Félix LUBIN
Diplôme Professionnel Son 2ème Année
2003-2004

LA COMPRESSION AUDIO-NUMERIQUE

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

SOMMAIRE

1. INTRODUCTION A L'AUDIO-NUMERIQUE 4

1.1. CONVERSION ANALOGIQUE/NUMERIQUE, NUMERIQUE/ANALOGIQUE. 4
1.2. PRINCIPE GENERAL DE LA CONVERSION (A/N,N/A). 4

2. L'ECOUTE HUMAINE 6

3. LA METHODE DE HUFFMAN 9

3.1. NOTION ALGORITHMES 9
3.2. CODAGE HUFFMAN/ SHANNON-FANO 9
3.3. L'ARBRE DE HUFFMAN 10

4. QU'EST CE QUE LA COMPRESSION 13

4.1. LES DIFFERENTS TYPES DE COMPRESSION ET DE CODAGE 13
4.2. INFORMATIONS SUR LA COMPRESSION NON DESTRUCTIVE 15
4.3. LA COMPRESSION DESTRUCTIVE. 15

5. L'ATRAC 16

5.1. INTRODUCTION 16
5.2 L'ENCODEUR ATRAC (1ERE ETAPE) 18
5.3. L'ENCODEUR ATRAC (2EME PARTIE) 20
5.4. L'EVOLUTION DE L'ATRAC. 21

6. LE MP3 23

6.1. L'HISTOIRE 23
6.2. LE CONCEPT 23
6.3. L'ALGORITHME POUR LA COMPRESSION MP3 24
6.3.1. La batterie de filtres Hybride 24
6.3.2. Le modèle psychoacoustique 24
6.3.3. Attribution de bit 25
6.4. NOTIONS DE COUCHES 25
6.5. UTILISATION DU BITRATE EN MP3 27

7. LE MP3PRO 28

8. LE MPEG-2 AAC 30

9. LE WMA 31

10. L'OGG VORBIS 32

10.1. L'HISTOIRE 32
10.2. LE CONCEPT 32
10.3. SON UTILISATION FACE A LA CONCURRENCE 33

11. ETUDE COMPARATIVE 34

11.1. ANALYSE 35
11.2. CONDITIONS DE L'ETUDE 36

12. CONCLUSION 37

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

LA COMPRESSION AUDIO-NUMERIQUE

La compression audio vise à réduire la quantité de données pour stocker la musique. Deux formats de compression sont possibles : la compression sans pertes de données et la compression avec perte de données. C'est cette dernière qui nous intéresse puisqu'elle occupe un espace 10 fois moins grand que la première.

Un certain nombre de normes de compression avec pertes de données est disponible aujourd'hui. Les plus communs sont MP3, VQF, OGG VORBIS. Mais il existe d'autres formats tels que ATRAC, MP3Pro, AAC, WMA….

Les compressions sans pertes de données sont le Wave et l'AIFF.

Je ferai une étude plus ou moins élaborés de chacun de ces formats, et une étude comparative entre le MP3, l'OGG VORBIS, le WMA et le MP3Pro.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

1. INTRODUCTION A L'AUDIO-NUMERIQUE

De nos jours, de plus en plus de médias sont numérisés (image, son, vidéo…) et stockés ou transmis dans un format numérique. Cela a plusieurs avantages, dont le plus important est l'absence de perte de qualité du média numérisé pendant son transport ou lors de traitements sur celui-ci (à l'aide de logiciel, par exemple). Il existe également quelques inconvénients intrinsèques au caractère numérique, notamment la perte de précision lors d'une conversion analogique/numérique et inversement.
Pour que le média numérisé soit rigoureusement le même que le signal d'origine, il faudrait numériser ce dernier avec une précision infinie, donc en stockant une quantité infinie d'informations. C'est évidemment très théorique et impossible à réaliser. En pratique, la limite de précision nécessaire est fixé généralement par celle du récepteur final de ce média : l'œil pour l'image, l'oreille pour le son.

1.1. Conversion Analogique/ Numérique, Numérique/ Analogique.

La grande majorité de systèmes d'enregistrement numérique travaillent largement de la même manière. Un signal audio entrant est introduit dans ce qui est connu comme un convertisseur Analogique-Numérique(A/N). Ce convertisseur prend une série de mesures du signal à intervalles réguliers, et stocke chacun comme un nombre. La longue série résultante de nombres est alors placée sur un certain genre de support de stockage duquel elle peut être recherchée.

La lecture du signal est essentiellement le même processus à l'envers, une longue série de nombres est recherchée d'un support de stockage, et passée à ce qui est connu comme un convertisseur Numérique-Analogique(N/A). Ce dernier prend les nombres obtenus en mesurant le signal original, et les emploie pour construire une approximation très étroite avec ce signal, qui peut alors être passée à un haut-parleur et être entendue comme un son.

Le nom générique pour ce système est la modulation de code d'impulsion (PCM). Il est employé dans tous les échantillonneurs modernes, enregistreurs numériques et interfaces audio d'ordinateur.

1.2. Principe général de la conversion (A/ N,N/ A).

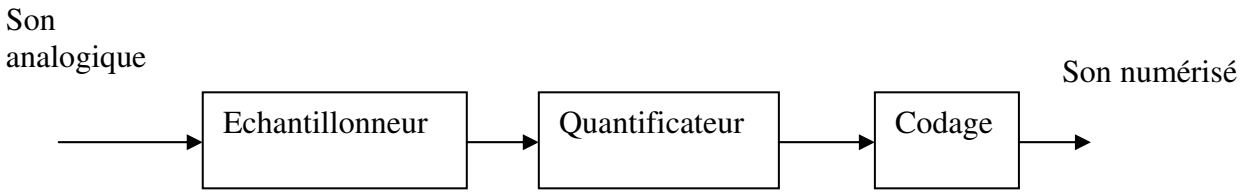
En prenant et en stockant une série de mesures très précises d'une forme d'onde, le PCM peut reconstruire une approximation très étroite du son qui correspond à cette forme d'onde. Le système d'enregistrement PCM utilise deux paramètres essentiels : la **fréquence d'échantillonnage** et la **quantification** .

La fréquence d'échantillonnage décrit le nombre de fois qu'un signal audio entrant est mesuré. On l'indique en kilohertz. Généralement on utilise une fréquence d'échantillonnage de 44.1 kHz.

La quantification concerne la précision de chaque mesure effectuée. Le convertisseur dans un système audio numérique de PCM mesure un signal, et stocke la mesure comme un nombre, ce nombre est représenté comme une série de 0 et de 1, également connue sous le nom de mot binaire exprimée en bit. Le codage sur 16 bits est très usité.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

On peut représenter schématiquement (en trois étapes principales) la chaîne qui permet de faire un enregistrement audio analogique-numérique.



Compte tenu des contraintes théoriques de restitution d'un signal analogique de bonne qualité (pour l'oreille humaine) à partir d'un signal échantillonné, on a imposé aux CD audio un format de stockage du son ayant les données suivantes :

- La fréquence d'échantillonnage (Fe) : 44.1 kHz,
- La quantification (Q) : 16 bits (2 octets),
- Son stéréo.

Avec ce format imposé on peut calculer la place qu'occupe une minute de musique.

La formule est la suivante : D= F* Q* P* T avec :

- D qui est la taille du fichier en kilo-octets,
- F qui est la fréquence d'échantillonnage,
- Q qui est le nombre d'octets utilisés pour coder le signal,
- P qui le nombre de piste. En stéréo on utilise deux pistes,
- T qui est le temps de l'enregistrement en seconde.

Numériquement on trouve :

D= 44.1* 2* 2* 60
D= 10 584* 10 exp. 6 soit environ 10 Mo

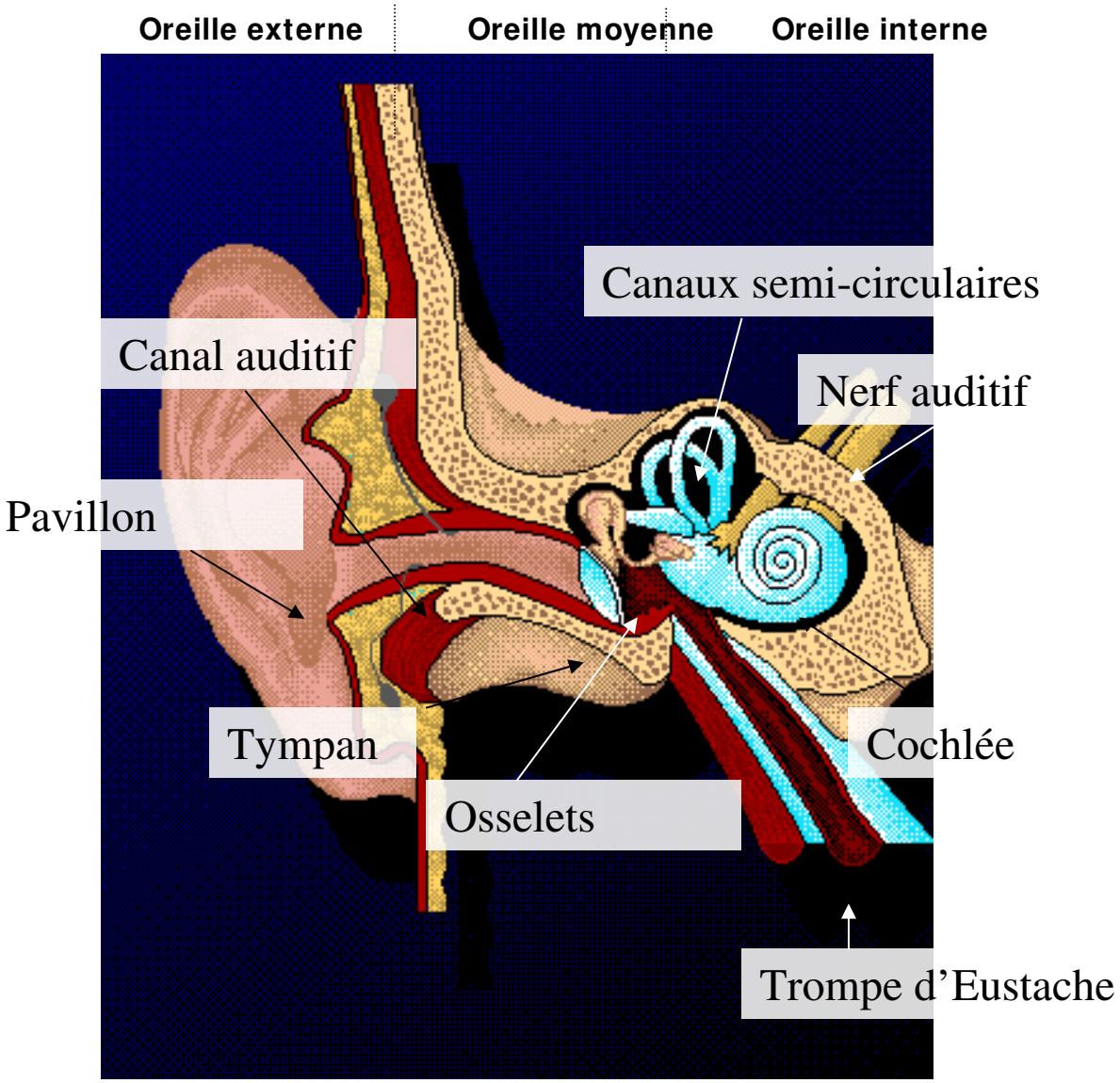
Il faut compter environ 40 Mo pour stocker 4 minutes de musique sur un CD, au format CDA (CD Audio).

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

2.L'ECOUTE HUMAINE

L'oreille est un organe extrêmement complexe, ultrasensible et est capable de s'adapter à une large plage d'intensités.

L'oreille est constituée en trois parties : L'oreille externe, moyenne et interne.



L'oreille externe comprend le pavillon et la première partie du conduit auditif. La forme du pavillon influe sur la perception acoustique, et notamment sur la localisation des sources sonores. Le conduit auditif sert à concentrer le son ambiant, tout en accentuant la plage de fréquences autour de 2500 Hz.

L'oreille moyenne joue le rôle d'amplification. Cette partie fonctionne comme un microphone, l'énergie acoustique extérieure est transformée en énergie mécanique qui fait vibrer

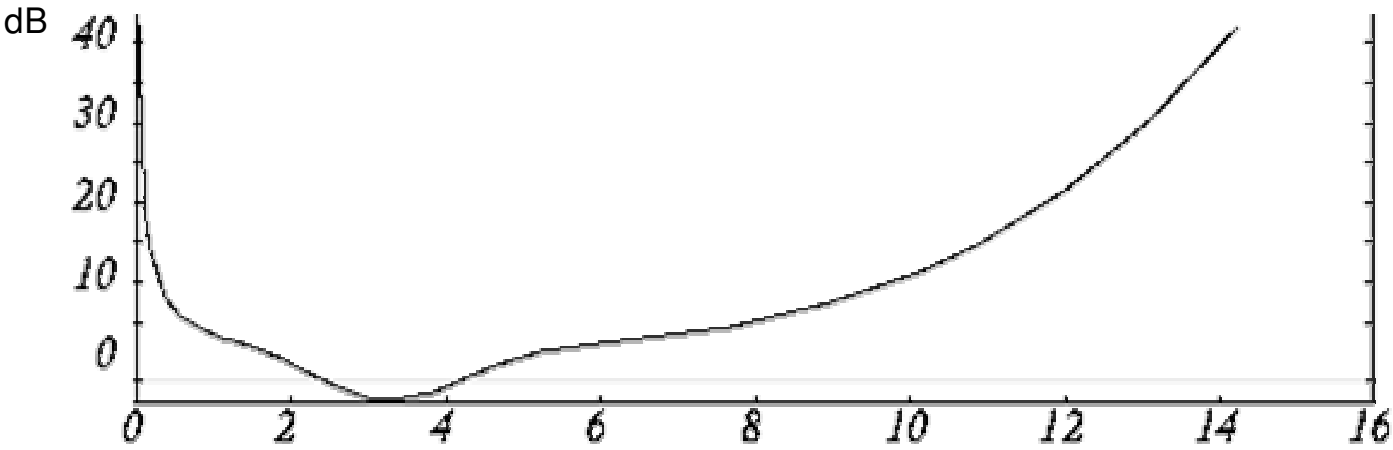
Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

le tympan. Cette vibration est transmise par la chaîne d'osselets (le marteau, l'enclume et l'étrier). La trompe d'Eustache assure l'équilibre entre l'oreille moyenne et l'arrière-gorge.

Les différents organes de l'oreille interne servent à transformer l'énergie mécanique en une énergie électrique directement analysable par le cerveau.

L'oreille est très sensible aux alentours de 2 kHz à 5kHz. Pour les fréquences alentours, la courbe des niveaux nécessaires à rendre une fréquence audible croit à vitesse variable en fonction de l'individu, de son âge etc.…

La courbe ci-dessous représente la courbe moyenne du seuil d'audibilité des fréquences en fonction de leur niveau et ce dans une chambre silencieuse.

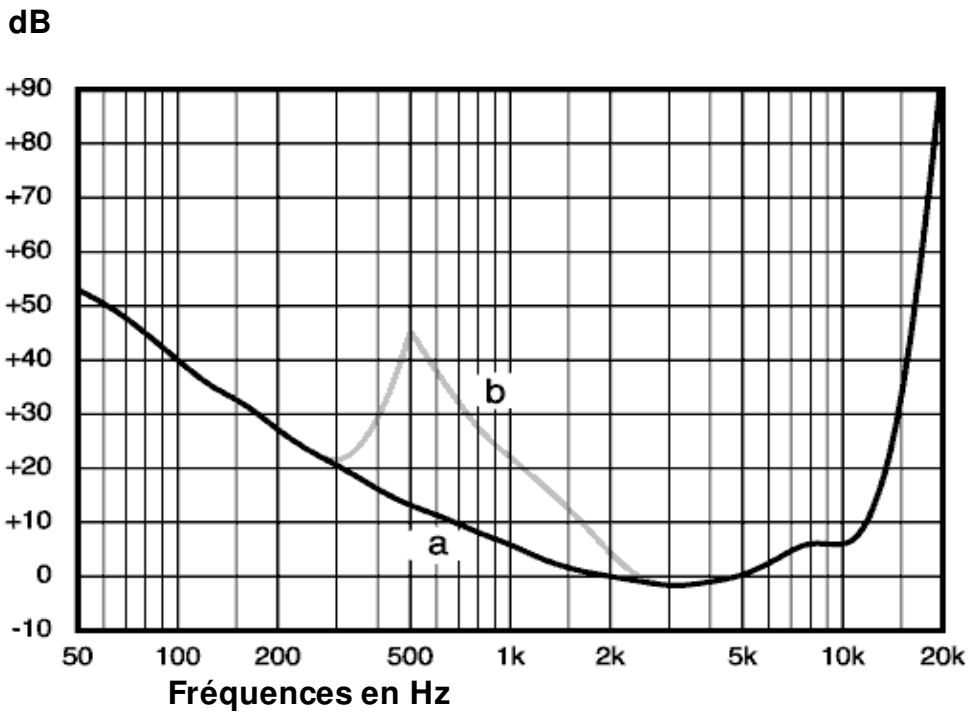


La compression audio-numérique utilise plusieurs techniques psychoacoustiques de façon à n'encoder que les informations utiles. L'une de ces techniques est le **masquage**. Le seuil à partir duquel l'oreille humaine perçoit un son dépend énormément de la fréquence de ce son (voir schéma ci-dessus). Par exemple, nous percevons beaucoup plus facilement un son faible à 4 kHz qu'à 50 Hz ou 15 kHz. De plus, à partir de 25 kHz, quelque soit le niveau sonore, l'oreille humaine ne perçoit plus aucun son.

La technique du masquage est la suivante : si deux sons de fréquences proches sont joués avec une intensité très différente, on pourra supprimé le son le plus faible qui sera de toute façon masqué et ignoré par l'oreille humaine.

Le schéma ci-dessous met en relief le phénomène de masquage de fréquences. En effet, en présence d'un signal de 500 Hz, un son d'une fréquence de 500 Hz est masqué et son niveau doit dépasser la courbe b pour être audible par l'oreille humaine. Si son niveau est en dessous, on n'est pas obligé d'en tenir compte dans le fichier compressé.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr



D'autres techniques psychoacoustiques seront explicitées ultérieurement dans le chapitre concernant la **compression**.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

3.LA METHODE DE HUFFMAN

3.1. Notion Algorithmes

Un algorithme est une procédure bien définie pour résoudre un problème en un nombre fini d'opérations. A l'origine, ce terme a été utilisé pour les règles permettant de faire les opérations arithmétiques dans la base dix. Ainsi, lorsque que l'on additionne ou multiplie deux nombres, on utilise un algorithme qui permet d'arriver à la réponse en un nombre fini d'étapes. Aujourd'hui le terme algorithme a pris un sens plus général pour inclure toutes les procédures pour résoudre des problèmes.

Lors d'une compression deux types d'algorithmes peuvent être utilisés.

L'algorithme **non destructif**. Sa propriété principale est qu'une fois décompressé, le fichier est rigoureusement le même que le fichier avant le compression. Il n'y a donc aucune perte d'information.

Son fonctionnement est le suivant : lors de la compression l'algorithme de compression recherche dans le fichier des occurrences multiples d'une suite d'octets. Il se crée ainsi ce que l'on appelle un dictionnaire, index qui référence ces suites d'octets répétitives. Ensuite, l'algorithme remplacera celles-ci, dans le fichier compressé, par ces codes qui prennent moins de place.

Si on prend l'exemple d'un fichier texte l'algorithme va créer une table de correspondance pour des mots qui sont utilisés plusieurs fois, 'X= 1', 'Y= 2', 'Z= 3', 'O= 4'. Ainsi, dans le texte, ces mots vont être remplacés par leur code, p)lus court, donc prenant moins de place dans le fichier compressé. Voilà, de manière simplifiée le fonctionnement de la compression non destructive.

Pour ce qui concerne l'audio on se servira d'une compression qui utilise un algorithme **destructif**. Un fichier texte peut se compresser très facilement, ce n'est pas le cas de l'audio, car il n'y a pas beaucoup de répétition.

3.2. Codage Huffman/ Shannon-Fano

Les ordinateurs codent généralement des caractères en utilisant le diagramme standard « ASCII », qui assigne un code de 8 bits à chaque symbole. Pour exemple, la lettre 'a' a pour valeur « ASCII » : 61 et est codé en tant que 00111101. Des caractères qui se produisent plus fréquemment comme 'e' sont traités de la même façon que les caractères rares tels que 'u'. Un dossier qui a 100 caractères exigeront 800 bits, cette valeur est fixe, si le dossier contient 100 caractères uniques ou s'il a 100 occurrences de même caractère. Les avantages de l'arrangement de codage ASCII est que des frontières entre les caractères sont facilement déterminées, et le modèle utilisé pour chaque caractère est fixe et universel.

Cependant, dans n'importe fichier texte, il y a des caractères qui se produisent plus fréquemment que d'autres. Aux caractères plus souvent usités ont été affecté des codes courts, et pour les caractères moins souvent utilisés ont été affecté des codes plus longs.

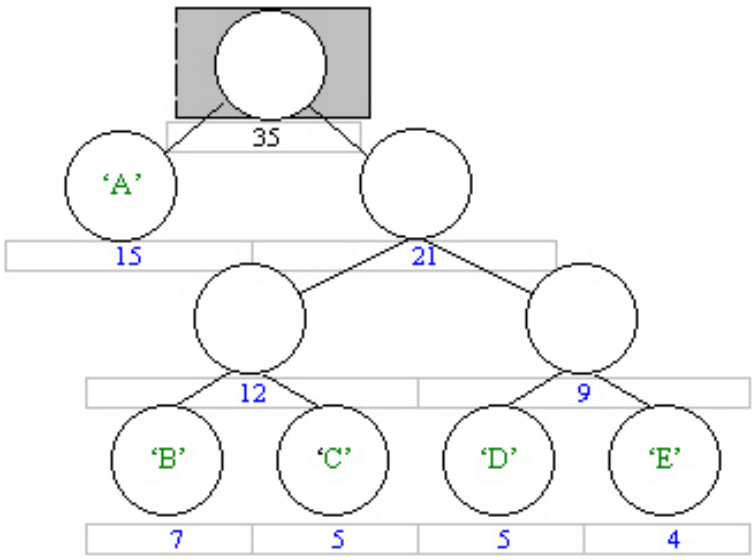
Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Cette idée n'est pas nouvelle. Un exemple de ce type de compression est le code Morse, développé par Samuel Morse au milieu du 19ème siècle. Des lettres envoyées par le télégraphe sont codées avec des points et des tirets. Morse a noté que certaines lettres se sont produites plus souvent que d'autres. Afin de réduire le temps moyen requis pour envoyer un message, il a assigné des ordres plus courts aux lettres qui se produisent plus fréquemment comme *e*(·) et *a*(·-), et de plus longs ordres aux lettres qui se produisent moins fréquemment comme *q*(--·-) et *j*(·---).

Cette idée d'employer des codes plus courts pour les caractères plus fréquemment utilisés a été prise dans le champ de calcul par Claude Shannon et Fano dans les années 50, quand ils ont développé l'algorithme de compression de Shannon-Fano. Cependant Huffman a édité un papier en 1952 qui améliore l'algorithme légèrement.

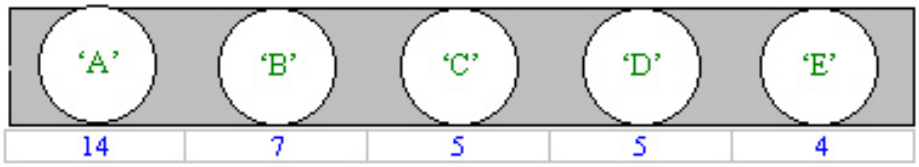
3.3. L'arbre de Huffman

Le codage d'Huffman crée ce qui s'appelle un arbre d'Huffman, qui est un arbre binaire tel que celui-ci :

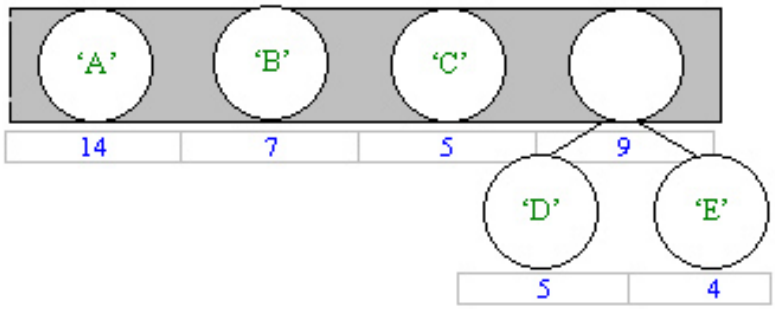


Je vais expliquer la construction de cet arbre par un exemple. Considérons une phrase composée de 35 caractères dont 14 A, 7 B, 5 C, 5 D, 4 E.

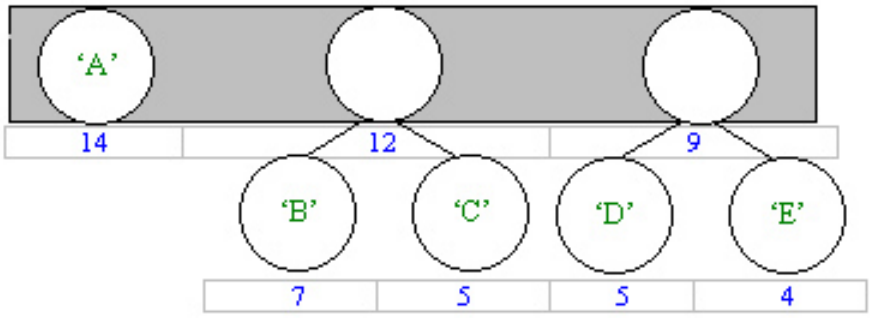
Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr



Nous commençons par choisir les deux plus petits nœuds, qui sont D et E. Nous combinons ces deux nœuds dans un nouvel arbre dont la racine est la somme des poids choisis dans ce cas 9 . Puis, nous remplaçons les deux nœuds par l'arbre combiné.

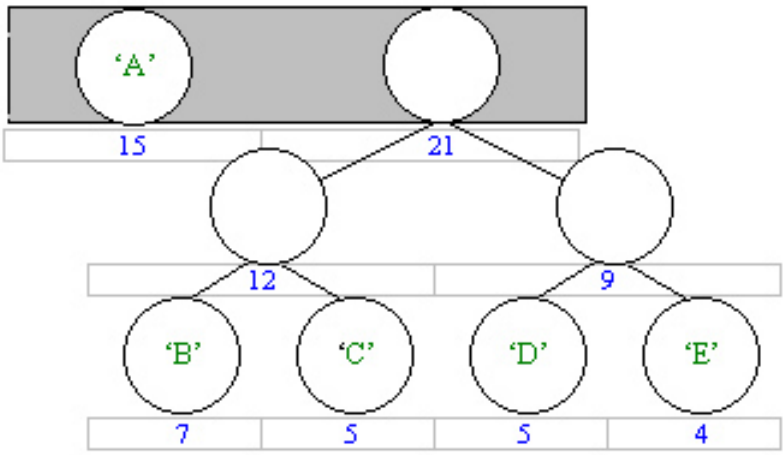


Après nous répétons cette étape, combinant B et C. Nous enlevons ces deux nœuds, et, comme dans la première étape, nous les combinons dans un arbre de poids 12. notez que sur chaque itération, le nombre de nœuds restants dans le choix se rétrécit par un, car nous enlevons deux nœuds et les remplaçons par un nœud simple de racine.

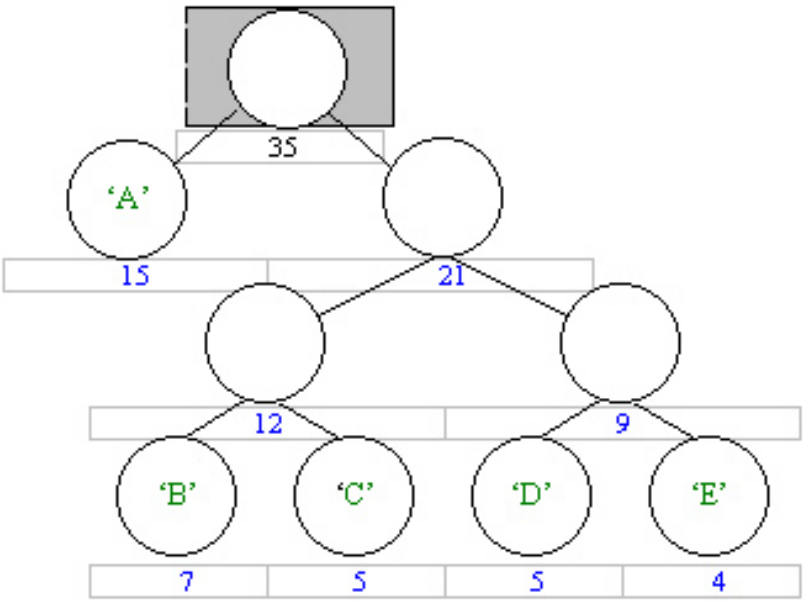


De nouveau, nous retirons les plus petits nœuds et construisons un arbre du poids 21.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr



Et en conclusion, nous combinons les deux derniers nœuds restants dans notre file d'attente pour obtenir notre arbre final, la racine de l'arbre final aura toujours un poids égal au nombre de caractères dans le dossier d'entrée, qui est dans cet exemple 35.



Pour lire les codes de cet arbre de Huffman, nous commençons par la racine et ajoutons un 0 à chaque fois que nous nous déplaçons vers la gauche, et ajoutons un 1 a chaque fois que nous nous déplaçons vers la droite.

Cet arbre nous donne le tableau ci-dessous :

Symbole	Code
A	0
B	100
C	101
D	110
E	111

Je vous ai présenté d'une manière simpliste le fonctionnement du codage de Huffman. Maintenant je vais passer à la compression.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

4. QU'EST-CE QUE LA COMPRESSION

De nos jours, la puissance des processeurs augmente plus vite que les capacités de stockage, et énormément plus vite que la bande passante des réseaux, car cela demande d'énormes changements dans les infrastructures de télécommunication.

Ainsi pour palier à se manque, il est courant de réduire la taille des données en exploitant la puissance des processeurs plutôt qu'en augmentant les capacités de stockage et de transmission des données.

La compression consiste à réduire la taille physique de blocs d'informations. Un compresseur utilise un algorithme qui sert à optimiser les données en utilisant des considérations propres au type de données à compresser. Un décompresseur est donc nécessaire pour reconstruire les données originelles grâce à l'algorithme inverse de celui utilisé pour la compression.

La méthode de compression déprend intrinsèquement du type de données à compresser ; on ne compressera pas de la même façon une image qu'un fichier audio.

4.1. Les différents types de compression et de codage

La **compression physique** agit directement sur les données, il s'agit ainsi de regarder les données redondantes d'un train de bits à un autre.

La **compression logique** par contre est effectuée par un raisonnement logique en substituant une information par une information équivalente.

La **compression symétrique** utilise la même méthode pour compresser et décompresser l'information. Il faut donc la même quantité de travail pour chacune de ces opérations. C'est ce type de compression qui est généralement utilisé dans les transmissions de données.

La **compression asymétrique** demande plus de travail pour l'une de ces opérations, on recherche souvent des algorithmes pour lesquels la compression est plus lente que la décompression. Des algorithmes plus rapides en compression qu'en décompression peuvent être nécessaire lorsque l'on archive des données auxquelles on n'accède peu souvent (pour des raisons de sécurité par exemple), car cela crée des fichiers compacts.

Les programmes ont besoin de conserver leur intégrité pour fonctionner, en effet il n'est pas concevable de reconstruire à l'à peu près un programme en omettant parfois des bits et en ajoutant là où il n'en faut pas.

La **compression avec pertes** se permet d'éliminer quelques informations pour avoir le meilleur taux de compression possible, tout en gardant un résultat qui soit le plus proche possible des données originales. C'est le cas par exemple de certaines compressions d'images ou de sons.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Certains algorithmes de compression sont basés sur des dictionnaires spécifiques à un type de données : ce sont des **encodeurs non adaptatifs**. Les occurrences de lettres dans un fichier texte par exemple dépendent de la langue dans laquelle celui-ci est écrit.

Un **encodeur adaptatif** s'adapte aux données qu'il va devoir compresser, il ne part pas avec un dictionnaire déjà préparé pour un type de données.

Un **encodeur semi-adaptatif** construira celui-ci en fonction des données à compresser : il construit le dictionnaire en parcourant le fichier, puis compresse ce dernier.

La **compression RLE** est utilisée par de nombreux formats d'images (BMP, PCX, TIFF). Elle est basée sur la répétition d'éléments consécutifs. Une première valeur (codée sur un octet) donne le nombre de répétitions, une seconde valeur donne la valeur à répéter (codée sur un octet).

La phrase suivante 'oooooohhhhhhhhhhh' donnerait '6o11h', elle est très utile dans ce cas là. Par contre dans 'onde' cela donne '1o1n1d1e', elle s'avère ici très coûteuse.

Le **codage Huffman** est une méthode qui permet d'attribuer un mot de code binaire aux différents symboles à composer (pixels ou caractères par exemple). La longueur de chaque mot de code n'est pas identique pour tous les symboles : les symboles les plus fréquents (qui apparaissent le plus souvent) sont codés avec de petits mots de code, tandis que les symboles les plus rares reçoivent de plus longs codes binaires. Ainsi la suite finale de mots codés à longueurs variables sera en moyenne plus petite qu'avec un codage de taille constante.

Le codeur de Huffman crée un arbre ordonné à partir de tous les symboles et de leur fréquence d'apparition. Les branches sont construites récursivement en partant des symboles les moins fréquents. Plus le symbole est 'profond' dans l'arbre, plus le mot de code sera long.

Les compressions basées sur ce type de codage donnent de bonds taux de compressions, notamment pour les images monochromes (les fax par exemple).

La **compression LZW** est un algorithme très rapide aussi bien en compression qu'en décompression.

Ses créateurs sont Abraham Lempel et Jacob Ziv, ils ont créé le compresseur LZ77 en 1977. Il était utilisé pour l'archivage (les formats ZIP, ARJ et LHA l'utilisent).

En 1978 ils créés le compresseur LZ78 spécialisé dans la compression d'images.

En 1984, Terry Welch le modifia pour l'utiliser dans les contrôleurs de disques durs, son initiale vint donc se rajouter à l'abréviation LZ pour donner LZW.

Le LZW substitue des motifs en construisant au fur et à mesure un dictionnaire. De plus il travaille sur des bits et non sur des octets, il ne dépend donc pas de la manière de laquelle le processeur code les informations. C'est un des algorithmes les plus populaires, il est utilisé notamment dans les formats TIFF et GIF.

Construction du dictionnaire :

Le dictionnaire est initialisé avec les 256 valeurs de la table ASCII. Le fichier est découpé en chaînes d'octets (ainsi pour des images monochromes -codées sur 1 bit. - cette compression est peu efficace), chacune de ces chaînes est comparée au dictionnaire et est ajoutée si jamais elle n'y est pas présente. Pour la décompression l'algorithme reconstruit le dictionnaire dans le sens inverse, il n'a donc pas besoin d'être stocké.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

4.2. Informations sur la compression non destructive

Il existe plusieurs algorithmes non destructifs (le ZIP, RAR , ACE). D'autres sont plus dédiés au son. Notons le WAVE (.wav) développé par Microsoft pour les PC et l'AIFF (.aif) développé par Apple pour les macintosh. Elle utilise comme mode de compression le PCM explicité dans le chapitre ' **Introduction à l'audionumérique'**).

De nombreux autres algorithmes propres à ce type de média existent, bien qu'ils soient moins connus. Citons WavArc (.arc), tiré du format de compression ARC de Dennis Lee, AudioZIP (.zip), tiré du ZIP de Lin Xiao, LPAC de Tilman Liebchen, Monkeys Audio de Matthew T. Ashland et enfin RKAU de Malcolm Taylor.

Sound Forge propose quant à lui un format propriétaire, le Perfect Clarrity Audio (dont l'extension des fichiers est le PCA).

Cependant, malgré la compression, les fichiers audio sont encore trop volumineux pour être utilisés dans certaines applications : dans l'utilisation de tels fichiers sur Internet, certes, mais aussi dans les canaux de transmission numériques comme le câble télévisé. On a donc recours à un autre type de compression : la compression destructive.

4.3. La compression destructive.

Il existe également de multiples formats de compression destructive. Il faut savoir que l'un d'entre eux, le très connu **MP3**, a été créé et breveté par Thomson Multimédia. Aussi théoriquement, chaque personne qui crée un fichier MP3 pour l'exploiter commercialement est censé reverser des droits à l'entreprise.

Le format imposé par Microsoft, le Windows Média Audio (**WMA**), ressemble au MP3. Son format est également breveté donc son utilisation se fait en contre partie de droit reversé à Microsoft.

Le format **OGG Vorbis** quant à lui résout ces problèmes de brevet. C'est un format « quasi identique » au MP3, à la différence que l'utilisation de celui-ci est libre de droit. Mais ce format n'est pas aussi répandu que le MP3 ou le WMA.

Parmi les autres formats de compression destructive, citons le **MP3Pro**, le **Real Audio** de Real Networks, le Yamaha **VQF**.

Les algorithmes utilisés sont principalement le **MPEG** (pour le format MP3), l'**AAC** (MP3Pro), l'**ATRAC** (**Sony Minidisc**), le **PASC** (Philips **DCC**), et enfin les **Dolby AC-1**, **AC-2** et **AC-3**.

Afin de définir les différentes techniques utilisées pour compresser le son, je vais principalement m'appuyer sur le format MP3, l'ATRAC, format audio utilisant typiquement la compression destructive, donc les limites psychoacoustiques de l'oreille humaine, afin de supprimer certaines parties du son inaudibles.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

5. L'ATRAC

5.1. Introduction

L'ATRAC signifie : Adaptive Transform Acoustic Coding for MiniDisc.

ATRAc est un système audio de codage basé sur des principes psychoacoustiques.

Le MiniDisc, sur lequel est stocké le signal codé au format ATRAC, est composé d'un disque optique ou magnéto-optique de 64 millimètres qui a approximativement 1/5 de la capacité de stockage de donnés d'un disque compact standard. En dépit de la capacité de stockage réduite, il était nécessaire que le MiniDisc maintiennent la qualité de son élevée et un temps de jeu de 74 minutes. Le système de la compression de donnés ATRAC a été pour cette fin conçu pour répondre aux critères suivants :

- Compression des 44.1 kiloHertz de 16 bits d'acoustique stéréo dans moins de 1/5 du début original avec la réduction minimale de la de son.
- Exécution de matériel simple et peu coûteuse appropriée aux joueurs et aux enregistreurs portatifs.

L'ATRAC emploie la psychoacoustique. Il utilise le seuil d'audibilité. Comme nous l'avons vu dans le chapitre 2, l'oreille est sensible aux fréquences comprises entre 2 kHz et 5 kHz. Il utilise le masking fréquentiel : pour deux sons de fréquences proches, le signal le moins fort ne sera pas ou peu entendu en fonction de son intensité (vu au chapitre 2).

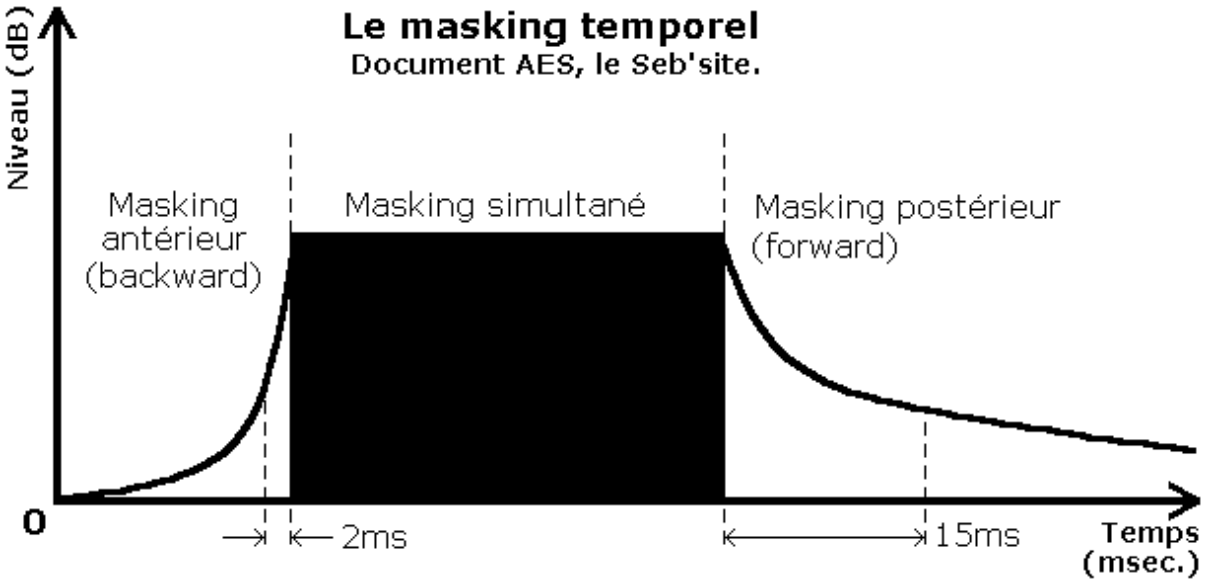
L'ATRAC met en œuvre une troisième notion de la psychoacoustique qui est le masking temporel.

Principe : après l'audition d'un son fort, l'oreille met un certain temps avant d'entendre un autre son de fréquence proche si son niveau est plus faible. L'expérience qui permet de le constater est délicate à réaliser. Il faut générer une fréquence 'masquante' (1 kHz à 60 dB par exemple), et une fréquence masquée (1.1 kHz à 40 dB pour cet exemple). Ce qui demande deux générateurs de fréquences. Pour mesurer le temps du masking temporel :

1. on coupe la fréquence 'masquante',
2. après un délai ***t,*** on coupe la fréquence masquée,
3. on recommence l'expérience en diminuant le délai ***t*** jusqu'à ce que le 1.1 kHz ne soit plus audible.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Sur la figure ci-dessous, c'est pendant les 2 et 15 milliseconde que le masking se produit, ces deux portions du signal ne seront pas analysées par l'oreille. Le système ATRAC ne les codera pas.



La dernière notion psychoacoustique sur laquelle repose le procédé ATRAC est la notion de la bande critique.

Pour comprendre cette partie, il faut partir de l'idée que l'oreille analyse le spectre audio en le découpant en bandes dites critiques. Dans chacune de ces bandes (dont la largeur varie en fonction de la précision de l'oreille à ces fréquences et non de la largeur de la courbe de masking), l'oreille assimile les fréquences et ne les différencie plus. Ces bandes de fréquences correspondent d'ailleurs à la répartition des cellules ciliées dans l'oreille interne.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Le tableau ci-dessous donne le détail de ces bandes. >On remarque que ces bandes sont plus étroites dans les basses fréquences. En fait, les trois quarts de ces bandes sont situés sous 5 kHz, ce qui montre bien que l'oreille est plus précise dans les basses fréquences que dans les hautes.

Bandes critiques discrètes

Critique Bande	Fréquence (Hertz) Bas	Haut	Largeur	Critique Bande	Fréquence (Hertz) Bas	Haut	Largeur
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

5.2 L'encodeur ATRAC (1ère étape)

Le cahier des charges du codeur ATRAC est de compresser les données numériques de façon à n'en garder que 1/5, tout en conservant une qualité sonore subjective avec les standards de la haute-fidélité. Pour cela, il va travailler en tenant compte des limites et des défauts de l'oreille et du cerveau de l'auditeur.

Les effets de masking rendent inutiles des portions entières du signal audio. De même, certaines bandes critiques sont plus larges que d'autres et méritent moins de définition. Le point de départ de la transformation est le signal stéréo PCM issu du disque compact. Il est codé sur 16 bits, avec une fréquence d'échantillonnage de 44.1 kHz.

Le codeur procède en trois étapes comme le montre la figure ci-dessous :

1. Création des **BFU**s (Bloc Floating Units) :
    a) Le signal est divisé en trois bandes par les filtres **QMF**.
    b) Chacune de ces trois bandes est transformée par les blocs **MDCT** en domaines de fréquences produisant ainsi des bandes (spectrales ou raies) avec chacune un coefficient ; il y en a 512 ; on parle de bandes critiques.
    c) Ces coefficients sont groupés non uniformément en blocs, les **BFU**s

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

2. Une allocation appropriée des bits les réorganise en mettant les bits de poids forts en début de mots
3. Une quantification réduit la taille des données en ne laissant passer finalement que les données non significatives, ce en vertu de l'effet de masking.

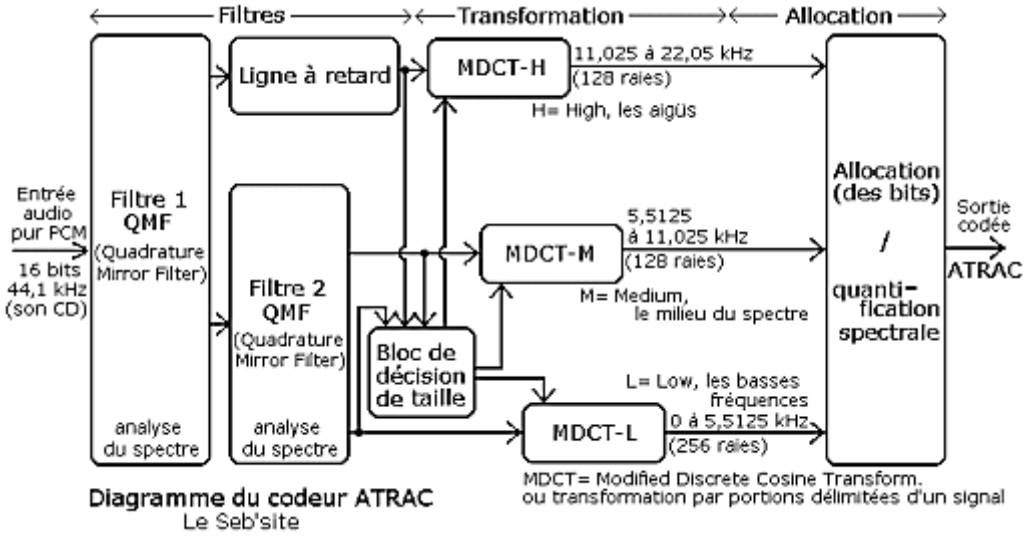


Diagramme du codeur ATRAC
Le Seb'site

Les filtres **QMF** divisent le signal en trois bandes :
- de 0 à 5.512 kHz,
- de 5.512 à 11.025 kHz,
- de 11025 à 22.05 kHz.

Les blocs **MDCT** (Modified Discrete Cosine Transform) transforment chacune de ces bandes en domaine de fréquence. C'est l'équivalent d'une transformée de Fourier (**F.F.T**, Fast Fourier Transform) afin d'orthogonaliser le signal et de travailler sur des axes différents, à savoir le temps et la fréquence. Excepté l'amplitude qui est traduite en un coefficient spectral pour chaque bande.

Le codeur choisi la longueur des blocs en fonction des caractéristiques du signal dans chaque bande. Il choisi entre deux modes :

- Le **mode long** : 11.6 millisecondes,
- Le **mode court** : 1.45 millisecondes dans les fréquences hautes (aiguës) et 2.9 pour les autres.

Le mode long donne une meilleure résolution fréquentielle. Toute fois, un problème intervient pendant les attaques (montés rapides) du signal. Plus précisément, le bruit de quantification augmente rapidement et n'est plus masqué. C'est le pré écho. Pour l'éviter, le codeur détecte les attaques et passe en mode court. Le bruit est toujours présent, mais sur un segment court. Il est masqué par l'effet de backward masking.

Le backward masking est certes effectif sur courte portion, mais puisque le mode court réduit la durée du bruit, l'oreille ne l'entend pas. L'ATRAC peut alors coder les passages plus stables, tout en ayant répondu aux attaques.

Le même problème se produit lorsque le signal décroît rapidement, mais le fowrad masking dure plus longtemps que le backward masking. Le phénomène inverse ne pose aucun problème.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

5.3. L'encodeur ATRAC (2ème partie)

Cette deuxième partie parlera de la **quantification**.
Les valeurs des BFUs sont quantifiées avec deux paramètres :

- un facteur d'échelle (scale factor) : il définit la largeur de la fenêtre de quantification et est choisit dans une liste fixe de possibilités traduites par des codes,
- la longueur du mot binaire (wordlength) : il définit la précision dans la fenêtre analysée et est obtenu par un algorithme comme celui de Huffman,

Le facteur d'échelle et la longueur de mot restent liés afin de refléter les propriétés psychoacoustiques de chaque BFU.

A ce stade on obtient les **sounds frames**, ou samples de données. Pour chaque sample, les informations suivantes sont enregistrées :

- Le mode (court ou long) pour la taille des blocs.
- La longueur de mot binaire pour chaque BFU.
- Le facteur d'échelle pour chaque BFU.
- Les coefficients spectraux quantifiés.

L'**allocation binaire** :
C'est maintenant que commence la compression des données. Mais le plus gros du travail a déjà été fait. On a des samples avec coefficients qui sont représentatifs de leurs propriétés psychoacoustiques. Et les mots binaires ont été construits de façon à ce que les bits de poids forts soient en premier.

Pour simplifier, il ne reste plus qu'au codeur qu'à réduire le nombre de BFUs, privilégiant les plus significatifs du signal, supprimant ceux qui ne seraient pas entendus par le cerveau et l'oreille.

Pour ce faire, on peut utiliser un algorithme de Huffman qui permet de stocker les octets les plus fréquemment rencontrés sur un nombre minimal de bits. On élimine aussi les redondances. Le signal est désormais compressé suivant le principe ATRAC. Mais cette formule n'est pas figée et l'ATRAC permet d'utiliser n'importe quel algorithme au codage, pourvu que l'enveloppe ATRAC soit respectée. D'où un énorme avantage : l'évolution du procédé ATRAC.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

5.4. L'évolution de l'ATRAC.

Comme nous l'avons vu précédemment, le procédé ATRAC n'est pas figé et évolue en fonction des recherches des fabricants.

Différentes versions ont donc vu le jour depuis le lancement du MiniDisc par Sony en 1992. Certains affirment que sony aurait sortit le **MD** avant que l'ATRAC soit réellement au point afin de ne pas être devancé par la **Digital Compact Cassette** (**DCC**) de Philips... Ceci explique les défauts audibles et les critiques des audiophiles qui ont désormais du mal à accepter le MiniDisc, lui préférant le **Digital Audio Tape** (**DAT**).

Voici l'historique sur les différentes versions de l'ATRAC, tirés du magazine allemand **Stereo**.

	ATRAC 1	**ATRAC 2**	**ATRAC 3**	ATRAC 3.5
Sortie	1992	1994	1995	1996
Seuil	15kHz !	18kHz	18kHz	18kHz
Bruit	Important	Moins que l' ATRAC 1	Filtre dynamique : pas de bruit lors de silences	= DAT
Son	Métallique	Proche du DAT	Égal au DAT en écoute "blind test*"	DAT
"Spratzeln"**	Audible	Pas mieux	Audible seulement dans les silences	Supprimé

*"Blind test" : les yeux bandés pour éviter d'être subjectif.
**"Spratzeln" : c'est le terme utilisé par le magazine allemand pour évoquer le bruit audible sur les premières versions. Assez difficile à décrire, il ressemble au bruit que font les bulles à la surface d'une boisson gazeuse.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Compatibilité entre versions.

	Enregistreur ATRAC 1	Enregistreur ATRAC 2	Enregistreur ATRAC 3	Enregistreur ATRAC 3.5
Lecture ATRAC 1	Voir tableau ci-dessus	15kHz, moins de bruit	15kHz, moins de bruit	15kHz, moins de bruit, bruit "Spratzeln" supprimé
Lecture ATRAC 2	15kHz, moins de bruit, son toujours métallique	Voir tableau ci-dessus	Aucune différence qu'avec un enr. ATRAC 2	Plus de "Spratzeln", un peu moins de bruit (comme avec un ATRAC 2)
Lecture ATRAC 3	Peu de différence avec ci-dessus	Pas de différence avec un lecteur ATRAC 2	Voir tableau ci-dessus	Plus de "Spratzeln", un peu moins de bruit (comme avec un ATRAC 3)
Lecture ATRAC 3.5	Peu de différence avec ci-dessus	"Spratzeln" toujours idem, mais moins de bruit	Il reste du "Spratzeln", mais moins de bruit qu'au dessus	Voir tableau ci-dessus

Le lecteur est un système fixe qui, contrairement au codeur, utilise une technique figée.
Les algorithmes de compression ne cessent d'être améliorés. Il faut donc privilégier la version la plus récente pour l'enregistrement.

Comme tout format numérique, l'ATRAC est censé permettre la copie de copie sans pertes de qualité. Ce n'est pas le cas en réalité.

Voici ce que donne la multi-copie avec le format ATRAC(ces informations sont tirés du magazine allemand **Stereo**).

ATRAC 1 vers ATRAC 1	Inacceptable après 5 générations, mauvais après 20 générations.
ATRAC 2 vers ATRAC 2	Pas de différence audible après 5 générations, légère distorsion après 20 générations.
ATRAC 3 vers ATRAC 3	Pas mieux qu'avec l' ATRAC 2.
ATRAC 3.5 vers ATRAC 3.5	Légèrement meilleur que l' ATRAC 3 pour le bruit.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

6. LE MP3

6.1. L'histoire

Le MP3 est une technique de compression des formats audio-numériques. Le brevet en a été déposé par l'institut allemand de recherche Fraunhofer. Au départ les chercheurs de cet institut travaillaient sur la compression de séquences vidéo en vue de la diffusion de la TV numérique. Le projet **Euréka** EU147 en 1987 a commencé à bien faire bouger les choses. Les chercheurs de l'institut Fraunhofer ont travaillé, dès le départ, en accord avec les normes établies par le **MPEG** (Moving Pictures Experts Group), un groupe d'experts au sein de l'ISO définissant le standard international pour la compression vidéo. Le MPEG-1 Layer 1 fut la 1ère norme à voir le jour. La 3ème version de cette technique est la MPEG-1 Layer 3 dont la variante spécifiquement audio est la norme MPEG-1 Audio Layer 3. Ce format a été standardisé par l'ISO, fin 1992. En novembre 1996, un brevet est déposé aux USA.

Les fichiers audio enregistrés sous ce format ont comme extension .mp3. Un fichier audio-numérique compressé à l'aide du format MPEG-1 Audio Layer 3 est appelé **MP3**.

L'institut allemand Fraunhofer détient 10 des 18 brevets MP3, Thomson Multimédia détient les 8 autres et gère les licences.

6.2. Le concept

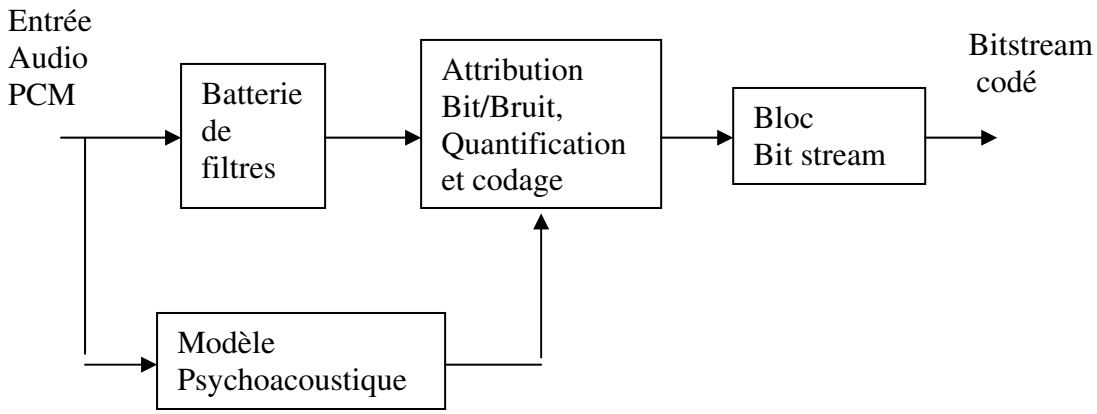
Sans n'importe quelle sorte de compression, il faut 1 méga octets pour représenter 1 seconde de musique CD de qualité. Mais en employant le codage audio MPEG, la taille peut être rétrécie par un facteur de douze sans perdre en terme de qualité de son.

La compression MP3 utilise tous les concepts psychoacoustique étudiés dans les chapitres précédents.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

6.3. L'algorithme pour la compression MP3

Le processus de la compression MP3 peut être décomposé en 3 étapes. D'abord, le jet audio d'entrée traverse une **batterie de filtres** qui divise le signal en sous-bandes de fréquence. Simultanément, il traverse un **modèle psychoacoustique** qui utilise le concept de masque auditif pour déterminer ce qui peut ou ne pas être entendu dans chaque sous-bande. Le **bloc d'attribution de bit** réduit au minimum l'audibilité du bruit. En conclusion, le bloc de train binaire accumule toutes les informations et processus. De ce bloc il en sort un bitstream codé.



6.3.1. La batterie de filtres Hybride

Le but de la batterie de filtres est de diviser le signal audio en 32 sous-bandes de fréquences d'égale largeur. L'oreille humaine a une résolution limité qui peut être exprimée en termes de largeurs de bande critiques moins que 100 Hz et plus de 4 kHz. Dans une largeur de bande critique l'oreille humaine brouille des fréquences. Ainsi la batterie de filtres crée les sous bandes de fréquence d'égale largeur qui se corrèlent avec les largeurs de bande critiques.

6.3.2. Le modèle psychoacoustique

La branche de la psychoacoustique examine le concept de l'auditoire masquant et son effet sur la compression. Dans chaque sous bande où le flou se produit, la présence d'un signal fort, par exemple peut masquer une région des signaux plus faibles.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

6.3.3. Attribution de bit

Par un algorithme itératif, l'attribution de bit emploie l'information du modèle psychoacoustique pour déterminer le nombre de bit de code à assigner à chaque sous bande. Ce processus peut être décrit en utilisant la formule suivante :

$$MNR\ (dB) = SNR\ (dB) - SMR\ (dB)$$

MNR est le rapport de masque à bruit.
SNR est le rapport signal/bruit, donné avec la norme audio MPEG.
SMR est le rapport de signal à masque, dérivé du modèle psychoacoustique.

Alors les sous bandes sont placées par ordre du plus bas au rapport de masque à bruit le plus élevé, et la plus basse sous-bande est assignée au plus petit nombre de bit de code et ce processus continue jusqu'à ce qu'il n'y ai plus de bit qui ne puisse être assigné.

Deux boucles d'itérations appelées **boucle de taux** et **boucle de bruit control/ distorsion** servent à quantifier et coder dans les encodeurs MP3. Les valeurs quantifier sont codées en utilisant les méthodes d'Huffman.

6.4. Notions de Couches

Le MP3 a trois couches distinctes pour la compression. La couche 1 forme l'algorithme de base et les autres couches augmentent la couche 1.

L'algorithme de la couche 1 code des données audio en groupant ensemble, 12 échantillons provenant de chacune des 32 sous-bandes créées dans l'étape « batterie de filtres » pour un total de 384 sous-bandes. Chaque groupe de 12 échantillons obtient une attribution de bit et un facteur de balance. L'attribution de bit indique au décodeur le nombre de bit employé pour représenter l'échantillon, tandis que le facteur de balance est un multiplicateur qui classe les échantillons.

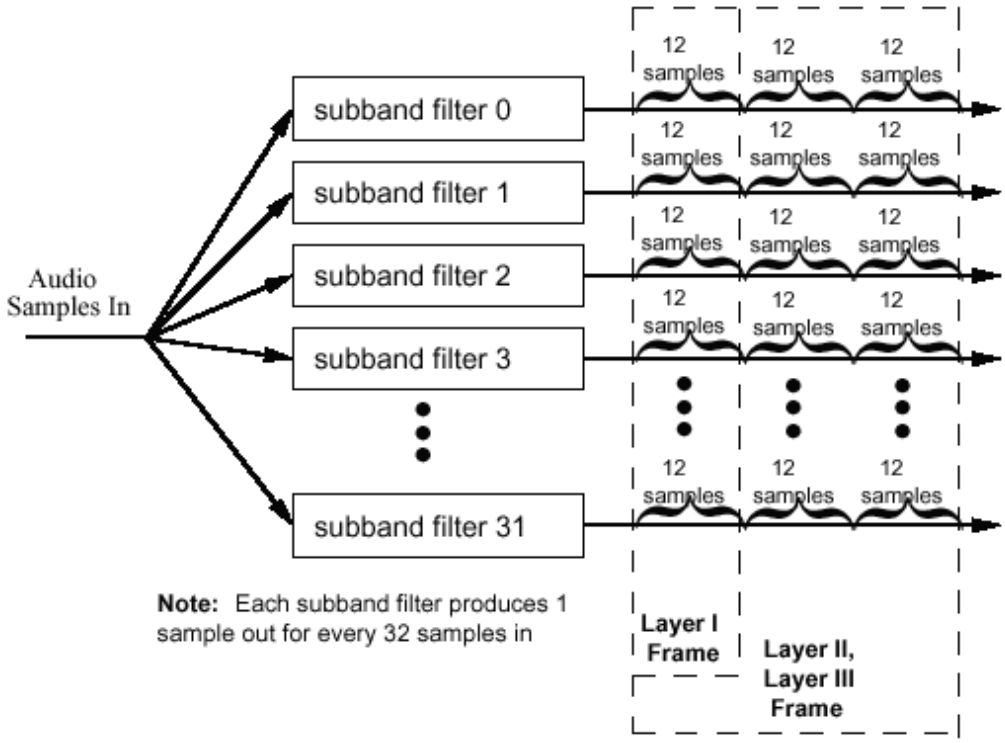
L'algorithme de la couche 2 augmente la couche 1 en codant des données dans de plus grands groupes avec des restrictions sur des attributions de bit pour des valeurs dans de plus grandes sous-bandes.

L'encodeur de la couche 2 rassemble trois groupes de douze échantillons comme organisé dans la couche 1.

La couche 2 sauvegarde un bit qui représente l'attribution de bit, les valeurs de facteur de balance, et les échantillons quantisés. Ceci tient compte pour que plus de bits soient consacrés à améliorer la qualité audio.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

La couche 3 est une amélioration au-dessus des deux autres couches parce qu'elle utilise une transformation connue sous le nom de **MDCT**, pour représenter la fréquence du signal à 32 bandes de fréquence différentes.



A débit égale, la qualité obtenue est proportionnelle au niveau de la couche utilisée. Le temps de compression/décompression l'est aussi. En pratique, le layer 1 convient à des applications où prime la qualité et/ou le temps d'encodage/décodage.

La plage de débits qu'offre le MPEG-1 s'échelonne de 32 à 448 kbps/s pour le Layer 1, de 32 à 384 pour le Layer 2 et de 32 à 320 pour le Layer 3.

Rien n'empêche d'encoder un signal en choisissant n'importe quel débit à l'intérieur de ces plages, mais à chacun de ces trois niveaux correspond un débit idéal, celui pour lesquels ils ont été conçus. 384 kbps/s par canal pour le Layer 1, 192 pour le Layer 2 et 128 pour le Layer 3 (très usité sur internet).

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

6.5. Utilisation du bitrate en MP3

On peut encore ajouter qu'il faut aussi faire attention à la technologie utilisée par l'encodeur. Il existe en effet trois grandes familles d'encodeurs qui se distinguent par leur algorithme de codage. Il y a les encodeurs développés par l'institut Fraunhofer, les encodeurs à base de routines développées par Xing Technologie et ceux à base de routines ISO. Grossièrement, on peut dire que Fraunhofer privilégie la qualité du signal compressé tandis que Xing privilégie la rapidité de ses algorithmes. Pour les autres cela dépend des créateurs.

Tableau de référence de l'utilisation du bitrate constant.

Bitrate	**Qualité**	**Compression**	**1 mn de son occupe**	**Notre avis sur l'utilisation**
Référence	CD audio	Aucune	10,094 Mo	Toutes (chaîne Hifi, mini-chaîne, ordinateur ou encore baladeur)
192 kbits/s ou plus	CD audio conservée	1:7 ou moins	1,373 Mo ou plus	Toutes (chaîne Hifi, mini-chaîne, ordinateur ou encore baladeur)
128 kbits/s	CD audio conservée (limite)	1:11	0,915 Mo	Toutes (limite pour bonne Hifi)
96 kbits/s	Supportable	1:15	703 Ko	Ordinateur ou baladeur
64 kbits/s	Limite	1:22	469 Ko	Baladeur
moins de 64 kbits/s	Non supportable	1:22 ou plus	moins de 469 Ko	Non conseillé

Pour l'encodage du signal nous avons fait référence au bitrate constant. Mais il est aussi possible d'encoder à bitrate variable (**VBR**). Ce système permet de faire varier dynamiquement le bitrate du codage en fonction de la complexité du son.

Ainsi un passage musical complexe sera encodé avec un haut bitrate et un simple passage musical avec un plus bas bitrate. L'avantage de cet méthode est qu'elle permet un meilleur codage avec une très large dynamique. L'inconvénient est qu'elle change invariablement la taille du fichier.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

7. LE MP3PRO

Le **MP3Pro** date de fin 2001. Ce format résulte de l'association de deux technologies : le codage MP3 d'une part et la technique de reconstruction de la bande passante d'autre part. Cette technique, nommée SBR (Spectral Bandwidth Replication), qui rétablit une partie des hautes fréquences d'un signal après la transmission de celui-ci, permet l'amélioration de la qualité sonore du signal reçu.

Pour expliquer l'amélioration apportée par le codage MP3Pro, prenons l'exemple d'une radio FM diffusé en numérique.

Le signal analogique a une bande passante d'environ 20 kHz. Avant d'envoyer ce signal sur les ondes, on va le numériser puis le coder pour qu'il occupe moins de place lors de la transmission. Utilisons le MP3 à 64 kbits/s. Le signal encodé n'a plus qu'une bande passante de 11 kHz car on a purement et simplement éliminé toutes les fréquences au-delà de 11 kHz). Le système de réception avant de délivrer le signal à l'utilisateur, va le décoder. Nous savons que le MP3 réalise une compression destructrice. Les données éliminées par l'encodage sont perdues. Le signal, une fois décodé, a donc une bande passante de 11kHz, soit une perte de la qualité audio, par rapport à l'original, pour l'utilisateur.

Le MP3Pro va en partie améliorer ce problème. Les 64 kbits/s disponibles pour réaliser la compression sont utilisés de la manière suivante. Le signal original est encodé à environ 60 kbits/s, en utilisant les techniques de codage MP3 habituelles. Les 4 kbits/s restant servent à véhiculer des informations utiles à la reconstruction de la bande passante après transmission du signal. Des informations telles que la fréquence à partir de laquelle il faut reconstruire la bande passante et le niveau des fréquences présentes dans le signal original mais qui vont être éliminés par codage MP3 sont codés par un encodeur distinct (l'encodeur SBR) et stockés de sorte qu'elles occupent les 4 kbits/s restant. Le signal encodé en MP3Pro véhicule bien 64 kbits/s mais il contient deux parties distinctes. Le signal est ensuite transmis puis décodé. La partie MP3 est décodée normalement et le décodeur SBR reconstruit la partie haute du spectre. L'utilisateur reçoit un signal ayant une bande passante de 16 kHz contre 11 kHz soit une bien meilleure qualité audio.

L'alliance du MP3 et de la technique SBR augmente considérablement la qualité audio d'un signal pour un bitrate donné. On peut raisonner dans l'autre sens et se dire qu'en encodant à un bitrate inférieur en MP3Pro, par exemple 96 kbits/s, on va obtenir la qualité que l'on avait en MP3 à 128 kbits/s. C'est en partie vrai. Le MP3Pro s'il est bien adapté aux bas bitrates, n'améliore pas beaucoup l'encodage à bitrate élevé. En effet, rappelons qu'à bitrates élevés, le MP3 le MP3 standard ne supprime guère les hautes fréquences. D'où l'inutilité de leur reconstruction par technique SBR. L'alliance d'un codage et de la technique SBR est donc surtout efficace pour un codage effectué à des bitrates peu élevés.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

Le MP3Pro peut encoder des sons issus de sources analogiques échantillonnées à une fréquence de 44.1 kHz, sur 16 bits, en mono ou stéréo. L'encodage se fait à bitrate constant. Celui ci peut être réglé de 24 kbits/s à 96 kbits/s. La nécessité de décoder deux parties distinctes fait que, globalement, le décodage d'un fichier MP3Pro est plus long que celui d'un fichier MP3 et surtout nécessite plus de puissance de la part du processeur. Si ceci ne représente pas une contrainte pour les PC actuels, cela pourrait être gênant pour les processeurs embarqués, en particulier pour les téléphones portables ou les baladeurs MP3.

La technique SBR a été mise au point par la société Coding Technologies et des accords ont été passé avec Fraunhofer et Thomsom Multimédia pour le développement des outils MP3Pro. Thomsom Multimédia gère les licences MP3Pro.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

8. LE MPEG-2 AAC

Etant donné le succès rencontré par le MPEG-1, il a été décidé de poursuivre dans cette voie et d'étendre le standard à d'autres types d'applications, de le rendre plus efficace mais aussi mieux armé pour faire face à de faibles débits. D'où la naissance du MPEG-2, publié sous la référence 13818. Cette nouvelle norme compte neuf parties numérotées de 13818-1 à 13818-9 (les cinq premières reprennent les intitulés de leurs homologues MPEG-1). Le développement de ces différentes parties s'est échelonné dans le temps. Deux d'entres elles concernent l'audio, à commencer par la troisième (ISO/IEC 13818-3). Finalisé en 1994 et publié en 1995, elle est non seulement très proche du standard MPEG-1 Audio, mais surtout compatible avec lui. C'est la raison pour laquelle on la surnomme **MPEG-2 BC** (Backward Compatible). Elle reprend le principe des trois Layers et propose une stucture de fichiers similaires.

Les différences portent essentiellement sur deux points.

Premièrement sur les fréquences d'échantillonnage des signaux à encoder : 16, 22.05, 24 kHz, soit la moitié de celles du MPEG-1. D'où des débits plus faibles, s'échelonnant de 32 à 256 kbps/s (Layer 1) et de 8 à 160 kbps/s (Layer 2 et 3). Notons qu'un logiciel de décodage MPEG-1 Audio ne pourra exploiter ces trois nouvelles fréquences.

Deuxièmement sur l'apport de canaux supplémentaires : jusqu'à sept, par exemple pour le surround, en 5.1. Stockés dans des zones auxiliaires ces canaux seront purement et simplement ignorés par un décodeur MPEG-1. Pour maintenir une compatibilité avec ce dernier dans le cadre d'une transmission surround, il est prévu de pouvoir scinder la modulation en deux flux : d'une part un mixage stéréo matricé, de l'autre des informations permettant d'en extraire nos cinq canaux. Concrètement, seul le mixage stéréo matricé sera lu par un décodeur MPEG-1. Tandis qu'un décodeur MPEG-2 tirera profit de toutes les informations.

Devant la nécessité de répondre plus efficacement aux exigences de la diffusion multicanal, mais aussi d'intégrer les progrès effectués en matière de compression, une extension au MPEG-2 Audio a été créée. Incompatible avec le MPEG-1, parfois surnommé MPEG-2 NBC (Non Backward Compatible), elle est officiellement baptisée MPEG-2 AAC.

Finalisé en avril 1997 et matérialisé en 1998 par le document ISO/IEC 13818-7, ce standard représente l'état de l'art en matière de compression. Plus élaboré que le **Dolby AC-3**, particulièrement approprié à la transmission d'audio sur internet.

Dans les faits, il est possible de véhiculer jusqu'à 48 canaux pleine bande, 15 canaux basse fréquence, des canaux multilingues.., et d'encoder des signaux aux fréquences d'échantillonnage comprises entre 8 et 96 kHz. Si la notion de Layer s'est envolée, on retrouve trois niveaux de complexité : **Main**, **LC** (Low Complexity) et **SSR** (Scaleable Sampling Rate). Plus la structure est complexe, plus les temps d'encodage/décodage sont longs, et meilleurs est la qualité à débit égal.

Ce format ne cesse d'évoluer le **MPEG-4 AAC** a aussi été développé (prévu pour la télévision haute définition le **MPEG-3** n'a pas vu le jour). Je n'en parlerai pas car mon étude de la compression s'arrête au MPEG-2 AAC.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

9.LE WMA

Le WMA exploite en 1er lieu les propriétés de l'oreille pour réduire la taille d'un morceau, puis on traite l'aspect stéréo et enfin on applique des codages en terminant par le codage de Huffman. L'utilisation de toutes les possibilités de réduction évoquées dépend de la place que l'on souhaite accorder à 1 minute de son compressé et donc aux taux de compression que l'on va devoir appliquer. Pour l'encodage de fichiers son en WMA on parle plutôt en terme de birate que de taux de compression.

Le bitrate est le nombre de bits que l'on autorise pour coder 1 seconde. On aura donc la relation suivante : plus on veut compresser un morceau (pour qu'il prenne le moins de place possible) plus le bitrate à choisir devra être faible.

Les fichiers encodés à l'aide du codeur WMA portent l'extension.wma. Dans sa forme actuelle, le WMA encode des sons numériques provenant de signazux sonores analogiques échantillonnés à une fréquence pouvant aller de /8 kHz à 48 kHz, sur 8 ou 16 bits, en mono ou stéréo. L'encodage se fait à bitrate constant. Ce dernier peut se régler à n'importe quelle valeur comprise entre 5 kbits/s et 192 kbits/s. Les techniques de compression WMA et MP3 sont semblabes.

Microsoft gère les licences WMA.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

10. L'OGG VORBIS

10.1. L'histoire

Sous l'appellation OGG se cache un certain nombre de projets liés au multimédia et au traitement du signal. La recherche engagée par ce groupe de projets touche à deux thèmes principaux : le codage audio et vidéo. OGG Vorbis s'intéresse au codage audio. C'est un codage avec perte comme le MP3 et le WMA. La spécificité de OGG est que c'est un projet totalement **Open Source** et que le format Vorbis est sans brevet et complètement libre de droits.

OGG est né en 1993 après que l'institut allemand Fraunhofer ait obtenu la standardisation ISO du MP3 et déposé son 1er brevet, faisant ainsi du MP3 un format audio propriétaire. Le groupe du de projets OGG a alors développé un format audio compressé devant rivaliser en taille et qualité avec le MP3 mais gratuit. Les spécifications du format Vorbis sont dans le domaine public.

10.2. Le concept

Le format OGG Vorbis délivre des fichiers audio dont l'extension est .ogg. Les techniques de compression du signal audio sont peu différentes de celles employés par les formats MP3 et WMA. Evidemment, on exploite toujours les caractéristiques de l'oreille humaine pour augmenter le taux de compression

Un encodeur OGG Vorbis procède en trois phases distinctes pour finir le fichier OGG. Celles ci :

- L'analyse.
- Le codage.
- Le streaming.

L'**analyse** consiste en deux étapes. Les données audio sont d'abord séparée en petits segments contigus puis ces segments sont traités séparément : ils sont transformés en une représentation plus propice au codage. La **phase de codage** consiste en deux étapes également. Les segments transformés précédemmen,t sont analysés à l'aide d'un modèle probabiliste (codage de Huffman par exemple) et codés en nombre de bits le plus faible possible. Les mots ainsi obtenus étant de longueur différente (on rappelle qu'un codage de Huffman produit le mot le plus court pour l'occurrence la plus fréquente), l'étape de 'bitpacking' les arranges en séquences d'octets sans perdre de place. L'ensemble des octets nécessaires au codage d'un segment (obtenu à la 1ère étape) est un **Vorbis packet**. Enfin, la troisième phase, le **streaming** assemble les Vorbis packets en un flux de données appelé un **Ogg bitsream**. Une table de correspondance entre la structure logique des Ogg bittstreams et leurs combinaisons en un flux physique de données est établie. Celle ci doit permettre au player ou au décodeur de retrouver comment assembler les Ogg bitstreams entre eux pour délivrer un signal audio image de l'original.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

10.3. Son utilisation face à la concurrence

Le format OGG Vorbis est particulièrement bien adapté à la diffusion de musique en streaming de par sa structure intrinsèque. Dans sa forme actuelle, Ogg encode des sons numériques provenant de signaux sonores analogiques échantillonnés à une fréquence pouvant aller de 8kHz à 48 kHz, sur 16 bits ou plus et en polyphonie. Ainsi, le son Surround peut être compressé en OGG sans perdre ses caractéristiques multi-voies alors qu'en MP3 ou WMA, il obligatoirement ramené à la stéréo.

L'encodage OGG se fait bitrate constant ou variable. Dans le cas d'un bitrate constant, ce dernier peut se régler à une valeur comprise entre 16 kbits/s et 128 kbits/s. Théoriquement, l'encodeur actuel supporte un bitrate pouvant aller jusqu'à 512 kbits/s. Les versions futures offriront un bitrate pouvant descendre en dessous de 16 kbits/s.

Dans la page d'accueil consacrée a Vorbis, le site xiph.org (Xiphophorus company est le nom du groupe de projets OGG) revendique une qualité compétitive avec celle du format AAC et supérieure à celle des formats MP3, WMA et VQF.

Le format OGG, est 100% '**backwards**' **compatible**. Cela signifie qu'un fichier encodé en OGG Vorbis avec l'encodeur actuel est lisible avec un ancien player Ogg sans dégradation de la qualité.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

11. ETUDE COMPARATIVE

J'ai fais une petite étude comparative entre différents formats de compression. J'ai pris comme cobaye un morceau que j'apprécie et donc que je connais très bien. Il s 'agit de Sensitive de Slaï qui dure 4 minutes et 17 secondes.

Vous trouverez ci-dessous un tableau récapitulatif des tests effectués.

Format de compression(bitrate)	Temps de compression en seconde	Taille du fichier kilo octets	Logiciel utilisé
WAVE 44.1 kHz, 16 bits	15.22	44 040	Nero (version 5.5.6.4)
AIFF 44.1 kHz, 16 bits	17.80	44 040	Nero
MP3 20 kbits/s	28.09	631	Play Center Creative(version 3.00.14)
MP3 96 kbits/s	35.48	3023	Play center
MP3 160 kbits/s	38.17	5037	Play center
MP3 192 kbits/s	38.76	6045	Play center
MP3 320 kbits/s	40.23	10 073	Play center
MP3Pro 40 kbits/s	85.32	1259	Nero
MP3Pro 96 kbits/s	94.76	3021	Nero
WMA 96 kbits/s	41.81	3057	Windows Media Player(version 7)
WMA 160 kbits/s	42.29	5080	Windows Media Player
WMA 96 kbits/s	39.15	3056	Play Center
WMA 160 kbits/s	40.68	5079	Play Center
OGG 96 kbits/s Set processing high priority	236	2878	OGG Vorbis (version 1.6.11c)
OGG 160 kbits/s Set processing high priority	250	4820	OGG Vorbis
OGG 160 kbits/s Set processing normal priority	249	4820	OGG Vorbis

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

11.1.Analyse

◄ En général on constate que la taille d'un fichier issu d'une compression destructive est en moyenne 10 fois plus petite qu'un fichier qui est compressé en WAVE ou en AIFF (non destructif).

◄ Il n'y a aucune différence entre les fichiers WAVE et AIFF, si bien au niveau de la taille du fichier, qu'à l'écoute.

◄ En passant au signal compressé en MP3, avec un bitrate de 20 kbits/s, on a certes un fichier intéressant en taille, mais on a une dégradation importante du signal.

Avec un bitrate de 320 kbits/s, nous avons un fichier 16 fois plus grand, mais avec une qualité nettement meilleure.

En le comparant avec un fichier WAVE le son est quasiment le même à la différence que ce dernier à plus de dynamique, surtout à l'attaque d'un son (exemple à 0.25 secondes du morceau).

En conclusion, en ce qui concerne le MP3, plus on utilise un bitrate bas, plus la qualité du son se dégrade et plus on perd en dynamique, mais on y gagne beaucoup en terme de taille de fichier.

Quand je compresse un morceau en MP3, j'utilise un bitrate de 192 kbits/s. C'est le meilleur compromis que j'ai trouvé. Et c'est d'ailleurs ce type de compression que l'on utilise sur Internet.

◄ Entre un fichier WMA et MP3 de même bitrate (160 kbits/s), le MP3 est légèrement avantageux en taille, mais le WMA est de meilleure qualité. Le son a plus de 'pêche'.

Un fichier WMA codé avec deux logiciels différents reste identiquement le même.

◄ En comparant un fichier WMA, MP3 et MP3Pro de même bitrate (96 kbits/s), le MP3Pro prend l'avantage au niveau taille, dynamique, mais la restitution des aigus, ce n'est pas encore ça. Il par contre meilleure que son cousin le MP3 compressé en 96 kbits/s.

◄ J'ai comparé ces trois derniers formats avec le concurrent, l'OGG Vorbis toujours à 96 kbits/s. L'OGG Vorbis a remporté de loin tous les suffrages.

La restitution du signal compressé à 160kbits/s est impressionnante. Il difficile de le distingué de l'original. Mais encore une fois je trouve que la dynamique prime légèrement.

A savoir qu'un fichier OGG Vorbis compressé à 160 kbits/s est bien de meilleure qualité qu'un fichier MP3 compressé à 192 kbits/s.

A noter qu'il y a une différence réelle entre deux compressions OGG Vorbis, qui ont le même bitrate et qui utilise une qualité de processing différente.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

11.2. Conditions de l'étude

Tous les fichiers ont été compressé sur un PC utilisant un Pentium 3 cadencé à 500 MHZ avec une RAM (SDRam) de 256 Mo.

La carte « son » utilisé est une Sound Blaster Audigy Platinium.

Le graveur est un Ricoh 8*4*32.

J'ai utilisé le lecteur (player) **Winamp** (version 5.01) afin de lire les différents formats qui ont été compressés.

N'ayant pas un laboratoire scientifique et informatique sophistiqué, mes **oreilles** de technicien du son étaient donc les seules juges.

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr

12. Conclusion

Alors, que choisir ? Tout dépend de l'utilisation que l'on veut faire du signal compressé. Lorsque que l'on veut travailler avec un format numérique tel qui l'est utilisé dans le milieu de l'audio professionnel. Il est rigoureusement conseiller de travailler avec un format de compression non destructif.

Mais si l'on veut faire du stockage massif de donnée ou de la transmission de signaux via les canaux dont nous disposons actuellement (internet), il préférable travailler avec un format de compression destructif.

Libre a vous de choisir…

Copyright - Enseignement des Métiers de la Communication
Malakoff 92240 - Tous droits réservés.
Site web : http://www.emc.fr