Master « Atiam » - Module ETS

Estimation de fréquences fondamentales multiples

Gaël RICHARD
TELECOM ParisTech
Département Traitement des signaux et des images
Janvier 2012

Merci à Roland Badeau pour un certain nombre de transparents

« Licence de droits d'usage"
http://formation.enst.fr/licences/pedago_sans.html

Détection de fréquence(s) fondamentale(s)

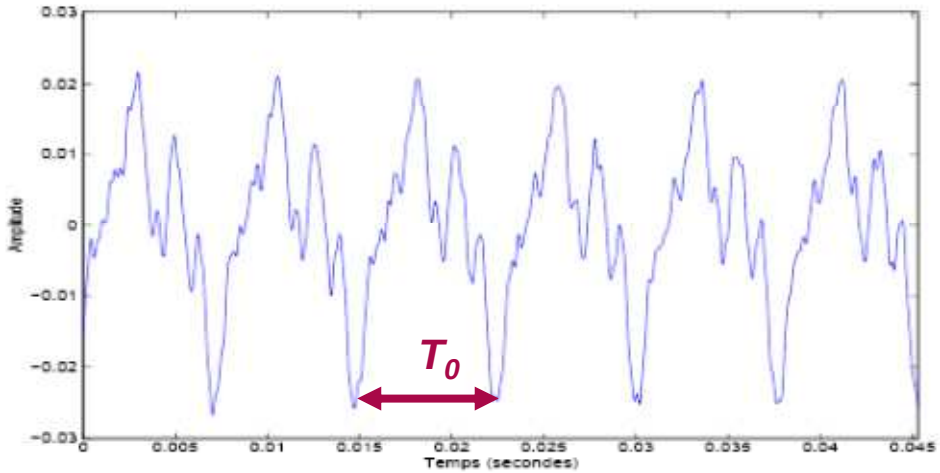
Droits d'usage autorisé

Contenu

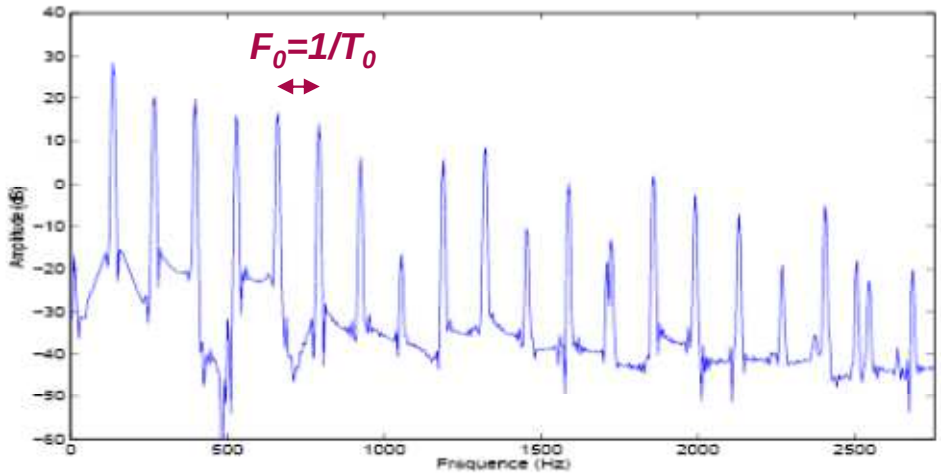
- **Introduction**
  - Sons quasi-périodiques
  - Modèle de son quasi-périodique
- **Méthodes temporelles**
- **Méthodes spectrales**
- **Extension à la détection de fréquences fondamentales multiples**

Droits d'usage autorisé

Un son quasi-périodique



Son de piano (C3)



Spectre du son de piano

Droits d'usage autorisé

Modèle de signal

$$x(n) = \sum_{k=1}^{H} 2A_k \cos(2\pi k f_0 n + \phi_k) + w(n)$$

- $f_0 = \frac{1}{T_0}$ est la fréquence fondamentale réduite
- H est le nombre d'harmoniques du signal
- Les amplitudes $\{A_k\}$ sont des réels > 0
- Les phases $\{\phi_k\}$ sont des v.a. indépendantes de loi uniforme sur [0, 2π [
- *w* est un bruit blanc centré de variance σ^2, indépendant des phases $\{\phi_k\}$
- x(n) est un processeur SSL centré d'autocovariance

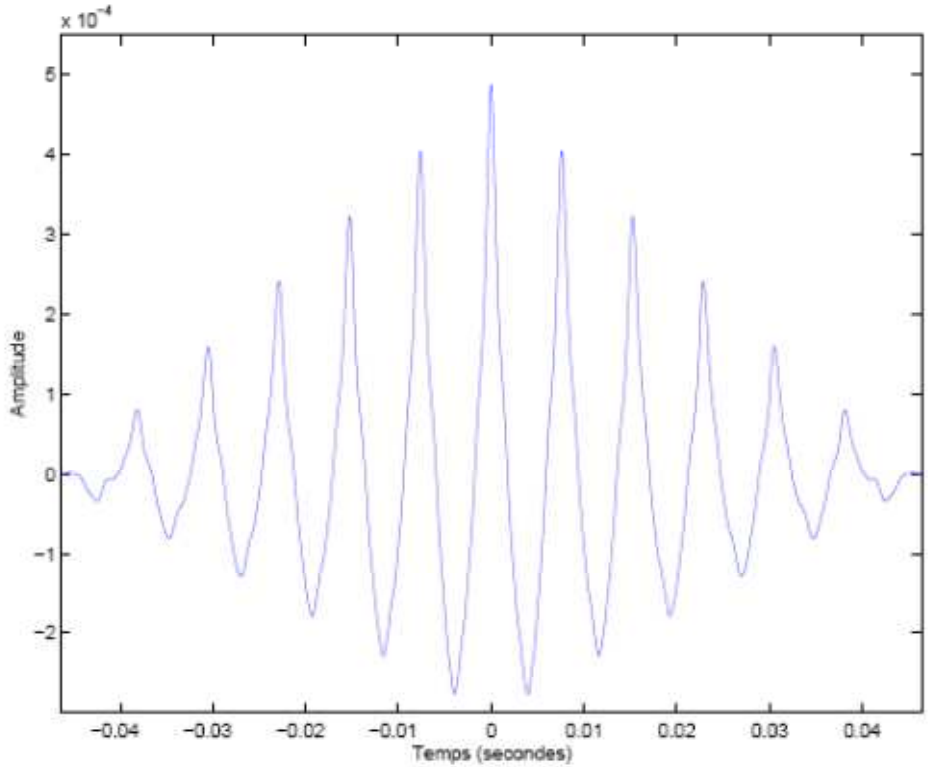
$$r_x(m) = \sum_{k=1}^{H} [2A_k^2 \cos(2\pi k f_0 m)] + \sigma^2 \delta[m]$$

Droits d'usage autorisé

Méthodes temporelles

- **Autocovariance biaisée** $\frac{1}{N}\sum_{n=0}^{N-1-m} x[n]\,x[n+m]$ si $m \geq 0$

$$\mathbf{E}(\hat{r}_x[m]) = \frac{N-|m|}{N}\, r_x[m] \qquad\qquad |\hat{r}_x[m]| \leq \hat{r}_x[0]$$



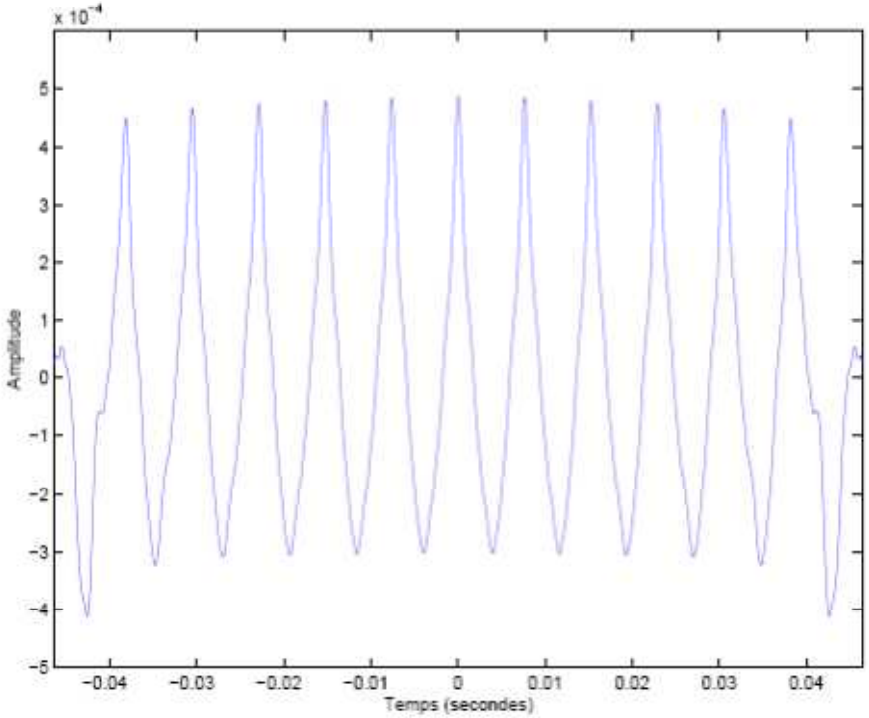
Droits d'usage autorisé

Méthodes temporelles

- **Autocovariance non biaisée**

$$\tilde{r}_x[m] = \frac{1}{N-m} \sum_{n=0}^{N-1-m} x[n]\, x[n+m] \text{ si } m \geq 0$$

$$\mathbf{E}(\tilde{r}_x[m]) = r_x[m] \qquad \mathrm{Var}(\tilde{r}_x[m]) = (\frac{N}{N-m})^2 \mathrm{Var}(\hat{r}_x[m])$$



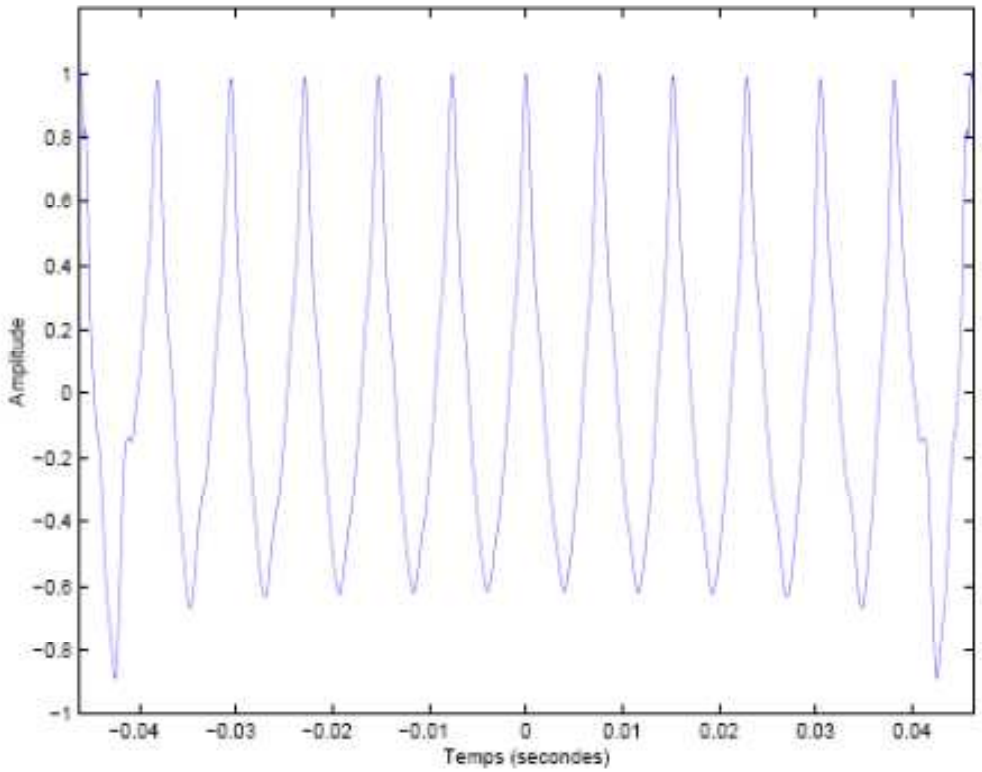
$$|\tilde{r}_x[m]| \not\leq \tilde{r}_x[0]$$

Droits d'usage autorisé

Méthodes temporelles

- **Autocorrélation** $\bar{r}_x[m] = \dfrac{\sum_{n=0}^{N-1-m} x[n]\, x[n+m]}{\sqrt{\sum_{n=0}^{N-1-m} x[n]^2}\sqrt{\sum_{n=0}^{N-1-m} x[n+m]^2}}$ si $m \geq 0$

$$|\bar{r}_x[m]| \leq \bar{r}_x[0] = 1 \qquad |\bar{r}_x[m]| = 1 \text{ ssi les vecteurs sont colinaires}$$



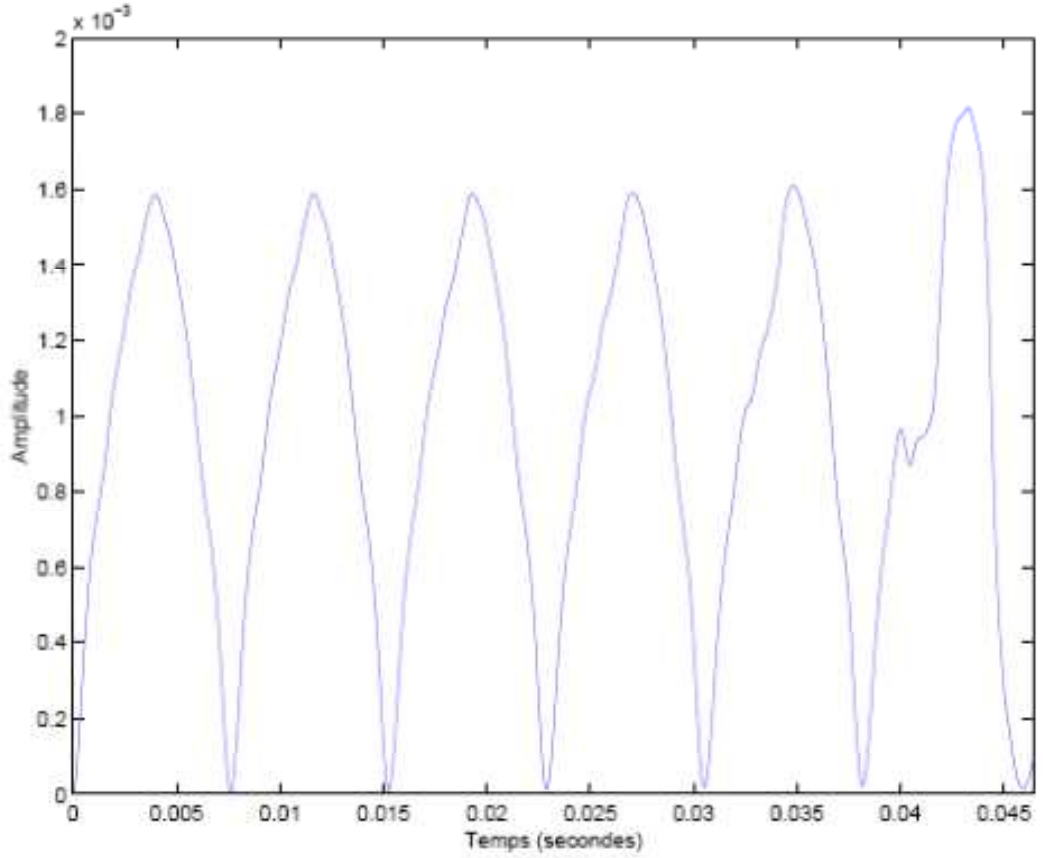
Droits d'usage autorisé

TELECOM ParisTech

Average square difference function (ASDF)

$$\mathrm{ASDF}[m] = \frac{1}{N-m} \sum_{n=0}^{N-1-m} (x[n] - x[n+m])^2$$

$$\mathrm{ASDF}[m] = 0 \text{ ssi } x \text{ est de période } T_0 = m$$



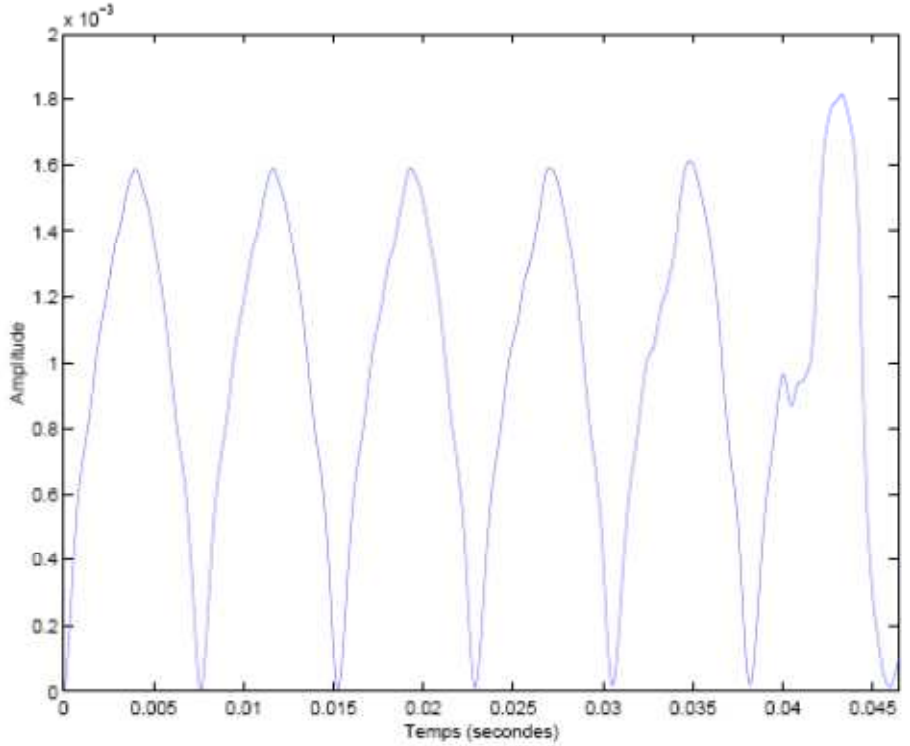
Droits d'usage autorisé

TELECOM ParisTech

Average square difference function (ASDF)

- La période T_0 peut être estimée en recherchant le minimum de l'écart quadratique entre les signaux x[n] et x[n+m] :

$$\mathbf{E}[\mathrm{ASDF}[m]] = 2(r_x[0] - r_x[m])$$

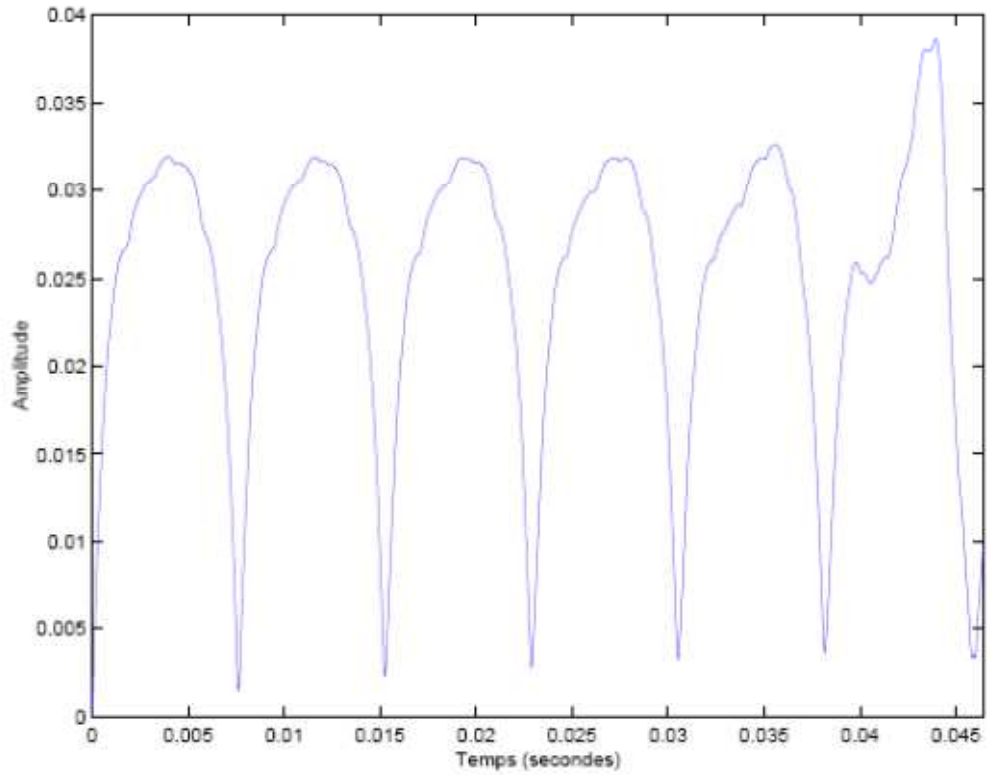


Droits d'usage autorisé

Average magnitude difference function (AMDF)

$$\mathrm{AMDF}[m] = \frac{1}{N-m} \sum_{n=0}^{N-1-m} |x[n] - x[n+m]|$$

$$\mathrm{AMDF}[m] = 0 \text{ ssi } x \text{ est de période } T_0 = m$$



Droits d'usage autorisé

TELECOM ParisTech

Un algorithme temporel performant: Yin

(merci à V. Emiya pour quelques transparents)

- **H. Kawahara A. de Cheveigné, *YIN, a fundamental frequency estimator for speech and music*,, JASA, 111(4), 2002**

- **Point de départ: Méthode de l'Autocorrélation (ACF)**
- **Améliorations successives:**
  - Utilisation de l'ASDF
  - Normalisation
  - Seuillage
  - Interpolation
  - Minimisation locale en temps

Version	Gross error (%)
Step 1	10.0
Step 2	1.95
Step 3	1.69
Step 4	0.78
Step 5	0.77
Step 6	0.50

Droits d'usage autorisé

YIN (2)

- **ASDF utilisée:**

$$\mathrm{d}_n[m] = \sum_{k=0}^{N-1}(x_n[k] - x_n[k+m])^2$$

- **Liens avec l'Autocorrélation**

$$\mathrm{d}_n[m] = r_n(0) + r_{n+m}(0) - 2r_n(m)$$

- **Gain net car l'ASDF est beaucoup moins sensible aux variations des amplitudes relatives que l'ACF (qui est sensible, par exemple, à l'accentuation des partiels d'ordre pair)**

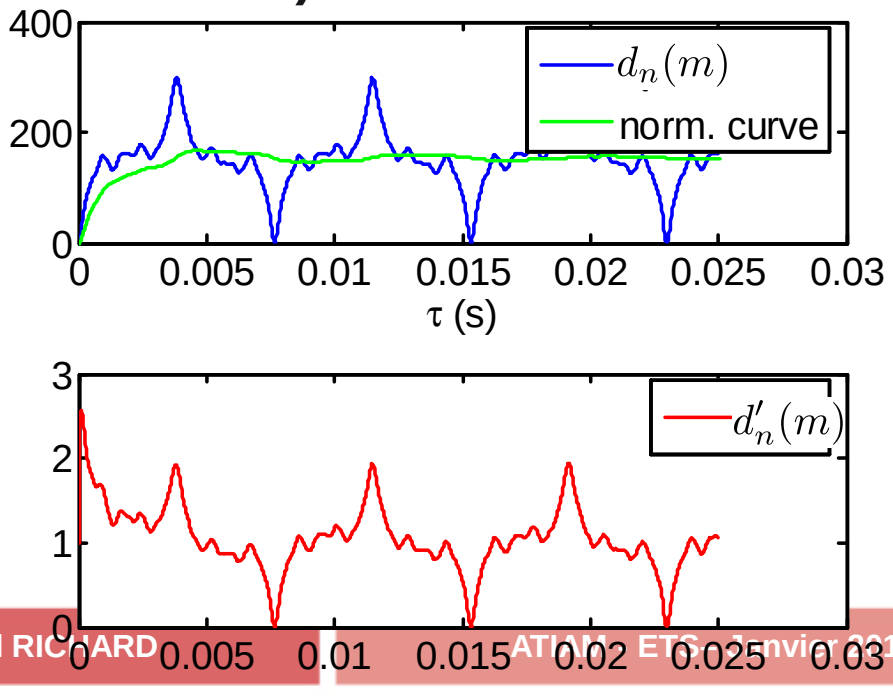
Droits d'usage autorisé

YIN (3)

- **Normalisation par la « moyenne cumulée »**

$$d'_n(m)(= \begin{cases} 1 & \text{si} \quad m = 0 \\ \dfrac{d_n(m)}{\frac{1}{m}\sum_{k=1}^{m} d_n(k)} & \text{sinon} \end{cases}$$

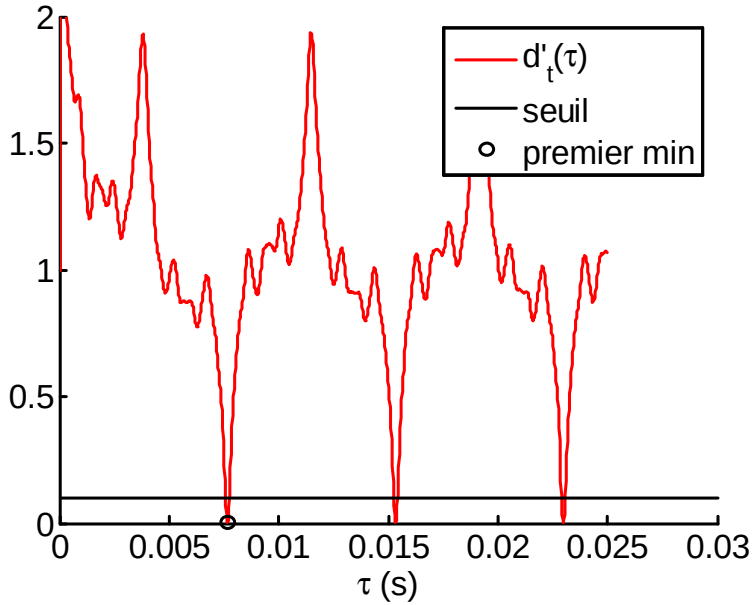
- **Gain net car permet d'éviter les erreurs pour les F0 élevées (suppression du lobe en 0)**



Droits d'usage autorisé
TELECOM ParisTech

YIN (4)

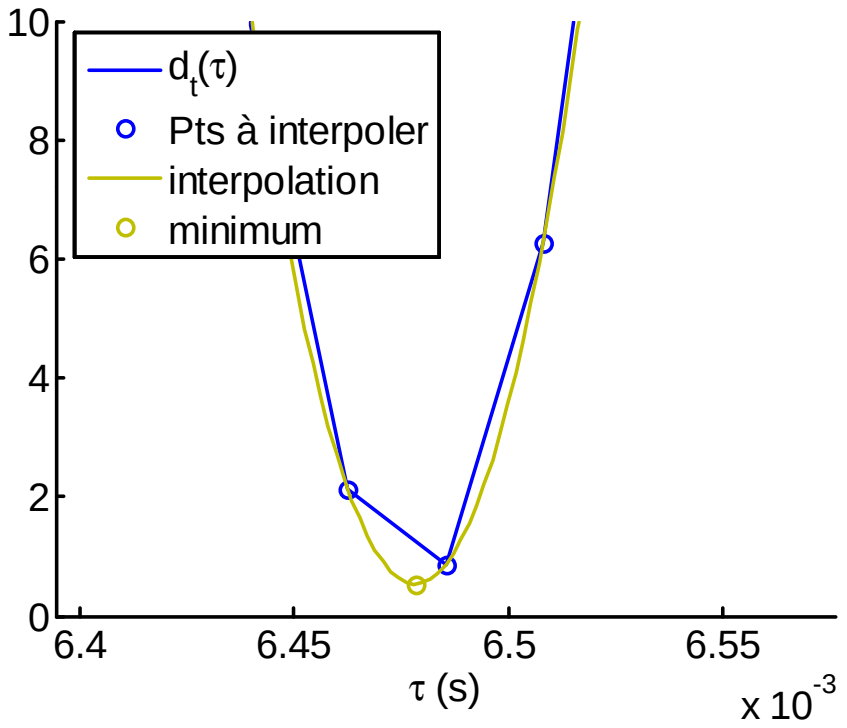
- **Seuillage absolu**
  - La plus petite période inférieure au seuil est choisie
  - Si aucune période n'est inférieure au seuil, alors le minimum global est choisi



Droits d'usage autorisé

YIN (4)

- **Interpolation parabolique autour du minimum**
  - ⇨ Réalisée sur $d_n(m)$ (i.e avant normalisation)
  - ⇨ Gain en précision sur la valeur de F0



Droits d'usage autorisé

YIN (5)

- **Minimisation locale en temps**
  - Période estimée: $T_n = argmin_n(d'_n(m))$
  - Minimisation autour du temps T_θ: $argmin_\theta(d'_\theta(T_\theta))$ avec

$$t - T_{max} < \theta < t + T_{max}, \qquad T_{max} = 25ms$$
$$0.8T_n < T_\theta < 1.2T_n$$

- **Gain en cas de fluctuations sur certains signaux; correspond à un effet de lissage (rappel l'effet du filtre médian ou programmation dynamique).**

Droits d'usage autorisé

YIN: Evaluation

- Sur quatre bases de données de parole, annotées automatiquement (par YIN, à partir du laryngographe) puis vérifiées et triées à la main

	Gross error (%)					
Method	DB1	DB2	DB3	DB4	Average	(low/high)
pda	10.3	19.0	17.3	27.0	**16.8**	(14.2/2.6)
fxac	13.3	16.8	17.1	16.3	**15.2**	(14.2/1.0)
fxcep	4.6	15.8	5.4	6.8	**6.0**	(5.0/1.0)
ac	2.7	9.2	3.0	10.3	**5.1**	(4.1/1.0)
cc	3.4	6.8	2.9	7.5	**4.5**	(3.4/1.1)
shs	7.8	12.8	8.2	10.2	**8.7**	(8.6/0.18)
acf	0.45	1.9	7.1	11.7	**5.0**	(0.23/4.8)
nacf	0.43	1.7	6.7	11.4	**4.8**	(0.16/4.7)
additive	2.4	3.6	3.9	3.4	**3.1**	(2.5/0.55)
TEMPO	1.0	3.2	8.7	2.6	**3.4**	(0.53/2.9)
YIN	0.30	1.4	2.0	1.3	**1.03**	(0.37/0.66)

Droits d'usage autorisé

Approche par le maximum de vraisemblance

- Modèle de signal: $x(n) = a(n) + w(n)$
  - *a* est un signal déterministe de période T_0
  - *w* est un bruit blanc gaussien de variance σ^2

- Vraisemblance des observations

$$p(x|T_0, a, \sigma^2)) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-a(n))^2}$$

- Log-vraisemblance

$$L(T_0, a, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n) - a(n))^2$$

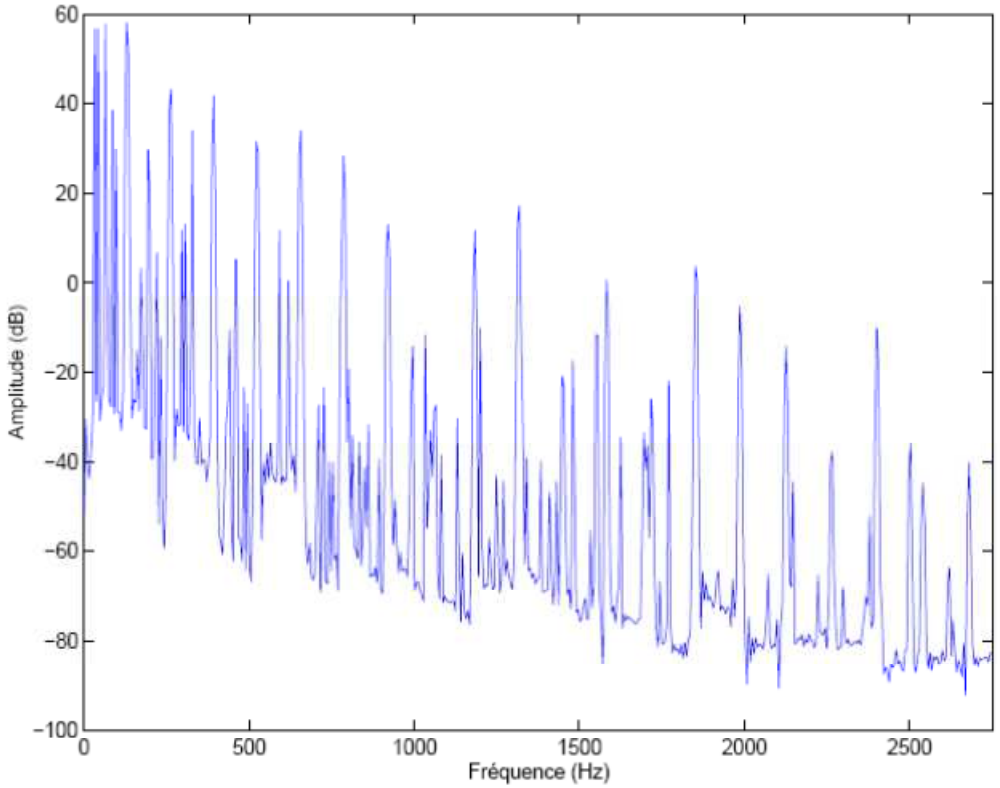
- Méthode: maximiser successivement L par rapport à a, puis σ^2 et enfin T_0

Droits d'usage autorisé

Approche par le maximum de vraisemblance

- On peut montrer que la maximisation de L par rapport à $F_0 = \frac{m}{N}$ revient à maximiser la somme spectrale

$$S(e^{j\,2\pi\,\frac{m}{N}}) = \sum_{k=1}^{H} \hat{R}_x(e^{j\,2\pi k\,\frac{m}{N}})$$

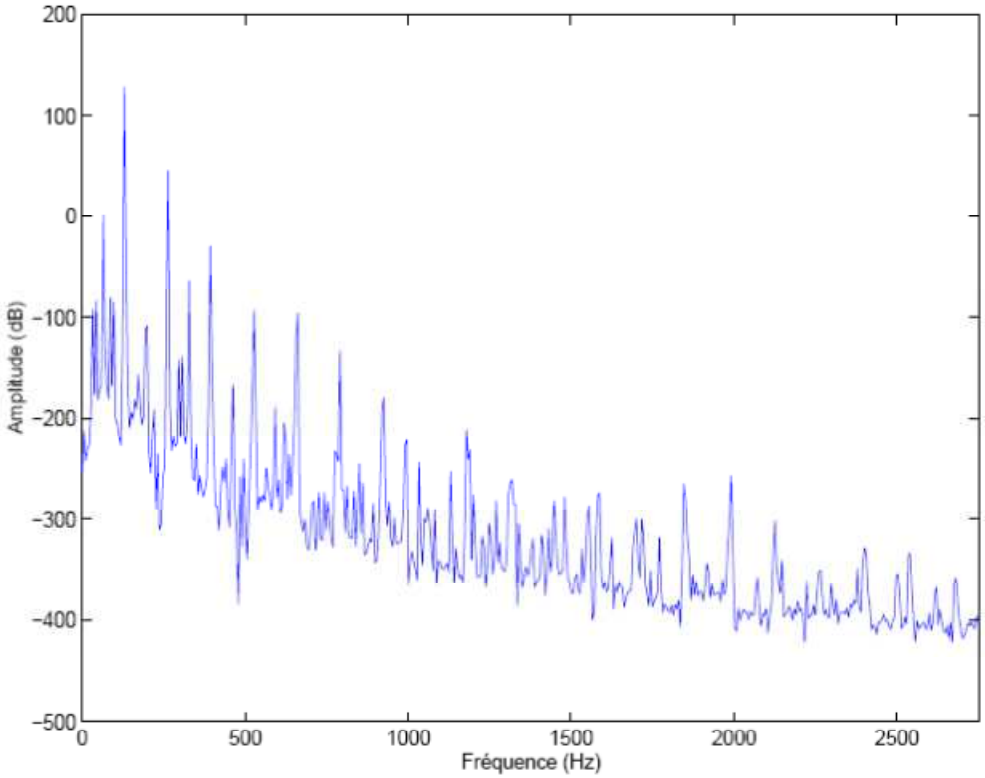


Droits d'usage autorisé

Produit spectral

- Par similitude avec la somme spectrale on peut définir le produit spectral (souvent plus robuste)

$$P(e^{j\,2\pi\,\frac{m}{N}}) = \prod_{k=1}^{H} \hat{R}_x(e^{j\,2\pi k\,\frac{m}{N}})$$



Droits d'usage autorisé

TELECOM ParisTech

Détection de fréquences fondamentales multiples

Droits d'usage autorisé

Détection de fréquences fondamentales multiples

- **Objectif: extraire l'ensemble des notes d'un enregistrement polyphonique**
- **Problème important lorsque les notes sont en rapport harmonique (ce qui est souvent le cas en musique…!!)**
- **Nécessité de traiter le caractère non parfaitement harmonique des notes jouées par un instrument.**

Droits d'usage autorisé

Détection de fréquences fondamentales multiples

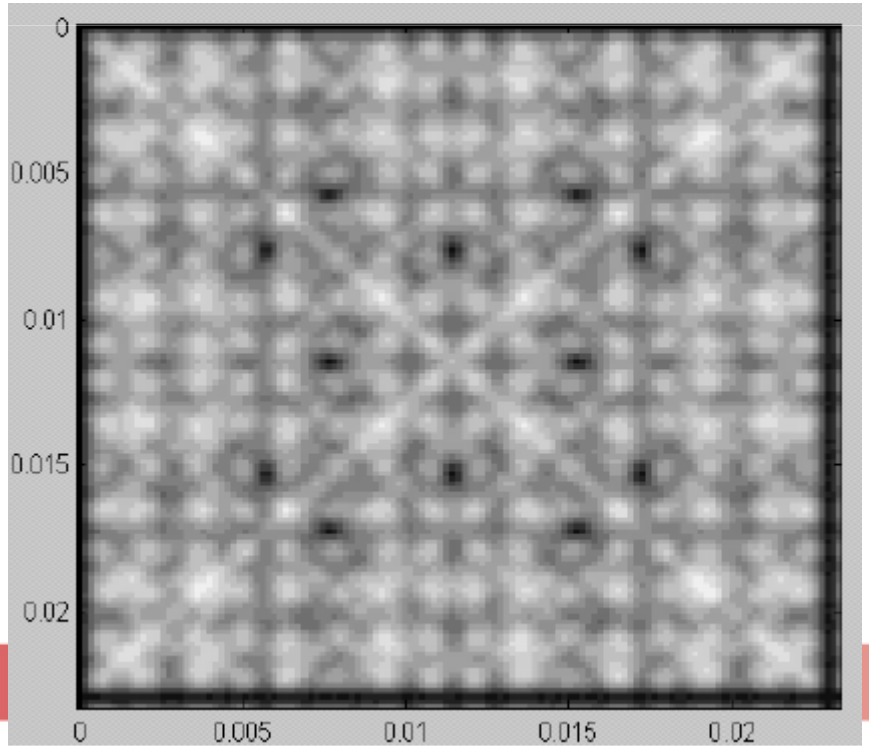
- **Approche par estimation/soustraction conjointe**
  - DMDF (*Double Magnitude Difference Function*)

$$DMDF(k_1, k_2) = \frac{1}{N - k_1 - k_2} \sum_{n=0}^{N-k_1-k_2-1} |d[n] - d[n+k_1] - d[n+k_2] + d[n+k_1+k_2]|$$

✓ **Son de piano** addition de deux notes:

T1=0.0076s

T2=0.0057s



Droits d'usage autorisé

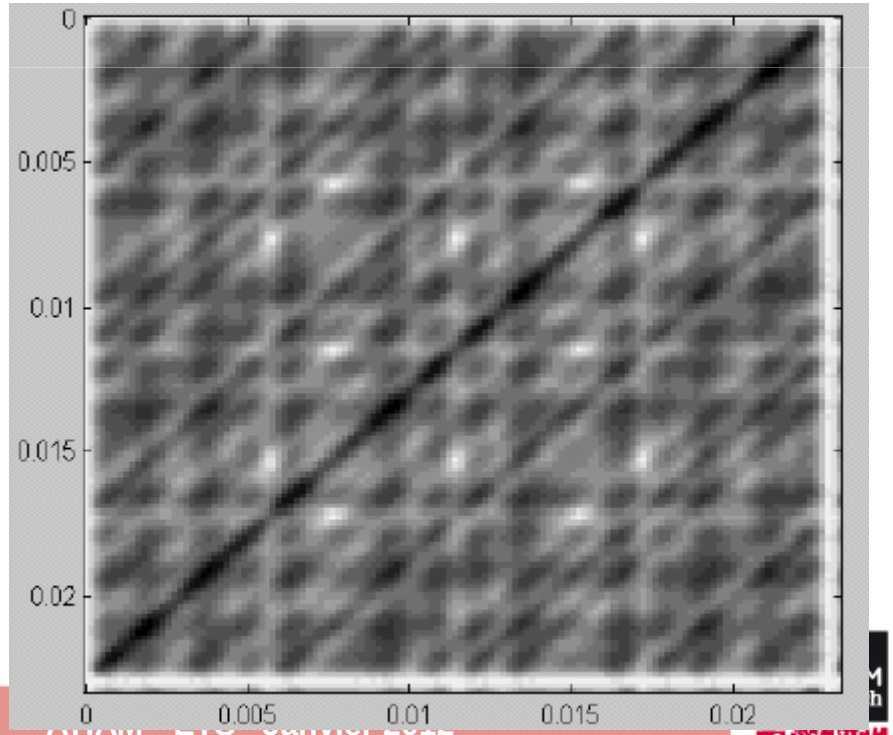
Détection de fréquences fondamentales multiples

- **Approche par corrélation bi-dimensionnelle**

$$\overline{r}\left(k_1,k_2\right)=\frac{\sum_{n=0}^{N-k_1-k_2-1} d\left[n\right]\left(d\left[n+k_1\right]+d\left[n+k_2\right]-d\left[n+k_1+k_2\right]\right)}{\left(\sum_{n=0}^{N-k_1-k_2-1} d\left[n\right]^2\right)^{1/2}\left(\sum_{n=0}^{N-k_1-k_2-1}\left(d\left[n+k_1\right]+d\left[n+k_2\right]-d\left[n+k_1+k_2\right]\right)^2\right)^{1/2}}$$

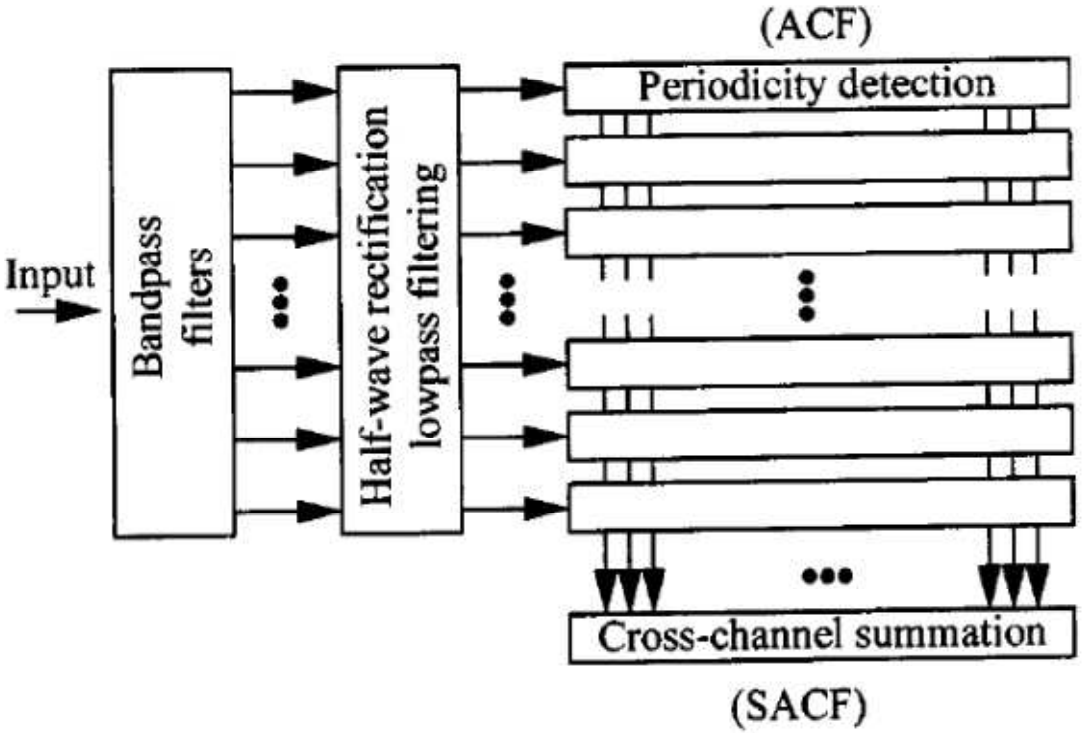
Mesure la « ressemblance » entre

- d(n) et
- d(n+k1) + d(n+k2)-d(n+k1+k2)



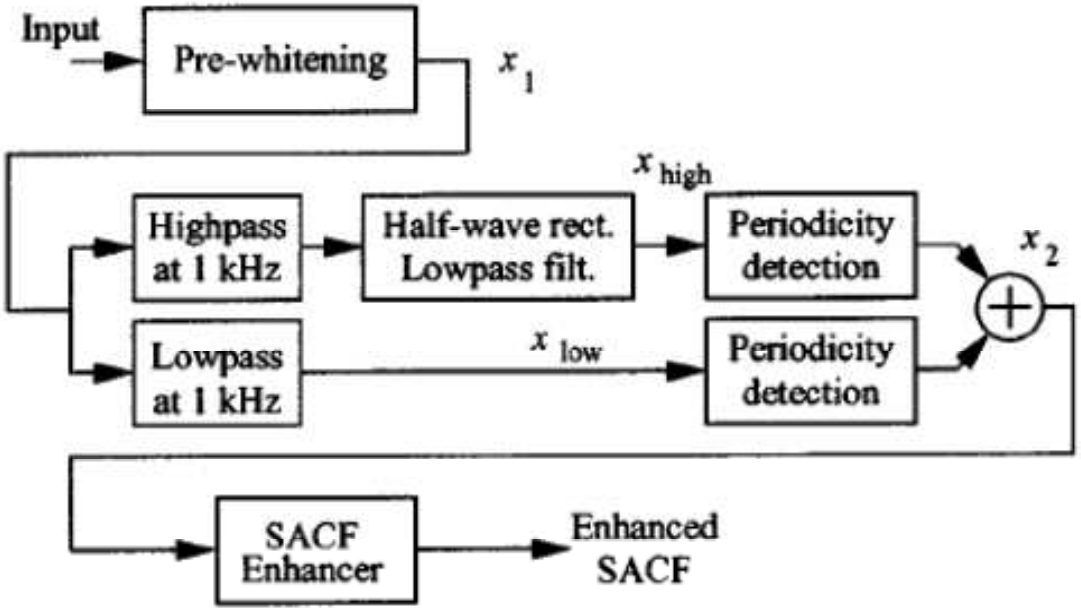
Droits d'usage autorisé

Une approche par banc de filtres



- R. Meddis and M. Hewitt, “Virtual pitch and phase sensitivity of a computer model of the auditory periphery—I: Pitch identification,” *J. Acoust. Soc. Am.*, vol. 89, pp. 2866–2882, June 1991.

Une approche plus simple inspirée de la précédente



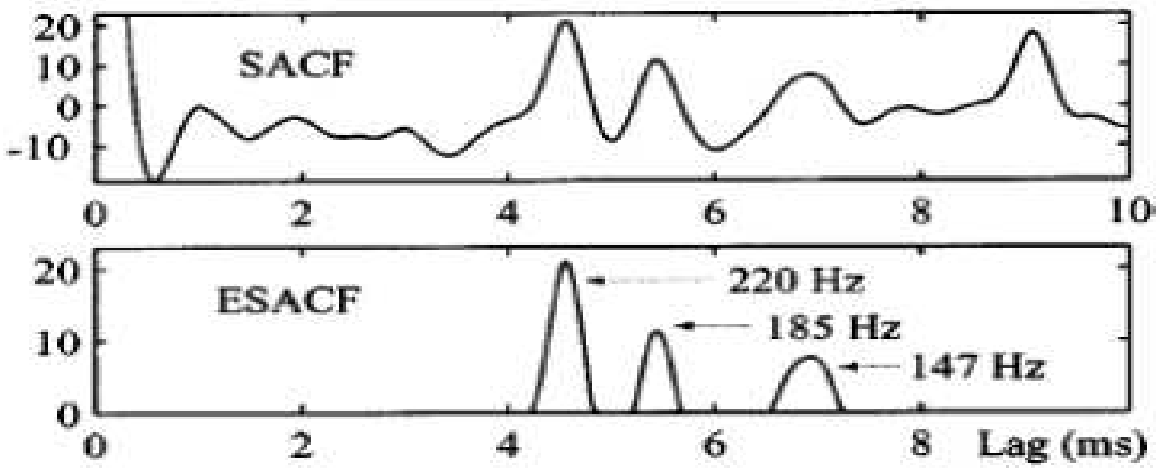
- T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. On Speech and Audio Processing*, vol. 8, no. 6, pp. 708–716, 2000.

Droits d'usage autorisé

Enhanced Summary ACF

- **Plusieurs étapes:**

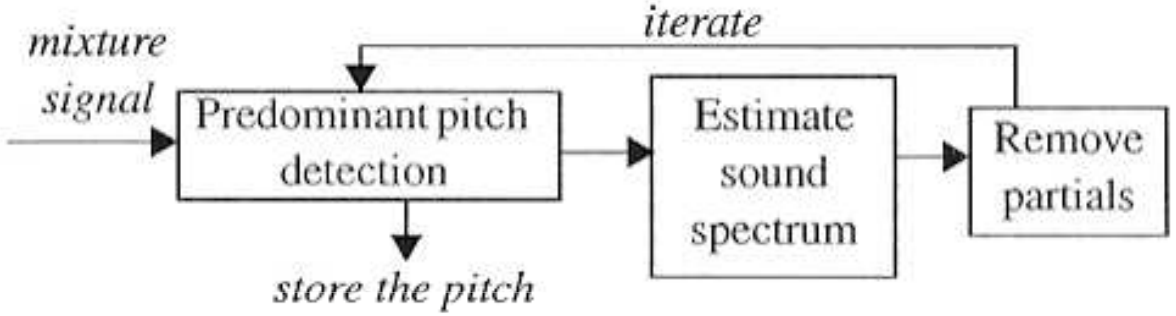
  - Redressement demi-onde
    - On ne conserve que les valeurs positives
  - Ralentie 2 (ou plus) fois puis déduite du SACF redressé
    - Permet de supprimer les pics doubles



Droits d'usage autorisé

Détection de fréquences fondamentales multiples

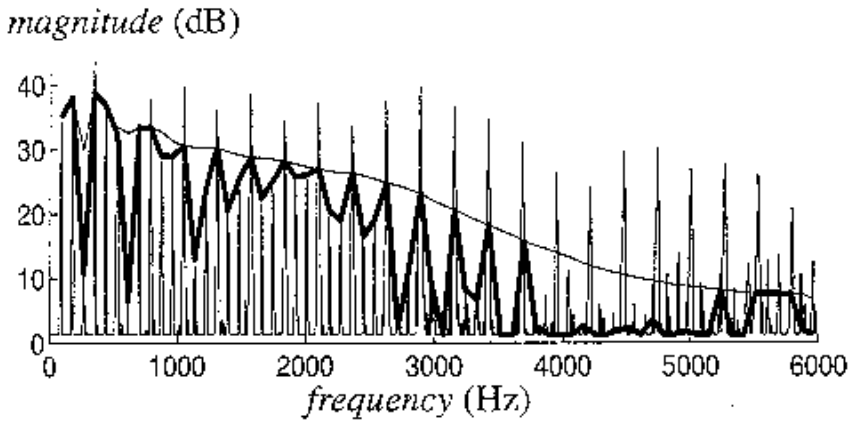
- **Approche par soustraction itérative (Klapuri, 2003)**



Principe de lissage spectral

$a_h = min(a_h, m_h)$

où m_h est la moyenne sur une fenêtre d 'un octave autour du partiel

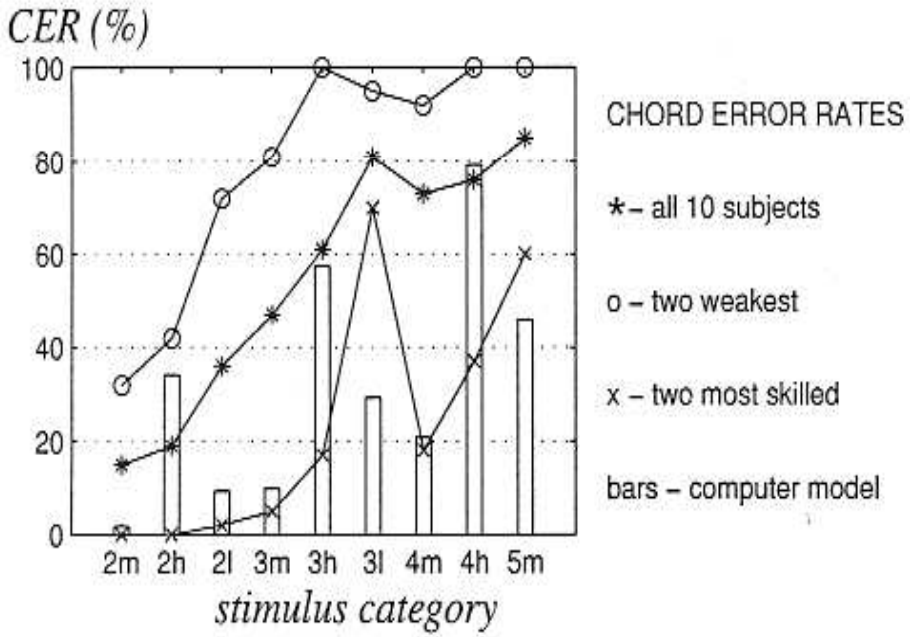


Droits d'usage autorisé

Détection de fréquences fondamentales multiples

Résultats: Comparaison aux performances humaines

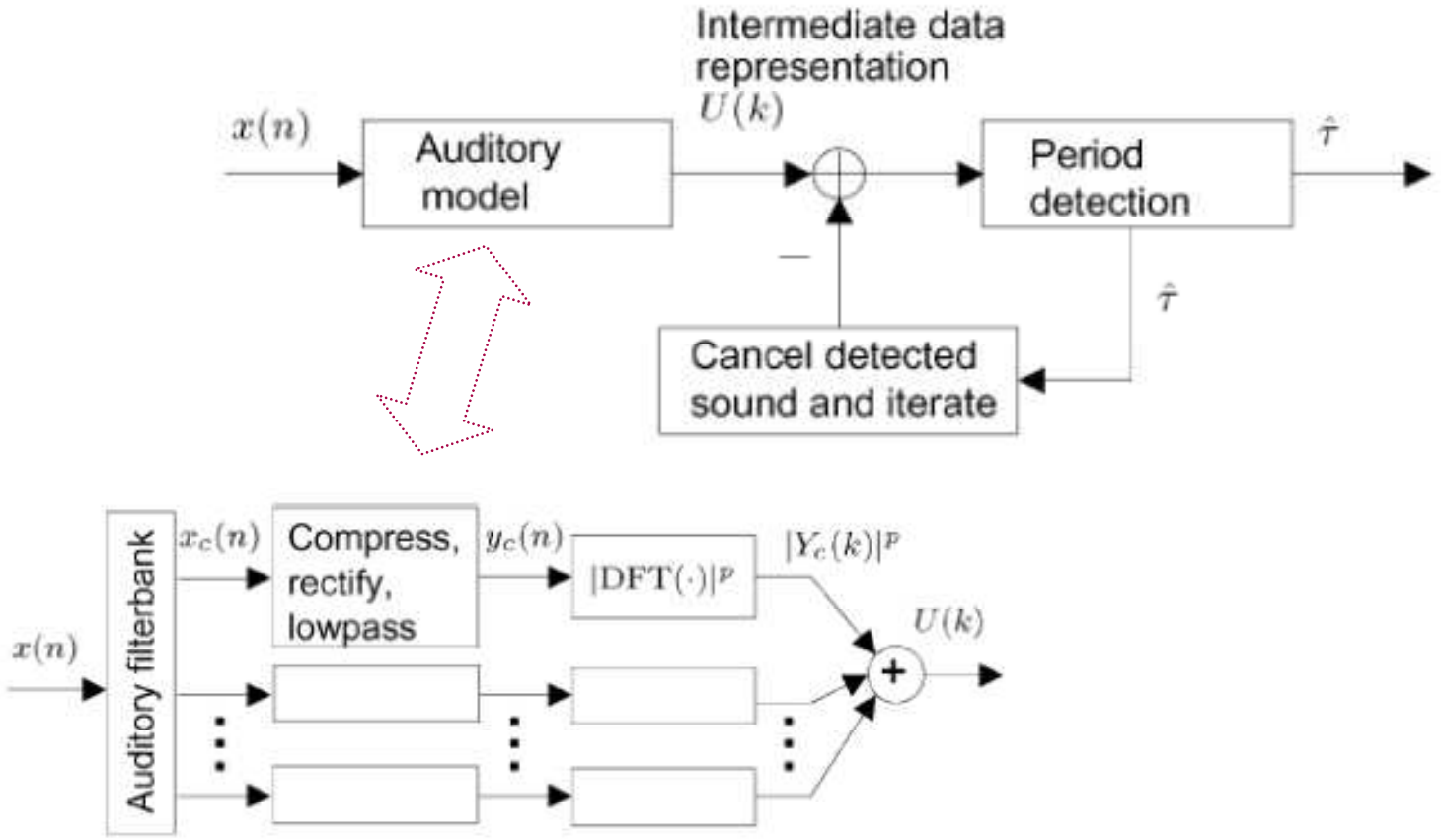
- **Registre bas** (l): 33 à 130 Hz
- **Registre médium** (m): 130 à 520 Hz
- **Registre haut:** 520 à 2100 Hz
- **200 stimuli sonores** (20 catégories)
- Sons polyphoniques générés par ordinateur à partir d 'échantillons de Piano Steinway provenant du *Master samples collection, Mc Gill University*



- Personnes ayant participé aux tests:
  - ⇨ Tous sont musiciens
  - ⇨ dont 2 ont l 'oreille absolue (musiciens quasi-professionnels

Droits d'usage autorisé

Une approche récente utilisant un modèle perceptuel

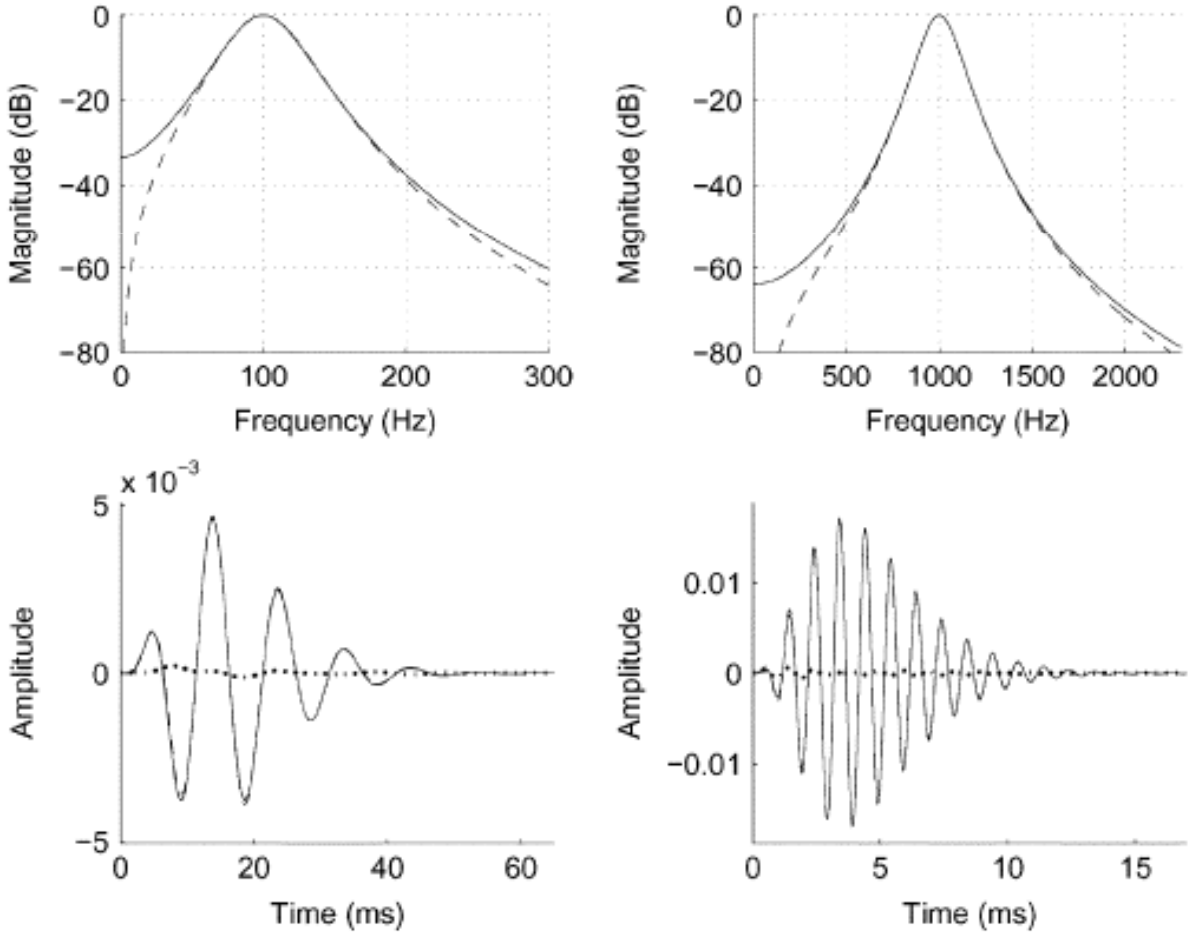


- **Anssi P. Klapuri "Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model", IEEE Trans. On ASLP, Feb. 2008**

Droits d'usage autorisé

Banc de filtres perceptuels

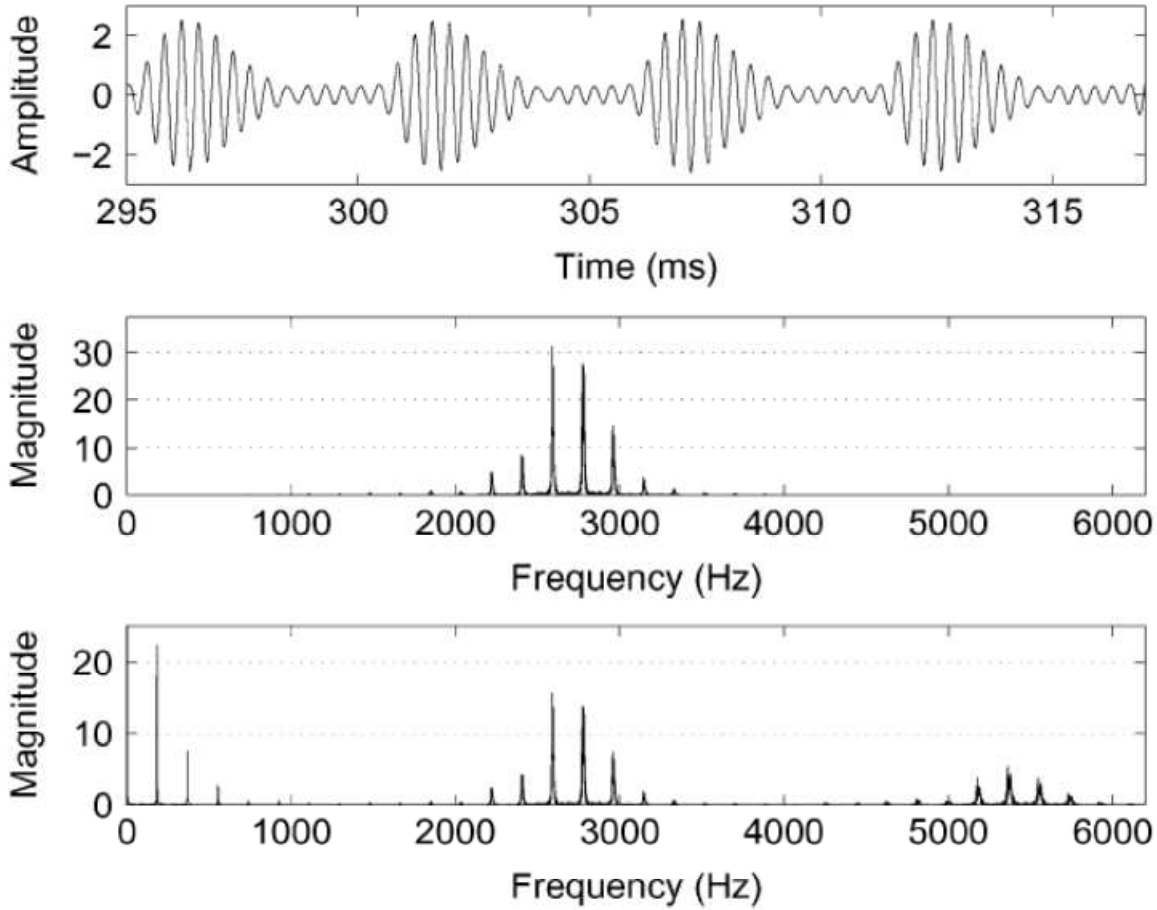
- **Une approximation d'un banc de filtres Gammatone**



Droits d'usage autorisé

Effet de la compression et redressement

- **Résultat sur une bande centrée at 2.7 kHz**



Droits d'usage autorisé

Détection de fréquences fondamentales multiples

- **Autres approches**
  - Approches bayesiennes
  - Méthodes haute-résolution
  - Factorisation en Matrices non-négatives (NMF) ou Analyse en composantes latentes (PLCA – équivalent probabiliste de la NMF)

Droits d'usage autorisé

Factorisation en Matrices Non-négatives

- Utilisation de méthodes de décomposition non supervisées (par exemple par factorisation en matrices non-négatives : NMF)

- **Principe de la NMF :**

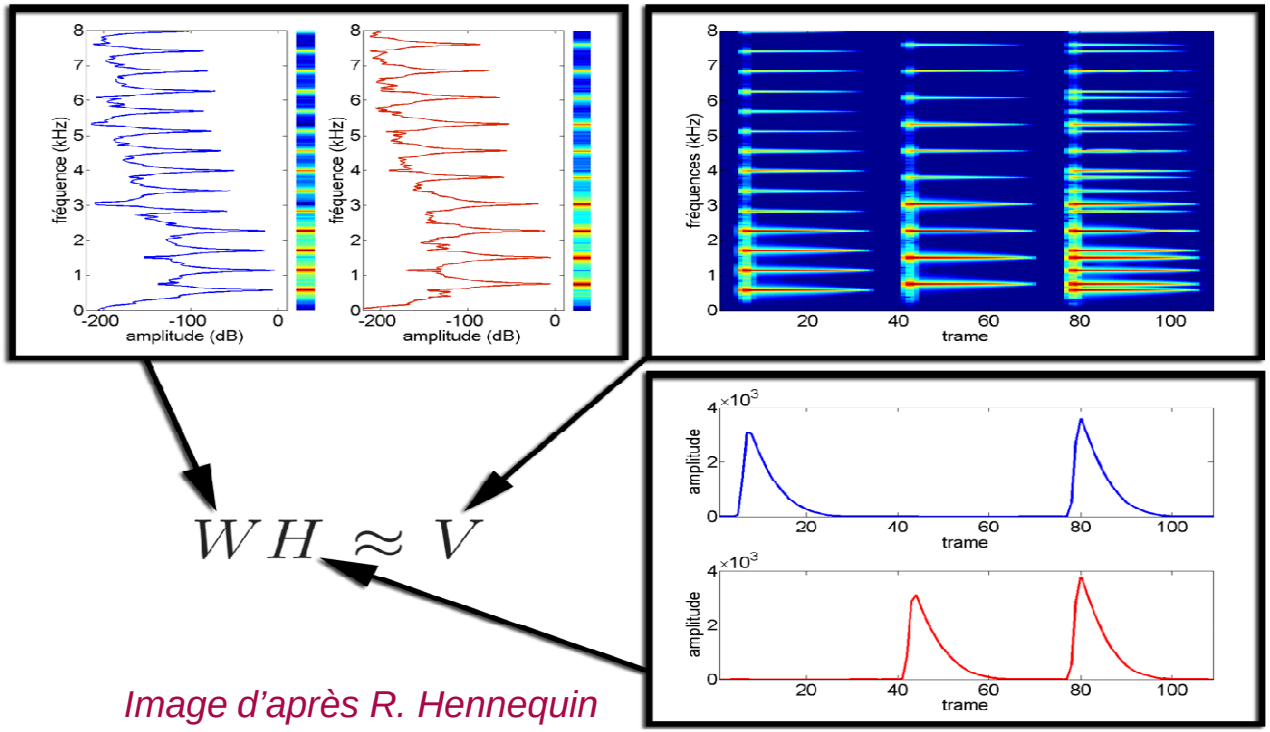
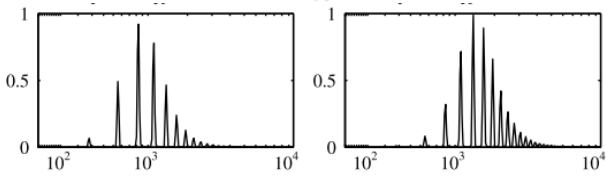


Image d'après R. Hennequin

Droits d'usage autorisé

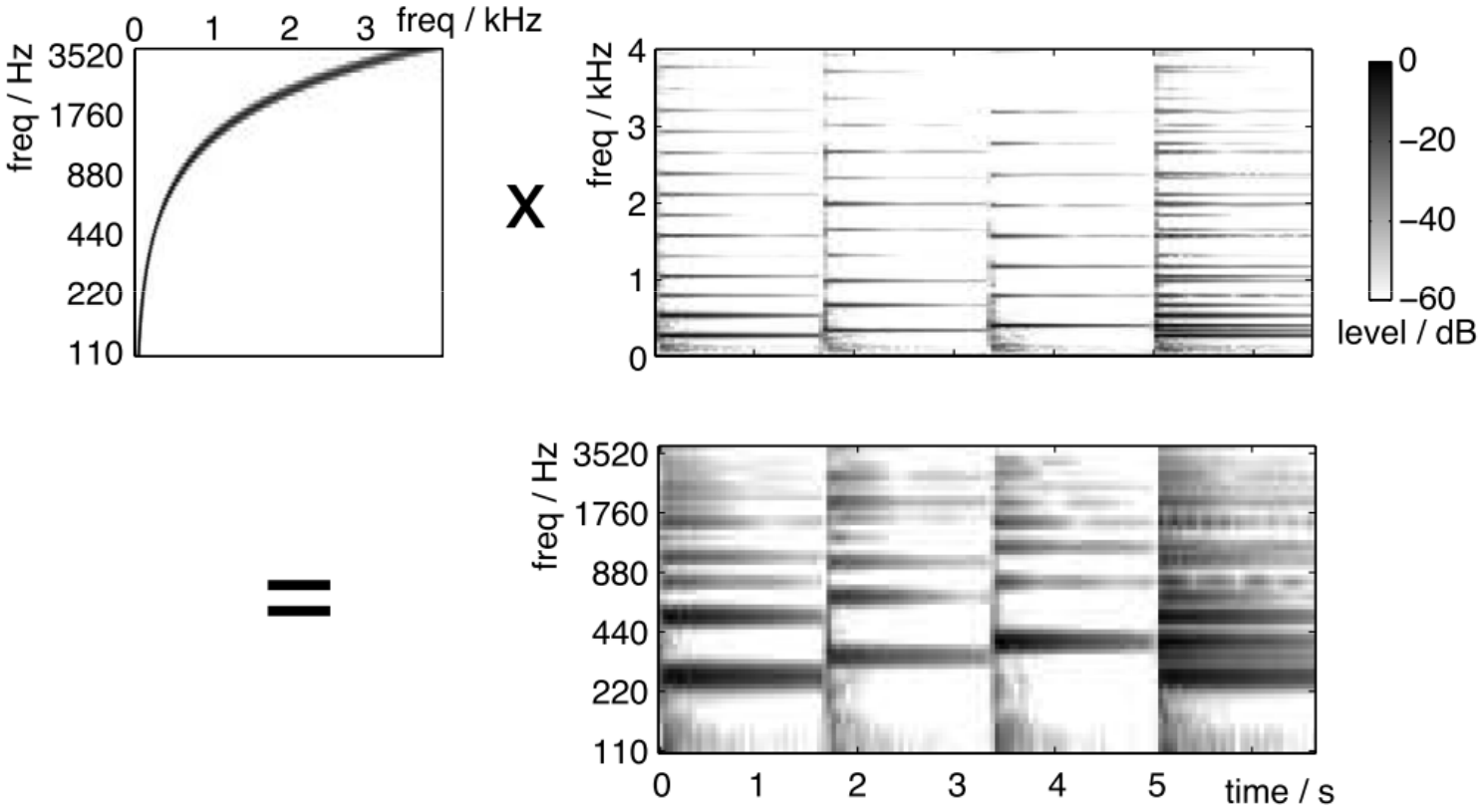
Factorisation en Matrices Non-négatives

- **Utilisation en estimation multi-pitch:**
  - Nécessité d'introduire des *a priori* (approche probabiliste) ou *des contraintes* (approche déterministe)
  - Exemple de contraintes (d'après Vincent & al, 2010):
    - NMF classique: $Y_{ft} = \sum_{i=1}^{I} A_{it} S_{if}$
    - NMF avec templates dépendantsdu pitch: $Y_{ft} = \sum^{p_{\text{high}}} \sum^{J_p} A_{pjt} S_{pjf}$
    - ..et avec contraintes sur les templates $S_{pjf} = \sum_{k=1}^{K_p} E_{pjk} N_{pkf}$
    - Exemples d'enveloppes locales



Droits d'usage autorisé

Utilisation d'une représentation à Q constant



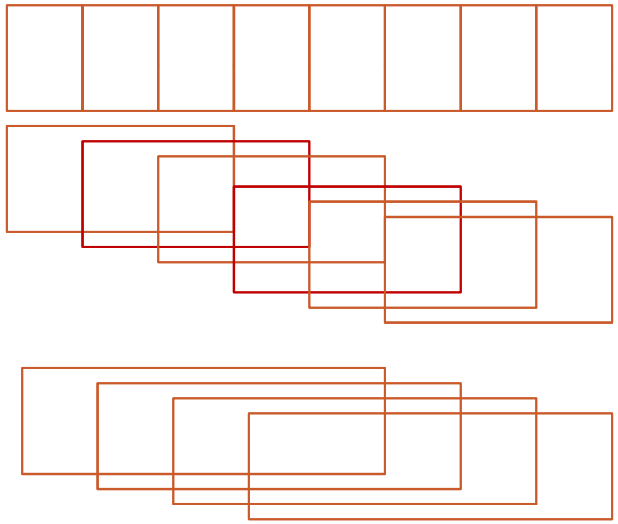
D'après M. Mueller & al. « Signal Processing for Music Analysis, IEEE Trans. On Selected topics of Signal Processing, oct. 2011

Droits d'usage autorisé

Utilisation d'une représentation à Q constant

- **En pratique:**
  - Solution peu satisfaisante

- **Solution souvent retenue: Utiliser des tailles de fenêtres différentes pour chaque nouveau bin fréquentiel k'**

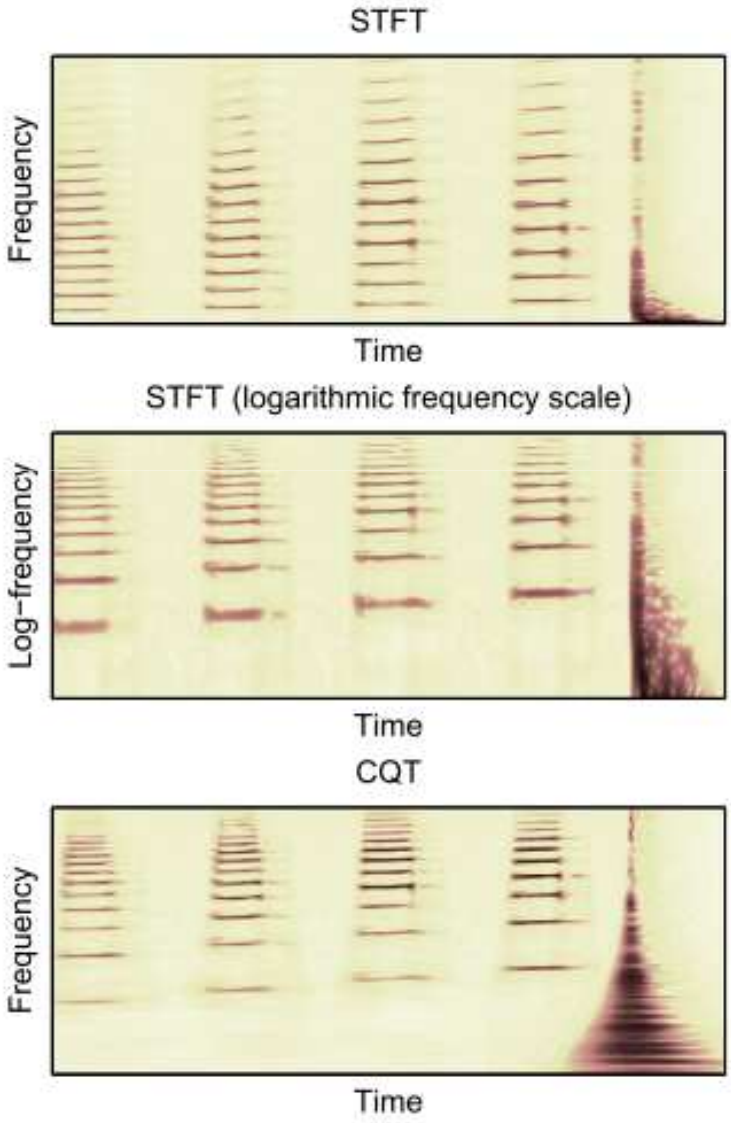


J. Brown and M. Puckette, An efficient algorithm for the calculation of a constant Q transform, JASA, 92(5):2698–2701, 1992.
J. Prado, Une inversion simple de la transformée à Q constant, technical report, 2011, http://www.tsi.telecom-paristech.fr/aao/en/2011/06/06/inversible-cqt/

Droits d'usage autorisé

Utilisation en estimation multipitch

- **Sur une transformée à Q constant:**
  - Une différence de pitch correspond à une translation sur l'axe des fréquences
  - Vers des modèles "Shift invariant PLCA (v. smaragdis2008 et Fuentes & al. 2011)



Droits d'usage autorisé

Quelques références en estimation de Fréquence(s) fondamentale(s)

Estimation de la fréquence fondamentale

- M. Schroeder, “Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement” The Journal of the Acoustical Society of America -- April 1968 -- Volume 43, Issue 4, pp. 829-834
- Alain de Cheveigné, *YIN, a fundamental frequency estimator for speech and music*, Hideki Kawahara, JASA, 111(4), 2002
- Geoffroy Peeters, *Music pitch representation by periodicity measures based on combined temporal and spectral representations*, ICASSP 2006

Estimation de fréquences fondamentales multiples

- B. Fuentes, R. Badeau, and G. Richard, “Adaptive harmonic time-frequency decomposition of audio using shift-invariant PLCA,” in Proc. of ICASSP, Prague, Czech Republic, May 2011, pp. 401–404.
- P. Smaragdis, B. Raj, and M.V. Shashanka, “Sparse and shift-invariant feature extraction from non-negative data,” in Proc. of ICASSP, Las Vegas, Nevada, USA, April 2008, pp. 2069–2072.
- E. Vincent, N. Bertin, and R. Badeau, “Adaptive harmonic spectral decomposition for multiple pitch estimation,” IEEE Transactions on Audio Speech and Language Processing, vol. 18, no. 3, pp. 528–537, Mar. 2010.
- T. Tolonen and M. Karjalainen, “A computationally efficient multipitch analysis model,” *IEEE Trans. On Speech and Audio Processing*, vol. 8, no. 6, pp. 708–716, 2000.
- Anssi P. Klapuri, *Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness*, IEEE Trans. On Speech and Sig. Proc., 11(6), 2003
- C. Yeh, A. Röbel, and X.Rodet, "*Multiple fundamental frequency estimation of polyphonic music signals*", IEEE ICASSP, pp. 225-228 (Vol. III), Philadelphia, Pennsylvvania, USA, 2005.
- Hirokazu Kameoka, Takuya Nishimoto, and Shigeki Sagayama, “A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering”, IEEE Trans. On ASLP, March. 2007
- V. Emiya, R. Badeau, B. David, “MULTIPITCH ESTIMATION OF QUASI-HARMONIC SOUNDS IN COLORED NOISE”, Proc. Of DAFX, Sept. 2007.
- V. Emiya, “Transcription automatique de la musique de piano », thèse de doctorat, Telecom ParisTech, 2008.
- Anssi P. Klapuri, *A perceptually motivated multiple-f0 estimation method*, WASPAA 2005
- Anssi P. Klapuri “Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model”, IEEE Trans. On ASLP, Feb. 2008

Droits d'usage autorisé