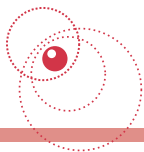




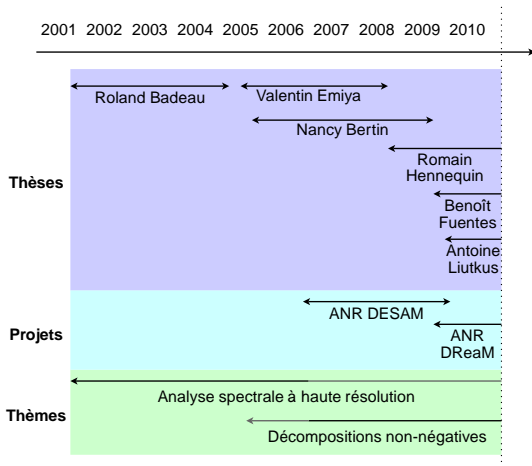
# Analyse spectrale à haute résolution et décompositions non-négatives appliquées au traitement des signaux de musique



Roland Badeau  
Soutenance d'Habilitation  
à Diriger des Recherches  
Télécom ParisTech, mardi 30 novembre 2010



# Chronologie





## Avant-propos

- CV détaillé (activités liées à la recherche, enseignements, etc.)
- Bibliographie détaillée de l'état de l'art du domaine
- Évaluation quantitative et comparative des algorithmes
- Sélection de thèmes extraits du mémoire
- Présentation synthétique de chaque thème, soulignant les idées essentielles et les contributions, et illustrée par des applications



# Avant-propos

- ~~CV détaillé (activités liées à la recherche, enseignements, etc.)~~
- ~~Bibliographie détaillée de l'état de l'art du domaine~~
- ~~Évaluation quantitative et comparative des algorithmes~~
- Sélection de thèmes extraits du mémoire
- Présentation synthétique de chaque thème, soulignant les idées essentielles et les contributions, et illustrée par des applications

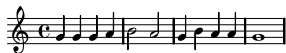


## Avant-propos

- ~~CV détaillé (activités liées à la recherche, enseignements, etc.)~~
- ~~Bibliographie détaillée de l'état de l'art du domaine~~
- ~~Évaluation quantitative et comparative des algorithmes~~
- Sélection de thèmes extraits du mémoire
- Présentation synthétique de chaque thème, soulignant les idées essentielles et les contributions, et illustrée par des applications



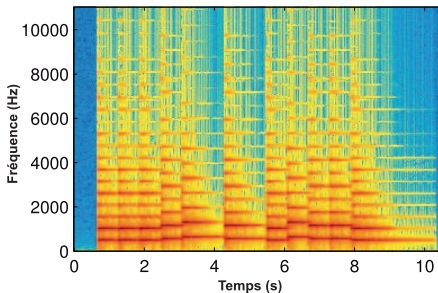
# Représentations de la musique



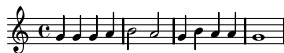
Partition musicale



# Représentations de la musique



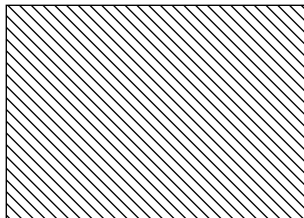
Spectrogramme de "Au clair de la lune"



Partition musicale



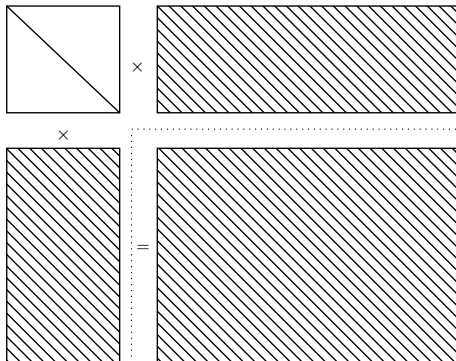
# Approximation par réduction de rang





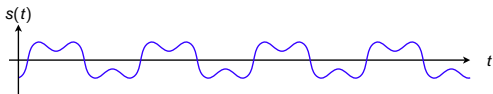


# Approximation par réduction de rang



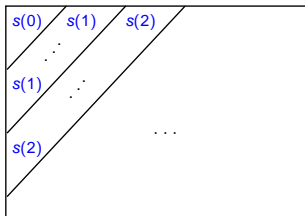
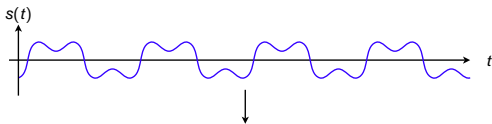


# Analyse spectrale à haute résolution (HR)



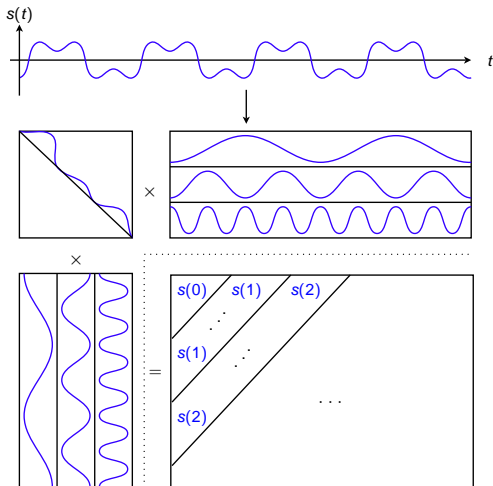


# Analyse spectrale à haute résolution (HR)



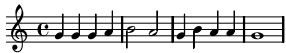


# Analyse spectrale à haute résolution (HR)

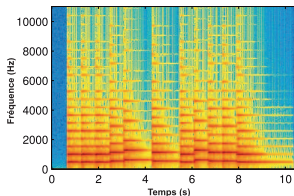




# Non-negative Matrix Factorization (NMF)



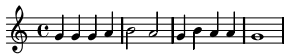
Partition musicale



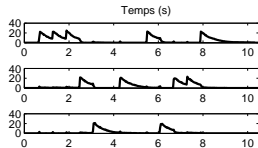
Spectrogramme



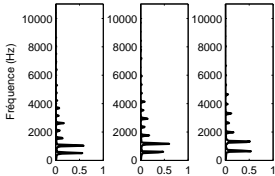
# Non-negative Matrix Factorization (NMF)



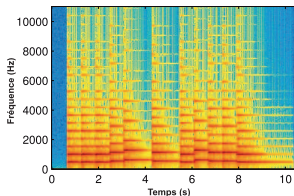
Partition musicale



Activations temporelles



Formes spectrales



Spectrogramme

## ■ Applications des méthodes HR

- Analyse spectrale (analyse modale, spectroscopie)
- Traitement d'antenne (formation de voies, localisation de sources)
- Communications numériques (identification de canaux)

## ■ Applications de la NMF

- Analyse d'images (reconnaissance de visages)
- Fouille de texte, spectroscopie, finance, etc.

## ■ Applications en traitement des signaux audio

- Analyse / synthèse, représentation, codage audio
- Estimation de hauteurs et de tempo, transcription automatique

## ■ Applications des méthodes HR

- Analyse spectrale (analyse modale, spectroscopie)
- Traitement d'antenne (formation de voies, localisation de sources)
- Communications numériques (identification de canaux)

## ■ Applications de la NMF

- Analyse d'images (reconnaissance de visages)
- Fouille de texte, spectroscopie, finance, etc.

## ■ Applications en traitement des signaux audio

- Analyse / synthèse, représentation, codage audio
- Estimation de hauteurs et de tempo, transcription automatique



- Applications des méthodes HR
  - Analyse spectrale (analyse modale, spectroscopie)
  - Traitement d'antenne (formation de voies, localisation de sources)
  - Communications numériques (identification de canaux)
- Applications de la NMF
  - Analyse d'images (reconnaissance de visages)
  - Fouille de texte, spectroscopie, finance, etc.
- Applications en traitement des signaux audio
  - Analyse / synthèse, représentation, codage audio
  - Estimation de hauteurs et de tempo, transcription automatique



## Partie I

# Analyse spectrale à haute résolution



## Exponential Sinusoidal Model (ESM)

- **Modèle réel** :  $s(t) = \sum_{k=1}^r a_k e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$ 
  - $a_k \in \mathbb{R}_+^*$  et  $\phi_k \in ]-\pi, \pi]$  sont l'**amplitude** et la **phase**
  - $\delta_k \in \mathbb{R}$  et  $\nu_k \in ]-\frac{1}{2}, \frac{1}{2}]$  sont le **facteur d'atténuation** et la **fréquence**
- **Modèle complexe** :  $s(t) = \sum_{k=1}^r \alpha_k z_k^t$ 
  - $\alpha_k = a_k e^{j\phi_k} \in \mathbb{C}^*$  est l'**amplitude complexe**
  - $z_k = e^{-\delta_k + j2\pi\nu_k} \in \mathbb{C}^*$  est le **pôle complexe**
- **Modèle bruité** :  $x(t) = s(t) + b(t)$  avec  $b(t)$  bruit blanc gaussien
- **Estimation du modèle**
  - Vecteur de données :  $\mathbf{s}(t) = [s(t), \dots, s(t+n-1)]^T$  avec  $n > r$
  - Analyse de Fourier : résolution spectrale de l'ordre de  $\frac{1}{n}$
  - Analyse sous-espace : **haute résolution** spectrale



## Exponential Sinusoidal Model (ESM)

- **Modèle réel** :  $s(t) = \sum_{k=1}^r a_k e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$ 
  - $a_k \in \mathbb{R}_+^*$  et  $\phi_k \in ]-\pi, \pi]$  sont l'**amplitude** et la **phase**
  - $\delta_k \in \mathbb{R}$  et  $\nu_k \in ]-\frac{1}{2}, \frac{1}{2}]$  sont le **facteur d'atténuation** et la **fréquence**
- **Modèle complexe** :  $s(t) = \sum_{k=1}^r \alpha_k z_k^t$ 
  - $\alpha_k = a_k e^{i\phi_k} \in \mathbb{C}^*$  est l'**amplitude complexe**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$  est le **pôle complexe**
- **Modèle bruité** :  $x(t) = s(t) + b(t)$  avec  $b(t)$  bruit blanc gaussien
- **Estimation du modèle**
  - Vecteur de données :  $\mathbf{s}(t) = [s(t), \dots, s(t+n-1)]^T$  avec  $n > r$
  - Analyse de Fourier : résolution spectrale de l'ordre de  $\frac{1}{n}$
  - Analyse sous-espace : **haute résolution** spectrale



# Exponential Sinusoidal Model (ESM)

- **Modèle réel** :  $s(t) = \sum_{k=1}^r a_k e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$ 
  - $a_k \in \mathbb{R}_+^*$  et  $\phi_k \in ]-\pi, \pi]$  sont l'**amplitude** et la **phase**
  - $\delta_k \in \mathbb{R}$  et  $\nu_k \in ]-\frac{1}{2}, \frac{1}{2}]$  sont le **facteur d'atténuation** et la **fréquence**
- **Modèle complexe** :  $s(t) = \sum_{k=1}^r \alpha_k z_k^t$ 
  - $\alpha_k = a_k e^{i\phi_k} \in \mathbb{C}^*$  est l'**amplitude complexe**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$  est le **pôle complexe**
- **Modèle bruité** :  $x(t) = s(t) + b(t)$  avec  $b(t)$  bruit blanc gaussien
- Estimation du modèle
  - Vecteur de données :  $\mathbf{s}(t) = [s(t), \dots, s(t+n-1)]^T$  avec  $n > r$
  - Analyse de Fourier : résolution spectrale de l'ordre de  $\frac{1}{n}$
  - Analyse sous-espace : **haute résolution** spectrale

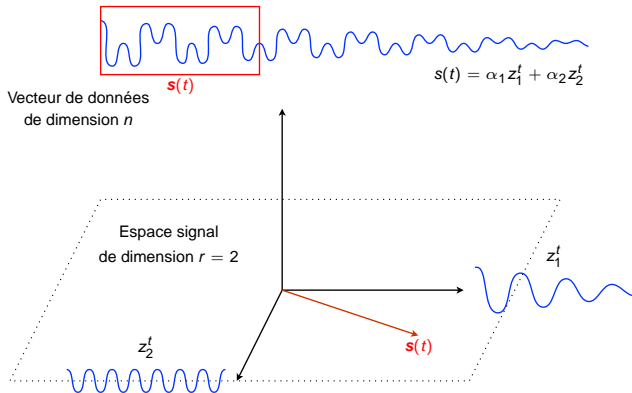


## Exponential Sinusoidal Model (ESM)

- **Modèle réel** :  $s(t) = \sum_{k=1}^r a_k e^{-\delta_k t} \cos(2\pi\nu_k t + \phi_k)$ 
  - $a_k \in \mathbb{R}_+^*$  et  $\phi_k \in ]-\pi, \pi]$  sont l'**amplitude** et la **phase**
  - $\delta_k \in \mathbb{R}$  et  $\nu_k \in ]-\frac{1}{2}, \frac{1}{2}]$  sont le **facteur d'atténuation** et la **fréquence**
- **Modèle complexe** :  $s(t) = \sum_{k=1}^r \alpha_k z_k^t$ 
  - $\alpha_k = a_k e^{i\phi_k} \in \mathbb{C}^*$  est l'**amplitude complexe**
  - $z_k = e^{-\delta_k + i2\pi\nu_k} \in \mathbb{C}^*$  est le **pôle complexe**
- **Modèle bruité** :  $x(t) = s(t) + b(t)$  avec  $b(t)$  bruit blanc gaussien
- **Estimation du modèle**
  - Vecteur de données :  $\mathbf{s}(t) = [s(t), \dots, s(t+n-1)]^T$  avec  $n > r$
  - Analyse de Fourier : résolution spectrale de l'ordre de  $\frac{1}{n}$
  - Analyse sous-espace : **haute résolution** spectrale

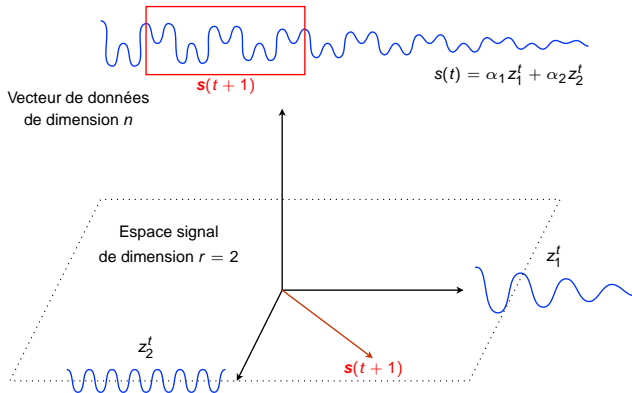


# Analyse sous-espace





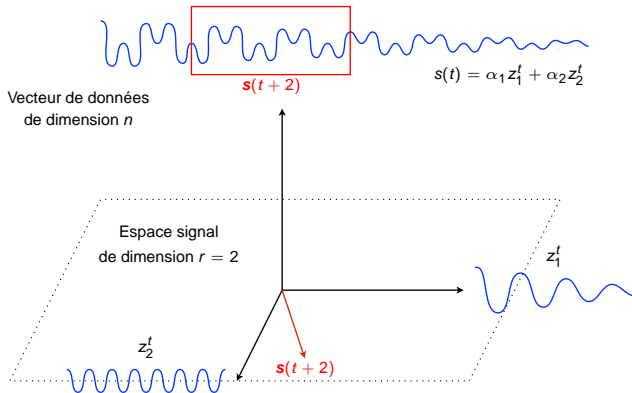
# Analyse sous-espace







# Analyse sous-espace





## Estimation du modèle

- Choix d'une fenêtre  $(\gamma_\tau)_{\tau \in \mathbb{N}}$  (exponentielle, rectangulaire, hybride)
- Calcul de la matrice de "corrélation"

$$\mathbf{C}_{xx}(t) = \sum_{\tau=0}^t \gamma_\tau \mathbf{x}(t - \tau) \mathbf{x}(t - \tau)^H$$

- Estimation des paramètres du modèle

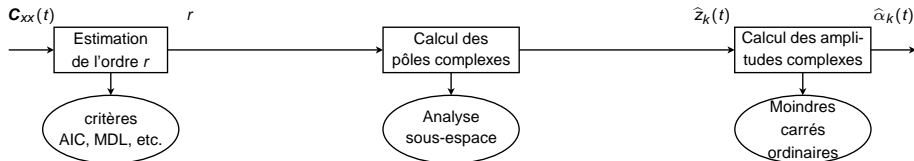


## Estimation du modèle

- Choix d'une fenêtre  $(\gamma_\tau)_{\tau \in \mathbb{N}}$  (exponentielle, rectangulaire, hybride)
- Calcul de la matrice de "corrélation"

$$\mathbf{C}_{xx}(t) = \sum_{\tau=0}^t \gamma_\tau \mathbf{x}(t - \tau) \mathbf{x}(t - \tau)^H$$

- Estimation des paramètres du modèle



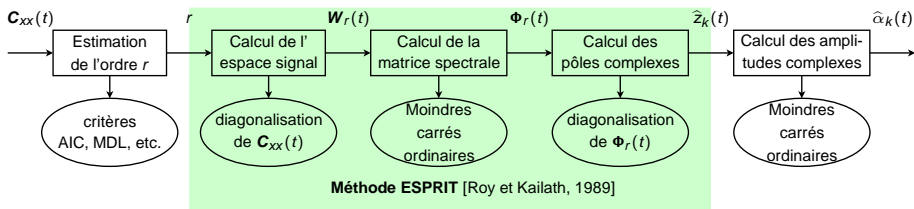


# Estimation du modèle

- Choix d'une fenêtre  $(\gamma_\tau)_{\tau \in \mathbb{N}}$  (exponentielle, rectangulaire, hybride)
- Calcul de la matrice de "corrélation"

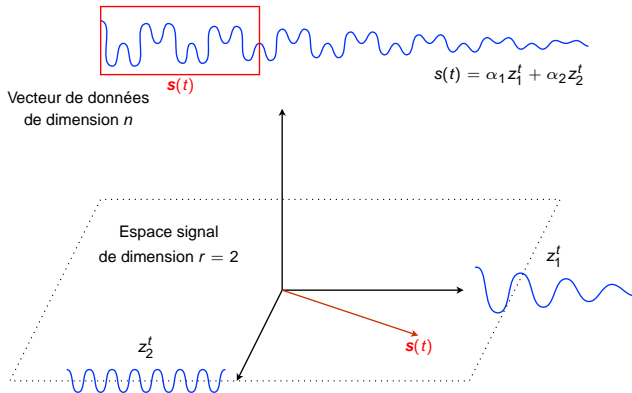
$$\mathbf{C}_{xx}(t) = \sum_{\tau=0}^t \gamma_\tau \mathbf{x}(t - \tau) \mathbf{x}(t - \tau)^H$$

- Estimation des paramètres du modèle



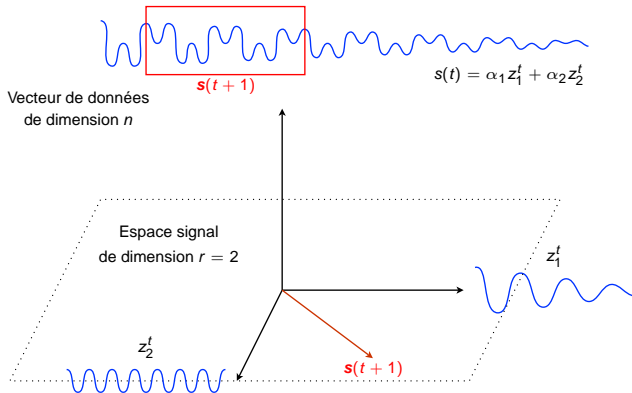


# Analyse sous-espace



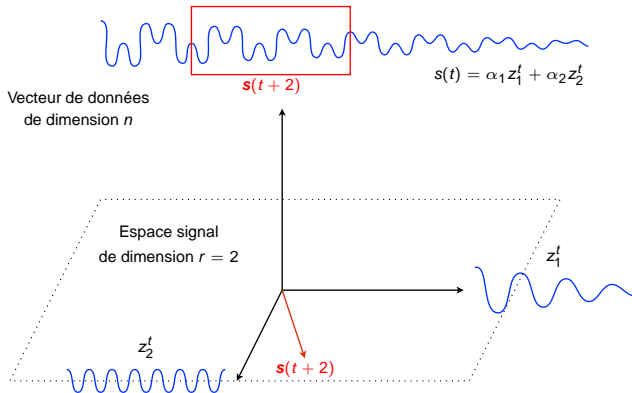


# Analyse sous-espace



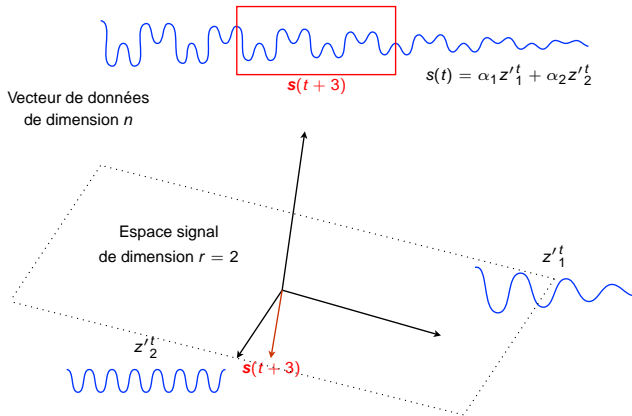


# Analyse sous-espace





# Analyse sous-espace





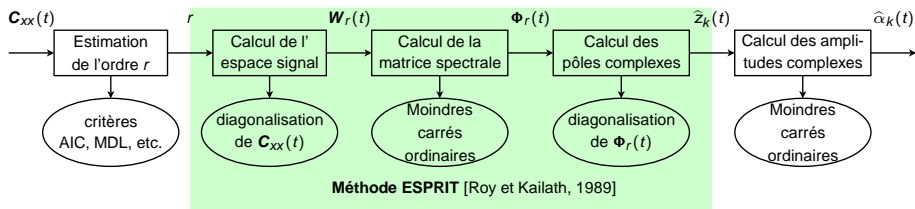


# Estimation du modèle

- Choix d'une fenêtre  $(\gamma_\tau)_{\tau \in \mathbb{N}}$  (exponentielle, rectangulaire, hybride)
- Calcul de la matrice de "corrélation"

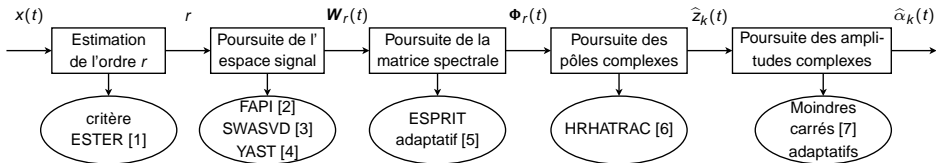
$$\mathbf{C}_{xx}(t) = \sum_{\tau=0}^t \gamma_\tau \mathbf{x}(t - \tau) \mathbf{x}(t - \tau)^H$$

- Estimation des paramètres du modèle





# Analyse temps-fréquence



[1] Roland Badeau, Bertrand David et Gaël Richard. "A new perturbation analysis for signal enumeration in rotational invariance techniques". *IEEE Transactions on Signal Processing*, 54(2) :450–458, février 2006.

[2] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.

[3] Roland Badeau, Gaël Richard et Bertrand David. "Sliding window adaptive SVD algorithms". *IEEE Transactions on Signal Processing*, 52(1) :1-10, janvier 2004.

[4] Roland Badeau, Gaël Richard et Bertrand David. "Fast and stable YAST algorithm for principal and minor subspace tracking". *IEEE Transactions on Signal Processing*, 56(8) :3437-3446, août 2008.

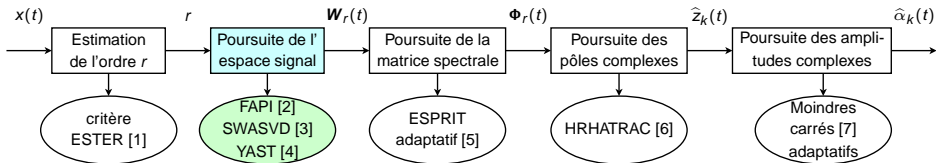
[5] Roland Badeau, Gaël Richard et Bertrand David. "Fast adaptive ESPRIT algorithm". Dans *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, juillet 2005.

[6] Bertrand David, Roland Badeau et Gaël Richard. "HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals". Dans *Proc. of IEEE ICASSP*, volume 3, pages 45-48, Toulouse, France, mai 2006.

[7] Bertrand David et Roland Badeau. "Fast sequential LS estimation for sinusoidal modeling and decomposition of audio signals". Dans *Proc. of IEEE WASPAA*, pages 211-214, New Paltz, New York, USA, octobre 2007.



# Analyse temps-fréquence



[1] Roland Badeau, Bertrand David et Gaël Richard. "A new perturbation analysis for signal enumeration in rotational invariance techniques". *IEEE Transactions on Signal Processing*, 54(2) :450–458, février 2006.

[2] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.

[3] Roland Badeau, Gaël Richard et Bertrand David. "Sliding window adaptive SVD algorithms". *IEEE Transactions on Signal Processing*, 52(1) :1-10, janvier 2004.

[4] Roland Badeau, Gaël Richard et Bertrand David. "Fast and stable YAST algorithm for principal and minor subspace tracking". *IEEE Transactions on Signal Processing*, 56(8) :3437-3446, août 2008.

[5] Roland Badeau, Gaël Richard et Bertrand David. "Fast adaptive ESPRIT algorithm". Dans *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, juillet 2005.

[6] Bertrand David, Roland Badeau et Gaël Richard. "HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals". Dans *Proc. of IEEE ICASSP*, volume 3, pages 45-48, Toulouse, France, mai 2006.

[7] Bertrand David et Roland Badeau. "Fast sequential LS estimation for sinusoidal modeling and decomposition of audio signals". Dans *Proc. of IEEE WASPAA*, pages 211-214, New Paltz, New York, USA, octobre 2007.



# Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )



# Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )



## Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )



## Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )



## Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )





## Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )

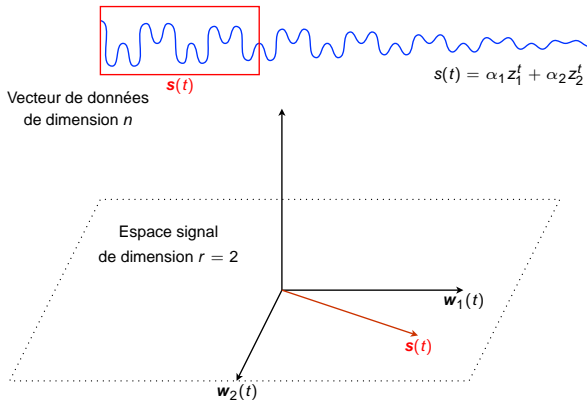


## Méthode des puissances itérées

- Méthode des puissances itérées (calcul récursif de  $\mathbf{W}_r(t)$ )
  - 1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx} \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )
  - 2)  $\mathbf{W}_r(t) \mathbf{R}(t) = \mathbf{C}_{xy}(t)$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )
    - $\text{Span}(\mathbf{W}_r(t))$  converge exponentiellement vers l'espace signal
    - Si 2) est une factorisation orthogonale-triangulaire,  $\mathbf{W}_r(t)$  converge vers les  $r$  principaux vecteurs propres de  $\mathbf{C}_{xx}$
- Poursuite de l'espace signal si  $\mathbf{C}_{xx}(t)$  varie au cours du temps
- Algorithme rapide [Strobach, 1996] (coût de  $nr^2$  au lieu de  $n^2r$ )

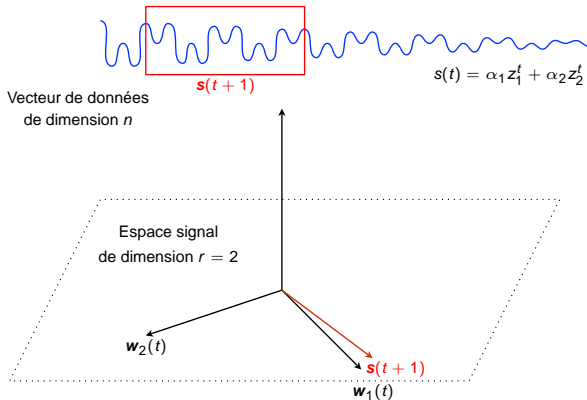


# Poursuite de sous-espace



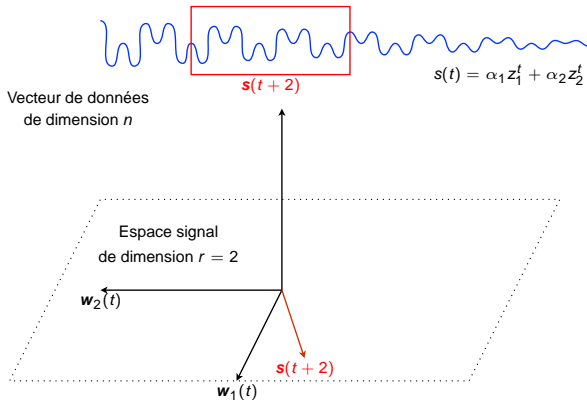


# Poursuite de sous-espace



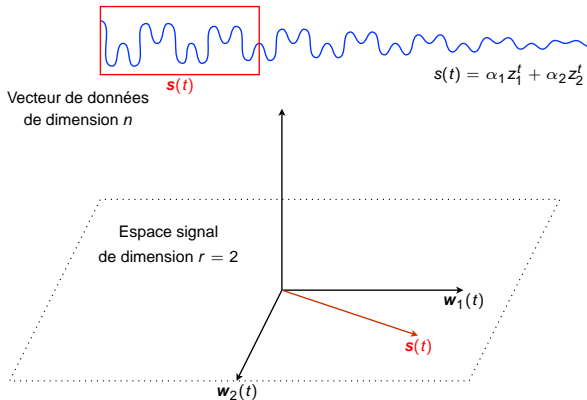


# Poursuite de sous-espace



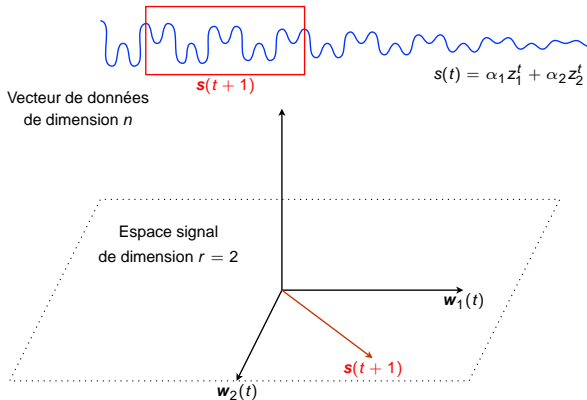


# Poursuite de sous-espace



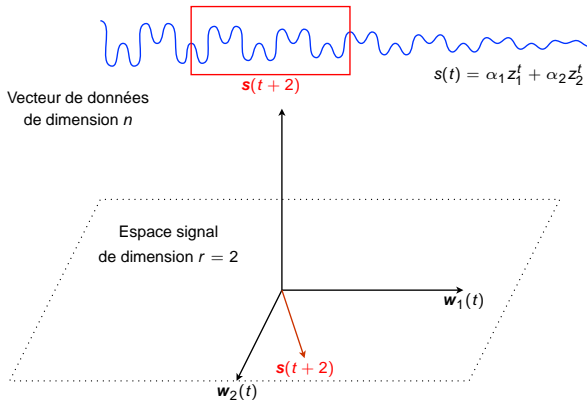


# Poursuite de sous-espace





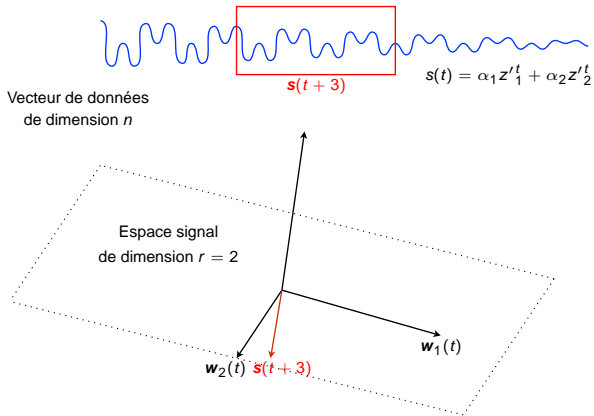
# Poursuite de sous-espace







# Poursuite de sous-espace





# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t) \mathbf{W}_r(t-1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t) \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t) \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t) \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t) \mathbf{W}_r(t - 1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.



# Méthode des puissances naturelles

## ■ Méthode des puissances naturelles

1)  $\mathbf{C}_{xy}(t) = \mathbf{C}_{xx}(t)\mathbf{W}_r(t-1)$  (compression de  $\mathbf{C}_{xx}$ )

2)  $\mathbf{W}_r(t) = \mathbf{C}_{xy}(t) (\mathbf{C}_{xy}(t)^H \mathbf{C}_{xy}(t))^{-\frac{1}{2}}$  (orthonormalisation de  $\mathbf{C}_{xy}(t)$ )

## ■ Algorithme FAPI [1] (coût de $3nr$ au lieu de $nr^2$ )

- atteint la borne inférieure de complexité ( $3nr$ )
- converge plus rapidement que PAST et ses variantes
- garantit l'orthonormalité de  $\mathbf{W}_r(t)$  et la stabilité numérique

[1] Roland Badeau, Bertrand David et Gaël Richard. "Fast Approximated Power Iteration Subspace Tracking". *IEEE Transactions on Signal Processing*, 53(8) :2931-2941, août 2005.





# Applications de l'analyse à haute résolution

## ■ Analyse / synthèse

- Représentation temps-fréquence à haute résolution
- Analyse des modes couplés d'une harpe
- Codage audio

## ■ Transcription automatique

- Estimation de hauteurs de notes de piano
- Estimation de tempo musical

## ■ Autres applications

- Estimation de canal en communications numériques
- SVD multilinéaire adaptative pour les tenseurs structurés



# Applications de l'analyse à haute résolution

## ■ Analyse / synthèse

- Représentation temps-fréquence à haute résolution
- Analyse des modes couplés d'une harpe
- Codage audio

## ■ Transcription automatique

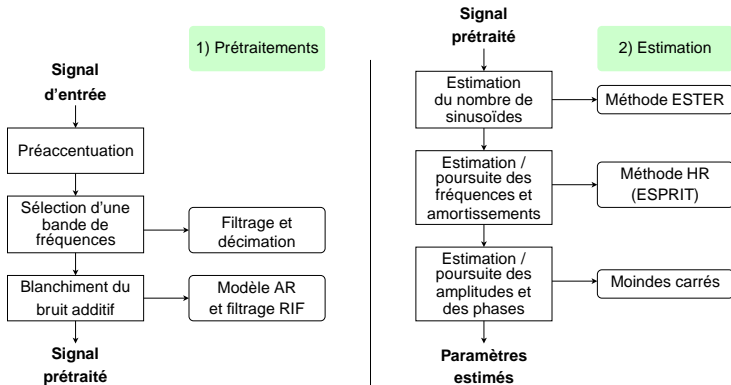
- Estimation de hauteurs de notes de piano
- Estimation de tempo musical

## ■ Autres applications

- Estimation de canal en communications numériques
- SVD multilinéaire adaptative pour les tenseurs structurés

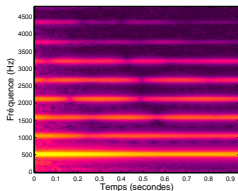


# Analyse / synthèse et représentation de signaux audio

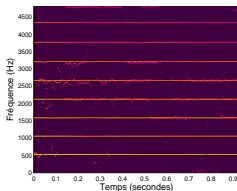




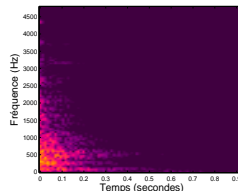
# Décomposition d'un son de piano



Spectrogramme



HR-ogramme



Résiduel



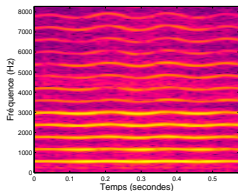
[1] Roland Badeau et Bertrand David. "Adaptive subspace methods for high resolution analysis of music signals". Dans Acoustics'08, Paris, France, juillet 2008.

[2] Bertrand David, Gaël Richard et Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". Dans Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pages 715-718, Stockholm, Suède, août 2003.

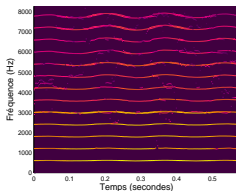
[3] Roland Badeau, Rémy Boyer et Bertrand David. "EDS parametric modeling and tracking of audio signals". Dans Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pages 139-144, Hambourg, Allemagne, septembre 2002.



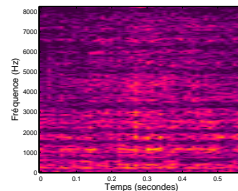
# Décomposition d'un son de violon



Spectrogramme



HR-ogramme



Résiduel



[1] Roland Badeau et Bertrand David. "Adaptive subspace methods for high resolution analysis of music signals". Dans Acoustics'08, Paris, France, juillet 2008.

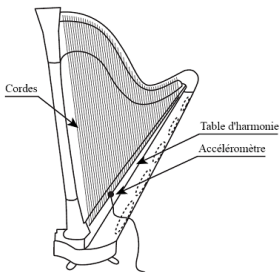
[2] Bertrand David, Gaël Richard et Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". Dans Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pages 715-718, Stockholm, Suède, août 2003.

[3] Roland Badeau, Rémy Boyer et Bertrand David. "EDS parametric modeling and tracking of audio signals". Dans Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pages 139-144, Hambourg, Allemagne, septembre 2002.

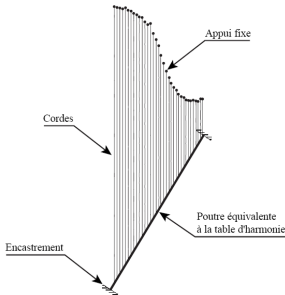


# Analyse des modes couplés d'une harpe

Modélisation de la vibration par sympathie des cordes d'une harpe [1]



Protocole expérimental



Modèle physique

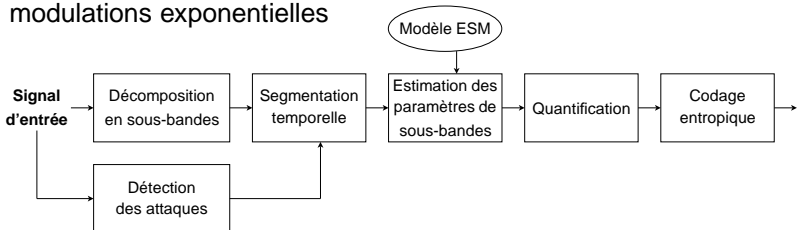
[1] Jean-Loïc Le Carrou, François Gautier et Roland Badeau. "Sympathetic string modes in the concert harp". *Acta Acustica united with Acustica*, 95(4) :744-752, juillet/août 2009.



# Codage audio

## ■ Codeur paramétrique basé sur le modèle ESM [1]

→ modulations exponentielles



## ■ Quantification scalaire conjointe avec contrainte entropique [2,3]

[1] Olivier Derrien, Gaël Richard et Roland Badeau. "Damped sinusoids and subspace based approach for lossy audio coding". Dans *Acoustics'08*, Paris, France, juillet 2008.

[2] Olivier Derrien, Roland Badeau et Gaël Richard. "Entropy-constrained quantization of exponentially damped sinusoids parameters". Dans *Proc. of IEEE ICASSP*, Prague, République Tchèque, mai 2011. Soumis pour publication.

[3] Olivier Derrien, Roland Badeau et Gaël Richard. "Calculation of an entropy-constrained quantizer for exponentially damped sinusoids parameters". Rapport Technique, Laboratoire de Mécanique et d'Acoustique, Marseille, France, juin 2010.



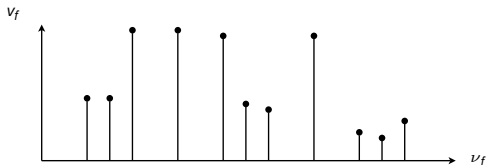
## Partie II

# Décompositions non-négatives



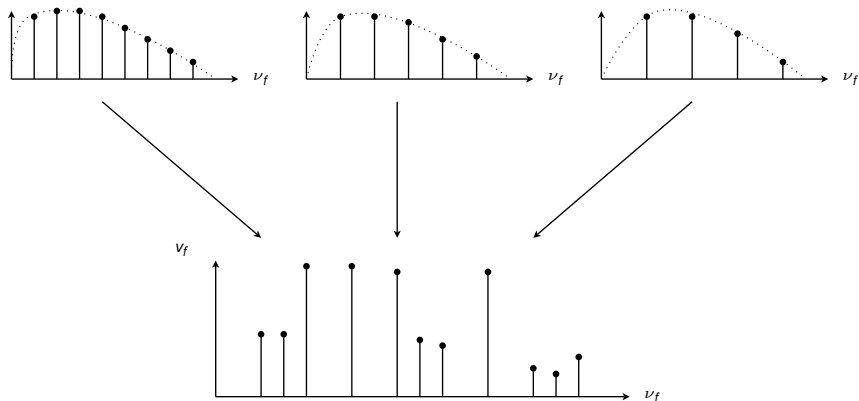


# Vers un modèle de plus haut niveau



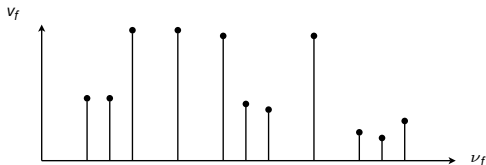


# Vers un modèle de plus haut niveau



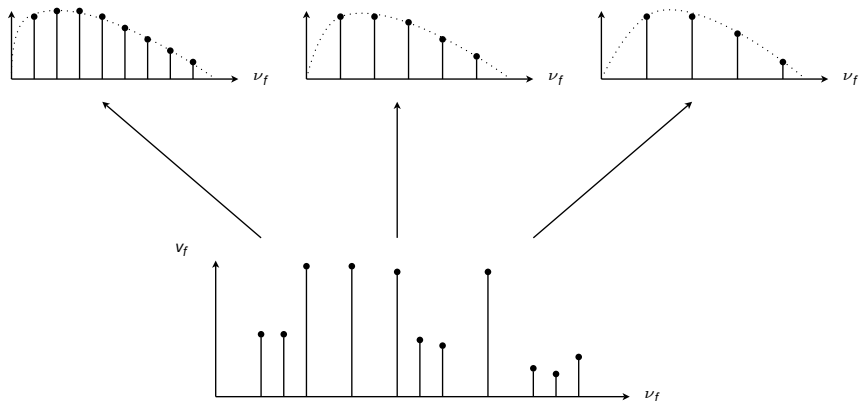


# Vers un modèle de plus haut niveau



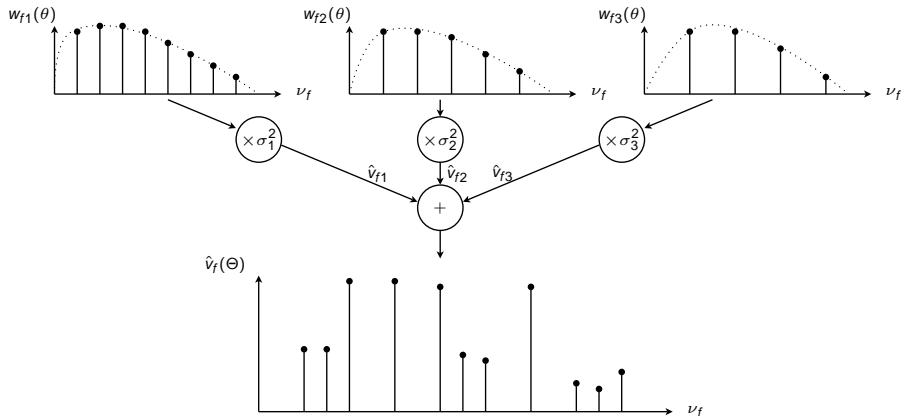


# Vers un modèle de plus haut niveau



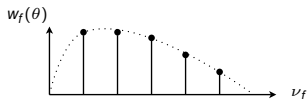


# Vers un modèle de plus haut niveau



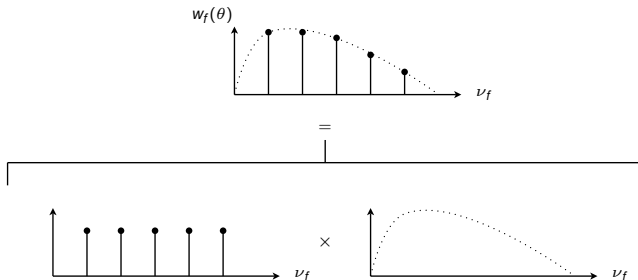


## Vers un modèle de plus haut niveau



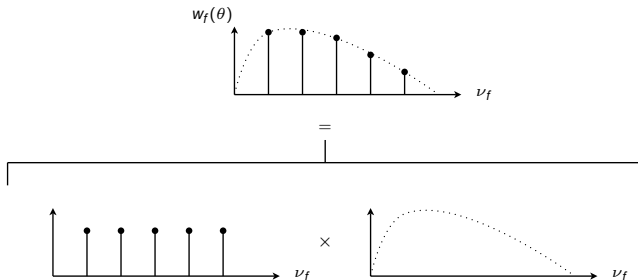


# Vers un modèle de plus haut niveau





# Vers un modèle de plus haut niveau



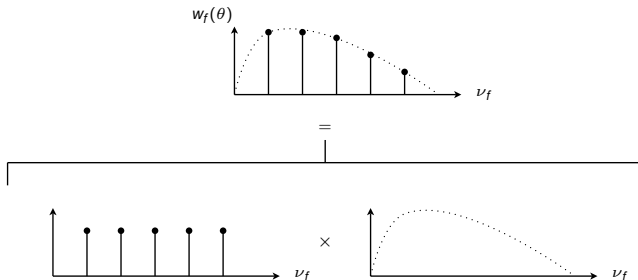
Modèle d'(in)harmonicit  :

$$\nu^h = h\nu^0 \sqrt{1 + \beta h^2}$$





# Vers un modèle de plus haut niveau



Modèle d'(in)harmonicit  :

$$\nu^h = h\nu^0 \sqrt{1 + \beta h^2}$$

Mod les d'enveloppe spectrale :

- Mod les lin aires :  $w_f(\theta) = \sum_{m=1}^M e_m P_m(f)$
- Mod les ARMA :  $w_f(\theta) = \left| \frac{1 + \sum_{q=1}^Q b(q) e^{-i2\pi\nu_f q}}{1 + \sum_{p=1}^P a(p) e^{-i2\pi\nu_f p}} \right|^2$



## Cadre statistique gaussien

- **Modèle de mélange** :  $\hat{v}_f(\Theta) = \sum_{k=1}^K \hat{v}_{fk}$  où  $\hat{v}_{fk} = \sigma_k^2 w_{fk}(\theta)$ 
  - $\sigma_k^2$  est un paramètre scalaire de puissance,
  - $w_{fk}(\theta)$  est la forme spectrale associée.
- **Modèle d'observations gaussiennes** :
  - Le spectre discret  $x_1 \dots x_F \in \mathbb{C}$  est défini pour  $\nu_1 \dots \nu_F \in [0, \frac{1}{2}]$ .
  - On suppose que  $x_f \sim \mathcal{N}(0, \hat{v}_f(\Theta))$  et  $x_1 \dots x_F$  sont indépendants.
- **Vraisemblance du modèle** :  $L(\Theta) \stackrel{C}{=} - \sum_{f=1}^F d_{IS}(\nu_f | \hat{v}_f(\Theta))$ 
  - $\nu_f = |x_f|^2$  est la valeur du périodogramme à la fréquence  $\nu_f$ ,
  - $d_{IS}(a|b) = \frac{a}{b} - \log \frac{a}{b} - 1$  est la **divergence d'Itakura-Saito**.
- **Maximum a posteriori** :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$



## Cadre statistique gaussien

- **Modèle de mélange** :  $\hat{v}_f(\Theta) = \sum_{k=1}^K \hat{v}_{fk}$  où  $\hat{v}_{fk} = \sigma_k^2 w_{fk}(\theta)$ 
  - $\sigma_k^2$  est un paramètre scalaire de puissance,
  - $w_{fk}(\theta)$  est la forme spectrale associée.
- **Modèle d'observations gaussiennes** :
  - Le spectre discret  $x_1 \dots x_F \in \mathbb{C}$  est défini pour  $\nu_1 \dots \nu_F \in [0, \frac{1}{2}]$ .
  - On suppose que  $x_f \sim \mathcal{N}(0, \hat{v}_f(\Theta))$  et  $x_1 \dots x_F$  sont indépendants.
- **Vraisemblance du modèle** :  $L(\Theta) \stackrel{C}{=} - \sum_{f=1}^F d_{IS}(\nu_f | \hat{v}_f(\Theta))$ 
  - $\nu_f = |x_f|^2$  est la valeur du périodogramme à la fréquence  $\nu_f$ ,
  - $d_{IS}(a|b) = \frac{a}{b} - \log \frac{a}{b} - 1$  est la **divergence d'Itakura-Saito**.
- **Maximum a posteriori** :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$



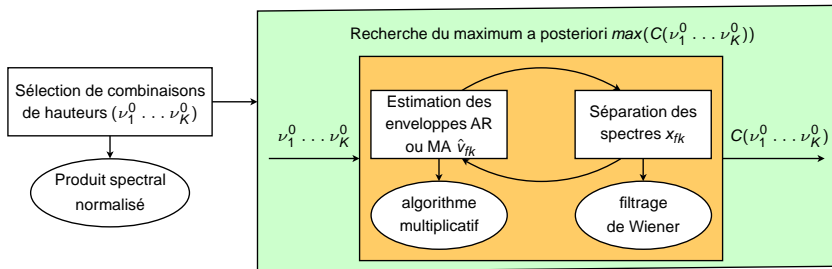
## Cadre statistique gaussien

- Modèle de mélange :  $\hat{v}_f(\Theta) = \sum_{k=1}^K \hat{v}_{fk}$  où  $\hat{v}_{fk} = \sigma_k^2 w_{fk}(\theta)$ 
  - $\sigma_k^2$  est un paramètre scalaire de puissance,
  - $w_{fk}(\theta)$  est la forme spectrale associée.
- Modèle d'observations gaussiennes :
  - Le spectre discret  $x_1 \dots x_F \in \mathbb{C}$  est défini pour  $\nu_1 \dots \nu_F \in [0, \frac{1}{2}]$ .
  - On suppose que  $x_f \sim \mathcal{N}(0, \hat{v}_f(\Theta))$  et  $x_1 \dots x_F$  sont indépendants.
- Vraisemblance du modèle :  $L(\Theta) \stackrel{C}{=} - \sum_{f=1}^F d_{IS}(v_f | \hat{v}_f(\Theta))$ 
  - $v_f = |x_f|^2$  est la valeur du périodogramme à la fréquence  $\nu_f$ ,
  - $d_{IS}(a|b) = \frac{a}{b} - \log \frac{a}{b} - 1$  est la **divergence d'Itakura-Saito**.
- Maximum a posteriori :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$



# Estimation de mélanges spectraux

## ■ Maximum *a posteriori* (MAP) pondéré [1]

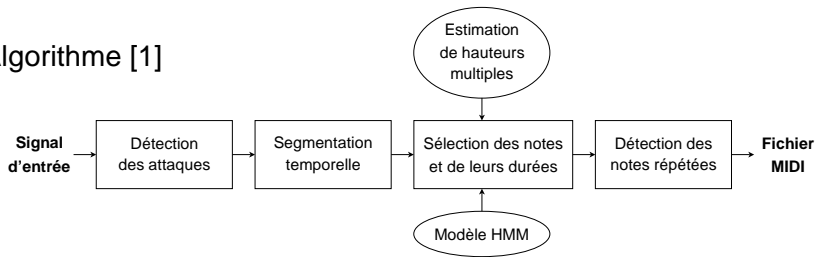


[1] Valentin Emiya, Roland Badeau et Bertrand David. "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6) :1643–1654, août 2010.



# Transcription automatique par HMM

## ■ Algorithme [1]



## ■ Démo

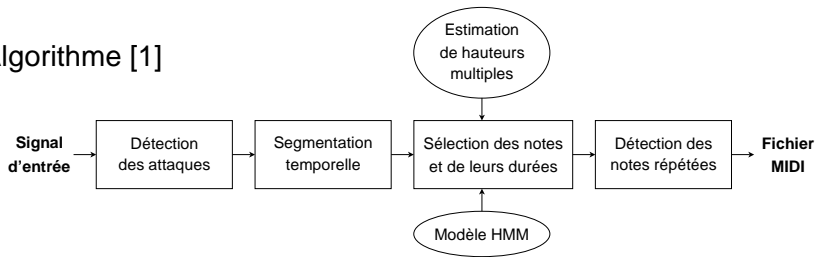
- Signal d'origine (Schumann) :
- Signal retranscrit :

[1] Valentin Emiya, Roland Badeau et Bertrand David. "Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches". Dans *Proc. of European Signal Processing Conference (EUSIPCO)*, Lausanne, Suisse, août 2008.





# Transcription automatique par HMM

## ■ Algorithme [1]



## ■ Démo

- Signal d'origine (Schumann) : 
- Signal retranscrit : 

[1] Valentin Emiya, Roland Badeau et Bertrand David. "Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches". Dans *Proc. of European Signal Processing Conference (EUSIPCO)*, Lausanne, Suisse, août 2008.



## Décompositions temps-fréquence

- **Modèle stationnaire** :  $x_f \sim \mathcal{N}(0, \hat{v}_f(\Theta))$  où  $\hat{v}_f(\Theta) = \sum_{k=1}^K \sigma_k^2 w_{fk}(\theta)$
- **Modèle non stationnaire** :  $x_{fn} \sim \mathcal{N}(0, \hat{v}_{fn}(\Theta))$  où

$$\hat{v}_{fn}(\Theta) = \sum_{k=1}^K w_{fk}(\theta) h_{kn}$$

- $h_{kn}$  est l'activation temporelle associée à la forme spectrale  $w_{fk}(\theta)$
- **Vraisemblance du modèle** :  $L(\Theta) \stackrel{c}{=} - \sum_{n=1}^N \sum_{f=1}^F d_{IS}(v_{fn} | \hat{v}_{fn}(\Theta))$ 
  - $v_{fn} = |x_{fn}|^2$  est le spectrogramme à la fréquence  $\nu_f$  et au temps  $n$
- **Non-negative Matrix Factorization (NMF)** [Lee et Seung, 1999] :

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





## Décompositions temps-fréquence

- Modèle stationnaire :  $\mathbf{x}_f \sim \mathcal{N}(0, \hat{\mathbf{v}}_f(\boldsymbol{\Theta}))$  où  $\hat{\mathbf{v}}_f(\boldsymbol{\Theta}) = \sum_{k=1}^K \sigma_k^2 w_{fk}(\boldsymbol{\theta})$
- Modèle non stationnaire :  $\mathbf{x}_{fn} \sim \mathcal{N}(0, \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$  où

$$\hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}) = \sum_{k=1}^K w_{fk}(\boldsymbol{\theta}) h_{kn}$$

- $h_{kn}$  est l'activation temporelle associée à la forme spectrale  $w_{fk}(\boldsymbol{\theta})$
- Vraisemblance du modèle :  $L(\boldsymbol{\Theta}) \stackrel{c}{=} - \sum_{n=1}^N \sum_{f=1}^F d_{IS}(v_{fn} | \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$ 
  - $v_{fn} = |\mathbf{x}_{fn}|^2$  est le spectrogramme à la fréquence  $\nu_f$  et au temps  $n$
- *Non-negative Matrix Factorization (NMF)* [Lee et Seung, 1999] :

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



## Décompositions temps-fréquence

- Modèle stationnaire :  $\mathbf{x}_f \sim \mathcal{N}(0, \hat{\mathbf{v}}_f(\boldsymbol{\Theta}))$  où  $\hat{\mathbf{v}}_f(\boldsymbol{\Theta}) = \sum_{k=1}^K \sigma_k^2 \mathbf{w}_{fk}(\boldsymbol{\theta})$
- Modèle non stationnaire :  $\mathbf{x}_{fn} \sim \mathcal{N}(0, \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$  où

$$\hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}) = \sum_{k=1}^K \mathbf{w}_{fk}(\boldsymbol{\theta}) h_{kn}$$

- $h_{kn}$  est l'activation temporelle associée à la forme spectrale  $\mathbf{w}_{fk}(\boldsymbol{\theta})$
- Vraisemblance du modèle :  $L(\boldsymbol{\Theta}) \stackrel{c}{=} - \sum_{n=1}^N \sum_{f=1}^F d_{IS}(\nu_{fn} | \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$ 
  - $\nu_{fn} = |\mathbf{x}_{fn}|^2$  est le spectrogramme à la fréquence  $\nu_f$  et au temps  $n$
- *Non-negative Matrix Factorization (NMF)* [Lee et Seung, 1999] :

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



## Décompositions temps-fréquence

- Modèle stationnaire :  $\mathbf{x}_f \sim \mathcal{N}(0, \hat{\mathbf{v}}_f(\boldsymbol{\Theta}))$  où  $\hat{\mathbf{v}}_f(\boldsymbol{\Theta}) = \sum_{k=1}^K \sigma_k^2 \mathbf{w}_{fk}(\boldsymbol{\theta})$
- Modèle non stationnaire :  $\mathbf{x}_{fn} \sim \mathcal{N}(0, \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$  où

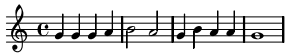
$$\hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}) = \sum_{k=1}^K \mathbf{w}_{fk}(\boldsymbol{\theta}) h_{kn}$$

- $h_{kn}$  est l'activation temporelle associée à la forme spectrale  $\mathbf{w}_{fk}(\boldsymbol{\theta})$
- Vraisemblance du modèle :  $L(\boldsymbol{\Theta}) \stackrel{c}{=} - \sum_{n=1}^N \sum_{f=1}^F d_{IS}(v_{fn} | \hat{\mathbf{v}}_{fn}(\boldsymbol{\Theta}))$ 
  - $v_{fn} = |\mathbf{x}_{fn}|^2$  est le spectrogramme à la fréquence  $\nu_f$  et au temps  $n$
- *Non-negative Matrix Factorization (NMF)* [Lee et Seung, 1999] :

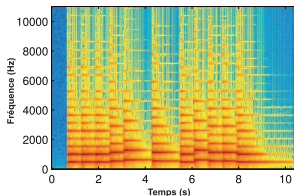
$$\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{W} \mathbf{H}$$



# Non-negative Matrix Factorization (NMF)



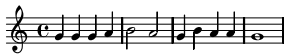
Partition musicale



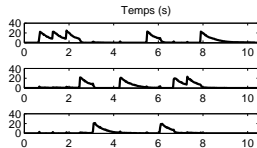
Spectrogramme  $V$



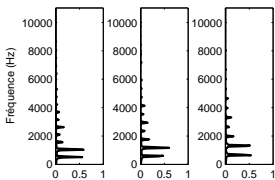
# Non-negative Matrix Factorization (NMF)



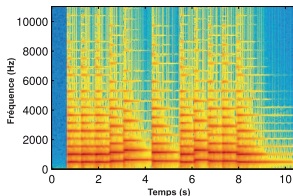
Partition musicale



Activations temporelles  $H$



Formes spectrales  $W$



Spectrogramme  $V$



## $\beta$ -divergence et règles multiplicatives

■ Minimisation du critère  $D(\mathbf{V}|\mathbf{WH}) = \sum_{n=1}^N \sum_{f=1}^F d(v_{fn} | \hat{v}_{fn})$

■  $\beta$ -divergence [Eguchi et Kano, 2001] :

$$d_{\beta}(a|b) = \frac{1}{\beta(\beta-1)} (a^{\beta} + (\beta-1)b^{\beta} - \beta ab^{\beta-1})$$

- $\beta = 2$  correspond à la **distance euclidienne** (EUC),
- $\beta = 1$  correspond à la **divergence de Kullback-Leibler** (KL),
- $\beta = 0$  correspond à la **divergence d'Itakura-Saito** (IS),
- $d_{\beta}(a|b)$  est convexe par rapport à  $b$  si et seulement si  $\beta \in [1, 2]$ ,

■ Règles multiplicatives [Kompass, 2007] :

$$\begin{cases} \mathbf{W} \leftarrow \mathbf{W} \otimes \frac{(\mathbf{V} \otimes (\mathbf{WH})^{\beta-2}) \mathbf{H}^T}{(\mathbf{WH})^{\beta-1} \mathbf{H}^T} \\ \mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T (\mathbf{V} \otimes (\mathbf{WH})^{\beta-2})}{\mathbf{W}^T (\mathbf{WH})^{\beta-1}} \end{cases}$$

- $D(\mathbf{V}|\mathbf{WH})$  décroît à chaque itération si et seulement si  $\beta \in [1, 2]$ .



## $\beta$ -divergence et règles multiplicatives

- Minimisation du critère  $D(\mathbf{V}|\mathbf{WH}) = \sum_{n=1}^N \sum_{f=1}^F d(v_{fn} | \hat{v}_{fn})$
- $\beta$ -divergence [Eguchi et Kano, 2001] :

$$d_{\beta}(a|b) = \frac{1}{\beta(\beta-1)} (a^{\beta} + (\beta - 1)b^{\beta} - \beta ab^{\beta-1})$$

- $\beta = 2$  correspond à la **distance euclidienne** (EUC),
  - $\beta = 1$  correspond à la **divergence de Kullback-Leibler** (KL),
  - $\beta = 0$  correspond à la **divergence d'Itakura-Saito** (IS),
  - $d_{\beta}(a|b)$  est convexe par rapport à  $b$  si et seulement si  $\beta \in [1, 2]$ ,
- Règles multiplicatives [KompPASS, 2007] :
- $$\begin{cases} W \leftarrow W \otimes \frac{(\mathbf{V} \otimes (\mathbf{WH})^{\beta-2}) \mathbf{H}^T}{(\mathbf{WH})^{\beta-1} \mathbf{H}^T} \\ H \leftarrow H \otimes \frac{\mathbf{W}^T (\mathbf{V} \otimes (\mathbf{WH})^{\beta-2})}{\mathbf{W}^T (\mathbf{WH})^{\beta-1}} \end{cases}$$
- $D(\mathbf{V}|\mathbf{WH})$  décroît à chaque itération si et seulement si  $\beta \in [1, 2]$ .



## $\beta$ -divergence et règles multiplicatives

- Minimisation du critère  $D(\mathbf{V}|\mathbf{WH}) = \sum_{n=1}^N \sum_{f=1}^F d(v_{fn} | \hat{v}_{fn})$
- $\beta$ -divergence [Eguchi et Kano, 2001] :

$$d_{\beta}(a|b) = \frac{1}{\beta(\beta-1)} (a^{\beta} + (\beta - 1)b^{\beta} - \beta ab^{\beta-1})$$

- $\beta = 2$  correspond à la **distance euclidienne** (EUC),
- $\beta = 1$  correspond à la **divergence de Kullback-Leibler** (KL),
- $\beta = 0$  correspond à la **divergence d'Itakura-Saito** (IS),
- $d_{\beta}(a|b)$  est convexe par rapport à  $b$  si et seulement si  $\beta \in [1, 2]$ ,
- Règles multiplicatives [Kompas, 2007] :
  - $$\begin{cases} \mathbf{W} \leftarrow \mathbf{W} \otimes \frac{(\mathbf{V} \otimes (\mathbf{WH})^{\beta-2}) \mathbf{H}^T}{(\mathbf{WH})^{\beta-1} \mathbf{H}^T} \\ \mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T (\mathbf{V} \otimes (\mathbf{WH})^{\beta-2})}{\mathbf{W}^T (\mathbf{WH})^{\beta-1}} \end{cases}$$
  - $D(\mathbf{V}|\mathbf{WH})$  décroît à chaque itération si et seulement si  $\beta \in [1, 2]$ .





## Stabilité des algorithmes multiplicatifs

- Introduction d'un pas d'exponentiation  $\eta$  dans les algorithmes multiplicatifs de NMF pour minimiser la  $\beta$ -divergence [1] :

$$W \leftarrow W \otimes \left( \frac{(V \otimes (WH)^{\beta-2}) H^T}{(WH)^{\beta-1} H^T} \right)^\eta \quad (1)$$

$$H \leftarrow H \otimes \left( \frac{W^T (V \otimes (WH)^{\beta-2})}{W^T (WH)^{\beta-1}} \right)^\eta \quad (2)$$

- Décroissance monotone du critère si  $\beta \in [1, 2]$  et  $\eta \in ]0, 1[$
- Stabilité exponentielle ou asymptotique de chacune des règles (1) et (2) si  $\eta \in ]0, \eta^*[$ , où  $\forall \beta \in \mathbb{R}, \eta^* \in ]0, 2]$  et si  $\beta \in [1, 2], \eta^* = 2$
- Instabilité si  $\eta \notin [0, 2] \forall \beta \in \mathbb{R}$
- Le pas  $\eta$  permet de contrôler le taux de convergence

[1] Roland Badeau, Nancy Bertin et Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to non-negativematrix factorization". *IEEE Transactions on Neural Networks*. A paraître.



## Stabilité des algorithmes multiplicatifs

- Introduction d'un pas d'exponentiation  $\eta$  dans les algorithmes multiplicatifs de NMF pour minimiser la  $\beta$ -divergence [1] :

$$W \leftarrow W \otimes \left( \frac{(V \otimes (WH)^{\beta-2}) H^T}{(WH)^{\beta-1} H^T} \right)^\eta \quad (1)$$

$$H \leftarrow H \otimes \left( \frac{W^T (V \otimes (WH)^{\beta-2})}{W^T (WH)^{\beta-1}} \right)^\eta \quad (2)$$

- Décroissance monotone du critère si  $\beta \in [1, 2]$  et  $\eta \in ]0, 1[$
- Stabilité exponentielle ou asymptotique de chacune des règles (1) et (2) si  $\eta \in ]0, \eta^*[$ , où  $\forall \beta \in \mathbb{R}$ ,  $\eta^* \in ]0, 2]$  et si  $\beta \in [1, 2]$ ,  $\eta^* = 2$
- Instabilité si  $\eta \notin [0, 2] \forall \beta \in \mathbb{R}$
- Le pas  $\eta$  permet de contrôler le taux de convergence

[1] Roland Badeau, Nancy Bertin et Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to non-negativematrix factorization". *IEEE Transactions on Neural Networks*. A paraître.



## Stabilité des algorithmes multiplicatifs

- Introduction d'un pas d'exponentiation  $\eta$  dans les algorithmes multiplicatifs de NMF pour minimiser la  $\beta$ -divergence [1] :

$$W \leftarrow W \otimes \left( \frac{(V \otimes (WH)^{\beta-2}) H^T}{(WH)^{\beta-1} H^T} \right)^\eta \quad (1)$$

$$H \leftarrow H \otimes \left( \frac{W^T (V \otimes (WH)^{\beta-2})}{W^T (WH)^{\beta-1}} \right)^\eta \quad (2)$$

- Décroissance monotone du critère si  $\beta \in [1, 2]$  et  $\eta \in ]0, 1]$
- Stabilité exponentielle ou asymptotique de chacune des règles (1) et (2) si  $\eta \in ]0, \eta^*[$ , où  $\forall \beta \in \mathbb{R}$ ,  $\eta^* \in ]0, 2]$  et si  $\beta \in [1, 2]$ ,  $\eta^* = 2$
- Instabilité si  $\eta \notin [0, 2] \forall \beta \in \mathbb{R}$
- Le pas  $\eta$  permet de contrôler le taux de convergence

[1] Roland Badeau, Nancy Bertin et Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization". *IEEE Transactions on Neural Networks*. À paraître.



## Stabilité des algorithmes multiplicatifs

- Introduction d'un pas d'exponentiation  $\eta$  dans les algorithmes multiplicatifs de NMF pour minimiser la  $\beta$ -divergence [1] :

$$W \leftarrow W \otimes \left( \frac{(V \otimes (WH)^{\beta-2}) H^T}{(WH)^{\beta-1} H^T} \right)^\eta \quad (1)$$

$$H \leftarrow H \otimes \left( \frac{W^T (V \otimes (WH)^{\beta-2})}{W^T (WH)^{\beta-1}} \right)^\eta \quad (2)$$

- Décroissance monotone du critère si  $\beta \in [1, 2]$  et  $\eta \in ]0, 1]$
- Stabilité exponentielle ou asymptotique de chacune des règles (1) et (2) si  $\eta \in ]0, \eta^*[$ , où  $\forall \beta \in \mathbb{R}$ ,  $\eta^* \in ]0, 2]$  et si  $\beta \in [1, 2]$ ,  $\eta^* = 2$
- Instabilité si  $\eta \notin [0, 2] \forall \beta \in \mathbb{R}$
- Le pas  $\eta$  permet de contrôler le taux de convergence

[1] Roland Badeau, Nancy Bertin et Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to non-negativematrix factorization". *IEEE Transactions on Neural Networks*. A paraître.



## Stabilité des algorithmes multiplicatifs

- Introduction d'un pas d'exponentiation  $\eta$  dans les algorithmes multiplicatifs de NMF pour minimiser la  $\beta$ -divergence [1] :

$$W \leftarrow W \otimes \left( \frac{(V \otimes (WH)^{\beta-2}) H^T}{(WH)^{\beta-1} H^T} \right)^\eta \quad (1)$$

$$H \leftarrow H \otimes \left( \frac{W^T (V \otimes (WH)^{\beta-2})}{W^T (WH)^{\beta-1}} \right)^\eta \quad (2)$$

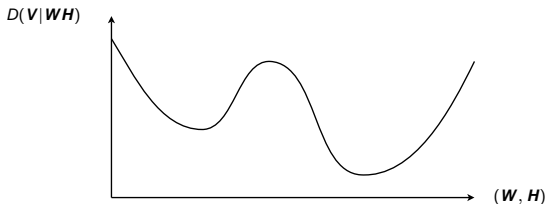
- Décroissance monotone du critère si  $\beta \in [1, 2]$  et  $\eta \in ]0, 1]$
- Stabilité exponentielle ou asymptotique de chacune des règles (1) et (2) si  $\eta \in ]0, \eta^*[,$  où  $\forall \beta \in \mathbb{R}, \eta^* \in ]0, 2]$  et si  $\beta \in [1, 2], \eta^* = 2$
- Instabilité si  $\eta \notin [0, 2] \forall \beta \in \mathbb{R}$
- Le pas  $\eta$  permet de contrôler le taux de convergence

[1] Roland Badeau, Nancy Bertin et Emmanuel Vincent. "Stability analysis of multiplicative update algorithms and application to non-negativematrix factorization". *IEEE Transactions on Neural Networks*. A paraître.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]

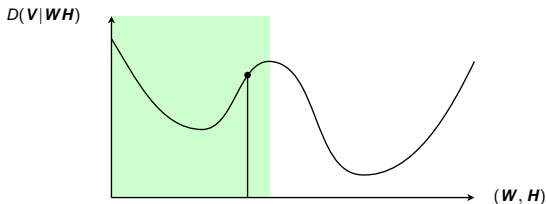


[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]

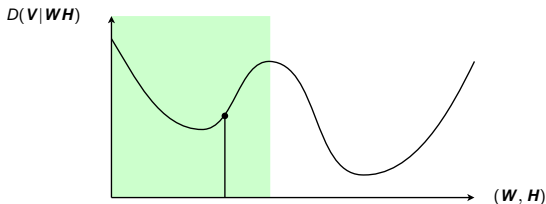


[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]



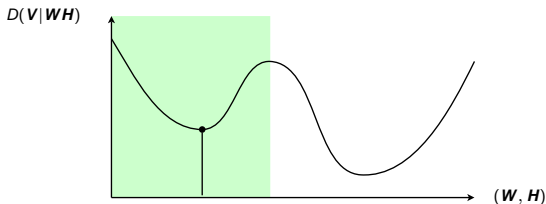
[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.





# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]

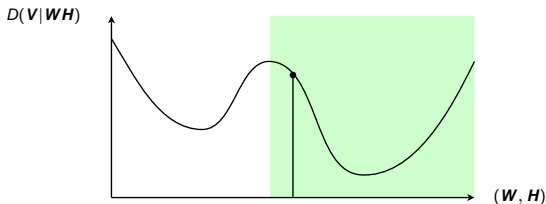


[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]

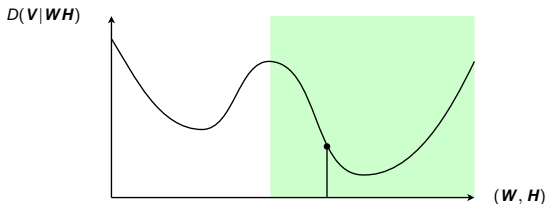


[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]

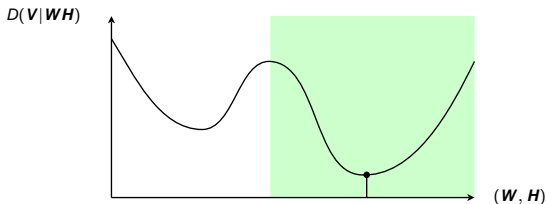


[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Les trois divergences EUC, KL et IS ont des minima locaux [1]



[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.



# Évitement des minima locaux

- Stratégies d'initialisation de l'algorithme [1]
  - Échec d'algorithmes issus de la classification automatique
- Algorithme de "refroidissement simulé" pour IS-NMF [2]
  - Le paramètre  $\beta$  devient fonction de l'indice d'itération  $p$  :
  - La meilleure transcription n'est pas obtenue pour la valeur la plus basse du critère, mais plutôt en contraignant la décomposition.

[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.

[2] Nancy Bertin, Cédric Févotte et Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". Dans *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, avril 2009.



# Évitement des minima locaux

- Stratégies d'initialisation de l'algorithme [1]
  - Échec d'algorithmes issus de la classification automatique
- Algorithme de "refroidissement simulé" pour IS-NMF [2]
  - Le paramètre  $\beta$  devient fonction de l'indice d'itération  $p$  :
  - La meilleure transcription n'est pas obtenue pour la valeur la plus basse du critère, mais plutôt en contraignant la décomposition.

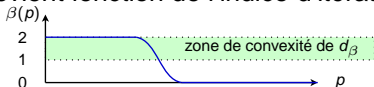
[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.

[2] Nancy Bertin, Cédric Févotte et Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". Dans *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, avril 2009.



# Évitement des minima locaux

- Stratégies d'initialisation de l'algorithme [1]
  - Échec d'algorithmes issus de la classification automatique
- Algorithme de "refroidissement simulé" pour IS-NMF [2]
  - Le paramètre  $\beta$  devient fonction de l'indice d'itération  $p$  :



- La meilleure transcription n'est pas obtenue pour la valeur la plus basse du critère, mais plutôt en contraignant la décomposition.

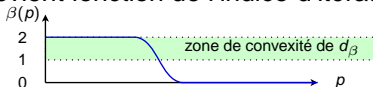
[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.

[2] Nancy Bertin, Cédric Févotte et Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". Dans *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, avril 2009.



# Évitement des minima locaux

- Stratégies d'initialisation de l'algorithme [1]
  - Échec d'algorithmes issus de la classification automatique
- Algorithme de "refroidissement simulé" pour IS-NMF [2]
  - Le paramètre  $\beta$  devient fonction de l'indice d'itération  $p$  :



- La meilleure transcription n'est pas obtenue pour la valeur la plus basse du critère, mais plutôt en contraignant la décomposition.

[1] Nancy Bertin et Roland Badeau. "Initialization, distances and local minima in audio applications of the non-negative matrix factorization". Dans *Acoustics'08*, Paris, France, juillet 2008.

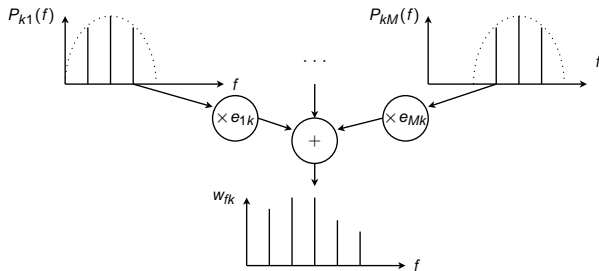
[2] Nancy Bertin, Cédric Févotte et Roland Badeau. "A tempering approach for Itakura-Saito nonnegative matrix factorization. With application to music transcription". Dans *Proc. of IEEE ICASSP*, pages 1545–1548, Taipei, Taiwan, avril 2009.





# Harmonicité et régularité spectrale

- Modèle [1] :  $\hat{v}_{fn} = \sum_{k=1}^K w_{fk}(\theta) h_{kn}$  où  $w_{fk}(\theta) = \sum_{m=1}^M e_{mk} P_{km}(f)$



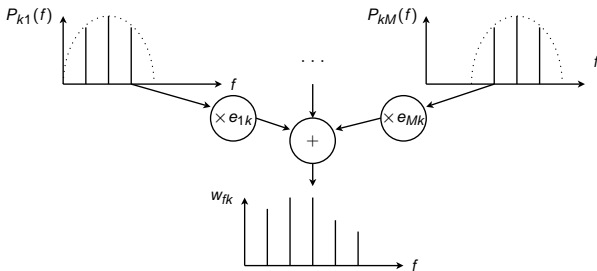
- $P_{km}(f)$  est un motif spectral harmonique prédéfini
- $e_{mk}$  et  $h_{kn}$  sont estimés par un algorithme multiplicatif

[1] Emmanuel Vincent, Nancy Bertin et Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :528- 537, mars 2010.



# Harmonicité et régularité spectrale

- Modèle [1] :  $\hat{v}_{fn} = \sum_{k=1}^K w_{fk}(\theta) h_{kn}$  où  $w_{fk}(\theta) = \sum_{m=1}^M e_{mk} P_{km}(f)$



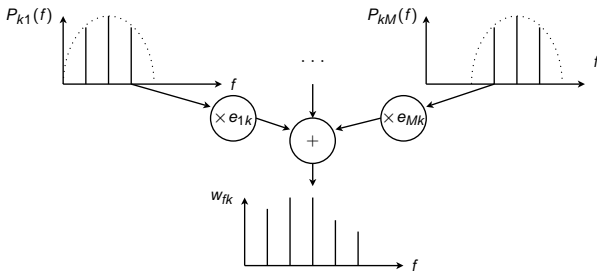
- $P_{km}(f)$  est un motif spectral harmonique prédéfini
- $e_{mk}$  et  $h_{kn}$  sont estimés par un algorithme multiplicatif

[1] Emmanuel Vincent, Nancy Bertin et Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :528- 537, mars 2010.



# Harmonicité et régularité spectrale

- Modèle [1] :  $\hat{v}_{fn} = \sum_{k=1}^K w_{fk}(\theta) h_{kn}$  où  $w_{fk}(\theta) = \sum_{m=1}^M e_{mk} P_{km}(f)$



- $P_{km}(f)$  est un motif spectral harmonique prédéfini
- $e_{mk}$  et  $h_{kn}$  sont estimés par un algorithme multiplicatif

[1] Emmanuel Vincent, Nancy Bertin et Roland Badeau. "Adaptive harmonic spectral decomposition for multiple pitch estimation". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :528- 537, mars 2010.



# Régularité temporelle

- Estimateur MAP :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$  où  $\Theta = \{e_{mk}, h_{kn}\}$
- Distribution a priori structurée par une chaîne de Markov :

$$p(\mathbf{H}) = \prod_{k=1}^K p(h_{k1}) \prod_{n=2}^N p(h_{kn}|h_{k(n-1)})$$

où  $p(h_{kn}|h_{k(n-1)})$  suit une distribution inverse-Gamma

- Algorithme SAGE [1] et règles multiplicatives [2] avec  $\eta = 0.5$

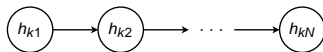
[1] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :538-549, mars 2010.

[2] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". Dans *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, octobre 2009.



# Régularité temporelle

- Estimateur MAP :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$  où  $\Theta = \{e_{mk}, h_{kn}\}$
- Distribution a priori structurée par une chaîne de Markov :



$$p(\mathbf{H}) = \prod_{k=1}^K p(h_{k1}) \prod_{n=2}^N p(h_{kn}|h_{k(n-1)})$$

où  $p(h_{kn}|h_{k(n-1)})$  suit une distribution inverse-Gamma

- Algorithme SAGE [1] et règles multiplicatives [2] avec  $\eta = 0.5$

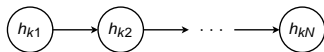
[1] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :538-549, mars 2010.

[2] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". Dans *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, octobre 2009.



## Régularité temporelle

- Estimateur MAP :  $C(\Theta) = L(\Theta) + \log(p(\Theta))$  où  $\Theta = \{e_{mk}, h_{kn}\}$
- Distribution a priori structurée par une chaîne de Markov :



$$p(\mathbf{H}) = \prod_{k=1}^K p(h_{k1}) \prod_{n=2}^N p(h_{kn}|h_{k(n-1)})$$

où  $p(h_{kn}|h_{k(n-1)})$  suit une distribution inverse-Gamma

- Algorithme SAGE [1] et règles multiplicatives [2] avec  $\eta = 0.5$

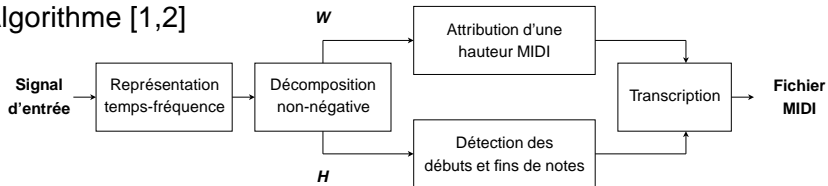
[1] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :538-549, mars 2010.

[2] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Fast Bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription". Dans *Proc. of IEEE WASPAA*, pages 29-32, New York, USA, octobre 2009.



# Transcription automatique par NMF

## ■ Algorithme [1,2]



## ■ Démo

- Signal d'origine (Liszt) :
- Signal retranscrit :

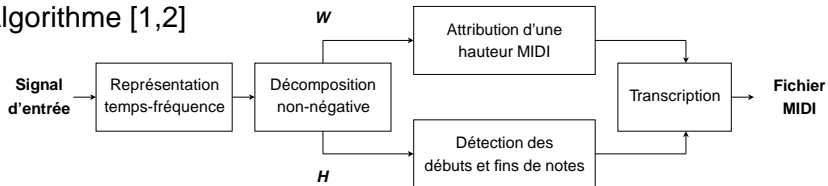
[1] Nancy Bertin, Roland Badeau et Gaël Richard. "Blind signal decompositions for automatic transcription of polyphonic music : NMF and K-SVD on the benchmark". Dans *Proc. of IEEE ICASSP*, volume 1, pages 65-68, Honolulu, Hawaii, USA, avril 2007.

[2] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :538-549, mars 2010.





# Transcription automatique par NMF

## ■ Algorithme [1,2]



## ■ Démo

- Signal d'origine (Liszt) : 
- Signal retranscrit : 

[1] Nancy Bertin, Roland Badeau et Gaël Richard. "Blind signal decompositions for automatic transcription of polyphonic music : NMF and K-SVD on the benchmark". Dans *Proc. of IEEE ICASSP*, volume 1, pages 65-68, Honolulu, Hawaii, USA, avril 2007.

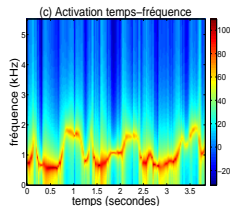
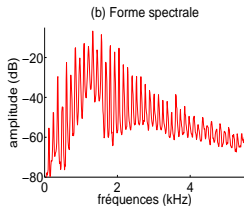
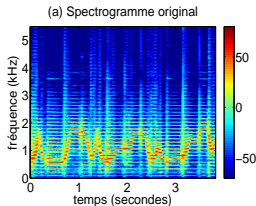
[2] Nancy Bertin, Roland Badeau et Emmanuel Vincent. "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription". *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3) :538-549, mars 2010.





# Activations temps-fréquence

$$\text{Modèle [1]} : \hat{v}_{fn} = \sum_{k=1}^K w_{fk} h_{kn}(f) \text{ où } h_{kn}(f) = \sigma_{kn}^2 \frac{\left| 1 + \sum_{q=1}^Q b_{kn}^{(q)} e^{-i2\pi\nu_f q} \right|^2}{\left| 1 + \sum_{p=1}^P a_{kn}^{(p)} e^{-i2\pi\nu_f p} \right|^2}$$



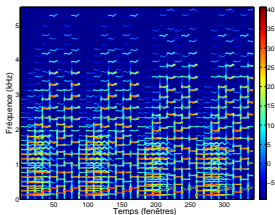
Son de guimbarde décomposé avec un ARMA d'ordre (1,1)

[1] Romain Hennequin, Roland Badeau et Bertrand David. "NMF with time-frequency activations to model non-stationary audio events". *IEEE Transactions on Audio, Speech, and Language Processing*, 2010. À paraître.

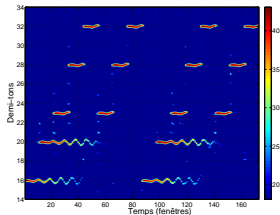


# Variations de fréquences fondamentales

$$\text{Modèle [1]} : \hat{\nu}_{fn} = \sum_{k=1}^K w_{fk}(\nu_{kn}^0) h_{kn} \text{ où } w_{fk}(\nu_{kn}^0) = \sum_{h=1}^{n_h(\nu_{kn}^0)} a_h g(\nu_f - h \nu_{kn}^0)$$



Spectrogramme original



Activations temporelles

## Décomposition d'un extrait du premier Prélude de J.S. Bach

[1] Romain Hennequin, Roland Badeau et Bertrand David. "Time-dependent parametric and harmonic templates in non-negative matrix factorization". Dans *Proc. of DAFX*, Graz, Autriche, septembre 2010.



# Conclusion

## ■ Modélisation de signaux non-stationnaires

- Analyse spectrale à court terme
  - Analyse spectrale à haute résolution
  - Modèles gaussiens incluant harmonicité et enveloppe spectrale
- Analyse temps-fréquence globale
  - Algorithmes sous-espace adaptatifs
  - Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

- Signaux audio et musicaux
  - Analyse / synthèse, représentation, codage,
  - Estimation de hauteurs et de tempo, transcription automatique
- Autres types de données
  - Communications numériques : transmission d'une suite de symboles
  - Analyse modale de la harpe : mesure de signaux d'accélération



# Conclusion

## ■ Modélisation de signaux non-stationnaires

### ■ Analyse spectrale à court terme

- Analyse spectrale à haute résolution
- Modèles gaussiens incluant harmonicité et enveloppe spectrale

### ■ Analyse temps-fréquence globale

- Algorithmes sous-espace adaptatifs
- Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

### ■ Signaux audio et musicaux

- Analyse / synthèse, représentation, codage,
- Estimation de hauteurs et de tempo, transcription automatique

### ■ Autres types de données

- Communications numériques : transmission d'une suite de symboles
- Analyse modale de la harpe : mesure de signaux d'accélération



# Conclusion

## ■ Modélisation de signaux non-stationnaires

### ■ Analyse spectrale à court terme

- Analyse spectrale à haute résolution
- Modèles gaussiens incluant harmonicité et enveloppe spectrale

### ■ Analyse temps-fréquence globale

- Algorithmes sous-espace adaptatifs
- Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

### ■ Signaux audio et musicaux

- Analyse / synthèse, représentation, codage,
- Estimation de hauteurs et de tempo, transcription automatique

### ■ Autres types de données

- Communications numériques : transmission d'une suite de symboles
- Analyse modale de la harpe : mesure de signaux d'accélération



# Conclusion

## ■ Modélisation de signaux non-stationnaires

- Analyse spectrale à court terme
  - Analyse spectrale à haute résolution
  - Modèles gaussiens incluant harmonicité et enveloppe spectrale
- Analyse temps-fréquence globale
  - Algorithmes sous-espace adaptatifs
  - Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

- Signaux audio et musicaux
  - Analyse / synthèse, représentation, codage,
  - Estimation de hauteurs et de tempo, transcription automatique
- Autres types de données
  - Communications numériques : transmission d'une suite de symboles
  - Analyse modale de la harpe : mesure de signaux d'accélération



# Conclusion

## ■ Modélisation de signaux non-stationnaires

- Analyse spectrale à court terme
  - Analyse spectrale à haute résolution
  - Modèles gaussiens incluant harmonicité et enveloppe spectrale
- Analyse temps-fréquence globale
  - Algorithmes sous-espace adaptatifs
  - Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

- Signaux audio et musicaux
  - Analyse / synthèse, représentation, codage,
  - Estimation de hauteurs et de tempo, transcription automatique
- Autres types de données
  - Communications numériques : transmission d'une suite de symboles
  - Analyse modale de la harpe : mesure de signaux d'accélération



# Conclusion

## ■ Modélisation de signaux non-stationnaires

- Analyse spectrale à court terme
  - Analyse spectrale à haute résolution
  - Modèles gaussiens incluant harmonicité et enveloppe spectrale
- Analyse temps-fréquence globale
  - Algorithmes sous-espace adaptatifs
  - Variantes de la NMF avec contraintes d'harmonicité et de régularité

## ■ Applications :

- Signaux audio et musicaux
  - Analyse / synthèse, représentation, codage,
  - Estimation de hauteurs et de tempo, transcription automatique
- Autres types de données
  - Communications numériques : transmission d'une suite de symboles
  - Analyse modale de la harpe : mesure de signaux d'accélération





# Conclusion

- Modélisation de signaux non-stationnaires
  - Analyse spectrale à court terme
    - Analyse spectrale à haute résolution
    - Modèles gaussiens incluant harmonicité et enveloppe spectrale
  - Analyse temps-fréquence globale
    - Algorithmes sous-espace adaptatifs
    - Variantes de la NMF avec contraintes d'harmonicité et de régularité
- Applications :
  - Signaux audio et musicaux
    - Analyse / synthèse, représentation, codage,
    - Estimation de hauteurs et de tempo, transcription automatique
  - Autres types de données
    - Communications numériques : transmission d'une suite de symboles
    - Analyse modale de la harpe : mesure de signaux d'accélération



## Quelques pistes exploratoires

- Associer méthodes HR et décompositions non-négatives
- Approfondir le lien entre modèles de signaux et NMF
- Approfondir conjointement les aspects
  - modélisation (complexité du modèle vs adaptabilité)
  - algorithmique (complexité, convergence globale et stabilité)
- Applications particulièrement visées :
  - Estimation de hauteurs et transcription automatique
  - Séparation de sources informée



## Quelques pistes exploratoires

- Associer méthodes HR et décompositions non-négatives
- Approfondir le lien entre modèles de signaux et NMF
- Approfondir conjointement les aspects
  - modélisation (complexité du modèle vs adaptabilité)
  - algorithmique (complexité, convergence globale et stabilité)
- Applications particulièrement visées :
  - Estimation de hauteurs et transcription automatique
  - Séparation de sources informée



## Quelques pistes exploratoires

- Associer méthodes HR et décompositions non-négatives
- Approfondir le lien entre modèles de signaux et NMF
- Approfondir conjointement les aspects
  - modélisation (complexité du modèle vs adaptabilité)
  - algorithmique (complexité, convergence globale et stabilité)
- Applications particulièrement visées :
  - Estimation de hauteurs et transcription automatique
  - Séparation de sources informée



## Quelques pistes exploratoires

- Associer méthodes HR et décompositions non-négatives
- Approfondir le lien entre modèles de signaux et NMF
- Approfondir conjointement les aspects
  - modélisation (complexité du modèle vs adaptabilité)
  - algorithmique (complexité, convergence globale et stabilité)
- Applications particulièrement visées :
  - Estimation de hauteurs et transcription automatique
  - Séparation de sources informée



## Quelques pistes exploratoires

- Associer méthodes HR et décompositions non-négatives
- Approfondir le lien entre modèles de signaux et NMF
- Approfondir conjointement les aspects
  - modélisation (complexité du modèle vs adaptabilité)
  - algorithmique (complexité, convergence globale et stabilité)
- Applications particulièrement visées :
  - Estimation de hauteurs et transcription automatique
  - Séparation de sources informée

# Remerciements

Merci ...

- aux membres du jury,
- à mes anciens directeurs de thèse,
- à mes doctorants,
- à mes collaborateurs (projets ANR DESAM, DReaM, etc.),
- à mes collègues du département TSI,
- aux institutions qui ont soutenu mes travaux de recherche :





## Partie IV

# Projet ANR DESAM





# Décompositions en Éléments Sonores et Applications Musicales (DESAM)

- Programme ANR Jeunes Chercheuses, Jeunes chercheurs
- Coordination : Télécom ParisTech (CNRS LTCI)
- Période : novembre 2006 - février 2010
- Site Web : <http://www.tsi.enst.fr/~rbadeau/desam/>
- Production scientifique
  - 8 articles de revues internationales
  - 23 articles de conférences internationales avec comité de lecture
  - 5 interventions dans une session spéciale *Acoustics'08*
  - Dépôt d'une boîte à outils (*DESAM Toolbox*)



# Partenaires du projet



## Télécom ParisTech

### **Permanents :**

Roland Badeau (MCF)  
Bertrand David (MCF)  
Gaël Richard (PROF)

### **Post-doctorant :**

Mathieu Lagrange

### **Doctorants :**

Valentin Emiya  
Nancy Bertin  
Jean-Louis Durrieu  
Romain Hennequin

# Partenaires du projet



Télécom ParisTech



Equipe Lutheries - Acoustique - Musique

**Permanent :**  
Laurent Daudet (PROF)

**Doctorant :**  
Emmanuel Ravelli

# Partenaires du projet



Télécom ParisTech



Equipe Lutheries - Acoustique - Musique



Laboratoire de Mécanique et d'Acoustique

**Permanent :**  
Olivier Derrien (MCF)

# Partenaires du projet



Télécom ParisTech



Equipe Lutheries - Acoustique - Musique



Laboratoire de Mécanique et d'Acoustique



Laboratoire Bordelais de Recherche en Informatique

**Permanent :**  
Sylvain Marchand (MCF)



# Contenu du projet

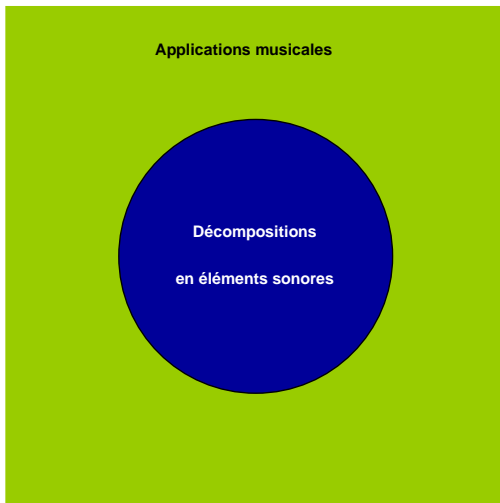


## Contenu du projet

Décompositions  
en éléments sonores



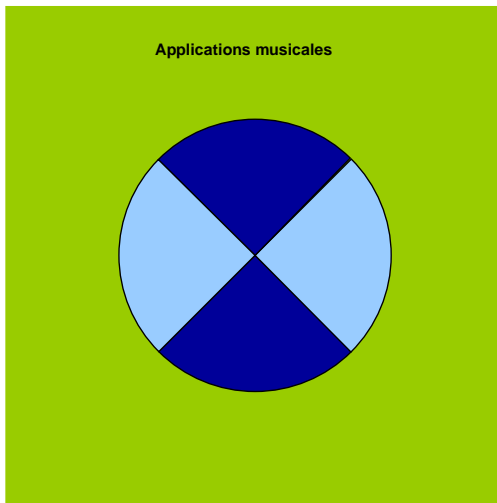
# Contenu du projet





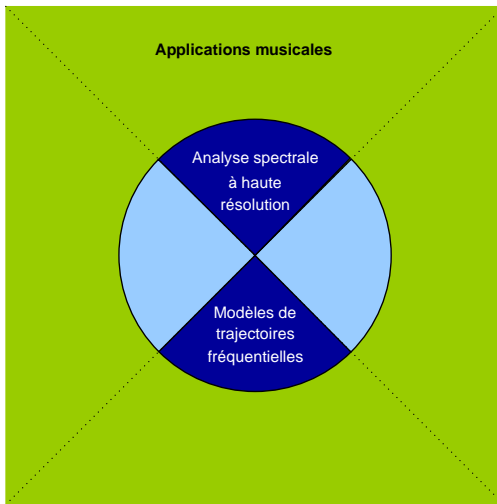


# Contenu du projet



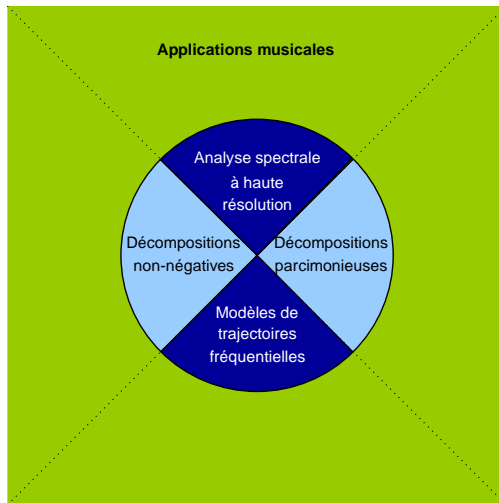


# Contenu du projet



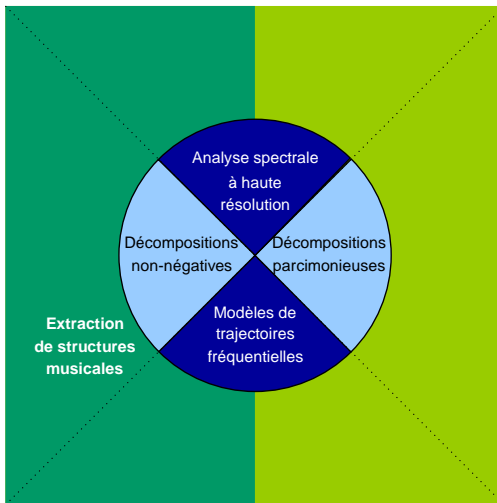


# Contenu du projet



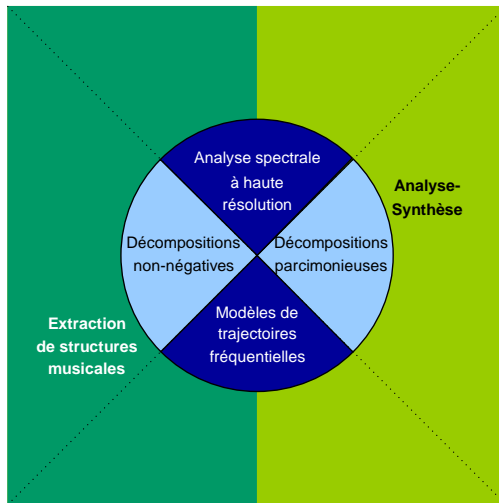


# Contenu du projet



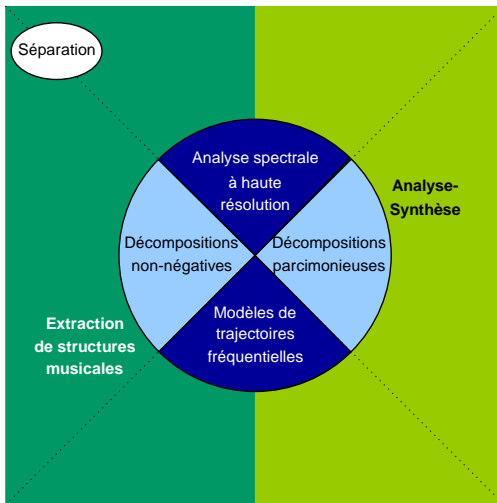


# Contenu du projet



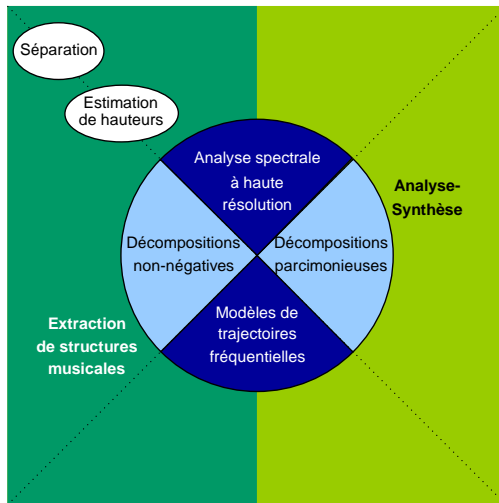


# Contenu du projet



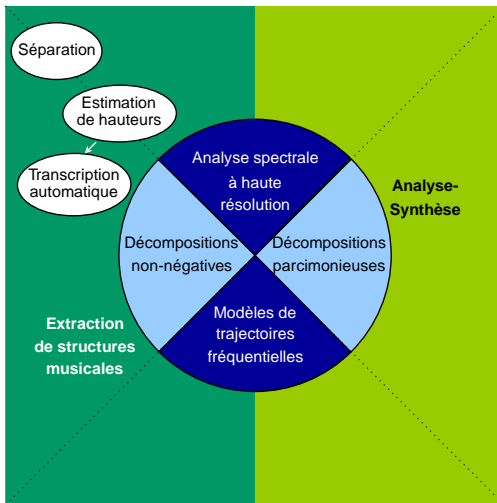


# Contenu du projet





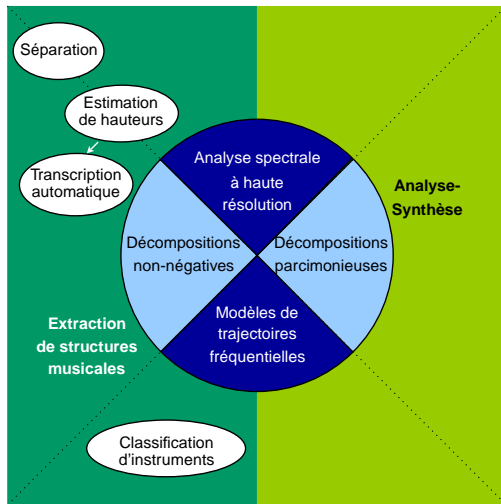
# Contenu du projet





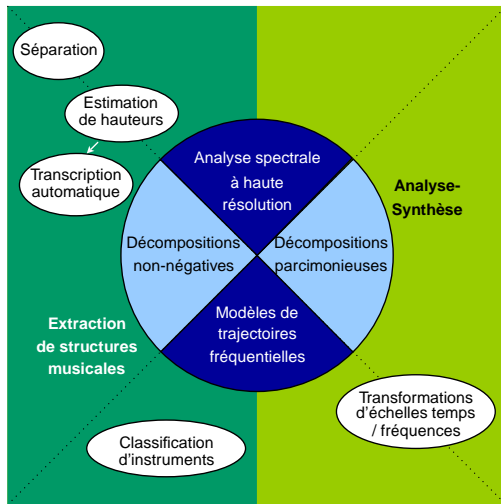


# Contenu du projet



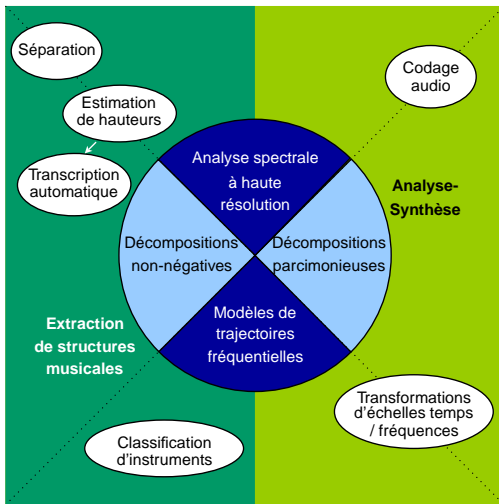


# Contenu du projet



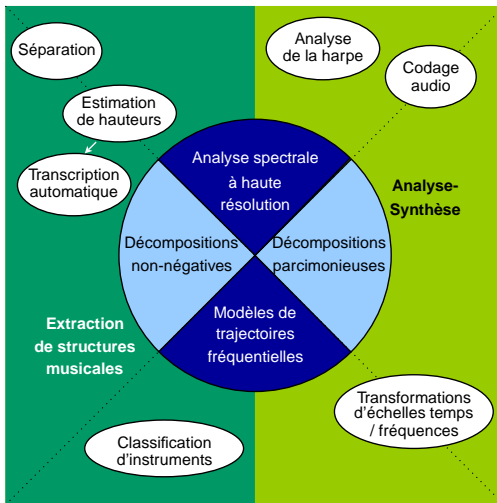


# Contenu du projet





# Contenu du projet





## Articles de revues

- "Explicit Modeling of Temporal Dynamics within Musical Signals for Acoustical Unit Similarity", *M. Lagrange, M. Raspaud, R. Badeau et G. Richard*, Pattern Recognition Letters (PRNSA), vol. 31, no. 12, pp. 1498-1506, sep. 2010.
- "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle", *V. Emiya, R. Badeau et B. David*, IEEE TASLP, vol. 18, no. 6, pp. 1643-1654, août 2010.
- "Enforcing Harmonicity and Smoothness in Bayesian Non-negative Matrix Factorization Applied to Polyphonic Music Transcription", *N. Bertin, R. Badeau et E. Vincent*, IEEE TASLP, vol. 18, no. 3, pp. 538-549, mars 2010.
- "Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation", *E. Vincent, N. Bertin et R. Badeau*, IEEE TASLP, vol. 18, no. 3, pp. 528-537, mars 2010.
- "Audio sparse decompositions in parallel - Let the greed be shared!", *L. Daudet*, IEEE Signal Processing Magazine, vol. 27, no. 2, pp. 90-96, mars 2010.
- "Sympathetic String Modes in the Concert Harp", *J.-L. Le Carrou, F. Gautier et R. Badeau*, Acta Acustica, vol. 95, no. 4, pp. 744-752, juillet/août 2009.
- "Union of MDCT bases for audio coding", *E. Ravelli, G. Richard et L. Daudet*, IEEE TASLP, vol. 16, no. 8, pp. 1361-1372, nov. 2008.



## Boîte à outils *DESAM Toolbox*

- Bibliothèque de fonctions Matlab et de scripts de démonstration
- Implémentation d'une partie des résultats du projet DESAM
- Diffusion en *GNU General Public License (GPL)*
- Mainteneur : Mathieu Lagrange (IRCAM)
- Documentation :
  - "The DESAM toolbox : spectral analysis of musical audio", *M. Lagrange, R. Badeau, B. David, N. Bertin, J. Echeveste, O. Derrien, S. Marchand et L. Daudet*, DAFx 2010.
- Téléchargement :
  - <http://www.tsi.enst.fr/ao/logiciels-et-bases-de-donnees>
  - <https://plato.enst.fr>



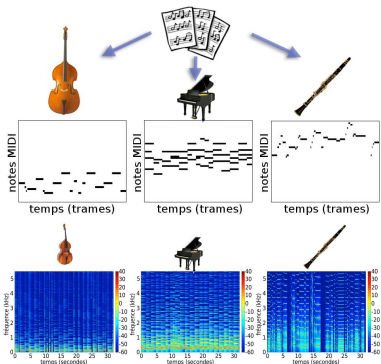
## Partie V

# Démonstrations



# Séparation informée par partition musicale

## ■ Algorithmme [1]



## ■ Round Midnight (Thelonious Monk) :

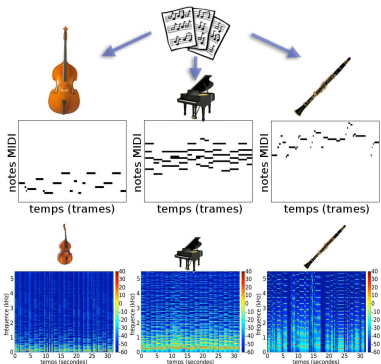
[1] Romain Hennequin, Bertrand David et Roland Badeau. "Score informed audio source separation using a parametric model of non-negative spectrogram". Dans *Proc. of IEEE ICASSP*, Prague, République Tchèque, mai 2011. Soumis pour publication.





# Séparation informée par partition musicale

## ■ Algorithm [1]



## ■ Round Midnight (Thelonious Monk) :



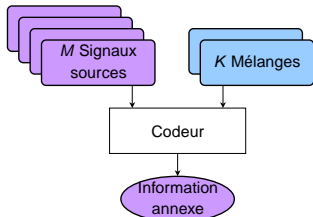
[1] Romain Hennequin, Bertrand David et Roland Badeau. "Score informed audio source separation using a parametric model of non-negative spectrogram". Dans *Proc. of IEEE ICASSP*, Prague, République Tchèque, mai 2011. Soumis pour publication.



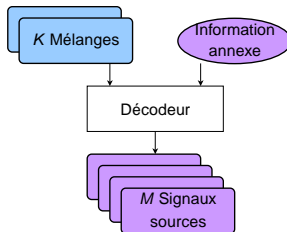
# Séparation informée par tatouage

## ■ Algorithme [1,2]

1) Codage



2) Décodage



## ■ Mix Tape (Jim's Big Ego) :

[1] Antoine Liutkus, Roland Badeau et Gaël Richard. "Informed source separation using multichannel NMF". Dans *Proc. of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Saint Malo, France, septembre 2010.

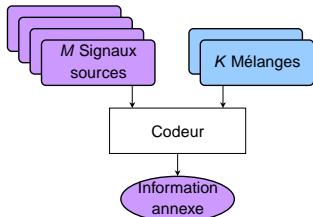
[2] Laurent Girin, Antoine Liutkus, Gaël Richard et Roland Badeau. "Procédé et dispositif de formation d'un signal mixé numérique audio, procédé et dispositif de séparation de signaux, et signal correspondant". Demande de brevet no. B10/3035FR / GBO, octobre 2010. Institut Polytechnique de Grenoble et Institut Télécom, Télécom ParisTech.



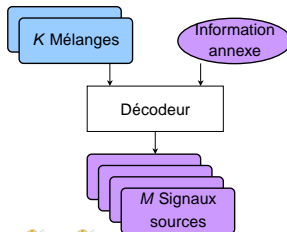
# Séparation informée par tatouage

## ■ Algorithme [1,2]

1) Codage



2) Décodage



## ■ Mix Tape (Jim's Big Ego) :



























[1] Antoine Liutkus, Roland Badeau et Gaël Richard. "Informed source separation using multichannel NMF". Dans *Proc. of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Saint Malo, France, septembre 2010.

[2] Laurent Girin, Antoine Liutkus, Gaël Richard et Roland Badeau. "Procédé et dispositif de formation d'un signal mixé numérique audio, procédé et dispositif de séparation de signaux, et signal correspondant". Demande de brevet no. B10/3035FR / GBO, octobre 2010. Institut Polytechnique de Grenoble et Institut Télécom, Télécom ParisTech.



## Séparation sinusoïdes et bruit

- Principe : projection sur l'espace signal ou sur l'espace bruit [1,2]

Instrument	Original	Sinusoïdes	Bruit
Piano			 
Guitare			 
Violon			 
Flûte			 
Saxophone			 
Cloche			 





[1] Roland Badeau, Rémy Boyer et Bertrand David. "EDS parametric modeling and tracking of audio signals". Dans Proc. of the 5th International Conference on Digital Audio Effects (DAFx), pages 139-144, Hambourg, Allemagne, septembre 2002.

[2] Bertrand David, Gaël Richard et Roland Badeau. "An EDS modelling tool for tracking and modifying musical signals". Dans Proc. of Stockholm Music Acoustics Conference (SMAC), volume 2, pages 715-718, Stockholm, Suède, août 2003.



# Séparation et estimation de rythmiques

## ■ Séparation de pistes de batterie [1]

- Original (Aerosmith) : 
- Batterie séparée : 
- Remix - plus de batterie : 
- Remix - moins de batterie : 

## ■ Estimation de la pulsation rythmique [2]

- Pink Floyd :
- Brad Mehldau :





[1] Olivier Gillet et Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3) :529-540, mars 2008.

[2] Miguel Alonso Arevalo, Roland Badeau, Bertrand David et Gaël Richard. "Musical tempo estimation using noise subspace projections". Dans *Proc. of IEEE WASPAA*, pages 95-98, New Paltz, New York, USA, octobre 2003.





# Séparation et estimation de rythmiques

## ■ Séparation de pistes de batterie [1]

- Original (Aerosmith) : 
- Batterie séparée : 
- Remix - plus de batterie : 
- Remix - moins de batterie : 

## ■ Estimation de la pulsation rythmique [2]

- Pink Floyd : 
- Brad Mehldau : 















[1] Olivier Gillet et Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3) :529-540, mars 2008.

[2] Miguel Alonso Arevalo, Roland Badeau, Bertrand David et Gaël Richard. "Musical tempo estimation using noise subspace projections". Dans *Proc. of IEEE WASPAA*, pages 95-98, New Paltz, New York, USA, octobre 2003.



## Codage audio

Son original : 

MDCT		ESM	
9 bits/spl		8.9 bits/spl	
8 bits/spl			
7 bits/spl		6.8 bits/spl	
6 bits/spl		6.4 bits/spl	
5 bits/spl		4.7 bits/spl	
4 bits/spl		4.4 bits/spl	
3 bits/spl		3.2 bits/spl	
2 bits/spl		2.1 bits/spl	