

Problèmes d'échantillonnage

René Bastian 2002–2008

Une expérience simple réalisable avec un vrai synthétiseur analogique (ou avec pythoneon) montre que l'oreille musicale peut distinguer des différences de hauteur très petites. C'est cette faculté qui permet aux musiciens de jouer juste. Pour que ce pouvoir de résolution puisse manifester, il faut que l'intensité soit raisonnable pour que l'oreille reste ouverte et ne cherche pas à se protéger contre les décibels. En outre, il faut - dans un premier temps - que les sons ne soient ni trop séparés dans le temps, ni trop dissociés par de grands sauts d'intervalles. Par exemple : deux tons t_0 et t_1 sont facilement discernables, bien que très proches, quand ils sont émis l'un après l'autre : (t_0, t_1, t_0, t_1) , mais si l'émission est telle que $(t_0, t_x, t_1, t_x, t_0, t_x, t_1)$ et que les tons t_x soient dans des tessitures éloignées de (t_0, t_1) ou très forts ou très longs, l'oreille aura de la peine à distinguer t_0 et t_1 .

Quand le son est numérisé, ce pouvoir de résolution dépend de la fréquence d'échantillonnage et du registre des sons (grave, médium, aigu). Dans les musiques traditionnelles de l'Iran ou de l'Inde, on fait usage d'intervalles qui ne se laissent pas enfermer dans le système dodécaphonique égalisé. Il est possible que ces intervalles autres soient perçues davantage comme des sortes de **coloration** que comme des différences de hauteurs(1).

Peu importe : il faut que ces colorations puissent être perçues exactement et enregistrées fidèlement. Le problème est le même en musique électronique : il est inutile de finasser si le résultat échantillonné est cabossé et distordu par une fréquence d'échantillonnage trop faible ou une résolution de l'échantillon trop faible (un codage sur 8 bits n'est acceptable que dans les cas où seule l'information sémantique est importante ; 16 bits sont acceptables mais un codage sur 24 ou 32 bits est souhaitable).

Soit sr la fréquence d'échantillonnage, f une fréquence du registre à étudier et i l'intervalle à étudier.

Un cycle à la fréquence f_0 occupe sr/f_0 échantillons et à la fréquence f_1 , un cycle occupe sr/f_1 échantillons. Si on souhaite que les deux cycles se distinguent dès que la première période ait été émise (c'est à dire après une durée égale à la durée d'un cycle), il faut que la différence entre le nombre d'échantillons soit de 1 en valeur absolue :

$$\text{abs}(sr/f_0 - sr/f_1) = 1$$

Disons $i = f_1 / f_0$. Après quelque gymnastique algébrique, on obtient :

$$sr = f_0 * i / (i-1)$$

ce qui signifie que, **si on souhaite qu'un intervalle i dans un registre de fréquences de l'ordre de f_0 soit musicalement perceptible, il faut que la fréquence d'échantillonnage soit égale ou supérieure à sr** . Quelques exemples :

Fréquences d'échantillonnages en fonction des intervalles et de la région des hauteurs musicales

Intervalle	formule	intervalle	f0 à 20 Hz	f0 à 440 Hz	f0 à 1000 Hz	f0 à 4000 Hz
demi-ton tempéré	$2^{1/12}$	1.059463094	356 Hz	7839 Hz	17817 Hz	71268 Hz
quart-de-ton	$2^{1/24}$	1.029302237	702 Hz	15455 Hz	35127 Hz	140508 Hz
12ème de ton	$2^{1/144}$	1.004825126	4164 Hz	91629 Hz	208248 Hz	832993 Hz
savart	$2^{1/300}$	1.004825126	8666 Hz	190655 Hz	433308 Hz	1733234 Hz
36ème de ton	$2^{1/432}$	1.001605795	12474 Hz	274447 Hz	623744 Hz	2494977 Hz
cent	$2^{1/1200}$	1.000577790	34634 Hz	761963 Hz	1731734 Hz	6926936 Hz

Exemple: si on veut percevoir un intervalle d'un 36e de ton aux environs de 1000 Hz, la fréquence d'échantillonnage devra être de au minimum 623 744 Hz (intersection de la colonne f0 à 1000 Hz avec la ligne 36ème de ton).

Comme le fait remarquer Jean-Jacques Sarton : la règle de Nyquist et Shanon dit qu'une oscillation peut être **reconstruite** à condition que le son soit suffisamment long – ce qui peut être lassant ...

Le débat concernant la fréquence d'échantillonnage ne pourra être clos, car il s'agit de compromis. Le propre d'un compromis est de rendre la vie supportable ici et maintenant, mais il peut être caduc si dans l'avenir la technologie offre une meilleure solution (96 kHz au lieu de 44.1 kHz ne sont qu'un compromis tout aussi transitoire). Que deviendront les partitions de musique électronique gravées une fois pour toutes en une fréquence d'échantillonnage donnée? Pour ne pas subir cette avanie, il suffit de concevoir les partitions et descriptions de musiques indépendantes de la fréquence d'échantillonnage et de ne procéder à la numérisation qu'en fin du processus de composition. C'est une des raisons pour lesquelles je préfère utiliser un langage informatique d'usage général au lieu de logiciels spécialisés pour décrire la musique. Ce langage d'usage général permet par exemple de calculer le tableau ci-dessus (le texte-source a 152 caractères).

L'opposition [digital vs. analogique] est similaire à l'opposition [intonation juste vs. tempéraments égalisés]. La musique électronique permet d'habituer l'oreille à davantage de finesse (ceci vaut autant pour la musique concrète que pour la musique artificielle), même si les techniques d'habituation actuelles vont en sens inverse.

Pour se rendre compte que la limitation de la fréquence d'échantillonnage n'est pas anodine, il suffit :

- d'écouter des enregistrements de chœurs: les voix humaines ne sont jamais tout à fait stables; ces petites fluctuations font la beauté des chœurs, mais leur enregistrement est un exercice périlleux et l'écoute d'un cd gravé à 44.1 kHz un moment pénible;

- de faire de la synthèse électronique : quand les fréquences sont proches, les phénomènes de battement attendus sont perturbés par le bruit de digitalisation ; on ne peut donc jouer avec ces battements qu'en augmentant considérablement la fréquence d'échantillonnage.

Un phonéticien néerlandais, Paul Boersma, prenant en considération la structure du conduit laryngé dans lequel se forme le signal acoustique de la voix humaine, est arrivé à la conclusion que la fréquence d'échantillonnage minimale devrait être d'au moins 500 kHz (voir Functional phonology section 3.8). En effet le tube le plus court que la voix doit traverser mesure 0.7 mm ; à la vitesse moyenne de 350 m/sec (compte tenu de la température du corps), le temps mis pour traverser ce conduit est de $350.0/0.0007$ soit $1./500000$ de seconde. Pour que l'effet de ce très court tube puisse être synthétisé il faut nécessairement que la fréquence d'échantillonnage ne soit pas inférieure à 500 kHz. On trouve assez facilement sur Internet le texte de Boersma avec un moteur de recherche et les mots clés : Boersma, phonology, praat.

On peut évidemment rétorquer qu'il suffit d'écouter de la musique normale: celle qui ne va pas au-delà de 500 Hz et n'a pas besoin d'intervalles hors des normes dont l'industrie musicale nous gratifie. Mais ce n'est pas donné à tout le monde : les « handicapés » qui ont l'oreille musicale ont aussi le droit de vivre.

1 Expérience

Les cinq exemples de sons glissés qui se trouvent sur ce site www.pythoneon.org montrent la dégradation progressive de la finesse des glissandos à mesure que la hauteur du son augmente:

- **songlisse88.wav** : les fréquences sonore est loin de la fréquence d'échantillonnage; aucune rupture n'est perceptible;
- **songlisse175.wav** : vers la fin on peut discerner comme une dissociation;
- **songlisse350.wav** : les ruptures de glissements apparaissent parce que la fréquence sonore se rapproche de la fréquence d'échantillonnage;
- **songlisse700.wav** : entre 700 Hz et 1400 Hz le son glissé est décomposé en une suite de sons isolés;
- **songlisse1400.wav** : là, on a l'impression qu'on pourrait faire passer un train entre les sons.

NB : les statistiques de consultation de cette page montrent qu'il arrive que seul l'exemple **songlisse88.wav** est joué; cela n'a pas de sens : il faut au moins comparer deux exemples pour se rendre compte de l'incidence de la fréquence d'échantillonnage.

Le sons sont réalisés par le petit python que voici :

```
from profils import recta
def exemples():
    si = SerieImpuls()
    duree = 8.0
    grain = numpy.array([1., -0.5, 0.5, -0.25, 0.25])
```

```

V = [
    [[recta, 1., 88., 176.]],
    [[recta, 1., 175., 350.]],
    [[recta, 1., 350., 700.]],
    [[recta, 1., 700., 1400.]],
    [[recta, 1., 1400., 2800.]]
]
for v in V:
    w = si.convolve(duree, v, grain)
    n = str(int(v[0][2]))
    nom = "songlisse" + n
    audio.fwav(w, w, nom)

```

Il combine un grain et une série d'impulsions selon un profil et produit ainsi une sonorité dont la hauteur varie. Selon le registre dans lequel se situe cette variation de hauteur on entend soit une variation continue soit une variation discontinue (en escalier ou comme une gamme). On constate qu'avec une fréquence d'échantillonnage à 44.1 kHz, on ne perçoit pas de discontinuités si la fréquence est inférieure à 175 Hz), que des discontinuités apparaissent entre 175 et 350 Hz et qu'elles sont évidentes si la fréquence est supérieure à 350 Hz.

Il est possible de réécrire ce script pour qu'il soit réalisable en Python de base (sans les modules de pythoneon).

Les résultats ont été présentés à Darmstadt-Wixhausen lors des journées **Grundlagenmusik** organisées par **Vemd - Hans Essel**: consulter [Vemd](#)

2 Explication

Résumons: Si la fréquence d'échantillonnage est trop basse, on constate auditivement un saut de hauteur quand la hauteur est censée varier continûment. Il convient de remarquer que de tels sauts de fréquence ne se manifestent pas

- quand la hauteur ne bouge pas (musique instrumentale ou pseudo-instrumentale habituelle);
- quand la variation de hauteur est rapide ou courte (inflexions de hauteur courantes dans l'intonation jazzistique).

Par contre, ils se manifestent dans le cas de variations de hauteur lentes principalement dans les registres de grande sensibilité auditive.

3 Discussions

Certains de mes propos ont été discutés sur [le forum d'Audiofanzine, Techniques du son](#)

Si les enregistrements de voix sont tellement épineux, ce n'est pas à cause de quelques battements – qui produiraient plutôt des très basses fréquences – mais à cause des nombreuses composantes aiguës, surtout des voix d'hommes. Le phénomène est aisément constatable sur certains enregistrements de ténors (p. ex. Fischer-Dieskau). L'échantillonnage trop bas de ces fréquences en fait du bruit erratique. Il n'y a alors qu'une solution : appliquer un filtre passe-bas.

En outre, il ne suffit pas de répéter que la fréquence d'échantillonnage est négligeable pour que ce soit vrai : un seul contre-exemple suffit pour prouver le contraire. Ceci dit quelques questions demeurent :

- l'enregistrement digital d'un son glissé de violon dans l'extrême-aigu semble être acceptable : effectivement, l'échantillonnage en continu cache le hoquet spectral sous un voile de bruit ;
- si la composition algorithmique du son révèle qu'à chaque fréquence d'échantillonnage correspond un contexte spectral particulier, il faut se rendre compte que – esthétiquement – on peut avoir besoin d'un grand nombre de fréquences d'échantillonnage. Donc que nous devons disposer d'un continuum de fréquences d'échantillonnage et peut-être même de fréquences d'échantillonnage glissantes ! Mais ceci est une autre histoire . . .

4 Esthétique

Qu'à chaque fréquence d'échantillonnage corresponde une certaine configuration du bruit d'échantillonnage se répercute sur l'esthétique : si on recherche une certaine configuration spectrale, elle n'est disponible qu'à une fréquence d'échantillonnage donnée – à moins de simuler les fréquences d'échantillonnages basses – ce qui, on s'en serait douté, n'est possible que si la fréquence d'échantillonnage est très élevée. Peu importe, pense-t-on. D'accord, mais alors pourquoi faire grand cas de telle manufacture de pianos ou de tel luthier ? Ce ne serait que du vent publicitaire ? Comme la philosophie des Heidegger, Derrida et consorts ?

Mises à jour

Mises à jour faites le 20.10.2002, 11.09.06, 2.03.08. La mise à jour de 2006 a ajouté le passage concernant Paul Boersma (qui confirme mes observations par d'autres arguments) et a modifié la présentation du tableau.

(1) Cette présomption de 2002 trouve un argument supplémentaire dans une expérience récente [2008].

