

Codage Audio

Mines d'Alès
Janvier 2024

Généralités

La codage audio est basée sur le comportement de l'audition.

La psycho-acoustique décrit la façon comment notre sens de l'audition (ouïe) Perçoit les sons. Notre ouïe est limité dans le temps, la fréquence et l'espace.

Si l'objectif d'un dispositif électronique de reproduction est la qualité (HiFi) Il est très important de « masquer » les imperfections lors de la reproduction.

Par déf : la qualité d'un son doit être mesurée par la perception dudit son.

Le sens de l'ouïe est chef et dicte sa loi !

Fait remarquable :

L'acuité de l'ouïe est surprenante !

Sens de l'ouïe

La bande de perception est très large (plusieurs dizaines d'octaves)

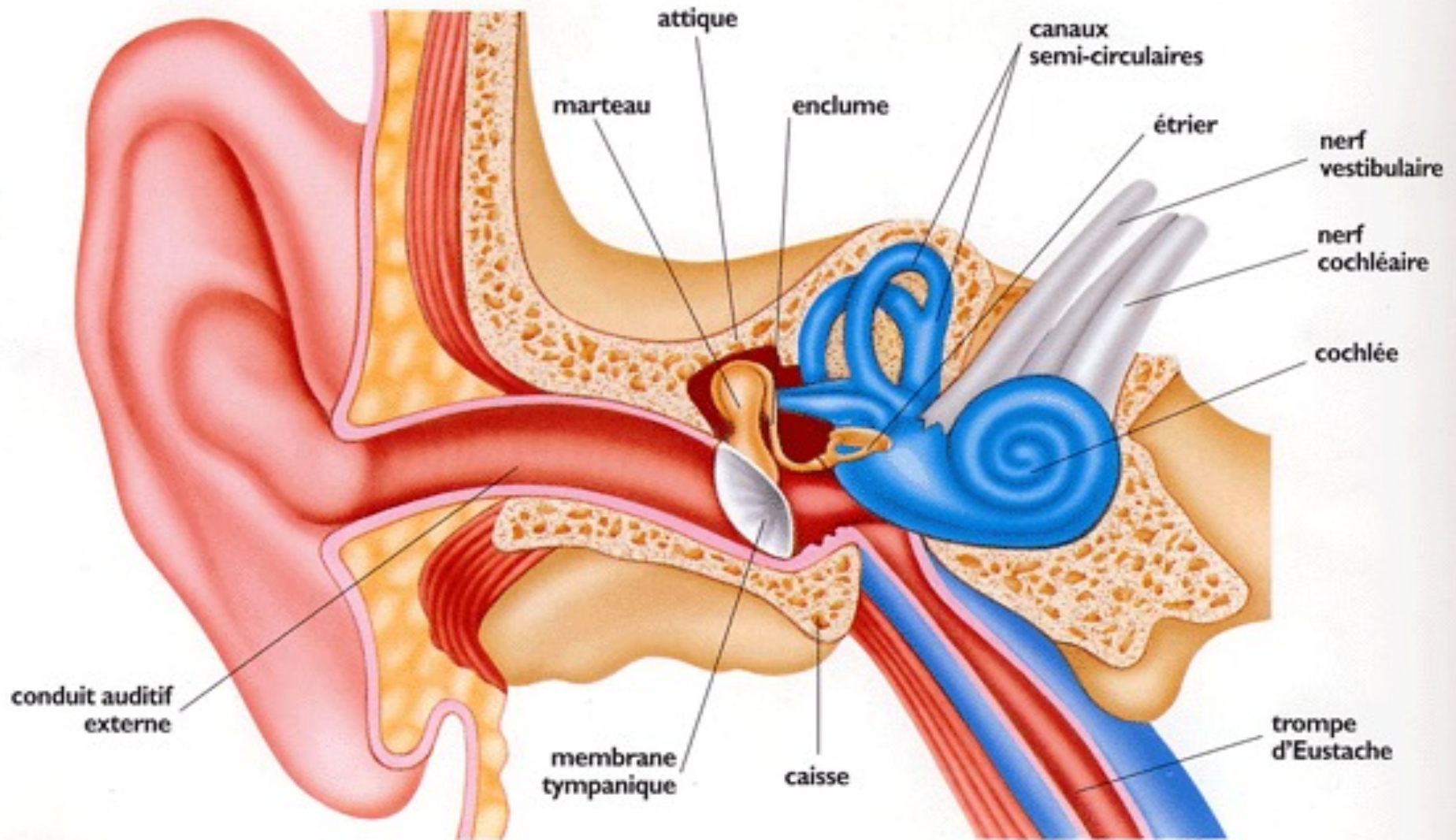
Il est capable de percevoir des quantités infimes de distorsion

Il a un rang dynamique énorme

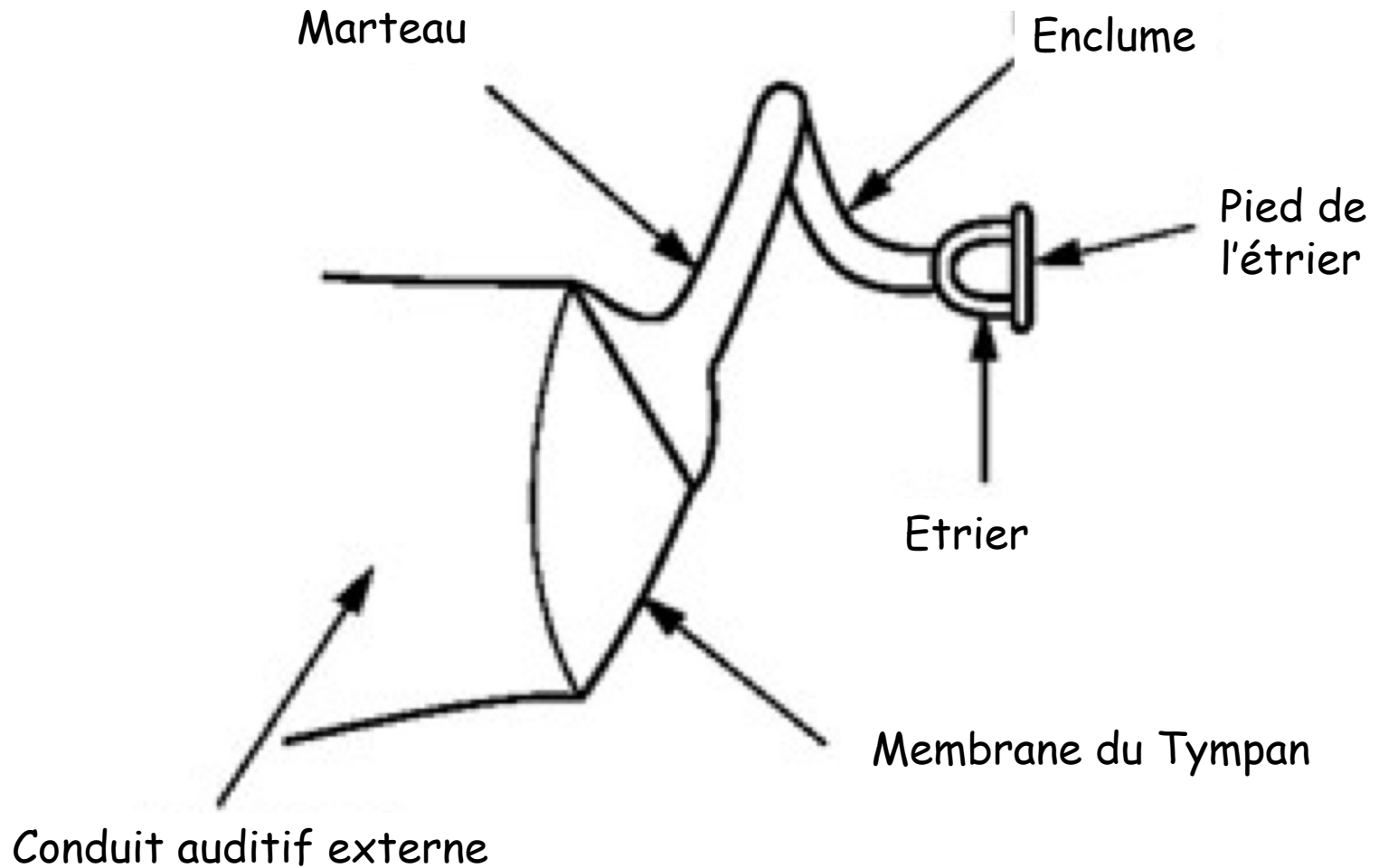
Il possède une résolution spatiale très fine.

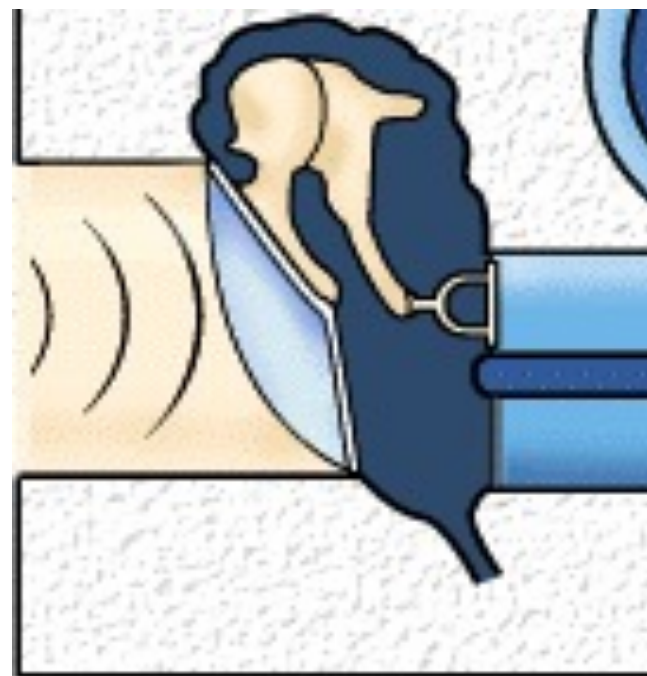
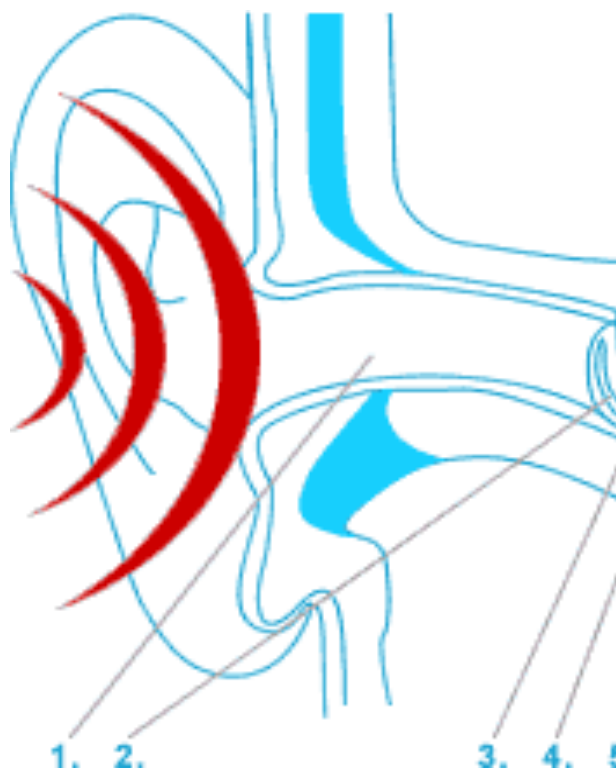
Comment fonctionne ce sens ?

Le système d'audition



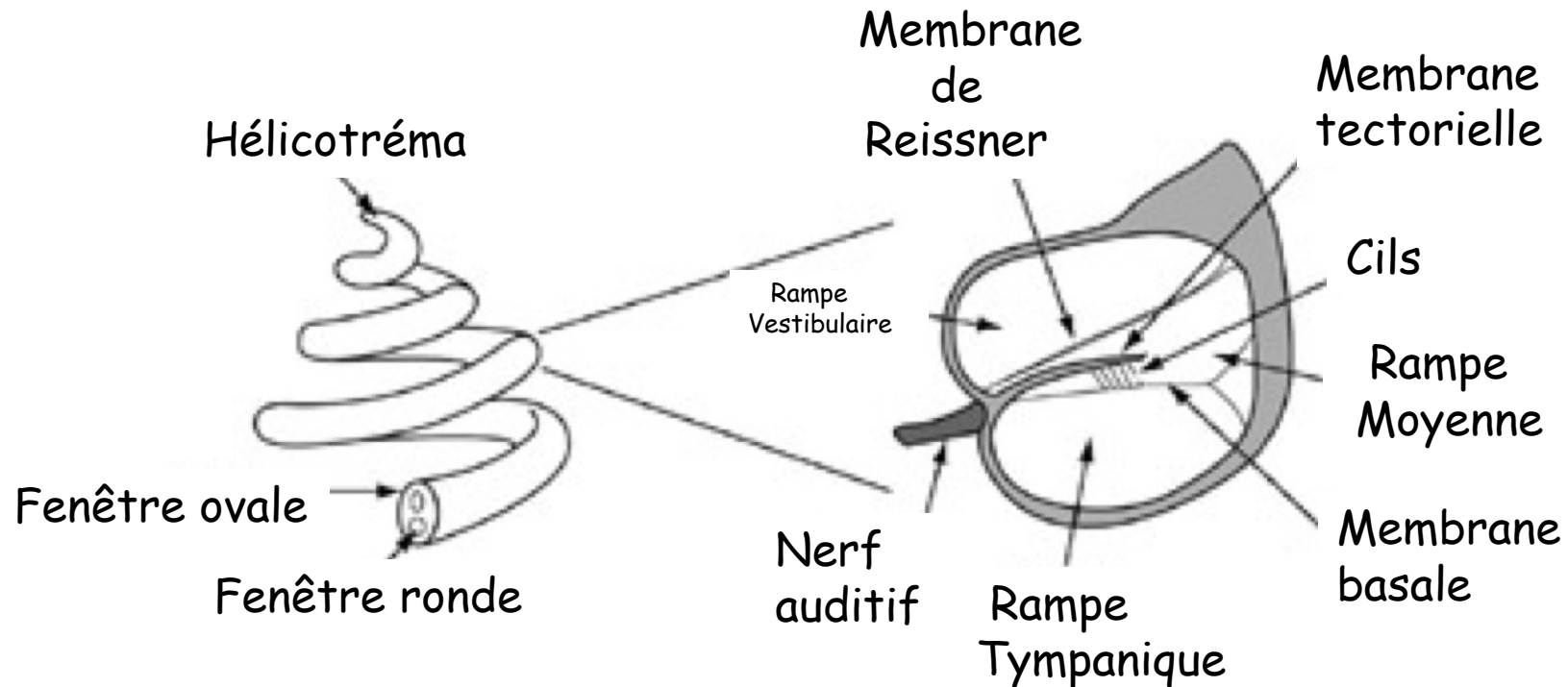
Le système d'audition ⁽²⁾





Le système d'audition ⁽³⁾

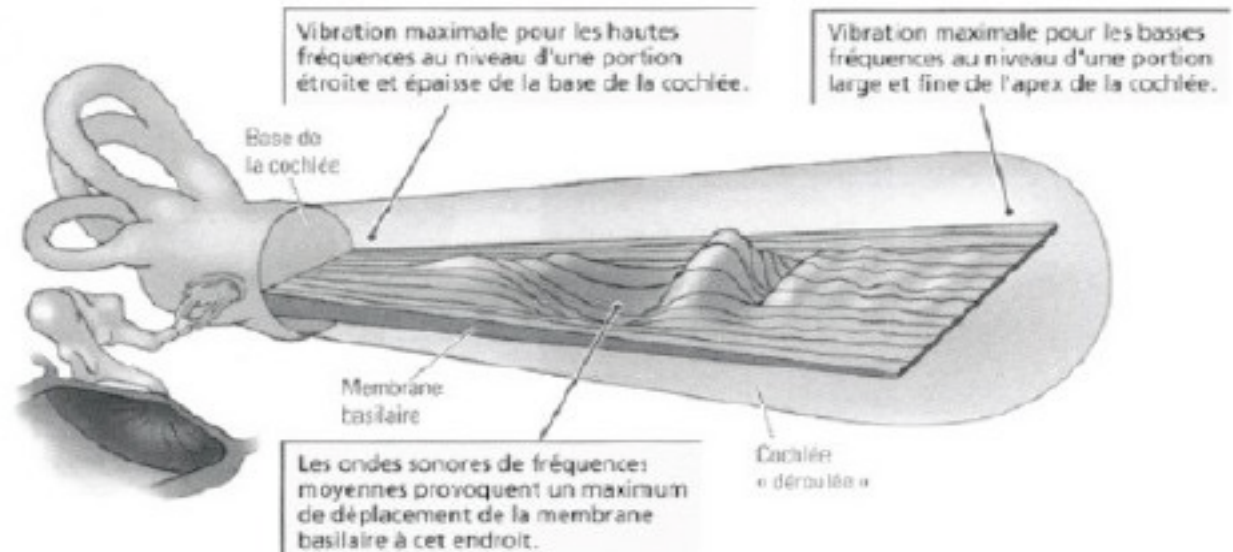
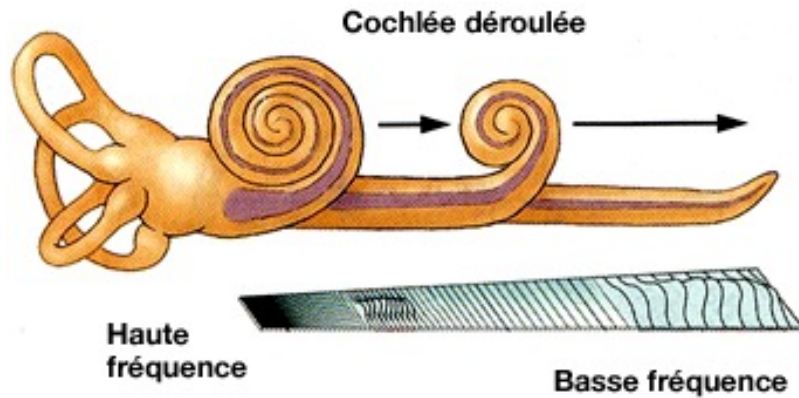
La cochlée



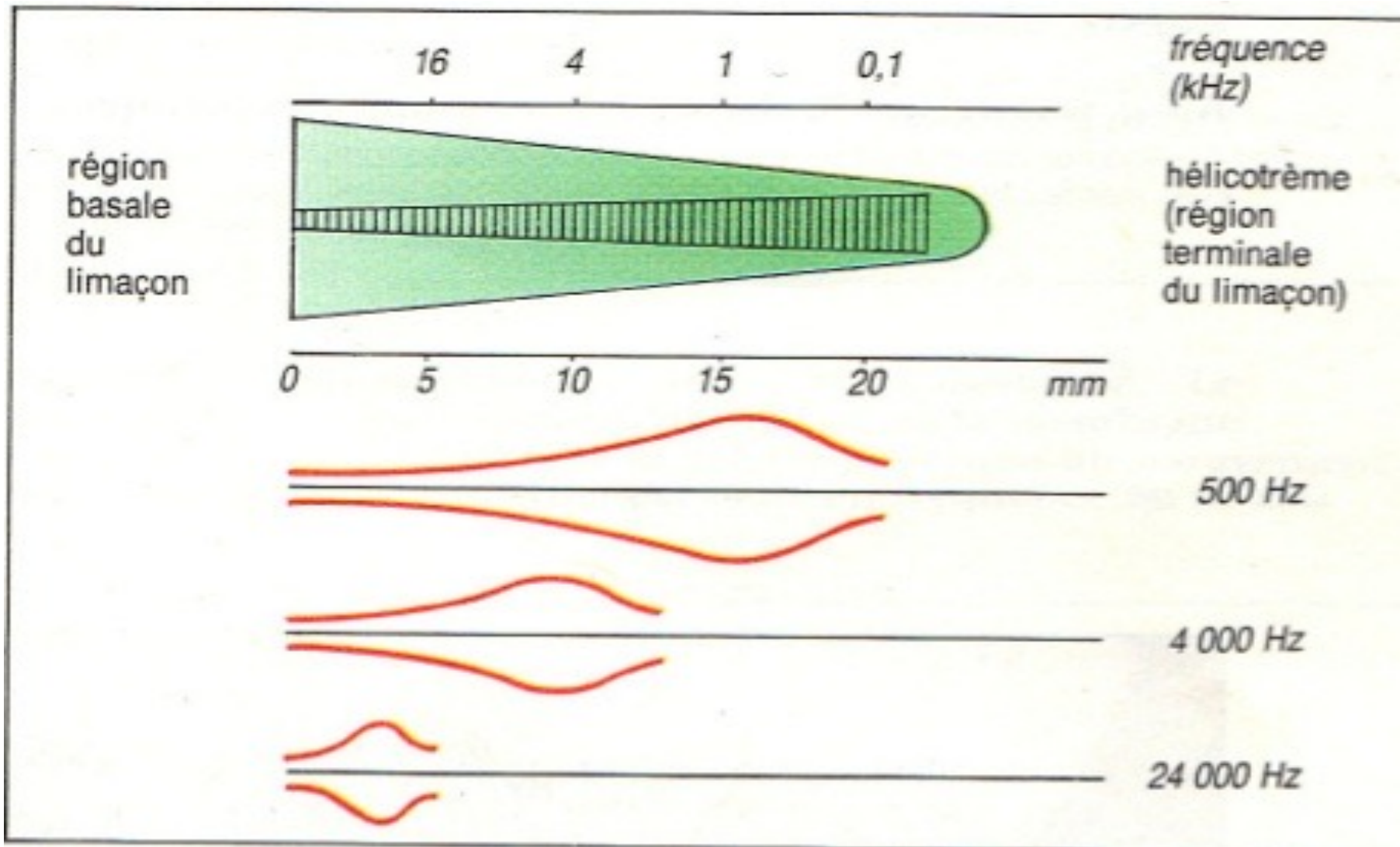
Le système d'audition ⁽⁴⁾



Le système d'audition (5)



Le système d'audition ⁽⁶⁾



Réponse de l'oreille interne

L'oreille interne peut détecter des variations de pression de l'ordre de

$2 \times 10^{-5} \text{ Pa}$
Seuil de sensibilité

La sensation d'intensité sonore est une fonction
log du niveau de pression acoustique (NPA).

Il sera mesuré en dB par rapport au seuil de sensibilité.

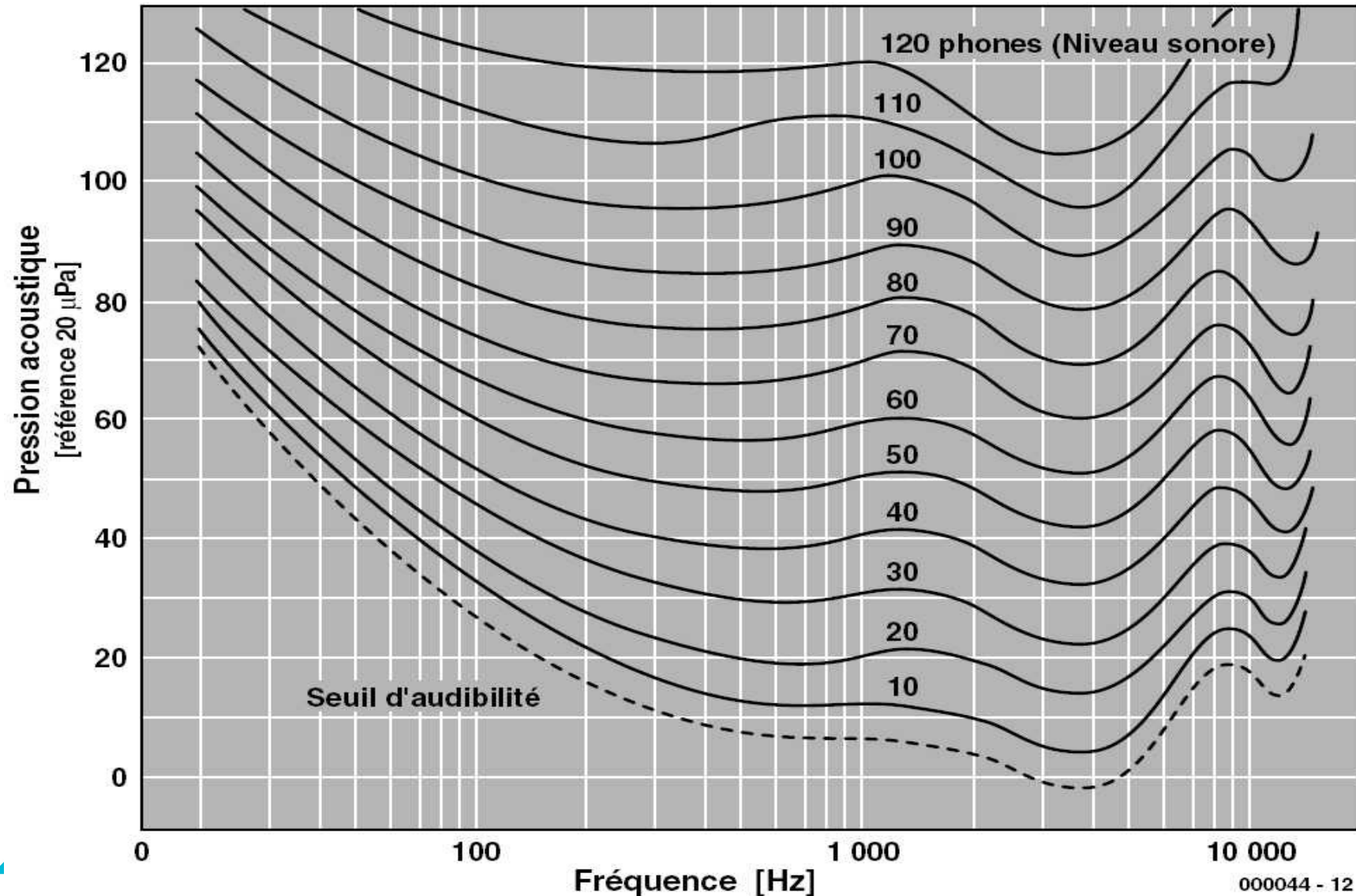
Le RD est de l'ordre de 130 dB

La réponse en fréquence est une fonction non-uniforme du NPA

La réponse subjective au niveau sonore s'appelle Intensité acoustique.
Elle est mesurée en PHONONS.

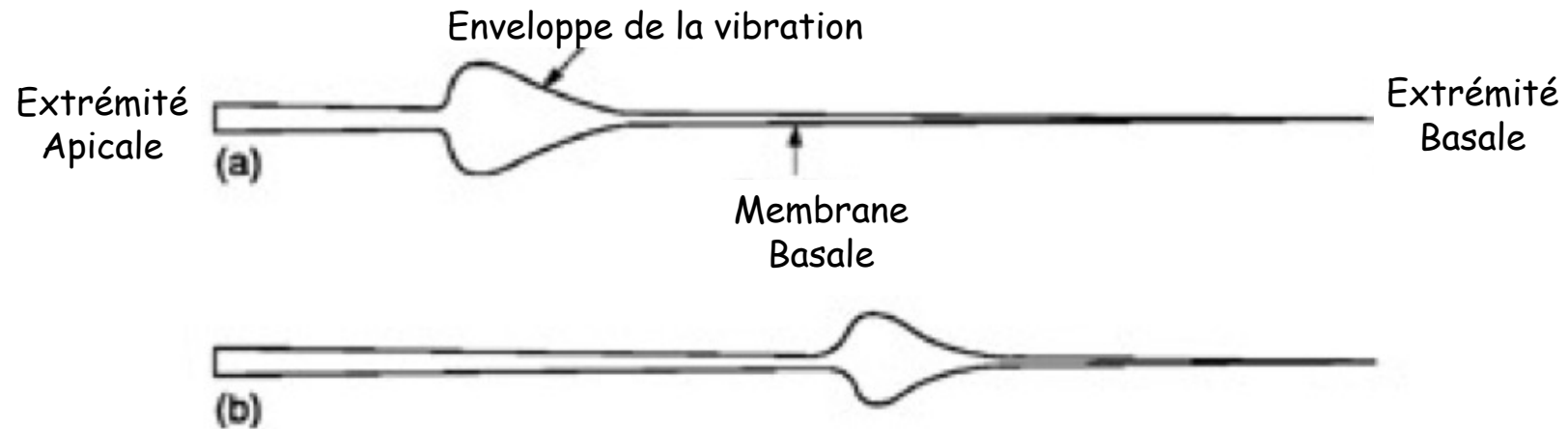
Les phonons et le NPA coïncident à 1kHz

Réponse de l'oreille interne ⁽²⁾



Discrimination de fréquence

Résonance de la membrane basale pour une seule fréquence



Remarques :

La résonance ne peut pas se produire dans un espace infinitésimal
(des fréquences voisines résonnent aussi)

L'enveloppe de résonance est asymétrique
(pertes d'énergie de l'onde à l'intérieur de la cochlée)

Si la fréquence change, le max se déplace (de façon continue)
L'espace entre cils détermine la capacité de discrimination des fréquences

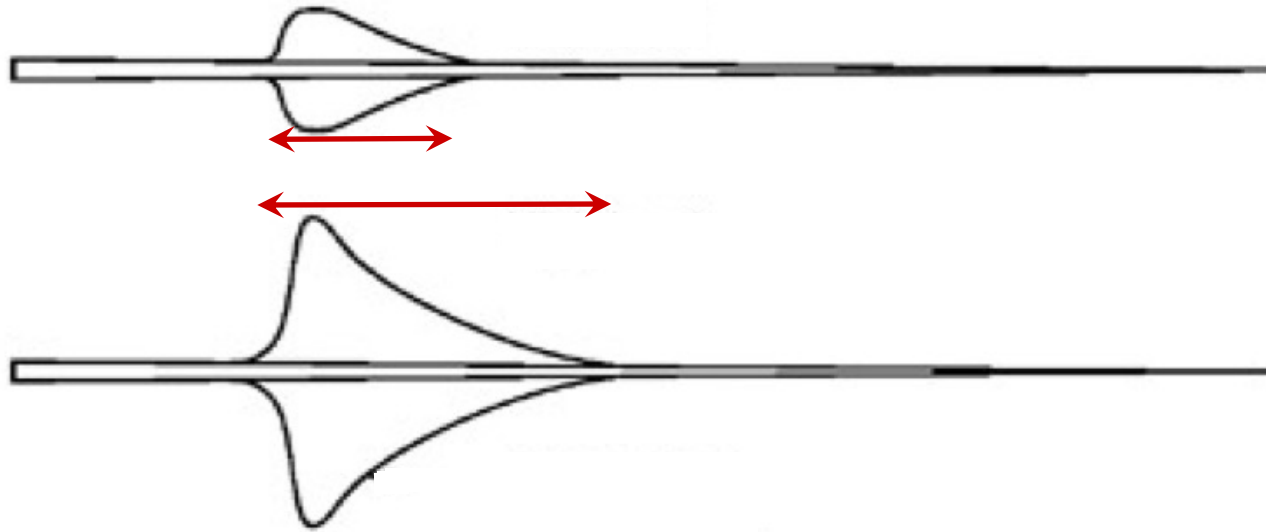
La largeur de l'enveloppe non nulle -> les cils ne résonnent pas à fréq. voisines de niveaux différents

Masquage audio

Largeur de bande critique

Le pouvoir de résolution en fréq. Dépend de la largeur de l'enveloppe

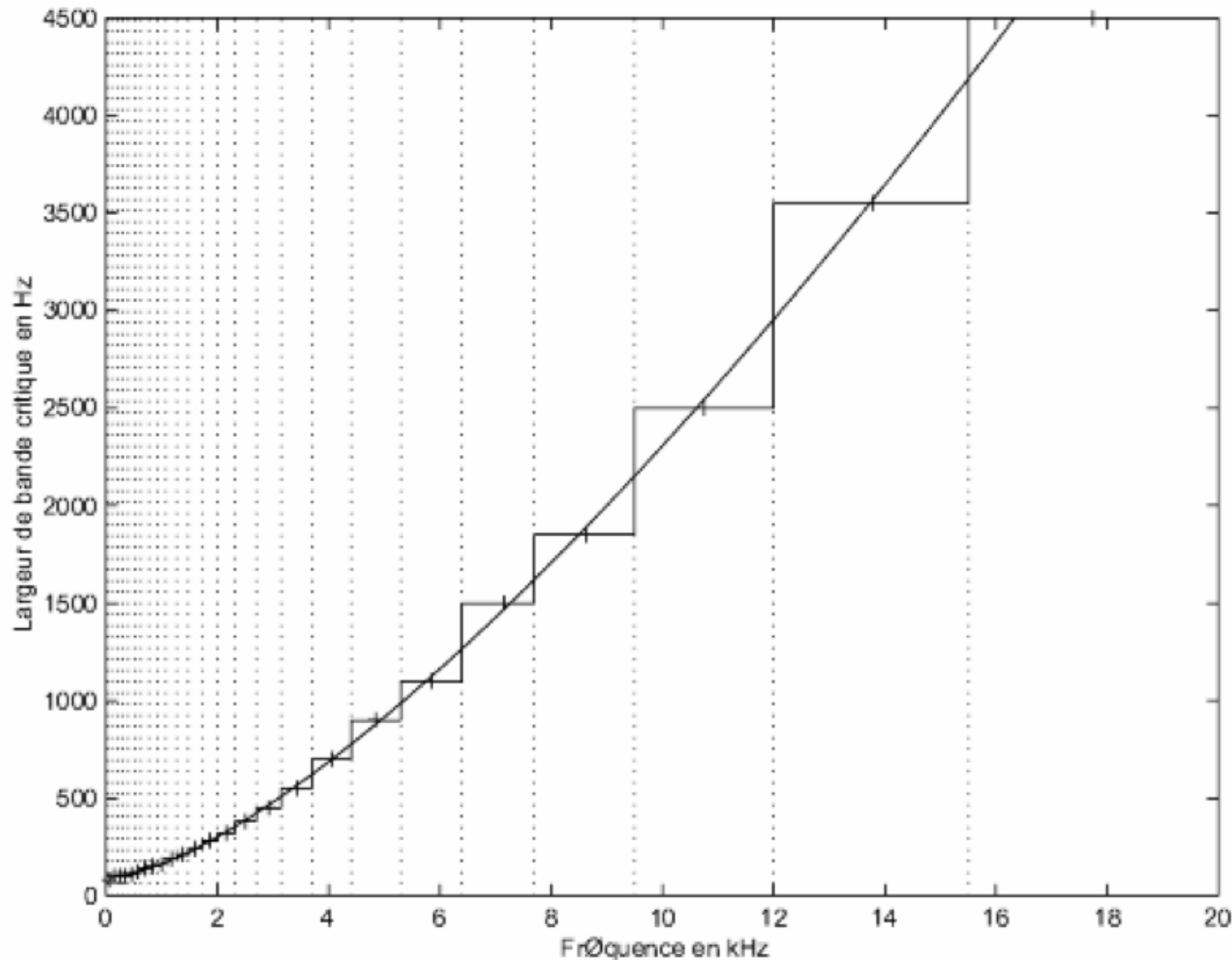
Intuitivement : elle dépend du niveau acoustique



Cette bande passante s'appelle :

LARGEUR DE BANDE CRITIQUE

Largeur de bande critique



Largeur de bande critique ⁽²⁾

La largeur de bande critique détermine le pouvoir de résolution dans le temps

Grande LBC → sélectivité temporelle petite (3ms)
Petite LBC → sélectivité temporelle grande (300 ms)

Ceci explique le masquage pré et post masqueur !

Il explique aussi la présence de battements
lorsque deux fréquences sont proches

Codage perceptif

L'effet de MASQUAGE induit le **codage perceptif**

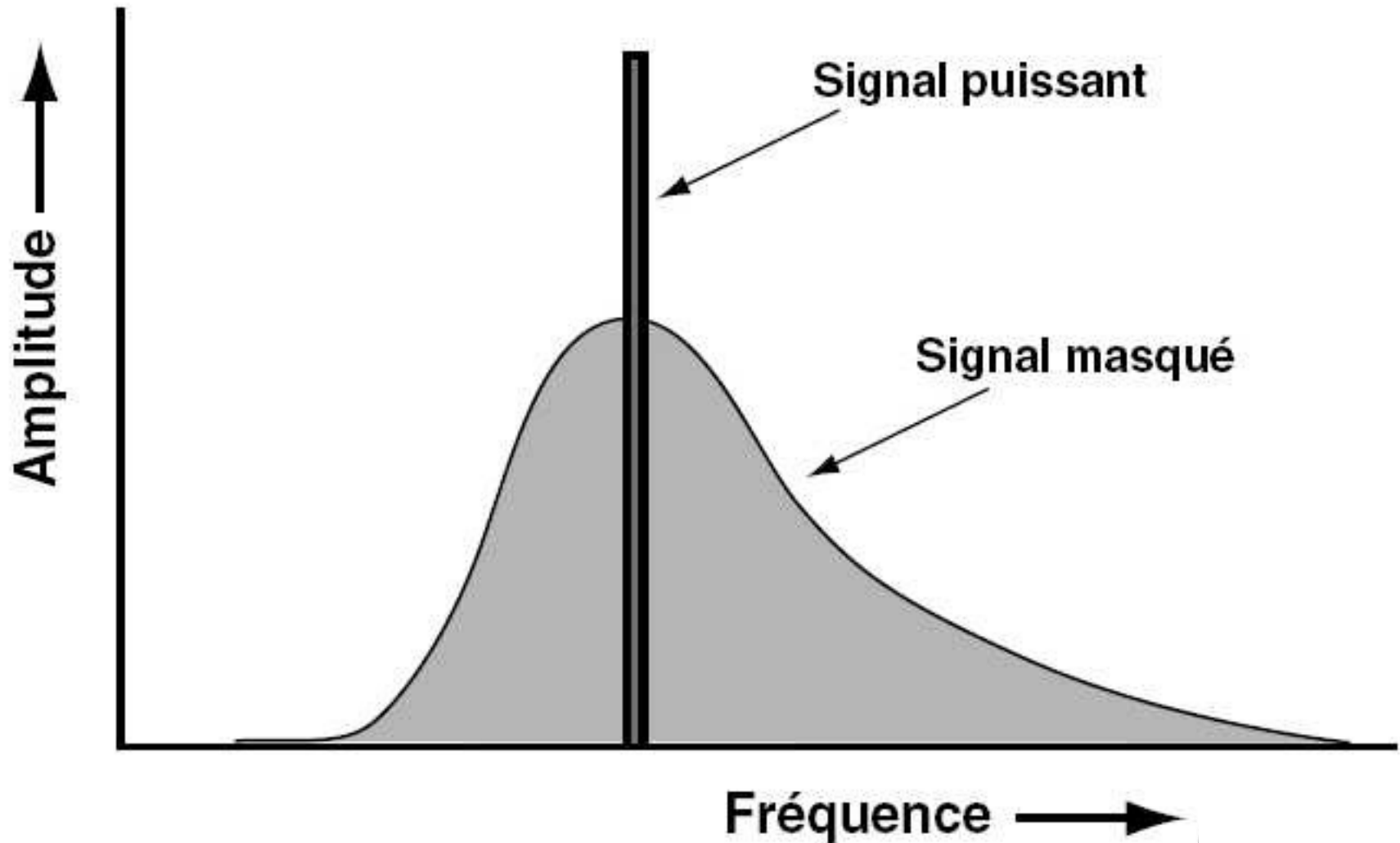
Comment ça fonctionne ?

Si une tonalité « **forte** » se présente à l'entrée,
elle « **masque** » **les fréquence voisines** d'amplitude inférieure

La quantification de cette fréquence (et les bandes voisines)
peut se faire de façon « grossière »,
c'est à dire avec **moins** de bits de résolution.

Il résulte des économies importantes en nombre
de bits de représentation → débit inférieur

Masquage audio



Codage en sous-bandes

Le codage en sous-bandes exploite le fait que le spectre d'énergie d'un signal acoustique n'est pas **uniforme** !

La résolution d'un codeur de type PCM se base sur le rang dynamique.
Ce nombre est indépendant de la fréquence.

EN général, les composantes de fréquences le plus fortes déterminent le RD.

Les autres sont codées avec des bits en trop !

Le codage en sous-bandes peut être appliqué en numérique ou en analogique.
Le système Dolby l'utilise intensivement

Reste en suspens la question : **combien de sous-bandes ?**

SB et masquage

On peut combiner la décomposition en SB avec le masquage audio

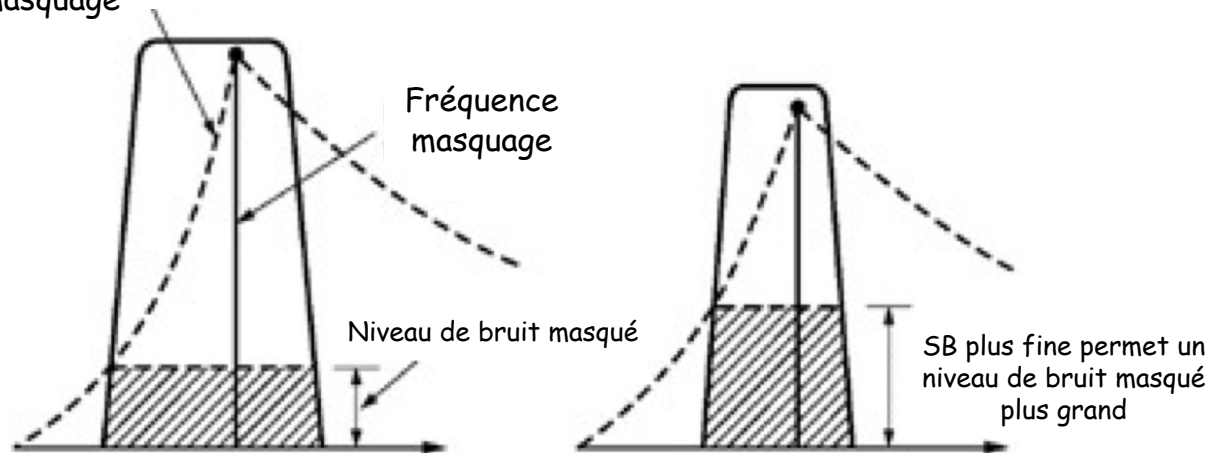
Il suffit de découper en SB de largeur \leq LBC !!

Problème : la LBC (f, NPA) !

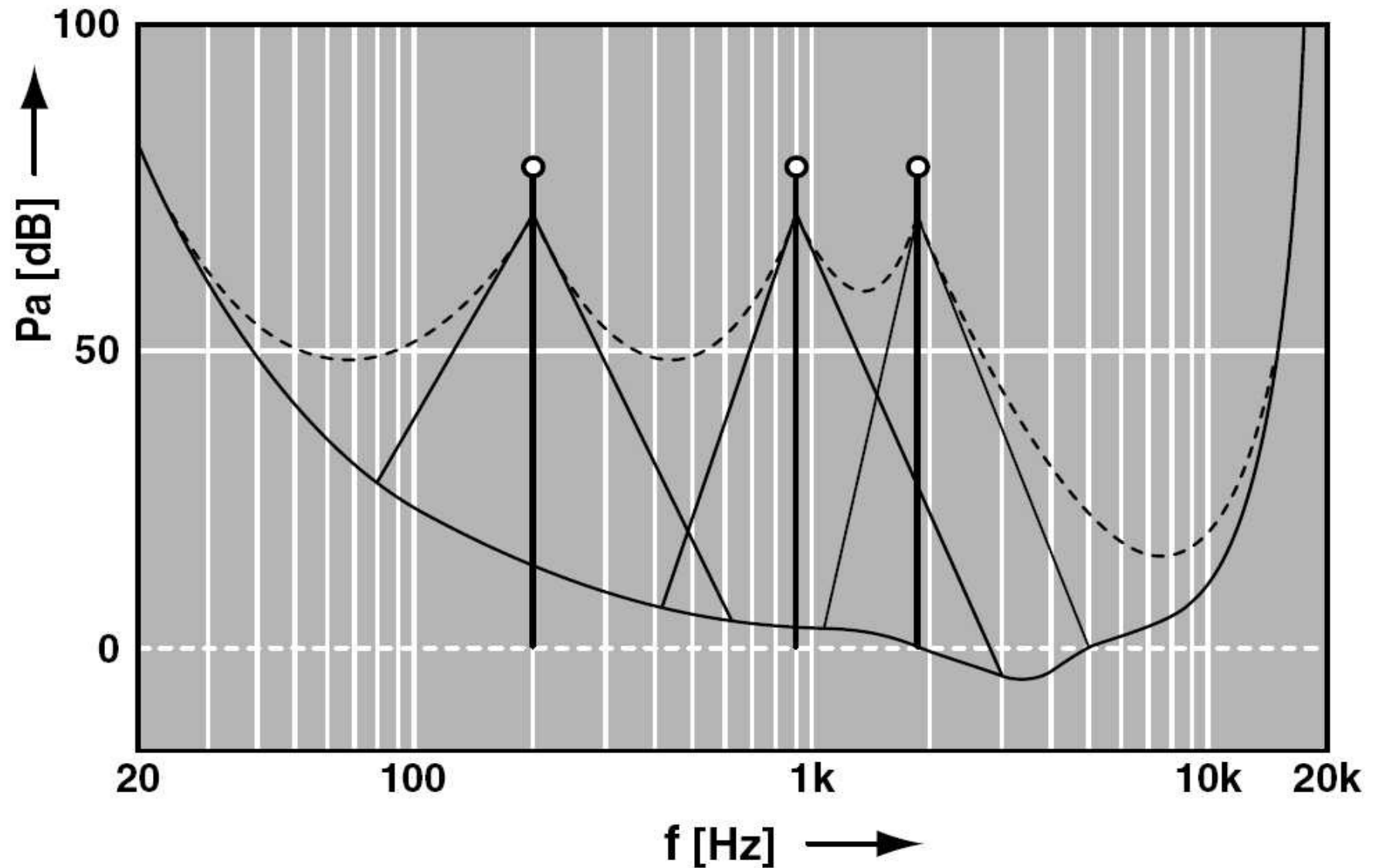
Cette contrainte n'est pas toujours appliquée

MPEG Layer I et Layer II utilise..... 32 sous-bandes !!

Profil de masquage



SB et masquage



Codage audio MPEG

Un peu d'histoire :

Le groupe de travail MPEG/Audio est né avec le Digital Audio Broadcasting (DAB) lancé par l'agence aérospatiale allemande Deutsche Luft und Raumfahrt

Financé par l'UE, il faisait partie du programme EUREKA (EU-147)(1987-1994)

Deux techniques de codage audio étaient disponibles :

MUSICAM (Mask pattern adapted Universal Subband Coding and Mux)(CCETT, IRT, Philips)
ASPEC (Adaptive Spectral Perceptual Entropy Coding)(ATT, Thomson, Inst Fraunhofer, CNET)

Le groupe de travail MPEG/Audio choisit **MUSICAM**
Structuration modulaire en plusieurs couches de codage [Layers],
sa simplicité de mise en œuvre côté décodeur
et sa grande tolérance aux erreurs de transmission.

MPEG Layer I/II - 192 kb/s (1992)

Codage audio MPEG ⁽²⁾

Une année plus tard :

Le groupe de travail MPEG/Audio fusionne les approches MUSICAM et ASPEC pour créer la couche III

MPEG Layer III - 128 kb/s (1993)

MPEG/Audio a trois niveaux de complexité (3 *layers*)

Un codeur audio peut fonctionner avec des taux de compression différents :

La musique classique stéréo nécessite plus de qualité que la parole monophonique

Codage audio MPEG ⁽⁴⁾

- ✓ **MPEG-I** version simplifiée de MUSICAM utilisée pour des applis à taux de compression moyen et faible coût
- ✓ **MPEG-II** = MUSICAM utilisé en DAB et pour l'audio du DVB.
- ✓ **MPEG-III** = le meilleur de MUSICAM et de ASPEC utilisé pour des taux de compression important pour des applis télécomm et enregistrement.

Caractéristiques MPEG

A chaque layer, MPEG/Audio supporte :

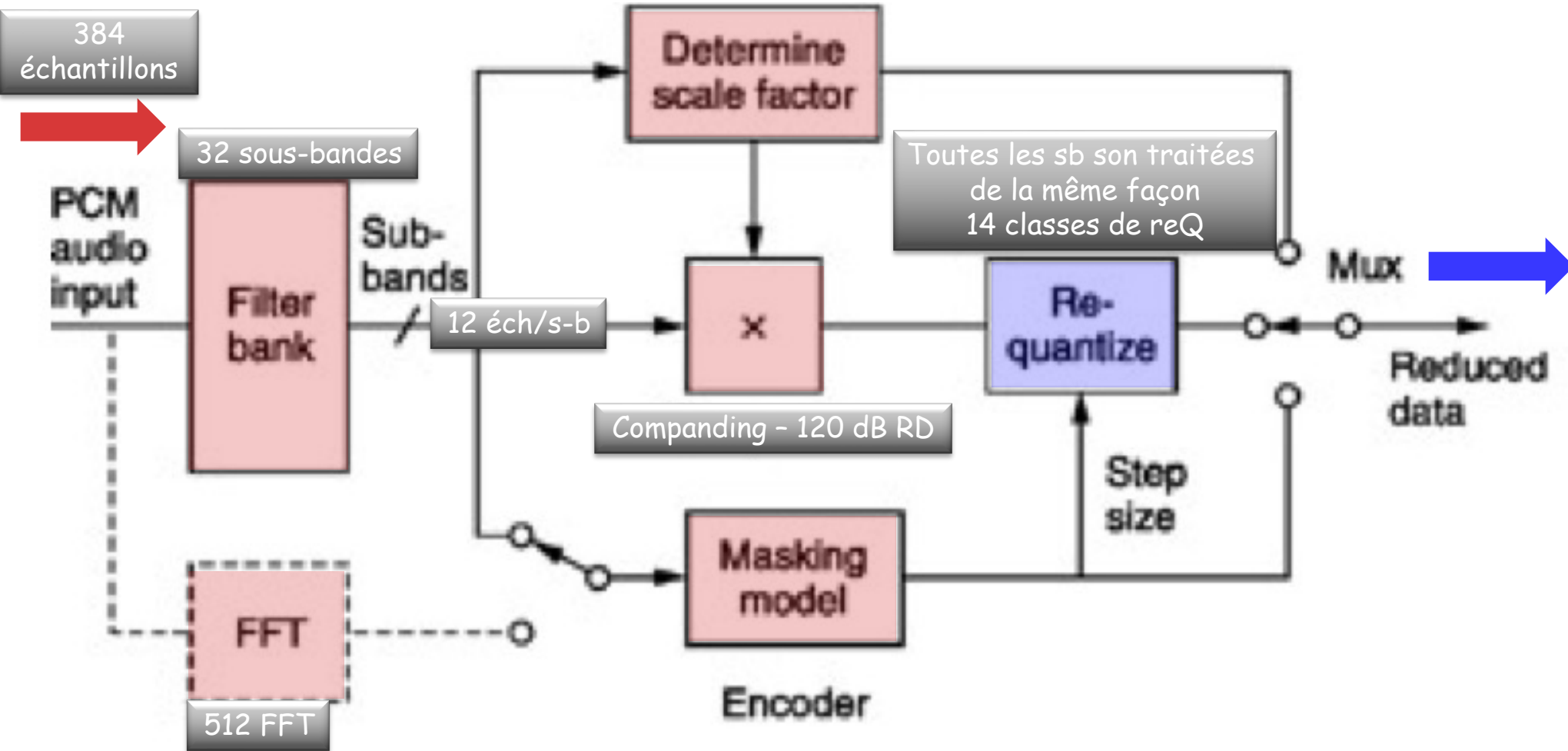
- * Entrée des fréq d'échantillonnage 32, 44.1 et 48 Kéch/s

- * Sortie

Taux de codage 32, 48, 56, 64 96, 112, 192, 256 et 384 kb/s

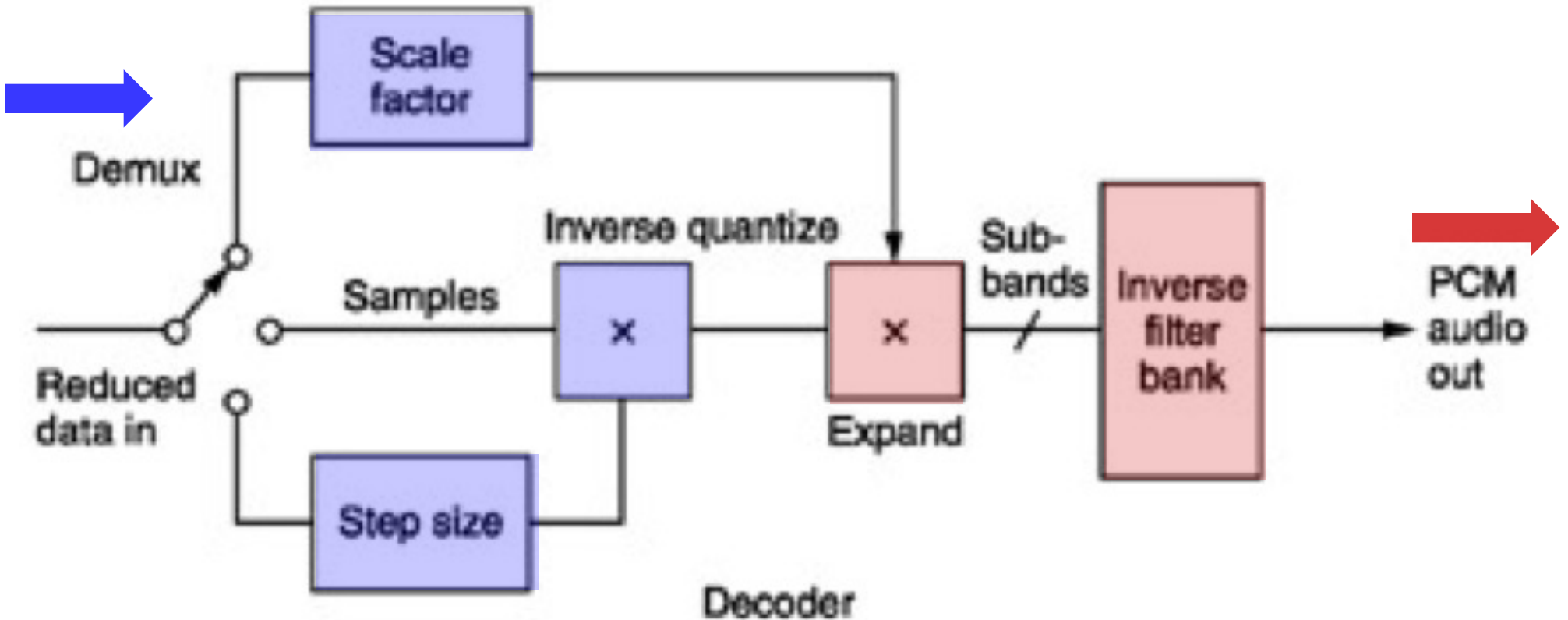
- * Mono, dual ou stéréo, stéréo sélectif en fréquence.

MPEG Layer I - Codeur

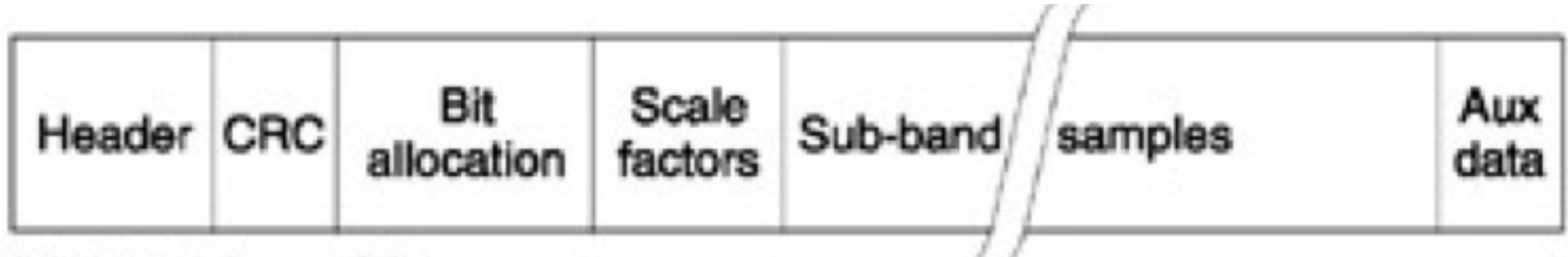


384 éch@48kéch/s → 8ms trame

MPEG Layer I - Décodeur



MPEG Layer I - Format d'une trame



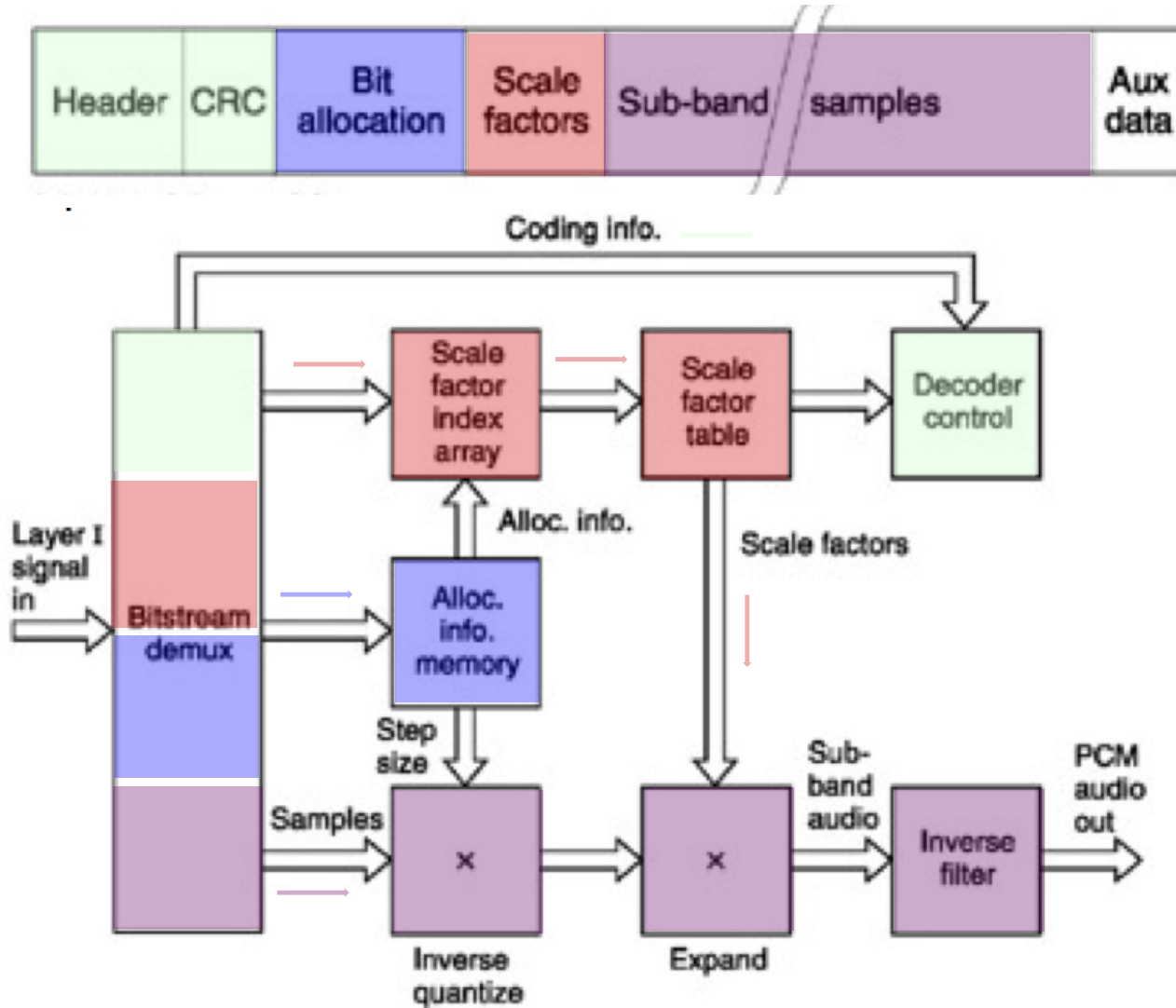
Header : synchro trame + descripteur d'éch

32x4 bit = descripteur de l'alloc de bits → longueur des mots de chaque sous-bande ;

32x6 bit = descripteur des facteurs d'échelle_(par pas de 2dB) de chaque sous-bande ;

32 ensembles de 12 éch codés par des mots de longueur variable _(selon l'alloc de bits) de 0 à 15 bits de longueur.

MPEG Layer I - Lecture d'une trame



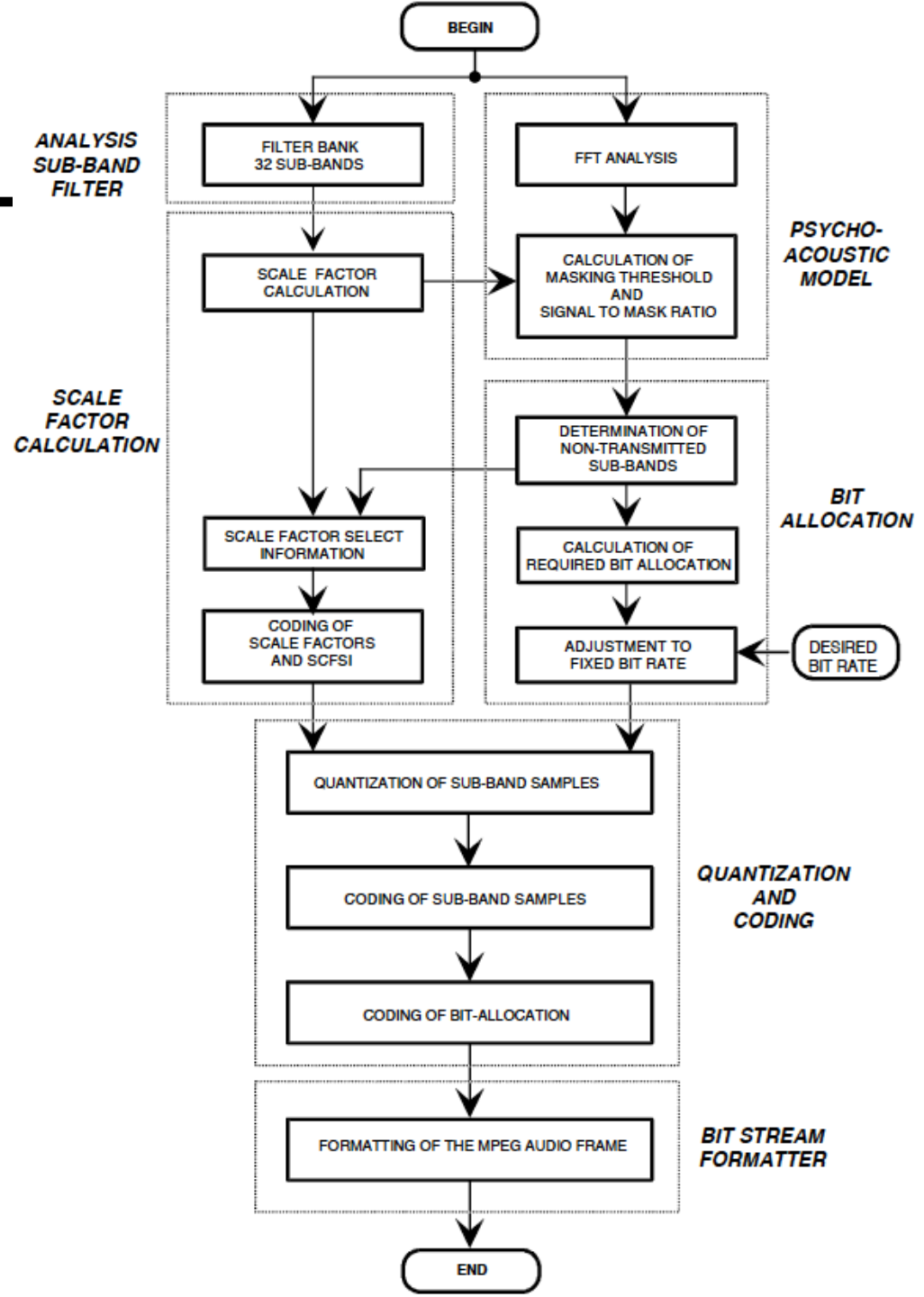
MPEG Layer II

Résolution spectrale améliorée :
FFT 1024 points !

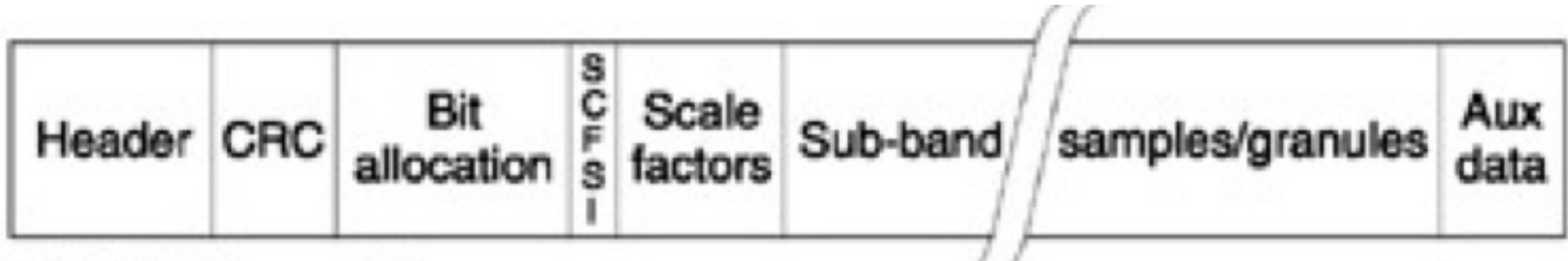
Block en entrée plus long :
1152 éch (@48kéch/s→24ms)

Q est plus fine, 3 interv de fréq :
low(15 wl-4b), **mid**(7wl-3b) et **high**(3wl-2b).
(on réserve 1 bit pour indiquer 0 amplitude dans la sous-bande)

3 sous-blocs de 384 éch sont traités ; les
facteurs d'échelle sont codés par
corrélation sur les 3 blocs (2b SCFSI-Scale
Factor Select Information)

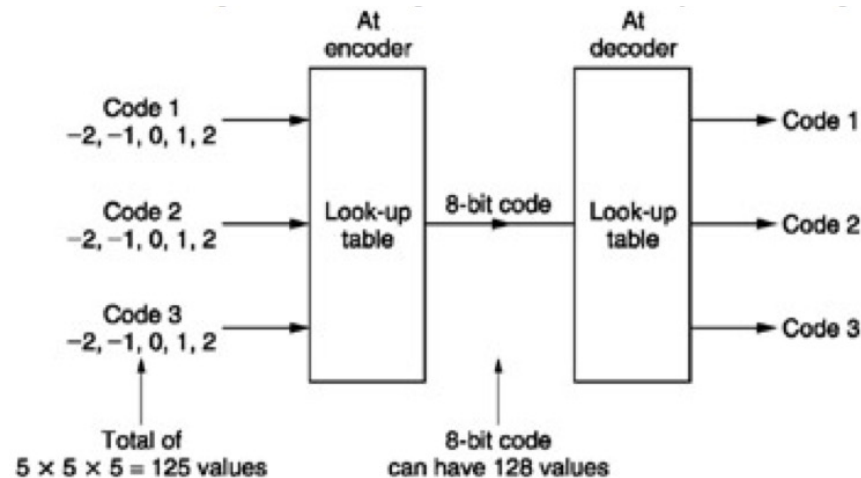


MPEG Layer II - Trame

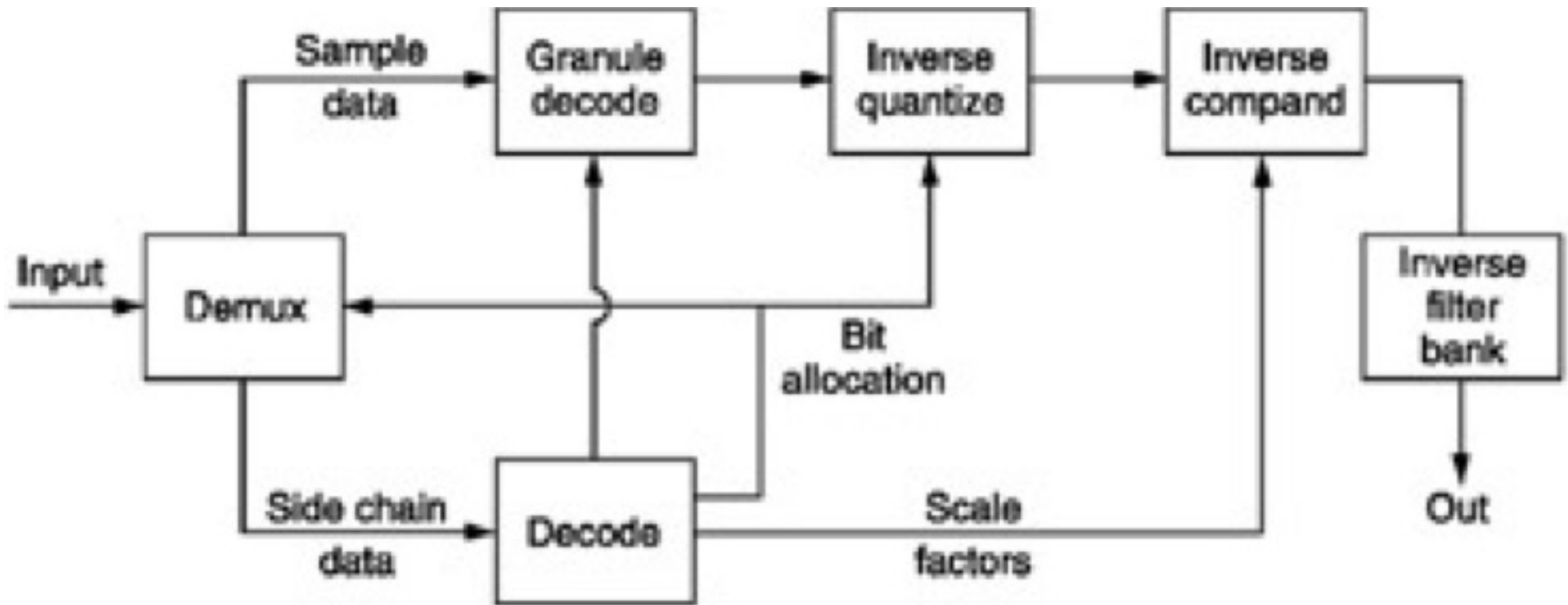


2b pour le SCFSI

Descripteur des « granules » d'amplitudes.



MPEG Layer II - Décodeur



MPEG Layer III - MP3

La couche III est la plus complexe → elle est utilisée dans les cas où on souhaite un fort taux de compression.

Elle est connue sous le nom de **MP3** dans les applications de transfert de musique sur INTERNET.

Il s'agit d'un codeur par transformée basé sur le modèle ASPEC, avec des modifications pour le rendre adaptable aux couches I et II.

Il assure une bonne qualité perceptuelle avec un taux de compression important basée sur le codage entropique (Huffman)

MPEG Layer III - Caractéristiques

32 bandes de fréquences (id L I et II)

Blocs de 1152 échantillons

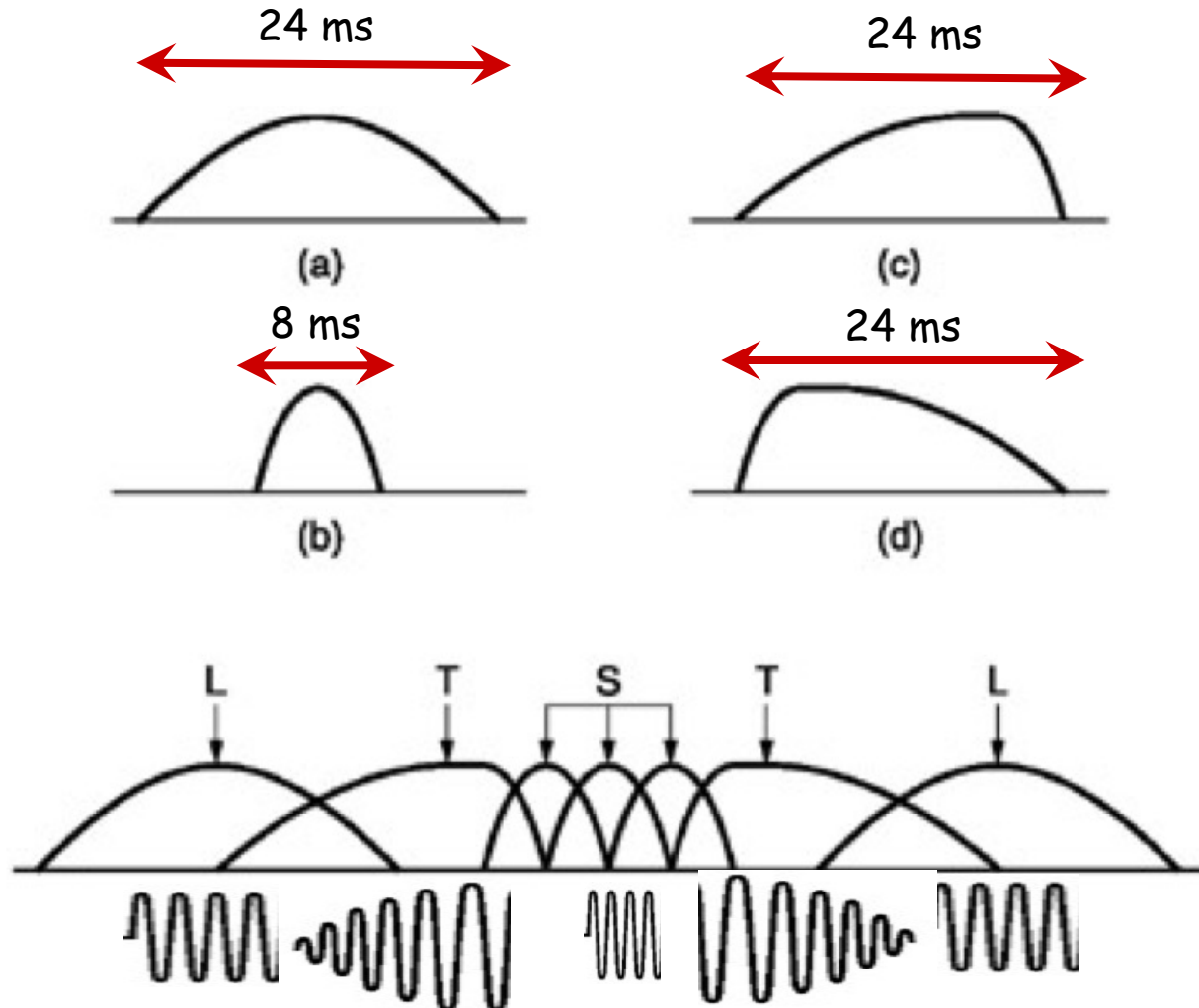
Les 32 sorties du filtre polyphase → MDCT (domaine de la féq)

Le passage dans le domaine de la fréquence impose des contraintes fortes sur le fenêtrage temporel.

Deux fenêtres différentes sont utilisées pour améliorer la résolution fréquentielle :

- * Une fenêtre longue de 24 ms → rés. 18 fréqs
- * Une fenêtre courte de 8 ms → rés 12 fréqs

MPEG Layer III - Fenêtres



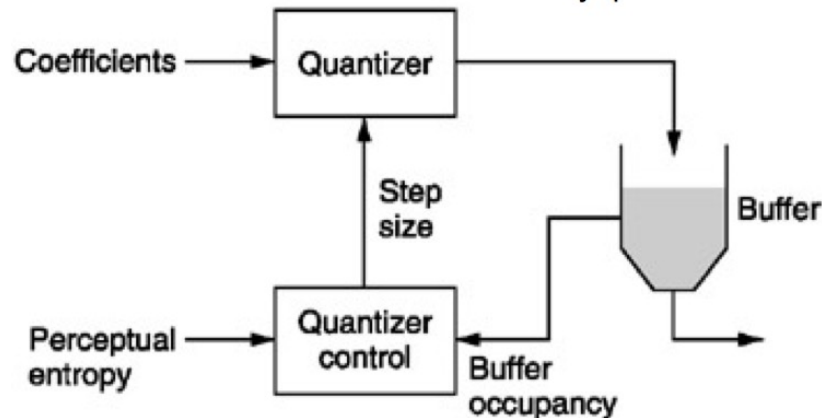
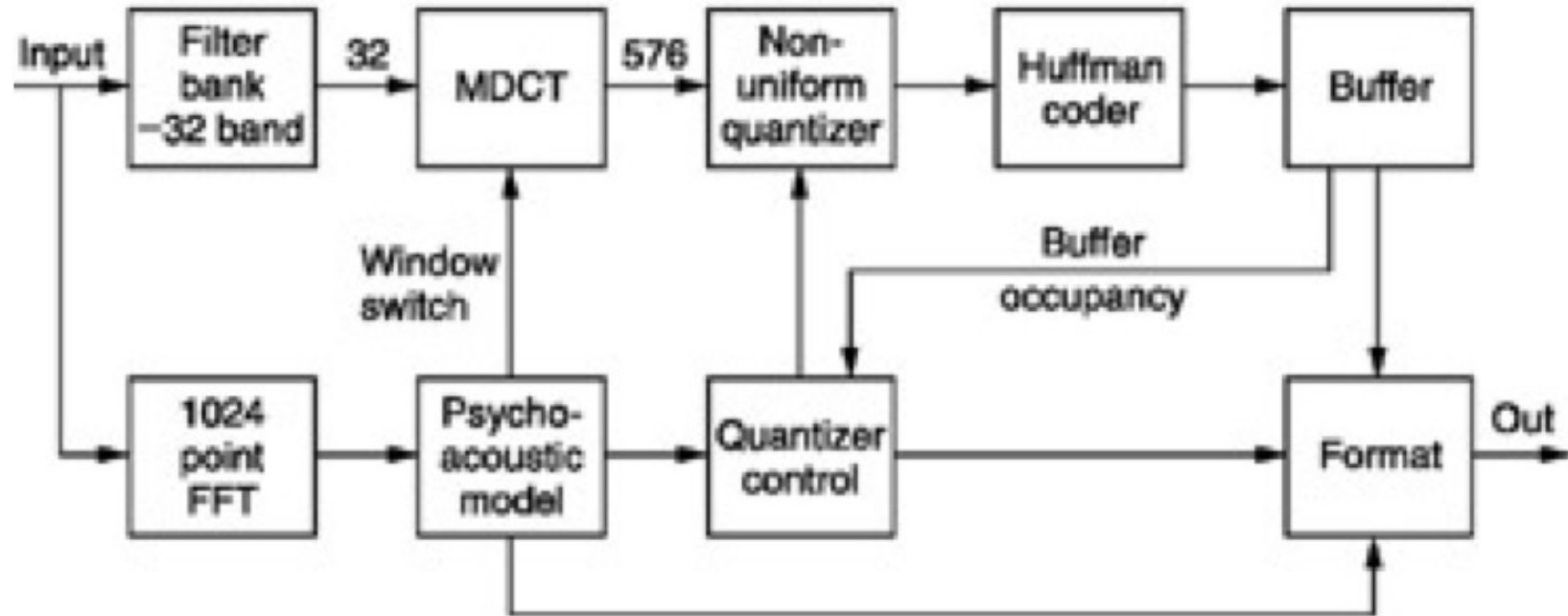
MPEG Layer III - Codeur

L'usage d'un modèle psycho acoustique est fort conseillé

La fine résolution fréquentielle permet de trouver des seuil de masquage mieux adaptés. Une FFT en // avec la DCT permet de résoudre l'ambiguïté de phase.

On utilise un Q non-uniforme. Les coefs quantifiés sont codés par un CLV de Huffman. La longueur variable est gérée par un buffer qui assure un débit constant à la sortie ! Cooooool !!

MPEG Layer III - Codeur ⁽²⁾



The End

A suivre : codage vidéo

