

1. Introduction

Le son est une chose familière dans notre vie quotidienne que l'on en oublie souvent la signification physique qui est loin d'être facile à comprendre.

D'un point de vue physique, un son est une énergie qui se propage sous forme de vibrations dans un milieu compressible (dans l'eau, dans l'air, dans les matériaux solides).

Le traitement du son est la branche du traitement du signal qui s'applique aux signaux audio, dans le but notamment d'améliorer la qualité, de les compresser, ou d'extraire de l'information.

Le traitement du signal est la discipline qui développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. Parmi les types d'opérations possibles sur ces signaux, on peut dénoter le filtrage, la compression de données, la numérisation, le codage, le chiffrement et la transmission de données.

La science qui étudie les sons s'appelle l'acoustique. La psychoacoustique combine l'acoustique avec la physiologie et la psychologie, pour déterminer la manière dont les sons sont perçus et interprétés par le cerveau.

Dans ce chapitre, nous présentons les notions de base de traitement de son.

2. Définition de son

Le **son** est une onde produite par la vibration mécanique d'un support fluide ou solide et propagée grâce à l'élasticité du milieu environnant sous forme d'*ondes longitudinales*. Par extension physiologique, le son désigne la sensation auditive à laquelle cette vibration est susceptible de donner naissance.

Lorsqu'on jette une pierre dans l'eau, on peut facilement observer le phénomène de propagation des ondes à la surface:



Figure 1. 1 : la propagation des ondes à la surface [s1]

Lors de la diffusion d'un son dans un concert, c'est l'air qui permet sa transmission jusque nos oreilles. De même que l'exemple de l'eau illustrée ci-dessus, les molécules d'air transmettent l'énergie et son donc un support pour le son. [s1]

➤ **Emission, Propagation, Réception**

Pour qu'un son soit émis, une énergie doit avant tout mettre en mouvement un corps pour produire une vibration. Ainsi, le muscle du larynx, la chute d'un objet sur le sol, ou la tension électrique dans un haut-parleur, provoqueront l'énergie nécessaire pour produire cette vibration. Ensuite, pour que ce son puisse se propager, il faut un milieu élastique favorable à la transmission de la vibration. En créant des surpressions ou des dépressions, l'air permet la propagation de l'onde. Les matériaux solides ont aussi cette capacité de transmettre le son. Dans le vide par contre, aucun son ne peut se propager, car il n'y a aucun de support.

Enfin, pour être perçue, il doit y avoir un récepteur sensible. Chez l'homme, l'oreille possède une membrane (le tympan) capable de transmettre les informations de vibration en signaux nerveux jusqu'au cerveau, grâce au nerf auditif. De même, le microphone possède également une membrane permettant de transformer les déplacements de l'air en signaux électriques. [s3]



Figure 1.2 : l'émission, la propagation et la réception du son [s3]

3. Caractéristiques du son

Comme tout phénomène vibratoire, le son peut être analysé comme un signal qui varie dans le temps. Les caractéristiques essentielles sont l'amplitude et la fréquence.

3.1.L'amplitude

La première caractéristique d'un son est son amplitude. Appelée aussi intensité ou volume sonore, c'est l'expression de la pression de l'air qui se mesure en décibels (dB). 0 dB correspond au minimum que l'oreille humaine puisse percevoir (seuil d'audibilité). Attention, une augmentation de 3db multiplie la puissance par deux!

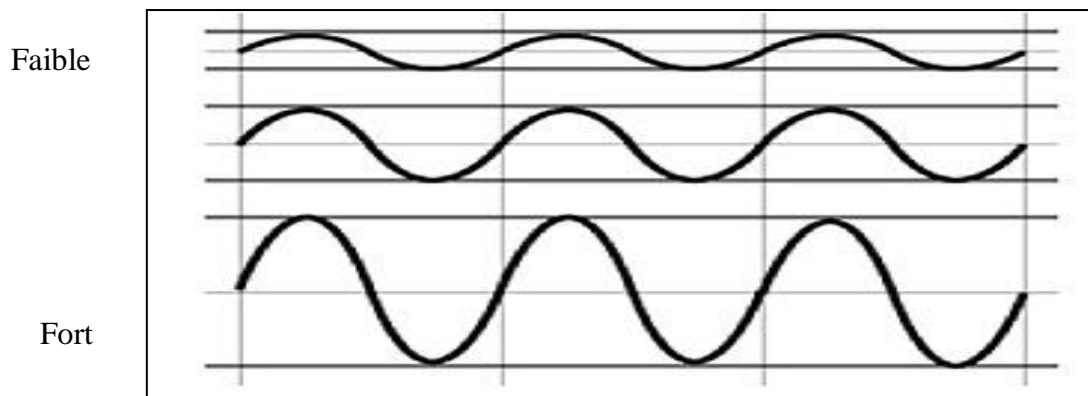


Figure 1.3 : L'évolution de l'amplitude sonore dans le temps. [s3]

Exemple concret:

- De 0 à 10 dB : Seuil d'audibilité, Désert
- De 30 à 40 dB : forêt
- De 60 à 70 dB : sonnerie de téléphone
- De 80 à 90 dB : tondeuse à gazon, klaxon de voiture
- 120 dB : seuil de la douleur, avion au décollage
- 180 dB : décollage de la fusée Ariane, lancement d'une roquette

3.2. Fréquence

La fréquence, exprimée en Hertz (Hz), est le nombre de répétition d'une période par seconde. Plus elle est élevée et plus le son paraîtra « aiguë », à l'inverse, il paraîtra « grave ». En musique, la fréquence définit donc la hauteur d'un son, soit, la note. (Ex: la note « LA » correspond à 440Hz, soit 440 vibration en une seconde). [s3]

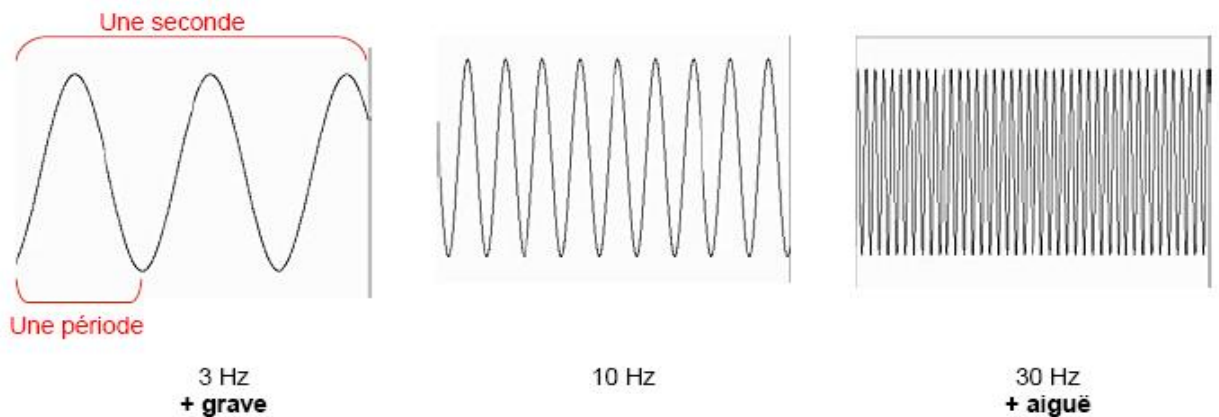


Figure 1.4: les types de fréquence [s3]

4. De l'analogique au numérique

4.1. Le son analogique : un signal continu

Lorsqu'on capte un son à partir d'un microphone, ce dernier transforme l'énergie mécanique (la pression de l'air exercée sur sa membrane), en une variation de tension électrique continue. Ce signal électrique dit « analogique » pourra ensuite être amplifié, et envoyé vers un haut parleur dont la fonction est inverse: transformer à nouveau le signal électrique en une énergie mécanique (on peut observer le déplacement de la membrane d'un haut parleur en marche).

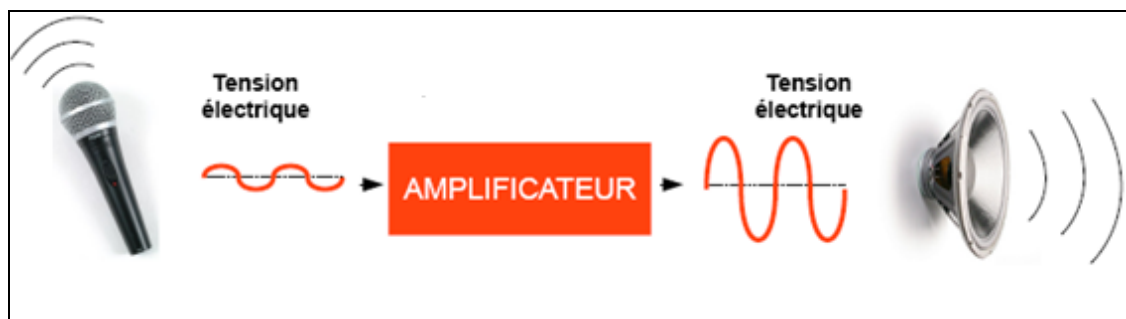


Figure 1.5: L'exemple d'une chaîne analogique. [s3]

Le son analogique est généralement fixé sur des supports comme les bandes magnétiques, disques vinyles etc. Le problème rencontré par ces supports réside dans l'usure physique des informations au cours de leur utilisation (lecture/écriture). A terme, le signal est affaibli et peut disparaître. [s3]

4.2. Le son en numérique: un signal discontinu

Avec l'informatique, lorsque ce même signal électrique est capturé à partir du micro, il est converti en une suite de nombre, on parle alors de numérisation du signal. C'est la carte son qui s'en charge, elle contient des entrées (convertisseurs analogique vers numérique) et des sorties (convertisseurs numérique vers analogique).

La première phase appelée numérisation consiste donc à passer d'un signal continu (une variation de tension électrique) en une suite de valeurs mesurées à intervalles réguliers, donc discontinu.

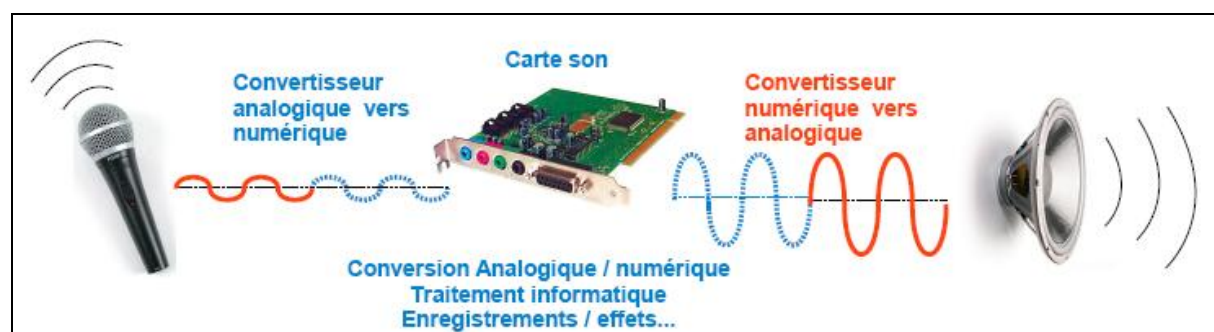


Figure 1.6 : Exemple d'une chaîne numérique [s3]

L'avantage du numérique, est la possibilité de lire et de dupliquer autant de fois ce signal sans aucune détérioration, puisqu'il a été réduit en une suite de nombres stockée dans un fichier informatique! Cela dit, la compression audio comme le MP3 peut provoquer une perte volontaire du signal afin d'économiser de l'espace de stockage. [s3]

◆ La fréquence d'échantillonnage

Échantillonner un signal audio analogique revient à prélever ses valeurs de tension électrique un certain nombre de fois par seconde. La fréquence de ces prélèvements est appelée fréquence d'échantillonnage. La fréquence d'échantillonnage est fixée avant l'opération de numérisation et ne varie pas pendant la numérisation.

Les fréquences d'échantillonnage couramment utilisées en audio sont 44100Hz et 48000Hz. Elles sont souvent imposées par des contraintes technologiques. Par exemple, la norme du disque compact audio (CD audio) impose une fréquence d'échantillonnage de 44100Hz.

L'échantillonnage est effectué par découpage temporel du signal audio analogique. Ce découpage temporel permet de reconstruire en données chiffrées la forme d'onde du signal numérisé. La numérisation ne repose que sur des séries de 0 et de 1 : il s'agit d'un codage binaire. [s4]

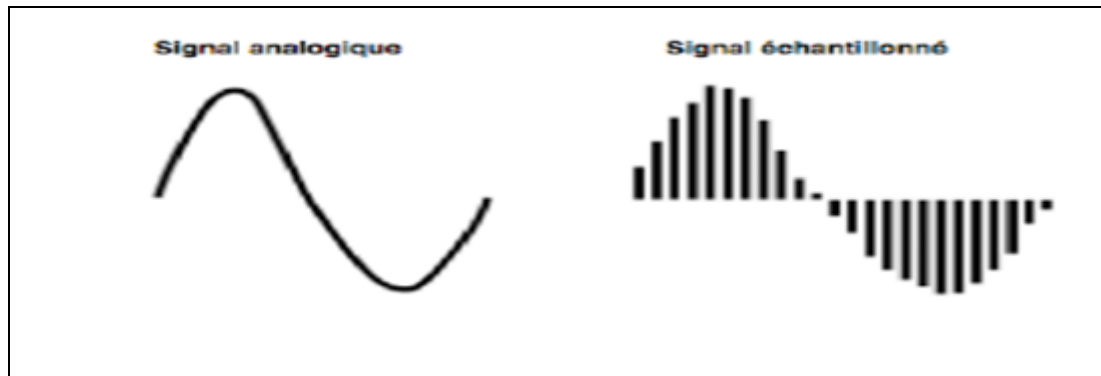


Figure 1.7: Echantillonnage d'un signal audio [s4]

◆ La quantification

Alors que l'échantillonnage opère un découpage temporel, l'opération de quantification crée une échelle de valeurs discrètes permettant d'attribuer à chaque échantillon une valeur d'amplitude. La quantification s'exprime en « bit » (un acronyme de binary digit). Les valeurs couramment utilisées en audio sont 16bit et 24bit.

L'amplitude de chaque échantillon doit impérativement prendre l'une des valeurs définies par l'échelle de quantification. Si la valeur d'amplitude de l'échantillon se situe entre deux paliers de l'échelle de quantification, elle est approximée au palier le plus proche. Cette approximation induit une erreur que l'on nomme « erreur de quantification ».

Par suite, plus le nombre de bits est élevé, plus le nombre de paliers est important et l'erreur de quantification faible. Autrement dit, les petites variations d'amplitude du signal échantillonné sont d'autant mieux approximées que la résolution de la quantification est élevée.

La fidélité de la forme d'onde numérisée à la forme d'onde du signal analogique dépend donc de la résolution (exprimée en bit) et de la fréquence d'échantillonnage (exprimée en kHz).

De même que pour la fréquence d'échantillonnage, le choix de la résolution de la quantification est soumis à des contraintes techniques. [s4]

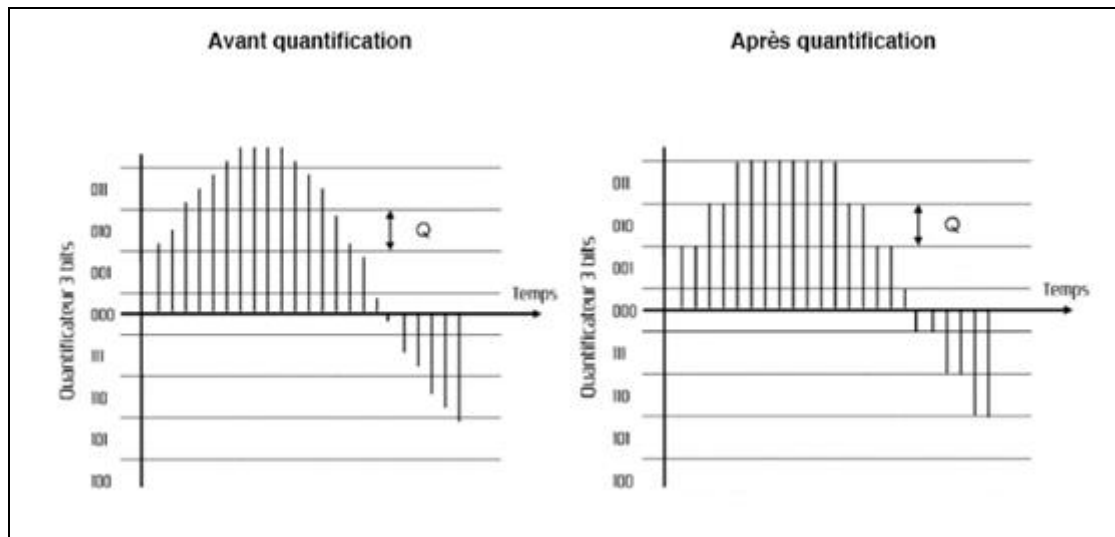


Figure 1.8: signal échantillonné avant et après quantification [s4]

❖ La compression du signal audio numérisé

Un signal audio numérisé est stocké sur des disques durs, des disques compacts, des DVD... La nature de l'information qu'ils contiennent rend ces fichiers relativement volumineux. L'intérêt de la compression de stockage limitée (ex : baladeur mp3).

Les techniques de réduction de débit sont déjà très largement employées dans les domaines du cinéma et de la radio, via le câble, le satellite. [s4]

✓ Les algorithmes de compression

Un algorithme est l'énoncé d'une suite d'opérations permettant de donner la réponse à un problème.

Dans le cas de la compression, l'algorithme a pour fonction de réduire la taille d'un fichier selon un certain nombre de contraintes que le programmeur spécifie. Par exemple, une des contraintes peut être de conserver toutes les fréquences inférieures à 20kHz afin de limiter les pertes de qualité sonore dans la zone audible du spectre.

Lors de l'étape de compression et de décompression d'un flux audio ou vidéo, on utilisera des algorithmes spécifiques rassemblés sous le terme commun de « CoDec ».

Un codec est constitué de deux éléments :

- le COdeur contient un algorithme destiné à coder l'information. Dans le cas de la compression ce sera pour effectuer une réduction du poids des données ;

• le DECodeur contient un algorithme destiné à décoder l'information. Dans le cas de la compression ce sera pour reconstruire un signal audionumérique. [s4]

✓ Le taux de compression

Compresser revient à réduire le débit du flux audio et/ou vidéo. Les algorithmes sont adaptés en fonction des applications (diffusion internet, télévision, cinéma) pour répondre aux besoins de chacun des médias. La réduction de débit (ou compression) s'exprime généralement sous la forme d'un taux dit « taux de compression ». Le taux de compression peut s'énoncer comme suit :

- soit comme le rapport entre le volume initial des données et le volume après réduction. Si le volume de données est deux fois plus faible après réduction (passant de 10Mo à 5Mo par exemple), on écrira qu'il s'agit d'un taux de 2:1 ;

- soit en pourcentage du volume après réduction par rapport au volume initial. Si le volume de données est deux fois plus faible après réduction, on écrira qu'il s'agit d'un taux de 50%.

Il existe par ailleurs deux types de compressions : la compression « destructive » et la compression « non destructive ». [s4]

5. Format de fichier audio

Un format de fichier audio est un format de données utilisé en informatique pour stocker des sons, (de la musique, des voix, etc.) sous forme numérique. De nombreux standards existent; certains s'appliquent à la production, au stockage et à la diffusion, d'autres (ceux qui utilisent des algorithmes de compression de données ou de débit), sont destinés, en principe, uniquement à la diffusion. Actuellement, le format le plus utilisé est de loin le mp3, suivi du wma, et de l'aac. [s2]

5.1. Caractéristique des formats audio

Les formats audio varient selon :

1. Le nombre de canaux sonores encodés.
2. Le nombre d'échantillons par seconde avec lequel on découpera numériquement, pour chaque canal, une onde sonore ou un signal électrique.
3. La résolution donnée à chaque échantillon et la grandeur physique qu'on lui donne.
4. l'application d'une compression ou non.

Chaque format audio présente aussi des caractéristiques découlant de l'algorithme de compression/décompression, ou codec (ou « codage-décodage » - COde-DECode en anglais), qu'il utilise ou non. Après la numérisation du son, le format utilisé est inscrit dans l'extension du fichier de données qui en stocke la transcription. Chaque format se caractérise aussi par sa propension à inclure et gérer des Métadonnées.

Dans un format donné, les fichiers peuvent être déclinés en plusieurs échelles de quantification (8, 16 ou 24 bits) avec différentes fréquences d'échantillonnage (p. ex. 22.05, 44.1, 48, 88.2 ; 96, 176.4, 192, kilohertz) appliqués à un certain nombre de voies (monophonique, stéréophonique).

Le nombre de canaux sonores peuvent être réels et séparés, ou mélangés discrètement aux signaux principaux; ils seront décodés et restitués par la suite à l'aide d'algorithmes spécifiques.

[s2]

5.2. Types de formats

5.2.1. Les formats audio compressés avec perte

La compression audio avec perte (*lossy*) se base sur des algorithmes spécialisés pour déterminer quelles transformations simplifient la représentation du son tout en étant perçue quasiment de la même manière par l'oreille humaine. Elle diminue la taille du fichier en éliminant les nuances perçues comme les moins utiles. L'élimination est définitive, créer un fichier dans un format de haute qualité à partir d'un fichier compressé avec perte ne sert strictement à rien.

Le format le plus connu est le MPEG-1/2 Audio Layer 3, dont le suffixe est .mp3. Ce format propose une qualité sonore très correcte pour un débit de 128 kbit/s. C'est ce format qui a été massivement utilisé pour transférer les musiques via internet dès la fin des années 1990. Rapidement, des baladeurs avec une mémoire réenregistrable et capables de lire directement ce format sont apparus.

Dans la décennie 2000, de nouveaux formats ont été proposés. Vu les progrès des algorithmes, ils surpassent largement le MP3 en termes de qualité à débit égal, et peuvent atteindre des qualités supérieures. De plus, certains sont moins contraignants que le MP3 quant aux droits d'utilisation (le Ogg est un format libre).

Mais le MP3 reste le plus utilisé, car l'arrivée en continu de nouveaux formats, apportant un avantage assez faible par rapport aux précédents, ne permet pas de mettre en place un standard meilleur que le MP3 et lisible par tous les baladeurs. En particulier, le fait que les iPod d'Apple ne lisent que le MP3 et l'AAC freine assez fortement les initiatives dans ce domaine.

Pour un même format de compression, il n'y a pas de manière unique de coder, car chaque algorithme cherche la meilleure manière de représenter le son d'origine suivant le langage de compression. En particulier, les codecs de MP3 ont réalisé des progrès très importants depuis le début de l'utilisation de ce format.

Elle permet typiquement un gain d'un facteur 10 de taille du fichier. Cela a rendu possible non seulement le stockage d'un temps d'écoute formidable sur les supports informatiques, mais aussi leur échange par Internet, souvent illégalement.

➤ AC3

Officialisée en 1992, la compression AC3 permet d'utiliser jusqu'à 6 canaux sonores indépendants avec un taux d'échantillonnage de 32, 44,1 ou 48 kHz et avec un taux de transfert allant de 32 à 640 kbit/s. Le Dolby Digital utilise ce principe de codage, c'est pourquoi on le désigne souvent sous ce nom. Format très courant dans les DVD.

➤ MP3

MP3 est l'abréviation de MPEG-1/2 Audio Layer 3. Cet algorithme de compression prend naissance en 1987. L'ISO en fera un standard dans les années 92-93. La couche (*Layer*) III est la couche la plus complexe. Elle est dédiée à des applications nécessitant des débits faibles (128 kbit/s) d'où une adhésion très rapide du monde Internet à ce format de compression. Les taux de compression (*ratio*) sont d'ordinaire de 1 pour 10 (1:10) (1:4 à 1:12). Très rapide à l'encodage. Des royalties importantes sont à payer pour exploiter la licence MP3.

Utiliser l'encodeur MP3 LAME dernière version, encodé à 130 kbit/s (V5) permet d'obtenir une *qualité comparable* au AAC (Advanced Audio Coding) encodé à 48 kbit/s.

Le suffixe des fichiers créés est **.mp3**

Type de compression : constant ou variable (VBR).

➤ **mp3PRO**

Le format mp3PRO, fruit de la collaboration entre Thomson Multimédia et l'Institut Fraunhofer, combine l'algorithme MP3 et un système améliorant la qualité des fichiers comprimés appelé **(en)SBR** pour *Spectral Bandwidth Replication*.

Ce format a été publié à la fin de 2001; un fichier MP3pro 64 kbit/s a une qualité équivalente à celle d'un MP3 à 128 kbit/s.

Le suffixe des fichiers créés est **.mp3**

➤ **Ogg Vorbis**

Le format Ogg Vorbis est un format libre, fruit de la fondation Xiph.org. Vorbis se différencie des MP3, WMA et autre AAC par son algorithme. Il segmente les sources audio en paquets successifs, l'algorithme de compression agissant dans un premier temps sur chaque paquet indépendamment des autres. Cela lui permet d'avoir très peu de faiblesses sur certaines fréquences et de conserver la même qualité quel que soit le type de musique.

Le suffixe des fichiers créés est **.ogg** ou parfois **.oga**.

Par abus de langage, on appelle 'fichier Ogg' des fichiers musicaux compressés par l'algorithme Vorbis. Ceci peut être particulièrement dérangeant à l'ère des baladeurs numériques supportant audio et vidéo. En fait, Ogg est un conteneur qui peut contenir des pistes sonores (Vorbis), audio sans perte (FLAC), audio parlées (Speex) et vidéo (Theora). Un 'fichier Ogg' peut donc contenir l'un ou l'autre (ou une combinaison) de pistes. Pour être plus clair, nous devrions parler de **fichier Ogg Vorbis** lorsque nous mentionnons un fichier. *Ogg* qui ne contient qu'une piste sonore au format Vorbis.

➤ **WMA**

Le format WMA (*Windows Media Audio*), créé par Microsoft à partir des recommandations MPEG-4 en 1999, est utilisé par le logiciel Windows Media Player. Ce format est lié à une gestion pointue des droits d'auteurs (Gestion numérique des droits, en anglais *Digital Right Management* ou DRM) qui permet de définir par exemple une durée de vie limitée pour les fichiers ou d'interdire les possibilités de gravure.

Il existe plusieurs versions du codec (wma7.1, wma9, wma pro).

Le suffixe des fichiers créés est **.wma**

➤ **AU**

Le format AU est assez bien répandu grâce à Unix et Linux. La fréquence d'échantillonnage est comprise entre 1 kHz et 200 kHz. Mais les applications de rendu audio ne lisent principalement que trois fréquences d'échantillonnage : 8012.821 (codec entré), 22050 et 44100 hertz.

Le suffixe des fichiers créés est **.au**

Les résolutions 8, 16, 20, 24 et 32 bits (flottant) sont acceptées.

5.2.2. Les formats audio compressés sans perte

La compression sans perte (*lossless*) signifie qu'on utilise un algorithme tel qu'on peut toujours retrouver les données d'origine. Dans l'absolu, il existe toujours un fichier d'origine tel que l'algorithme ne ferait pas gagner d'espace disque.

Typiquement, la compression sans perte permet de diviser la taille des fichiers par deux ou trois. Elle est relativement peu utilisée, car ce gain est très faible en comparaison de ceux permis par la compression avec perte (ce qui est un gros handicap pour les échanges de fichiers), et assez gourmande en temps de calcul. Aucun standard n'a donc suffisamment convaincu pour devenir universellement lisible.

➤ **ATRAC**

L'ATRAC (*Adaptive Transform Acoustic Coding*) est une technique de compression audio avec et sans pertes développée par Sony en 1992. Ce format a subi plusieurs évolutions : ATRAC3, ATRAC3plus (familièrement écrit ATRAC3+) et ATRAC Advanced Lossless se sont succédé respectivement en 1999, 2002 et 2006.

➤ **FLAC**

Le format FLAC (*Free Lossless Audio Codec*), est un format libre de compression audio sans perte. Maintenu par la fondation Xiph.org, il est apprécié pour conserver la qualité des fichiers sonores originaux en alternative aux formats de compression avec perte type MP3.

5.2.3. Les formats audio sans compression

Il existe un format audio non compressé, PCM, qui est généralement stocké sous forme de WAV sur Windows ou sous .Aiff sur Mac OS. WAV et AIFF sont des formats de fichiers flexibles conçus pour stocker plus ou moins n'importe quelle combinaison de taux d'échantillonnage ou de bitrates (bit rate). Ce sont les formats de fichier appropriés pour le stockage et la réalisation d'enregistrements originaux.

➤ **RAW**

RAW (*Real Audio Wrapper*) est un format audio utilisé pour représenter les données de son en modulation d'impulsion codée sans en-tête ni métadonnées.

➤ **WAV**

Le format WAV (ou WAVE),(WAVEform audio format) est une extension de fichiers audio, il s'agit d'un conteneur capable de recevoir des formats variés. Il est basé sur le format de fichier RIFF, lequel est semblable au format IFF.

Mono ou stéréo, il a été mis au point par Microsoft et IBM.

Le suffixe des fichiers créés est **.wav**

➤ **BWF**

Le BWF (*Broadcast Wave Format*) est un format audio standard créé par l'European Broadcasting Union en tant que successeur du WAV. Le BWF permet de stocker des métadonnées dans le fichier. Voir *European Broadcasting Union : Spécification du Broadcast Wave Format* (EBU Technical document 3285, juillet 1997). Il s'agit du format d'enregistrement usuel utilisé dans de nombreuses stations de travail audio professionnel de la télévision et du cinéma. Les Fichiers BWF incluent une référence standardisée Timestamp qui permet et facilite la synchronisation avec un élément d'image distincte.

L'AIFF est un format de stockage de sons sur les ordinateurs de Apple. C'est l'équivalent du format WAV dans le monde Windows.

Les résolutions 8, 16, 20, 24 et 32 bits (à virgule flottante) sont acceptées.

Le suffixe des fichiers créés est **.aif**

➤ **CAF**

Le CAF (*Core audio format*) a été développé par Apple pour s'affranchir des limitations de conteneur audio plus ancien comme le AIFF ou le WAV.

Il est compatible avec le système Mac OS X d'Apple depuis la version 10.3 et est lisible par Quicktime 7.

➤ **CDA**

Le CDA (*Compact Disc Audio*), est un format Microsoft spécifique à Windows, des pistes des CD audio, telles qu'elles apparaissent lorsqu'elles sont insérées dans le lecteur CD-ROM. Les CD audio du commerce répondent à la norme professionnelle "Red Book". La technique d'échantillonnage du son utilisée pour les disques compacts est la modulation d'impulsion codée (en anglais PCM, pour *Pulse Coded Modulation*).

Le suffixe des fichiers créés est **.cda [s2]**

6. Conclusion

Depuis la découverte de la synthèse numérique des sons, et avec l'arrivée d'ordinateurs personnels équipés en standard d'une carte son, il est devenu à la portée de tous d'enregistrer et de traiter les sons. De nombreux professionnels se tournent vers des solutions numériques, de moins en moins onéreuses, qui offrent, avec la progression de la capacité des ordinateurs, une foule de possibilités. Dans notre projet de fin d'étude, nous sommes intéressés par le cryptage des fichiers son non compressé représenté par le format Wave.