

2-Comment fonctionne l'audio numérique

Une explication détaillée du fonctionnement de l'audio numérique va bien au-delà de la portée de ce manuel. Ce qui suit est une explication très brève qui vous donnera le minimum de compréhension nécessaire pour utiliser MSP avec succès.

Pour une explication plus complète du fonctionnement de l'audio numérique, nous vous recommandons le *didacticiel d'informatique musicale* de Curtis Roads, publié en 1996 par MIT Press. Il comprend également une bibliographie complète sur le sujet.

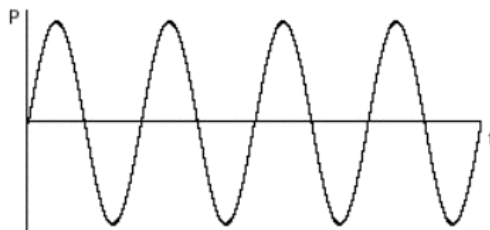
Son

Mouvement harmonique simple

Les sons que nous entendons sont des fluctuations de la pression atmosphérique - de minuscules variations par rapport à la pression atmosphérique normale - causées par des objets en vibration. (techniquement, il pourrait s'agir de la pression de l'eau si vous écoutez sous l'eau, mais veuillez à garder votre ordinateur hors de la piscine.)

Lorsqu'un objet se déplace, il déplace des molécules d'air à côté de lui, ce qui déplace à leur tour des molécules d'air, et ainsi de suite, ce qui crée un «front de haute pression» momentané qui s'éloigne de l'objet en mouvement (vers vos oreilles). Ainsi, si nous faisons vibrer un objet - nous frappons un diapason, par exemple - et mesurons ensuite la pression de l'air en un point proche avec un microphone, ce dernier détectera une légère augmentation de la pression de l'air au fur et à mesure que "front de haute pression" se déplace. Étant donné que la fourche du diapason est assez rigide et est fixée à l'une de ses extrémités, il existe une force de rappel qui la ramène à sa position normale. Comme cette force de rappel lui donne de l'élan, elle dépasse sa position normale, se déplace ensuite vers la position extrême opposée, et continue de vibrer d'avant en arrière de cette manière jusqu'à ce qu'elle perde finalement son élan et s'arrête dans sa position normale. Par conséquent, notre microphone détecte une augmentation de la pression, suivie d'une chute de pression, suivie d'une augmentation de pression, et ainsi de suite., correspondant aux vibrations de va-et-vient de la fourche du diapason.

Si nous devons tracer un graphique de la variation de la pression atmosphérique détectée par le microphone au fil du temps, nous verrions une forme sinusoïdale (une onde sinusoïdale) montant et descendant correspondant aux vibrations de va-et-vient du diapason.



Changement sinusoïdal de la pression atmosphérique provoqué par une simple vibration de va-et-vient

Cette augmentation et cette diminution continues de la pression créent une onde sonore. L'ampleur de la variation de la pression de l'air, par rapport à la pression atmosphérique normale est appelée **amplitude** de l'onde (littéralement sa 'taille'). Nous utilisons le plus souvent le terme «amplitude» pour désigner **l'amplitude maximale**, c'est-à-dire la plus grande variation de pression obtenue par l'onde.

Ce type de mouvement simple de va-et-vient (que l'on observe également dans le balancement d'un pendule) est appelé **mouvement harmonique simple**. Il est considéré comme la forme la plus simple de vibration car l'objet effectue un cycle complet d'aller-retour à une vitesse constante. Même si sa vitesse change lorsqu'il ralentit pour changer de direction et reprend de la vitesse dans l'autre sens - comme le montre la courbe de l'onde sinusoïdale - sa vitesse moyenne d'un cycle à l'autre est la même. Chaque cycle vibratoire complet se produit donc dans un intervalle de temps égal (dans une période donnée), de sorte que l'onde est dite **périodique**. Le nombre de cycles qui se produisent en une seconde est appelé la fréquence de la vibration. Par exemple, si la fourche du diapason va et vient 440 fois par seconde, sa fréquence est de 440 cycles par seconde et sa période est de **1/440** seconde par cycle.

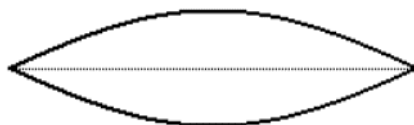
Pour que nous puissions entendre de telles fluctuations de pression:

- Les fluctuations doivent être suffisamment importantes pour affecter notre membrane tympanique (tympan), mais pas au point de nous blesser. En pratique, l'intensité des changements de pression atmosphérique doit être supérieure à environ 10^{-9} fois la pression atmosphérique, mais pas supérieure à environ 10^{-3} fois la pression atmosphérique. Vous n'aurez jamais besoin de cette information, mais elle est là. Cela signifie que le son le plus faible que nous puissions entendre a environ un millionième de l'intensité du son le plus fort que nous puissions supporter. C'est un assez large éventail de possibilités.
- Les fluctuations doivent se répéter à un rythme régulier suffisamment rapide pour que nous les percevions comme un son (plutôt que comme des événements individuels), mais pas trop vite pour que cela dépasse notre capacité à l'entendre. Les manuels présentent généralement cette gamme de fréquences audibles allant de 20 à 20 000 cycles par seconde (cps, également appelé **hertz**, en abrégé **Hz**). Votre propre étendue peut varier. Si vous approchez de l'âge mûr ou si vous avez trop écouté de musique forte, vous pouvez plafonner à environ 17 000 Hz, voire moins.

Sons complexes

Un objet qui vibre dans un mouvement harmonique simple est dit avoir un mode de vibration résonant – c'est-à-dire une fréquence à laquelle il aura naturellement tendance à vibrer lorsqu'il sera mis en mouvement. Cependant, la plupart des objets du monde réel ont **plusieurs** modes de vibration résonants et vibrent donc à plusieurs fréquences à la fois. Tout son qui contient plus d'une seule fréquence (c'est-à-dire tout son qui n'est pas une simple onde sinusoïdale) est appelé un son **complexe**. Prenons par exemple une corde de guitare tendue.

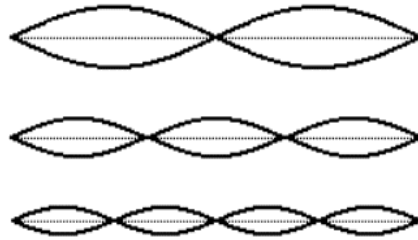
Une corde de guitare a une masse uniforme sur toute sa longueur, une longueur connue puisqu'elle est fixée aux deux extrémités (au «sillet» et au «chevalet») et une tension donnée qui dépend de la précision avec laquelle elle est accordée. la cheville de réglage. Comme la corde est fixée à ses deux extrémités, elle doit toujours être immobile en ces points, et c'est donc naturellement en son centre qu'elle vibre le plus.



Une corde pincée vibrant dans son mode de résonance fondamentale

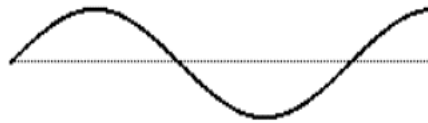
La fréquence à laquelle elle vibre dépend de sa masse, de sa tension et de sa longueur. Ces caractéristiques restent assez constantes dans le cours d'une même note, de sorte qu'elle a une fréquence fondamentale à laquelle elle vibre.

Cependant, d'autres modes de vibration sont encore possibles.



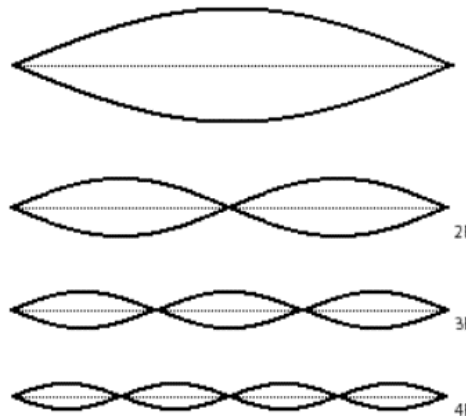
Quelques autres modes de résonance d'une corde tendue

Les modes de vibration possibles sont limités par le fait que la corde doit rester immobile à chaque extrémité. Cela limite ses modes de résonance aux divisions entières de sa longueur.



Ce mode de résonance serait impossible car la corde est fixée à chaque extrémité

Comme la tension et la masse sont définies, les divisions entières de la longueur de la corde résultent en multiples entiers de la fréquence fondamentale.



Chaque mode de résonance entraîne une fréquence différente

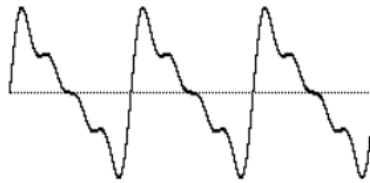
En fait, une corde pincée vibre simultanément dans tous ces modes de résonance possibles, créant de l'énergie à toutes les fréquences correspondantes. Bien entendu, chaque mode de vibration (et donc chaque fréquence) aura une amplitude différente. (Dans l'exemple de la corde de guitare, les segments les plus longs ont plus de liberté pour vibrer.) Le son résultant sera la somme de toutes ces fréquences, chacune avec sa propre amplitude.

À mesure que les vibrations de la corde s'atténuent en raison de la force d'amortissement de la fixation à chaque extrémité, chaque fréquence peut s'atténuer à un rythme différent. En fait, dans de nombreux sons, les amplitudes des différentes fréquences composant peuvent varier séparément et différemment les unes des autres. Cette variété semble être l'un des facteurs fondamentaux de notre perception des sons comme ayant une *couleur de son* différente (c'est-à-dire un *timbre*), et le timbre d'une simple note peut changer radicalement au cours de la durée de la note.

Sons harmoniques

La combinaison de fréquences - et de leurs amplitudes - présentes dans un son s'appelle son *spectre* (tout comme les différentes fréquences et intensités de la lumière constituent un spectre de couleur). Chaque fréquence individuelle entrant dans la composition d'un son complexe est appelée un *partiel*. (C'est une partie du son entier.)

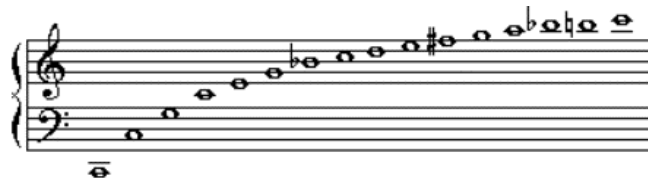
Lorsque les partiels (fréquences composantes) d'un son complexe sont tous des multiples entiers de la même fréquence fondamentale, comme dans notre exemple de corde de guitare, on dit que le son a un *spectre harmonique*. Chaque composante d'un spectre harmonique est appelée un *partiel harmonique* ou simplement un *harmonique*. La somme de toutes ces fréquences harmoniques donne toujours une onde périodique ayant une fréquence fondamentale. Les fréquences multiples entières fusionnent donc «harmonieusement» en un seul son.



La somme des fréquences liées harmoniquement se répète toujours à la fréquence fondamentale

Cette fusion est étayée par le célèbre théorème mathématique de Jean-Baptiste Joseph Fourier, selon lequel toute onde périodique, aussi complexe soit elle, peut être démontrée comme étant la somme de différentes fréquences harmoniquement liées (ondes sinusoïdales), chacune ayant sa propre amplitude et sa propre phase. (La *phase* est un décalage dans le temps d'une fraction de cycle.)

Les fréquences harmoniquement liées définissent un ensemble particulier de hauteurs liées dans notre perception musicale.



Partiels harmoniques d'une fréquence fondamentale f , où $f = 65,4 \text{ Hz}$ = la hauteur du C grave

Chaque fois que la fréquence fondamentale est multipliée par une puissance de 2 - 2, 4, 8, 16, etc. - la hauteur musicale perçue augmente d'une octave. Toutes les cultures semblent partager la perception selon laquelle il existe une certaine «similitude» de classe de hauteur entre ces fréquences liées à l'octave. Les autres multiples entiers de la fondamentale produisent de nouvelles hauteurs musicales. Chaque fois que vous entendez un son complexe harmonique, vous entendez réellement un accord! Comme nous l'avons vu, le résultat combiné se répète à la fréquence fondamentale, et nous avons donc tendance à fusionner ces fréquences de manière à percevoir une seule hauteur.

Sonorités inharmoniques et bruit

Certains objets, comme une cloche par exemple, vibrent de manière encore plus complexe, avec de nombreux modes de vibration différents qui ne pas produire un ensemble de partiels en relation harmonique. Si les fréquences présentes dans un son ne sont pas des multiples entiers d'une seule fréquence fondamentale, l'onde ne se répète pas périodiquement. Par conséquent, un ensemble *inharmonique* de partiels ne se fond pas si facilement dans notre perception. Nous pouvons être en

mesure de distinguer plus facilement les partiels individuels et, surtout lorsque ceux-ci sont nombreux et complètement inharmoniques, nous risquons de ne pas percevoir le son comme ayant une seule hauteur fondamentale discernable.

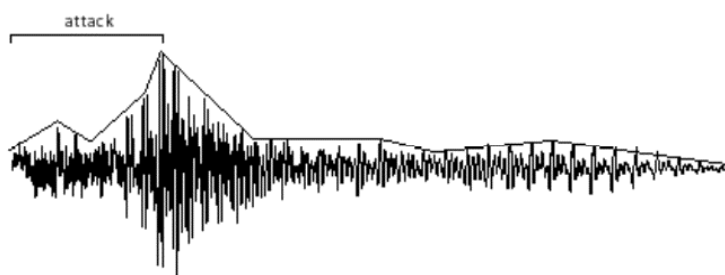
Lorsqu'un son est si complexe qu'il contient de très nombreuses fréquences différentes sans relation mathématique apparente, nous le percevons comme un bruit. Un son dont les fréquences et les amplitudes sont complètement aléatoires - toutes les fréquences sont présentes en proportion égale - est le son statique connu sous le nom de **bruit blanc** (analogue à la lumière blanche qui contient toutes les fréquences de la lumière).

Il peut donc être utile de considérer que les sons se situent sur un continuum allant de la pureté et de la prévisibilité totales (une onde sinusoïdale) au caractère totalement aléatoire (le bruit blanc). La plupart des sons se situent entre ces deux extrêmes. Un son harmonique - une note de trompette ou une note de guitare, par exemple - se situe à l'extrémité la plus pure du continuum, tandis qu'un bruit de cymbale est plus proche de l'extrémité la plus bruyante du continuum. Les timbales et les cloches peuvent être suffisamment évocatrices d'un spectre harmonique pour que nous puissions identifier une hauteur fondamentale, mais elles contiennent d'autres partiels inharmoniques. D'autres tambours produisent davantage un bruit à bande limitée - des fréquences liées de façon aléatoire, mais restreintes dans une certaine gamme de fréquences - donnant une impression de gamme de hauteur, ou de hauteur non spécifique, plutôt qu'une fondamentale identifiable. Il est important de garder à l'esprit ce continuum lorsque l'on synthétise des sons.

Enveloppe d'amplitude

Un autre facteur important dans la variété presque infinie de sons est la modification de l'amplitude globale d'un son au cours de sa durée. La forme de cette variation globale macroscopique de l'amplitude est appelée **l'enveloppe d'amplitude**. La partie initiale du son, à mesure que l'enveloppe d'amplitude passe du silence à l'audibilité, pour atteindre son amplitude maximale, est connue est appelé **l'attaque** du son. L'enveloppe, et surtout l'attaque d'un son, sont des facteurs importants dans notre capacité à distinguer, reconnaître et comparer les sons. Nous avons très peu comment lire la représentation graphique d'une onde sonore et entendre le son dans notre tête comme le ferait un bon lecteur à vue avec la notation musicale.

Cependant, l'enveloppe d'amplitude peut au moins nous renseigner sur l'évolution générale de l'intensité du son dans le temps.



L'enveloppe d'amplitude est l'évolution de l'amplitude d'un son dans le temps

Amplitude et intensité

La relation entre l'amplitude d'un son mesurée objectivement et notre impression subjective de son intensité est très complexe et dépend de nombreux facteurs. Sans essayer d'expliquer tous ces facteurs, nous pouvons au moins souligner que notre perception de l'intensité relative de deux sons est liée au rapport de leurs intensités, plutôt qu'à la différence mathématique entre leurs intensités. Par exemple, sur une échelle de mesure arbitraire, la relation entre un son d'amplitude 1 et un son

d'amplitude 0,5 est la même pour nous que le rapport entre un son d'amplitude 0,25 et un son d'amplitude 0,125. La différence soustractive entre les amplitudes est de 0,5 dans le premier cas et de 0,125 dans le second, mais ce qui nous intéresse sur le plan perceptif c'est le rapport, qui est de 2:1 dans les deux cas.

Est-ce qu'un son dont l'amplitude est deux fois plus grande paraît-il deux fois plus fort ? En général, la réponse est non. Tout d'abord, notre perception subjective de «volume» n'est pas directement proportionnelle à l'amplitude. Des expériences ont montré que pour la plupart des auditeurs, la sensation (extrêmement subjective) d'un son est «deux fois plus fort» nécessite une augmentation de l'amplitude bien supérieure au double. De plus, notre perception de l'intensité sonore varie considérablement en fonction de la fréquence des sons considérés. Nous sommes beaucoup plus sensibles aux fréquences comprises entre 300 Hz et 7 000 Hz environ qu'aux fréquences situées en dehors de cette plage. (Cela peut éventuellement être dû à l'évolution de l'importance de l'audition de la parole et de beaucoup d'autres sons importants qui se situent principalement dans cette gamme de fréquences.)

Néanmoins, il existe une corrélation - même si elle n'est pas parfaitement linéaire - entre l'amplitude et l'intensité sonore, de sorte qu'il est donc certainement instructif de connaître l'amplitude relative de deux sons. Comme mentionné précédemment, le son le plus faible que nous puissions entendre a une amplitude d'environ un millionième de celle du son le plus fort que nous pouvons supporter. Plutôt que de discuter de l'amplitude en utilisant une gamme aussi large de nombres, allant de 0 à 1 000 000, il est plus courant de comparer les amplitudes sur une échelle logarithmique.

Le rapport entre deux amplitudes est généralement discuté en termes de *décibels* (abrégé en dB). Un *niveau* exprimé en décibels est l'énoncé d'une relation de rapport entre deux valeurs - et non une mesure absolue. Si nous considérons une amplitude comme une référence que nous appelons A0, alors l'amplitude relative d'un autre son en décibels peut être calculée avec l'équation suivante:

$$\text{niveau en décibels} = 20 \log_{10} (A / A_0)$$

Si nous considérons l'amplitude maximale possible comme une référence avec une valeur numérique de 1, alors un son d'amplitude 0,5 a la moitié de l'amplitude (égale à $10^{-0,3}$) et son niveau est donc de

$$20 \log_{10} (0,5 / 1) = 20 (-0,3) = -6 \text{ dB}$$

Chaque réduction de moitié de l'amplitude correspond à une différence d'environ -6 dB; chaque doublement de l'amplitude correspond à une augmentation d'environ 6 dB. Ainsi, si une amplitude est supérieure de 48 dB à une autre, on peut estimer qu'elle est environ 2^8 (256) fois supérieure.

Résumé

Une compréhension théorique des ondes sinusoïdales, des sons harmoniques, des sons complexes inharmoniques et du bruit, telle que discuté ici, est utile pour comprendre la nature du son. Cependant, la plupart des sons sont en fait des combinaisons complexes de ces descriptions théoriques, qui changent d'un instant à l'autre. Par exemple, une corde jouée à l'archet peut inclure du bruit de raclement de l'archet contre la corde, des variations d'amplitude dues aux variations de pression et de vitesse de l'archet, des changements dans la prééminence des différentes fréquences dues à la position de l'archet, des changements d'amplitude et de fréquence fondamentale (et de toutes ses harmoniques) dus aux mouvements de vibrato dans la main gauche, etc. Un son de batterie peut être bruyant, mais peut évoluer de manière à avoir des accentuations dans certaines régions de son spectre qui impliquent un son harmonique, donnant ainsi une impression de hauteur

fondamentale. L'examen des sons existants et l'expérimentation de la synthèse de nouveaux sons peuvent donner un aperçu de la manière dont les sons sont composés. L'ordinateur fournit cette opportunité.

Représentation numérique du son

Échantillonnage et quantification d'une onde sonore

Pour comprendre comment un ordinateur représente le son, considérons comment un film représente le mouvement. Un film est réalisé en prenant des images fixes en séquence rapide à une vitesse constante, généralement vingt-quatre images par seconde. Lorsque les images sont affichées en séquence à la même vitesse, nous pensons voir un mouvement continu, alors qu'il s'agit en fait de vingt-quatre images *distinctes* par seconde. L'enregistrement numérique du son fonctionne sur le même principe. Nous prenons de nombreux échantillons discrets de l'amplitude instantanée de l'onde sonore, nous stockons cette information, puis nous reproduisons ultérieurement ces amplitudes au même rythme pour créer l'illusion d'une onde continue.

Le rôle d'un microphone est de convertir (transformer une forme d'énergie en une autre) la variation de pression atmosphérique en une variation analogue de la tension électrique. Cette tension changeant continuellement peut ensuite être échantillonnée périodiquement selon un processus appelé (sample and hold) *échantillonnage et maintien*. À des moments régulièrement espacés dans le temps, la tension à cet instant est échantillonnée et maintenue constante jusqu'à la prise de l'échantillon suivant. Cela réduit la quantité totale d'informations à un certain nombre de tensions discrètes.



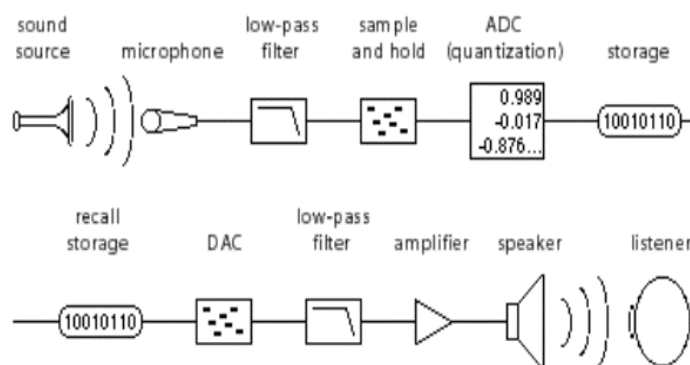
Tension variable dans le temps échantillonnée périodiquement

Un dispositif appelé *convertisseur analogique-numérique* (ADC) reçoit les tensions discrètes du dispositif d'échantillonnage et de maintien, et attribue une valeur numérique à chaque amplitude. Ce processus de conversion de tensions en nombres est appelé *quantification*. Ces nombres sont exprimés dans l'ordinateur sous la forme d'une chaîne de chiffres binaires (1 ou 0). Les nombres binaires obtenus sont stockés en mémoire, généralement sur une bande audio numérique, un disque dur ou un disque laser. Pour lire le son, nous lisons les chiffres de la mémoire et les transmettons à un convertisseur numérique-analogique (DAC) à la même vitesse que celle à laquelle ils ont été enregistrés. Le DAC convertit chaque nombre en tension et communique ces tensions à un amplificateur pour augmenter l'amplitude de la tension.

Pour qu'un ordinateur puisse représenter le son avec précision, il faut prélever de nombreux échantillons par seconde, bien plus que pour filmer une image visuelle. En fait, nous devons prélever deux fois plus d'échantillons que la fréquence la plus élevée que nous souhaitons enregistrer. (Pour une explication des raisons de ce phénomène, reportez-vous à la section suivante, *Limitations de l'audio numérique*.) Si nous voulons enregistrer des fréquences aussi élevées que 20 000 Hz, nous devons échantillonner le son au moins 40 000 fois par seconde. La norme pour les enregistrements sur disque compact (et pour l'audio informatique de qualité «CD») consiste à prélever 44 100 échantillons par seconde pour chaque canal audio. Le nombre d'échantillons prélevés par seconde est connu sous le nom de *taux d'échantillonnage*.

Cela signifie que l'ordinateur ne peut représenter avec précision les fréquences que jusqu'à la moitié de la fréquence d'échantillonnage. Toutes les fréquences du son qui dépassent la moitié de la fréquence d'échantillonnage doivent être filtrées avant que le processus d'échantillonnage n'ait lieu. Pour se faire, le signal électrique est envoyé à travers un *filtre passe-bas* qui supprime toutes les fréquences supérieures à un certain seuil. De plus, lorsque le signal numérique (le flux de chiffres binaires représentant les échantillons quantifiés) est envoyé au DAC pour être reconverti en un signal électrique continu, le son sortant du DAC contient des hautes fréquences parasites créées par le processus de sample and hold lui-même. (Ces fréquences sont dues aux "bords saillants" créés par les échantillons discrets, comme on le voit dans l'exemple ci-dessus). Par conséquent, nous devons également envoyer le signal de sortie à travers un filtre passe-bas.

Le processus d'enregistrement et de lecture numérique est donc une chaîne d'opérations, comme illustré dans le diagramme suivant.



Enregistrement numérique et processus de lecture

Limites de l'audio numérique

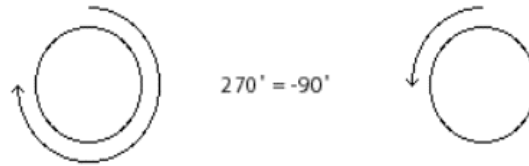
Fréquence d'échantillonnage et taux de Nyquist

Nous avons constaté qu'il est nécessaire de prélever au moins deux fois plus d'échantillons que la fréquence la plus élevée que nous souhaitons enregistrer. Cela a été prouvé par Harold Nyquist et est connu sous le nom de *théorème de Nyquist*. En d'autres termes, l'ordinateur ne peut représenter avec précision les fréquences que jusqu'à la moitié de la fréquence d'échantillonnage. La moitié de la fréquence d'échantillonnage est souvent appelée *fréquence de Nyquist* ou *taux de Nyquist*.

Si nous prenons, par exemple, 16 000 échantillons d'un signal audio par seconde, nous ne pouvons capturer que des fréquences allant jusqu'à 8 000 Hz. Les fréquences supérieures au taux de Nyquist sont «repliées» de manière perceptible dans la plage située au-dessous de la fréquence de Nyquist. Ainsi, si le son que nous essayions d'échantillonner contient de l'énergie à 9 000 Hz, le processus d'échantillonnage déformera cette fréquence en la faisant passer à 7 000 Hz - une fréquence qui n'aurait peut-être pas été présente dans le son original. Cet effet est connu sous le nom de *repliement* ou *aliasing*. Le principal problème de l'aliasing est qu'il peut ajouter au son numérisé des fréquences qui ne figuraient pas dans le son d'origine. À moins de connaître le spectre exact du son d'origine, il n'y a aucun moyen de savoir quelles fréquences appartiennent réellement au son numérisé et lesquelles sont le résultat du repliement. C'est pourquoi il est essentiel d'utiliser le filtre passe-bas avant le processus de Sample and Hold afin de supprimer toutes les fréquences supérieures à la fréquence de Nyquist.

Pour comprendre pourquoi ce phénomène de repliement de spectre se produit, reprenons l'exemple d'une caméra de cinéma, qui filme 24 images par seconde. Si nous tournons le film d'une voiture et que la roue de la voiture tourne à une vitesse supérieure à 12 tours par seconde, elle dépasse la

moitié du «taux d'échantillonnage» de la caméra. La roue effectue plus d'un demi-tour par image. Si, par exemple, elle effectue en réalité 18/24 tours par image, elle semblera reculer à une vitesse de 6 tours par seconde. En d'autres termes, si nous ne sommes pas témoins de ce qui se passe entre les échantillons, une rotation de 270° de la roue est indiscernable depuis une révolution de -90° . Les échantillons que nous obtenons dans les deux cas sont exactement les mêmes.



Pour la caméra, une révolution de 18/24 n'est pas différente d'une révolution de -6/24

Pour l'échantillonnage audio, le phénomène est pratiquement identique. Toute fréquence qui dépasse le taux de Nyquist est impossible à distinguer d'une fréquence négative inférieure d'autant au taux de Nyquist. (Et nous ne faisons pas de distinction perceptive entre les fréquences positives et négatives.) Dans la mesure où une fréquence dépasse le taux de Nyquist, elle est repliée de la même fréquence par rapport à la fréquence de Nyquist.

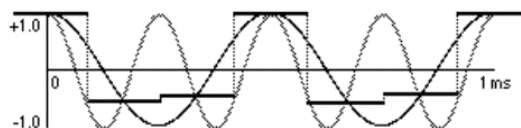
Pour le démontrer, considérez les deux exemples suivants. L'exemple suivant montre un graphique d'une onde cosinusoidale de 4 000 Hz (énergie uniquement à 4 000 Hz) échantillonnée à une fréquence de 22 050 Hz. 22 050 Hz correspond à la moitié de la fréquence d'échantillonnage d'un CD, et constitue une fréquence d'échantillonnage acceptable pour les sons qui n'ont pas beaucoup d'énergie dans l'octave supérieure de notre gamme d'audition.

Dans ce cas, le taux d'échantillonnage est tout à fait adéquate car la fréquence maximale que nous essayons d'enregistrer est bien inférieure à la fréquence de Nyquist.



Une onde cosinusoidale de 4 000 Hz échantillonnée à 22 050 Hz

Considérons maintenant la même onde cosinusoidale de 4 000 Hz échantillonnée à une fréquence inadéquate, par exemple 6 000 Hz. L'onde accomplit plus d'un demi-cycle par échantillon, et les échantillons résultants ne peuvent pas être distingués de ceux qui seraient obtenus à partir d'une onde cosinusoidale de 2000 Hz.



Une onde cosinusoidale de 4 000 Hz sous-échantillonnée à 6 000 Hz

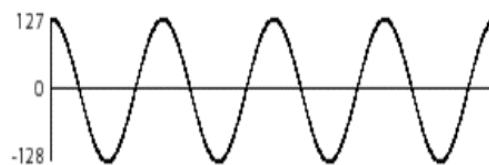
La leçon simple à tirer du théorème de Nyquist est que l'audio numérique ne peut représenter avec précision une fréquence supérieure à la moitié de la fréquence d'échantillonnage. Toute fréquence de ce type sera mal représentée en étant repliée dans la plage inférieure à la moitié de la fréquence d'échantillonnage.

Précision de quantification

A chaque échantillon d'un signal audio doit être attribuée une valeur numérique à stocker dans l'ordinateur. La valeur numérique exprime l'amplitude instantanée du signal au moment où il a été échantillonné. La plage des nombres doit être suffisamment large pour exprimer de manière adéquate toute la plage d'amplitude du son échantillonné.

La gamme des nombres possibles utilisés par un ordinateur dépend du nombre de chiffres binaires (*bits*) utilisés pour stocker chaque nombre. Un bit peut avoir l'une des deux valeurs possibles: 1 ou 0. Deux bits ensemble peuvent avoir l'une des quatre valeurs possibles: 00, 01, 10 ou 11. Plus le nombre de bits augmente, plus la gamme de nombres possibles qu'ils peuvent exprimer augmente par une puissance de deux. Ainsi, un seul octet (8 bits) de données informatiques peut exprimer l'un des $2^8 = 256$ nombres possibles. Si nous utilisons deux octets pour exprimer chaque nombre, nous obtenons une plage beaucoup plus large de valeurs possibles, car $2^{16} = 65\ 536$.

Le nombre de bits utilisés pour représenter le nombre dans l'ordinateur est important car il détermine la résolution avec laquelle nous pouvons mesurer l'amplitude du signal. Si nous n'utilisons qu'un octet pour représenter chaque échantillon, nous devons alors diviser l'ensemble des amplitudes possibles du signal en 256 parties, car nous ne disposons que de 256 façons de décrire l'amplitude.



En utilisant un octet par échantillon, chaque échantillon peut avoir l'une des 256 valeurs possibles différentes.

Par exemple, si l'amplitude du signal électrique échantillonné varie de -10 volts à +10 volts et que nous utilisons un octet pour chaque échantillon, chaque nombre ne représente pas une tension précise, mais plutôt une portion de 0,078125 V de la plage totale. Tout échantillon compris dans cette portion se verra attribuer le même nombre. Cela signifie que chaque description numérique de la valeur d'un échantillon peut s'écarter de sa valeur réelle jusqu'à 0,078125V - 1/256 de la plage d'amplitude totale. Dans la pratique, chaque échantillon sera différent d'une quantité aléatoire comprise entre 0 et 1/256 de la plage d'amplitude totale. L'erreur moyenne sera 1/512 de la plage totale.

C'est ce qu'on appelle l'**erreur de quantification**. Elle est inévitable, mais elle peut être réduite à un niveau acceptable en utilisant plus de bits pour représenter chaque nombre. Si nous utilisons deux octets par échantillon, l'erreur de quantification ne sera jamais supérieure à 1/65 536 de la plage d'amplitude totale et l'erreur moyenne sera de 1/131 072.

Comme l'erreur de quantification de chaque échantillon est généralement aléatoire (parfois un peu trop élevée, parfois un peu trop faible), nous entendons généralement l'effet de l'erreur de quantification comme un bruit blanc. Ce bruit n'est pas présent dans le signal d'origine. Il est ajouté dans le signal numérique par la nature imprécise de la quantification. C'est ce qu'on appelle le **bruit de quantification**.

Le rapport entre l'amplitude totale et l'erreur de quantification est appelé **rapport de signal sur bruit de quantification** (SQNR). Il s'agit du rapport entre l'amplitude maximale possible du signal et le niveau moyen de quantification du bruit de quantification. Il est généralement exprimé en décibels.

En règle générale, chaque bit de précision utilisé dans la quantification ajoute 6 dB au SQNR. Par conséquent, un son quantifié avec une précision numérique de 8 bits aura, dans le meilleur des cas un SQNR d'environ 48 dB. C'est suffisant dans les cas où la fidélité n'est pas importante, mais n'est certainement pas souhaitable pour la musique ou à d'autres utilisations critiques. Le son échantillonné avec une précision de 16 bits («qualité CD») a un SQNR de 96 dB, ce qui est plutôt bon - il est beaucoup plus silencieux que les salles d'écoute habituelles.

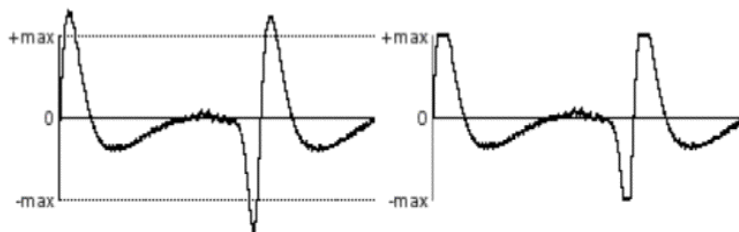
En bref, plus l'ordinateur utilise de bits pour stocker chaque échantillon, meilleur est le rapport potentiel signal / bruit.

Mémoire et stockage

Nous avons vu que le taux d'échantillonnage standard pour l'audio haute-fidélité est de 44 100 échantillons par seconde. Nous avons également vu que 16 bits (2 octets) sont nécessaires par échantillon pour obtenir un bon rapport signal / bruit. Avec ces informations, nous pouvons calculer la quantité de données nécessaires pour l'audio numérique: 44 100 échantillons par seconde, multipliés par 2 octets par échantillon, multipliés par 2 canaux pour la stéréo, multipliés par 60 secondes par minute, cela représente plus de 10 mégaoctets de données par minute de son de qualité CD. Il est clair que les besoins en mémoire et en stockage de l'audio numérique sont considérables. Heureusement, un disque compact contient plus d'une heure de son stéréo et un disque dur d'ordinateur de plus d'un gigaoctet est suffisant pour la plupart des enregistrements et traitements audio.

Écrêtage

Si l'amplitude du signal électrique entrant dépasse l'amplitude maximale pouvant être exprimée numériquement, le signal numérique sera une version écrêtée du son réel.



Un signal qui dépasse l'amplitude maximale sera écrêté lors de la quantification

L'échantillon écrêté aura souvent un son très différent de l'original. Parfois, ce type d'écrêtage n'entraîne qu'une légère distorsion du son, qui se traduit par un changement de timbre. Mais le plus souvent, il s'agit d'un bruit très désagréable ajouté au son. Pour cette raison, il est très important de prendre des précautions pour éviter l'écrêtage. L'amplitude du signal électrique ne doit pas dépasser le maximum attendu par l'ADC.

Il est également possible de produire des nombres dans l'ordinateur qui dépassent le maximum attendu par le DAC. Le son sortant du DAC sera alors une version écrêtée du signal numérique. L'écrêtage par le DAC est aussi grave que l'écrêtage par l'ADC. Il faut donc veiller à ne pas générer un signal numérique qui dépasse la plage numérique que le DAC est capable de gérer.

Avantages de l'audio numérique

Synthèse de l'audio numérique

Puisqu'une représentation numérique du son n'est qu'une liste de nombres, toute liste de nombres peut théoriquement être considérée comme une représentation numérique d'un son. Pour qu'une liste de nombres soit audible en tant que son, les valeurs numériques doivent fluctuer vers le haut et vers le bas à un taux audio. Nous pouvons écouter une telle liste en envoyant des nombres à un DAC où ils sont convertis en tensions. C'est la base de la synthèse sonore en informatique. Tous les nombres que nous pouvons générer avec un programme informatique peuvent être écoutés sous forme de son.

De nombreuses méthodes ont été découvertes pour générer des nombres qui produisent des sons intéressants. L'une de ces méthodes consiste à écrire un programme qui résout de manière répétée une équation mathématique contenant deux variables. A chaque répétition, la valeur en augmentation constante est entrée pour l'une des variables, représentant le passage du temps. La valeur de l'autre variable au moment de la résolution est utilisée comme amplitude pour chaque instant dans le temps. Le résultat du programme est une amplitude qui varie de haut en bas au fil du temps.

Par exemple, une onde sinusoïdale peut être générée en résolvant de manière répétée l'équation algébrique suivante, en utilisant une valeur croissante pour n :

$$y = A \sin (2\pi f n / R + \phi)$$

où A est l'amplitude de l'onde, f la fréquence de l'onde, n le numéro de l'échantillon (0,1, 2,3, etc.), R le taux d'échantillonnage et ϕ la phase. Si nous entrons des valeurs pour A , f et ϕ tout en augmentant la valeur de n , la valeur de y (l'échantillon de sortie) variera de manière sinusoïdale.

On peut produire un son complexe en ajoutant des sinusoïdes - une méthode appelée **synthèse additive**:

$$y = A1 \sin (2\pi f1 n / R + \phi1) + A2 \sin (2\pi f2 n / R + \phi2) + \dots$$

Ceci est un exemple de la manière dont une seule expression algébrique peut produire un son. Naturellement, de nombreux autres programmes plus compliqués sont possibles. Quelques méthodes de synthèse telles que la synthèse additive, la synthèse par table d'ondes, la modulation de fréquence et la mise en forme d'onde sont présentées dans le **didacticiel MSP**.

Manipulation des signaux numériques

Tout son sous forme numérique - qu'il ait été synthétisé par l'ordinateur ou quantifié à partir d'un son «réel» - n'est qu'une série de nombres. Toute opération arithmétique effectuée avec ces nombres devient une forme de traitement audio.

Par exemple, la multiplication équivaut à l'amplification audio. En multipliant chaque nombre d'un signal numérique par 2, l'amplitude du signal est doublée (augmentée de 6 dB). Multiplier chaque nombre d'un signal par une valeur comprise entre 0 et 1 réduit son amplitude.

L'addition équivaut au mixage audio. Etant donné deux signaux numériques, un nouveau signal peut être créé en ajoutant les premiers chiffres de chaque signal, puis les deuxièmes chiffres, puis les troisièmes chiffres, etc.

Un écho peut être créé en rappelant des échantillons qui se sont produits plus tôt et en les ajoutant aux échantillons actuels. Par exemple, le signal qui a été envoyé 1000 échantillons plus tôt peut être renvoyé à nouveau, combiné à l'échantillon actuel.

$$y = xn + A y_{n-1000}$$

En fait, les effets que de telles opérations peuvent avoir sur la forme d'un signal (audio ou autre) sont si nombreux et variés qu'ils constituent une branche complète de l'ingénierie électrique appelée traitement numérique du signal (DSP). Le DSP s'intéresse aux effets des filtres numériques - des formules permettant de modifier les signaux numériques par des combinaisons de retard, de multiplication, d'addition et d'autres opérations numériques.

Résumé

Ce chapitre a décrit comment le phénomène continu du son peut être saisi et reproduit fidèlement sous la forme d'une série de nombres, et finalement stocké dans la mémoire d'un ordinateur sous la forme d'un flux de chiffres binaires. De nombreux avantages ne peuvent être obtenus qu'en vertu de cette représentation numérique du son: enregistrement d'une plus grande fidélité que ce qui était possible auparavant, synthèse de nouveaux sons par des procédures mathématiques, application de techniques de traitement numérique du signal aux signaux audio, etc.

MSP fournit une boîte à outils pour explorer cet éventail de possibilités. Il intègre l'enregistrement, la synthèse et le traitement audio numérique avec le contrôle MIDI et la programmation par objets de Max.