Introduction aux terminologies numériques

Par opposition aux signaux analogiques, l'information numérique est constituée de valeurs discrètes. C'est à dire que l'on ne connaît la valeur du signal qu'à certains instants seulement. Un signal numérique est une suite temporelle de valeurs binaires.

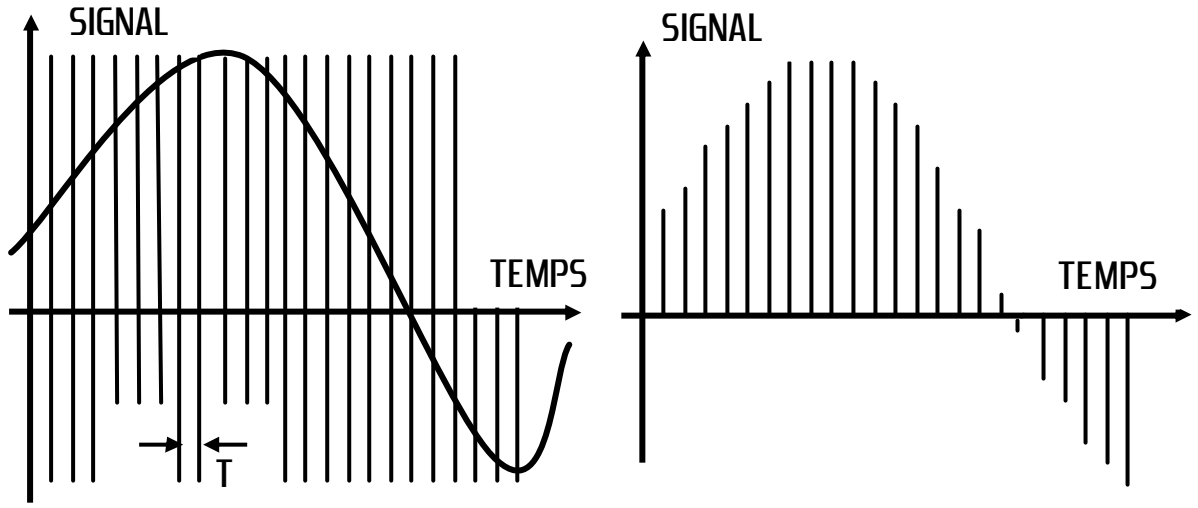
Une unité d'information binaire s'appelle un bit (de l'anglais binary digit), et un bit ne peut emprunter que les valeurs 1 ou 0.

Continuons avec quelques notions sur les nombres binaires. Dans le système décimal usuel, chaque chiffre représente une puissance de dix alors qu'il représente une puissance de deux dans le système binaire. Un nombre constitué de plus d'un chiffre binaire (bit) est appelé "mot" binaire (c'est un peu la même distinction qu'entre chiffre et nombre). Plus un mot contient de bits et plus le nombre d'états qu'il peut représenter est important : un mot de 8 bits admet 256 (2 à la puissance 8) états et un mot de 16 bits en admet 65536 (2 à la puissance 16). Le bit de plus faible poids (2 puissance 0) est appelé élément binaire de poids faible (LSB ou Least Significant Bit) alors que le bit de plus grand poids est appelé élément binaire de poids fort (MSB ou Most Significant Bit). Electriquement, il est possible de représenter un mot binaire soit sous forme "sérielle" soit sous forme "parallèle". La communication sérielle ne requiert qu'une seule connexion, le mot étant transmis bit par bit au cours du temps. En communication parallèle, chaque bit du mot est pris en charge par une connexion séparée et tous les bits du mot sont donc transmis simultanément.

L'information sonore analogique sous forme électrique est convertie sous forme électrique numérique par l'intermédiaire d'un système appelé convertisseur analogique-numérique (A/N ou CAN). Il est établi qu'un signal analogique peut emprunter un nombre infini de valeurs, alors qu'un signal numérique ne peut emprunter qu'un nombre limité de valeurs fixées. Le nombre de valeurs fixées possibles pour un signal numérique dépend de la longueur des mots binaires utilisés, autrement dit du nombre de bits. Afin de convertir un signal analogique en signal numérique, il est nécessaire de mesurer son amplitude à intervalles de temps spécifiques (c'est l'échantillonnage) et d'affecter une valeur binaire à chacune des mesures (c'est la quantification). Le processus de conversion analogique-numérique a une incidence majeure sur la qualité finale du signal audionumérique. En effet, la qualité du message musical, une fois converti, ne peut jamais s'améliorer, mais plutôt empirer. Pour les applications audionumériques, l'offre s'étend aujourd'hui du convertisseur 8 bits/32 kHz jusqu'au convertisseur 24 bits/192 kHz très haut de gamme en passant par le traditionnel convertisseur 16 bits/44,1 kHz. Comme la suite de cette partie le démontre, le taux d'échantillonnage et le nombre de bits par échantillon sont les principaux facteurs qui influent sur la qualité audio. La qualité des convertisseurs détermine quant à elle la différence entre la qualité sonore obtenue et la qualité théorique fixée par ces deux facteurs.

Fréquence d'Echantillonnage

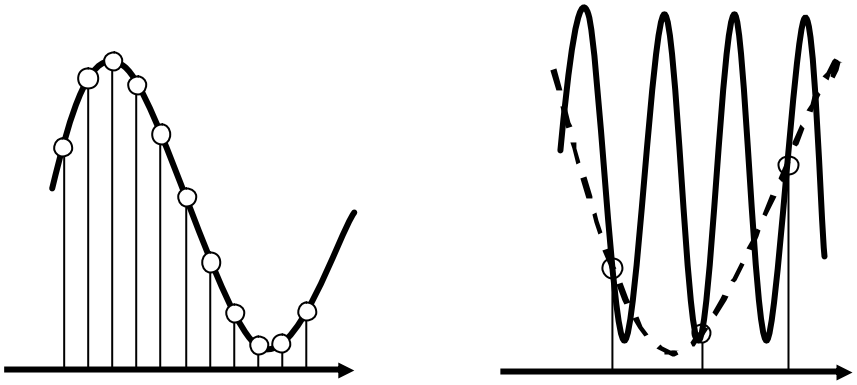
Un signal audio sous forme analogique est une forme d'onde électrique continue dans le temps. La tâche d'un convertisseur analogique-numérique est de traduire ce signal en une séquence de nombres binaires. La méthode d'échantillonnage employée dans un convertisseur analogique-numérique consiste à mesurer, ou encore "quantifier", l'amplitude de la forme d'onde à des intervalles de temps réguliers :



Ici Un signal quelconque est échantillonné à intervalles de temps réguliers t (à gauche) afin de générer de courtes impulsions (à droite) dont les amplitudes représentent l'amplitude instantanée du signal.

Sur ce diagramme, il apparaît clairement que les impulsions représentent les amplitudes instantanées du signal à chaque instant t. La période T est appelée période d'échantillonnage. Les échantillons peuvent être considérés comme des images instantanées du signal audio qui, assemblés dans une séquence, donnent une représentation de la forme d'onde continue, de la même manière que la séquence d'images d'un film, projetée en succession rapide, donne l'illusion d'une image en mouvement continu.

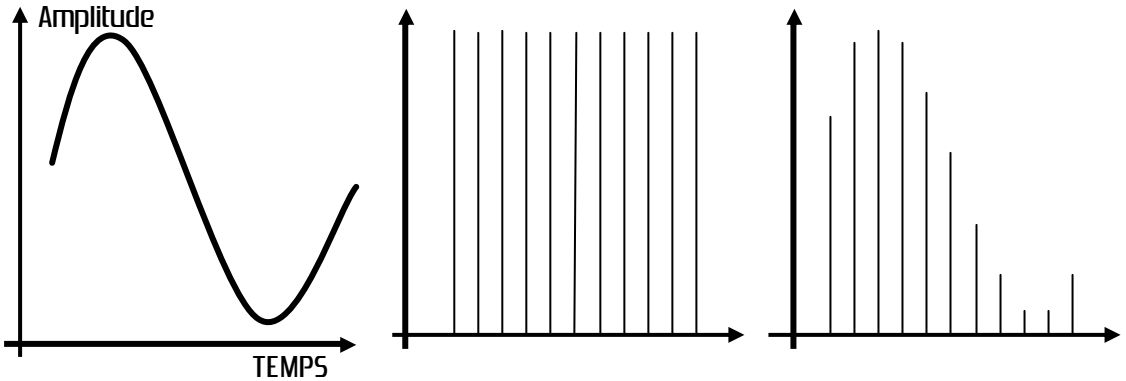
Afin de représenter les détails fins du signal, il est nécessaire de prélever un grand nombre de ces échantillons à chaque seconde. Comme on peut le voir dans la figure suivante, si on prélève trop peu d'échantillons par cycle, ils peuvent alors être interprétés comme la représentation d'une forme d'onde différente de la forme d'onde d'origine échantillonnée. Ce problème est en fait un exemple de phénomène connu sous le nom de repliement de spectre (ou Aliasing). Un Alias est un produit indésirable du signal d'origine survenant lors de sa reconstruction en conversion numérique-analogique.



A gauche on prélève de nombreux échantillons par cycle de l'onde alors qu'à droite on prélève moins de deux échantillons par cycle. Il est alors impossible de reconstruire une forme d'onde de fréquence plus haute à partir des échantillons, c'est un exemple de repliement du spectre.

Il nous faut donc trouver la bonne période d'échantillonnage. C'est à dire une période qui permette de restituer assez fidèlement le signal d'origine. On pourrait être tenté de diminuer le plus possible cette période mais on se confronterait alors à des problèmes de stockage ou de bande passante. En effet, prendre plus d'échantillons que nécessaire va impliquer plus d'informations et donc un besoin accru de ressources. Les mathématiques nous indiquent que, pour obtenir les informations nécessaires à la caractérisation du signal, il faut prélever au moins deux échantillons par cycle audio. Afin de justifier ce résultat, nous pouvons considérer le processus d'échantillonnage en termes de modulation :

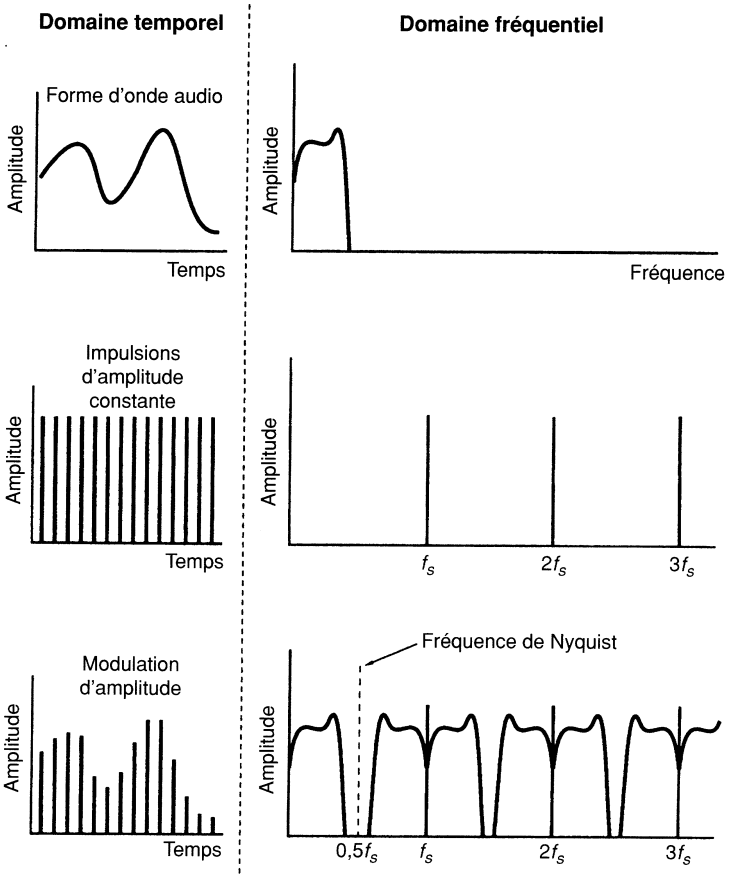
La forme d'onde continue que l'on souhaite échantillonner est utilisée pour moduler une chaîne régulière d'impulsions. La fréquence de ces impulsions est appelée fréquence d'échantillonnage. Avant modulation, toutes les impulsions ont la même amplitude. Après modulation, l'amplitude des impulsions est modifiée en fonction de l'amplitude instantanée du signal audio. Ce procédé est appelé modulation d'impulsions en amplitude.



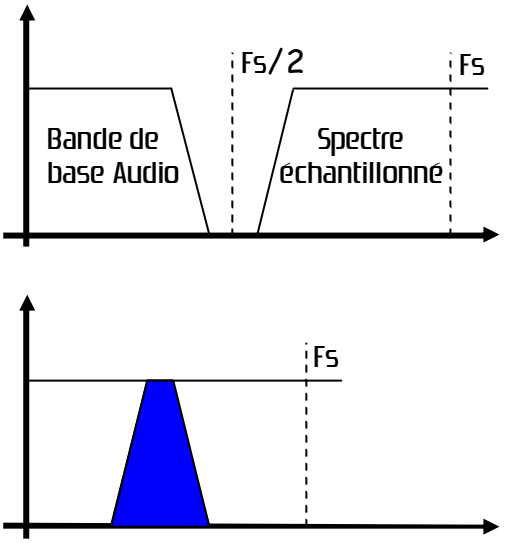
En modulation d'impulsions par amplitude, l'amplitude instantanée des impulsions (au milieu) est modulée par l'amplitude instantanée du signal audio (à gauche)

Poursuivons en considérant le domaine fréquentiel (ou spectre). Tout signal est décomposable en une somme de sinusoïdes de différentes fréquences. Par exemple un extrait musical est usuellement composé de fréquences basses, medium et aigues. Ces groupements représentent en fait des familles d'ondes harmoniques (sinusoïdales) de fréquences proches. Mais il suffit de continuer le raisonnement en continuant à découper les bandes de fréquences en bandes plus petites jusqu'à considérer toutes les fréquences séparément. Parler du spectre d'un signal c'est décrire la puissance de chaque pulsation (ou fréquence).

Le spectre de fréquences du signal modulé est montré sur la figure suivante. On remarque qu'en plus du spectre d'origine avant échantillonnage, apparaît maintenant un certain nombre de spectres additionnels, chacun étant centré sur des multiples de la fréquence d'échantillonnage. Des bandes secondaires résultant de la modulation d'amplitude ont été produites de chaque côté de la fréquence d'échantillonnage et de ses multiples. Celles-ci s'étendent en dessous et au-dessus de la fréquence d'échantillonnage et de ses multiples sur des largeurs équivalentes à celle de la bande de base. En d'autres termes, ces bandes secondaires sont des paires d'images miroirs de la bande audio.



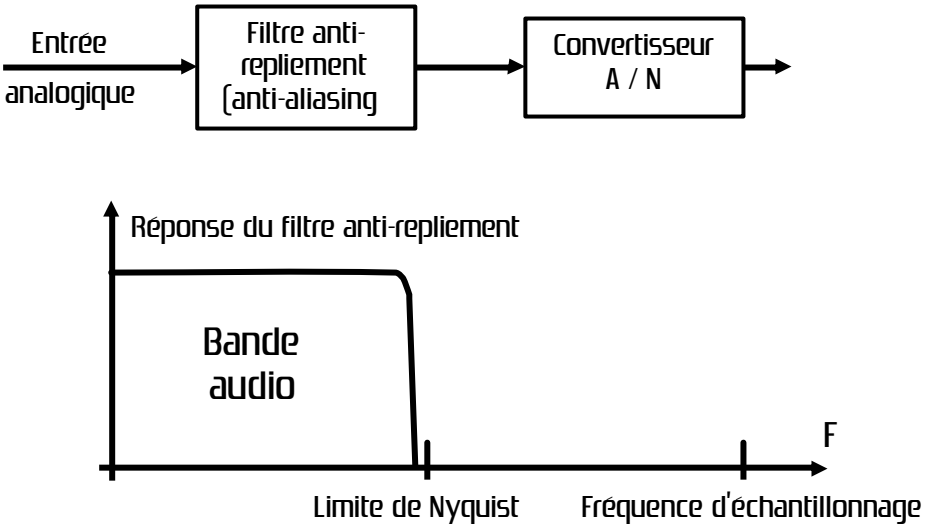
Ce sont ces spectres additionnels de part et d'autre de la fréquence d'échantillonnage qui vont imposer d'échantillonner à une fréquence au moins double de la plus haute fréquence présente dans le signal. Il suffit d'imaginer que la fréquence d'échantillonnage ne soit pas suffisamment grande : le spectre original et le spectre image vont entrer en collision créant des alias (zone bleue ci dessous), c'est à dire ajoutant des sons initialement non présents.



En étudiant la figure précédente, on comprend pourquoi la fréquence d'échantillonnage doit être supérieure au double de la fréquence audio la plus élevée du signal. On note qu'une extension de la bande de base au-dessus de la fréquence de Nyquist (moitié de la fréquence d'échantillonnage Fs) produit un recouvrement de la bande secondaire inférieure de la première répétition spectrale par la limite supérieure de la bande de base. Conjointement, une diminution de la fréquence d'échantillonnage a les mêmes conséquences. Dans le premier exemple, la hauteur tonale de la bande de base reste suffisamment basse pour que les bandes secondaires échantillonnées restent au-dessus du spectre audio ; dans le second, une fréquence d'échantillonnage trop faible entraîne une collision des spectres, générant ainsi des alias du spectre original dans la bande de base, autrement dit des distorsions (déformations du signal en fonction de lui même).

Dans la mesure où l'image cinématographique constitue également un exemple de signal échantillonné, l'effet cinématographique bien connu de "la roue qui tourne à l'envers" rend le phénomène de repliement du spectre (ou aliasing) visible et donc plus concret. Pour le film, les images sont en principe mises à un taux de 24 par seconde. Si une roue marquée est filmée, elle semblera tourner dans le sens de la marche tant que sa vitesse de rotation reste inférieure au taux de la caméra. Si la vitesse de rotation augmente, la roue semblera ralentir, s'arrêter, puis se mettre à tourner en sens inverse, et cette impression de mouvement rétrograde augmentera si la vitesse de rotation de la

roue augmente encore. Ce mouvement rétrograde est en fait l'alias généré par un échantillonnage trop faible. En audionumérique, si le phénomène de repliement de spectre n'est pas contrôlé, on aperçoit auditivement l'équivalent du mouvement rétrograde d'une roue filmée sous la forme de composantes sonores (originellement absentes) dans le spectre audible. Leur fréquence décroît à mesure que la fréquence du signal d'origine augmente. Avec des convertisseurs basiques, il est donc nécessaire de filtrer le signal audio avant échantillonnage afin de supprimer toute composante dont la fréquence excède la fréquence de Nyquist (la moitié de la fréquence d'échantillonnage).

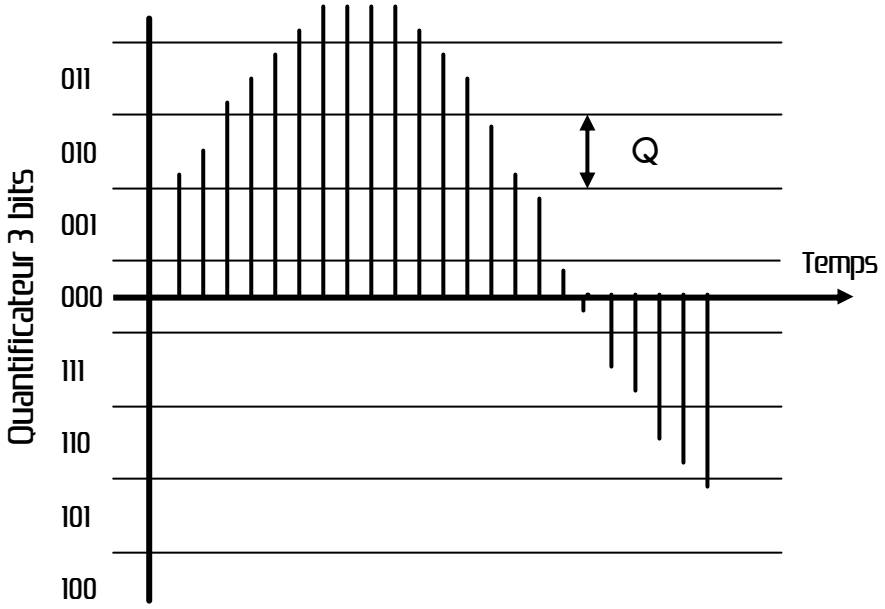


En réalité, comme les filtres ne sont pas parfaits, on choisit une fréquence d'échantillonnage légèrement supérieure au double de la fréquence audio la plus élevée devant être représentée. On peut ainsi accepter des filtres qui coupent de façon un peu plus douce. Les filtres intégrés aux convertisseurs analogique-numérique et numérique-analogique ont un effet prononcé sur la qualité sonore, puisqu'ils déterminent la linéarité de la réponse en fréquence dans la bande audio ainsi que la linéarité du système. Dans un convertisseur classique, le filtre doit rejeter tous les signaux au-dessus de la moitié de la fréquence d'échantillonnage (fréquence de Nyquist) avec une atténuation d'au moins 80dB.

Le procédé de suréchantillonnage (voir plus loin) qui échantillonne à des fréquences plus élevées a contribué à atténuer les problèmes du filtrage analogique dans la mesure où la première répétition de la bande de base est rejetée à une fréquence beaucoup plus élevée, permettant ainsi l'emploi d'un filtre de pente moins raide.

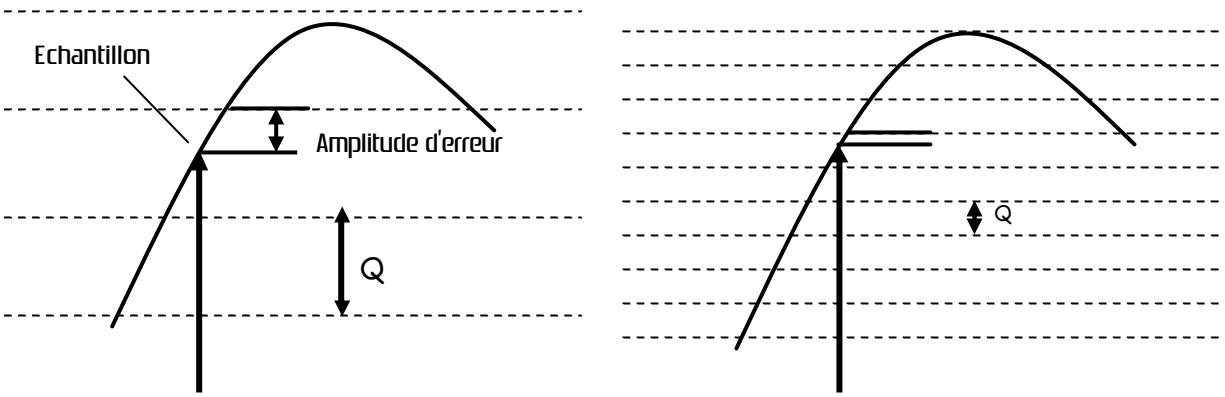
La Quantification

Après l'échantillonnage, la chaîne d'impulsions modulées est quantifiée. Quantifier un signal échantillonné consiste à placer les amplitudes des échantillons sur une échelle de valeurs à intervalles fixes (voir figure suivante).



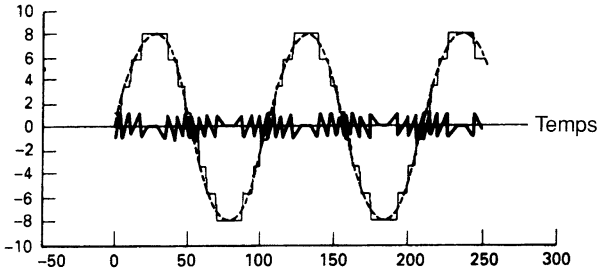
Le quantificateur détermine dans quel intervalle de quantification (de taille Q) l'échantillon se situe, et lui affecte une valeur qui représente le point central de cet intervalle. Ce procédé permet d'attribuer à l'amplitude de chaque échantillon un mot binaire unique. En quantification linéaire, chaque pas de quantification représente une incrémentation identique de la tension du signal. De plus, dans le système binaire le nombre de pas de quantification est égal à 2 puissance n, où n est le nombre de bits des mots binaires utilisés pour représenter chaque échantillon. En conséquence, un quantificateur 4 bits offre seulement 2 puissance 4 (16) niveaux de quantification, alors qu'un quantificateur 16 bits en offre 2 puissance 16 soit 65 536.

Il apparaît clairement qu'une erreur intervient dans la quantification, puisqu'on ne dispose que d'un nombre limité de valeurs différentes pour représenter l'amplitude du signal à chaque instant. La valeur maximale de l'erreur est de 0.5 Q. En conséquence, plus le nombre de bits par échantillon est important, et plus l'erreur est petite (voir figure suivante).



L'erreur maximale de quantification est égale à la moitié de l'intervalle de quantification (Q). A gauche le nombre d'intervalle est faible et l'erreur est importante alors qu'à droite le nombre d'intervalle est plus grand et l'erreur devient plus petite.

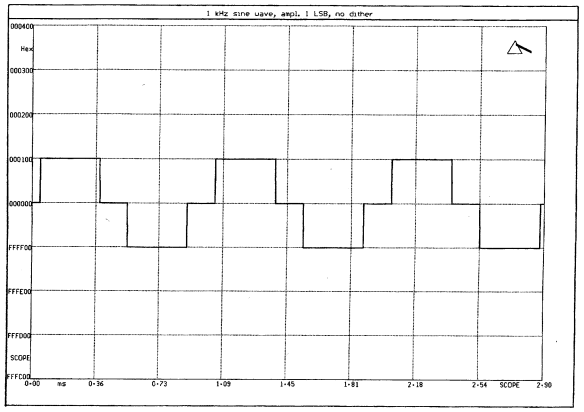
L'erreur de quantification peut être considérée comme un signal indésirable ajouté au signal utile (voir figure suivante). Les signaux indésirables sont classifiés comme distorsion ou bruit en fonction de leurs caractéristiques. La nature du signal d'erreur de quantification dépend du niveau et de la nature du signal audio qui lui est rattaché.

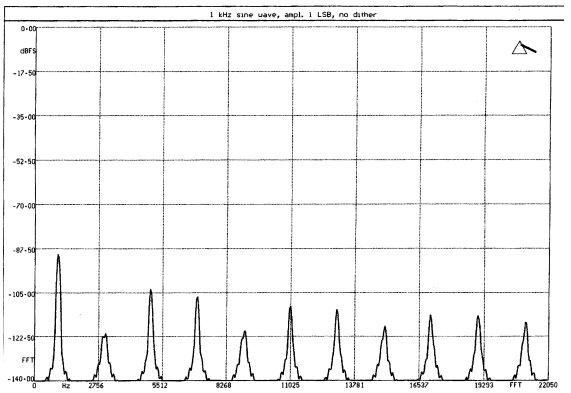


L'erreur de quantification vue comme signal indésirable ajouté aux valeurs d'échantillons d'origine. Ici, l'erreur est directement corrélée au signal et apparaîtra comme de la distorsion.

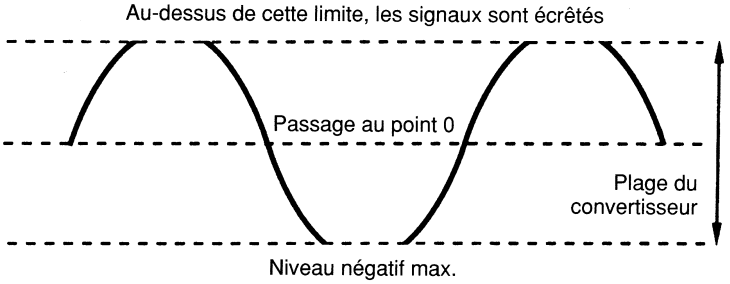
Pour plus de clarté, considérons l'exemple d'une quantification 16 bits d'un signal sinusoïdal, échantillonné, de très bas niveau. Son niveau est tout juste suffisant pour affecter la valeur du bit le moins significatif à son niveau maximal (voir figure suivante à gauche). Un tel signal aura une erreur de quantification périodique et fortement corrélée au signal, apportant de la distorsion harmonique. La figure de droite montre le spectre d'un signal de ce type analysé dans le domaine numérique. La distorsion engendrée apparaît clairement (avec une prédominance des harmoniques impaires) en addition de la fondamentale d'origine. Une fois le signal descendu au-dessous du niveau auquel l'élément binaire de poids faible (LSB) se déclenche, il n'y a plu aucune modulation. Du point de vue audible, on constate alors la disparition soudaine d'un signal très fortement distordu. Un signal sinusoïdal de plus haut niveau traverserait un plus grand nombre d'intervalles de quantification et générerait une plus grande quantité de valeurs d'échantillon non nulles.

Quand le niveau du signal augmente, l'erreur de quantification (toujours avec une valeur maximale de 0.5Q), devient de plus en plus petite comparée au niveau total du signal. La corrélation entre l'erreur et le signal diminue graduellement.





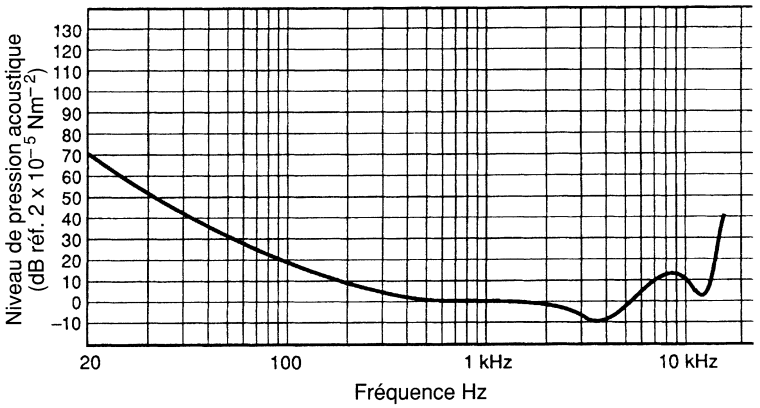
On considère maintenant un signal musical d'un niveau raisonnable. Les caractéristiques spectrales et l'amplitude d'un tel signal varient beaucoup : cela confère à l'erreur de quantification une nature assez aléatoire. En d'autres termes, l'erreur de quantification ressemble plus à du bruit qu'à de la distorsion, d'ou le terme de "bruit de quantification" couramment employé pour en décrire l'effet audible. L'analyse de la puissance de l'erreur de quantif ication (en assumant que sa nature se rapproche du bruit) montre une amplitude efficace électrique de Q sur racine de 12, où Q est l'incrément de tension représenté par un intervalle de quantification. Ainsi, le rapport signal/bruit d'un signal idéal de n bits peut être approximativement donné par : (6.02n + 1.76) dB. Ceci implique un rapport signal/bruit théorique approché d'un peu plus de 6 dB par bit. Un convertisseur 16 bits doit afficher un rapport signal/bruit autour de 98 dB, et un convertisseur 8 bits autour de 50 dB. Par ailleurs, d'autres erreurs sont introduites si le signal à échantillonner présente une amplitude supérieure à la plage de conversion. Les signaux dépassants cette limite sont durement écrêtés ce qui induit une distorsion très sévère.



Les signaux qui dépassent le niveau de pic sont durement écrêtés dans un système audionumérique. Il n'existe pas de valeurs disponibles pour représenter les échantillons.

Limitations Psycho-Acoustiques et Choix Techniques

La question de savoir quel taux d'échantillonnage et quelle résolution sont requis pour accéder à une qualité audio donnée trouvent certaines réponses en relation avec les capacités de l'oreille humaine, laquelle doit certainement être considérée comme l'arbitre ultime. L'audionumérique permet d'atteindre les limites de l'oreille humaine en termes de qualité sonore. Cependant, l'audionumérique, mal maîtrisé, peut "sonner" de façon très médiocre, et le terme numérique n'implique pas automatiquement une haute qualité sonore. Le choix des paramètres d'échantillonnage et des méthodes de mise en forme de bruit affecte la réponse en fréquence, la distorsion et la dynamique perçue. Les capacités de l'oreille humaine pourraient être considérées comme le standard en regard duquel la qualité des systèmes numériques serait évaluée. On peut en effet défendre l'idée que seuls comptent les distorsions et les bruits perceptibles par l'oreille. Il pourrait, par exemple, sembler pertinent de concevoir un convertisseur dont le plancher de bruit correspondrait au seuil de sensibilité de l'oreille. La figure suivante montre une courbe typique du seuil de sensibilité de l'oreille aux niveaux bas, indiquant le niveau de pression acoustique (SPL, Sound Pressure Level) requis pour qu'un son soit tout juste audible.



Il faut noter que l'oreille est plus sensible au milieu du spectre, autour de 4kHz, et moins sensible dans les zones limites inférieure et supérieure. Cette courbe est généralement appelée "champ audible minimum" ou encore "seuil de l'audition". Elle présente un niveau de pression acoustique de 0dB (réf. 20 Pa) à 1kHz. Il est toutefois important de se rappeler que le seuil d'audition de l'oreille humaine n'est pas une valeur absolue mais une valeur statistique. Cette notion est capitale pour toute recherche qui tente d'établir des critères d'audibilité, puisque certains sons, bien que 10dB inférieurs aux seuils admis, conservent une probabilité de perception qui peut avoisiner la certitude.

On peut définir la plage dynamique comme étant égale à la plage dynamique située entre le seuil d'audibilité et le plus fort son tolérable. Le plus fort son tolérable dépend de la personne ; toutefois, on considère généralement que le seuil de la douleur se situe entre les niveaux de la pression acoustique de 130 et 140 dB. La plage dynamique maximale absolue de l'oreille humaine se situe donc autour de 140dB à 1kHz, mais bien en deçà aux basses et hautes fréquences. On peut ensuite débattre pour savoir s'il est nécessaire d'enregistrer et de produire une plage dynamique aussi importante. Les travaux menés par Louis Fielder et Elizabeth Cohen ont tentés de définir la plage dynamique requise pour les systèmes audio de haute qualité : ils ont exploré les pressions extrêmes produites par des sources acoustique diverses et les ont comparées avec les planchers de bruit perceptible dans des conditions acoustiques réelles. En s'appuyant sur la théorie psycho-acoustique, Fielder a pu établir ce qui a une probabilité d'être entendu à diverses fréquences en termes de bruit et de distorsion, et a localisé les éléments limitant d'une chaîne acoustique typique. Il a défini la plage dynamique comme étant le "rapport entre le niveau de la valeur efficace d'un signal (RMS) maximal d'une onde sinusoïdale non distordue produisant des pics de niveau égaux à un niveau donné, et le niveau de la valeur efficace d'un signal (RMS) d'un bruit blanc limité à 20kHz dont le niveau sonore apparent serait le même que le bruit d'une chaîne audio donnée en l'absence de signal". Après quoi il a établi que le niveau tout juste audible d'un bruit dont la largeur de bande est de 20kHz est d'un niveau de pression acoustique d'environ 4 dB et que le nombre de prestations musicales produisent des niveaux de pression acoustique entre 120 et 129 dB au point d'écoute optimal. Il en a déduit que la plage dynamique nécessaire à une reproduction naturelle était de 122 dB. En prenant en compte les performances des microphones et les limitations des enceintes grand public, cette spécification est tombée à 115dB.

Le choix du taux d'échantillonnage détermine la largeur maximale de bande audio disponible. Un débat sévit concernant le choix d'un taux ne dépassant pas le strict nécessaire, à savoir le double de la fréquence audio la plus élevée pouvant être représentée. D'où le débat secondaire portant sur la plus haute fréquence audio utile. Par convention, il a été posé que la bande de fréquence audio s'étendait jusqu'à 20 kHz, ce qui entraîne des taux tout juste supérieurs à 40 kHz. Le choix s'est en fait porté sur deux fréquences d'échantillonnage standard comprises entre 40 et 50 kHz : le taux de 44,1 kHz du disque compact et le taux de 48 kHz dit "professionnel" bien qu'étant largement dépassé aujourd'hui. Ces fréquences sont entérinées par le standard AES5 de 1984. AES est l'abréviation d'Audio Engineering Society, organisme indépendant chargé de normaliser l'ensemble des applications audio. En fait le taux d'échantillonnage de 48 kHz avait été choisi pour offrir une certaine variation des vitesses de défilement des bandes électromagnétiques encore utilisées au début de l'audionumérique pour stocker les données numériques ; ainsi les risques de repliement du spectre étaient amoindris du fait de la marge offerte. La fréquence de 44,1 kHz a été établie plus tôt avec le lancement du disque compact. Par ailleurs ce taux génère 10 % de données en moins que le taux de 48 kHz, d'où une certaine économie.

On peut d'ailleurs ici s'interroger sur la provenance de cette valeur de 44,1 kHz au combien exotique dans la mesure où l'on cherchait simplement une fréquence supérieure au double de la plus haute fréquence audible. La réponse se trouve simplement dans le matériel dont disposaient les chercheurs à cette époque. En effet, aux premiers temps de la recherche audionumérique, les débits requis pour le stockage des données, d'environ 1 Mbit/seconde, étaient difficiles à atteindre. Les lecteurs de disquette les rendaient possible, mais leurs capacités étaient insuffisantes pour des enregistrements d'une certaine durée; aussi se tourna-t-on vers les enregistreurs vidéo. Ceux-ci furent adaptés en vue du stockage d'échantillons audio, en créant un signal dit pseudo-vidéo qui transportait des données binaires sous forme de niveau de noir et de blanc. La fréquence d'échantillonnage de tels systèmes fut conditionnée par le fait d'être en relation simple avec la structure et la fréquence des trames du standard vidéo utilisé, de façon qu'un nombre entier d'échantillons soient enregistrés par ligne utile. Les standards vidéo ont ainsi imposés cette fréquence de 44,1 kHz.

Comme on a pu le voir plus haut, le nombre de bits par échantillon défini le rapport signal/bruit ainsi que l'étendue dynamique d'un système audionumérique. On ne prend en compte que les systèmes en modulation par impulsions codées (PCM) linéaires. Depuis de nombreuses années, la modulation par impulsions codées linéaire 16 bits est considérée comme la norme pour les applications audio de qualité. C'est en effet le standard du disque compact, capable d'offrir une dynamique satisfaisante supérieure à 90 dB. Ce standard convient pour la plupart des cas mais ne satisfait pas à l'idéal de Fielder d'une dynamique de 122 dB pour une reproduction subjectivement exempte de bruit dans les systèmes professionnels. Accéder à cette dynamique requiert une résolution d'environ 21 bits. Il arrive souvent qu'une certaine "marge" avant saturation soit requise en enregistrement professionnel. En d'autres termes, une plage dynamique excédant le niveau d'enregistrement maximum nominal doit être disponible pour encaisser un éventuel dépassement. C'est une des raisons pour lesquelles les professionnels réclament des résolutions supérieures à 16 bits. Le passage à une résolution de 24 bits est aujourd'hui fortement engagé même si l'étendue dynamique excède les besoins psycho-acoustiques.