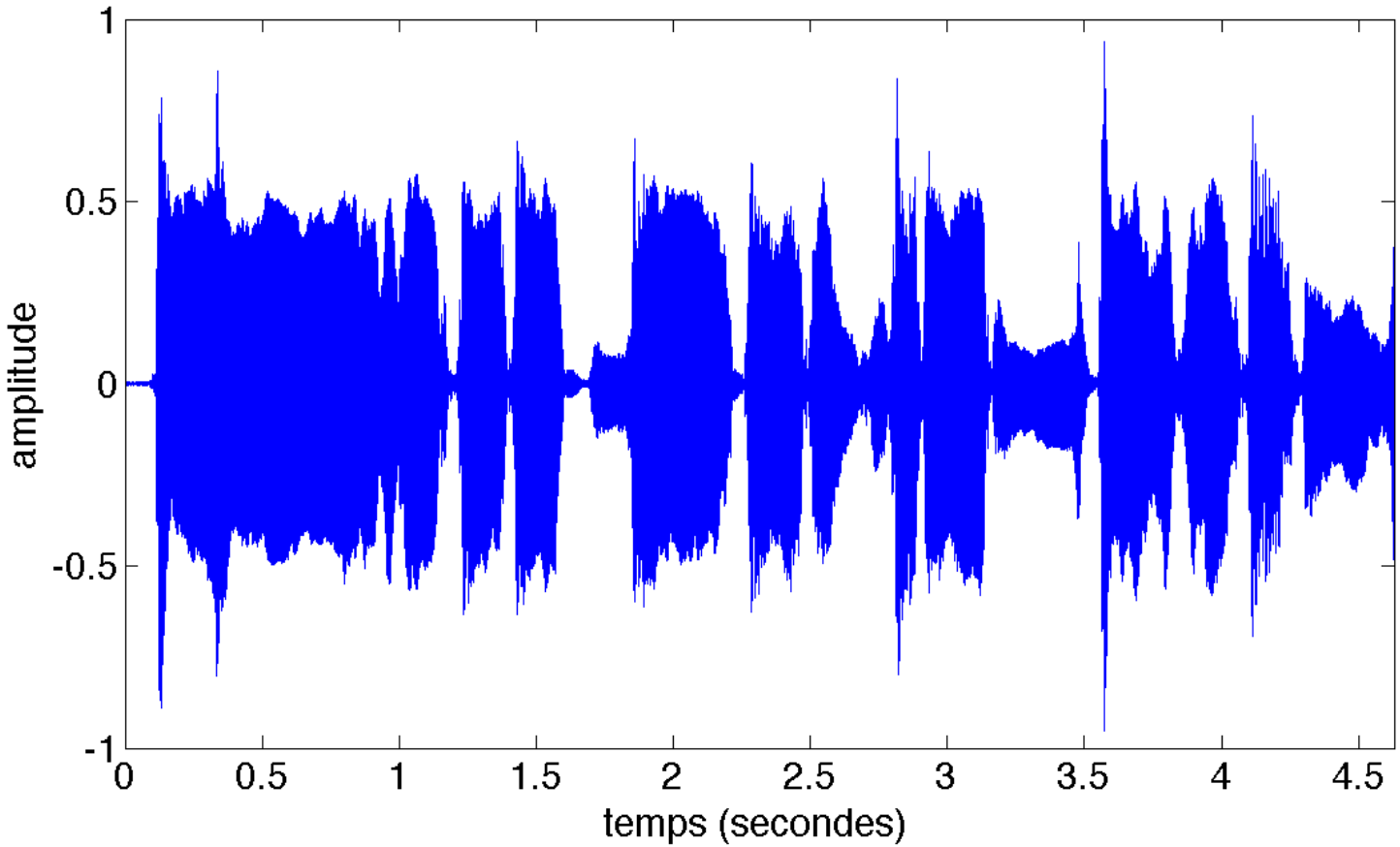
Romain Hennequin
Telecom ParisTech

21 november 2011

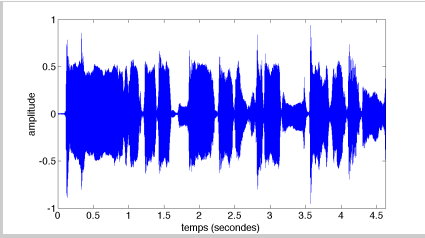
The labels of the figure are unfortunately not translated in english, so here is a small french lexicon to understand them (most of the words are transparent):
Transparent ones:
- temps = time
- secondes = seconds
- fréquence = frequency
- spectrogramme = spectrogram
- spectrogramme original = original spectrogram
- spectrogramme reconstruit = reconstructed spectrogram
- filtre = filter
- trame = frame
- atome = atom
- notes MIDI = MIDI notes
- demi-tons = semitones
- translation = shift, translation

Not transparent:
- motif = pattern, template
- motif fréquentiel = frequency pattern
- hauteur de note = pitch
- facteur d'homothétie = homothety ratio, scaling ratio

Introduction



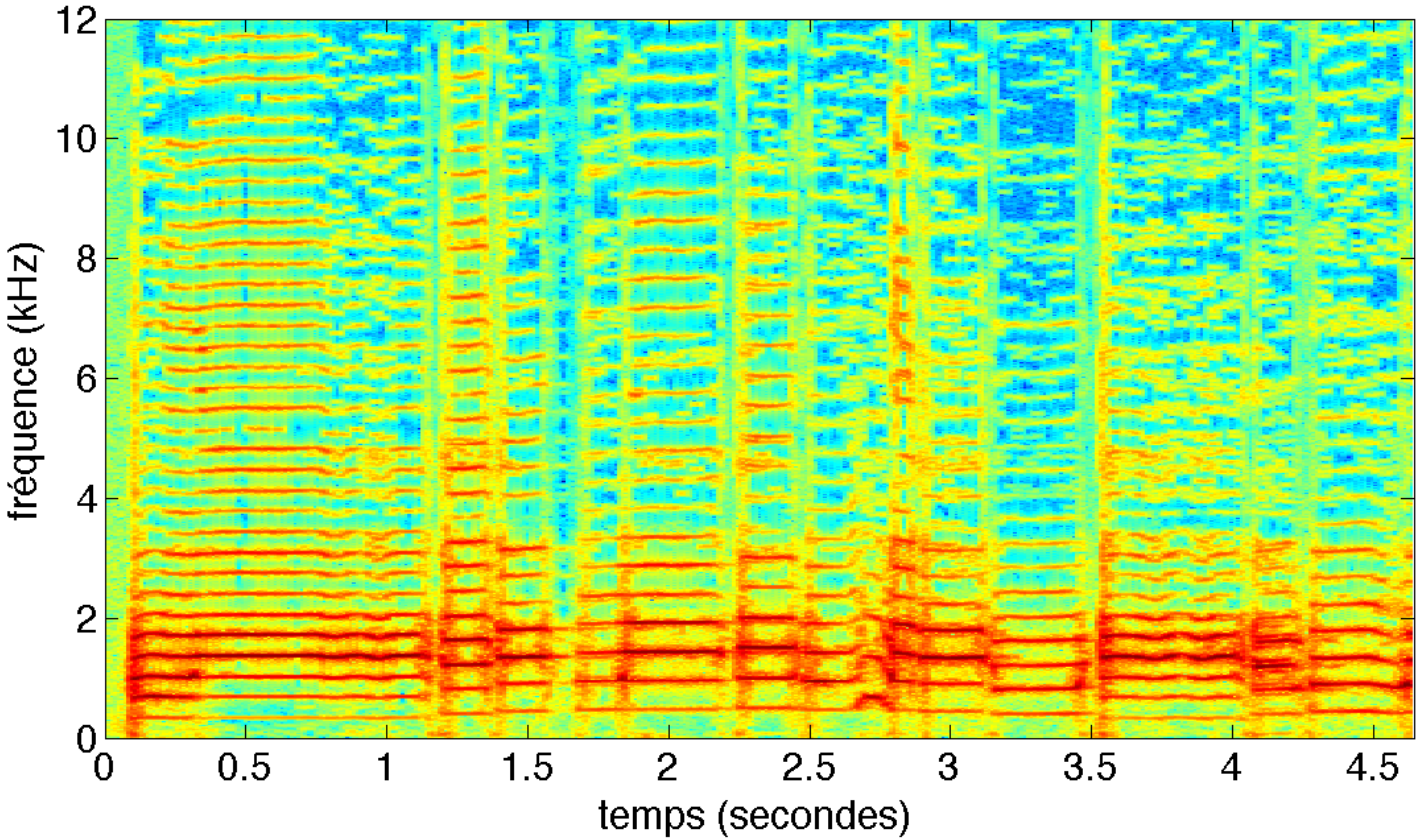
Waveform



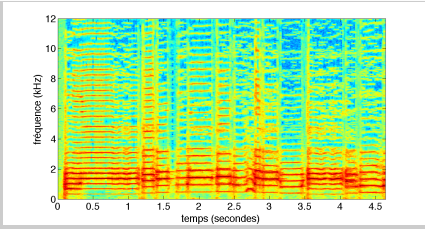
Waveform

We work with sounds. They are originally in the form of a *waveform* which physically corresponds to a pressure variation. Here is an example of such a sound (trumpet sound).

Introduction



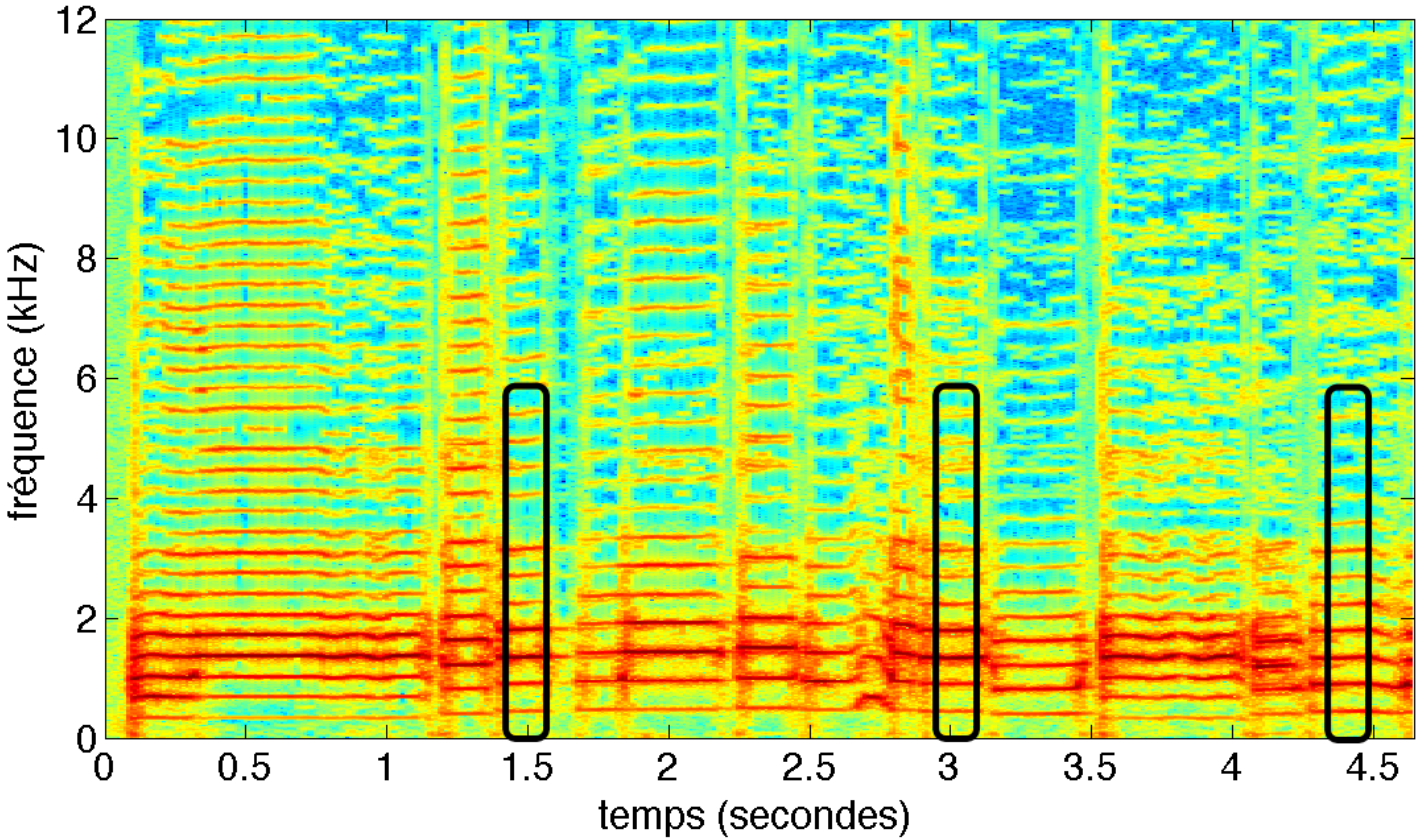
spectrogram



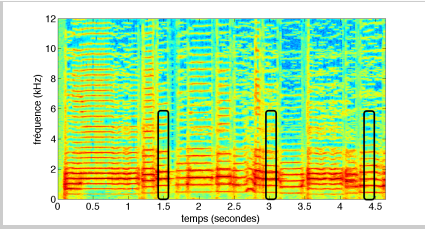
spectrogram

As the waveform is usually not very informative (we cannot see much directly in the waveform), we generally transform it to a time/frequency representation called spectrogram by means of a Short Time Fourier Transform.

Introduction



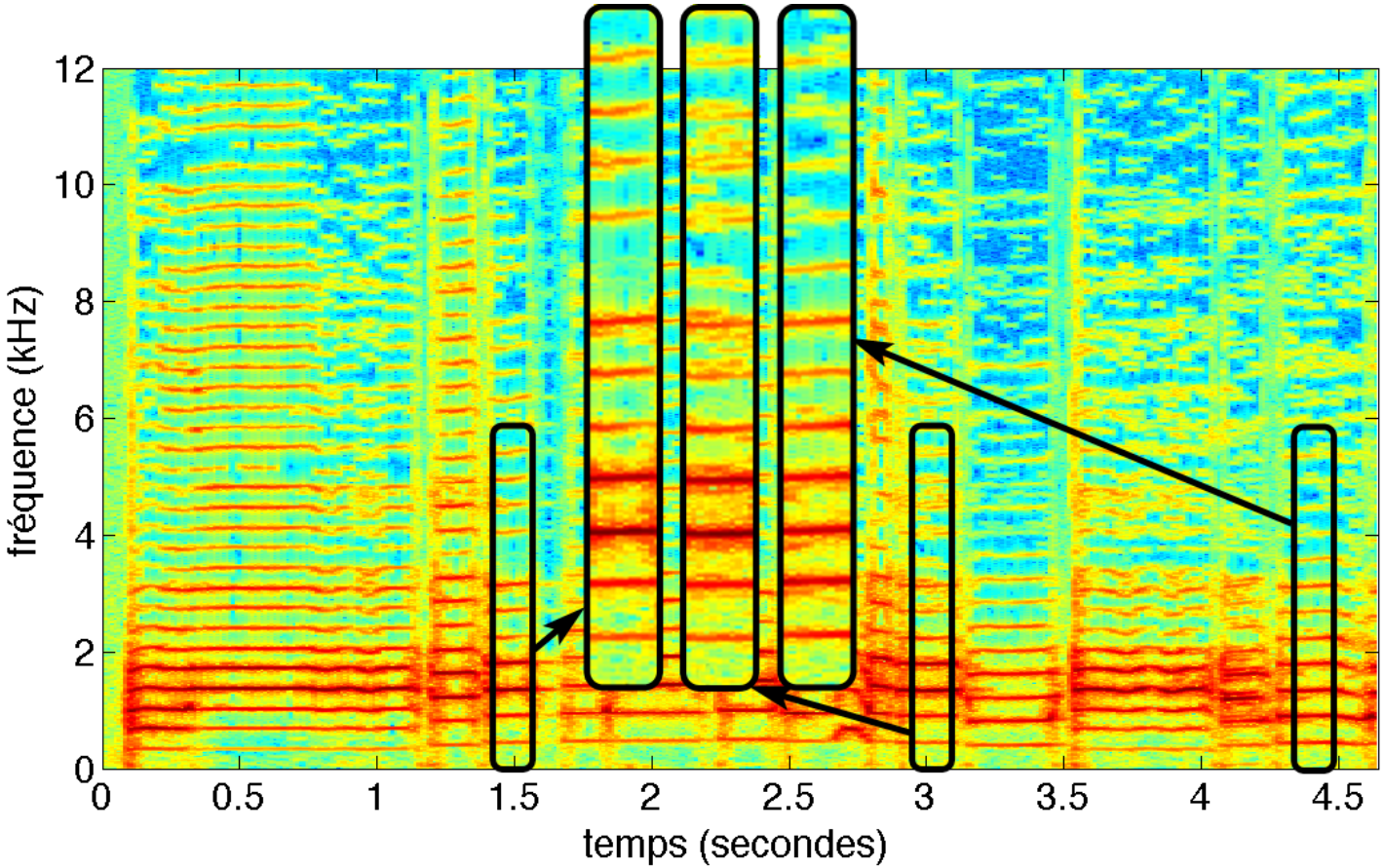
spectrogram



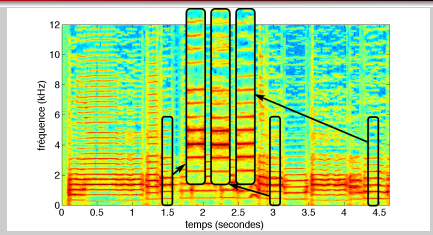
spectrogram

This kind of representation highlights redundancies.

Introduction



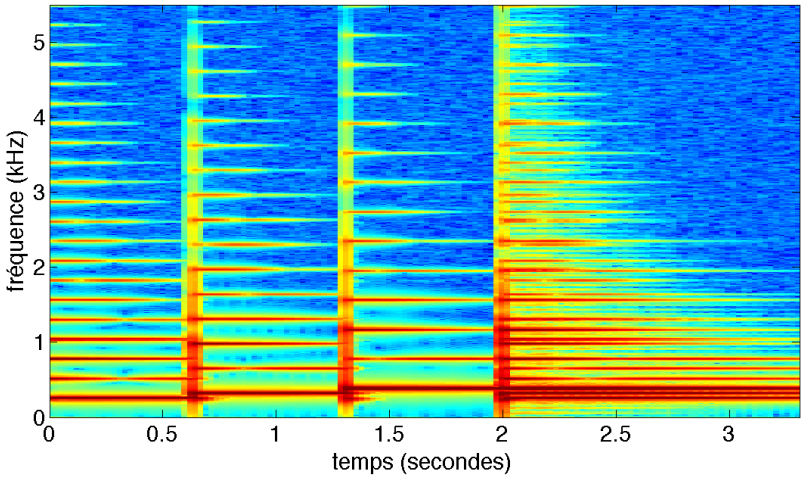
spectrogram



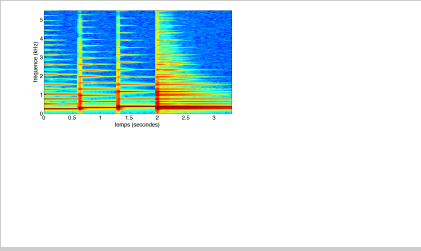
spectrogram

For instance, we can notice that the three selected parts of the spectrogram are very similar (since they correspond to the same note). The human perception of the sounds is partly based on these redundancies. We work on automatic decompositions of (non-negative) spectrograms based on these redundancies, thus trying to mimic the human perception.

Redundancy extraction: Non-negative Matrix Factorization (NMF)

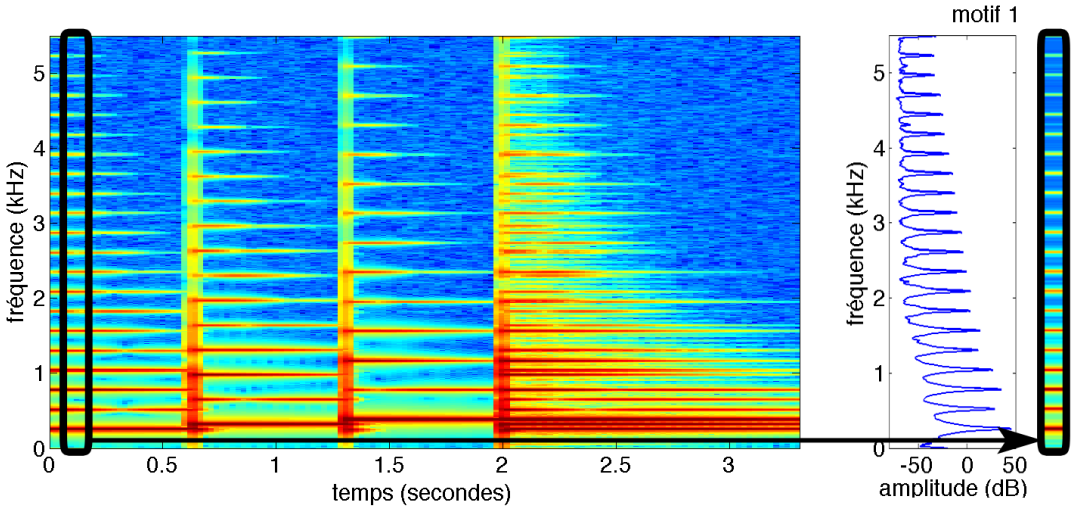


Redundancy extraction: Non-negative Matrix Factorization (NMF)

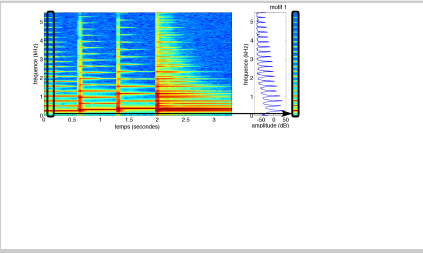


NMF makes it possible to extract straight redundancies. This is illustrated with this simple extract of electric piano.

Redundancy extraction: Non-negative Matrix Factorization (NMF)

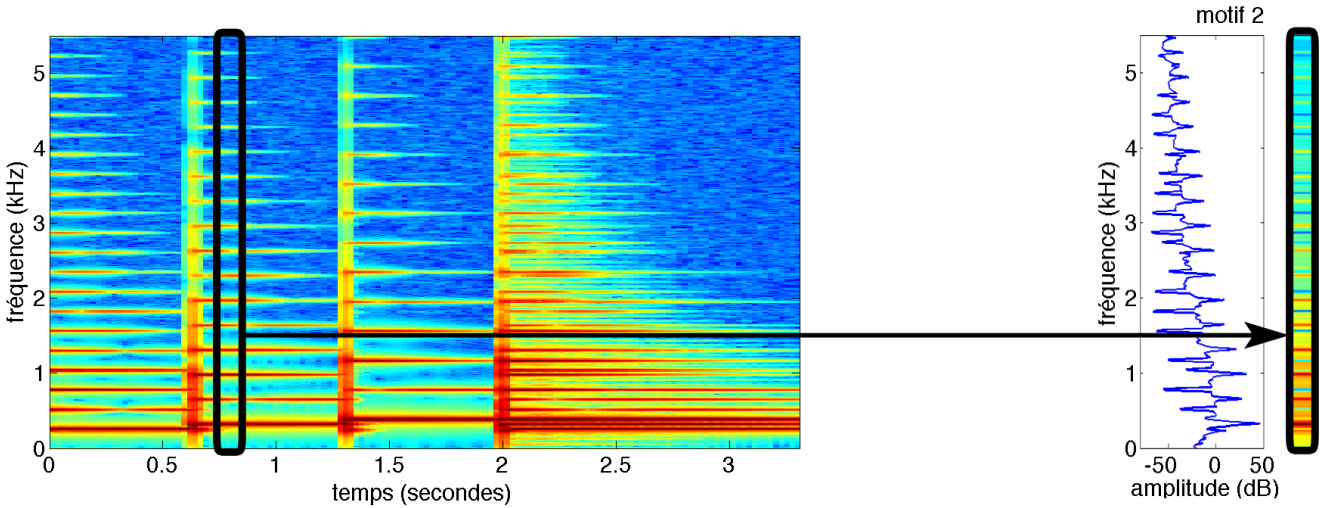


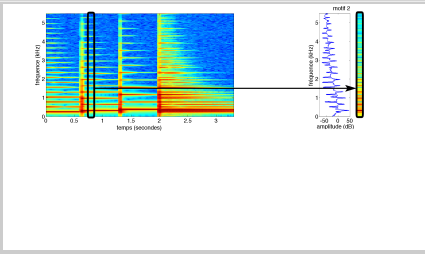
Redundancy extraction: Non-negative Matrix Factorization (NMF)



We can extract a first pattern corresponding to the first note,

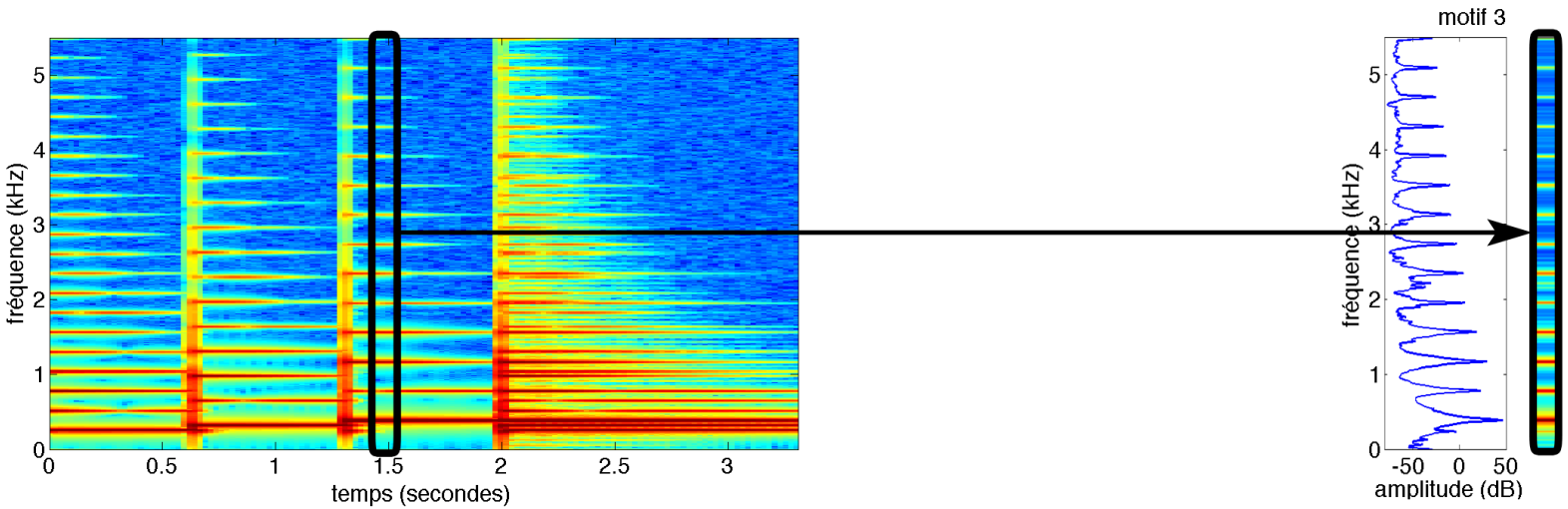
Redundancy extraction: Non-negative Matrix Factorization (NMF)

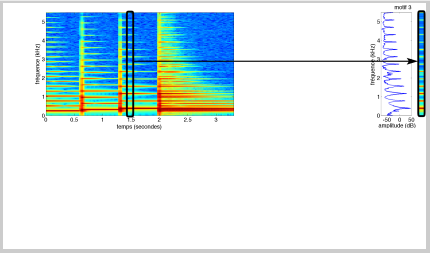




a second pattern corresponding to the second note,

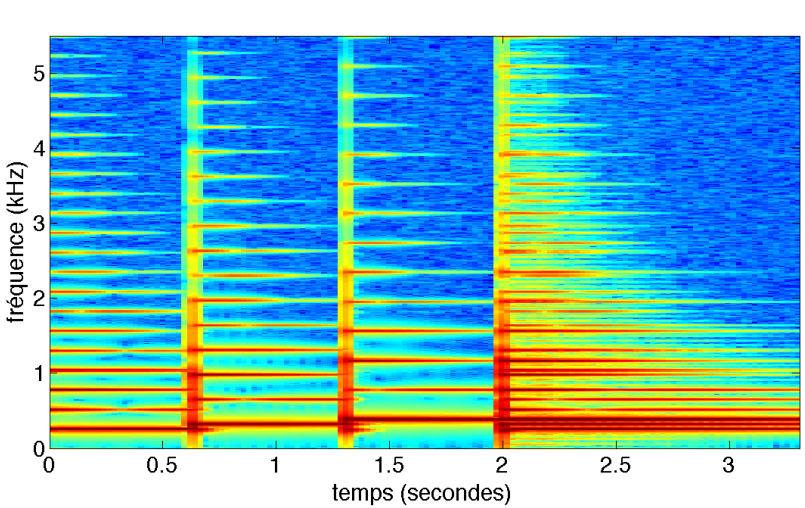
Redundancy extraction: Non-negative Matrix Factorization (NMF)

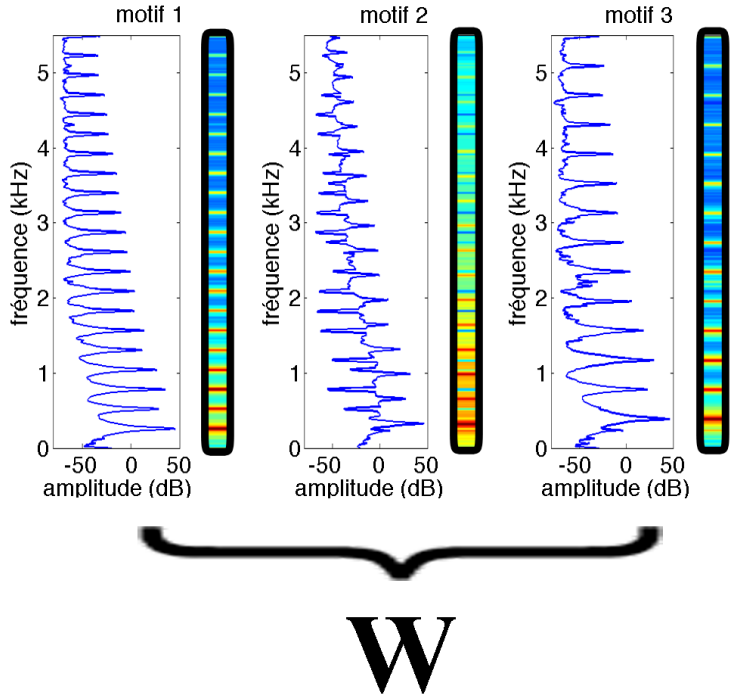


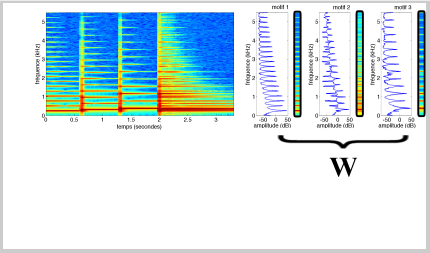


and a third pattern corresponding to the third note.

Redundancy extraction: Non-negative Matrix Factorization (NMF)

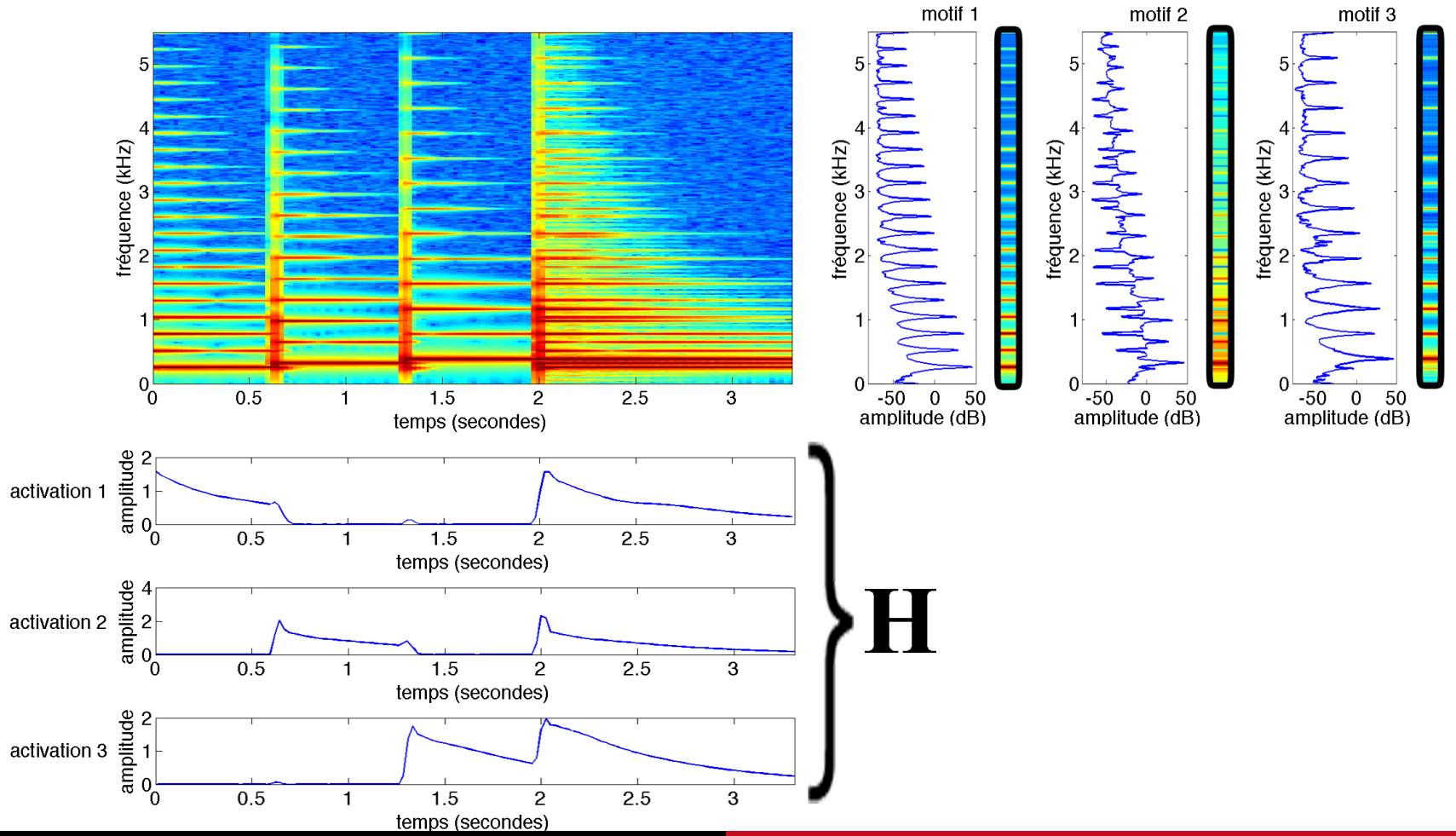


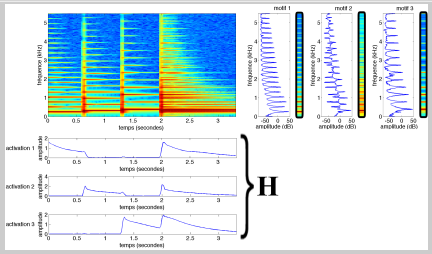




We thus get a matrix **W** of atoms (patterns)

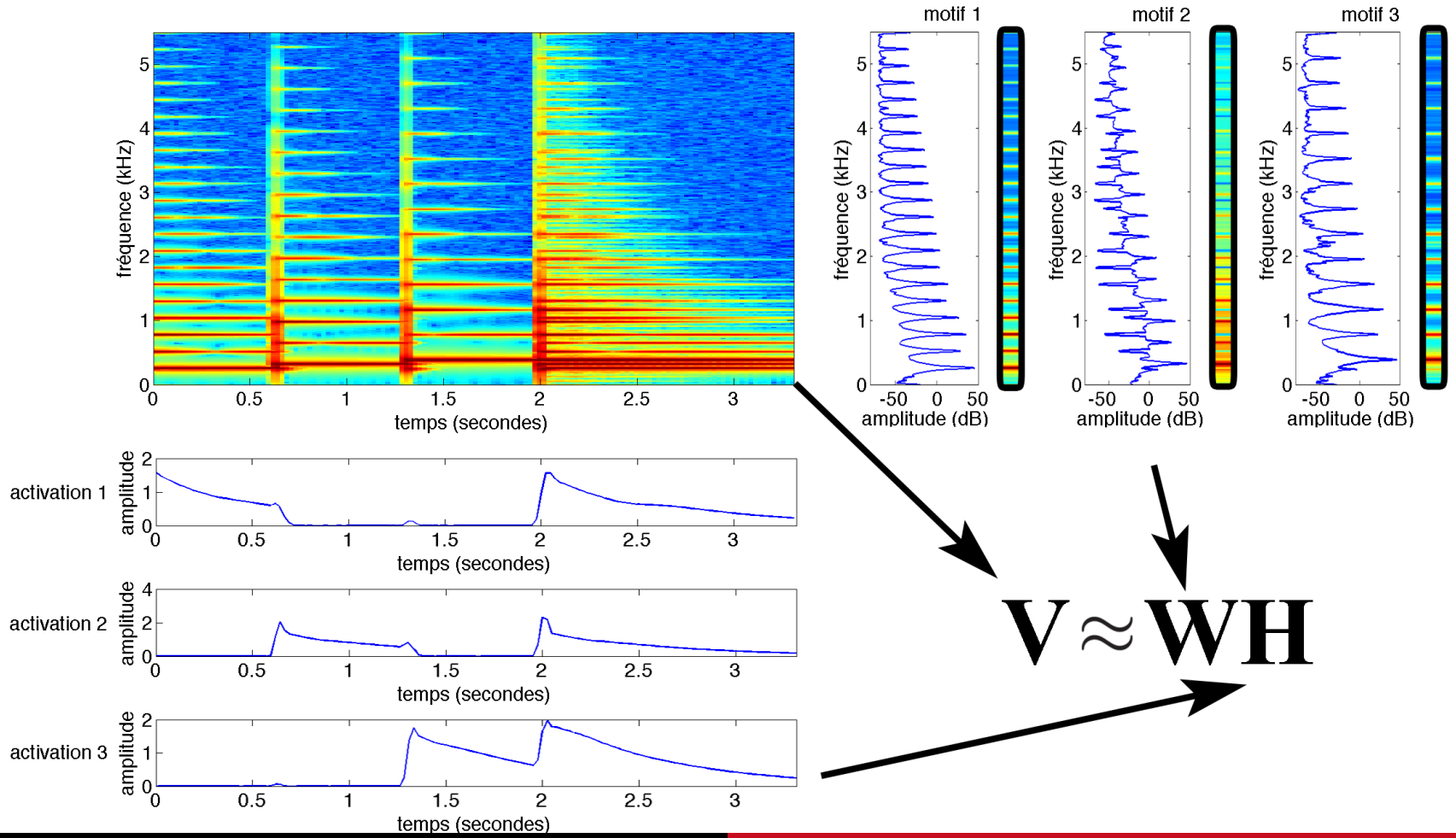
Redundancy extraction: Non-negative Matrix Factorization (NMF)



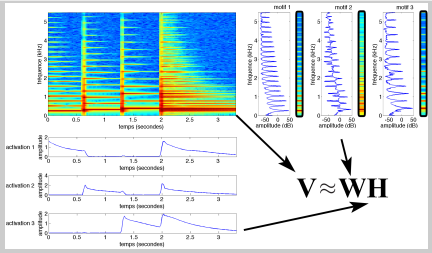


and we can straightforwardly decompose the given spectrogram on these atoms: we get the matrix **H** called activation matrix.

Redundancy extraction: Non-negative Matrix Factorization (NMF)

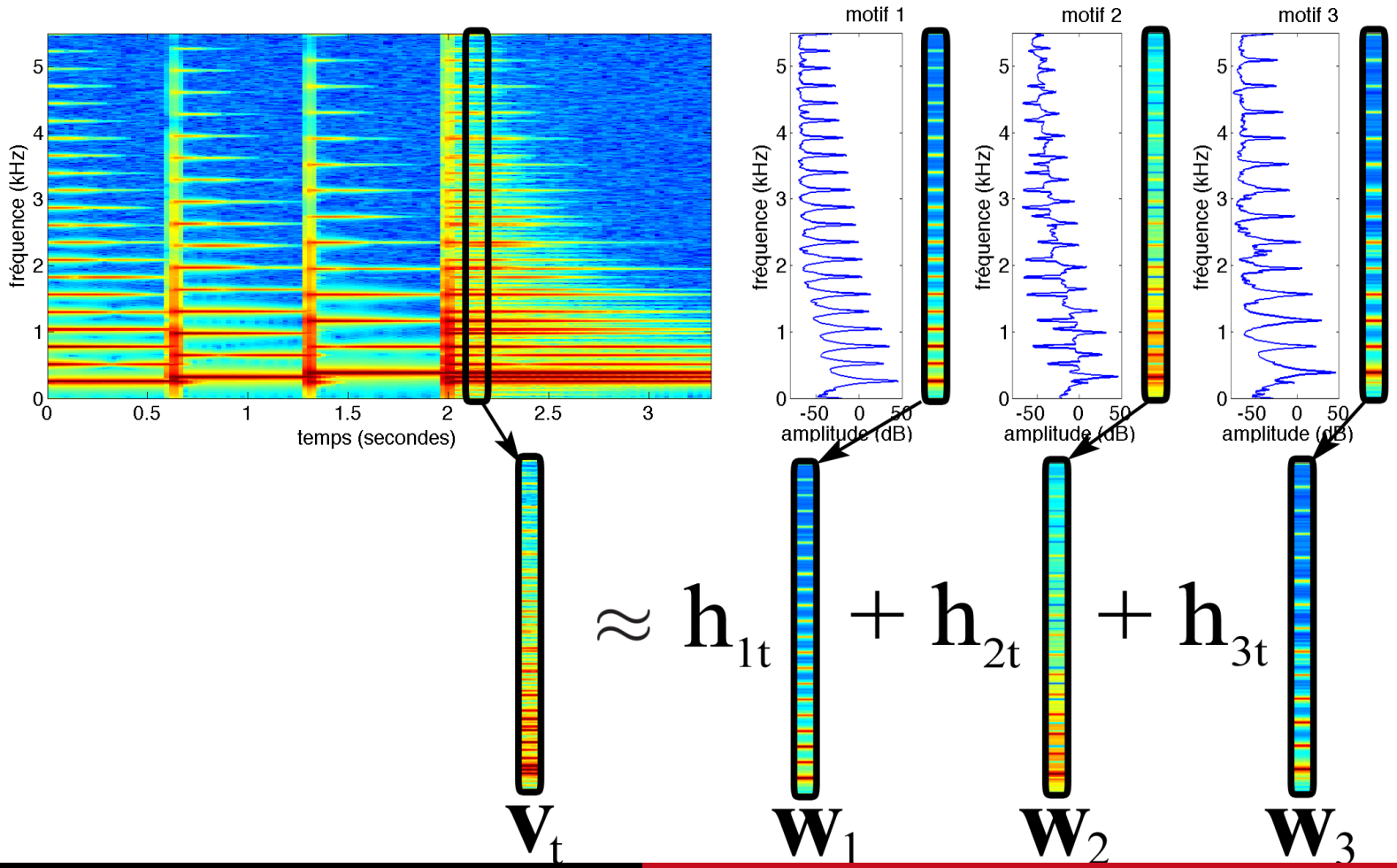


Redundancy extraction: Non-negative Matrix Factorization (NMF)

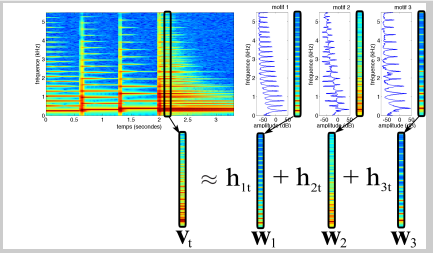


This operation corresponds to a matrix factorization (approximation) of our original spectrogram.

Redundancy extraction: Non-negative Matrix Factorization (NMF)

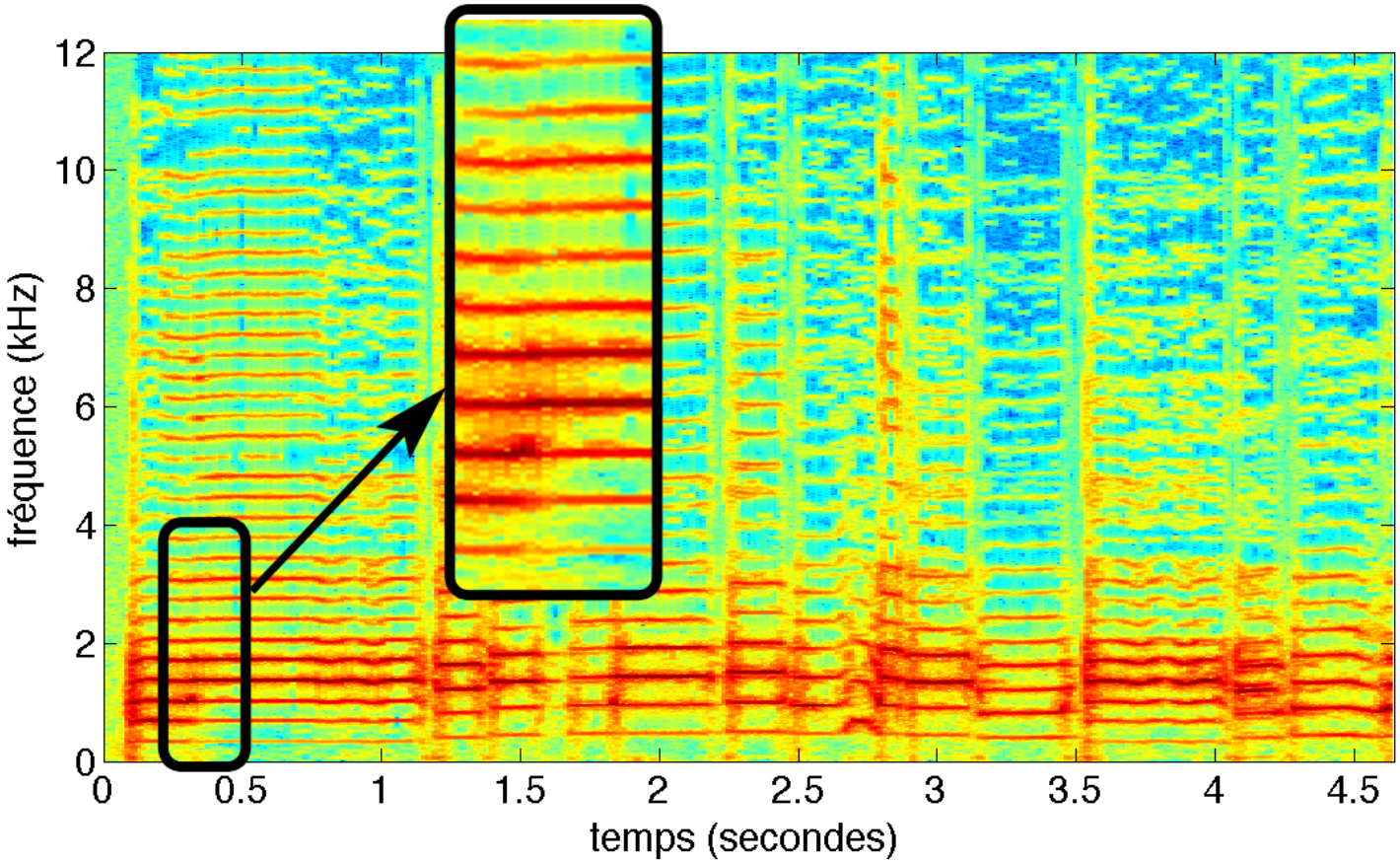


Redundancy extraction: Non-negative Matrix Factorization (NMF)

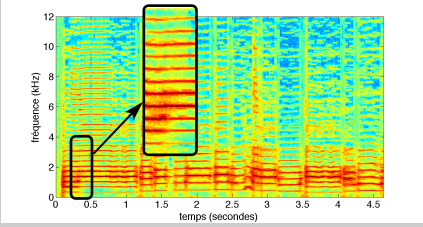


Each frame of the spectrogram is decomposed on three “good” patterns.

Introduction



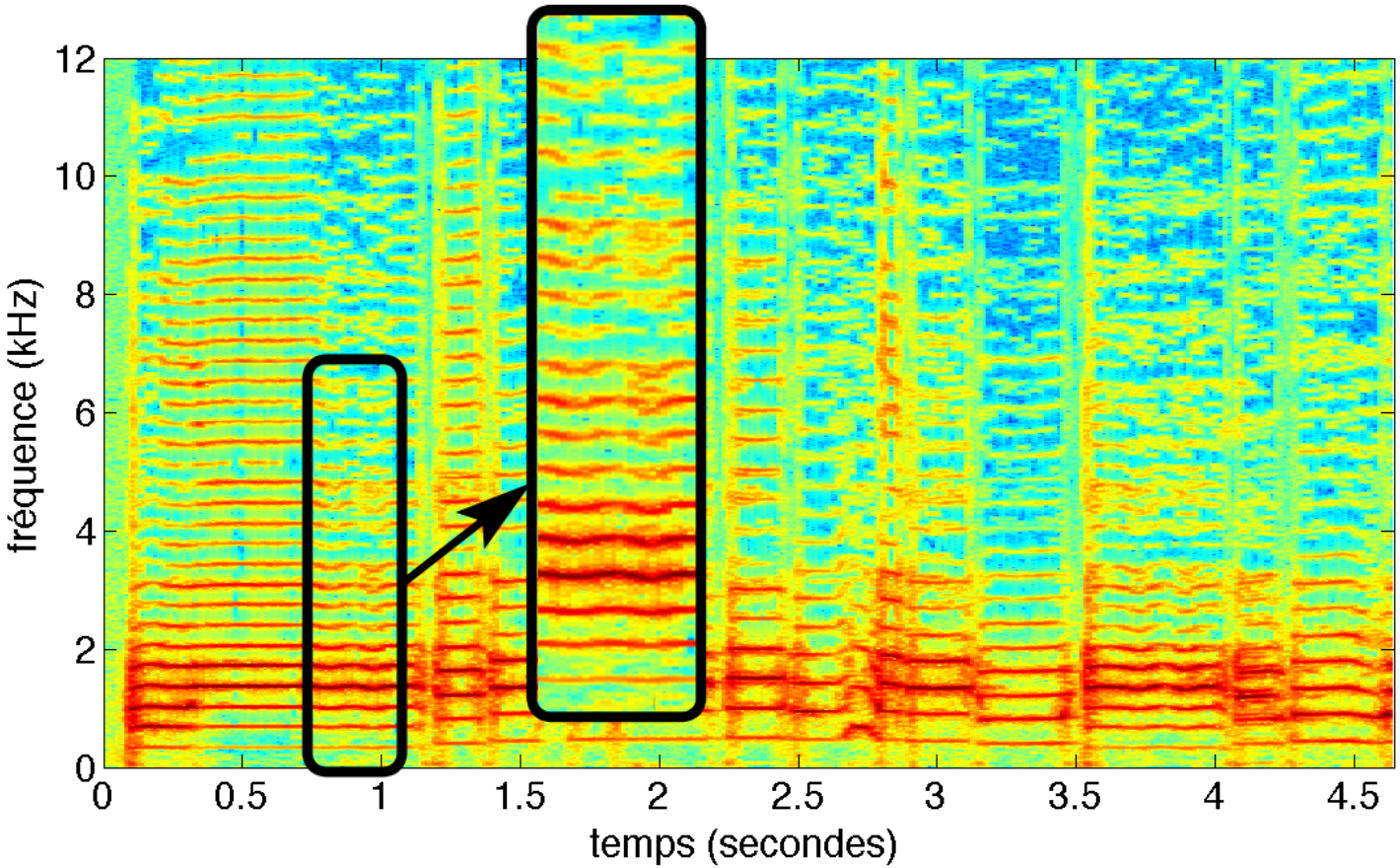
Spectral envelope variations



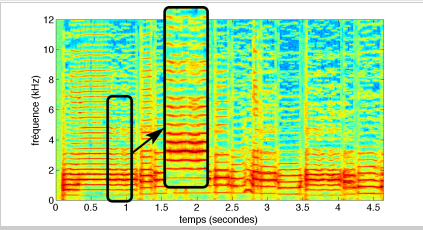
Spectral envelope variations

Unfortunately, NMF does not permit to take efficiently time variations within a single note into account. In this thesis we focused on two kinds of very common time variations: the first one is the variations of the spectral envelope (for instance the "wah" sound in the trumpet extract),...

Introduction



Variations de fréquence fondamentale



Variations de fréquence fondamentale

... and the second one is the variations of the fundamental frequency (for instance, the vibrato in the trumpet extract).

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
- ...

Possible applications:

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

Although this thesis is more focused on the representation aspect, decompositions of musical spectrograms can be used in several applications and we will present some of them.

Contents

Contents

Principle of NMF

Low rank approximation:

V being a non-negative matrix (amplitude or power spectrogram), NMF approximates **V** in the following way:

$$\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{WH}$$

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

- **W** matrix $F \times R$ and **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:

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

Multiplicative algorithm

- Alternated optimization with respect to (wrt) $\mathbf{W}$ and $\mathbf{H}$.
- Decomposition of the gradient as a difference of two positive terms:
  $\nabla_{\mathbf{W}}\mathcal{C}_{\mathbf{V}}(\mathbf{W},\mathbf{H}) = \mathbf{P}_{\mathbf{W}} - \mathbf{M}_{\mathbf{W}}$ where $\mathbf{P}_{\mathbf{W}} > 0$ and $\mathbf{M}_{\mathbf{W}} > 0$
  $\nabla_{\mathbf{H}}\mathcal{C}_{\mathbf{V}}(\mathbf{W},\mathbf{H}) = \mathbf{P}_{\mathbf{H}} - \mathbf{M}_{\mathbf{H}}$ where $\mathbf{P}_{\mathbf{H}} > 0$ and $\mathbf{M}_{\mathbf{H}} > 0$
- Update rules:

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

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

Most commonly used algorithms for NMF are multiplicative algorithms. The optimization is made alternatively wrt **W** and **H**. A simple approach to get the update rules consist in decomposing the gradient as a difference of two positive terms. The update rules is then simply an element-wise multiplication with the ratio of these two terms.

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.

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.

This kind of approach ensure several good properties to deal with non-negative data: the update rules ensure that the parameters remains non-negative, the direction of the update rule is a descent direction (even if the cost function is not necesarly decreasing) and zeros of the gradient are fixed point of the update rules.

Majoration/Minimization algorithms form a more rigorous framework for multiplicative algorithms: these algorithms notably ensure the decrease of the cost function.

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] ...

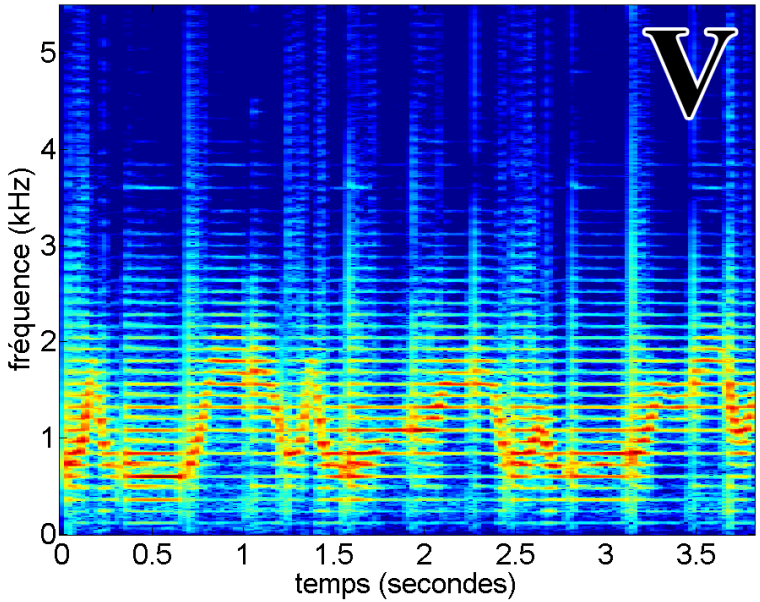
Issues with NMF

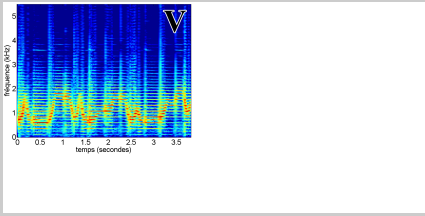
Time variations

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

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

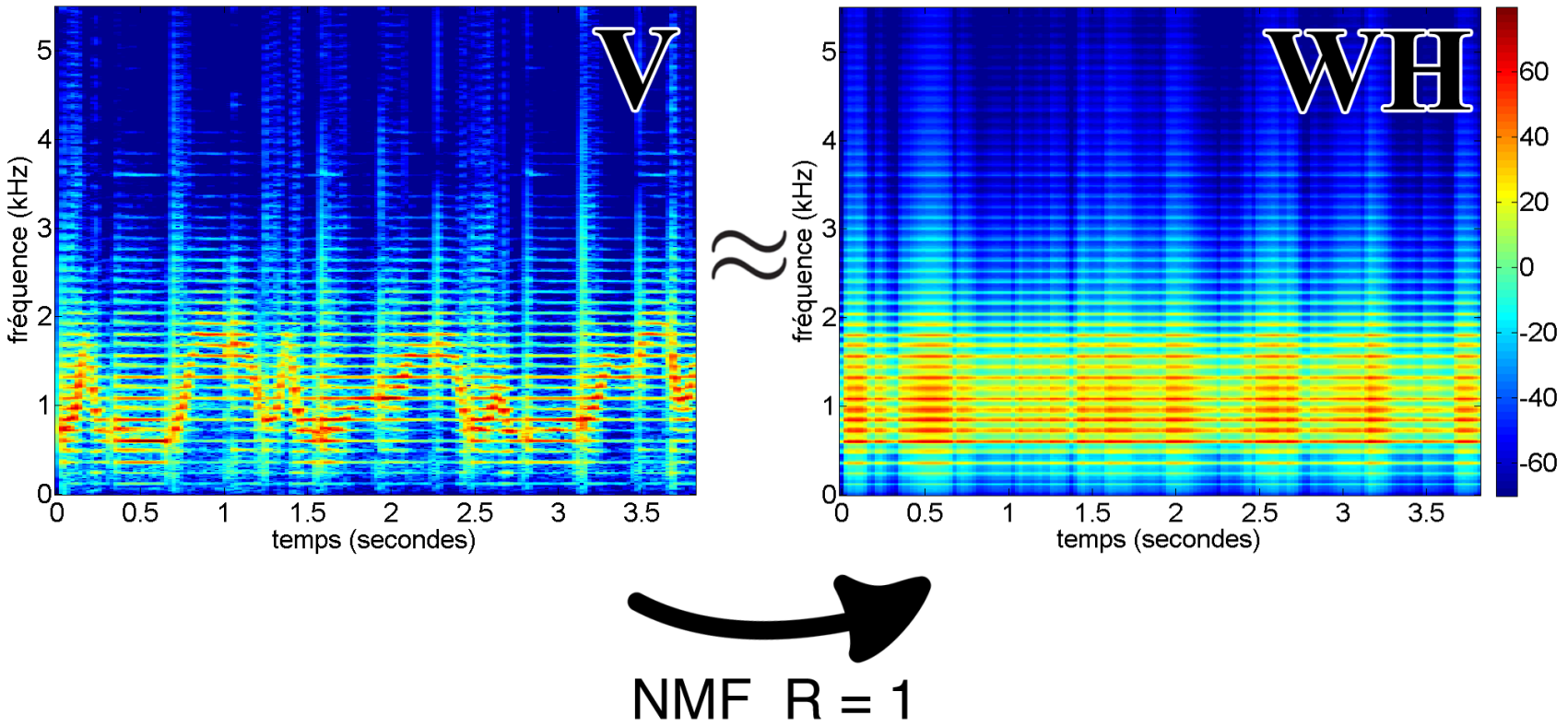
Issues with NMF: spectral envelope variations

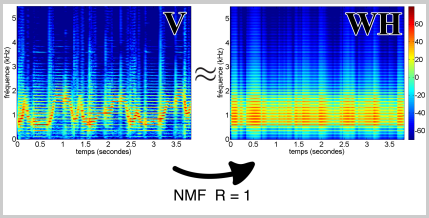




A jew harp sound is considered to illustrate issues when using NMF on sound elements with strong spectral envelope variations: the sound of a jew harp is harmonic and presents a strong resonance produced by the mouth of the instrumentist.

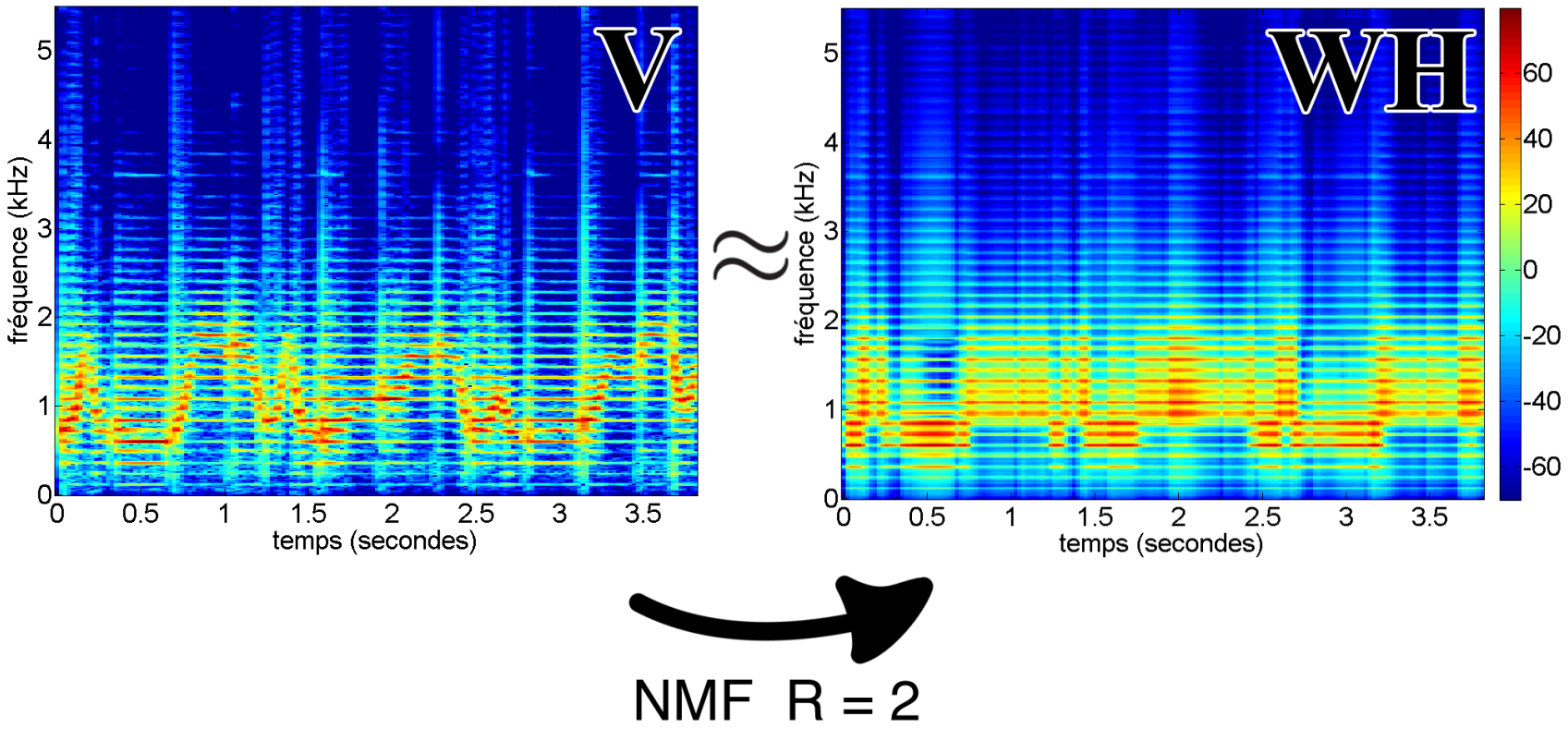
Issues with NMF: spectral envelope variations

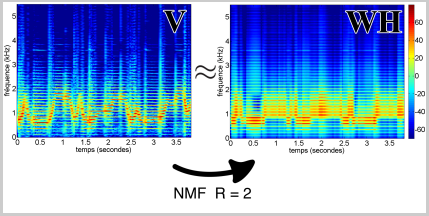




Decomposing the spectrogram of the jew harp sound using a NMF with a single atom does not permit to model the strong resonance, even if there is a strong latent redundancy inside the spectrogram (harmonic spectral template with fixed fundamental frequency).

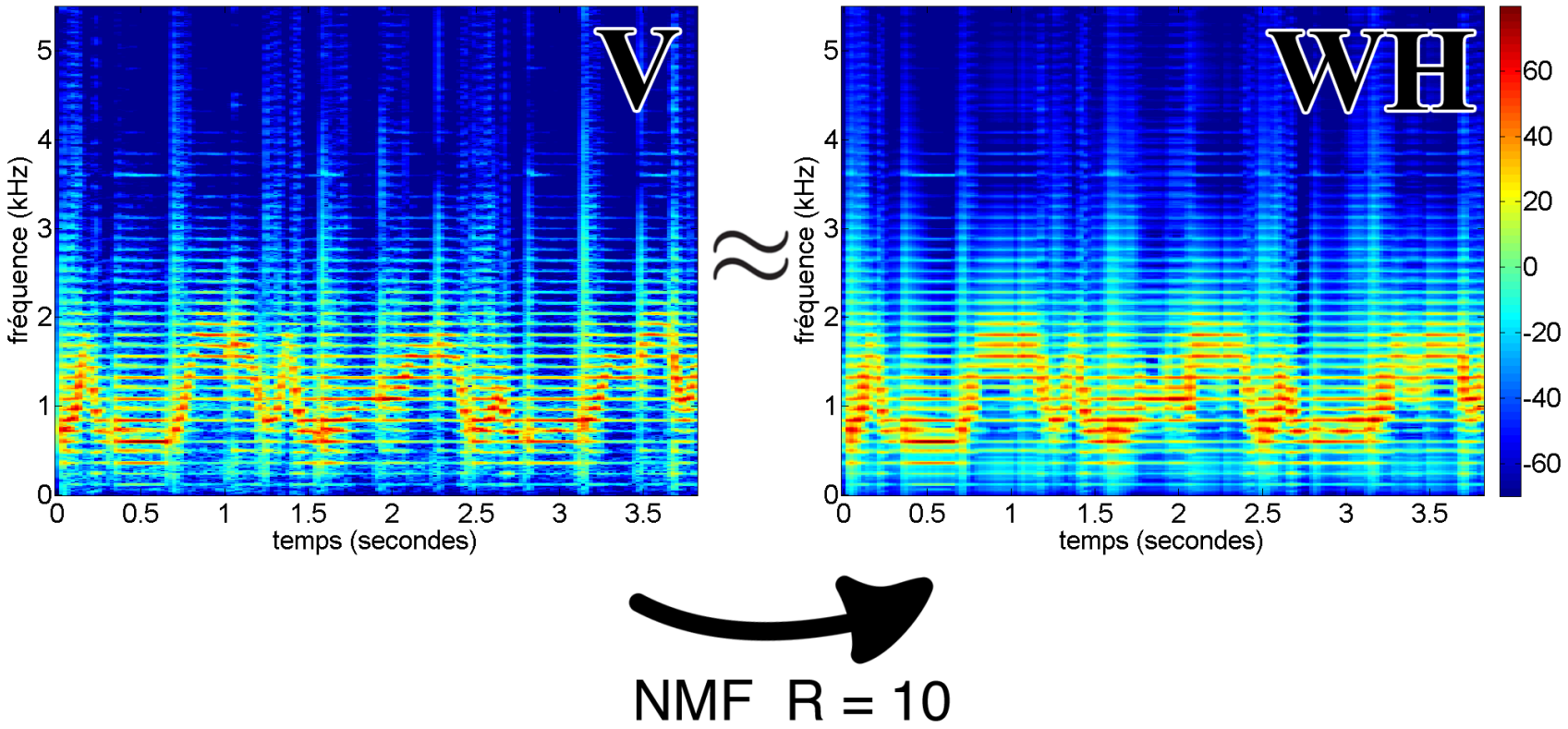
Issues with NMF: spectral envelope variations

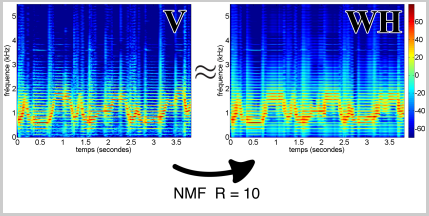




With two atoms, the resonance is loosely modeled: there is some sort of a two-state behavior.

Issues with NMF: spectral envelope variations





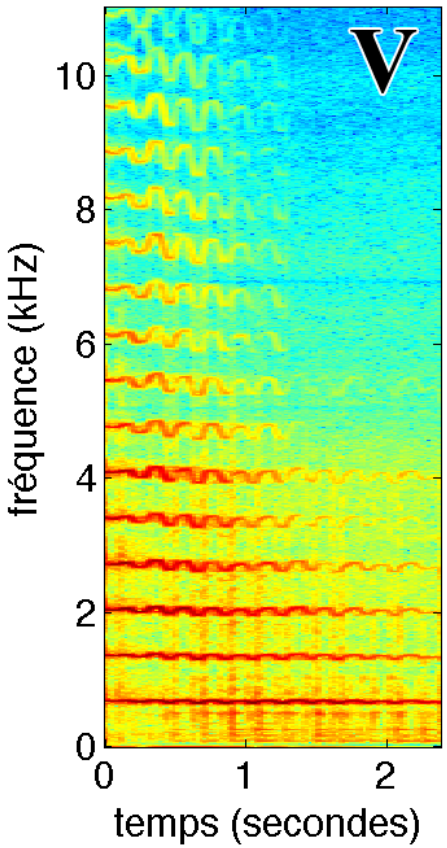
With ten atoms, the resonance is finely modeled, but the atoms no longer have a meaning individually.

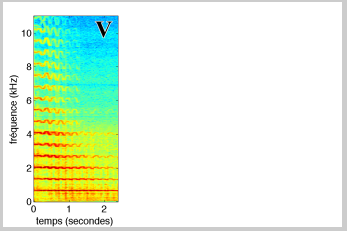
Issues with NMF: spectral envelope variations.

Issues

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

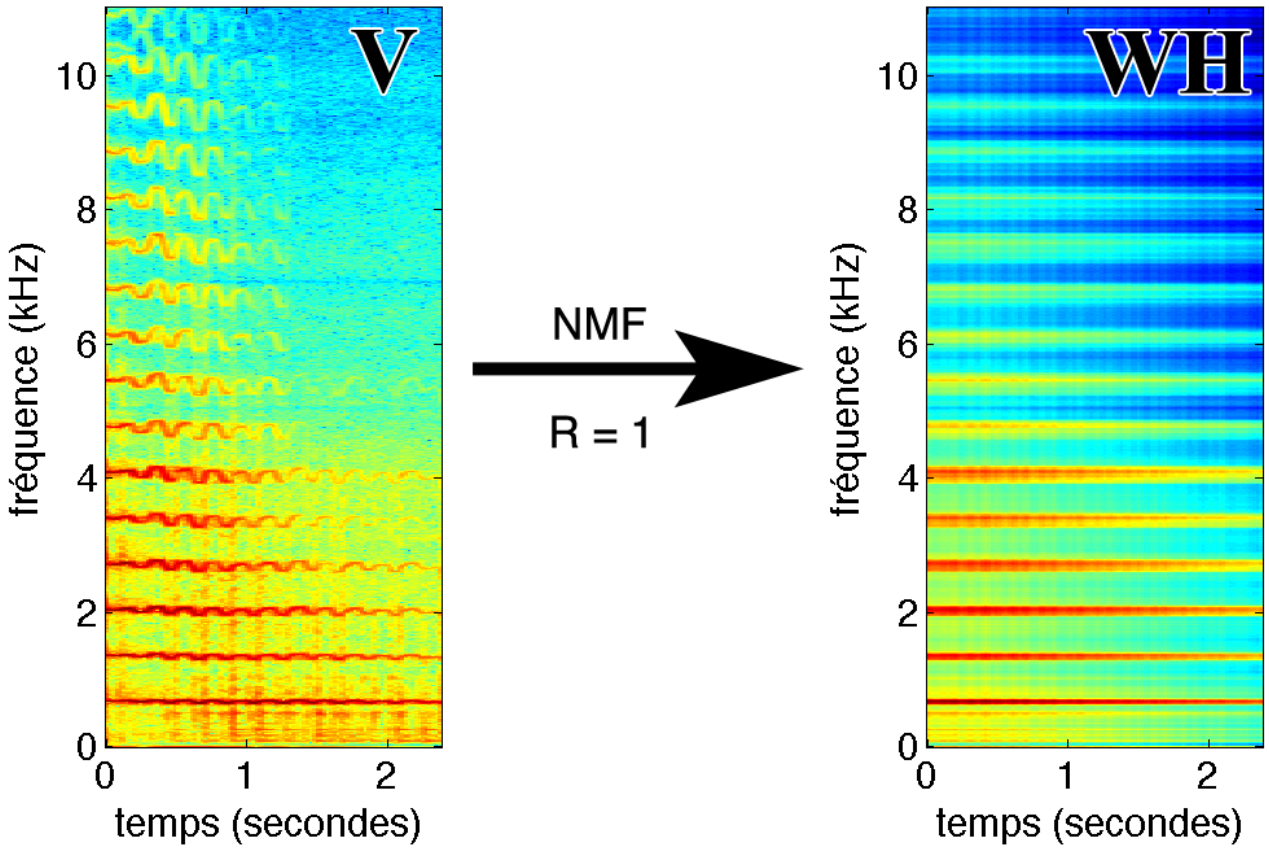
Issues with NMF: fundamental frequency variations

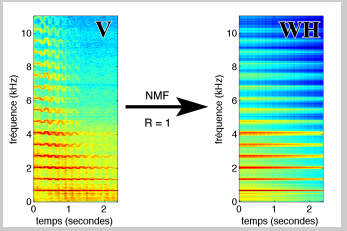




Here is the sound of a note played by an electric guitar with vibrato to illustrate issues with fundamental frequency variations.

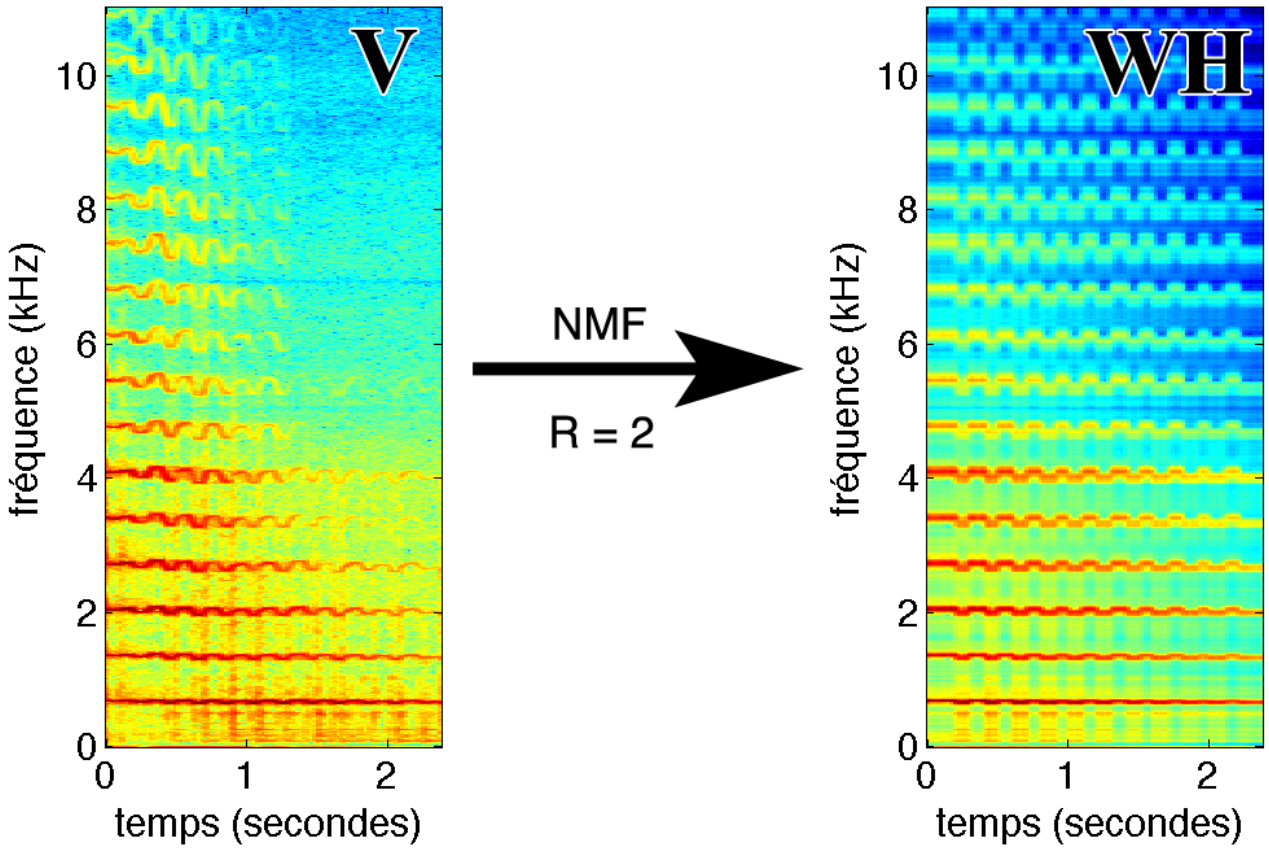
Issues with NMF: fundamental frequency variations

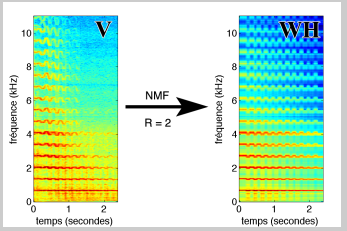




With a single atom, the vibrato is discarded.

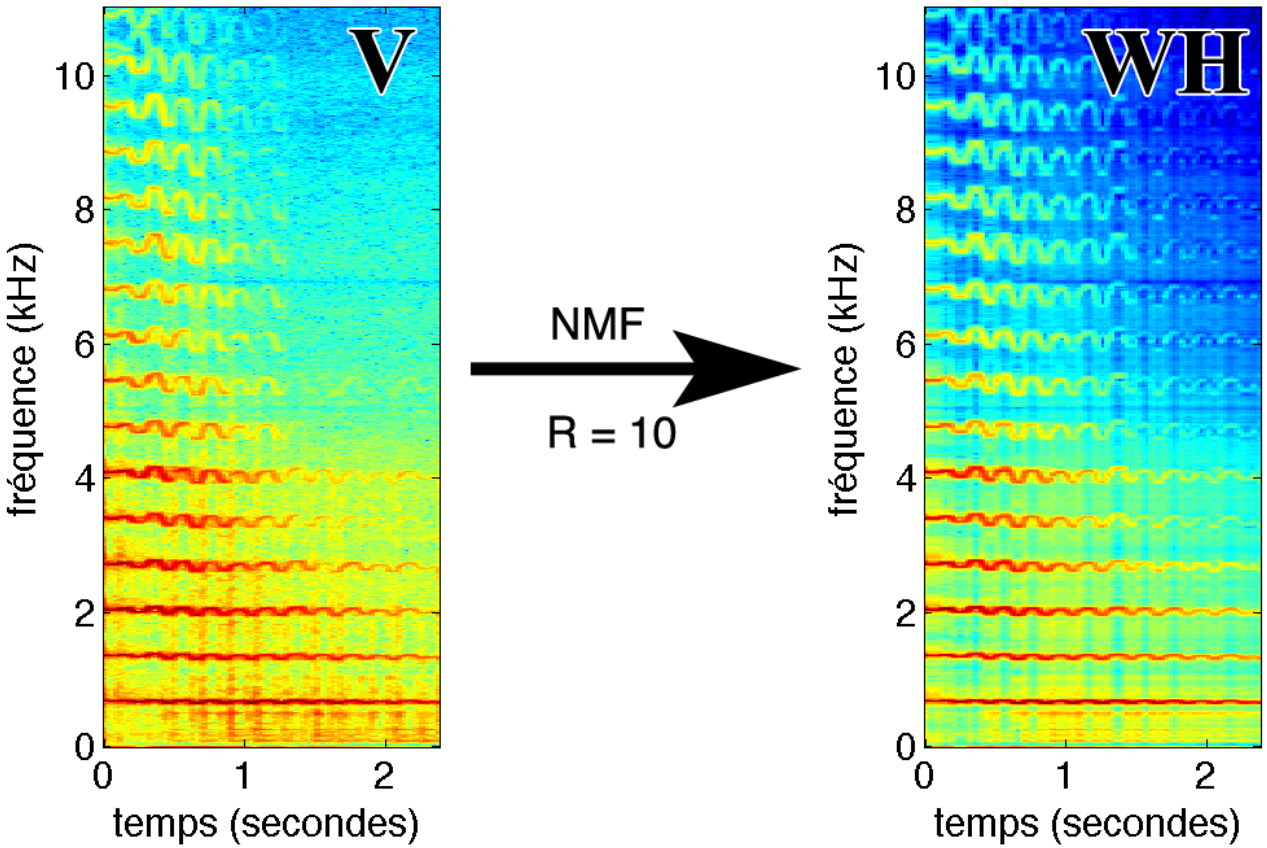
Issues with NMF: fundamental frequency variations

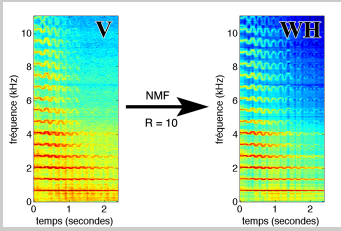




With two atoms, the vibrato is quite well modeled for low frequency partials but still appears very fuzzy for high frequency partials.

Issues with NMF: fundamental frequency variations





Then, one needs a lot of atoms to model finely the vibrato: with ten atoms the vibrato is accurately modeled, but each atom no longer has a perceptual meaning.

Solutions

Proposed solutions

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

Proposed solutions

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

The main idea of the thesis is to introduce into NMF models of spectrogram that makes it possible to extract underlying redundancies. Models proposed are 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).

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).

We propose to introduce source/filter synthesis in NMF: in source filter synthesis, the source is a stationary sound harmonically rich which is filtered by a time-varying filter. Then, in the source filter synthesis model the source is the redundant part of the sound and the filter the varying part. We thus propose to add more freedom in the activation by introducing a frequency dependence in **H**. Then **W** contains the spectral patterns of the source and **H** acts as a filter on the these source patterns. To keep a compact decomposition, one should parameterized this frequency dependence.

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.)

Filters parametrization

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

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

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

(ν_f: normalized frequency.)

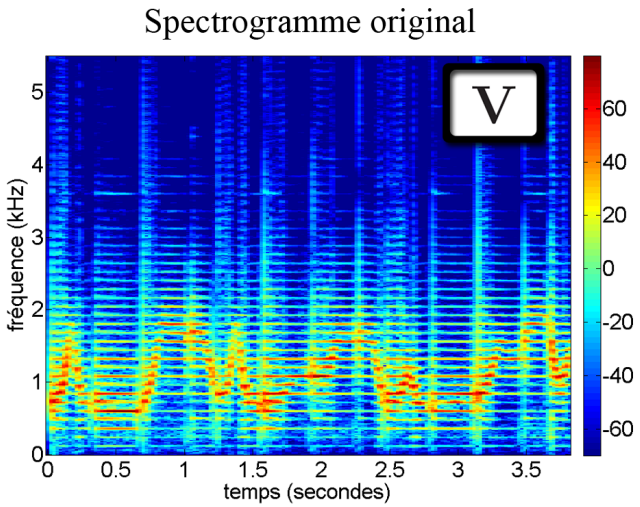
The chosen parametrization is the AutoRegressive Moving-Average (ARMA) filter since it models a large class of filters with a few parameters: the parameters to estimate in order to compute the time-frequency activations are then the MA coefficients, the AR coefficients and the global gain. When $P = 0$ and $Q = 0$, there is no frequency dependence and the decomposition is a standard NMF.

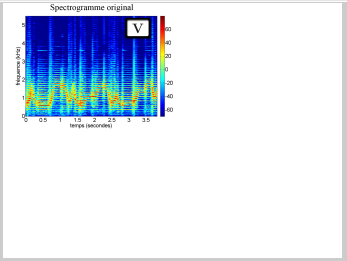
Source/filter decomposition

Decomposition

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

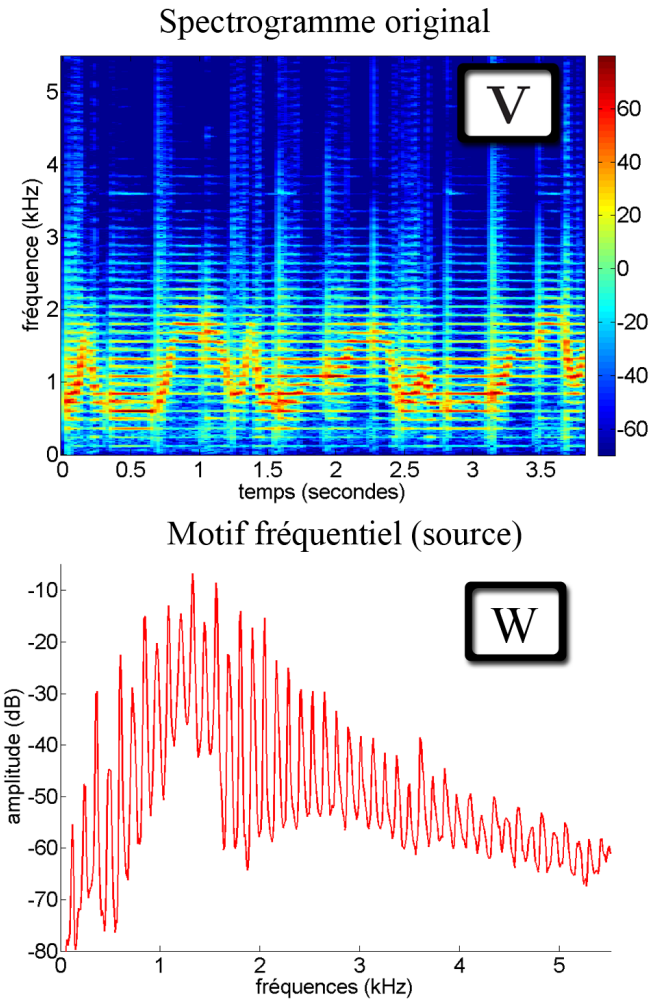
Decomposition of the jew harp sound, 2nd order AR filter

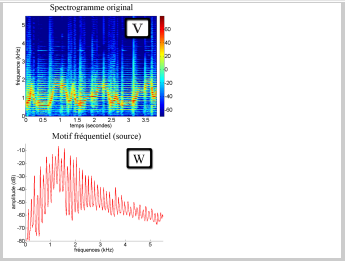




The same jew harp sound as presented in the previous part is here decomposed with the proposed source/filter decompositon using a single atom and an AR-2 filter.

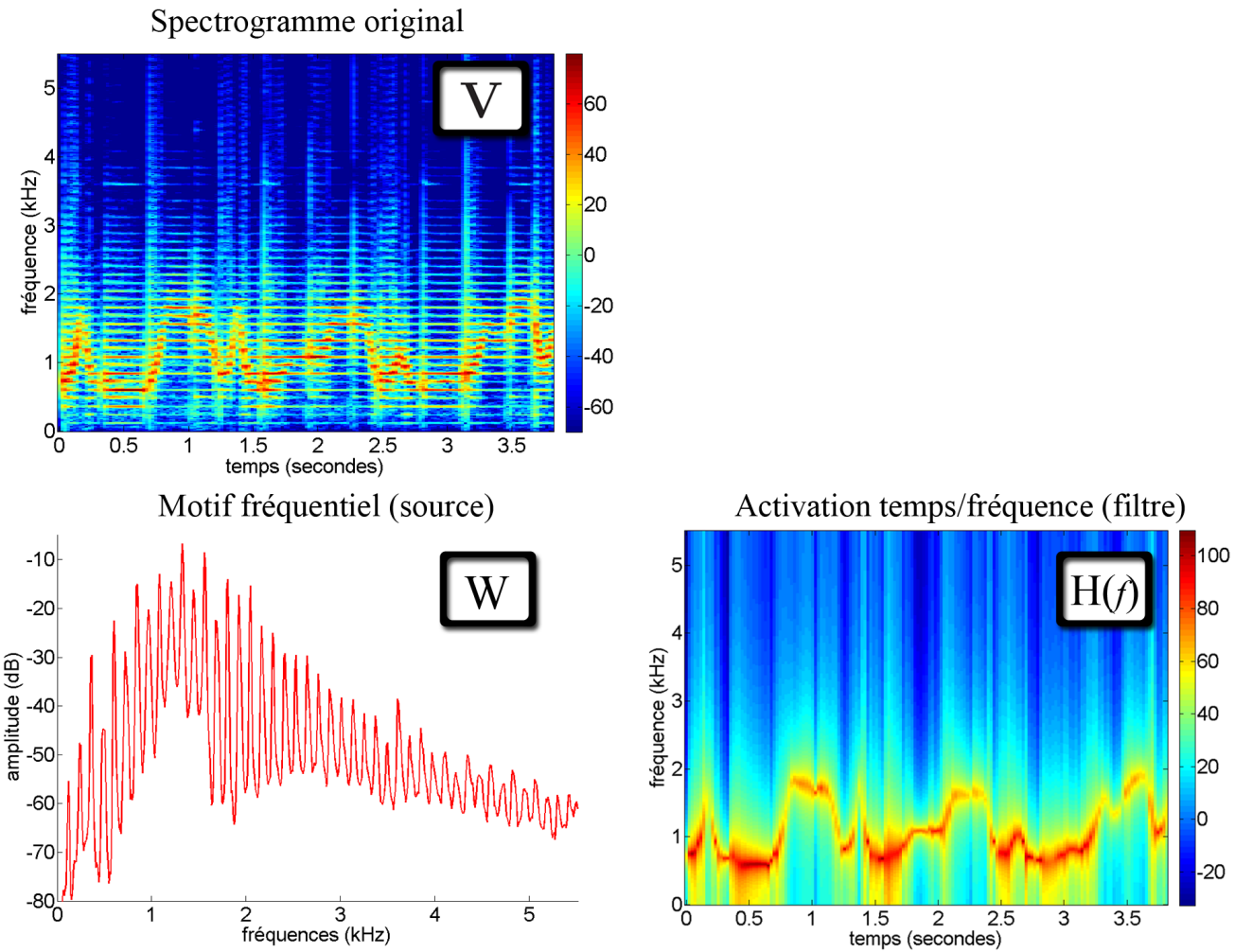
Decomposition of the jew harp sound, 2nd order AR filter

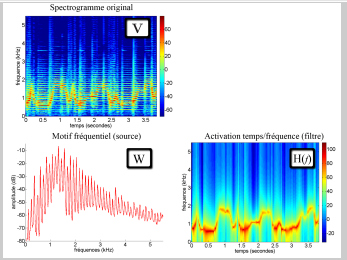




The decomposition provides an harmonic pattern which correspond to the source in the source filter/model, which is the redundant part in the sound,

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

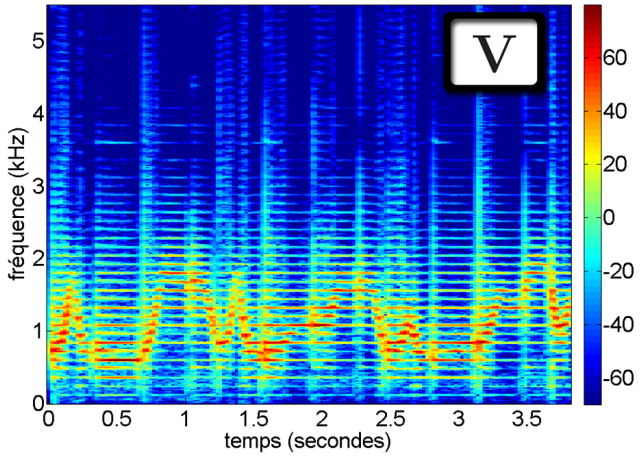




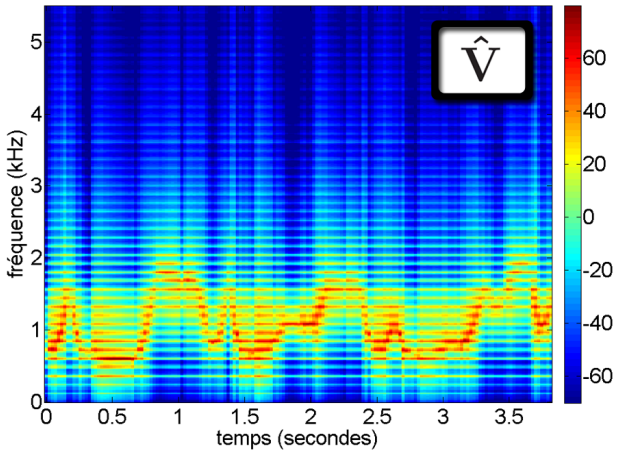
and the time/frequency activation reveals the strong resonance.

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

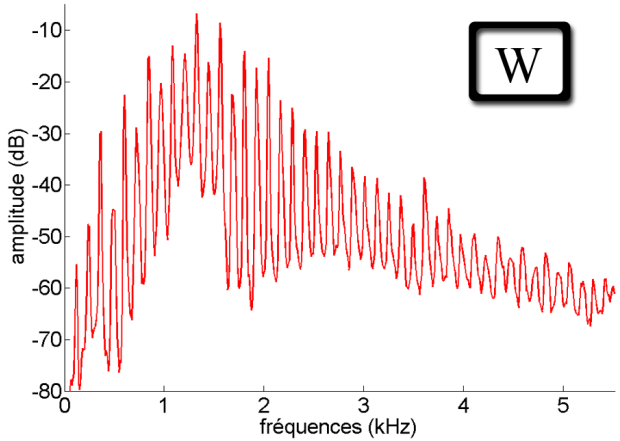
Spectrogramme original



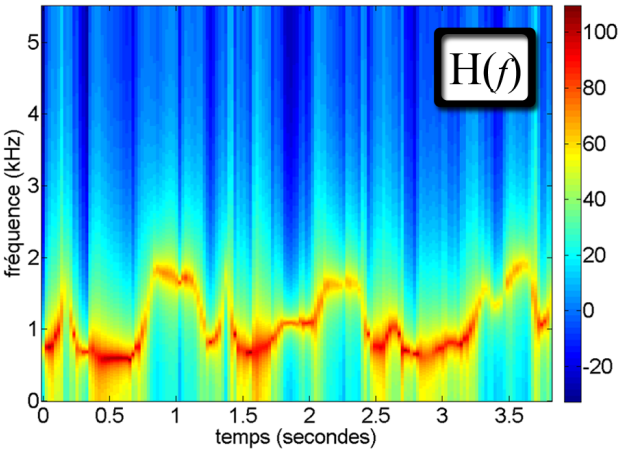
Spectrogramme reconstruit

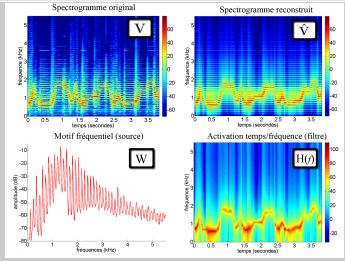


Motif fréquentiel (source)



Activation temps/fréquence (filtre)





The reconstructed spectrogram is then very similar to the orignal one.

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.

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.

The proposed decomposition based on the source/filter paradigm is thus efficient since it permits an accurate modeling of the spectrogram with a single element for a single atom.

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.

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.

This second part focuses on the problem of modeling fundamental frequency variations in sound elements (one considers here, that there are no spectral envelope variations in these elements). The idea is to synthesize an harmonic atom as it is done in additive synthesis: the harmonic atom is build as a sum of its harmonics from a fundamental frequency parameter (which can vary over time) and the amplitudes of these harmonics.

Parametric spectrogram

Time varying atoms

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

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

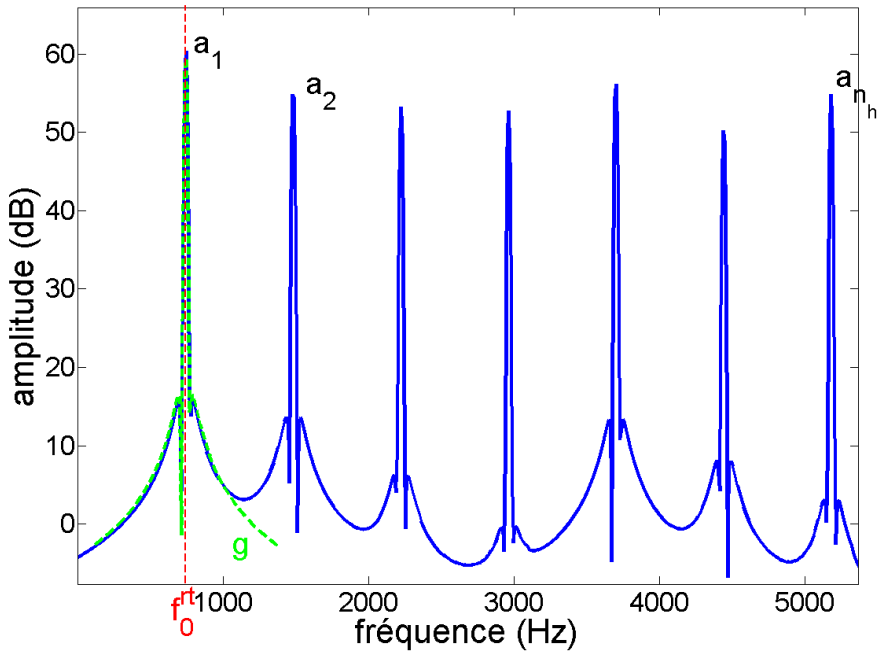
Time varying atoms

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

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

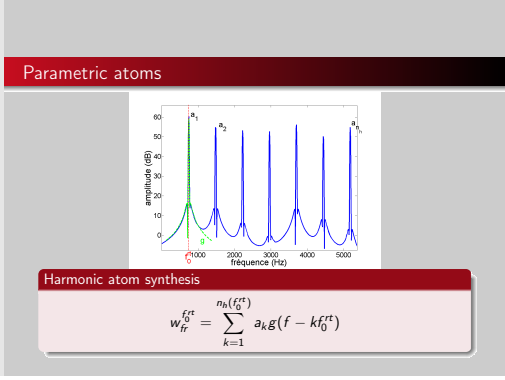
A time dependence is introduced in the atoms with a fundamental frequency parameter which can vary over time.

Parametric atoms



Harmonic atom synthesis

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



Parametric atoms are simply build using the Fourier transform g of the analysis window used to compute the considered spectrogram. This function g is centered on f_0^{rt} and the frequency of the harmonics kf_0^{rt}, and is multiplied by a parameter corresponding to the amplitude of the harmonic. The atom r at time t is then simply the sum of all these peaks.

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

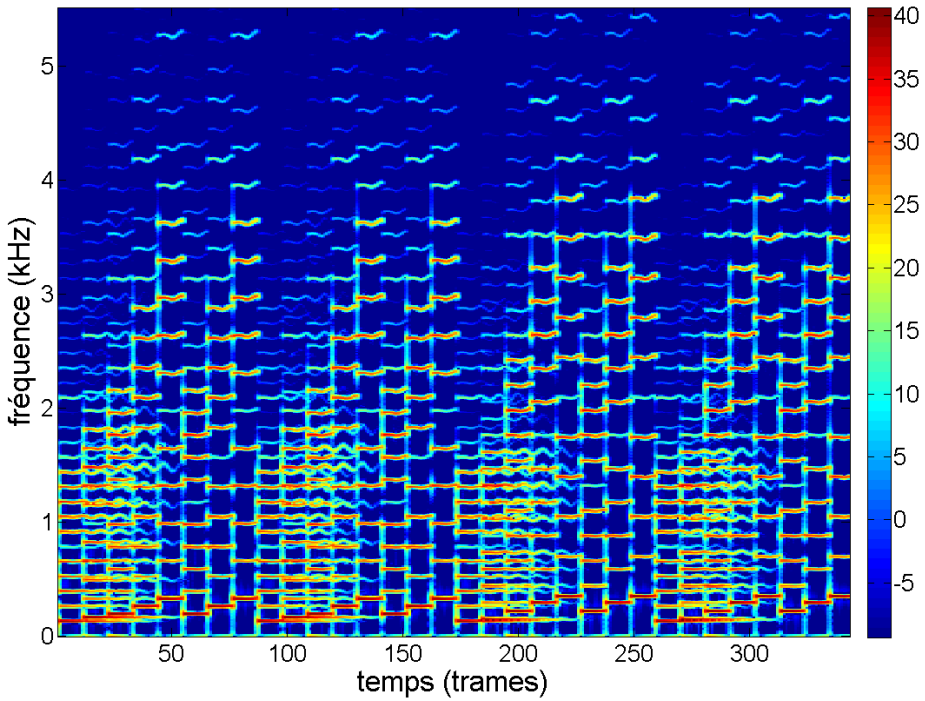
Cost function: $\mathcal{C}(f_0^{rt}, a_k, h_{rt}) = D(\mathbf{V}|\hat{\mathbf{V}})$

Algorithm

Minimization

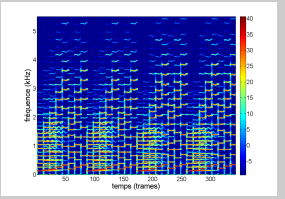
- Global optimization with respect to f_0^{rt} is impossible (numerous local minima in $\mathcal{C}$).
  $\Rightarrow$ Introduction of an atom for each note of the chromatic scale (*i.e.* for each MIDI note)
  $\Rightarrow$ Local optimization wrt f_0^{rt} (fine estimation of f_0^{rt}).
- Minimization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

Decomposition of a synthetic spectrogram



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

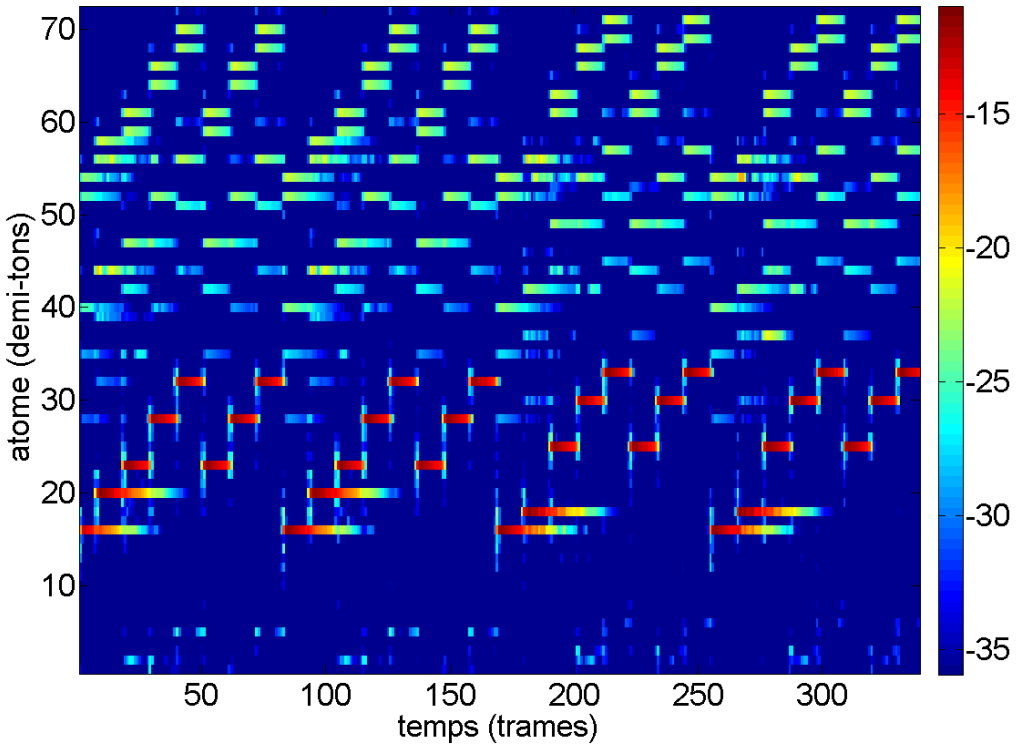
Decomposition of a synthetic spectrogram



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

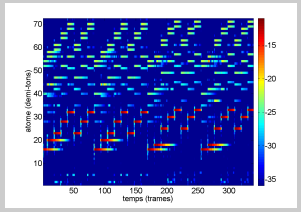
To illustrate the proposed decomposition, we decompose the spectrogram of the first bars of Bach's first prelude played by a synthesizer. Each note was played with a slight vibrato, as it can be seen in the spectrogram.

Obtained decomposition



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

Obtained decomposition



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

The obtained activations can be represented as something which is very similar to a “piano-roll” since there is one atom for each note of the chromatic scale.

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

Onsets

Standard NMF atoms are added to the decomposition:

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

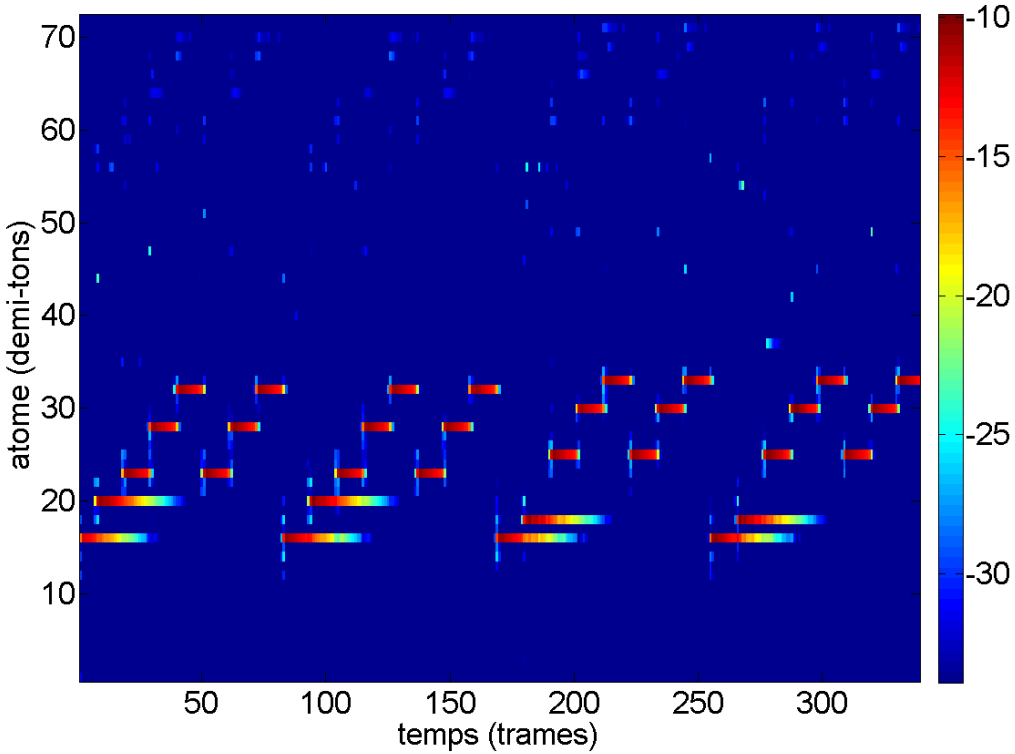
Octave, twelfth...

Penalization is added to the cost function:

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

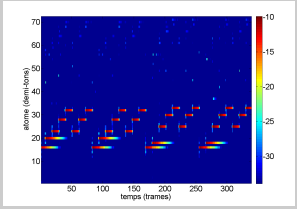
In order to improve the decomposition and deal with the issues of onsets and replicas, we can add a few standard NMF atoms to model the onsets and add penalizations in the cost function to reduce the activation of octave (twelfth...) of played notes. For instance, one can use a correlation penalization that reduce the co-occurence of a note and its octave.

Improvement



Decomposition with right penalizations and a few standard NMF atoms.

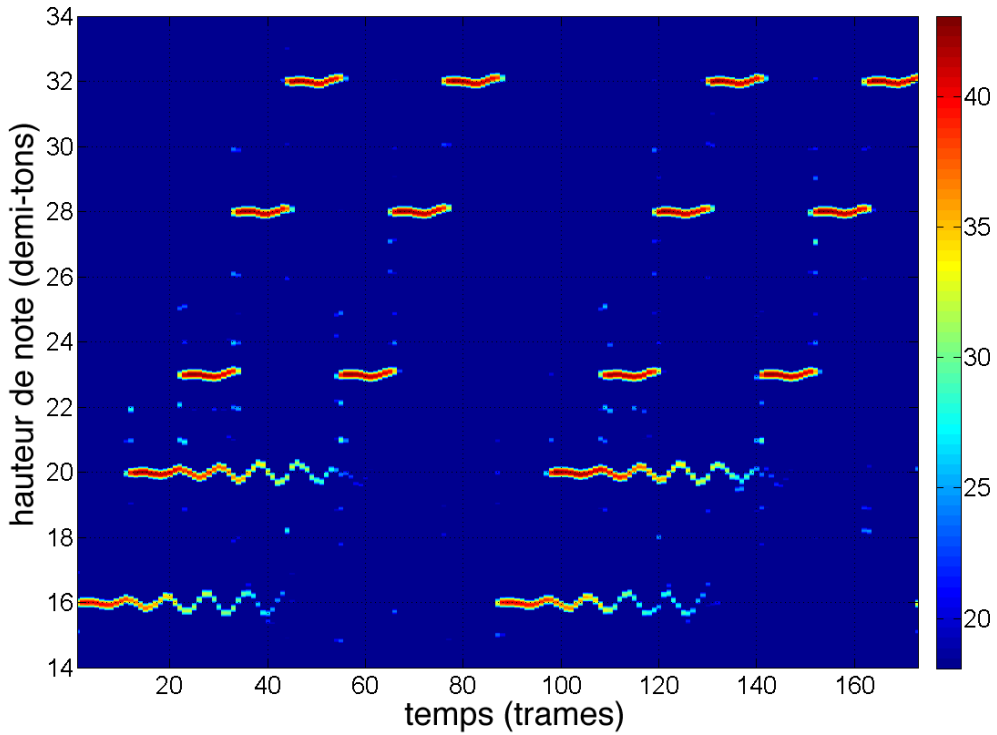
Improvement



Decomposition with right penalizations and a few standard NMF atoms.

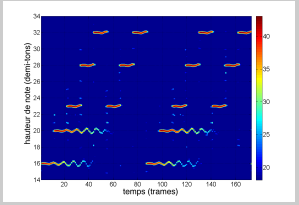
This makes it possible to reduce the replicas and the onset problems.

Representation with estimated fundamental frequencies



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

Representation with estimated fundamental frequencies



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

So far, we just presented the obtained activations. But the algorithm also estimate fundamental frequency of each atom at each time. To illustrate this, we propose a synthetic time/frequency representation where activations are centered on the estimated fundamental frequencies: we thus can see the vibrato in all the notes.

Conclusion

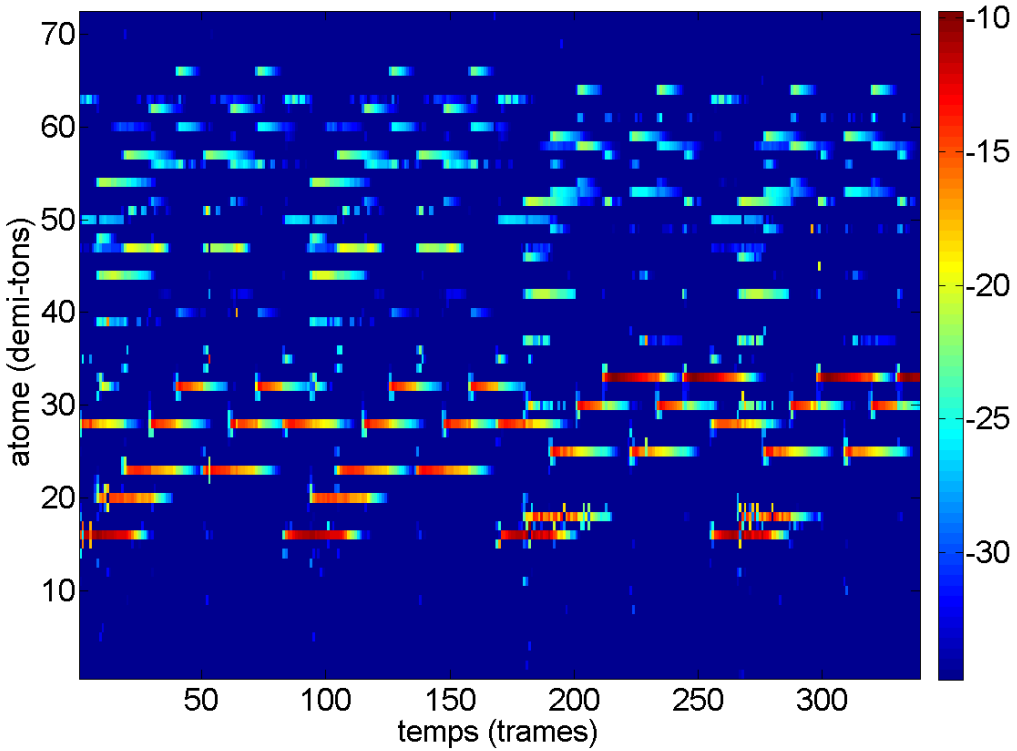
Summary

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

Issues

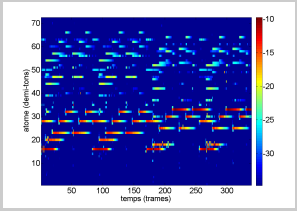
- The shape of the atoms is very constrained.
- Robustness issues to decompose real signals.

Real signal



Same piece but played by a piano.

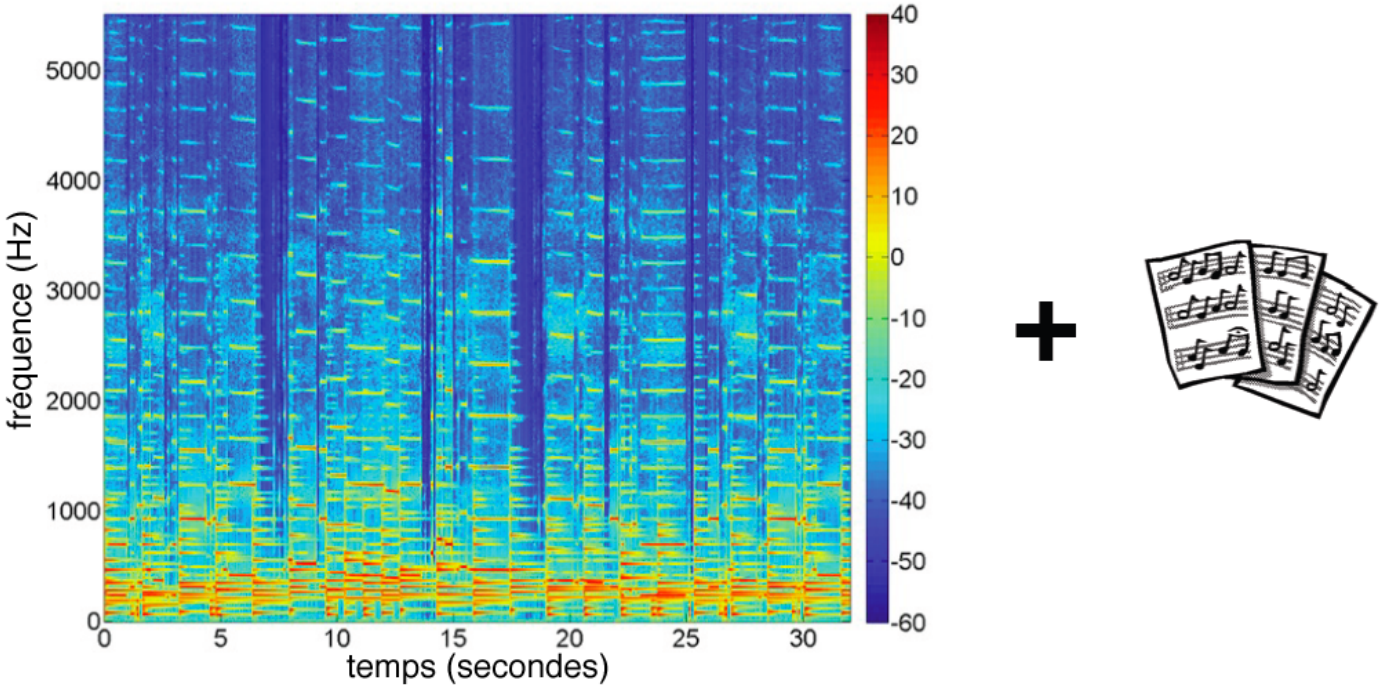
Real signal



Same piece but played by a piano.

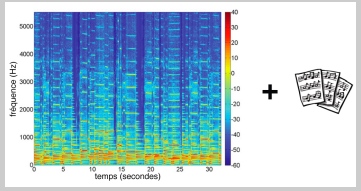
This figure illustrates the issues encountered with real sounds: even with penalizations, there can remain a lot of replicas. Thus the model seems a bit too constrained to decompose blindly real sounds. However with a good initialization, it can be used for informed source separation, which will be presented in the next slides.

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



Mixture signal + Aligned score (MIDI)

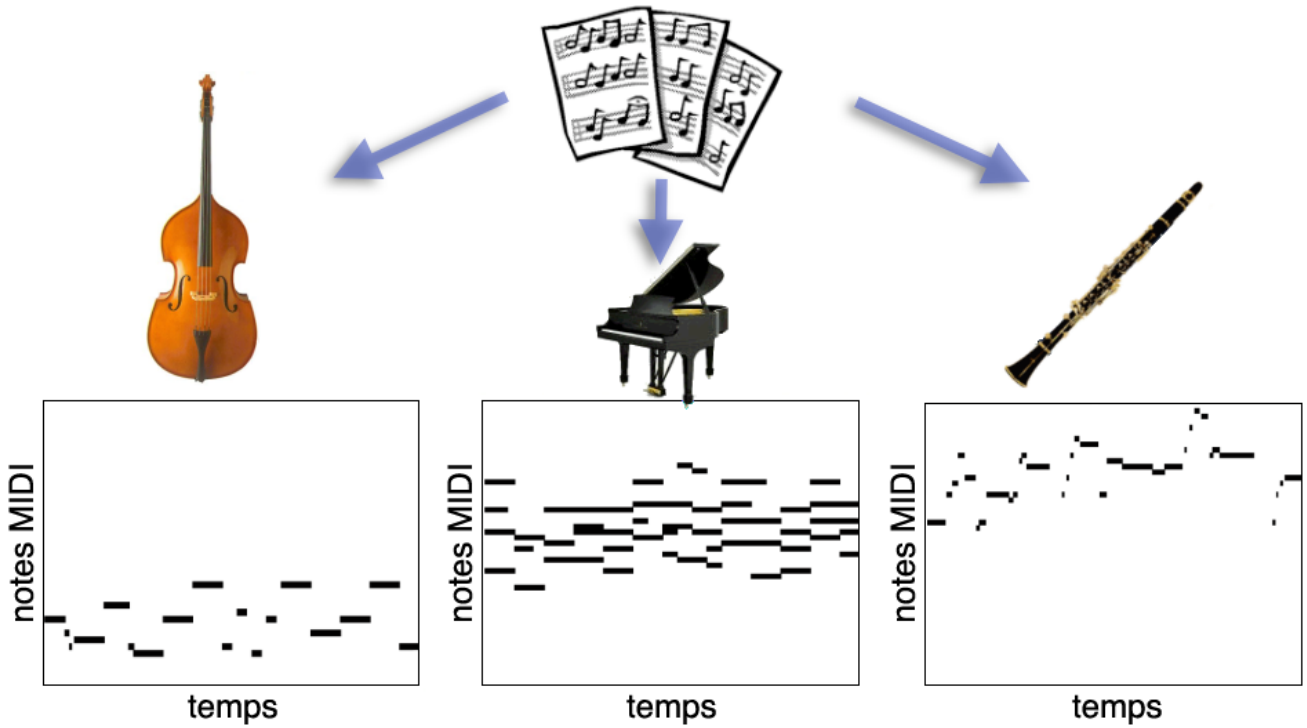
Score informed source separation [Hennequin et al., 2011c]



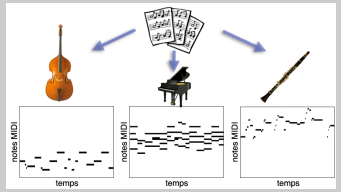
Mixture signal + Aligned score (MIDI)

We have a mixture signal and a MIDI file which is supposed to be aligned to the mixture signal: we do not deal with alignement issues which is a large and complex field of research (Cyril Joder, a former student of our lab, did its PhD about the problem of alignement).

Score informed source separation



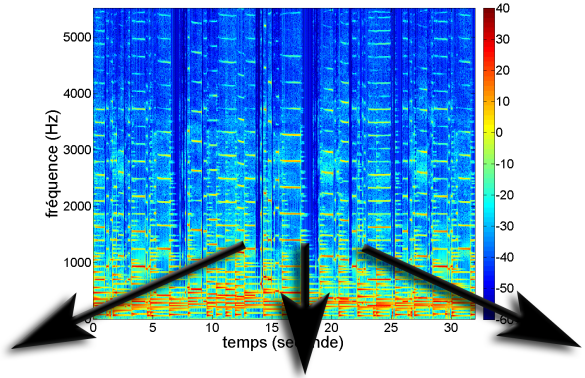
The score is used to initialize (and thus constrain) the decomposition.

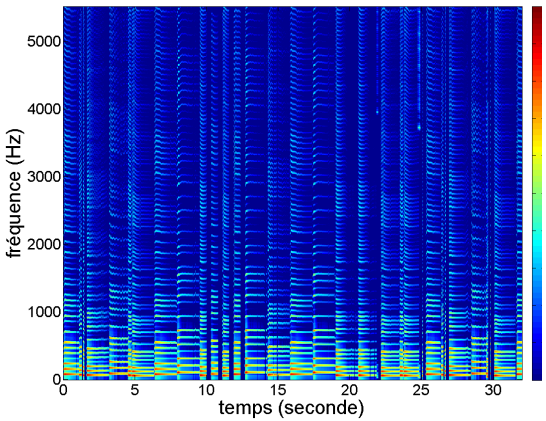


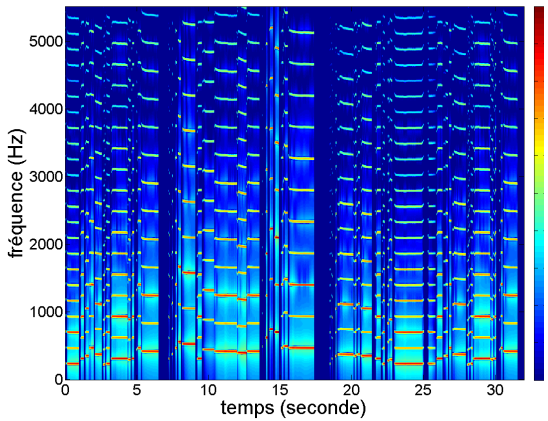
The score is used to initialize (and thus constrain) the decomposition.

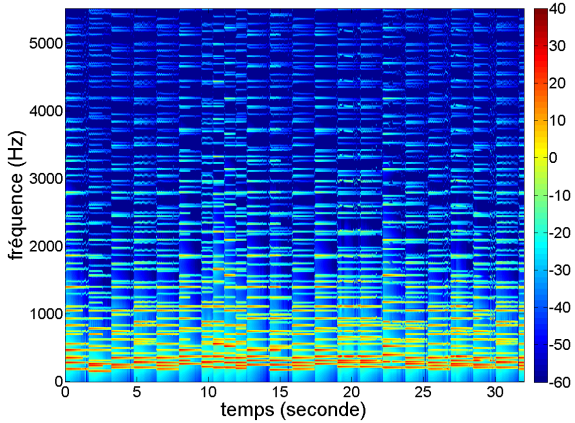
We use a generalization to multi-instrument sounds of the proposed model. As the activations of each instrument is very similar to a "Piano-Roll", the idea is to generate binary piano-roll from the aligned MIDI file and to use them as initialization for the activations. As we use multiplicative algorithms, initializing with zero is a hard constraint (Coefficients initialized to zero will remain zero).

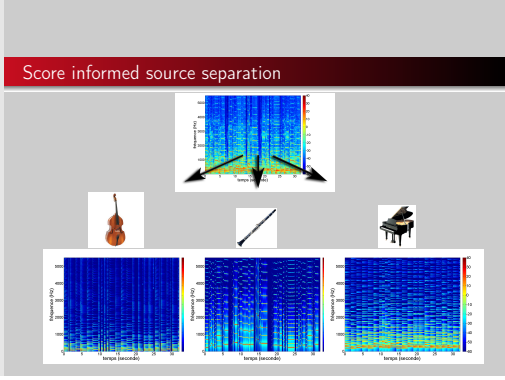
Score informed source separation











From the proposed initialization, the following parameters are estimated: the remaining activations, the fundamental frequency of each atom at each time were it is active, and the spectral shape (amplitudes of the harmonics) of the atoms. We thus obtain a time/frequency mask for each instrument that we can use to separate the signal of each instrument with Wiener filtering.

Contents

Scale-invariant decomposition

Model inspired by wavetable synthesis:

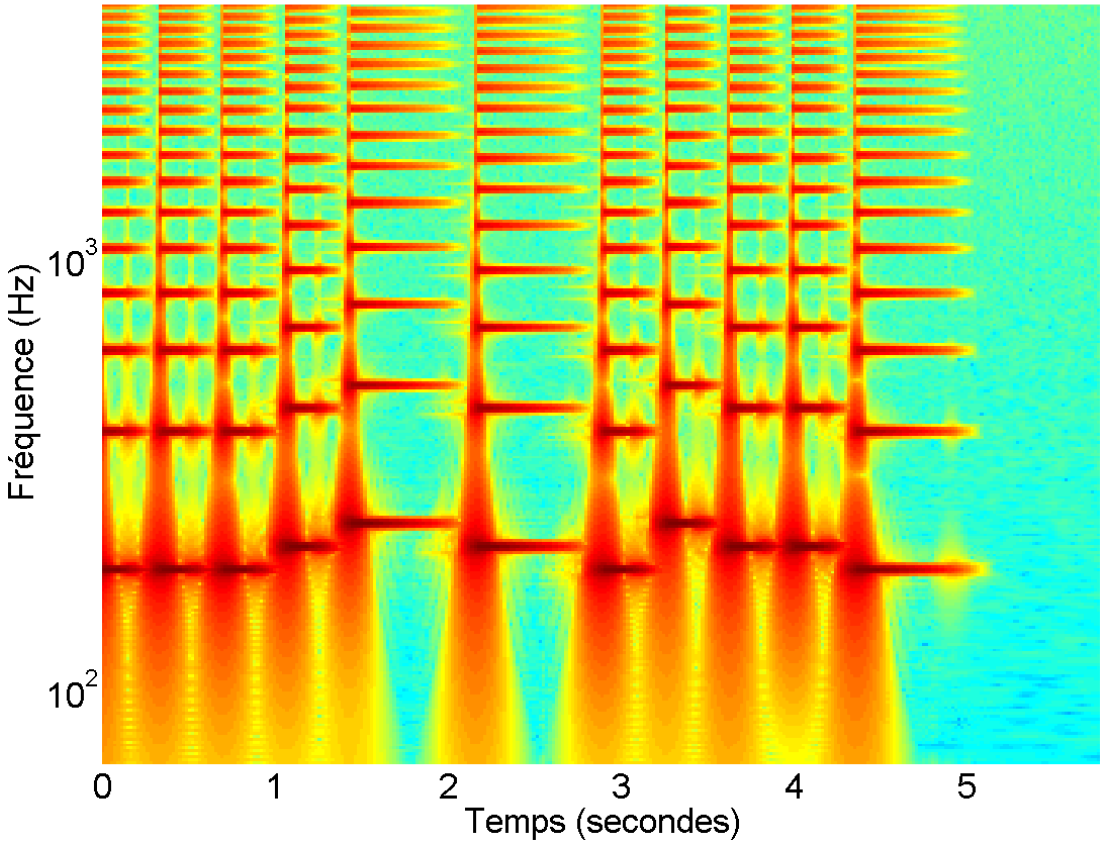
- A single atom to model all the notes of an instrument.
- Transformation of the atom to rebuild all the range of the instrument.

Model inspired by wavetable synthesis:

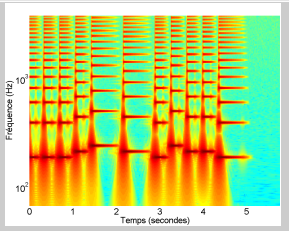
- A single atom to model all the notes of an instrument.
- Transformation of the atom to rebuild all the range of the instrument.

In this part, we propose another solution to the problem of modeling fundamental frequency variations in sound elements, which is more robust and permits to decompose more easily real signals. The proposed method is inspired by wavetable synthesis: in wavetable synthesis, a single short piece of waveform permits to synthesize the whole range (or a part of the range) of an instrument, by transforming it: the waveform is played at different speeds to create the different notes. In the proposed model, we try to extract a single atom to model all the notes (or at least a part of the notes) of an instrument: The notes are rebuilt by transforming the atom. This kind of method was already used on constant-Q spectrograms were the transformation used was a shift. We propose here a method which decomposes standard spectrograms (with a linear frequency resolution).

Shift-invariant decomposition



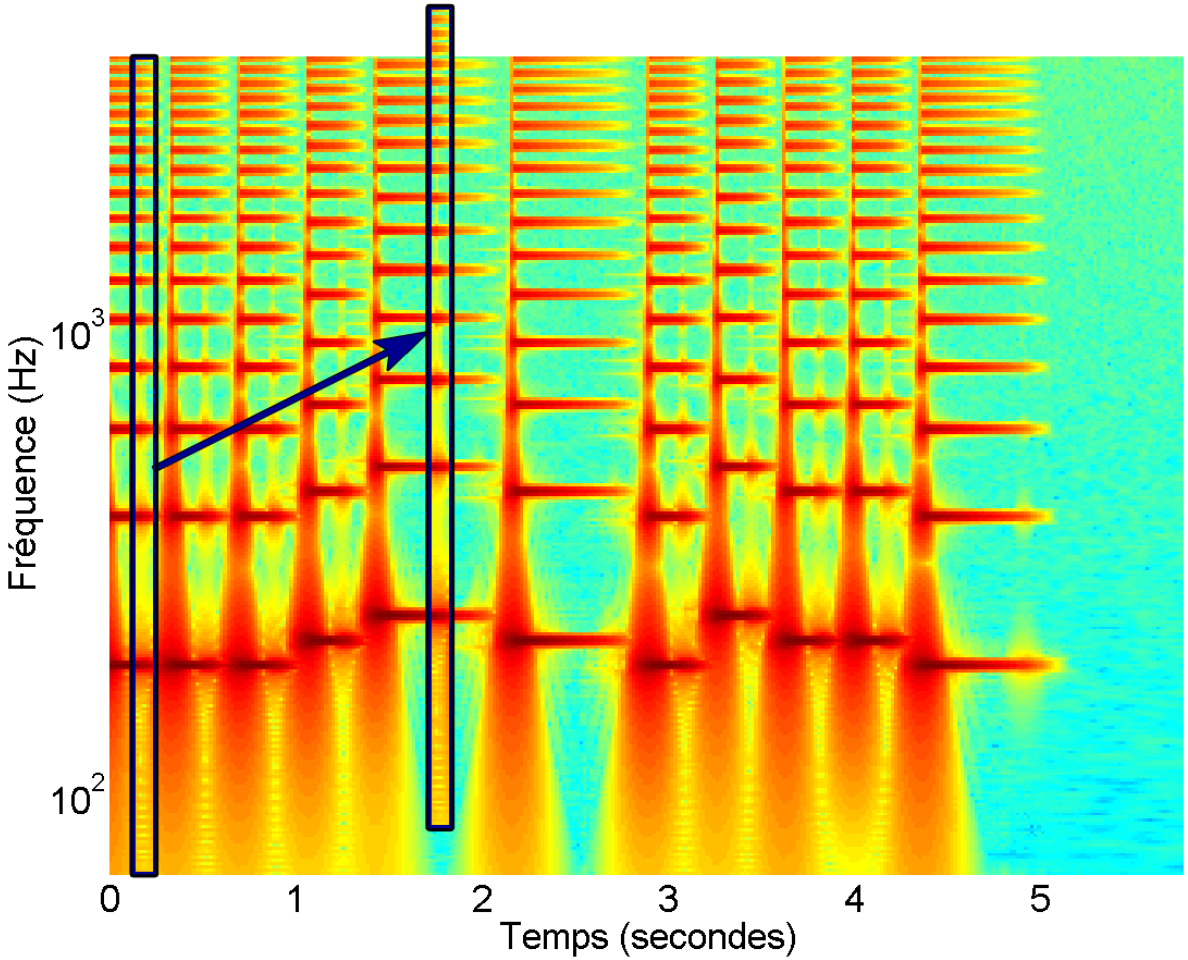
Constant-Q spectrogram of the french lullaby *Au Clair De La Lune* (played by a synthesizer).



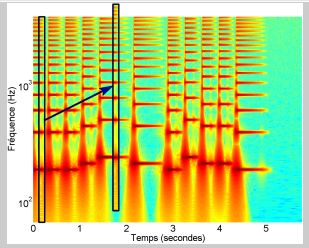
Constant-Q spectrogram of the french lullaby *Au Clair De La Lune* (played by a synthesizer).

To illustrate the shift-invariance in constant-Q spectrograms, we use a very simple extract.

Shift-invariant decomposition



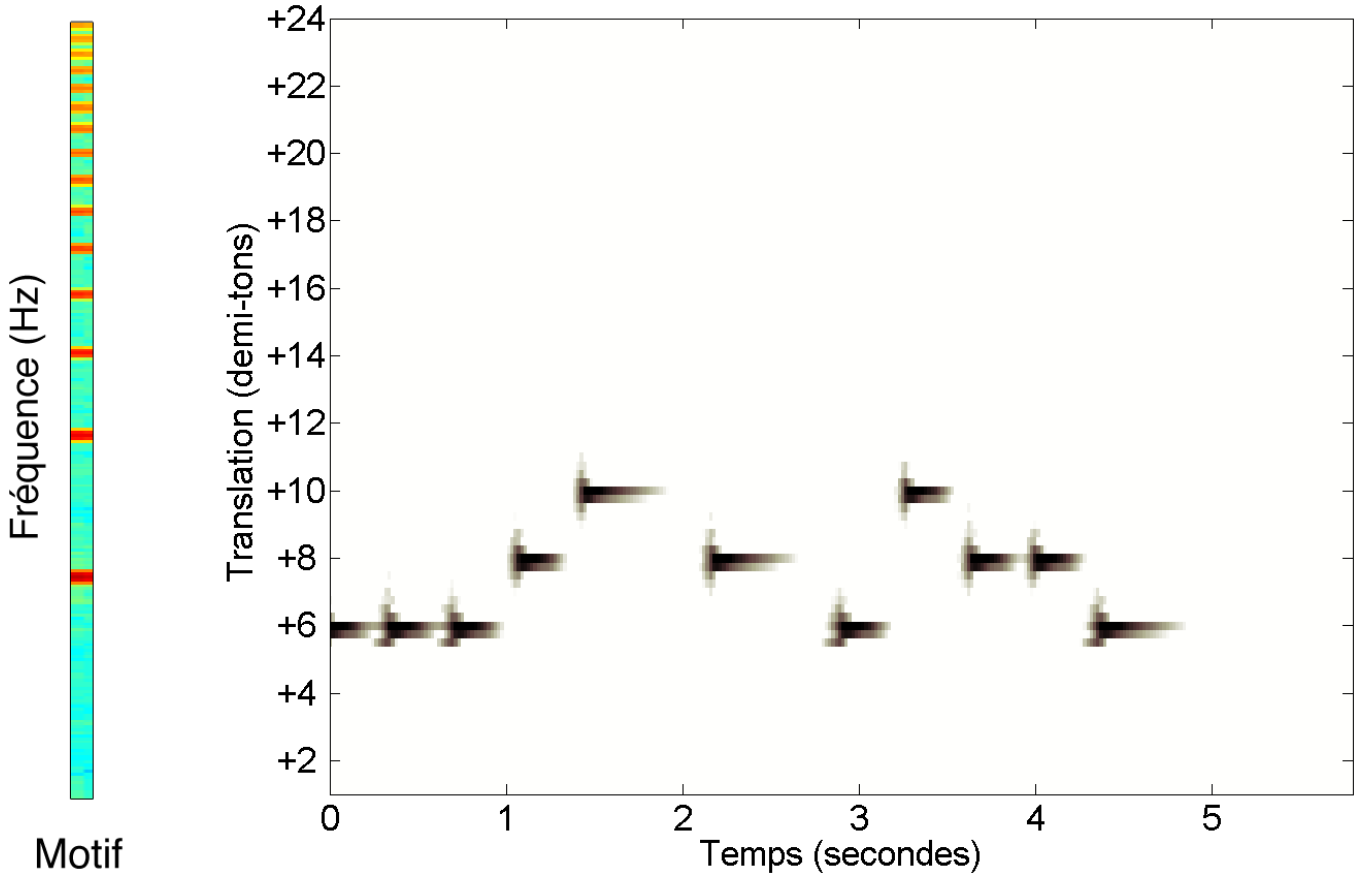
Shift-invariance in a constant-Q spectrogram.



Shift-invariance in a constant-Q spectrogram.

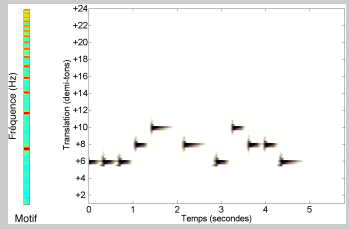
Here is the illustration of the shift-invariance: the shifted pattern of a note fits other notes.

Shift-invariant decomposition



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

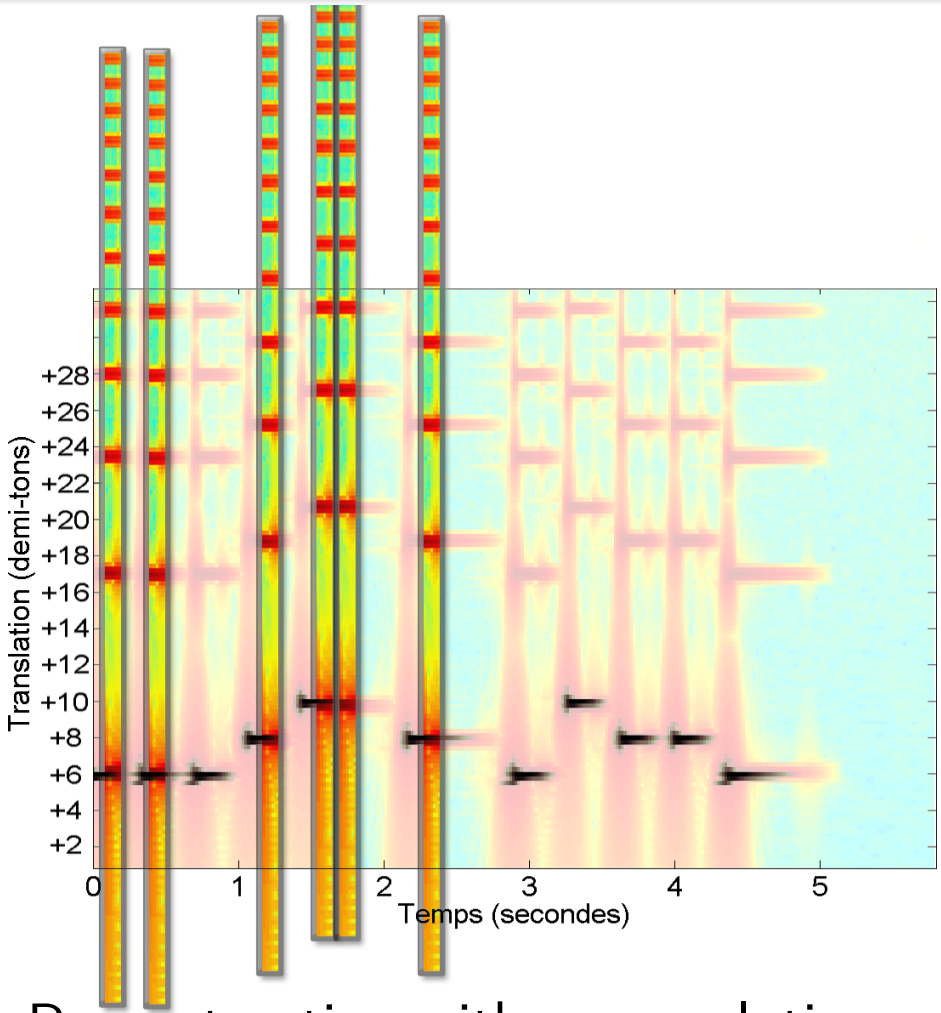
Shift-invariant decomposition



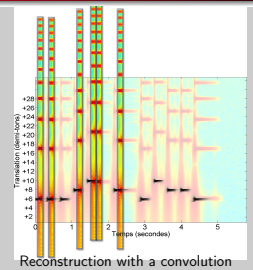
Decomposition: a single pattern and a map of the shifts (impulse distribution).

Thus, a compact decomposition consists in keeping only a single atom and building a map which provides the position where you must put these patterns.

Shift-invariant decomposition



Reconstruction with a convolution



Reconstruction with a convolution

The model spectrogram is then reconstructed with a simple 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]

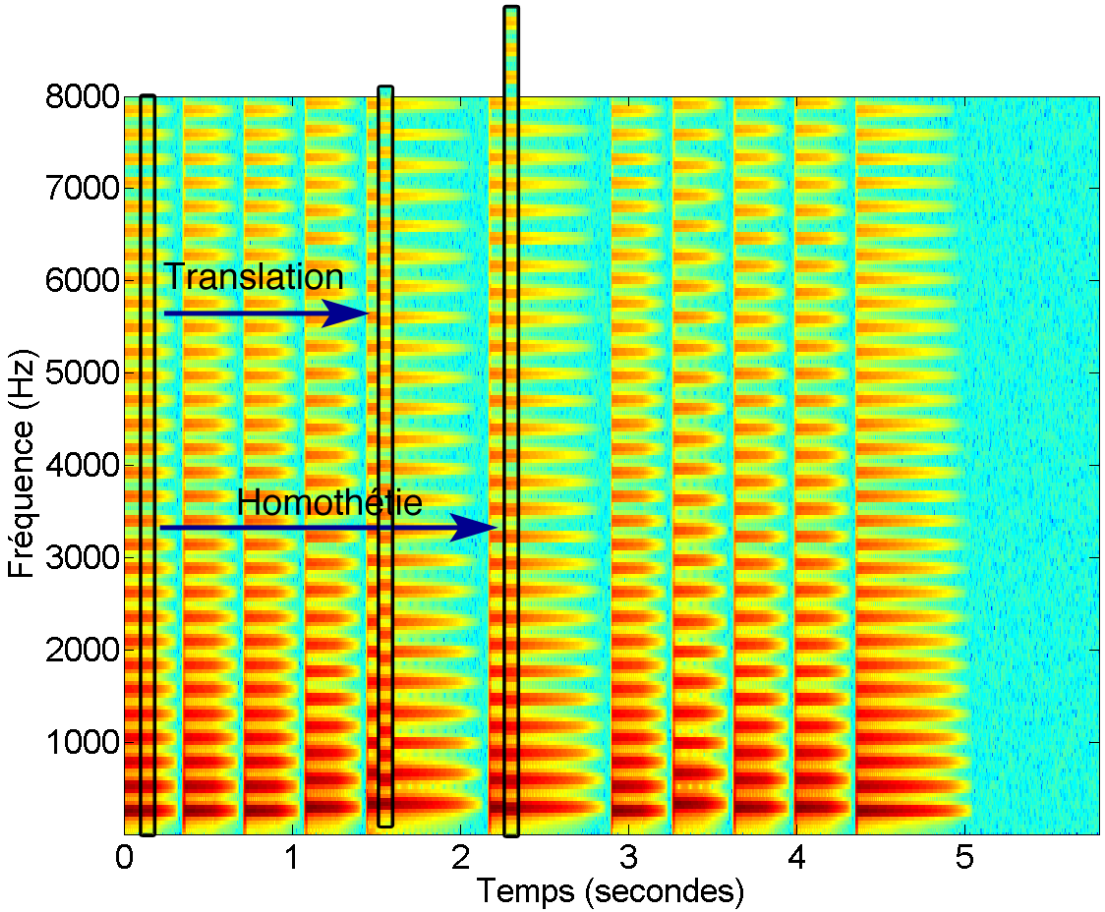
Principle

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

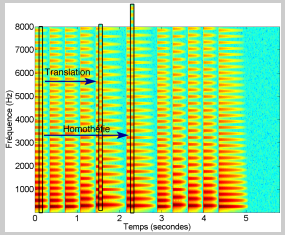
Probabilist Latent Component Analysis (PLCA)

- 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



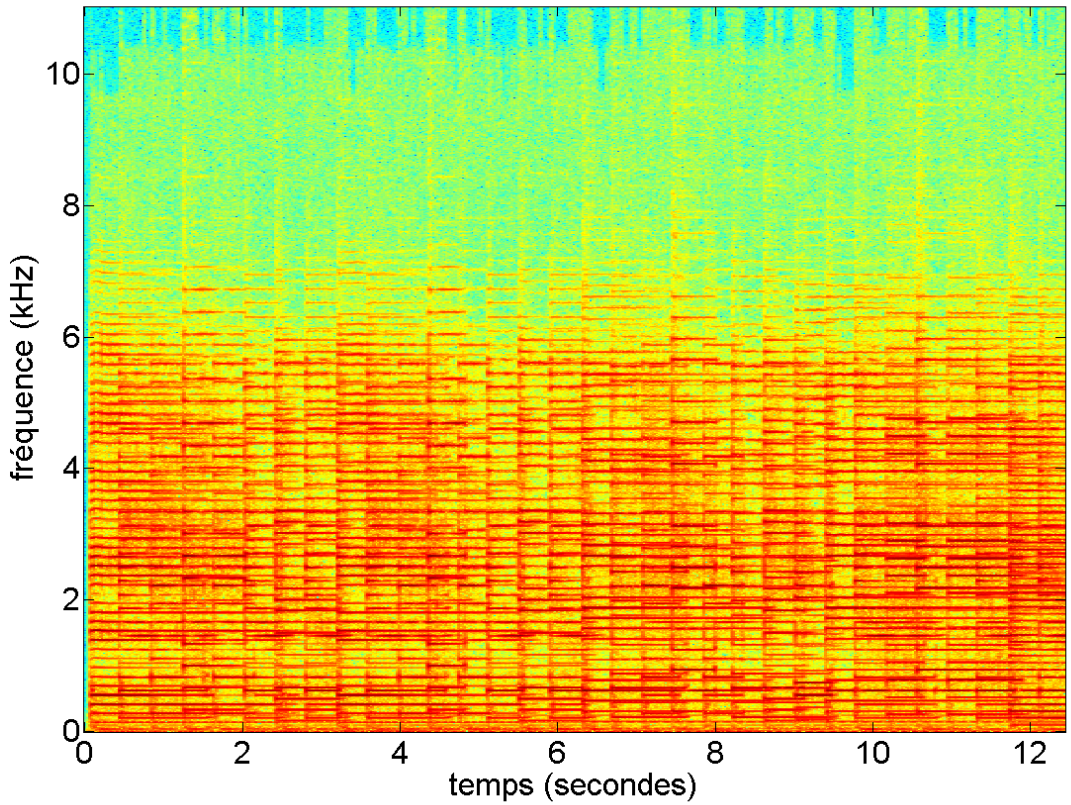
Scale-invariance in STFT spectrograms



Scale-invariance in STFT spectrograms

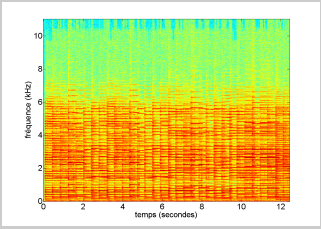
In a standard STFT spectrogram we can observe a scale invariance: the rescaled pattern of a note fits other notes.

Example



Spectrogram of the introduction of *Because* by the Beatles

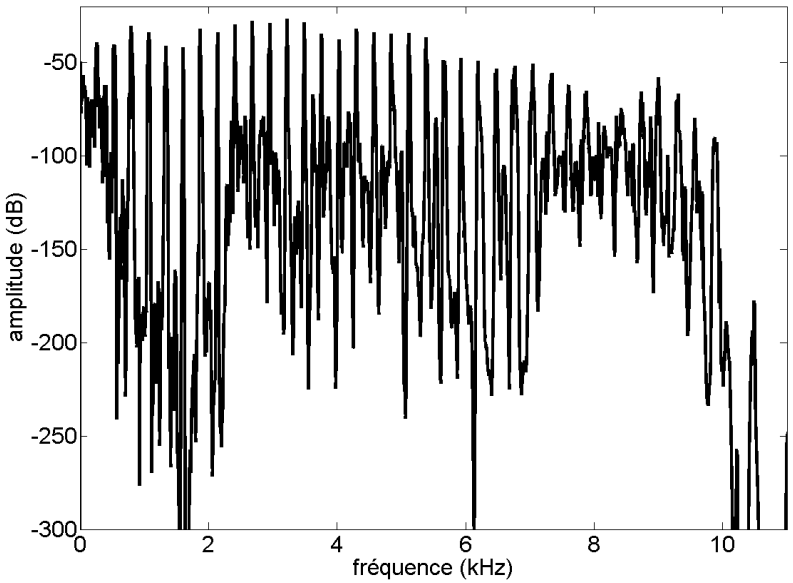
Example



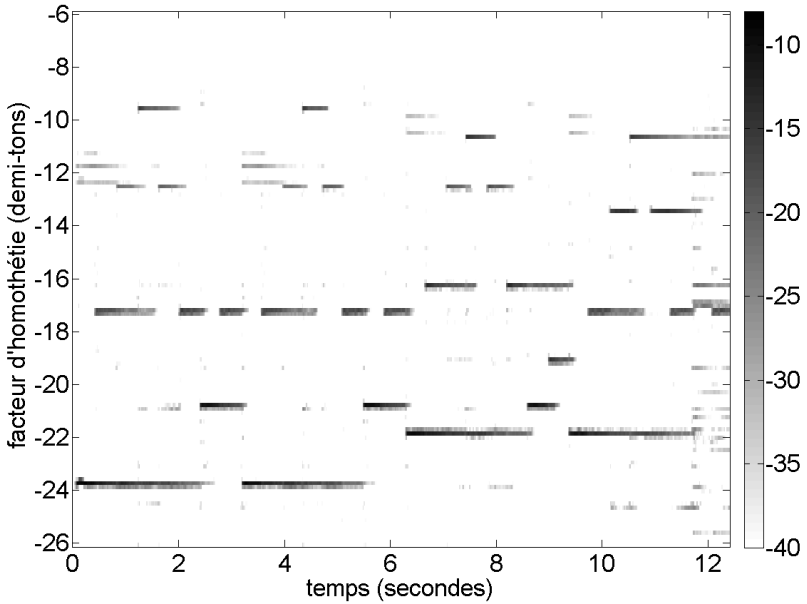
Spectrogram of the introduction of *Because* by the Beatles

To illustrate the decomposition, we decompose the spectrogram of the introduction of the song Because by the Beatles with a single atom.

Example

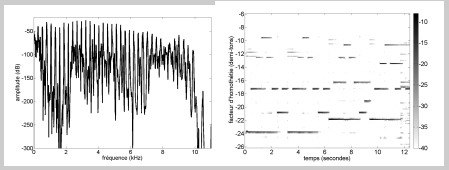


Pattern



Impulse distribution

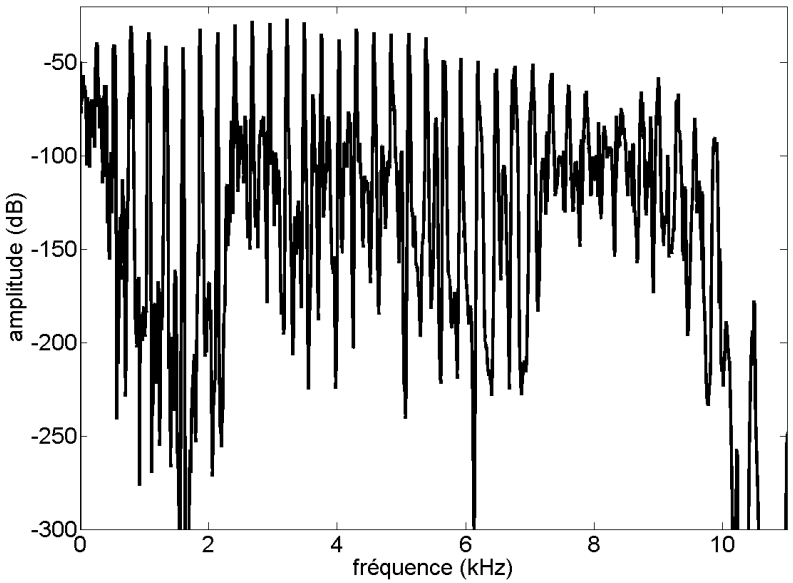
Example



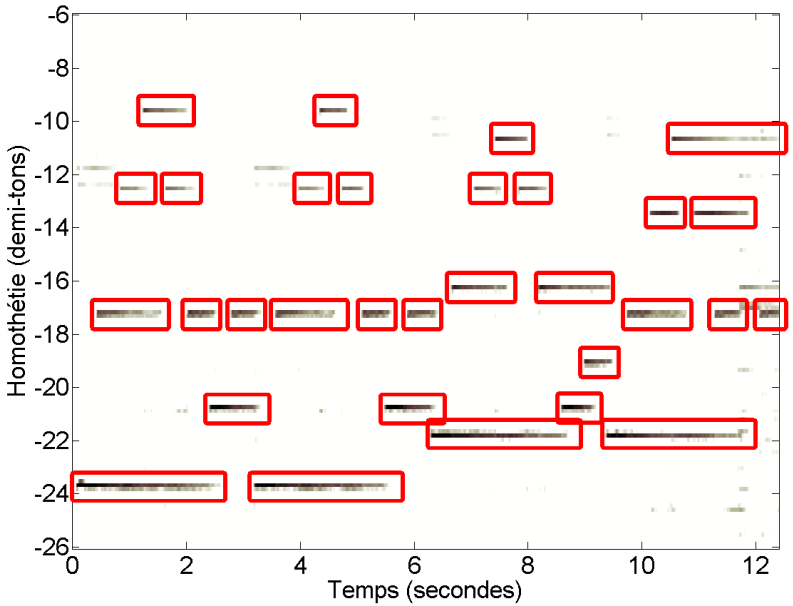
Pattern Impulse distribution

The decomposition provides an harmonic pattern and a “map of scaling” (an impulse distribution). The resolution of the homothety is logarithmic in order to match the natural repartition of musical notes.

Example

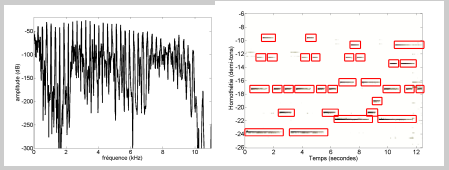


Pattern



Impulse distribution

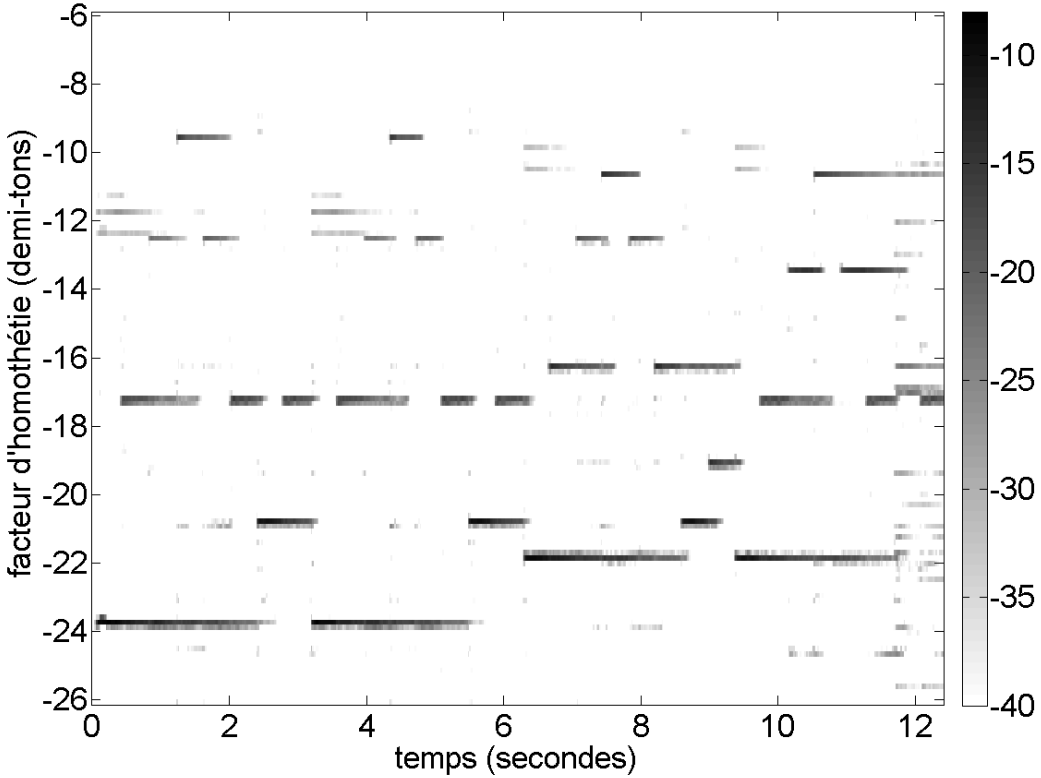
Example



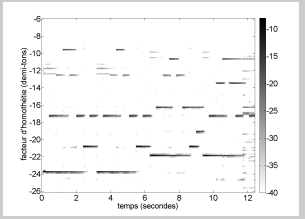
Pattern Impulse distribution

Actual notes are materialized with red rectangles. We can see that high values of the impulse distribution are mainly in these rectangles.

Modification of isolated notes in a polyphonic mixture



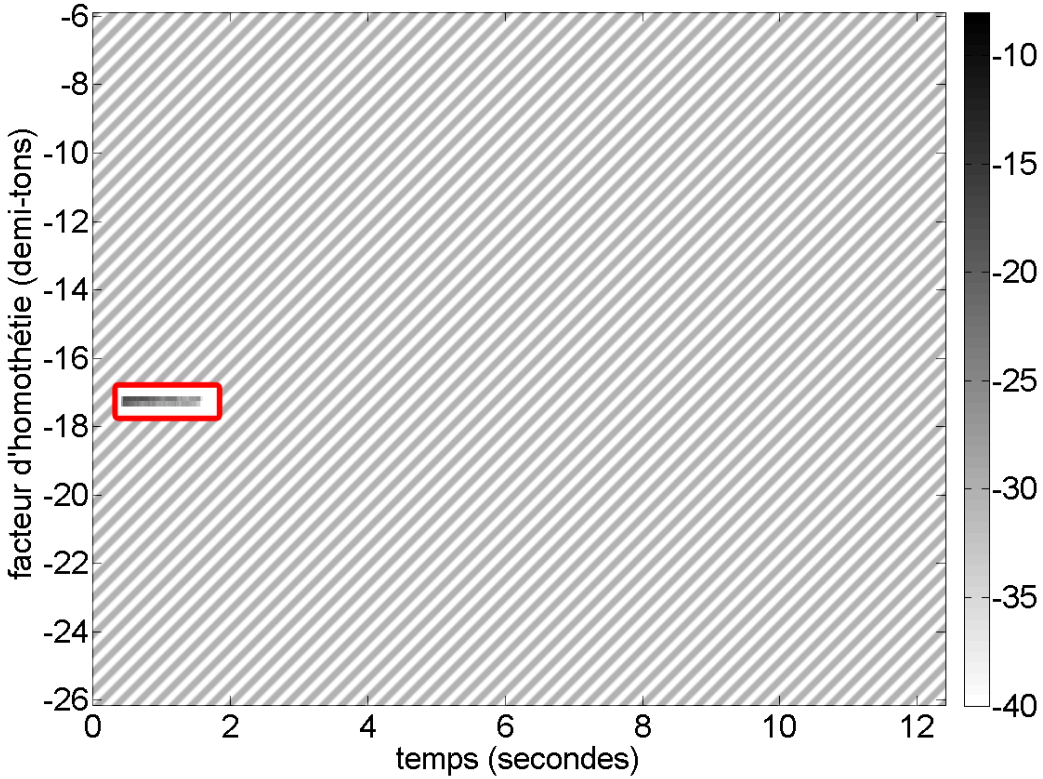
Impulse distribution of the introduction of *Because*



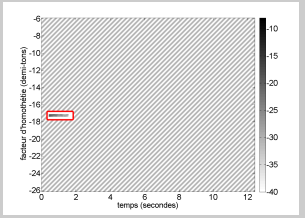
Impulse distribution of the introduction of *Because*

A straightforward application of this decomposition is proposed here: isolated notes are modified in a polyphonic mixture (again the introduction of Because).

Modification of isolated notes in a polyphonic mixture



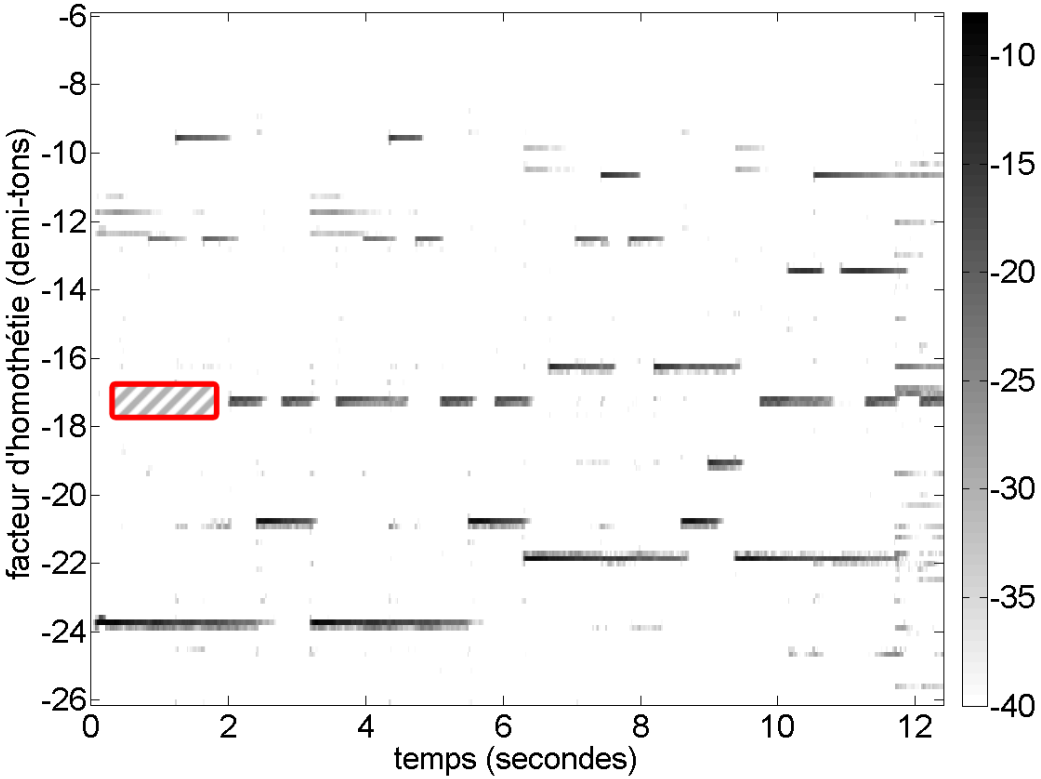
Isolated note



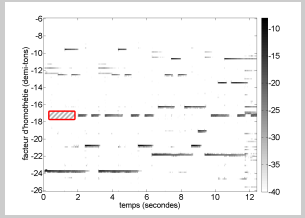
Isolated note

One can easily isolate a single note in the impulse distribution...

Modification of isolated notes in a polyphonic mixture



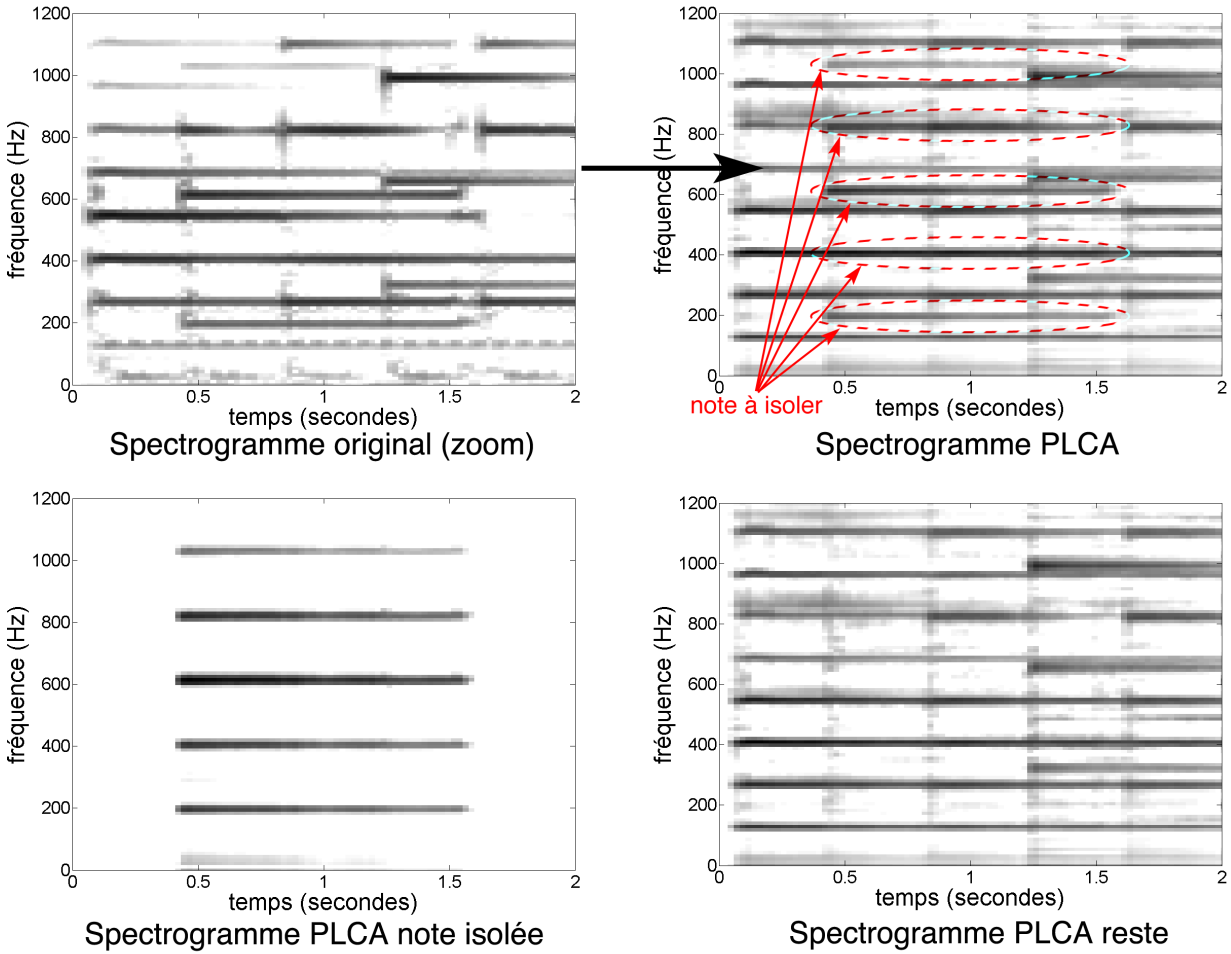
Rest

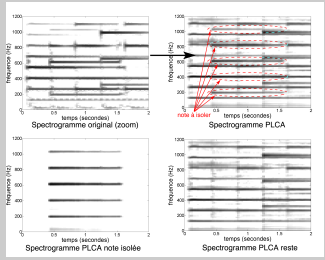


Rest

... from the rest.

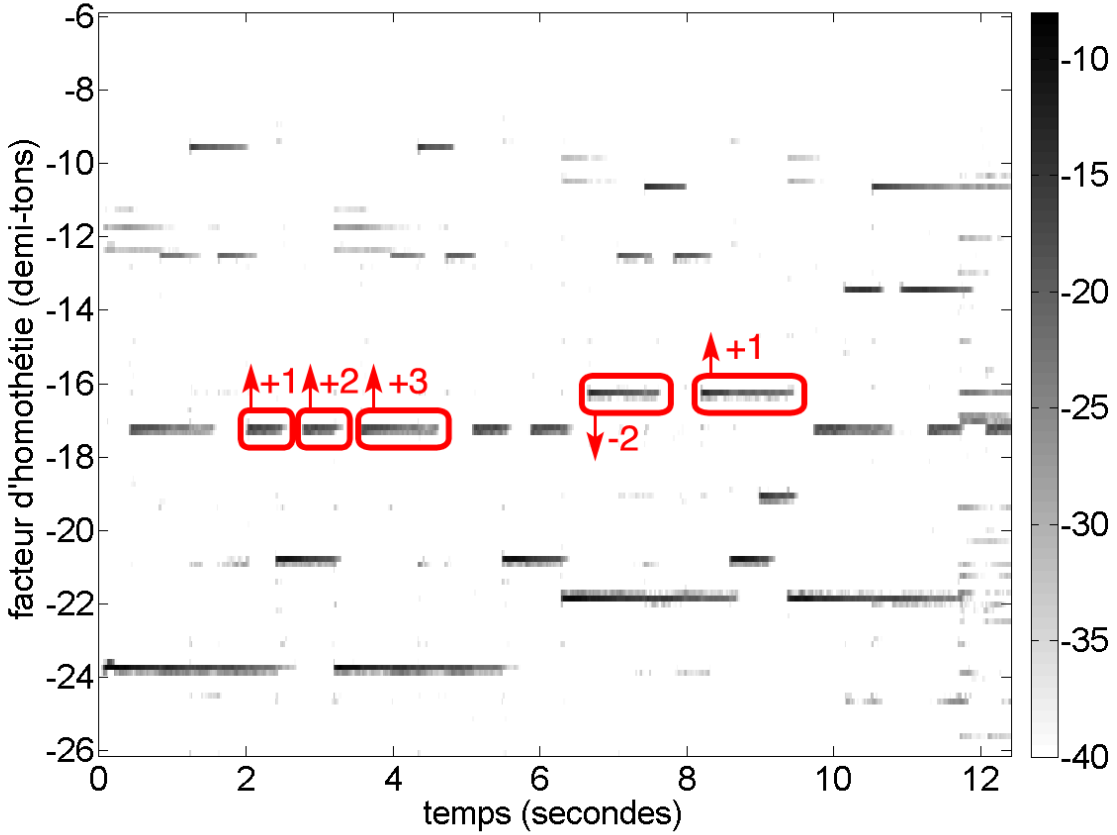
Modification of isolated notes in a polyphonic mixture

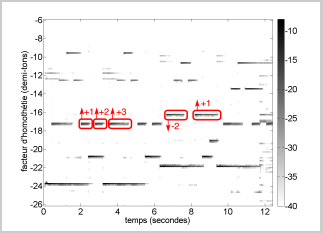




From these two impulse distributions, one can rebuild two time/frequency masks: one for the isolated note, the other for the rest. It makes it possible to separate the sound of this note from the rest with Wiener filtering.

Modification of isolated notes in a polyphonic mixture





Then, isolated notes can be modified independently. Here is an example of what we can do with this application: some notes were repitched individually in the introduction of Because (the modification is materialized by a red arrow. The amount of the modification in semi-tones is given for each modified note by the red number.).

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.