



# Applications des méthodes Haute-Résolution

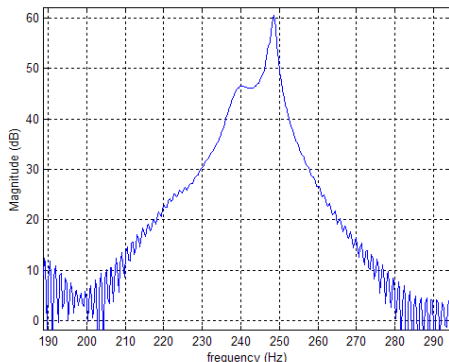
*En acoustique musicale et signaux audiofréquences*

B. DAVID

Télécom ParisTech  
Département Traitement du Signal et des Images  
Groupe Audio

Cours LIESSE  
5 novembre 2012

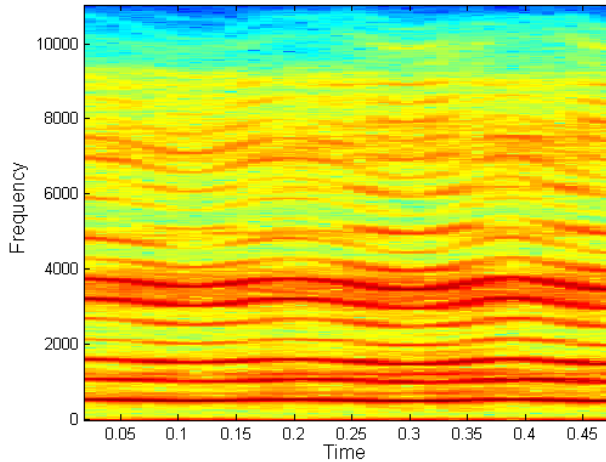
## Problématiques



- Observé couramment pour piano, guitare, clavecin,...
- Comment estimer les fréquences et amortissements ?
- Comment automatiser les mesures ?
- Validité du modèle (non-linéarités ?), ordre du modèle
- attaques / extinctions / vibratos / glissanos ...

# Problématiques

voice



la plupart des paysages sonores sont *instationnaires*.

## Problématique : décomposition

### Son de piano

- Une note de piano
- Modèle de signal : harmoniques + bruit

## Autres utilisations

- estimation : paramètres pour la synthèse, estimation de hauteur (M. Cristensen par exemple)
- analyse modale,
- segmentation,
- décompositions (H+N) : pour l'analyse/synthèse, pour le rythme (M. Alonso),...

## Journée méthodes HR / SFA / TP, décembre 2011

- 10h00 Introduction de la journée (B. David)
- 10h30 Méthodes HR et représentations adaptatives : R. Badeau, B. David
- 11h30 Prétraitements : R. Badeau, B. David
- 11h45 Première table ronde : estimation de l'ordre des modèles
  - principe d'Ester R. Badeau
  - B. Elie et les plaques vibrantes
- 12h45 Repas
- 14h00 Table ronde modes proches
  - Intro B. David
  - Mélodie Monteil : séparation de modes proches dans les sons de steelpans
  - J-L Le Carrou, identification de modes sympathiques : cas de la harpe de concert
- 15h30 Interlude : Méthodes HR avec échantillonnage non uniforme (G. Chardon)
- 16h00 Table ronde analyse modale :
  - intro et contexte : Jean-Loic Le Carrou
  - méthode d'identification modale en moyennes fréquences : Kerem Ege
  - applications : B. Elie
  - K. Ege : Comportement vibratoire d'une table d'harmonie de piano en BF et MF
  - M. Rébillat : Estimation des paramètres d'élasticité et d'amortissement d'une plaque "sandwich"
  - A. Sirdey
- 17h00 Une autre application aux vibrations : J. Drouet (LVA, INSA de Lyon) : Estimation paramétrique d'un filtre de Wiener mesuré sur moteur Diesel

## Quelques jalons

- Méthodes Haute-résolution : Prony (1795), Kumaresan (années 80), ESPRIT (Roy, 86),
- application à la musique : Laroche (93)
- contexte adaptatif : Duvaut (SinTrack, 96), Strobach (Loraf, 97  $nr^2$ ), Abed-Meraim (OPAST, 2000,  $nr$ ), Badeau (FAPI, 2004,  $nr$ )

## Structure de l'exposé

1. Rappels : analyse spectrale HR et modèle sous-jacent
2. Application à l'estimation de modes propres proches
3. Décomposition Harmoniques/Bruit
4. Méthodes adaptatives



## Modélisation

### Modèle de signal avec bruit additif

- signal non bruité :  $x(t) = \sum_{k=0}^{r-1} a_k e^{j2\pi f_k t + \phi_k} = \sum_{k=0}^{r-1} b_k z_k^t$
- bruit blanc, gaussien, complexe,  $p_b(\eta(t)) = \frac{1}{\pi\sigma^2} e^{-\frac{|\eta(t)|^2}{\sigma^2}}$
- modèle de signal  $s(t) = x(t) + \eta(t)$

### Notations vectorielles

$$\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}, \quad \mathbf{v}(z) = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{N-1} \end{bmatrix}, \quad \mathbf{V} = [\mathbf{v}(z_0) \quad \mathbf{v}(z_1) \quad \dots \quad \mathbf{v}(z_{r-1})]$$

$\mathbf{V}$  Matrice de Vandermonde

on en déduit  $\mathbf{x} = \mathbf{V}\mathbf{b}$

## Maximum de vraisemblance

- vraisemblance  $L(\theta) = f_{\theta}(\mathbf{x}|\theta)$
- Maximiser la vraisemblance peut se ramener à maximiser la fonction

$$\mathcal{J}(z_0, \dots, z_{(K-1)}) = \mathbf{s}(t)^H \mathbf{V} \left( \mathbf{V}^H \mathbf{V}^N \right)^{-1} \mathbf{V}^N \mathbf{s}(t)$$

- en général très difficile (optimiser une fonction de  $K$  variables), pas de solution analytique,
- nombreux maxima locaux
- MV = Moindres carrés pour estimer les amplitudes  $\mathbf{b}$  (MC linéaires) et  $\sigma^2$
- Pour des composantes bien séparées :  
Maximiser  $\mathcal{J}$  c'est trouver les  $K$  plus grands maxima du périodogramme.

## Méthodes HR

- la limite de résolution de Fourier contraint la méthode par ML,
- intérêt des méthodes HR : affranchies de cette limitation, peuvent distinguer les composantes proches,
- dans la pratique :
  - analyse/synthèse : pouvoir réduire les tailles de fenêtres,
  - estimation de composantes amorties.
- défauts :
  - coût de calcul ( $\propto N^3$ )
  - bruit non blanc, preprocessing
  - pb de .... "résolution" (lié à RSB  $\neq +\infty$ )

## Principes de base

### Prony et prédiction linéaire (1795)

- rappel du modèle non bruité :  $x(t) = \sum_{k=0}^{r-1} b_k z_k^t$
- une seule composante :  $x(t) = z_0 x(t-1)$
- à  $r$  composantes :  $x(t) = -\sum_{k=1}^r p_k x(t-k)$
- polynôme annulateur :  $P(z) = z^r + \sum_{k=1}^r p_k z^{r-k} = \prod_{k=0}^r (z - z_k)$

### Idée de Prony

- le pb de trouver les  $z_k$  est non-linéaire
- le pb de trouver les  $p_k$  est *linéaire* (prédiction)
- algo :
  1. minimiser l'erreur de prédiction (en présence de bruit)  $\epsilon(t) = \sum_{k=0}^r p_k s(t-k)$ ,  $\hookrightarrow \hat{p}_k$
  2. chercher les racines du polynôme  $P \hookrightarrow \hat{z}_k$

## Méthodes sous-espace : réduction de rang

### Exemple simple

- signal non bruité :  $x(t) = b_0 z_0^t + b_1 z_1^t$ ,  $z_k = e^{-\delta_k + j2\pi f_k}$

- en matriciel  $\mathbf{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(n-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ z_0 & z_1 \\ \vdots & \vdots \\ z_0^{n-1} & z_1^{n-1} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{V}\mathbf{b}$

- sous-espace signal :  $\mathbf{x} \in \text{span}\mathbf{V} = \mathcal{E}_{\text{sig}}$
- $z_0 \neq z_1$ ,  $\dim \mathcal{E}_{\text{sig}} = 2$

## Décomposition de la matrice de covariance

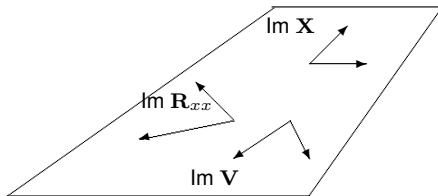
- pour une fenêtre d'analyse de  $l + n$  points la matrice de données est maintenant de dimension  $n \times l$  avec  $n, l > r$ .

$$\mathbf{X} = \begin{bmatrix} x(0) & \cdots & x(l-1) & x(l) \\ x(l) & \cdots & x(l) & x(l+1) \\ \vdots & \cdots & \vdots & \vdots \\ x(n-1) & \cdots & x(l+n-1) & x(l+n-1) \end{bmatrix}$$

- cette matrice est de rang  $r$
- matrice de covariance des données non bruitées  $\mathbf{R}_{xx} = \frac{1}{l} \mathbf{X} \mathbf{X}^H : n \times n$

$$\mathbf{R}_{xx} = \mathbf{V} \mathbf{P} \mathbf{V}^H$$

Sous-espace signal :



## Méthodes sous-espace : décomposition sous-espace signal/bruit

### Base propre orthonormée de l'espace signal

- $\mathbf{W}$  base propre o.n. de  $\mathbf{R}_{xx}$  (en général  $\neq \mathbf{V}$ )
- $\mathbf{W}$  associée à  $r$  v.p.  $\lambda_k > 0$
- $\mathbf{W}_\perp$ , complémentaire orth. associée à  $n - r$  v.p. = 0
- $\mathbf{W}$  base propre o.n. de  $\mathbf{R}_{ss} = \mathbb{E}\hat{\mathbf{R}}_{ss}$  associée aux v.p.  $\lambda_k + \sigma^2$

### MUSIC

- MUSIC : 
$$\frac{1}{\|\text{proj. sur l'espace bruit}\|^2}$$
- $$P_{\text{music}} = \frac{1}{\|\mathbf{W}_\perp^H \mathbf{v}(z)\|^2}$$
- souvent plus efficace :  $\mathbf{W}_\perp \mathbf{W}_\perp^H = \mathbf{I} - \mathbf{W}\mathbf{W}^H$

### ESPRIT

- Utilise une propriété supplémentaire : l'invariance rotationnelle,
- Estime directement les pôles.

## Méthode ESPRIT

### Invariance rotationnelle

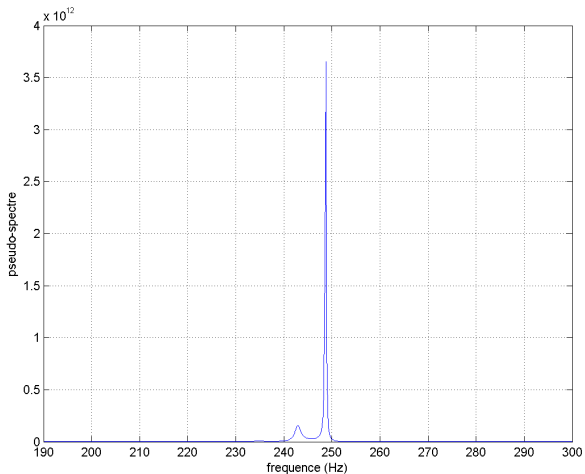
- $$\begin{bmatrix} 1 & 1 \\ z_0 & z_1 \\ \vdots & \vdots \\ z_0^{n-2} & z_1^{n-2} \end{bmatrix} \begin{bmatrix} z_0 & 0 \\ 0 & z_1 \end{bmatrix} = \begin{bmatrix} z_0 & z_1 \\ z_0^2 & z_1^2 \\ \vdots & \vdots \\ z_0^{n-1} & z_1^{n-1} \end{bmatrix}$$
- $$\mathbf{V}_\downarrow \mathbf{D} = \mathbf{V}_\uparrow$$
- $\mathbf{V}$  et  $\mathbf{W}$  engendrent le même sous-espace  $\Rightarrow \exists \Phi$  telle que  $\mathbf{W}_\downarrow \Phi = \mathbf{W}_\uparrow$  avec  $\Phi = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$

### Algorithme

- on estime  $\mathbf{W}$  à partir de la SVD de  $\hat{\mathbf{R}}_{SS}$
- on estime  $\Phi = \mathbf{W}_\downarrow^\dagger \mathbf{W}_\uparrow$
- on estime les  $z_k$  comme v.p. de  $\Phi$  (matrice spectrale)



## Application de MUSIC à un partiel de guitare

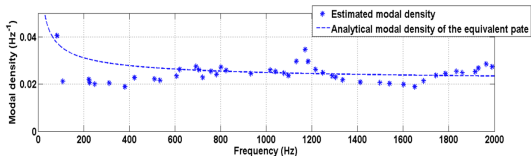
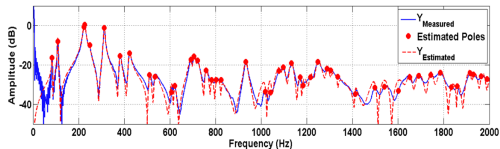


## ANR PAFI

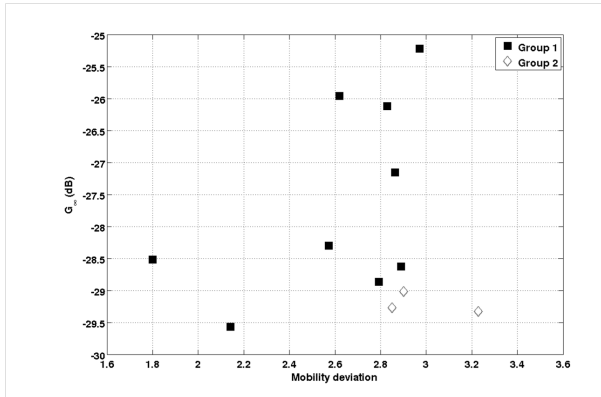
- travail de plus de 10 ans avec les facteurs d'instrument
- cadre d'un projet ANR : PAFI (Plate Forme d'Aide à la Facture Instrumentale)
- développement d'outil d'investigation et d'analyse
- *Démonstration*

## Analyse en moyennes fréquences

- Mesures d'admittance au chevalet :  $\gamma(\nu) = \frac{\hat{v}(\nu)}{\hat{F}(\nu)}$
- modélisation par ESPRIT



## Analyse en moyennes fréquences



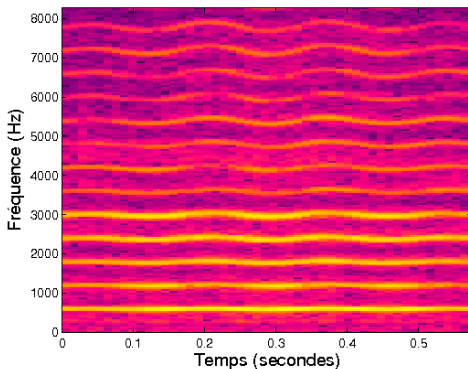
## Application de ESPRIT à un son de piano

L'un des deux est une resynthèse

- son1
- son2

## Exemple du vibrato de violon

son de violon



### Transformée, peak picking, lien a posteriori

- TFCT, Serra (sms, 90)
- HMM, Depalle (93)
- filtrage particulière, Dubois (2005)

## mise à jour directe par LMS adaptatif, SINTRACK (Duvaut 96, David 2002)

Rappel, équation de prédiction :  $\epsilon(t) = \sum_{k=0}^r p_k s(t-k) = \mathbf{p}^T \mathbf{s}(t)$

### Algorithme

Init : Matrix-pencil + calcul des  $p_k$  (coeff. du polynôme prédicteur)

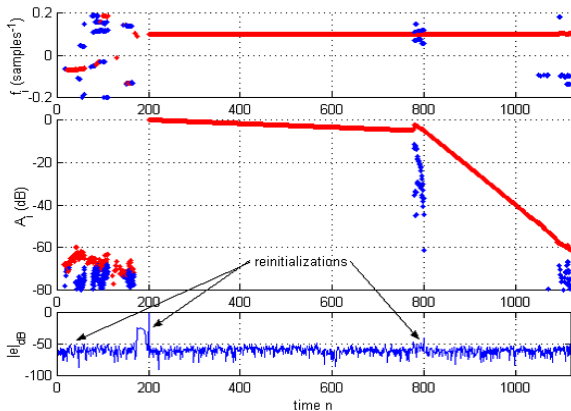
1. calcul de l'erreur de prédiction  $\epsilon(t)$
2. si  $\epsilon(t) > \text{seuil}$  : reinitialisation
3. gradient adaptatif sur  $\mathbf{p}$  :  $\mathbf{p}(t+1) = \mathbf{p}(t) - \mu \epsilon(t) \mathbf{s}(t)$
4. calcul des poles comme racines du polynôme

### avantages et inconvénients

- très rapide (pour la partie lms)
- pas une "vraie" mise à jour des pôles
- réglages de l'algo : seuil, nombre de composantes, pas du lms.

## Simulations

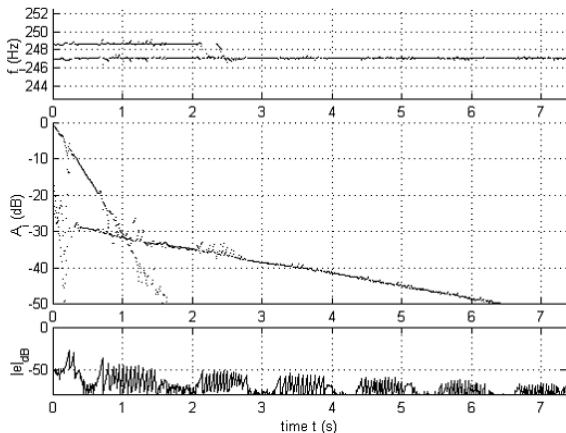
### Cas d'un composante simple à double décroissance





## Résultats sur signaux réels

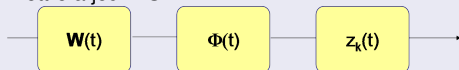
Résultats pour le Mi grave d'une guitare :  
48 kHz, 3ème partiel, pincé, prétraitement (filtrage)



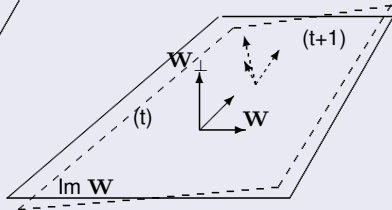
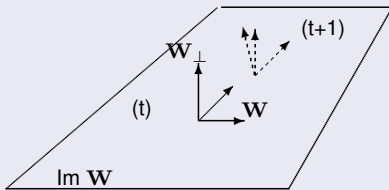
## Suivre les poles

### Mises à jour

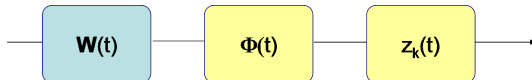
- mettre les pôles à jour directement ?
- mettre à jour ESPRIT :



### Mise à jour de l'espace signal



## Rendre ESPRIT adaptatif



### Références

- Thèse R. Badeau (2005), FAPI (IEEE TSP, 2004) : mise à jour de  $\mathbf{W}(t)$ ,
- B. David : mise à jour des  $z_k$  (Icassp, 2006)

### Suivi du sous-espace

- mise à jour de la matrice de corrélation :

$$l\hat{\mathbf{R}}_{ss} = \mathbf{C}_{ss}(t) = \beta \mathbf{C}_{ss}(t-1) + \mathbf{s}(t)\mathbf{s}(t)^H$$

- Itération orthogonale appliquée *séquentiellement*
- Approximation de projection  $\mathbf{W}(t)\mathbf{W}(t)^H \approx \mathbf{W}(t-1)\mathbf{W}(t-1)^H$

## FAPI (Badeau, 2004)

**Initialization :**

$$W(0) = \begin{bmatrix} I_r \\ \mathbf{0}_{(n-r) \times r} \end{bmatrix}, \quad Z(0) = I_r$$

For each time step do

Input vector : $\mathbf{x}(t)$	
<b>FAPI main section</b>	<b>Cost (MAC)</b>
$\mathbf{y}(t) = \mathbf{W}(t-1)^H \mathbf{x}(t)$	$nr$
$\mathbf{h}(t) = \mathbf{Z}(t-1) \mathbf{y}(t)$	$r^2$
$\mathbf{g}(t) = \frac{\mathbf{h}(t)}{\beta + \mathbf{y}(t)^H \mathbf{h}(t)}$	$2r$
$\epsilon^2(t) = \ \mathbf{x}(t)\ ^2 - \ \mathbf{y}(t)\ ^2$	$n + r$
$\tau(t) = \frac{\epsilon^2(t)}{1 + \epsilon^2(t) \ \mathbf{g}(t)\ ^2 + \sqrt{1 + \epsilon^2(t) \ \mathbf{g}(t)\ ^2}}$	$r$
$\eta(t) = 1 - \tau(t) \ \mathbf{g}(t)\ ^2$	$1$
$\mathbf{y}'(t) = \eta(t) \mathbf{y}(t) + \tau(t) \mathbf{g}(t)$	$2r$
$\mathbf{h}'(t) = \mathbf{Z}(t-1)^H \mathbf{y}'(t)$	$r^2$
$\epsilon(t) = \frac{\tau(t)}{\eta(t)} (\mathbf{Z}(t-1) \mathbf{g}(t) - (\mathbf{h}'(t)^H \mathbf{g}(t)) \mathbf{g}(t))$	$r^2 + 3r$
$\mathbf{Z}(t) = \frac{1}{\beta} (\mathbf{Z}(t-1) - \mathbf{g}(t) \mathbf{h}'(t)^H + \epsilon(t) \mathbf{g}(t)^H)$	$2r^2$
$\mathbf{e}(t) = \eta(t) \mathbf{x}(t) - \mathbf{W}(t-1) \mathbf{y}'(t)$	$nr + n$
$\mathbf{W}(t) = \mathbf{W}(t-1) + \mathbf{e}(t) \mathbf{g}(t)^H$	$nr$

## FAPI (Badeau, 2004)

$$\mathbf{y}(t) = \mathbf{W}(t-1)^H \mathbf{x}(t)$$

⋮

$$\mathbf{W}(t) = \mathbf{W}(t-1) + \mathbf{e}(t) \mathbf{g}(t)^H$$

## Applications de la mise à jour de $\mathbf{W}$

### Décomposition harmoniques/bruit

- projecteurs :  
sur l'espace signal  $\mathcal{P}_s = \mathbf{W}\mathbf{W}^H$ ,  
sur l'espace bruit  $\mathcal{P}_b = \mathbf{W}_\perp \mathbf{W}_\perp^H$
- permet de séparer sans calcul explicite des amplitudes

### Application : sons musicaux

- piano : piano, piano bruit. piano bruit. ech., piano vibrato  
saxo, saxo harmo. saxo bruit., saxo bruit. ech.

## Application : remix de la partie bruit

Démo du logiciel minimix de O. Gillet

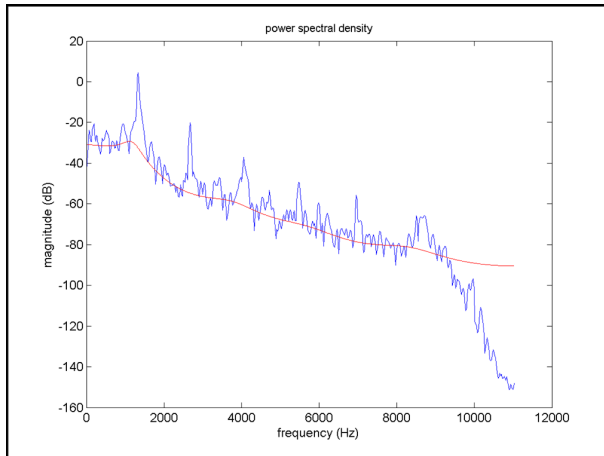
## Applications à la détection du rythme

- jazz, Brad Mehldau
- Salsa, Maraca
- Country, Lyle Lovet



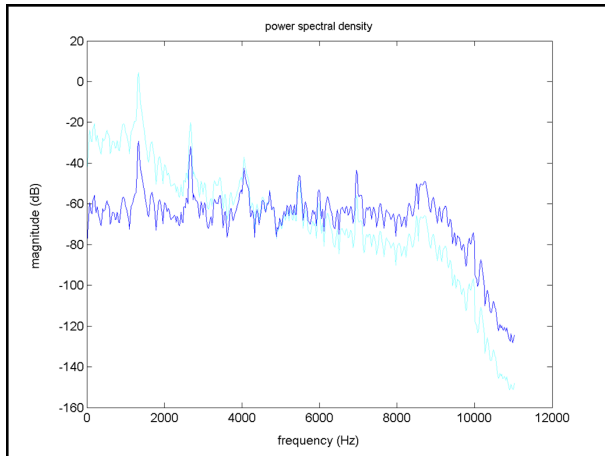
# Prétraitement

## Filtrage médian + modélisation AR-12

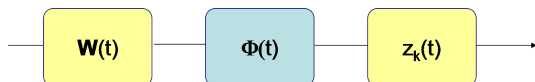


# Prétraitement

## Filtrage médian + modélisation AR-12



## Mise à jour de la matrice spectrale

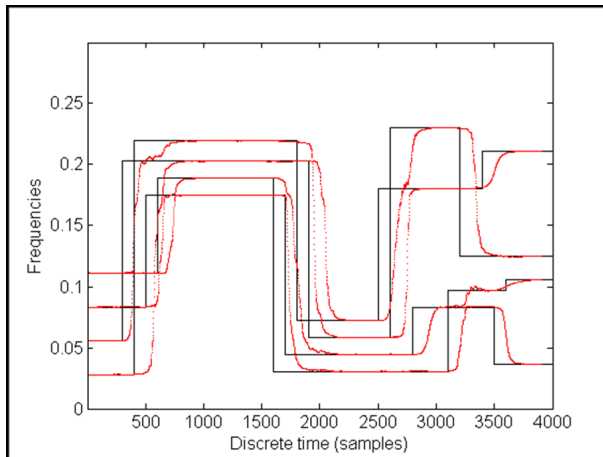


### Décomposition et mise à jour

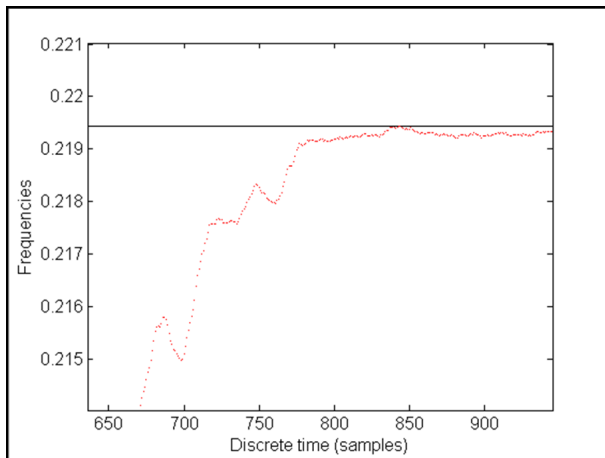
- Décomposition  $\Phi(t) = \underbrace{(\mathbf{W}_{\downarrow}(t)^H \mathbf{W}_{\downarrow}(t))^{-1}}_{\Omega(t)} \underbrace{\mathbf{W}_{\downarrow}(t)^H \mathbf{W}_{\uparrow}(t)}_{\Psi(t)}$
- Récursion sur  $\mathbf{W} \leftrightarrow$  récursion sur  $\Psi \leftrightarrow$  récursion sur  $\Phi$  :  

$$\Phi(t) = \Psi(t) + \frac{1}{1 - \|\nu(t)\|^2} \nu(t) \varphi(t)^H$$
- coût :  $2nr$  (pour  $\Psi$ ) et  $r^2$  pour le produit  $\nu(t) \varphi(t)^H$ .

## Résultats de simulation

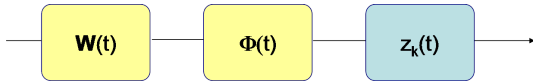


## Résultats de simulation



# Lignes spectrale adaptative

## Suivi des pôles



### Gradient adaptatif pour $\Phi(t)\mathbf{V}(t) = \mathbf{V}(t)\mathbf{\Lambda}(t)$

- mise à jour séquentielle de
  1. la matrice des valeurs propres :  $\mathbf{\Lambda}(t) = \text{diag}\{z_k(t)\}$
  2. la matrice des vecteurs propres  $\mathbf{V}(t)$

- fonctions de coût

$$J(\mathbf{\Lambda}) = \text{tr}\{\mathbf{E}_L(\mathbf{\Lambda})^H \mathbf{E}_L(\mathbf{\Lambda})\}$$

$$J(\mathbf{V}) = \text{tr}\{\mathbf{E}_V(\mathbf{V})^H \mathbf{E}_V(\mathbf{V})\}$$

## Calcul du gradient

### Pour les valeurs propres

- Matrice d'erreur :  $\mathbf{E}_L(\boldsymbol{\Lambda}) = \boldsymbol{\Lambda} - \text{diag}(\mathbf{V}(t-1)^{-1}\boldsymbol{\Phi}(t)\mathbf{V}(t-1))$
- Calcul du gradient de  $J(\boldsymbol{\Lambda})$  :

$$\boldsymbol{\Lambda}(t) = (1 - \mu_L)\boldsymbol{\Lambda}(t-1) + \mu_L \text{diag}(\mathbf{V}(t-1)^{-1}\boldsymbol{\Phi}(t)\mathbf{V}(t-1))$$

### Pour les vecteurs propres

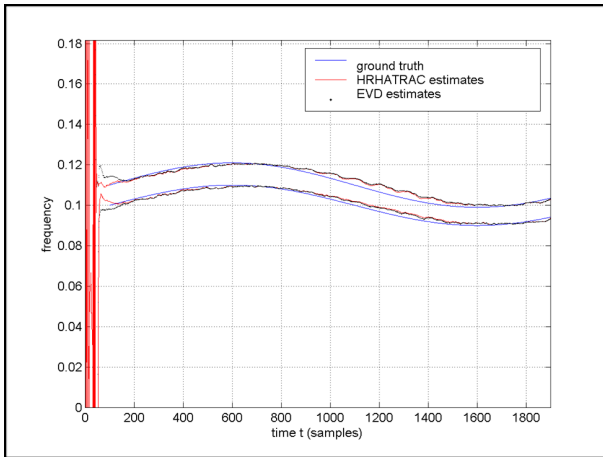
- Matrice d'erreur :  $\mathbf{E}_V(\mathbf{V}) = \mathbf{V} - \boldsymbol{\Phi}(t)\mathbf{V}\boldsymbol{\Lambda}(t)^{-1}$
- calcul du gradient :

$$\begin{aligned} \mathbf{V}(t) = & (1 - \mu_V)\mathbf{V}(t-1) + \mu_V ( \boldsymbol{\Phi}(t)\mathbf{V}(t-1) \\ & \boldsymbol{\Lambda}(t)^{-1} \\ & + \boldsymbol{\Phi}(t)^H \mathbf{E}_V(\mathbf{V}(t-1))\boldsymbol{\Lambda}(t)^{-H} ) \end{aligned}$$



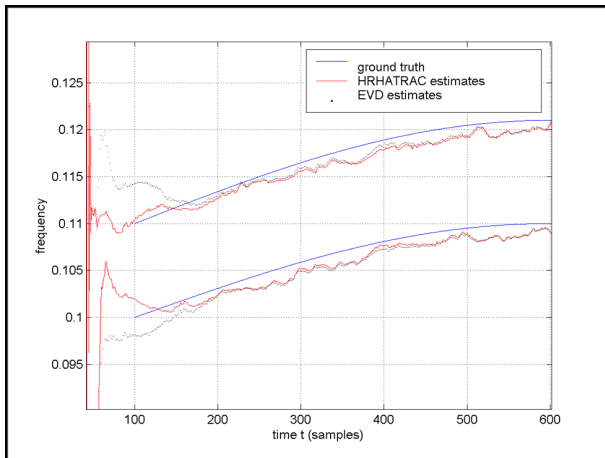
## Résultats de simulation sur 2 sinus modulés

$\nu_1 = 0.1, \nu_2 = 0.11$ , Trans. de Hilbert,  $n=31, r=2$ .

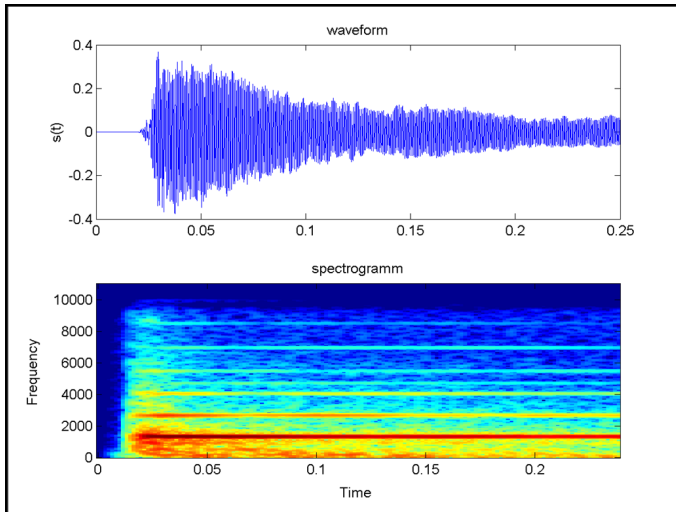


## Résultats de simulation sur 2 sinus modulés

$\nu_1 = 0.1$ ,  $\nu_2 = 0.11$ , Trans. de Hilbert,  $n=31$ ,  $r=2$ .

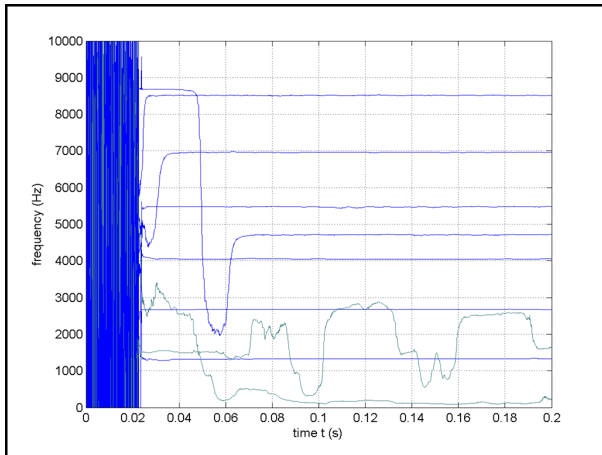


## Application à une note de piano



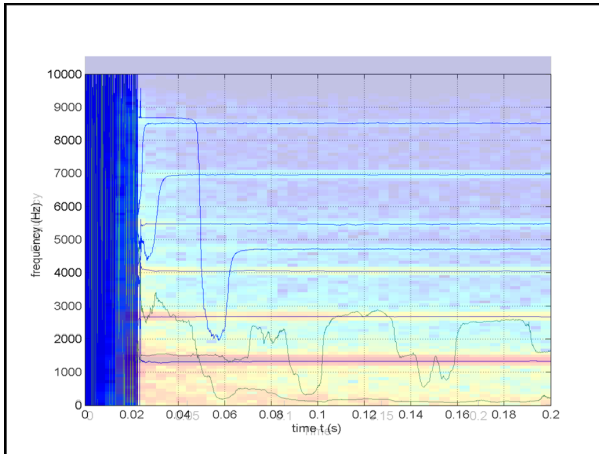
## Résultats du suivi de lignes

$n=101$  (5ms),  $r=9$

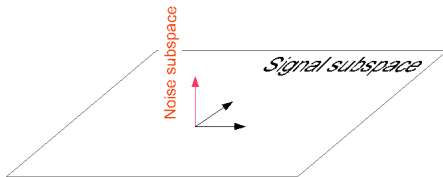


## Résultats du suivi de lignes

$n=101$  (5ms),  $r=9$



## Représentation temps-fréquence adaptative



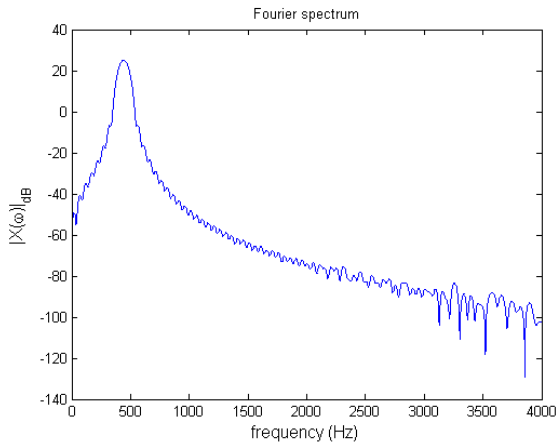
### Principles

- low rank approximation of signals (sines + noise)
- (projection onto noise subspace)<sup>-1</sup> : MUSIC pseudo-spectrum

## Illustration

### A two components example

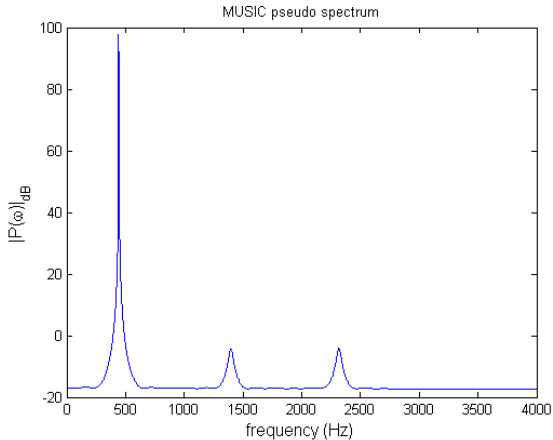
- 440 Hz, 445Hz
- 20 ms, sampling at 8kHz



## Illustration

### A two components example

- 440 Hz, 445Hz
- 20 ms, sampling at 8kHz

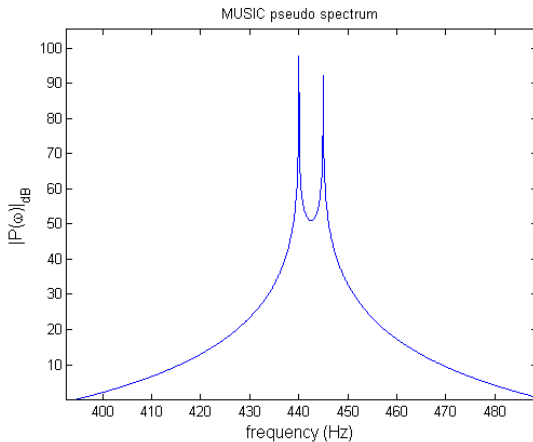




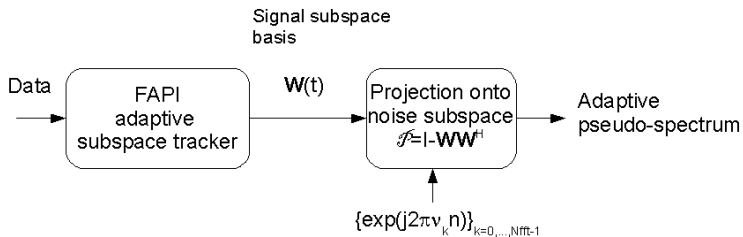
## Illustration

### A two components example

- 440 Hz, 445Hz
- 20 ms, sampling at 8kHz

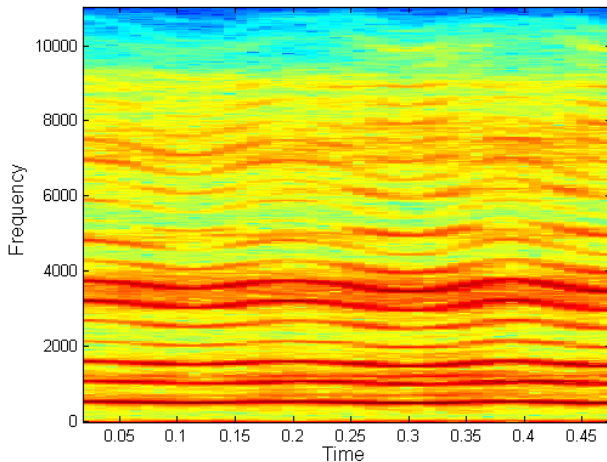


## Adaptive MUSIC : block Diagram



## Enhancing the readability of tf representation : singing voice

voice



## Enhancing the readability of tf representation : singing voice

voice

Adaptive pseudo-spectrum

