Les CODECs audio et plus particulièrement Ogg Vorbis

Philippe Teuwen
<phil@teuwen.org>

Présentation basée sur les matériaux de
Werner Oomen & Nicolas Werner
que je remercie au passage

Contenu de la présentation

Introduction aux CODECs audio

Ingrédients nécessaires:

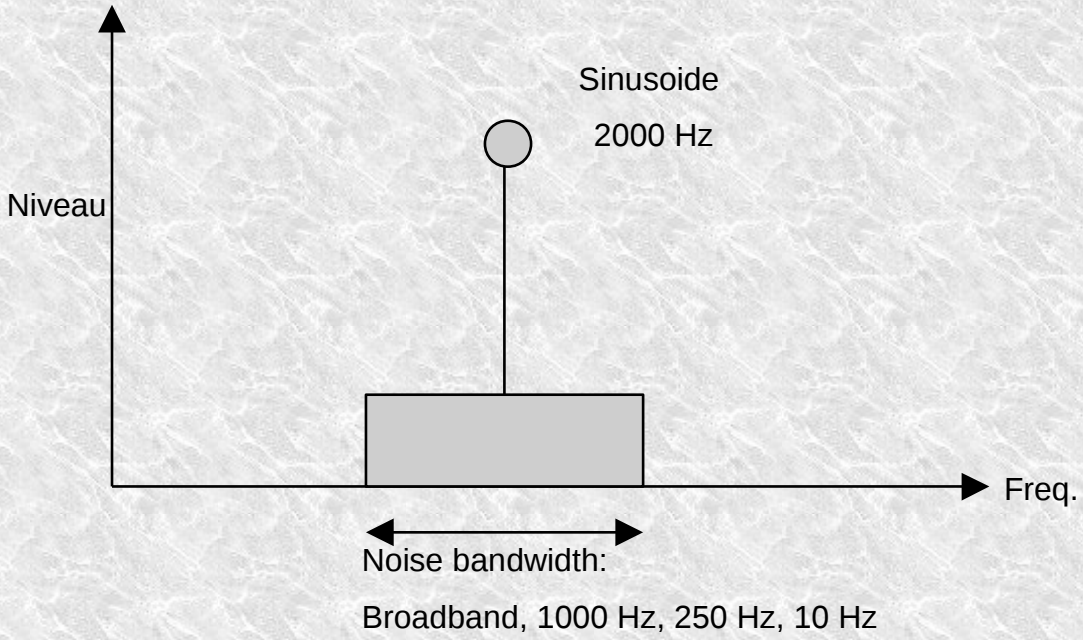
- Une bonne connaissance de la psycho-acoustique
- Des algorithmes de Digital Signal Processing (DSP)
  - Transformations du signal
  - Quantification
- La théorie de l'information
  - Codage sans pertes (Loss-less Coding)

Et enfin...

Psycho-acoustique:
1) L'effet de masquage (masking)

Un son est-il toujours audible?

Interaction entre les sons...

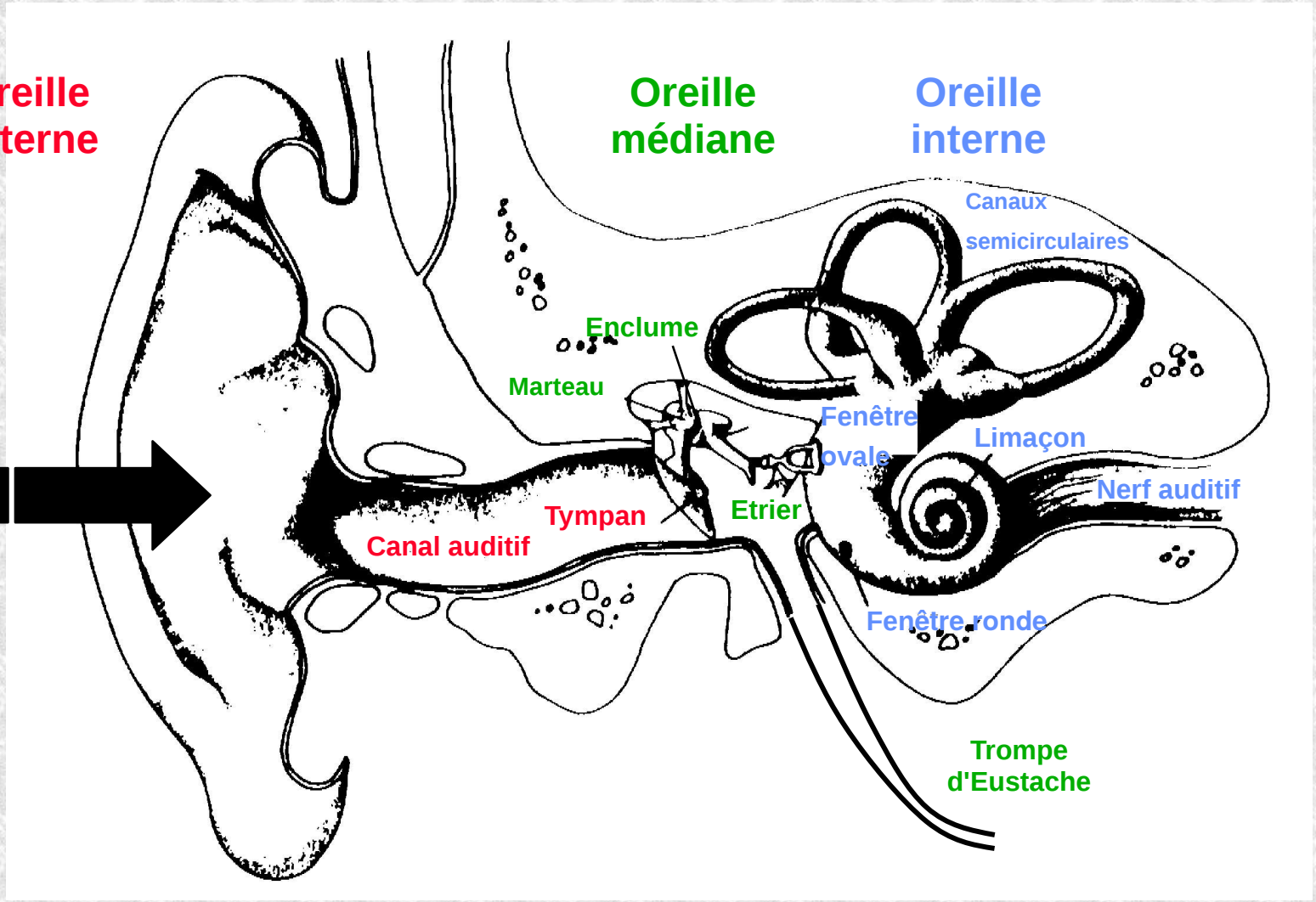


Avant d'aller plus loin...

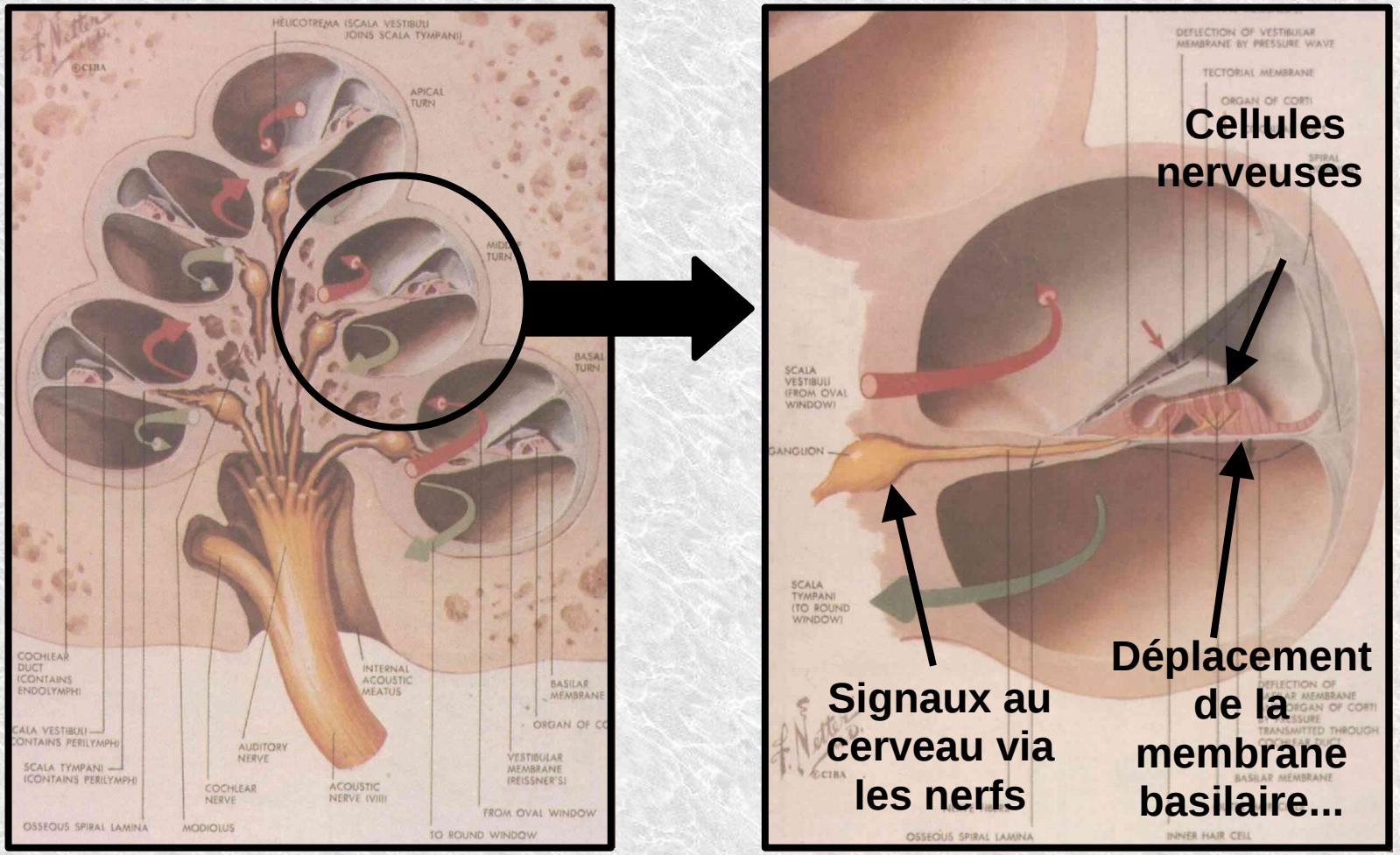
Pourquoi les fréquences sont représentées selon une échelle logarithmique?

Pourquoi les niveaux de pression sonore sont représentés selon une échelle logarithmique?
(le dB est une représentation logarithmique)

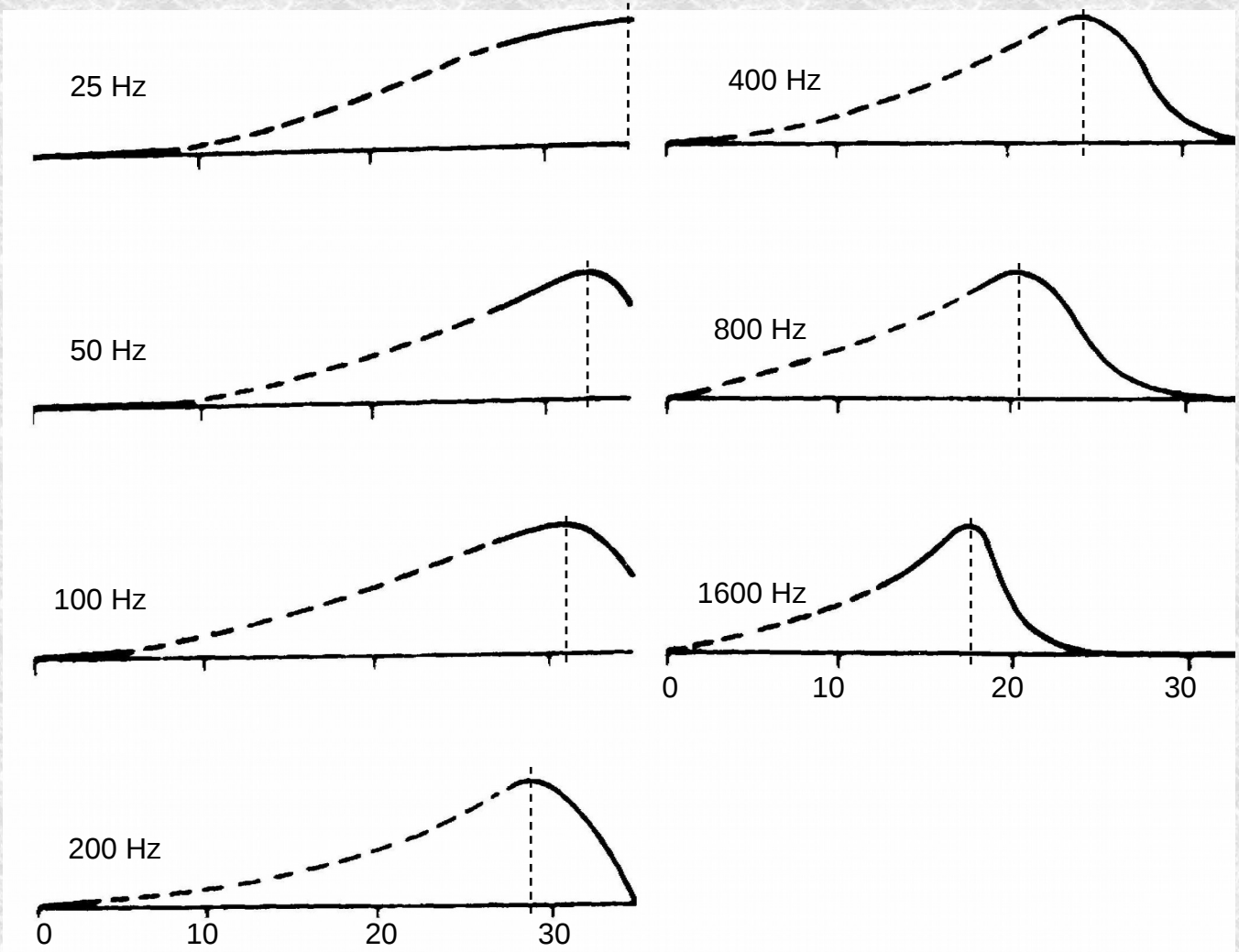
L'appareil auditif



Limaçon et membrane basilaire



Membrane basilaire

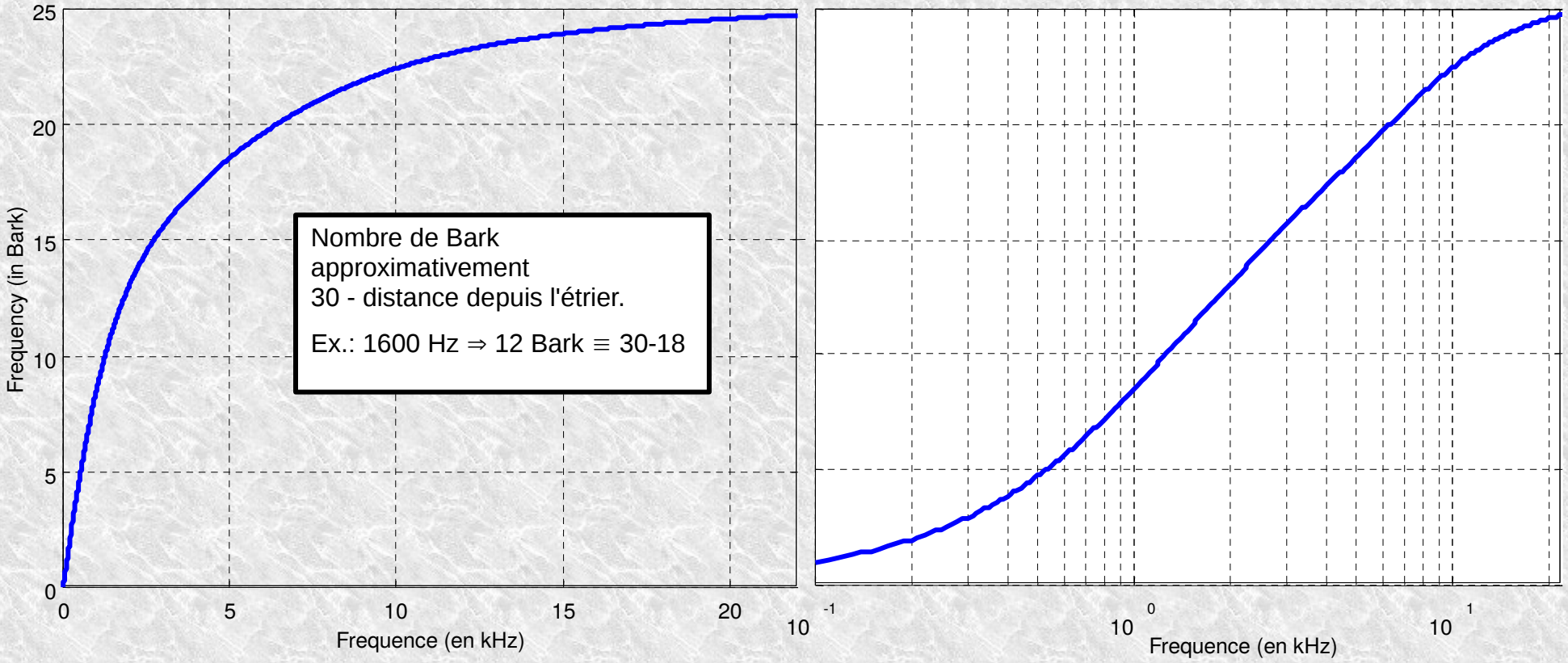


Lorsque la fréquence double, la position de l'excitation maximale varie linéairement

Logarithme...

Echelle de Bark

Relie la fréquence à la position de l'excitation de la membrane basilaire:



Niveau de pression sonore (SPL)

Deux tondeuses à gazon ne font pas deux fois plus de bruit qu'une seule tondeuse...

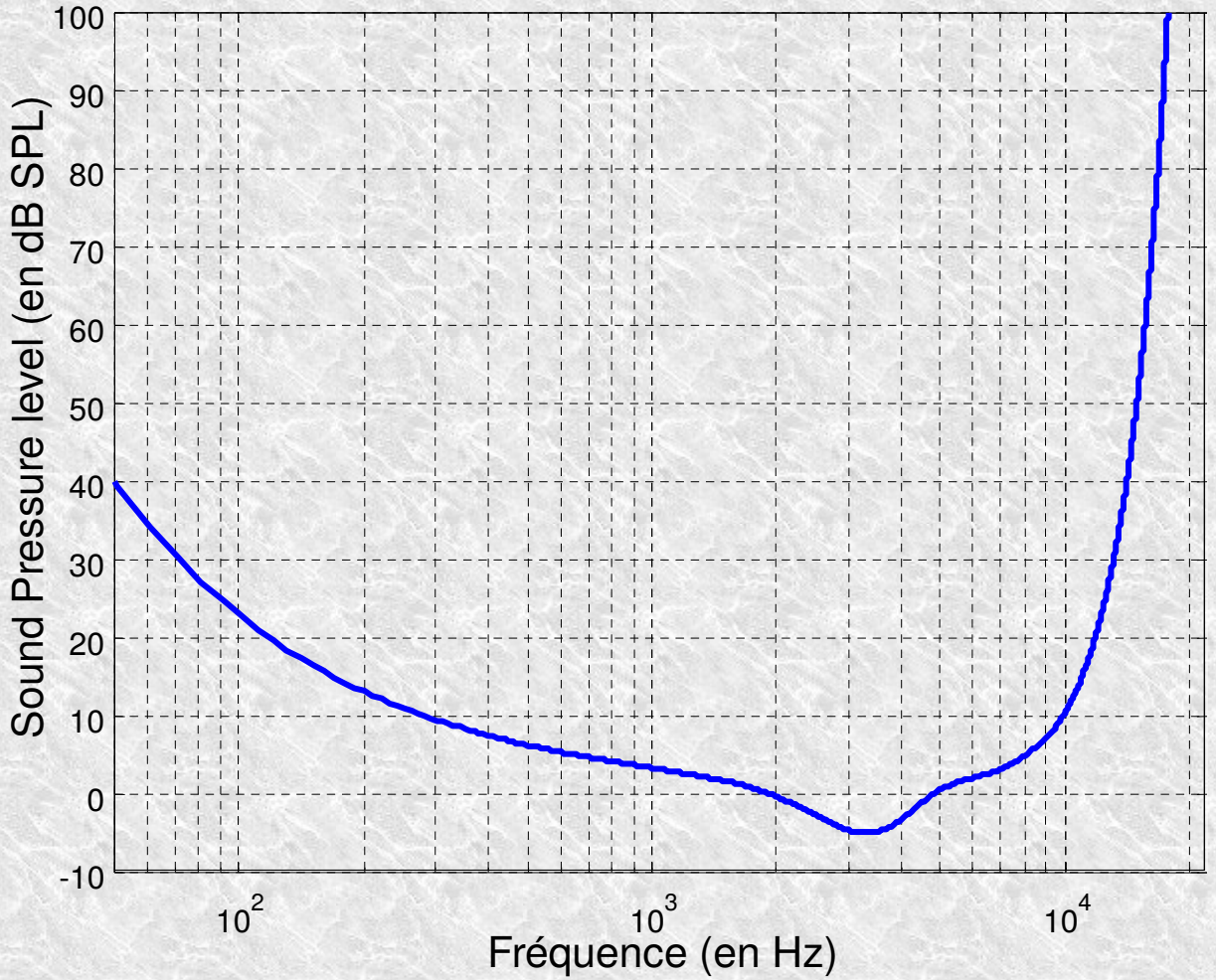
Lorsque la pression sonore double, la sensation ne double pas.

Logarithme...

SPL: Sound Pressure Level sur une échelle en dB relative à une pression de référence.

Les seuils de perception sont représentés sur une échelle de SPL en dB.

Le seuil d'audibilité



La pression de référence:
seuil d'inaudibilité à 2000 Hz

Représentation numérique

Le problème:

Comment faire le lien entre des nombres entiers représentant une amplitude et les dB SPL?

Une solution:

Une sinusoïde avec la plus petite amplitude (1 LSB) a un niveau correspondant au seuil absolu d'audibilité à 4kHz, soit -5 dB SPL.

$\frac{1}{2}A^2$ pour A=1 équivaut à -3dB + K $\equiv$ -5 dB SPL

$\Rightarrow$ K=-2 dB

16 bits couvrent environ 90 dB de dynamique.

Revenons au masquage

Les sons qu'on n'entend pas:

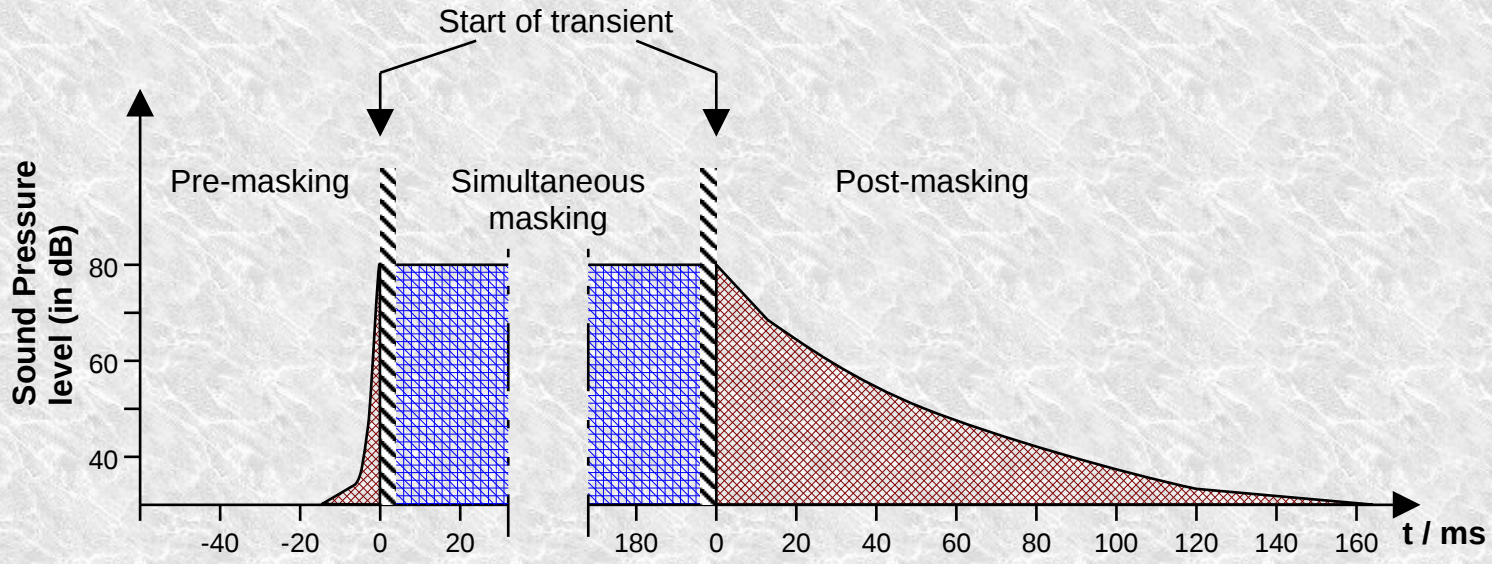
- Ceux qui sont sous le seuil d'audibilité
- Ceux qui sont masqués temporellement
- Ceux qui sont masqués fréquentiellement

Masquage temporel

Pré et post masquages

Différence entre le temps de perception des sons faibles et des sons forts

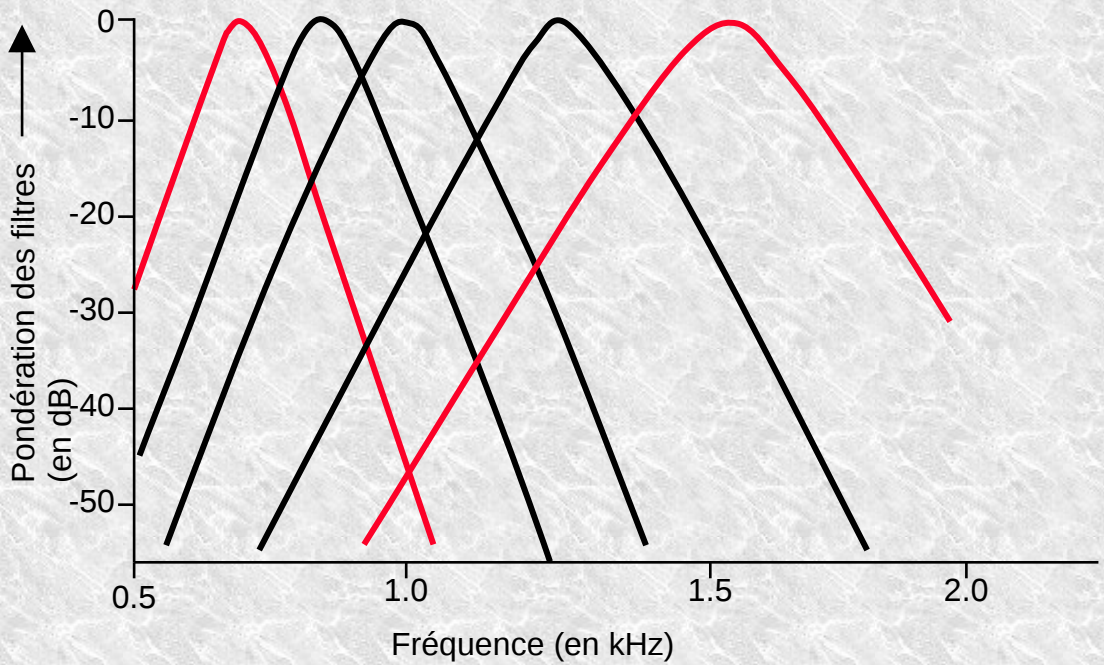
L'effet est implicitement utilisé dans les CODECs



Masquage spectral

Un son arrivant dans l'oreille se propage dans un ensemble de filtres passe-bande (les cils)

A chaque filtre correspond l'excitation d'une cellule nerveuse.

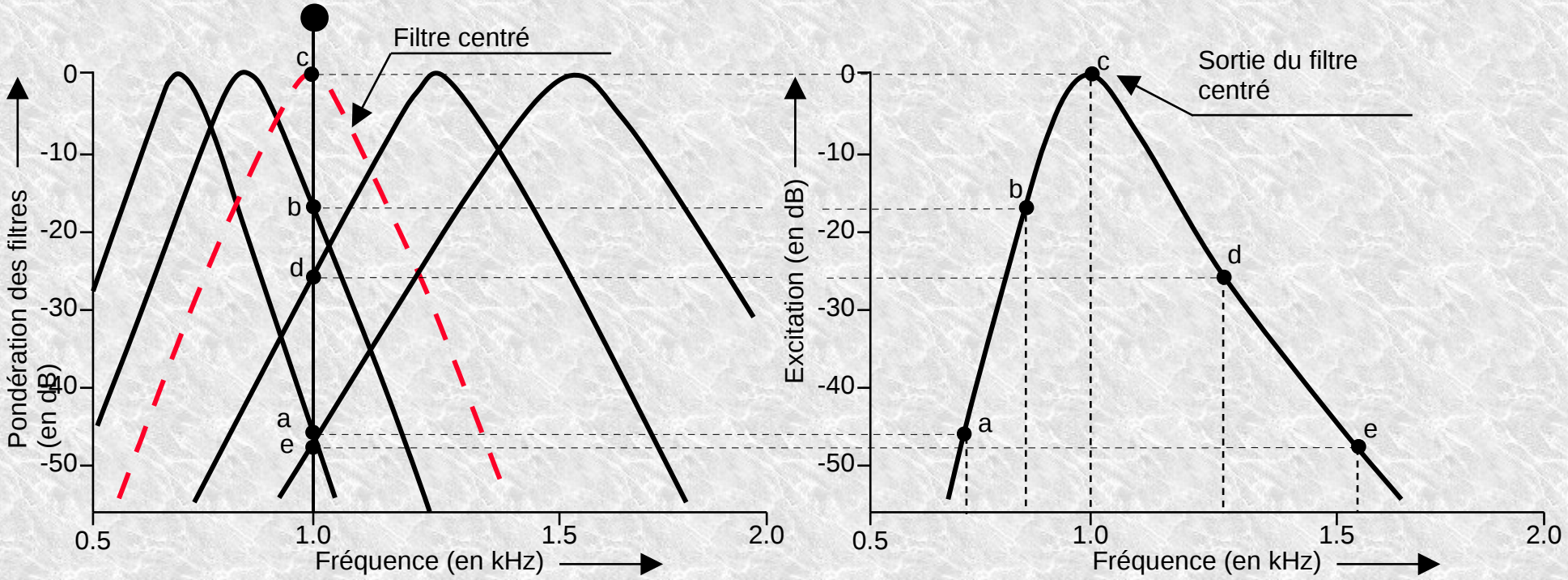


Masquage spectral

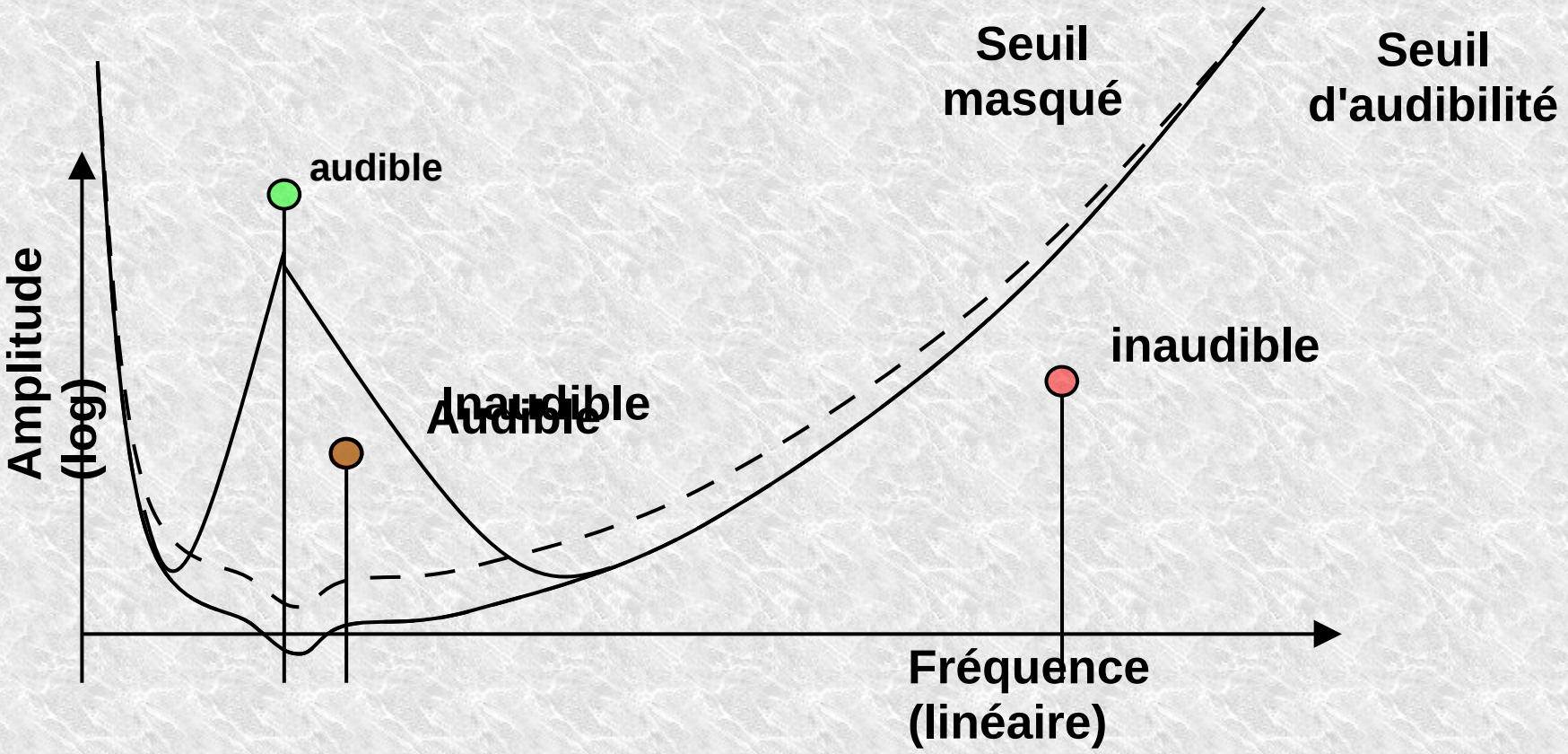
Une sinusoïde excite plusieurs filtres adjacents.

Surtout le filtre centré sur cette fréquence.

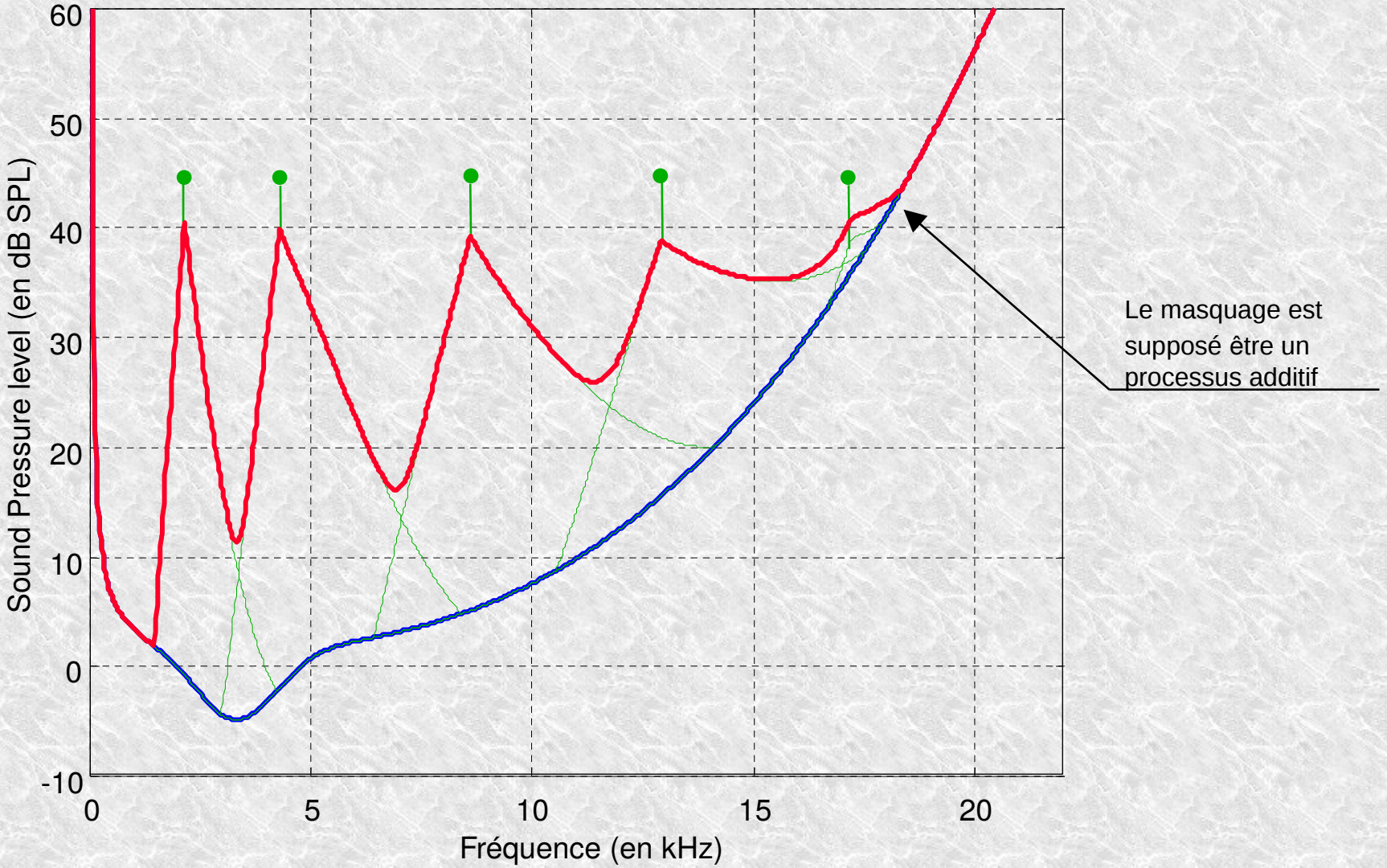
L'excitation causée se répand en partie.



Masquage spectral



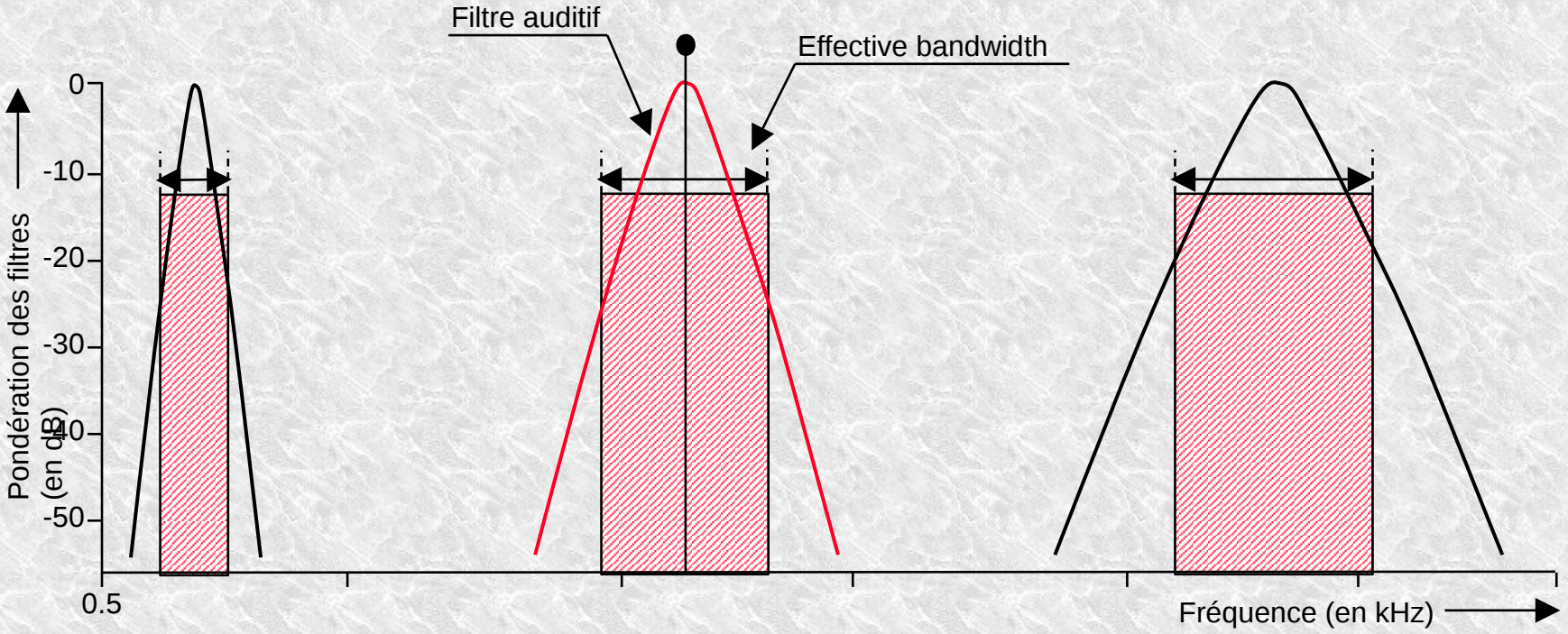
Seuil de masquage



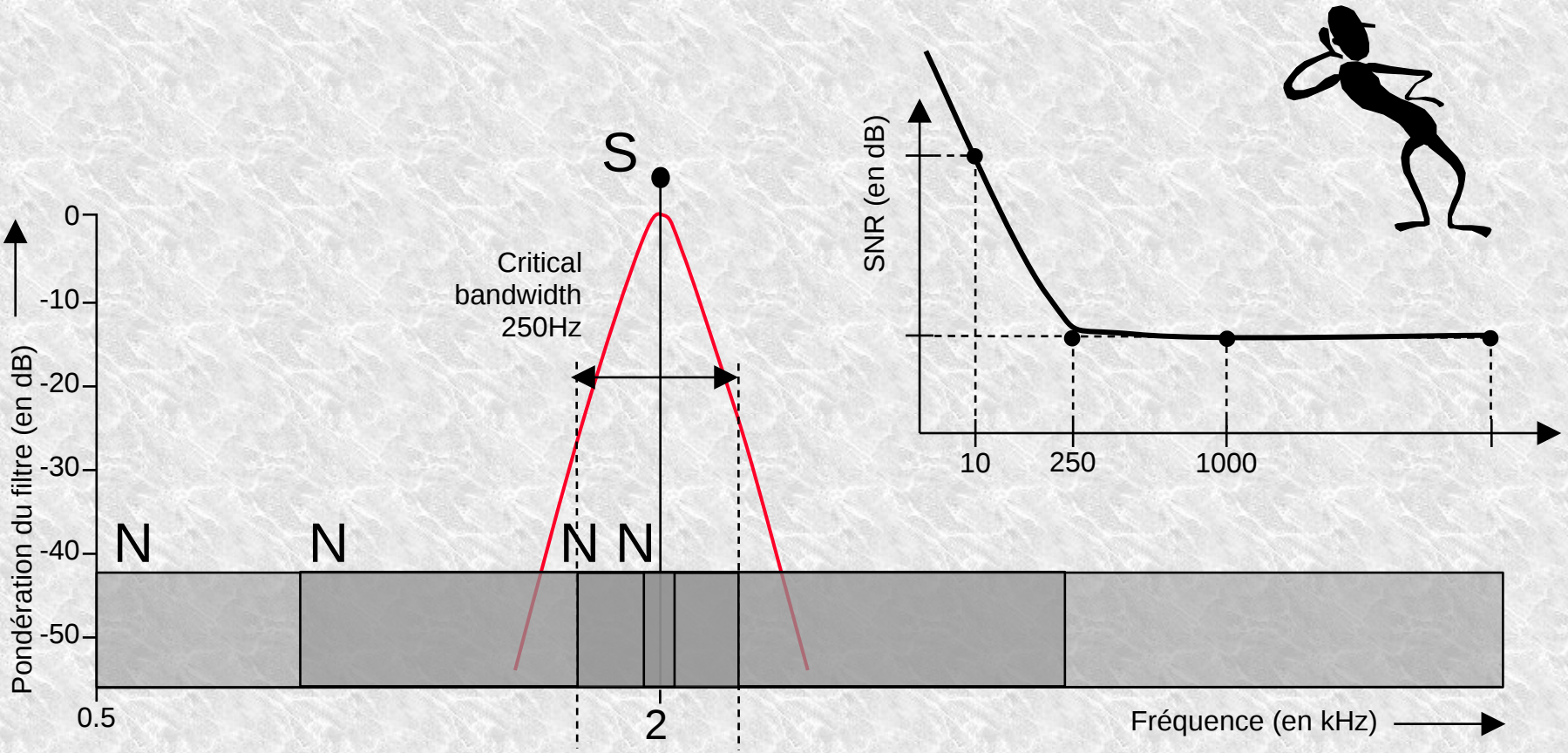
Largeur de bande critique

Le système auditif humain est capable d'intégrer la puissance reçue par chaque filtre auditif (cil).

Un bruit (large bande) faible peut masquer un son (sinusoïde) pourtant plus fort.



Largeur de bande critique

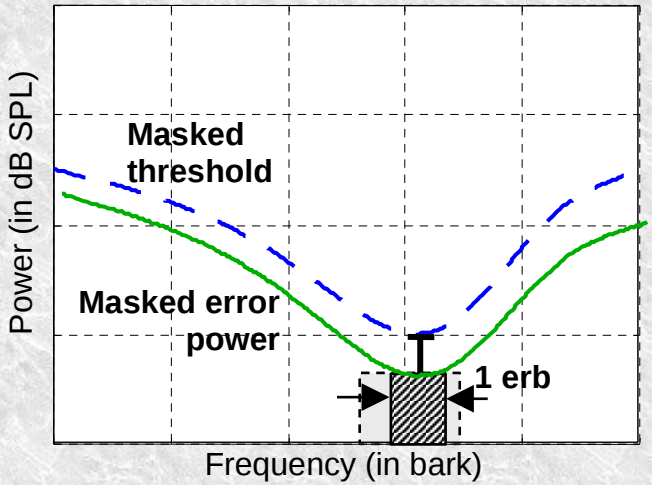


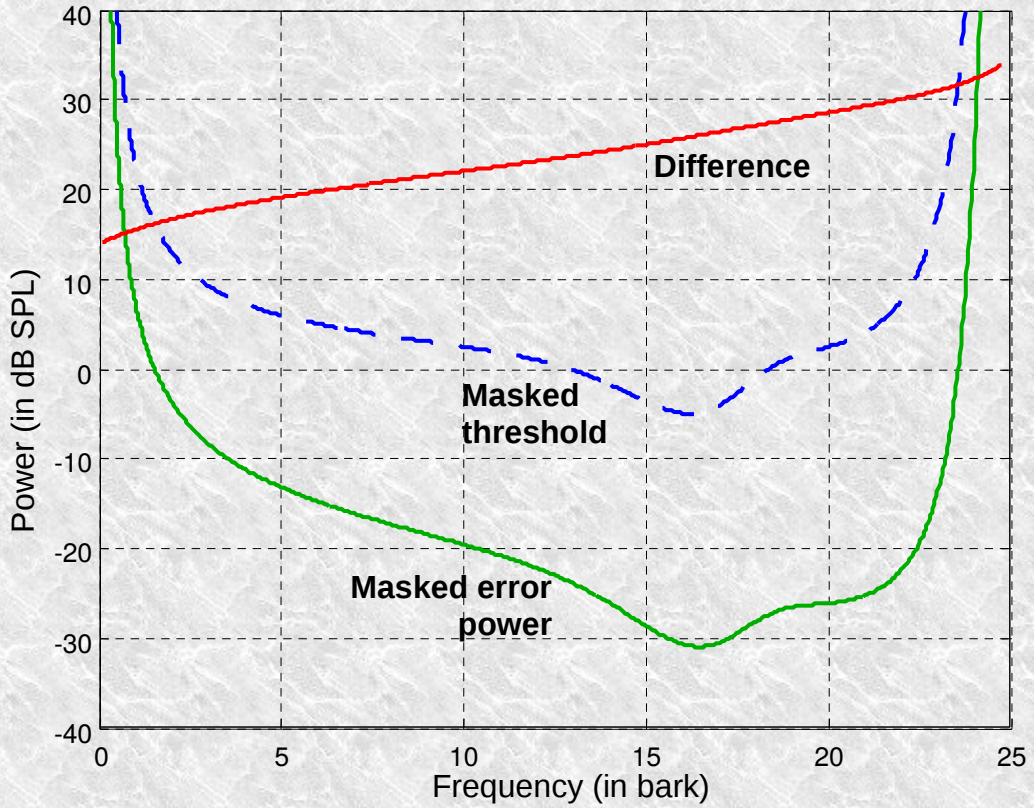
SNR s'améliore seulement si la bande a une largeur inférieure à la largeur de bande critique

Largeur de bande critique

La puissance intégrée dans une largeur de bande critique doit être inférieure au seuil d'audibilité

Nouvelle courbe corrigée





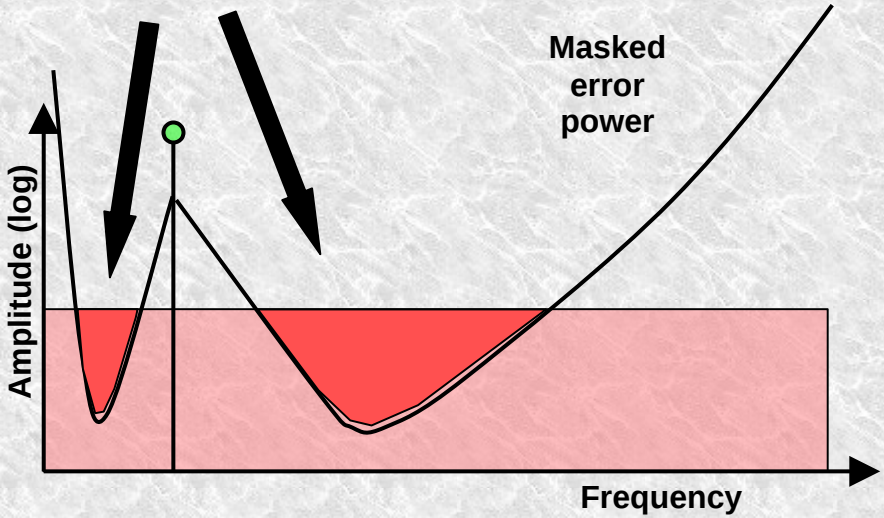
Application: Buried data

Example d'application du seuil de masquage:

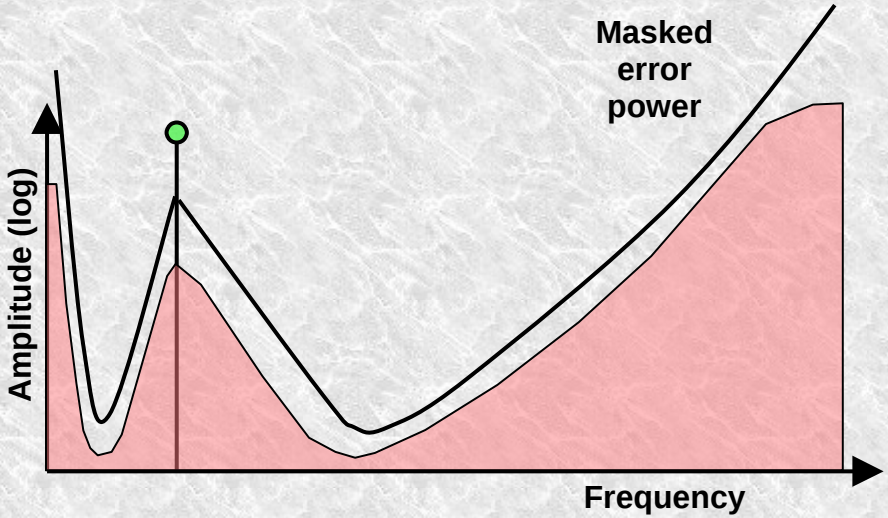
- On peut ajouter un bruit au signal jusqu'au seuil de masquage corrigé.
- Le signal audio reste perceptuellement identique au signal audio original.
- Le bruit représente en fait un data-stream.
- Ce data-stream peut être récupéré par après.
- Jusqu'à 1/3 de la capacité d'un CD peut être ainsi utilisée sans différence de perception.

Buried data

audible

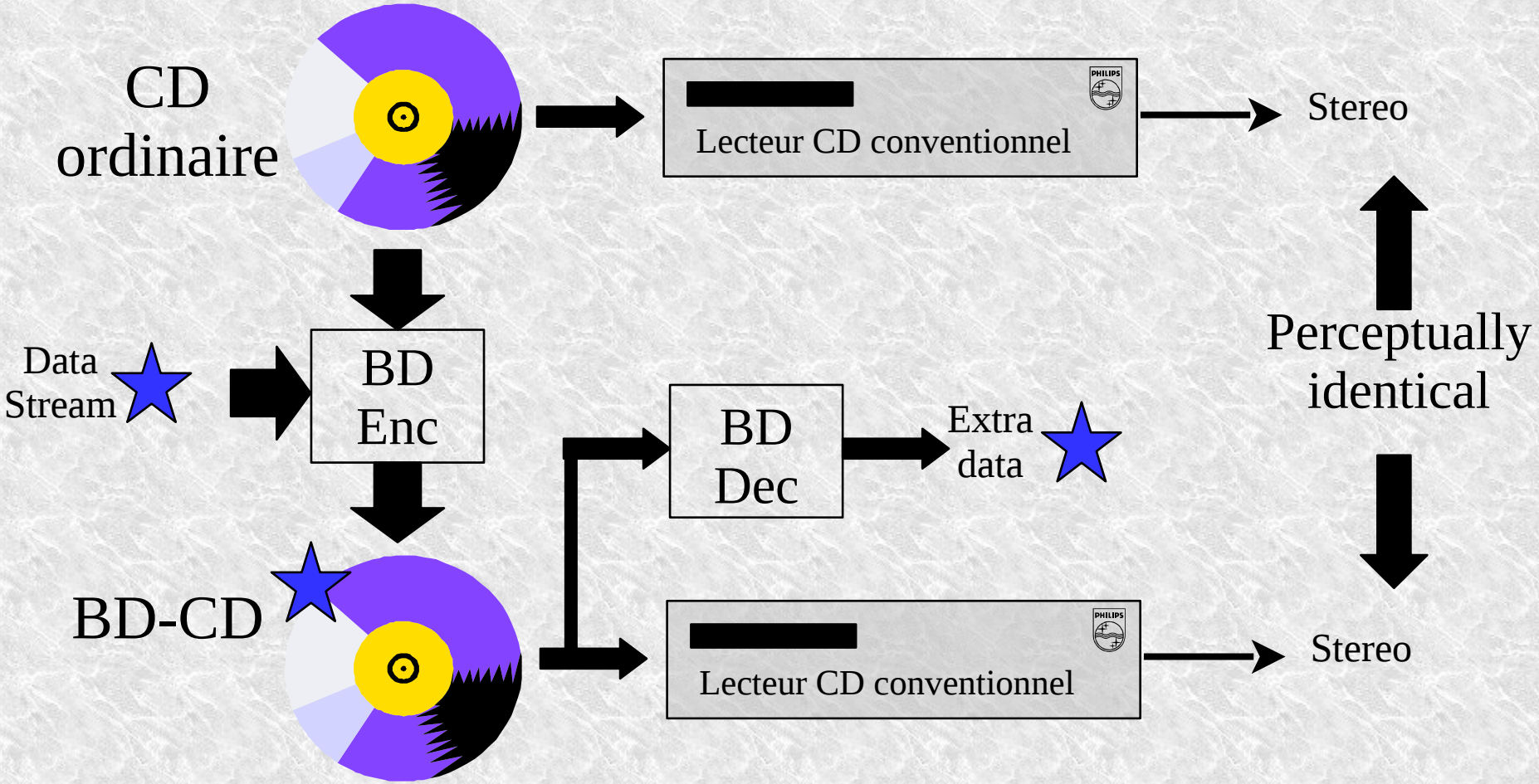


13dB SNR white noise



13dB SNR perceptually shaped noise

Buried data process



Références

Brian C.J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1997

IEEE ASSP Magazine, *Cochlear Modeling*, January 1985, Vol.2, No.1

JAES, *Buried Data Channels*, Vol.43, No.1/2, 1995 Jan/Feb.

Les principes utilisés pour le codage audio

Redondance et irrelevance

Redondance:

Information qui n'est pas nécessaire pour comprendre totalement le message

Infrmtn qi n'e pa ncesair pr comprndr ttlmnt l msg

Irrélévance:

Information qui fait partie du message, mais qui ne change pas la perception de ce message

Codage sans pertes <> avec pertes

	Avec pertes	Sans pertes
Propriété	Output ≈ Input	Output ≡ Input
Exploite	Redondance & Irrélévance	Redondance seulement
Tests perceptuels	Oui	Non
Compression ratio η	> 30	> 2
Bit rate	Fixe ou variable	Variable
Qualité	Subjective	Objective (max)

Les paramètres caractéristiques d'un CODEC audio

Qualité (bande passante (voix: 4kHz <> HQ Audio:24kHz), résolution perceptuelle (8 <> 24 bits))

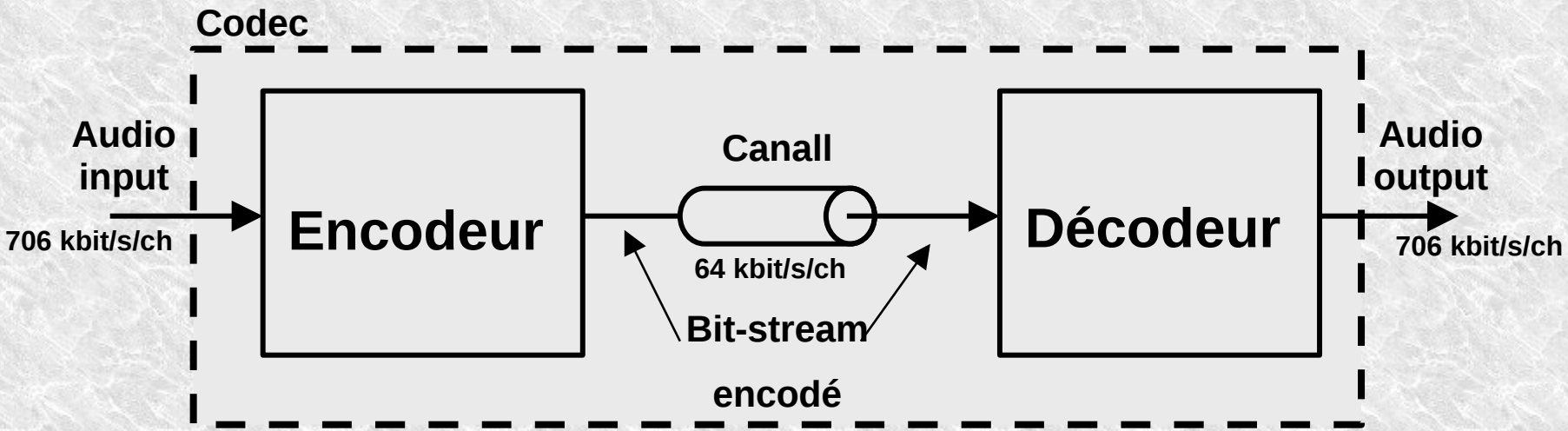
Compression ratio ($\eta_{lossless}$:2 … η_{lossy}>30 ...)

Complexité (ROM, RAM, MIPS, Vitesse, Symétrie enc/déc)

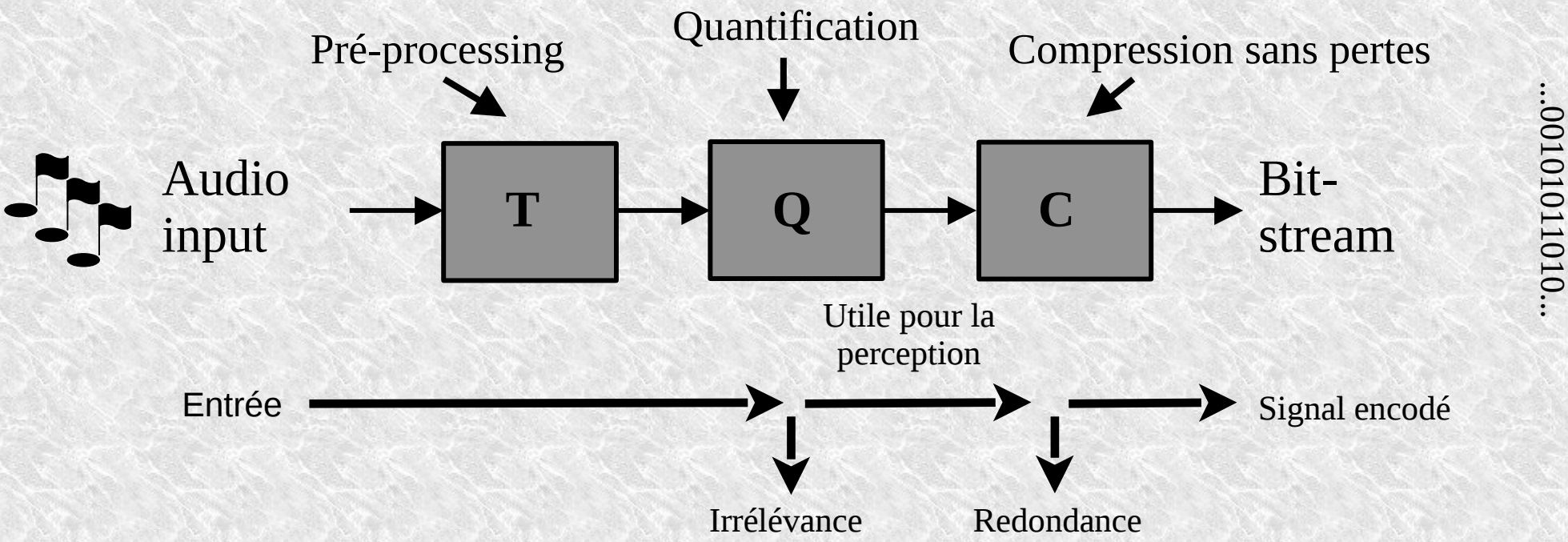
Flexibilité (Scalability, Retard, Canaux, Fonctionnalités suppl.)

CODEC

Codec ≡ encodeur + décodeur



Encodeur


Pré-processing
Quantification
Compression sans pertes
Audio input
T
Q
C
Bit-stream
...0010101011010...
Utile pour la perception
Entrée
Signal encodé
Irrélévance
Redondance

Encodeur

Pré-processing:

Banc de filtres

- Facilite l'analyse psychoacoustique
- Les puissances du signal intégrées en bandes de fréquence peuvent être mises facilement en relation avec le seuil de masquage.

'Noise-shaping' perceptuel

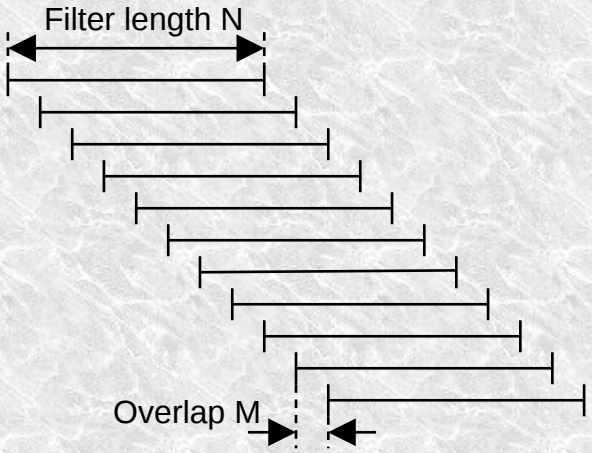
- Le bruit de quantification de chaque bande peut être choisi séparément selon le seuil de masquage.

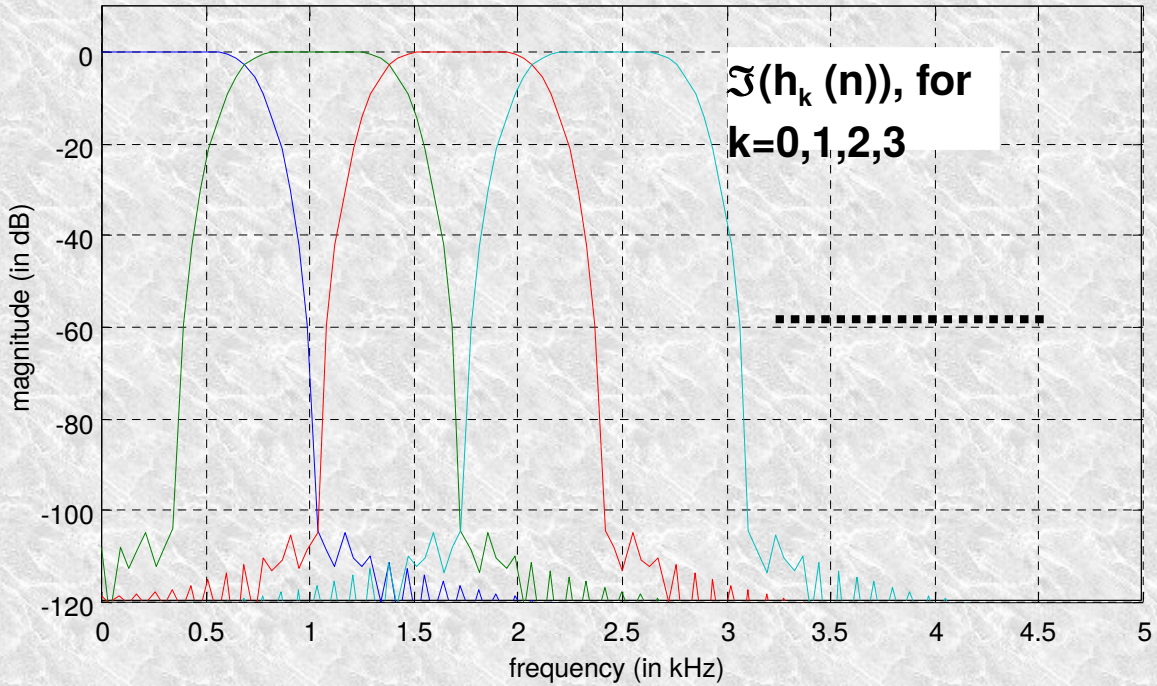
Subband coding

Large overlap dans le domaine temporel

Faible overlap dans le domaine spectral

⇒ **Les échantillons dans une bande sont corrélés**

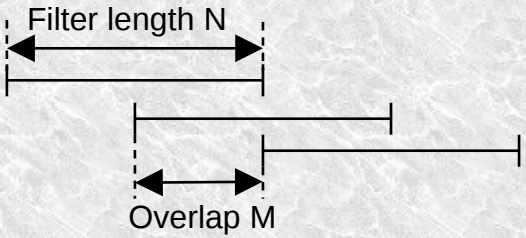


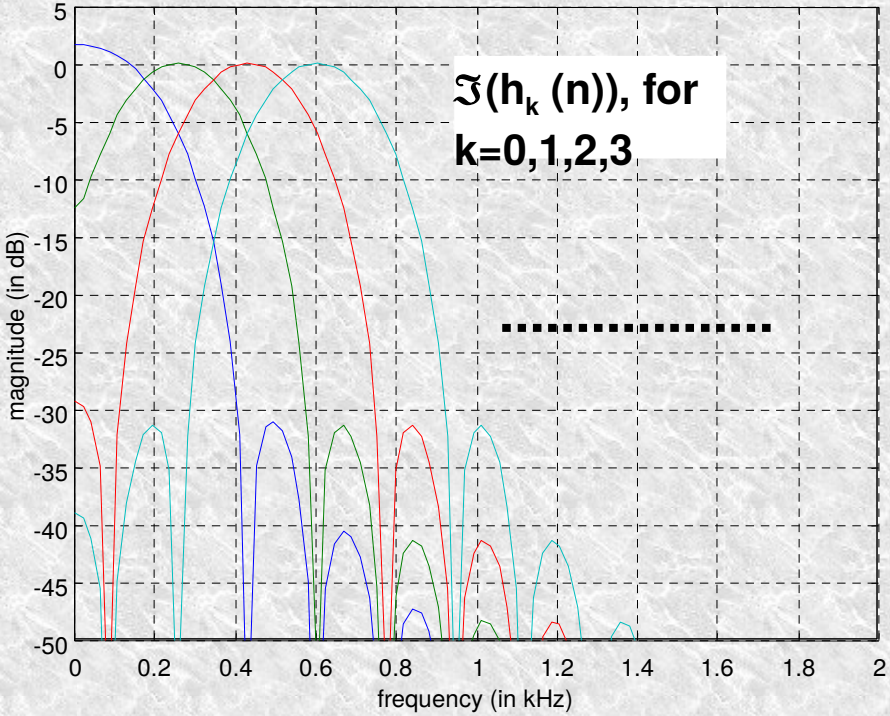


Transform coding

Large overlap dans le domaine spectral

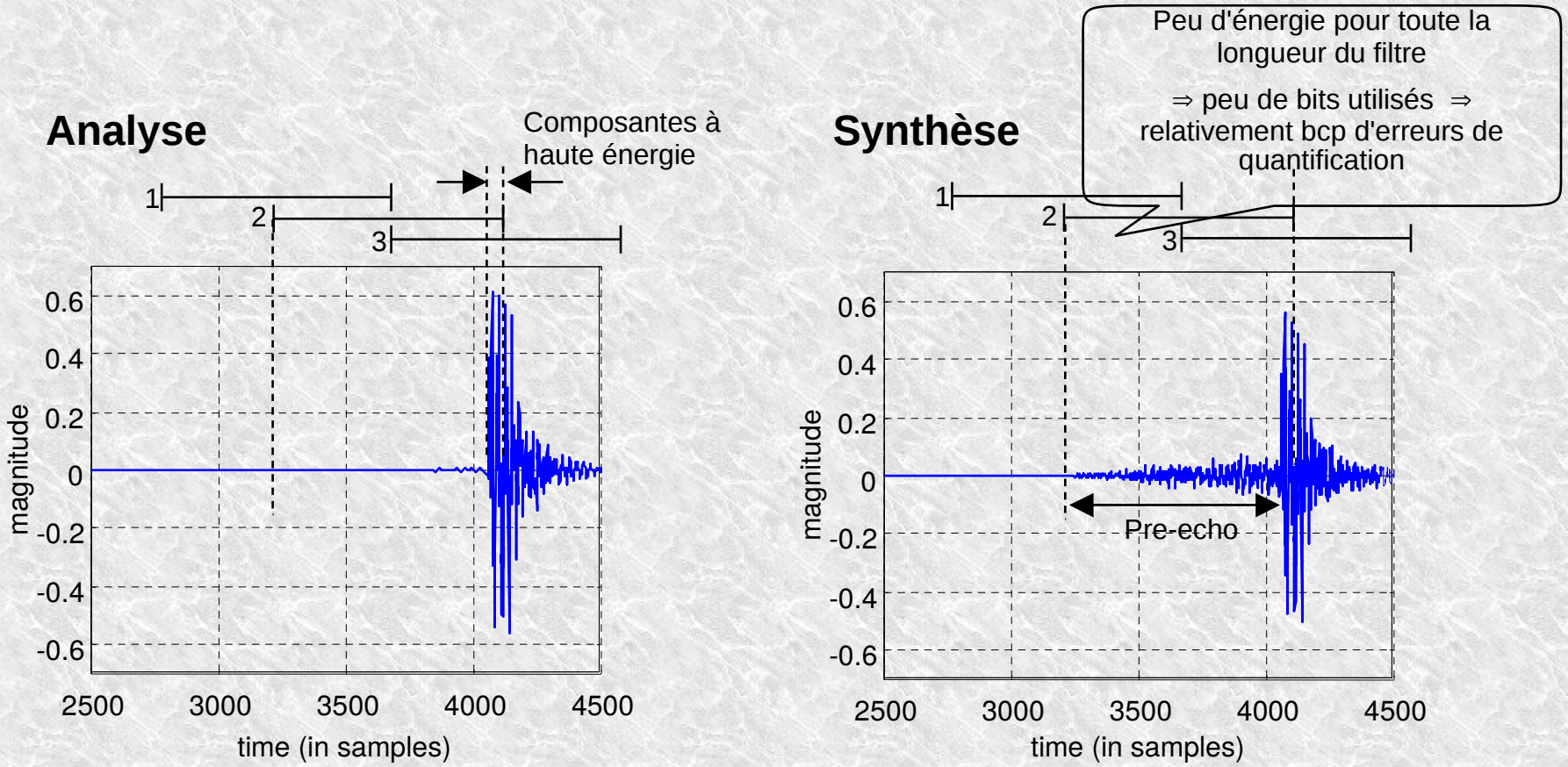
Faible overlap dans le domaine temporel





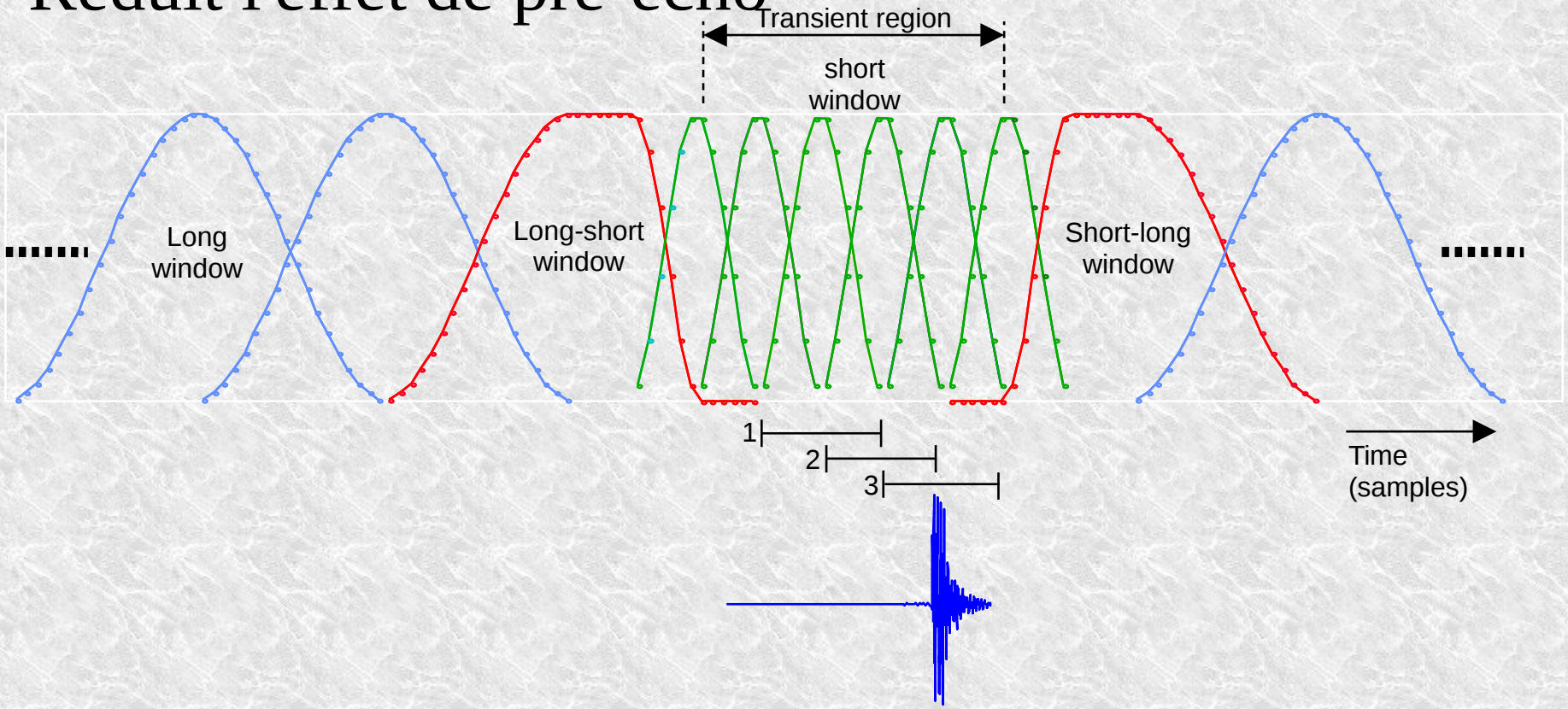
Erreurs au pré-écho

Erreurs de quantification en fréquentiel se répandent en temporel lors de la synthèse du son:



Changement de fenêtres

- Basculement vers des fenêtres plus courtes lors des transitoires
- Réduit l'effet de pré-écho



Subband vs. Transform

Valeurs typiques

Type	N/M	M	Implémentation
SBC	2...64	2...64	Polyphase matrix
TC	2	256...2048	FFT

Avec les outils adéquats (quantification, compression sans pertes, ...) subband coding et transform coding sont comparables en termes de performances/complexité

Codage sans pertes

Enlever la redondance

Huffman:

- Les mots les plus fréquents reçoivent les symboles les plus courts

L'alphabet Morse international

A	.-	J	.---	S	...	0	-----
B	-...	K	-.-	T	-	1	.----
C	-.-.	L	.-..	U	..-	2	..---
D	-..	M	--	V	...-	3	...--
E	.	N	-.	W	.--	4	....-
F	..-.	O	---	X	-..-	5	.....
G	--.	P	.--.	Y	-.--	6	-....
H	....	Q	--.-	Z	--..	7	--...
I	..	R	.-.			8	---..
						9	----.

36 symboles Morse nécessitent 6 bit -> 64 possibilités
(A=000001, B=000010, ...)

En exploitant les probabilités des lettres dans les mots, la longueur moyenne des symboles en Morse (A .. Z) ≈ 3.2

Codage stéréo

Pour les fréquences élevées, l'oreille humaine:

- Devient moins sensible aux différences inter-aurales
- Détermine la direction à partir des différences d'intensité uniquement

Intensity Stereo Coding

Codage stéréo

- Sum difference (MS) coding
  - Code L-R et L+R séparément (efficace pour les basses fréquences)
    - L+R: év. 1 bit en plus
    - L-R: nettement moins de bits si canaux fort corrélés
  - Ajustement du modèle perceptuel et de la quantification nécessaires

$$L = [(L-R) + (L+R)] / 2$$

$$R = [(L+R) - (L-R)] / 2$$

D'autres outils

TNS: Temporal noise shaping

Exploite post-masquage autour des transitoires

PNS: Perceptual noise shaping

Substitue les bandes de fréquence par une synthèse de bruit blanc dans le décodeur (applaudissements=bruit blanc...)

LTP: Long term prediction

Réduction supplémentaire de la corrélation dans les sous-bandes par prédiction dans le domaine temporel

TWIN-VQ

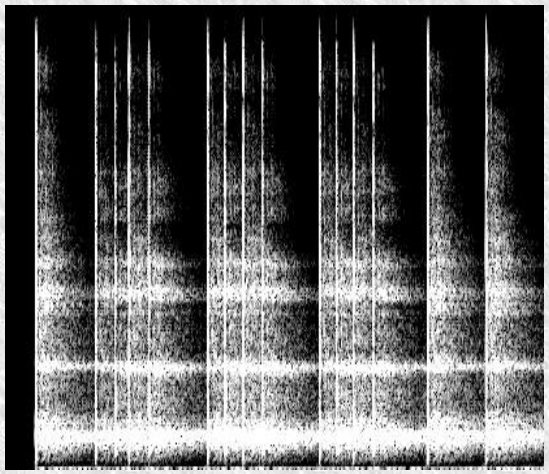
Autre type de quantification dans les sous-bandes

...

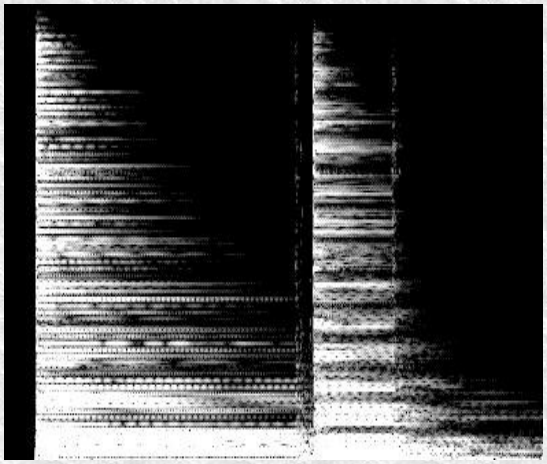
Différentes façons d'encoder...

Pour différents types de sons

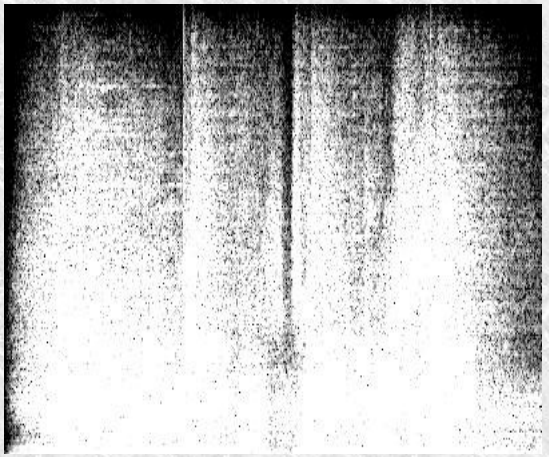
Castagnettes



Clavecin



Heavy metal



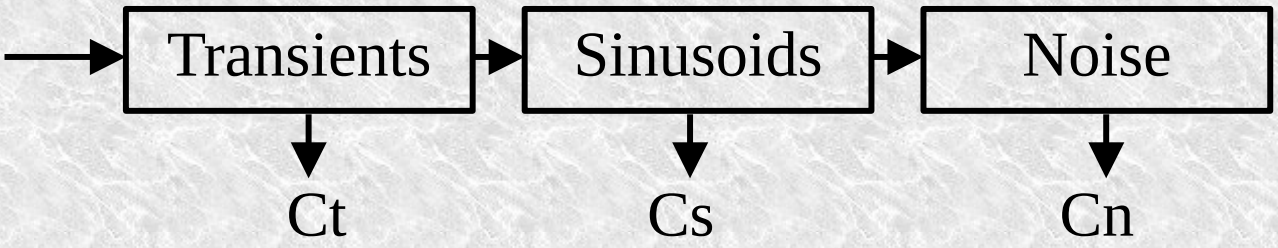
Model based coding

Classification pour tous signaux musicaux/voix

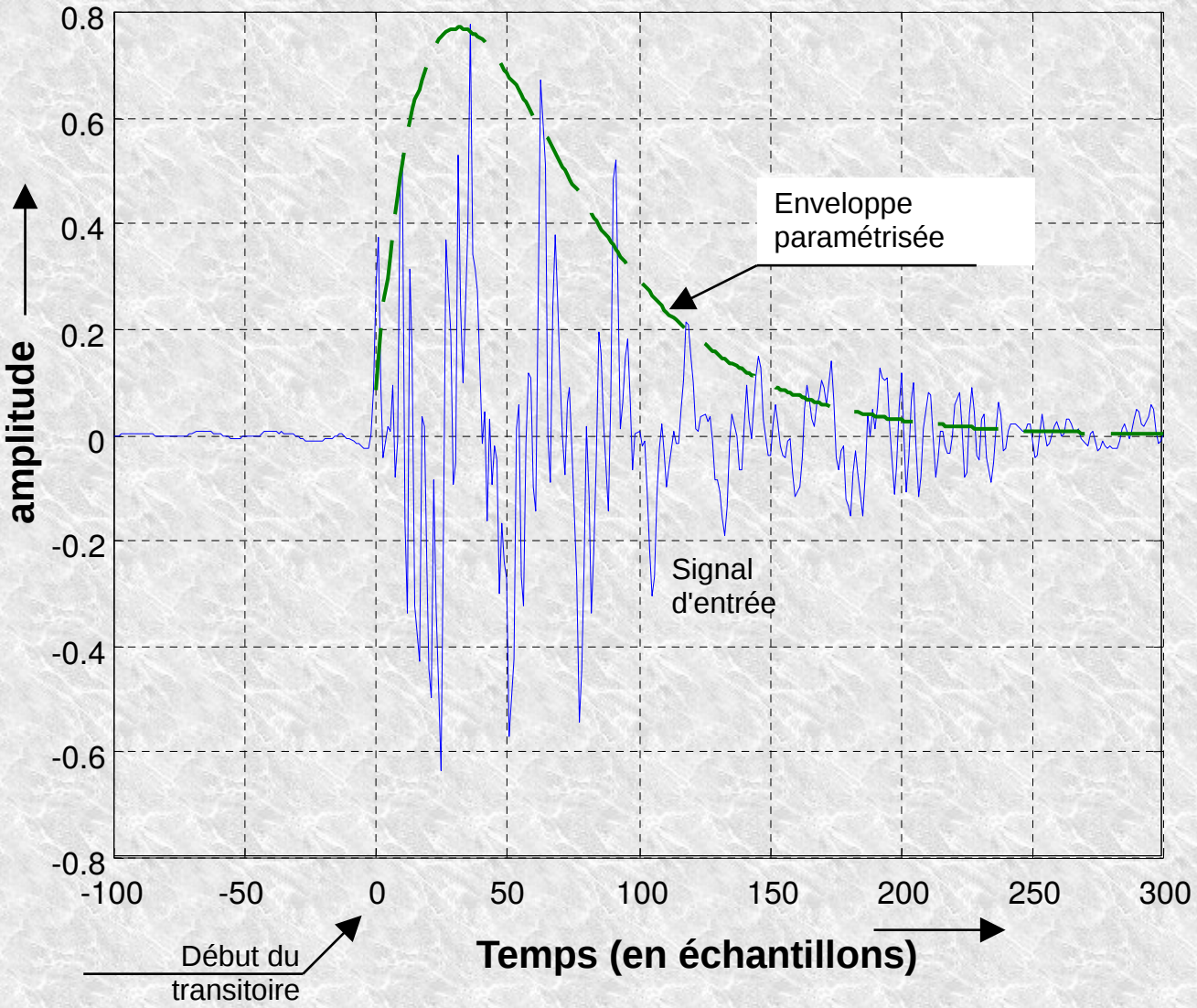
Description paramétrique des transitoires

Bruit généré par le décodeur (cf PNS)

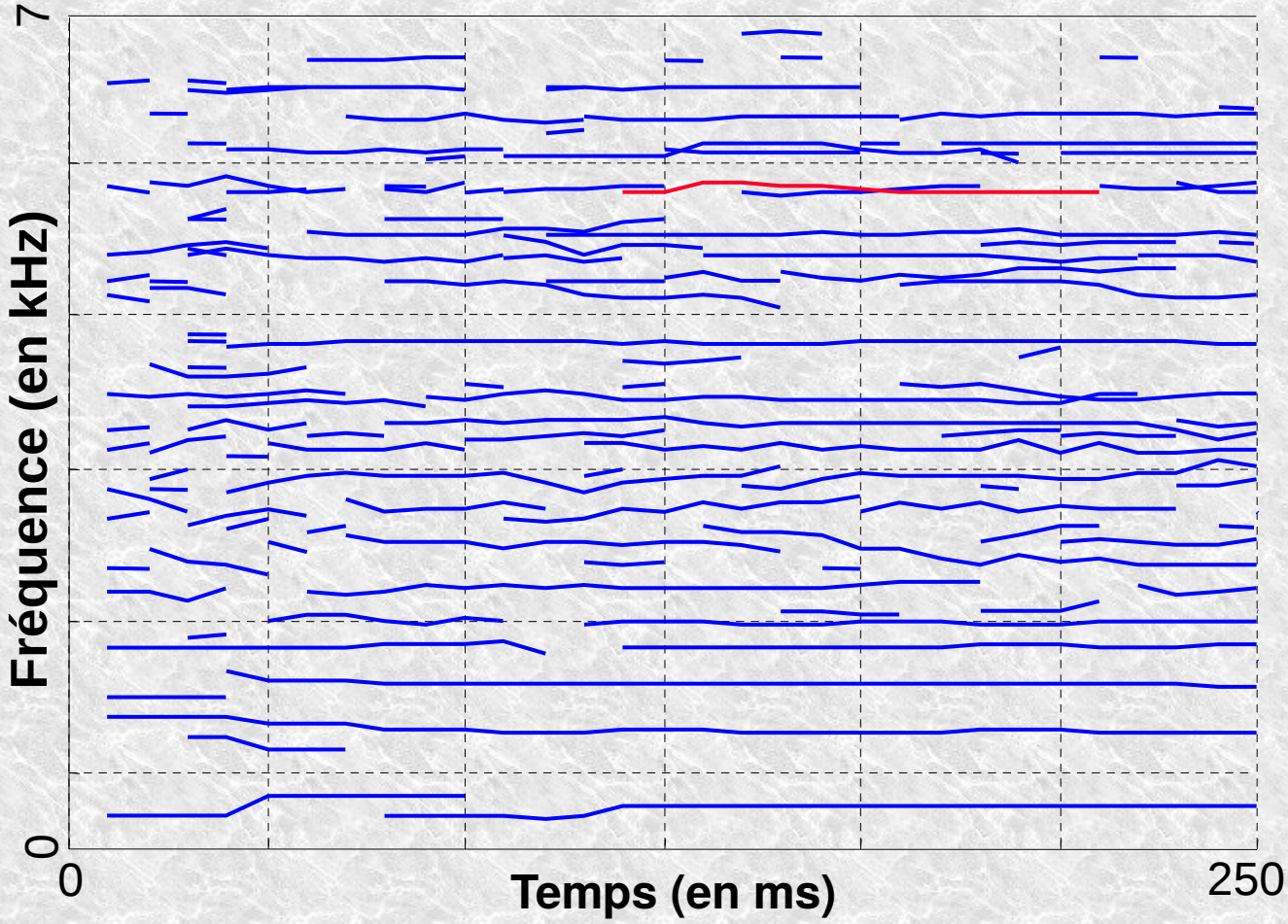
Sinusoïdes peuvent être représentées efficacement



Les transitoires



Les sinusoïdes



Les CODECs disponibles

MPEG Layer III

Fait partie du standard MPEG-1 depuis 1993

Développé par FhG et AT&T

Basé en partie sur le MPEG Layer II

THE Internet codec ⇒ mp3

De-facto mode is stereo at 128 kbit/s

High quality at 160 kbit/s

Le MP3 n'est PAS libre!

- De nombreux brevets couvrent le MP3
  - La plupart de Thompson et Fraunhofer
- Si l'utilisateur peut en bénéficier gratuitement,
  - Rien ne garantit que cela le restera dans le futur
  - Il n'en est pas de même pour l'encodage...

Royalty Rates

Home
Overview
PC Software Applications
Hardware Products
ICs / DSPs
Games
Electronic Music Distribution / Broadcasting / Streaming

Overview

PC Software Applications		
mp3	Decoder	· **US$ 0.75** per unit or **US$ 50 000.00 - US$ 60 000.00** one-time paid-up
	Encoder / Codec	· **US$ 2.50 - US$ 5.00** per unit
mp3PRO	Decoder	· **US$ 1.25** per unit or **US$ 90 000.00** one-time paid-up
	Encoder / Codec	· **US$ 5.00** per unit

Hardware Products		
mp3	Decoder	· **US$ 0.75** per unit
	Encoder / Codec	· **US$ 2.50 - US$ 5.00** per unit
mp3PRO	Decoder	· **US$ 1.25** per unit
	Encoder / Codec	· **US$ 5.00** per unit

ICs / DSPs
For available software, supported platforms, porting and licensing options, please contact us at info@mp3licensing.com.

Games	
mp3	· **US$ 2 500.00** per title
mp3PRO	· **US$ 3 750.00** per title

Electronic Music Distribution / Broadcasting / Streaming	
mp3	· **2.0 %** of related revenue
mp3PRO	· **3.0 %** of related revenue

Advanced audio coding: AAC

Addendum au standard MPEG-2 depuis 1997

Développé par Dolby, FhG, AT&T, Sony et d'autres

Trois profils: Main, Low complexity, Sampling rate scalable* (Sony))

Twin-VQ (MPEG4 tool) pour les faibles bit-rates*

Qualité stéréo HiFi à 128 kbit/s

Qualité stéréo mp3-like à 80-96 kbit/s

* peu de preuves de réelle efficacité

Autres CODECs propriétaires

AC 3 (Dolby)

ATRAC (Sony) ⇒ MD, ATRAC3

Apt-x

E-PAC (ATT)

Qsound (Qdesign)

...

Qualité <> bit-rate

	Qualité mp3 @128 kb/s	Transparence
MPEG Layer I	224	384
MPEG Layer II	160	192 à 224
MPEG Layer III	128	160
AAC	80 - 96	128
MPEG-4 CELP	16 (uniq voix)	n.a.
AC-3	160	192 à 224
ATRAC3	128 ?	160 ?
ePAC	80 - 96	128
QDMC	n.a.	n.a.
WMA	?	?
SSC	Cible 37	?

Jugement de qualité objectif

- Les spécifications traditionnelles de la HiFi comme le SNR sont sans valeur marketing pour les CODECs audio (SNR 13 dB !)
- Il faut faire des mesures qui tiennent compte du caractère perceptuel utilisé dans les CODECs
- Méthode rapide et bon marché pour évaluer les CODECs

Où est le meilleur modèle?

Dans le CODEC?
ou dans l'outil de mesure objectif?

Jugement de qualité subjectif

Le seul outil correct mais...

- Cher
- Lent
- Interne ou externe?
- Formel ou informel?
- Demande une bonne préparation
- L'interprétation des résultats est délicate (stats)

Références

Ted Painter and Andreas Spanias, Perceptual Coding of Digital Audio, IEEE , 2000


OggVorbis

Ogg Vorbis

Ogg Vorbis est du logiciel libre, donc gratuit et ouvert pour l'éternité!

L'idée initiale a germé en 1993

Le projet Ogg Vorbis a débuté à l'automne 1998, peu après que Fraunhofer décida de reprendre le mp3 et de poursuivre tous les projets libres autour du mp3

Ogg Vorbis

La license mp3 était prohibitive pour ces [musiciens un peu programmeurs] qui se sont transformés peu à peu en [programmateurs un peu musiciens]!

C'est un exemple typique de la communauté des logiciels libres: Si quelquechose manque, on ne s'en plaint pas, on le crée...

Le projet continue à évoluer: meilleur encodeur, outils plus rapides, error correction, etc.

Ogg? Vorbis? Xiph?

Ogg est un format de bitstream très complet, aussi bien pour l'audio que la vidéo (cf Mpeg4)

Vorbis et le nom du CODEC audio

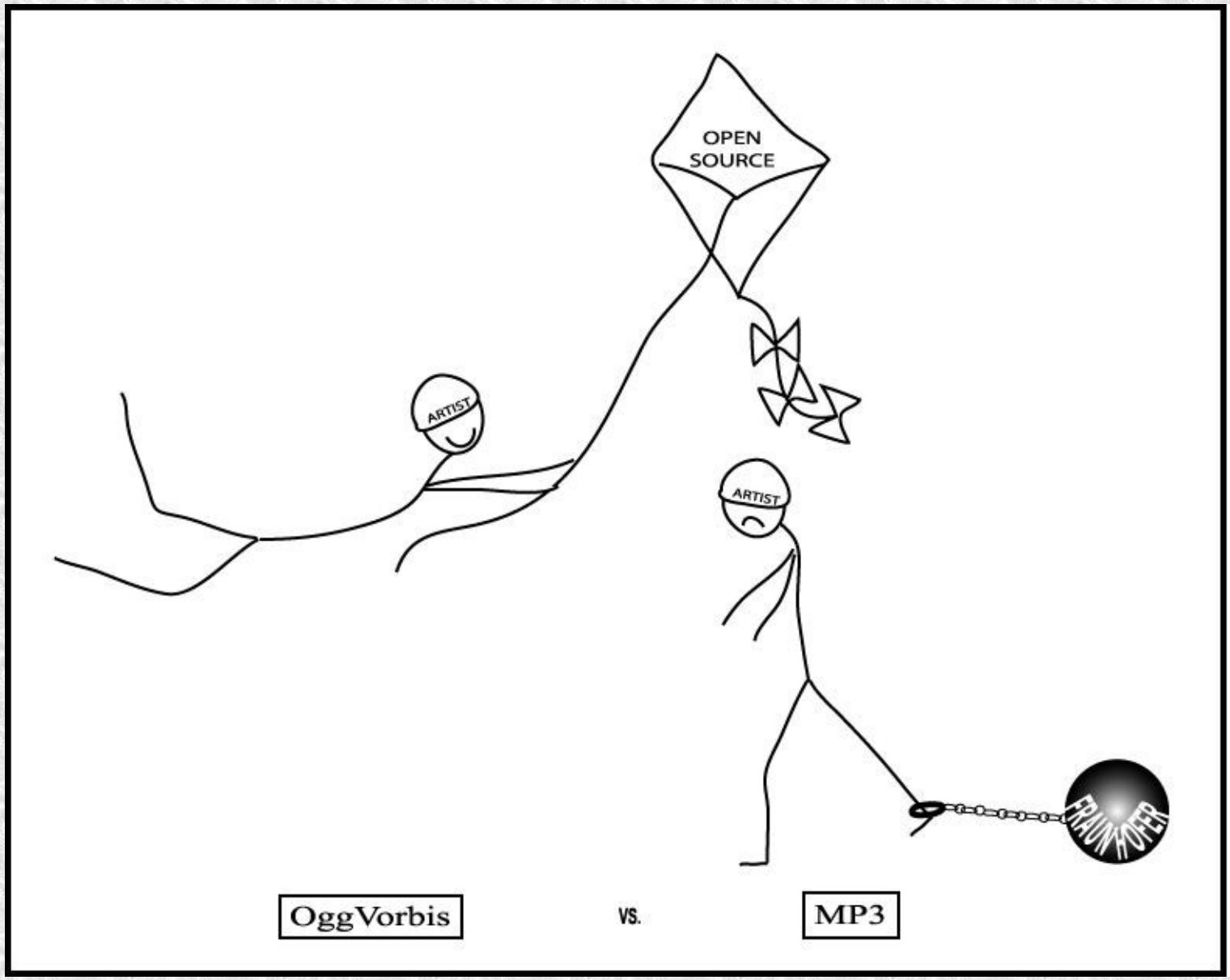
Xiph est la société hébergeant les principaux développeurs, Xiph nous a déjà donné l'excellent cdparanoia, outil d'extraction audio, sous la license libre GPL.

Tarkin est le prochain grand projet de Xiph: avoir également un CODEC libre pour la vidéo

Tous ces noms et symboles tordus sont expliqués:
http://www.xiph.org/xiphname.html

Et qu'est-ce que ça donne?

- Meilleur que le mp3
- Equivalent à l'AAC et au mp3pro?
- Exemples...
- Streaming radio avec la BBC...
- Supporté par de plus en plus de softs: Winamp, Freeamp, JAVA, etc, etc
- VBR: penser en terme de qualité plutôt que de bitrate (CBR ou ABR)


OPEN
SOURCE
ARTIST
ARTIST
FRAUNHOFER
OggVorbis
VS.
MP3

Références

Le site officiel

http://www.xiph.org/ogg/index.html

http://www.xiph.org/about.html

Why artists should be using Ogg Vorbis

http://itw.itworld.com/GoNow/a14724a50163a75981044a4

Ogg Vorbis en profondeur

Principales composantes

Un modèle psycho-acoustique

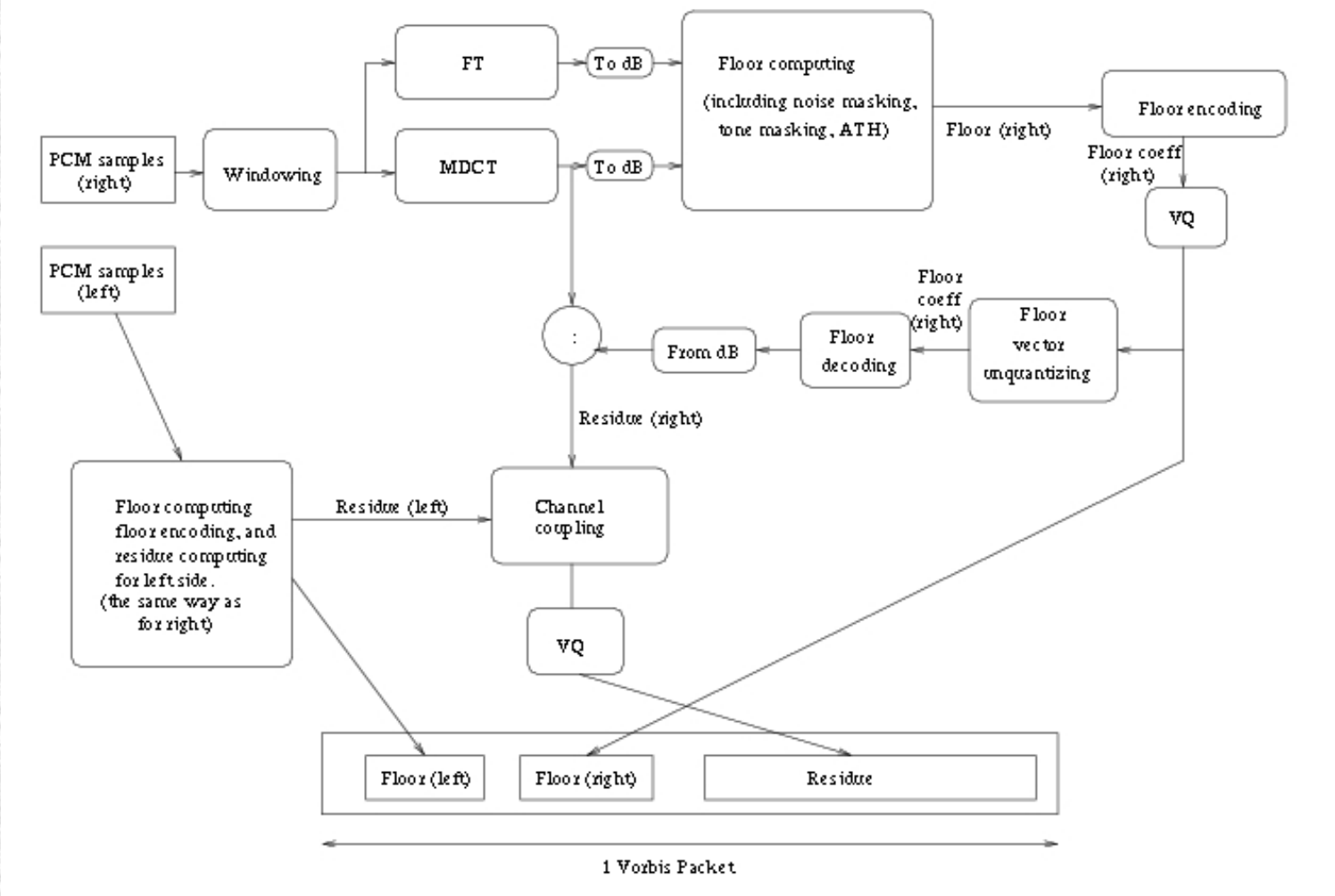
Une courbe de masquage

La quantification vectorielle (VQ)

Le couplage entre canaux (channel coupling)

Le codage sans pertes de Huffman

Fonctionnement de l'encodeur



Windowing

Problème du pré-écho

Emploi de fenêtres plus courtes lors des transitoires

Pourquoi pas tout le temps?

La résolution fréquentielle est moins bonne

C'est le compromis
résolution temporelle <> résolution spectrale

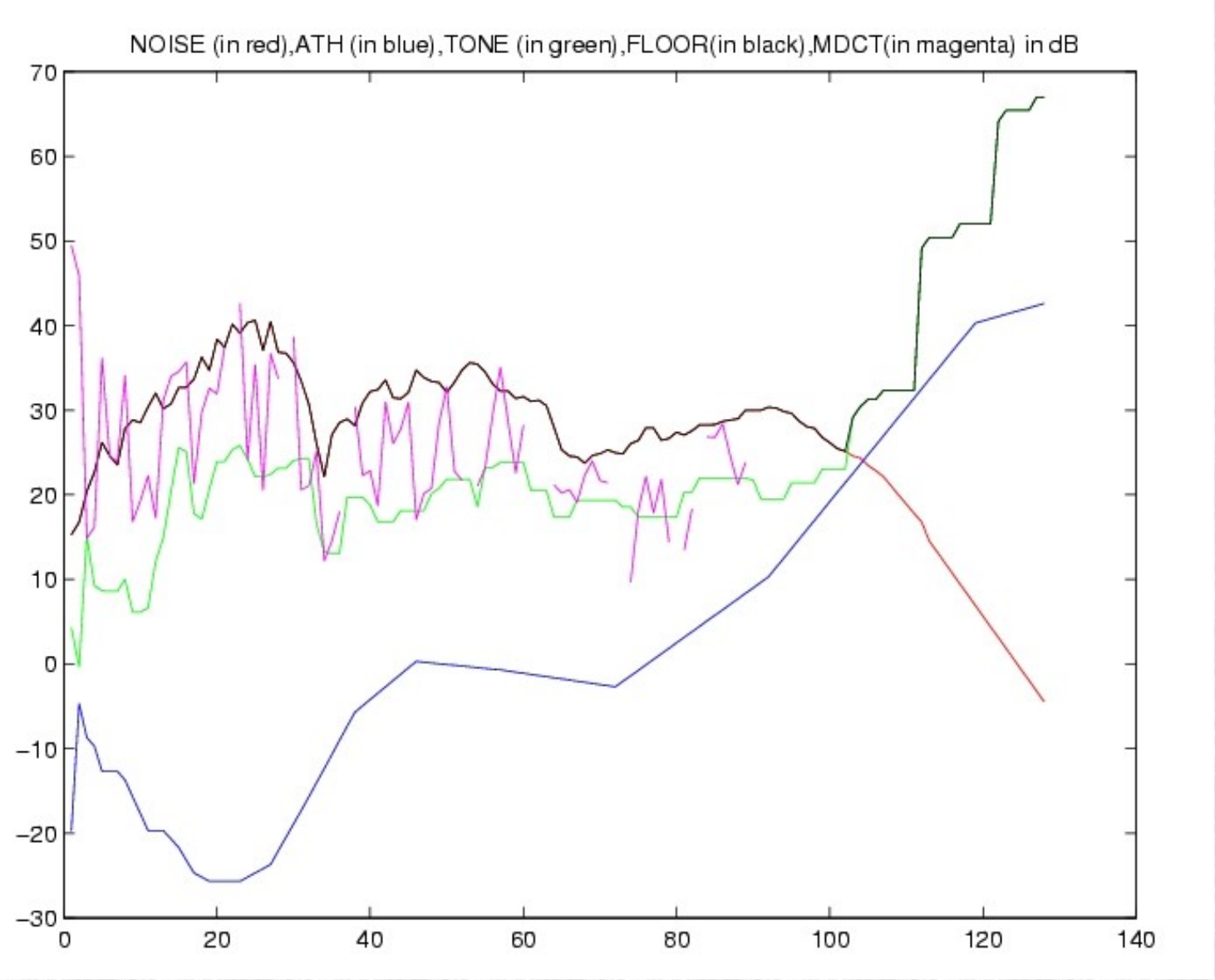
Floor

Le modèle psycho-acoustique détermine les fréquences inaudibles et le bruit de quantification permis (modèle basé sur les travaux de Robert Ehmer, années '50)

Le floor est la courbe maximale de 3 autres:

- Le max. de masquage par le bruit (cf expérience)
- Le masquage dû aux sons (tone masking)
- L'ATH: absolute threshold of hearing en dernier recours, tout en tenant compte que l'utilisateur peut augmenter le son lors des 'blancs'

Exemple de floor



Encodage et décodage du floor

Le floor étant déterminé, il reste à coder les résidus, différence entre le signal original et le floor.

Mais le floor sera quantifié dans le bitstream donc

- on va d'abord calculer cette quantification,
- se servir de la sortie quantifiée, la décoder
- puis seulement la soustraire au signal
- Ainsi l'erreur de quantification du floor ne se répercutera pas sur le calcul des résidus.

Les résidus

Ils ont une très faible dynamique

Obtenus séparément pour chaque canal

On peut leur appliquer le channel coupling:

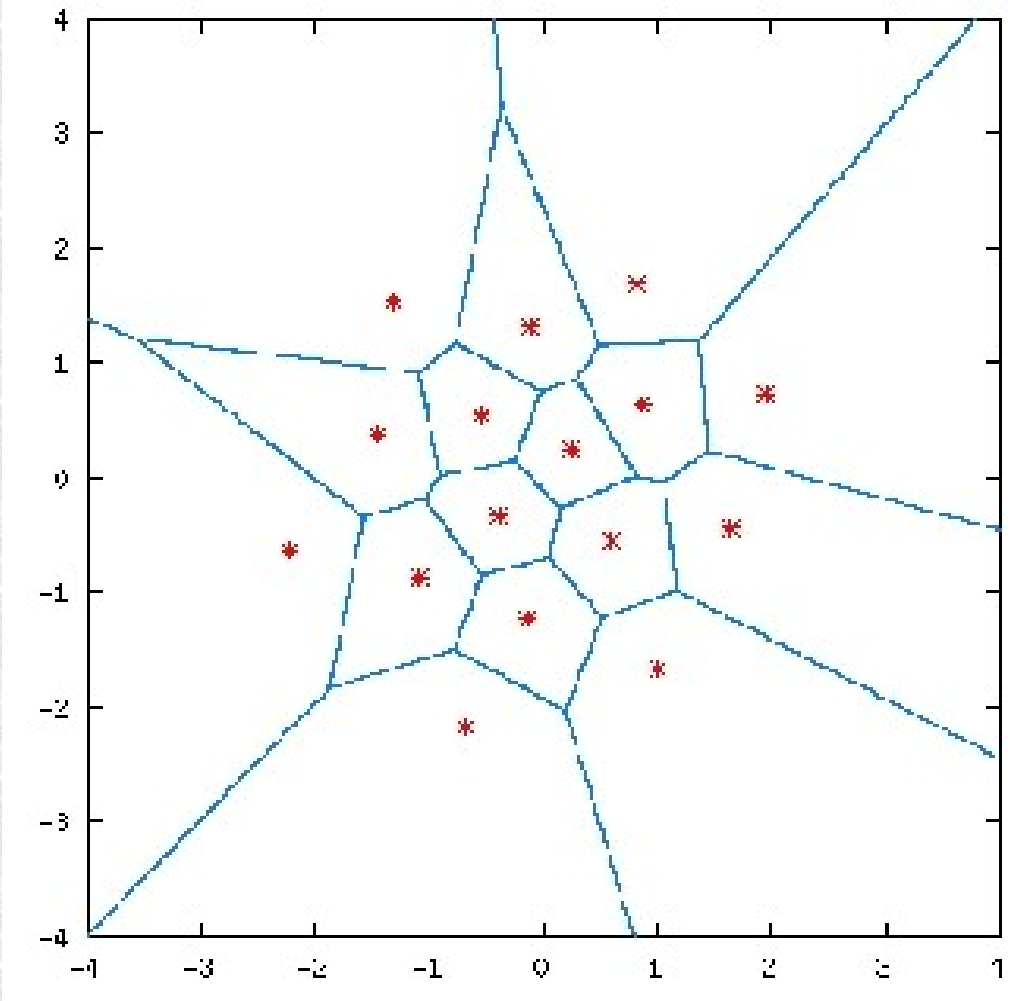
- Passage en coordonnées polaires rectangulaires
- L'angle (la phase) est fortement quantifié

Puis la quantification vectorielle:

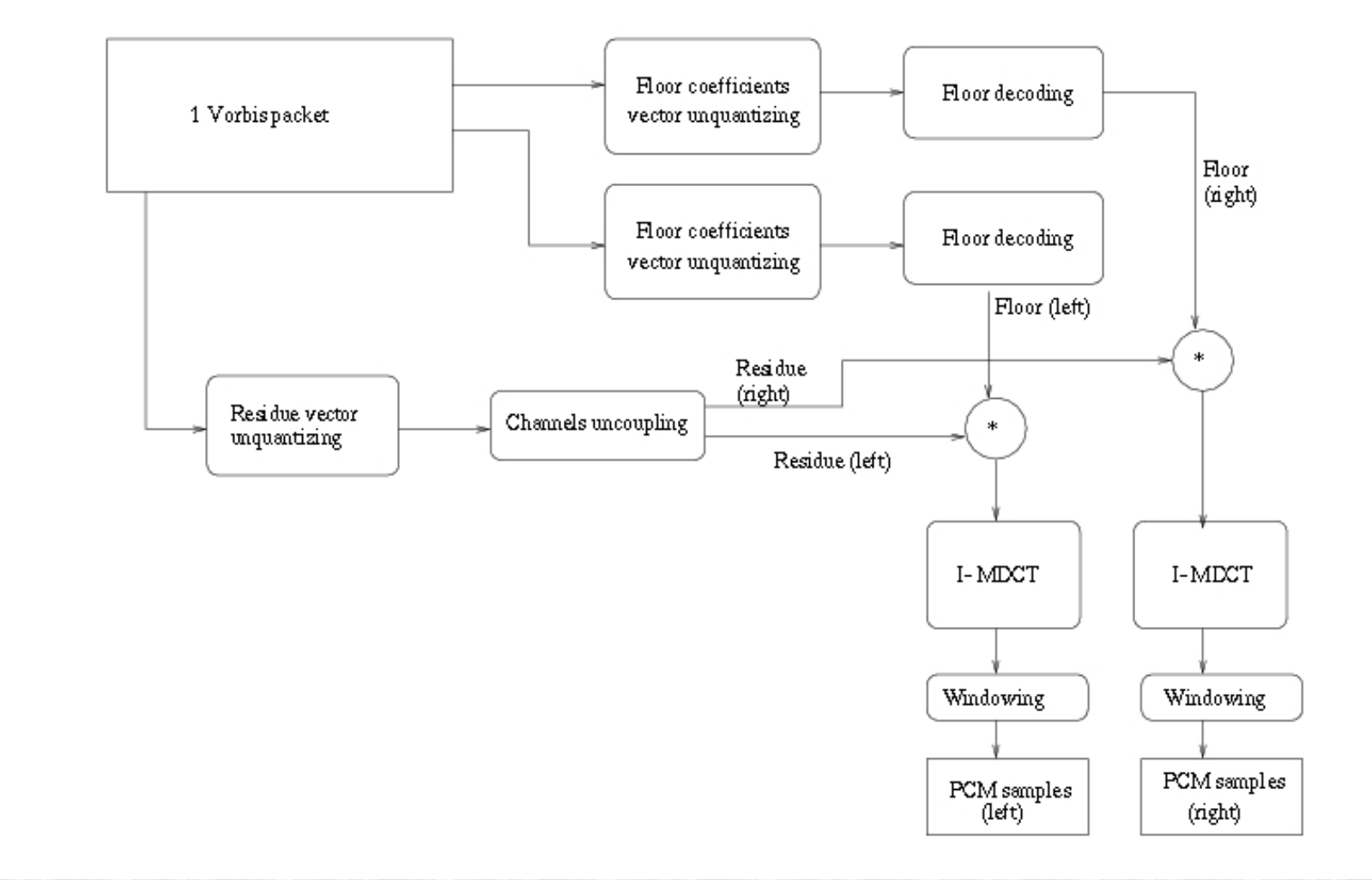
- Emploi de codebooks joints au bitstream

Les mots des codebooks sont choisis selon un arbre d'Huffman

VQ



Fonctionnement du décodeur



On emballe...

Paquet Vorbis:

- Header avec fréquence, channels, etc
- Un tag avec des infos diverses (cf id3 sur mp3)
- Les codebooks
- Floor gauche
- Floor droit
- Les résidus

Mais pas d'infos de synchro

Ogg: container format

C'est un format générique multimédia

- Audio, voix, video, ...

Ajoute ses propres headers, sorte d'enveloppe supplémentaire autour du bitstream Vorbis

Permet le streaming et la re-synchronisation

Il est possible de s'en passer: utiliser l'UDP

- Cf Icecast

THE END