

Chapitre 3. Audio

3.1 Les bases de l'audio

3.2 Le son numérique

3.3 Les erreurs de quantification.

3.4 Compression audio

3.4.1 Les étapes de la compression MP3

3.4.2 Psycho-acoustique et sons masqués

3.5 Quelques formats

3.5.1 MPEG-1/2 Audio Layer 3

3.5.2 Le format mp3PRO

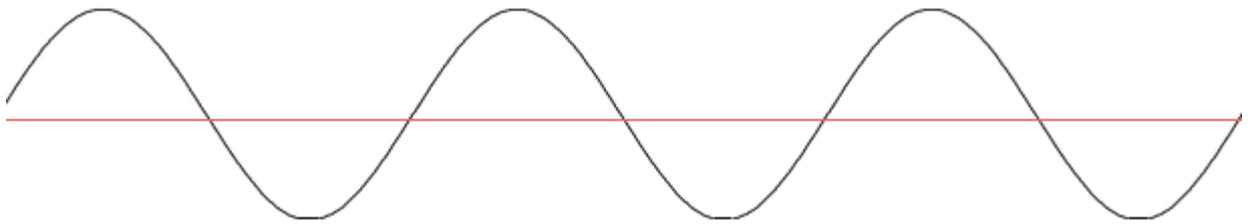
3.5.3 Le format audio .WAV

3.1 Les bases de l'audio

Si quelque chose vibre, ça fait vibrer l'air, l'air fait vibrer les os sensibles de l'oreille et cette sensation est interprétée par le cerveau comme du son.

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

Les grandes vibrations produisent un son fort, et ainsi de suite. Ces vibrations peuvent être représentées graphiquement par un signal:



Dans ce diagramme, l'amplitude des vibrations est mesurée sur l'axe vertical (y) et l'instant où sont prises les mesures est représenté sur l'axe horizontal (x).

Une note à tonalité plus élevée aurait plus de cycles sur une même durée plus de pics et de creux. Un son plus faible aurait des pics et des creux d'amplitude plus petite.

3.2 La numérisation du son : échantillonnage et quantification

La numérisation est une opération qui consiste à transformer un signal analogique en un signal numérique. Elle se déroule en deux temps :

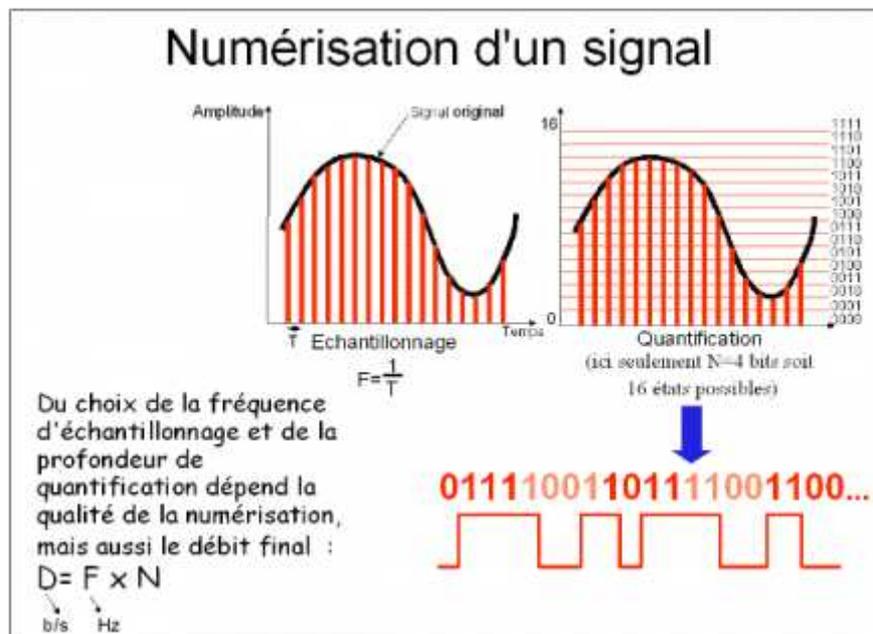
- Un découpage temporel et régulier du signal analogique (échantillonnage) : La première étape consiste à découper temporellement le signal analogique en « fines tranches » : c'est l'opération dite d'échantillonnage.

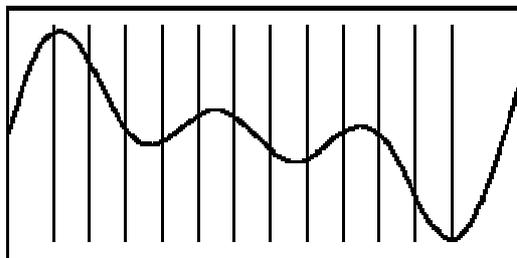
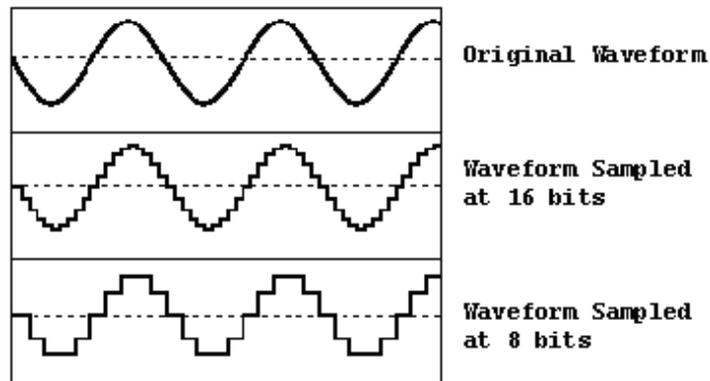
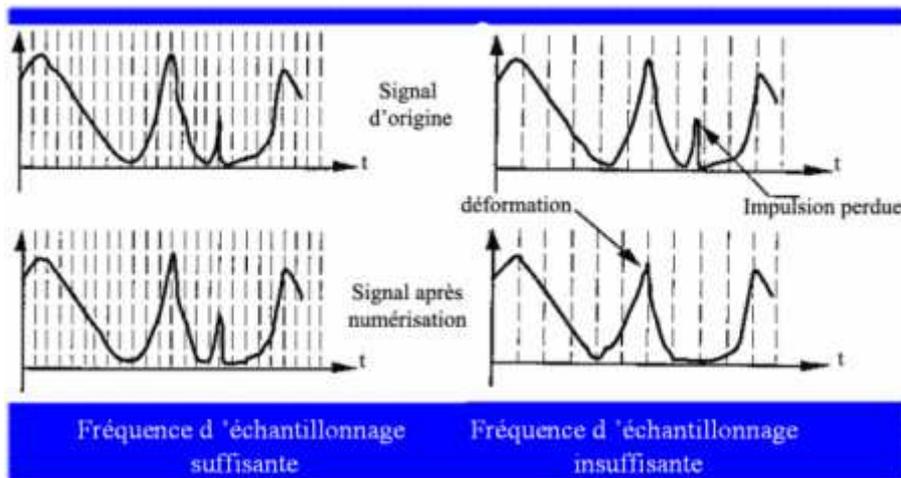
La fréquence d'échantillonnage est très simple : c'est le nombre de fois par seconde où le son est échantillonné, c'est-à-dire qu'on lui assigne une valeur numérique. Les CD sont échantillonnés à 44,1 kHz, soit 44100 fois par seconde.

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 et reproduire.

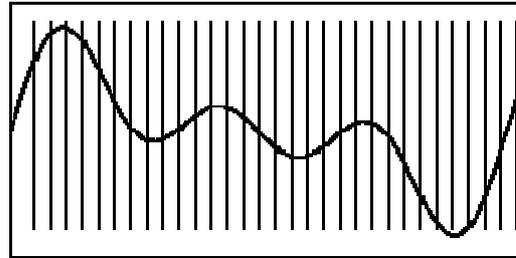
Pour pouvoir retransmettre intégralement l'étendue du spectre audible (donc jusqu'à 20 000 Hz), il faudra une fréquence d'échantillonnage au moins égale à 40 000 Hz. Si on souhaite une bande passante moins étendue (pour la restitution de la parole par exemple), on pourra opter pour des fréquences d'échantillonnage moins élevées, ce qui, corollaire immédiat, impliquera des fichiers moins volumineux (et donc des débits moins élevés).

- Une évaluation de l'amplitude de chacun de ces échantillons par rapport à des références selon une échelle précisément établie (la quantification).

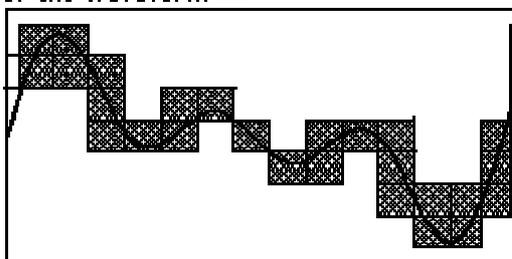




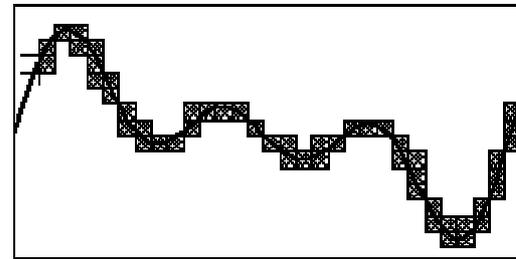
lower sample rates take fewer snapshots of the waveform.....



faster sample rates take more snapshots....



resulting in a rough recreation of the waveform.



resulting in a smoother and more detailed recreation of the waveform.

Interprétation : "Audio 16 bits 44,1 kHz" - signifie 65536 niveaux possibles mesurés 44100 fois par seconde.

Ce système de numérisation de l'audio est appelé :

Pulse Code Modulation (PCM)

MIC - Modulation par impulsions codées

3.3 Les erreurs de quantification

Affecter une valeur numérique à un signal analogique peut poser des problèmes. Etant donné qu'il n'y a qu'un certain nombre de valeurs (niveaux de quantification) que l'on peut assigner aux amplitudes, il y a un nombre considérable d'erreurs d'arrondi.

Evidemment, choisir un échantillonnage à 16 bits plutôt que 8 bits atténuera ces erreurs d'arrondi, puisqu'il y a plus de niveaux, comme il a été expliqué précédemment.

Normalisation

Dans le cadre de l'audio, il s'agit d'un processus à deux étapes qui vise à maximiser le volume de l'extrait sonore. Le processus de normalisation fait ce qui suit:

- 1) Trouve la plus haute amplitude dans le signal audio
- 2) Amplifie tout le signal de manière à ce que ladite amplitude ne dépasse pas le maximum qui puisse être stocké en numérique.

3.4 Compression audio

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

1) La compression de données 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.

2) 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.

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).

3.4.1 Les étapes de la compression MP3

Le signal est découpé en petites sections appelées frames

Ces chiffres sont ensuite comparés à des tables de données propres au Codec, qui contiennent des informations sur les modèles psycho-acoustiques.

Dans le codec mp3, ces modèles sont très avancés, et une grosse partie de la modélisation se fonde sur un principe du nom de Masking.

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.

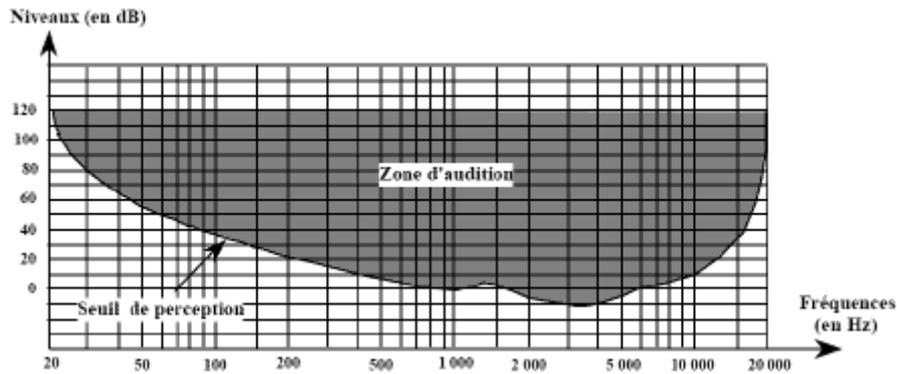
Ceci étant fait, le résultat passe par la compression sans pertes Huffman, qui réduit encore la taille de 10%.

La technique principale du codec mp3 pour enlever de l'information est en détectant quels sons ne sont pas détectables, ou 'masqués', et qui du coup ne peuvent pas être entendus.

Ces sons sont ensuite enlevés sans perte audible dans le signal.

3.4.2 Psycho-acoustique et sons masqués:

L'oreille humaine peut percevoir en théorie toutes les fréquences comprise entre 20 et 20 000 Hz



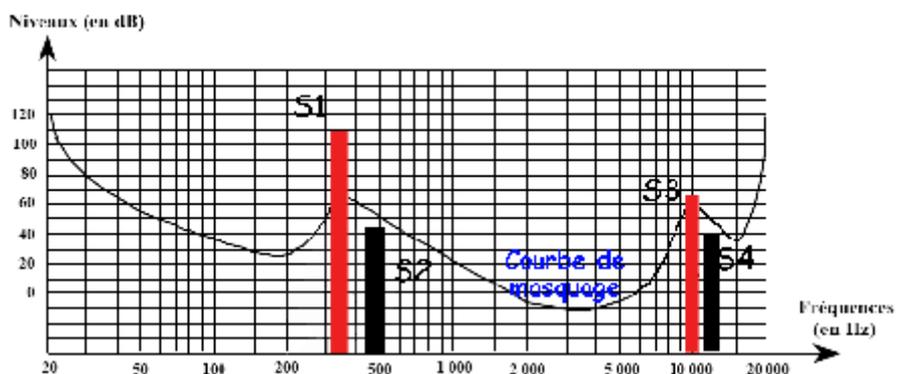
Seuil d'audition de l'oreille humaine

Il y a deux principaux types d'effets de masquage - le masquage simultané et le masquage temporel.

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.



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 masquage temporel

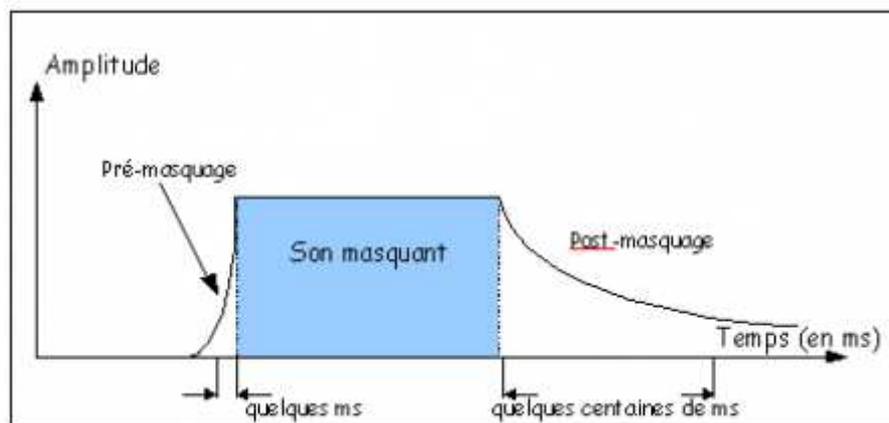
Le terme de masquage temporel fait référence au masquage réalisé après l'apparition d'un son masquant de forte intensité, mais également, au masquage avant la perception de ce son.

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.



Le masquage temporel

3.5. Quelques formats

3.5.1 MPEG-1/2 Audio Layer 3

Plus connu sous son abréviation de **MP3**, est la spécification sonore du standard MPEG-1, du Moving Picture Experts Group (MPEG). C'est un algorithme de compression audio (voir aussi codec) capable de réduire drastiquement la quantité de données nécessaire pour restituer de l'audio, mais qui, pour l'auditeur, ressemble à une reproduction du son original non compressé, c'est-à-dire avec perte de qualité sonore significative mais acceptable pour l'oreille humaine.

Ce format populaire de compression audio permet une compression approximative de 1:4 à 1:12.

Un fichier audio occupe ainsi quatre à douze fois moins d'espace une fois compressé au format MP3.

La réduction de taille facilite le téléchargement et le stockage de données musicales sur un support numérique, tel qu'un disque dur ou une mémoire flash.

Taux de compression

L'efficacité de la compression des codeurs avec perte est habituellement définie par le débit binaire, puisque le taux de compression dépend de la taille de l'échantillon et de la fréquence d'échantillonnage du signal d'origine. Toutefois, les paramètres du disque compact sont souvent utilisés comme référence (44,1 kHz, 2×16 bits). Le taux de compression pour cette référence est plus élevé, ce qui montre la complexité de la définition du terme « taux de compression » pour les codeurs avec perte.

Les taux de compression et les débits de données pour le MPEG-1 *Layer I*, II et III :

- *Layer I* : 384 kbit/s, compression 4:1
- *Layer II* : 160..256 kbit/s, compression 6:1..10:1
- *Layer III* : 112..128 kbit/s, compression 12:1..14:1

3.5.2 Le format mp3PRO

CBR et 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, mais en définitive, si vous encodez à bit-rate constant, ils doivent tous être décrits avec le même nombre de bits.

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.

Présentation

La compression mp3Pro se présente comme un successeur avantageux au mp3. En effet, les fichiers générés par un encodeur mp3PRO sont compatibles avec tous les lecteurs mp3. Cependant seuls les lecteurs certifiés mp3PRO bénéficient des avantages supplémentaires de cet encodage.

L'encodage mp3Pro présente trois particularités qui en font un système de compression offrant de meilleurs résultats sonores que ceux d'un simple mp3 standard :

- L'utilisation du VBR (pour *Variable Bit Rating*) qui adapte en temps réel le débit en fonction de la densité et de la complexité sonore ;
- Une qualité sonore équivalente au 320 kbit/s du layer III en VBR (qui varie entre 64 kbit/s et 128 kbit/s) si bien qu'à qualité égale un fichier mp3Pro permet un gain de place substantiel par rapport à un fichier mp3 standard (entre 25% et 50% de gain).

3.5.3 Les formats WAV et AIFF

Sur un plan purement informatique, les deux principaux formats de fichiers audios 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.