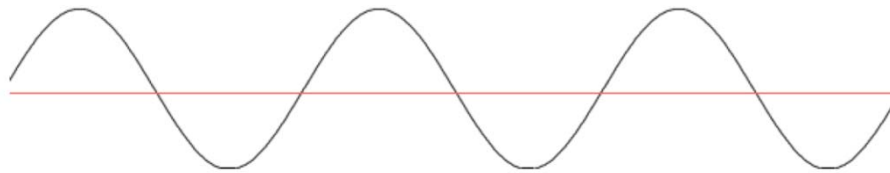


Le son

Définition et caractéristiques

- Le son résulte d'une perturbation ou vibration dans un milieu élastique. La déformation (compression et dilatation) est transmise de particule en particule jusqu'à arriver à l'oreille et cette sensation est interprétée par le cerveau comme du son. La science qui étudie les sons est l'acoustique.



- La vibration produit une onde sinusoïde (ou une combinaison de sinusoïdes) qui possède les caractéristiques suivantes :

Définition et caractéristiques

- La fréquence (f)

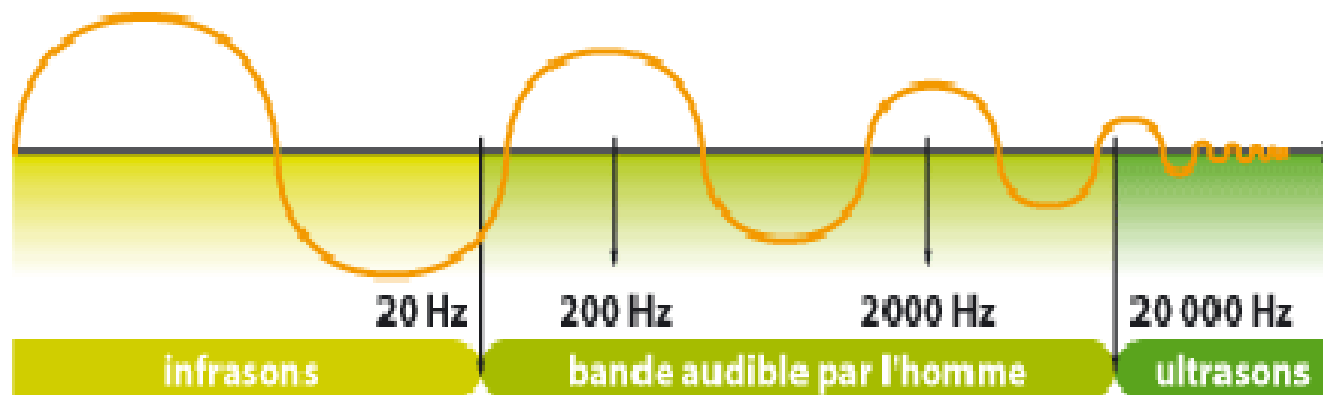
Un cycle étant une oscillation, la fréquence représente le nombre de cycles de l'onde par seconde et se mesure en Hz (1cycle/seconde).

Ce qui vibre rapidement (grande fréquence) est perçu comme un son à haute tonalité (aigu), et ce qui vibre lentement (petite fréquence) produit un son basse tonalité (grave).

Définition et caractéristiques

- La fréquence (f)

L'oreille humaine peut distinguer le son entre 20 et 20000 Hz.



Définition et caractéristiques

- La Période (T)

C'est le temps nécessaire pour effectuer un cycle $f = \frac{1}{T}$

Définition et caractéristiques

- La célérité (C)

C'est la vitesse de propagation du son, qui dépend de la nature, de la température et de la pression du milieu. Les ondes sonores se déplacent à environ 344 mètres par seconde dans de l'air. Dans des milieux solides le son peut se propager encore plus rapidement. Ainsi dans l'eau, sa vitesse est de 1482 m/s et dans l'acier de 5050 m/s.

Définition et caractéristiques

- La longueur d'onde (λ)

C'est l'étendu de l'onde ou la distance entre deux pics (ou deux creux).
Plus la fréquence est élevée, plus la longueur d'onde est courte

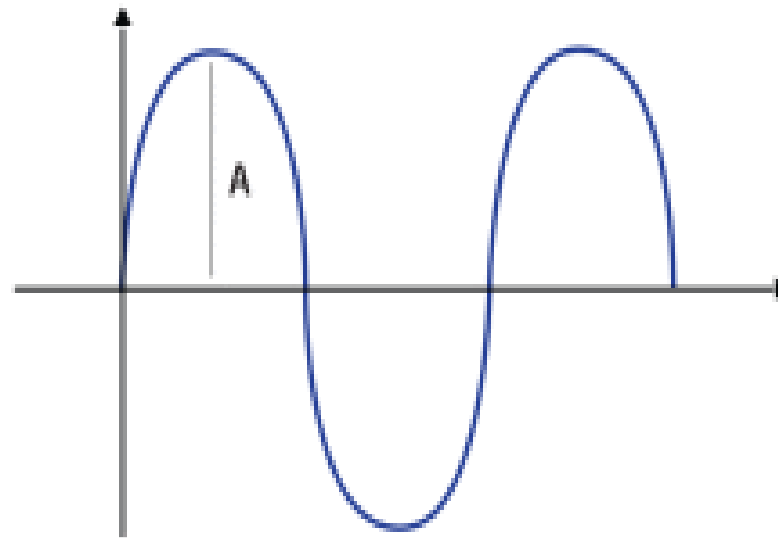
$$\lambda = CT = \frac{C}{f}$$

Cette relation lie la célérité à la longueur d'onde et à la fréquence de la vibration acoustique.

Définition et caractéristiques

- L'intensité du son (A) :

C'est l'amplitude de l'onde ou sa hauteur. L'intensité se mesure en décibels (dB).



Définition et caractéristiques

- Le timbre du son :

Un son est en général un mélange d'ondes appelées "harmoniques" qui ont des fréquences différentes mais toutes sont des multiples de la fréquence de base.

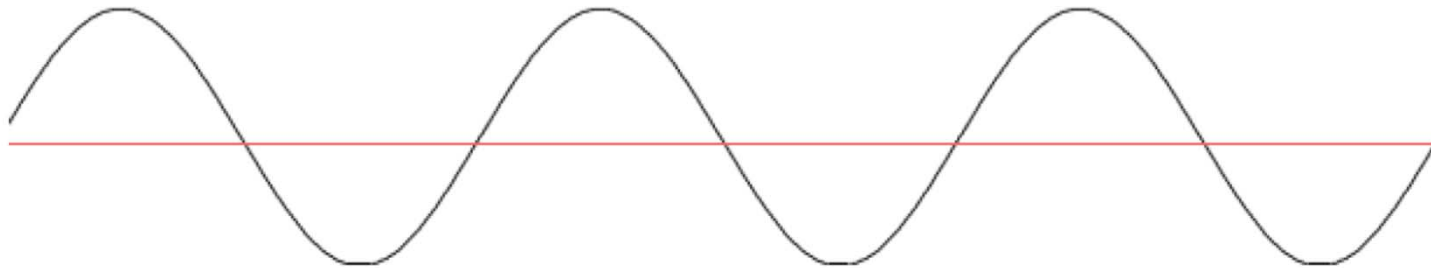
Le timbre ou la couleur est donné par le nombre et l'intensité des harmoniques qui accompagne le son fondamental.

Lorsque le son contient une seule fréquence, ce qui est rare, il est appelé son pur ou simple.

Les formes d'ondes

- Sinusoïde :

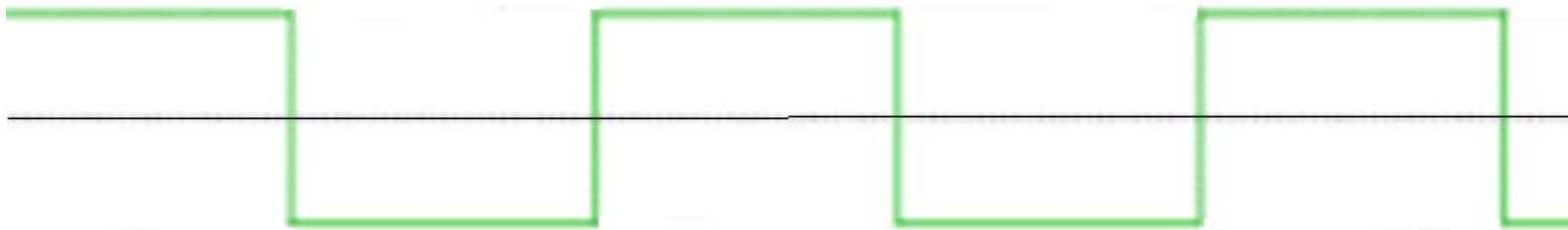
C'est la Sinusoïde pure ou simple qui est à la base de tout son.



Les formes d'ondes

- Carrée :

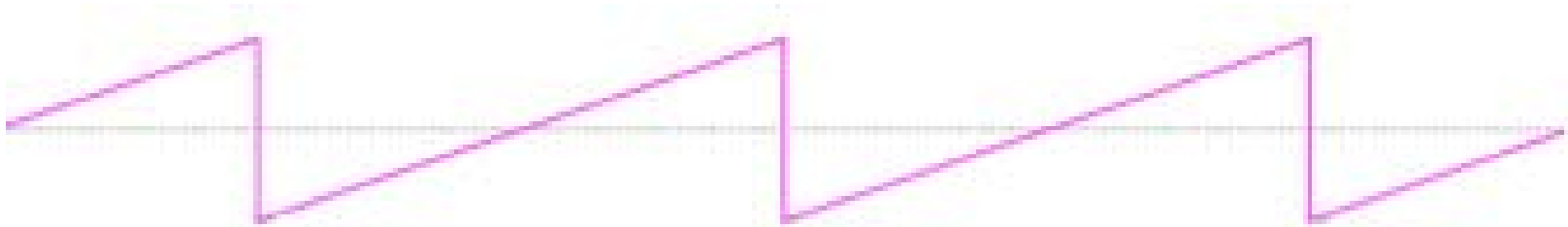
C'est la somme de plusieurs harmoniques d'ordre impair. L'ampleur décroît de cette manière : l'harmonique avec la fréquence $f' = 3f$ a une amplitude de $1/3$ de la fondamentale, celle avec $5f$ a une amplitude de $1/5$ de la fondamentale, ainsi de suite.



Les formes d'ondes

- En dents de scie :

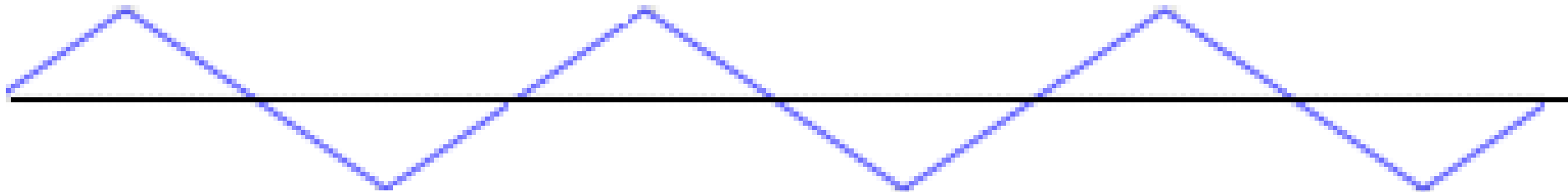
Elle est constituée de toutes les harmoniques (pairs et impairs). Son ampleur décroît de la manière que l'onde carrée.



Les formes d'ondes

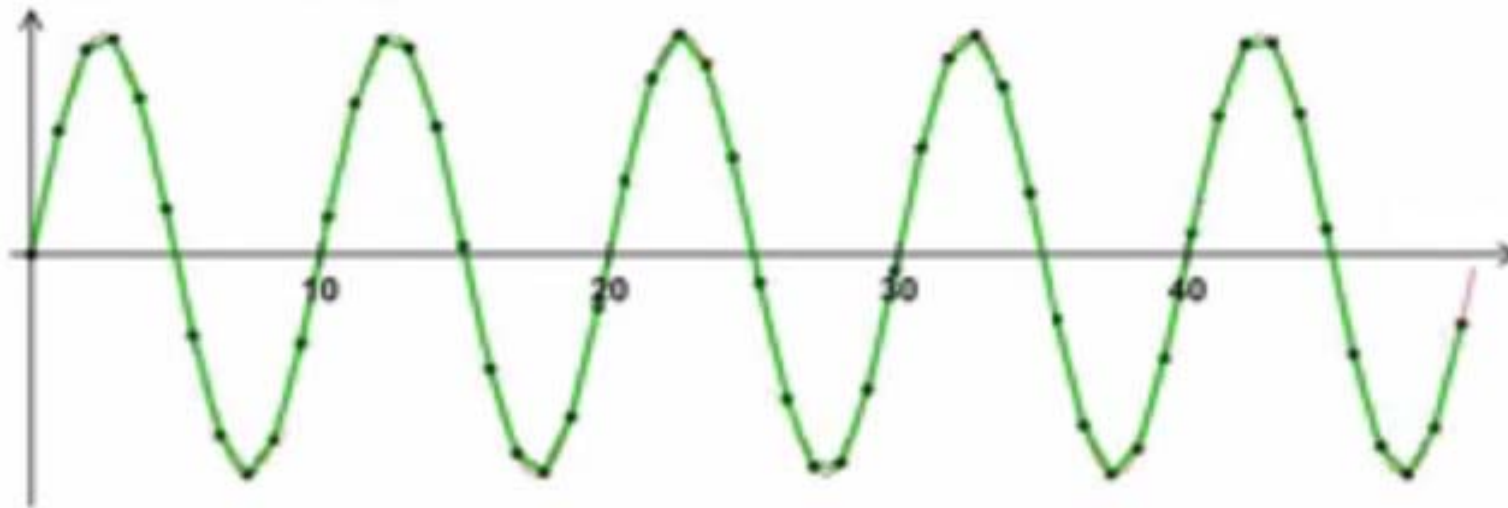
- Triangulaire :

Contient toutes les harmoniques impairs (comme l'onde carrée) par contre son amplitude décroît plus vite.



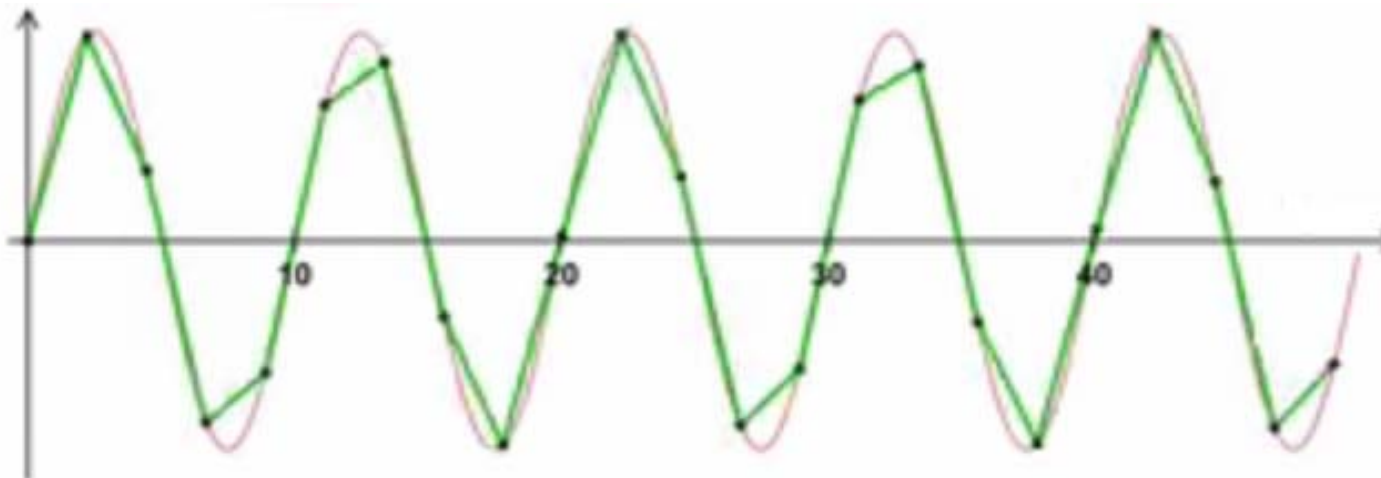
La numérisation du son: échantillonnage

Cette étape consiste à découper temporellement le signal analogique en «tranches» : c'est l'opération dite d'échantillonnage.



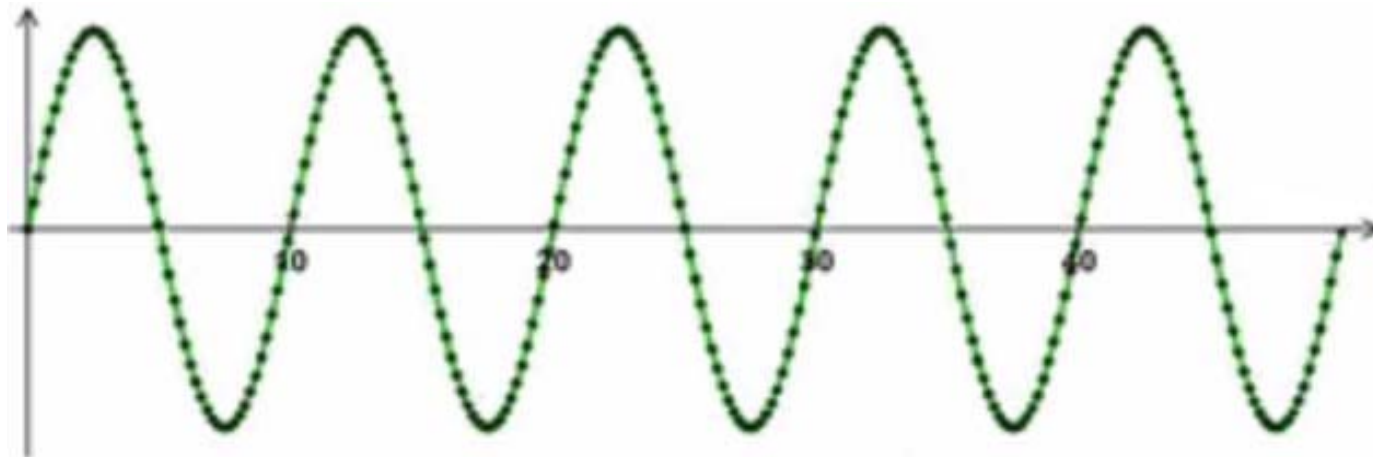
La numérisation du son: échantillonnage

La fréquence d'échantillonnage F_e est le nombre de fois par seconde où le son est échantillonné. Le choix de la fréquence d'échantillonnage est important. Elle s'exprime en hertz et doit être en rapport avec la fréquence maximum que l'on souhaite transmettre.



La numérisation du son: échantillonnage

La fréquence d'échantillonnage F_e est le nombre de fois par seconde où le son est échantillonné. Le choix de la fréquence d'échantillonnage est important. Elle s'exprime en hertz et doit être en rapport avec la fréquence maximum que l'on souhaite transmettre.



La numérisation du son: échantillonnage

Théorème de Nyquist-Shannon

Pour reconstruire un signal de sortie de manière fidèle au signal d'entrée, il faut choisir une fréquence d'échantillonnage au moins deux fois supérieure à la fréquence maximale contenue dans le signal d'entrée.

La numérisation du son: échantillonnage

- La fréquence d'échantillonnage F_e : nombre d'échantillons par seconde

$$F_e = \frac{1}{T_e}$$

- La période d'échantillonnage T_e : le temps d'un échantillon

$$T_e = \frac{\text{durée d'une séquence}}{\text{nombre d'échantillons}}$$

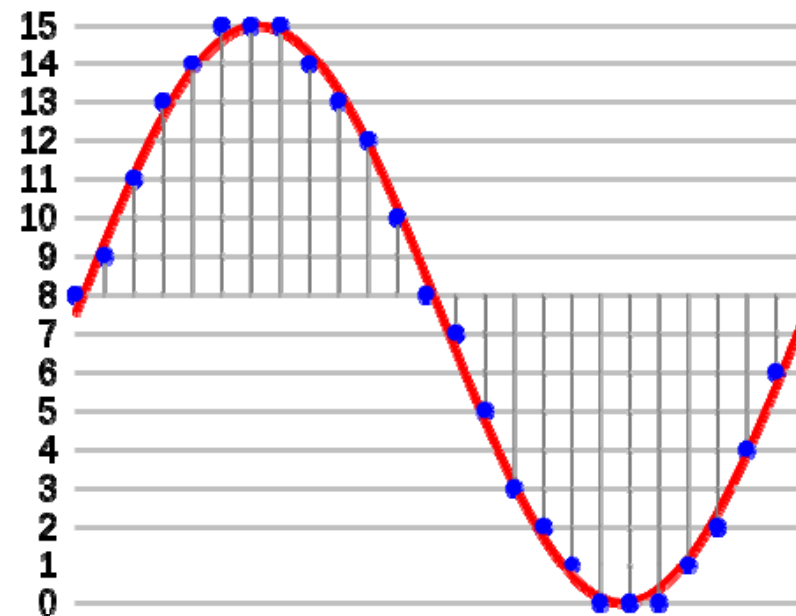
La numérisation du son: échantillonnage

Exemples

- Pour numériser un son à 3000Hz, il faut une fréquence d'échantillonnage d'au moins
- Avec une fréquence d'échantillonnage de 7MHz, il y aura une perte pour les sons de fréquence ...

La numérisation du son: quantification

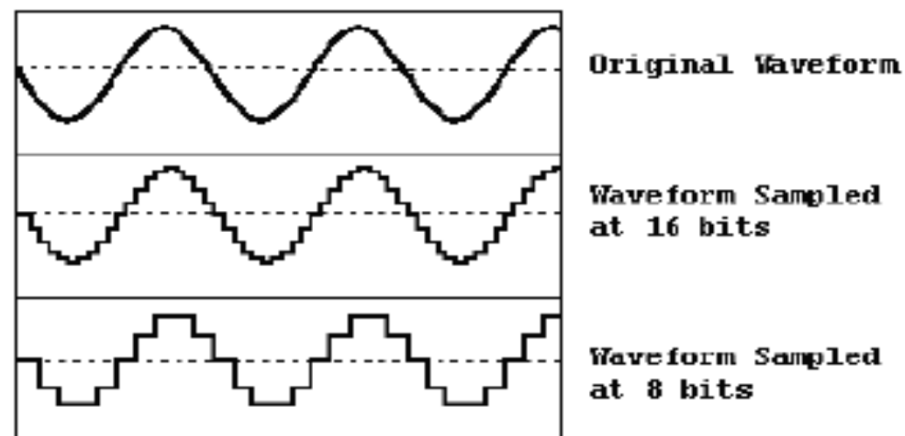
C'est la conversion de valeurs continues vers un nombre fini de valeurs possibles



La numérisation du son

- La qualité de la numérisation et le débit final dépendent de la fréquence d'échantillonnage et de la profondeur de quantification:

$$D = Fe * N$$



La numérisation du son

Le CBR et le VBR :

- En général l'audio est compressé à Bit-rate Constant (Constant Bitrate, CBR), ce qui signifie que chaque seconde aura le même nombre de bits qui lui sera alloué. Cependant, il est évident qu'un signal audio est tout sauf constant. Il y a des passages calmes, des passages à volume élevé, des passages compliqués et des passages simples.
- Un système d'encodage à bit-rate variable (Variable Bit-rate - VBR), dans lequel le bit-rate pour chaque frame est adapté suivant le principe que certaines sections nécessitent moins de bits, et que d'autres en nécessitent plus. Ça signifie que pour la même taille de fichier, une meilleure qualité peut être encodée.

La numérisation du son

- La taille des fichiers audio

Les CD audio : 44,1 KHz à 16 bits (2 octets) en stéréo (2 pistes) → environ 10Mo pour une minute au format CDA (CD Audio).

La numérisation du son

- Ce système de numérisation de l'audio est appelé PCM (Pulse Code Modulation) ou en français Modulation par Impulsions Codées. C'est la première phase de la numérisation d'un signal, elle ne contient pas de compression.

La compression du son

Il y a deux étapes essentielles à la compression audio (codec).

- La compression de données sans perte. C'est le type de compression "zip" (LZW), qui recherche des séquences afin de réduire la quantité de données qui doivent être stockées.
- 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.

La compression du son

- Un encodeur va jeter de l'information pour réduire la taille. C'est fondé sur un modèle mathématique qui tente de décrire ce que l'oreille humaine entend réellement - i.e. avec l'objectif de se débarrasser de l'information qu'on ne peut pas entendre. L'information exacte à éliminer dépend du codec utilisé.
- Certains codecs sont faits pour enlever certaines fréquences pour que la compression soit meilleure pour les voix. Divers modèles ont été formulés au fil des années afin de réduire la taille des fichiers audio. Mais le plus marquant ces dernières années est sans aucun doute le modèle psycho acoustique utilisé dans la compression mpeg1 layer 3 (mp3).

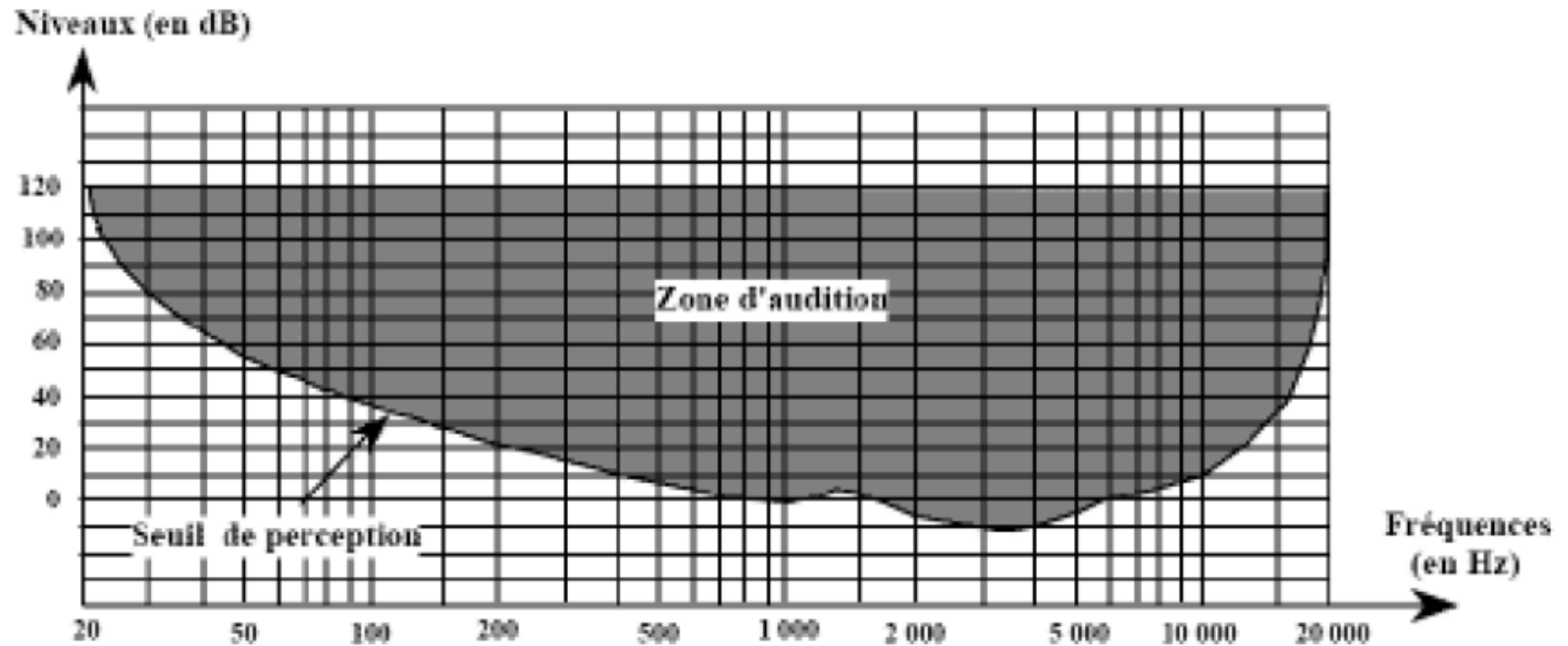
La compression du son: le MP3

- C'est une technique de compression de l'audio numérique. Son brevet a été déposé par l'institut de recherche allemand Fraunhofer en 1996. La variante audio de la 3^{ème} version de MPEG1, MPEG-1 Audio Layer 3 a été standardisée par l'ISO en 1992.
- Est basé sur un modèle psycho-acoustique qui détermine ce qui est audible de ce qu'il ne l'est pas.

La compression du son: le MP3

- Le signal audio est divisé en sous bandes , chaque sous bande est comparée à des tables de données propres au codec qui contiennent des informations sur les modèles psycho-acoustiques qui utilisent le concept de masque auditif « Masking » pour déterminer ce qui peut ou ne peut pas être entendu.
- Toute information qui correspond au modèle psycho-acoustique est conservée et le reste est rejeté. Ce sont les bases de la compression audio.
- En fonction du bit-rate (nombre de bits par seconde), le codec utilise la taille allouée pour stocker ces données. Le rapport du masque à bruit est utilisé pour assigner un nombre de bits donné à chaque bande.
- Les valeurs obtenues sont codées en utilisant la compression de Huffman qui réduit encore la taille.

La compression du son: le seuillage



La compression du son: le masquage

Phénomène psycho-acoustique qui fait qu'un son peut masquer un autre. Le son masqué est inaudible, donc inutile de le conserver.

- Masquage fréquentiel
- Masquage temporel

La compression du son: le masquage

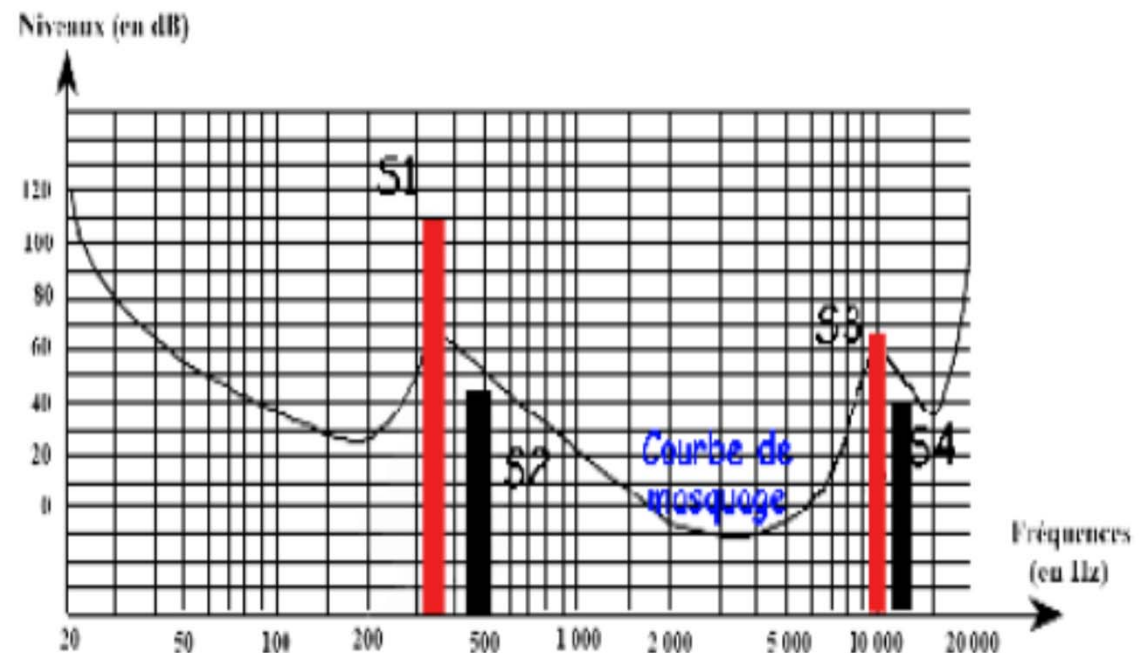
- Le masquage fréquentiel

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

La compression du son: masquage fréquentiel

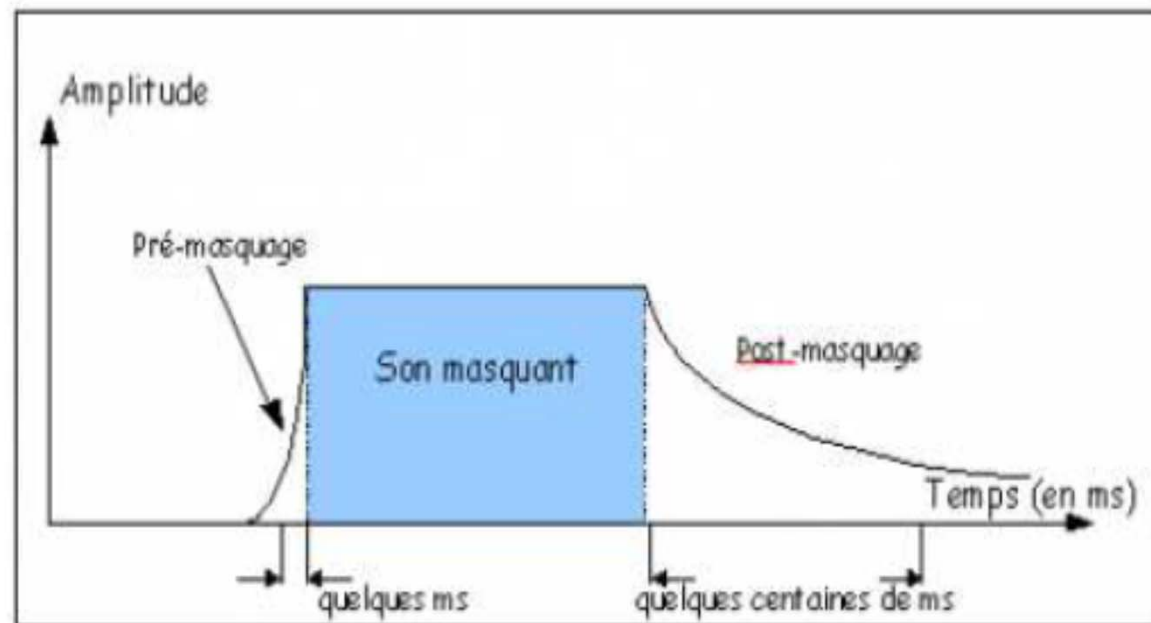
Cette courbe de « masquage » varie à chaque instant en fonction du contenu spectral du signal.

Dans l'exemple suivant, les fréquences S2 et S4 ne sont pas perceptibles, donc, inutile de les coder.



La compression du son: masquage temporel

Un son qui suit immédiatement un son de plus forte intensité n'est pas perçu par l'oreille. Ce type de masquage est moins exploité que le premier.



Quelques formats audio

- WAV et AIFF

Sur un plan purement informatique, les deux principaux formats de fichiers audio non compressés sont le format WAV développé par Windows et dont l'extension est justement « .wav » et le format AIFF (Audio Interchange File format) son équivalent chez Apple.

Quelques formats audio

- MPEG 1 Audio :

Le signal est filtré sur le modèle psycho-acoustique suivant :

- Seuillage : lorsque l'amplitude est inférieure au seuil d'écoute la fréquence est supprimée.
- Masquage : toute composante masquée par une autre est supprimée.

Quelques formats audio

- MPEG 1 Audio :

Il existe plusieurs niveaux:

- Le niveau 1 : la quantification est uniforme et le masquage est fréquentiel.
- Le niveau 2 : quantification uniforme et masquage fréquentiel et temporel.
- Le niveau 3 : quantification adaptative en fonction du rapport du signal à masquer, le masquage est fréquentiel et temporel et le codage de Huffman est ajouté.