

RAVERDINO Julien
GUATTERI Jean Marc

Encadrement : Jean-Marie Munier

Techniques multimédia pour le son

Table des matières

I) INTRODUCTION À LA COMPRESSION	4
1) Compressions sans perte	4
2) Compression avec perte	4
a) Principes de numérisation	6
b) Conversion analogique-numérique	6
II) CODAGE ET APPLICATIONS	12
1) Le Codage MPEG	16
2) Le codage PCM	20

INTRODUCTION

A l'heure actuelle, si les espaces de stockage ne posent pas de réels problèmes, ce n'est pas le cas des flux de données sur les réseaux. En effet, les réseaux (Lan, Internet ...) sont de plus en plus utilisés pour échanger des informations. Cela pose cependant un problème important : celui du temps de transmission. Le débit des réseaux étant limité, on cherche à réduire la taille des données à envoyer afin de rendre la transmission plus rapide et par conséquent plus satisfaisante. Pour cela, il existe deux méthodes.

La première est dite « lossless », c'est à dire sans perte et elle est appliquée dans tous les domaines où la perte d'information est inacceptable : exécutable, textes ... Elle consiste à analyser des données et à en déduire une écriture de taille inférieure. L'une de ces méthodes est appelée compression de Huffman (inventée par le mathématicien du même nom). Nous allons l'étudier ci-après.

La seconde, qui engendre une perte ou une altération de l'information, est utilisée pour tout ce qui concerne l'image et le son. Elle consiste dans un premier temps à enlever toute l'information à laquelle les capteurs humains ne sont pas sensibles (fréquences inaudibles, surfaces insignifiantes ...) puis à écrire une nouvelle donnée qui ne pourra plus reprendre son format (qualité) original. Il y a eu suppression de l'information de manière irréversible. C'est cette forme de compression que nous détaillerons principalement ici.

Malgré tout, on est quand même en droit de se poser une question : La compression a-t-elle réellement un avenir, dans ce domaine qu'est l'informatique, lorsque le progrès nous inonde chaque jour de nouvelles technologies toujours plus performantes? En effet, si l'on revient ne serait-ce que dix ans en arrière, les disques dur atteignaient difficilement 500 ou 600 Mo tandis que les connections internet, aussi rares qu'onéreuses, étaient loin de permettre de manipuler de telles données. Dans 10 ans, il ne serait pas étonnant de trouver des disques dur de plusieurs milliers de Go, et des connections permettant de transmettre, via Internet, plusieurs Mo par seconde. La relative perte de qualité occasionnée par la compression pourrait être à l'origine d'un retour à une qualité supérieure quitte à perdre un peu de place.

I) Introduction à la compression

1) Compressions sans perte

La compression sans perte, bien qu'étant moins performante que sa concurrente destructive, n'en est pas moins aussi utile. En effet, certains types de fichiers, notamment les fichiers de données, ne peuvent se permettre de perdre le moindre bit de donnée sous peine de perdre l'intégralité de leur sens. Effectivement, une feuille de comptabilité donc un chiffre aurait disparu ne serait pas utilisable. C'est en cela que ce type de compression est toujours très utile. De plus, elle est aussi utilisée, en second rideau, dans les compressions de fichiers multimédia. Une fois la compression destructive terminée, pourquoi ne pas encore réduire la taille du fichier comprimé de 20 ou 30% ? Le principal algorithme de compression sans perte est l'algorithme de Huffman que nous allons détailler ici.

La méthode de compression Huffman consiste à diminuer au maximum le nombre de bits utilisés pour coder un fragment d'information. Prenons l'exemple d'un fichier de texte : Le fragment d'information sera un caractère ou une suite de caractères. Plus le fragment sera grand, plus les possibilités seront grandes et donc la mise en œuvre complexe à exécuter. L'algorithme de Huffman se base sur la fréquence d'apparition d'un fragment pour le coder : plus un fragment est fréquent, moins on utilisera de bits pour le coder. Dans notre exemple de fichier texte, si on considère que notre fragment est la taille d'un caractère, on peut remarquer que les voyelles sont beaucoup plus fréquentes que les consonnes : par exemple la lettre 'e' est largement plus fréquente que la lettre 'x' par conséquent la lettre 'e' sera peut-être codée sur 2 bits alors que la lettre 'x' en prendra 10.

Pour pouvoir compresser puis décompresser l'information, on va donc devoir utiliser une table de fréquences et deux choix s'offrent à nous : calculer une table et l'intégrer au fichier ou utiliser une table générique intégrée dans la fonction de compression. Dans le premier cas, la compression est meilleure puisqu'elle est adaptée au fichier à compresser, mais elle nécessite le calcul d'une table de fréquences et le fichier sera plus important également du fait de la présence de cette table dans le fichier. Dans le second cas, la compression sera plus rapide puisque elle n'aura pas à calculer les fréquences, par contre l'efficacité de la compression sera moindre et le gain obtenu par la première méthode (ratio de compression + taille de la table) peut être supérieur à celui de la deuxième (ratio de compression).

On va donc, par création de la table des fréquences, puis de l'arbre dit de Huffman, réussir à comprimer le fichier.

2) Compression avec perte

Les algorithmes de compression sans pertes permettent de réduire considérablement la taille mémoire occupée par les données. Cependant, il est encore possible d'améliorer ces taux de compression grâce à des algorithmes dit avec perte ou encore destructifs. Cela a d'autant plus d'intérêt que les images en haute résolution ou un son dont la définition est de qualité CD prennent énormément de place.

L'idée qui introduit cette nouvelle forme de compression vient des possibilités humaines. Effectivement, un œil humain ne fera pas la différence entre deux couleurs très proches parmi les 16 millions disponibles. Pourquoi dans ce cas, ne pas réduire le nombre de couleurs employées par l'image? C'est ce que font les formats de compression JPEG, PING et GIF, et, de ce fait, ils réduisent de manière radicale la taille des images.

De même, aucun homme normalement constitué ne peut entendre un son à la fréquence de 20 Khz. En supprimant ces informations inaudibles par le commun des mortels, on diminue significativement le nombre de données à prendre en compte et on obtient donc un fichier plus léger.

Poursuivant cette chasse aux informations inutiles, les psycho acousticiens ont établis qu'aucun être humain n'était capable de percevoir la stéréo dans les extrêmes graves. Autant donc passer toutes les fréquences trop basses en mono quitte à rétablir une pseudo stéréo au moment de la lecture du fichier. Et encore quelques octets de gagnés! Mais le plus malin reste à venir : les mêmes psycho acousticiens ont observé que certaines fréquences en écrasaient d'autres qui, du coup, devenaient superflues. Imaginons un enregistrement proposant la voix d'une petite fille sur le bord d'un circuit de formule1. Lors du passage à pleine vitesse des bolides, il y a fort à parier que l'on ne distinguera plus rien de ses paroles. Autant donc qu'elle ne parle plus.

Fort de ce constat, l'algorithme de compression repère donc les sons "dominants" et retire toutes les données relatives aux sons "dominés". Puisque de toutes façons, on ne les aurait pas entendus, cela ne fait guère de différence au niveau sonore tout en réduisant notablement la taille du fichier.

C'est sur ces principes que reposent de nombreux algorithmes de compression du son tels que le mp3 : ils allègent le fichier de toutes les informations qu'ils estiment superflues avant de le passer par un algorithme de compression classique (type Winzip). Parce qu'ils retirent de la matière sonore au fichier traité et dégradent ses qualités d'origine, on dit qu'ils opèrent une compression "destructive". Evidemment, plus on veut un petit fichier, plus la dégradation du son sera grande.

Voici un petit exemple du potentiel de compression de différents algorithmes :

Compression	Taille du fichier	Gain de place
Aucune	1273 ko	0 %
WinZip 7.0	1173 ko	8 %
WinRar 2.50	1154 ko	10 %
MP3 haute qualité	288 ko	77 %
MP3 basse qualité	58 ko	95 %

Voyons maintenant en détail les différentes étapes qui conduisent à une compression :

a) Principes de numérisation

Sous le nom général de "traitement du signal" sont regroupées des activités diverses telles que :

- Enregistrement, reproduction, transmission et filtrage du son musical, de la voix, de l'image fixe ou vidéo, reconnaissance vocale, etc.
- Correction d'images fixes ou vidéo - élimination d'artefacts - modifications colorimétriques - montages, etc.
- Traitement de signaux industriels : déparasitage, lissage, régulation, analyse spectrale etc.

Jusqu'aux années 1960-1970 le traitement des signaux se faisait par voie purement analogique grâce à des systèmes matériels électroniques. Cela imposait de nombreux inconvénients tels qu'un manque de fiabilité due à l'inévitable dérive des caractéristiques des composants, un coût élevé ou encore un manque de fiabilité.

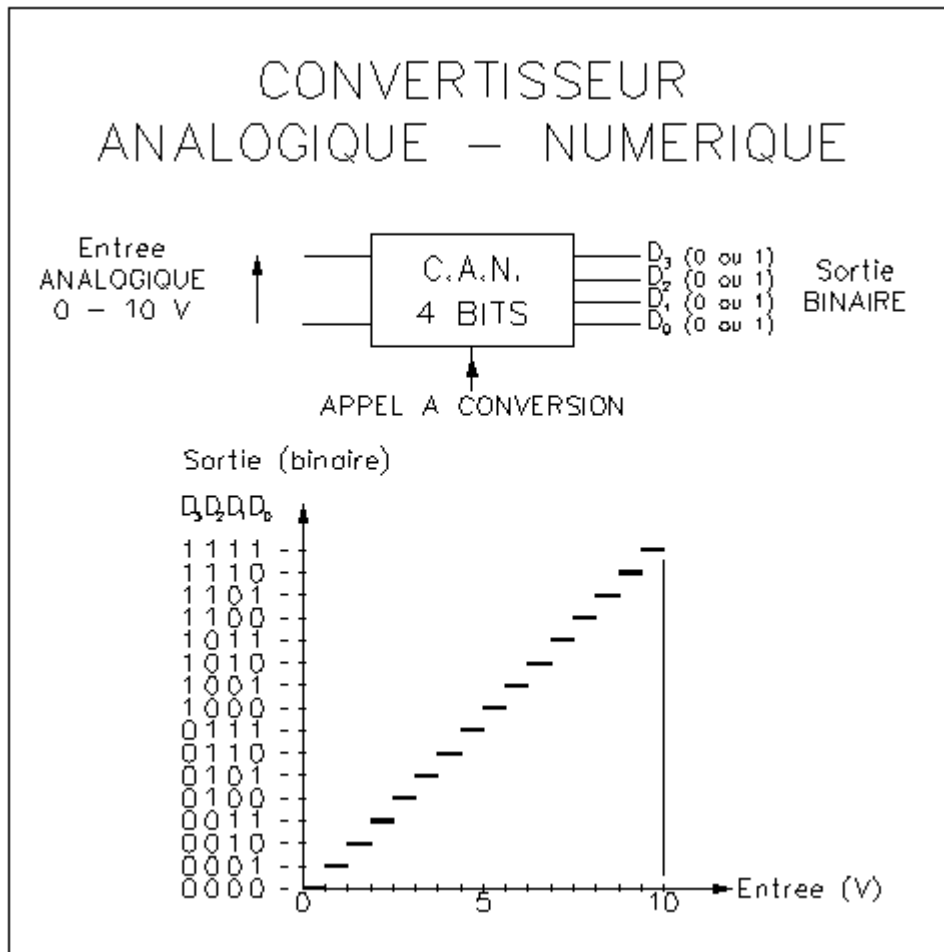
L'unité centrale, grâce à un programme approprié peut effectuer des calculs sur les valeurs instantanées d'un signal et en déduire les corrections souhaitées. On y transféra les compétences de la physique vers les mathématiques.

Etapes du traitement numérique :

- Transformer le signal analogique à traiter en un signal électrique proportionnel : c'est la capture ou l'acquisition du signal. Les instruments en sont divers : microphones pour les sons, tubes analyseurs d'images pour la vidéo, capteurs industriels etc. Cette étape est encore purement analogique. Elle se déroule en deux étapes : l'échantillonnage et la quantification.
- Convertir le signal électrique en une suite de valeurs numériques binaires, seules compréhensibles par les calculateurs numériques. C'est la conversion analogique-numérique.
- Lancer les différents algorithmes de codage et de compression.

b) Conversion analogique-numérique

La conversion est obtenue grâce à un circuit électronique intégré appelé Convertisseur Analogique Numérique. (CAN) .Celui-ci permet de transformer un signal sinusoïdal en un relevé de valeurs binaires afin de rendre utilisable les données par un outil informatique. Voici un petit schéma de ce que réalise un CAN.



Le paramètre primordial de cette opération est la résolution choisie. Celle-ci dépend du nombre de bits du convertisseur utilisé, et ce en fonction de l'utilisation que l'on souhaite en faire. Une conversion réalisée avec un convertisseur 8bits (celle utilisée dans la téléphonie en Europe) permettra de rendre 256 valeurs différentes, alors qu'une conversion 16bits (typiquement utilisé pour les cd audio) permettra de rendre 65536 valeurs. Nous constatons aisément que plus la résolution est élevée, plus les échelons sont nombreux pour une même étendue de valeurs extrêmes. On obtiendra donc des écarts entre deux valeurs successives d'autant plus faible et par conséquent, le signal initial sera rendu avec plus de précision.

Exemples :

- Résolution relative pour un convertisseur de 8 bits : $1 / 255 = 0,003921 = 0,3921 \%$
- Avec un convertisseur 16 bits : $1/65535 = 15,2590 \text{ E-6} \%$

Cette conversion se déroule en deux étapes :

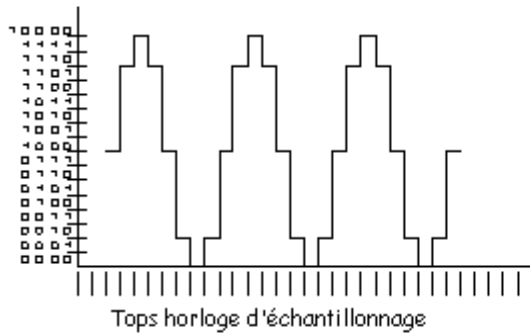
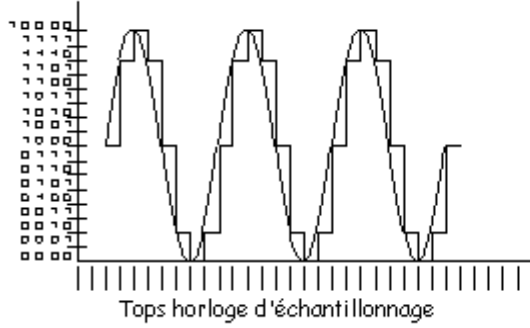
- L'échantillonnage qui consiste à effectuer un relevé de valeurs chacune séparé par un laps de temps très court.
- la quantification qui consiste à donner des paliers de valeurs pour les différents sons.

i) Echantillonnage :

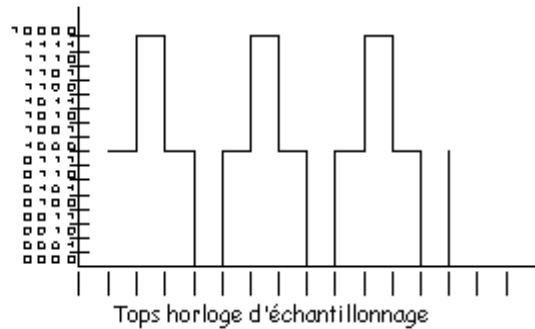
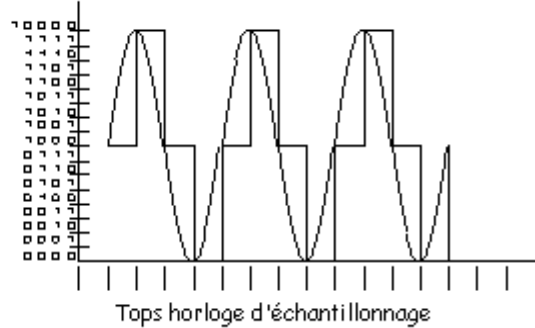
Si le signal à transmettre est une fonction du temps ; c'est entre autre le cas du son ou de l'image vidéo. Combien de fois par seconde devons-nous relever ses valeurs successives pour le restituer fidèlement ?

Nous comprenons bien que si les échantillons sont "rares" le signal analogique sera grossièrement traduit et donc grossièrement restitué : on le dira sous-numérisé. Il semble bien qu'il faudra un nombre "assez élevé" d'échantillonnages par seconde si l'on souhaite une "bonne restitution" par la suite. La figure ci-dessous présente le même signal sinusoïdal échantillonné 8, puis, 4, puis 3, puis 2 fois par période. Cette figure suggère intuitivement que la limite de sous-échantillonnage se situe à deux échantillonnages par période pour un signal sinusoïdal. Certes, la sinusoïde est devenue un signal rectangulaire de même période, mais il suffira d'en soustraire toutes les harmoniques en ne laissant subsister que la fondamentale pour récupérer le signal sinusoïdal initial. Cette limite inférieure pour la fréquence d'échantillonnage est mathématiquement, donc rigoureusement confirmée par le théorème l'échantillonnage énoncé plus bas.

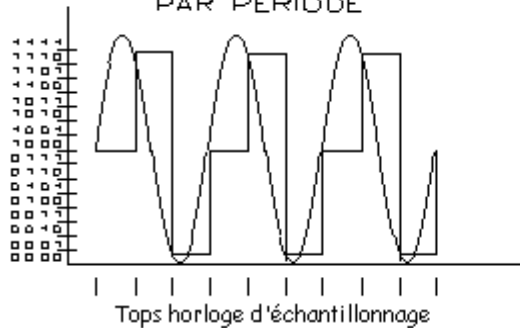
8 ECHANTILLONNAGES
PAR PERIODE



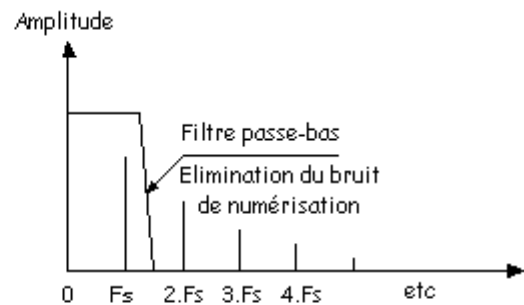
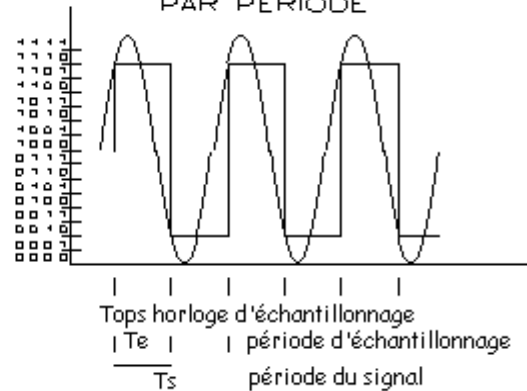
4 ECHANTILLONNAGES
PAR PERIODE



3 ECHANTILLONNAGES
PAR PERIODE



2 ECHANTILLONNAGES
PAR PERIODE



ii) Théorème de l'échantillonnage :

Le théorème de l'échantillonnage précise la fréquence minimale d'échantillonnage pour un signal sinusoïdal analogique de fréquence donnée :

La fréquence d'échantillonnage minimale requise pour pouvoir ensuite restituer un signal sinusoïdal est le double de la fréquence de ce signal.

Pour un signal quelconque, il suffira d'appliquer ce théorème à toutes ses composantes spectrales, qui sont par définition des sinusoïdes.

D'où le théorème suivant :

La fréquence d'échantillonnage minimale requise
pour pouvoir ensuite restituer un signal
est le double de la fréquence
de la plus haute des harmoniques de ce signal que l'on souhaite restituer.

Exemples :

Le son téléphonique est contenu dans la bande théorique maximale de 0 - 4 kHz.

L'harmonique la plus élevée a une fréquence de 4 kHz.

Si nous voulons restituer toutes ses harmoniques, il nous faudra donc prélever 8 000 échantillons par seconde.

(En fait, la bande passante pratique de la boucle terminale analogique d'abonné est de 300 Hz - 3,5 kHz.

Soit 3,2 kHz.)

La musique de qualité exige une bande passante de 20 Hz à 20 kHz.

L'échantillonnage se fera donc à 40 kHz.

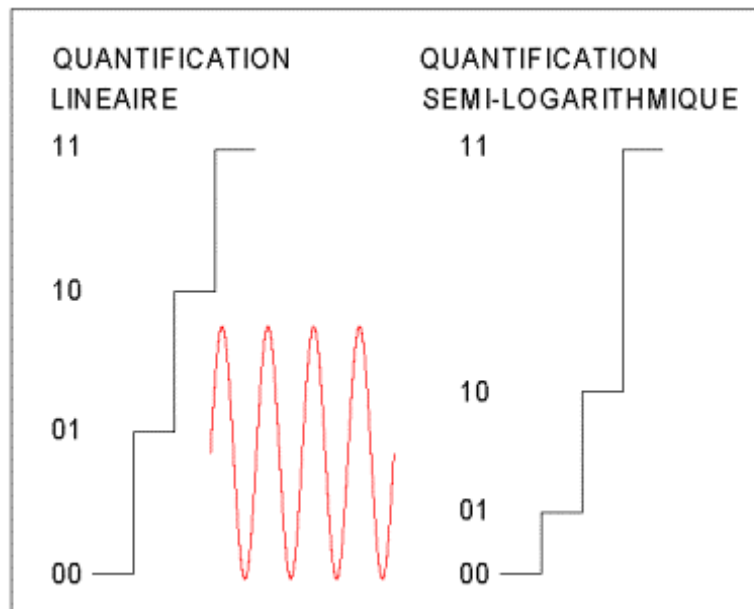
L'échantillonnage standard pour les CD est de 44,1 kHz (44 100 échantillons par seconde).

iii) Quantification :

Chaque échantillon représente une valeur proportionnelle à la valeur instantanée du signal sonore au moment de l'échantillonnage. La traduction binaire la plus simple consiste en une transposition linéaire. Par exemple, si la variable sonore à échantillonner est un signal électrique de 0 à 1 V., nous pourrions attribuer les valeurs binaires comme suit :

Signal échantillonné	Valeurs binaires (sur 2 bits pour simplifier)
1 V	1 1
0,666 V	1 0
0,333 V	0 1
0 V	0 0

On constate que les sons atteignant le maximum d'intensité sont rares et ponctuels. Il est donc avantageux de réserver aux sons moyens le maximum de bits de numérisation au détriment des éclats de voix dont le rendu n'est pas très intéressant. La figure ci-dessous compare très schématiquement deux lois de quantification : une linéaire, l'autre semi-logarithmique.

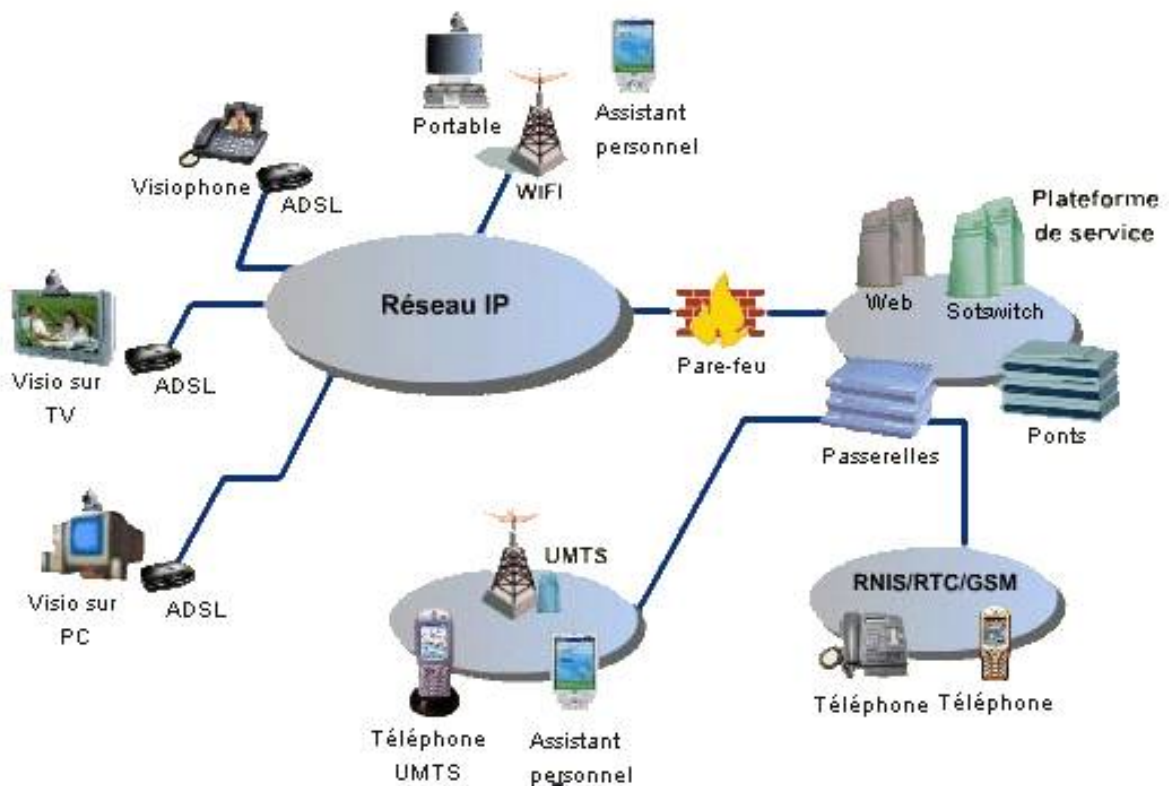


On observe que pour le signal d'intensité moyenne dessiné, la quantification semi-logarithmique attribue plus d'échelons que la quantification linéaire. Avec des convertisseurs 8 bits (typiquement utilisés dans la téléphonie), l'utilisation d'une l'échelle semi-logarithmique est très utile dans la mesure où le nombre de valeurs quantifiable est très faible. De cette manière, on obtient une plus grande quantité de valeurs différentes pour les sons de faible valeur. En revanche, pour les utilisations plus précises avec des convertisseurs 16bits (cd audio), une échelle linéaire conviendra parfaitement du fait que la différence entre chaque valeur possible est quasiment insignifiante. On utilisera donc un algorithme de codage différent selon le type d'utilisation que l'on fera du son ainsi traité. Voyons maintenant les principaux algorithmes de codage ainsi que des exemples d'utilisation les faisant intervenir.

II) Codage et applications

Ce mot est utilisé de manière très diverse (souvent à contresens). Dans la littérature technique, il englobe indifféremment toutes les méthodes de compression, les paramètres d'échantillonnage et la résolution...

Cet hiatus persiste pour les logiciels ou matériels procédant à la compression des sons et vidéos qui sont appelés des codecs et dont la traduction est tantôt "codeurs - décodeurs", tantôt "compresseurs - décompresseurs". En principe, le codage désigne le type de correspondance que l'on souhaite établir entre chaque valeur du signal analogique et le nombre binaire qui représentera cette valeur. On peut voir à partir du schéma ci-dessous que les algorithmes de compressions sont utilisés dans tous outils d'interconnexion du réseau mondial (du simple téléphone fixe au visiophone le plus perfectionné). Le monde entier est donc concerné par ce phénomène.



Pour choisir le type de codage adapté, plusieurs caractéristiques entrent en jeu :

i) la qualité du codeur :

En général, les techniques de codage offrant des faibles débits exigent des temps de traitement plus long, augmentant ainsi le délai de transit. Un critère de notation est établi en standard pour caractériser la qualité du codeur. C'est la note moyenne d'opinion MOS (Mean Opinion Score) qui classe les codeurs en cinq grandes classes :

- 1 = Mauvais,
- 2 = Médiocre,
- 3 = Moyen assez bon,
- 4 = Bon,
- 5 = Excellent.

ii) Retard :

Le délai est le temps écoulé entre l'émission de la parole et sa restitution à l'arrivée. Pour permettre un échange interactif, la voix doit être transmise avec des contraintes de délai. Les chiffres suivants (tirés de la recommandation UIT-T G114) sont donnés à titre indicatif pour préciser les classes de qualité et d'interactivité en fonction du délai de transmission dans une conversation téléphonique.

Classe n°	Délais par sens	Commentaires
1	0 à 150 ms	Acceptable pour la plupart des conversations.
2	150 à 300 ms	Acceptable pour les communications à faible interactivité.
3	300 à 700 ms	Deviens pratiquement inutilisable
4	Au-delà de 700 ms	Inutilisable sans une bonne pratique du half duplex

Les causes de délai sont nombreuses :

- retard dû au codage, au décodage et à la mise en paquets de la voix,
- retard dû à la sérialisation
- retard dans les files d'attente des routeurs
- retard de propagation
- retard dû à la compensation de gigue.

iii) Débit :

C'est la quantité d'information envoyé en Kbits/sec. Selon l'application, cette quantité doit être plus ou moins importante. Par exemple :

- le téléphone à la norme européenne (PCM) envoie 64Kbits/sec.
- Un cd audio délivre lui 1.41 Mbits/sec.

iv) Complexité de l'algorithme :

Cette complexité est exprimée en MIPS requis pour le DSP.

Voici un petit tableau illustrant ces différentes caractéristiques sur les algorithmes les plus utilisés :

Codeur	PCM (Temporel)	MICDA (Temporel)	RPE-LTP (Analyse/Synthèse)	CELP (Analyse/Synthèse)	LPC (Paramétrique)
Standard/home	G.711	G.726	ETSO GSM 06-10	DOD FS1016	DOD LPC10 FS1015
Débit	64 Kbit/s	32 Kbits/s	13 Kbit/s	4.8 Kbit/s	2.4 Kbit/s
Qualité (MOS)	4.2	4.0	3.6	3.5	2.3
Retard codeur + décodeur	125 ms	300 ms	50 ms	50 ms	50 ms
Complexité Mips	0.1	12.0	2.5	16	7.0

Prenons par exemple le son téléphonique ; les américains ont opté un codage sur 7 bits, les européens sur 8. La vitesse d'échantillonnage étant fixée à 8 000 échantillons par seconde, le flux numérique américain pour la parole téléphonique s'établit à 56 k bit/s, alors que l'Europe a adopté 64 k bit/s - bande d'un canal RNIS par exemple.

Les standards d'enregistrement sonore pour CD-ROM codent sur 16 bits, ce qui leur permet de différencier 65 535 échelons d'intensité sonore.

Voici un tableau présentant un petit panorama non exhaustif des principaux algorithmes de compression du son utilisés :

Codeur	Description
CCITT A-law CCITT u-law	Type de quantification et de compression audio recommandée par le CCITT et supportée par les applications téléphoniques sur le Web de Windows 95. Parmi d'autres applications, A-law et u-law ont été initialement développés comme standards de communication téléphonique.
Codecs TrueSpeech	Développé par la firme "DSP Group's TrueSpeech" ce codec échantillonne le signal à 8 kbit/s en monophonique avec une profondeur de codage équivalente à 1, ce qui réduit le flux à 1 ko/s. C'est, en volume et en bande passante, le 1/8 ème des résultats obtenus par les plus réduits des formats PCM audio : WAV ou AIFF. Où le 1/4 du plus réduit des formats ADPCM. En contrepartie, la qualité du son est proche de celle du son téléphonique d'abonné, c'est à dire : bonne pour la voix, médiocre pour la musique.

AD-PCM	<p>Adaptive Differential Pulse Code Modulation MIC-DA - Modulation par Impulsions Codées - Différentiel Adaptatif - Cette méthode de codage du son exige moins d'espace de stockage que les formats PCM utilisés dans les fichiers WAV et AIFF ou le format CD Audio. ADPCM de la firme IMA est utilisé dans les mini disques "Sony" pour faire tenir plus de données dans un petit espace. ADPCM de Microsoft est inclus dans les codecs utilisés par Windows 95 98.</p>
AIFF	<p>Audio Interchange File Format Format audio développé par Apple Computer pour l'enregistrement musical de haute qualité. Egalement utilisé par Silicon Graphics et d'autres professionnels du son. Peut être joué par toute une gamme d'utilitaires téléchargeables sur PC ou MAC.</p>
Lernout & Hauspie Speech Products	<p>Firme spécialisée dans la reconnaissance vocale, la conversion texte-parole, le codage digital de la voix et de la musique. Techniques Utilisées: SBC (sub-band coding), LPC (codebook excited linear predictive coding), harmonic coding. Ils couvrent une large gamme de rapports de compression liés à diverses qualités du son.</p>
WAV	<p>Format de fichier son non compressé codé PCM. Particulier à Windows.</p>
AU	<p>Format de fichiers son de Sun Microsystems pour Unix. Les clients http tels que Netscape Navigator lisent les fichiers d'extension *.au.</p>
MIDI	<p>musical instrument digital interface Standard permettant de relier un ordinateur de type PC ou MAC à des instruments de musique, des synthétiseurs pour réaliser des sons solistes ou orchestraux. Les fichiers MIDI ont pour extension *.mid et comportent uniquement des indications musicales : instruments - notes - sonie - tempos - etc. Le matériel et le logiciel contenu dans chaque élément d'un ensemble MIDI interprète ces informations de manière interactive par un échange de messages. MIDI est très largement utilisé dans les bandes sonores de jeux et studios d'enregistrement.</p>
VOC	<p>Format créé il y a longtemps par Creative Labs pour être utilisé dans les premières cartes Sound Blaster sous DOS. Format en perte de vitesse au profit de WAV.</p>
AVI	<p>Audio / Video Interleave. Format pour vidéos sous Windows: Video for Windows. Les deux autres formats sont MPEG et QuickTime. Dans un fichier AVI le son et l'image sont alternativement enregistrés. On peut "jouer" ces vidéos grâce à "Window's Media Player. inclus dans W95 et suivants.</p>
MOV	<p>Extension de fichier utilisée par QuickTime multimedia technology to run it. MOV files can be movie clips, such as Video for Windows' AVI files, or still images, such as GIFs.</p>
MPEG	<p>Moving Pictures Experts Group MPEG est un standard de compression du son et de vidéos permettant un téléchargement avec utilisation en temps réel à flot continu (streaming) sur Internet. MPEG-1 délivre vidéo et son à 150 ko/s - à la même vitesse qu'un lecteur de CD-ROM x1 Une des techniques consiste à gérer séparément les parties de l'image qui changent.</p>

	<p>MPEG-2 améliore grandement les choses en ramenant la qualité presque au niveau de celle des disques laser et du son CD-stéréo.</p> <p>MPEG-2 a été adopté par les services satellitaires et les DVD.</p> <p>MPEG-3 - plus connu sous le nom de MP3 - est le standard de compression de son assurant à la fois une excellente qualité de reproduction et un excellent taux de compression. Bande passante : 16 ko/s 128 kbit/s échantillonnée à 44,1 kHz.</p>
DSP	<p>Digital Signal Processor.</p> <p>Microprocesseur spécialement développé pour le traitement rapide du signal.</p> <p>Les DSP sont utilisés dans les cartes son, les modems, téléphones portables, pour réaliser des compressions décompressions en temps réel.</p>
Streaming	<p>Osons une traduction française : "flot continu"</p> <p>On utilise ce qualificatif pour indiquer que des données passent très rapidement d'un dispositif à un autre de telle manière que le dispositif récepteur puisse les utiliser en même temps qu'il les reçoit, sans attendre qu'elles soient toutes arrivées.</p> <p>C'est le cas des échanges de données son lors d'une conversation téléphonique interactive.</p> <p>Ce n'est pas le cas du transfert d'une vidéo QuickTime sur Internet.</p>

Nous allons maintenant étudier plus en détail les algorithmes les plus utilisés.

1) Le Codage MPEG

Sans compression, des signaux audio digitaux consistent typiquement en des échantillons 16 bits enregistré à un taux d'échantillonnage plus de deux fois plus important que les bandes passantes audio (44,1 kHz pour les CDS). Il faut alors plus de 200 Ko pour représenter juste une seconde de musique stéréo de qualité CD. En utilisant le codage audio MPEG on peut réduire les données du son original d'un CD d'un facteur 12 sans perte de qualité. Des facteurs de 24 et même plus maintiennent encore une qualité de son significativement meilleur que celle obtenue en réduisant simplement le taux d'échantillonnage et la résolution des échantillons. En général, cela est réalisé par des techniques de codage perceptuel adressant la perception d'ondes sonores par l'oreille humaine. En utilisant l'audio MPEG on peut atteindre une réduction des données de :

- 1:4 pour la couche 1, qui correspond à 384 kbps pour un signal stéréo
- 1:6 à 1:8 pour la couche 2, qui correspond à 256 à 192 kbps pour un signal stéréo
- 1:10 à 1:12 pour la couche 3, qui correspond à 128 à 112 kbps pour un signal stéréo

En maintenant toujours une qualité CD.

En exploitant des effets stéréos et en limitant la bande passante audio, le codage peut procurer un son de qualité acceptable à des taux encore plus bas. La couche 3 est la plus performante des techniques de la famille MPEG.

Voyons donc ce qu'il est possible de réaliser simplement à partir d'une norme de codage telle que le MPEG-4 :

Trois critères déterminent le type d'application :

Contraintes temporelles

Les applications sont classées selon qu'elles sont en temps réel ou non. Dans une application temps réel, l'information est simultanément acquise, analysée, transmise et potentiellement exploitée par l'utilisateur.

Symétrie des moyens de transmission

Les applications sont soit symétriques, soit asymétriques. Elles sont symétriques lorsque des moyens de transmission équivalents sont exploitables de part et d'autre de la communication.

Interactivité

Les applications peuvent être interactives ou non interactives. Elles sont interactives lorsque l'utilisateur a un contrôle individuel de la présentation, soit seulement au niveau du système de stockage des données, soit également sur les séquences du flux d'information selon les choix de l'utilisateur.

Ces trois critères mènent à huit classes possibles d'applications. Cependant, seules cinq d'entre elles s'avèrent réellement intéressantes.

Temps réel	symétrique	interactive	classe 1
		non interactive	sans application
	non symétrique	interactive	classe 2
		non interactive	sans application
non temps réel	symétrique	interactive	classe 3
		non interactive	sans application
	non symétrique	interactive	classe 4
		non interactive	classe 5

Classe 1

Visio téléphonie

Communication personne à personne sans limitations liées à l'environnement ou au réseau utilisé. Cela peut également inclure le transfert éventuel de données, pour des documents auxiliaires par exemple.

Vidéotéléphonie multiple

Communication impliquant plus de deux personnes, chacune en un lieu différent. Les fonctionnalités de contrôle peuvent être soit propre au réseau soit attaché à un terminal.

Vidéoconférence

Communication impliquant plus d'une personne par site.

Travail de groupe

Implique simultanément l'interaction entre des personnes (vidéotéléphonie) et des moyens de communication de données au moins aussi importantes que l'information audiovisuelle.

Classe distante symétrique

Cette application ressemble à la téléphonie et à la visioconférence. Elle diffère de la téléphonie dans le fait que le contenu audiovisuel peut inclure des scènes sans animateur et il y a en général un point de contrôle central fixe qui envoie la scène à tout le monde et sélectionne la réception d'une scène venant de quelqu'un.

Expertise distante symétrique

Des experts sont consultés à partir d'un site distant.

Classe 2

Expertise distante asymétrique

Les experts sont consultés à partir d'un site distant. Il semble raisonnable de s'attendre à ce que nombre de ces communications soient asymétriques avec un canal audiovisuel et un canal exclusivement audio (les experts montrent des documents pour illustrer).

Contrôle distant et télésurveillance

Les données audiovisuelles sont collectées en temps réel à partir d'un site distant pour une utilisation privée, typiquement par une communication homme machine. Il y a en général un canal audio-visuel à partir du site distant et un canal audio et/ou un canal de contrôle dans l'autre sens puisque ces applications incluent en général le contrôle distant par caméras et/ou microphones.

Collecte d'informations

Les informations sont collectées dans des sites distants où il est difficile d'établir une bonne qualité de connection en un délai court. L'interaction est en général assez faible, typiquement limitée à un canal de retour audio.

Classe distante asymétrique

La situation la plus courante est lorsqu'il y a un point central qui envoie des informations audiovisuelles et des postes de réception, en général moins onéreux, qui ont comme seule possibilité de participer par un canal audio.

Classe 3

Messages multimédia

Des messages contenant texte, audio, vidéo et graphiques sont envoyés à travers tout réseau ou groupe de réseaux à une boîte aux lettres située dans le réseau ou à la destination locale. L'E-mail en est une application courante.

Classe 4

Récupération de bases d'informations

Des informations audiovisuelles sont récupérées à partir d'une base d'information selon des critères personnels. La récupération à distance implique des situations de transmission plus ou moins complexes. Cela offre de nombreuses applications comme la vidéo à la demande, téléshopping, tourisme, encyclopédies, presse électronique, etc...

Jeux

Des jeux interactifs audiovisuels tournent à partir d'un serveur distant ou avec d'autres personnes via un serveur distant. Un canal retour envoie au serveur les réactions du joueur.

Classe 5

Présentations multimédia

Présentations multimédia locales ou distantes où aucune interactivité n'est possible. L'utilisateur n'a aucun contrôle sur les séquences du flux d'information.

Radiodiffusion multimédia pour portables ou récepteurs mobiles

Radiodiffusion de programmes multimédia (basse résolution) pour portables et récepteurs mobiles.

L'utilisation sans doute la plus connue et la plus utilisée de la norme de la norme MPEG est sans doute le MP3 (MPEG - Layer 3). Voici une présentation détaillée de cette méthode de compression :

a) Présentation du format MP3

Le MP3 "MPEG Audio layer 3" est un format de compression de données audio par destruction de données audio développée par l'organisation de standardisation internationale (ISO). Ce format permet de compresser à un taux de 1:12 les formats audio habituels (WAV ou CD audio). Il permet de faire tenir l'équivalent en fichiers de douze albums de musique sur un seul CD-ROM. De plus, le format mp3 n'altère que faiblement le son pour l'oreille humaine.

b) Le contexte

En fait la compression MPEG layer 3 consiste à retirer des données audio les fréquences inaudibles pour l'auditeur moyen dans des conditions habituelles d'écoute. La compression vise donc à analyser les composantes spectrométriques d'un signal audio, et de leur appliquer un modèle psycho-acoustique pour ne conserver que les sons "audibles". L'oreille humaine est capable de discerner, en moyenne, des sons entre 0.02kHz et 20kHz, sachant que sa sensibilité est maximale pour des fréquences entre 2 et 5kHz (la voix humaine est entre 0.5 et 2kHz), suivant une courbe donnée par la loi de Fletcher et Munson. La compression consiste à déterminer les sons que nous n'entendons pas et à les supprimer, il s'agit donc d'une compression destructive, c'est-à-dire avec une perte d'information. Nous allons voir les différents procédés de la compression.

i) L'effet de masque

Cette technique consiste à ne pas enregistrer tous les sons, c'est-à-dire de ne pas tenir compte des sons inaudibles par l'homme car ils sont camouflés par des sons de plus fort volume sonore.

ii) Le réservoir de bytes

Souvent, certains passages d'une musique ne peuvent pas être encodés sans altérer la qualité. Le mp3 utilise donc un petit réservoir de bytes qui agit en utilisant des passages qui peuvent être encodés à un taux inférieur au reste des données.

iii) Le joint Stéréo

Dans beaucoup de chaînes hi-fi, il y a un boomer unique (qui produit les basses). Cependant on n'a pas l'impression que le son vient de ce boomer mais plutôt des haut-parleurs satellites. En effet, en dessous d'une fréquence donnée l'oreille humaine est incapable de localiser l'origine du son. Le format mp3 peut exploiter (en option) cette astuce en utilisant la méthode du joint stéréo. C'est-à-dire que certaines fréquences sont enregistrées en mono mais elles sont accompagnées d'informations complémentaires afin de restituer un minimum d'effet spatial.

iv) Le code de Huffman

La technique de l'algorithme Huffman est un algorithme de codage (et non de compression), qui agit à la fin de la compression, en créant des codes de longueurs variables sur un grand nombre de bits. Les codes ont l'avantage d'avoir un préfixe unique, ils peuvent toutefois être décodés correctement malgré leur longueur variable, et rapidement grâce à une correspondance de tables. Ce type d'encodage permet de gagner en moyenne un peu moins de 20% d'espace.

Lorsque les sons sont "purs" (lorsqu'il n'y a pas de masquage) l'algorithme Huffman est très efficace car le son digitalisé contient de nombreux sons redondants.

Voici quelques résultats de compressions mp3 avec différentes qualités de compression :

Bande passante	Mode	Débit	Qualité	Compression
11.025	Mono	8 Kbps	Mauvaise	200: 1
22.050	Stéré o	64 Kbps	Acceptable	25: 1
44.100	Stéré o	96 Kbps	Acceptable	16: 1
44.100	Stéré o	128 Kbps	Bonne	12: 1
44.100	Stéré o	196 Kbps	Très bonne	12: 1

On voit que le mp3 nous permet d'obtenir des taux de compression incroyablement élevés (jusqu'à 200 fois !). Cependant, la baisse de qualité est rapidement décelable pour une personne spécialisée. Ainsi, on est en droit de se demander s'il est toujours nécessaire de se priver d'un son de qualité pour gagner quelques Mo.

2) Le codage PCM

Le codage PCM (ou MIC, Modulation par Impulsions et Codage) est utilisé par le réseau numérique à intégration de services (RNIS ou ISDN, Integrated Services Digital Network). PCM est basé sur un échantillonnage préalable, suivi d'une quantification non uniforme privilégiant les amplitudes faibles relativement aux grandes amplitudes. Le but final est de rendre le rapport signal-sur-bruit de quantification pratiquement constant sur une dynamique de 40 dB. Ce rapport devrait en principe être meilleur que 38 dB sur l'intégralité de la dynamique. On peut montrer, mathématiquement, qu'un codage selon une loi uniforme entraîne une quantification à 12 bits (ou 4096 niveaux). En relation avec la fréquence d'échantillonnage nécessaire, qui est de 8 kHz, ceci entraîne un débit de 96 kbit/s pour une simple conversation téléphonique; ceci est jugé excessif. Le codage non uniforme utilisé par PCM permet de réduire ce débit à 64 kbit/s, et corollairement, se contente de 8 bit (ou 256 niveaux) pour le codage.

La loi A (norme européenne CEPT) implémente une caractéristique logarithmique par une suite de segments linéaires. Le rapport signal-sur-bruit ne sera pas tout à fait constant dans la plage où l'on désirerait qu'il le soit.

D'un segment à l'autre, le taux de compression varie d'un facteur 2, ce qui représente également un bit de moins pour le codage, ou des pas de quantification deux fois plus grossiers. Le rapport

signal-sur-bruit de quantification se dégrade par conséquent de 6 dB lorsque l'on passe d'un segment au segment immédiatement suivant.

La discontinuité du rapport signal-sur-bruit lors de la transition d'un segment à un autre ne représente pas une gêne pour l'auditeur, ainsi qu'il est possible de s'en rendre compte par le biais des échantillons de parole.

La modulation :

La modulation PCM, ou MIC (Modulation par Impulsions et Codage) est la première modulation numérique appliquée à grande échelle, dans le cas de la téléphonie, essentiellement. Son domaine d'applications est si vaste que chaque coup de téléphone (ou presque) en Europe Occidentale utilise cette modulation pour aboutir.

Le but de la modulation PCM est de transmettre aussi fidèlement que possible un signal, indépendamment de sa signification. Dans cet ordre d'idées, le réseau RNIS (Réseau Numérique à Intégration de Services, ISDN pour les anglo-saxons, ou Integrated Services Digital Network), qui utilise PCM comme modulation pour la transmission de parole, offre à chaque utilisateur RNIS un débit brut de 64 kbit/s, qui correspond à une fréquence d'échantillonnage de 8 kHz, quantifié avec 8 bits par échantillon.

Dans le cadre de RNIS, PCM représente la technique permettant de transmettre la voix humaine sur le réseau; RNIS de son côté est indifférent (ou presque) à la nature de l'information transmise: il est en effet difficile de différencier des bits émis par un modulateur (et corollairement, générés par un être humain!) de ceux émis par un ordinateur.

La loi de quantification est non uniforme, avec un codage sur 8 bits. Cette quantification permet d'obtenir un rapport signal-sur-bruit de quantification équivalent à celui obtenu avec une quantification à 12 bits uniformes pour les basses amplitudes.

Ce rapport est pratiquement constant sur une plage de 40 dB au-dessous du niveau nominal du signal.

L'Union Internationale des Télécommunications a défini deux lois de quantification:

- La loi μ , utilisée aux Etats-Unis et au Japon.
- La loi A, utilisée en Europe.

La loi A est postérieure à la loi μ , et est légèrement plus facile à implémenter. La loi μ en revanche est légèrement meilleure, au niveau du rapport signal-sur-bruit pour les faibles amplitudes. Ces deux lois sont incompatibles: il n'est pas possible de coder selon la loi A et de décoder selon la loi μ . Ceci interdit toute compatibilité directe entre les systèmes européens et américains.

La téléphonie étant en pleine évolution, notamment avec la technologie UMTS. S'il est déjà possible de faire des vidéoconférences, Les nouveaux systèmes sont prévus pour accepter la multi vidéoconférence (comme le montre l'appareil ci-dessous). Il est donc évident qu'une amélioration des techniques de compression ne pourra qu'améliorer ce type de technologie.



CONCLUSION

Dans ce domaine qu'est la compression, deux avis seront toujours en contradiction. Du point de vue du musicien, on voudra privilégier la qualité quitte à perdre un peu de performance ou de mémoire. En effet, pourquoi se donner du mal à réaliser des enregistrements de grande qualité si c'est pour détruire ce travail d'orfèvre avec des compressions destructives ?

A l'inverse, les scientifiques privilégieront toujours l'avancée technologique plutôt que la qualité.

La compression du son est donc massivement utilisée dans tous les domaines. Ainsi nous avons à notre disposition de très nombreuses techniques, chacune optimisée pour une application bien précise. Si dans certains cas on peut se poser la question de son intérêt dans le futur, il paraît évident que ces techniques vont encore progresser et toujours nous permettre d'améliorer les réseaux. En gagnant des octets par la compression, on se donne les moyens de rajouter toujours plus de fonctionnalités et ainsi d'offrir de nouvelles perspectives aux utilisateurs.

BIBLIOGRAPHIE

Le son et les limites de l'audition :

<http://perso.wanadoo.fr/f5zv/RADIO/RM/RM11/RM11a01.html>

Audiofanzine :

<http://fr.audiofanzine.com/apprendre/dossiers/index,idossier,3.html>

http://fr.audiofanzine.com/apprendre/dossiers/index,idossier,31_page,1.html

Cours d'une université américaine :

<http://www.cs.sfu.ca/CourseCentral/365/li/material/notes/Chap4/Chap4.4/Chap4.4.html>

MP3 la révolution :

<http://www.figer.com/publications/mp3.htm>

Travaille sur la compression en générale :

<http://theangelmax.free.fr/compression/principale.html>

Page de Jean Marie Mounier :

<http://deptinfo.unice.fr/~jmunier/>