par S. Lamrous

Son :
Du codage à la compression

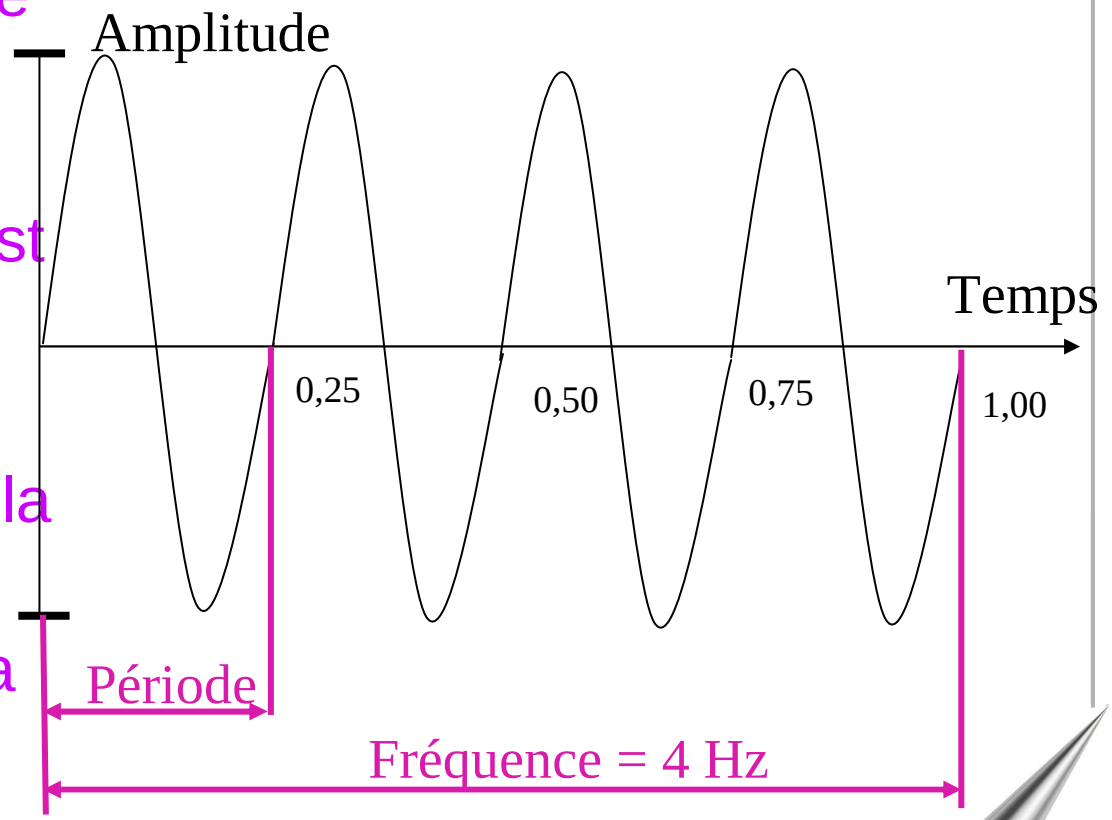
Transmission d'un son

- La vibration de l'air constitue un phénomène périodique car elle se reproduit identiquement à elle-même au bout d'un certain temps appelé *période*.
- La période (T), exprimée en secondes, désigne le temps nécessaire à un cycle vibratoire complet, c'est-à-dire à une *oscillation*
- *La fréquence* (f) définit le nombre de périodes par unité de temps : il s'agit de cycles par secondes ou Hertz (Hz)

Période, Fréquence

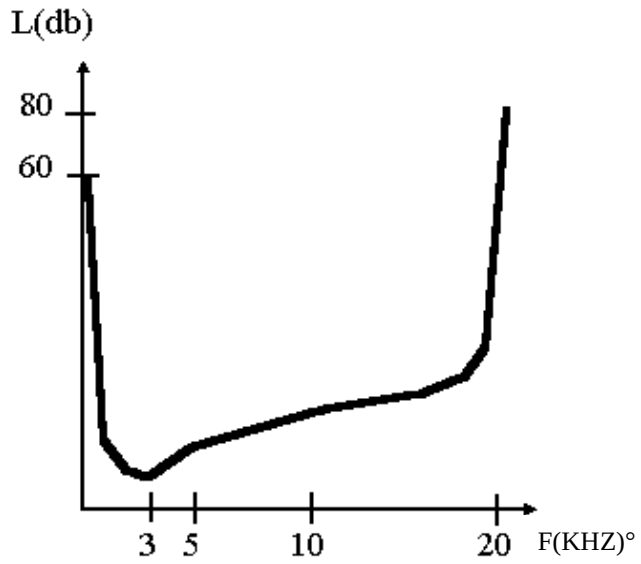
- Caractéristiques temporelles de la vibration génératrice du son
- Exemple : mouvement périodique dont la fréquence est de 4Hz
- La Période est 0,25 s
- Plus la période est petite, plus la fréquence est grande
- La fréquence est l'inverse de la période :

$$f=1/T$$



Amplitude et Audibilité

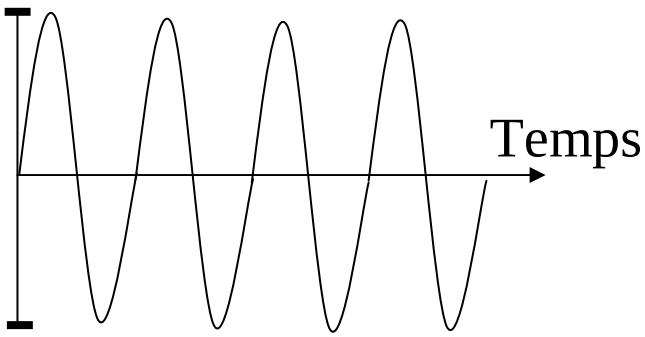
- L'amplitude est la grandeur de la vibration
- Si par exemple une source émettait une vibration de fréquence 4Hz, celle-ci serait inaudible quelle que soit son amplitude car on considère généralement que la perception auditive humaine s'étale dans le meilleur des cas, entre 20 et 20 000Hz
- Cette fourchette 20-20000Hz définit la bande de fréquences ou bande passante utile qu'un système de reproduction sonore de haute qualité doit s'efforcer de reproduire le plus linéairement possible



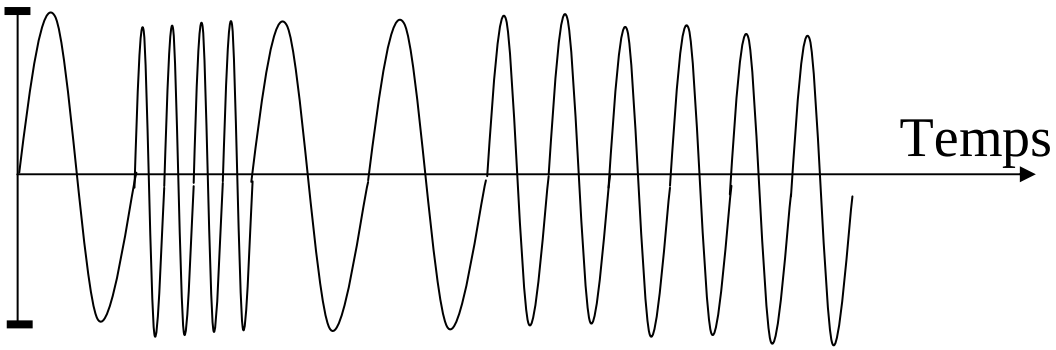
Exemples de bandes passantes

Système	Bande passante (Hz)
Téléphone	300 - 4 000
Radio AM	50 - 6 000
Radio FM	50 - 15 000
Magnétophone K7	40 - 15 000
CD audio et matériel professionnel	20 - 20 000

Signal pur, signal complexe



Son pur



Son complexe

- Un son pur est défini par une seule fréquence
- Les sons que l'on rencontre généralement ne sont pas purs, constitués de plusieurs fréquences, ils sont dits complexes
- Dans tout son complexe, à la fréquence *fondamentale* qui définit la hauteur sonore viennent s'ajouter une série de signaux dits *harmoniques.*

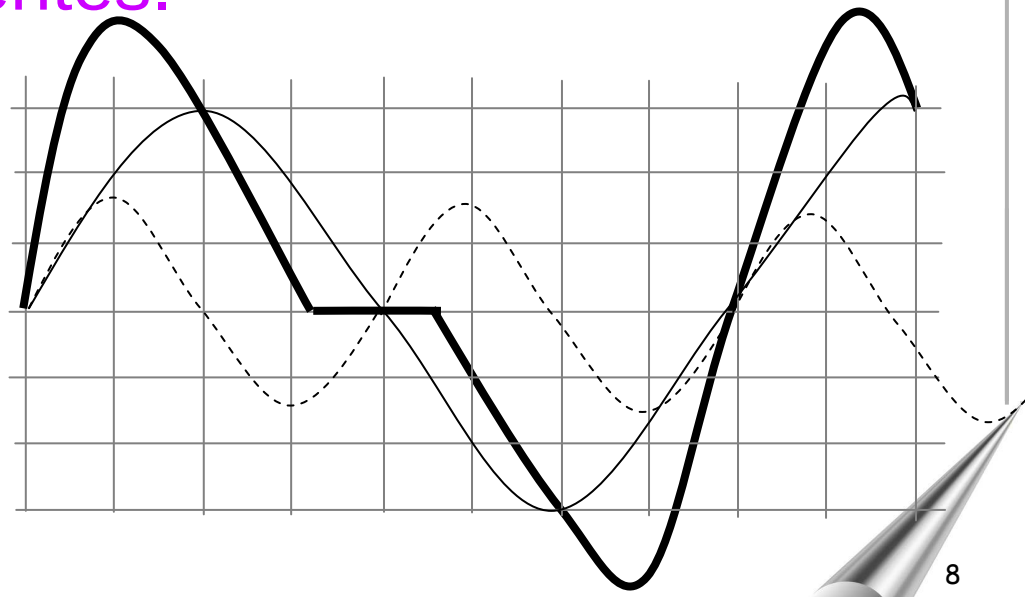
Signal fondamental, signal harmonique (1)

- Fourier a démontré mathématiquement que tout phénomène périodique pouvait être décomposé en une série de sinusoïdes élémentaires dont les fréquences sont des multiples entiers de la fréquence la plus grave dite fondamentale
- Si par exemple nous frappons la touche d'un piano correspondant au « la », nous obtenons un son complexe constitué par l'addition à la fréquence fondamentale (440Hz) –et à des amplitudes variables- des fréquences dites harmoniques à des valeurs double, triple, quadruple, etc. de cette fondamentale (harmonique 2 à 880Hz, harmonique 3 à 1320 Hz, 4 à 1760 Hz, etc.)

Signal fondamental, signal harmonique (2)

- Le nombre et l'amplitude des diverses fréquences harmoniques sont spécifiques à chaque source sonore et déterminent le *timbre* propre de cette source.
- Les harmoniques caractéristiques de chaque source qui permettent à l'oreille et au cerveau humains de distinguer des sources sonores différentes.

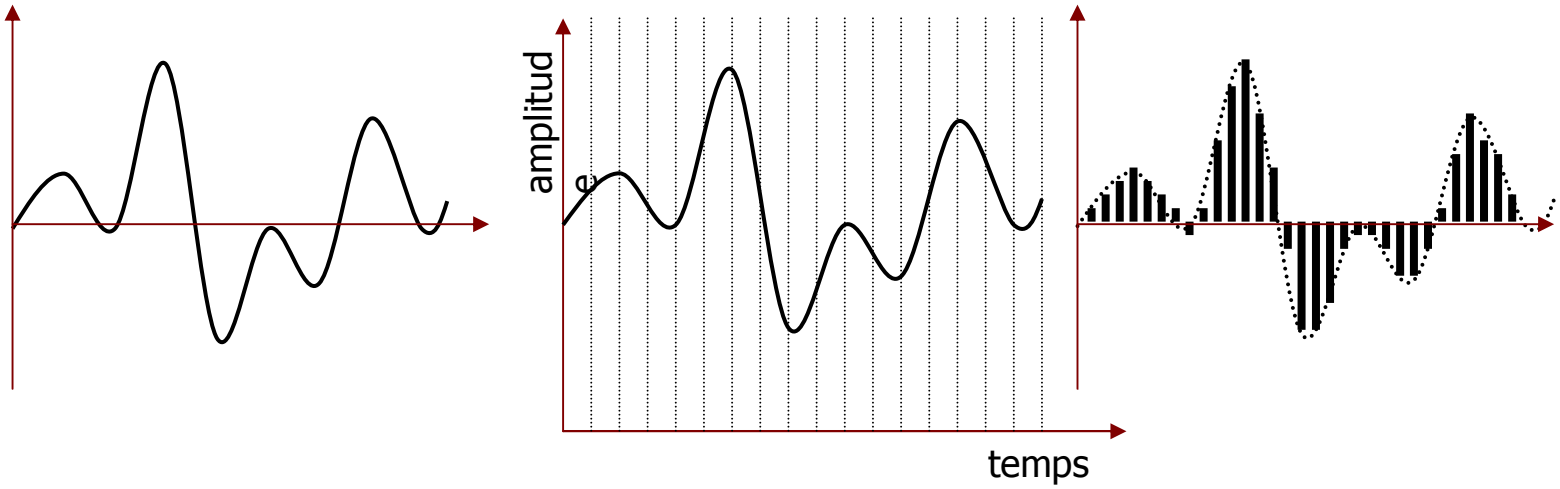
Le timbre d'une voix émettant par exemple à une fréquence fondamentale de 400Hz sera perçu comme différent de celui d'une autre voix émettant pourtant à la même fréquence fondamentale



De l'analogique au numérique

- La représentation analogique du son consiste à passer d'une onde sonore à son image électrique.
- La représentation numérique du son nécessitera quant à elle la quantification des échantillons prélevés sur cette image.

Du continu au discret



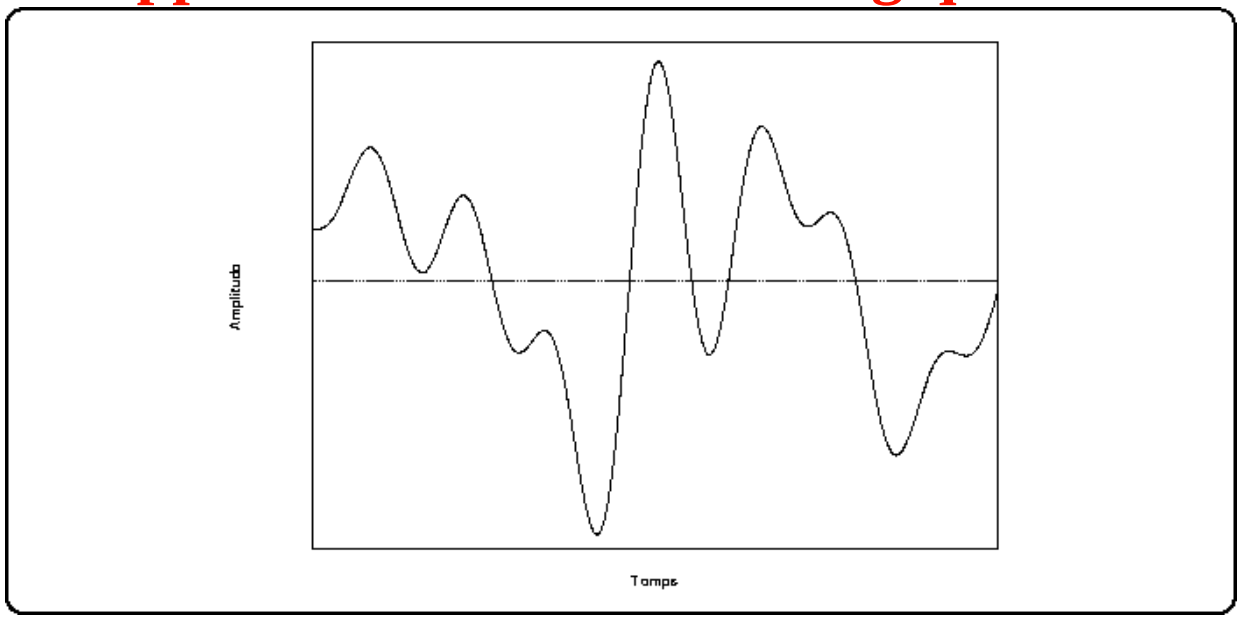
Onde sonore → échantillonnage → Quantification

Format du son

- 2 types de formats pour stocker des sons ou de la musique :
  - Les formats numériques.
  - Les formats musicaux.
- Formats numériques :
  - Stocker une approximation du son analogique.
  - Constitué d'une suite d'échantillons.
- Formats musicaux :
  - Stockent essentiellement des informations sur quand et comment jouer tel ou tel instrument.

Formats numériques

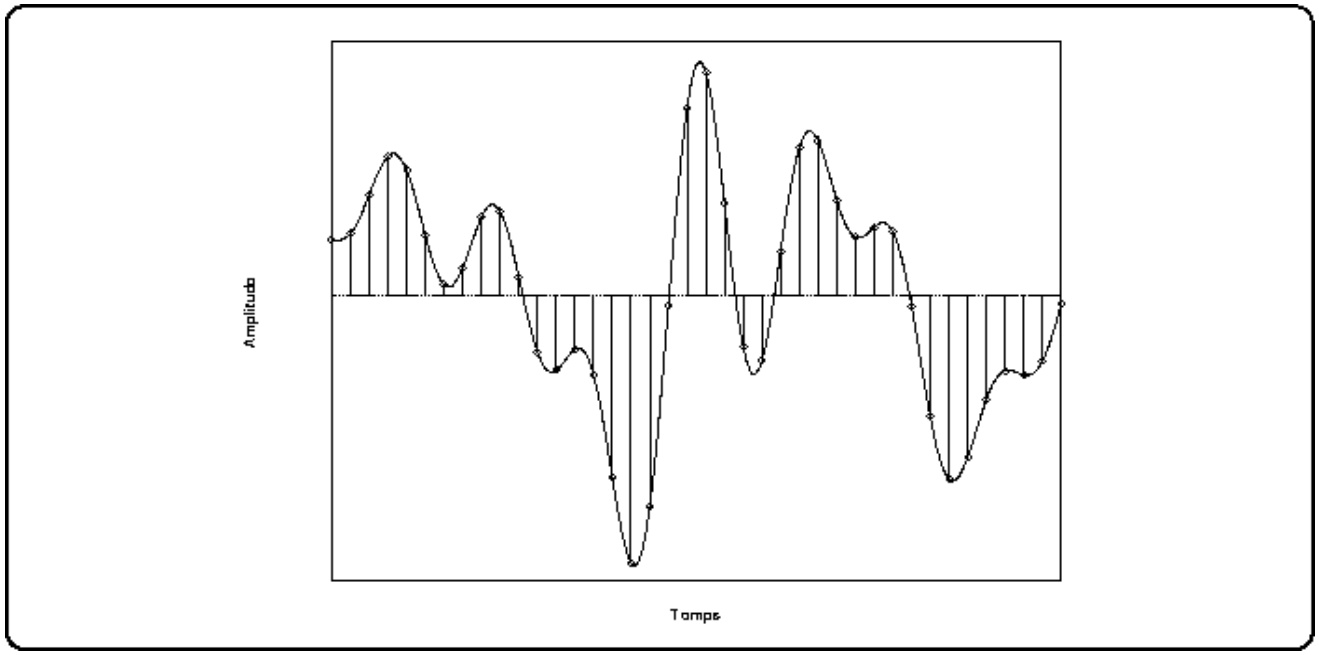
- Stocker des sons, des textes parlés ainsi que des morceaux de musique.
- 2 Dimensions : temps et amplitude.
- Stocker une approximation du son analogique :



Numérisation du son

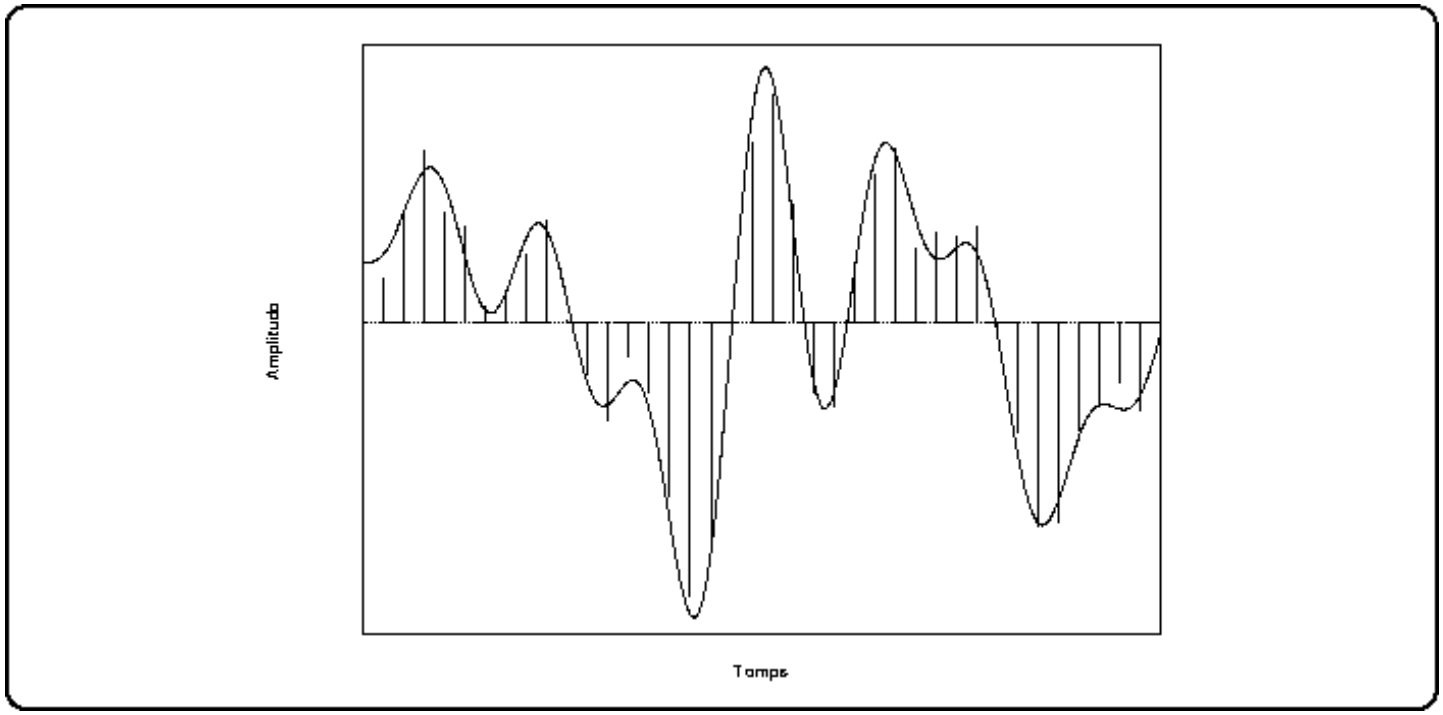
- Numérisation d'un signal sonore : 2 étapes.
- Echantillonnage :
  - Discrétisation dans le temps.
  - Choix du nombre d'échantillons par seconde.
- Quantification :
  - Discrétisation de l'amplitude.
  - Choix de la précision de représentation d'un échantillon.

Echantillonnage du son



- Discretisation dans le temps :
  - Découpage en "tranches".
  - Choix du nombre d'échantillons par seconde.
- Ex. : échantillonage à 11 KHz.
  - Le son est analysé 11.000 fois par seconde.

Quantification du son



- Discrétisation de l'amplitude.
  - Choix de la précision de représentation d'un échantillon.
- Ex. : représentation sur 1 octet.
  - Un échantillon est codé sur 256 niveaux.

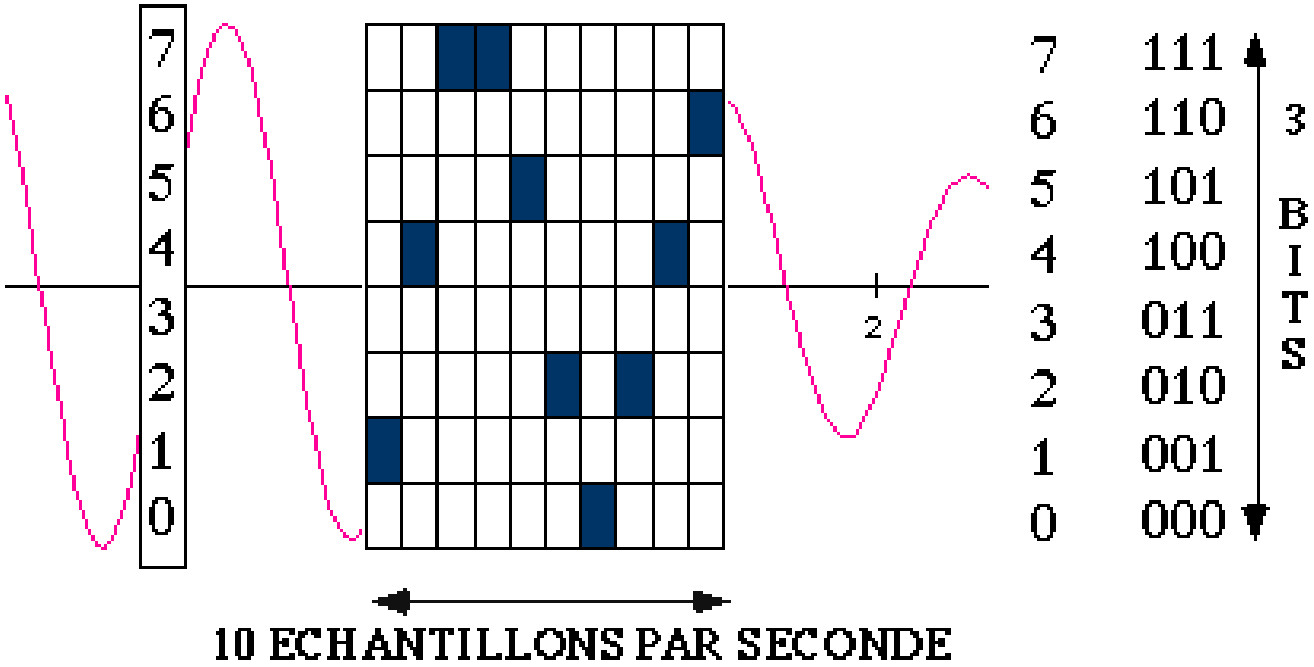
Fonctions de quantifications

- Quantification par arrondi (valeur approchée)
  - L'arrondi permet tantôt de supprimer, tantôt d'ajouter : la somme de toutes ces petites « erreurs » peut être souvent voisine de zéro et donc avoir une incidence moindre que la seconde procédure suivante
- Quantification par troncature
  - Conduit toujours à des suppressions de valeurs

Echantillonnage et quantification

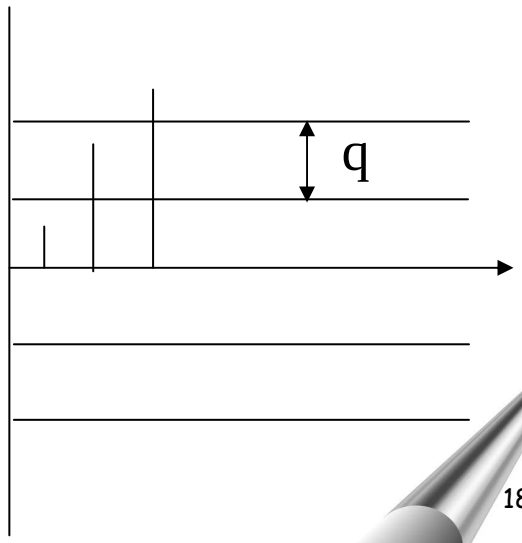
- Exemple :
  - Echantillonage : 10 Hz.
  - Quantification : 3 bits

Exemple animé



Distorsion de quantification

- Le bruit de quantification affecte gravement les signaux de faible amplitude
- Lorsque le niveau du signal augmente, l'erreur de quantification (dont la valeur maximale est toujours égale à q/2) devient de plus en plus petite relativement à l'amplitude du signal

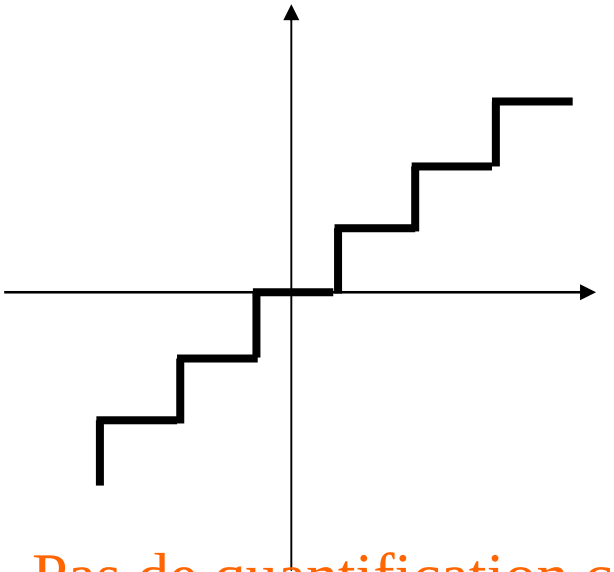


Quantification non uniforme (1)

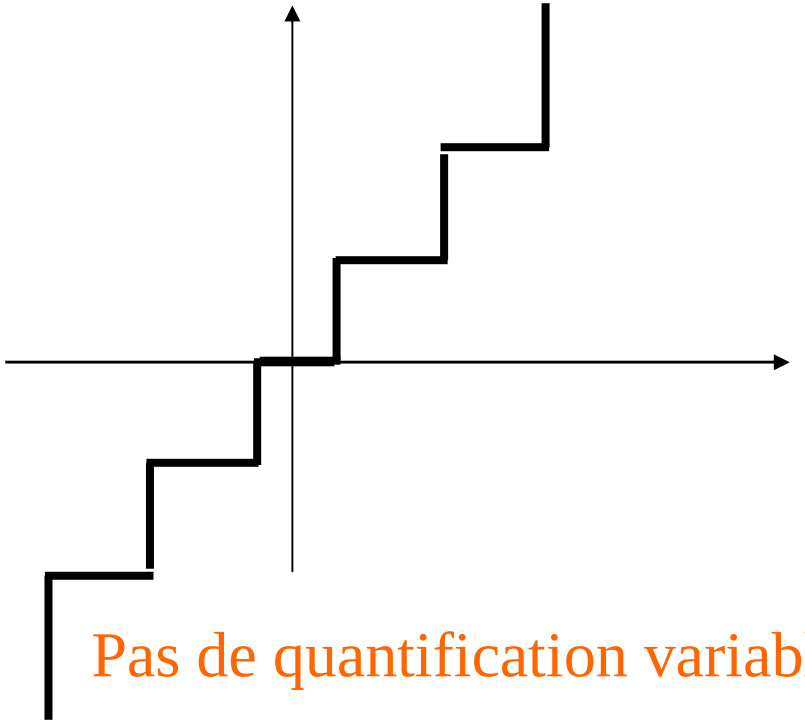
- Jusqu'à présent nous avons évoqué le cas d'une quantification uniforme
  - *q* est constant quelle que soit l'amplitude du signal
    - ie : erreur de quantification constante et indépendante de l'amplitude du signal
- Or, les signaux de faible amplitude seront davantage perturbés par le bruit de quantification (rapport S/B défavorable)
- Il serait intéressant d'obtenir, pour les signaux de faible amplitude, un rapport S/B aussi favorable que celui des signaux d'amplitude plus importante.

Quantification non uniforme (2)

- Solution : faire varier le pas de quantification en fonction de l'amplitude instantanée du signal : plus le signal est faible, plus on affine le pas de quantification et plus on améliore le rapport S/B



Pas de quantification constant



Pas de quantification variable

Echantillonnage : fréquence optimale

- Théorème de Shannon :

  Si un signal est échantillonné périodiquement à une fréquence au moins égale à deux fois la fréquence la plus élevée contenue dans ce signal, l'ensemble des échantillons contient toute l'information du signal d'origine.

Fréquences d' échantillonnage

Fréquence (kHz)	Applications
8	Téléphonie
22,05	PC
32	HTDV (high definition TV) Radio numérique NICAM
44,1	CD audio
48	DAT (digital audio tape)
192	DVD audio *2 canaux stéréo*
2822,4	SACD (super audio CD) DSD (Direct stream digital)

Echantillonnage : fréquence optimale (2)

- On peut alors déterminer la fréquence d'échantillonnage optimale.
- Par exemple :
  - Le domaine perceptible par l'oreille humaine : jusqu'à 20 kHz.
  - La qualité CD est donc fixée à 44100Hz.
  - Celui de la parole est restreint à 150 Hz - 5 kHz.
  - En téléphonie, on utilise 8 kHz.

Qualité de la numérisation

- Dépend de 3 paramètres.
- Fréquence d'échantillonnage (nombre d'échantillons/seconde).
  - Qualité téléphonique : 8 kHz.
  - Qualité CD : 44,1 kHz.
- Le nombre de bits par échantillon :
  - 8 bits (qualité médiocre).
  - 16 bits (qualité CD).
- Le nombre de canaux :
  - 1 canal : son mono.
  - 2 canaux : son stéréophonique.
  - 4 canaux : son quadriphonique.

Numérisation et taille des données (1)

- Exemple : numérisation qualité CD.
  - Fréquence d'échantillonnage : 44 KHz.
  - Nombre de bits par échantillon : 16 bits.
  - Nombre de canaux : 2 (stéréo).
- 44000 échantillons par seconde, chacun nécessitant 2 octets, le tout stocké en double :
  - 44.000 x 16 x 2 = 1.408.000 bits/s.
  - C'est-à-dire 172 Ko/s.

Principe de compression de son

- Le principe général de ce type de compression est fondé sur la recherche des occurrences multiples d'une même suite d'octets. Ces suites étant rassemblées dans un dictionnaire où elles seront représentées par les codes les plus courts possibles.
- Les codes dans le dictionnaire remplaceront dans le fichier compressé les longues suites d'octets du fichier original.
- Exemple 1 : algorithmes généralistes

  WinZip, WinRar et WinAce (*.zip, *.rar, *.ace) doivent se satisfaire d'une réduction du volume du fichier d'origine (selon la complexité de la source) de 5 à 35%
- Exemple 2 : algorithmes spécifiques

  WavArc, RKAU, Perfect Clarity Audio (*.wa, *.rka, *.pca) sont spécifiquement dédiés au son et non destructifs, permettent d'atteindre une réduction de volume de 50%.

Le format WAV
(Waveform audio File Format)

- Le format propriétaire de l'environnement windows sur PC
- Lu directement par le lecteur multimédia de windows ou le lecteur QuickTime ainsi que la quasi-totalité des logiciels d'édition ou de composition musicale
- Le format de données comprend :
  - 44 octets d'en-tête permettant une identification.
  - Une structure les organisant.
  - Le codage des données.

Le format AIF ou AIFF (ou SND)

- AIFF (Audio Interchange File Format).
- Apple 1988 :
  - Fichiers son de haute qualité.
  - Initialement développé pour les ordinateurs Macintosh.
- Jusqu'à 44,1 kHz en 16 bits.
- Equivalent Macintosh du format WAV.
- Devenu le premier format d'échange multi-plateformes.

Le format AU (Audio Unix)

- Format développé par UNIX
- Lu par la quasi-totalité des logiciels professionnels de traitement audio

Real Audio (RA, RM ou RAM)

- Format propriétaire adapté aux débits limités de Internet pour la diffusion de sons ou de vidéos en streaming.
- Peuvent être lus grâce à l'application spécifique RealPlayer (en version commerciale ou shareware).

Quicktime (MOV ou QT)

- Destiné à l'origine à l'environnement Macintosh.
- Il est aujourd'hui disponible sur PC et peut être utilisé pour réaliser du streaming.

MP3 (MPEG-Layer 3)

- Brevet de l'algorithme de compression déposé en 1996 par l'institut de recherche allemand Fraunhofer, qui fait partie du groupe Thomson Multimédia.
- Recommandé par le MPEG (Mouving Picture Experts Group)
- Autorise un taux de compression important sans altérer notablement la qualité sonore (du moins jusqu'à un taux de l'ordre de 12:1)
- De nombreux lecteurs permettent la restitution des fichiers .mp3 : WinAmp, Nad mp3 Player, UnrealPlayer, JetAudio, etc.
- Taux de compression très intéressant, d'autant plus que ceux-ci sont adaptables en fonction de la bande passante disponible pour la diffusion en temps réel (streaming) du fichier compressé.

MP3 (MPEG-Layer 3)

- **MP3 (MPEG Audio Layer 3) : Attention cela n'a rien à voir avec MPEG-3 (norme ratée).**
- **La compression Mpeg Layer-3 permet de diminuer d'environ 12 fois la taille d'un fichier audio habituel. Ceci est obtenu par suppression de zones dans la bande passante dont les fréquences sont inaudibles par l'oreille humaine.**
- **L'oreille ne fait aucune différence entre le son original d'un CD et le son de ce même CD après compression MP3.**
- **Qualité soit disant équivalente au CD, a un débit de 192Kbits/s = 1,44 Mo/min [CD : 10 Moctets/min]**

Formats musicaux

- Les fichiers de sons de synthèse.
- Permettre aux musiciens de contrôler plusieurs instruments électroniques à partir d'un seul.
- Stocker des messages musicaux permettant de piloter un périphérique MIDI :
  - Quelle note jouer,
  - Sur quel instrument,
  - A quel volume, etc.
- Les instruments ne sont pas définit dans le format, ce qui permet un gain de place considérable.
- Faible volume d'information requis (1000 fois moins que le signal audio en qualité CD)
- La qualité de la musique dépend du périphérique utilisé pour la jouer.
- Ex. : le format MIDI (Musical Instrument Digital Interface).

Le format MIDI

- Le format MIDI (Musical Instrument Digital Interface).
  - 16 canaux : 16 instruments peuvent jouer en même temps.
  - Indications sur quelle note jouer, sur quel instrument, à quel volume, etc.
- Gain de place : 2 minute :
  - Format MIDI : 80 Ko.
  - Format WAV : 20 Mo (qualité CD).
- Fichier facilement éditable avec un séquenceur MIDI.
- Il est facile de changer les instruments.

Le format MIDI : limitations

- Utilisable que pour la musique.
- La qualité de la musique :
  - Dépend du périphérique utilisé pour la jouer.
  - (En particulier le nombre de canaux).
- Le codage des instruments n'est pas normalisé :
  - Par exemple : un morceau de violon peut être interprété par un saxophone sur un autre synthétiseur.
- Obligation d'utiliser des sons déjà existants :
- "Impossible " d'intégrer un son créé par le musicien (échantillonné ou réellement synthétisé).

Algorithme de compression du son

- Quant il s'agit de diffuser du matériel sonore de haute qualité en temps réel (streaming) par voie hertzienne ou par câble, le débit exigé pose de sérieux problèmes…
- La diffusion d'un programme musical de qualité CD exige de disposer d'un débit d'environ 180ko par seconde !
- D'où la nécessité de compresser ces fichiers.

Les modulations différentielles

- Nous avons vu précédemment la méthode classique appelée modulation par codage d'impulsions (en anglais PCM), chaque échantillon est codée sur 1 ou 2 octets
- Pour cette méthode (modulations différentielles) ou DPCM, au lieu de coder la valeur absolue de chaque échantillon, se contente de coder la différence constatée entre deux échantillons
- Modulation delta :
  - Il s'agit de substituer à la valeur définissant l'amplitude de l'échantillon le seul signe (positif ou négatif) de la variation du signal par rapport à sa valeur précédente.
  - Cette information ne nécessite qu'un seul bit par échantillon
  - 1 = différence positive
  - 0 = différence négative

Modulation delta et saturation de pente

- Cette technique suppose que chaque valeur quantifiée ne diffère de la précédente que de plus ou moins 1 incrément de quantification
  - Limité aux signaux dont le variations sont relativement lentes
  - Des variations rapides excédent en effet les capacité de codage
    - Incapable de suivre le signal d'entrée avec une exactitude satisfaisante :
    - On parle de saturation de pente

Modulations différentielles adaptatives (1)

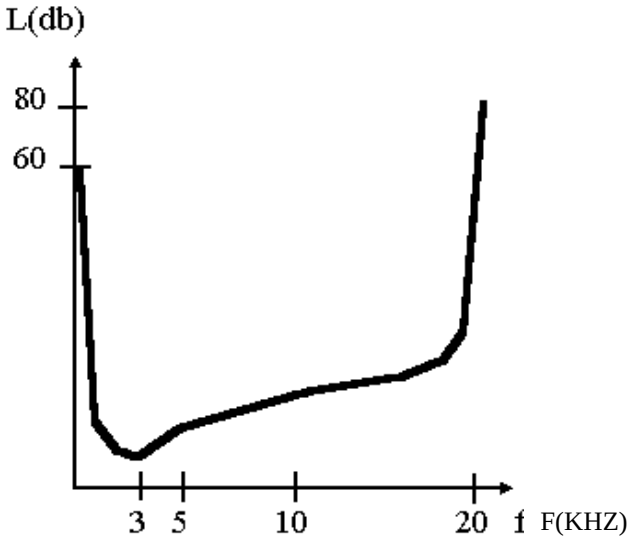
- ADPCM (Adaptative Differential Pulse Code Modulation) va non seulement, comme DPCM, coder la différence entre deux échantillons successifs (en général sur 4 bits), mais encore tenir compte de l'évolution du signal sur les échantillons précédents pour prédire le pas de quantification à utiliser.
- Ceci permet d'augmenter les capacités du codage du modulateur et éviter la saturation de pente, ou du moins en limiter l'incidence.

Modulations différentielles adaptatives (2)

- Par, exemple, le signal passe brusquement d'une tension élevée à une tension faible, la valeur de pas retenue sera grande ; au contraire, si le signal d'entrée présente de faibles variations de tension, la valeur de pas retenue sera petite. On stocke ensuite ce pas dans une table de pas.

Compression des sons algorithmes destructifs : le codage psychoacoustique

- Principale méthode de compression de sons :
  - Retirer des données audio les fréquences inaudibles par l'oreille humaine



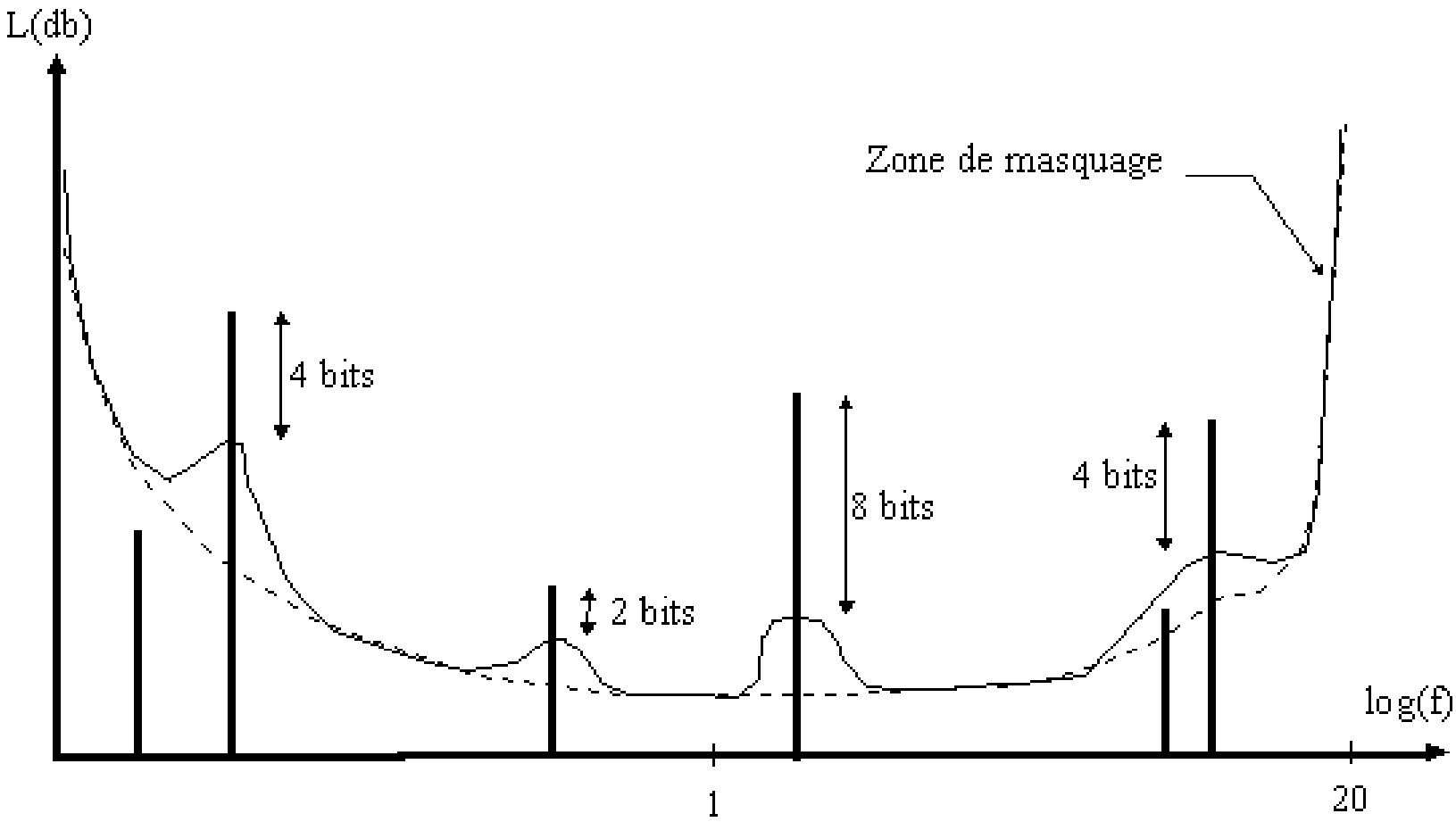
- L'oreille humaine :
  - Capable de discerner des sons entre 0.20Khz et 20Khz.
  - Sensibilité est maximale entre 2 et 5Khz.

- Principale méthode de compression (MPEG, PASC) :
  - Modèle psycho-acoustique.
  - Déterminer les sons que nous n'entendons pas.
  - Les supprimer : "masquage".
- Mais aussi...
- Elimination de la redondance :
  - Utilisation de l'Algorithme de Huffman.
- Autres traitements :
  - Suppression des silences.

Algorithme PASC

- 2 tâche distinctes :
- Un rôle filtreur :
  - Eliminer du signal toutes les fréquences inférieures à 15 Hz et supérieures à 20 KHz.
- Un rôle échantillonneur :
  - Diviser le spectre audible en 32 bandes de 720 Hz chacune.
  - Comparer la valeur moyenne de chaque bande à la valeur correspondant sur la courbe de réponse de l'oreille.
  - Toute bande en dessous de la courbe de réponse est éliminée.
  - Chaque bande retenue est divisée en 12 sous-bandes échantillonnées sur 16 bits.
  - On compare encore une fois la valeur des sous-bandes à la courbe de référence.
  - On quantifie les différences à la courbe, et on code ces différences sur un nombre minimum de bits.
  - Ces opérations sont renouvelées tous les 8 millièmes de seconde.

Algorithme PASC (2)



Quantification des niveaux de fréquences audibles

Algorithme PASC (3)

- 6 bandes (traits verticaux) :
  - Traversent ou non la zone de masquage.
  - 2 restent sous la courbes : rejetées.
  - Les 4 autres sont échantillonnées en 12 autres sous-bandes.
  - Pour chaque sous-bande : un nombre de bits.

MPEG

- Audio MPEG (Moving Picture Experts Group).
  - Echantillonnage à 32, 44.1 ou 48 kHz.
  - Précision équivalant à une quantification à 16 bits par échantillon.
- 3 niveaux de compression possibles :
  - Layer 1 : 192 kbit/s (compression 4:1).
  - Layer 2 : 128 kbit/s (compression 6:1).
  - Layer 3 : 64 kbit/s (compression 12:1) (le célèbre MP3!!!).
- Spectre divisé en 32 sous-bandes.
- Application du modèle psycho-acoustique :
  - Calcul des masquages (ne conserver que les sons audibles).
  - Allocation du nombre de bits.
  - Quantification.