

MUS 1321 - Analyse, synthèse et traitement des sons

Caroline A. Traube et Olivier Bélanger

12 janvier 2011

Table des matières

1	Introduction à l'audio numérique	5
1.1	Rappels	5
1.1.1	Le son comme variation de la pression atmosphérique	5
1.1.2	Les son purs et les sons complexes	5
1.1.3	Représentation du son dans le domaine fréquentiel	6
1.2	De l'analogique au numérique	6
1.2.1	Le signal analogique	6
1.2.2	Le signal numérique	8
1.2.3	Les systèmes décimal, binaire et hexadécimal	9
1.2.4	La conversion analogique/numérique/analogique	12
1.2.5	L'échantillonnage	15
1.2.6	La quantification	18
1.2.7	L'avantage du numérique sur l'analogique	21
1.3	Introduction à l'analyse de Fourier	22
1.3.1	Représentation du signal dans le domaine fréquentiel	22
1.3.2	Les nombres complexes	24
1.3.3	Pourquoi y a-t-il des fréquences négatives dans un spectre?	27
1.3.4	Les transformées de Fourier	28
1.4	La théorie de l'échantillonnage	30

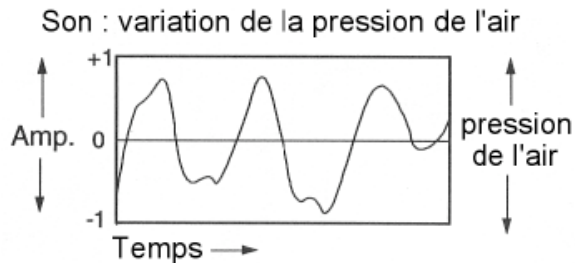
Chapitre 1

Introduction à l'audio numérique

1.1 Rappels

1.1.1 Le son comme variation de la pression atmosphérique

Le son est un phénomène psychophysique qui se produit lorsqu'un objet entre en vibration mécanique et cause une variation de la pression atmosphérique dans l'air entourant cet objet. Les variations de pression se propagent dans l'air (à la vitesse du son) et, le cas échéant, atteignent l'oreille de l'auditeur. L'amplitude (ou l'ambitus) de la variation de la pression atmosphérique donne la sensation de volume (ou intensité, force) au son.



La figure ci-dessus montre une représentation d'un son dans le domaine temporel (amplitude en fonction du temps). Lorsque la courbe est en haut du graphique, la pression atmosphérique est à son maximum (compression) et lorsqu'elle est au bas du graphique, elle est au minimum (raréfaction). L'amplitude de la forme d'onde est donc la quantité de variation de la pression atmosphérique. En général, l'amplitude se calcule à partir de l'axe 0 (position d'équilibre ou valeur moyenne des variations) jusqu'au point d'amplitude maximum de l'onde.

1.1.2 Les son purs et les sons complexes

La nature des vibrations acoustiques audibles peut être périodique ou non-périodique, ou encore une combinaison de ces deux natures. Les variations périodiques suivent un motif particulier, la forme d'onde, qui se répète dans le temps et donnent, en général, une sensation de hauteur alors que les variations non-périodiques donnent une sensation de bruit. Il est à noter

que les sons naturels sont presque toujours semi-périodiques, c'est-à-dire qu'ils correspondent à une combinaison de variations périodiques et non-périodiques.

Dans le cas de sons périodiques, la durée du motif récurrent de la forme d'onde est la période T . Le nombre de fois que cette période se répète en une seconde nous donne sa fréquence fondamentale f que l'on exprime en Hertz (Hz). Lorsque le nombre de « périodes-par-seconde » décroît, la sensation de hauteur du son baisse et vice-versa. Mathématiquement, on exprime la fréquence comme étant l'inverse de la période :

$$f = \frac{1}{T}$$

Si un son ne contient qu'une seule fréquence, on le dit sinusoïdal ou son pur. Bien que le son de la flûte s'en rapproche, les sons purs (strictement sinusoïdaux) ne se retrouvent pas dans la nature. Ce sont des sons de laboratoire. En général, les sons acoustiques contiennent des fréquences autres que la fréquence fondamentale. Celles-ci s'additionnent à la fréquence fondamentale pour former un son complexe, ce qui donne lieu à la perception du timbre d'un son. Les fréquences additionnelles sont appelées harmoniques ou partiels selon qu'elles sont des multiples entiers de la fréquence fondamentale ou pas.

1.1.3 Représentation du son dans le domaine fréquentiel

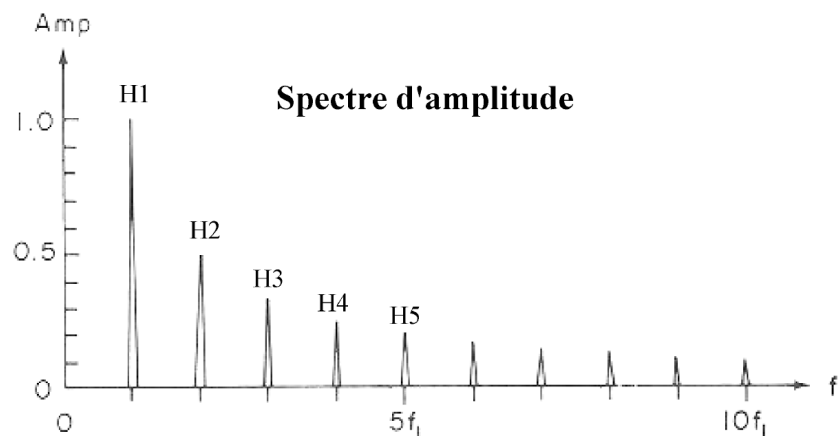
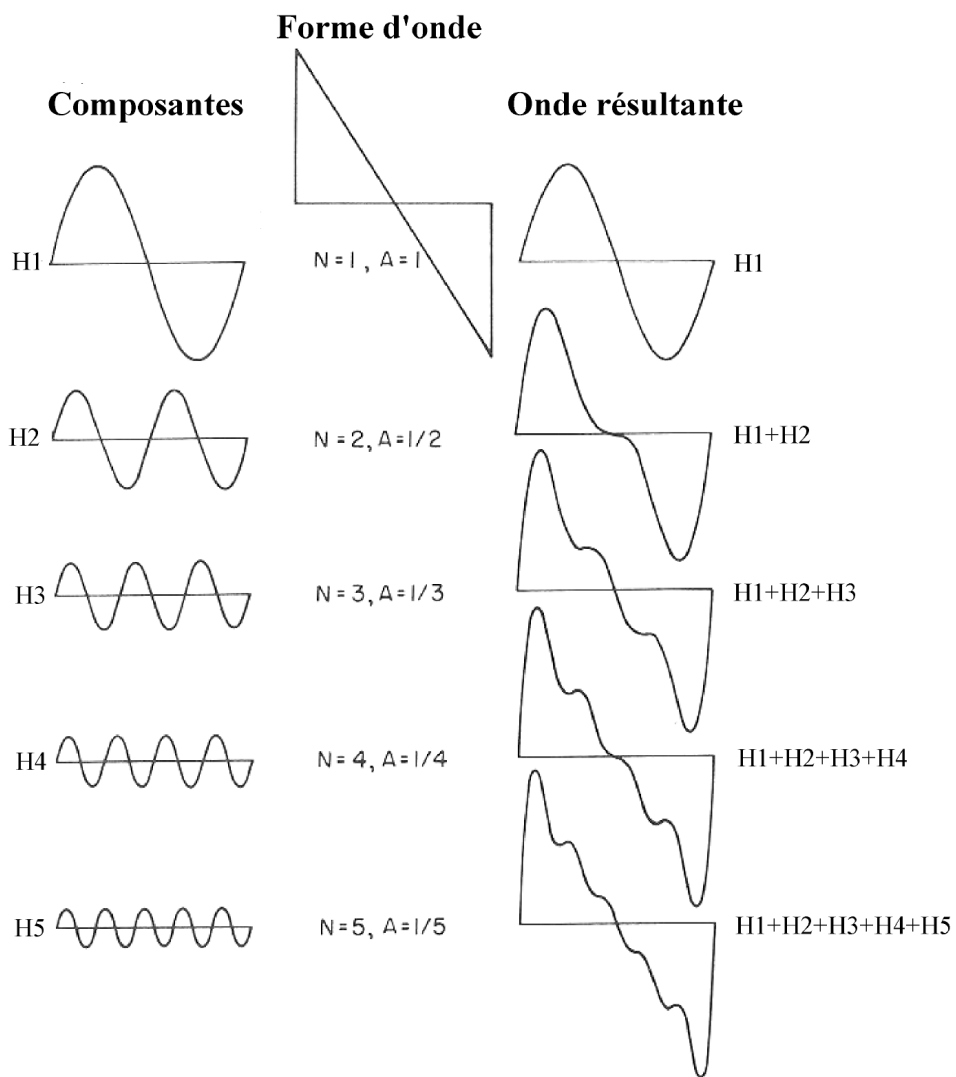
La représentation fréquentielle permet d'afficher les caractéristiques spectrales d'un signal audio. La figure ci-contre donne les représentations temporelle (forme d'onde) et fréquentielle (spectre d'amplitude) d'une onde en dent de scie. Comme la variation est périodique, on obtient un spectre de raies. La représentation fréquentielle permet de voir le contenu fréquentiel de la forme d'onde. Dans cette représentation, l'amplitude de chaque raie indique l'amplitude de chacune des fréquences présentes dans le signal (les harmoniques). La figure illustre aussi la reconstruction progressive de l'onde par superposition de ses composantes harmoniques (d'ordre 1 à 5).

1.2 De l'analogique au numérique

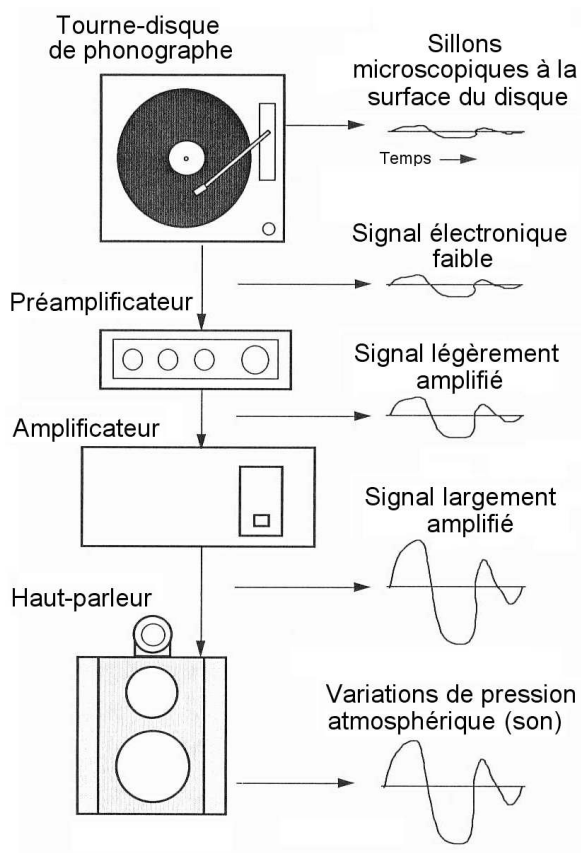
1.2.1 Le signal analogique

La variation de pression atmosphérique causée par un corps en vibration et produisant un son peut être captée par un microphone. Dans un microphone électrodynamique, une membrane liée à un aimant se déplace dans un mouvement de va-et-vient sous l'effet de la pression. L'aimant, à son tour, engendre un courant dit analogique dans le fil du microphone. La variation de la tension électrique correspondante est donc analogue à la variation de la pression de l'air. La caractéristique principale de ce signal en est sa continuité. On dira que c'est un signal continu.

La reproduction analogique du signal audio, bien qu'adéquate pour plusieurs applications, possède un défaut difficile à contourner : les différentes opérations effectuées sur un signal analogique s'accompagnent généralement d'une quantité non-négligeable de bruit. Ce bruit est notamment dû au fait que l'enregistrement analogique doit se faire via la magnétisation d'un support métallique, qui présente toujours un bruit de fonds dû à l'agitation thermique des particules magnétisées. D'autre part, quand on amplifie le signal, le bruit présent dans le signal est



également amplifié. Le problème est qu'il n'est pas vraiment possible de parfaitement séparer le bruit d'un signal analogique de manière à restituer ce dernier dans sa version originale non bruitée.



La figure ci-dessus illustre la chaîne de reproduction audio analogique.

1.2.2 Le signal numérique

La numérisation du signal peut aider à résoudre le problème de bruit de fond qui accompagne les signaux analogiques de façon souvent perceptible. La numérisation consiste à obtenir une représentation symbolique du signal analogique en lui appliquant successivement les opérations d'échantillonnage et de quantification. Le signal quantifié est représenté par une série de valeurs nommées bits (de *Binary Digit*), prenant les valeurs 0 ou 1 et se succédant sur le support magnétique ou autre (cf. sur disque compact).

1.2.3 Les systèmes décimal, binaire et hexadécimal

Le système décimal

Le système décimal que nous utilisons régulièrement pour compter utilise 10 symboles (0, 1, 2, 3, 4, 5, 6, 7, 8, 9). Un nombre décimal est composé d'une suite de ces symboles, chacun d'eux représentant une puissance de 10. Par exemple :

$$1984 = (1 \times 10^3) + (9 \times 10^2) + (8 \times 10^1) + (4 \times 10^0) = 1000 + 900 + 80 + 4$$

Le fait qu'il y ait 10 symboles dans le système le plus communément utilisé n'est pas dû au hasard : il vient du nombre de doigts que nous possédons (c'est d'ailleurs de là que vient le terme « digital », qui se traduit par « numérique » en français). On dira qu'on compte en « base 10 ».

Le système binaire

Quant à lui, le système binaire utilise seulement deux symboles : 0 et 1. On dira alors qu'on compte en « base 2 ». Un nombre binaire est composé d'une suite de bits, prenant les valeurs 0 ou 1, chaque bit représentant une puissance de 2 (0, 1, 2, 4, 8, 16, 32, ... de droite à gauche). On peut facilement calculer que le nombre binaire 1101 correspond au nombre décimal 13 :

$$(1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0) = 8 + 4 + 0 + 1 = 13$$

Avec n bits, on peut représenter 2^n valeurs différentes (de 0 à $2^n - 1$). Ainsi, avec 2 bits, on peut représenter $2^2 = 4$ valeurs différentes (de 0 à 3) :

Valeur décimale	0	1	2	3
Valeur binaire	00	01	10	11

Avec 8 bits, on peut représenter $2^8 = 256$ valeurs différentes (de 0 à 255). Un byte ou octet est un mot de 8 bits¹. Avec 16 bits, on peut représenter $2^{16} = 65\,536$ valeurs différentes (de 0 à 65535). Le tableau suivant donne la correspondance entre quelques nombres binaires et décimaux :

1. Contrairement à ce qu'on pourrait penser, un kilobyte ou kilo-octet vaut 1024 octets et non 1000. Ce facteur 1024 correspond à 2^{10} .

Valeur décimale	Décomposition en une somme de puissances de 2	Valeur binaire
	... 2 ⁶ 2 ⁵ 2 ⁴ 2 ³ 2 ² 2 ¹ 2 ⁰	
	... 64 32 16 8 4 2 1	
0		0
1		1
2		2 + 0
3		2 + 1
4		4 + 0 + 0
5		4 + 0 + 1
8		8 + 0 + 0 + 0
15		8 + 4 + 2 + 1
16		16 + 0 + 0 + 0 + 0
32		32 + 0 + 0 + 0 + 0 + 0
65535	32768 + ... + 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1	1111111111111111

MSB et LSB : Most Significant Bit et Least Significant Bit

Dans un nombre binaire, les bits les plus significatifs (Most Significant Bits), ou bits de poids fort, sont à gauche et les moins significatifs (Least Significant Bits), ou bits de poids faible, sont à droite. Les plus significatifs sont ceux qui entraînent les plus grands changements. Pour illustrer cela, considérons par exemple le nombre binaire 101 qui correspond à 5 en décimal. Le MSB est le 1 de gauche. S'il passe à 0, on obtient 001, c'est-à-dire 1 en décimal, ce qui représente une grande variation de valeur. Le LSB est le 1 de droite. S'il passe à 0, on obtient 100, c'est-à-dire 4 en décimal, ce qui représente une plus petite variation de valeur.

Cette notion est extrêmement importante pour comprendre le traitement numérique des données qui ne se fait pas toujours sur le mot en entier mais parfois sur les MSB seulement.

Application : le protocole de communication MIDI

Apparu en 1982, le « Musical Instrument Digital Interface » ou MIDI est un protocole de communication et de commande permettant l'échange de données entre instruments de musique électronique, un ou plusieurs de ces « instruments » pouvant être des ordinateurs. Il est géré par un comité international, l'International Midi Association.

Sous le terme MIDI sont regroupées plusieurs normes, relatives au protocole logique, à l'interface physique, au format de fichier et à l'attribution des sons. Les caractéristiques de communication du protocole de communication MIDI sont une vitesse de communication à 31,25 kilobits par seconde (Kbps)² et des mots de 8 bits de donnée.

Tous les messages MIDI sont constitués d'un octet de statut – *status byte* en anglais (dont le bit de poids fort, bit 7, est toujours à 1) suivi d'octets de données (dont le bit de poids fort, bit 7, est toujours à 0). De ce fait, les valeurs MIDI sont codées en réalité sur 7 bits et sont comprises entre 0 et 127 ($2^7 = 128$ valeurs).

2. Dans ce cas, comme il s'agit d'un taux de transfert de données, un kilobit vaut bien 1000 bits. Mais si on considère le stockage de données sur un disque, le kilobit correspond à 1024 bits.

Le système hexadécimal

Le système de numérotation hexadécimal est un système en base 16. Chaque chiffre peut être représenté par l'un des 16 symboles de la base. Ces symboles sont les dix premiers chiffres du système de numérotation décimal : 0, 1, 2, 3, 4, 5, 6, 7, 8 et 9 et les six premières lettres de l'alphabet : A, B, C, D, E et F (car on ne pourrait pas prendre les seize premiers nombres de la base 10 vu que les six derniers nombres sont composés de deux chiffres). En hexadécimal,

- le nombre 10 est donc représenté par la lettre A,
- le nombre 11 par la lettre B,
- le nombre 12 par la lettre C,
- le nombre 13 par la lettre D,
- le nombre 14 par la lettre E,
- le nombre 15 par la lettre F.

Comme pour le décimal et le binaire, une fois que tous les symboles de la base sont épuisés, deux chiffres sont nécessaires pour représenter le prochain nombre de la séquence. Le nombre décimal 16 est donc représenté par 10 en hexadécimal. En effet,

$$(1 \times 16) + 0 = 16$$

De même, 17(décimal)= 11(hexadécimal), et 18(décimal) = 12(hexadécimal) ainsi de suite jusqu'au nombre 31(décimal), correspondant à 1F(hexadécimal) car $(1 \times 16) + 15 = 31$. Le nombre hexadécimal suivant 1F dans la séquence est 20 qui vaut $(2 \times 16) + 0 = 32$ en décimal.

Ainsi, la suite des nombres exprimé en hexadécimal s'écrit :

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F,
 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 1A, 1B, 1C, 1D, 1E, 1F,
 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 2A, 2B, 2C, 2D, 2E, 2F,
 30, 31, 32, 33, etc.

Avec deux symboles, le plus grand nombre représentable est

FF

c'est-à-dire

$$(15 \times 16) + 15 = 255$$

en décimal. En binaire, ce nombre est représenté sur 8 bits :

1111 1111

On comprend alors l'utilité du système hexadécimal qui permet de représenter les nombres binaires avec beaucoup moins de symboles. Par exemple, le nombre hexadécimal

9A

représente le nombre binaire

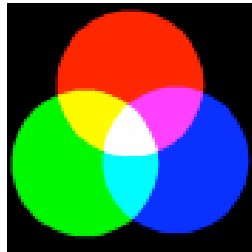
1001 1010

qui vaut

$$(128 + 16) + (8 + 2) = 144 + 10 = 154$$

en décimal. On notera aussi que 1001 0000 (144 en décimal) vaut 16 fois 0000 1001 (9 en décimal).

Notons finalement que l'hexadécimal est utilisé pour le codage des couleurs, à l'aide de 3 nombres hexadécimaux à deux symboles.



Le codage RVB (Rouge-Vert-Bleu ou RGB en anglais, pour Red-Green-Blue) affiche directement les couleurs sur un moniteur par synthèse additive qui consiste à composer une couleur via les 3 couleurs primaires par addition de lumière. Les trois couleurs primaires sont Rouge, Vert, Bleu et les trois couleurs secondaires sont Cyan, Magenta, Jaune.

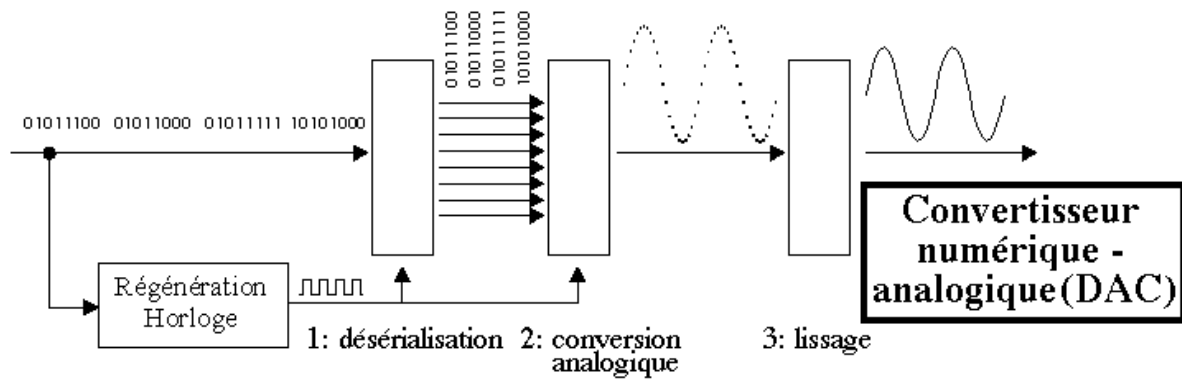
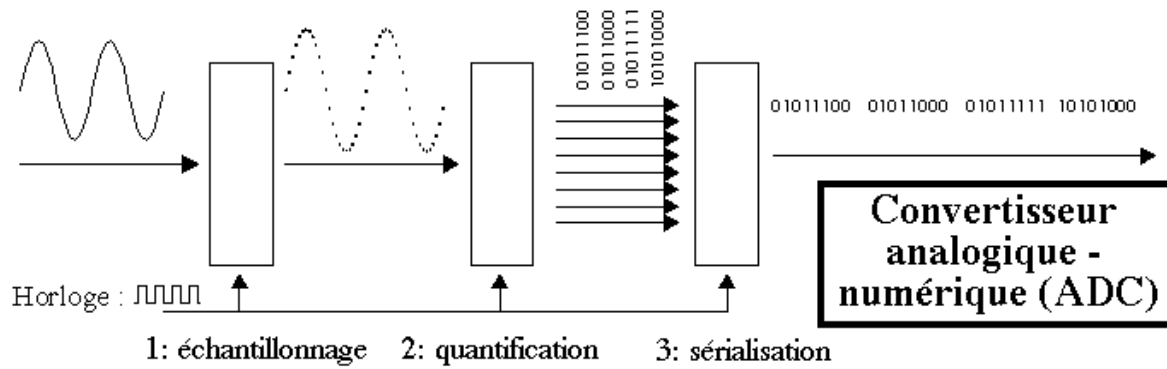
Toute couleur peut être notée selon sa proportion en rouge, en vert et en bleu. L'addition des trois couleurs primaires donne une lumière blanche (100% pour chaque composante) et le noir est une absence de couleur (0% pour chaque composante).

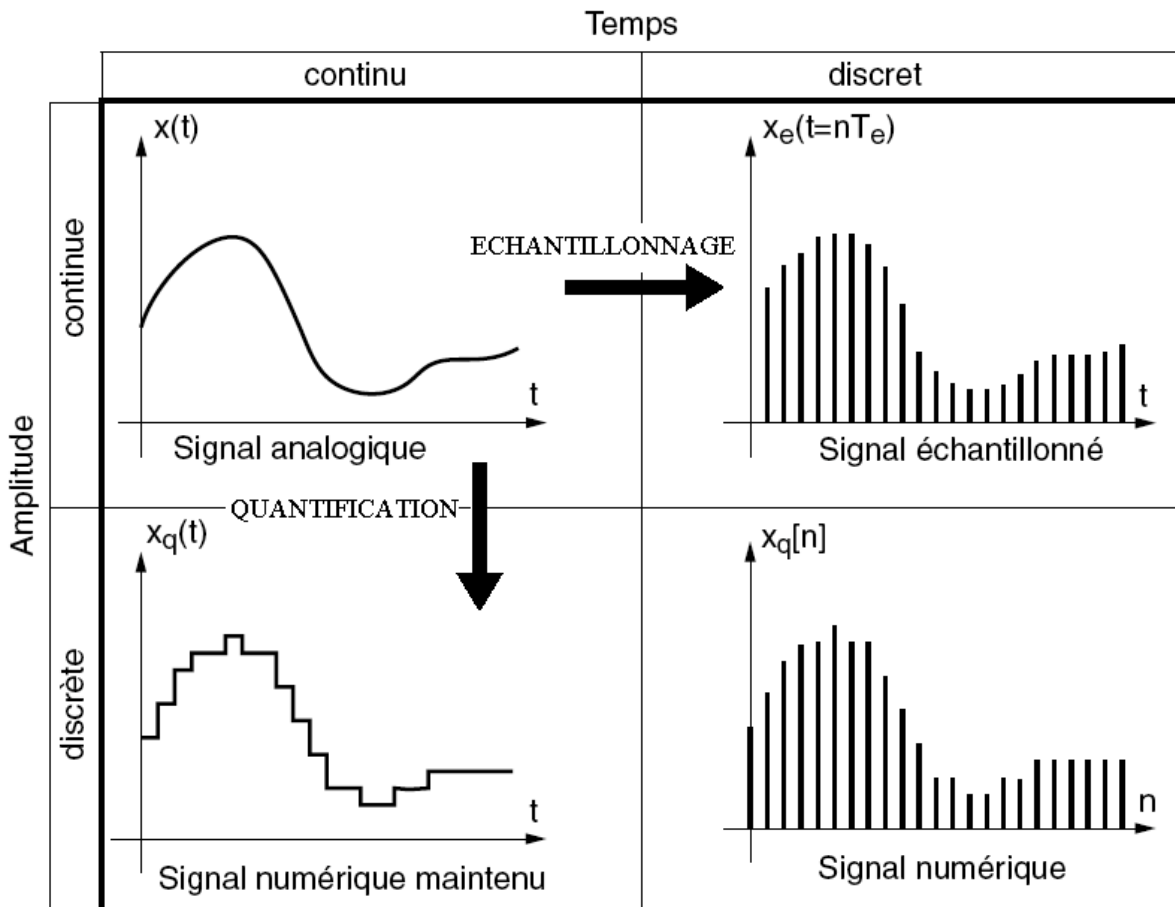
Chaque composante est codée sur un octet ; elle peut prendre une valeur comprise entre 0 (0%) et 255 (100%), c'est-à-dire, en adoptant la notation hexadécimale, entre 00 et FF. On va généralement coller les trois valeurs l'une contre l'autre, et faire précéder l'ensemble d'un symbole dièse pour indiquer qu'il s'agit de valeurs en hexadécimal : #xxyyzz. Ainsi, le rouge primaire est représenté par #FF0000, le vert primaire par #00FF00 et le bleu primaire par #0000FF.

1.2.4 La conversion analogique/numérique/analogique

Pour convertir un signal analogique en un signal numérique, on a recours à un convertisseur analogique-numérique ou ADC (Analog-to-Digital Converter). Cette conversion comprend deux étapes principales : l'échantillonnage et la quantification. Ces deux opérations « discrétisent » le signal analogique, c'est-à-dire qu'elles transforment un signal continu en un signal discret, autant suivant la dimension du temps (échantillonnage, au rythme d'une horloge) que celle de l'amplitude (quantification).

Pour restituer le signal analogique à partir de sa représentation binaire, on a recours à un DAC (Digital-to-Analog Converter). Ces convertisseurs (DAC et ADC) ont pour caractéristique de mesurer, en général plusieurs milliers de fois par seconde, la valeur du signal et de convertir cette valeur soit en un courant électrique (DAC) ou un nombre binaire (ADC).





1.2.5 L'échantillonnage

L'échantillonnage, première étape d'une conversion analogique-numérique, est défini par sa fréquence : la fréquence d'échantillonnage (exprimée en Hz) qui est l'inverse de l'intervalle de temps entre chaque prélèvement sur le signal analogique. C'est donc la fréquence de l'horloge numérique qui contrôle l'opération d'échantillonnage. Les symboles de la fréquence ou taux d'échantillonnage sont f_e (e pour échantillonnage) ou f_s (s pour sampling), ou encore $\omega_s = 2\pi f_s$ (la fréquence angulaire) ou encore SR (pour Sampling Rate). La période d'échantillonnage T_s est l'inverse de la fréquence d'échantillonnage : $T_s = 1/f_s$.

Le théorème de Nyquist

Le théorème de Nyquist, théorème fondamental de l'audio numérique, nous apprend que pour pouvoir représenter numériquement un signal de fréquence maximale f_{max} , on doit l'échantillonner à un taux qui vaut au moins le double de cette fréquence :

$$f_s > 2f_{max}$$

En effet, chaque période d'un signal sinusoïdal doit être représentée par au moins deux points : un point pour la valeur maximale de la période et un second point pour la valeur minimale. Ainsi, puisque la plage fréquentielle de l'oreille humaine s'étend de 20 Hz à 20 000 Hz, un système de conversion idéal devra prélever des échantillons au moins 40 000 fois par seconde.

Fréquences d'échantillonnage courantes

Les deux fréquences d'échantillonnages les plus courantes en audio actuellement sont 44100 Hz (CD) et 48000 Hz (DAT). À 48000 Hz, le signal analogique est échantillonné toutes les $1/48000 = 0,00002083$ s = 0,02083 ms = 20,83 μ s.

Fréquences (Hz)	Applications
8000	Téléphonie, basse qualité audio standard
11025	Un quart du taux CD (Macintosh pour audio basse qualité)
16000	Certaines applications téléphoniques
18900	Standard CD-ROM/XA et CD-I, qualité basse à moyenne
22050	Moitié du taux CD (Macintosh)
32000	Systèmes Nicam 3, Nicam 728 et mode DAT longue durée
37800	Taux du CD-ROM/XA et CD/I, qualité intermédiaire
44056	Légère modification du taux 44100, utilisé pour s'aligner sur le taux NTSC
44100	Taux du CD, très exploité en milieu professionnel sur de nombreux formats
47952	À éviter, rencontré avec de la vidéo NTSC
48000	DAT, taux professionnel certifié par l'AES5-1984
88200 et 96000	Double des standards 44100 et 48000
176400 et 192000	Triple des standards 44100 et 48000

88200 et 96000 Hz, les doubles des standards 44100 et 48000 Hz, sont en train de devenir le standard pour la très haute qualité professionnelle.

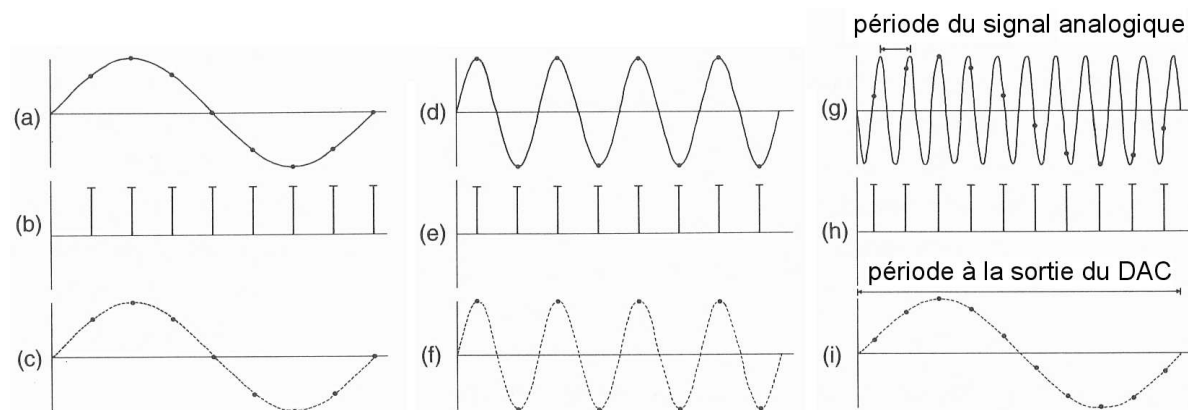
Les fréquences de 44056 Hz et 44100 kHz ont été choisies en fonction de leur rapport avec les fréquences de balayage des standards de la télévision. En effet, à cause de l'utilisation fréquente des magnétoscopes comme magnétophones maîtres, il est devenu essentiel que la même horloge effectue le contrôle du balayage horizontal de l'image et de l'échantillonnage. Ainsi :

- NTSC (Amérique du Nord) : $525\text{-}35 \text{ lignes} \times 29,97 \text{ trames/s} \times 3 \text{ (sync vertical)} = 44056$
- PAL/SECAM (Europe) : $625\text{-}37 \text{ lignes} \times 25 \text{ trames/s} \times 3 \text{ (sync vertical)} = 44100$

[Note : NTSC = National Television Standard/Systems Comittee, PAL = Phase Alternating Line, SECAM = Séquentiel Couleur Avec Mémoire]

Repliement du spectre ou aliasing

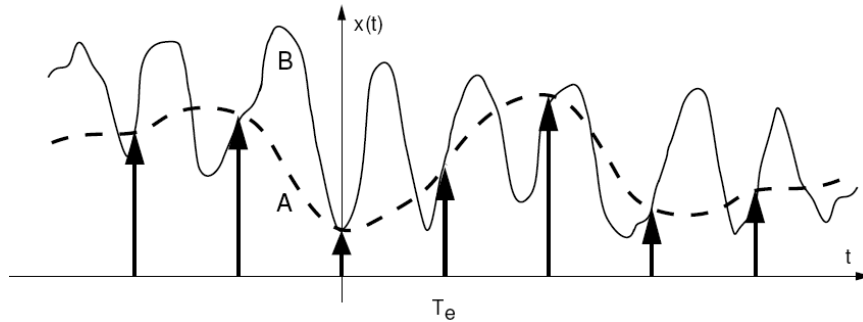
Selon le théorème de Nyquist, dans le cas de la fréquence d'échantillonnage du standard CD qui vaut 44100 Hz, le contenu fréquentiel des signaux audio échantillonnés à cette fréquence devra donc être limité à 22050 Hz pour assurer une représentation numérique adéquate. Si la condition de Nyquist n'est pas respectée, il se produit un repliement du spectre (aliasing en anglais). Le spectre original s'en trouve altéré par l'apparition de nouvelles composantes. En voici une illustration. Sur la figure ci-dessous, la première ligne (a, d, g) représente des signaux



sinusoïdaux à l'entrée d'un convertisseur analogique-numérique (ADC), la deuxième ligne (b, e, h) représente les trains d'impulsions à la fréquence d'échantillonnage, et la troisième ligne (c, f, i) donnent les signaux reconstitués à la sortie du convertisseur numérique-analogique (DAC). Considérons un signal complexe composé des trois fréquences représentées sur les figures (a), (d) et (g) respectivement.

- Dans le premier cas (à gauche), la fréquence du signal est bien inférieure à la fréquence d'échantillonnage. Le signal est parfaitement reconstruit par le convertisseur DAC.
- Le deuxième cas (au centre) est le cas limite : la fréquence du signal vaut la moitié de la fréquence d'échantillonnage (c'est la fréquence de Nyquist). Il y a deux échantillons par période, ce qui est juste assez pour permettre la reconstruction du signal de départ.
- Dans le troisième cas (à droite), la condition de Nyquist n'est pas satisfaite. Le prélèvement des échantillons ne se fait pas assez vite. La fréquence d'échantillonnage devrait valoir au

moins le double de la fréquence du signal. Comme ce n'est pas le cas, il en résulte que le signal échantillonné n'est pas reconstruit correctement. En lieu et place du signal de départ apparaît une fréquence fantôme. C'est ce que l'on appelle une « fréquence de repliement », résultat d'un sous-échantillonnage du signal.



La figure ci-dessus illustre bien l'ambiguïté qui peut résulter de l'échantillonnage d'un signal qui varie beaucoup trop rapidement par rapport au taux d'échantillonnage. Le signal B échantillonné sera reconstruit par le DAC sous la forme du signal A, qui est assez différent.

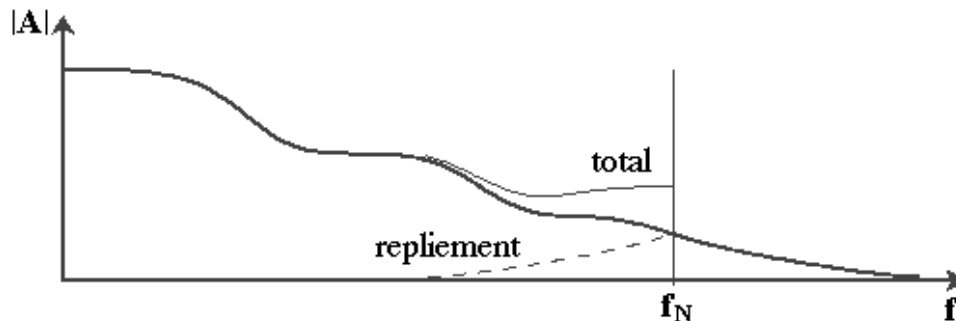
Les fréquences de repliement sont calculées en soustrayant les fréquences du signal (f_x) qui dépassent la fréquence de Nyquist de la fréquence d'échantillonnage (f_s) :

$$f_r = f_s - f_x$$

Ainsi, si on tente d'échantillonner un signal de 26 000 Hz à un taux de 40 000 Hz, on obtiendra une fréquence de repliement de

$$40000 - 26000 = 14000 \text{ Hz}$$

La figure ci-dessous illustre l'effet du repliement sur l'ensemble d'un spectre d'amplitude dont la partie supérieure dépasse la fréquence de Nyquist.



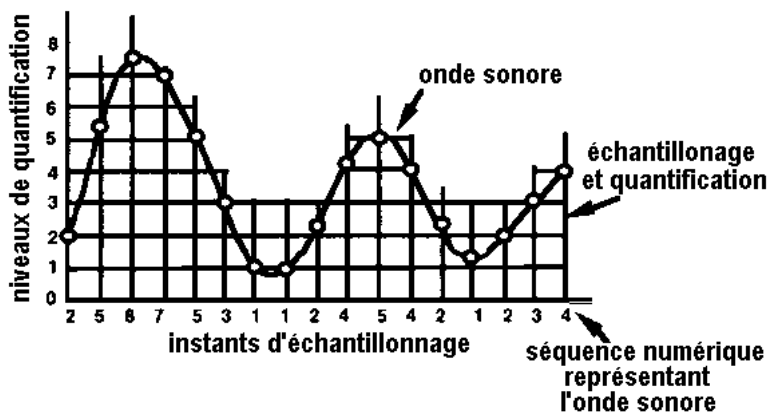
Dans le domaine visuel, le repliement de spectre peut se produire sous l'effet d'une lumière stroboscopique ou au cinéma. Dans la mesure où l'image cinématographique constitue également un exemple de signal échantillonné, l'effet bien connu de « la roue qui tourne à l'envers » rend

le phénomène de repliement du spectre visible et donc plus concret. Au cinéma, les images sont en principe présentées 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 nombre d'images par secondes. 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 (qui vaut, rappelons-le, la moitié de la fréquence d'échantillonnage). On effectue cela à l'aide d'un filtre passe-bas, appelé filtre de garde ou filtre anti-repliement (ou encore filtre anti-aliasing).

1.2.6 La quantification

La seconde étape d'une conversion analogique-numérique est la quantification. La représentation numérique d'un échantillon utilise un nombre fini de bits. Avec n bits, on peut représenter 2^n valeurs différentes. La quantification a pour effet d'arrondir l'amplitude de chaque échantillon à l'une de ces 2^n valeurs, comme l'illustre la figure ci-dessous.



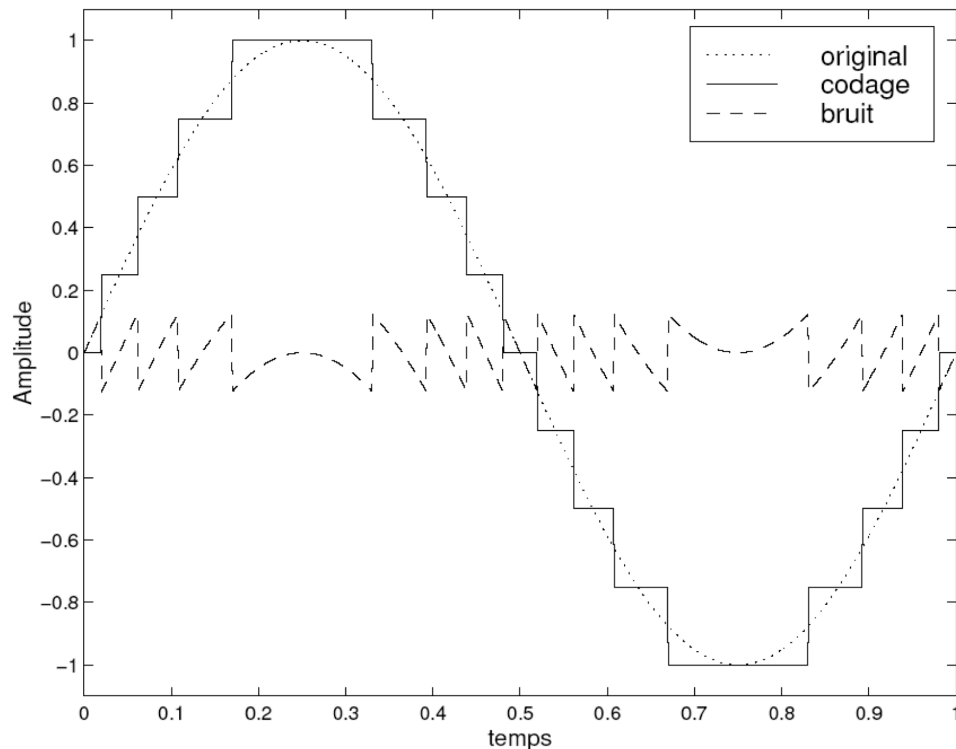
Avec 16 bits (format CD), on peut représenter $2^{16} = 65\,536$ valeurs dont 32 767 valeurs positives et 32 768 valeurs négatives (et une valeur pour le zéro). Si on code une valeur supérieure à ces limites, il se produit un écrêtage (clipping). Ce phénomène s'accompagne toujours d'une forte distorsion et se produit lorsque le niveau d'entrée est trop élevé. La technologie numérique est beaucoup moins tolérante face aux écrêtages que l'analogique (qui sature au lieu d'écrêter de façon radicale).

Le nombre de bits de quantification détermine donc la précision en amplitude ou la dynamique de la conversion (alors que la fréquence d'échantillonnage détermine la précision tempo-

relle de la conversion). En principe, chaque bit donne un ambitus dynamique de 6 dB. En effet, en ajoutant un bit à la représentation binaire de l'amplitude, on peut représenter deux fois plus de valeurs distinctes (on passe de $2n$ à $2n + 1$), ce qui donne une dynamique augmentée de 6 dB. Dans le cas du standard CD, la résolution est de 16 bits La plage dynamique d'un CD est donc d'environ 96 dB ($6 \times 16 = 96$). Ce chiffre est toutefois théorique et est sujet à des limitations.

Le bruit de quantification

Pour que le signal soit fidèlement reproduit, on doit non seulement l'échantillonner à un taux suffisant mais aussi avec un nombre suffisant de bits pour pouvoir rendre l'amplitude de chaque valeur la plus précise possible. Si les valeurs sont représentées dans un ambitus trop faible, les valeurs quantifiées diffèrent significativement des valeurs échantillonnées et le signal est entaché d'un bruit de quantification, comme illustré sur la figure suivante.



Une solution à ce problème est évidemment de s'assurer que les signaux sont échantillonnés avec un nombre suffisant de bits pour chaque prélèvement.

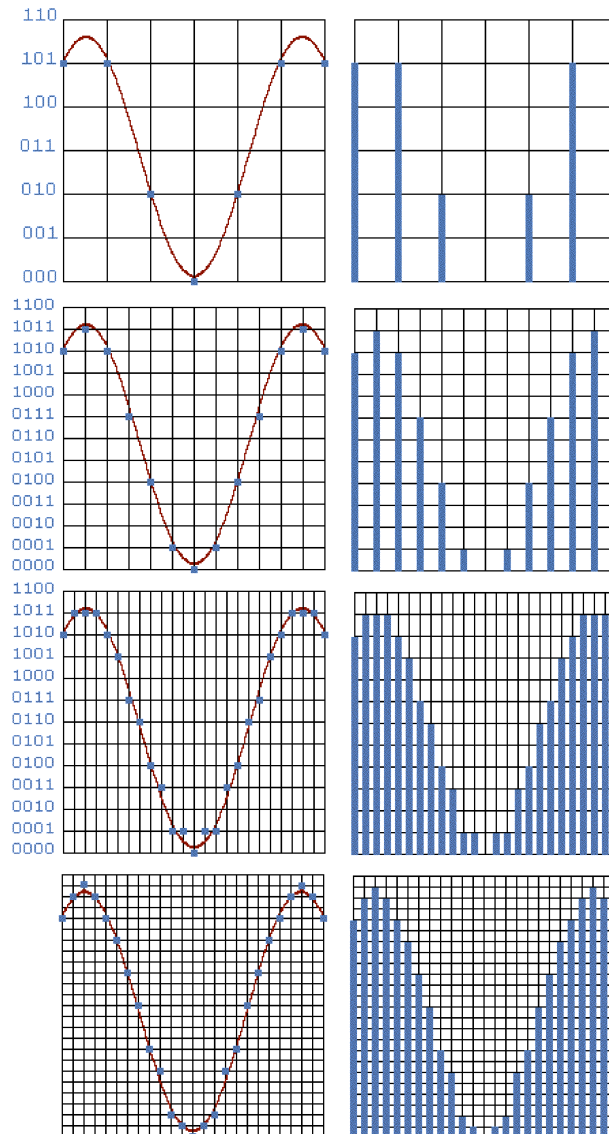
La réduction du bruit de quantification par dithering

La bruit de quantification devient particulièrement problématique pour les signaux de basses fréquences et de faibles amplitudes. La quantification peut avoir pour effet de transformer un signal sinusoïdal de basse fréquence en une onde carrée (dans le cas où l'amplitude est inférieure au

pas de quantification), ce qui introduirait de la distorsion harmonique (apparition des multiples de la fréquence).

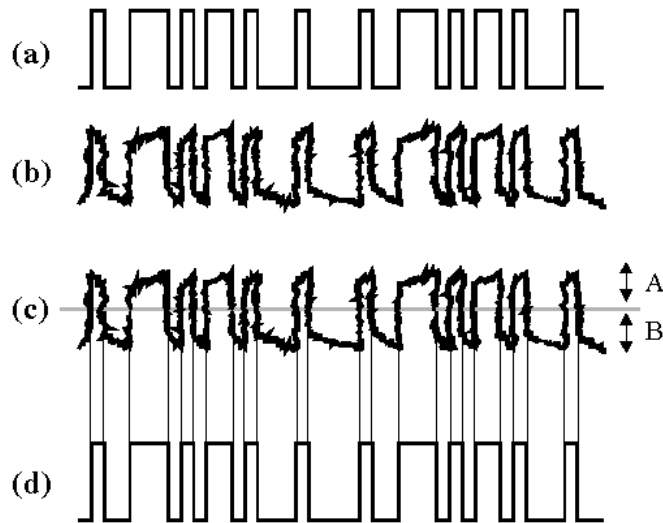
Il est possible de réduire le bruit de quantification en ajoutant au signal numérique un bruit blanc à très faible niveau (généralement à $3/4$ du pas de quantification). C'est ce que l'on appelle le bruit de dispersion ou *dither*. L'opération de dithering permet de linéariser le convertisseur numérique/analogique. Sous l'effet du bruit de dispersion, la distorsion de quantification devient ainsi un signal aléatoire de type bruit, quelque soit le signal, ce qui est beaucoup moins gênant et perceptible qu'une distorsion.

La figure ci-dessous illustre différents taux d'échantillonnage et de quantification sur la numérisation d'un signal sinusoïdal.



1.2.7 L'avantage du numérique sur l'analogique

Une fois numérisé, le signal devient plus robuste et est moins affecté par les bruits et parasites. C'est l'avantage du numérique sur l'analogique. La figure suivante en donne une illustration.



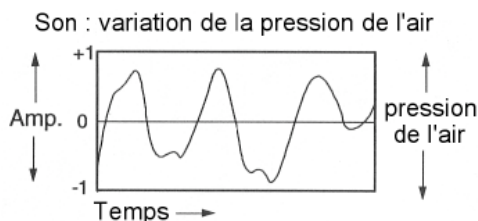
- Le signal (a) est un signal analogique contenant l'information correspondant à une série de bits (0100111101011 ...).
- En (b), c'est le même signal auquel s'est ajouté un bruit de fond (soit le signal a été transmis dans un canal bruité, comme un câble qui capte les ondes électromagnétiques environnantes, soit le signal a été enregistré sur un support qui comporte un bruit de fond, comme une bande magnétique de DAT par exemple).
- En (c), on procède à la régénération du signal numérique en le comparant à un seuil qui peut, par exemple, être à mi-distance entre la valeur 1 (palier supérieur) et la valeur 0 (palier inférieur) : si le signal est au-dessus du seuil (zone A), c'est un 1, et si le signal est en-dessous du seuil (zone B), c'est un 0.
- En (d), on retrouve le signal de départ, parfaitement régénéré. Le bruit a pu être complètement supprimé.

Bien sûr, il peut arriver que la dégradation du signal soit telle qu'un 0 soit pris pour un 1 et vice versa. Pour se prévenir de ces erreurs, on introduit des bits de correction (bits de parité à la suite de chaque octet par exemple).

1.3 Introduction à l'analyse de Fourier

1.3.1 Représentation du signal dans le domaine fréquentiel

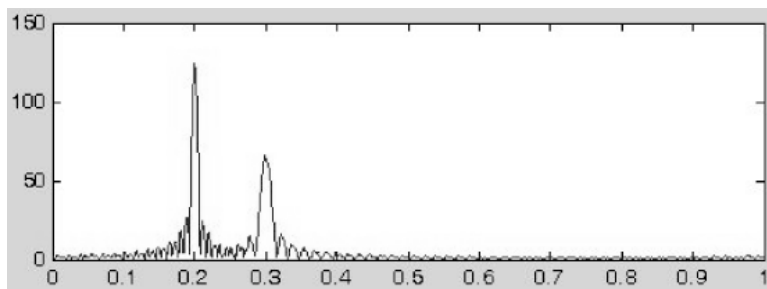
Dans le domaine temporel, une onde acoustique est représentée par un signal réel, c'est-à-dire qu'on a une variable réelle (qui peut prendre des valeurs à décimales) qui évolue au cours du temps. Au cours d'une période, on décrit la forme d'onde. Sur un graphique, dans le domaine temporel, on représente donc l'amplitude (A) en fonction du temps (t).



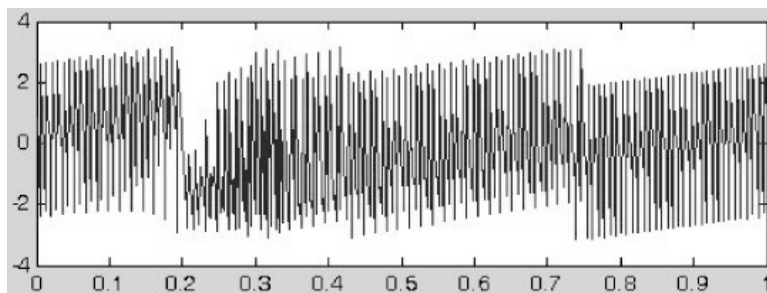
Grâce à la transformée de Fourier, on peut obtenir une représentation de l'onde dans le domaine fréquentiel. La transformée de Fourier d'un signal réel est un signal complexe, c'est-à-dire un signal à deux dimensions car chaque composante est définie par deux paramètres : son amplitude (A) et sa phase (ϕ).

Sur un graphique à deux dimensions, dans le domaine fréquentiel, on représente donc

- l'amplitude (A) en fonction de la fréquence (f) : c'est le **spectre d'amplitude**

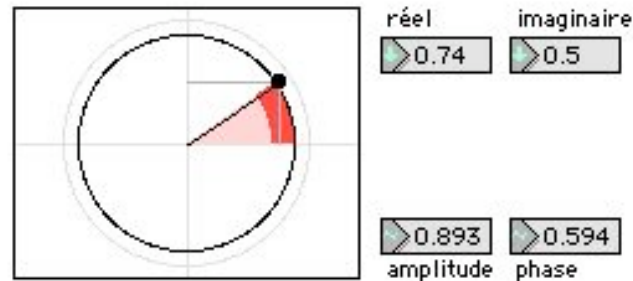


- la phase (ϕ) en fonction de la fréquence (f) : c'est le **spectre de phase**



Mais si on veut tout représenter sur le même graphique, on a besoin de deux dimensions au lieu d'une. C'est ainsi qu'on utilisera le plan complexe dont l'axe horizontal est l'axe réel et l'axe

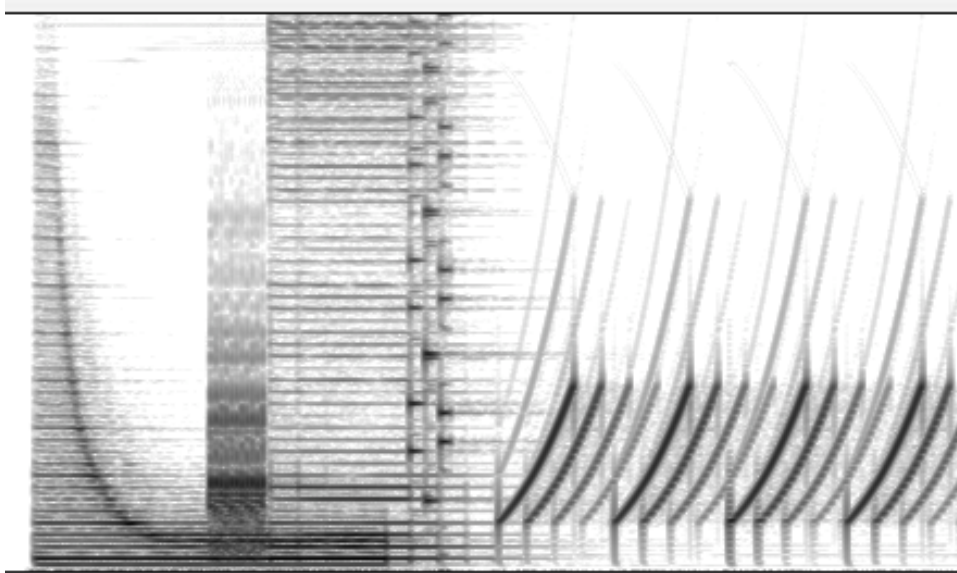
vertical est l'axe imaginaire. Sur la figure ci-dessus, représentant un vecteur délimité par un



point sur un cercle :

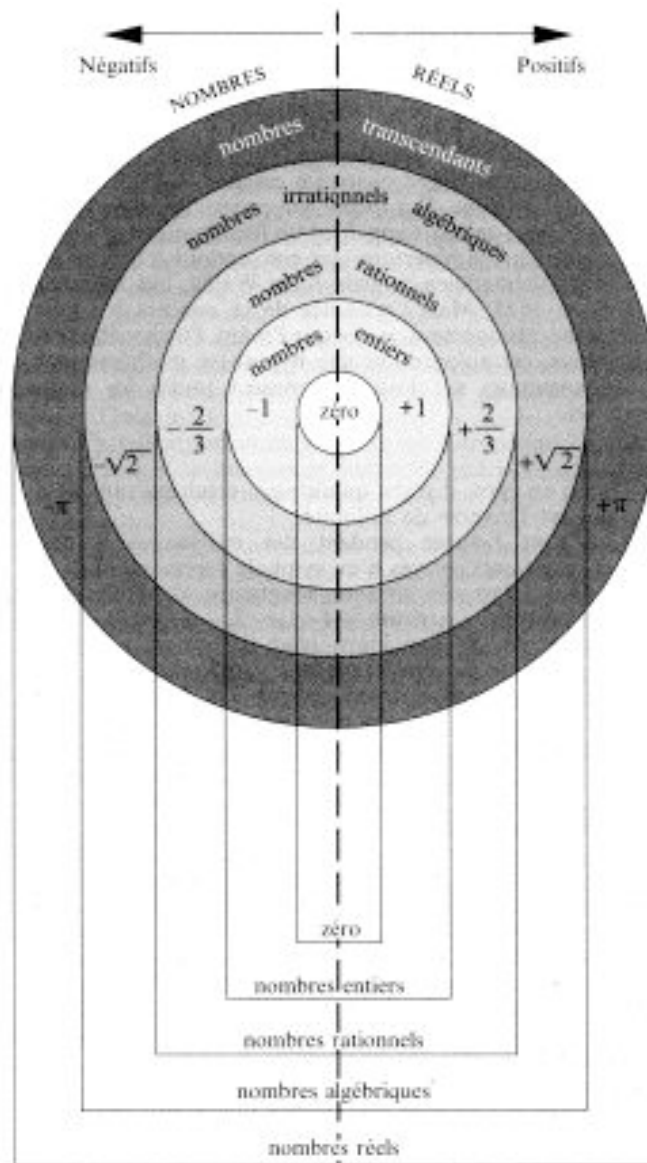
- l'amplitude du vecteur est l'amplitude de la composante spectrale
- l'angle initial du vecteur par rapport à l'horizontale correspond à la phase initiale
- la vitesse avec laquelle le vecteur tourne correspond à la fréquence de la composante (en Hz ou radians par seconde).

Il est aussi possible de représenter trois dimensions du signal - la fréquence, le temps et l'intensité - sur un même graphique, et ce, sous la forme d'un spectrogramme (ou sonagramme), comme l'illustre la figure ci-dessous (analyse des premières secondes de "The Robots" de Kraftwerk). Le spectrogramme représente la fréquence sur l'axe vertical et le temps sur l'axe horizontal, de manière similaire à une partition (le temps s'écoule de gauche à droite). L'intensité est quant à elle représentée par le niveau de gris des points du spectrogramme (un point plus foncé correspond à une intensité plus élevée). Notons que l'information de phase n'est pas représentée par le spectrogramme.



1.3.2 Les nombres complexes

Avant de présenter les nombres complexes, rappelons comment les mathématiciens ont défini les ensembles ou familles de nombres, regroupant des nombres ayant les mêmes caractéristiques. Ils ont été imaginés vers le début du siècle par des mathématiciens allemands et italiens. Pour représenter le système des ensembles de nombres, on a recours à une représentation imagée du principe des poupées russes : chaque ensemble est compris dans un autre ensemble, des nombres naturels aux nombres complexes, en passant par les entiers relatifs, les nombres décimaux, les nombres rationnels et les nombres réels.



- **N** : de *naturale* en italien, l'ensemble des entiers naturels est l'ensemble des nombres entiers consécutifs supérieurs ou égaux à 0. L'ensemble N fut défini par le mathématicien italien Peano (1858-1932).
- **Z** : de *zahl* (nombre) en allemand, l'ensemble des entiers relatifs défini par le mathématicien allemand Dedekind (1831-1916) est l'ensemble des entiers positifs ou négatifs.
- **D** : les nombres décimaux - c'est l'ensemble des nombres avec un nombre fini de décimales. L'ensemble D est une notation française issue de la pédagogie des années 1970.
- **Q** : de *quotiente* en italien, l'ensemble des nombres rationnels, défini par Peano, est l'ensemble des nombres pouvant s'écrire sous la forme d'un quotient.
- **R** : les nombres réels - c'est l'ensemble de tous les nombres, incluant le nombre π qui est irrationnel (il ne peut s'exprimer sous forme d'un quotient)
- **C** : les nombres complexes - c'est l'ensemble des nombres de la forme $a + ib$ où i est le nombre imaginaire.

Le nombre imaginaire i est défini comme étant la racine carrée de -1 :

$$i = \sqrt{-1}$$

Ce nombre est dit imaginaire car la racine carrée d'un nombre négatif n'existe pas dans l'ensemble des nombres réels (Note : le nombre imaginaire est parfois représenté par la lettre j , surtout en électricité car le symbole I est déjà utilisé pour représenter le courant électrique).

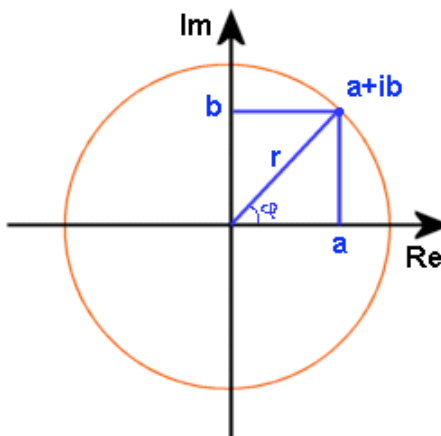
Un nombre complexe est un nombre qui comprend une composante réelle et une composante imaginaire. Il est ainsi représenté par l'équation :

$$z = a + ib$$

où z est un point de coordonnées cartésiennes (a , b)

a est la coordonnée suivant l'axe réel (horizontal)

b est la coordonnée suivant l'axe imaginaire (vertical).



Autrement dit, multiplier par le nombre imaginaire i revient à appliquer une rotation de 90 degrés (dans le sens anti-horlogique). On peut aussi représenter un nombre complexe suivant ses coordonnées polaires. Les coordonnées sont alors l'amplitude r et l'angle ϕ .

En résumé :

(a, b) sont les coordonnées cartésiennes

a est la partie réelle "Re" (coordonnée horizontale)

b est la partie imaginaire "Im" (coordonnée verticale)

(r, ϕ) sont les coordonnées polaires :

r est l'amplitude (rayon du cercle où se trouve le point z)

ϕ = phase (angle du vecteur par rapport à l'horizontale)

Ces coordonnées permettent de spécifier un nombre complexe dans le plan complexe. Pour convertir des coordonnées polaires (r, ϕ) en coordonnées cartésiennes (a, b) , on applique les formules suivantes :

$$a = r \cos \phi$$

$$b = r \sin \phi$$

À l'inverse, pour convertir des coordonnées cartésiennes (a, b) en coordonnées polaires (r, ϕ) , on applique les formules suivantes :

$$r = \sqrt{a^2 + b^2}$$

qu'on déduit du théorème de Pythagore, et

$$\phi = \arctg\left(\frac{b}{a}\right)$$

qui est l'arctangente du rapport de b sur a . Par exemple, un nombre complexe de coordonnées polaires $(1, 60^\circ)$ a pour partie réelle

$$a = \cos 60^\circ = 0,5$$

et pour partie imaginaire

$$b = \sin 60^\circ = \sqrt{3}/2 = 0,866$$

Ainsi, ce nombre complexe a pour expression

$$0,5 + i0,866$$

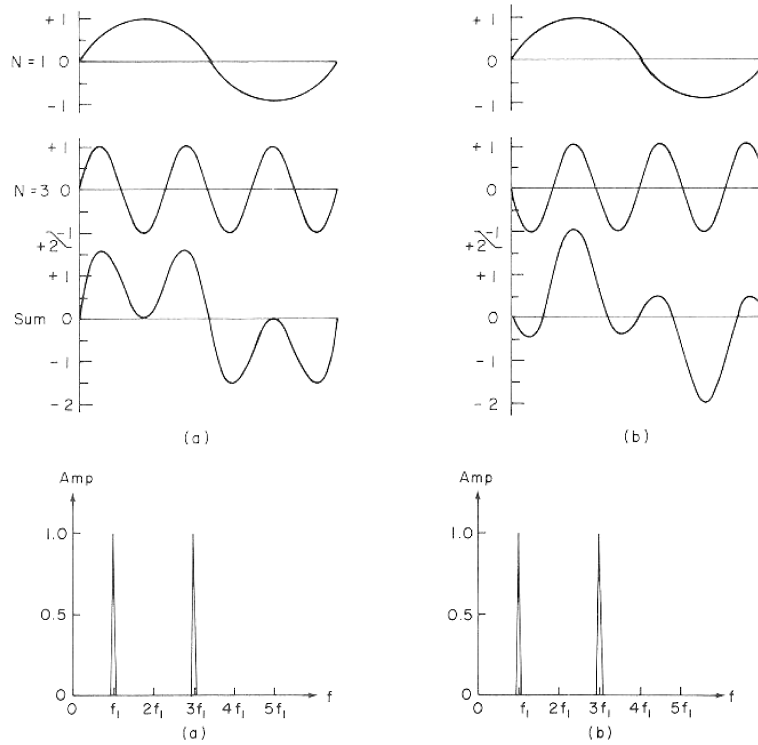
De même, le nombre complexe opposé (symétrique par rapport à l'axe réel) de coordonnées polaires $(1, -60^\circ)$ a pour expression

$$0,5 - i0,866$$

Et le nombre complexe symétrique par rapport à l'axe imaginaire, de coordonnées polaires $(1, 120^\circ)$, a pour expression

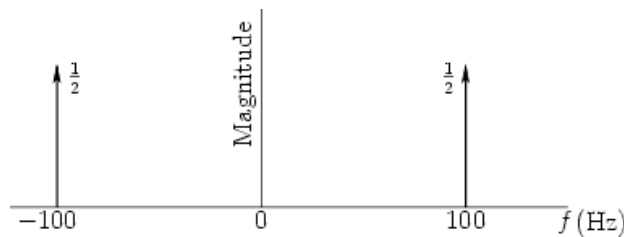
$$-0,5 + i0,866$$

Notons qu'il est important de tenir compte de la phase car la forme d'onde peut être très différente si l'on modifie ce paramètre, comme l'illustre la figure ci-dessus, représentant la somme de deux ondes sinusoïdales de même amplitude. Dans le cas (a), les deux ondes ont une phase initiale nulle. Dans le cas (b), la seconde onde est inversée. On peut constater que la forme d'onde de la somme est affectée par le changement de phase, bien que les spectres d'amplitude demeurent identiques dans les deux cas.



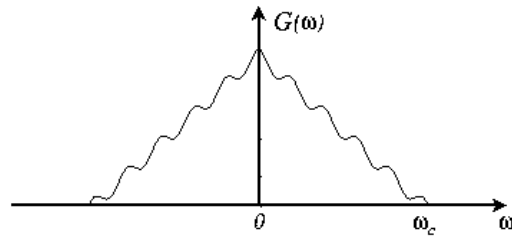
1.3.3 Pourquoi y a-t-il des fréquences négatives dans un spectre ?

Dans le cours d'acoustique, on a toujours représenté les spectres avec des fréquences positives seulement. Par exemple, on sait qu'une sinusoïde est représentée par un spectre avec une raie spectrale à la fréquence de cette onde. En fait, il y en a deux : une onde sinusoïdale de fréquence 100 Hz présente un spectre avec deux raies, une à 100 Hz et l'autre à -100 Hz.

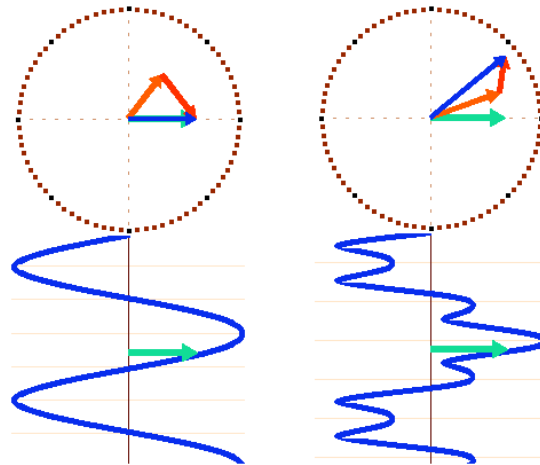


En réalité, le spectre d'un signal réel est toujours symétrique autour de la fréquence 0 Hz : l'enveloppe spectrale du côté des fréquences positives se retrouve en miroir du côté des fréquences négatives.

Ainsi, si la fréquence d'échantillonnage est de 44100 Hz, le spectre s'étend donc de -22050 Hz à +22050 Hz. Rappelons que si ces fréquences sont dépassées, il se produit un repliement de spectre.



Sur la figure suivante, à gauche, un signal sinusoïdal est généré par la rotation de deux vecteurs (dans le plan complexe) correspondant aux deux fréquences composant le spectre (positive et négative). Les deux vecteurs complexes (en rouge), tournant à la même vitesse mais dans des sens opposés, s'additionnent pour donner un vecteur réel (en bleu) qui est donc toujours horizontal. Sur la figure de droite, on voit comment une composante harmonique peut se combiner à la composante fondamentale. Dans le cas de cet exemple, le vecteur qui démarre à l'extrémité du vecteur principal tourne deux fois plus vite que ce dernier.



Voir démo sur le site :

<http://ptolemy.eecs.berkeley.edu/eecs20/berkeley/phasors/demo/phasors.html>

1.3.4 Les transformées de Fourier

Suivant la nature du signal temporel (analogique ou numérique) et le type d'analyse, on distingue :

- la **transformée de Fourier continue** (Continuous-Time Fourier Transform – CTFT) définie par une intégrale :

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt$$

- la **transformée de Fourier discrète** (Discrete-Time Fourier Transform – DFT) définie par une somme :

$$X(\omega_k) \sum_{n=0}^{N-1} x(t_n)e^{-i\omega_k t_n}, \quad k = 0, 1, 2, \dots, N-1$$

- la **transformée de Fourier rapide** (Fast Fourier Transform – FFT)
- la **transformée de Fourier à court terme** (Short-term Fourier Transform – STFT)

C'est généralement l'algorithme de la FFT qui est implémentée dans les outils informatiques d'analyse, qui est une version rapide de la DFT (Transformée de Fourier Discrète), optimisée pour des tailles de fenêtre d'analyse qui sont des puissances de 2 (par exemple : 512, 1024, 2048, ...). La transformée de Fourier à court terme (STFT) est utilisée pour générer des spectrogrammes, à partir de l'analyse de très courts extraits du signal formant des fenêtres qui se chevauchent.

Quand on effectue une analyse spectrale à l'aide de la transformée de Fourier discrète (DFT ou FFT), il faut spécifier les données suivantes :

- La taille de la fenêtre d'analyse en nombre d'échantillons (par exemple 1024). Ce facteur détermine la précision fréquentielle de l'analyse.
- Le type de fenêtre (rectangulaire, Hanning, Blackman, ...). L'utilisation de fenêtres autres que la fenêtre rectangulaire peut aider à éliminer les artefacts dans le signal analysé. C'est souvent la fenêtre de Hanning qui est choisie par défaut.
- Le facteur de zeropadding, opération qui consiste à ajouter des points à l'analyse sans augmenter la taille des fenêtres.
- Le taux de recouvrement des fenêtres (par exemple 50 %) dans le cas d'une transformée de Fourier à court terme

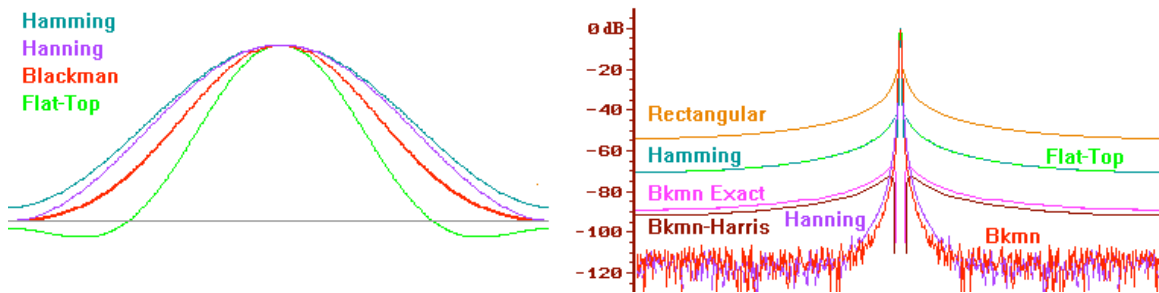
Si on effectue une analyse de Fourier sur une fenêtre de 1024 échantillons extraits du signal, on obtiendra une analyse en 1024 points ou bins ou paniers de fréquences. On comprend donc que plus la fenêtre est large, plus on a de points dans l'analyse spectrale et plus elle est précise au niveau fréquentiel. Par contre, l'information étant estimée sur toute la longueur de l'analyse, les événements très rapides, comme les attaques, seront adoucis par la moyenne effectuée sur plusieurs échantillons. Le compromis est inévitable :

Petites fenêtres d'analyse → plus de précision temporelle
Grandes fenêtres d'analyse → plus de précision fréquentielle

Par exemple, si on utilise une fenêtre de 1024 échantillons, on obtient un spectre de 1024 points. Mais attention : ces 1024 points sont répartis de 0 à 44100 Hz (la valeur de la fréquence d'échantillonnage). Donc, en réalité, on n'a que 512 points de 0 à 22050 Hz (fréquences positives). Les points sont ainsi espacés de

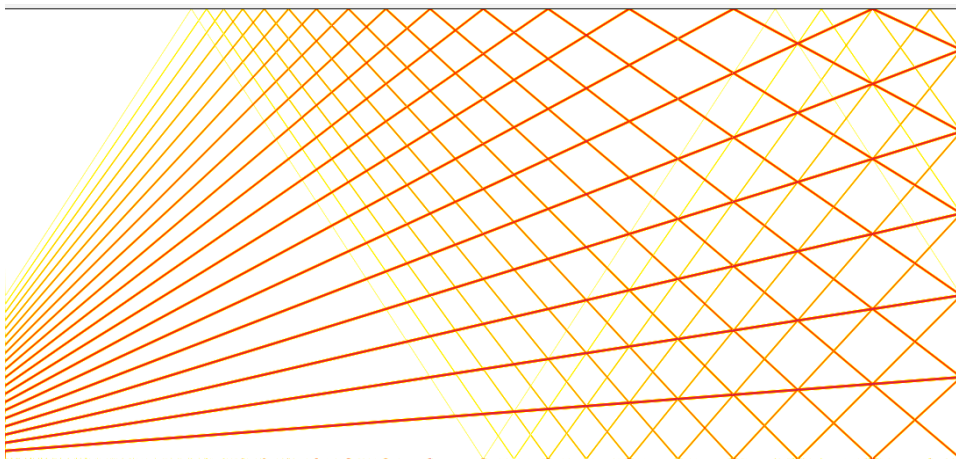
$$44100/1024 = 43 \text{ Hz}$$

La figure suivante donne, à gauche, l'allure de différentes fenêtres d'analyse, et à droite, leurs spectres d'amplitude (sur une échelle en dB).



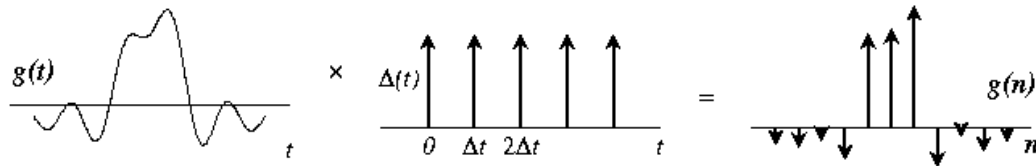
1.4 La théorie de l'échantillonnage

Rappelons d'abord que si la fréquence maximale contenue dans un signal sonore dépasse la moitié de la fréquence d'échantillonnage (la fréquence de Nyquist), il se produit un repliement de spectre. Ceci est facilement illustré par un son synthétique harmonique dont on augmente la fréquence fondamentale progressivement et linéairement, jusqu'à ce que la fréquence de Nyquist soit dépassée. Le repliement de spectre ou *aliasing* se manifeste par une « réflexion » des composantes sur la fréquence de Nyquist et peut être très bien visualisé sur un sonagramme (cf. figure ci-dessous).



Dans cette section, nous allons revenir sur les conséquences de l'échantillonnage d'un signal analogique, en examinant de plus près ce qui se passe dans le domaine fréquentiel.

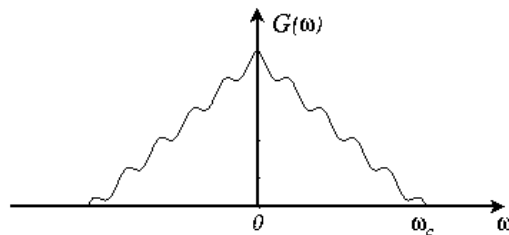
Considérons un signal analogique $g(t)$. Échantillonner ce signal revient à le multiplier par un train d'impulsions, comme illustré sur la figure ci-dessous.



Le temps qui s'écoule entre deux impulsions successives est la période d'échantillonnage $\Delta t = T_s$. Supposons que le signal analogique possède un spectre limité en fréquence (de fréquence maximale f_c) et supposons que la condition de Nyquist est satisfaite, c'est-à-dire que la fréquence d'échantillonnage est supérieure à $2 \times$ la fréquence maximale présente dans le spectre du signal :

$$f_s > 2f_c$$

La transformée de Fourier de la fonction-temps $g(t)$ nous fournit la fonction-fréquence $G(\omega)$ correspondante (son spectre), qui pourrait ressembler à ceci :



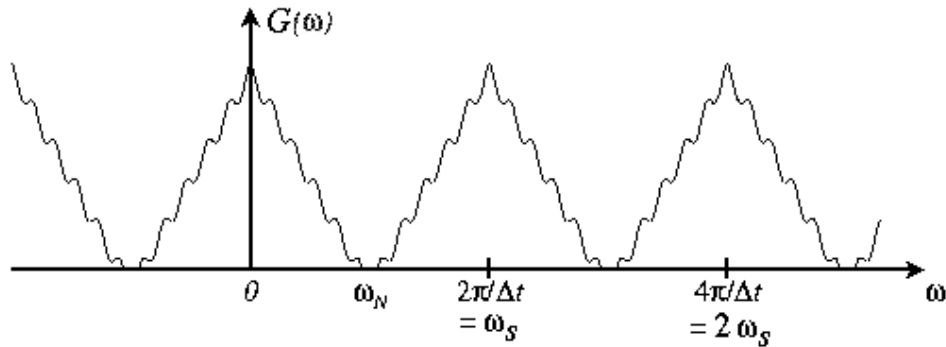
Pour connaître l'effet de l'échantillonnage dans le domaine fréquentiel, il faut savoir que :

- une multiplication (\times) dans le domaine temporel correspond à une convolution ($*$) dans le domaine fréquentiel
- la convolution entre deux fonctions $g(t)$ et $h(t)$ est la somme des produits d'une fonction par l'autre après inversion et délai dans le temps
- la transformée de Fourier d'un train d'impulsions (de période T_s) est aussi un train d'impulsions (de période $1/T_s = f_s = \omega_s/2\pi$).

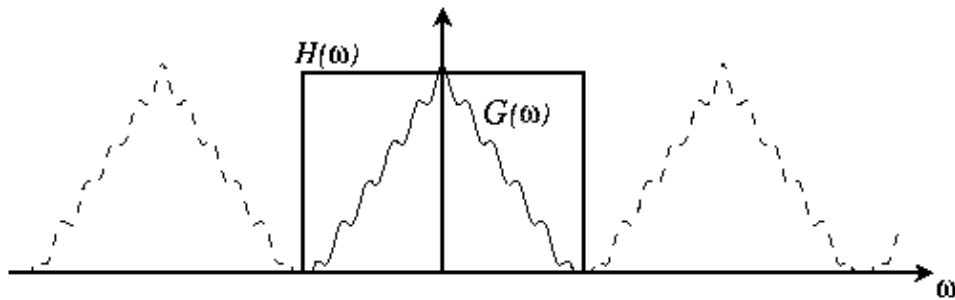
On comprend alors que l'échantillonnage provoque une convolution entre le spectre du signal et un train d'impulsions de période f_s :

$$G(\omega) \text{ échantillonné} = G(\omega) * \text{ train d'impulsions}$$

Il en résulte l'apparition de plusieurs "copies" du spectre centrées sur tous les multiples entiers de la fréquence d'échantillonnage :

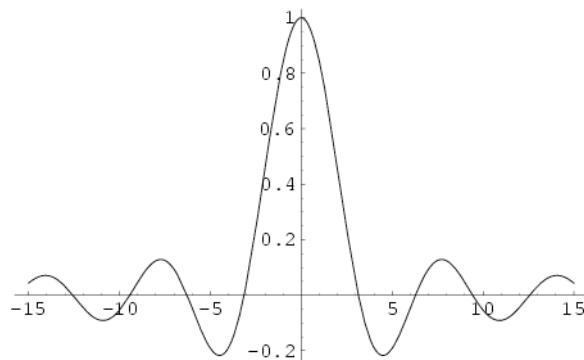


Maintenant, si l'on veut reconstruire le signal analogique à partir du signal échantillonné, on voit qu'il suffit de se débarrasser des copies latérales du spectre. Pour ce faire, on peut théoriquement utiliser un filtre passe-bas idéal (une fonction rectangulaire) qui vient isoler le spectre original de ses copies latérales.



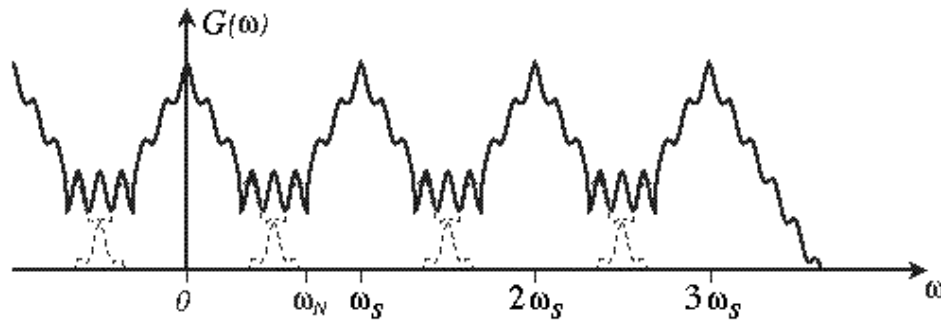
Pour connaître l'effet de la multiplication par cette fonction rectangulaire dans le domaine temporel, il faut d'abord savoir que :

- une multiplication (\times) dans le domaine fréquentiel correspond à une convolution ($*$) dans le domaine temporel,
- la transformée de Fourier d'une fonction rectangulaire est la fonction **sinc** illustrée ci-dessous (qu'on appelle parfois la "fonction pieuvre") :



Multiplier par une fonction rectangle dans le domaine fréquentiel revient donc à convoluer par la fonction sinc dans le domaine temporel. Une copie de la fonction sinc se retrouve au sommet de chaque échantillon, et la somme de toutes ces fonctions sinc pondérées par l'amplitude des échantillons donne le signal analogique original. Cette opération a donc pour effet d'interpoler entre les échantillons, de manière à retrouver le signal analogique de départ. Si la condition de Nyquist a été respectée, le signal de départ est parfaitement reconstitué³.

Maintenant, envisageons le cas d'un spectre qui ne satisfait pas la condition de Nyquist. Dans ce cas, les copies latérales qui apparaissent dans le spectre lors de l'échantillonnage se superposent et s'additionnent. Il n'est plus possible d'isoler le spectre du signal original à l'aide d'un filtre passe-bas. De nouvelles composantes sont introduites à cause de l'invasion des copies du spectre adjacentes dans l'intervalle de fréquences du spectre de départ. On dira qu'il y a repliement du spectre. La figure suivante illustre le phénomène. Le spectre d'origine et ses copies latérales sont tracés en pointillés. Le spectre résultant de leur somme est tracé en gras.



Pour éviter ce problème, il faut veiller à ce que le signal à échantillonner soit limité en fréquence. C'est pour cette raison qu'un convertisseur analogique-numérique comprend un filtre passe-bas anti-repliement à son entrée.

En guise de conclusion, voici une figure illustrant la chaîne complète de la conversion analogique/numérique/analogique.

3. Notons que cette reconstruction mathématique du signal ne peut être implémentée dans des convertisseurs DAC, car elle est trop coûteuse en temps de calcul. Au lieu de ça, un signal "en escalier" est généré à partir des échantillons, qui ensuite subit un lissage par filtrage

