

Compression de données

ReDS
heig-vd

1. Pourquoi compresser?
2. Principe de la compression de données audio
3. Compression sans pertes
4. FLAC
5. Compression avec pertes
6. Mpeg (MP3)

Pourquoi compresser?

Avant: les disques dur atteignaient difficilement 500 ou 600 Mo tandis que les connections internet, aussi rares qu'onéreuses, étaient loin de permettre de manipuler de telles données.

=> La compression audio vise à réduire la quantité de données pour stocker la musique

Souvenez-vous!

Télécharger un fichier de plusieurs méga prenait plusieurs minutes voire des heures.

Pourquoi compresser?

Maintenant: La capacité de stockage ne pose plus forcément de problème, même pour des fichiers audio non-compressé (les supports devenant de plus en plus performants à ce niveau).

Pour l'audio, il n'y a plus véritablement de problème de bande passante, car l'écoute peut se faire en streaming.



les réseaux (Lan, Internet ...) sont de plus en plus utilisés pour échanger des informations. Cela pose cependant un problème important : celui du temps de transmission. Le débit des réseaux étant limité, on cherche à réduire la taille des données à envoyer afin de rendre la transmission plus rapide et par conséquent plus satisfaisante

Pourquoi compresser?

La deuxième contrainte va être l'économie d'énergie!

- + vous stockez de données,
- + vous avez besoin de place
- + vous consommez d'énergie

=> La compression fera économiser de l'énergie (comparaison avec le stockage du génome)

Compression audio

1. Pourquoi compresser l'audio?
2. **Principe de la compression de données audio**
3. Compression sans pertes
4. Compression avec pertes
5. Mpeg (MP3)

- Est un phénomène vibratoire
- Se propage dans un milieu particulier (air, eau, matière...)
- 3 phases pour exister: émission / transmission / réception

Ses caractéristiques sont:

- L'amplitude, ou l'intensité du son exprimé en décibel (db)
- La fréquence, répétition d'une période définissant ainsi la hauteur (grave ou aiguë)
- La nature de son signal: analogique (variation électrique) ou numérique (codage binaire)

Un son numérique est traduit par :

- Sa fréquence d'échantillonnage, le nombre de « relevés » effectués chaque seconde.
- le nombre de bits des échantillons, les valeurs numériques disponibles pour traduire l'amplitude du signal (8bit = 256 valeurs, 16bit = 65535 valeurs).
- le nombre de voies utilisées (Mono = une voie, Stéréo = 2 voies, Quadriphonie = 4 voies...)

Mémoire requise pour stocker un son non compressé en octet:

- Fréquence d'échantillonnage(Hz) x Nombre de bits / 8 (octet) x Durée (sec) x Nombre de voies

R. Mosqueron (HES-SO / HEIG-VD / REDS), 2017

- Son qualité téléphone: 8000 Hz 8bit
- Son qualité radio FM: 22050 Hz 16bit
- Son qualité CD: 44100 Hz 16bit
- Son qualité DVD: 48000 Hz 24bit
- Son audio professionnel: 96000 et 192000 Hz 24 et 32bit

Mémoire requise pour stocker un son:

$\text{poids (octet)} = \text{Fréquence d'échantillonnage (Hz)} \times \text{Résolution (octet)} \times \text{Durée (seconde)} \times \text{Nombre de voies}$

Exemple: Calcul d'une seconde d'audio qualité CD

Rappel: 1octet = 8bit et 1kilo-octet (ko) = 1024 octet

⇒ Calculer le poids d'1 minute audio en 44100Hz, 16bit, stéréo.

⇒ $44100(\text{hz}) \times 16 (\text{bit}) \times 60 (\text{sec}) \times 2 (\text{voies})$

⇒ $44100 \times 2 \times 60 \times 2 = 10584000 \text{ octet}$

Conversion de octet vers Kilo-octet (Ko): $10584000 / 1024 = 10335 \text{ Ko}$

Conversion de octet vers Mega-octet (Mo): $10335 / 1024 = 10 \text{ Mo}$

R. Mosqueron (HES-SO / HEIG-VD / REDS), 2017

Production (Studio):

Fech = 48 kHz ; NQ = 16 à 24 bits ; Mono/Stéréo/...

Débit = 768 kbit/s à 2,3 Mbit/s

Diffusion:

Fech = 32 kHz ; NQ = 14 à 16 bits ; Mono/Stéréo

Débit = 448 kbit/s à 1 Mbit/s

Téléphone (Studio):

Fech = 8 kHz ; NQ = 12 bits ; Mono

Débit = 96 kbit/s

Compact Disc:

Fech = 44.1 kHz ; NQ = 16 bits ; Stéréo

Débit = 1.4 Mbit/s

C'est quoi la compression?

La compression de donnée consiste à obtenir des fichiers plus léger, afin d'améliorer la vitesse de transfert sur internet ou limité l'espace de stockage utilisé sur un disque dur.

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 $x(n)$



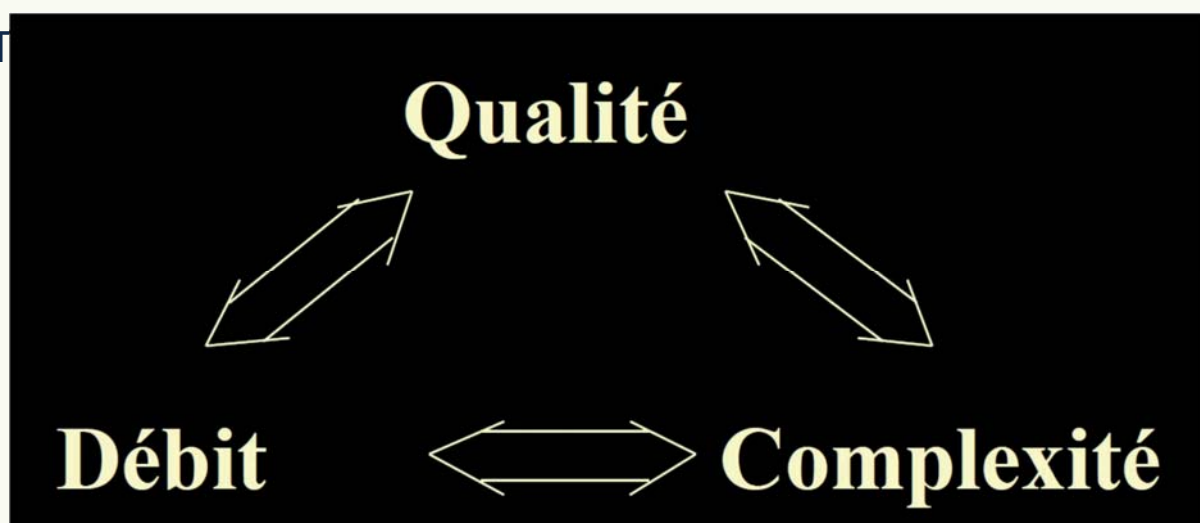
Ne pas confondre compression dynamique et compression de données.

C'est quoi la compression?

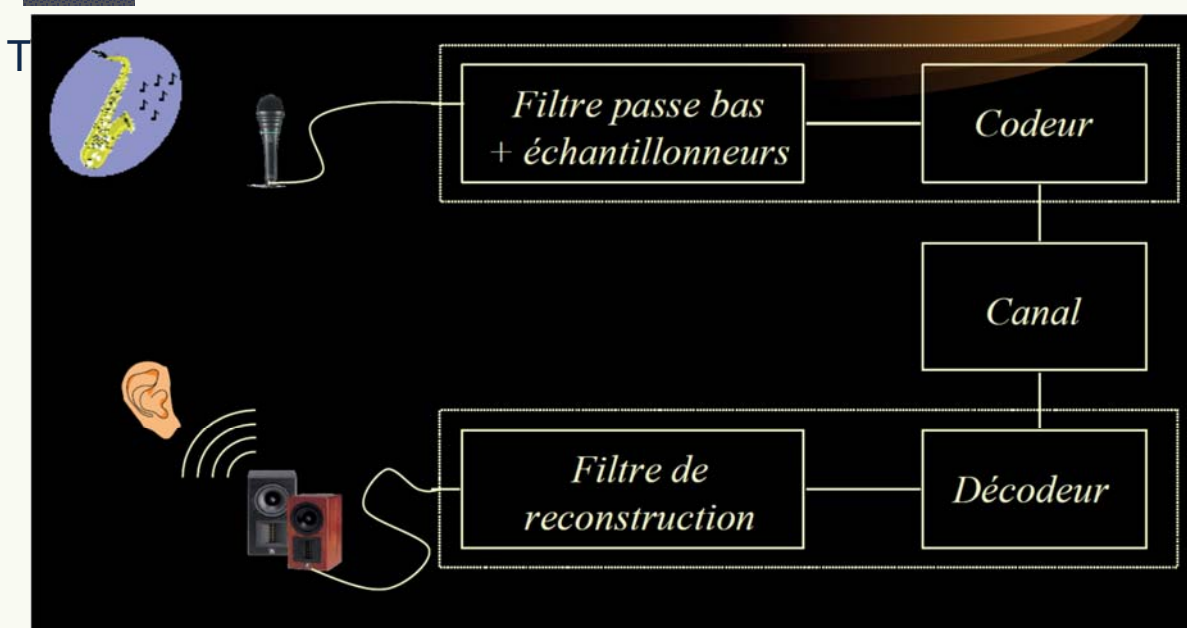
- Techniques permettant la réduction des coûts de transmission ou de stockage des signaux numériques
- par exploitation des caractéristiques statistiques des signaux
 - Par exploitation des phénomènes de masquage des systèmes perceptuels humains.
 - Éventuellement par exploitation d'un modèle de production (ex. codage de la parole)

Compromis

T



- 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



Normalisé?

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)...

Comment?

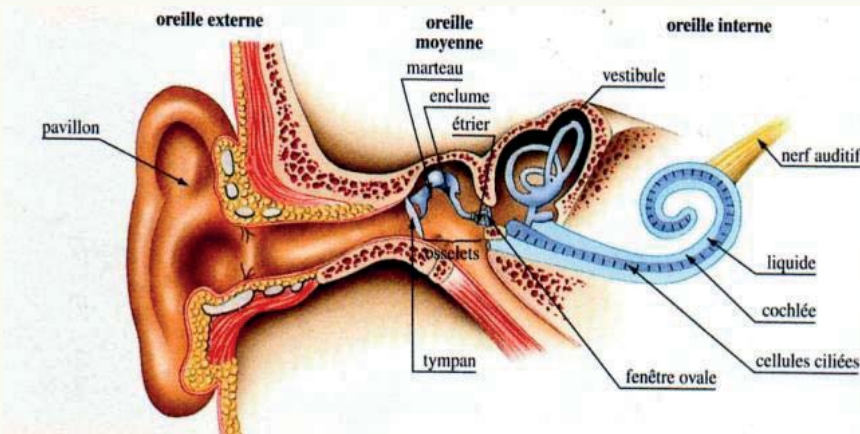
- But: diminuer le nombre de bits de quantification en rendant l'augmentation de bruit de quantification inaudible
- Principe : Exploiter les propriétés de l'audition (en particulier les effets de masque)

L'oreille

Rappel

ReDS
heig-vd

- L'oreille externe contient : le pavillon, le conduit auditif ainsi que le tympan.
- Les vibrations acoustiques « sont captées » par le pavillon, puis canalisés dans le conduit auditif pour atteindre le tympan.



06/12/2017

Systèmes audio-visuel / RMQ / V1.0

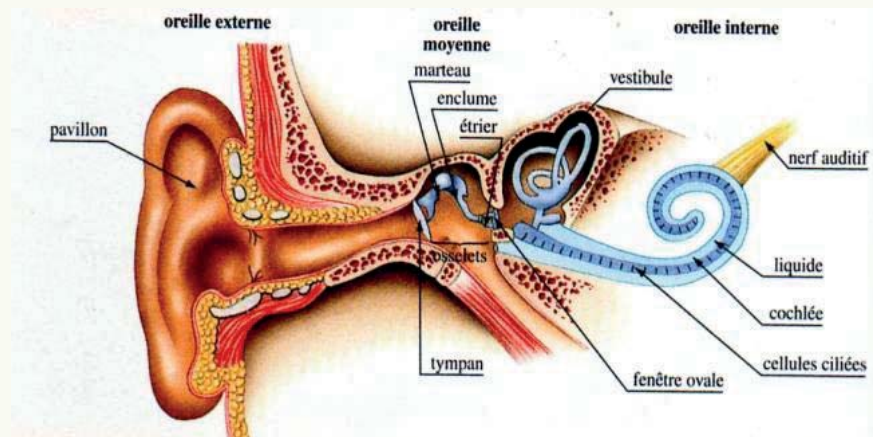
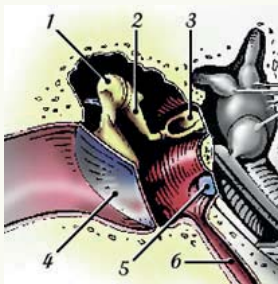
17

L'oreille

Rappel

ReDS
heig-vd

- L'oreille moyenne est composée d'une chambre d'air dans laquelle se trouve un dispositif mécanique constitué de trois osselets : le marteau, l'enclume et l'étrier.

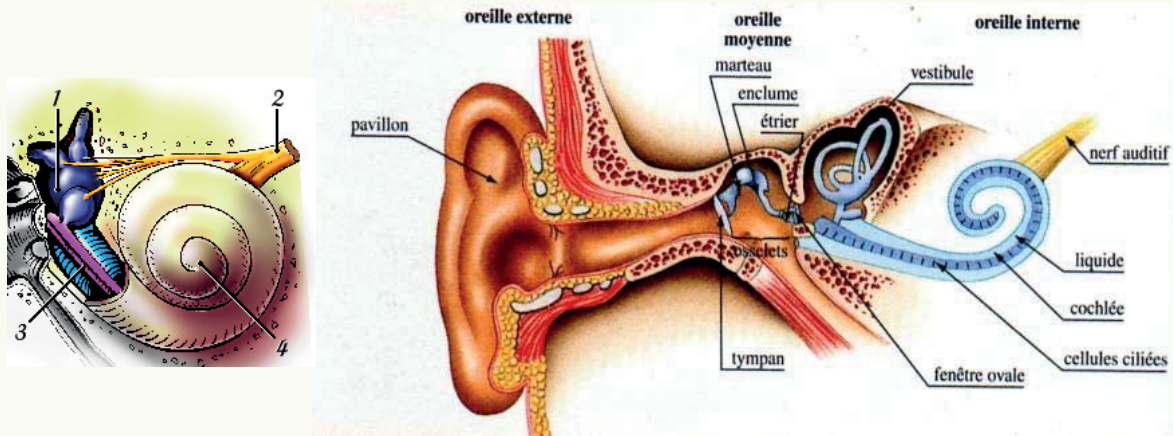


06/12/2017

Systèmes audio-visuel / RMQ / V1.0

18

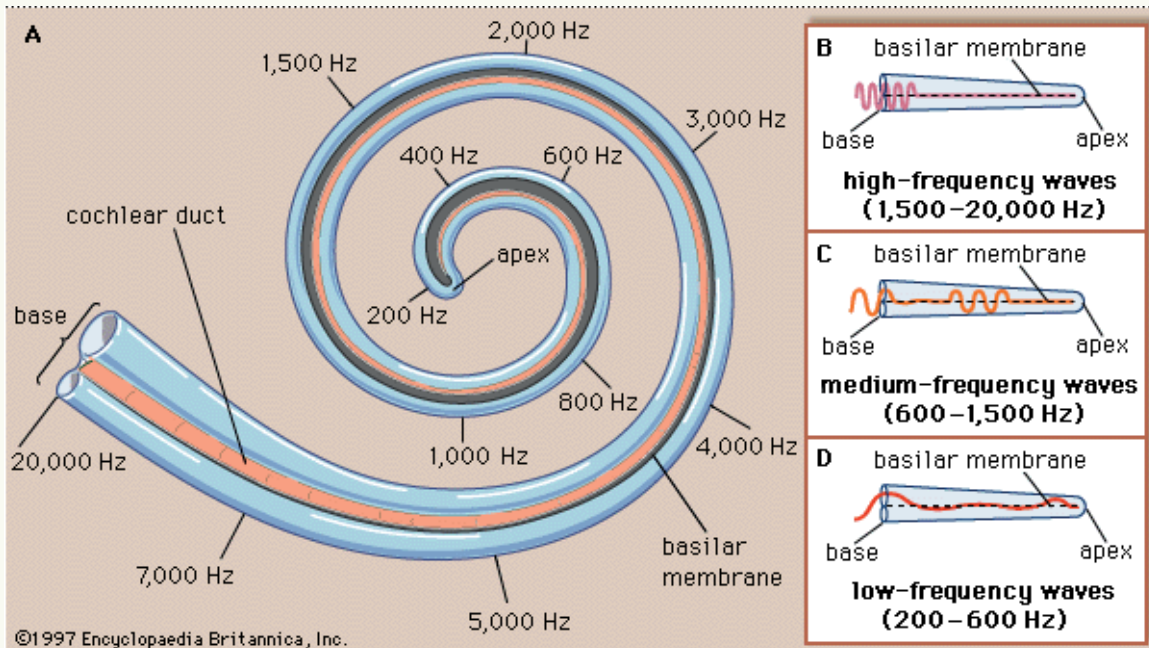
- L'oreille interne est composée de 2 parties distinctes : le vestibule, qui est l'organe de l'équilibration, et la cochlée, organe de l'audition.



La cochlée est formée de 3 parois à l'intérieur desquelles est contenu un liquide : l'endolymphe. Une de ces 3 parois, appelée membrane basilaire, est plus épaisse que les autres car elle renferme l'organe de Corti. Cet organe est l'élément sensible de l'ouïe. Sur cet organe se trouve ce que l'on appelle des cellules ciliées réparties sur quatre rangées. 3 rangées contiennent les cellules ciliées dites externes et la dernière rangée des cellules ciliées internes.

Sous l'effet du son, la membrane basilaire se met en mouvement. Les cellules ciliées situées à cet endroit sont excitées sous l'effet de la membrane en mouvement. Ces cellules sont elles-mêmes reliées au nerf auditif qui transmet l'information au cerveau. On peut alors se demander quelles cellules se mettent en mouvement selon la ou les fréquences qui composent le son.

Caractéristiques utiles de l'oreille

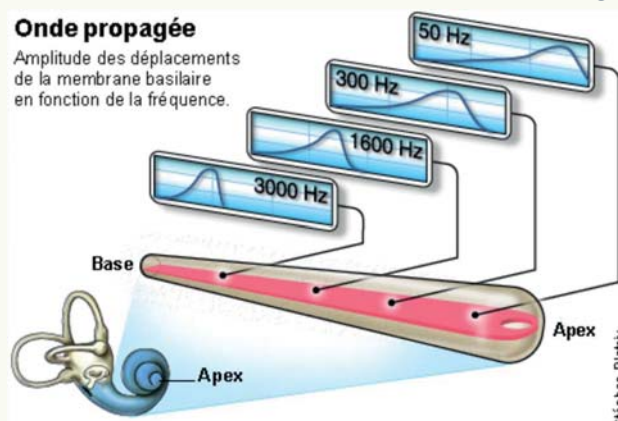


R. Mosqueron (HES-SO / HEIG-VD / REDS), 2017

Caractéristiques utiles de l'oreille

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



R. Mosqueron (HES-SO / HEIG-VD / REDS), 2017

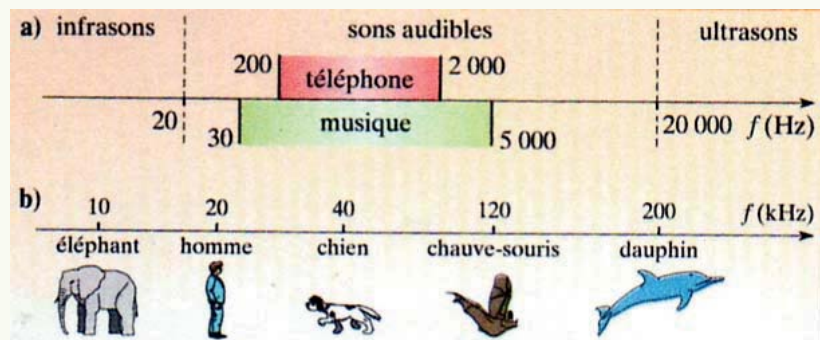
Fréquences audibles:

Une oreille normale perçoit les sons de fréquences comprises entre environ 20 Hz et 20 000 Hz.

Cet intervalle se réduit avec l'âge.

a) Domaine des fréquences audibles pour l'oreille humaine.

b) Limites supérieures des fréquences audibles chez l'homme et chez certains animaux.



Sensation auditive

- Dépend de l'intensité des sons reçus.

On appelle intensité acoustique (ou intensité sonore) notée I , la puissance acoustique reçue par unité de surface de récepteur. Elle se mesure en watt par mètre carré (symbole: $W.m^{-2}$).

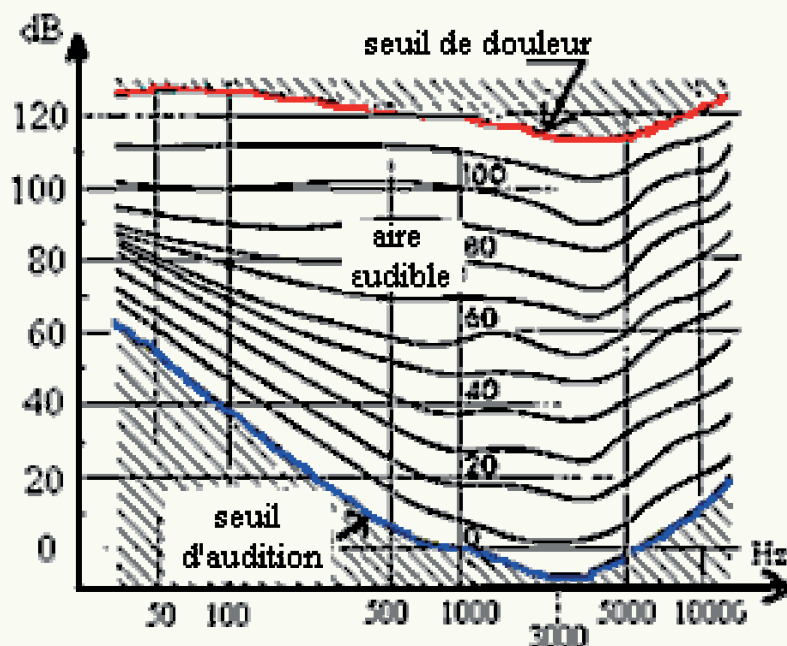
- N'est pas proportionnelle à l'intensité acoustique. Son niveau L est mesuré en décibel (dB) avec un sonomètre.

Lorsque l'intensité acoustique I est multipliée par 2, le niveau d'intensité acoustique L augmente seulement de 3 dB.

- Dépend également de la fréquence.

L'oreille présente une sensibilité maximale autours de 3 000Hz.

Seuil audible:



06/12/2017

Systèmes audio-visuel / RMQ / V1.0

25

2 possibilités

- **Compression sans perte**

C'est le type de compression "zip" (Huffman), qui recherche des séquences afin de réduire la quantité de données qui doivent être stockées

- **Compression avec pertes**

Les modèles psycho acoustiques, c'est le domaine principal pour la compression audio.

C'est la partie où la compression s'effectue avec des pertes, durant laquelle un encodeur va jeter de l'information pour réduire la taille.

1. Pourquoi compresser l'audio?
2. Principe de la compression de données audio
3. **Compression sans pertes**
4. Compression avec pertes
5. Mpeg (MP3)

Sous Windows est le WAV (« .wav », WAVEform audio format),
Sous Mac est le AIFF (« .aiff », Audio Interchange File Format).

Flux audio au format PCM.

Le « BMP » des formats audio.

Avantages :

- Pas de perte, donc parfait pour le traitement.
- Lisible partout.

Inconvénients :

- Pas du tout fait pour être stocké sur son baladeur, ou même sur son disque dur. De plus, le WAV ne gère pas les « tags » contrairement à tous les futurs formats listés plus bas. Donc il n'est pas fait pour réaliser une audiothèque.

La compression sans perte appelée aussi « compactage », consiste simplement à coder les données binaires de manière plus concise dans un fichier. Elle permet ainsi de retrouver la totalité des informations après une procédure de décompactage.

- Avantage de réduire le débit ou la capacité de stockage nécessaire, sans perdre d'informations de la source audio
- La réduction de taille est dépendante de la source : plus le signal est constitué d'ondes régulières (sons naturels), meilleure est la compression. Par contre, un signal aléatoire (bruit) ne se compresse pas très bien.

La compression entropique (de données sans perte), type *zip* (Huffman), qui recherche des séquences afin de réduire la quantité de données qui doivent être stockées

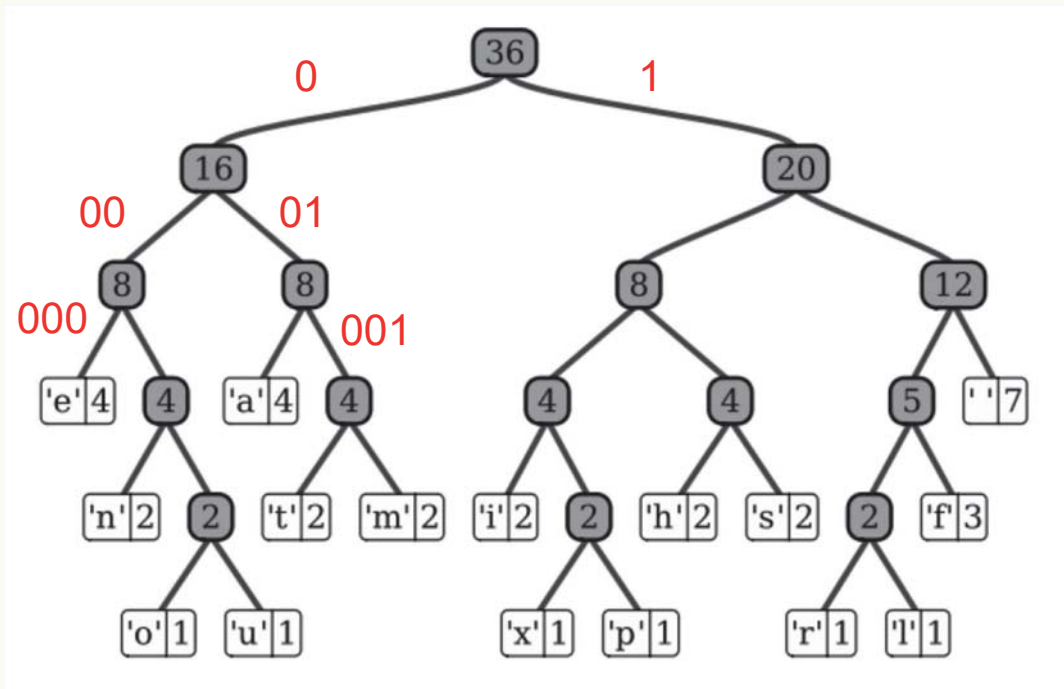
Compresser-décompresser en utilisant une table de fréquences:

- Calculer une table et l'intégrer au fichier
 - Meilleure compression
 - Fichier plus important
- Utiliser une table générique intégrée dans la compression
 - Compression plus rapide
 - Efficacité réduite

Ratio 1 > Ratio 2

Huffman

this is an example of a huffman tree



Huffman

phrase :

(a b r a c a d a b r a)

liste des frequences

((a 5) (r 2) (b 2) (d 1) (c 1))

arbre :

((11 (a 5) (6 (b 2) (4 (2 (d 1) (c 1)) (r 2))))))

code :

((a (0)) (b (1 0)) (d (1 1 0 0)) (c (1 1 0 1)) (r (1 1 1)))

phrase codee :

(0 1 0 1 1 1 0 1 1 0 1 0 1 1 0 0 0 1 0 1 1 1 0)

decodage :

(a b r a c a d a b r a)

RLE

L'encodage RLE (Run Length Encoding) consiste à remplacer toutes les suites de caractères ou bits pareils par un nombre représentant le nombre de répétitions du caractère ou bits suivit de celui-ci.

Généralement, pour moins de gaspillage d'espace, on ne remplace une chaîne que si le nombre de répétition du caractère ou du bit est supérieur à deux.

Visuelle, un encodage RLE sur un texte ressemblerait à ceci :

Chaîne d'origine :

WWWWWWWWWWWWWWBWWWWWWWWWWWWWWWWWWWWBBBWWWWW
WWWWWWWWWWWWWWWWWWWWWWWWBWWWWWWWWWWWWWWWW

Chaîne encodée :

12WB14W3B23WB11W

LZW

LZW

Free Lossless Audio Codec

un fichier FLAC est composé de blocs successifs d'environ 100 ms de données, et peut être décompressé à la volée durant la lecture (quel que soit le niveau de compression), Ce sont bien sûr les coefficients de la transformée de Fourier qui vont être compressés

Avantage:

- Rapide
- Open-source
- «Friendly»

Inconvénient:

- Moins bon taux de compression (comparé avec APE)

https://xiph.org/flac/documentation_format_overview.html

Le APE (Monkey's Audio)

Un format légèrement plus lent que le FLAC à l'encodage et au décodage, mais qui offre un meilleur taux de compression. Il est également légèrement moins supporté au niveau matériel et logiciel (par exemple, il n'est pas supporté nativement par le lecteur foobar2000 et certaines distributions du libre, contrairement au FLAC).

Avantages :

- Bon taux de compression, rapide...

Inconvénients :

- Une licence « spéciale » qui visiblement n'a pas aidé à son expansion dans le monde du libre et au niveau des périphériques.

WavPack (« .wv »)

Entre le FLAC et l'APE au niveau de la compression, et est même meilleurs que le FLAC en terme de vitesse. Seulement, il est moins supporté au niveau matériel. Il propose également une caractéristique intéressante, nommée

Hybrid, qui consiste à proposer un mode de compression avec perte. Le fichier compressé avec perte peut être complété d'un fichier de correction permettant de rétablir les informations perdues.

Avantages :

- Bon taux de compression, et en plus très rapide à l'encodage et au décodage.

Inconvénients :

- Aucun à part le fait d'être moins bien supporté que le FLAC.

R. Mosqueron (HES-SO / HEIG-VD / REDS), 2017

Comparaison

Nom	Musepack	FLAC	APE	WavPack
Paramètres/indice de qualité	(original)	-5 (défaut)	-c2000 (défaut)	(défaut)
Taille du fichier	52,3Mo	41,4Mo	40,5Mo	41,2Mo
Temps d'encodage		Moins de 5s	4,4s	3,91s
Domaine d'application	Traitement audio, CD audio	Stockage	Stockage	Stockage

1. Pourquoi compresser l'audio?
2. Principe de la compression de données audio
3. Compression sans pertes
4. **Compression avec pertes**
5. Mpeg (MP3)

L'idée qui introduit cette nouvelle forme de compression vient des possibilités humaines.

Un oeil humain ne fera pas la différence entre deux couleurs très proches parmi les 16 millions disponibles. Pourquoi dans ce cas, ne pas réduire le nombre de couleurs employées par l'image? C'est ce que font les formats de compression JPEG, PING et GIF, et, de ce fait, ils réduisent de manière radicale la taille des images.

De même, aucun homme normalement constitué ne peut entendre un son à la fréquence de 20 Khz. En supprimant ces informations inaudibles par le commun des mortels, on diminue significativement le nombre de données à prendre en compte et on obtient donc un fichier plus léger.

Chasse aux informations inutiles:

- Pas capable de percevoir la stéréo dans les extrêmes graves.
⇒ toutes les fréquences trop basses en mono
- Certaines fréquences en écrasaient d'autres -> superflues.

L'algorithme de compression repère donc les sons "dominants" et retire toutes les données relatives aux sons "dominés". Puisque de toutes façons, on ne les aurait pas entendus, cela ne fait guère de différence au niveau sonore tout en réduisant notablement la taille du fichier.

Plusieurs techniques psycho-acoustiques => encoder que les informations utiles

Etude de la perception auditive chez l'homme,

Caractérisation,

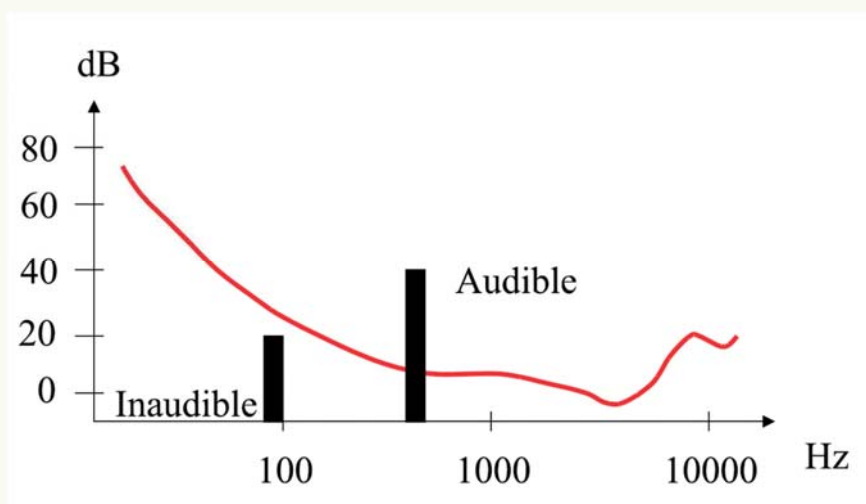
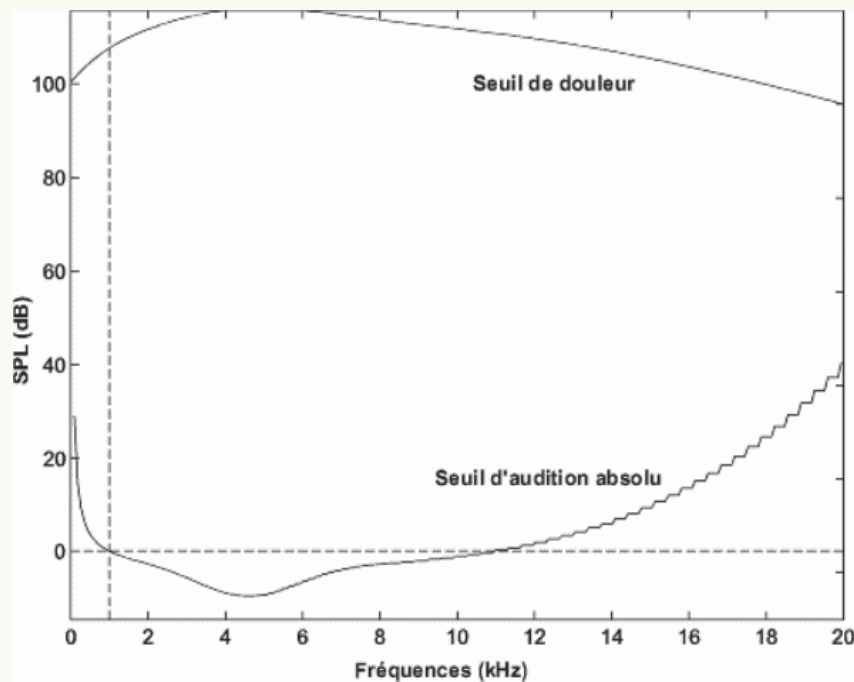
Analyse temps-fréquence des capacités de l'oreille humaine,

Relation entre grandeurs physiques et grandeurs perceptuelles,

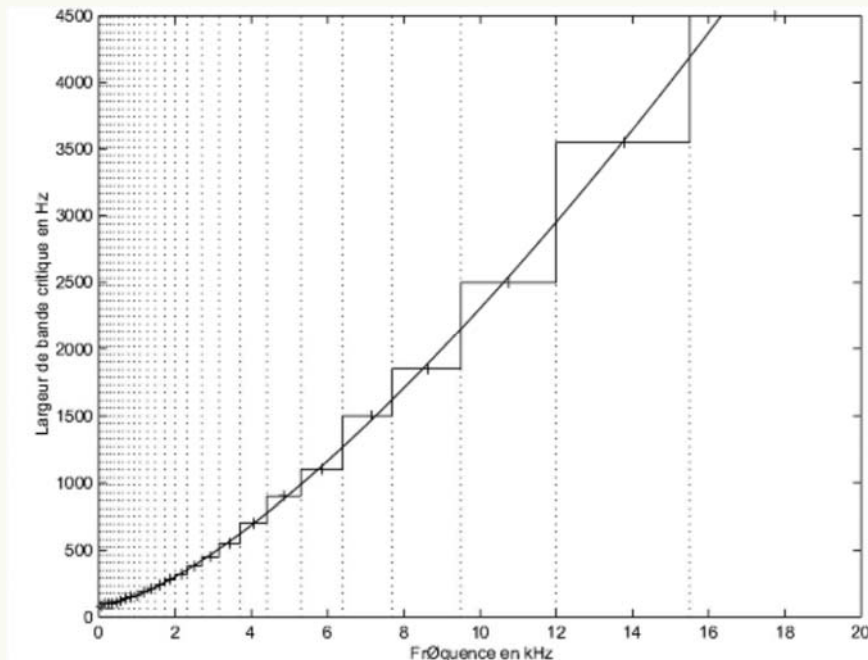
Protocoles d'expérimentation pour tests psycho-acoustiques.

Paramètres

- La fréquence
- Le masquage fréquentiel
- Le masquage temporel



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

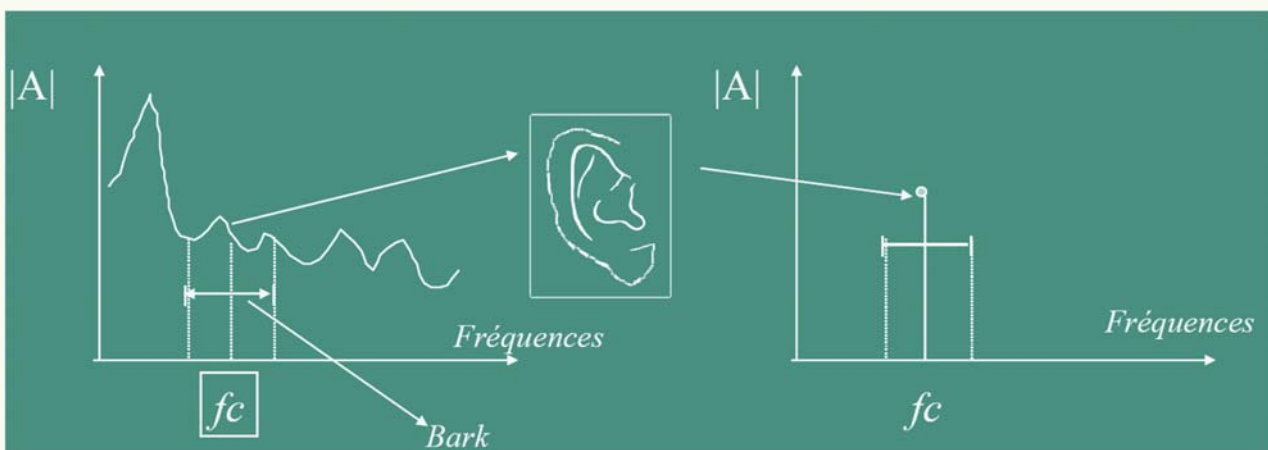


06/12/2017

Systèmes audio-visuel / RMQ / V1.0

45

Caractérisent les bandes passantes des « filtres cochléaires » et représentent ainsi les régions où l'oreille est capable d'effectuer une intégration fréquentielle (1 bande critique = 1 bark)



06/12/2017

Systèmes audio-visuel / RMQ / V1.0

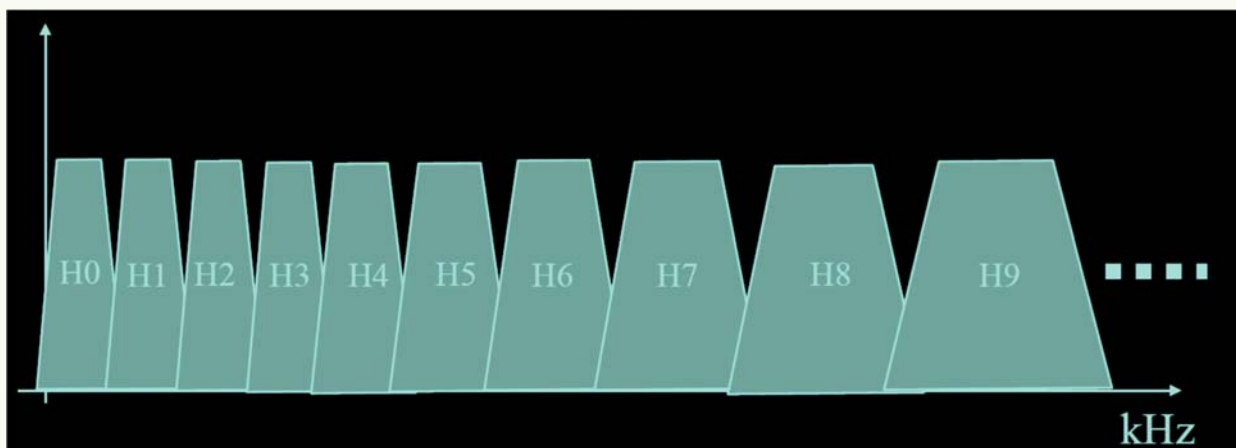
46

Psycho-acoustiques

Barks	Fréq. centr. [Hz]	Larg. bande [Hz]	Fréq. super. [Hz]
1	50	- 100	100
2	150	100	200
3	250	100	300
4	350	100	400
5	450	110	510
6	570	120	630
7	700	140	770
8	840	150	920
9	1000	160	1080
10	1170	190	1270
11	1370	210	1480
12	1600	240	1720
13	1850	280	2000
14	2150	320	2320
15	2500	380	2700
16	2900	450	3150
17	3400	550	3700
18	4000	700	4400
19	4800	900	5300
20	5800	1100	6400
21	7000	1300	7700
22	8500	1800	9500
23	10500	2500	12000
24	14000	4000	16000

Psycho-acoustiques

L'oreille se comporte comme un banc de filtres répartis sur une échelle Bark



Plusieurs techniques psycho-acoustiques => encoder que les informations utiles

Etude de la perception auditive chez l'homme,

Caractérisation,

Analyse temps-fréquence des capacités de l'oreille humaine,

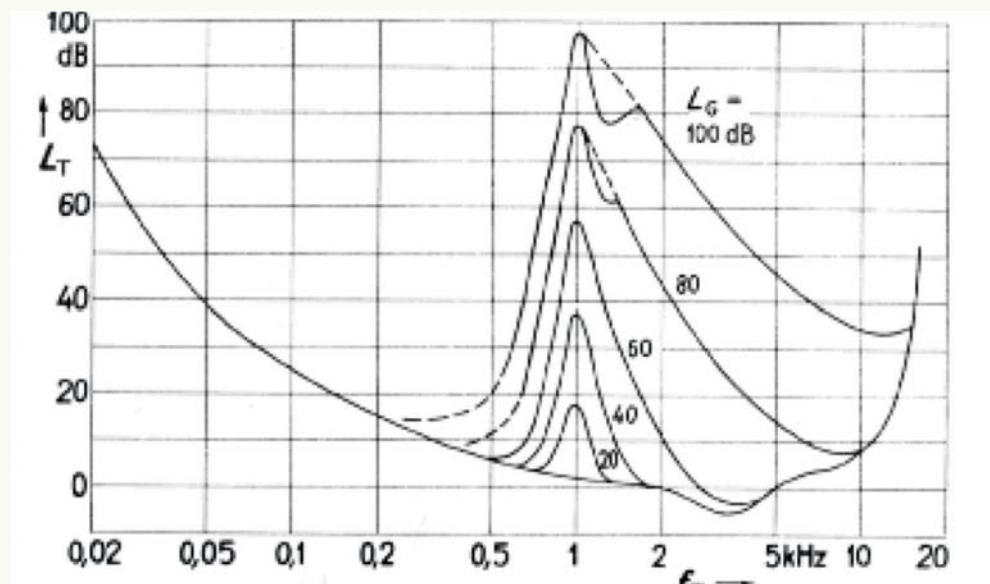
Relation entre grandeurs physiques et grandeurs perceptuelles,

Protocoles d'expérimentation pour tests psycho-acoustiques.

Paramètres

- La fréquence
- Le masquage fréquentiel
- Le masquage temporel

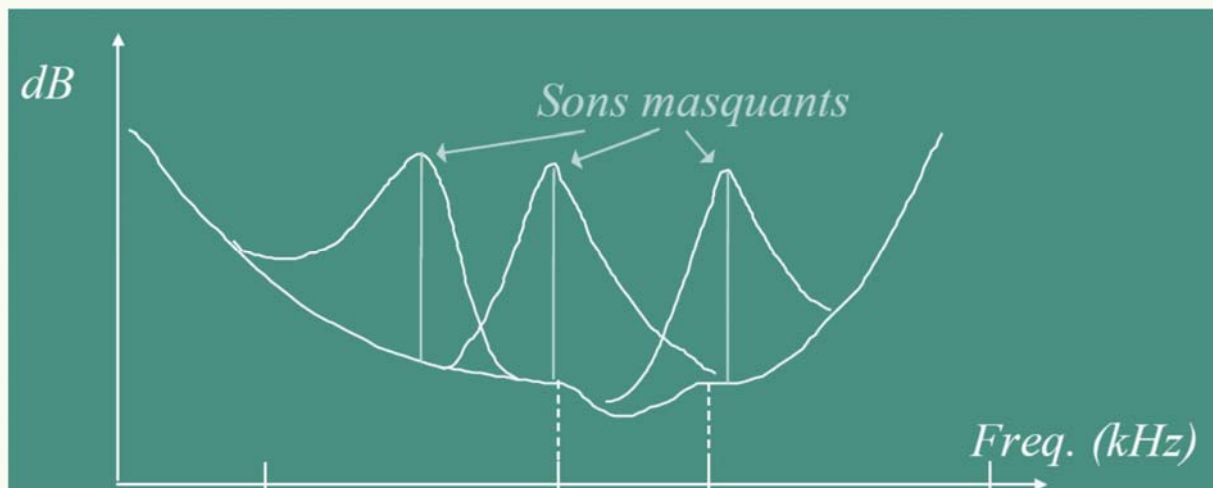
Niveau SPL nécessaire à un son "masqué" pour être entendu en présence d'un "masquant" donné (fonction de la fréquence)



Masquage fréquentiel

« Attention, un son peut en cacher un autre.... »

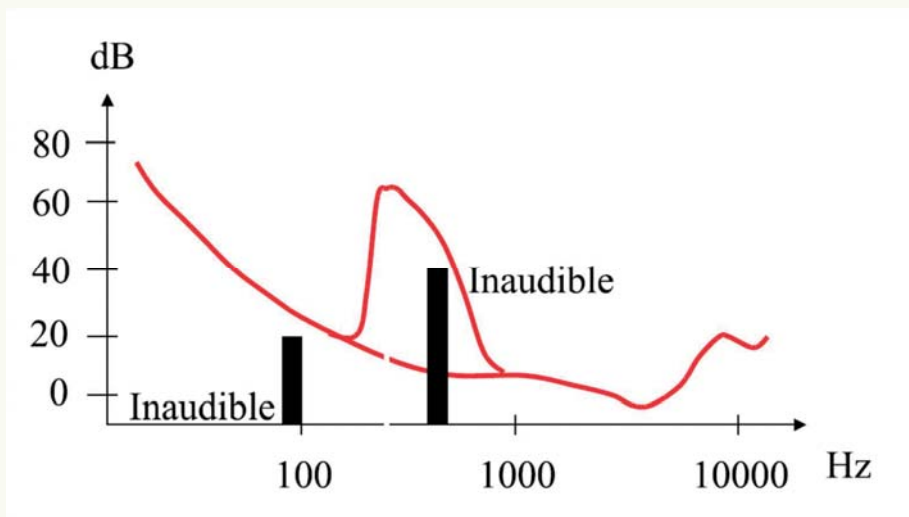
Le seuil d'audibilité est modifié en présence d'un son masquant



Masquage fréquentiel

« Attention, un son peut en cacher un autre.... »

Le seuil d'audibilité est modifié en présence d'un son masquant



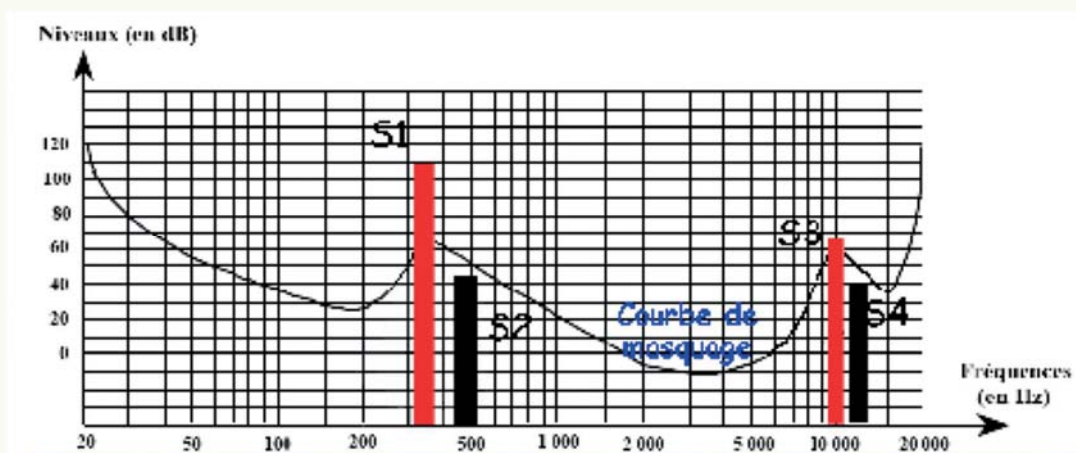
Masquage fréquentiel

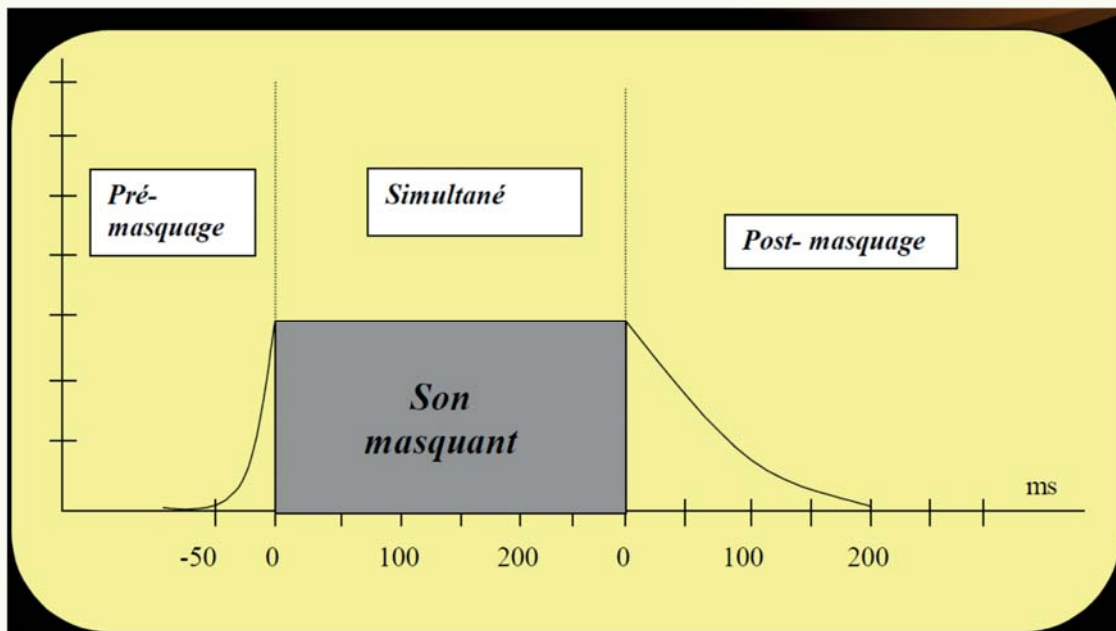
On parle de masquage fréquentiel lorsqu'un son faible - qui serait parfaitement audible s'il était émis seul - est masqué parce qu'il se trouve accompagné simultanément par un son fort de fréquence voisine (son « masquant »).

Il est inutile de coder les signaux qui sont situés en dessous. Cette courbe de « masquage » (les variations du seuil d'audition donc) variant à chaque instant en fonction du contenu spectral du signal, c'est donc une véritable analyse en temps réel qui doit être réalisée par les circuits de codage.

Masquage fréquentiel

Les variations du seuil d'audibilité avec le masquage fréquentiel: les deux sons S2 et S4 seraient parfaitement audibles s'ils étaient émis seuls. En présence des sons S1 et S3, ils ne sont plus perceptibles.





Le terme de masquage temporel fait référence au masquage réalisé après l'apparition d'un son masquant de forte intensité, mais également, au masquage.

Le terme de temporel indique que le son masquant et le son masqué sont décalés dans le temps par opposition au masquage fréquentiel qui ne concerne que des sons simultanément présents.

Le pré-masquage est très court et ne dure que quelques millisecondes (2 à 5 ms). Les sons ayant une durée trop courte ne seront pas perçus.

Le post-masquage est beaucoup plus long (100 à 200 ms) et dépend des caractéristiques du son masquant.

Après un son fort, l'oreille ne pourra percevoir un son plus faible qu'au terme de ce laps de temps

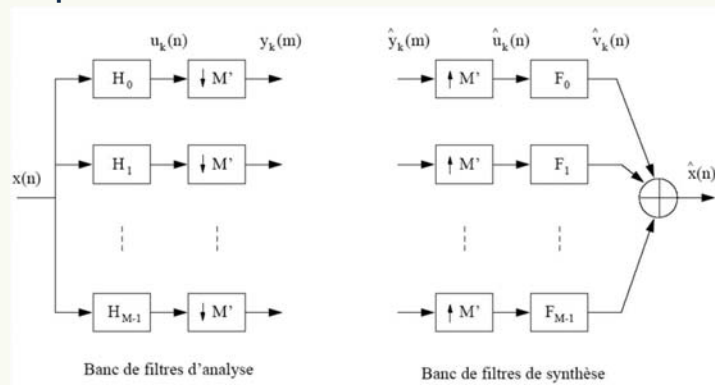
Filtres passe-bande : en général, filtre prototype passe-bas modulé

Décimation dans chaque sous-bande

Nombre de bandes : M ;

Facteur de décimation : M' ;

Décimation critique $M = M'$



06/12/2017

Systèmes audio-visuel / RMQ / V1.0

57

Compression audio

1. Pourquoi compresser l'audio?
2. Principe de la compression de données audio
3. Compression sans pertes
4. Compression avec pertes
5. **Mpeg (MP3)**

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

Codage audio MPEG -> réduire les données du son original d'un CD d'un facteur 12 sans perte de qualité auditive.

Facteurs de 24 possible -> une qualité de son significativement meilleur que celle obtenue en réduisant simplement le taux d'échantillonnage et la résolution des échantillons.

Réalisé par des techniques de codage perceptuel adressant la perception d'ondes sonores par l'oreille humaine.

En utilisant l'audio MPEG on peut atteindre une réduction des données de :

- **1:4 pour la couche 1**, qui correspond à 384 kbps pour un signal stéréo
- **1:6 à 1:8 pour la couche 2**, qui correspond à 256 à 192 kbps pour un signal stéréo
- **1:10 à 1:12 pour la couche 3**, qui correspond à 128 à 112 kbps pour un signal stéréo

En maintenant toujours une qualité CD.

En exploitant des effets stéréos et en limitant la bande passante audio, le codage peut procurer un son de qualité acceptable à des taux encore plus bas. La couche 3 est la plus performante des techniques de la famille MPEG.

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

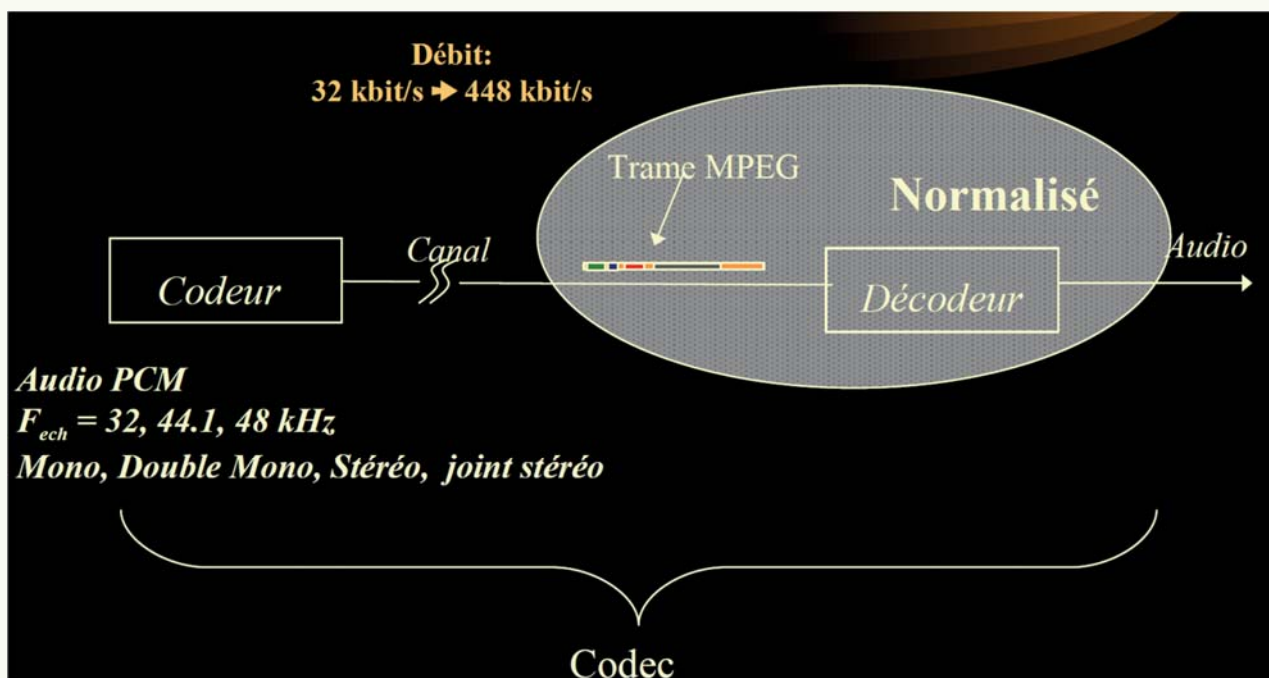
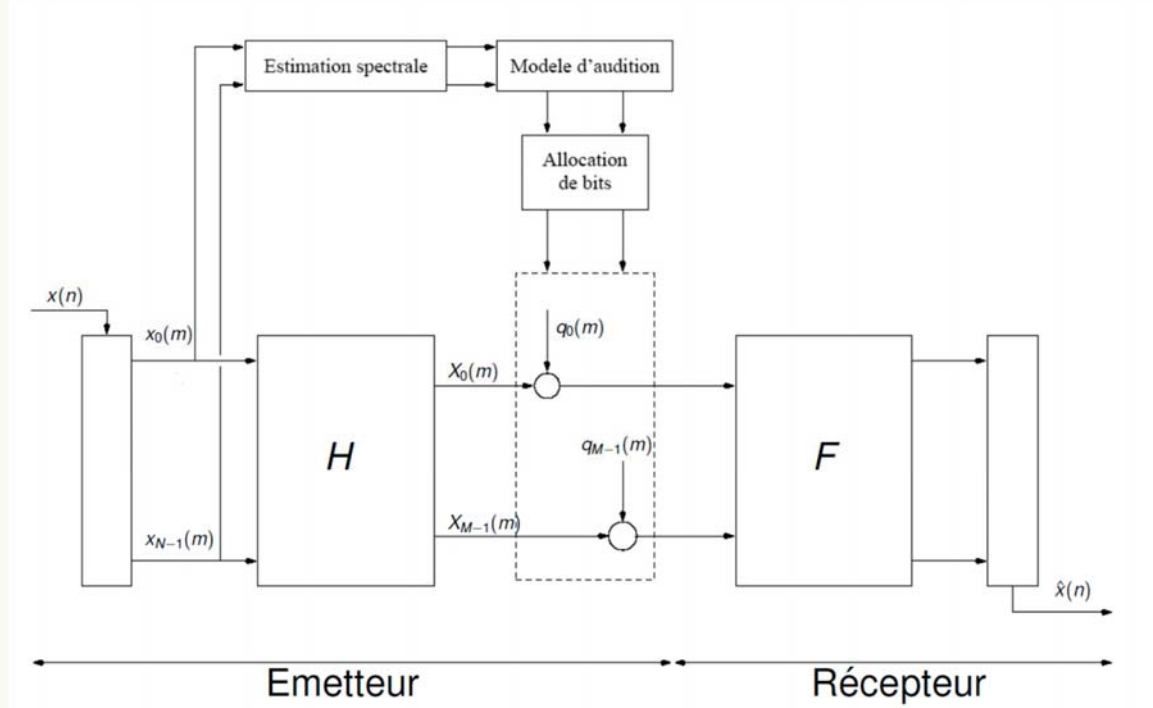
mp3 = MPEG-1 couche III, stockage de données

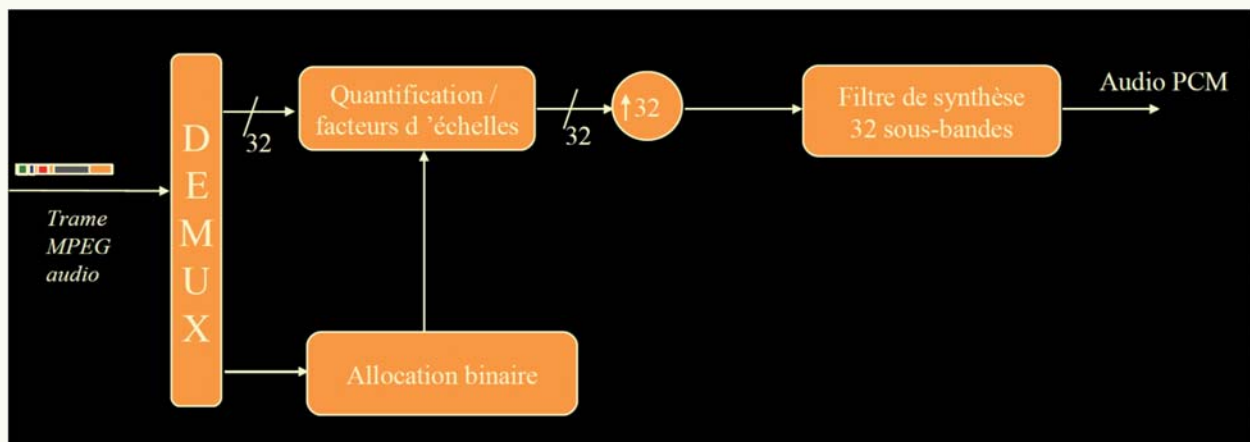
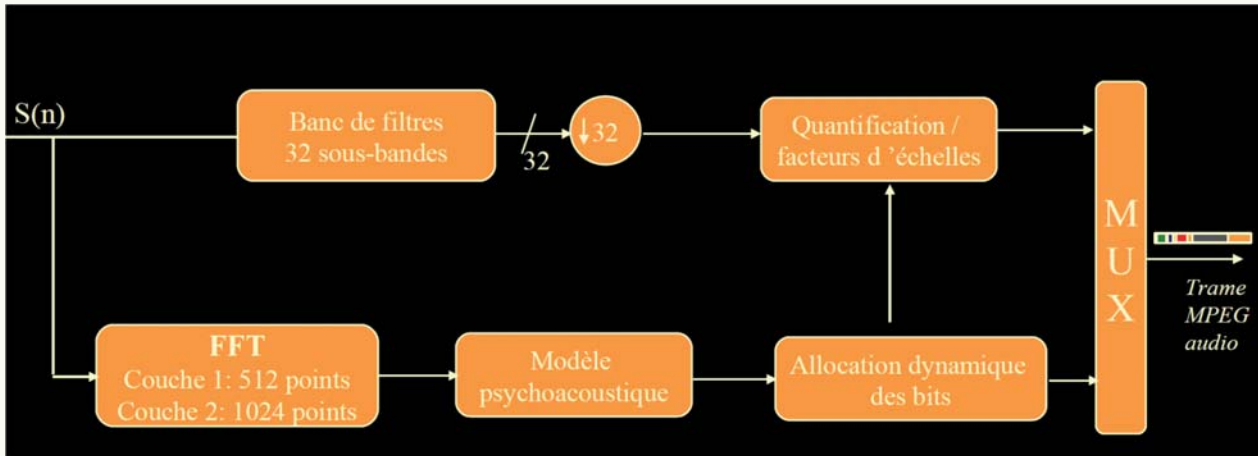
Débits : 32, 40, 48 . . . , 96, . . . , 224,256,320 kbits/s

Retard minimum théorique de codage/décodage : 59 ms

MPEG

- 1) Le signal est découpé en petites sections appelées frames
- 2) Ces chiffres sont ensuite comparés à des tables de données propres au Codec, qui contiennent des informations sur les modèles psycho-acoustiques.
- 3) Toute information qui correspond au modèle psycho-acoustique est conservée et le reste est rejeté. Ce sont les bases de la compression audio.
- 4) En fonction du bit-rate (nombre de bits par seconde), le codec utilise la taille allouée pour stocker ces données.
- 5) Le résultat passe par la compression sans pertes Huffman, qui réduit encore la taille de 10%.





Description

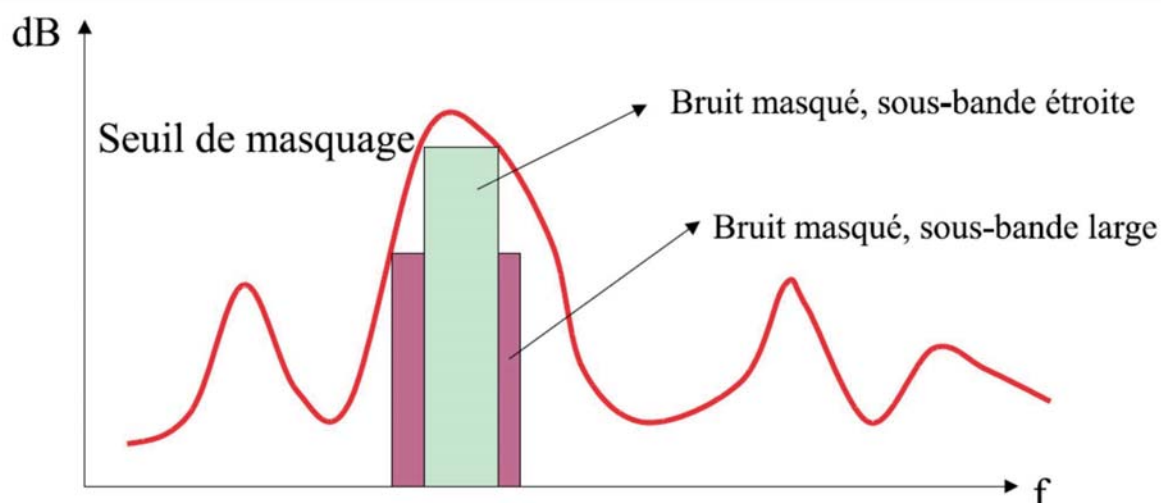
Existence de bandes critiques (oubarks):

Le niveau du bruit masqué dépend uniquement de l'énergie du signal dans une bande critique, et non de sa répartition spectrale dans la bande.

Une analyse par bandes critiques compliquerait le traitement du signal. De plus, elle n'est pas absolument nécessaire. Plusieurs codes perceptuels découpent le spectre audio en bandes de largeurs égales, par exemple en 32 sous-bandes de largeur uniforme.

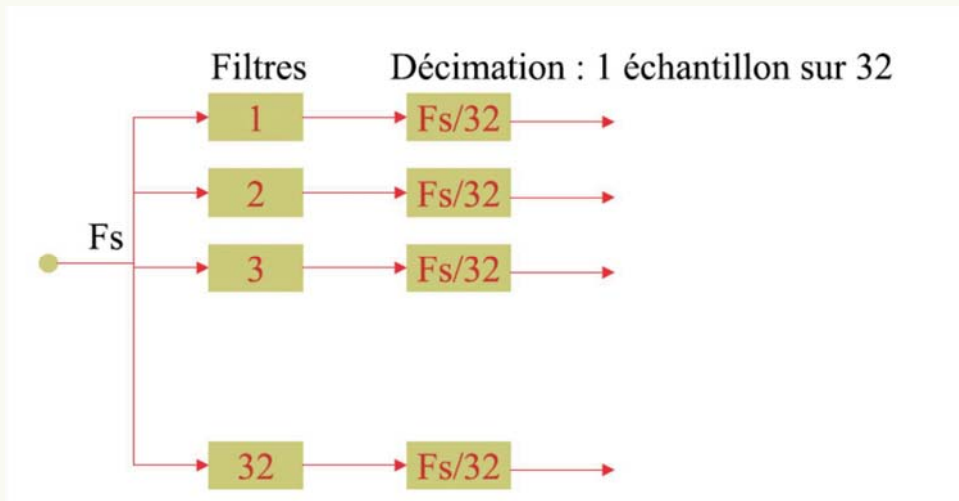
Description

Plus la sous bande est étroite, plus le bruit admissible sera élevé
=> Codage plus efficace



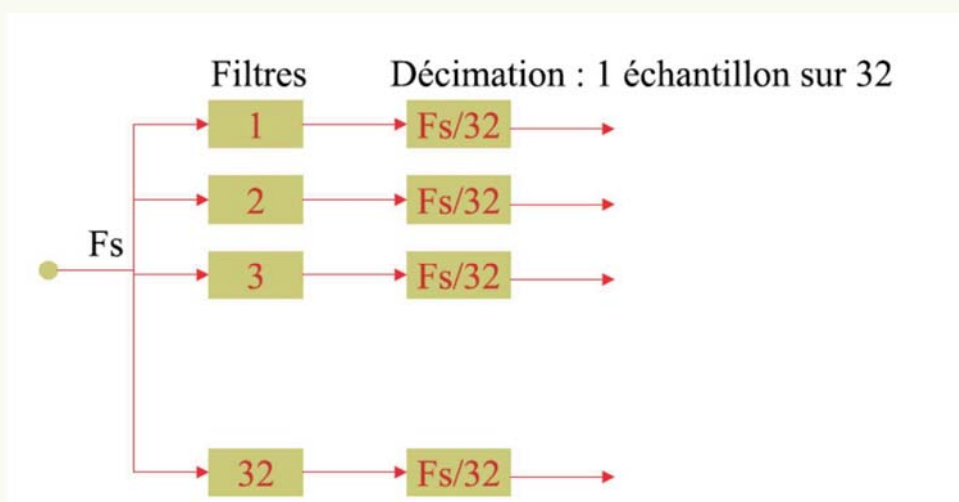
Banc de filtres

- 32 sous bandes de largeur égale
- 32 filtres passe-bande (théorie ne se recouvrent pas)



Banc de filtres

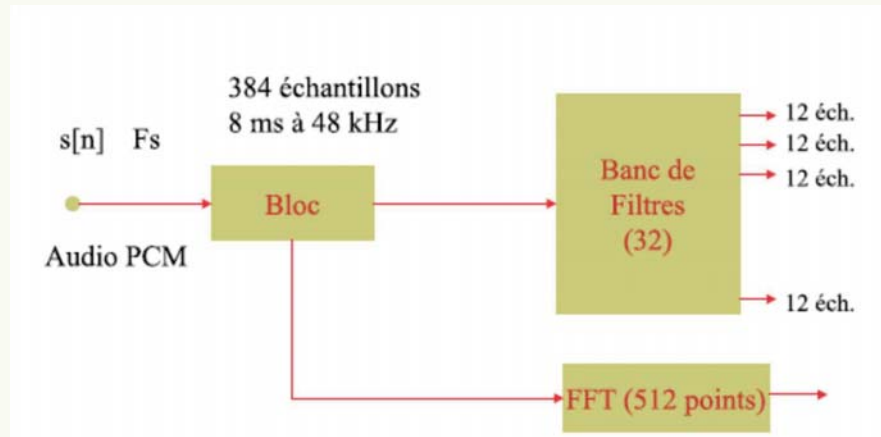
- 32 sous bandes de largeur égale
- 32 filtres passe-bande (théorie ne se recouvrent pas)



384 échantillons/frame \Rightarrow 8ms pour 48kHz
 \Rightarrow 12 échantillons par bloc à la sortie

En parallèle:

Spectre avec FFT de 512 points \Rightarrow alimente le module psycho-acoustique.



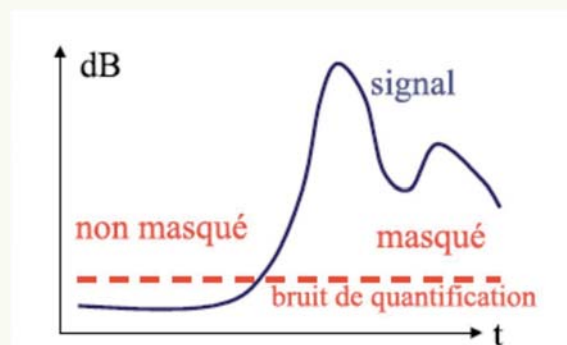
06/12/2017

Systèmes audio-visuel / RMQ / V1.0

71

8ms trop long pour assurer un pré-masquage en cas de changement brutal:

- Détecter ces apparitions
- Identifier les blocs influencés
- Eviter quantification trop grossière



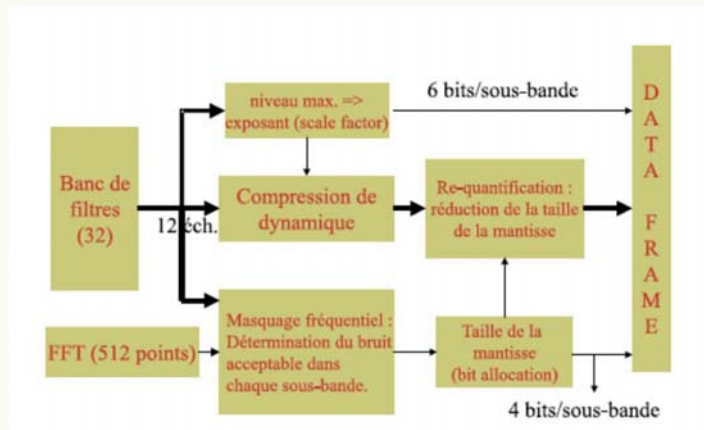
06/12/2017

Systèmes audio-visuel / RMQ / V1.0

72

Un seul bloc considéré:

- Niveau max -> facteur d'échelle => 1 seule valeur par sous bloc
6 bits donc 128dB par pas de 2 dB (exposant dans notation floating point)
- Mantisse réduite au nombre de bits déterminé par le bruit de quantification
- 4 bits pour que le décodeur soit informé de ce nombre de bits



06/12/2017

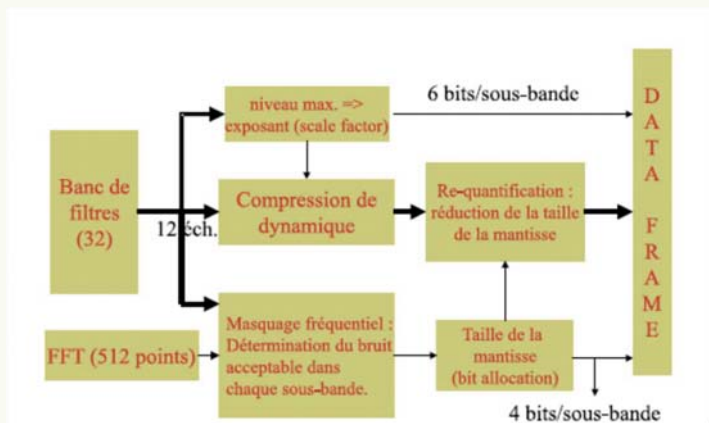
Systèmes audio-visuel / RMQ / V1.0

73

Data frame

- 32 facteurs d'échelle (6*32 bits)
- 32 nombres de bits alloués à chaque sous-bande
- Mantisses des 384 échantillons

Bitstream = ensemble des data-frames



06/12/2017

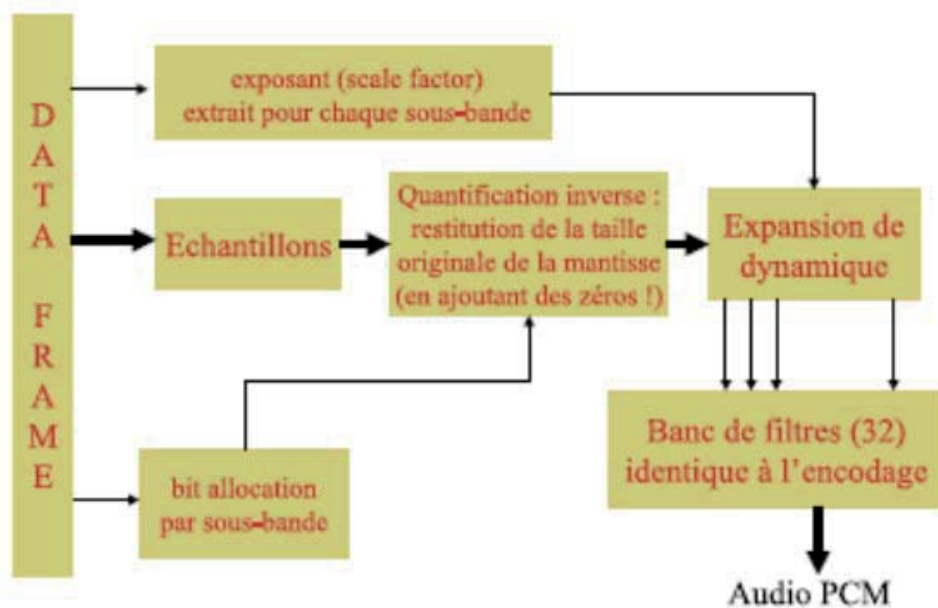
Systèmes audio-visuel / RMQ / V1.0

74

Décodage: bitstream découpé en data frames, correspondant chacune à un bloc temporel.

Extraction de:

- facteurs d'échelle pour chacune des 32 sous-bandes,
- les nombres de bits adoptés pour les mantisses dans chaque sous-bande,
- les mantisses des 12 échantillons dans chaque sous-bande.
(portées au nombre de bits initial (en ajoutant des zéros),
- multipliées par le facteur d'échelle correspondant pour revenir à la notation PCM-16bits (expansion de dynamique).
- 32 signaux sont interpolés (à F_s) repassent dans le banc de filtres initial pour recréer le signal "large bande»



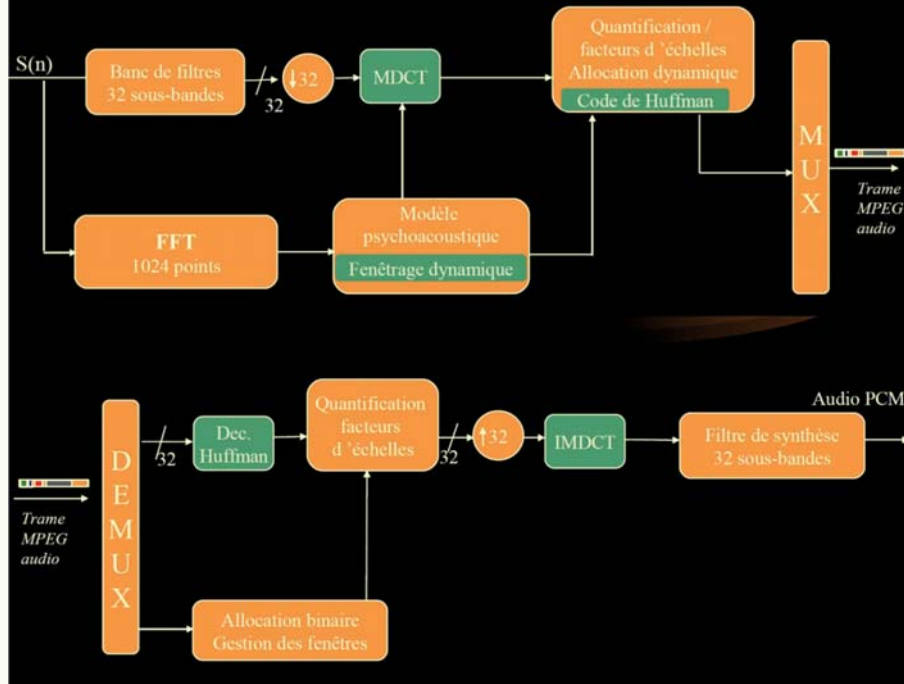
Layer 2:

info principale = 32 x 36 échantillons = 1152 PCM échantillons =
24 ms à 48 kHz

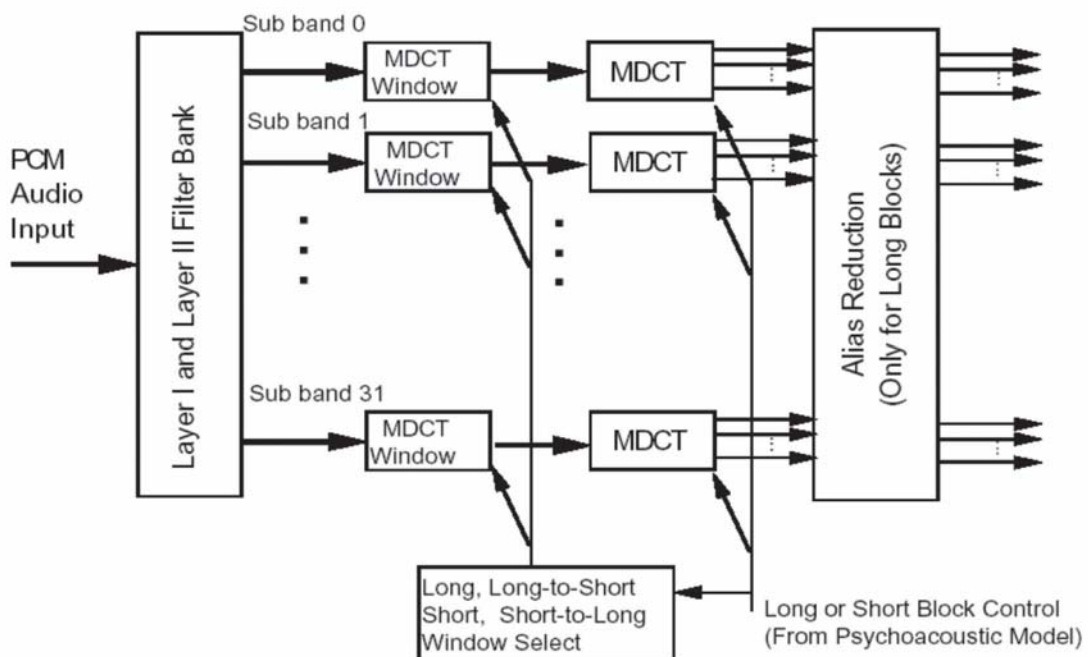
FFT à 1024 points

- 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

MPEG1-Layer3 (MP3)



MPEG1-Layer3 (MP3)



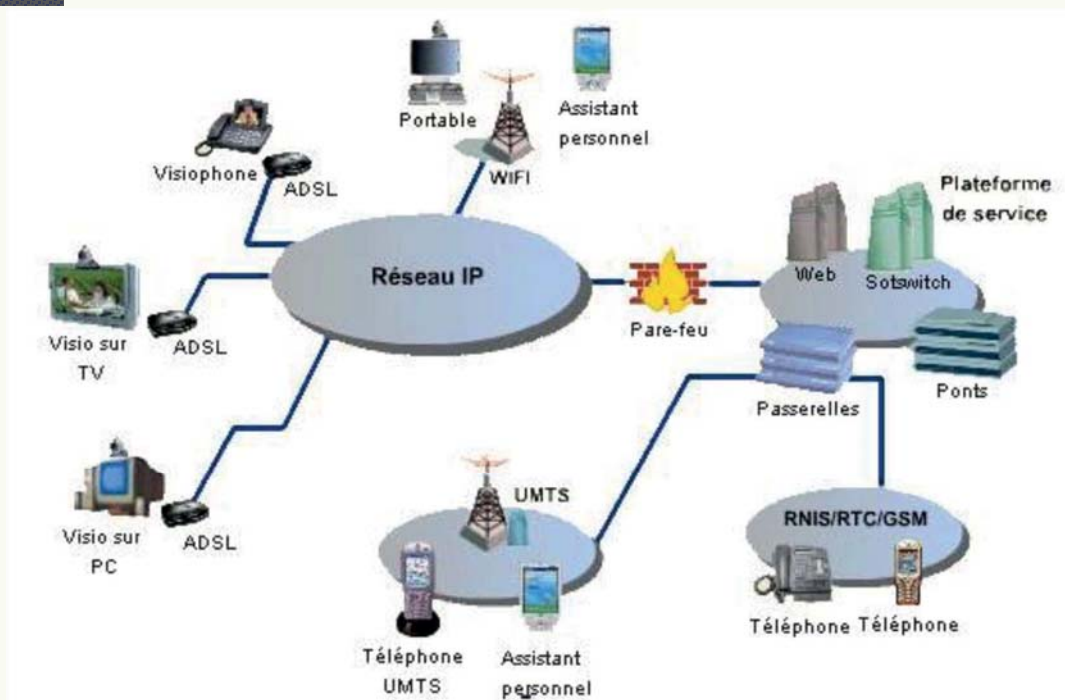
- Fenêtres longues (36 échantillons), bonne résolution fréquentielle, segments stationnaires
- Fenêtres courtes (12 échantillons), bonne résolution temporelle, segments transitoires
- Décision de changement de fenêtre:
 - Calcul de PE (entropie perceptuelle) pour fenêtres longues
 - Si $PE > \text{seuil}$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

Qualité maximum : CBR à 320kbps (Constant Bit Rate , débit binaire constant). Chaque échantillon est encodé à un bitrate constant.

A 320kbps, on peut considérer que c'est pour du stockage, mais avec perte. On peut aller jusqu'à 128kbps. - Qualité correcte pour l'écoute, pour une audiothèque par exemple : VBR V2 (Variable Bit Rate, débit binaire variable).

Le VBR a un débit qui varie, comme son nom l'indique. Par exemple, une zone de « silence » (ou avec très peu d'information) sera encodée à un débit très faible. Les zones plus complexes seront par contre encodées à un débit « maximal » décidé en fonction du profil choisi. Ici, « V2 » indique le type de profilé (de V0 à V9, V0 correspondant la meilleure qualité).

Conclusion



Conclusion

- La qualité du codeur
- Retard
- Débit
- Complexité

Classe n°	Délais par sens	Commentaires
1	0 à 150 ms	Acceptable pour la plupart des conversations.
2	150 à 300 ms	Acceptable pour les communications à faible interactivité.
3	300 à 700 ms	Deviens pratiquement inutilisable
4	Au-delà de 700 ms	Inutilisable sans une bonne pratique du half duplex

Codeur	PCM (Temporel)	MICDA (Temporel)	RPE-LTP (Analyse/Synthèse)	CELP (Analyse/Synthèse)	LPC (Paramétrique)
Standard/home	G.711	G.726	ETSO GSM 06-10	DOD FS1016	DOD LPC10 FS1015
Débit	64 Kbit/s	32 Kbits/s	13 Kbit/s	4.8 Kbit/s	2.4 Kbit/s
Qualité (MOS)	4.2	4.0	3.6	3.5	2.3
Retard codeur + décodeur	125 ms	300 ms	50 ms	50 ms	50 ms
Complexité Mips	0.1	12.0	2.5	16	7.0