Technique de compression MP3

Céline CATTOËN
Nicolas PRESTAT

10 janvier 2003

1 Introduction

1.1 Historique :

Le MP3 (MPeg Audio Layer 3) est un format de fichier son compressé obtenu par suppression de données. La norme MPEG a été élaborée à la suite de la norme JPEG par un groupe d'experts (Movie Picture Expert Group) réunis par l'ISO (International Standards Organisation) et l'IEC (International Electrotechnical Commission). MP3 fait appel a des algorithmes mis au point par l'institut Fraunhofer à Erlangen mettant à profil les spécificités de l'ouïe humaine. Le codages Mpeg Layer-3 permet de diminuer d'environ 12 fois la taille d'un fichier audio habituel. Le principal intérêt de ce format est d'atteindre un taux de compression très important sans perte de qualité sonore . Exemple : Il compresse tellement le son, qu'il est grâce à lui possible de faire des compilations de plus de 150 titres sur un seul CD (soit l'équivalent de plus ou moins 10 albums complets).

La qualité d'un morceau sonore est déterminé par son flux de données. Par exemple, le téléphone a un flux de 1 ko/s contre 4 Ko/s pour la radio. La qualité CD est très bonne puisque son flux est de 176 ko/s. La taille d'une chanson de 3 mn en format WAV sera donc de 32 Mo. C'est à cause du volume important des fichiers audio de bonne qualité qu'un logiciel de compression s'est révélé nécessaire.

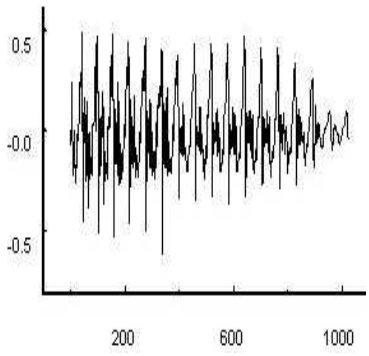
2 Principes généraux :

2.1 Transformée de Fourier et représentation fréquentielle :

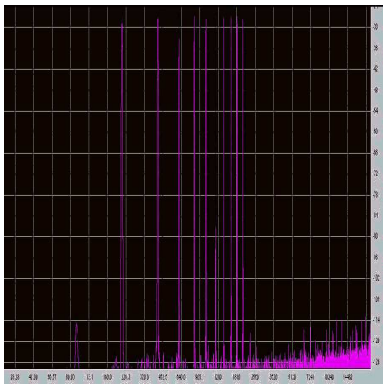
Pour comprendre le signal vocal et les signaux musicaux il est nécessaire d'en faire une analyse en fréquences.
La transformée de Fourier :

$$\hat{f}(\lambda) = \int_{-\infty}^{+\infty} f(t)e^{-2i\pi\lambda t}dt$$

est une formule très importante, car elle est utilisée dans tous les domaines de traitement des signaux et permet d'introduire la notion du spectre ou représentation en fréquence du signal. Le spectre révèle la distribution des fréquences dans un signal.



Représentation temporelle du signal


Représentation en fréquence

La transformée de Fourier permet d'échantillonner un signal à temps continu afin d'obtenir un signal discret. L'un des intêrets des filtres discrets est de pouvoir réaliser sous les hypothèses du théorème de Shannon [1] un filtrage équivalent à celui effectué sur des signaux à temps continu. Une application interressante des filtres discrets est le codage en

sous-bandes réalisé au travers de banc de filtres. Ce procédé permet de décomposer un signal discret en ses composantes basses fréquences et hautes fréquences, et a des applications dans la compression de données.

Or un son numérique est l'échantillonnage d'un son analogique à une certaine fréquence, le signal correspondant à un son est donc un signal discret.

La transmission et le stockage de sons haute fidélité se fonde sur une utilisation de la transformée de Fourier rapide qui permet d'accélerer les calculs de la transformée de Fourier discète.

On utilise notamment la Modified Discrete Cosine Transform dont la formule est la suivante :

$$X(m) = \sum_{k=0}^{n-1} f(k)x(k)cos(\frac{\pi}{2n}(2k+1+\frac{n}{2})(2m+1)), \text{pour } m = 0..\frac{n}{2}-1$$

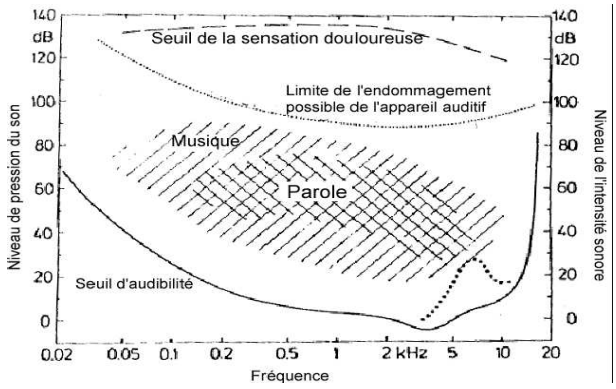
Le codage MP3 se fonde sur la décomposition en différentes bandes de fréquences (filtrage numérique et transformée de Fourier discrète) et sur la prise en compte de phénomènes psycho-acoustiques.

2.2 Propriétés psychoacoustiques :

Les caractéristiques de l'ouïe humaine d'une part et le traitement des informations acoustiques (ce que l'on appelle la psycho-acoustique) d'autre part jouent un rôle important dans le cas d'un processus de compression intelligent.

2.2.1 L'oreille humaine :

Le seuil d'audibilité est fonction de la fréquence correspondant à l'intensité sonore nécessaire pour la perception d'un stimulus. Cette fonction est caractéristique pour chaque personne et sa forme dépend de plusieurs facteurs. Sur le schéma ci-dessous on peut distinguer les différentes frontières de la perception des stimuli (son pur dont la durée dépasse 100ms) par l'appareil auditif.



Le format de compression MP3 utilise les "défauts" naturels de l'oreille afin de reduire la taille des sons stockés dans ce format.

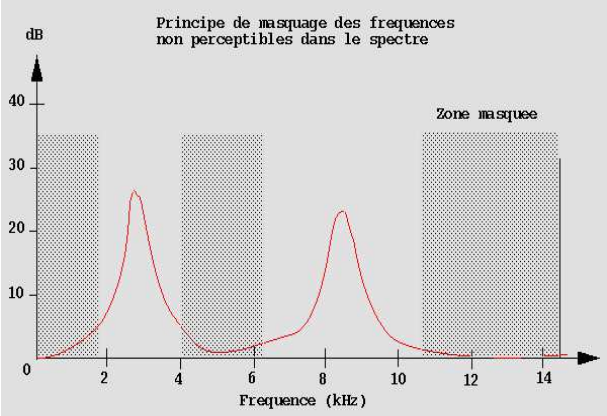
L'oreille humaine est capable de discerner des sons entre 0,20 kHz et 20 kHz, sachant que la sensibilité maximale est entre 2 et 4 kHz, et que la voix humaine se situe entre 0,5 et 2 kHz. Les sons graves ont une majorité de fréquences faibles et les sons aigus une majorité de fréquences élevées. Le Mpeg utilise le fait que des fréquences élevées peuvent masquer des fréquences basses. Le codage Mpeg supprime donc ces sons que l'oreille humaine ne perçoit pas, réduisant ainsi la taille d'un fichier sonore.

2.2.2 Sub Band Coding et masquage :

Le principe de Sub Band Coding (codage en sous-bande) qui dépend du principe de masquage permet de supprimer les sons que l'oreille ne perçoit pas.

L'idée est de couper le domaine de fréquence audible pour l'appareil auditif en petites bandes (comme l'oreille le fait). L'oreille humaine est sensible à un large spectre de fréquences. Mais pour être audibles, ces fréquences devront avoir une intensité sonore supérieure à leur valeur limite. Il est donc inutile de prendre en compte les sous-bandes pour lesquelles cette conditions n'est pas remplie. De plus, les fréquences proches peuvent se masquer entre elles. Quand une grande quantité d'énergie est présente sur une fréquence (pic), l'oreille ne peut pas distinguer de plus basses énergies présentes aux fréquences voisines. Les fortes énergies (appelées masqueurs) masquent les petites. L'idée est de sauvegarder la bande passante en supprimant les informations des fréquences masquées, pour cela on peut par exemple utiliser des filtres passe-bandes. Le résultat ne sera pas le même que le signal original (du point de vue de la bande passante), mais l'oreille ne verra pas la différence si le calcul est bien fait.

La figure suivante illustre le principe du masquage des fréquences non perceptibles :



3 Le fonctionnement de la compression MP3

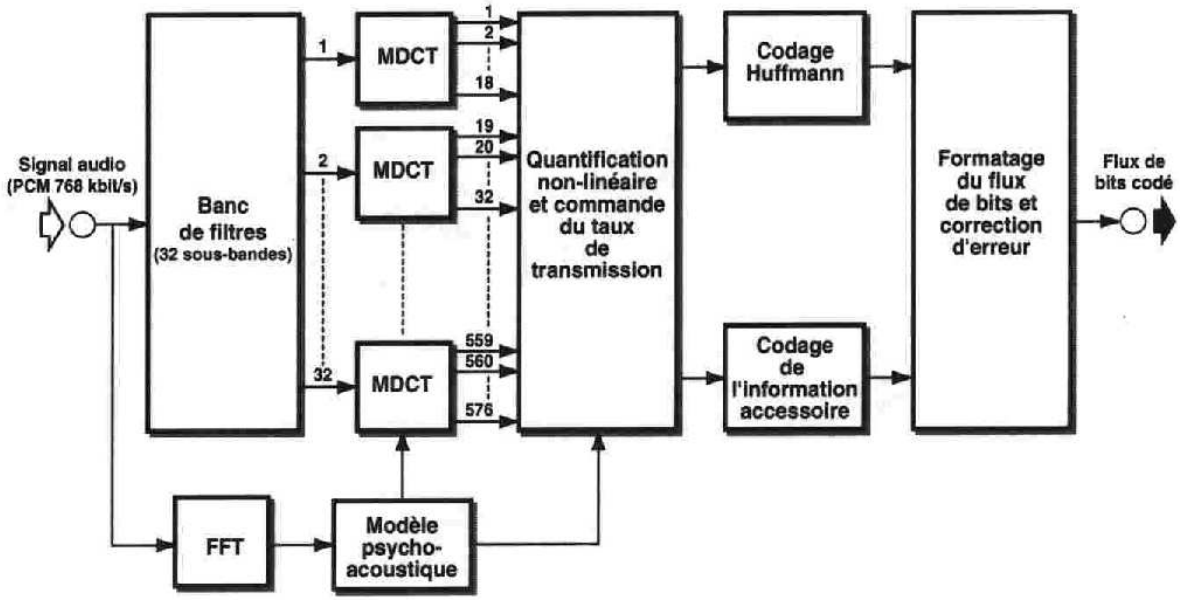
3.1 Présentation des différentes étapes

Avant d'entamer tout processus de compression, il est nécessaire de transformer le son en fichier informatique. On utilise pour cela un convertisseur analogique/numérique (CAN) qui échantillonne et enregistre les données ainsi obtenues. Comme on l'a vu, l'inconvénient de ce type de fichier (comme les fichiers WAV) est leur taille (environ 10 Mo pour une minute d'enregistrement).
A partir de ces données, on procède à la compression proprement dite :

Le filtrage : les données sont tout d'abord filtrées de manière à éliminer les fréquences inaudibles pour l'oreille humaine. Les données *temporelles* sont transformées en données *fréquentielles* à partir desquelles le travail de compression sera effectué.

Le codage : cette phase consiste à déterminer le nombre de bits alloués pour chaque donnée. La perte d'information que l'on s'autorise entre le signal original et le signal compressé est donnée par le modèle psychoacoustique.

Le processus de compression se déroule comme suit :



Nous allons à présent expliciter ces différentes étapes.

3.2 Filtrage

Passage par un banc de filtre *(Filter bank)* : on procède tout d'abord au traitement des données temporelles. On utilise un banc de filtres passe-bande pour découper le spectrogramme en 32 *sous-bandes équivalentes*. Cette transformation est effectuée à l'aide d'une *Modified Discrete Cosine Transform* (MDCT) (voir la définition plus haut). La fréquence maximale que l'on sélectionnera est la fréquence limite d'audition par l'oreille humaine, à savoir 22.05 MHz. Cette perte d'information sera indétectable lors de l'écoute puisque de toute manière inaudible par l'utilisateur. Le théorème de Shannon [1] nous dit que la fréquence d'échantillonnage maximum devra alors être de 44.1 MHz, soit 44 100 données par seconde.
Parallèlement, il faut définir les seuils de masquage. Le calcul est effectué là aussi à l'aide d'un spectrogramme, obtenu par une transformée de Fourier rapide. On dispose alors pour chaque sous-bande de l'erreur maximale que l'on s'autorise lors de la compression.

Elimination des redondances *(Joint stereo coding)* : Lorsque le signal est stéréo, certaines données sont les mêmes sur les deux canaux. Ces informations superflues peuvent être supprimées sans altérer la qualité de l'enregistrement.

3.3 Codage

Il s'agit de la partie la plus importante de la compression. En effet, c'est durant cette phase que l'on décide des informations que l'on peut occulter pour réduire la taille du fichier. Dans le cas de la compression MP3, les données ainsi supprimées doivent être imperceptibles lors de l'écoute du fichier compressé. On procède donc ainsi :

Réduction de l'échelle des données *(Scale factors)* : Afin de faciliter le processus de compression et de codage, on multiplie nos valeurs numériques par $3\frac{}{4.}$

On procède alors à la *quantification*. Ce processus détermine le nombre de bits que l'on alloue au codage de chaque sous-bande. Il s'agit de trouver le taux de compression

maximal que permet le modèle psychoacoustique. L'algorigramme est donc constitué de deux boucles imbriquées :

Codage *(Noiseless coding)* : la boucle intérieure permet le codage des données en binaire. Elle utilise le *codage de Huffman* qui attribue peu de données aux mots fréquents et beaucoup aux mots rares. Cette méthode est optimale, c'est à dire qu'elle donne en moyenne la plus petite longueur de code de toutes les techniques de codage, et sans perte. C'est pourquoi elle est utilisée pour tous les formats Layer III, par exemple MPEG et JPEG.

Quantification *(Quantizer)* : la quantification définit l'échelle à partir de laquelle sera effectuée le codage. Plus cette échelle est fine, plus les données seront codées avec précision mais avec peu de gain. A l'inverse, un pas de quantification plus grand donnera une meilleure compression au dépend de sa qualité. La boucle extérieure est une boucle de calcul et de contrôle. Elle vérifie que les erreurs dues à la compression restent en deçà des seuils de masquage pour chacune des sous-bandes. Si tel n'est pas le cas, le pas de quantification est modifié et le codage de Huffman est réeffectué de manière à augmenter la précision (au dépend du taux de compression). Cette boucle s'arrête lorsque le gain de compression est optimal pour chacune des sous-bandes sans que les seuils de masquage ne soient dépassés.

On obtient alors un fichier MP3 compressé. Ce dernier n'est bien entendu pas audible directement. Il faut tout d'abord le décompresser à l'aide d'un décodeur qui effectue l'algorithme en sens inverse.

Conclusion : avantages et inconvénients du MP3

Parmi les différents formats de compression audio, le MP3 est le plus répendu auprès du grand public grâce à ses nombreux avantages. Le très bon taux de compression permet d'obtenir des fichiers téléchargeables par un réseau ou sur Internet. Cette performance ne se fait pas au dépend de la qualité puisque le modèle psychoacoustique garantit que les modifications seront indétectables par l'utilisateur. Mais le MP3 reste un format de compression. Pour écouter les morceaux musicaux, il est nécessaire de décompresser les données auparavant.

Le futur de la compression audio est en phase de conception : le format MPEG-4 Audio. Ce format audio intégrera la plupart des techniques actuelles de codage audio et comportera des outils pour la modélisation de sons 3D à partir de sources naturelles ou artificielles. Il sera complètement multi-canaux (5 canaux séparés et donc la possibilité de supprimer un ou plusieurs instruments lors de l'écoute d'un concert et de jouer à la place de celui-ci). La qualité CD sera obtenue avec un bitrate de 64kb/s (le MP3 est à 128 kb/s) et ce taux pourra varier de 2kb/s à plus de 64kb/s. Ce taux et cette qualité seront obtenus grâce à un savant mélange des nouvelles techniques de codage suivant différents critères (fréquences, bitrate, ratio de compression, type de sons, etc...).

Références

[1] Théorème de SHANNON :
Soit f un signal dont la transformée de Fourier soit à support inclus dans $[-B, B]$, $B0$. Si $\frac{1}{T} \geq 2B$, alors

$$rect_{2B}(\lambda)\hat{f}_T(\lambda) = \hat{f}(\lambda)$$

Si $f \in L^2$, alors on a la formule de Shanon :

$$f(t) = \sum_{n \in Z} f(nT) sinc(t/T - n)$$

Si de plus $\sum_{n \in Z} |f(nT)| \infty$, l'inégalité a lieu pour tout $t \in R$.

[2] R. Vaillancourt, *Applied Analysis : Fourier analysis, data compression (MPEG,JPEG), wavelets*, 2000

[3] Sultan El Turrah, *Le MP3, le principe, les avantages, le fonctionnement*, SETh, 2001

[4] M. Aron, J.B. Murat, *Caractéristiques et défauts de la compression MP3*

[5] *MP3 : how it works*, Intel Corporation 2002

[6] Silvia Pfeiffer Thomas Vincent *Formalisation of MPEG-1 compressed domain audion features*, 2001