



HAL
open science

Des courbes et des sons

Daniel Arfib

► **To cite this version:**

Daniel Arfib. Des courbes et des sons. Journées d'informatique musicale 1994, Mar 1994, Bordeaux, France. pp.68-73. hal-03347048

HAL Id: hal-03347048

<https://hal.science/hal-03347048>

Submitted on 17 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Des courbes et des sons.

Daniel ARFIB
CNRS-LMA
31 Chemin Joseph Aiguier
13402 Marseille Cedex 20
ARFIB@lma.cnrs-mrs.fr

Cet exposé tente de décrire brièvement comment le compositeur de sons sur ordinateur peut s'appuyer sur des courbes pour non seulement décrire des sons, mais aussi en créer d'autres, qu'ils proviennent d'une synthèse ex-abrupto ou d'une resynthèse à partir de sons naturels.

Décrire un son.

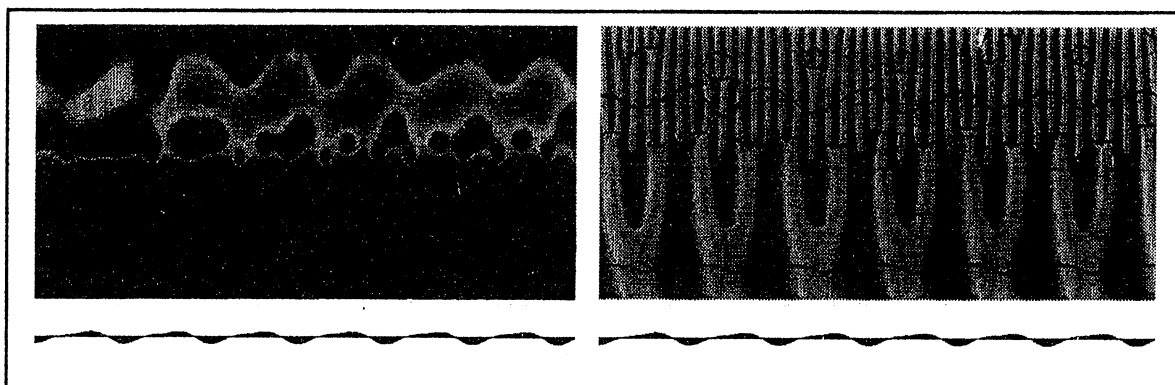


fig 1 et 2: représentation temps-fréquence en module (gauche) et phase et arêtes (droite)

Il est évident que décrire un son de synthèse par des appréciations subjectives est intéressant mais délicat à utiliser lorsqu'il s'agit de fournir des données à un ordinateur. Ce même ordinateur est d'ailleurs un outil permettant d'obtenir de telles courbes à partir d'analyses de sons naturels. Une des représentations les plus classiques est celle du sonagramme, mais même avec une telle représentation temps-fréquence, il est possible de découvrir des aspects insoupçonnés d'un son. A titre d'exemple, les figures 1 et 2 sont obtenues par transformée de Gabor d'un son vocal, dont le signal est représenté sur la ligne inférieure. Les diagrammes des figures 1 et 2, recomposés correctement, permettent de réobtenir le son d'origine par une transformée inverse. Les arêtes, indiquées sur la figure 2, sont des points du plan temps-fréquence où l'énergie est spécialement localisée, ce qui permettrait une reconstruction à partir d'une vue partielle. De ces diagrammes, il est possible d'obtenir différentes courbes, qui présentent un intérêt certain, surtout si on veut ensuite alimenter un système de synthèse à partir de ces courbes.

mesure globale de l'amplitude en fonction du temps.

Il s'agit ici de décrire l'énergie du signal en fonction du temps. La seule définition mathématiquement valable de l'amplitude d'un signal serait le module du signal rendu analytique en tenant compte de sa transformée de Hilbert. Malheureusement cette définition introduit des oscillations à l'intérieur de l'amplitude qui le rendent impropre à son utilisation conventionnelle. Il faut alors passer par des définitions comme celle du RMS (root mean square) qui font intervenir une fenêtre temporelle, qu'elle soit récursive ou non.

amplitude instantanée par Transformée de Hilbert:

Vue du côté musicien, la transformée de Hilbert consiste à faire pivoter les composantes d'un son d'une phase de $\pi/2$. Cela peut se réaliser de plusieurs manières: l'une est de prendre la transformée de Fourier du signal entier, et d'ajouter $\pi/2$ aux phases de sa représentation polaire (ou d'opérer $x'=-y$ et $y'=x$ sur les coordonnées cartésiennes) avant de pratiquer une transformée de Fourier inverse; on peut aussi obtenir une bonne approximation de la transformée de Hilbert d'un son en le filtrant à l'aide d'un filtre non récursif, qui représente la réponse impulsionnelle d'un filtre déphaseur de $\pi/2$, ou encore en combinant des filtres récursifs qui déphasent globalement de $\pi/4$ et de $-\pi/4$.

Une fois ce signal transformé: $sh(t) = H(s(t))$

il suffit de réaliser l'opération: $a(t) = \sqrt{s^2(t) + sh^2(t)}$

Malheureusement pour nous, l'amplitude obtenue par transformée de Hilbert possède des oscillations, c'est à dire qu'on sépare mal la partie enveloppe de celle du contenu sonore.

amplitude par RMS:

La majorité des systèmes électroniques de détection d'amplitude utilisent le calcul du RMS. Voici en quoi cela consiste: il s'agit de filtrer par un filtre passe-bas le carré du signal sonore. Le filtre le plus simple est un filtre passe-bas à un pôle du genre:

$$y_n(t) = s^2(t) + k y_{n-1}(t)$$

$$a(t) = \sqrt{y_n(t)}$$

Des systèmes plus sophistiqués combinent deux filtres pour permettre un suivi des attaques plus rapides que celui des résonances. On voit par là combien certaines notions à priori simples font intervenir, et c'est logique, la perception.

Il est aussi possible, notamment lorsqu'on a pas d'impératif de suivi en temps réel, de réaliser le filtrage du carré du signal par une fenêtre temporelle, telle la fenêtre de Hanning. On a alors la notion d'un épaisseur de présent, qui correspond à la largeur de cette fenêtre. Choisie trop petite, elle fait que l'amplitude suit la valeur instantanée du signal, trop grande elle amollit les transitoires. Le compromis réside en fait à ce qu'une expérience simple montre: la répétition d'une impulsion basse fréquence est perçue comme un rythme, puis un grain (entre 10 et 25 par seconde), puis une fréquence (au delà de 20), et ce même comportement doit être enregistré sur la détection d'amplitude. C'est pourquoi des fenêtres d'analyse de 20 à 50 millisecondes permettent une détection d'amplitude évitant les oscillations parasites et les amollissements trop inconvenants.

détection de fréquence.

Là aussi il faut s'entendre sur ce que l'on appelle la fréquence d'un son. La seule définition théorique serait la fréquence instantanée du signal analytique, c'est à dire la dérivée de la phase du signal obtenu en prenant pour partie réelle le signal et pour signal imaginaire la transformée de Hilbert.

par transformée de Hilbert:

On peut écrire:

$$sh(t) = a(t) e^{2\pi jft}$$

on calcule la phase instantanée: $\phi(t) = \text{Arctg}(sh(t)/s(t))$

et sa dérivée par rapport au temps permet de calculer la fréquence instantanée. mais celle-ci varie à l'intérieur même d'une période et par là-même répond peu à ce que l'on appelle usuellement la fréquence fondamentale du son.

systèmes de détection électroniques ou numériques de fréquence fondamentale:

Pour détecter une fréquence fondamentale, celle que l'on "entend" il faut bien sûr qu'il y en ait une. Deux cas particulièrement délicats sont celui où un harmonique supérieur est prépondérant. L'autre est celui où les partiels sont inharmoniques, même légèrement. Lorsqu'au contraire une quasi-périodicité existe on peut rechercher l'évolution d'une période avec le temps.

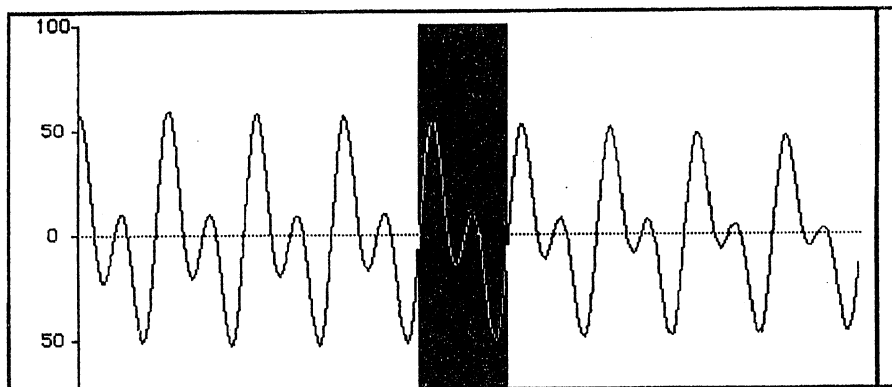


fig 3: extraction d'une période

Les détecteurs de fondamental peuvent en gros se classer en trois domaines:

- le domaine temporel: si la forme d'onde est suffisamment lisse pour que l'on puisse détecter les passages à zéro, le tour est joué. Par contre, et c'est le cas général, il faut souvent filtrer le son pour empêcher la forme d'onde de traverser plusieurs fois la ligne du zéro. Il est possible aussi dans ce domaine de détecter les crêtes, et ceci est souvent employé en analyse de la parole. Cette fois il ne faut pas rater une crête, ce qui nécessite une courbe de décroissance du détecteur, et il faut aussi ne pas détecter des crêtes doubles.

-l'auto-corrélation à court terme: il y a une brusque remontée de l'autocorrélation autour de la valeur de la période, mais la valeur du seuil doit éviter les fausses périodes, et ne pas rater la bonne.

-le domaine fréquentiel: cela consiste à prendre une transformée de Fourier à court terme et d'essayer de découvrir une structure harmonique. Il faut alors tenir compte du fait qu'une FFT ne tombe pas pile sur les valeurs exactes, et que d'autre part il faut user d'astuce pour éviter par exemple de tomber dans l'erreur quand un harmonique est prédominant.

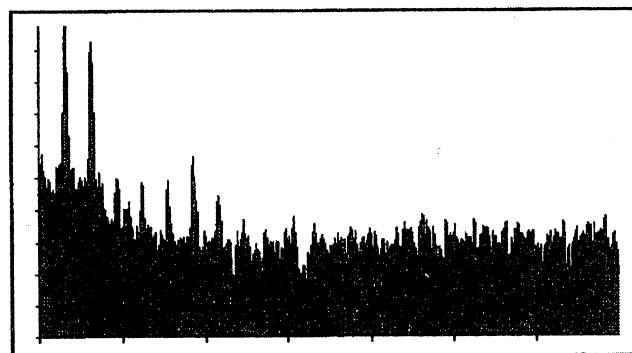


fig 4: Transformée de Fourier (les pics sont les harmoniques)

Rien n'est donc parfait, mais en ajustant quelque peu certains paramètres, on peut distinguer une évolution du fondamental. Ainsi sur un son de voix chantée, on pourra obtenir la courbe du vibrato (fig 5). Pour en connaître la fréquence, on peut prendre la transformée de Fourier de cette courbe d'évolution de fréquence, et ceci nous indique une fréquence de vibrato prépondérante à 8.1 Hz.

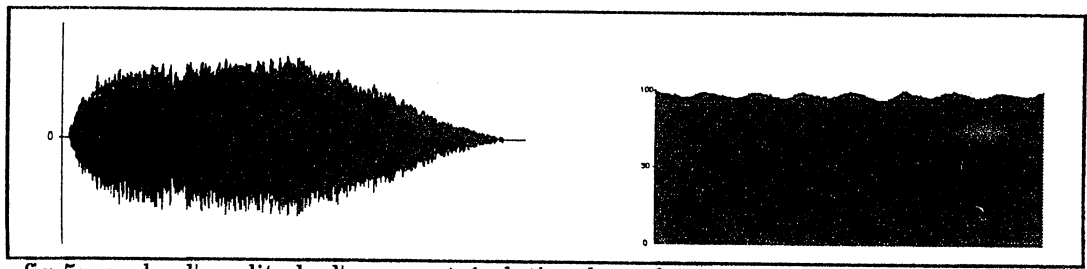


fig 5: courbe d'amplitude d'un son et évolution de sa fréquence fondamentale (vibrato)

Les autres paramètres.

Il est évident qu'une synthèse qui ne se préoccuperait que de ces deux courbes serait un peu défailante. Il faudrait au minimum une forme d'onde, et d'une manière plus générale l'évolution d'une forme d'onde.

extraction d'une forme d'onde.

Extraire une forme d'onde d'un signal sonore relève de la micro-chirurgie. Il faut non seulement repérer une période (ce que l'être humain fait bien avec un ordinateur) mais ensuite trouver les passages à zéro par interpolation et étendre cette période sur un nombre déterminé de points. Cela l'ordinateur peut le faire.

On peut aussi, sans extraire une période, pratiquer une transformée de Fourier fenêtrée, si possible avec un grand nombre de points et une fenêtre de type gaussienne assez étroite. On peut alors relever les valeurs des harmoniques qui donnent ainsi le spectre d'une forme d'onde.

évolution de la forme d'onde.

On désire maintenant disposer d'un moyen de décrire l'évolution du son avec le temps. Il est possible de pratiquer une dissection du son harmonique; l'évolution de la somme des harmoniques rendra alors le son. Ceci présente quelques difficultés théoriques, basées sur le fait que si l'on veut être précis en temps pour analyser chaque enveloppe d'un harmonique, alors il y a interférence des autres harmoniques. Il faut alors "démêler" les contributions des autres.

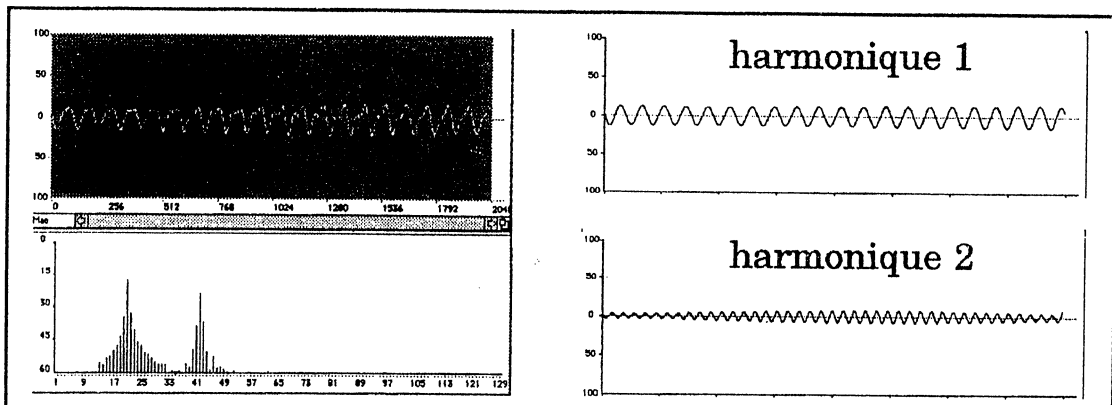


fig 6: transformée de Fourier à court terme (gauche) et décomposition selon les harmoniques (droite)

Quelle que soit la puissance potentielle d'une telle analyse, elle donne lieu à un nombre de courbes impressionnant, et on peut se demander s'il n'y a pas un moyen de gérer plus globalement ces évolutions.

Certains auteurs ont proposé de corrélérer ces enveloppes soit entre elles, soit en fonction de

l'enveloppe globale. Il semblerait qu'au moins les cuivres bénéficieraient d'une réduction de données substantielles moyennant la donnée de ces courbes. On peut aussi essayer de repérer s'il n'y a pas un comportement formantique. Sur une note, cela est délicat, mais si l'on possède plusieurs exemplaires joués sur des hauteurs différentes, un formant se repère par le fait qu'une résonance de hauteur fixe se trouve sur toutes ces notes. Lorsqu'on dispose d'une gamme entière, une technique a été développée qui permet de calculer un filtre global qui modélise ces formants. Par ailleurs un indice d'importance est le barycentre du spectre, car la "richesse", la "brillance" d'un son évolue avec le temps à l'intérieur d'un même son, et cette évolution est caractéristique. Des techniques spécifiques de synthèse comme la distorsion non-linéaire n'ont d'ailleurs besoin que d'un paramètre, l'index, pour gouverner la force de cette distorsion. On ne peut bien sûr imaginer retranscrire toute la véracité du son, mais au moins son allure générale.

Une méthode particulière de synthèse mérite d'être citée, c'est la modulation de fréquence, la FM, et il serait intéressant de pouvoir retrouver, sur un son naturel, l'équivalent de cet index de modulation qui rend magique cette méthode. Cela n'est pas évident, pour un problème de "bulles": on espèrerait avoir une méthode d'analyse qui délivre une sorte de fréquence instantanée qui soit oscillante par rapport à une fréquence porteuse. Mais une représentation temps-fréquence, qu'elle soit de Gabor ou en ondelette, si elle délivre parfois de telles courbes, engendre aussi des configurations plus tumultueuses qui empêchent toute modélisation. Un certain nombre de pistes sont suivies dans ce domaine, mais à ce jour aucune n'est garantie ni automatique.

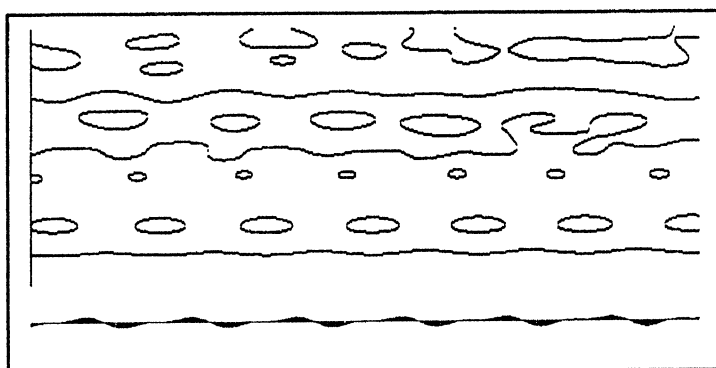


fig 7: extraction d'arêtes

Les synthèses à partir de courbes issues d'analyse.

Resynthétiser un son en parfaite imitation est un jeu intéressant certes, mais guère plus que la restitution d'un son à partir de ses échantillons si on ne peut rien en faire d'autre. Ce qui est intéressant musicalement, c'est de pouvoir dévier, soit par valeurs infimes qui donnent des variétés allotropiques, soit par grands débordements, ce qui donne des mutations génétiques. Lorsqu'on veut modifier un son, on peut en gros agir par trois différents moyens:

- l'un est de partir de la définition du signal lui-même et de lui appliquer un traitement global. Parmi ces méthodes, le filtrage et la réverbération font merveille. Il s'agit alors de convoluer un son avec une réponse impulsionnelle, qui représente la coloration qu'obtiendra le son d'origine. On peut mettre en oeuvre cela soit par convolution directe, soit en utilisant un algorithme de FFT, mais en aucun cas on ne se soucie du contenu du son.

- le second est d'effectuer une décomposition temps-fréquence, c'est à dire d'évaluer sur un plan de référence la valeur de la convolution du son avec un système d'ondelettes ou de gaborettes transposées en fréquence et déplacées en temps. Tout le jeu

musical consiste alors à déformer ce plan temps-fréquence pour apporter des transformations sonores. Comme je l'ai longuement indiqué par ailleurs, ceci n'est pas aussi simple qu'il n'apparaît, du à l'effet transverse d'un noyau reproduisant qui fait que ce que l'on obtient n'est pas ce que l'on désire mais ce que l'on désire plus l'effet d'une sorte de filtrage de l'image par un noyau de convolution (évolutif en temps et fréquence, ce n'est pas un pur filtrage) qui agit sur l'image temps-fréquence.

- le troisième, qui autorise tous les débordements, est d'extraire des courbes de sons, et de les appliquer à un quelconque paramètre de resynthèse. Il est évident qu'une conduite vocale de fréquence, cela ne s'invente pas facilement, et qu'il est possible de partir d'une conduite naturelle. A titre d'exemple, on peut voir qu'en distinguant dans une telle conduite la partie "lisse" et la partie "vibrato", on rend indépendantes ces deux parties, permettant de dissocier deux éléments à priori liés. De même en séparant dans le domaine spectral la contribution de la source et de la résonance, on peut agir indépendamment sur eux, par exemple en "octaviant" une parole sans en détruire son intelligibilité.

Il n'en reste pas moins certain que l'étape d'une reconstruction par similitude est importante, car qui veut le plus doit pouvoir le moins. Néanmoins certaines méthodes de synthèse globales, comme la FM ou la distorsion non-linéaire, dont on sait à priori qu'elles limitent l'imitation peuvent se révéler performantes justement par la possibilité d'ajuster quelques paramètres pour en trouver des clones pas bien ajustés mais musicalement intéressants.

Conclusion.

Ce tour d'horizon n'est là que pour mettre en valeur certaines possibilités habituellement inconnues des compositeurs, celles de travailler au coeur du son, de travailler le son et non pas seulement avec le son, en le décrivant en sa totalité. Cet aspect plutôt rébarbatif devient plus souple et alléchant avec les techniques numériques d'analyse-synthèse actuelles, dans la mesure où l'objet sonore, palpable et audible, est aussi décrit dans des termes qui le rendent graphiquement manipulables. Toutefois même ces recherches avancées ne sont que des balbutiements, notamment au niveau des manettes (réelles ou virtuelles) de contrôle musical des sons. Cela ressemble encore à un meccano où il faut assembler soi-même les pièces, et encore quand elles veulent bien s'assembler, mais c'est à coup sûr une voie intéressante qui se développera de plus en plus.

D. Arfib (1991). Analysis, transformation, and resynthesis of musical sounds with the help of a time-frequency representation. In G. de Poli, A. Picciali & C. Roads, editors, *The representation of musical sounds*, M.I.T Press, 87-118.

D. Arfib, R. Kronland-Martinet (1993). Transformer le son: modeler, modéliser. *Les Cahiers de l'IRCAM*, n° 2: la synthèse sonore, 67-82.

D.Arfib, N. Delprat. Musical Transformations using the modification of time-frequency images, *Computer Music Journal*, vol 17, n°2, summer 93, pp 66-72
