

Licence, département Informatique, CNAM
Multimédia & interaction humain-machine (2004-5)

Le son numérique

P. Cubaud <cubaud@cnam.fr>

- 1. généralités**
- 2. acquisition, traitements (analyse)**
- 3. synthèse**
- 4. transport, compression**

Bibliographie

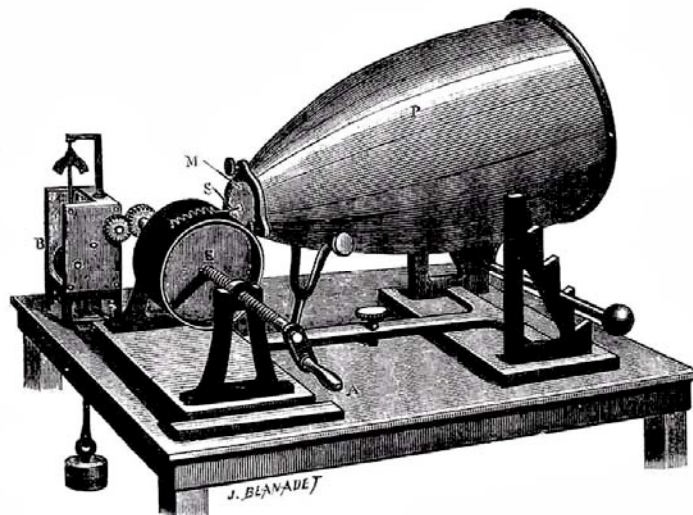


1. Généralités

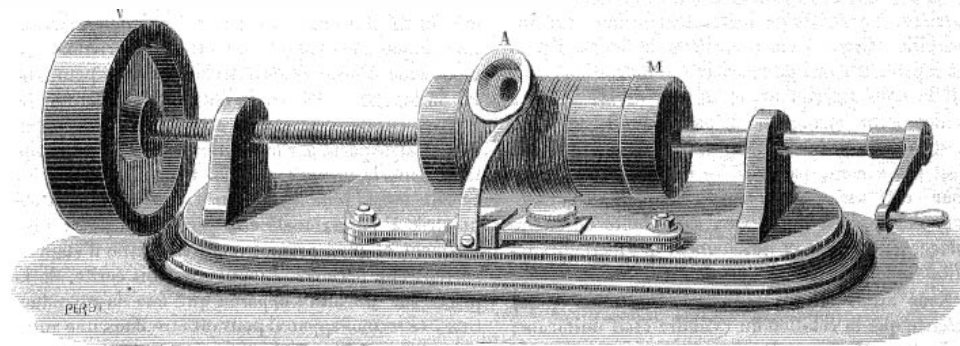
- voir des sons
- formes d'ondes élémentaires
- représentation spectrale
- la parole
- le son musical
- temps et calcul

Voir des sons...

- Léon Scott de Martinville. Phonautographe (phono-autographe ?), 1857
 - Rudolph Kœnig "Quelques expériences d'accoustique" Paris, 1882
- => Charles Cros, Thomas Edison : le phonographe



www.phonautograph.com

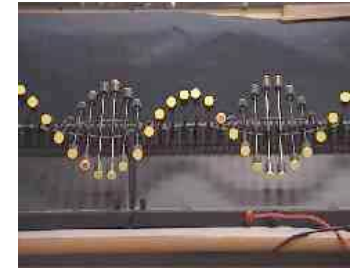
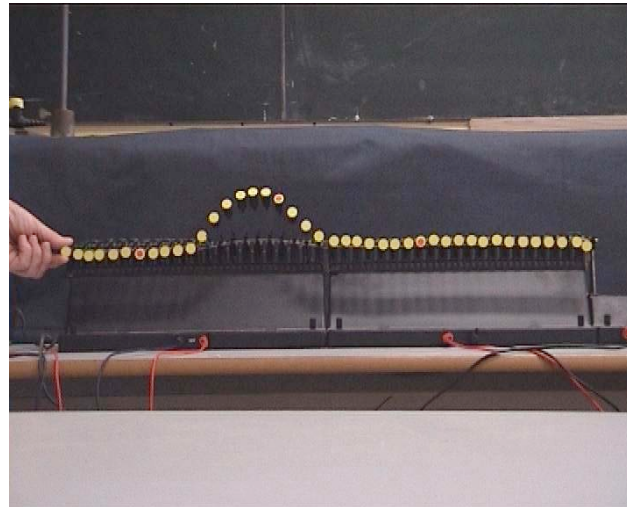


cnum.cnam.fr

Le phonographe ou machine parlante d'Edison. (1/6 grandeur d'exécution.)

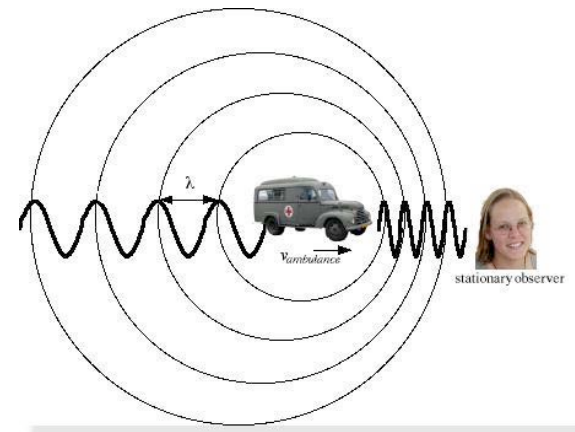
Onde :

- propagation d'une énergie dans un milieu élastique
- période T (s)
- longueur d'onde L (m)
- fréquence $F = 1/T$ (Hz)
- célérité $c = L/T$



Onde sonore :

- milieu : air (ou gaz, eau... $\neq$ StarWars)
- célérité : 330 (340) m/s dans l'air à 16° (20°)
- réflexion et réfraction, niveau ambiant ...
- gamme de fréquence : 20Hz à 20KHz
- énergie min pour l'oreille : 10^{-16} W/m²
- durée min de perception d'un timbre : 50 ms ???



Intensité sonore : le décibel (dB)

	intensité (W/m ²)	rapport I/I ₀	niveau (dB)
seuil de douleur	10 ⁰ =1	10 ¹²	120
fff	10 ⁻²	10 ¹⁰	100
f	10 ⁻⁴	10 ⁸	80
p	10 ⁻⁶	10 ⁶	60
ppp	10 ⁻⁸	10 ⁴	40
seuil d'audibilité	10 ⁻¹²	1	0

coup de marteau sur plaque acier à 70 cm : 114 dB

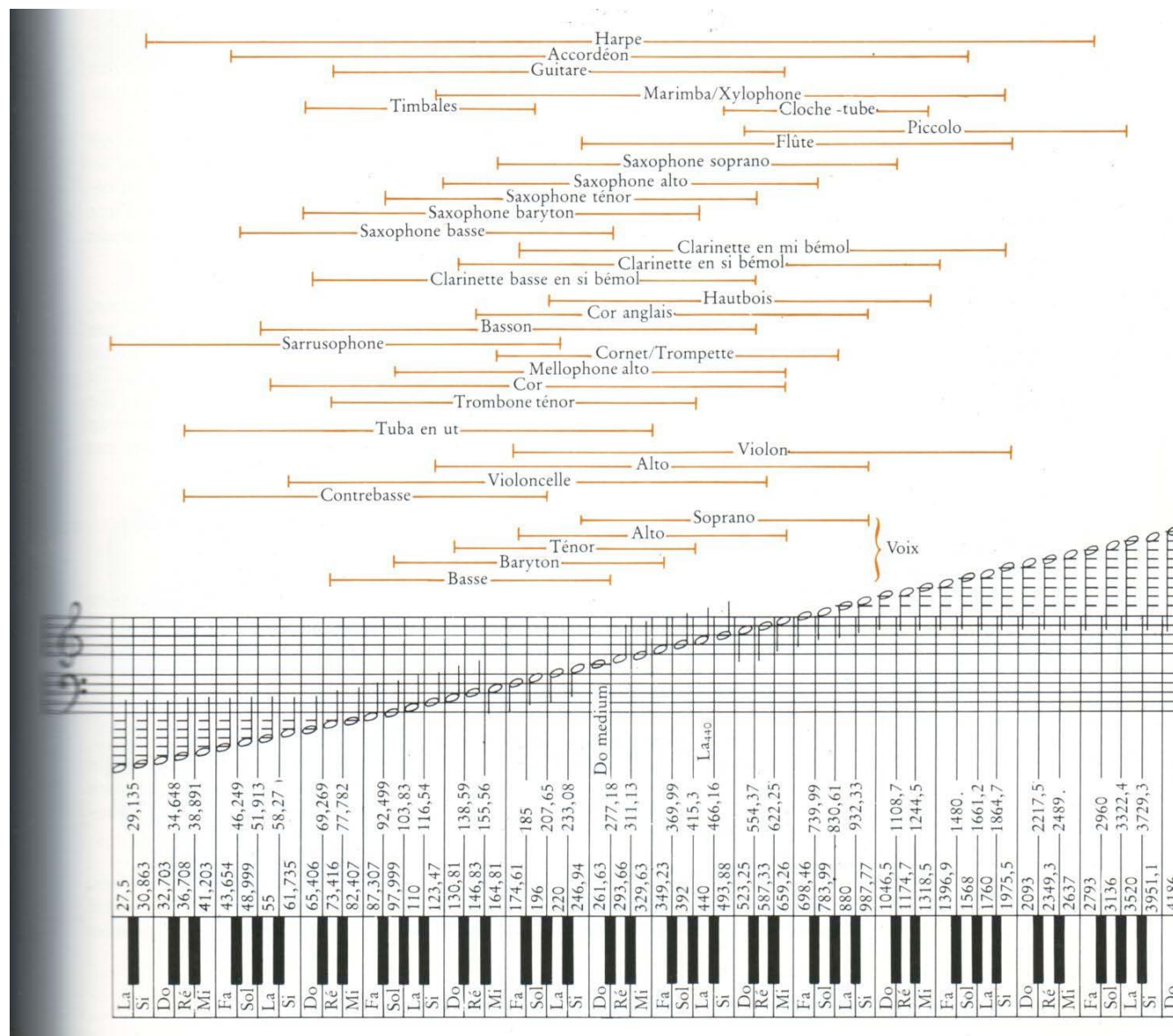
piano à 3m : 100 dB

conversation à 1m (moyen) : 65 dB

ma machine à laver la vaisselle à 1m : 43 dB => **et la votre ???**

murmure : 30 dB

Gammes de fréquence [Pierce, p. 21]



Sons élémentaires



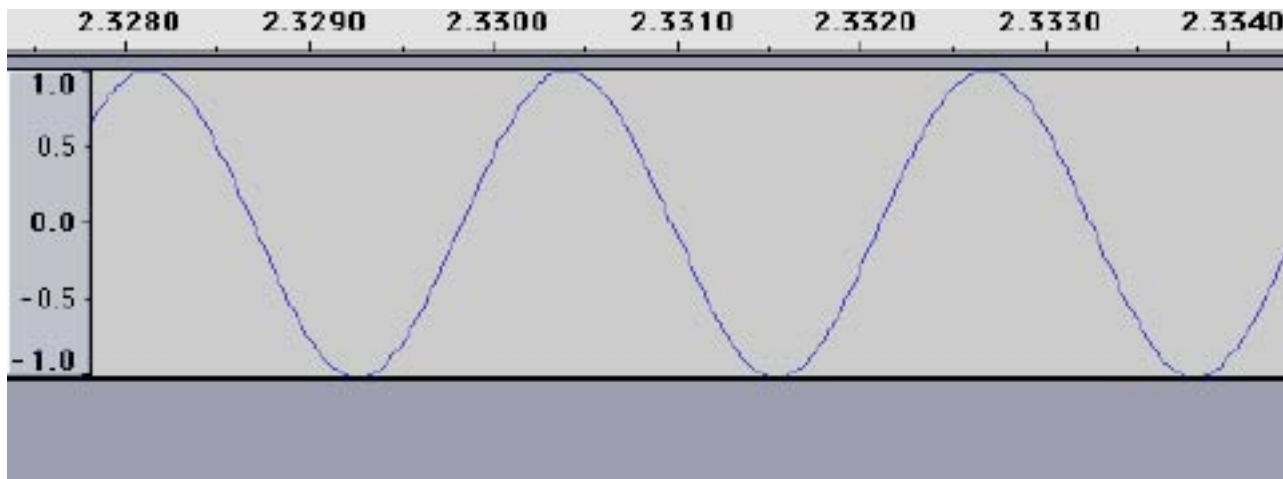
Sinusoïde 100 Hz



Sinusoïde 440 Hz



Sinusoïde 4400 Hz

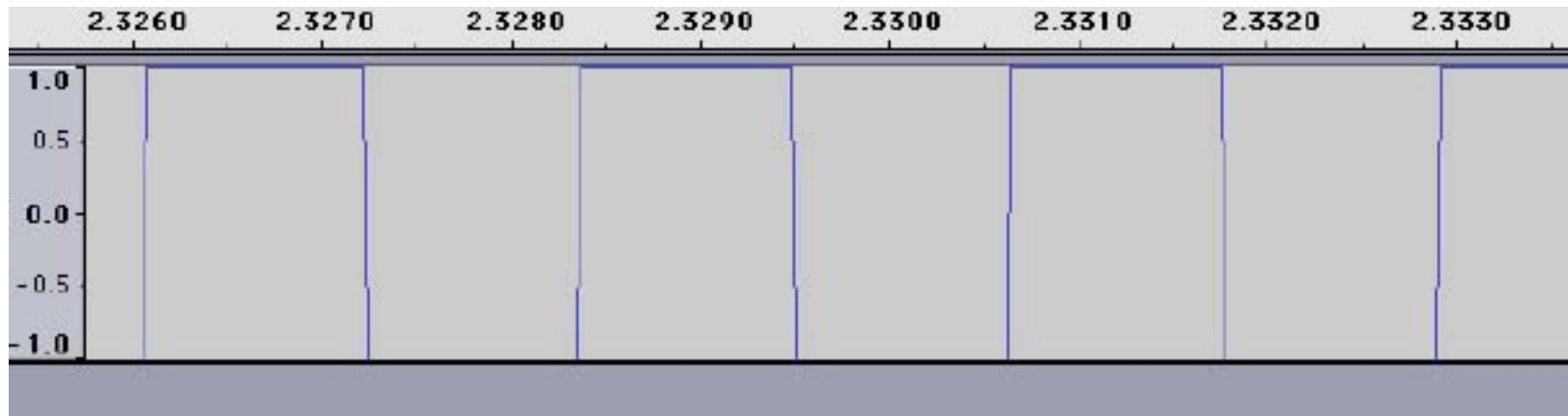




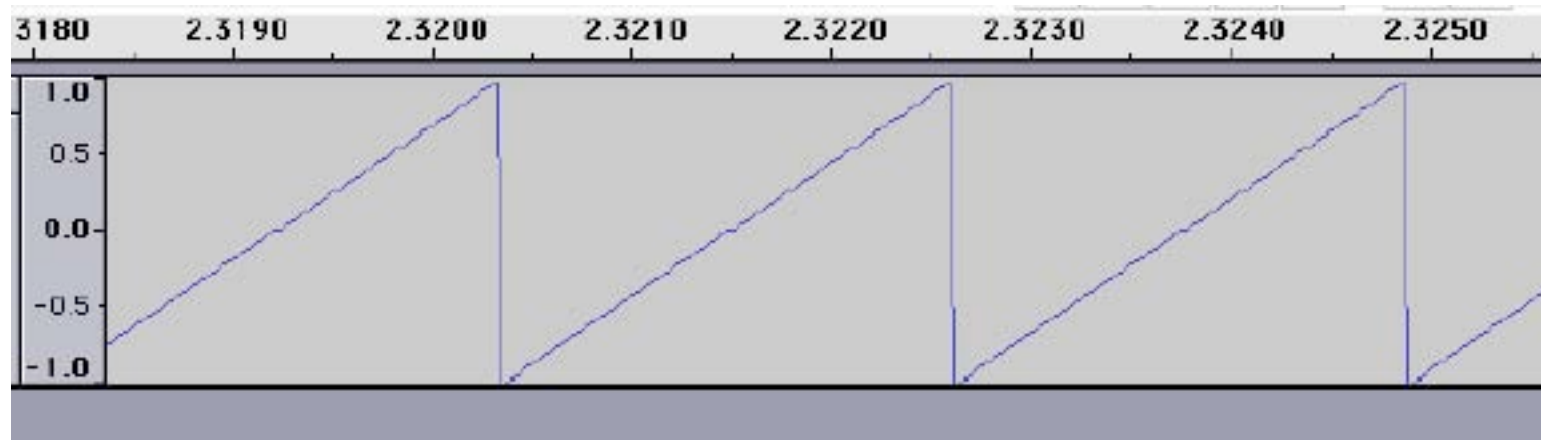
Signal carré 440 Hz



(Sinusoïde 440 Hz)



dent de scie 440 Hz

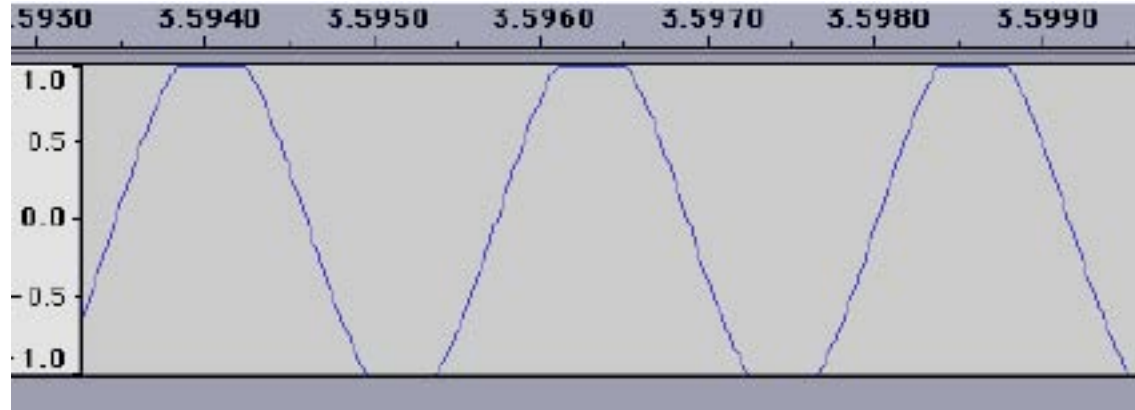




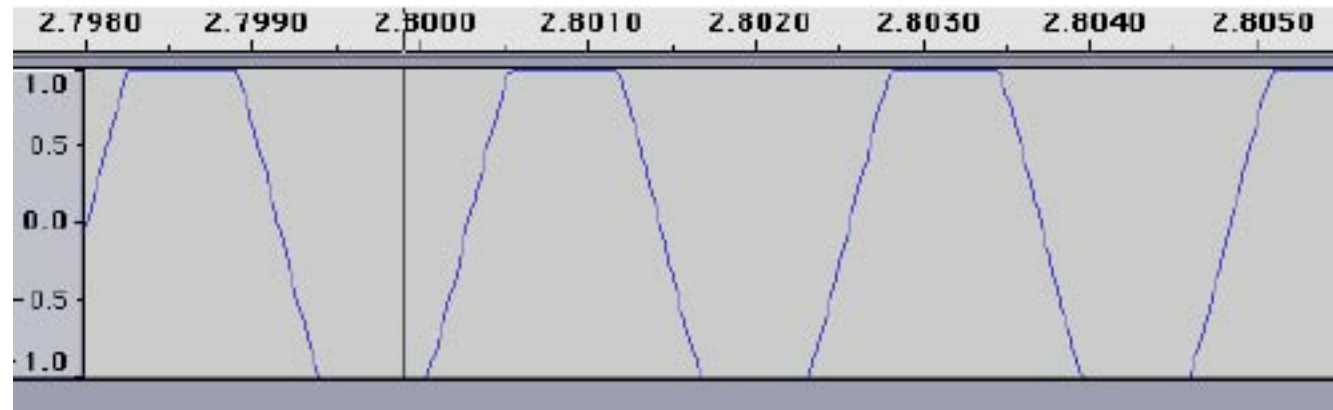
Sinusoïde 440 Hz



Sinusoïde 440 Hz écrettée (+1.5 dB)

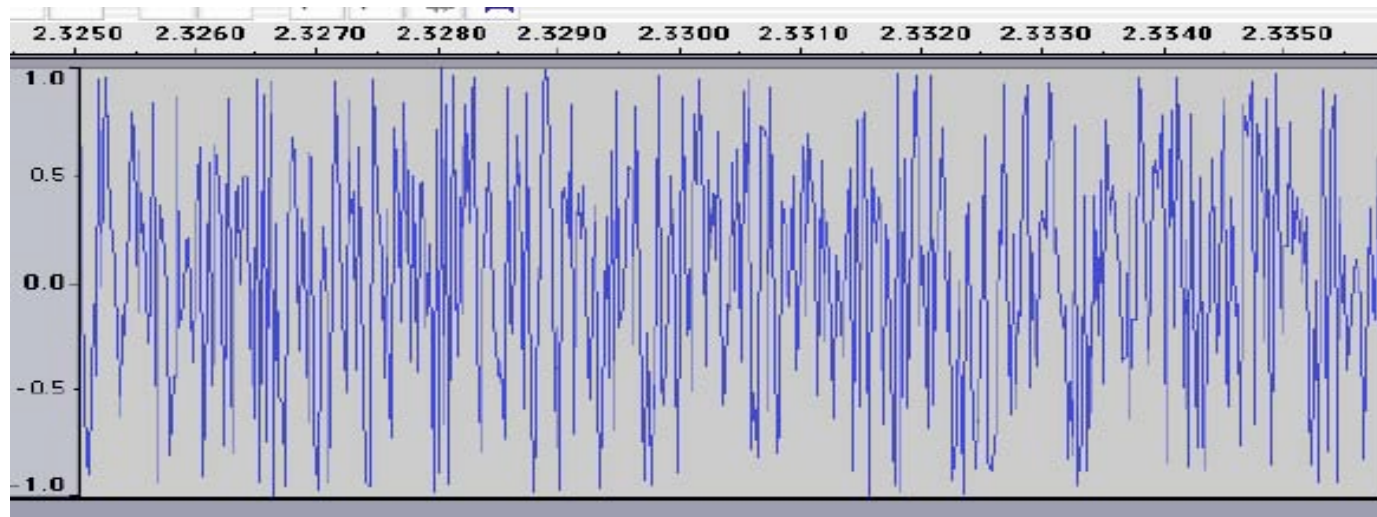


Sinusoïde 440 Hz écrettée (+3 dB)

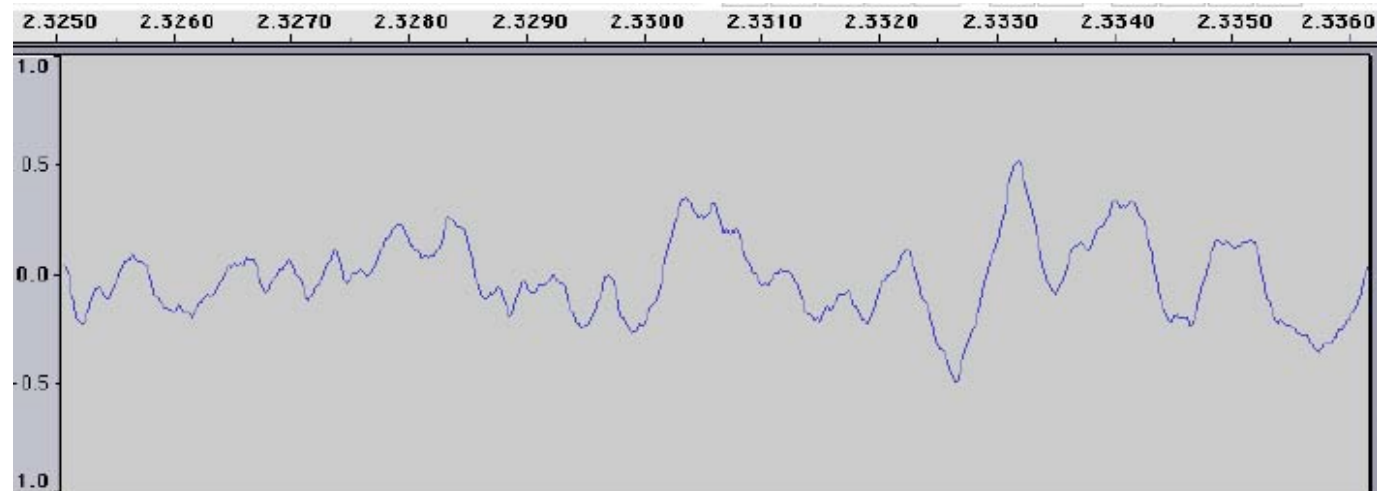




Bruit blanc

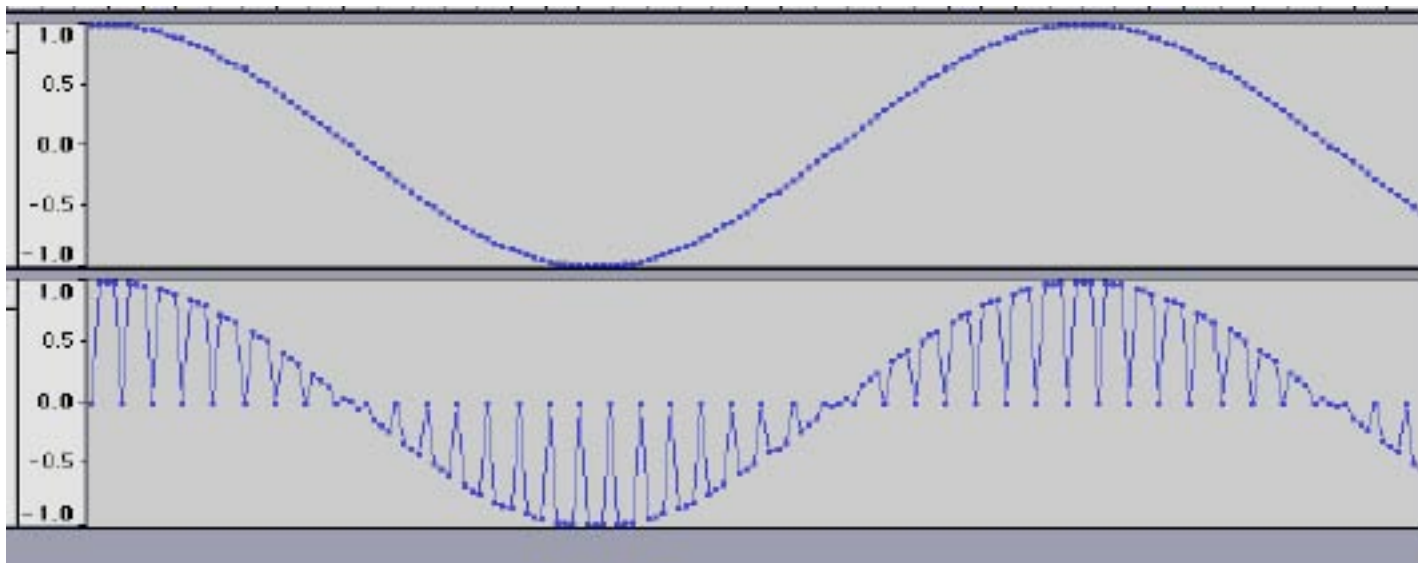


Bruit rose 430-450 Hz - comparer avec sinus 440 :



Représentation spectrale

Une première expérience : on annule des échantillons dans la sinusoïde.



Sinus d'origine



1 sur 4 annulé



1 sur 10 annulé

=> Deux sons perçus

Analyse harmonique (Fourier)

- $f(t)$ signal périodique borné (etc.) de pulsation $\omega=2\pi/T$

$$f(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos n\omega t + \sum_{n=1}^{\infty} B_n \sin n\omega t = A_0 + \sum_{n=1}^{\infty} C_n \cos(n\omega t - \varphi_n)$$

$$A_n = \frac{1}{T} \int_0^T f(t) \cdot \cos n\omega t \cdot dt \quad B_n = \frac{1}{T} \int_0^T f(t) \cdot \sin n\omega t \cdot dt$$

$$C_n = \sqrt{A_n^2 + B_n^2} \quad \varphi_n = \arctan \frac{B_n}{A_n}$$

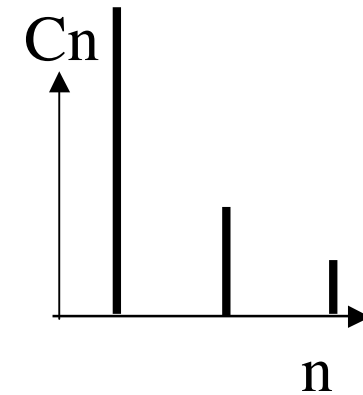
f est paire $\Rightarrow B_n$ nuls

f est impaire $\Rightarrow A_n$ nuls

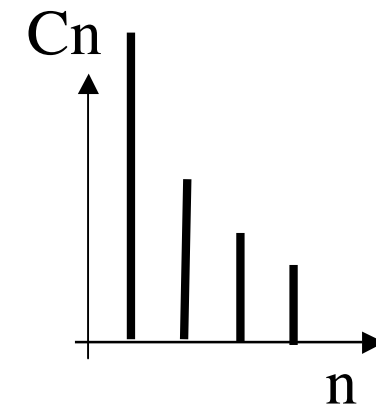
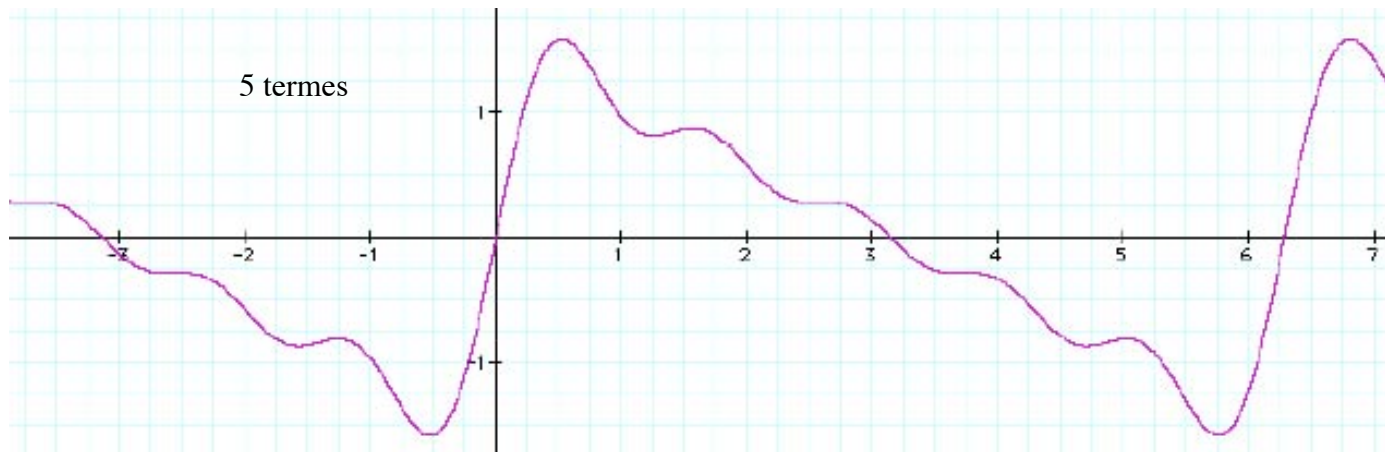
$f(x+T/2)=-f(x) \Rightarrow A_{2n}$ et B_{2n} nuls

- signal carré : $f(t) = \frac{4a}{\pi} \left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots \right)$

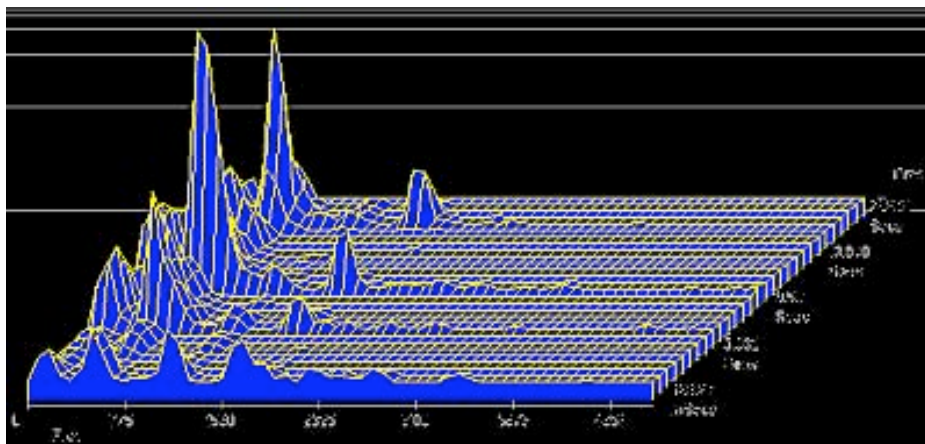
Spectre :



- dents de scie : $f(t) = -\frac{8a}{\pi} \left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \dots \right)$



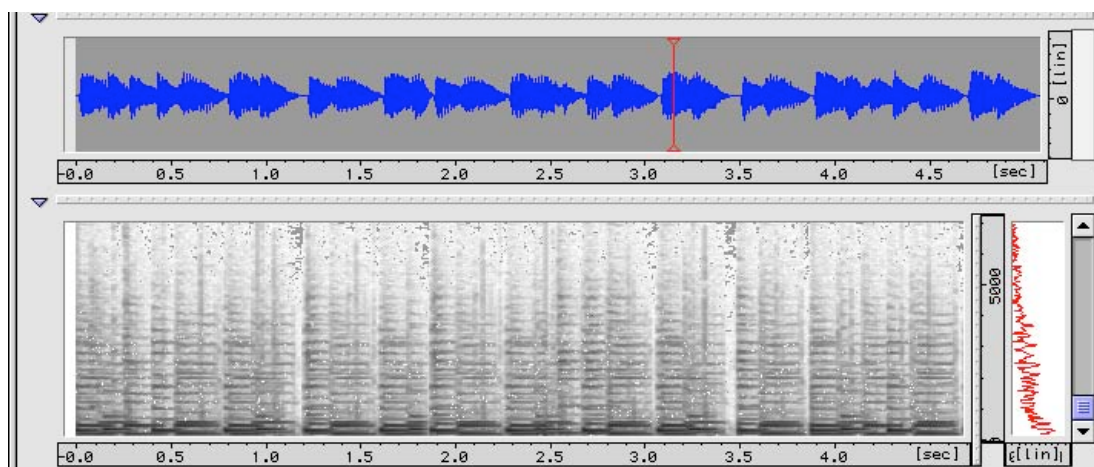
Sonagramme : le spectre en fonction du temps



http://www.inrp.fr/Acces/JIPSP/phymus/m_techni/sonagra/ac_sona.htm

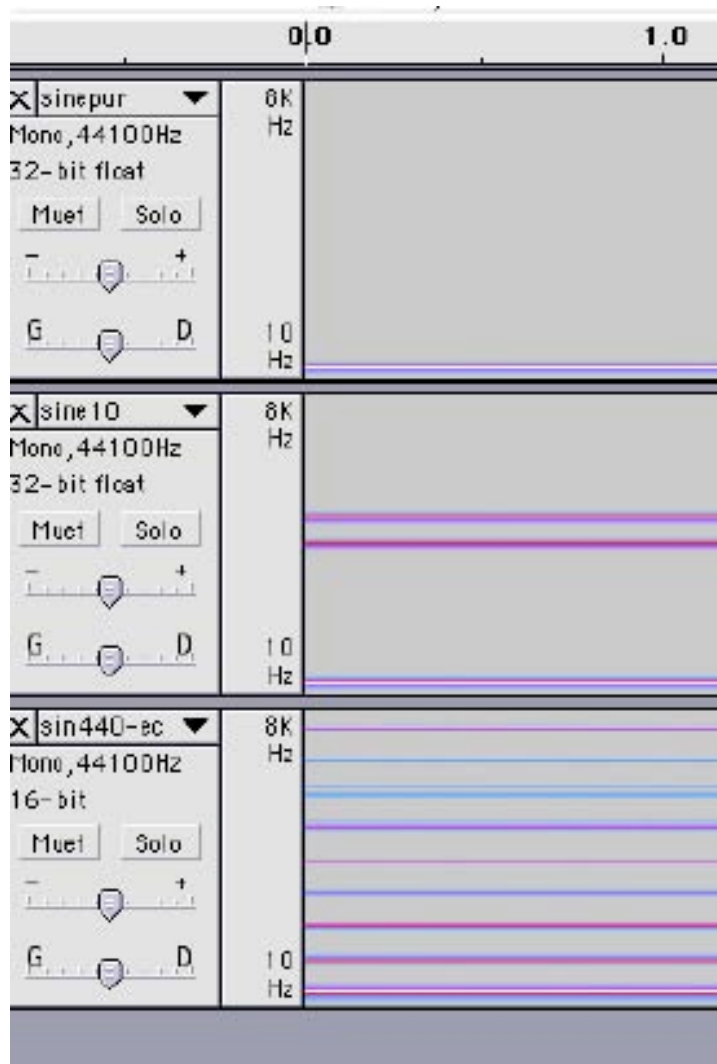
Virtual waves

AudioSculpt



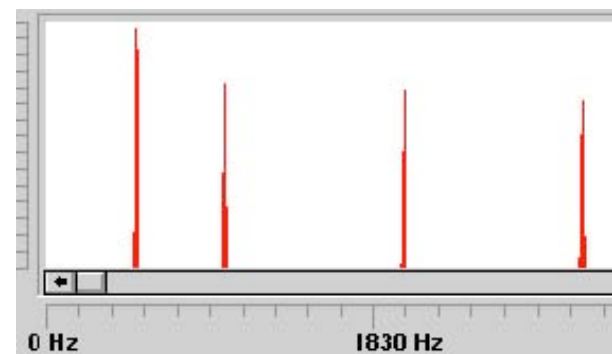
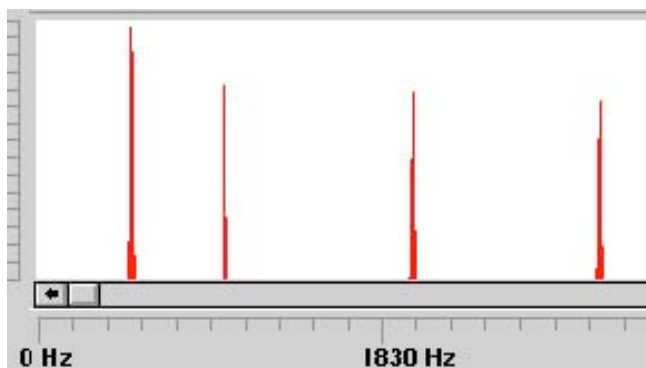
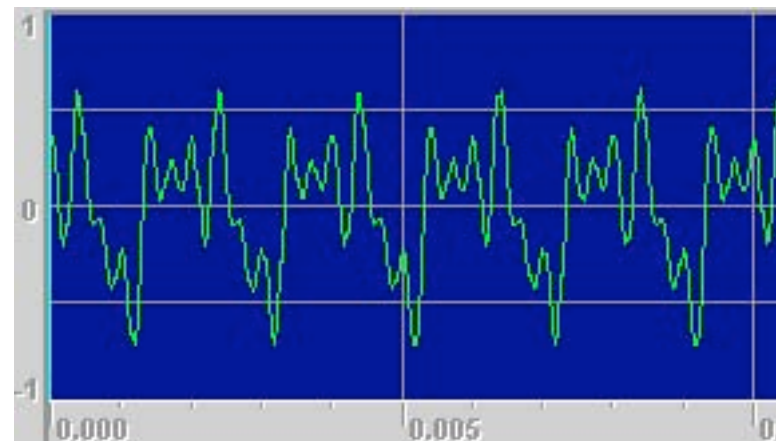
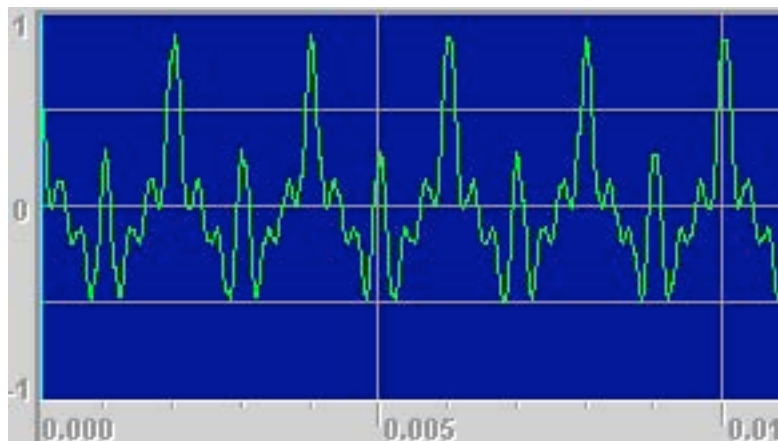
<http://recherche.ircam.fr/equipes/analyse-synthese/ralley/AS2Doc/AudioSculpt2.html>

ex. d'Audacity (macosX)



<= bruit

Et la phase ?



LE SONAGRAMME UNE REPRÉSENTATION INCONTOURNABLE POUR L'ANALYSE SONORE

Daniel BEAUFILS, Jean-Claude LE TOUZÉ© INRP - TECNE Dernière mise à jour 11/05/1998

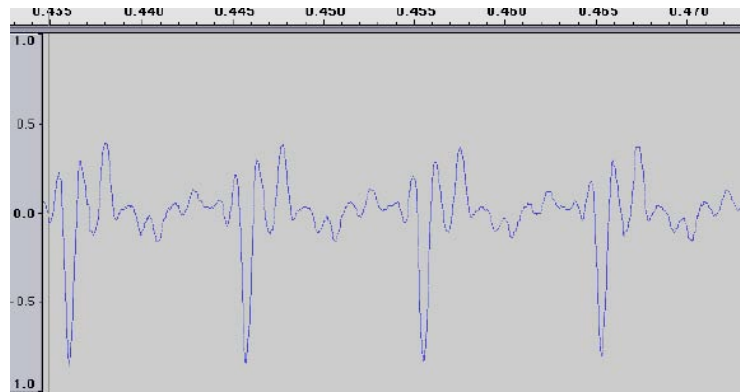
http://www.inrp.fr/Acces/JIPSP/phymus/m_pedag/sonagr/ac_sonag.htm

La parole

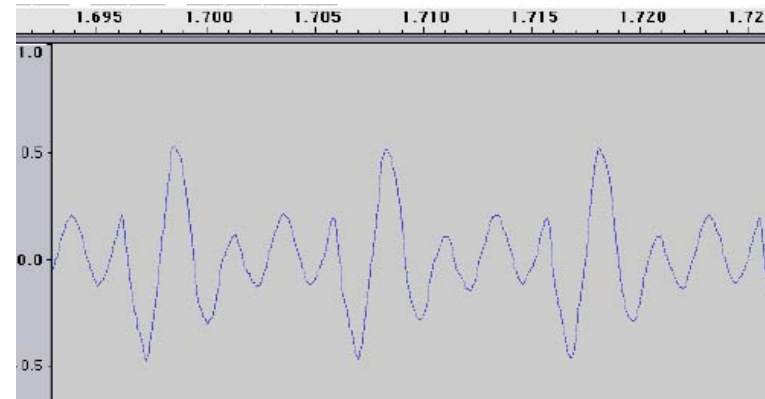
mes voyelles



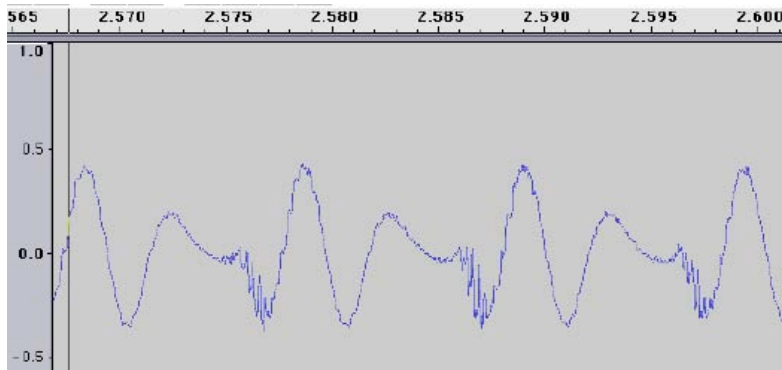
aaa



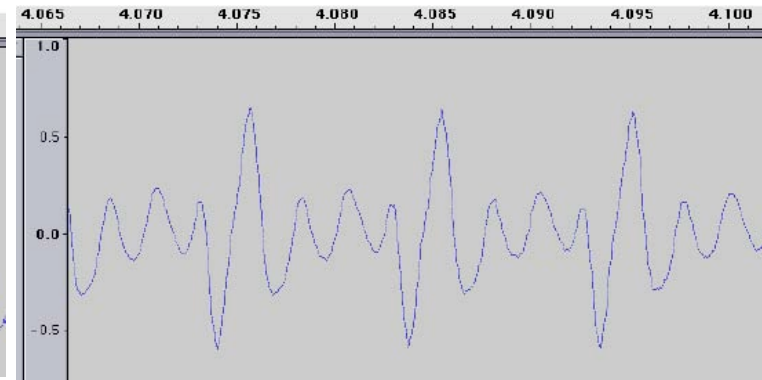
eee



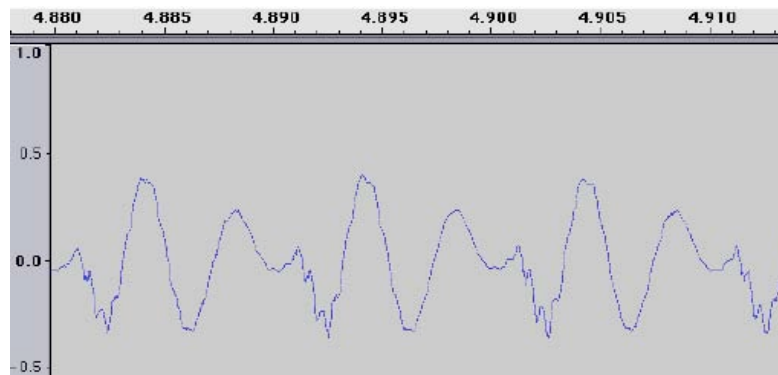
iii



ooo



uuu



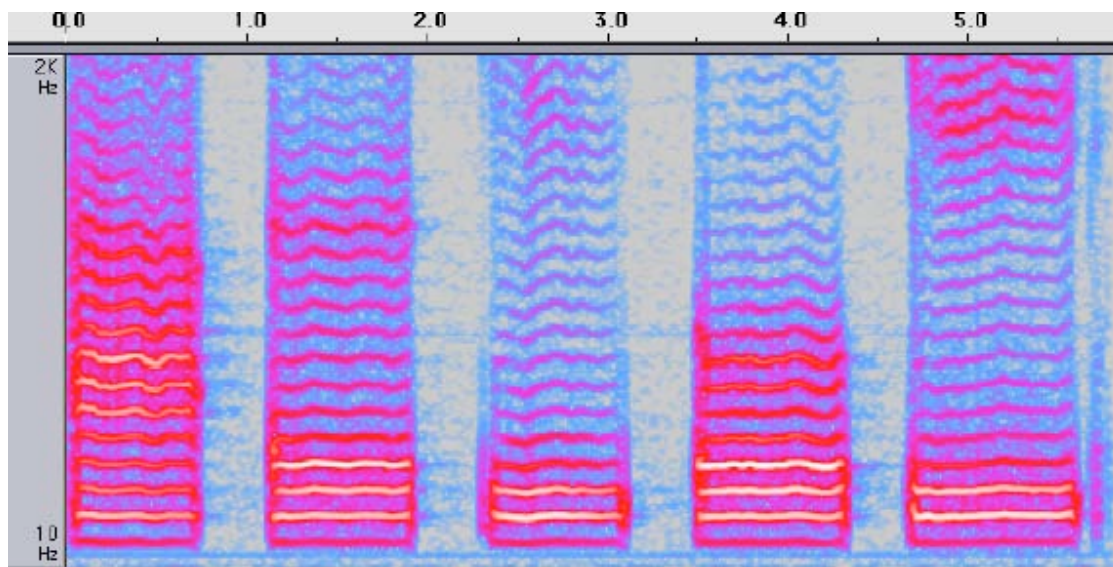
un "iii" synthétique :



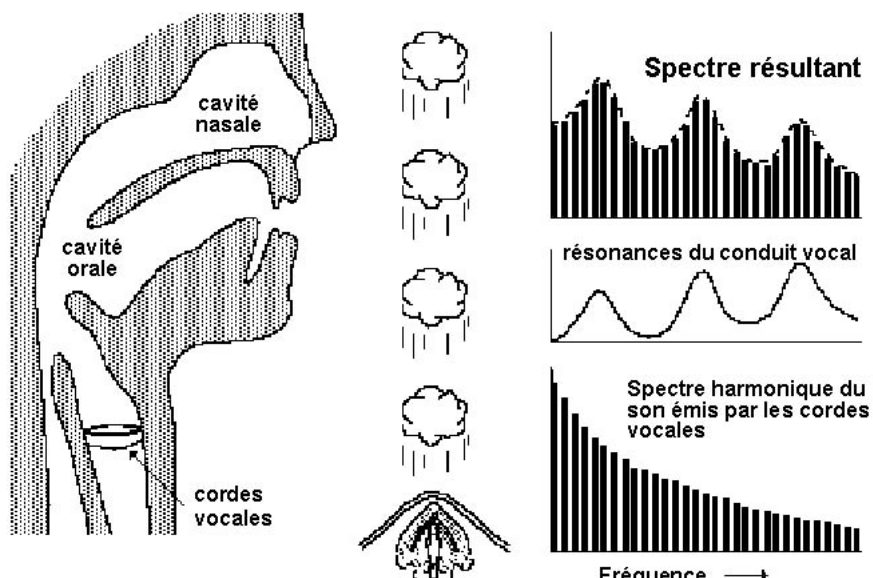
un "uuu" : "iii" -30%



sonagrammes a-e-i-o-u : les formants



ex. de formants :



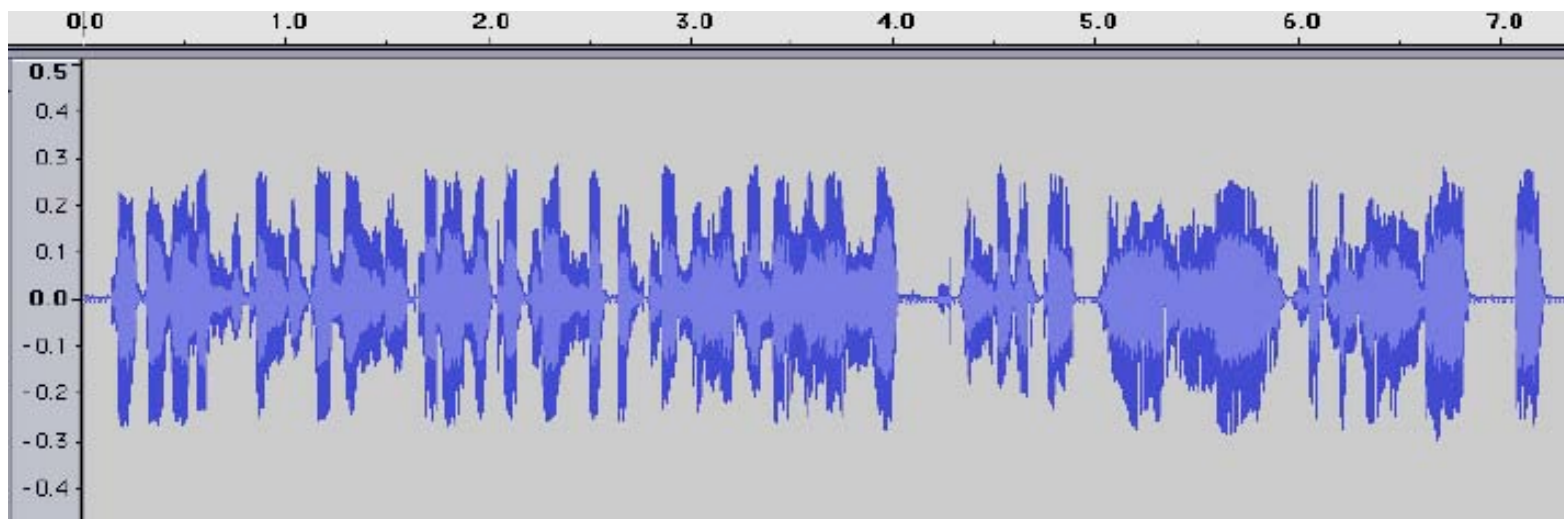
(Hz)	F1	F2	F3
aa	690	1660	2490
oo	640	1190	2390
ii	400	1900	2550

<http://cslu.cse.ogi.edu/tutordemos/SpectrogramReading/ipa/ipahome.html>

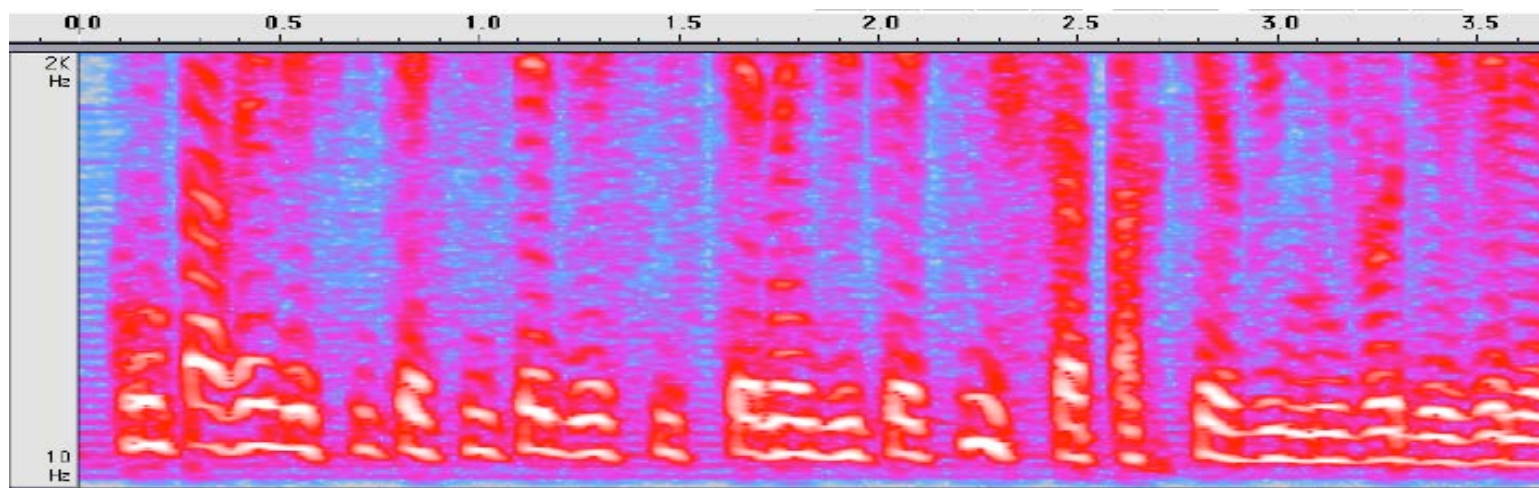
Une phrase



extrait France-Info 6/12/05 10h



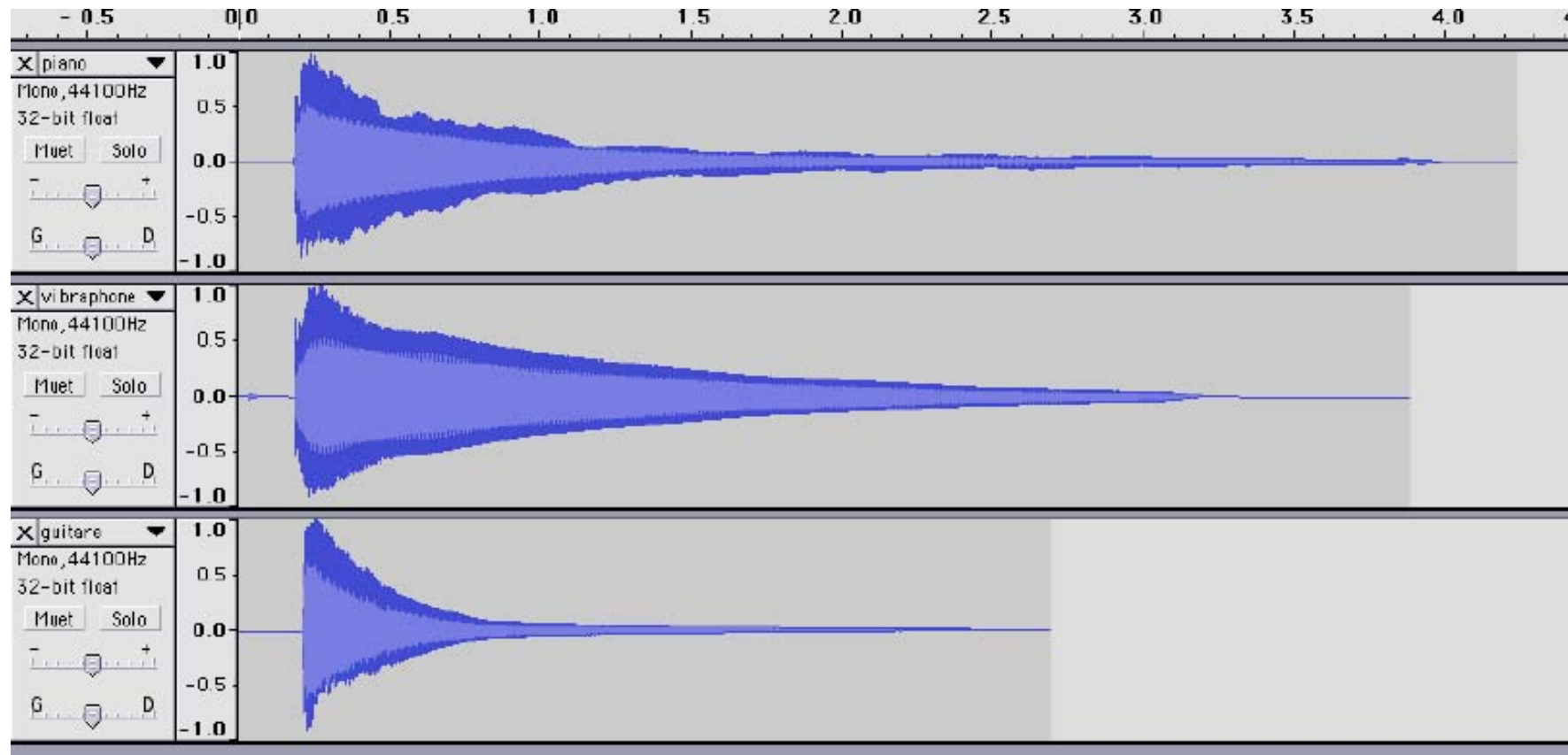
beaucoup de silence... et du bruit (plosives)



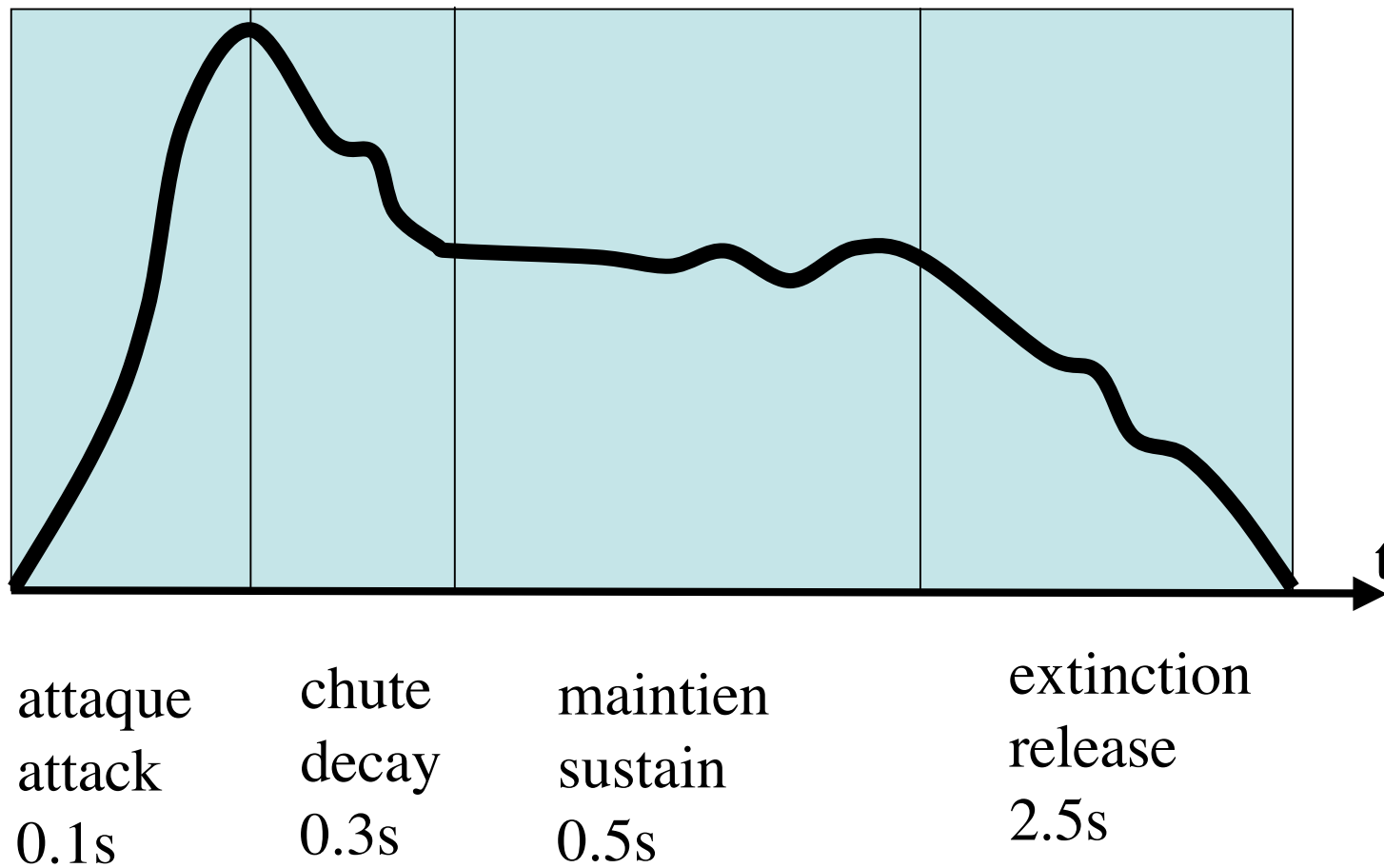
Instruments de musique

 piano (La 440 Hz)  vibraphone  guitare

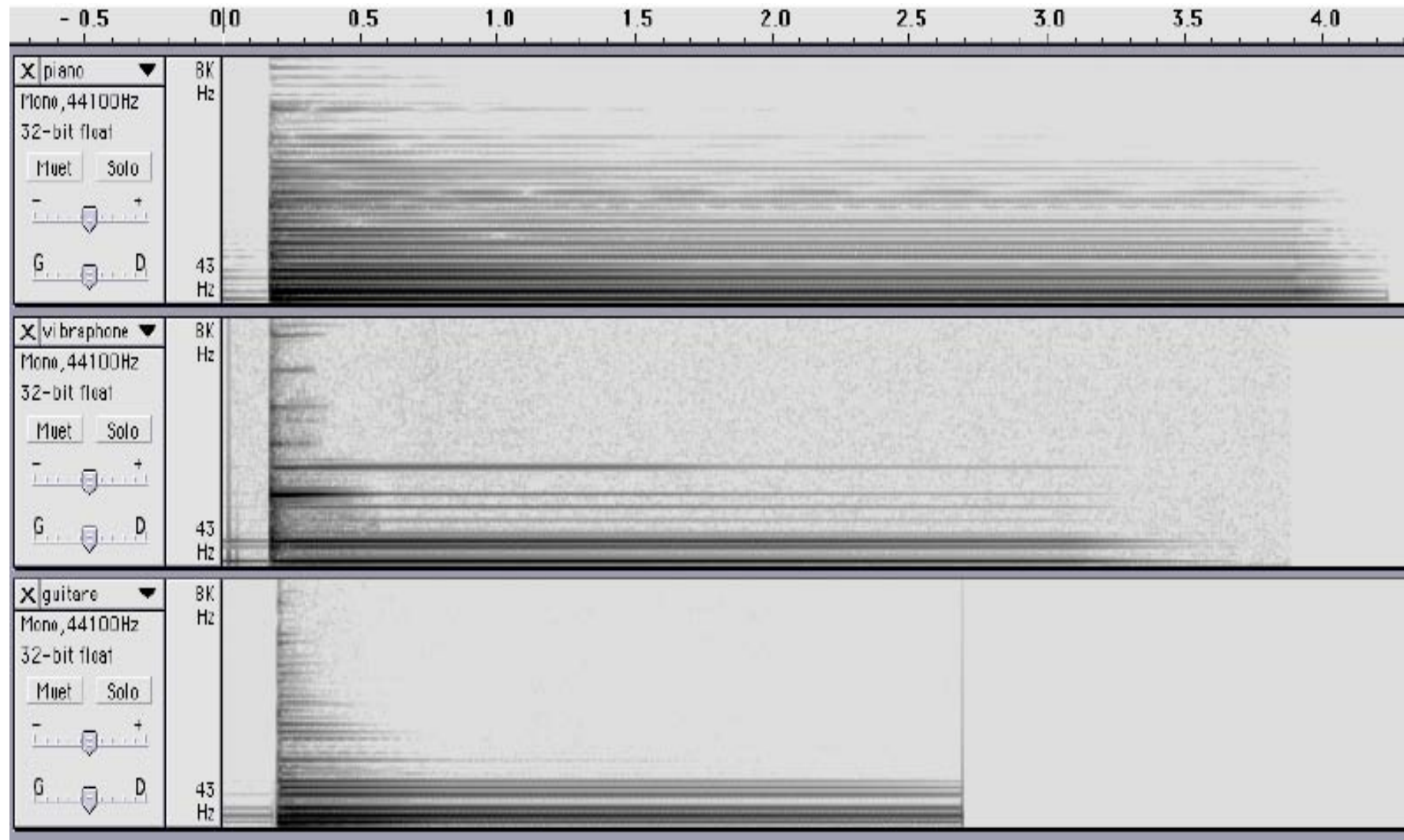
Des **enveloppes** différentes :



Modèle ADSR de l'enveloppe :



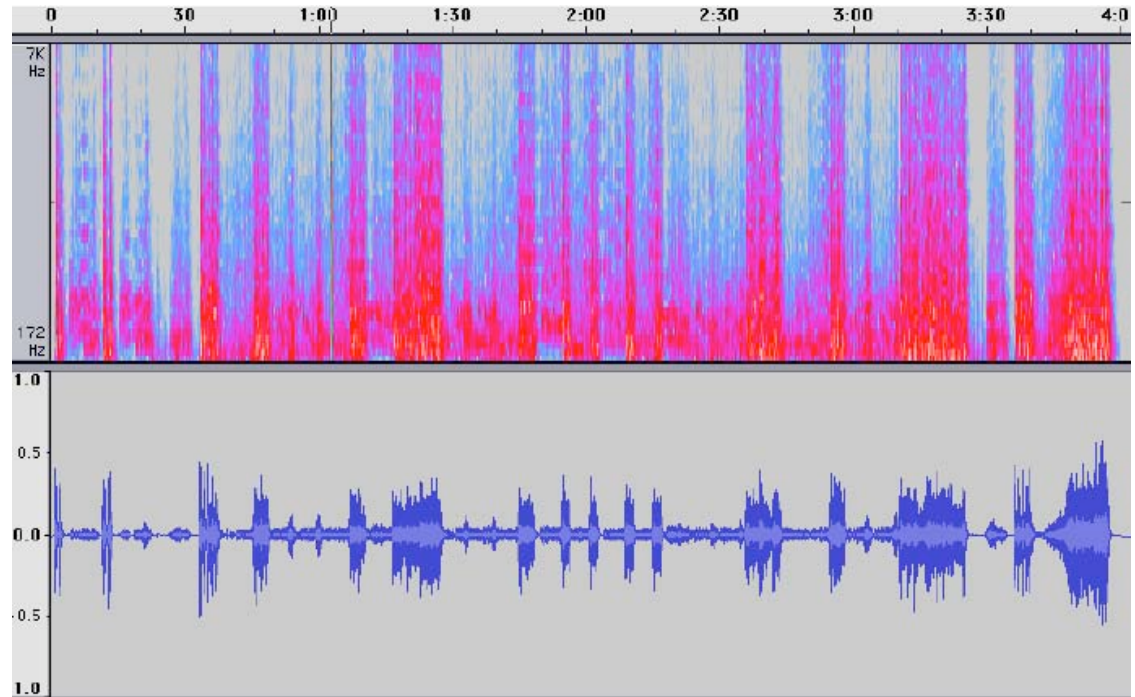
Sonagrammes : ($\neq$ du timbre de l'instrument)



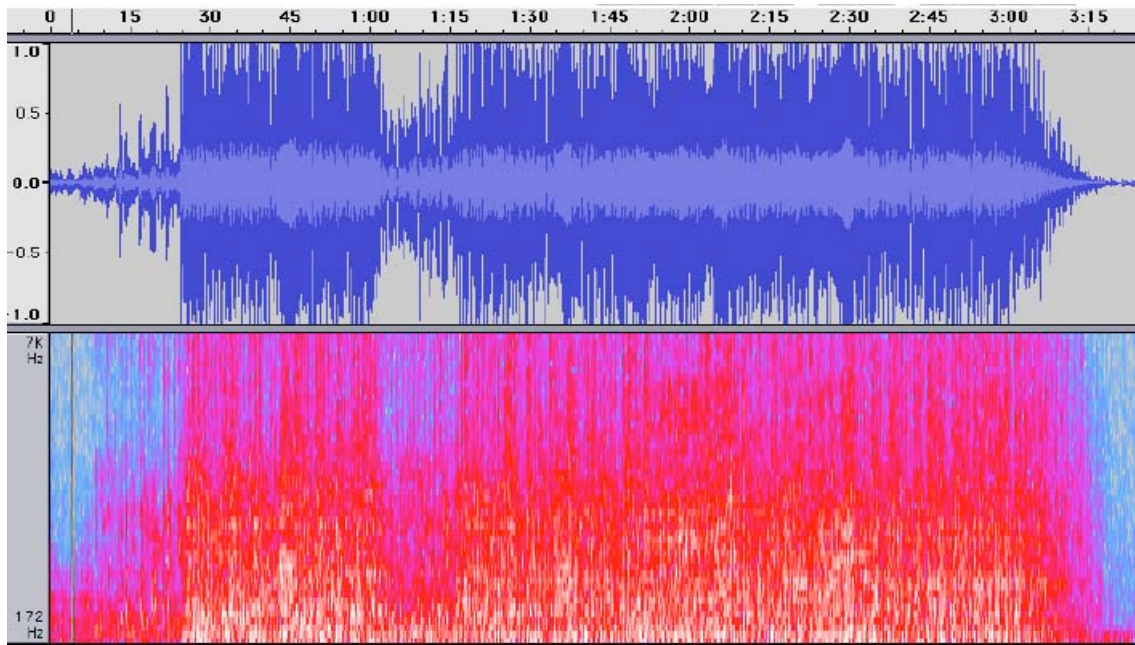
Musique(s)



Mozart
Cosi fan tutte (début)
(CD audio DDD)



P. Smith
Because the night
(33 tours)



Temps et calcul

-temps réel $\neq$ temps différé

Son 20000 Hz $\Rightarrow$ 40000 ech/s

$\Rightarrow$ 1/40000 s pour traiter l'échantillon

Sur machine 1 Gflops $\Rightarrow 10^9/(4 \cdot 10^5)$ flop/ech = 2500 flop/ech

-Traitement au fil de l'eau $\neq$ traitement a posteriori

Son enregistré pendant une heure : $3600 \cdot 40000 = 144$ Mech

Si un échantillon est un réel 32 bits : 576 Moctets

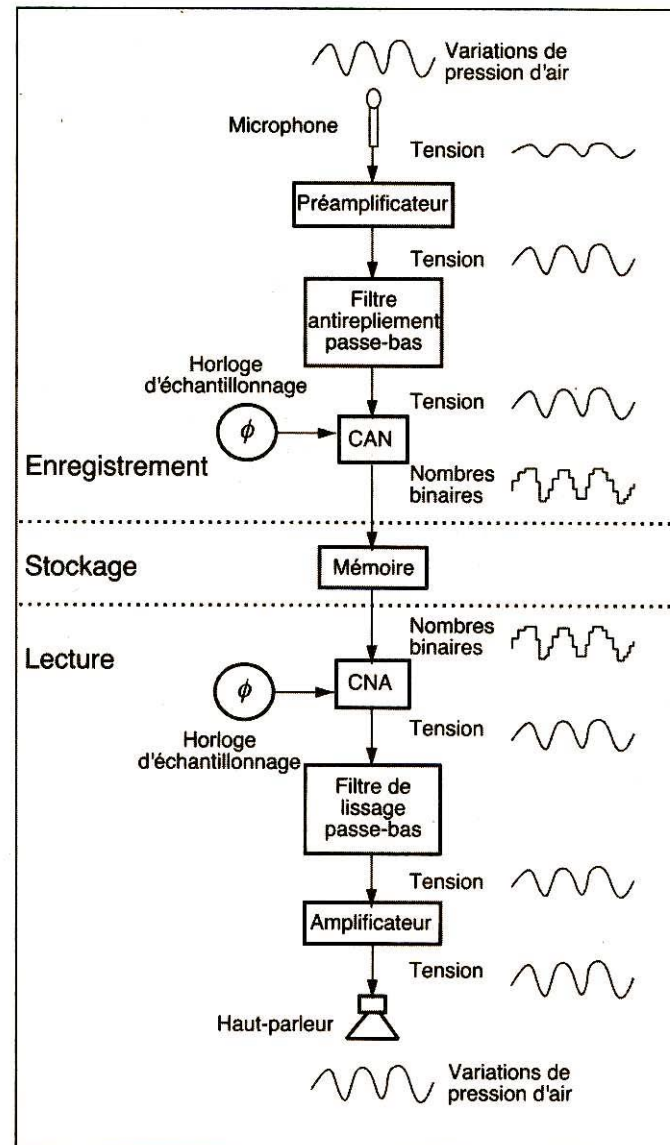
$\Rightarrow$ Les deux aspects sont à la portée des machines bureautiques
en 2005

2. Acquisition, traitements

- numérisation
- analyse harmonique
- ex. domaine temporel : modulation, retard
- ex. domaine fréquentiel : les filtres linéaires

Numérisation

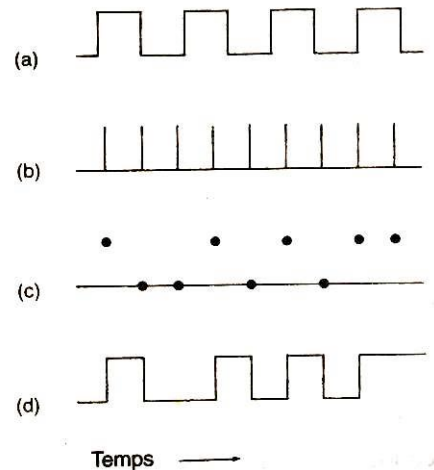
[Roads, p. 23]



Illust. 1.13 : Vue globale de l'enregistrement et de la lecture numérique.

Fréquence d'échantillonnage (bis)

[Roads, p. 28]



Nyquist : $f_e \geq 2 \text{ BP}$

Si BP oreille = 22 KHz : travailler à 44100 échantillons/s

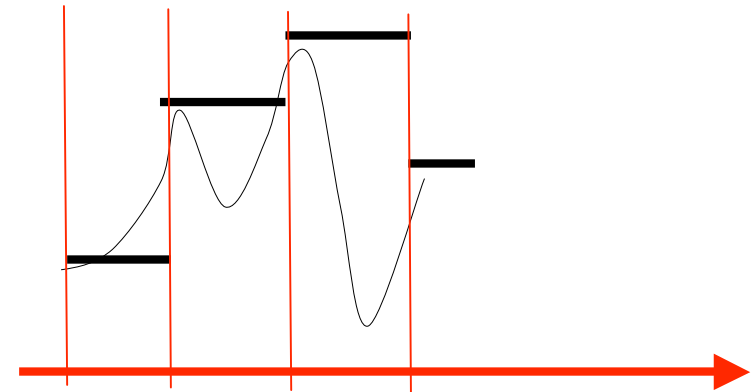
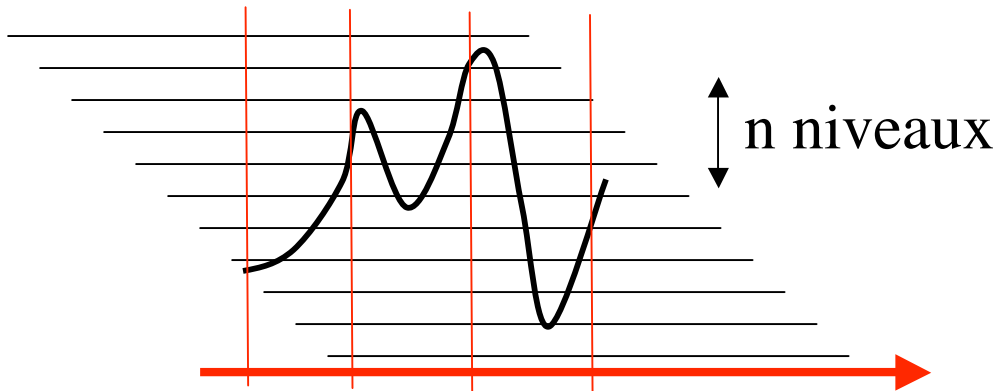
Plus prudent : 48000 ech/s

Cartes acquisition : 96000 ech/s

Audio (très) professionnelle : 192000 ech/s

En téléphonie (fixe) : 8000 ech/s

Quantification (re-bis)



en pratique, $n = 2^b$

- cartes son pro : 24 ou 32 bits
- CD audio : 16 bits
- téléphonie : 8 bits
- PC WAV mono : 8 bits

Dynamique correspondante ?

$$2^{16} = 10^n \text{ donc}$$

$$n = \text{Log}(2^{16}) = 4.82 \Rightarrow 48 \text{ dB}$$

24 bits : 72 dB

32 bits : 96 dB

8 bits : 24 dB

$$\text{en fait : dB} = 6,11 \times \text{nb bits}$$

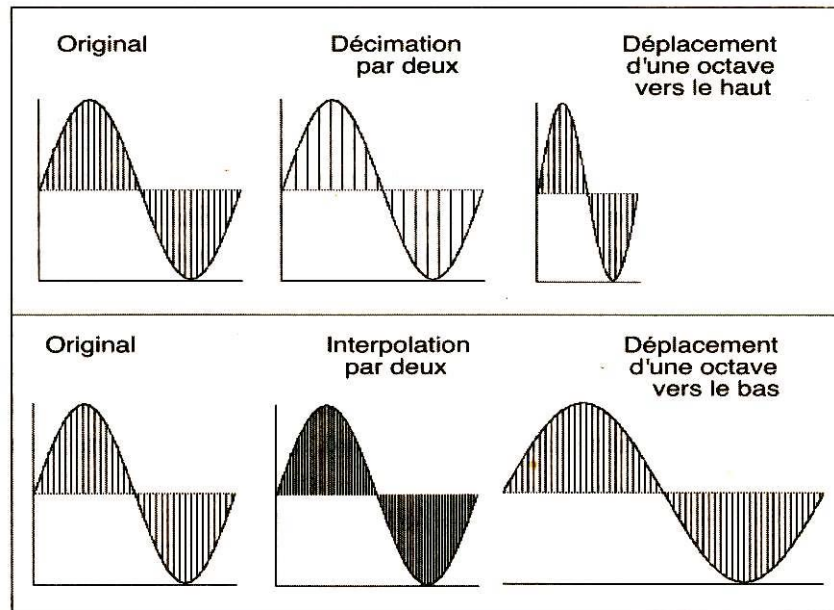
problème du **bruit de quantification** si $q < 18$ bits : effet de marche d'escalier des échantillons dans les décroissances lentes

Traitements dans le domaine temporel

- Normalisation : cf "point noir /blanc" en traitement d'image
- changer la vitesse => ré-échantillonner

Così fan tutte : $33t \Rightarrow 78t$ ($\times 1,34$) 

Because the night : $78t \Rightarrow 33t$ ($\div 1,34$) 

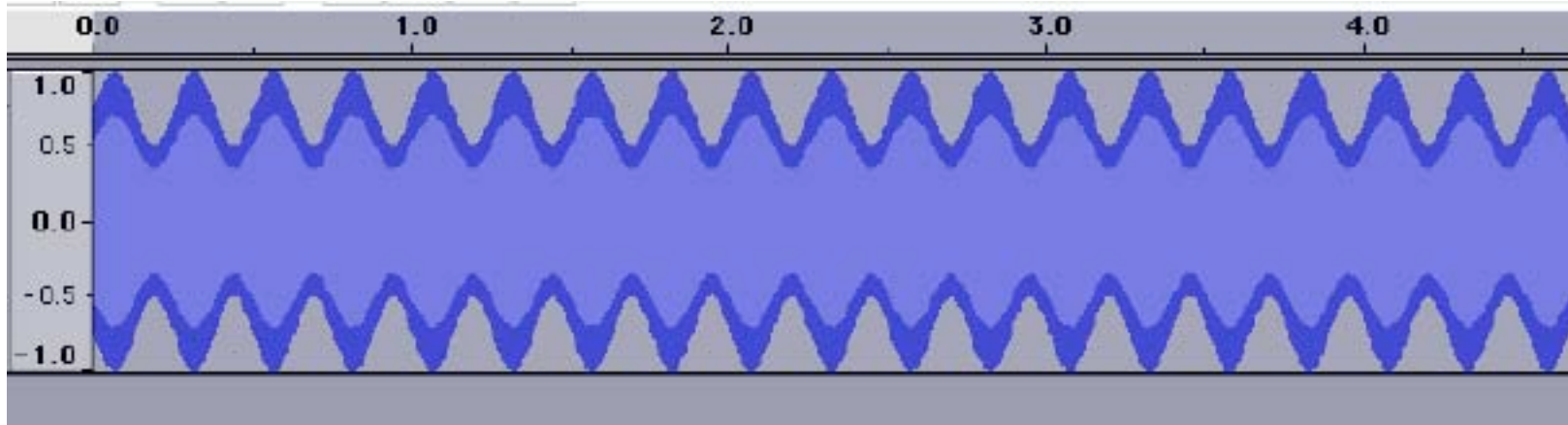


[Roads, p.93]

Modulation de l'amplitude

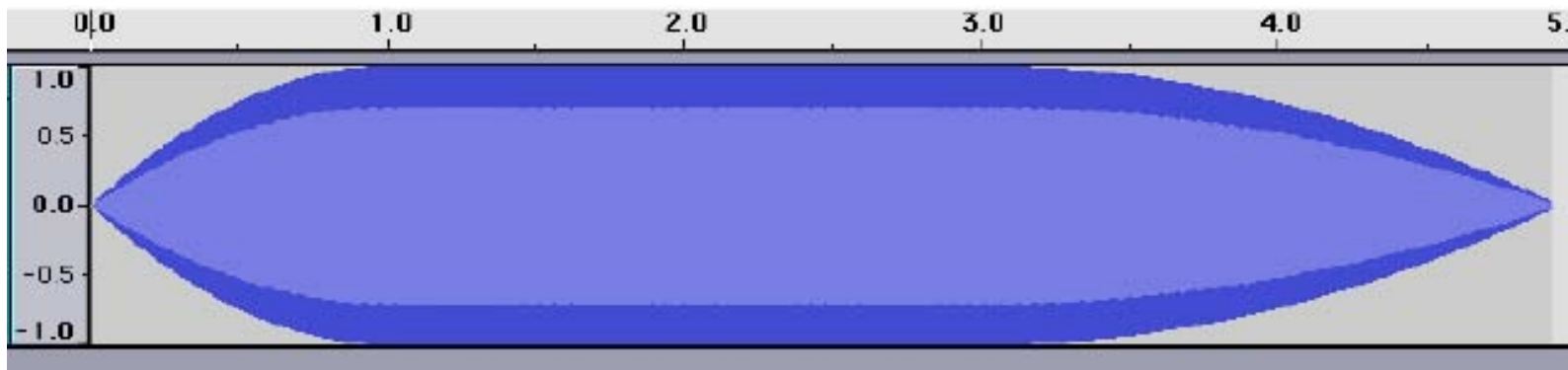
$$y_n = f_m(t_n) \times x_n$$

- Effet "tremolo" $f_m = \text{sinusoïde}$



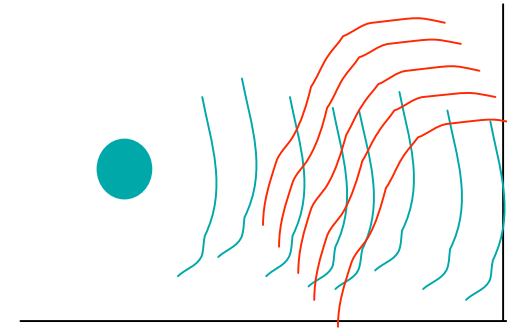
si $\text{freq} > 18 \text{ Hz}$: la modulation devient audible. 

- Fade-in et fade-out 



Les effets de retard

$$y_n = a(x_n + \beta x_{n-K})$$



à 20° : 1 ms de retard $\Leftrightarrow$ 30 cm de distance

- retard moyen de 10 à 50 ms : effet de "présence" utile pour rehausser un signal "mince"



a-e-i-o-u



retard 50 ms - décroissance 50 %

- au-delà : c'est l'écho

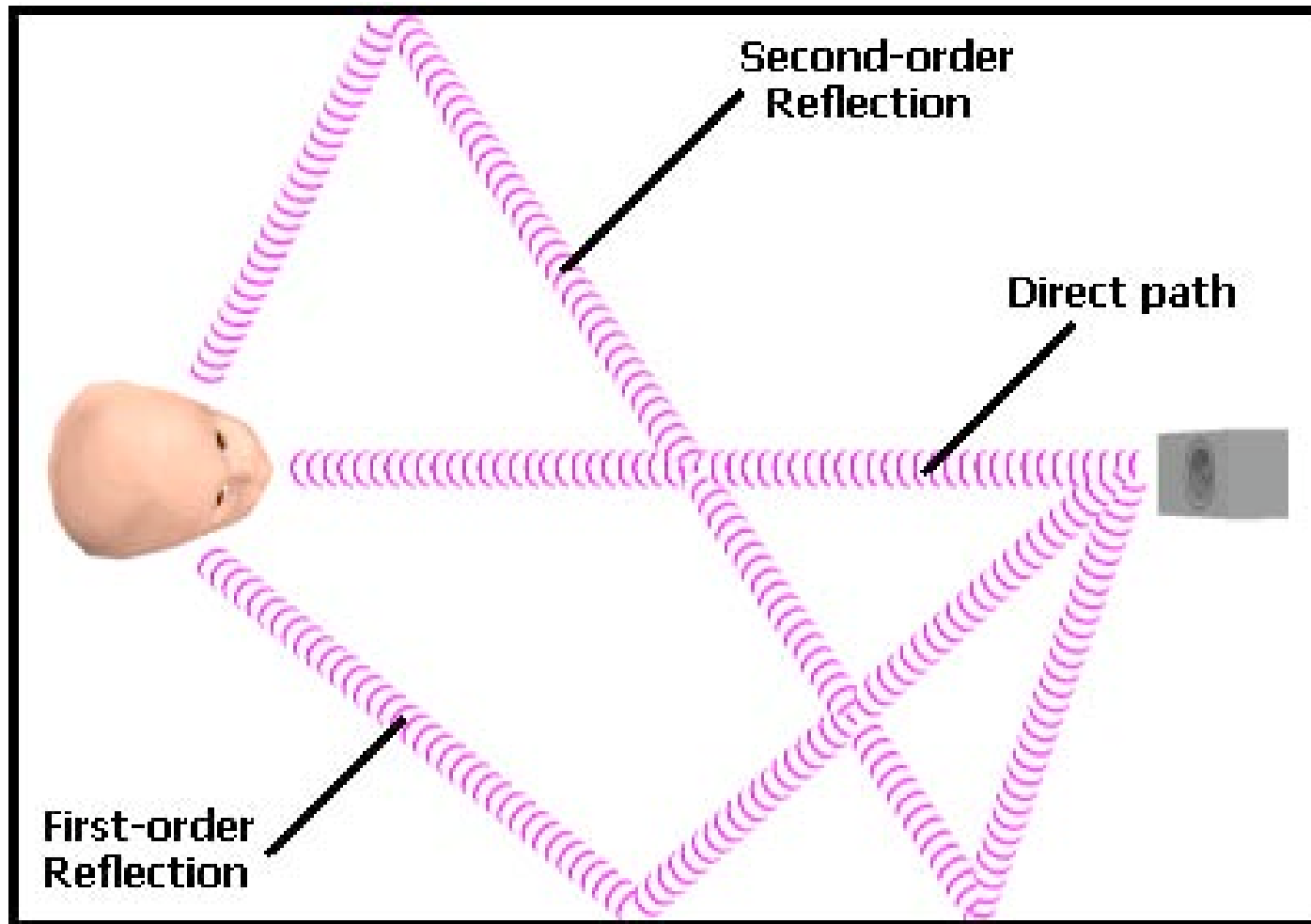


retard 500 ms - décroissance 50 %



retard 800 ms - décroissance 90 %

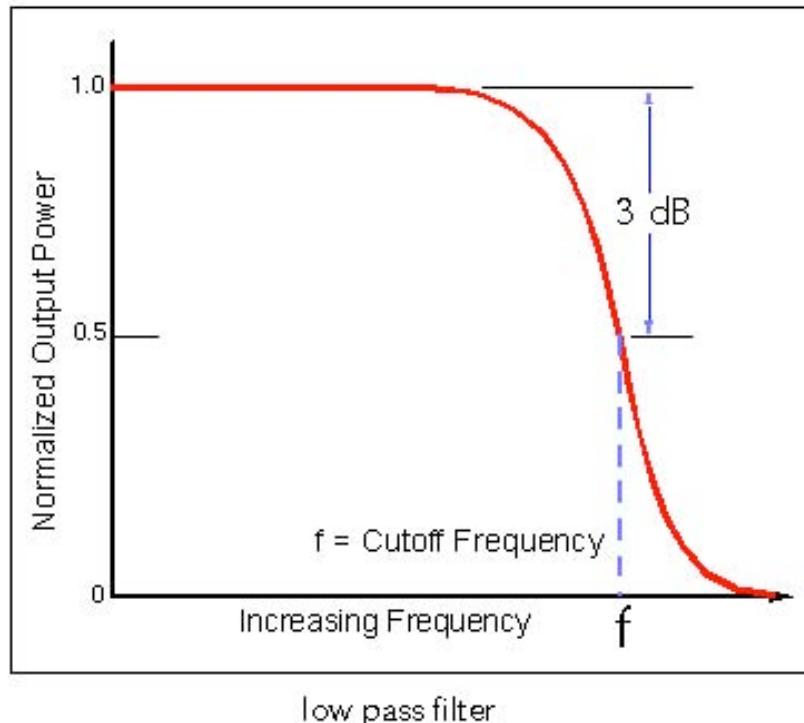
Application à la spatialisation sonore



Traitements dans le domaine fréquentiel : les filtres

(rque : modif domaine frequentiel $\Rightarrow$ modif domaine temporel)
 $\Leftarrow$

Filtre passe-bas (low pass) :



freq. de coupure et pente

Filtre passe-haut (high pass) :

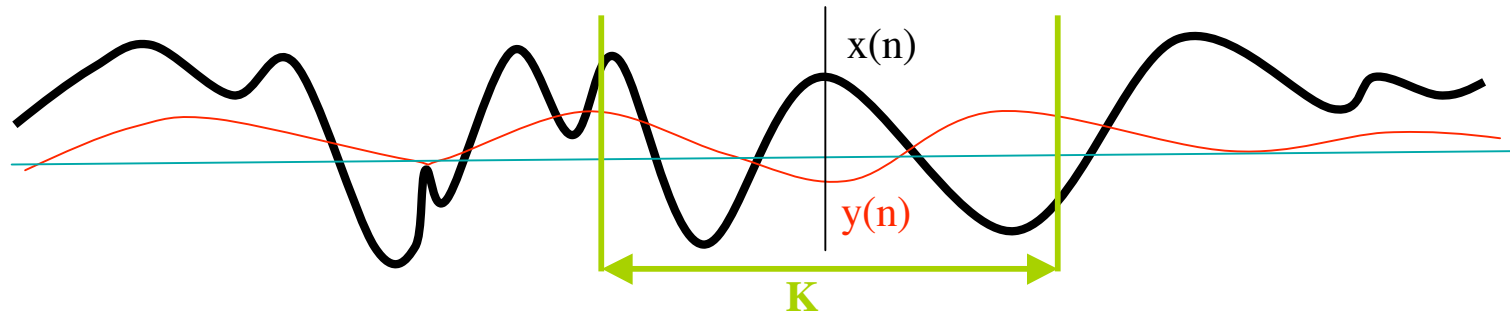
Le signal d'origine auquel
on ôte un bout par filtre
Passe-bas

Filtre passe-bande (band pass) :

passe-haut(F_{min}) suivi
d'un passe-bas(F_{max}) ou
l'inverse

Filtrage par moyenne mobile

- Si on connaît toute la séquence d'échantillons, on peut lisser :



le nombre de points K pris en compte influe sur la fréquence de coupure (K aug. $\Rightarrow$ F_c diminue)

- Si on ne connaît que le passé du signal : 2 méthodes

- filtre **FIR** (finite impulse response)

$$y(n) = b_0 \cdot x(n) + b_1 \cdot x(n-1) + b_2 \cdot x(n-2) + \dots + b_K \cdot x(n-K)$$

pas de déphasage, pas de cumul d'erreur

- filtre **IIR** (infinite impulse response)

$$y(n) = x(n) - a_1 \cdot y(n-1) - a_2 \cdot y(n-2) - \dots - a_K \cdot y(n-K)$$

plus général, moins de mémoire, pas toujours stable

Analyse harmonique par la transformée de Fourier discrète

- cas d'un signal périodique

Si on a $2m$ échantillons :

$$A_N = \frac{1}{2m} \sum_{i=0}^{2m-1} y_i \cos \frac{Ni}{m} \quad N = 0 \dots m$$

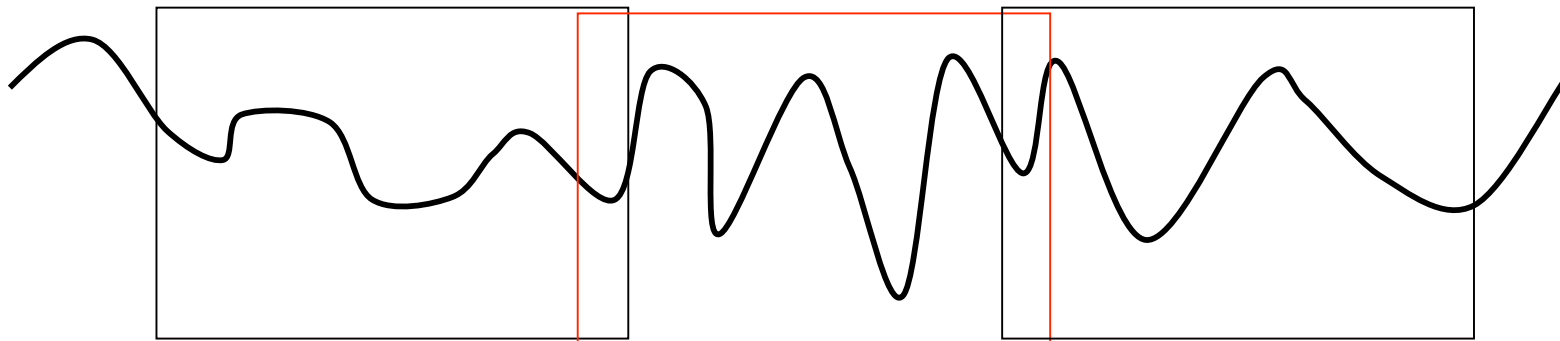
$$B_N = \frac{1}{2m} \sum_{i=0}^{2m-1} y_i \sin \frac{Ni}{m} \quad N = 1 \dots m-1$$

$4m^2$ opérations : trop dur pour un humain !!

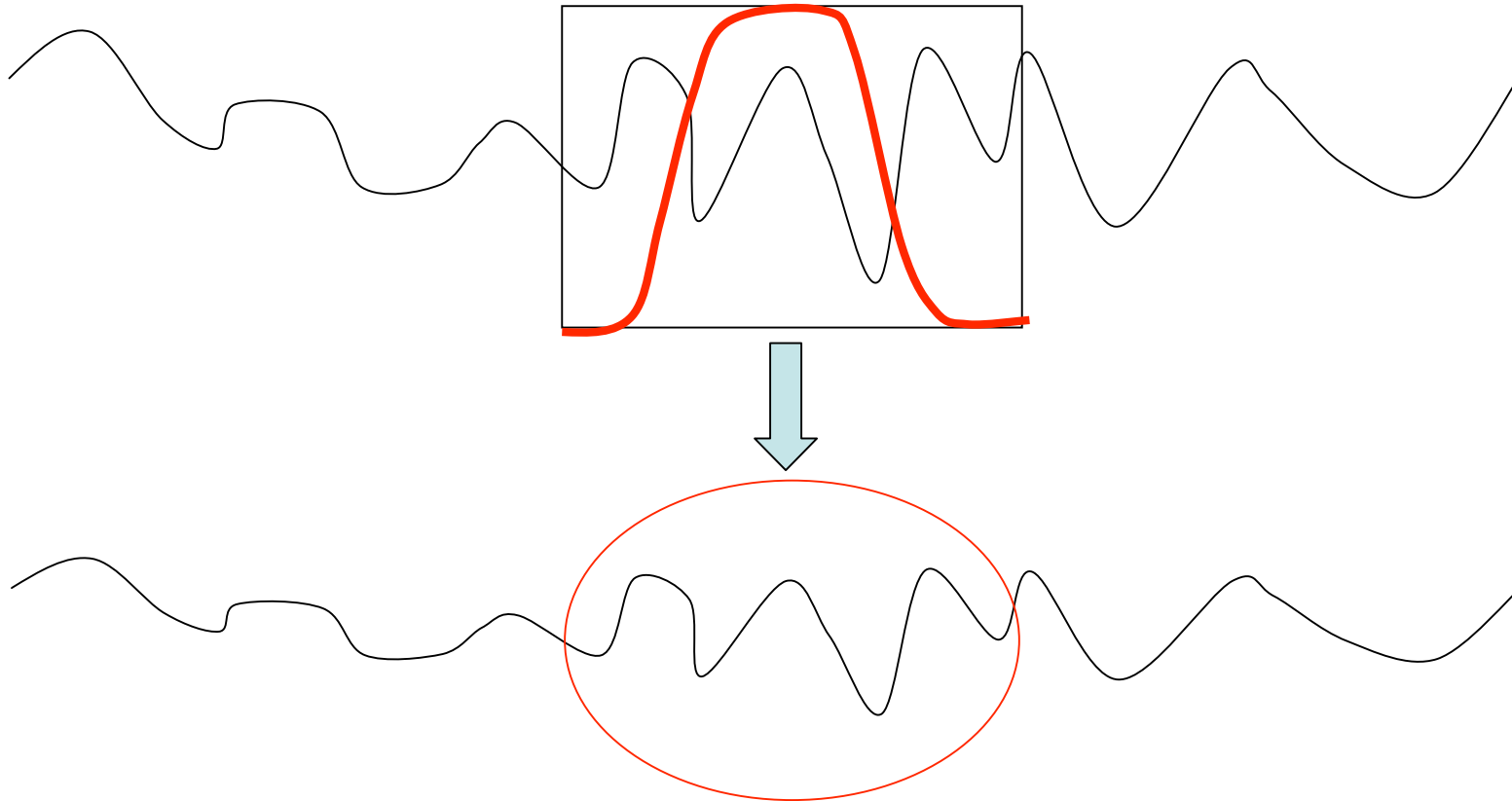
et longtemps aussi pour les ordinateurs en temps réel

=> calcul par récurrence (FFT, $\log_2(2m)$ ops.)

- cas d'un signal quelconque : utilisation d'une fenêtre glissante



Traitement du signal dans la fenêtre : fonction de **lobe**



Nombreuses variantes : Gaussienne, Hanning, Hamer, Kaiser...

Choix de la taille de fenêtre :

Avec la TFD, les analyseurs sont répartis linéairement sur le spectre : pas précis pour les BF et trop pour les HF

-valeurs usuelles : 64, 128, 256, 512, 1024, 2048, 4096

Avec 256 échantillons et $f_e = 44100$ Hz
la fenêtre représente $256/44100 = 6$ ms de son
pour $44100/256 = 172$ Hz par analyseur
 $\Rightarrow$ bonne résolution temporelle

*compromis
à faire a priori...*

Avec 1024 échantillons :
la fenêtre représente $1024/44100 = 23$ ms de son
pour $44100/1024 = 43$ Hz par analyseur
 $\Rightarrow$ bonne résolution fréquentielle

3. Synthèse sonore

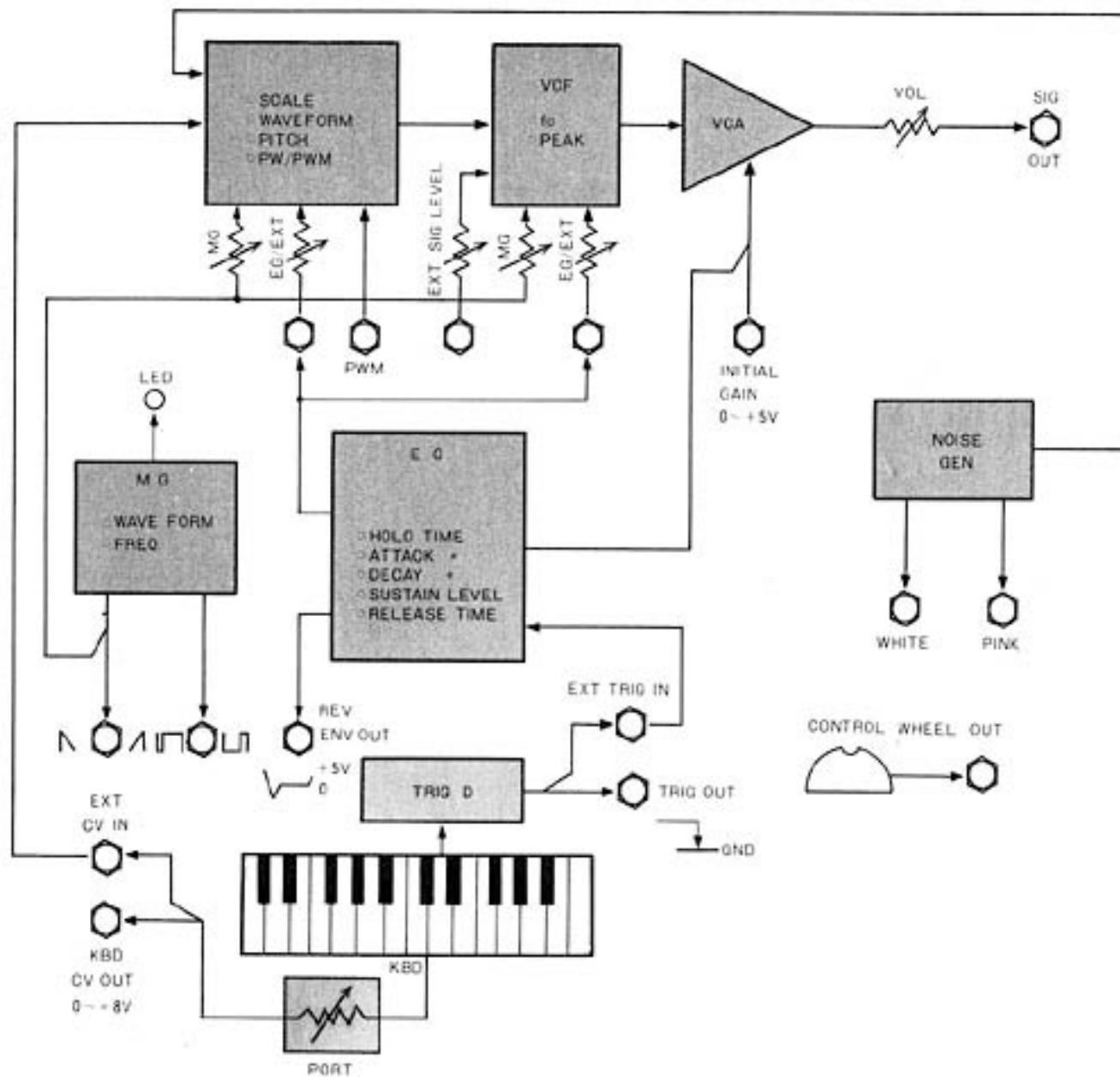
- de l'analogique au numérique
- modulation de signaux élémentaires
- synthèse additive
- synthèse physique
- du son à la musique

Mon Korg MS10 (circa 1981)



- un VCO
- un VCF
- un VCA
- un ADSR

- + un générateur de bruit
- + un LFO
- + une commande manuelle



Moog modular...



Mon Yamaha P120 (circa 2001)

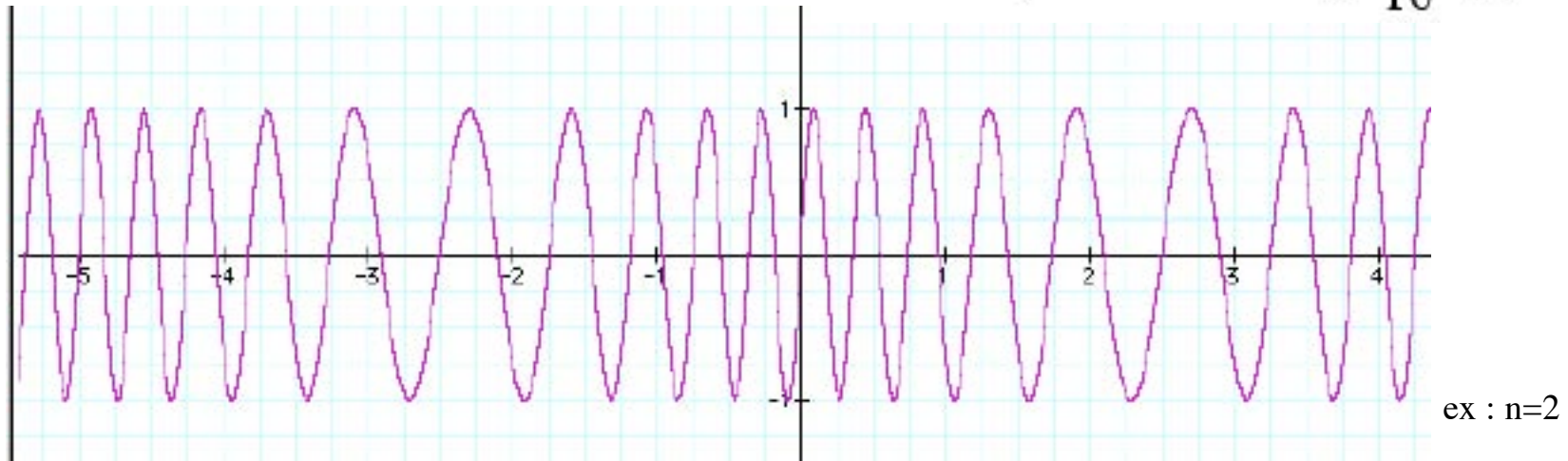


- synthèse par tables d'ondes (samples)
- polyphonique 64 notes
- 14 voix (piano, orgue, clavecin, "guitare" ...)
- reverb, chorus, phaser, tremolo, delay
- MIDI + interface série
- clavier sensitif

Modulation de fréquence (FM, Chowning, 1967)

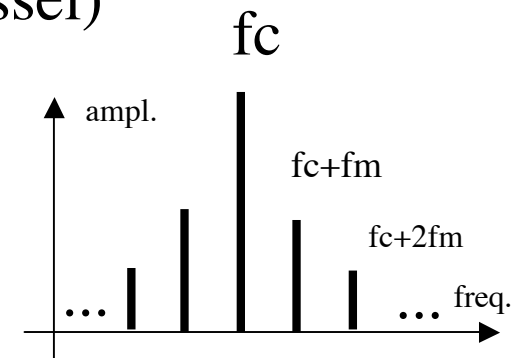
- une fréquence porteuse (f_c)
- un signal modulant (f_m)

$$y = \sin \left(2\pi n x + 4 \sin \left(\frac{2\pi n x}{10} \right) \right)$$



⇒ Un spectre très riche (harmonique = fct de Bessel)

⇒ facile à fabriquer

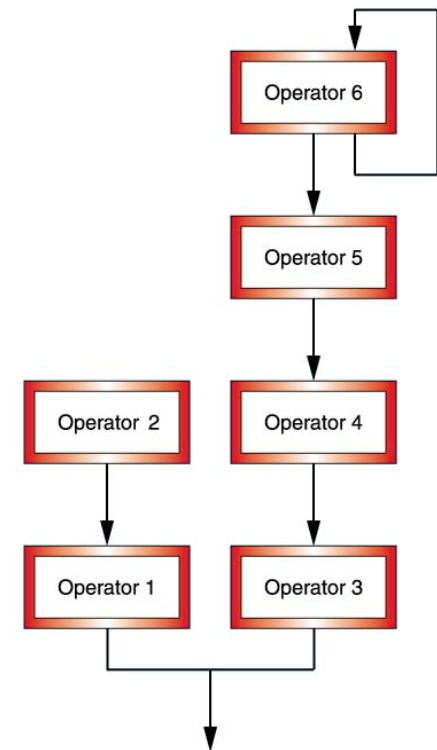


Ex. du synthétiseur Yamaha DX7 (1983)

- 6 oscillateurs controlables en freq. et amplitude ("opérateurs")
- un générateur d'enveloppe associé à chaque opérateur pour varier l'amplitude (si modulateur) ou volume (si porteur)
- plusieurs schémas d'interconnexion ("algorithmes")



Plein de sons sur <http://www.synthmania.com/dx7.htm>



<http://www.soundonsound.com/sos/may00/articles/synth.htm>

Music III (Max Mathews et al., 1960) et dérivés

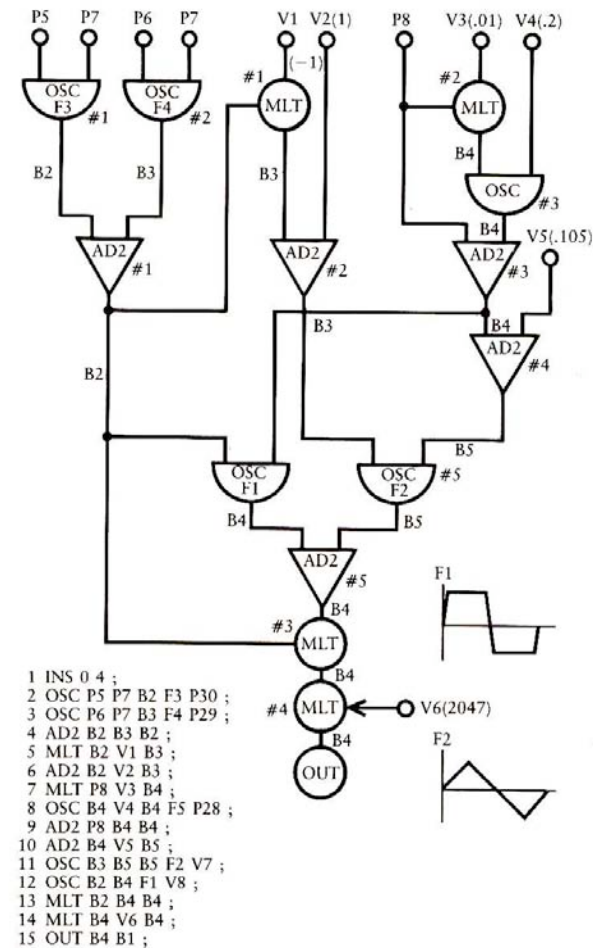
OSC : oscillateur (ampl, freq)

AD2 : additionneur

MLT : multiplicateur

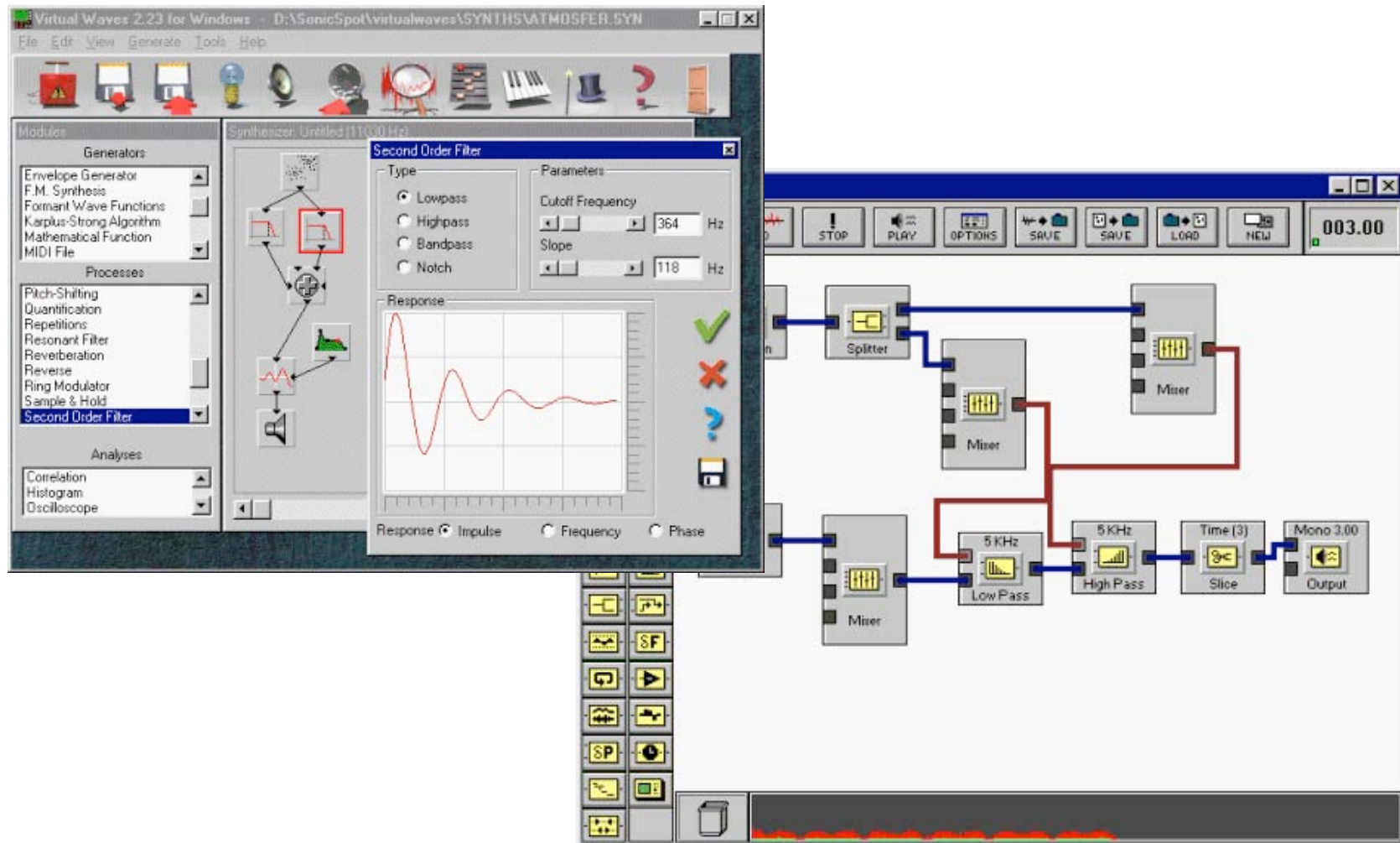
OUT : module de sortie des échantillons

INS : définition d'instrument



[Pierce p.214]

Virtual Waves



AudioFX

A étudier (1) : CSound

The image displays the 'scrubber' application interface, which is divided into three main sections: a code editor, an output window, and a control panel.

Code Editor (Left): The code is written in a C-like syntax and includes comments in French. It defines variables for pitch, gain, and scrub, and implements a function to estimate the new position in a file based on these parameters.

```

; disk in scrubber
;
; -na++ may 2003

nchnls = 2

instr 1
  kpitch   invalue "pitch"
  kgain     invalue "gain"
  kscrub    invalue "scrub"

  kpos      init i(kscrub)

  if kscrub != kpos then
    kpos = kscrub
    reinit reset
  endif

  ifn filename "vine.aiff" ; substitute a long file!
  outvalue "filename", filename

  reset:
  ar1, ar2 diskin "vine.aiff", kpitch, i(kscrub)*ilen, 1
  freturn

; estimate new position in file
kpos = kpos + kpitch/(ilen * kn)
if (kpitch < 0 && kpos < 0) then
  kpos = 1
elseif (kpitch > 0 && kpos > 1) then
  kpos = 0
endif

outvalue "scrub", kpos
outvalue "pos secs", kpos*ilen

; handle gain
ar1 = ar1*kgain
ar2 = ar2*kgain

; display rms
kleft rms ar1
kright rms ar2

```

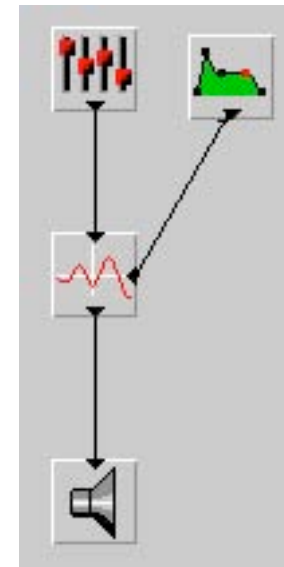
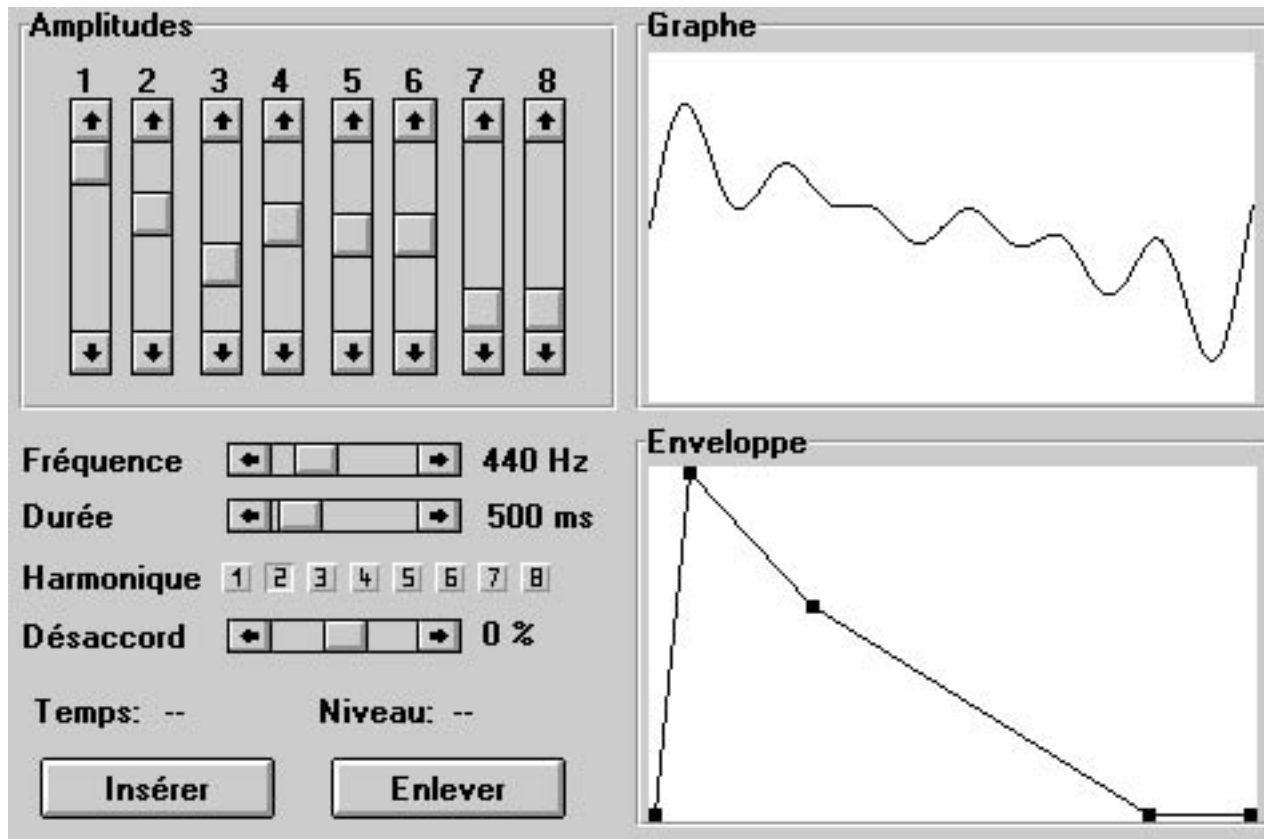
Output Window (Top Right): This window shows the execution log of the application. It includes messages about the initialization of the Csound library, the creation of the orchestra and score, and the loading of the audio file 'vine.aiff'. It also displays the file length (8.823 seconds) and the current position (7.853 seconds).

Control Panel (Bottom Right): This panel provides a graphical interface for controlling the scrubber. It features a 'Position' slider set to 7.853, a 'File Length' display showing 8.823 seconds, and two knobs for 'Am' (Amplitude) and 'Pit' (Pitch) with values 2.962 and -2.307 respectively. The panel is titled 'disk in scrubber by na++'.



resultat...

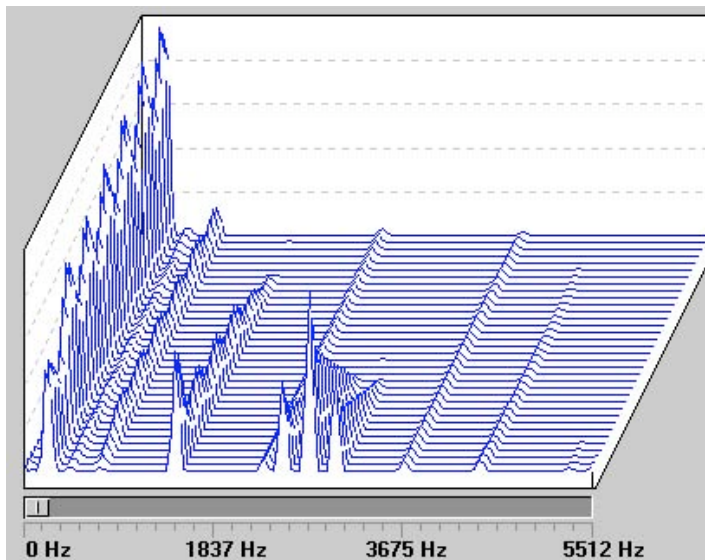
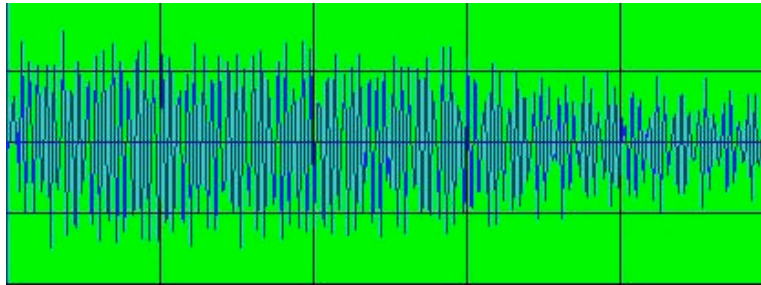
Synthèse additive : ex. avec VirtualWaves



Daniel BEAUFILS, Gérard SERRA © INRP - TECNE

http://www.inrp.fr/Acces/JIPSP/phymus/m_pedag/syntadd/ac_synt.htm

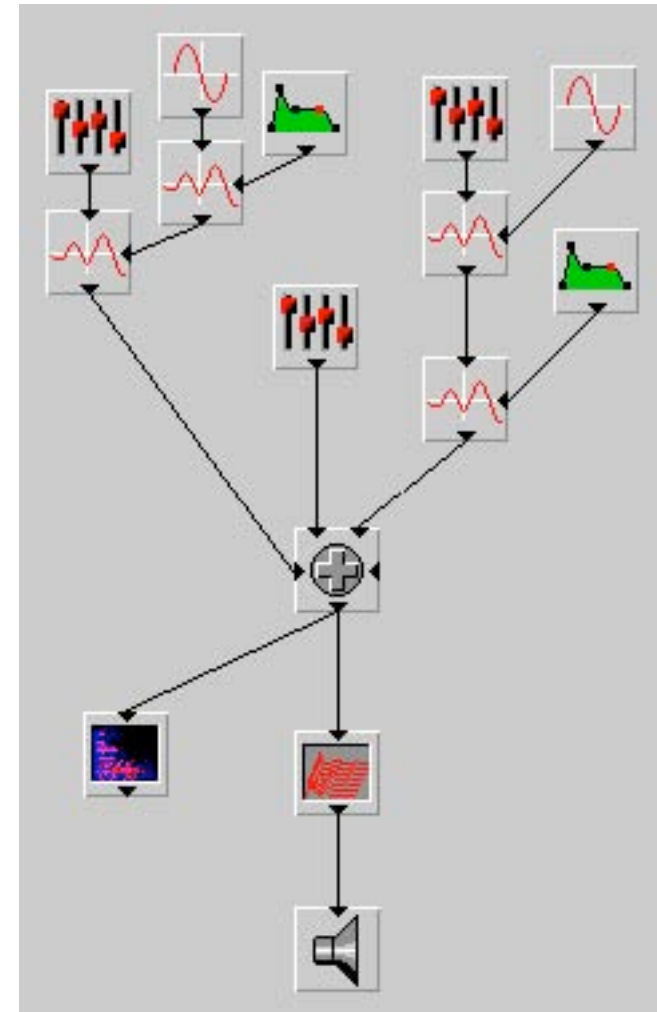
Analyse et re-synthèse additive :



Un couvercle métallique frappé légèrement

Daniel BEAUFILS © INRP - TECNE

http://www.inrp.fr/Acces/JIPSP/phymus/m_sons/etude4/ac_etud4.htm



3 synthèses mixées :

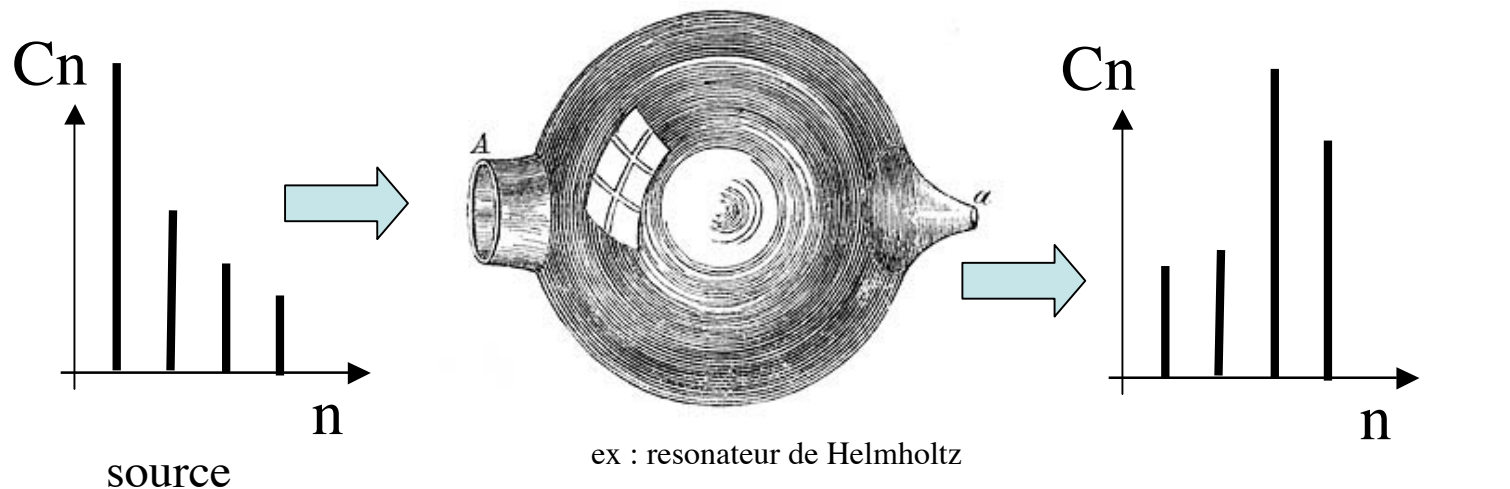
- vibrations BF
- choc en MF (2000 Hz)
- résonnances secondaires



Synthèse physique : des effets aux causes

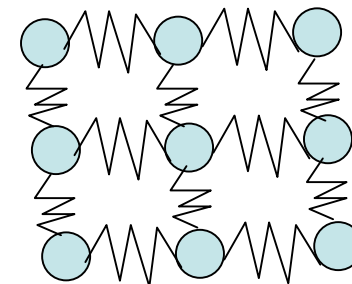
Modéliser les sources sonores et en déduire l'évolution de l'intensité (pression) des ondes sonores

Source sonore : un excitateur + un résonateur (+ interférences)
ex : corde pincée-relâchée + caisse = guitare, clavecin, piano...

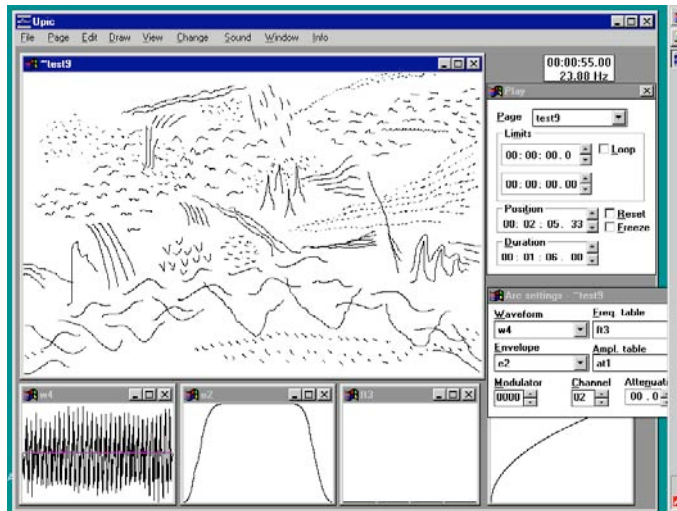


Modèle masse-ressort-amorti

=> logiciels CORDIS et MODALYS



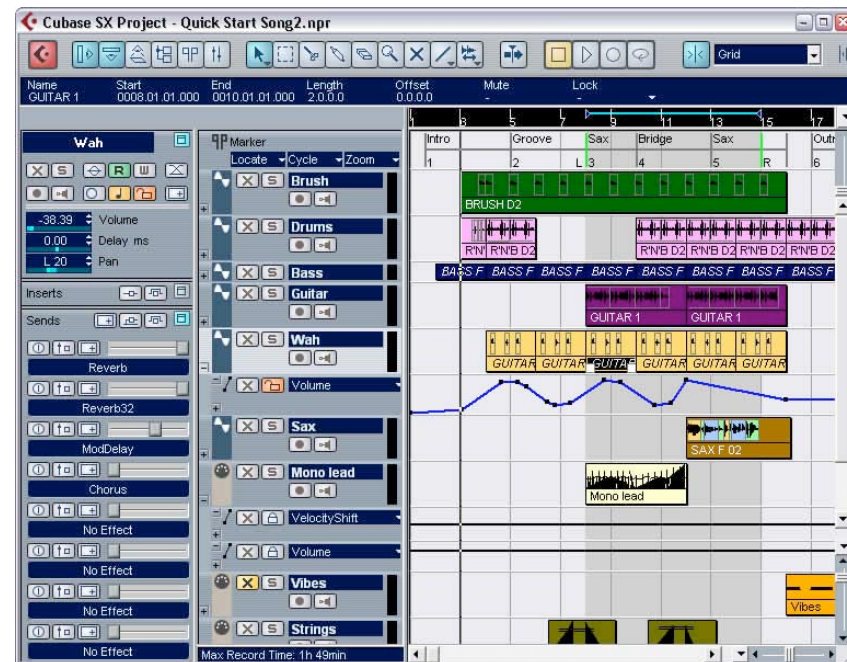
Du son à la musique



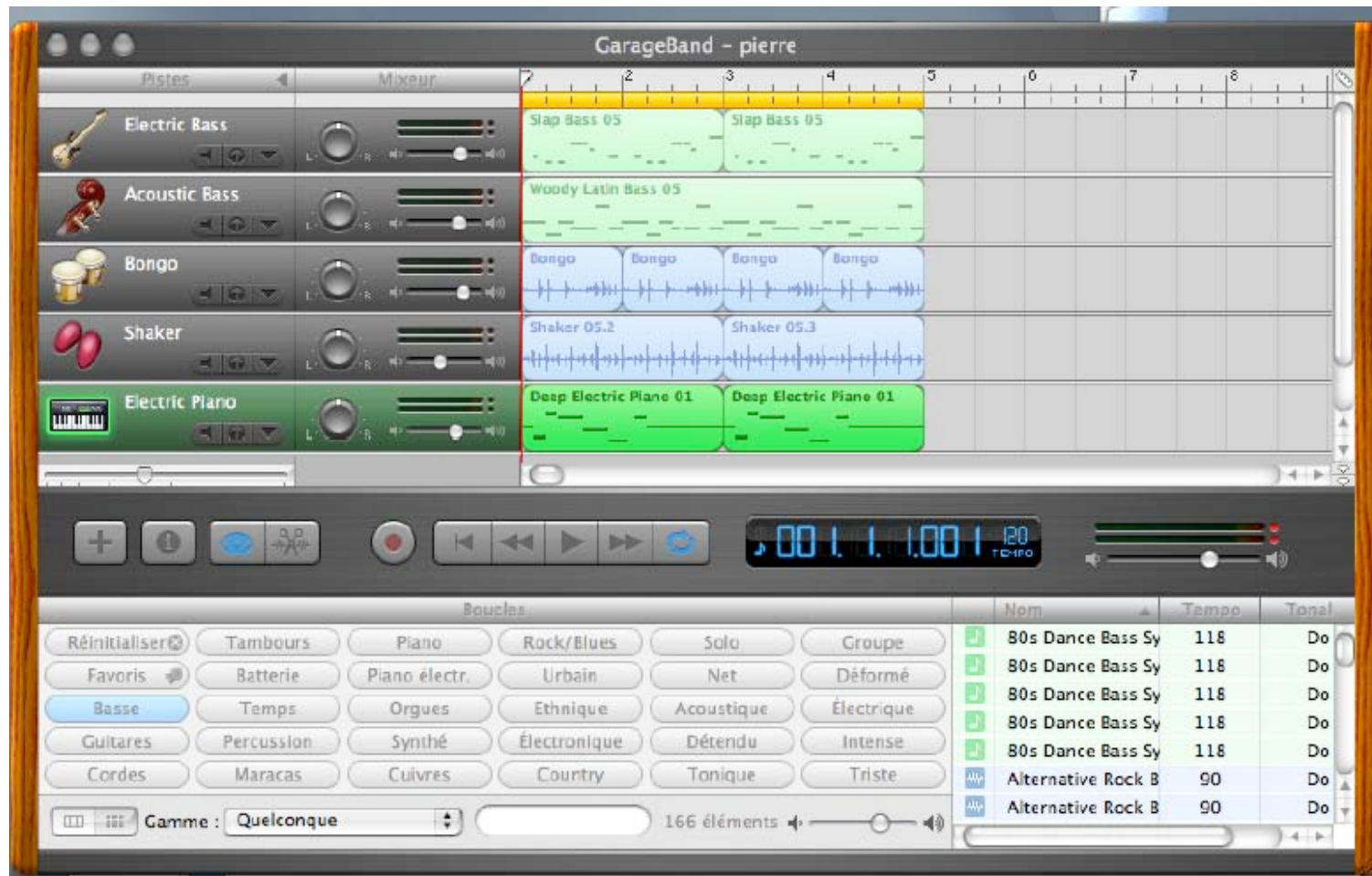
UPIC (Y. Xenakis et al. circa 1985)



CUBASE (SX 2002)



A étudier (2) : GarageBand



ma création

4. Transport, compression

- codages PCM, PDM
- flux synchrone, asynchrone
- compression avec perte : ex du MP3
- compression sans perte : ex. de FLAC
- formats de fichiers et conversion