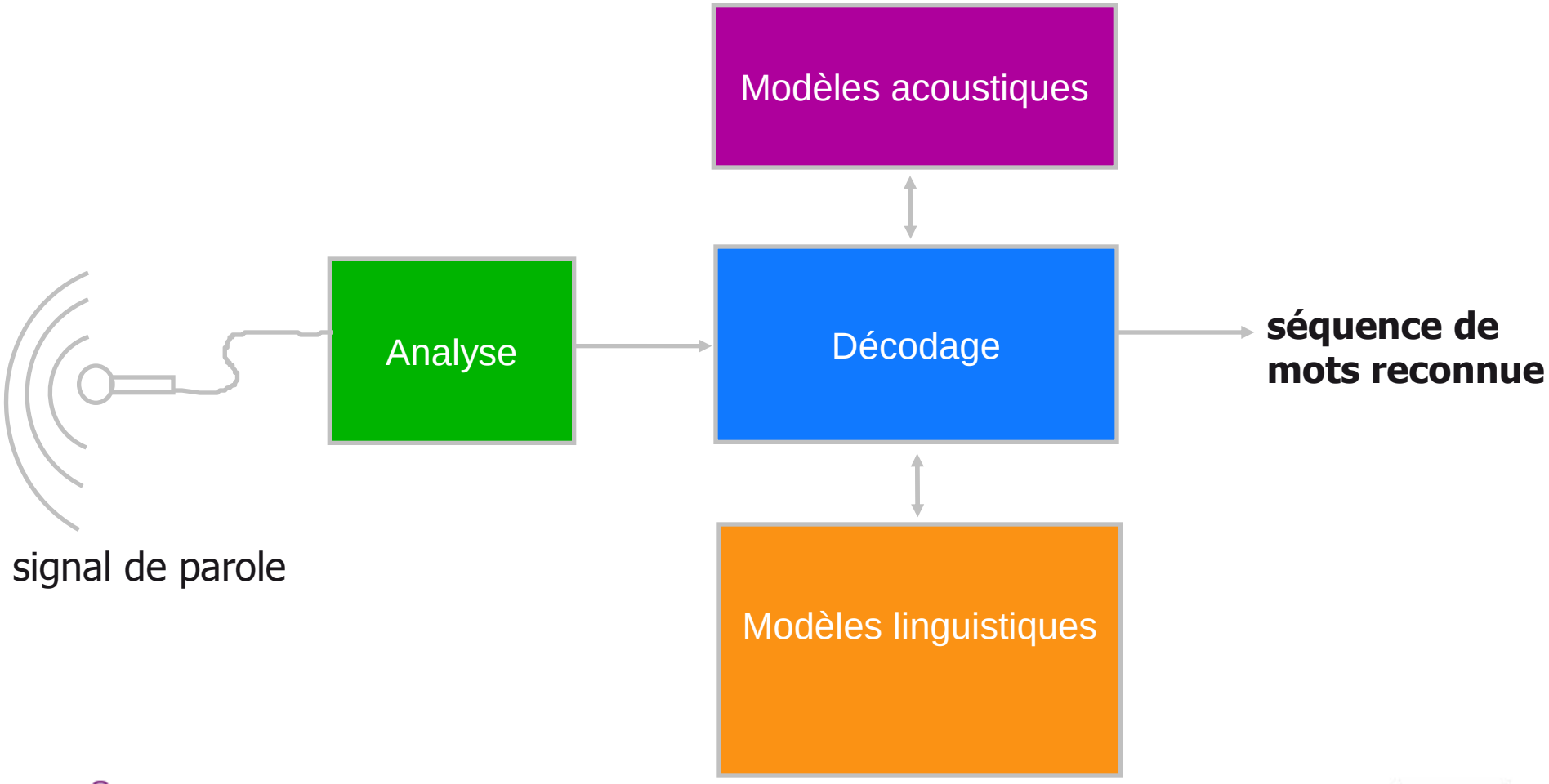
Master 2 Recherche AIC & SETI

Reconnaissance et interaction vocale

Quelques bases de traitement du signal

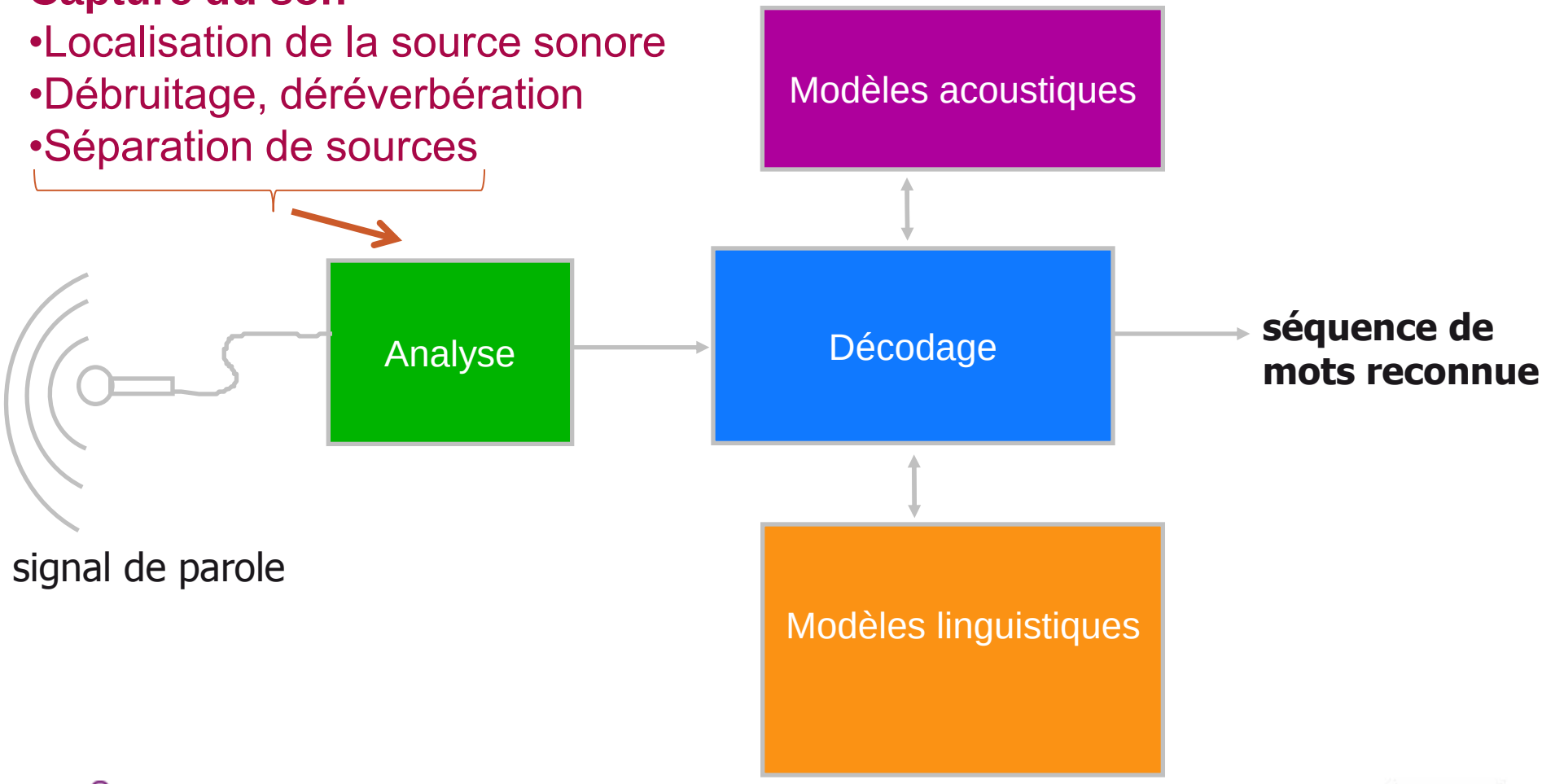
G. Richard

Le « traitement du Signal » dans la reconnaissance vocale



Le « traitement du Signal » dans la reconnaissance vocale

Capture du son
- Localisation de la source sonore
- Débruitage, déréverbération
- Séparation de sources



Le « traitement du Signal » dans la reconnaissance vocale

Capture du son
- Localisation de la source sonore
- Débruitage, déréverbération
- Séparation de sources

signal de parole

Analyse

Paramétrisation
- MFCC, LPCC,..
- DNN,…

Modèles acoustiques

Décodage

Modèles linguistiques

séquence de mots reconnue

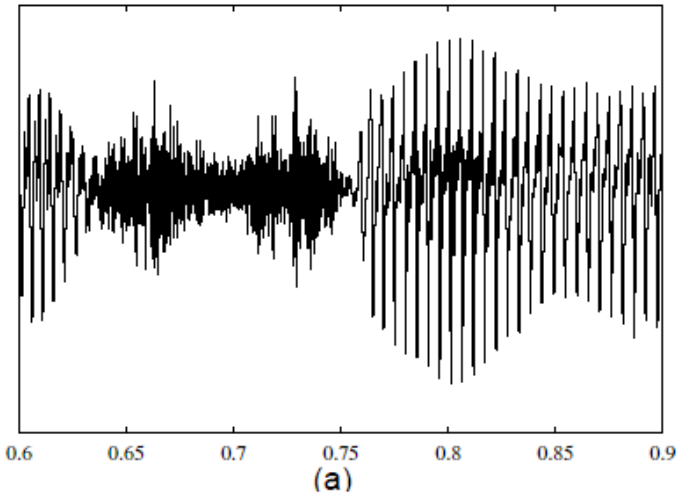
Contenu

- **Objectif du cours:**
  - Présenter quelques bases du traitement du signal

- **Contenu**
  - Représentation de Fourier
  - Échantillonnage
  - Transformée en Z
  - Transformée de Fourier Discrète
  - Filtrage
  - La représentation cepstrale

Représentation des signaux

- **Qu'est-ce qu'un signal ?**



- **Signal déterministe:** $x(t) = A\cos(2\pi f_0 t)$
- **Signal aléatoire**

Représentation de Fourier

- **Séries de Fourier**

  Tout signal périodique *x(t)* de période *T* peut être décomposé sous la forme d'une série de Fourier :

$$x(t) = \sum_{-\infty}^{\infty} X_n e^{2j\pi nt/T}$$

$$X_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t) e^{-2j\pi nt/T} dt$$

Formule de Parseval

Soit *x(t)* et *y(t)* deux signaux périodiques de période T

Soit $z(t) = x(t).y^*(t)$ **Alors** $Z_n = \sum_{-\infty}^{\infty} X_k Y_{k-n}^*$

(Exercice)

Formule de Parseval

Soit *x(t)* et *y(t)* deux signaux périodiques de période T

Soit $z(t) = x(t).y^*(t)$ **Alors** $Z_n = \sum_{-\infty}^{\infty} X_k Y_{k-n}^*$

En faisant *n=0*, on obtient

$$\sum_{k=-\infty}^{\infty} X_k Y_{k-n}^* = \frac{1}{T} \int_{-T/2}^{T/2} x(t) y^*(t) dt$$

En faisant *x(t)* = *y(t)* on obtient

$$P = \sum_{k=-\infty}^{\infty} |X_k|^2 = \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 dt$$

Interprétation: La puissance d'un signal est égale à la somme des puissances élémentaires de chacune de ses composantes.

Composante = signal « sinusoidal » $X_n e^{2j\pi nt/T}$

Représentation de Fourier (temps continu)

- Soit x(t) appartenant à $L_2 \cap L_1$ la transformée de Fourier existe et appartient à L_2

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2j\pi f t} dt$$

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{2j\pi f t} df$$

Propriétés

Properties	$x(t)$	$X(f)$
Convolution	$x(t) \star y(t)$	$X(f)Y(f)$
Similitude	$x(at)$	$\frac{1}{\lvert a \rvert} X(f/\lvert a \rvert)$
Translation	$x(t-t_0)$	$X(f)\exp(-2j\pi t_0 f)$
Modulation	$x(t)\exp(2j\pi f_0 t)$	$X(f-f_0)$
	real	$X(f) = X^*(-f)$

Important ?

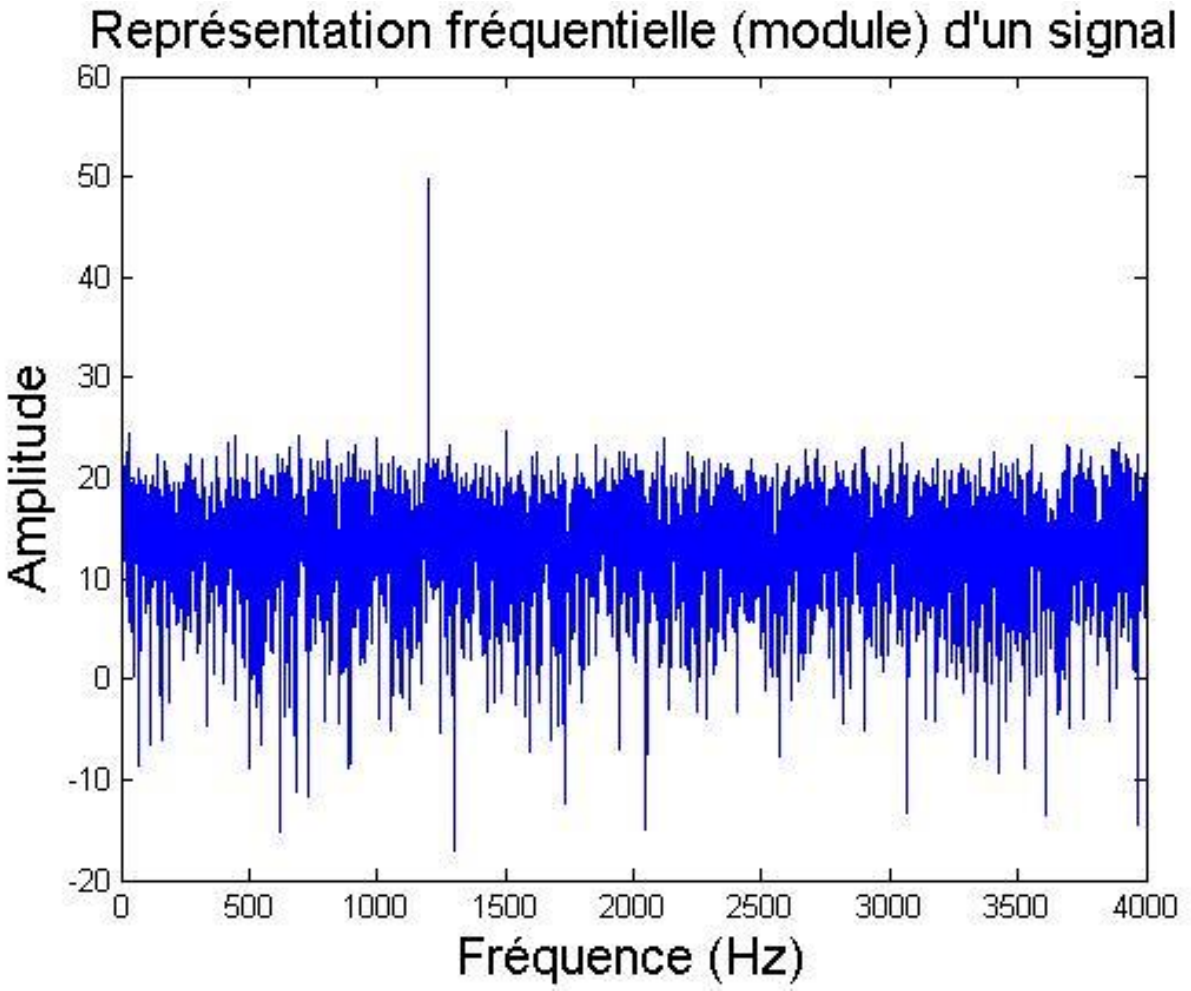
Exercice

- **Parseval**

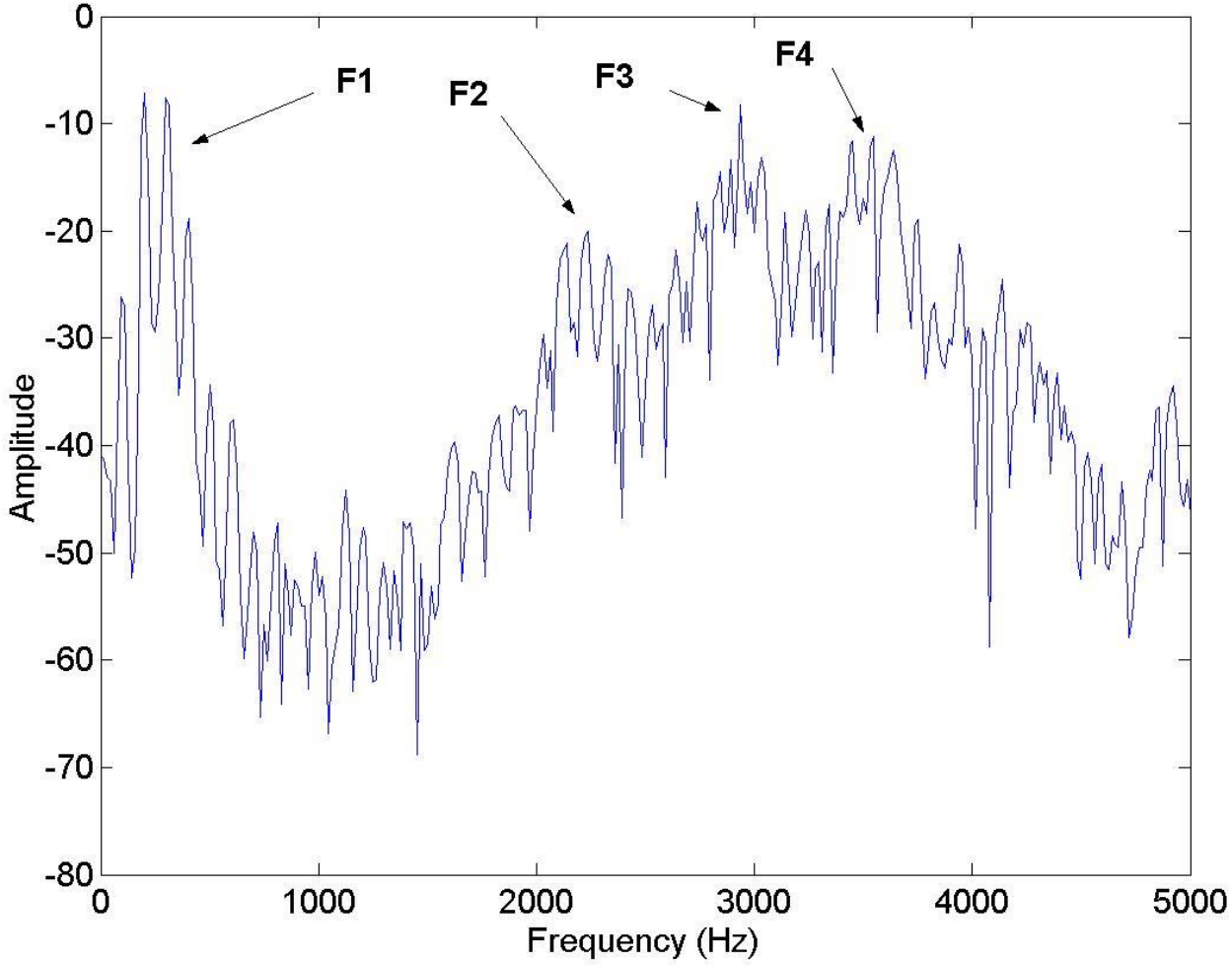
$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(f)|^2 df$$

- **Spectre (ou densité spectrale d'énergie):** $|X(f)|^2$

Exemple: Spectre de quel signal ?

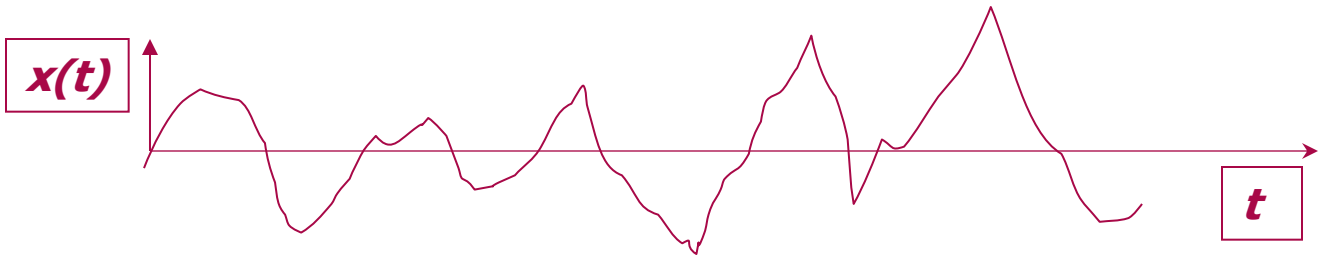


Exemple: Spectre d'un segment de /i/

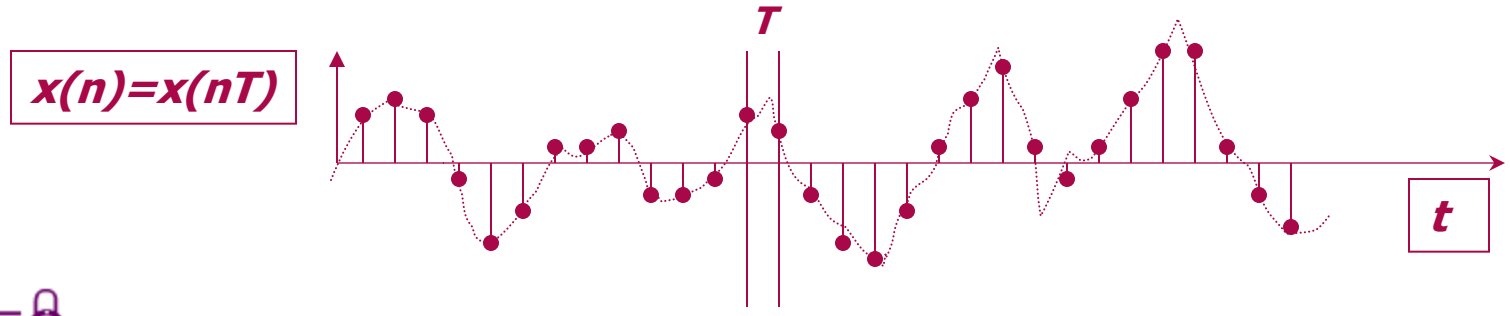


Représentation du signal

- Soit un signal x(t) à valeurs continues dans le temps:



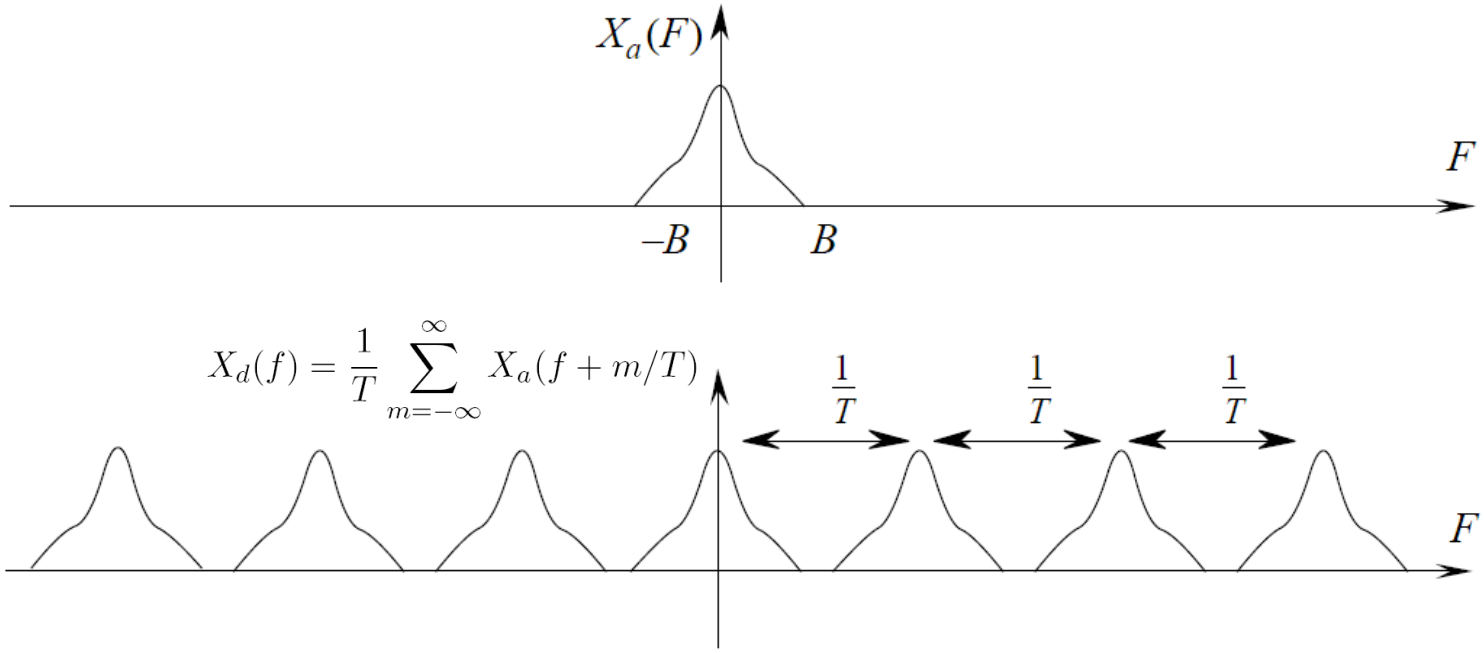
- Soit x(nT) le signal échantillonné à des valeurs discrêtes *t=nT*



Échantillonnage: Formule de Poisson

- **Interprétation: Echantillonnage ➡ périodisation du spectre**

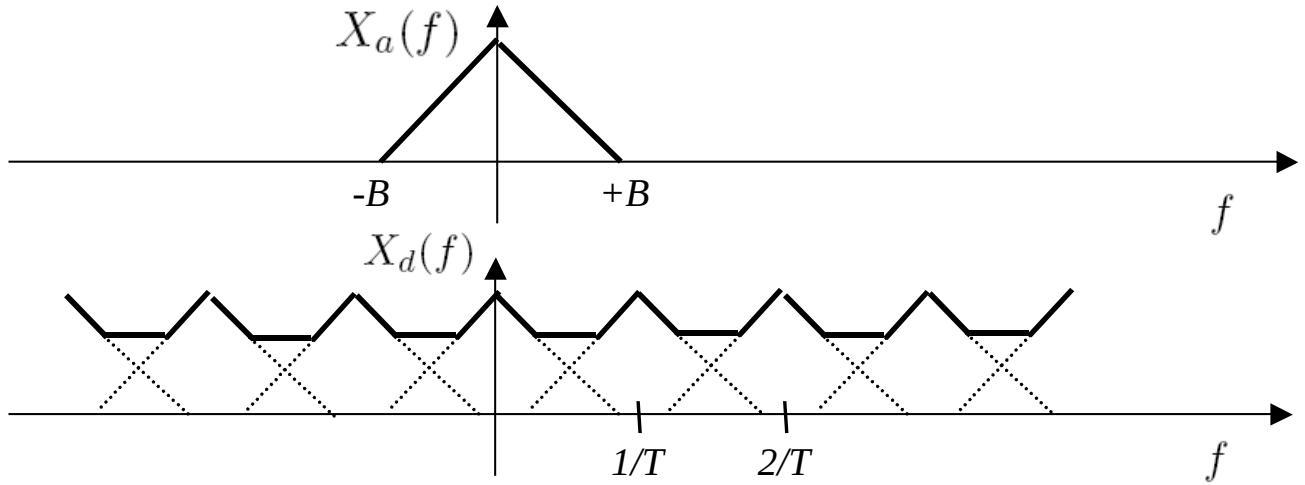
$$X_d(f) = \frac{1}{T} \sum_{m=-\infty}^{\infty} X_a(f + m/T) = \sum_{n=-\infty}^{\infty} x(n) e^{-2j\pi f n T}$$



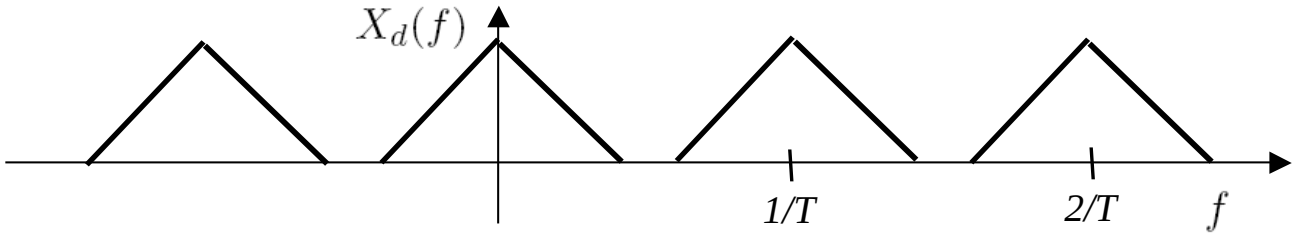
Reconstruction

- **2 situations:**

$\frac{1}{T} < 2B$

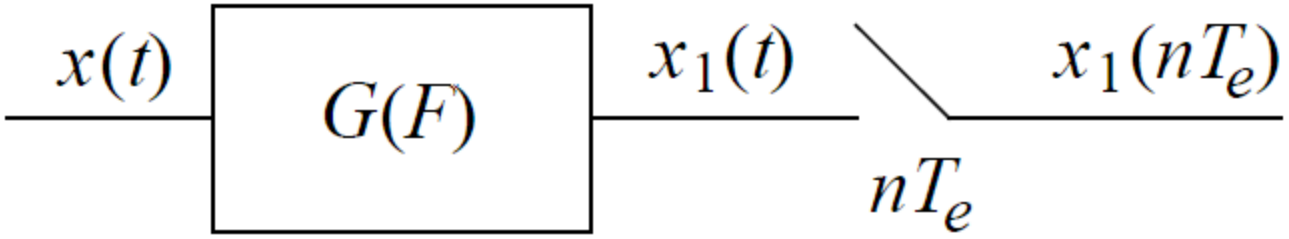


$\frac{1}{T} > 2B$



Échantillonnage d'un signal à bande illimitée

- Nécessité de filtrer le signal analogique pour obtenir un signal à bande limitée avant échantillonnage



Transformée en Z / TFTD

- La transformée en Z d'un signal x(n) est donnée par:

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} \quad \text{avec} \quad z \in \mathcal{C} = \{z \in \mathbb{C} : R_1 < |z| < R_2\}$$

- La Transformée de Fourier à Temps Discrêt (TFTD) est donnée par:

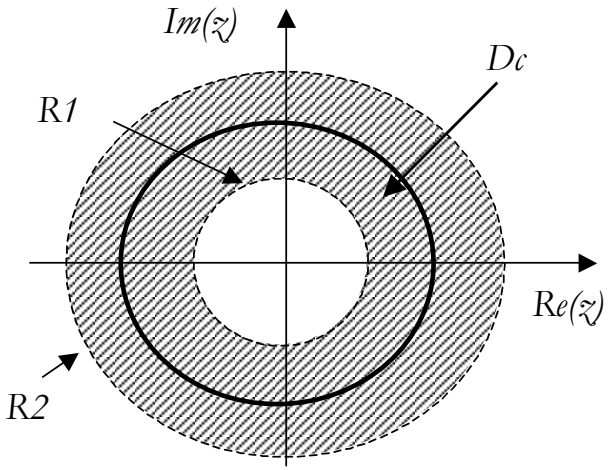
$$X(e^{2j\pi f}) = \sum_{n=-\infty}^{\infty} x(n) e^{-2j\pi n f}$$

$$x_n = \int_{-1/2}^{1/2} X(e^{2j\pi f}) e^{2j\pi n f} df$$

- ➡ $X(e^{2j\pi f})$ est périodique de période 1

Quelques résultats

- **Le domaine de convergence D_c est une couronne circulaire**



- **Si x(n) est de durée finie D_c est le plan tout entier** $D_c = R_1 \leq |z| < \infty$
- **Causalité: si *x(n)* est nul à gauche (*x(n)= 0 pour n<0*) on a**
  - Exemple:

$$x(n) = a^n u(n) \longrightarrow X(z) = \frac{1}{1 - az^{-1}} \quad \text{Converge pour} \quad |z| > a$$

Quelques propriétés

- **Linéarité**
- **Symétrie hermitienne**

$$x(n) \text{ real } \rightleftarrows X(e^{2j\pi f}) = X^*(e^{-2j\pi f})$$

- **Convolution**

$$y(n) = h(n) \star x(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) \rightleftarrows \begin{cases} Y(z) = H(z)X(z) \\ Y(e^{2j\pi f}) = H(e^{2j\pi f})X(e^{2j\pi f}) \end{cases}$$

- **Décalage fréquentiel**

$$y(n) = x(n)\exp(2j\pi f_0 n) \rightleftarrows Y(e^{2j\pi f}) = X(e^{2j\pi(f-f_0)})$$

- **Décalage temporel (retard)**

$$y(n) = x(n-n_0) \rightleftarrows Y(e^{2j\pi f}) = e^{2j\pi f n_0} X(e^{2j\pi f})$$

$$Y(e^{2j\pi f}) = e^{-2j\pi f n_0}.X(e^{2j\pi f})$$

Transformée de Fourier Discrète (TFD)

- **Par définition, la TFTD est une fonction périodique de période 1.**
- **En pratique, nous prenons N échantillons, et on discrétise l'intervalle de fréquences [0-1] en L valeurs telles que:**

$$f = k/L \quad \text{avec} \quad k \in \{0,1,..L-1\}$$

- **On obtient:**

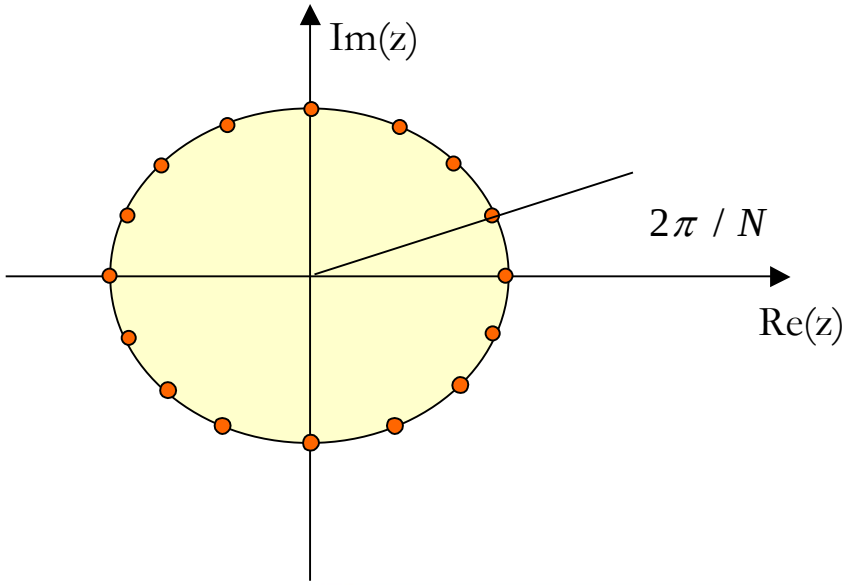
$$X_e(k/L) = \sum_{n=0}^{N-1} x_n e^{-2j\pi nk/L}$$

$$\{x_0, \cdots, x_{N-1}\} \rightleftarrows \{X_0, \cdots, X_{N-1}\} \quad with \quad \text{avec} \begin{cases} X_k = \sum_{n=0}^{N-1} x_n e^{-2j\pi nk/N} \\ x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2j\pi nk/N} \end{cases}$$

Relation TZ <-> TFD

- Cela correspond à un échantillonnage de la transformée en z en N points régulièrement espacés autour du cercle unité

$$X(k) = X(z)\big|_{z=e^{2j\pi k/N}}$$



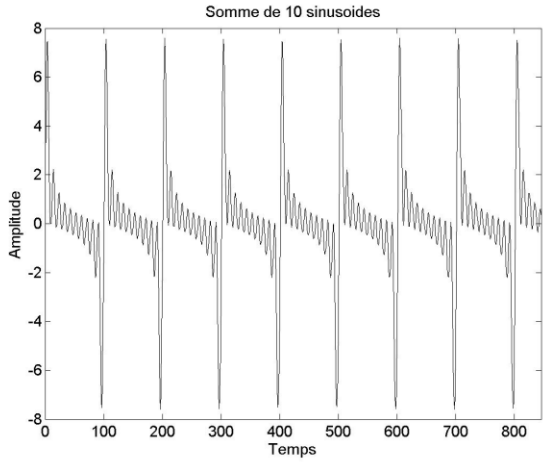
Représentation temps-fréquence

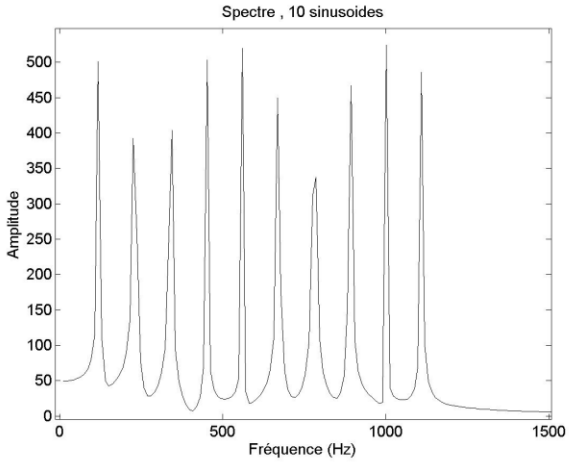
- **Transformée de Fourier discrête**

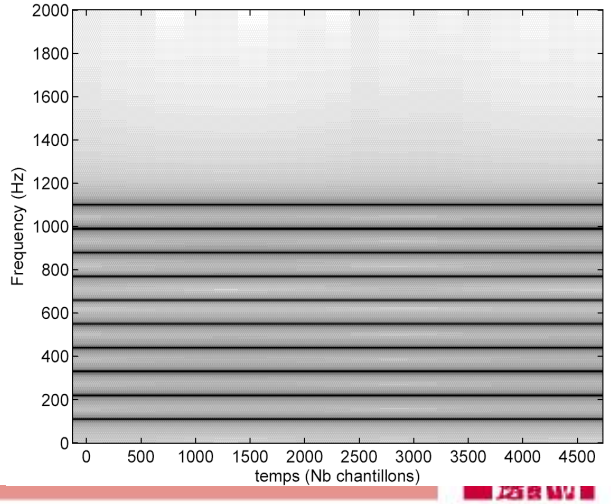
$$X_k = \sum_{n=0}^{N-1} x_n e^{-2j\pi nk/N}$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2j\pi nk/N}$$

x_n $|X_k|$ *Spectrogramme*

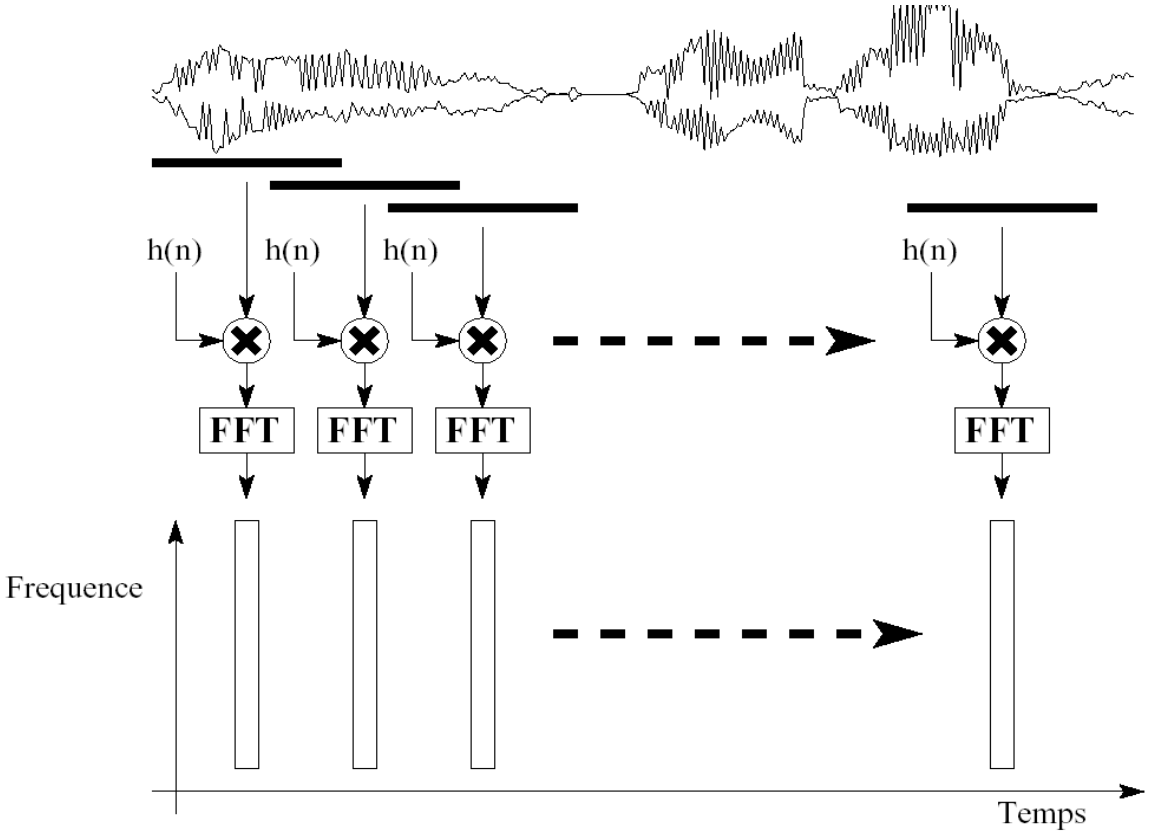






Paramétrisation: paramètres spectraux

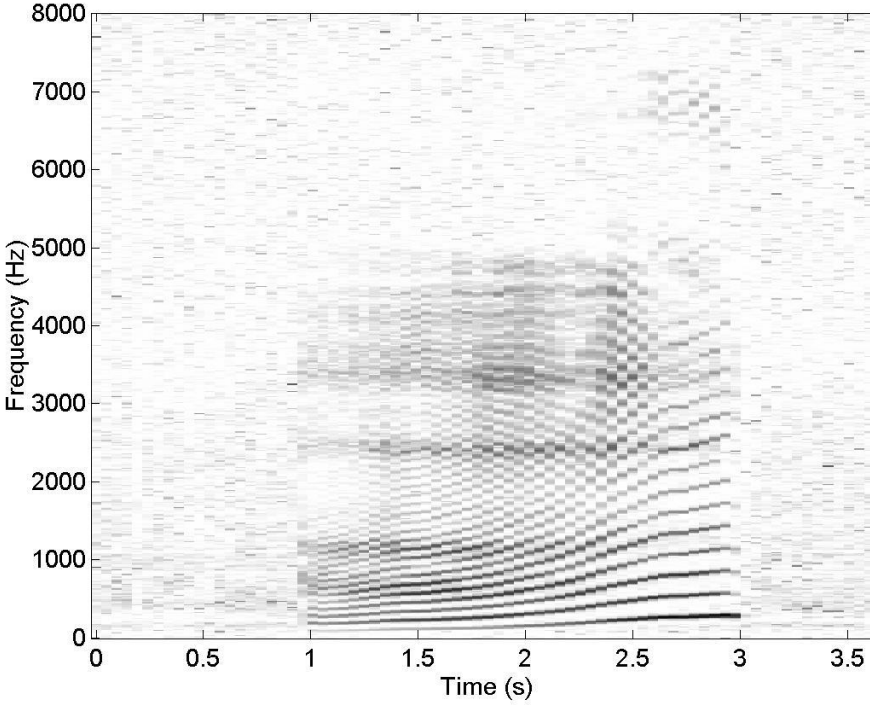
- **Paramétrisation spectrale: analyse d'un signal audio**
  (*d'après Laroche*)



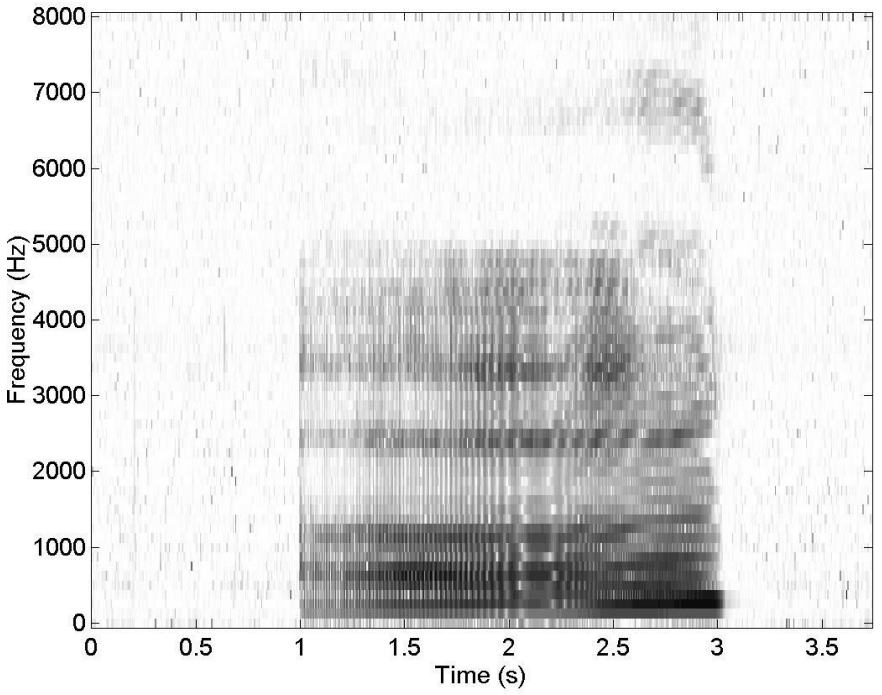
Description du signal de parole

Importance de la taille de la fenêtre d'analyse

Bande étroite

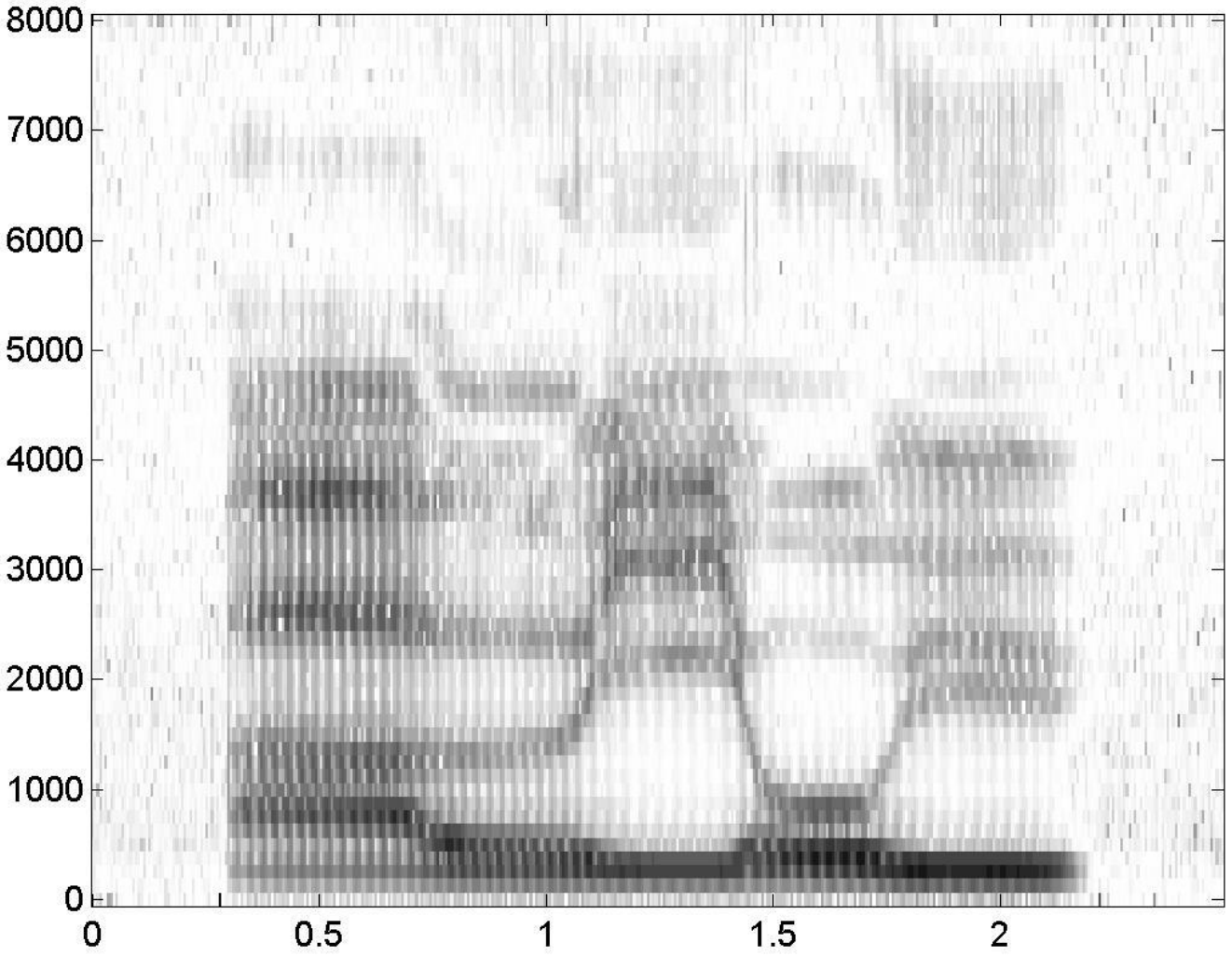


Large bande



Spectrogrammes sur une voyelle /a/ avec un pitch montant

Spectrogramme des voyelles / a e i o u/

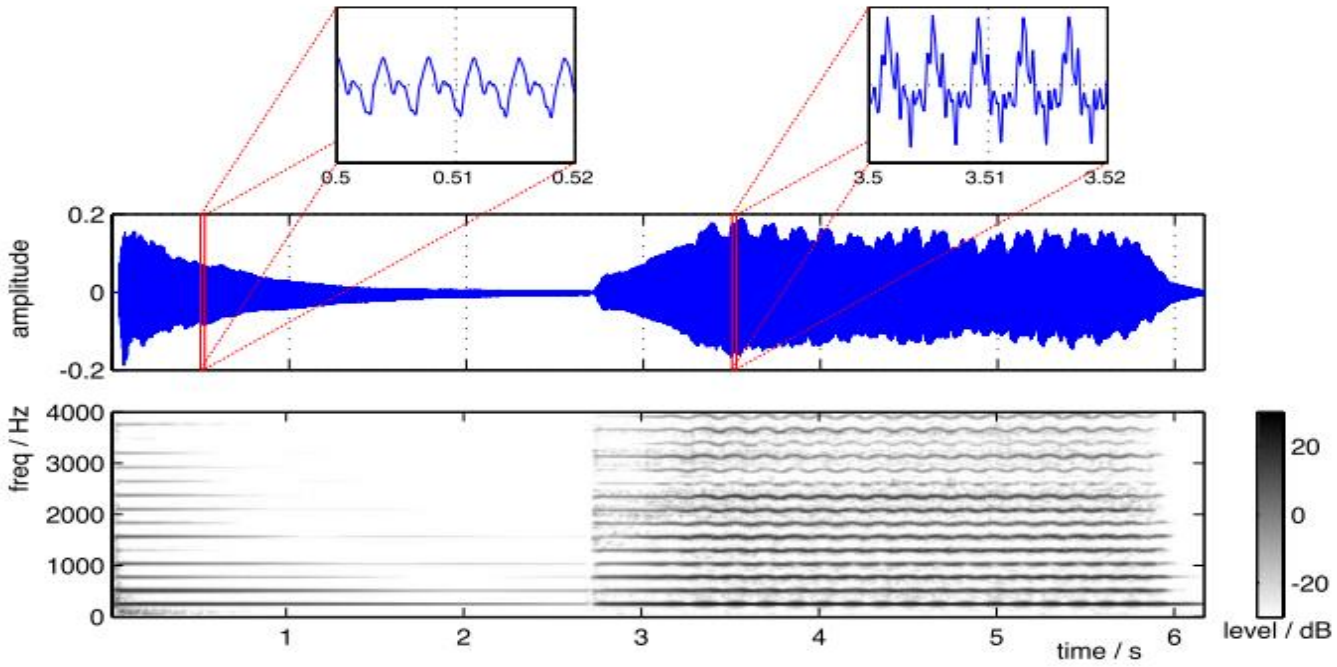


Représentations du signal audio

- **Exemple sur un signal audio: note Do (262 Hz) jouée par un piano et un violon.**

Signal temporel

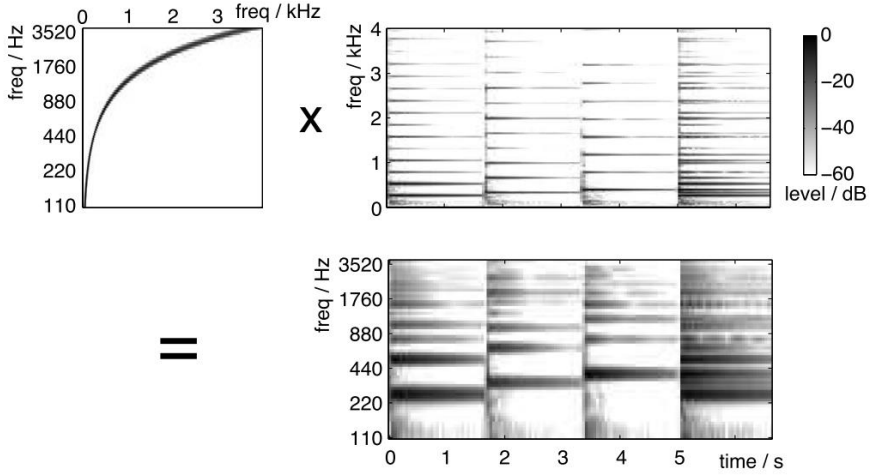
Spectrogramme



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

Représentations du signal audio

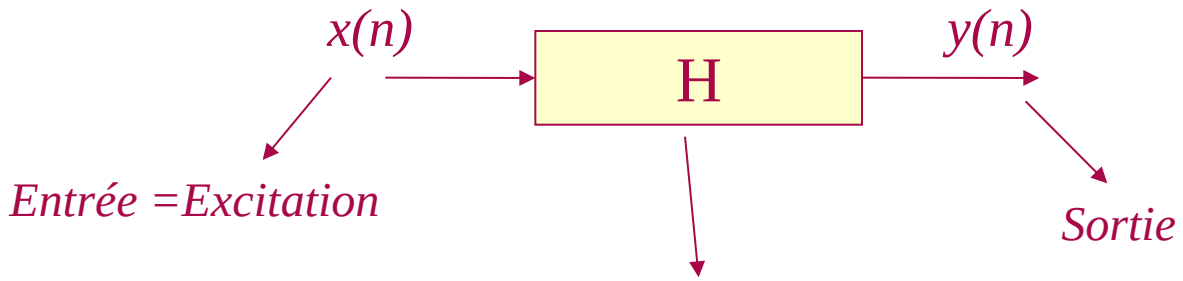
- **Exploitation de propriétés perceptives: Echelles fréquentielles non linéaires**
  - Transformée à « Q » constant
  - Transformée temps- log(fréquence)



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

Notions de Filtrage

- **Système linéaire invariant dans le temps**



Filtre est caractérisé par sa réponse impulsionnelle *h(n)*
et sa fonction de transfert H(z)

- La convolution permet de caractériser la transformation entrée/sortie réalisée par un filtre linéaire invariant.

$$y(n) = \sum_{-\infty}^{\infty} x(k)h(n-k) = \sum_{-\infty}^{\infty} x(n-k)h(k)$$

$$y(n) = x(n) * h(n)$$

Notions de Filtrage (2)

- **Equation récurrente entrée/sortie (pour un filtre RIF)**

$$y(n) = h(n) * x(n) = \sum_{i=0}^{N-1} a_i x(n-i)$$

- **Par transformée en Z:**

$$Y(z) = \sum_{-\infty}^{\infty} y(n) z^{-n} \longrightarrow \boxed{Y(z) = H(z)X(z)}$$

- **Réponse en fréquence**

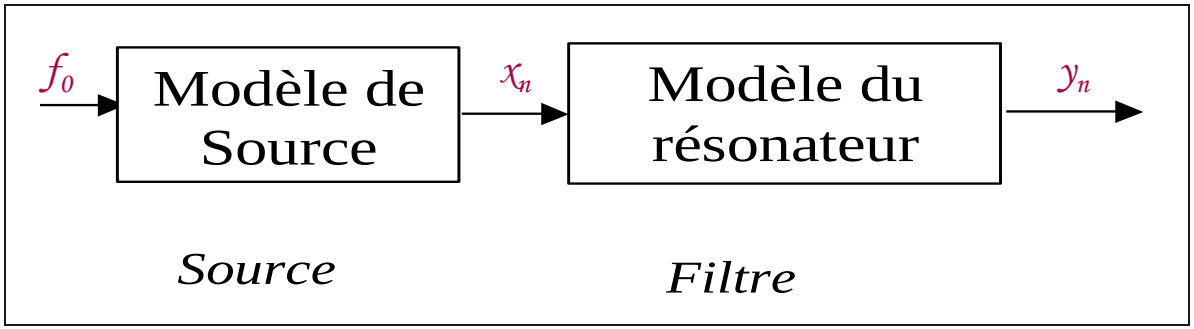
$$H(e^{2j\pi f}) = \sum_{k=-\infty}^{\infty} h(k) e^{-2j\pi k f} = H(z)|_{z=e^{2j\pi f}}$$

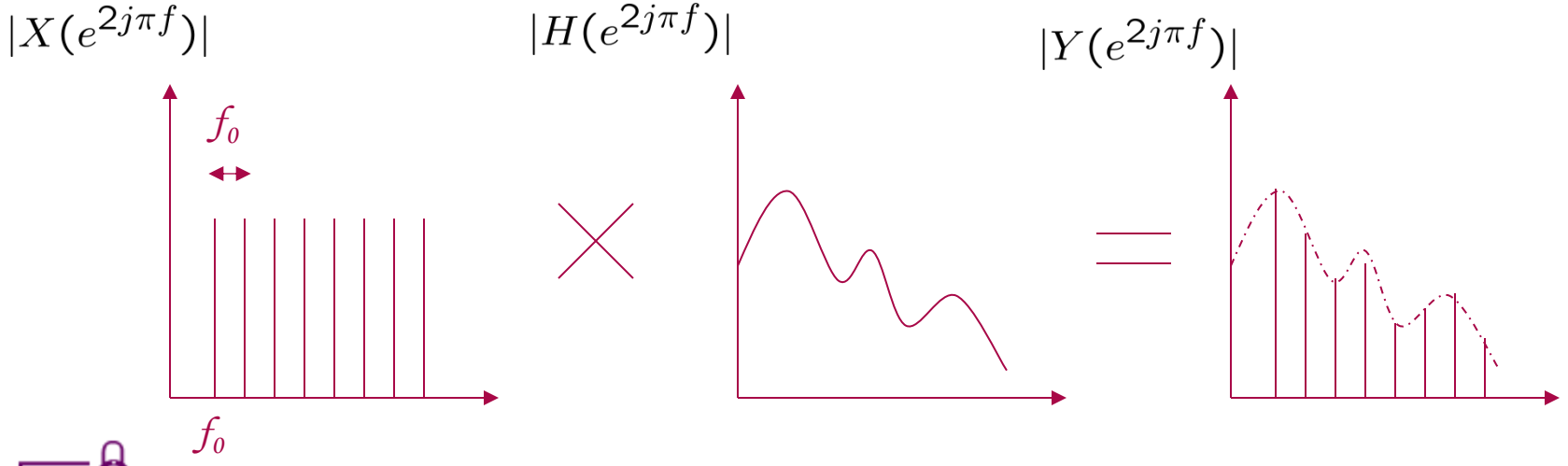
$$H(e^{2j\pi f}) = |H(e^{2j\pi f})| e^{j\phi(f)}$$

Module Phase

Modèle source-filtre

- enveloppe spectrale, source





Echelle Mel

- **Correspond à une approximation de la sensation psychologique de hauteur d'un son (Tonie)**

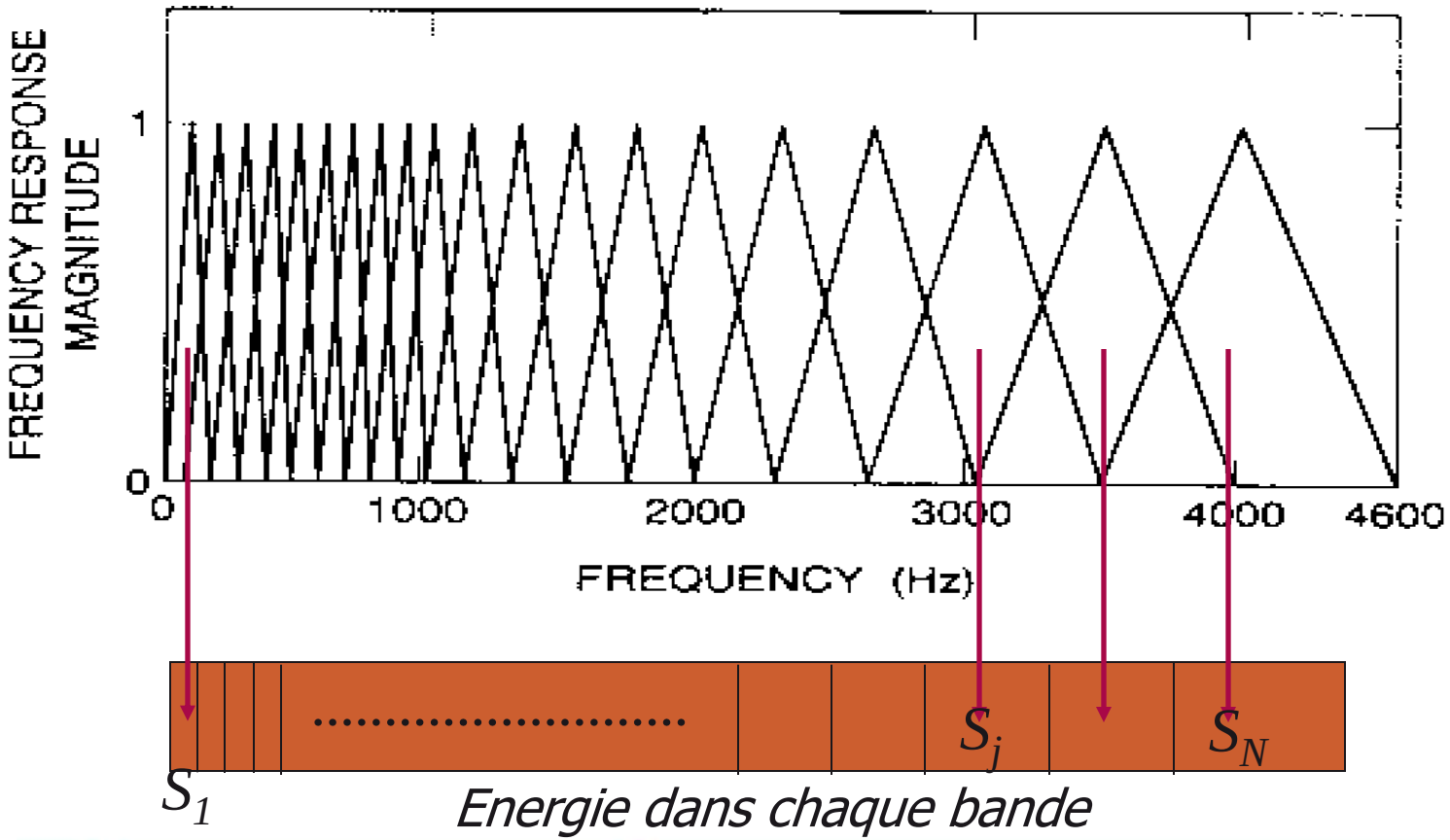
- **Existence de formules analytiques:**

$$mel(f) = 1000 \log_2(1 + \frac{f}{1000})$$

- **Exemples:**
  - Gamme mel
  - Gamme Hertz

Filtre en échelle Mel

- **Filtrage Mel (d'après Rabiner93)**



Représentation cepstrale

- **Intérêt**
  - Modèle source filtre de la parole

$$s(t) = g(t) * h(t)$$

✓ Modèle source filtre dans le domaine spectral

$$S(\omega) = G(\omega)H(\omega)$$

✓ Cepstre (réél): somme de 2 termes

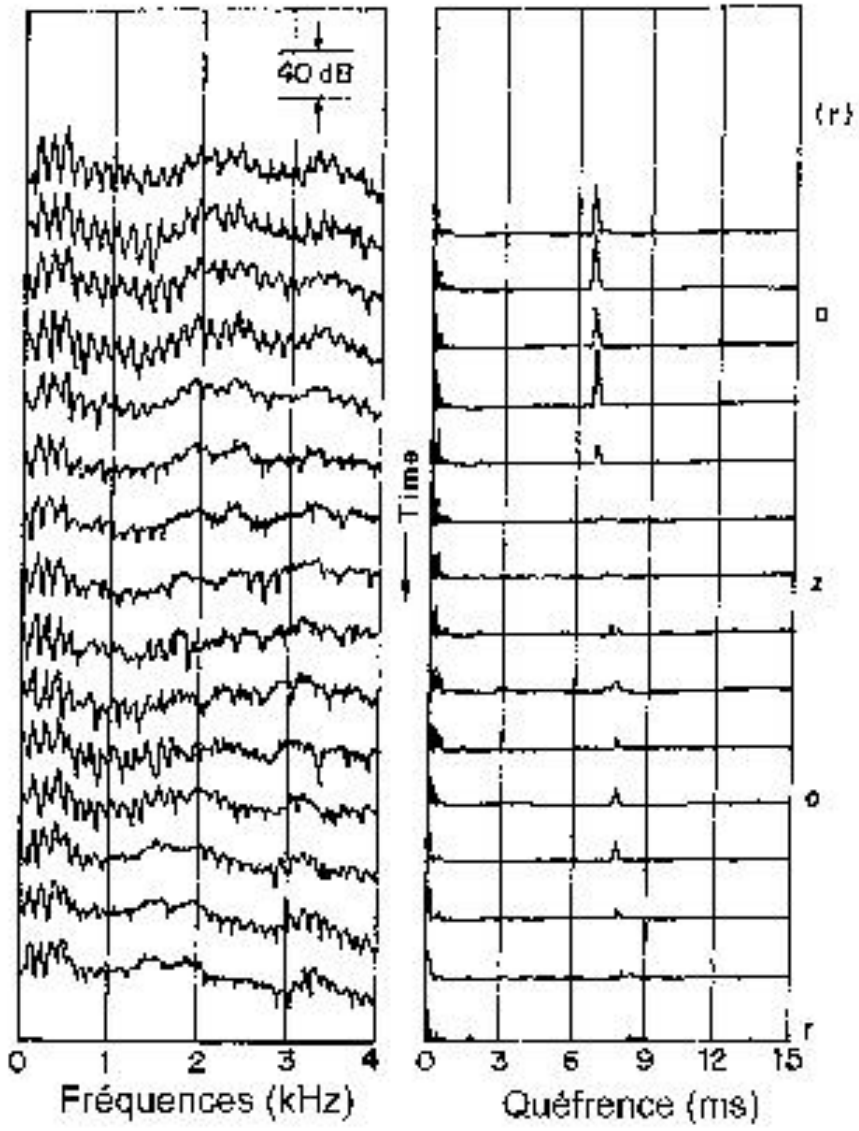
$$c(\tau) = FFT^{-1} \log |S(\omega)| = FFT^{-1} \log |G(\omega)| + FFT^{-1} \log |H(\omega)|$$

$$c_n = \frac{1}{N} \sum_{k=0}^{N-1} \log |X(k)| e^{2j(\pi)kn/N}$$

Représentation cepstrale (d'après Furui2001)

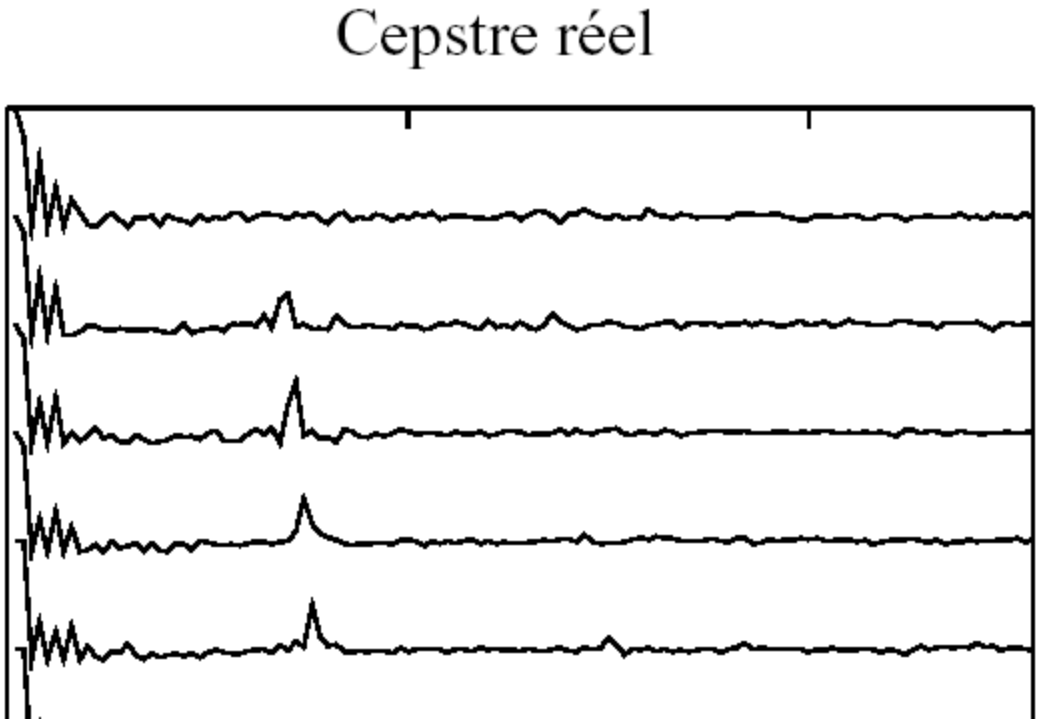
- **Exemples:**
  - de Spectres à court terme (gauche)
  - et de cepstre $c(\tau)$ (droite)

- **τ est homogène à un temps et est appelé quéfrence**



Représentation cepstrale

- **Séparation de la contribution du conduit vocal et de la source par liftrage**



Représentation cepstrale

- **Contribution de la source**

$$p_n = \sum_{i=-\infty}^{\infty} \alpha_i \delta(n - iT) \longrightarrow \hat{p}_n = \sum_{i=-\infty}^{\infty} \beta_i \delta(n - iT)$$

- **Contribution du conduit vocal**
- **(hypothèse: filtre causal, stable, minimum de phase)**

$$F(z) = K \frac{\prod_{j=1}^{M}(1 - a_j z^{-1})}{\prod_{j=1}^{N}(1 - b_j z^{-1})} \qquad \begin{aligned} |a_j| &< 1 \\ |b_j| &< 1 \end{aligned}$$

Représentation cepstrale

- **Contribution du conduit vocal**

$$\log(F(z)) = \sum_{n=0}^{\infty} c_n z^{-n}$$

- **Développement en série**

$$\log(1-a) = -\sum_{n=1}^{\infty} a^n/n \quad pour \quad |a| < 1$$

$$\hat{c}_n = \begin{cases} \log(K) & n = 0 \\ -\sum_{j=1}^{M} \frac{a_j^n}{n} + \sum_{j=1}^{N} \frac{b_j^n}{n} & n > 0 \end{cases} \qquad |z| > |a_j|, |b_j|$$

Représentation cepstrale

- **Exemples de liftres (d'après Calliope89)**

(1) filtre rectangulaire

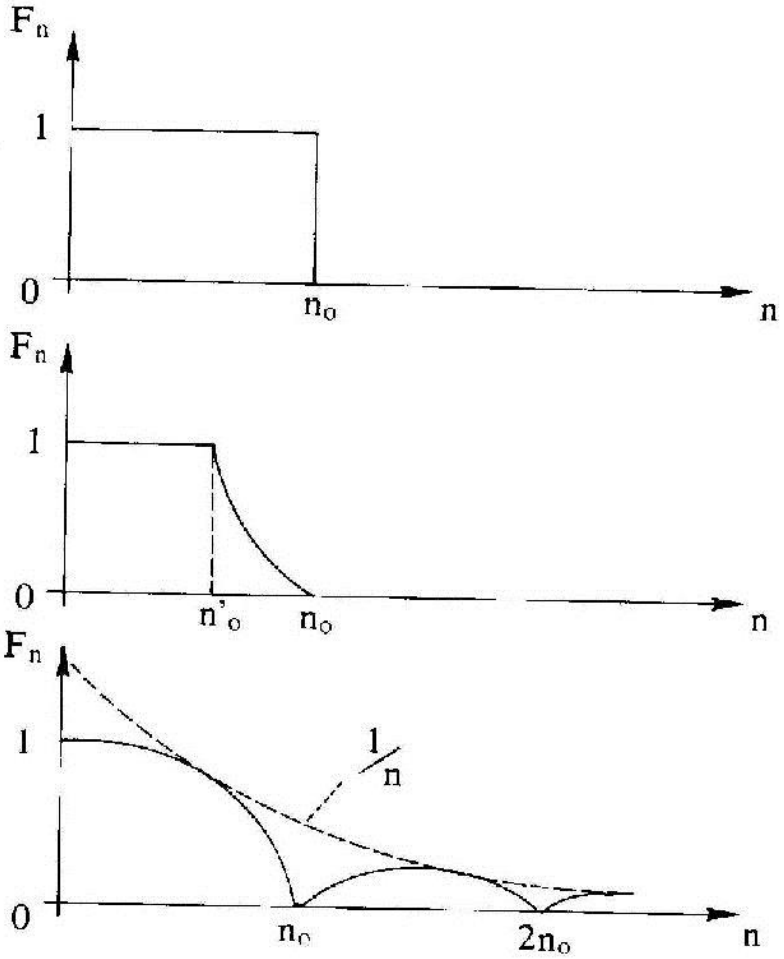
$$\begin{cases} F_n = 1 & \text{si } n < n_0 \\ F_n = 0 & \text{si } n \geqslant n_0 \end{cases}$$

ou (2) filtre adouci

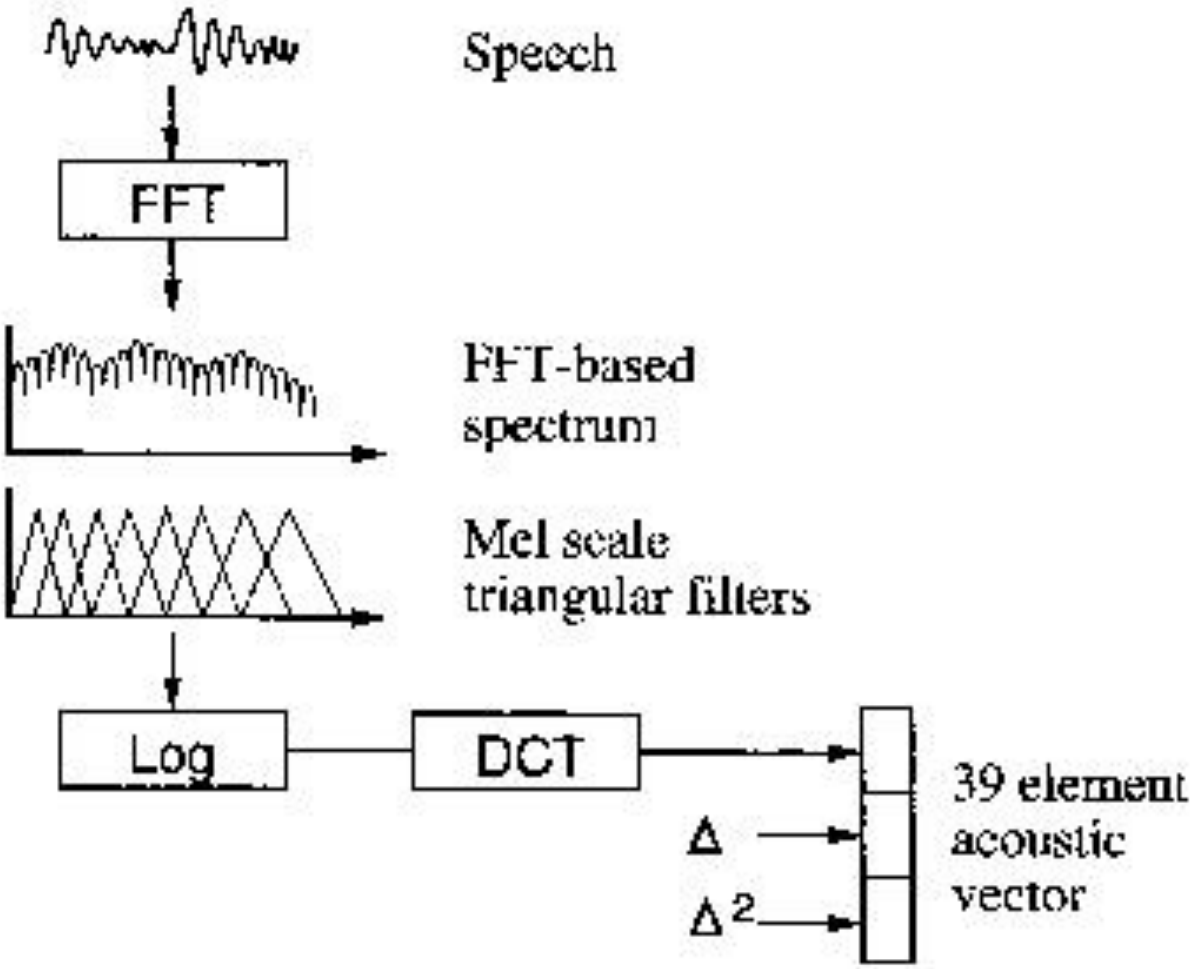
$$\begin{cases} F_n = 1 & \text{si } n < n'_0 < n_0 \\ F_n = 1 - e^{-\alpha(n-n'_0)} & \text{si } n \geqslant n'_0 \end{cases}$$

ou (3) filtre de Combs

$$F_n = \hat{C}_n - C_{n-n_0}$$



Paramétrisation MFCC
« Mel-Frequency Cepstral Coefficients »



Paramétrisation MFCC

- **Calcul des coefficients MFCC**

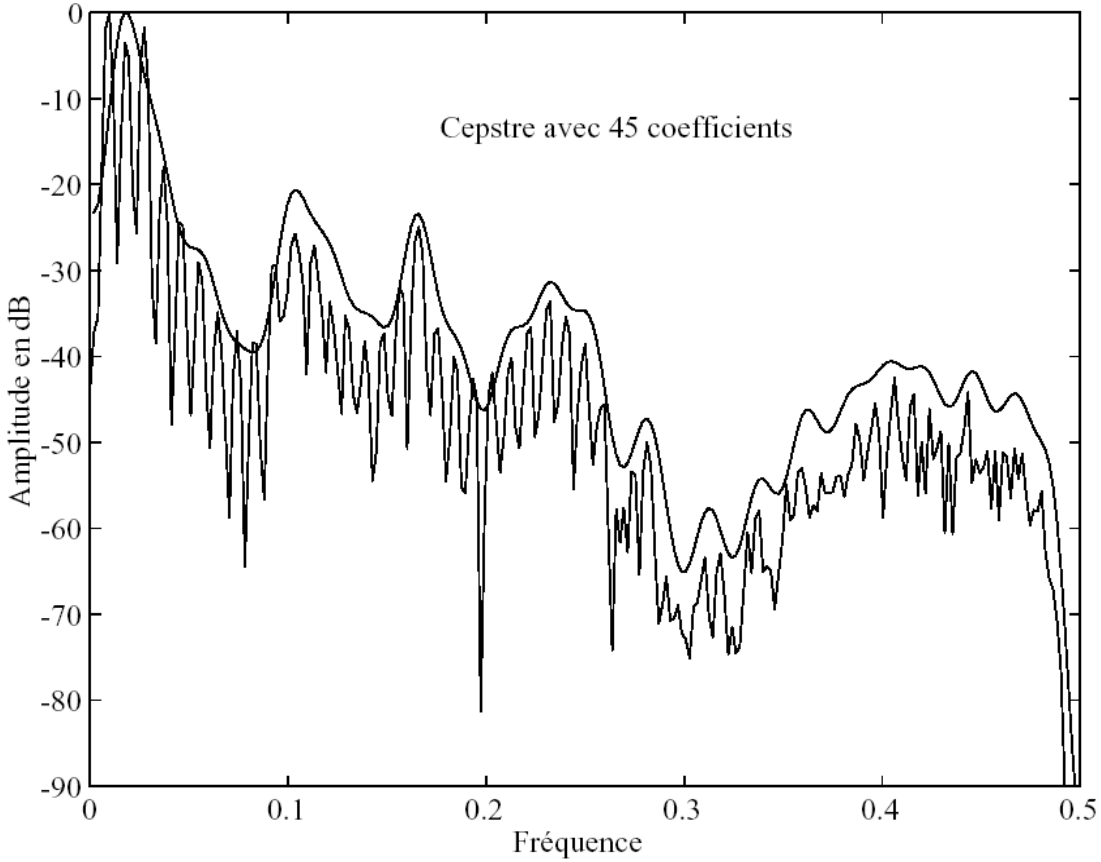
$$\tilde{c}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos \left[n(k - \frac{1}{2}) \frac{\pi}{K} \right]$$

$$\text{pour} \quad n = 1,2,\ldots,L$$

- **Une implémentation classique:**
  - 13 Coefficients (sans C0)
  - Filtres Mels espaces de 150 Mel (largeur de bandes 300 Mels)
  - Utilisation des dérivées premières et secondes
  - Soit des vecteurs de 39 paramètres acoustiques

Lissage cepstral

- **Estimation de l'enveloppe par le cepstre:**
  - Calcul du cepstre réel Cn, puis lifrage basses quéfrences
  - Reconstruction de l'enveloppe spectrale d'amplitude E =FFT(Cn)



Quels paramètres aujourd'hui pour la reconnaissance avec DNN

- **MFCC toujours possible mais souvent remplacé par :**
  - Spectrogramme
  - Mel-spectrogramme (plusieurs fenêtres successives autour de la fenêtre courante comme entrée du DNN): le plus courant
  - Des « bancs de filtres perceptifs »
  - Un réseau spécifique pour des features discriminants
  - … voir le signal de parole brut *(mais pas encore aussi performant)*

Compléments

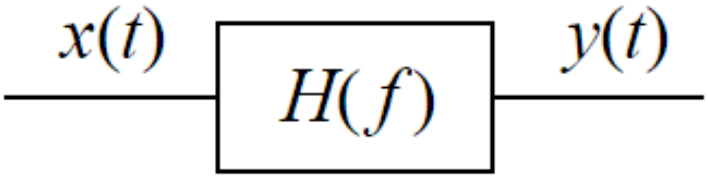
- **Pour en savoir plus:**
  - G. Blanchet, M. Charbit, « Signaux et images sous Matlab », Ed. Hermès, 2001
  - *(existe en anglais chez ISTE, 2006)*

Compléments

Quelques transparents supplémentaires pour le théorème d'échantillonnage

Système linéaire invariant dans le temps

- Soit *x(t)* u



$$x(t) \rightarrow y(t) = \int_{\mathbb{R}} x(u)h(t-u)du = \int_{\mathbb{R}} h(u)x(t-u)du = y(t) \star h(t)$$

$$\Leftrightarrow Y(f) = H(f)X(f)$$

Échantillonnage

- Soit *x(n)* la version échantillonnée de $x_a(t)$:

$$x(n) = x_a(nT)$$

- Peut-on reconstruire $x_a(t)$ à partir de x(n) ?

$$x_a(t) = \sum_n x(n)h(t - nT)$$

- En prenant la Transformée de Fourier

$$\begin{aligned} X_a(f) &= \int_{-\infty}^{\infty} \sum_n x(n)h(t - nT)e^{-2j\pi ft}dt \\ &= \sum_n x(n)e^{-2j\pi fnT}.H(f) \\ &= X_d(f).H(f) \end{aligned}$$

- où

$$X_d(f) = \sum_n x(n)e^{-2j\pi fnT}$$

Échantillonnage (2)

- **Or $X_d(f)$ est périodique:**

$$X_d(f) = X_d(f + 1/T)$$

- **Et est donc développable en série de Fourier**

$$X_d(f) = \sum_{-\infty}^{\infty} X_n e^{-2j\pi fnT} = \sum_{n} x(n) e^{-2j\pi fnT}$$

- **avec**

$$x(n) = \frac{1}{(1/T)} \int_{-1/2T}^{1/2T} X_d(f) e^{2j\pi fnT} df$$

Échantillonnage (3)

- **Or** $x_a(t) = \int_{-\infty}^{\infty} X_a(f) e^{2j\pi ft} df$

- **Posons t=nT**

$$\begin{aligned} x_a(nT) &= \int_{-\infty}^{\infty} X_a(f) e^{2j\pi fnT} df \\ &= \sum_{m=-\infty}^{\infty} \int_{\frac{2m-1}{2T}}^{\frac{2m+1}{2T}} X_a(f) e^{2j\pi fnT} df \end{aligned}$$

- **posons** $\nu = f - m/T$

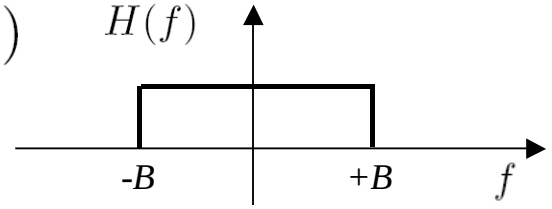
$$\begin{aligned} x_a(nT) &= \sum_{m=-\infty}^{\infty} \int_{-1/2T}^{1/2T} X_a(\nu + m/T) e^{2j\pi\nu nT} d\nu \\ &= \int_{-1/2T}^{1/2T} \sum_{m=-\infty}^{\infty} X_a(\nu + m/T) e^{2j\pi\nu nT} d\nu \end{aligned}$$

$$\Rightarrow \quad X_d(f) = \frac{1}{T} \sum_{m=-\infty}^{\infty} X_a(f + m/T)$$

Reconstruction (2)

- **Sans perte d'information possible uniquement si** $\frac{1}{T} > 2B$
- **En choisissant**

$$H(f) = Trect_{2B}(f)$$



$$H(f) = Trect_{2B}(f) \quad \rightleftharpoons \quad h(t) = T.\frac{\sin(2\pi Bt)}{\pi t}$$

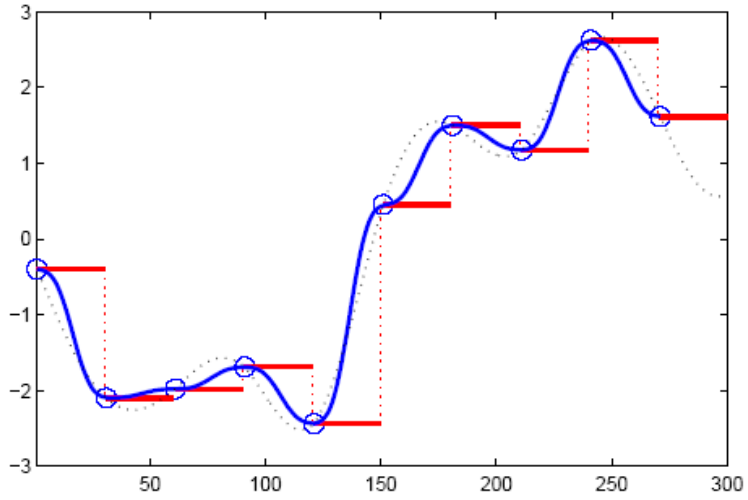
- **Formule de reconstruction**

$$x_a(t) = T \sum_{n=-\infty}^{\infty} x_a(nT) \frac{\sin(2\pi B(t-nT))}{\pi(t-nT)}$$

Reconstruction pratique

- **Bloqueur d'ordre zéro**

$$x_0(t) = \sum_{k=-\infty}^{+\infty} x_a(kT_e)h(t-kT_e) \quad \text{où} \quad h(t) = \text{rect}_{(0,T_e)}(t)$$



$$X_0(F) = \frac{H(F)}{T_e} \sum^{+\infty} X_a(F - \frac{n}{T_e}) \quad \text{où} \quad H(F) = \frac{\sin(\pi F T_e)}{\pi F} e^{-j\pi F T_e}$$