Codage Audio et Normes

Jean-Louis Durrieu, Télécom ParisTech

April 17, 2008

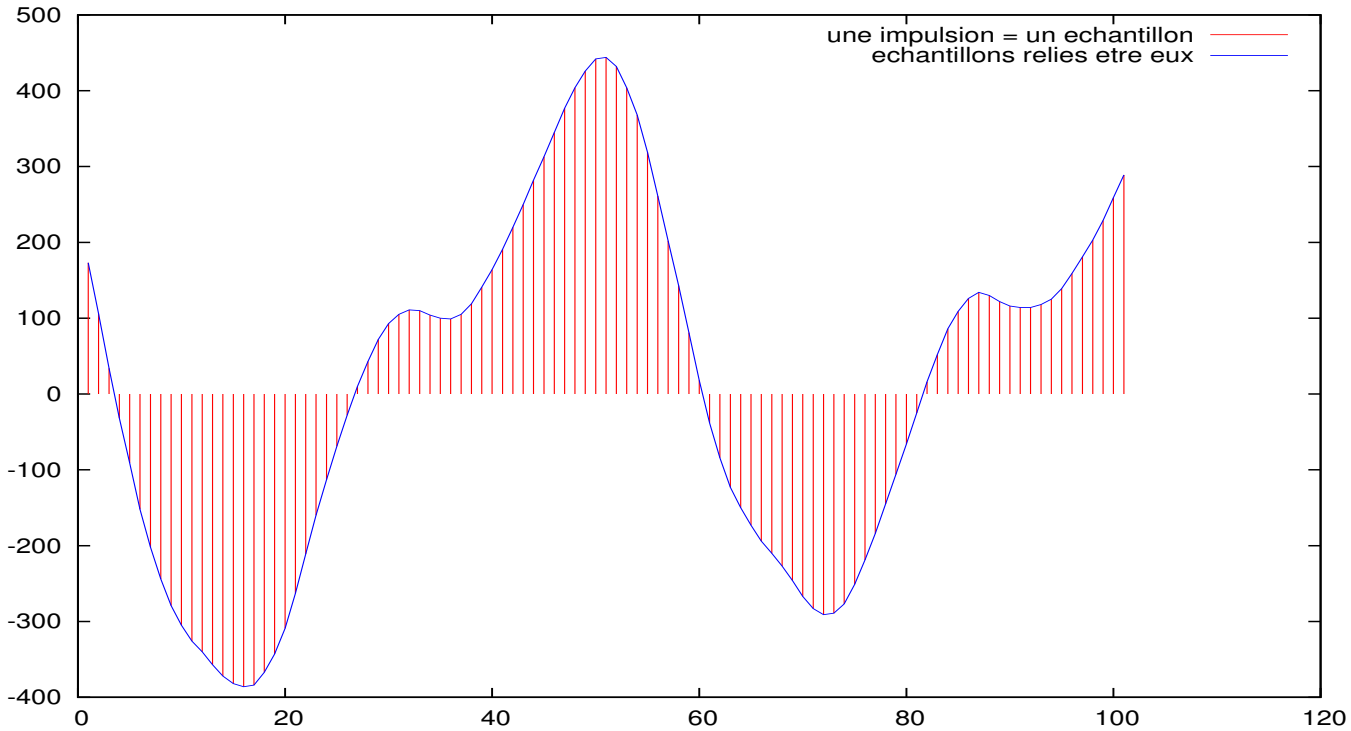
Plan

Plan

Objectifs (1/2) :

Codage des signaux de musique, introduction aux normes

- Signal d'entrée $x(n)$: signal audio **numérique** (ex : format PCM - Pulse Code Modulation, comme le WAV), musique



Signal de musique au format WAV.

Objectifs (2/2) :

- Signal de sortie (du codeur) $y(n)$: représentation **réduite** de $x(n)$, où l'on veut un débit $<$ débit nominal et $<$ débit donné
- Signal **reconstruit** (au décodeur) $\hat{x}(n)$: aussi "semblable" que possible à $x(n)$
- Exemples principaux issus des **normes** ISO MPEG 1 et 2 (MP3 et AAC).

Applications : nécessité du codage

	F_e	R	Déb. nom.	Déb. us.	compression
Parole					
Bande téléphonique	8	13	104	4 - 64	26 - 1.6
Bande élargie	16	14	224	16 - 64	14 - 3.5
Musique					
Qualité “FM”	16	16	512	64 - 192	8 - 2.6
Qualité “CD” stéréo	**44.1**	**16**	. . .	56 - 192	. . .
“Transparence”	96	24	13824	1000	13.8

F_e fréquence d’échantillonnage (kHz), R le taux de codage en bits/échantillon, débit nominal et usuel en kbits/s.

Applications : nécessité du codage

	F_e	R	Déb. nom.	Déb. us.	compression
Parole					
Bande téléphonique	8	13	104	4 - 64	26 - 1.6
Bande élargie	16	14	224	16 - 64	14 - 3.5
Musique					
Qualité “FM”	16	16	512	64 - 192	8 - 2.6
Qualité “CD” stéréo	**44.1**	**16**	**1411**	56 - 192	12 - 3.6
“Transparence”	96	24	13824	1000	13.8

F_e fréquence d’échantillonnage (kHz), R le taux de codage en bits/échantillon, débit nominal et usuel en kbits/s.

Un “bon” codeur : résoudre des compromis

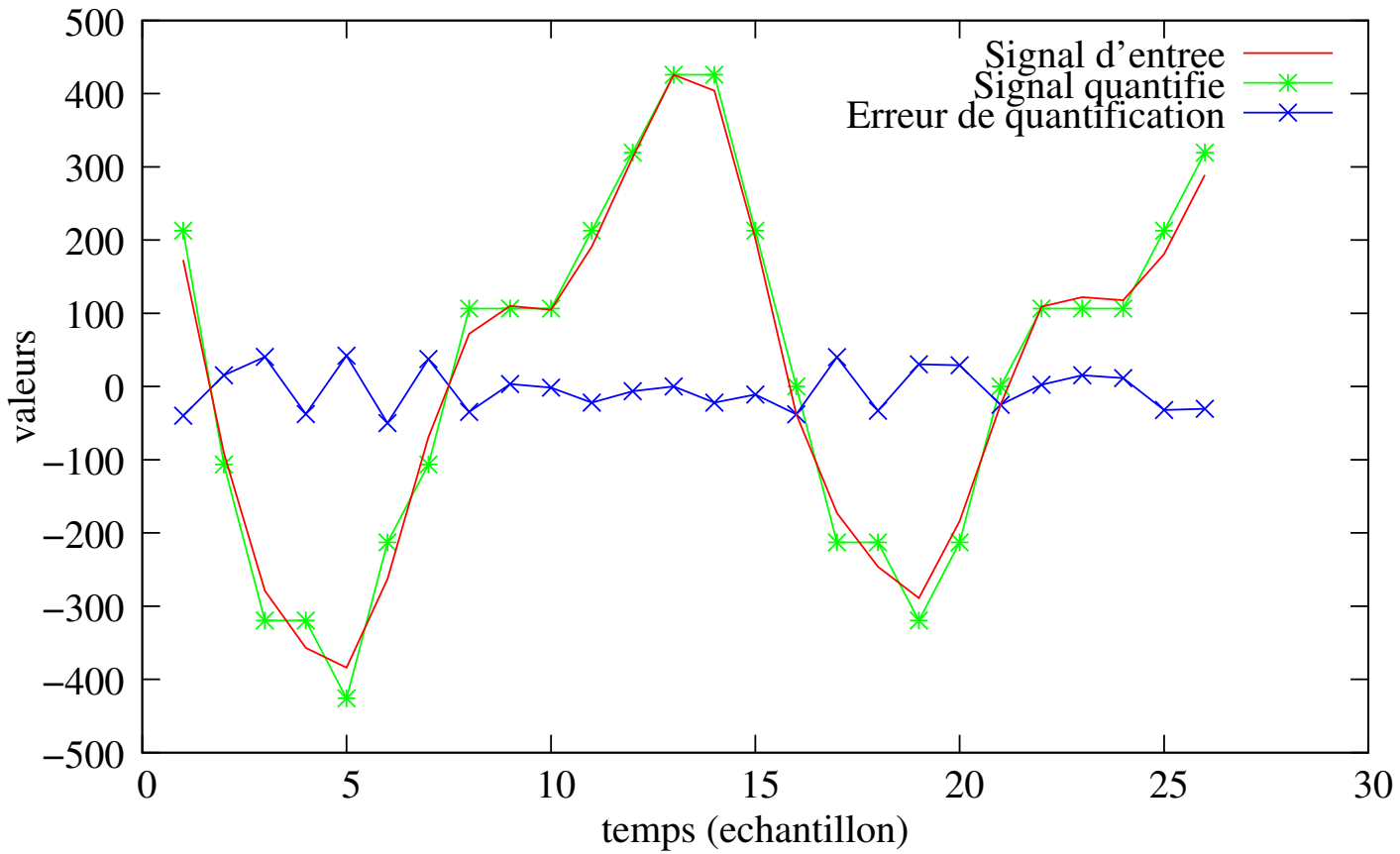
- **Débit**, selon l’application
- **Complexité**
  - coût et puissance consommée, calcul en MIPS
- **Retard de reconstruction**
  - essentiellement pour applications téléphoniques
  - $<$ 150ms, perte d’interactivité au-dessus de 400 ms
- **Erreurs de reconstruction**
  - pour les mobiles : codes correcteurs d’erreur
  - voix sur IP : pertes de trames
- **Qualité**
  - tests subjectifs
  - dépend du type de signal

Nécessité de la normalisation

- **Avantages** ...
  - ... pour le consommateur : pas de limitation dans l'acquisition de morceaux pour son appareil d'écoute de musique;
  - ... pour les développeurs et constructeurs de codeurs/décodeurs : interopérabilité;
- **Inconvénients** ... pour les constructeurs : pas de verrouillage technologique possible;
- Existence de **formats propriétaires** : AC3 (ou Dolby Digital), Mini-Disc (Sony)...

Plan

Quantification scalaire : exemple



Quantification scalaire

Soit $x(n)$ un signal à temps discret et à valeurs dans $[-A, +A]$,

- **Quantificateur scalaire** sur b bits/éch. :
  - une **partition** de $[-A, +A]$ en $L = 2^b$ intervalles $\{\Theta_1 \dots \Theta_L\}$, de longueur $\{\Delta_1 \dots \Delta_L\}$,
  - **numérotation** des éléments de la partition $\{i_1 \dots i_L\}$
  - choix d'un **représentant** pour chaque Θ_l, définissant un dictionnaire $C = \{\hat{x}_1 \dots \hat{x}_L\}$.
- **Quantification** du signal d'entrée :
  - $i(n) \in [1, L]$ tel que $x(n) \in \Theta_{i(n)}$,
  - transmission de $i(n)$.
  - transformation à perte pour $L < K = \text{card}([\text{-A,+A}])$
- **Décodage** (reconstruction) du signal : opération de quantification "inverse"
  - $\hat{x}(n) = \hat{x}_{i(n)} \in \{\hat{x}_1 \dots \hat{x}_L\}$
  - Mesure de la distorsion, par ex. puissance de l'erreur de quantification = erreur quadratique moyenne (EQM) : $\sigma_Q^2 = \mathrm{E}[|X - \hat{X}|^2]$

Quantification scalaire

Soit $x(n)$ un signal à temps discret et à valeurs dans $[-A, +A]$,

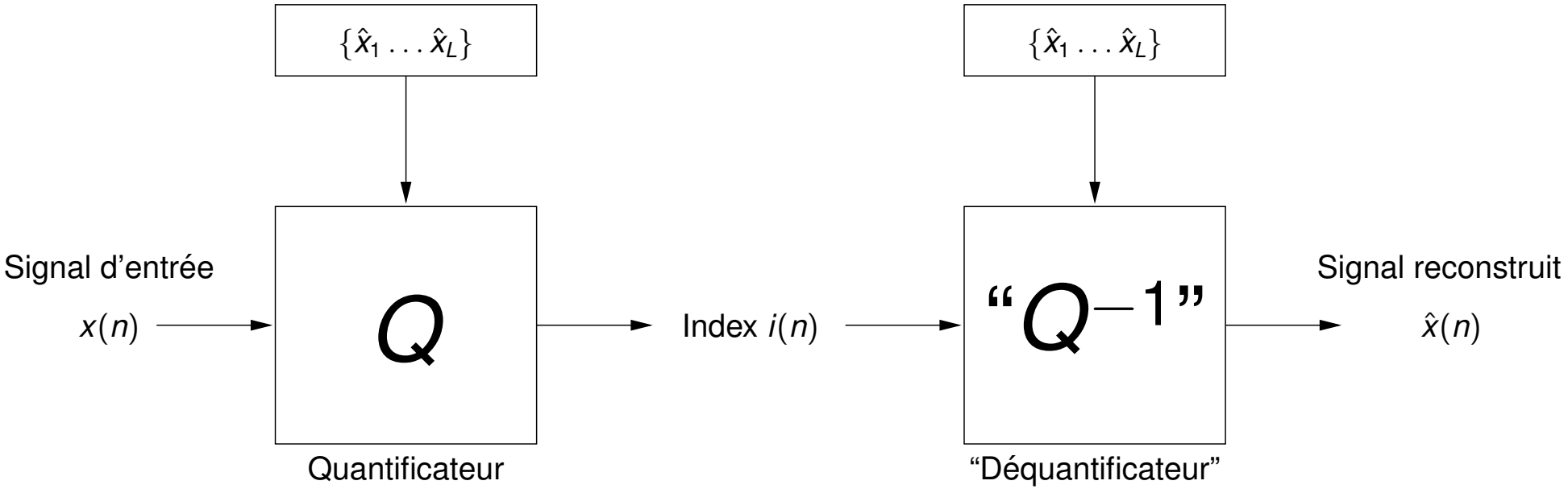
- **Quantificateur scalaire** sur b bits/éch. :
  - une **partition** de $[-A, +A]$ en $L = 2^b$ intervalles $\{\Theta_1 \dots \Theta_L\}$, de longueur $\{\Delta_1 \dots \Delta_L\}$,
  - **numérotation** des éléments de la partition $\{i_1 \dots i_L\}$
  - choix d'un **représentant** pour chaque Θ_l, définissant un dictionnaire $C = \{\hat{x}_1 \dots \hat{x}_L\}$.
- **Quantification** du signal d'entrée :
  - $i(n) \in [1, L]$ tel que $x(n) \in \Theta_{i(n)}$,
  - transmission de $i(n)$.
  - transformation à perte pour $L < K = \text{card}([\text{-A,+A}])$
- **Décodage** (reconstruction) du signal : opération de quantification "inverse"
  - $\hat{x}(n) = \hat{x}_{i(n)} \in \{\hat{x}_1 \dots \hat{x}_L\}$
  - Mesure de la distorsion, par ex. puissance de l'erreur de quantification = erreur quadratique moyenne (EQM) : $\sigma_Q^2 = \mathrm{E}[|X - \hat{X}|^2]$

Quantification scalaire

Soit $x(n)$ un signal à temps discret et à valeurs dans $[-A, +A]$,

- **Quantificateur scalaire** sur b bits/éch. :
  - une **partition** de $[-A, +A]$ en $L = 2^b$ intervalles $\{\Theta_1 \dots \Theta_L\}$, de longueur $\{\Delta_1 \dots \Delta_L\}$,
  - **numérotation** des éléments de la partition $\{i_1 \dots i_L\}$
  - choix d'un **représentant** pour chaque Θ_l, définissant un dictionnaire $C = \{\hat{x}_1 \dots \hat{x}_L\}$.
- **Quantification** du signal d'entrée :
  - $i(n) \in [1, L]$ tel que $x(n) \in \Theta_{i(n)}$,
  - transmission de $i(n)$.
  - transformation à perte pour $L < K = \text{card}([\text{-A,+A}])$
- **Décodage** (reconstruction) du signal : opération de quantification "inverse"
  - $\hat{x}(n) = \hat{x}_{i(n)} \in \{\hat{x}_1 \dots \hat{x}_L\}$
  - Mesure de la distorsion, par ex. puissance de l'erreur de quantification $=$ erreur quadratique moyenne (EQM) : $\sigma_Q^2 = \mathrm{E}[|X - \hat{X}|^2]$

Quantification scalaire : schéma de principe



Quantification vectorielle, caractérisation et critères

Généralisation du cas scalaire : définitions identiques, avec $\mathbf{x}(n)$ vecteur.

- $N =$ longueur des vecteurs $\mathbf{x}(n) \in [-A, +A]^N$, $N = 1 \Rightarrow$ cas scalaire,
- **Taux de codage** $R = \log_2(L)/N$, bit/éch.
- **Débit** $B = RF_e$, bit/s
- **Facteur de compression** $\tau = \log_2(K)/R$, où $K = \text{card}([-A, +A])$
- **Erreur de quantification** : $q(n) = x(n) - \hat{x}(n)$
- **Mesure de distorsion** :

$$D = \frac{1}{N}\mathrm{E}[|Q|^2] = \frac{1}{N}\mathrm{E}[|X - \hat{X}|^2]$$

$$= \frac{1}{N}\int_{\mathbb{R}^N} ||\mathbf{x} - \hat{\mathbf{x}}||^2 \mathrm{p}(\mathbf{x})d\mathbf{x} = \frac{1}{N}\sum_{i=1}^{L}\int_{\Theta_i} ||\mathbf{x} - \hat{\mathbf{x}}_i||^2 \mathrm{p}(\mathbf{x})d\mathbf{x}$$

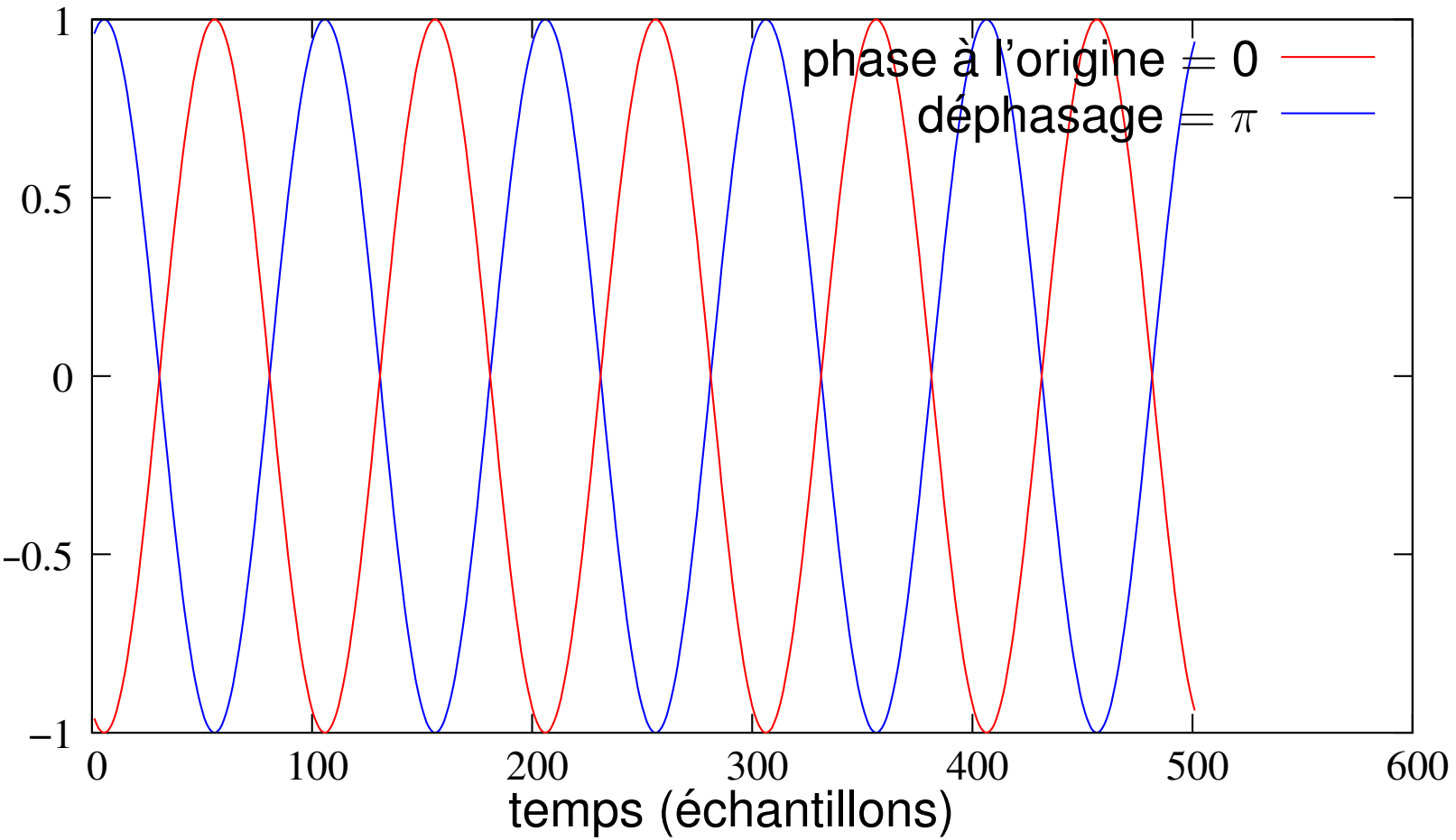
Quantifications scalaire et vectorielle

- Gains possibles avec **prédiction linéaire** (linear prediction coding - LPC)
- Minimiser EQM : perceptuellement peu significatif (voir la suite)
- plus de détails : poly de cours de N. Moreau (http://www.tsi.enst.fr/ moreau/enseignement.html)

Nécessité d'une mesure de distorsion **perceptuelle**

Deux cosinus avec un déphasage de π

EQM $= 2$, perçues de la même façon!

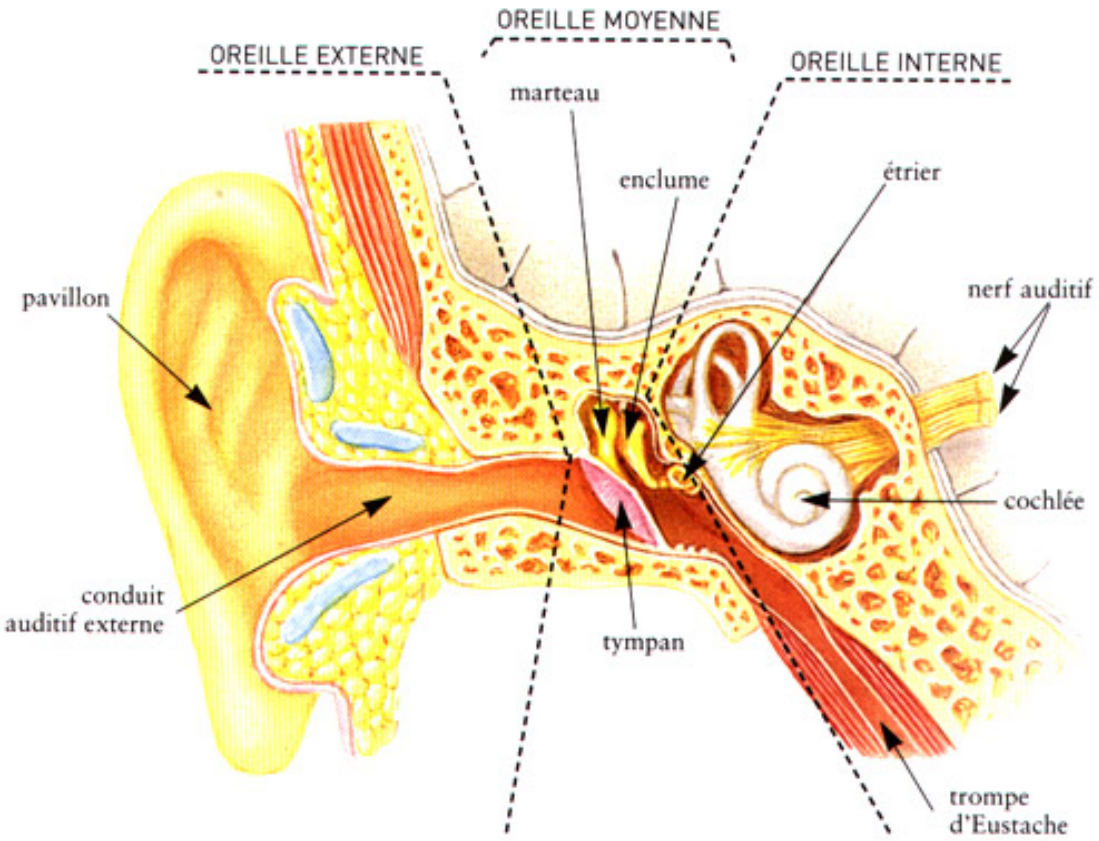


Niveau de pression acoustique

- Son : **onde acoustique**, caractérisée par la pression de l'air,
- **Modèle acoustique** : $P(t) = H_0 + p(t)$ où
  - $P(t)$ la pression instantanée
  - H_0 la pression atmosphérique (pression moyenne de l'air 10^5 Pa (N/m^2))
  - $p(t)$ vibrations
- $p_r = 2 \times 10^{-5}$ Pa, **pression de référence** $\sim$ pression minimale audible (pour un sujet moyen, son de référence à 1 kHz)
- Niveau de pression acoustique :
  - **SPL (Sound Pressure Level)** $= 20 \log_{10}(p/p_r)$ (en dB SPL)
  - 0 dB SPL $\sim$ silence, douleur $>$ 140 dB SPL.

Le système auditif : structure de l'oreille

- **Oreille externe** : amplification (dépendant de l'angle d'arrivée → spatialisation)
- **Oreille moyenne** : adaptation d'impédance entre l'air et le liquide
- **Oreille interne** : sensibilité aux fréquences → domaine fréquentiel adapté à un travail sur la perception



Tonotopie passive et pattern d'excitation

Tonotopie passive : localisation fréquentielle sur MB

- 2 hypothèses pour expliquer la *tonotopie passive* : la **résonance** (des cellules cillées CC) ou l'**onde propagée** (le long de la membrane basilaire MB)
- sur membrane basilaire : **pattern d'excitation** → masquage fréquentiel

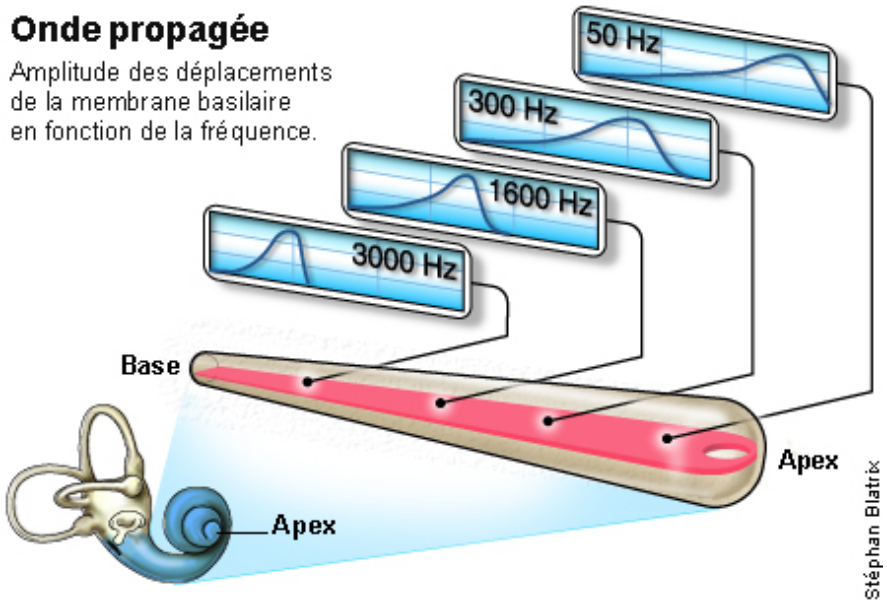


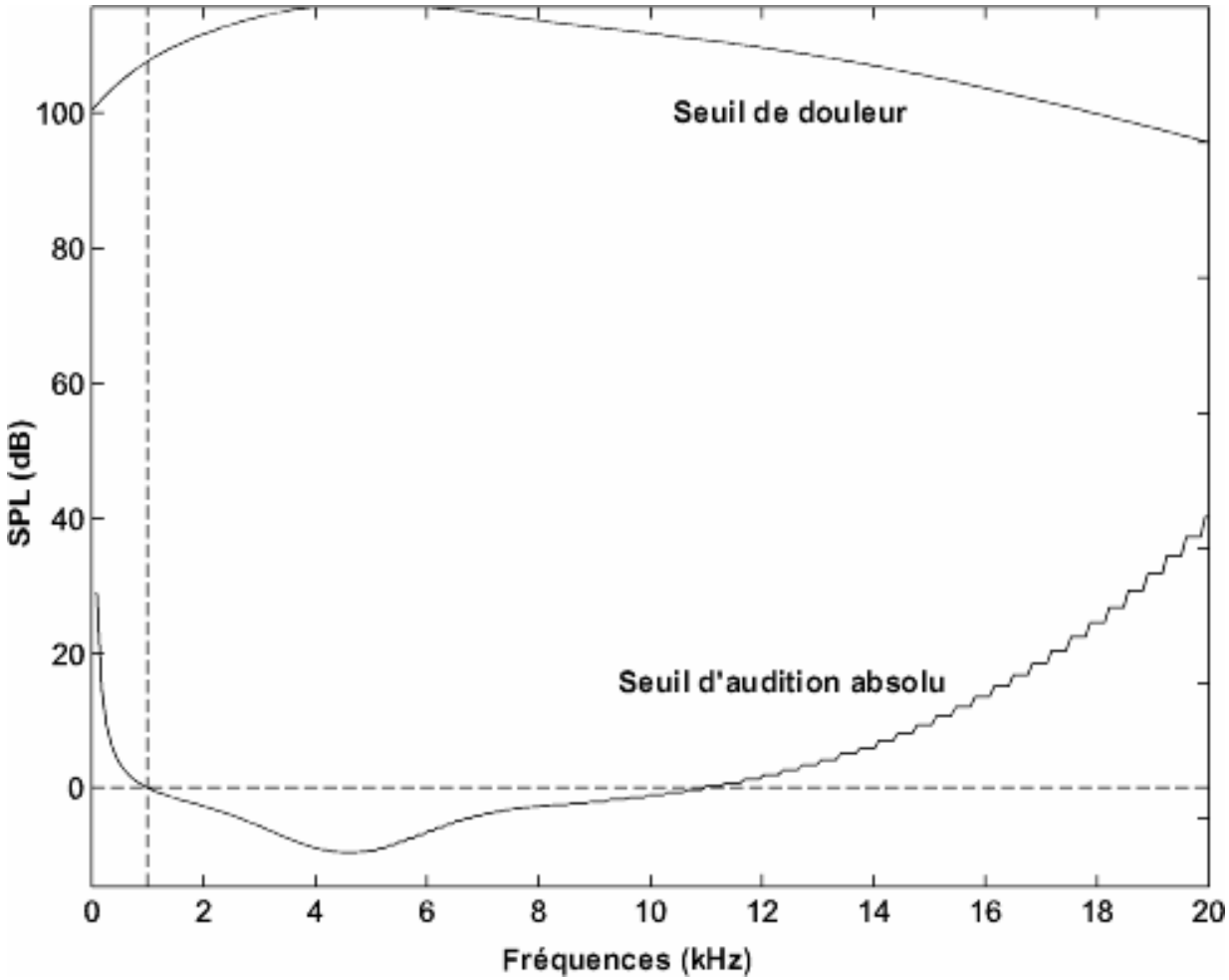
Image de A. Blatrix, issue de "Promenade autour de la cochlée" (http://www.cochlee.info), Rémy Pujol et al., Montpellier

Notions de psychoacoustique

- **Etude de la perception auditive chez l'homme**,
- Caractérisation,
- Analyse **temps-fréquence** des capacités de l'oreille humaine,
- **Relation** entre grandeurs physiques et grandeurs perceptuelles,
- Protocoles d'expérimentation pour **tests psychoacoustiques**.

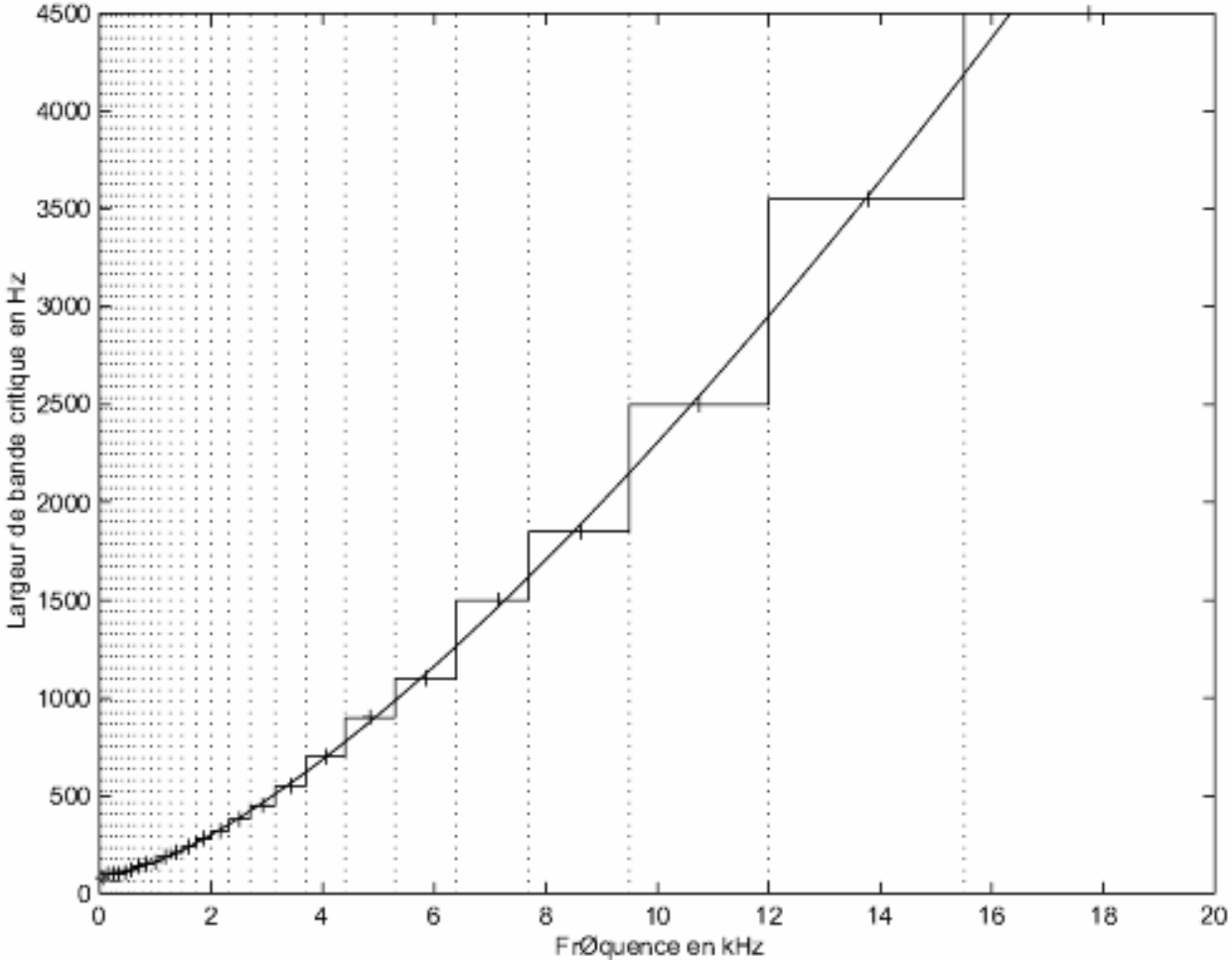
Seuil d'audition absolu

- SPL minimal pour entendre un son, dépend de la fréquence



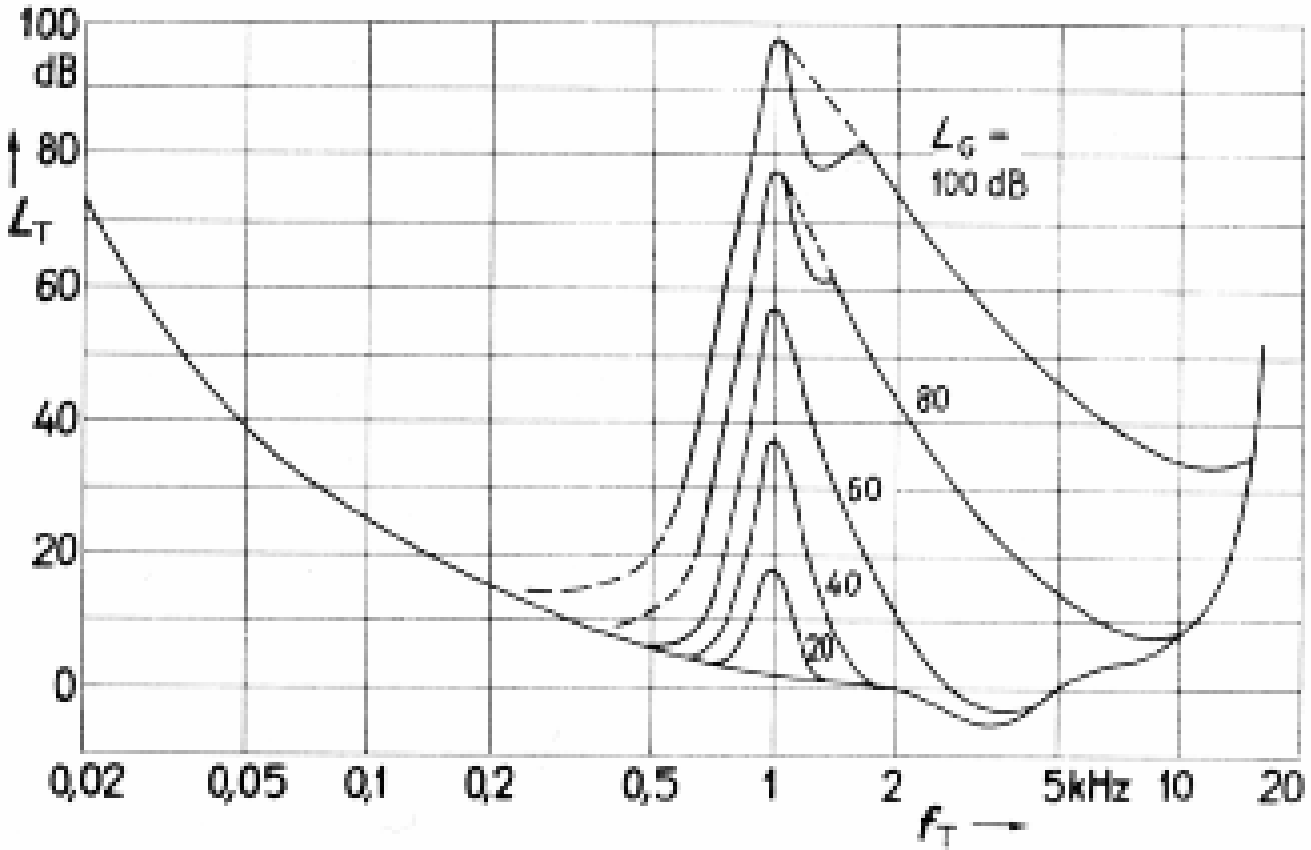
Bandes critiques

- bandes de fréquences sur lesquelles l'oreille "intègre" l'information



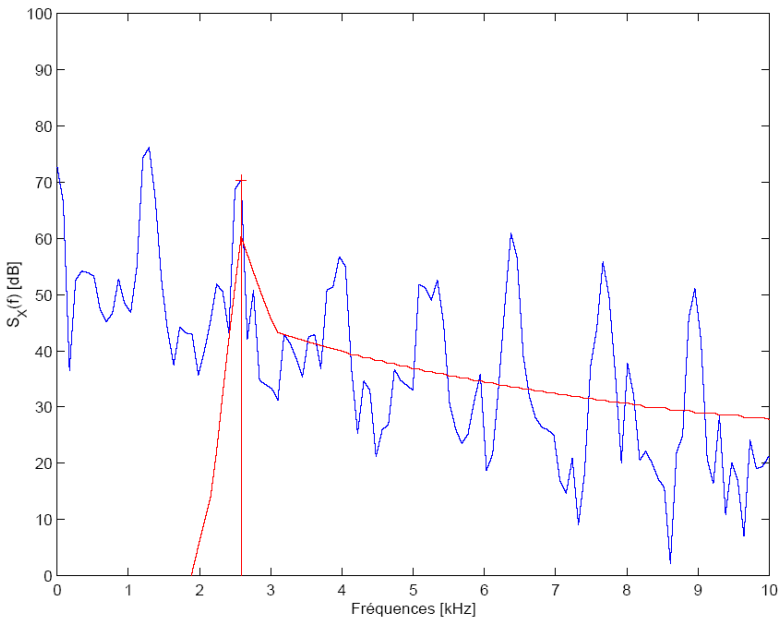
Masquage en fréquence : courbe de masquage

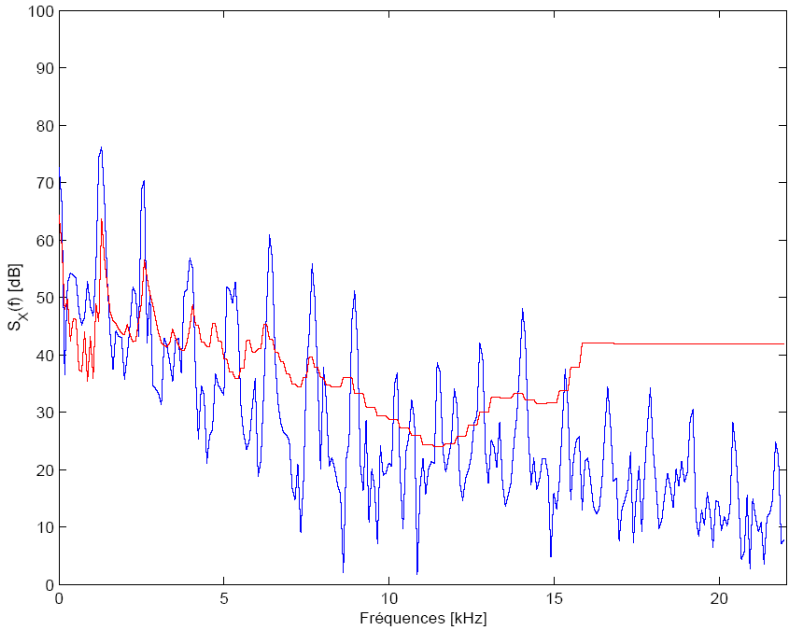
- Niveau SPL nécessaire à un son “masqué” pour être entendu en présence d’un “masquant” donné (fonction de la fréquence)



Masquage en fréquence : calcul de masque

- Généralisation à plusieurs masquants et plusieurs masqués :



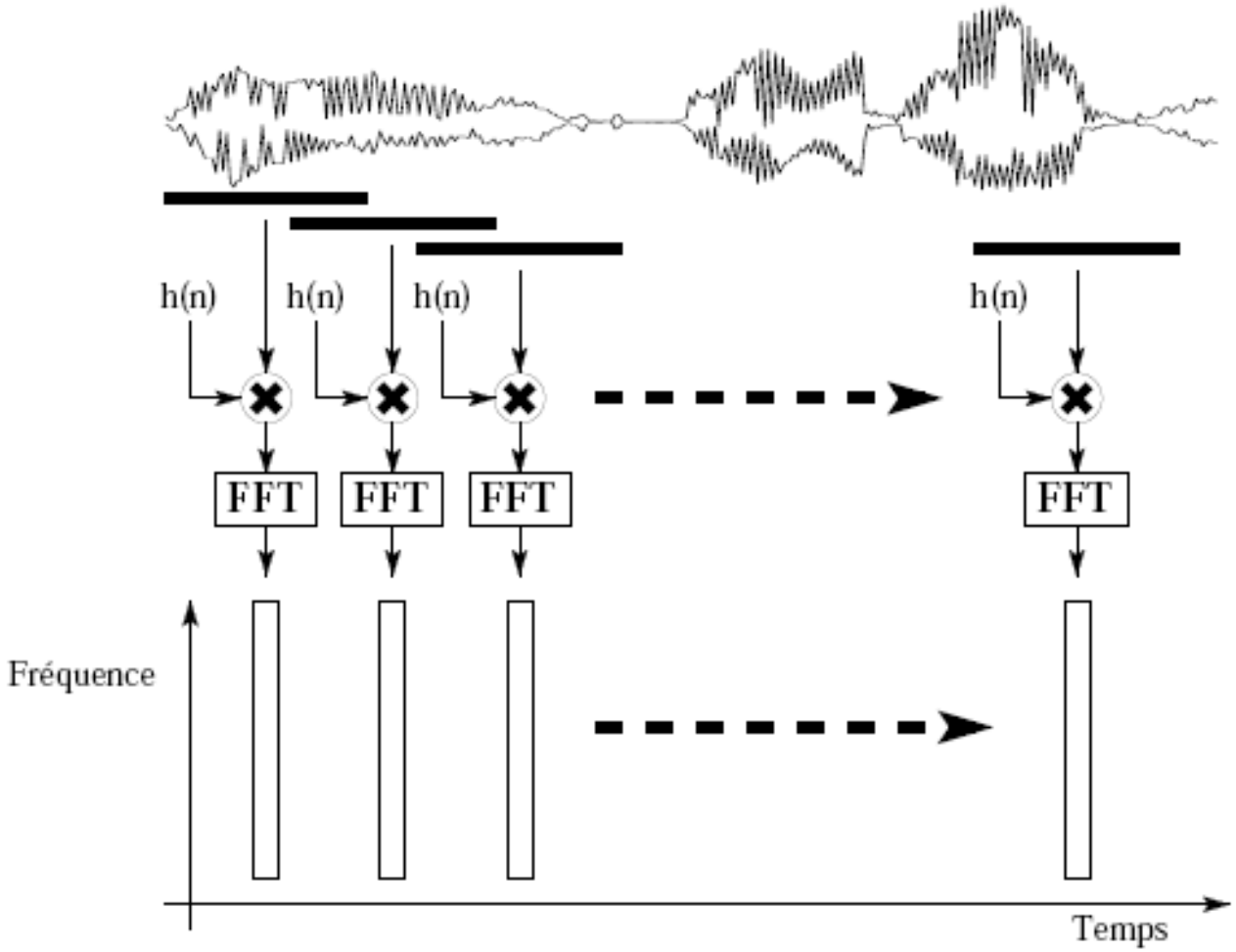


Masquage en fréquence : conjectures

Calcul de masque et principe d'allocation de bits

- **Double généralisation** à plusieurs masquants et plusieurs masqués : correctes, même si impossible d'effectuer des tests subjectifs directs;
- **Erreur inaudible** si $\sigma_Q^2(f) < \phi(f)$, avec $\phi(f)$ masque et $\sigma_Q^2(f)$ puissance de l'erreur de quantification dans la bande centrée sur f;
- **Allocation de bits** dans les bandes critiques : rendre l'erreur de quantification $<$ masque (souvent algorithme itératif, sous-optimal)

Transformée de Fourier à court terme (TFCT) (1/3)

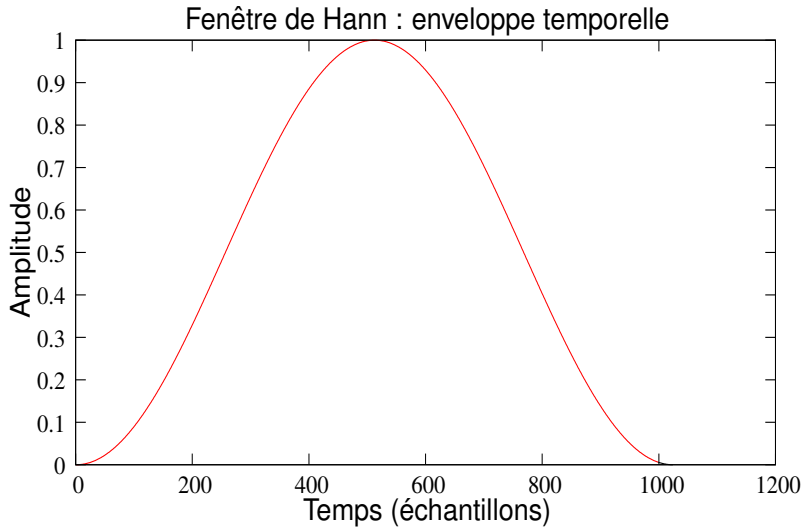


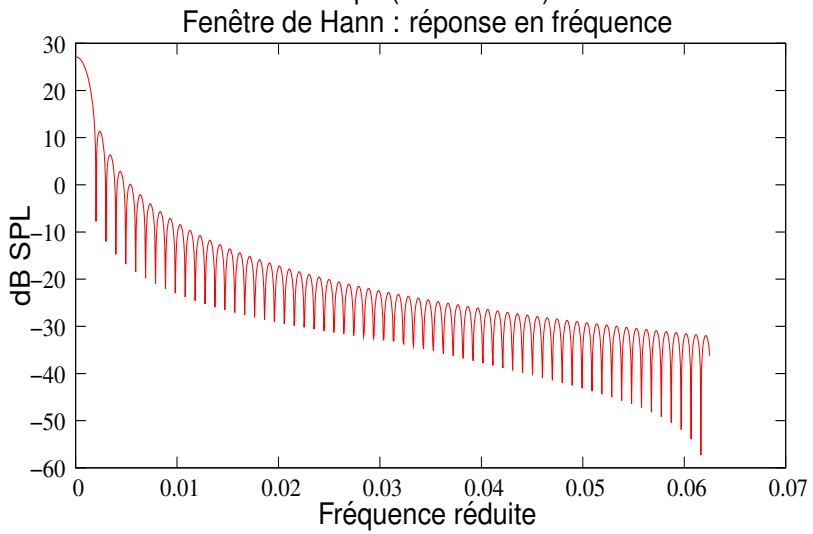
Transformée de Fourier à court terme (TFCT) (2/3)

- **Version discrète de la TFCT**, au décalage temporel t_a, bande de fréquence centrée sur $\nu_p = \frac{p}{N}$ (fréquence réduite) :

$$X(t_a, \nu_p) = \sum_{n=0}^{N-1} x(n + t_a) w_a(n) e^{-j2\pi \frac{pn}{N}}$$

- **Fenêtre d'analyse** $w_a(n)$: propriétés spectrales (Hann, Hamming, ...)
- Equivalent à une opération de filtrage : $X(t_a, \nu_p) = \sum_{n=0}^{N-1} x(n) h(t_a - n)$ avec $h(n) = \ldots$?



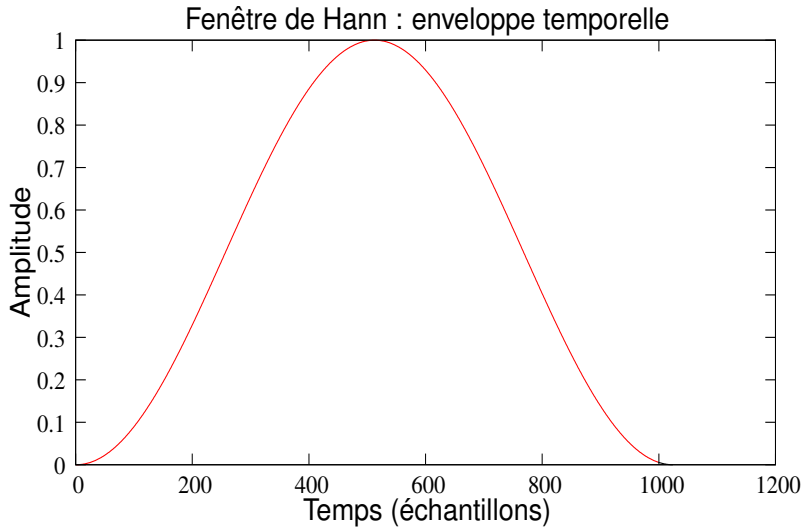


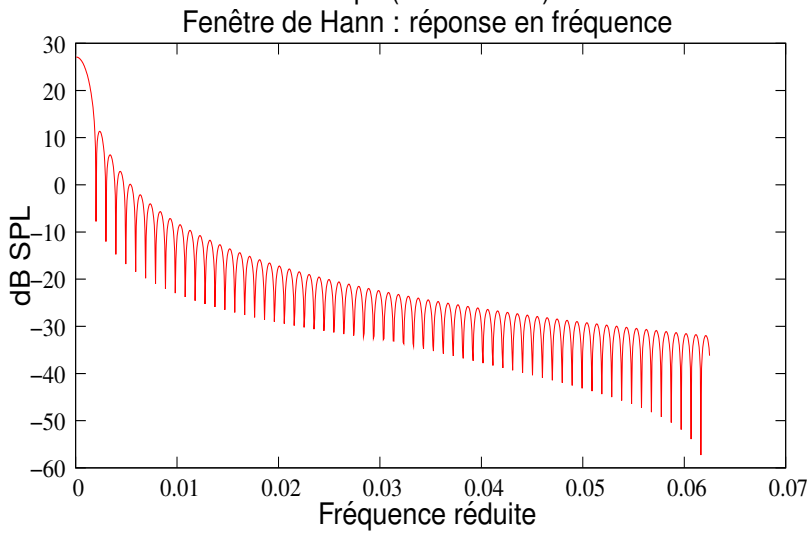
Transformée de Fourier à court terme (TFCT) (2/3)

- **Version discrète de la TFCT**, au décalage temporel t_a, bande de fréquence centrée sur $\nu_p = \frac{p}{N}$ (fréquence réduite) :

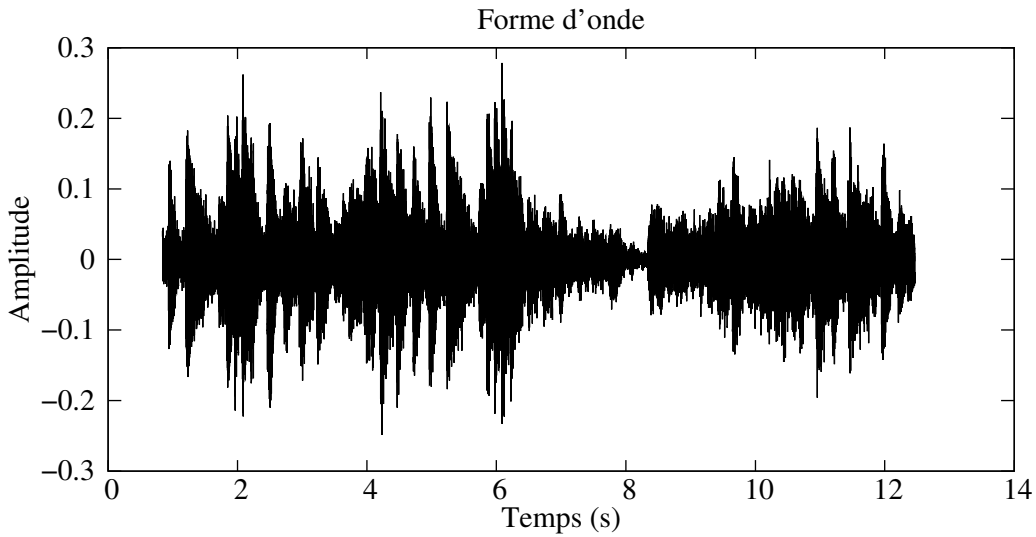
$$X(t_a, \nu_p) = \sum_{n=0}^{N-1} x(n + t_a) w_a(n) e^{-j2\pi \frac{pn}{N}}$$

- **Fenêtre d'analyse** $w_a(n)$: propriétés spectrales (Hann, Hamming, ...)
- Equivalent à une opération de filtrage : $X(t_a, \nu_p) = \sum_{n=0}^{N-1} x(n) h(t_a - n)$ avec $h(n) = w_a(-n) e^{j2\pi \frac{pn}{N}}$

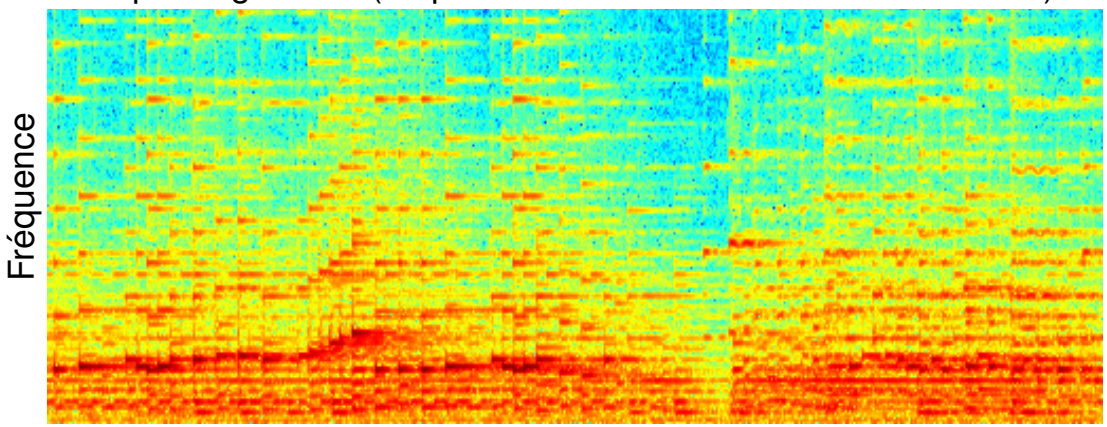




Transformée de Fourier à court terme (TFCT) (3/3)

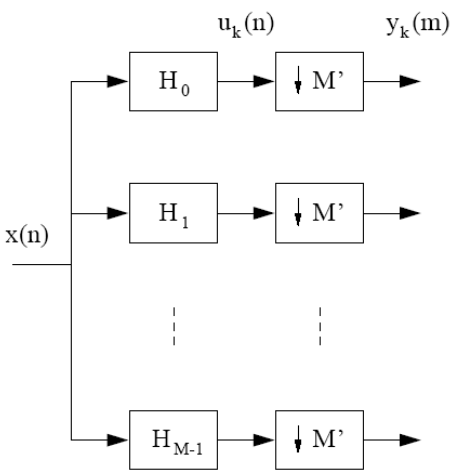


Spectrogramme (amplitude au carré de la TFCT en dB SPL)

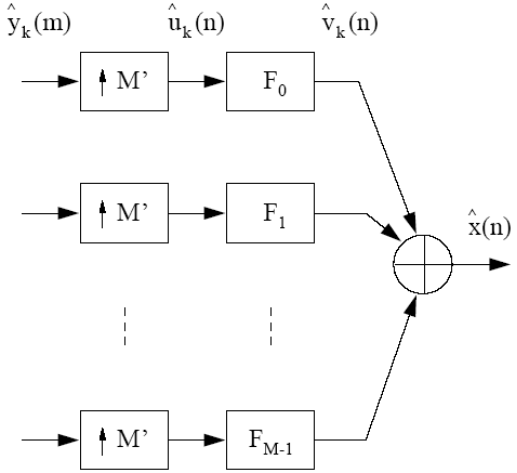


Banc de filtres :

- **Filtres passe-bande** : en général, filtre prototype passe-bas modulé
- **Décimation** dans chaque sous-bande
- Nombre de bandes : M; facteur de décimation : M'; **décimation critique** $M = M'$



Banc de filtres d'analyse



Banc de filtres de synthèse

Analyse et synthèse:

Transformée du signal :

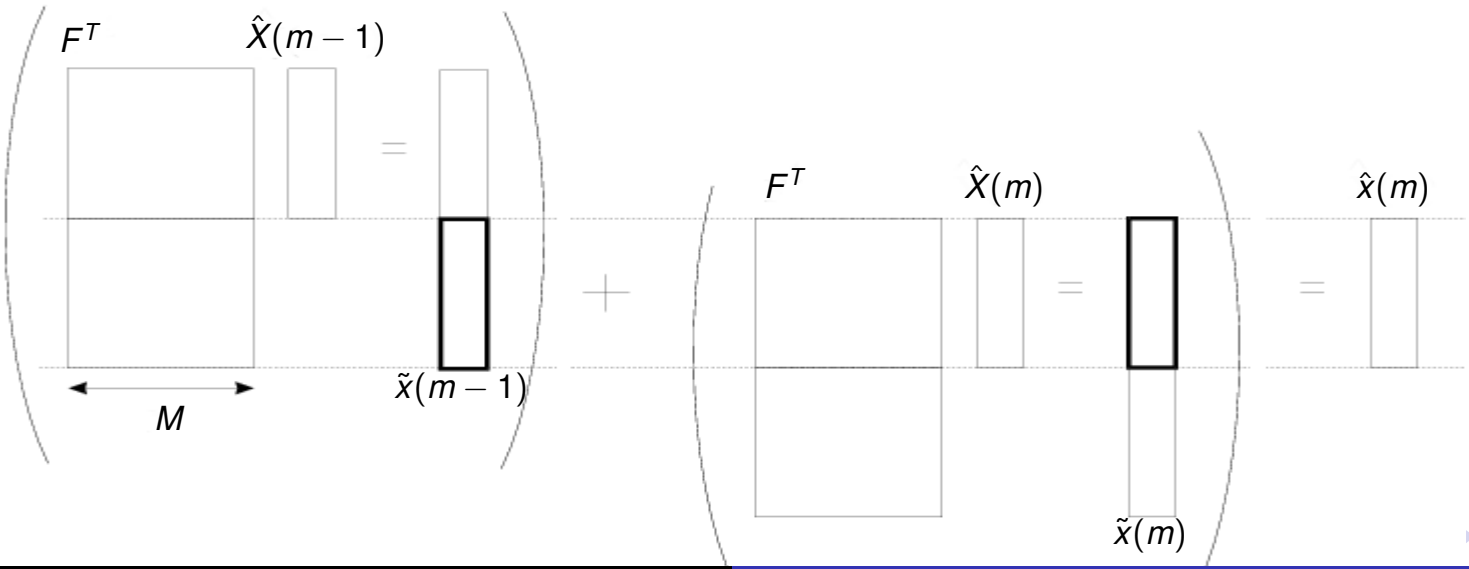
- $\mathbf{X}(m) = H\mathbf{x}(m)$ avec : $H = \begin{bmatrix} h_0(N-1) & \dots & h_0(0) \\ \vdots & \dots & \vdots \\ h_{M-1}(N-1) & \dots & h_{M-1}(0) \end{bmatrix}$,
  $\mathbf{x}(m) = [x(mM - N + 1) \dots x(mM)]^T$ et
  $\mathbf{X}(m) = [X_0(m) \dots X_{M-1}(m)]^T$
- **Recouvrement** des vecteurs $\mathbf{x}$: $N - M$ échantillons entre $\mathbf{x}(m)$ et $\mathbf{x}(m+1)$

Synthèse à partir de la transformée

- **Addition/Recouvrement** (OverLap-Add OLA)
- $\tilde{\mathbf{x}}(m)$ transformée inverse par opérateur F^T sur $\mathbf{X}(m)$:

$$F = \begin{bmatrix} f_0(N-1) & \dots & f_0(0) \\ \vdots & \dots & \vdots \\ f_{M-1}(N-1) & \dots & f_{M-1}(0) \end{bmatrix}$$

- addition des parties se recouvrant dans $\tilde{\mathbf{x}}(m-1)$ et $\tilde{\mathbf{x}}(m)$ (cas où $N = 2M$)



Paramètres de la transformée

Reconstruction parfaite

- Importance du **recouvrement** : $N > M$
- Conditions sur les filtres utilisés : banc de filtres modulés

Dimensionnement de N et M

- **Résolution fréquentielle** :
  - M grand $\rightarrow$ bonne résolution ($\Delta\nu \sim \frac{1}{M}$)
  - N grand $\rightarrow$ filtres sélectifs
- **Résolution temporelle** :
  - N petit $\rightarrow$ détection de variations temporelles rapides (transitoires)

Compromis : $N = 2M = 512$, 2048 ou 256...

Codage par transformée : allocation de bits

Cadre théorique

- **Transformée optimale** pour allocation de bits : transformée de Karhunen-Loève (KLT)
- Décomposition en **valeurs propres** de la matrice d'autocovariance du signal : complexité prohibitive

Choix pratique

- **Modified Discrete Cosine Transform (MDCT)**
- Perception humaine sensible aux **bandes critiques**
- Algorithmes itératifs d'**allocation de bits** : tant qu'il y a des bits à allouer :
  - Calcul du **masque** $\Phi(f)$ dans chaque sous-bande,
  - Déterminer la bande où l'erreur rapportée à $\Phi(f)$ est la plus grande et augmenter le nombre de bits pour cette sous-bande

Evaluation de la qualité (perceptuelle) d'un codeur audio

Tests subjectifs

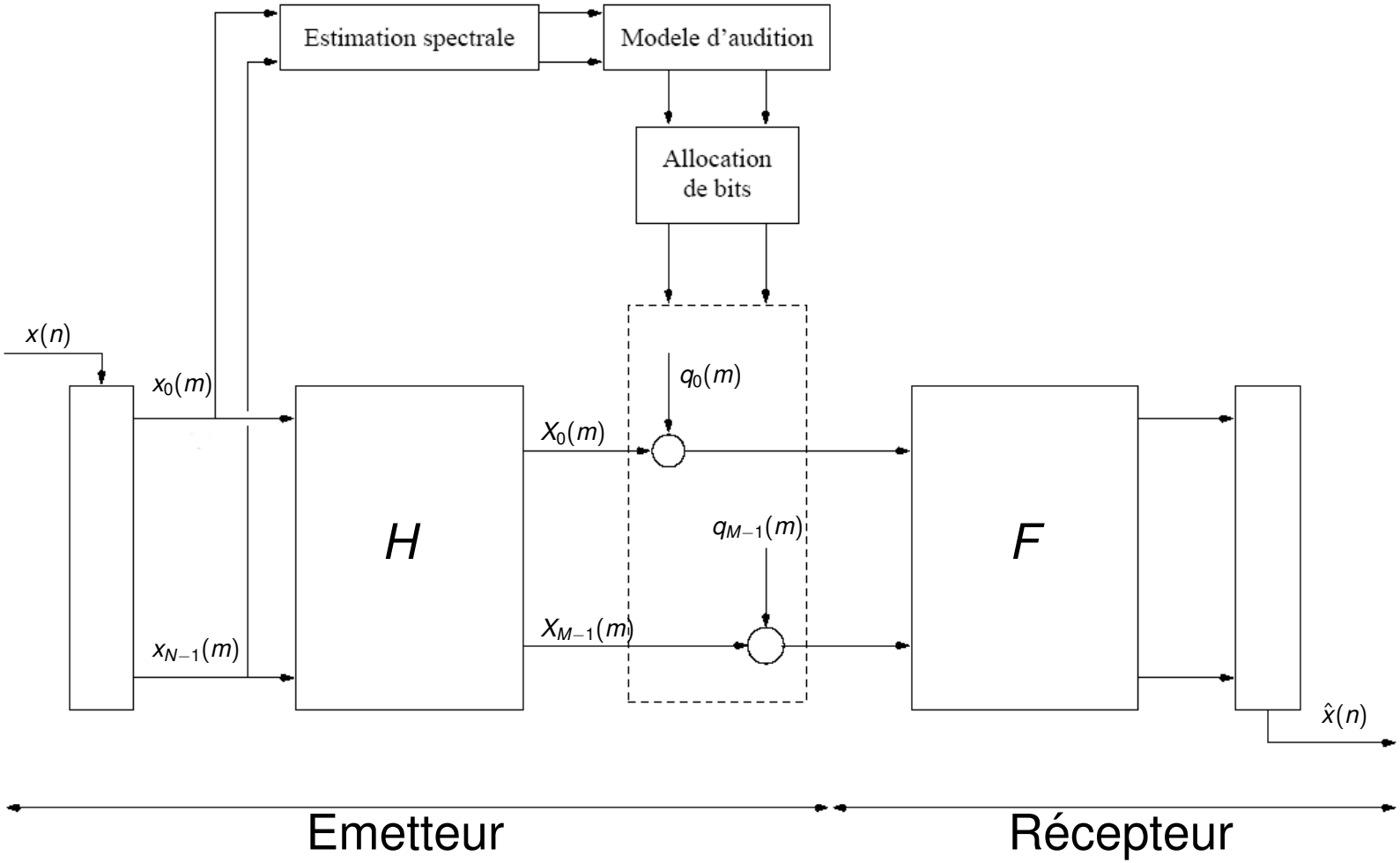
- Codeurs de parole : tests d'intelligibilité
- Codeurs audio, débits entre 20 et 64 kbits/s : qualité "acceptable", méthode **MUSHRA** (MUlti Stimulus with Hidden Reference and Anchor)
- Codeurs haute qualité : "transparence"

Recommandations UIT-R BS.1116 : "doublement aveugle à triple stimulus et référence dissimulée"

- Enregistrements courts (5 à 10s) répéptés 3 fois
- Deux possibilités : A B A ou A B B
- Réponses demandées :
  - B en 2ème ou 3ème position ?
  - Opinion sur B de 5 (bruit totalement inaudible) à 0 (très mauvais)

Plan

Codage par transformée, schéma général



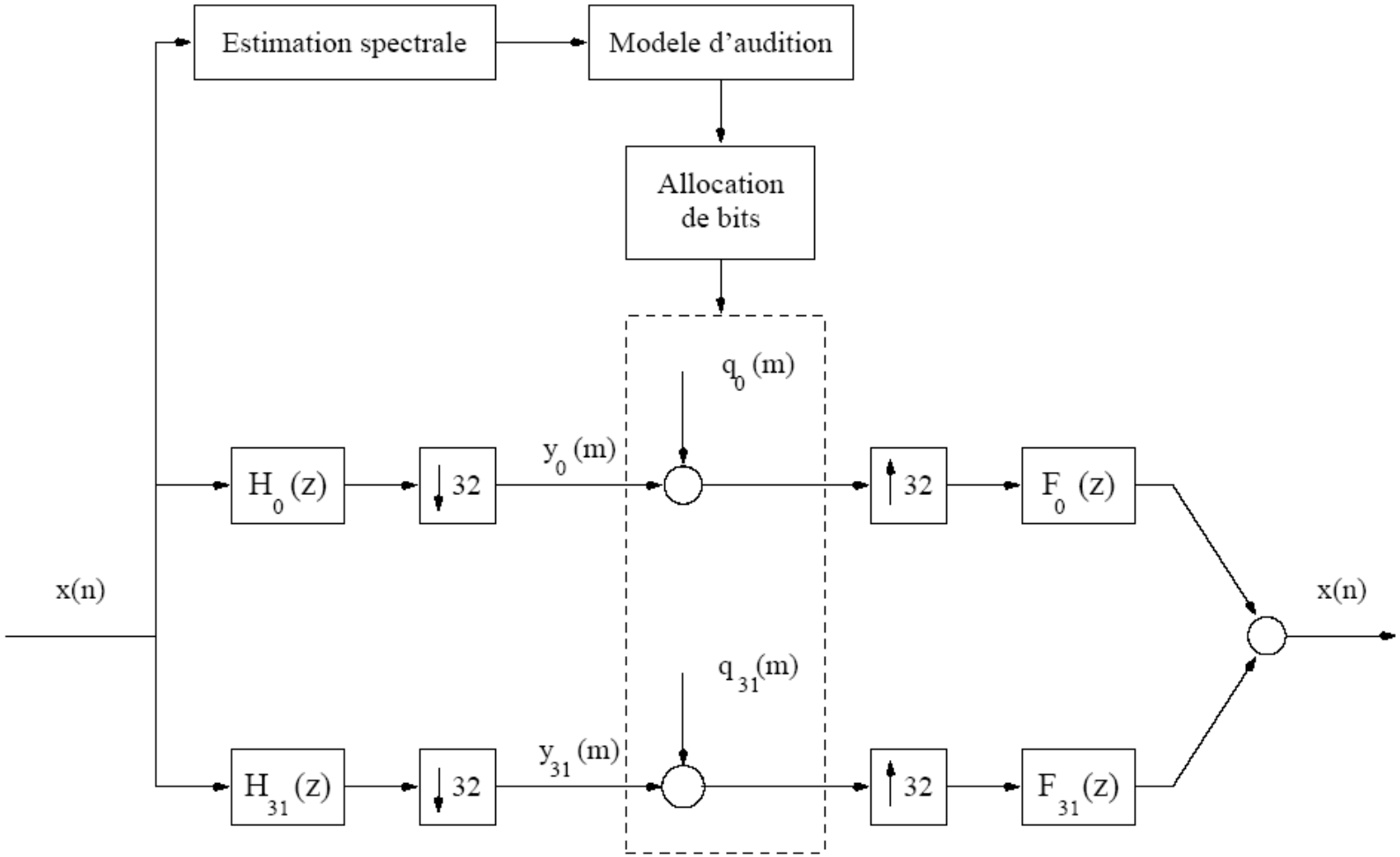
La norme MPEG-1 : historique

- **MPEG-1 (ISO/IEC 11172)** : Moving Picture Experts Group, “Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s”
  - 5 parties : *systems, video,* ***audio****, compliance testing, software simulation*
  - **Audio** : ISO/IEC 11172-3:1993 + correctifs en 1996
- **Normalisation du décodeur**, annexes informatives (1992)
  - Fréquences d’échantillonnage (44.1, 48 et 32kHz)
  - Modes : stéréophonique, stéréo combiné (exploitation des redondances entre les canaux), “dual monophonic” (2 canaux indépendants : programmes bilingues, ...)

MPEG-1 : 3 couches

- **Couche I**
  - Studio, diffusion par satellite,...
  - Débits : 32, 64, 96, . . ., **192**, . . ., 384,416,448 kbits/s
  - Retard minimum théorique de codage/décodage : 19 ms
- **Couche II**
  - DAB (Digital Audio Broadcasting) en Europe
  - Débits : 32, 48,56, . . ., **128**, . . ., 256,320,384 kbits/s
  - Retard minimum théorique de codage/décodage : 35 ms
- **Couche III**
  - mp3 = MPEG-1 couche III, stockage de données
  - Débits : 32, 40, 48 . . ., **96**, . . ., 224,256,320 kbits/s
  - Retard minimum théorique de codage/décodage : 59 ms

MPEG-1 couche I : Schéma du codeur/décodeur



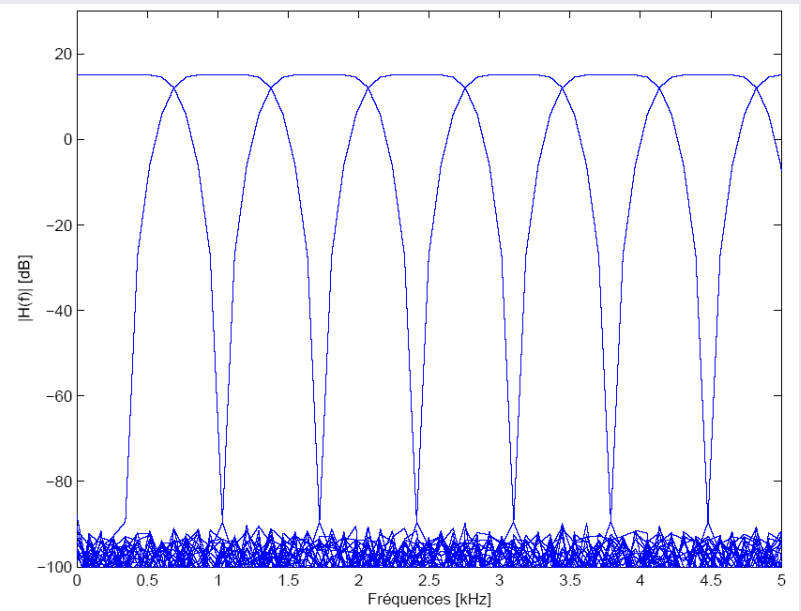
MPEG-1 couche I : Transformée utilisée

MPEG-1 couche I : Banc de filtres

- **Banc de $M = 32$ filtres modulés** (exploitation d'un filtre prototype) : $H_k(f) = H(f - \frac{2k+1}{4M}) + H(f + \frac{2k+1}{4M})$,

$$h_k(n) = 2h(n)cos(2\pi\frac{2k+1}{4M}n + \phi_k),\ n = 0 \ldots N-1$$

- Longueur du **filtre prototype** $N = 512$
- **Sous-échantillonnage critique** : facteur $M' = 32$
- **Pas de reconstruction parfaite**, mais RSB $>$ 90dB



MPEG-1 couche I : codeur

Codeur pour la norme MPEG-1 couche I

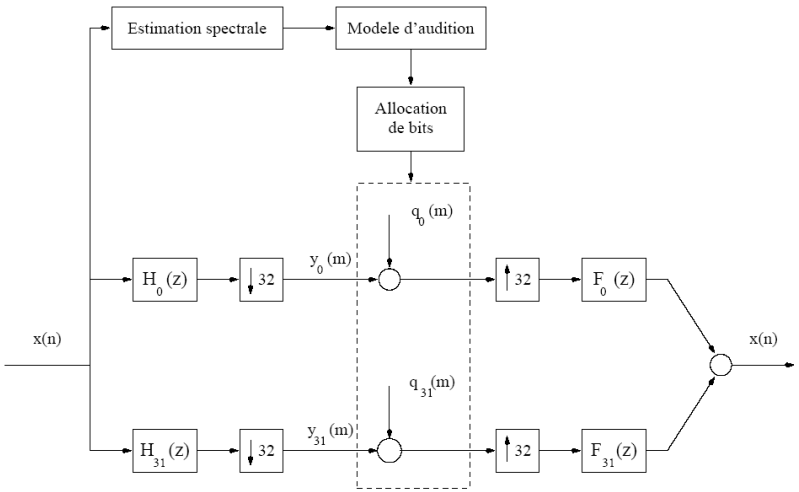
- **Banc de $M = 32$ filtres** pseudo-QMF
- Pour chaque sous-bande $k \in \{0 \dots M-1\}$:
  - Construction du **vecteur** $\mathbf{y}_k = [y_k(0) \dots y_k(11)]$
  - Détermination d'un **"facteur d'échelle"** $g_k = \max\{y_k(0) \dots y_k(11)\}$
  - **Normalisation** des composantes $[y_k(0) \dots y_k(11)]/g_k$
  - **QS uniforme** des composantes normalisées sur b_k bits
- Détermination des $b_0 \dots b_{M-1}$: **allocation de bits** sous le contrôle d'un modèle d'audition

MPEG-1 : Modèles psychoacoustiques

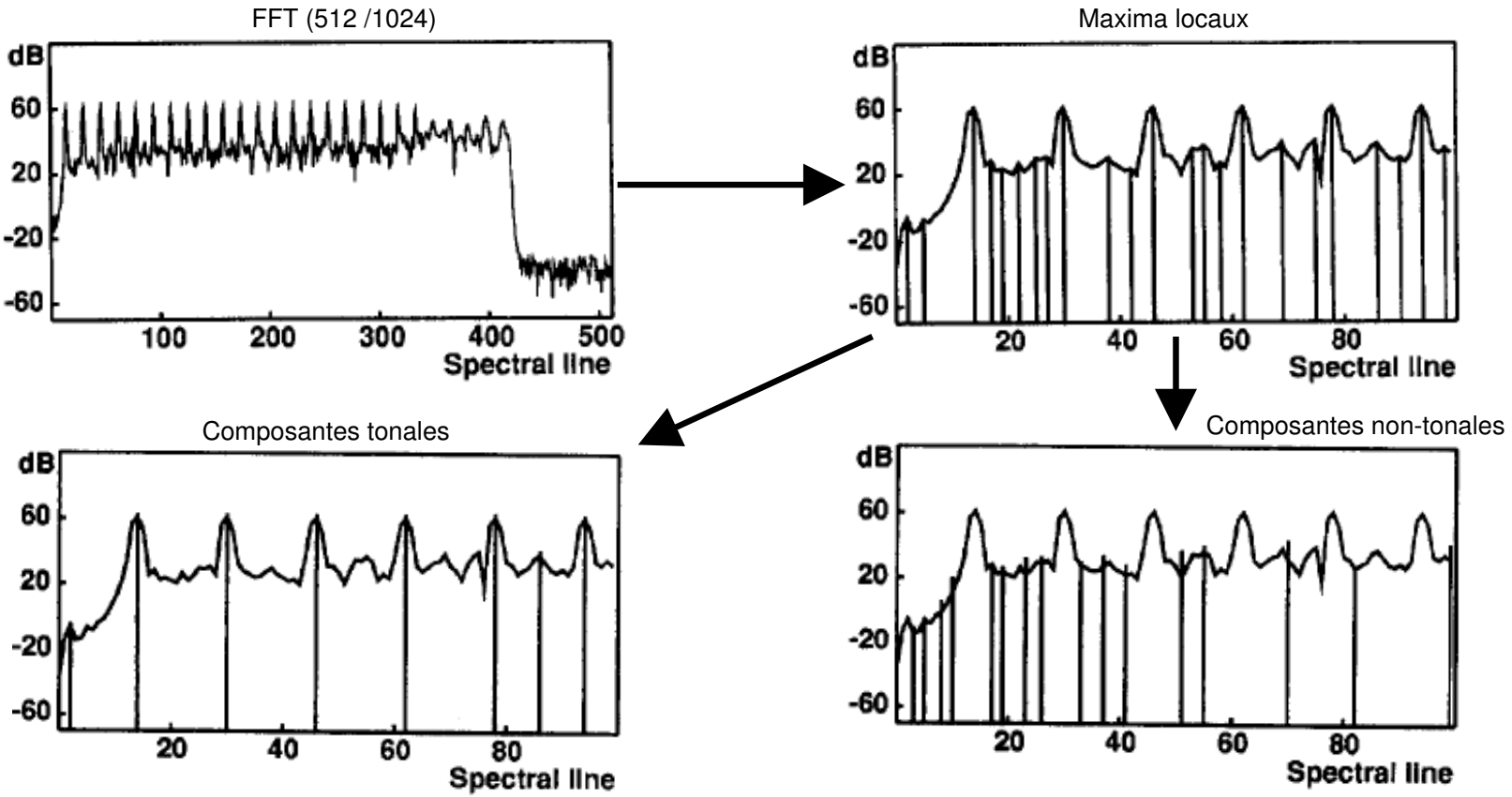
- **Complexité** du codeur : essentiellement due au modèle d'audition
- Contrôle par modèle psychoacoustique **seulement au codeur**

Norme : 2 propositions de modèles psychoacoustiques

- **Modèle I** : faible complexité, bons résultats à débits élevés
- **Modèle II** : complexité supérieure, meilleur pour bas débits

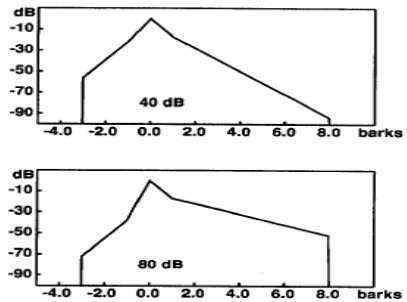


MPEG-1 : modèle psychoacoustique I (1/2)



MPEG-1 : modèle psychoacoustique I (2/2)

- **FFT** du signal dans la trame considérée (les $32 * 12 = 384$ échantillons correspondants)
- Détermination des **composantes tonales et non-tonales** de la FFT
- Calcul de l'**effet masquant** des différentes composantes dans leur voisinage fréquentiel, dépend de :
  - un **indice de masquage**, fonction du bark (unité pour les bandes critiques)
  - une **fonction de masquage**, étalement de l'effet de la composante sur les bandes fréquentielles voisines



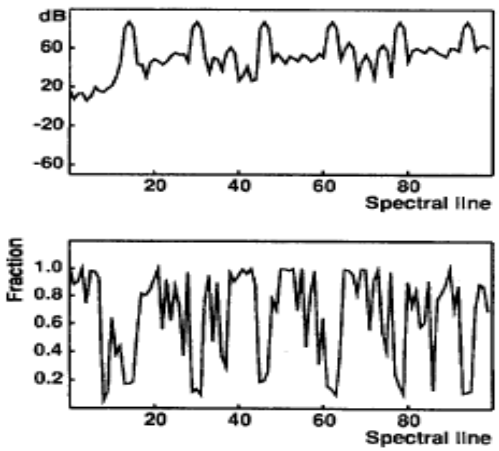
- Seuil de **masquage global** : pour chaque fréquence du spectre, on somme tous les masques calculés précédemment
- **Domaine des sous-bandes** du banc de filtres : seuil de masquage $=$ min. des seuils globaux dans chaque sous-bande
- **SMR (Signal-to-Mask-Ratio)** $=$ max$\{$Signal$\}$ - min$\{$Seuil masquage$\}$

MPEG-1 : modèle psychoacoustique II (1/2)

- **Pas de distinction** composante tonale/non-tonale
- Définition de **partitions** ($\sim 1/3$ BC)
- Une partition $\rightarrow$ **indice de "tonalité"**
- Détermination du **masque** en fonction de cet indice

MPEG-1 : modèle psychoacoustique II (2/2)

- **TFD** sur 1024 points (fenêtre de Hann)
- **Mesure d' "imprévisibilité"** : extrapolation de phase et amplitude de la ligne spectrale avec les 2 trames précédentes ($\rightarrow$ proche de 0 = "tonalité" de la composante)

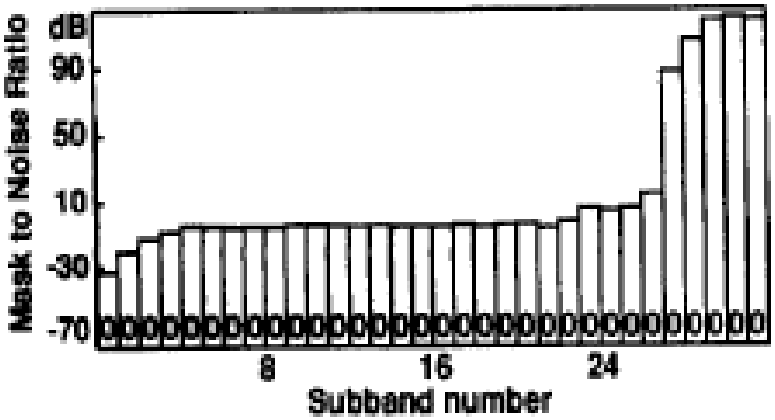


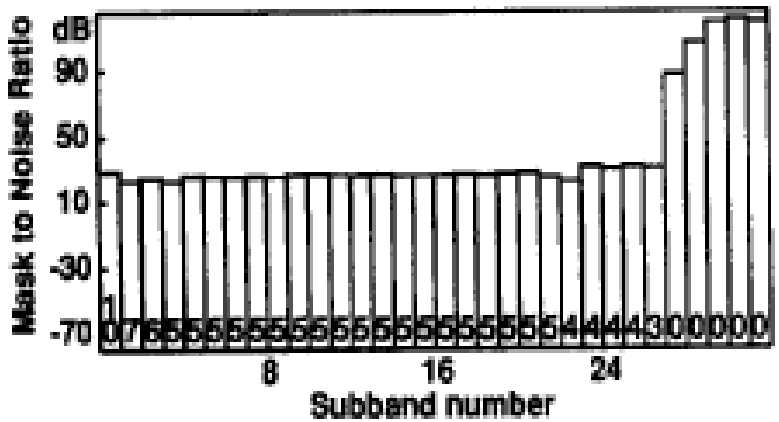
- **Mapping** vers partitions
- Calcul du **seuil de masquage** : convolution entre TFD (puis TFD pondérée par imprévisibilité) et fonction d'étalement
- **Mesure de "tonalité"** = log rapport du masque "imprévisibilité" sur le masque "signal"
- **Seuils de masquage globaux**
- **SMR** par sous-bande (MP2) / par facteur d'échelle (MP3)

MPEG-1 couche 1 : allocation de bits

- Nombre de bits **alloués à une trame**
  - bits/trame = (bits/seconde) / (trames/seconde)
  - trames/secondes = (echantillons/seconde)/(echantillons/trame)
  - Prendre en compte les bits d'entête et de données
- Allouer les bits restants de façon à **maximiser le MNR** (Mask-to-Noise Ratio) minimum par sous-bande
- **Modèle psychoacoustique** : SMR par sous-bande
- **Algorithme d'allocation** : MNR
  - MNR = SMR - SNR
  - SNR en fonction de l'allocation des bits, données par tables du standard

MPEG-1 couche 1 : algorithme d'allocation





Algorithme d'allocation de bits (MP1)

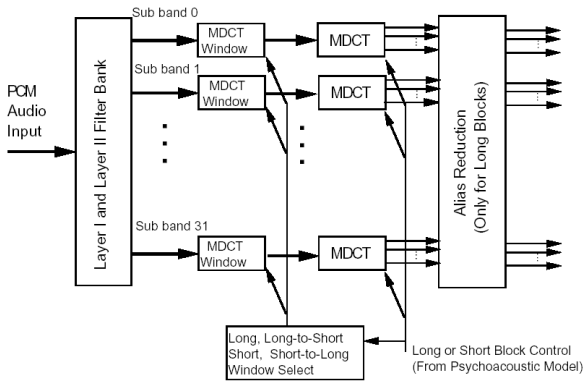
- **Initialisation** : 0 bits par sous-bande
- Calculer le **MNR** dans chaque sous-bande
- Déterminer la sous-bande où le **MNR est min** et où le nombre de bits est $<$ limite max.
- **Incrémenter** de 1 bit dans cette sous-bande

MPEG-1 couche 1 : transmission des données

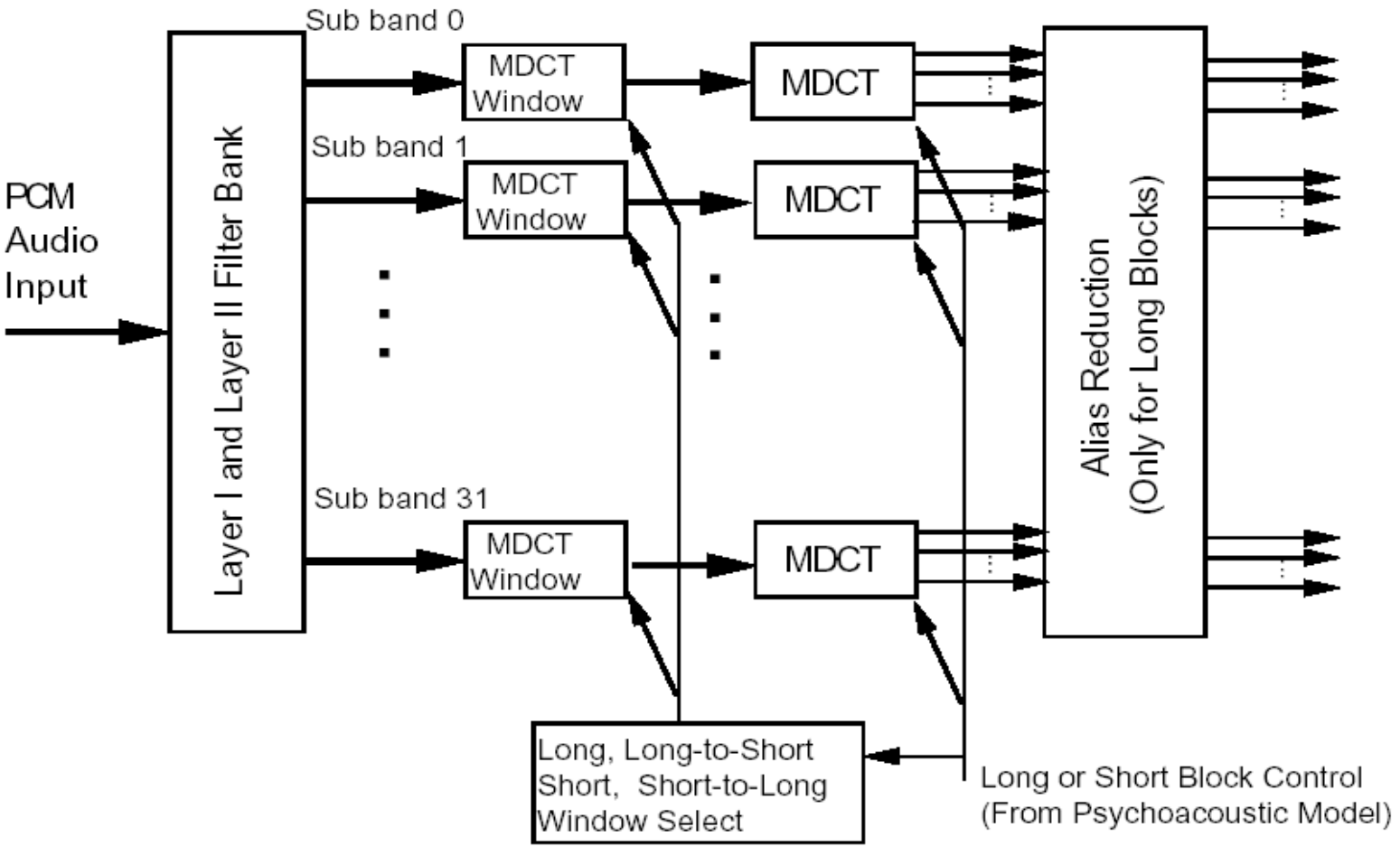
- Dans chaque fenêtre d'analyse ($32 \times 12 = 384$ échantillons $\Rightarrow$ 9 ms à $F_e = 44100$Hz) :
  - **Allocation de bits** $b_0 \ldots b_{M-1}$ explicitement transmise (88 bits)
  - Chaque **facteur d'échelle** g_k codé sur 6 bits si $b_k > 0$
  - Bits restants : **mots de code** associés à chaque composante normalisée
- **Bits restants** : à 96 kbits/s, $96000 * 384/44100 - 88 - 120 = 628$bits; à 64 kbits/s, 350 bits; etc. $\Rightarrow$ très vite impossible de diminuer plus le débit

MPEG-1 couche 3 : caractéristiques

- **Banc de filtres hybride**, MDCT
- Fenêtres courtes/ fenêtres longues
- Quantification **non uniforme**
- Bandes de **facteur d'échelle**
- Codage entropique des données
- Utilisation de réservoir de bits



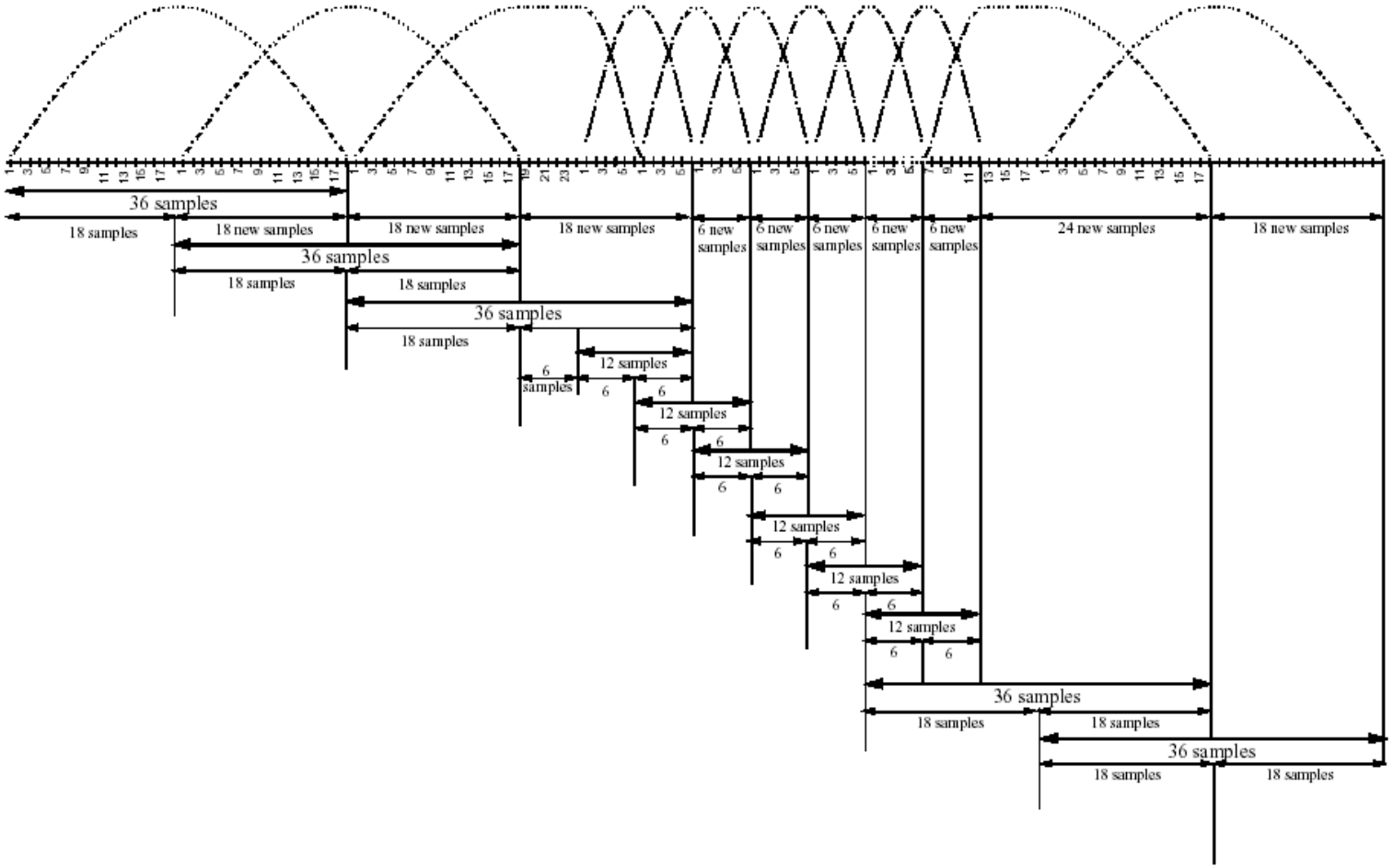
MPEG-1 couche 3 : structure hybride



MPEG-1 couche 3 : changement de fenêtre

- **Fenêtres longues** (36 échantillons), bonne résolution fréquentielle, segments stationnaires
- **Fenêtres courtes** (12 échantillons), bonne résolution temporelle, segments transitoires
- **Décision de changement** de fenêtre :
  - Calcul de PE (entropie perceptuelle) pour fenêtres longues
  - Si PE $>$ seuil alors changement
- **50 % de recouvrement** entre fenêtres
- Utilisation de **fenêtres de transitions**

MPEG-1 couche 3 : changement de fenêtre



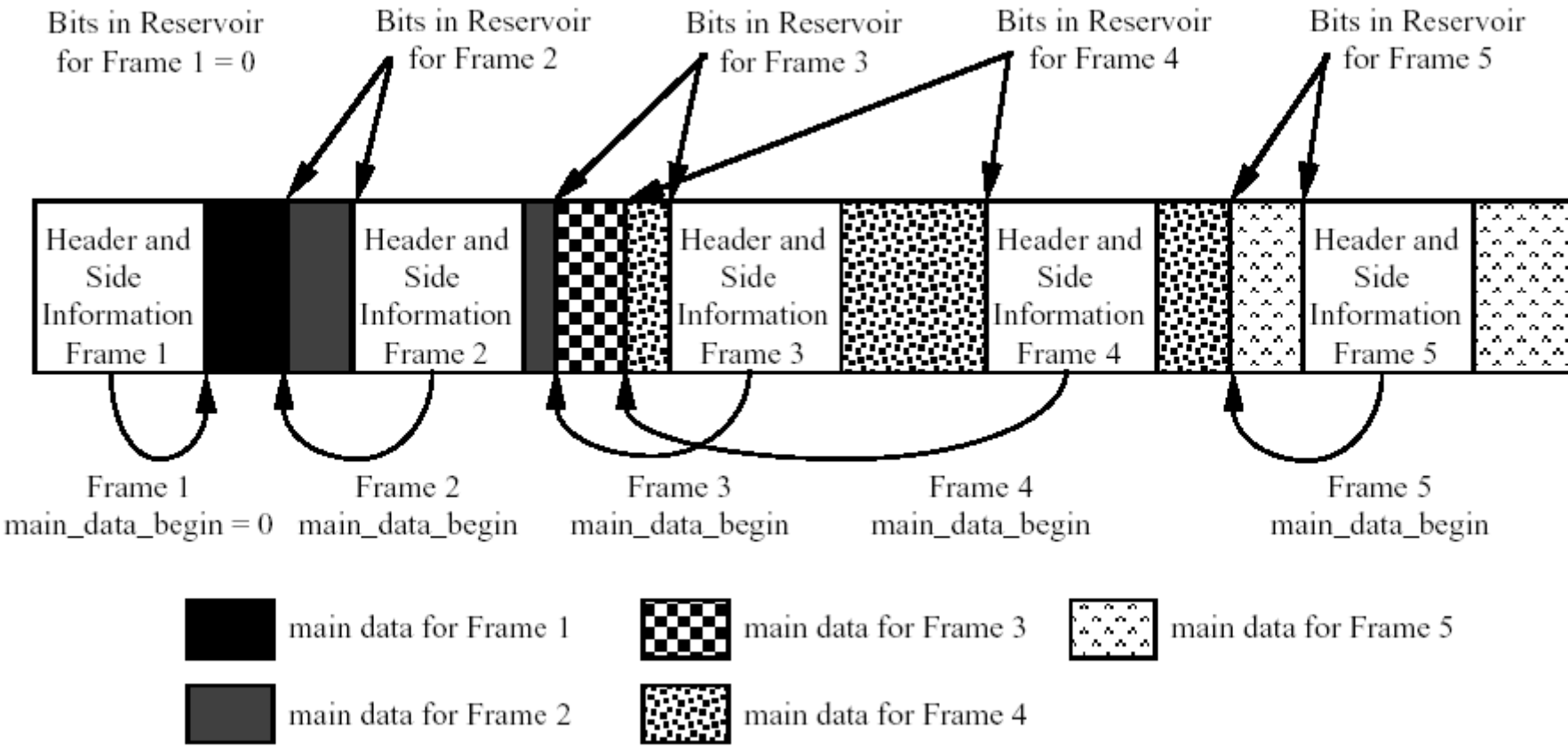
MPEG-1 couche 3 : Quantification

- **2 contraintes** : compromis débit / contrainte psychoacoustique (masque)
- $X_q(\text{sfb}, k) = \text{sign}(X(sfb, k))\text{round}\left[\left(|X(\text{sfb}, k)|2^{(a(\text{sfb})-b)/4}\right)^{3/4}\right]$:
  - X_q échantillon de MDCT **quantifié**
  - X échantillon de MDCT **original**
  - sfb : **bande de facteur d'échelle** ($\Leftrightarrow$ bande critique), 21 bandes pour fenêtres longues, 12 pour fenêtres courtes
  - k : indice de la raie spectrale
  - a(sfb) : **facteur d'échelle**, amplifier les raies spectrales pour satisfaire la contrainte de masquage
  - b : **pas de quantification** (constant pour les 576 raies)
- a et b estimé dans **2 boucles d'itération** (boucle externe et boucle interne)
- $\frac{3}{4}$: meilleur SNR.

MPEG-1 couche 3 : Codage de Huffman

- Classement par ordre de fréquences croissantes des 576 coefficients MDCT (18x32)
  - Basses freq. $>$ valeurs assez importantes
  - Freq. plus élevées $>$ valeurs proches de 0
  - Hautes freq. $>$ grand nombre de 0
- **Délimiter les coefficients** en trois régions à coder par des tables différentes
  - **Région des 0** : non codée, étendue déduite à partir des 2 autres
  - **Région où "abs $<$ 1"** : codage de quadruplets de valeurs
  - **Big_values** : codage par paires, sous-régions et tables adaptées
- 34 tables de Huffman au total

MPEG-1 couche 3 : Réservoir de bits



MPEG-2 : historique

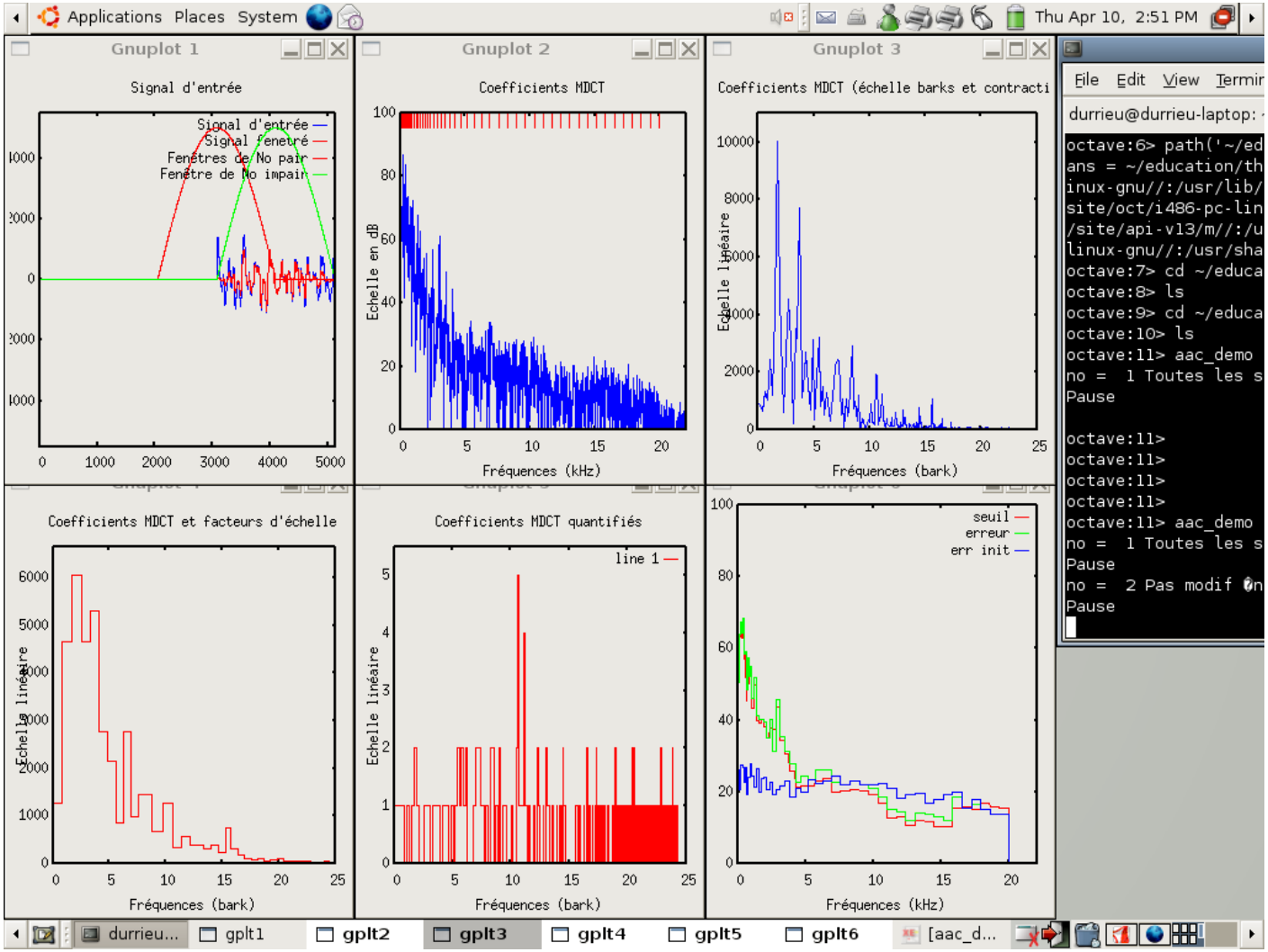
- **MPEG-2 audio** (1994)
  - **Extension** du MPEG-1 : multi-canal (5.1), F_e réduites
  - **Diffusion** : TVHD, stockage (DVD)
- **MPEG AAC** (Advanced Audio Coding) (1997 - 1998)
  - Codage multi-canal à des débits raisonnables
  - **Transparence** à 384 kbits/s - 5.1
  - 1 à 48 canaux, 8 à 96 kHz, 8 à 160 kbits/s
  - **Performances** nettement supérieures
  - **Complexité** moindre
  - 3 profils : “main profile”, “low complexity profile”, “scalable rate profile”
  - **Etat de l’art** en codage audio

MPEG-2 : AAC (Advanced Audio Coding)

MPEG-2 : Caractéristiques du AAC

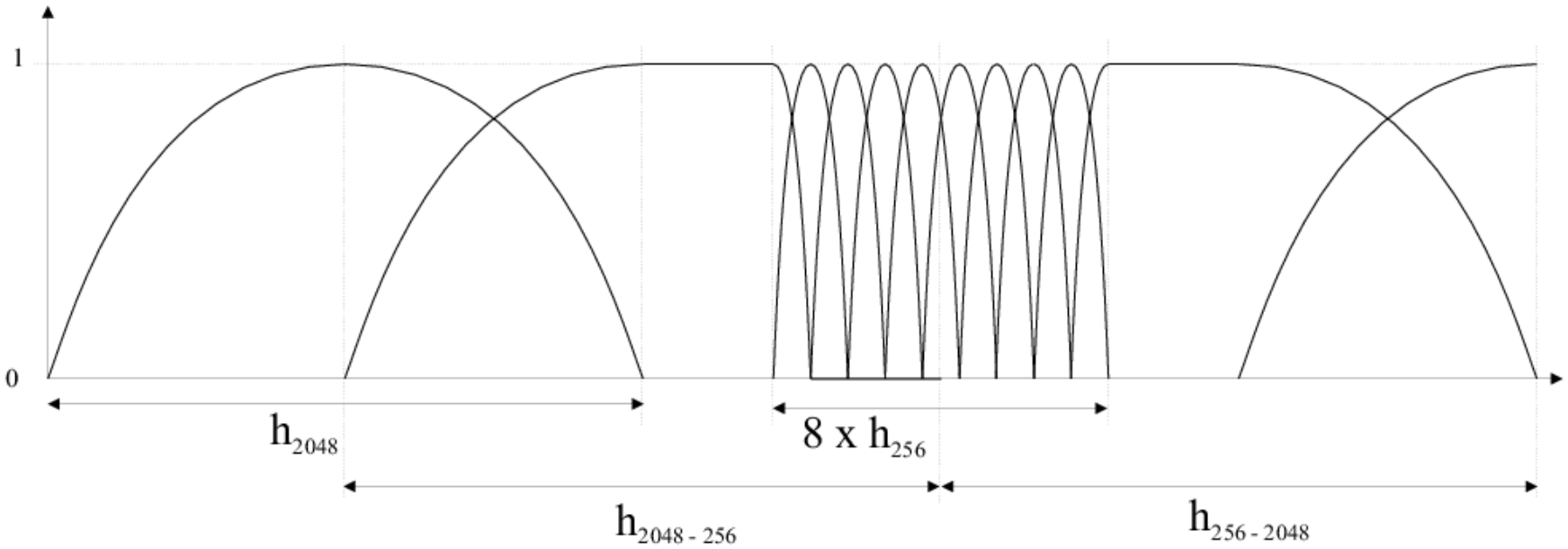
- $\mathbf{x}(m) = [x(mM) \dots x(mM + N - 1)] \Rightarrow \mathbf{X}(m) = [X_0(m) \dots X_{M-1}(m)]$
- MDCT : $N = 2048$, $M = 1024$
- Transformée à recouvrement : problème de reconstruction

MPEG-2 : les différentes étapes du AAC



MPEG-2 : fenêtrage dynamique

- Fenêtres longues $N = 2048$, $M = 1024$
- Fenêtres courtes (par 8) $N = 256$, $M = 128$



MPEG-2 : codeur AAC

- Facteurs d'échelle $\mathbf{g} = [g(0) \dots g(M-1)]$
- Emetteur : calcul de

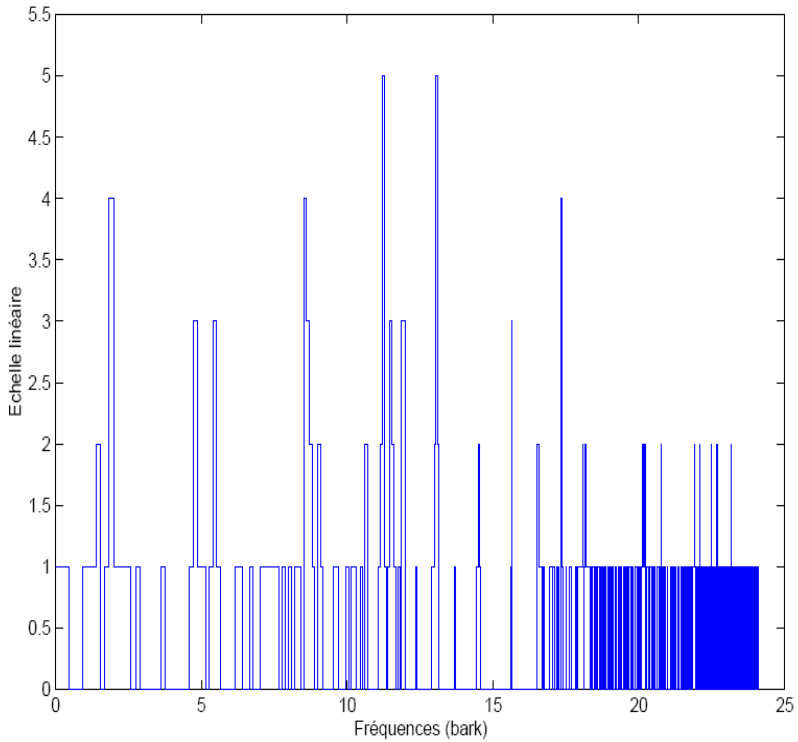
$$\mathbf{i}(m) = \text{round}\left(\left[\frac{X_0(m)}{g_0(m)} \dots \frac{X_{M-1}(m)}{g_{M-1}(m)}\right]\right)$$

- Récepteur : reconstruction de $\hat{\mathbf{X}}(m)$

$$[\hat{X}_0(m) \dots \hat{X}_{M-1}(m)] = [g_0(m) \times i_0(m) \dots g_{M-1}(m) \times i_{M-1}(m)]$$

MPEG-2 : codage du vecteur $\mathbf{i}(m) = [i_0(m) \dots i_{M-1}(m)]$

- Si QS à 3 bits sur une bande, $M = 1024$ entiers tq $0 \leqslant i_k(m) \leqslant 8$.
  Bits nécessaires $B = 1024 \times 3 >$ nombre de bits disponibles ≈ 750
- **Codage de Huffman** : Valeurs plus probables (apparaissant souvent) $\longrightarrow$ moins de 3 bits



- En pratique : **partitionnement** en 51 sous-bandes dans chaque bande, $\max\{i_k(m)\}$, puis codage séparé
- Dans le **"bit stream"** : mots de code des $\mathbf{i}_k(m)$, des $\max\{i_k(m)\}$ et mots de des $g_k(m)$

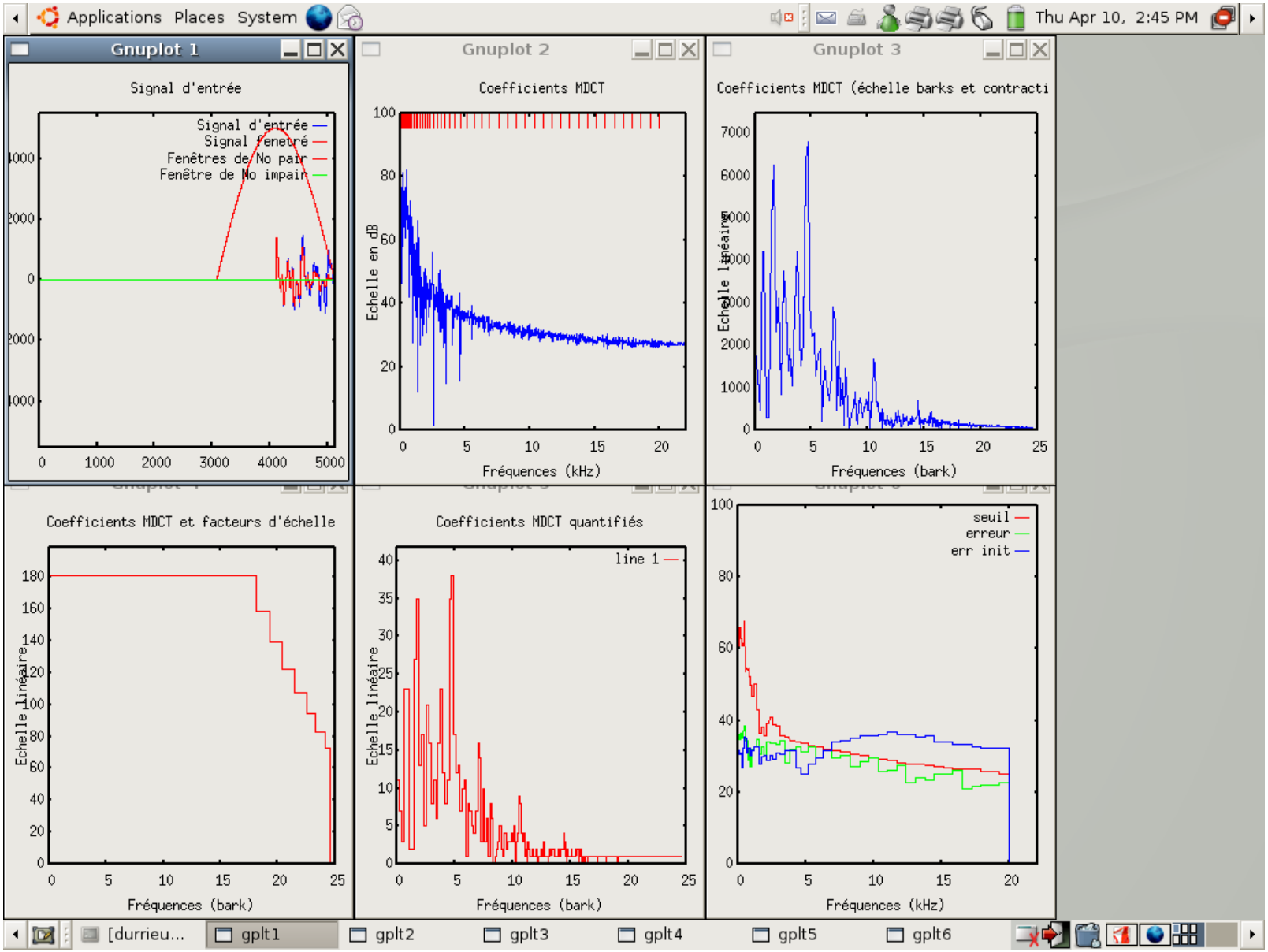
MPEG-2 : détermination des facteurs d'échelle $\mathbf{g}(m)$

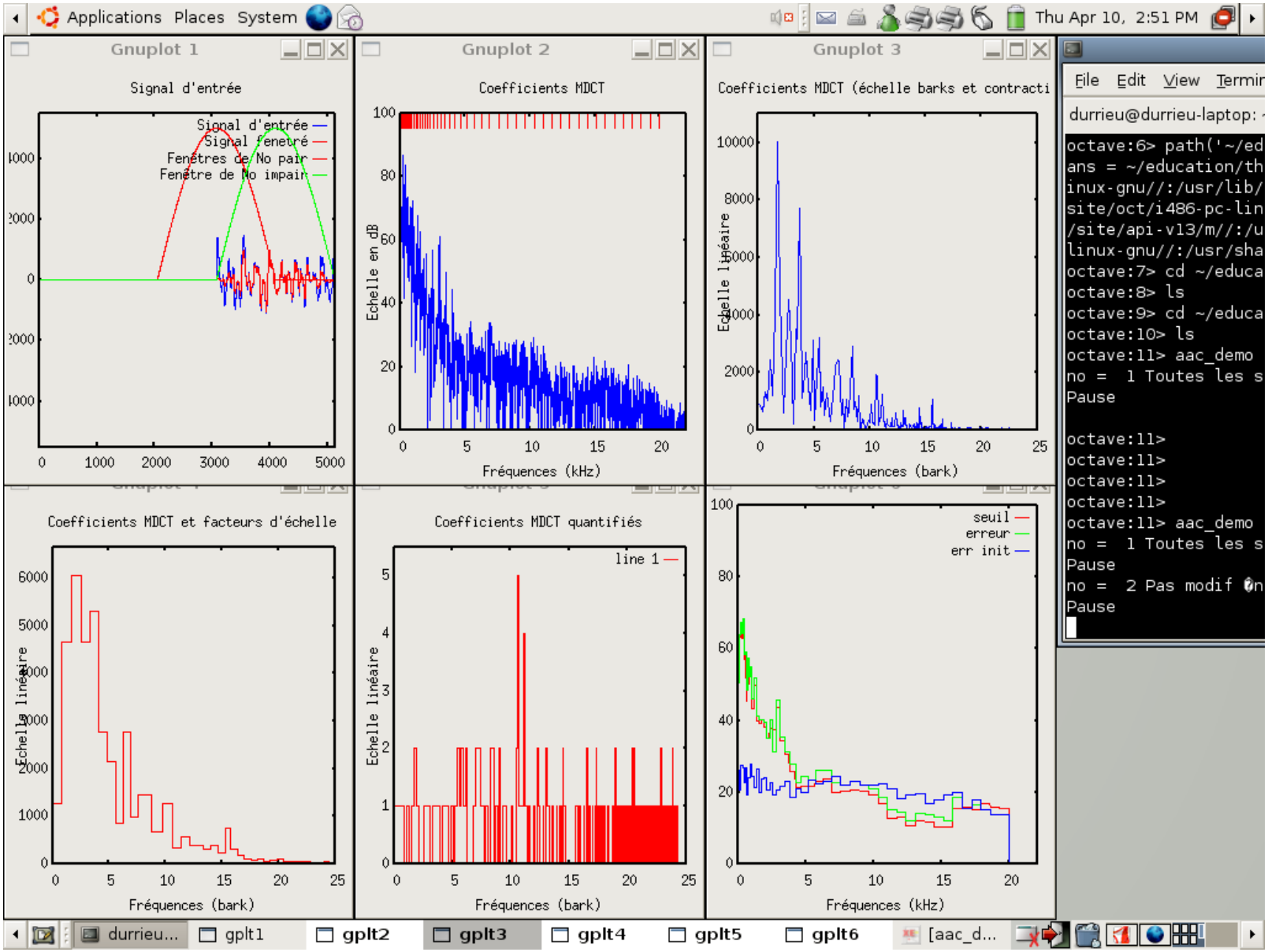
- Taux de compression élevé $\rightarrow$ exploitation des résultats de psychoacoustique
- **"Mise en forme spectrale"** du bruit de reconstruction
- problème d'**optimisation sous contrainte** :
  - contrainte de débit : contrôle par une **composante globale** dans le facteur d'échelle (b)
  - contrainte "psychoacoustique" : $\sigma_Q^2(sfb) < \Phi(sfb)$, **composante propre à chaque sous-bande**

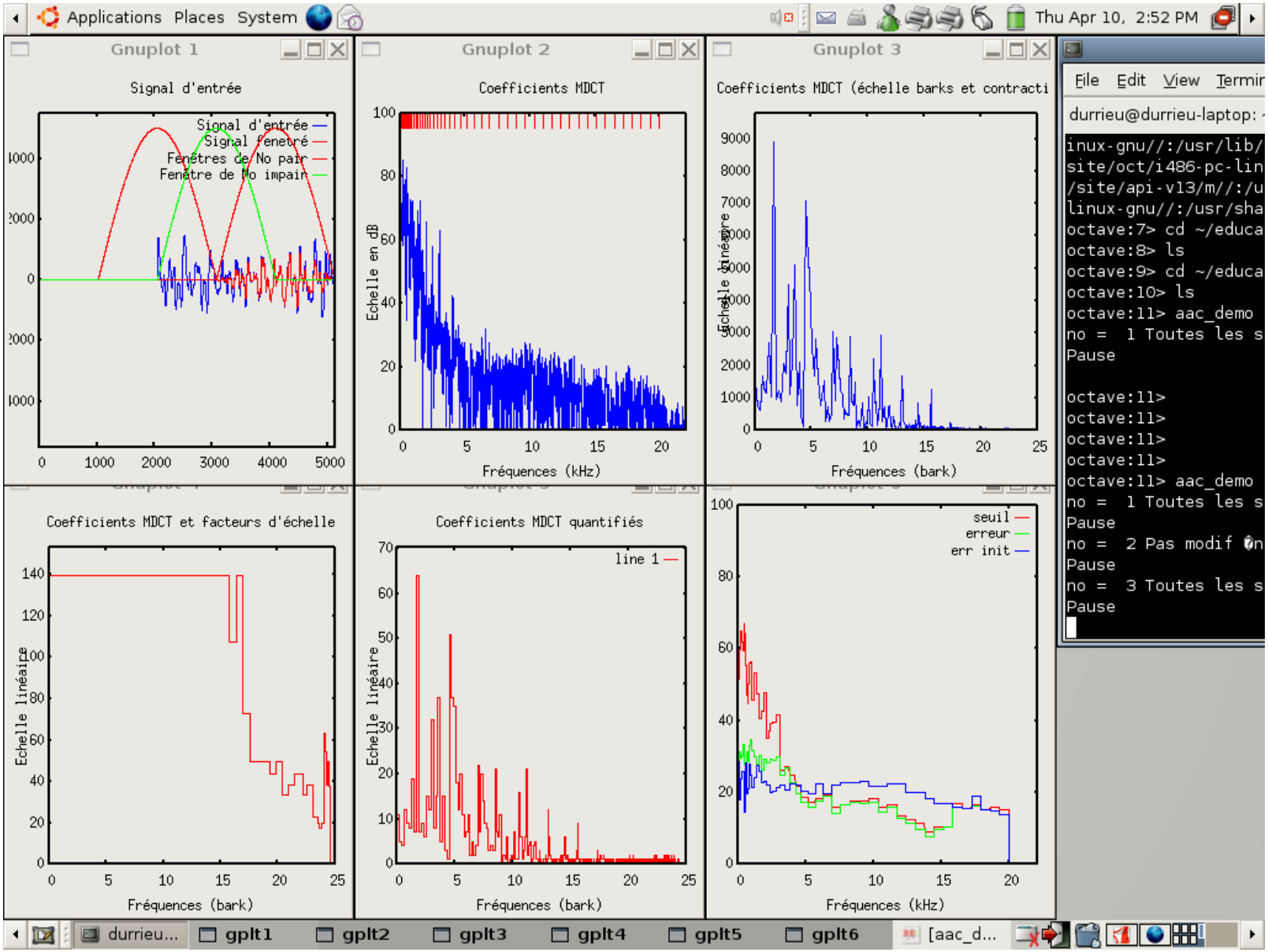
MPEG-2 : informations transmises

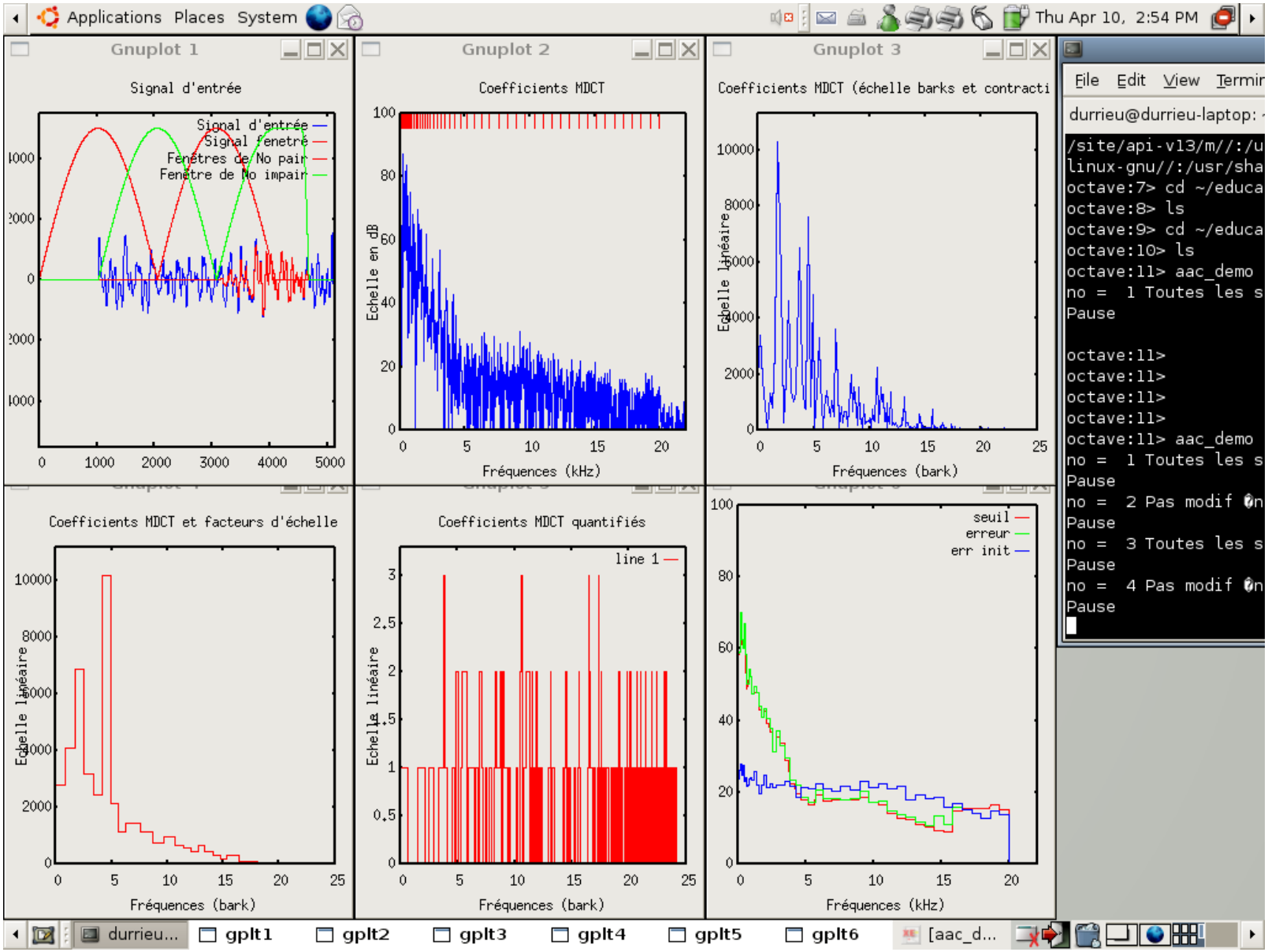
Pour chaque fenêtre d'analyse :

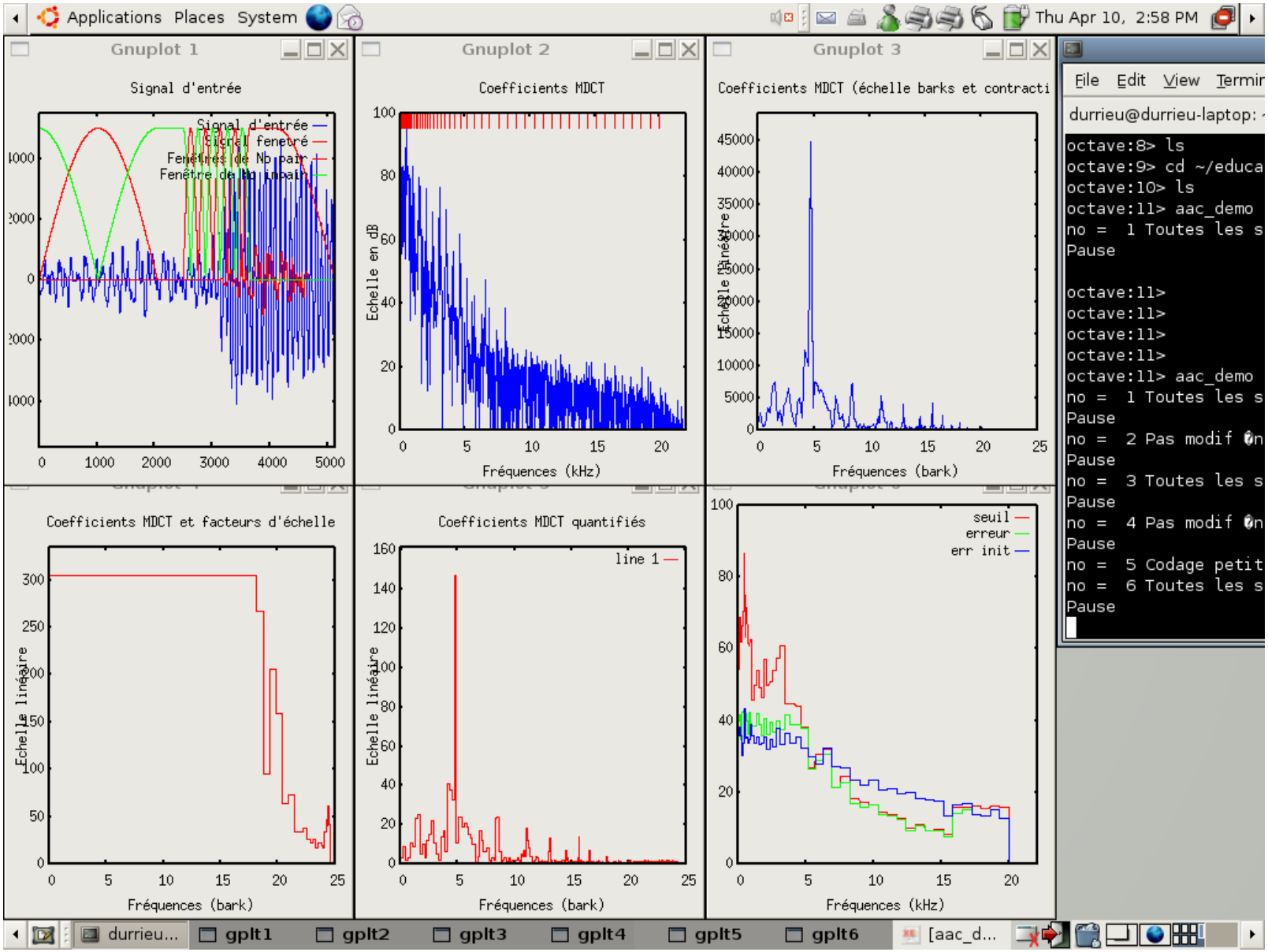
- **Partition** de l'axe des fréquences en 51 sous-bandes
- **"Facteurs d'échelle"**:
  - Codage du 1er directement sur 8 bits
  - Codage de $\Delta_k(m) = g_k(m) - g_{k-1}(m)$ pour les 50 suivants
  - Table de Huffman
- **Coefficients de la MDCT** :
  - 4 composantes dans la 1ère sous-bande, 32 dans la dernière
  - Codage du signe à part
  - Détermination du max dans chaque sous-bande, choix d'une table de Huffman parmi 11
  - Codage des $i_k(m)$

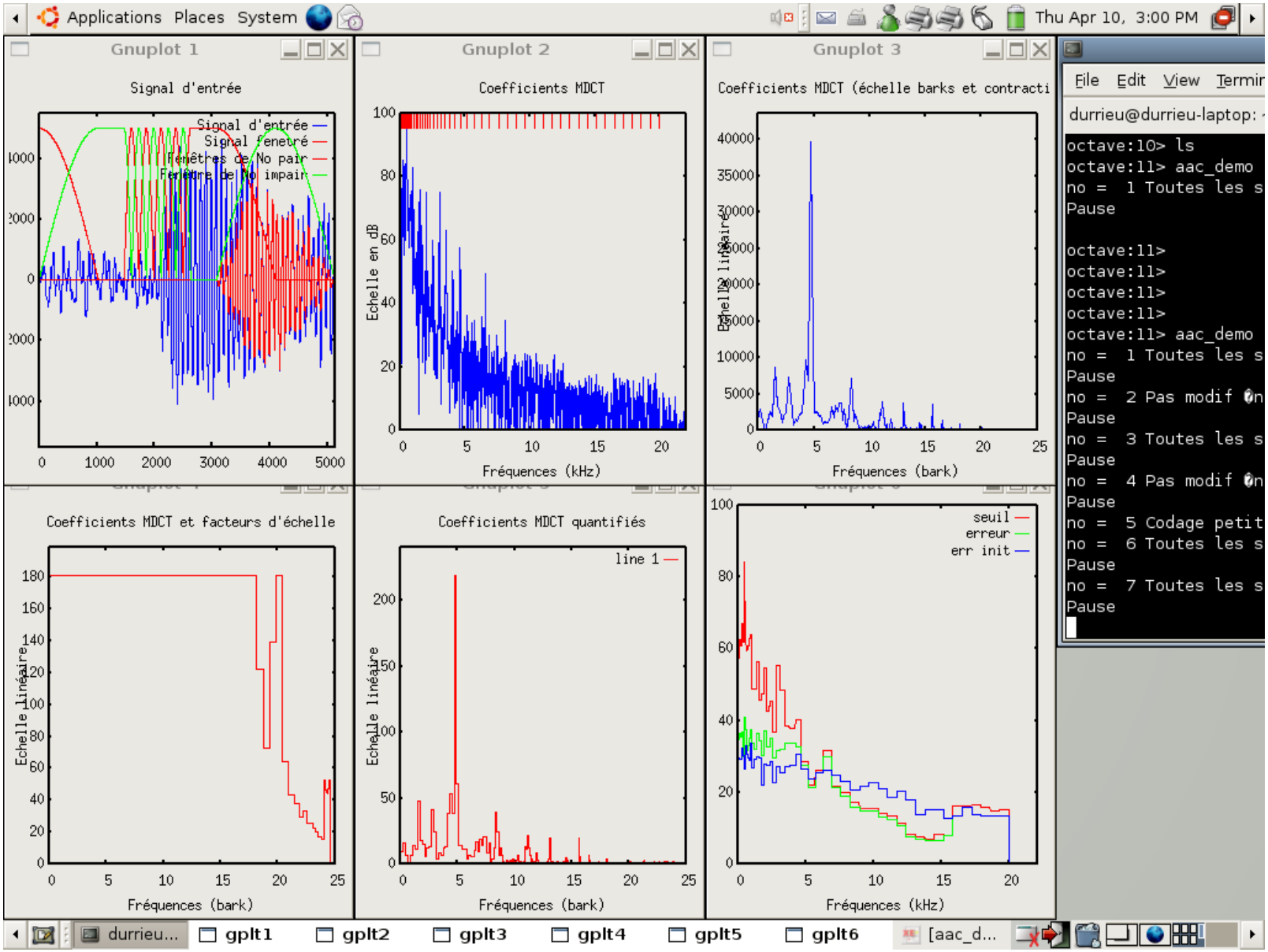

Applications Places System
Thu Apr 10, 2:45 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
[durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6

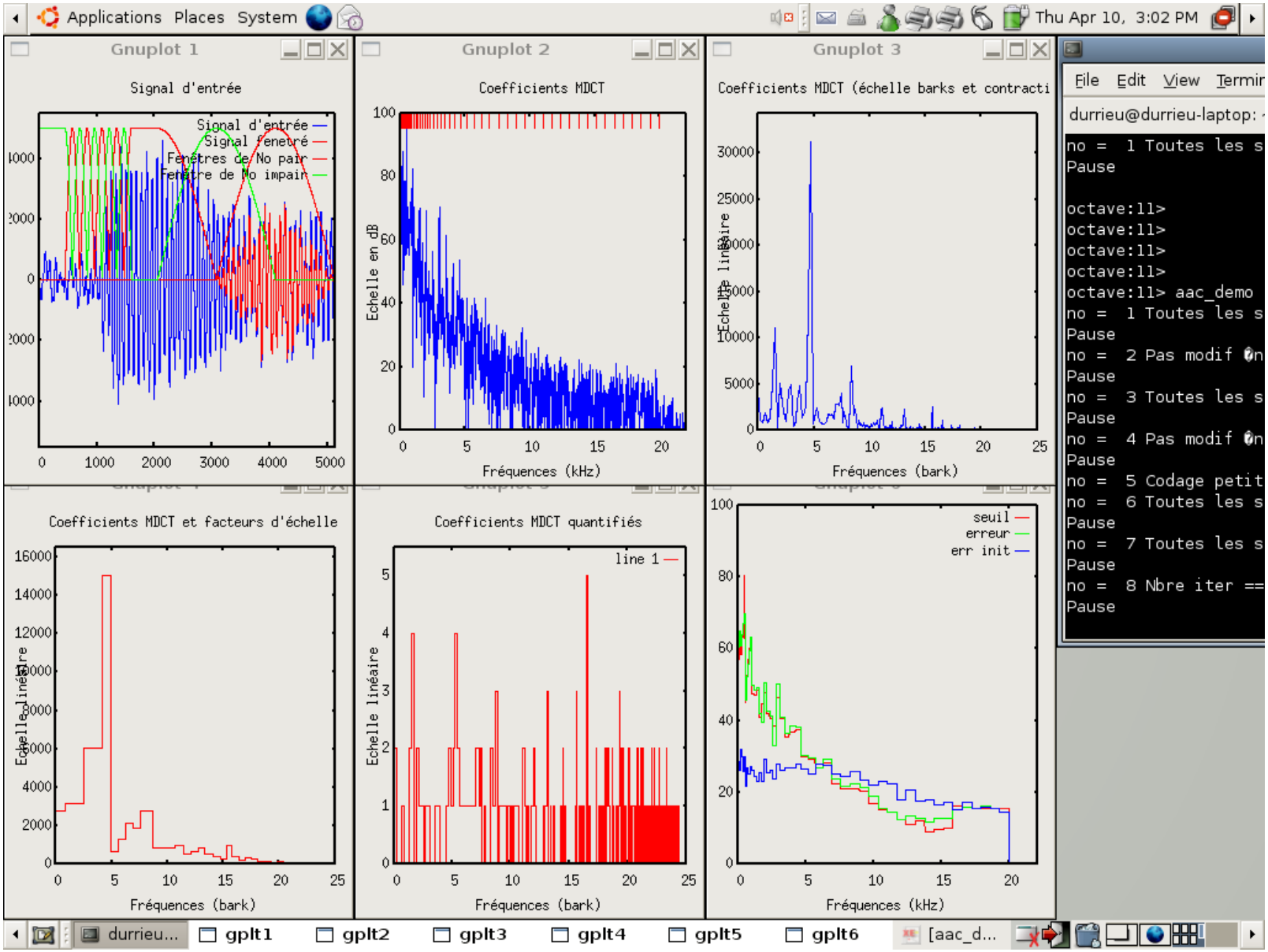

Applications Places System
Thu Apr 10, 2:51 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
File Edit View Termir
durrieu@durrieu-laptop:
octave:6> path('~/ed
ans = ~/education/th
inux-gnu//:/usr/lib/
site/oct/i486-pc-lin
/site/api-v13/m//:/u
linux-gnu//:/usr/sha
octave:7> cd ~/educa
octave:8> ls
octave:9> cd ~/educa
octave:10> ls
octave:11> aac_demo
no = 1 Toutes les s
Pause
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...

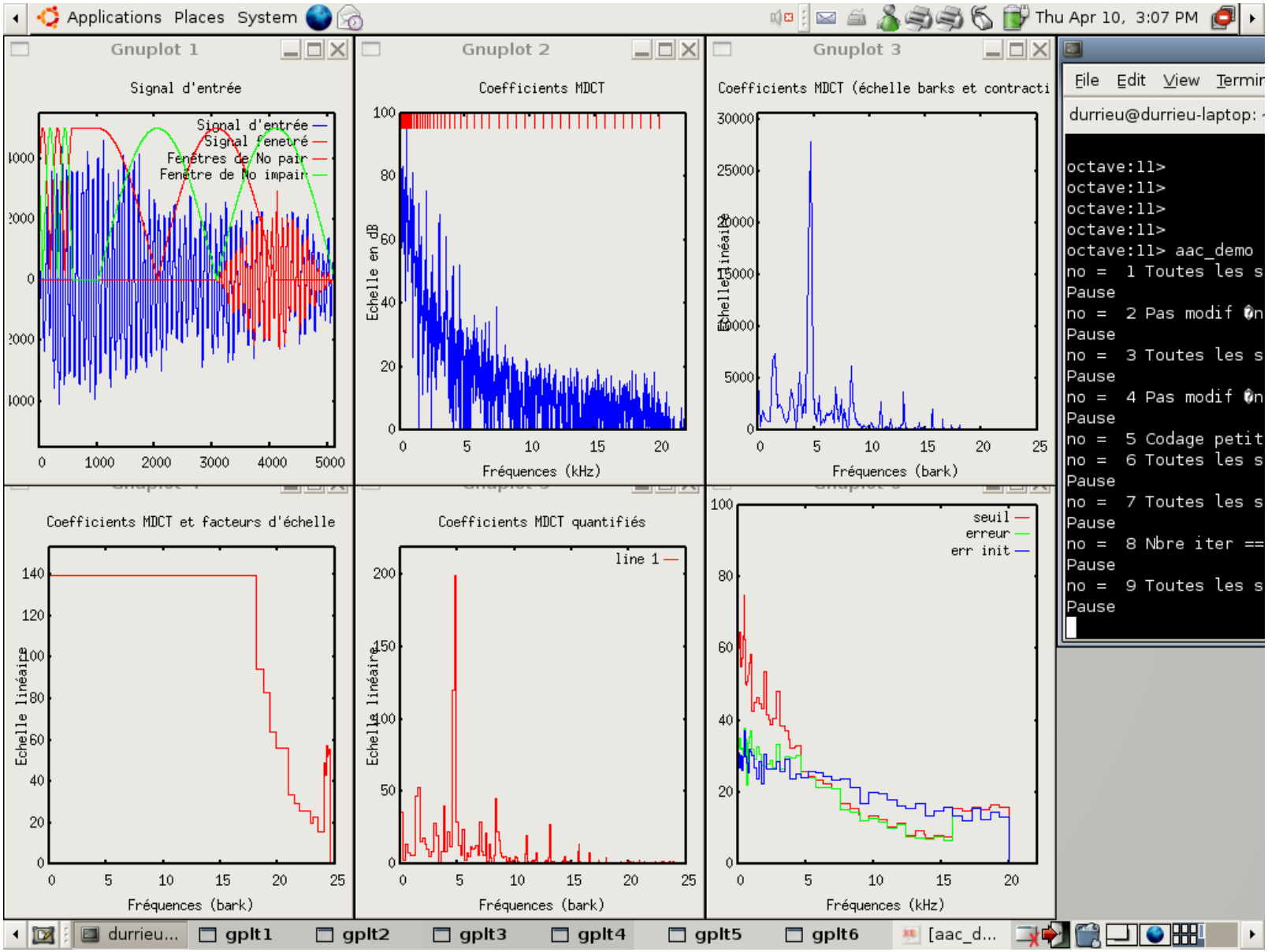

Applications Places System
Thu Apr 10, 2:52 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
octave:7> cd ~/educa
octave:8> ls
octave:9> cd ~/educa
octave:10> ls
octave:11> aac_demo
no = 1 Toutes les s
Pause
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...

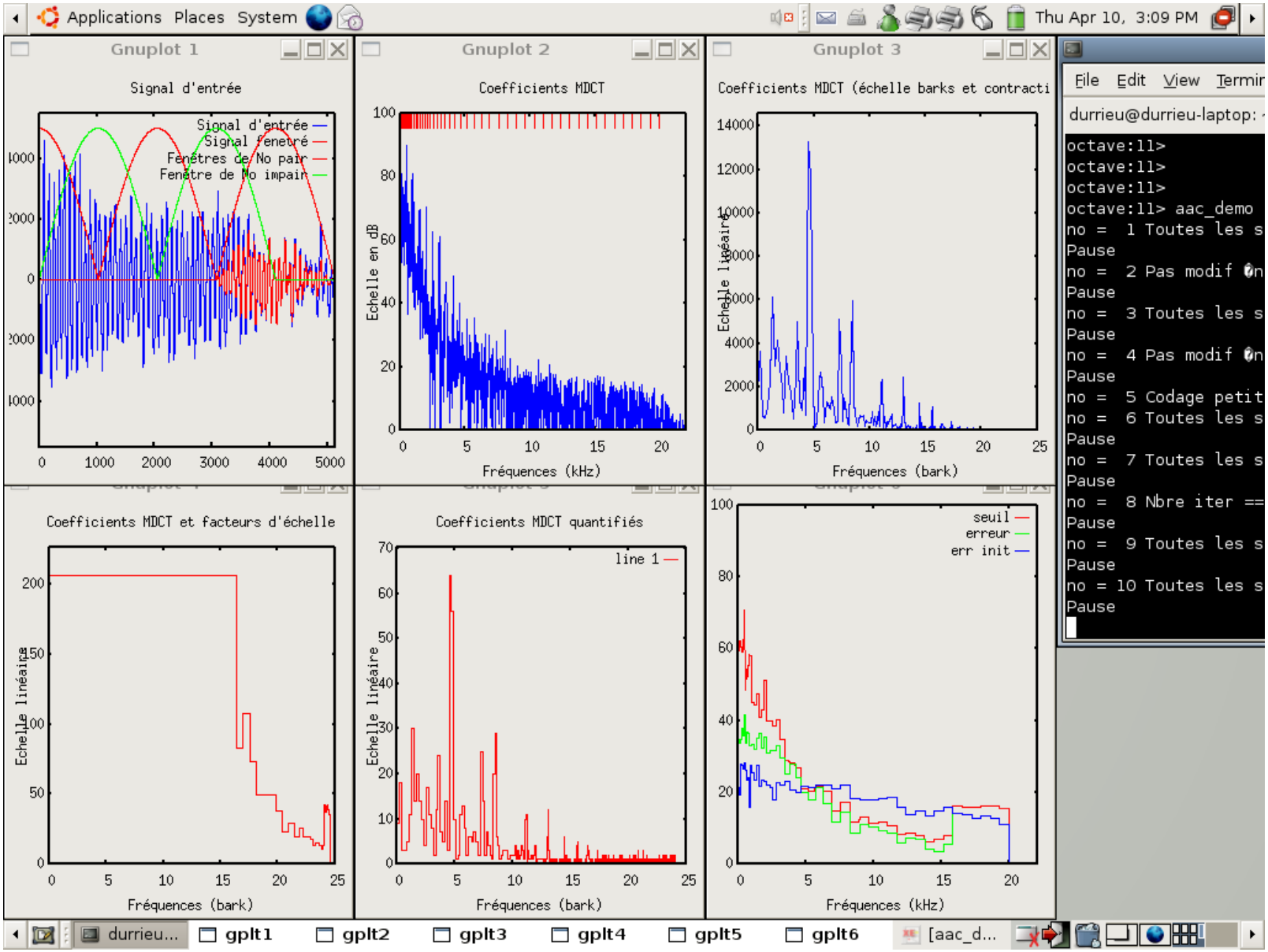

Applications Places System
Thu Apr 10, 2:54 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
/site/api-v13/m//:/u
linux-gnu//:/usr/sha
octave:7> cd ~/educa
octave:8> ls
octave:9> cd ~/educa
octave:10> ls
octave:11> aac_demo
no = 1 Toutes les s
Pause
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...


Applications Places System
Thu Apr 10, 2:58 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de Nb pair
Fenêtre de Nb impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
octave:8> ls
octave:9> cd ~/educa
octave:10> ls
octave:11> aac_demo
no = 1 Toutes les s
Pause
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
no = 5 Codage petit
no = 6 Toutes les s
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...


Applications Places System
Thu Apr 10, 3:00 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtres de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
octave:10> ls
octave:11> aac_demo
no = 1 Toutes les s
Pause
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
no = 5 Codage petit
no = 6 Toutes les s
Pause
no = 7 Toutes les s
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...


Applications Places System
Thu Apr 10, 3:02 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
no = 1 Toutes les s
Pause
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
no = 5 Codage petit
no = 6 Toutes les s
Pause
no = 7 Toutes les s
Pause
no = 8 Nbre iter ==
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...


Applications Places System
Thu Apr 10, 3:07 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
octave:11>
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
no = 5 Codage petit
no = 6 Toutes les s
Pause
no = 7 Toutes les s
Pause
no = 8 Nbre iter ==
Pause
no = 9 Toutes les s
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...


Applications Places System
Thu Apr 10, 3:09 PM
Gnuplot 1
Signal d'entrée
Signal d'entrée
Signal fenetré
Fenêtres de No pair
Fenêtre de No impair
Gnuplot 2
Coefficients MDCT
Echelle en dB
Fréquences (kHz)
Gnuplot 3
Coefficients MDCT (échelle barks et contracti
Echelle linéaire
Fréquences (bark)
File Edit View Termin
durrieu@durrieu-laptop:
octave:11>
octave:11>
octave:11>
octave:11> aac_demo
no = 1 Toutes les s
Pause
no = 2 Pas modif
Pause
no = 3 Toutes les s
Pause
no = 4 Pas modif
Pause
no = 5 Codage petit
no = 6 Toutes les s
Pause
no = 7 Toutes les s
Pause
no = 8 Nbre iter ==
Pause
no = 9 Toutes les s
Pause
no = 10 Toutes les s
Pause
Coefficients MDCT et facteurs d'échelle
Echelle linéaire
Fréquences (bark)
Coefficients MDCT quantifiés
line 1
Echelle linéaire
Fréquences (bark)
seuil
erreur
err init
Fréquences (kHz)
durrieu...
gplt1
gplt2
gplt3
gplt4
gplt5
gplt6
[aac_d...

Plan

Autres normes :

- **MPEG 4** : v1 (98), v2 (99), représentation des sons
  - **Sons naturels** : codeurs hiérarchiques, boite à outils pour la compression
  - **Sons synthétiques** : synthèse de parole, format MIDI, description de scène
- **MPEG 7** : **descripteurs** pour les données multimedia, faciliter les **requêtes par le contenu** dans les bases de données
- **MPEG 21** : sécurité, tatouage

Références

- Poly de cours de codage de N. MOREAU : http://www.tsi.enst.fr/ moreau/enseignement.html
- Support de cours de S. Essid : http://www.tsi.enst.fr/ essid/teach/cours-int06.pdf
- Articles sur les normes :
  - S. Shlien, *Guide to MPEG-1 Audio Standard*, IEEE Trans. on Broadcasting, vol.40, n. 4, 1994
  - E. Ambikairajah *et al*, *Auditory masking and MPEG-1 audio compression*, Electronics & Communication Engineering Journal, 1997
  - O. Derrien *et al*, *Le codeur MPEG-2 AAC expliqué aux traiteurs de signaux*, Annales des télécommunications, 2000
- Sites internet sur le fonctionnement de l'oreille : http://www.iurc.montp.inserm.fr/cric51/audition/ (http://www.cochlee.info), http://mediatheque.ircam.fr/articles/textes/Groscarret99a/