Probabilistic modelling of time-frequency representations with application to music signals

Roland Badeau
roland.badeau@eecs.qmul.ac.uk
C4DM, Wednesday, March 6, 2013

Introduction

- NMF applied to time-frequency distributions:
  - is a powerful tool for modelling music signals
  - has many applications in audio signal processing
- Most probabilistic models for NMF:
  - \+ permit to exploit some a priori knowledge
  - \- do not take phase into account
  - \- assume that all time-frequency bins are independent
- The proposed HR-NMF model:
  - takes phases and local correlations into account
  - achieves high spectral resolution

Introduction

- NMF applied to time-frequency distributions:
  - is a powerful tool for modelling music signals
  - has many applications in audio signal processing
- Most probabilistic models for NMF:
  - \+ permit to exploit some a priori knowledge
  - \- do not take phase into account
  - \- assume that all time-frequency bins are independent
- The proposed HR-NMF model:
  - takes phases and local correlations into account
  - achieves high spectral resolution

Introduction

- NMF applied to time-frequency distributions:
  - is a powerful tool for modelling music signals
  - has many applications in audio signal processing
- Most probabilistic models for NMF:
  - + permit to exploit some a priori knowledge
  - - do not take phase into account
  - - assume that all time-frequency bins are independent
- The proposed HR-NMF model:
  - takes phases and local correlations into account
  - achieves high spectral resolution

TELECOM ParisTech

Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Outline

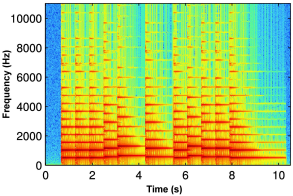
- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Non-negative Matrix Factorization (NMF)

Musical score

Non-negative Matrix Factorization (NMF)

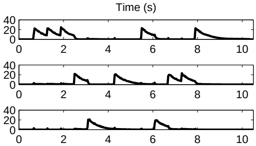
Musical score



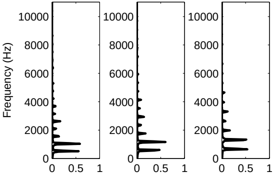
Spectrogram ***V***

Non-negative Matrix Factorization (NMF)

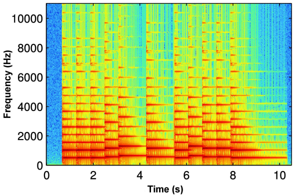
Musical score



Temporal activations ***H***



Spectral templates ***W***



Spectrogram ***V***

Non-negative Matrix Factorization (NMF)

- Factorization of a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F \times T}$ as a product $\boldsymbol{V} \approx \boldsymbol{W}\,\boldsymbol{H}$
- Rank reduction: $\boldsymbol{W} \in \mathbb{R}_+^{F \times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K \times T}$ where $K < \min(F, T)$
- Usual applications:
  - Image analysis, data mining, spectroscopy, finance, *etc.*
  - Audio signal processing:
    - Multi-pitch estimation, onset detection
    - Automatic music transcription
    - Musical instrument recognition
    - Source separation
    - Audio inpainting

Non-negative Matrix Factorization (NMF)

- Factorization of a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F \times T}$ as a product $\boldsymbol{V} \approx \boldsymbol{W}\,\boldsymbol{H}$
- Rank reduction: $\boldsymbol{W} \in \mathbb{R}_+^{F \times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K \times T}$ where $K < \min(F, T)$
- Usual applications:
  - Image analysis, data mining, spectroscopy, finance, *etc.*
  - Audio signal processing:
    - Multi-pitch estimation, onset detection
    - Automatic music transcription
    - Musical instrument recognition
    - Source separation
    - Audio inpainting

Non-negative Matrix Factorization (NMF)

- Factorization of a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F \times T}$ as a product $\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}$
- Rank reduction: $\boldsymbol{W} \in \mathbb{R}_+^{F \times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K \times T}$ where $K < \min(F, T)$
- Usual applications:
  - Image analysis, data mining, spectroscopy, finance, *etc.*
  - Audio signal processing:
    - Multi-pitch estimation, onset detection
    - Automatic music transcription
    - Musical instrument recognition
    - Source separation
    - Audio inpainting

Non-negative Matrix Factorization (NMF)

- Factorization of a matrix $\boldsymbol{V} \in \mathbb{R}_+^{F \times T}$ as a product $\boldsymbol{V} \approx \boldsymbol{W}\,\boldsymbol{H}$
- Rank reduction: $\boldsymbol{W} \in \mathbb{R}_+^{F \times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K \times T}$ where $K < \min(F, T)$
- Usual applications:
  - Image analysis, data mining, spectroscopy, finance, *etc.*
  - Audio signal processing:
    - Multi-pitch estimation, onset detection
    - Automatic music transcription
    - Musical instrument recognition
    - Source separation
    - Audio inpainting

NMF probabilistic models

- Mixture models with (hidden) latent variables
  - + can exploit a priori knowledge
  - + can use well-known statistical inference techniques
- Probabilistic models of time-frequency distributions:
  - Additive Gaussian noise [Schmidt 2008],
  - Probabilistic Latent Component Analysis [Smaragdis 2006],
  - Mixture of Poisson components [Virtanen 2008],
  - Mixture of Gaussian components [Févotte 2009],
    - + Only model taking the existence of phase into account, and justifying the use of Wiener filtering for separating the components

NMF probabilistic models

- Mixture models with (hidden) latent variables
  - + can exploit a priori knowledge
  - + can use well-known statistical inference techniques
- Probabilistic models of time-frequency distributions:
  - Additive Gaussian noise [Schmidt 2008],
  - Probabilistic Latent Component Analysis [Smaragdis 2006],
  - Mixture of Poisson components [Virtanen 2008],
  - Mixture of Gaussian components [Févotte 2009],
    - + Only model taking the existence of phase into account, and justifying the use of Wiener filtering for separating the components

Gaussian model (IS-NMF) [Févotte 2009]

$$c_{ft}^{(k)} \sim \mathcal{N}\left(\mathbf{0},\ \hat{v}_{ft}^{(k)} = w_{fk}\, h_{kt}\right)$$

$\boldsymbol{C}^{(k)}$ — w_{fk} — h_{kt} — $\hat{\boldsymbol{V}}^{(k)}$

all time-frequency bins are independent

Gaussian model (IS-NMF) [Févotte 2009]

$$\mathbf{C}^{(k)}: \quad c_{ft}^{(k)} \sim \mathcal{N}\left(\mathbf{0},\; w_{fk}\, h_{kt}\right), \qquad \hat{\mathbf{V}}^{(k)}: \quad \hat{v}_{ft}^{(k)} = w_{fk}\, h_{kt}$$

all time-frequency bins are independent

$$\Downarrow$$

$$\mathbf{X}: \quad x_{ft} = \sum_{k=1}^{K} c_{ft}^{(k)} \sim \mathcal{N}\left(\mathbf{0},\; \hat{v}_{ft} = \sum_{k=1}^{K} \hat{v}_{ft}^{(k)}\right), \qquad \hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$$

Gaussian model (IS-NMF) [Févotte 2009]

$$\underbrace{c_{ft}^{(k)}}_{\mathbf{C}^{(k)}} \sim \mathcal{N}\left(0,\; w_{fk}\, h_{kt}\right), \qquad \underbrace{\hat{v}_{ft}^{(k)} = w_{fk}\, h_{kt}}_{\hat{\mathbf{V}}^{(k)}}$$

all time-frequency bins are independent

$$\Downarrow$$

$$\underbrace{x_{ft} = \sum_{k=1}^{K} c_{ft}^{(k)}}_{\mathbf{X}} \sim \mathcal{N}\left(0,\; \hat{v}_{ft}\right), \qquad \underbrace{w_{fk}}_{\mathbf{W}}, \quad \underbrace{h_{kt}}_{\mathbf{H}}, \qquad \hat{v}_{ft} = \sum_{k=1}^{K} \hat{v}_{ft}^{(k)}, \qquad \hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$$

$$\max L(\mathbf{X}) \;\Updownarrow\; \min D_{IS}(\mathbf{V} = |\mathbf{X}|^2 \,|\, \hat{\mathbf{V}})$$

A priori knowledge in probabilistic models

- Various kinds of a priori knowledge:
  - Harmonicity [Virtanen 2008, Vincent 2008. . .]
  - Smoothness of spectral envelopes [Schmidt 2008, Vincent 2008. . .]
  - Smoothness of temporal activations [Virtanen 2008, Févotte 2009. . .]
  - Spectral or temporal sparsity [Schmidt 2008, Smaragdis 2009. . .]
- Standard approaches:
  - Parametrisation of $\boldsymbol{W}$ and / or $\boldsymbol{H}$
  - Use of a predefined dictionary $\boldsymbol{W}$ (parametric or non-parametric, learned beforehand)
  - Bayesian methods (a priori distribution of the parameters)

A priori knowledge in probabilistic models

- Various kinds of a priori knowledge:
  - Harmonicity [Virtanen 2008, Vincent 2008...]
  - Smoothness of spectral envelopes [Schmidt 2008, Vincent 2008...]
  - Smoothness of temporal activations [Virtanen 2008, Févotte 2009...]
  - Spectral or temporal sparsity [Schmidt 2008, Smaragdis 2009...]
- Standard approaches:
  - Parametrisation of $\boldsymbol{W}$ and / or $\boldsymbol{H}$
  - Use of a predefined dictionary $\boldsymbol{W}$ (parametric or non-parametric, learned beforehand)
  - Bayesian methods (a priori distribution of the parameters)

Analysis levels

Time-frequency distribution	Probabilistic model of time-frequency distribution	A priori knowledge
Transformation of the data	Low-level model	High-level model

- The low-level model raises several issues:
  - Phase is not (or insufficiently) taken into account
  - Sinusoids are not modelled as such (they cannot be properly separated by Wiener filtering)
  - All time-frequency bins are assumed independent

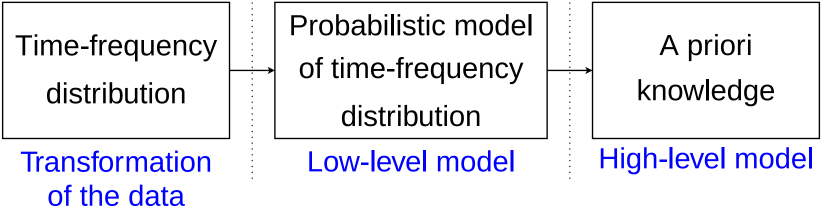
Analysis levels

Time-frequency distribution → Probabilistic model of time-frequency distribution → A priori knowledge

Transformation of the data | Low-level model | High-level model

- The low-level model raises several issues:
  - Phase is not (or insufficiently) taken into account
  - Sinusoids are not modelled as such (they cannot be properly separated by Wiener filtering)
  - All time-frequency bins are assumed independent

Analysis levels



- The low-level model raises several issues:
  - Phase is not (or insufficiently) taken into account
  - Sinusoids are not modelled as such (they cannot be properly separated by Wiener filtering)
  - All time-frequency bins are assumed independent

Analysis levels

Time-frequency distribution → Probabilistic model of time-frequency distribution → A priori knowledge

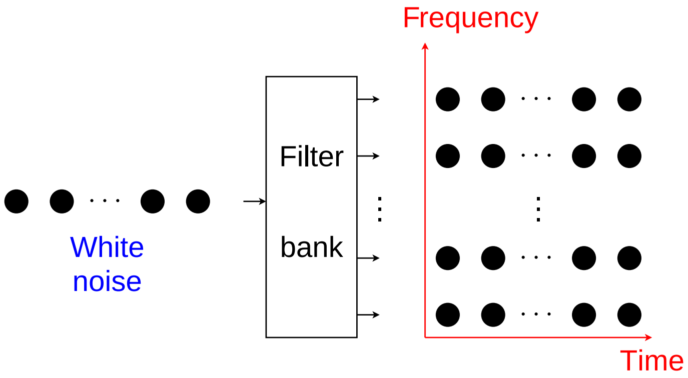
Transformation of the data | Low-level model | High-level model

- The low-level model raises several issues:
  - Phase is not (or insufficiently) taken into account
  - Sinusoids are not modelled as such (they cannot be properly separated by Wiener filtering)
  - All time-frequency bins are assumed independent

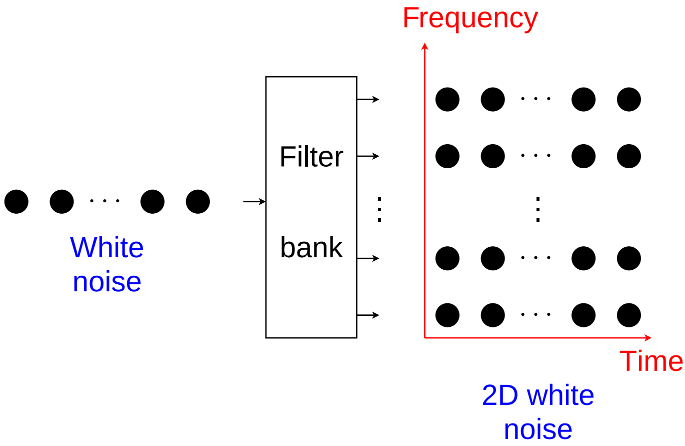
Outline

- Advantages and drawbacks of NMF probabilistic models
- **Choosing an appropriate TF representation**
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

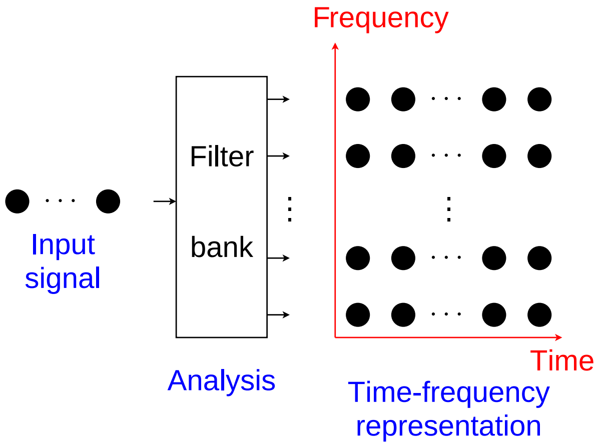
Preservation of whiteness (PW)



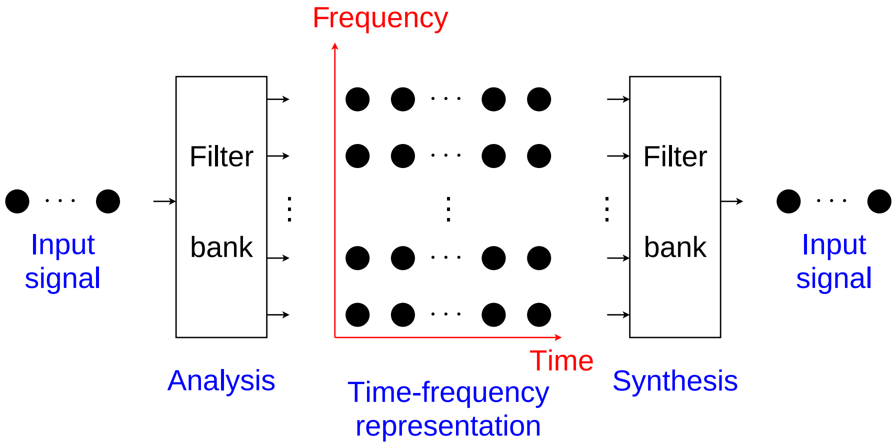
Preservation of whiteness (PW)



Perfect reconstruction (PR)

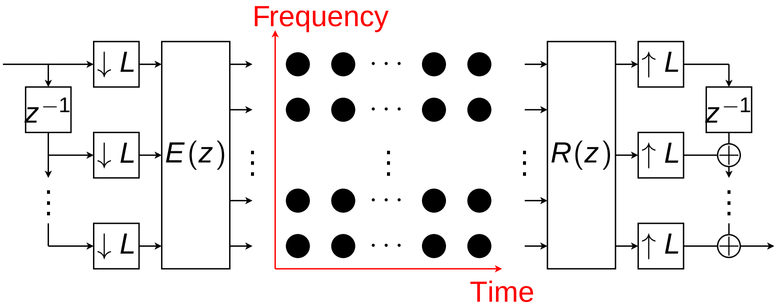


Perfect reconstruction (PR)



Solution of (PW) + (PR)

- Critically sampled paraunitary filter banks: $R(z) = \widetilde{E}(z)$

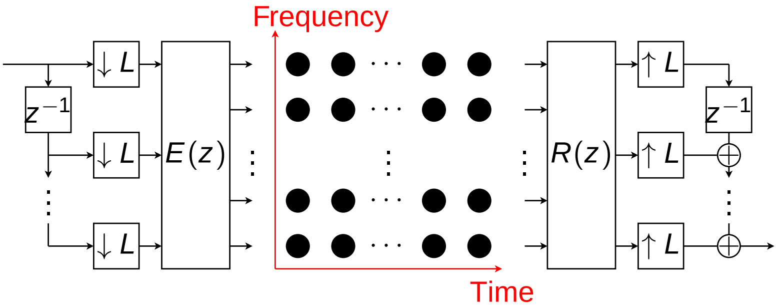


$\longrightarrow$ DFT (complex Gaussian processes)
$\longrightarrow$ MDCT (real Gaussian processes)

- "Decorrelating" effect onto a stationary process

Solution of (PW) + (PR)

- Critically sampled paraunitary filter banks: $R(z) = \widetilde{E}(z)$



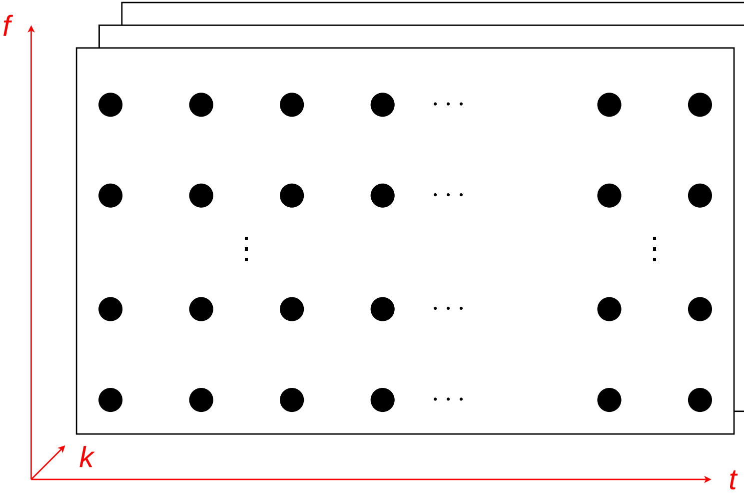
$\longrightarrow$ DFT (complex Gaussian processes)
$\longrightarrow$ MDCT (real Gaussian processes)

- "Decorrelating" effect onto a stationary process

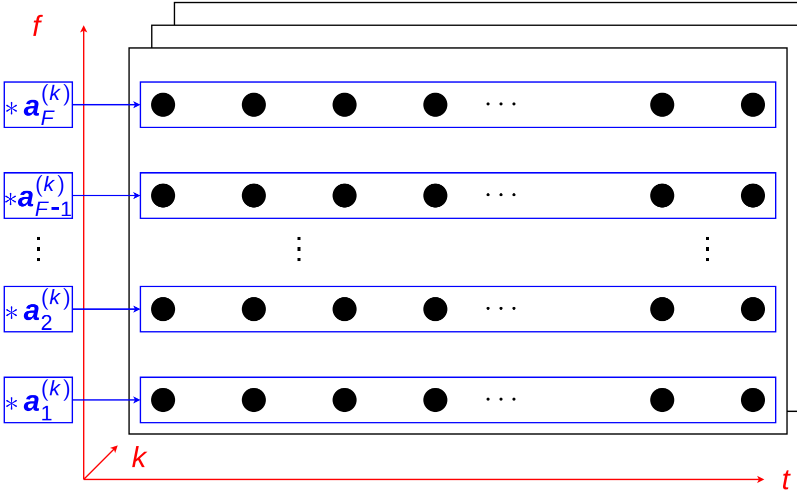
Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- **Modelling phases and correlations in the TF domain**
  - **HR-NMF model**
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

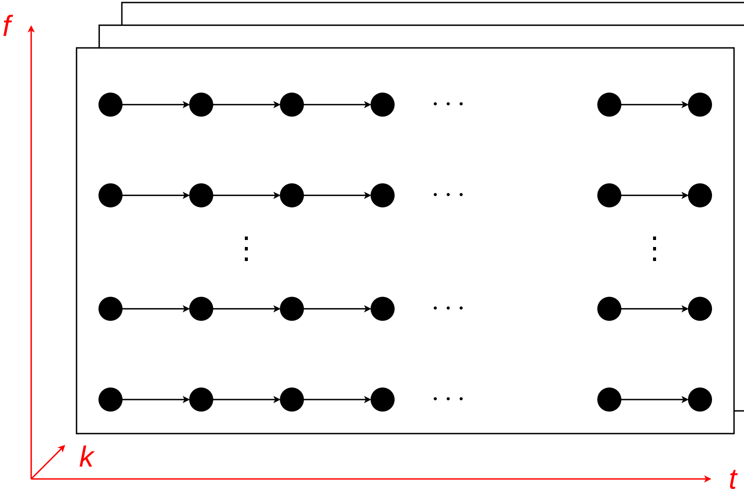
Graphical model of IS-NMF ($X \sim \mathcal{N}(0, WH)$)



Autoregressive filtering of the channels

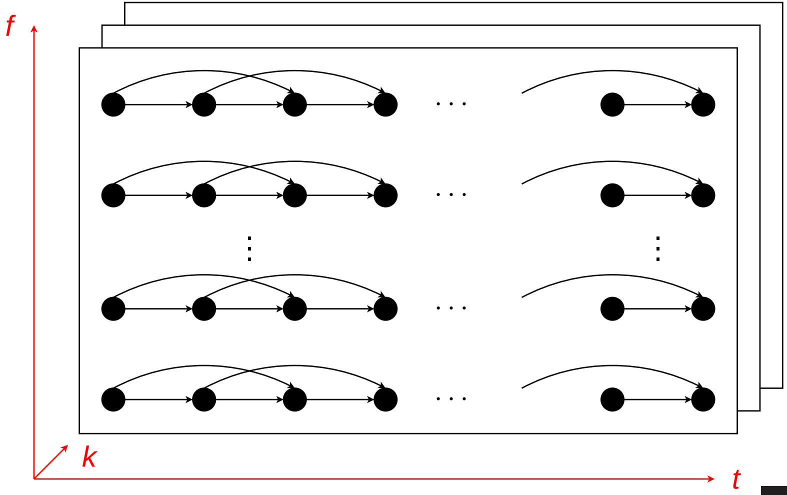


Graphical model of HR-NMF (AR1)



TELECOM ParisTech

Graphical model of HR-NMF (AR2)



HR-NMF model

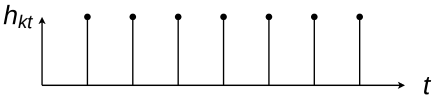
- Frequency bands are independent and non-stationary
- Particular cases:
  - IS-NMF model
  - Autoregressive process
  - Exponential Sinusoidal Model (ESM)

HR-NMF model

- Frequency bands are independent and non-stationary
- Particular cases:
  - IS-NMF model
  - Autoregressive process
  - Exponential Sinusoidal Model (ESM)

HR-NMF model

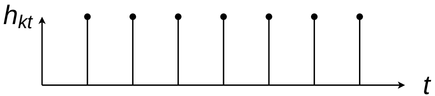
- Frequency bands are independent and non-stationary
- Particular cases:
  - IS-NMF model
  - Autoregressive process



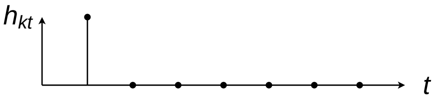
  - Exponential Sinusoidal Model (ESM)

HR-NMF model

- Frequency bands are independent and non-stationary
- Particular cases:
  - IS-NMF model
  - Autoregressive process



  - Exponential Sinusoidal Model (ESM)



Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- **Modelling phases and correlations in the TF domain**
  - HR-NMF model
  - **Algorithms**
- Preliminary results
  - Audio source separation
  - Audio inpainting
- Conclusions

Maximum likelihood estimation

- Expectation-Maximization (EM) algorithm:
  - E-step:
    - Kalman filtering with smoothing (forward-backward)
    - Complexity: $O(FTK^3P^3)$
  - M-step:
    - Iterative algorithm which switches between $(\boldsymbol{W}, \boldsymbol{a})$ and $\boldsymbol{H}$
    - Complexity: $O(FTKP^2)$
- Processing realistic data requires faster algorithms:
  - Improve the convergence speed
  - Reduce the computational complexity

[1] Roland Badeau. "Gaussian modelling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October 2011.

Maximum likelihood estimation

- Expectation-Maximization (EM) algorithm:
  - E-step:
    - Kalman filtering with smoothing (forward-backward)
    - Complexity: $O(FTK^3P^3)$
  - M-step:
    - Iterative algorithm which switches between $(\boldsymbol{W}, \boldsymbol{a})$ and $\boldsymbol{H}$
    - Complexity: $O(FTKP^2)$
- Processing realistic data requires faster algorithms:
  - Improve the convergence speed
  - Reduce the computational complexity

[1] Roland Badeau. "Gaussian modelling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October 2011.

Maximum likelihood estimation

- Expectation-Maximization (EM) algorithm:
  - E-step:
    - Kalman filtering with smoothing (forward-backward)
    - Complexity: $O(FTK^3P^3)$
  - M-step:
    - Iterative algorithm which switches between $(\boldsymbol{W}, \boldsymbol{a})$ and $\boldsymbol{H}$
    - Complexity: $O(FTKP^2)$
- Processing realistic data requires faster algorithms:
  - Improve the convergence speed
  - Reduce the computational complexity

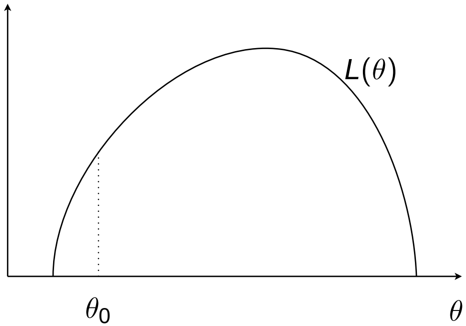
[1] Roland Badeau. "Gaussian modelling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October 2011.

Maximum likelihood estimation

- Expectation-Maximization (EM) algorithm:
  - E-step:
    - Kalman filtering with smoothing (forward-backward)
    - Complexity: $O(FTK^3P^3)$
  - M-step:
    - Iterative algorithm which switches between $(\boldsymbol{W}, \boldsymbol{a})$ and $\boldsymbol{H}$
    - Complexity: $O(FTKP^2)$
- Processing realistic data requires faster algorithms:
  - Improve the convergence speed
  - Reduce the computational complexity

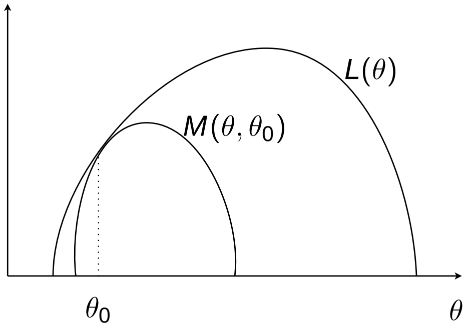
[1] Roland Badeau. "Gaussian modelling of mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October 2011.

EM as Minorize-Maximize (MM) method



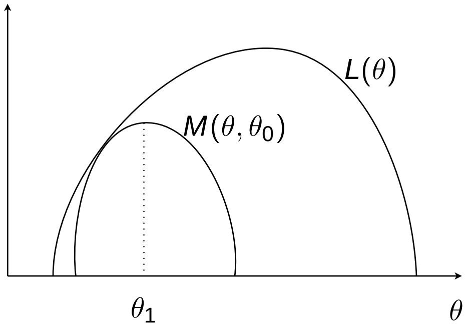
$$L(\theta) = \ln(p(x;\theta))$$

EM as Minorize-Maximize (MM) method



$$Q(\theta, \theta_0) = \int \ln(p(x, c; \theta)) p(c|x; \theta_0) dc$$
$$M(\theta, \theta_0) = L(\theta_0) + Q(\theta, \theta_0) - Q(\theta_0, \theta_0)$$

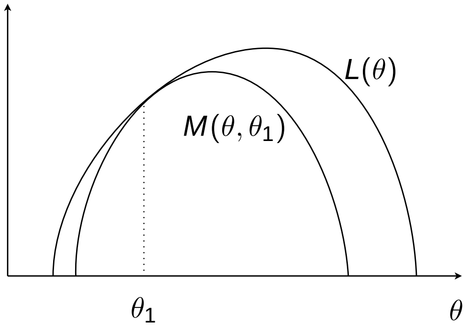
EM as Minorize-Maximize (MM) method



$$\theta_1 = \underset{\theta}{\operatorname{argmax}}\, M(\theta, \theta_0)$$

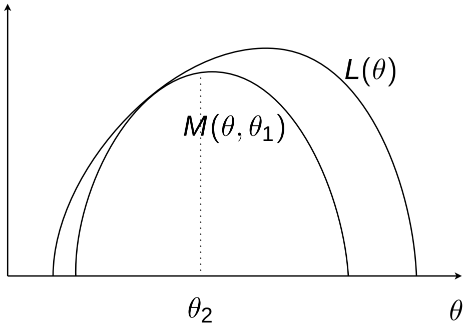
TELECOM ParisTech

EM as Minorize-Maximize (MM) method



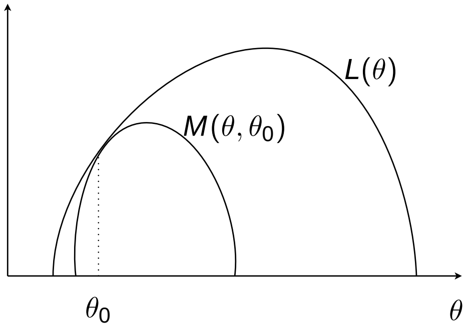
$$Q(\theta, \theta_1) = \int \ln(p(x, c; \theta)) p(c|x; \theta_1) dc$$
$$M(\theta, \theta_1) = L(\theta_1) + Q(\theta, \theta_1) - Q(\theta_1, \theta_1)$$

EM as Minorize-Maximize (MM) method



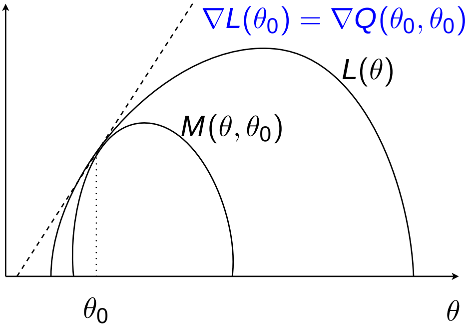
$$\theta_2 = \underset{\theta}{\operatorname{argmax}}\, M(\theta, \theta_1)$$

Computing the gradient of L



$$Q(\theta, \theta_0) = \int \ln(p(x, c; \theta)) p(c|x; \theta_0) dc$$
$$M(\theta, \theta_0) = L(\theta_0) + Q(\theta, \theta_0) - Q(\theta_0, \theta_0)$$

Computing the gradient of L



$$Q(\theta, \theta_0) = \int \ln(p(x, c; \theta)) p(c|x; \theta_0) dc$$
$$M(\theta, \theta_0) = L(\theta_0) + Q(\theta, \theta_0) - Q(\theta_0, \theta_0)$$

Multiplicative update rules

- Purpose: improve the convergence rate of EM
- Observation: the E-step permits to efficiently compute the gradient of the log-likelihood function
- Principle: replace the M-step by any gradient-based optimizer
- New update rules parametrized by $\varepsilon \geq 0$, which generalize both IS-NMF multiplicative updates ($\varepsilon = 0$) and EM ($\varepsilon = 1$)
- Enhanced convergence speed obtained with a "simulated cooling" strategy (make ε decrease over iterations)

[1] Roland Badeau, Alexey Ozerov. "Multiplicative update rules for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". Submitted to *EUSIPCO*, Marrakech, Morocco, September 2013.

Multiplicative update rules

- Purpose: improve the convergence rate of EM
- Observation: the E-step permits to efficiently compute the gradient of the log-likelihood function
- Principle: replace the M-step by any gradient-based optimizer
- New update rules parametrized by $\varepsilon \geq 0$, which generalize both IS-NMF multiplicative updates ($\varepsilon = 0$) and EM ($\varepsilon = 1$)
- Enhanced convergence speed obtained with a "simulated cooling" strategy (make ε decrease over iterations)

[1] Roland Badeau, Alexey Ozerov. "Multiplicative update rules for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". Submitted to *EUSIPCO*, Marrakech, Morocco, September 2013.

Multiplicative update rules

- Purpose: improve the convergence rate of EM
- Observation: the E-step permits to efficiently compute the gradient of the log-likelihood function
- Principle: replace the M-step by any gradient-based optimizer
- New update rules parametrized by $\varepsilon \geq 0$, which generalize both IS-NMF multiplicative updates ($\varepsilon = 0$) and EM ($\varepsilon = 1$)
- Enhanced convergence speed obtained with a "simulated cooling" strategy (make ε decrease over iterations)

[1] Roland Badeau, Alexey Ozerov. "Multiplicative update rules for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". Submitted to *EUSIPCO*, Marrakech, Morocco, September 2013.

Multiplicative update rules

- Purpose: improve the convergence rate of EM
- Observation: the E-step permits to efficiently compute the gradient of the log-likelihood function
- Principle: replace the M-step by any gradient-based optimizer
- New update rules parametrized by $\varepsilon \geq 0$, which generalize both IS-NMF multiplicative updates ($\varepsilon = 0$) and EM ($\varepsilon = 1$)
- Enhanced convergence speed obtained with a "simulated cooling" strategy (make ε decrease over iterations)

[1] Roland Badeau, Alexey Ozerov. "Multiplicative update rules for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". Submitted to *EUSIPCO*, Marrakech, Morocco, September 2013.

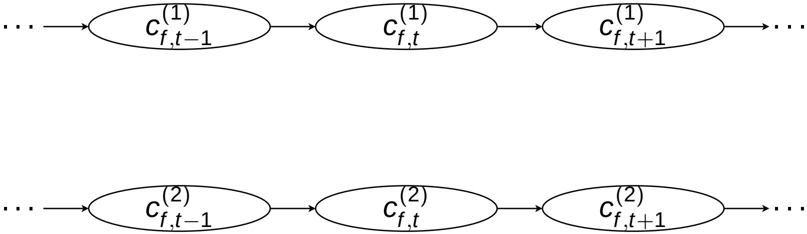
Multiplicative update rules

- Purpose: improve the convergence rate of EM
- Observation: the E-step permits to efficiently compute the gradient of the log-likelihood function
- Principle: replace the M-step by any gradient-based optimizer
- New update rules parametrized by $\varepsilon \geq 0$, which generalize both IS-NMF multiplicative updates ($\varepsilon = 0$) and EM ($\varepsilon = 1$)
- Enhanced convergence speed obtained with a "simulated cooling" strategy (make ε decrease over iterations)

[1] Roland Badeau, Alexey Ozerov. "Multiplicative update rules for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". Submitted to *EUSIPCO*, Marrakech, Morocco, September 2013.

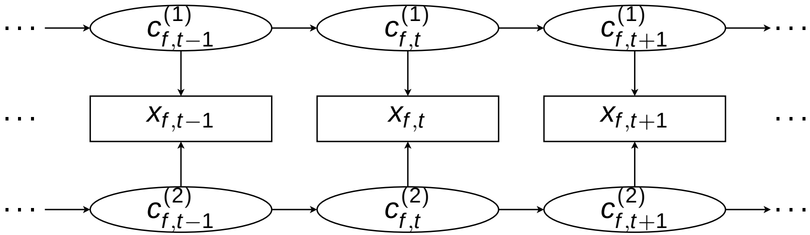
Variational Bayesian EM algorithm

- Prior distribution of latent variables in band f $(P = 1, K = 2)$



Variational Bayesian EM algorithm

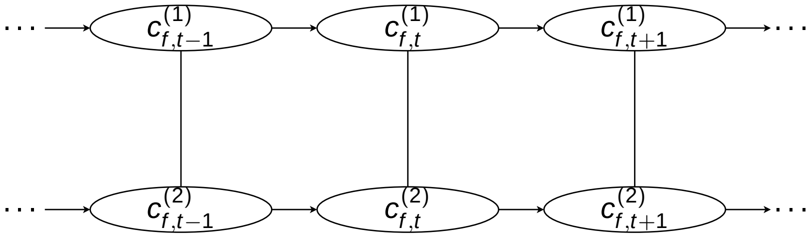
- Joint distribution of complete data in band f $(P = 1, K = 2)$



TELECOM ParisTech

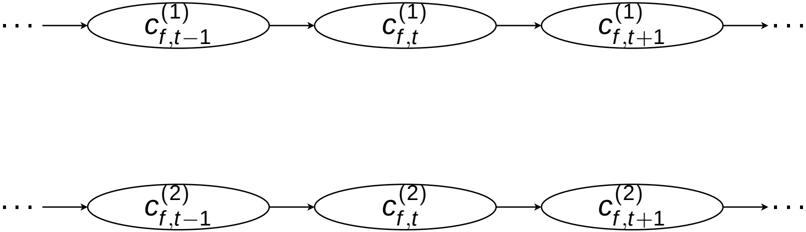
Variational Bayesian EM algorithm

- Posterior distribution of latent variables in band f ($P = 1, K = 2$)



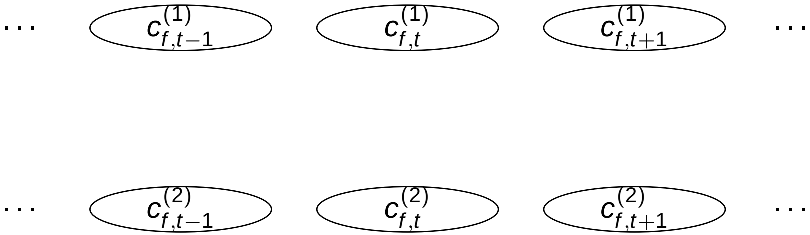
Variational Bayesian EM algorithm

- Structured mean field approximation in band f $(P = 1, K = 2)$



Variational Bayesian EM algorithm

- Mean field approximation in band f ($P = 1$, $K = 2$)



Variational Bayesian EM algorithm

- Purpose: reduce the computational complexity of EM
- Principle: the posterior distribution of the latent variables is approximated by a factorized distribution
- Complexity reduction:
  - Exact E-step: $O(FTK^3(1+P)^3)$
  - Structured mean field (no dependency over k) : $O(FTK(1+P)^3)$
  - Mean field (no dependency over k and t) : $O(FTK(1+P))$
- Performance loss:
  - The increase of log-likelihood function is no longer guaranteed
  - In practice, no perceptual difference

[1] Roland Badeau, Angélique Drémeau. "Variational Bayesian EM algorithm for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". To appear in *IEEE ICASSP*, Vancouver, Canada, May 2013.

Variational Bayesian EM algorithm

- Purpose: reduce the computational complexity of EM
- Principle: the posterior distribution of the latent variables is approximated by a factorized distribution
- Complexity reduction:
  - Exact E-step: $O(FTK^3(1+P)^3)$
  - Structured mean field (no dependency over k) : $O(FTK(1+P)^3)$
  - Mean field (no dependency over k and t) : $O(FTK(1+P))$
- Performance loss:
  - The increase of log-likelihood function is no longer guaranteed
  - In practice, no perceptual difference

[1] Roland Badeau, Angélique Drémeau. "Variational Bayesian EM algorithm for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". To appear in *IEEE ICASSP*, Vancouver, Canada, May 2013.

Variational Bayesian EM algorithm

- Purpose: reduce the computational complexity of EM
- Principle: the posterior distribution of the latent variables is approximated by a factorized distribution
- Complexity reduction:
  - Exact E-step: $O(FTK^3(1+P)^3)$
  - Structured mean field (no dependency over k) : $O(FTK(1+P)^3)$
  - Mean field (no dependency over k and t) : $O(FTK(1+P))$
- Performance loss:
  - The increase of log-likelihood function is no longer guaranteed
  - In practice, no perceptual difference

[1] Roland Badeau, Angélique Drémeau. "Variational Bayesian EM algorithm for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". To appear in *IEEE ICASSP*, Vancouver, Canada, May 2013.

Variational Bayesian EM algorithm

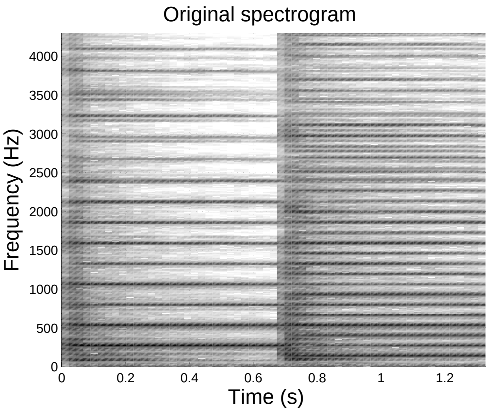
- Purpose: reduce the computational complexity of EM
- Principle: the posterior distribution of the latent variables is approximated by a factorized distribution
- Complexity reduction:
  - Exact E-step: $O(FTK^3(1+P)^3)$
  - Structured mean field (no dependency over k) : $O(FTK(1+P)^3)$
  - Mean field (no dependency over k and t) : $O(FTK(1+P))$
- Performance loss:
  - The increase of log-likelihood function is no longer guaranteed
  - In practice, no perceptual difference

[1] Roland Badeau, Angélique Drémeau. "Variational Bayesian EM algorithm for modelling mixtures of non-stationary signals in the time-frequency domain (HR-NMF)". To appear in *IEEE ICASSP*, Vancouver, Canada, May 2013.

Outline

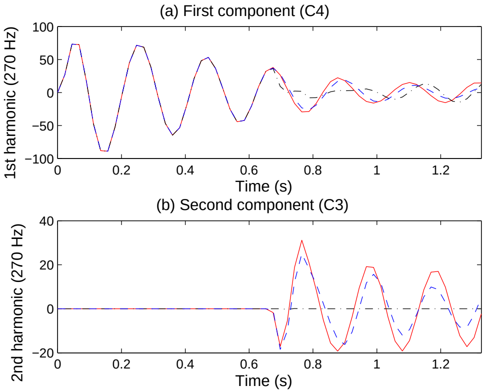
- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- **Preliminary results**
  - **Audio source separation**
  - **Audio inpainting**
- Conclusions

Application to piano tones



Spectrogram of the input piano sound (C4 + C3)

Source separation



IS-NMF:

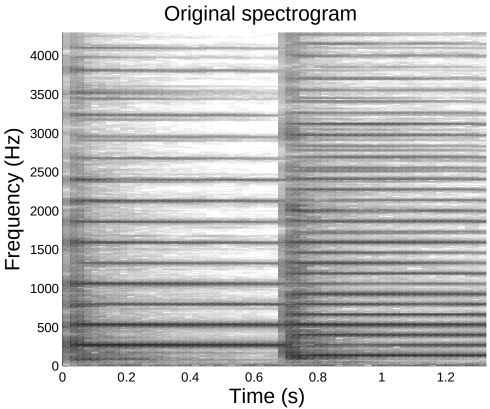
HR-NMF:

IS-NMF:

HR-NMF:

Separation of two sinusoidal components

Audio inpainting



C4+C3:

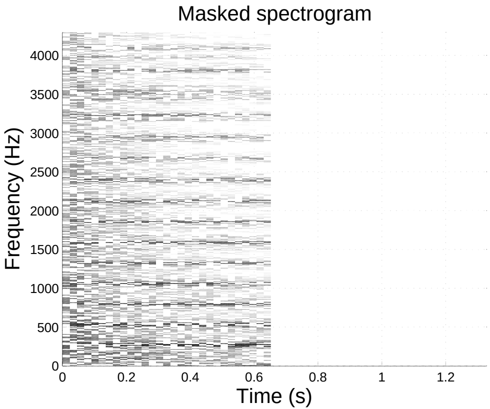
C4 alone:

IS-NMF:

HR-NMF:

Spectrogram of the input piano sound (C4 + C3)

Audio inpainting



C4+C3:

C4 alone:

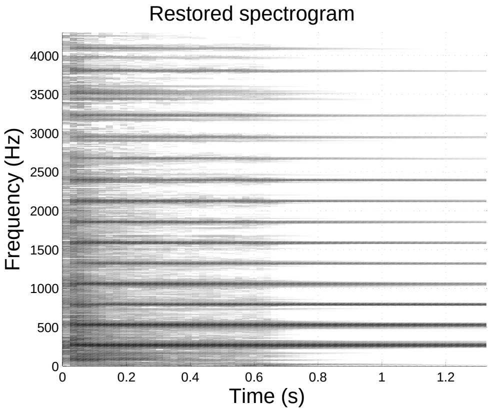
IS-NMF:

HR-NMF:

Masked spectrogram of the input piano sound

TELECOM ParisTech

Audio inpainting



C4+C3:

C4 alone:

IS-NMF:

HR-NMF:

Recovery of the full C4 piano tone

Outline

- Advantages and drawbacks of NMF probabilistic models
- Choosing an appropriate TF representation
- Modelling phases and correlations in the TF domain
  - HR-NMF model
  - Algorithms
- Preliminary results
  - Audio source separation
  - Audio inpainting
- **Conclusions**

Contributions

- Critically sampled paraunitary filter banks satisfy both PW and PR
- HR-NMF time-frequency model:
  - models phases and local correlations in each frequency band
  - generalizes IS-NMF, mixtures of AR processes, and ESM models
- Algorithms:
  - EM algorithm: too computationally demanding
  - Multiplicative update rules: improved convergence speed
  - Variational Bayesian EM algorithm: lower computational complexity
- Preliminary results:
  - Separation of overlapping sinusoids without perceptible artefacts
  - Restoration of missing observations without perceptible artefacts

Contributions

- Critically sampled paraunitary filter banks satisfy both PW and PR
- HR-NMF time-frequency model:
  - models phases and local correlations in each frequency band
  - generalizes IS-NMF, mixtures of AR processes, and ESM models
- Algorithms:
  - EM algorithm: too computationally demanding
  - Multiplicative update rules: improved convergence speed
  - Variational Bayesian EM algorithm: lower computational complexity
- Preliminary results:
  - Separation of overlapping sinusoids without perceptible artefacts
  - Restoration of missing observations without perceptible artefacts

Contributions

- Critically sampled paraunitary filter banks satisfy both PW and PR
- HR-NMF time-frequency model:
  - models phases and local correlations in each frequency band
  - generalizes IS-NMF, mixtures of AR processes, and ESM models
- Algorithms:
  - EM algorithm: too computationally demanding
  - Multiplicative update rules: improved convergence speed
  - Variational Bayesian EM algorithm: lower computational complexity
- Preliminary results:
  - Separation of overlapping sinusoids without perceptible artefacts
  - Restoration of missing observations without perceptible artefacts

Contributions

- Critically sampled paraunitary filter banks satisfy both PW and PR
- HR-NMF time-frequency model:
  - models phases and local correlations in each frequency band
  - generalizes IS-NMF, mixtures of AR processes, and ESM models
- Algorithms:
  - EM algorithm: too computationally demanding
  - Multiplicative update rules: improved convergence speed
  - Variational Bayesian EM algorithm: lower computational complexity
- Preliminary results:
  - Separation of overlapping sinusoids without perceptible artefacts
  - Restoration of missing observations without perceptible artefacts

Outlooks

- Design consistent TF representation and TF probabilistic model
- Extensions of HR-NMF:
  - Extension to multichannel signals (*e.g.* stereo)
  - Correlations between frequency bands ($\rightarrow$ attacks, vibratos, chirps)
  - Correlations between components ($\rightarrow$ sympathetic modes)
  - Replace NMF by other parametric models, or priors enforcing harmonicity, sparsity, smoothness...
- Algorithms:
  - Variational Bayesian methods
  - Markov Chain Monte Carlo (MCMC)
  - Belief propagation methods (message passing algorithm)
- Applications:
  - Source coding, source separation, audio inpainting...

Outlooks

- Design consistent TF representation and TF probabilistic model
- Extensions of HR-NMF:
  - Extension to multichannel signals (*e.g.* stereo)
  - Correlations between frequency bands ($\rightarrow$ attacks, vibratos, chirps)
  - Correlations between components ($\rightarrow$ sympathetic modes)
  - Replace NMF by other parametric models, or priors enforcing harmonicity, sparsity, smoothness...
- Algorithms:
  - Variational Bayesian methods
  - Markov Chain Monte Carlo (MCMC)
  - Belief propagation methods (message passing algorithm)
- Applications:
  - Source coding, source separation, audio inpainting...

Outlooks

- Design consistent TF representation and TF probabilistic model
- Extensions of HR-NMF:
  - Extension to multichannel signals (*e.g.* stereo)
  - Correlations between frequency bands ($\rightarrow$ attacks, vibratos, chirps)
  - Correlations between components ($\rightarrow$ sympathetic modes)
  - Replace NMF by other parametric models, or priors enforcing harmonicity, sparsity, smoothness...
- Algorithms:
  - Variational Bayesian methods
  - Markov Chain Monte Carlo (MCMC)
  - Belief propagation methods (message passing algorithm)
- Applications:
  - Source coding, source separation, audio inpainting...

Outlooks

- Design consistent TF representation and TF probabilistic model
- Extensions of HR-NMF:
  - Extension to multichannel signals (*e.g.* stereo)
  - Correlations between frequency bands ($\rightarrow$ attacks, vibratos, chirps)
  - Correlations between components ($\rightarrow$ sympathetic modes)
  - Replace NMF by other parametric models, or priors enforcing harmonicity, sparsity, smoothness...
- Algorithms:
  - Variational Bayesian methods
  - Markov Chain Monte Carlo (MCMC)
  - Belief propagation methods (message passing algorithm)
- Applications:
  - Source coding, source separation, audio inpainting...