

# Introduction aux techniques audio-numériques

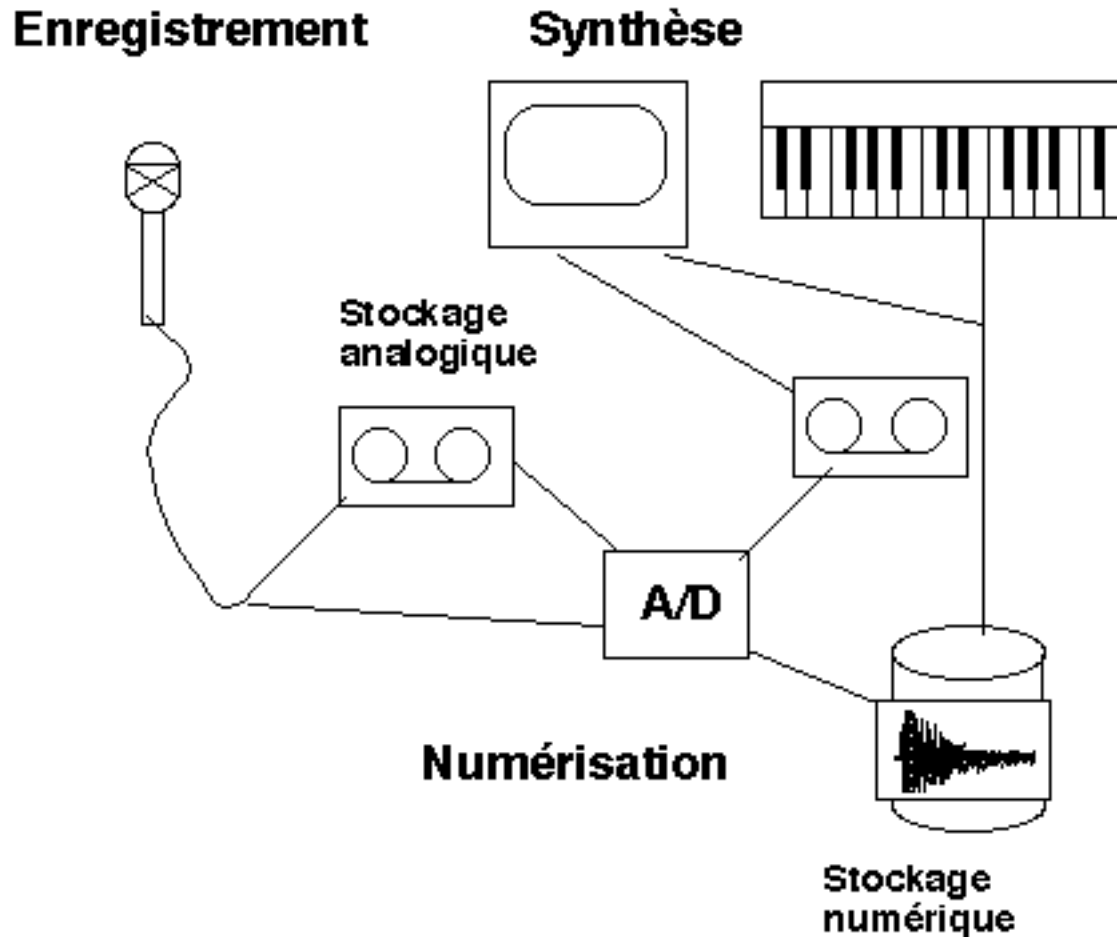
S. Natkin

Décembre 2005

# Audio Numérique

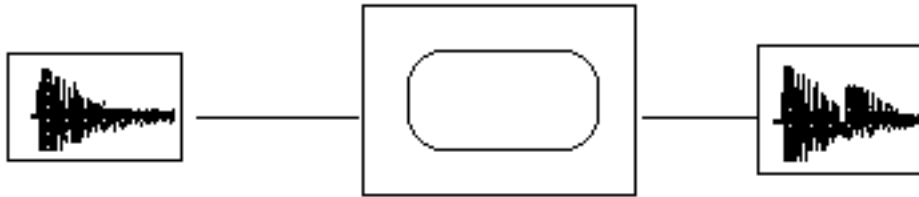
- Introduction: la chaîne des outils audionumériques
- Numérisation du son
- Transformations et effets
- Exemple d'outils

# La chaîne de traitement du son

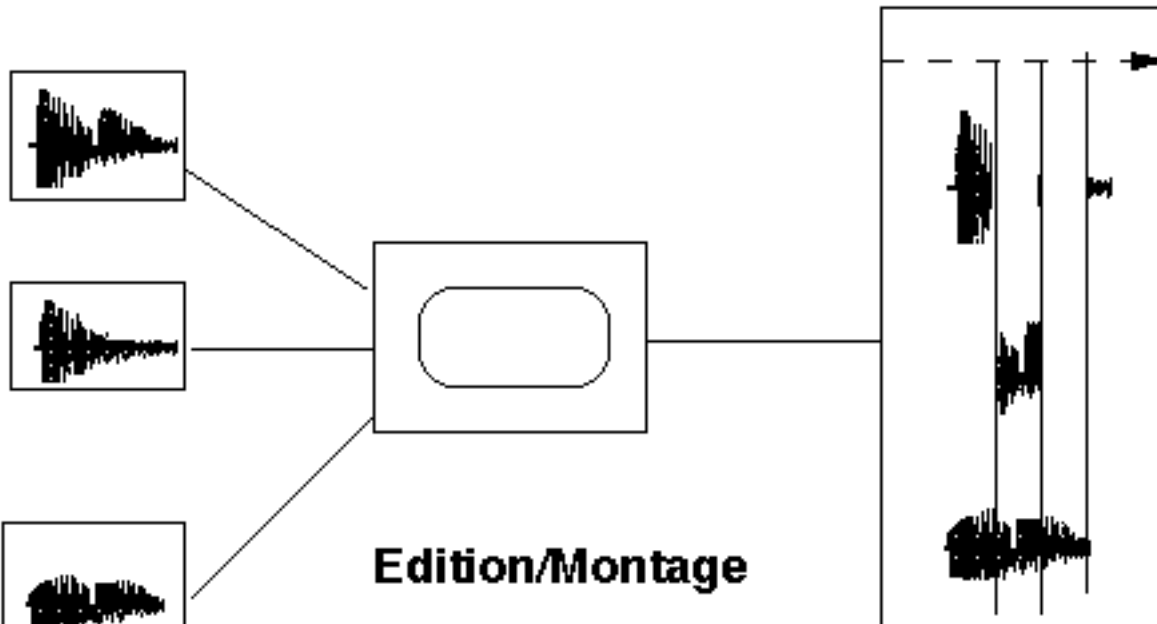


# La chaîne de traitement du son

(2)

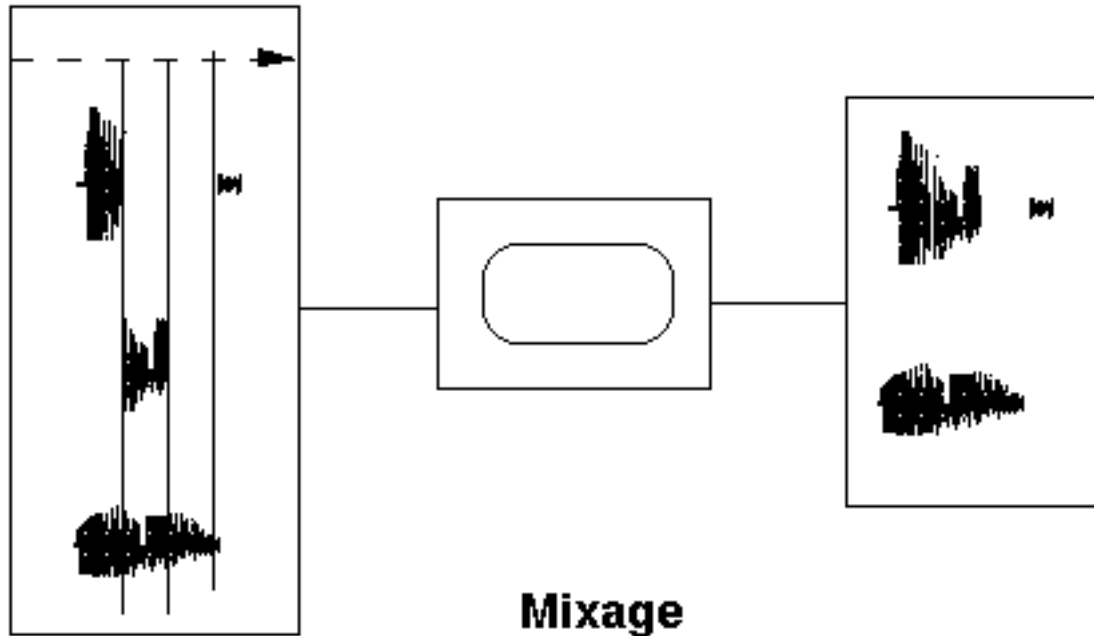


**Transformations/Effets**

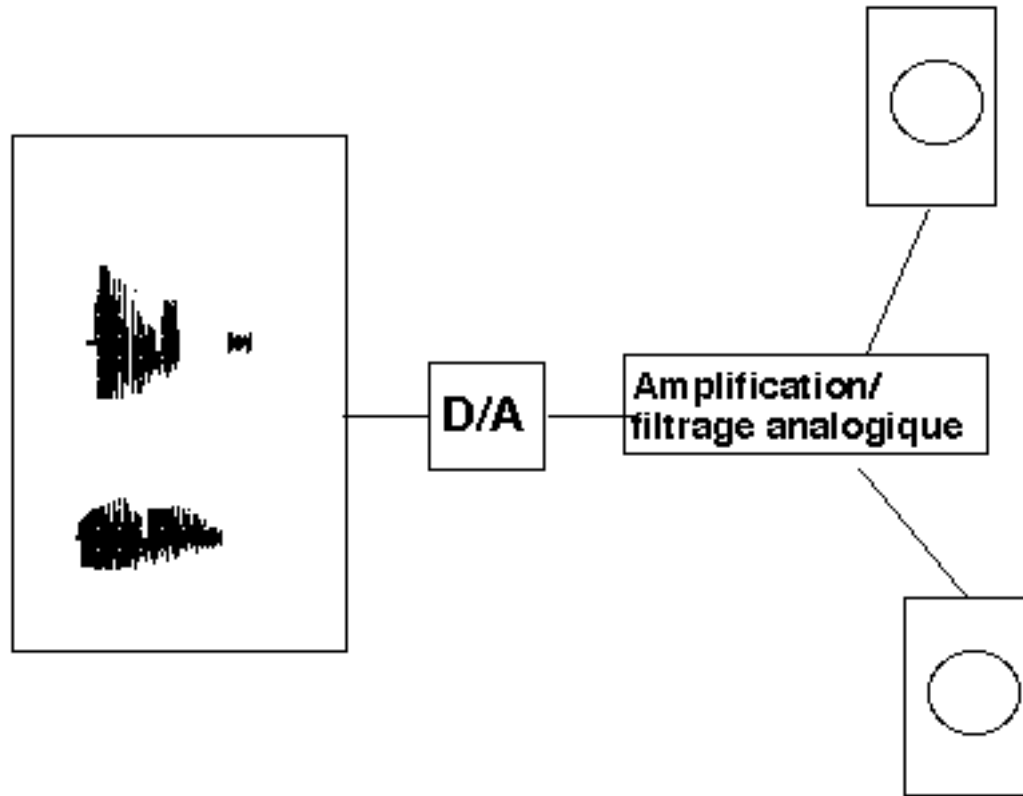


**Edition/Montage**

# La chaîne de traitement du son (3)



# La chaîne de traitement du son (4)



# Limites

Capacité de traitement

(Fft/convolution) => la réalisation des étapes complexes

en temps réel nécessite soit des processeurs dédiés (dsp), soit une capacité de traitement relativement importante

Capacité de stockage

1 s de son mono a 44.1 mpeg layer2 = 128kb

1h de son stéréo qualité cd=500 à 600 mo

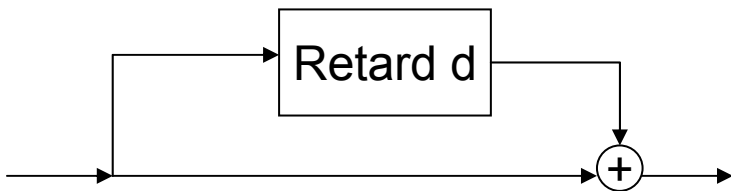
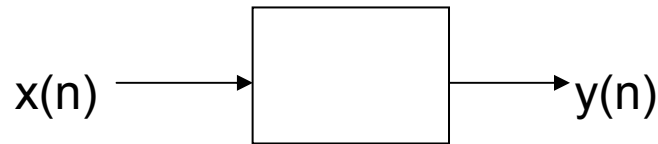
# Effets et transformations



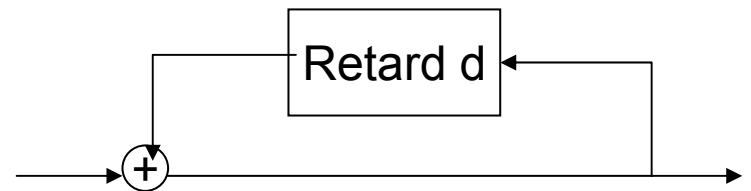
# Filtrage numérique

Très nombreuses techniques de transformation portant soit sur le signal temporel soit sur le spectre

Principales transformations basées sur le filtrage numérique linéaire



$$y(n) = x(n) + x(n-d)$$



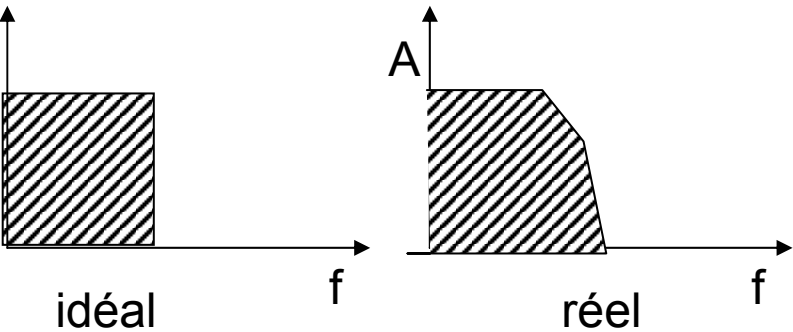
$$y(n) = x(n) + y(n-d)$$

Exemples de base d'un filtre FIR et IIR

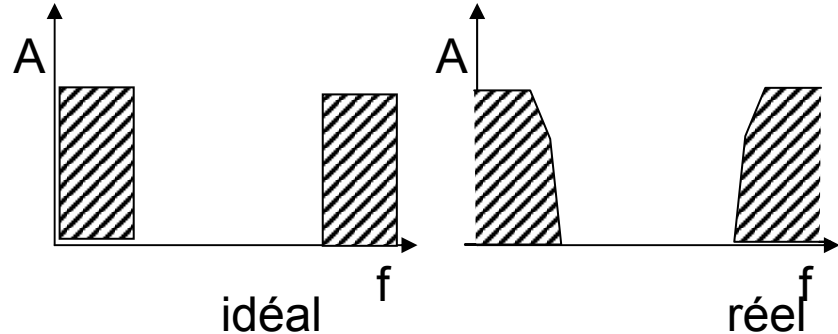
# Effet d'un filtre sur le spectre d'un signal

- Modifie la phase (effet de retard)
- Modifie le spectre d'amplitude

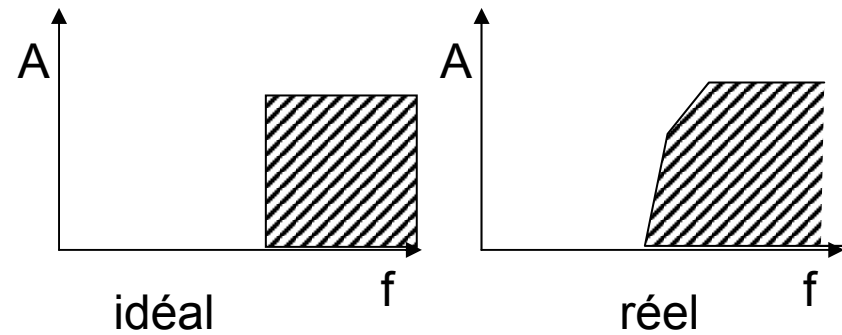
**Passé haut**



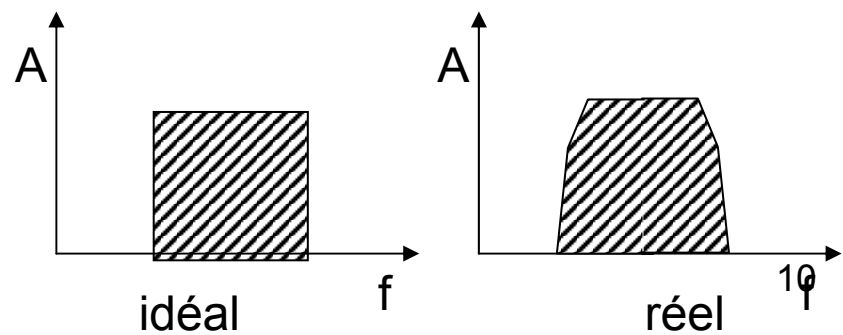
**Passé bande**



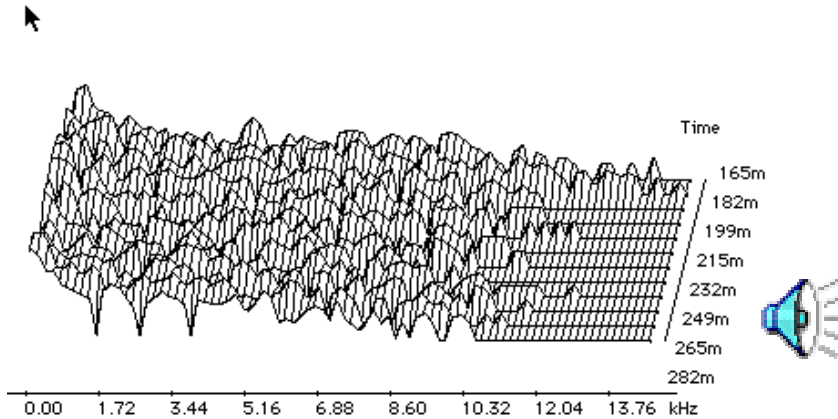
**Passé bas**



**Coupe bande**



# Filtrage numérique audio



**Parametric EQ**

Channel(s):  Left  Right  Both

Soundfile:

Filter Type: **Low Shelf**

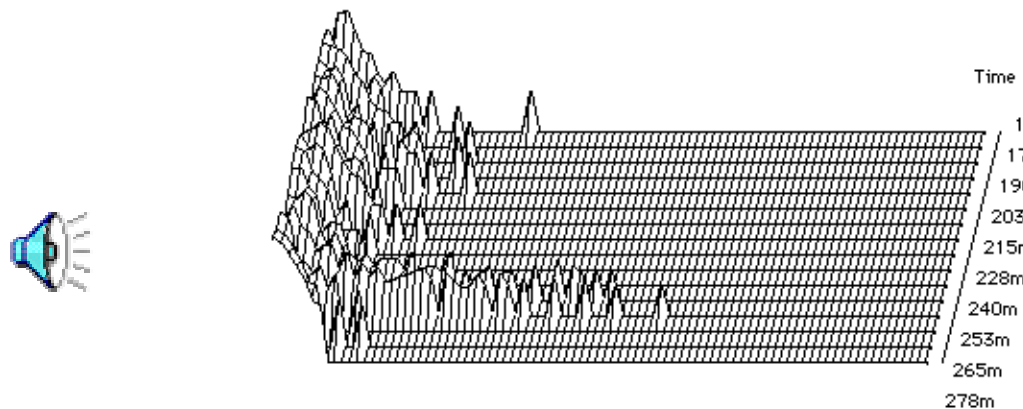
Sample Rate: 44100 Hz

Corner Freq.  Hz

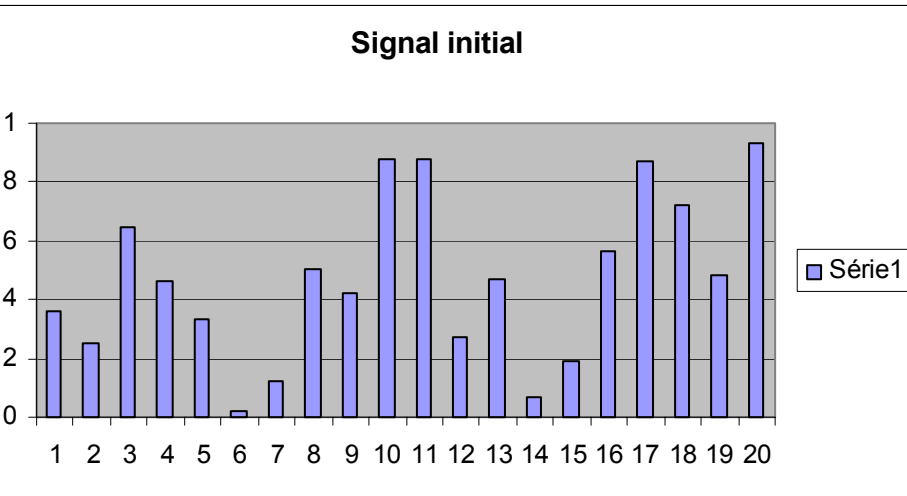
Boost/Cut:  dB

Use for playback  Bypass

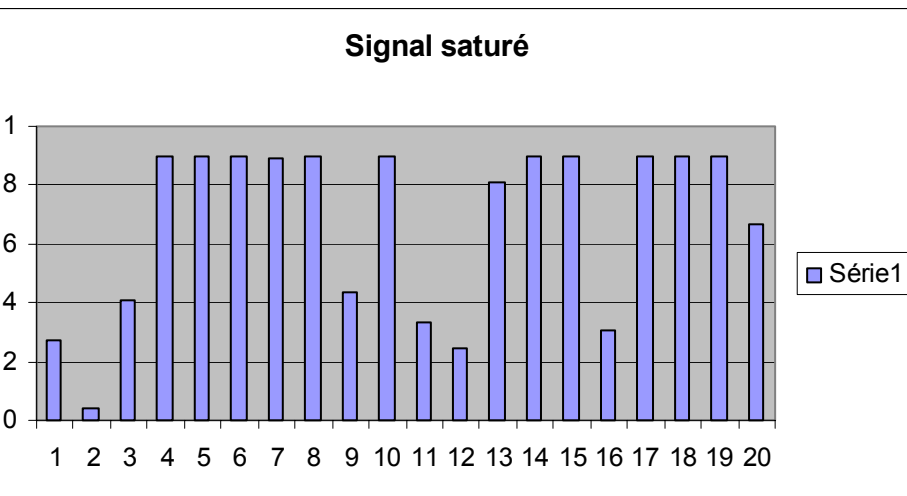
The EQ window displays a frequency response graph with a low shelf filter curve. The x-axis is labeled 'Uol' and ranges from 0k to 22k Hz. The y-axis represents gain in dB, with a knob set to 1000 dB. The curve shows a low-frequency shelf that rises and then levels off at higher frequencies.



# Un filtre non linéaire: effet de saturation

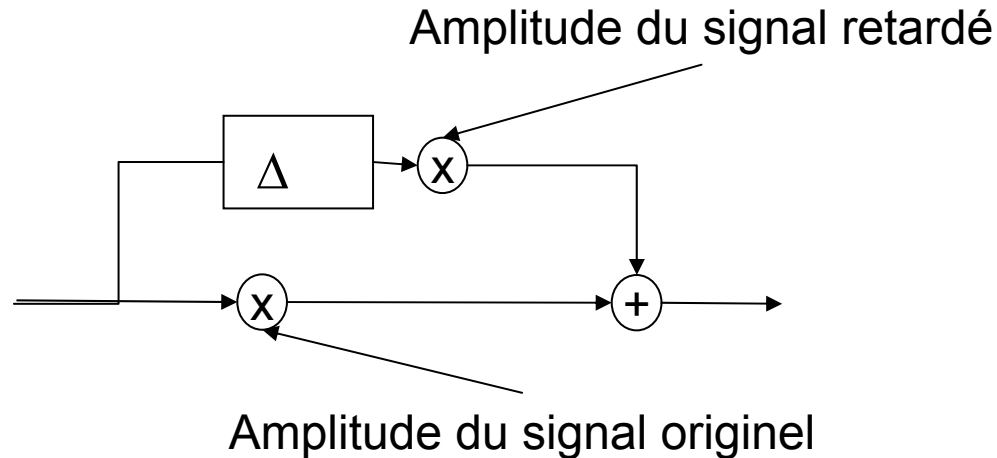


$$y(n) = \min(1,5.s(n),0,9)$$



# Les effets de retards

Schéma d'une ligne à retard simple



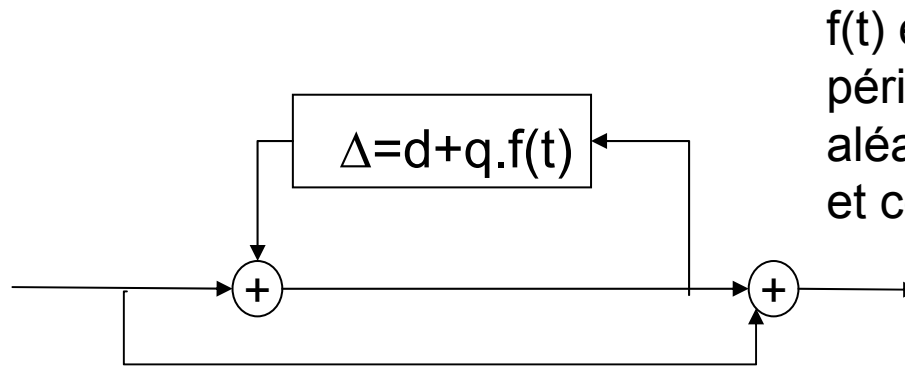
- Court  $< 10$  ms modification du spectre (un ou deux échantillons passe bas FIR, plus long filtre en peigne)
- Moyen  $> 10$  ms et  $< 50$  ms effet d'ambiance, rehausse un signal mince
- Long  $> 50$  ms effet d'écho



Utiliser aussi pour donner un effet de localisation dans l'espace



# Flanger



$f(t)$  est une fonction périodique ou une fonction aléatoire de moyenne nulle et comprise entre -1 et +1



Rolling Stone

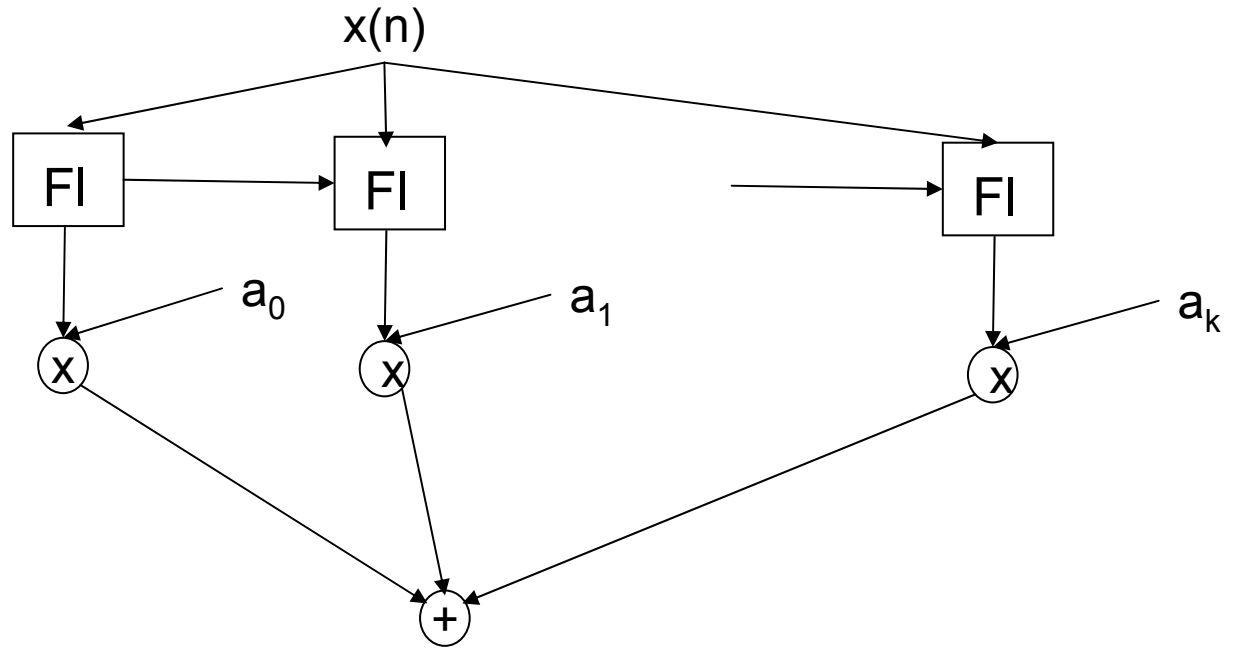


Rolling Stone Flangé

Version plus élaborée le phasing

# Effet Chorus

Une technique

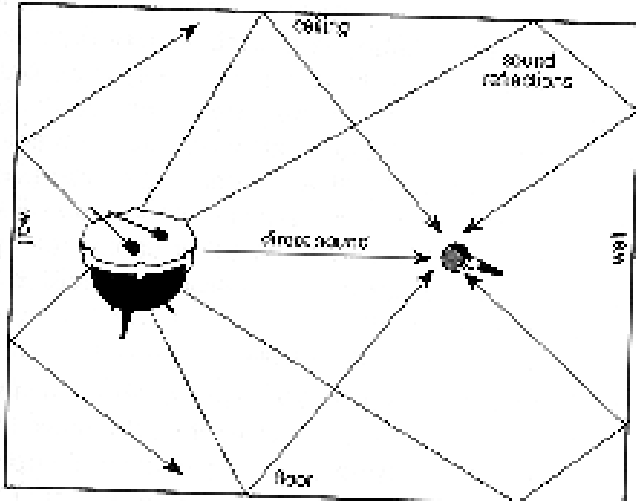


Sans

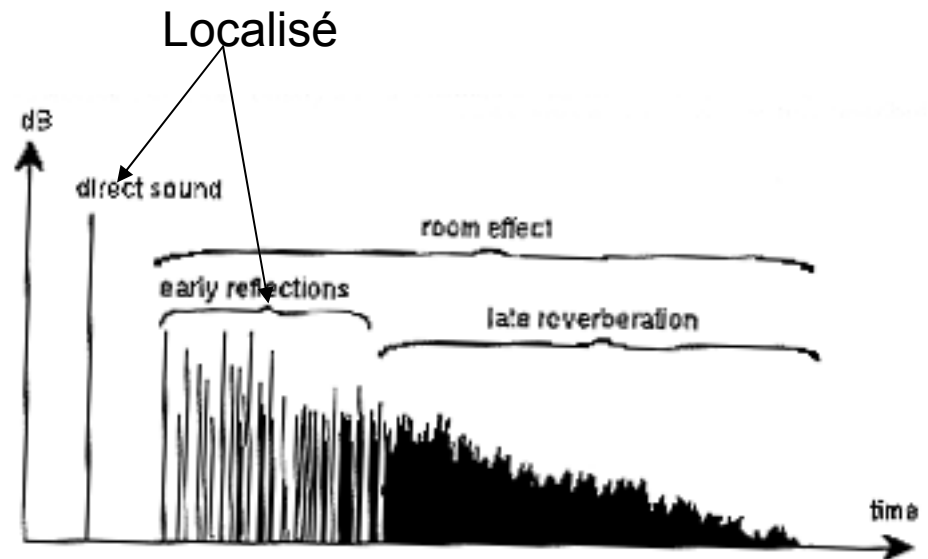


Avec

# Spatialisation



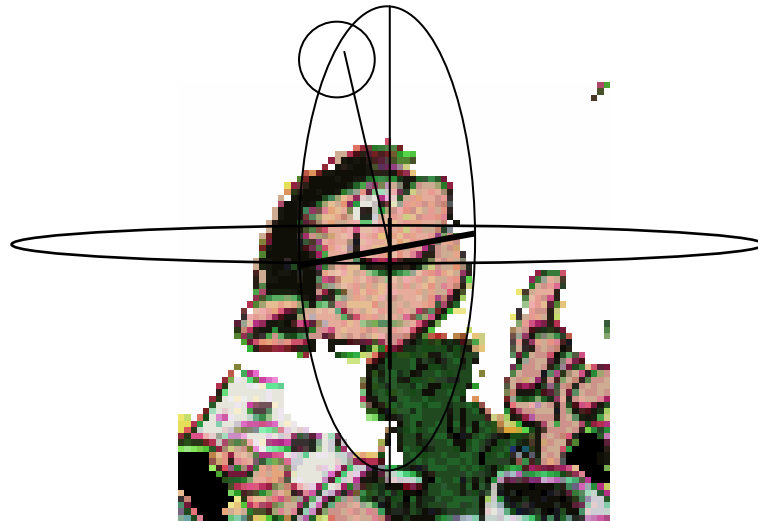
Effet de réverbération  
Effet de localisation



Spectre du signal réverbéré (JM)



# Détermination de la position d'un son dans l'espace



## Détermination de l'azimut et zenith

- Différence de temps entre les deux oreilles
- Différence d'amplitude avant arrière (ombre de la tête)
- Différence spectrale liées aux réflexions asymétriques (pavillons, épaules, torses)

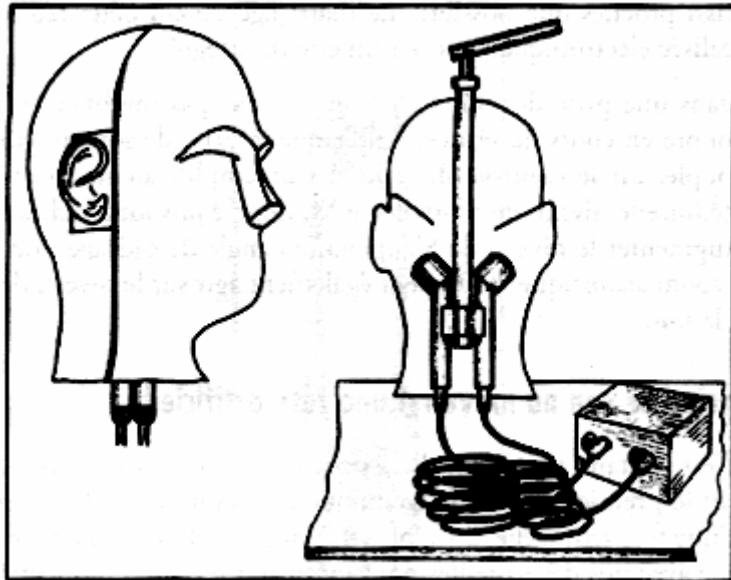
## Détermination de la distance

- Rapport entre le signal direct et le signal réverbéré
- Pertes des composantes hautes fréquences
- Pertes des détails

# Restitution

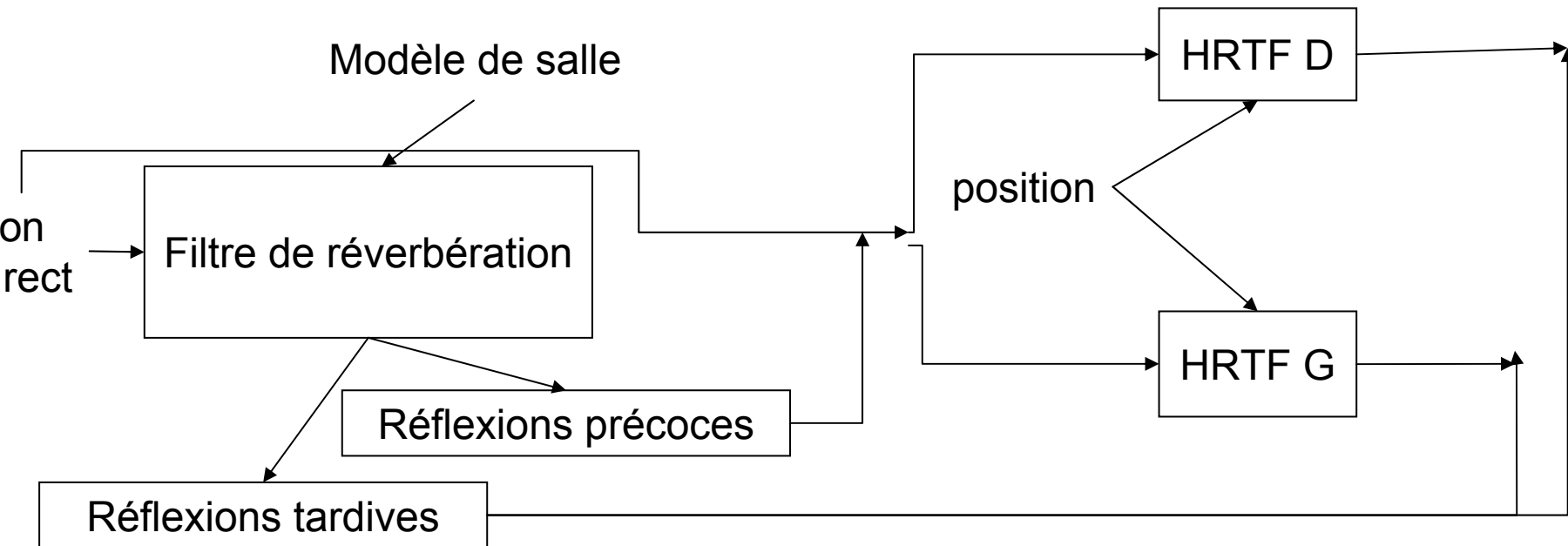
- Mono (distance)
- Stéréo (distance et localisation sur un plan)
- Spatial avec deux hauts parleurs (transaural)
- Spatial avec un casque (binaural)
- Spatial avec N haut parleurs

# HRTF

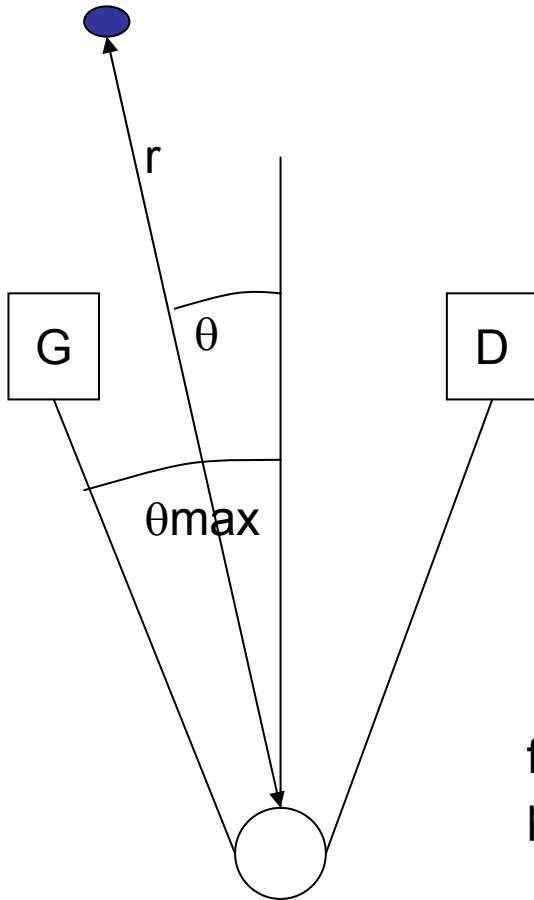


Détermination expérimentale des filtres définissant la différences de perception entre les deux oreilles pour une position d'un objet sonore.  
Dépend de la tête...

# Exemple de processus de spatialisation binaural



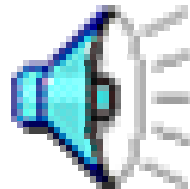
# Pan Pot d'intensité (2D)



$$A_G = A.f(r). \frac{\sqrt{2}}{2} [\cos \theta + \sin \theta]$$

$$A_D = A.f(r). \frac{\sqrt{2}}{2} [\cos \theta - \sin \theta]$$

$f(r)$  décroît avec la distance. Peut être différente pour le son direct, les réflexions précoces et tardives



# Codage du son numérique

# Eléments de codage

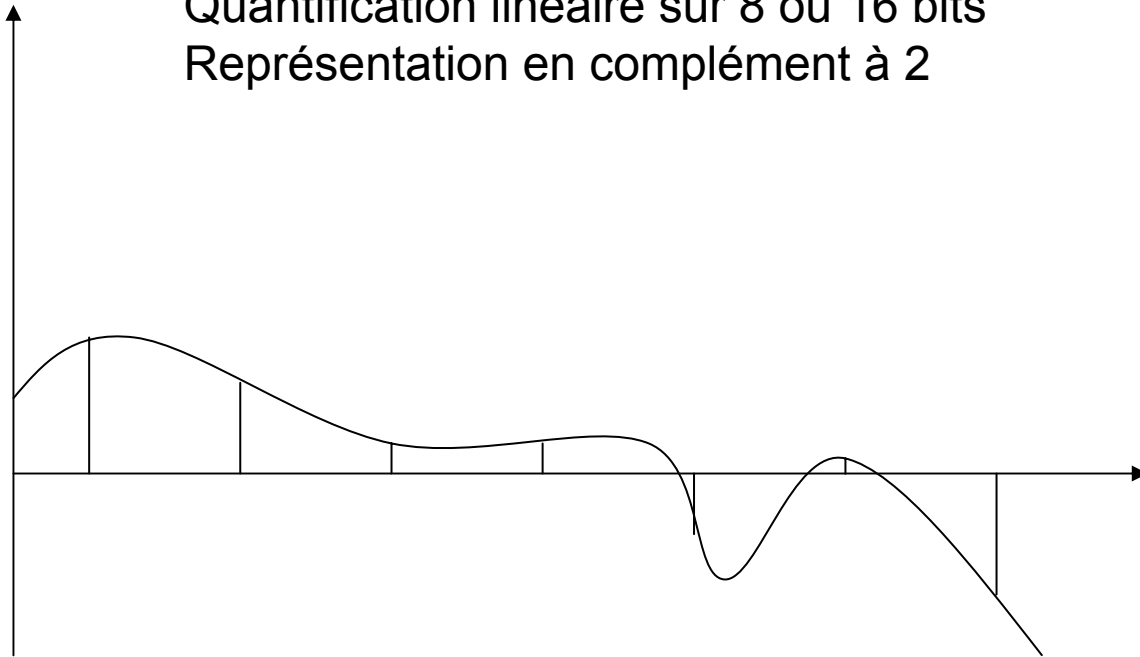
- Fréquence d'échantillonnage (2Khz-80 Khz)
- La loi de quantification (8-16 bits)
- La (les) algorithmes de compression (symétrique ou asymétrique)
- La représentation binaire

Exemple : le codage sur un CD fréquence 44.1 Khz échantillonnage sur 16 bits

soit 700 kb/s

# Pulse Code Modulation

Représentation temporelle du signal échantillonné  
Fréquences d'échantillonnage  
de 4Khz à 44Khz  
Quantification linéaire sur 8 ou 16 bits  
Représentation en complément à 2





# PCM différentiel

DCMP codage différentiel: on code  $x(0)$  et  $d_n = x(n) - x(n-1)$

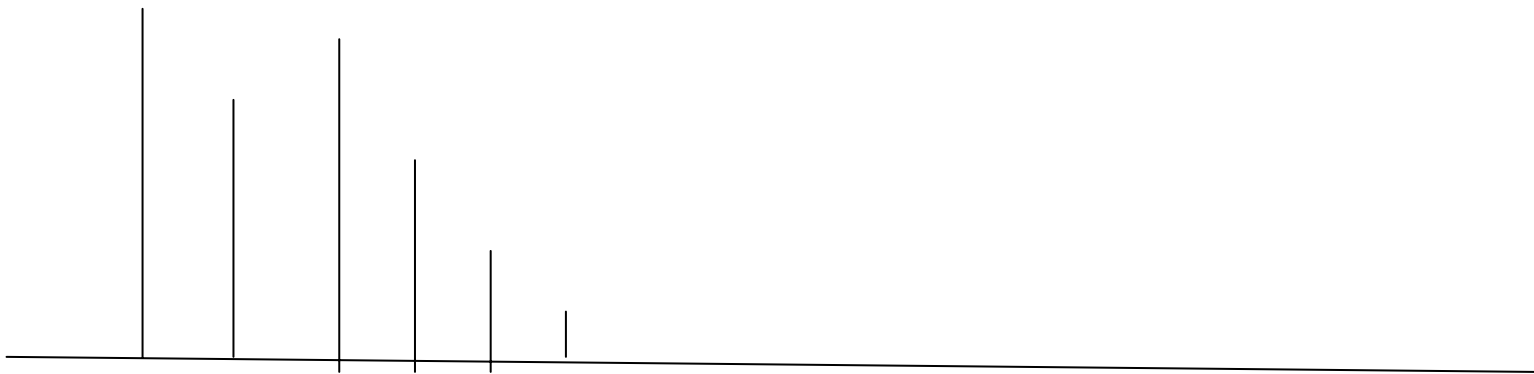
Efficace si le signal varie lentement.

Par exemple si  $x(n)$  est sur 16 bits on code la longueur de  $d(n)$  sur 3 bits puis  $d_n$

Efficace si  $d_n$  tient sur moins de 13 bits

ADCMP: coder la variation des différences.

Par exemple (schéma IMA) 16 bits  $\rightarrow$  4 bits



# Adaptative PCM : ADPCM

Principe: coder la variation des différences. Par exemple (schéma IMA)

$$x(n-1) = x(n-2) + d_{n-1} \text{ avec } d_{n-1} = D(k)$$

Notons  $i_{n-1} = k$

$$x(n) = x(n-1) + d_n \text{ avec } d_n = D(t)$$

$i_n = t$

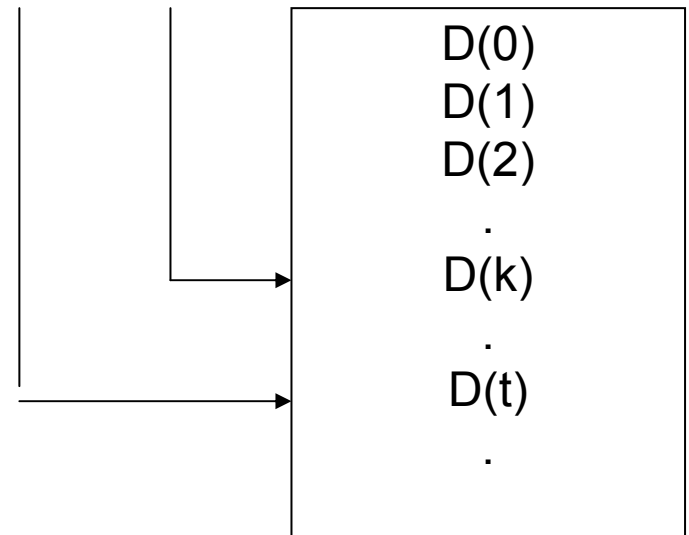
On code

$x(0)$  sur 16 bits

$i_1$  sur 8 bits

$di_n = i_n - i_{n-1}$  sur 4 bits

Table des différences D



# Compression psycho perceptive

- La perception du son est logarithmique

Utiliser une quantification logarithmique

- L'oreille distingue mieux les sons dans la bande 300 – 6000 Hz

Echantillonner différemment selon les bandes de fréquence

- Certains sons en masque d'autre

Supprimer les informations non perçues

# Modulation par Impulsion et Codage

Codage utilise pour le téléphone numérique (RNIS)

Bande passante de 4000hz

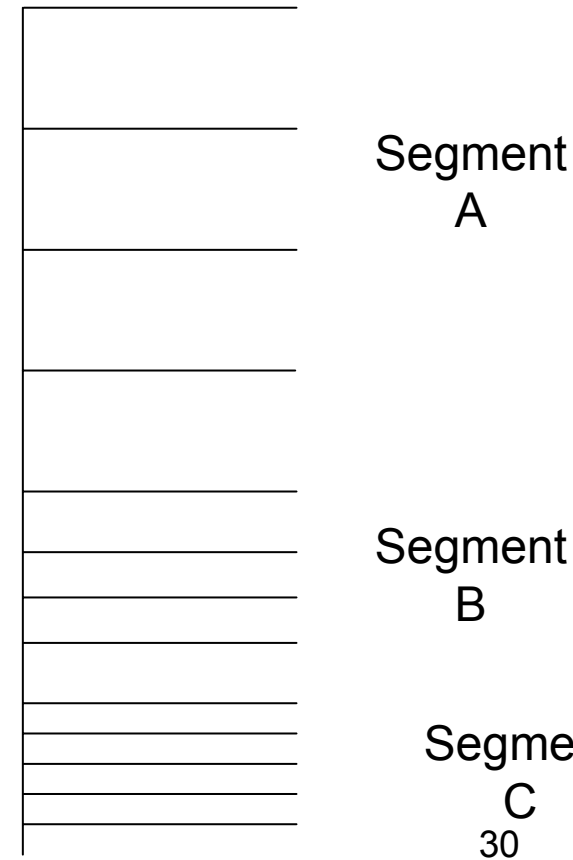
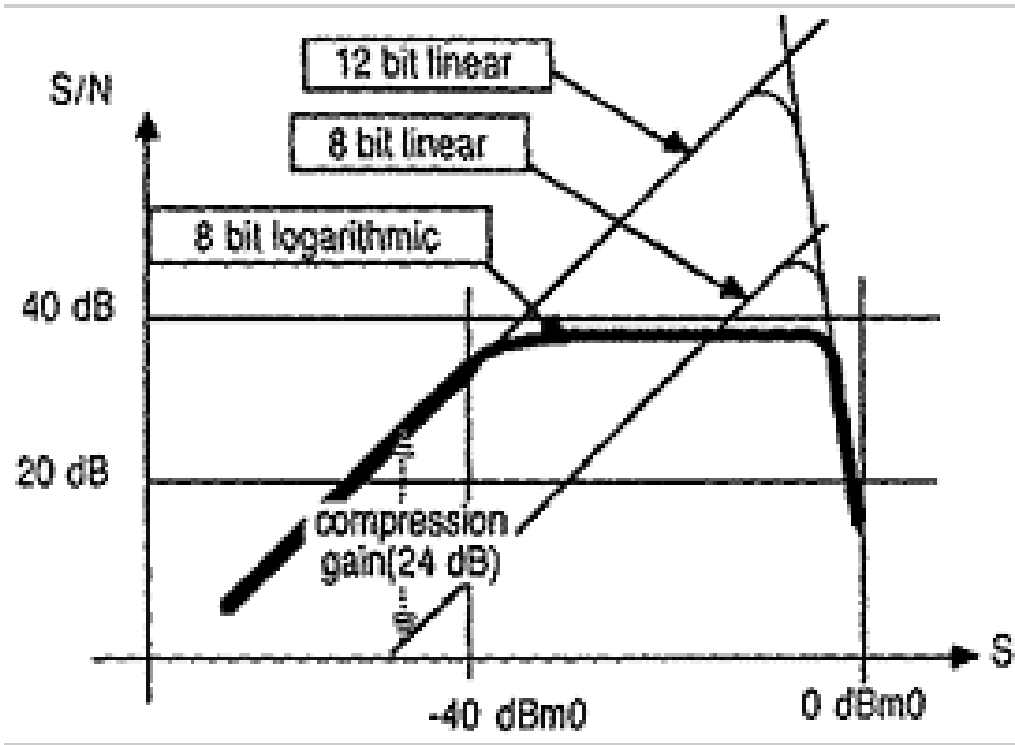
=> 8000 échantillons par secondes (th de niqyst)

Un échantillon toutes les 125 microsecondes

Une numérisation sur 8 bits (loi de quantification non linéaire tenant compte des spectres d'audition (mu law USA/ A law en Europe)

Débit  $8000 \times 8 = 64 \text{ kb/s}$

# Quantification perceptive: lois de quantification mu et A



Mu: 8 segments de 16 Valeurs,  
incrément de 2 de à 256  
A : 7 segments le premier de  
32 valeurs et les autres de 16  
valeurs

# Notion de bande critique

- Dans un intervalle de fréquence autour d'une fréquence centrale, l'oreille ne distingue pas nettement deux sons.
- Si un son a une intensité forte il masque, pendant une courte période les sons d'intensité plus faible dans sa bande critique
- La notion de bande critique est lié à un effet de filtrage en peigne à Q constant effectué par l'oreille
- En première approximation il est possible de décomposer un son audible sur une vingtaine de bande de fréquences (pendant du RVB visuel)

# Compression fréquentielle perceptive (ex MP3)

MPEG 1 est une norme de transmission de vidéo numérique (Image animée+ son) intégrant différentes qualités selon les types D'application (visiophonie...Video a la demande)

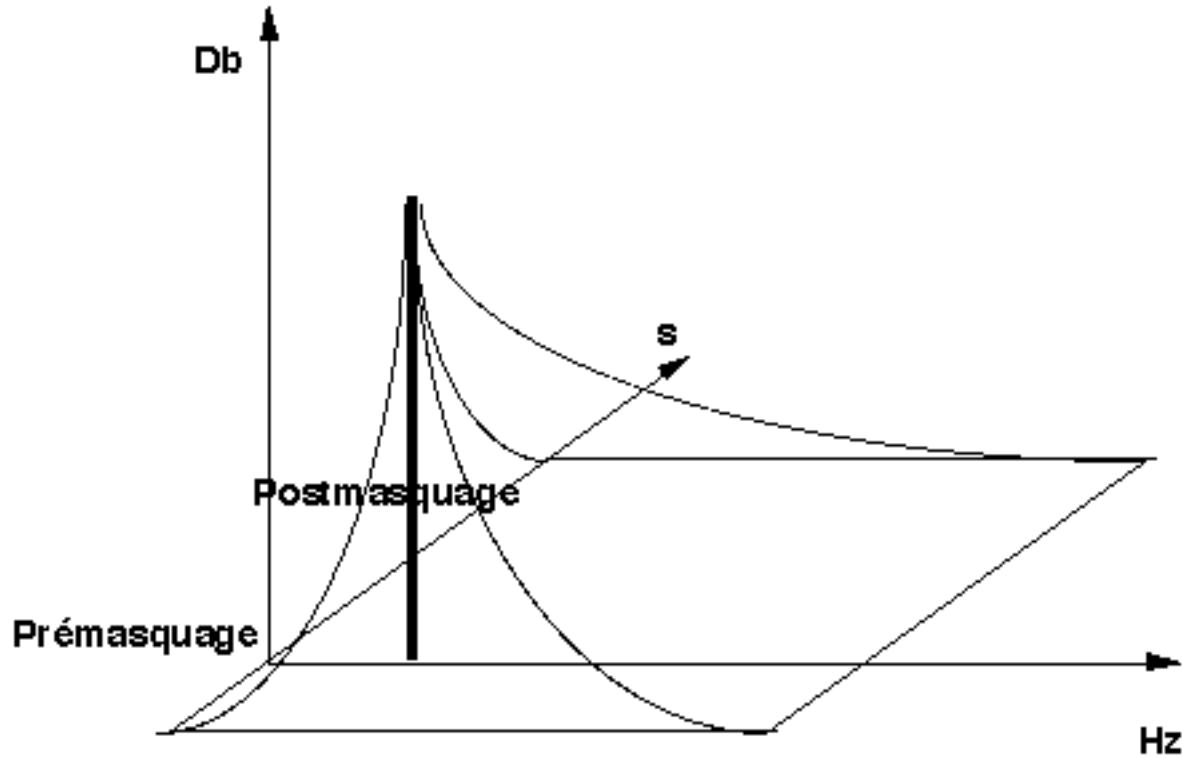
Le son est échantillonné a 32, 44,1 kHz ou 48 kHz

MPEG 1 Layer 3 (MP3) est un codage du son , il s'agit d'une compression non conservative basée sur le masquage

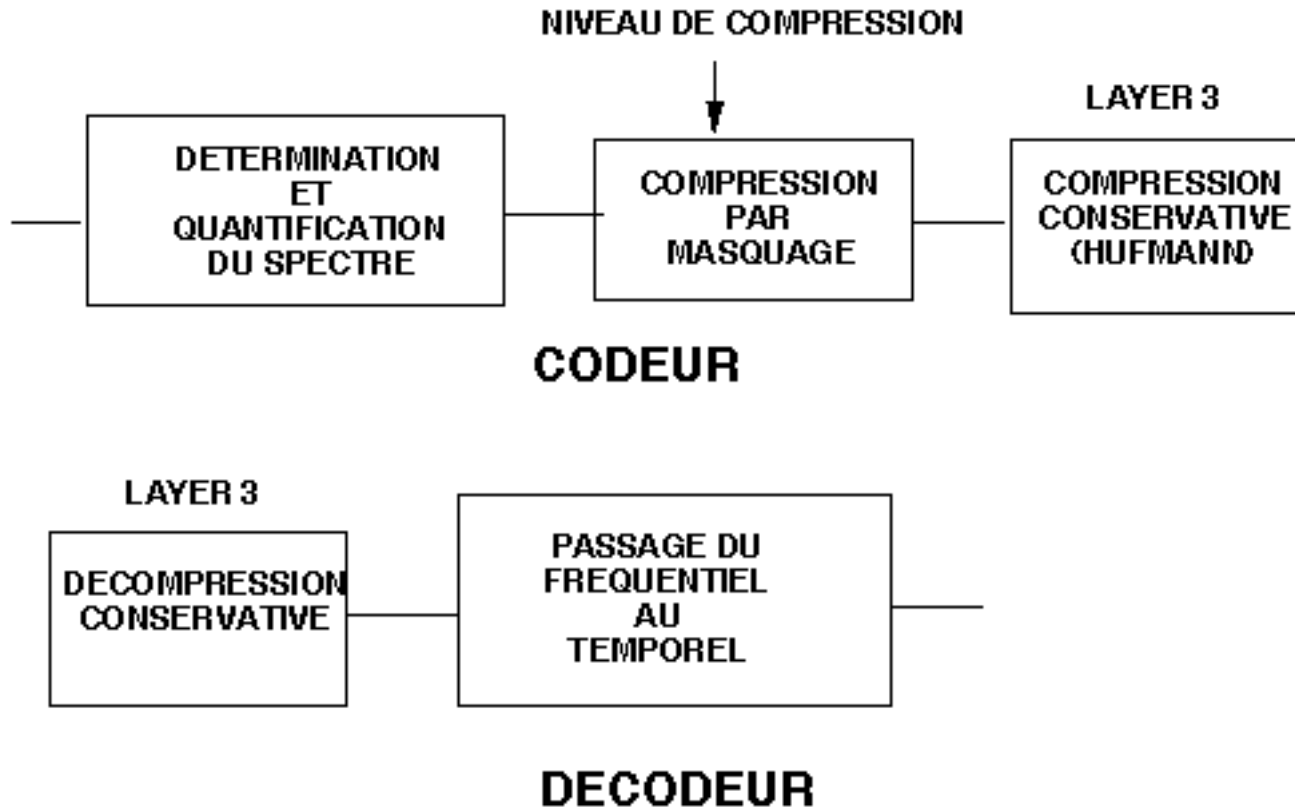
3 niveaux avec des codeurs de plus en plus complexes en  
Fonction de la qualité de son visée



# Utilisation des effets de masquage



# Codeur MP3



# Compression MPEG1

**Table 6.6 Summary of MPEG audio**

	<i>Bit rate range (Kbit/s)</i>	<i>Target bit rate (Kbit/s)</i>	<i>Typical compression ratio</i>
Layer 1	32–448	192 (mono)	1:4
Layer 2	32–384	128 (mono)	1:6 to 1:8
Layer 3 (MP3)	32–320	64 (mono) 128 (stereo)	1:10 to 1:12

# Compression dynamique de la voix

Mode de compression utilisé pour un traitement en temps réel de la voix sur un canal à faible débit (GSM, RealAudio...)

Basé sur un modèle acoustique de la gorge qui permet de caractériser le son initial et l'effet de résonance (formants)

# LPC: Linear Predictive Coding

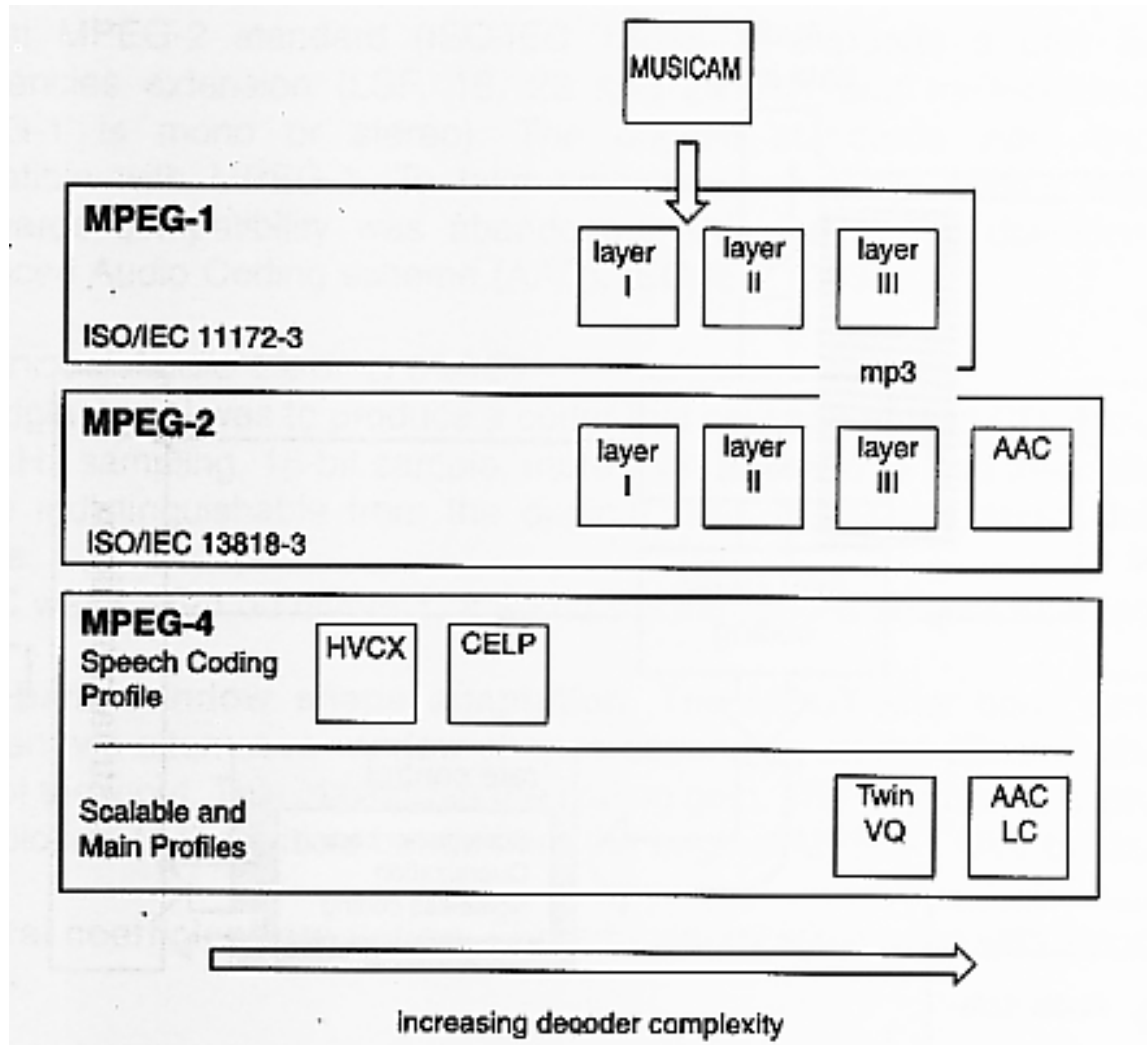
Les formants sont prédits par un filtre linéaire adaptatif qui induit le formant courant par rapport aux précédents.

Une fois le formant calculé il est soustrait du signal initial ce qui donne le résidu qui doit en théorie contenir le son direct

En pratique il faut transmettre une partie du formant ou une référence à un formant voisin

Amélioration le CELP

# MPEG Audio



# Codec MPEG 4

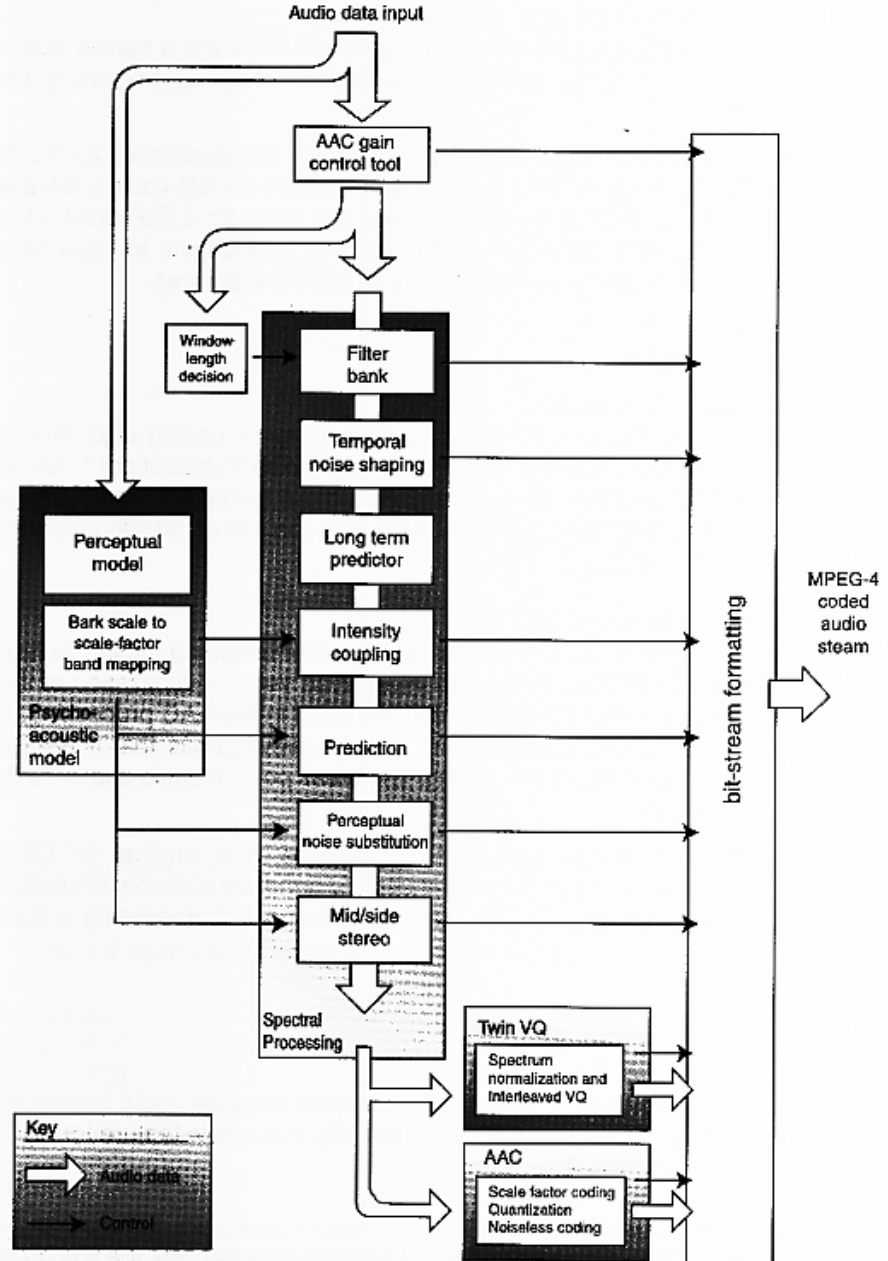


Figure 6.15 MPEG-4 general audio coder.

# Débits MPEG 4

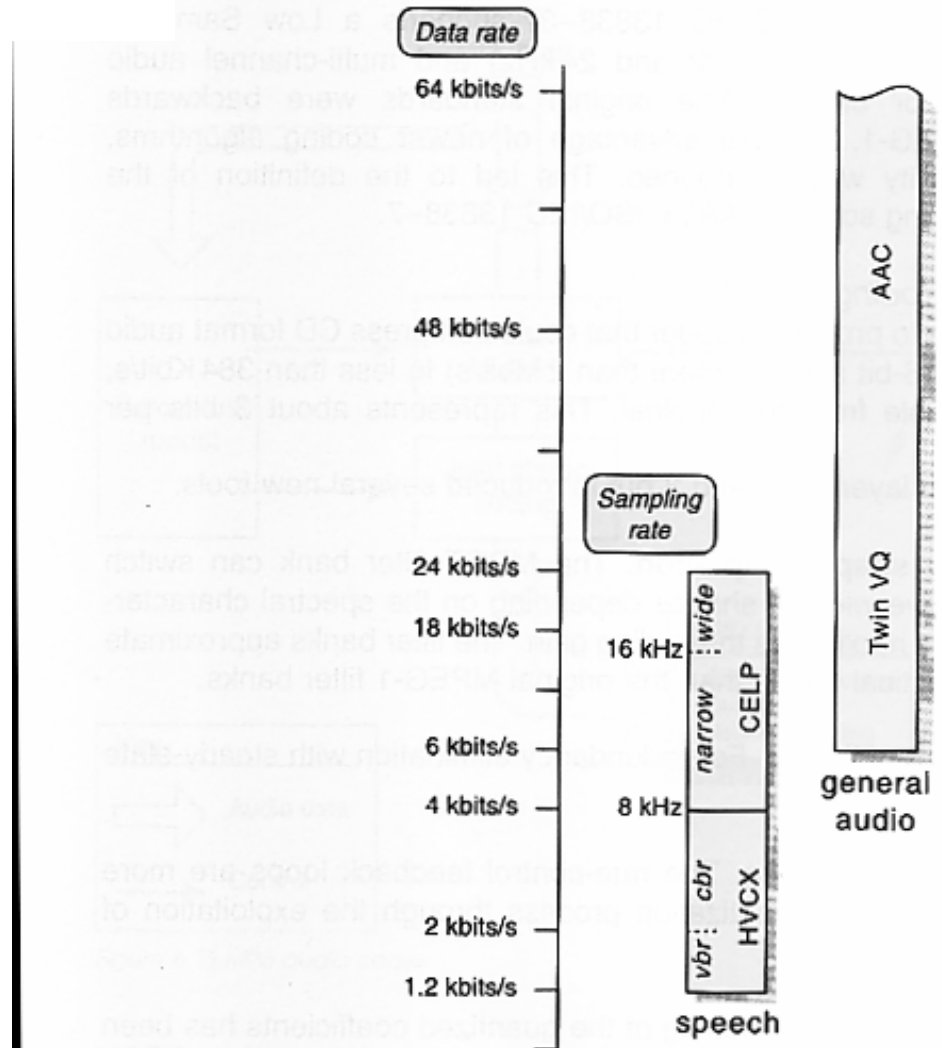


Figure 6.14 MPEG-4 audio coding algorithms.



# Formats Audio en général

Table 6.5 Summary of codec algorithms

<i>Organization/family</i>	<i>Algorithm</i>	<i>Codec</i>
MPEG		MPEG-1, layers 1, 2, 3 MPEG-2, layer 2 and AAC MPEG-4, layer
ITU-T	PCM ADPCM	G.711 PCM ( $\mu$ -law, A-law) G.726 ADPCM at 16, 24, 32 and 40 Kbit/s (replaces G.721 at 32 Kbit/s and G.723 for 24 and 40 Kbit/s)
Interactive Multimedia Association	ADPCM	IMA/DVI
US DOD Federal Standard	Linear predictive CELP	LPC-10E (1015) (1016)
ETSI	Linear predictive	GSM 06.10 (global système mobile)
Proprietary	ACELP ACELP ATRACS ADPCM	Voiceage Sipro Sony ATRAC3 Microsoft ADPCM

# Comparaison superficielle

<b>Codage</b>	<b>F Ech</b>	<b>Q</b>	<b>Débit Kb/s</b>	<b>Facteur de comp</b>	<b>Sym/Assym</b>
PCM/CD	44100	16	705600	1	S
ADCMP	44100	16	176400	4	A
MP3	44100	16	70560	10	A
MIC	8000	10	64000	11	S
LPC (Cell)	8000	8	8000	88,2	S

# Introduction aux techniques de synthèse du son

S. Natkin 2003

# Généralités

Construction d'un son par un dispositif électronique (ici numérique) .

## **Synthèse par tables d'ondes**

Par exploration et altérations d'échantillons enregistrés

## **Synthèse additive**

Par addition de signaux simples (sinusoidaux par exemple)

## **Synthèse granulaire**

Multiplexage temporel de petits éléments dont l'ensemble des éléments sont contrôlés

## **Synthèse soustractive**

Par filtrage de signaux riches en harmoniques

## **Synthèse par modulation**

Par variation temporelle d'un paramètre d'un signal se décomposant selon une série de Fourier dont les harmoniques dépendent de ce paramètre

## **Synthèse par modèles physiques**

Par simulation des lois acoustiques

# Synthèse par table d'onde

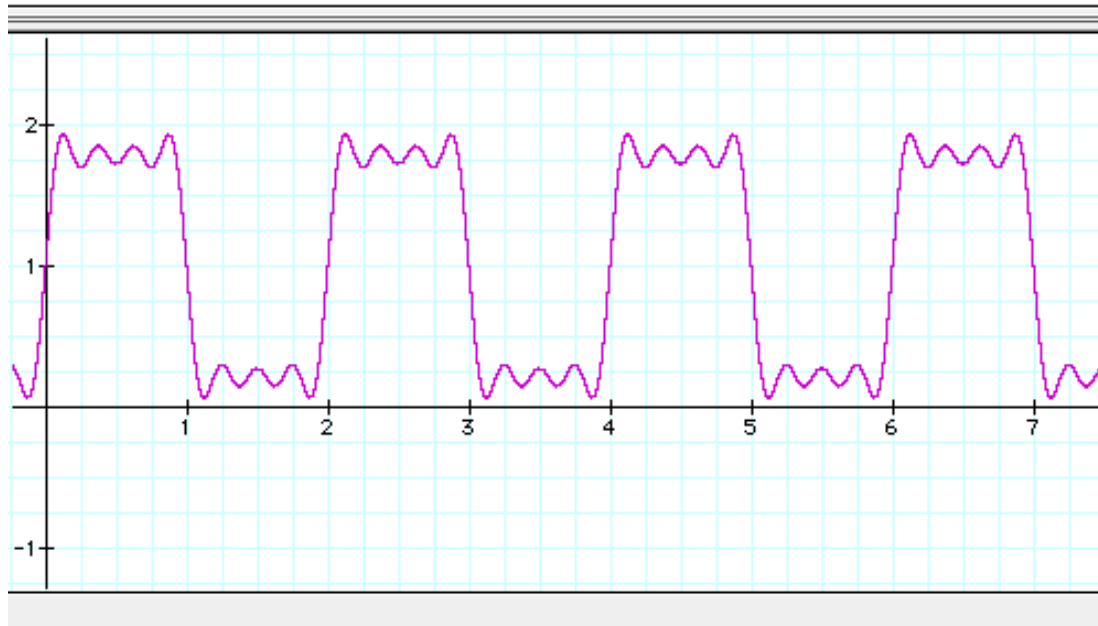
- On part d'une table d'un son enregistrés qui est lue périodiquement
- On modifie le fréquences soit en « sautant des échantillons », soit en créant artificiellement (avec éventuellement un processus d'interpolation)
- Les échantillons peuvent être altérés en hauteur en suivant une enveloppe d'amplitude

# Utilisation de plusieurs tables

- Enchaînées (pour des périodes déterminées)
- Additionnée (Somme des constructions précédentes)
- Version plus complexes (terrains d'onde)
- Standard Microsoft Downloadable Sound Level (DSL)

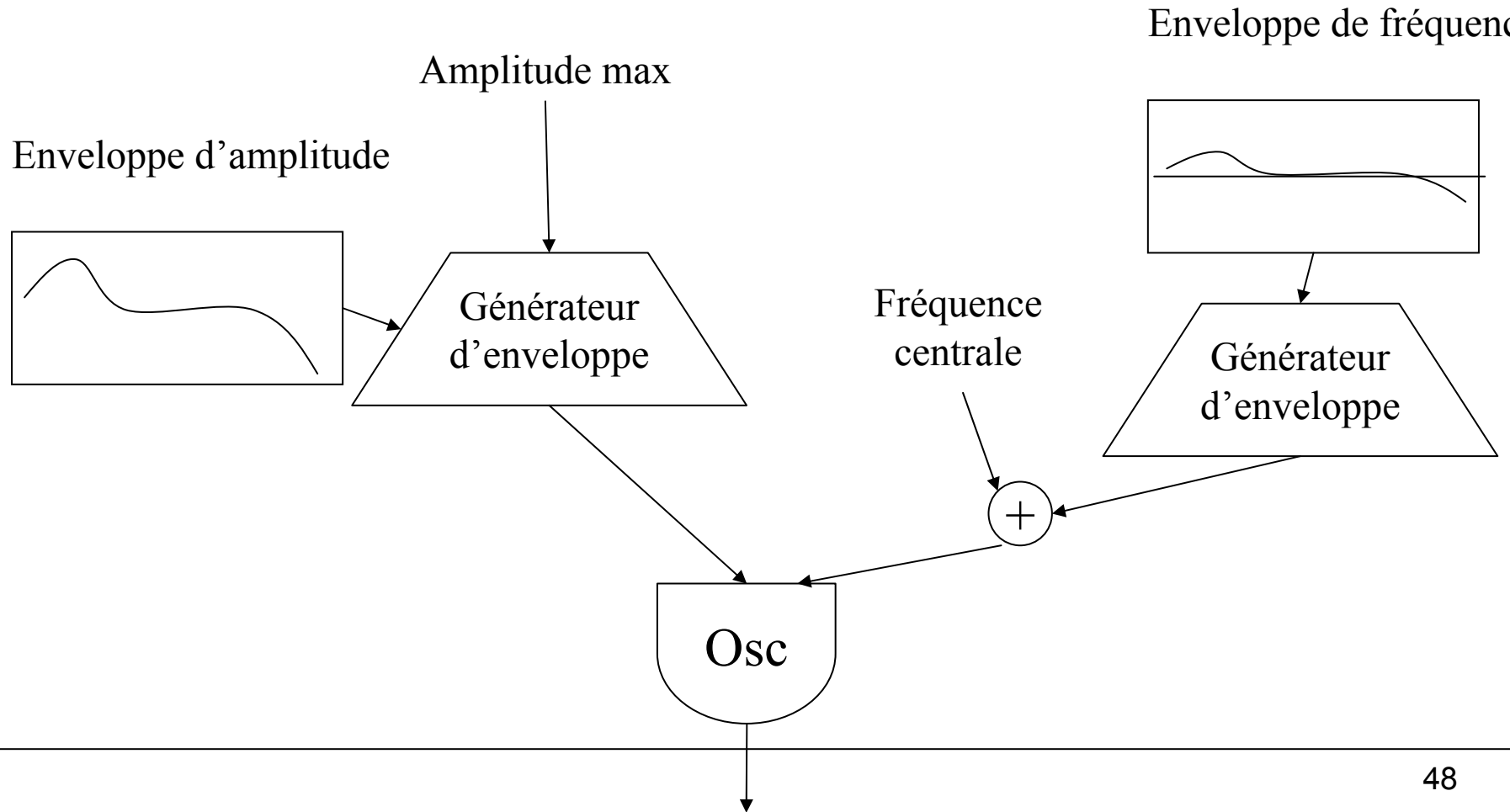
# Synthèse additive

$$y = 1 + \sin \pi x + \frac{1}{3} \sin (3 \pi x) + \frac{1}{5} \sin (5 \pi x) + \frac{1}{7} \sin (7 \pi x)$$



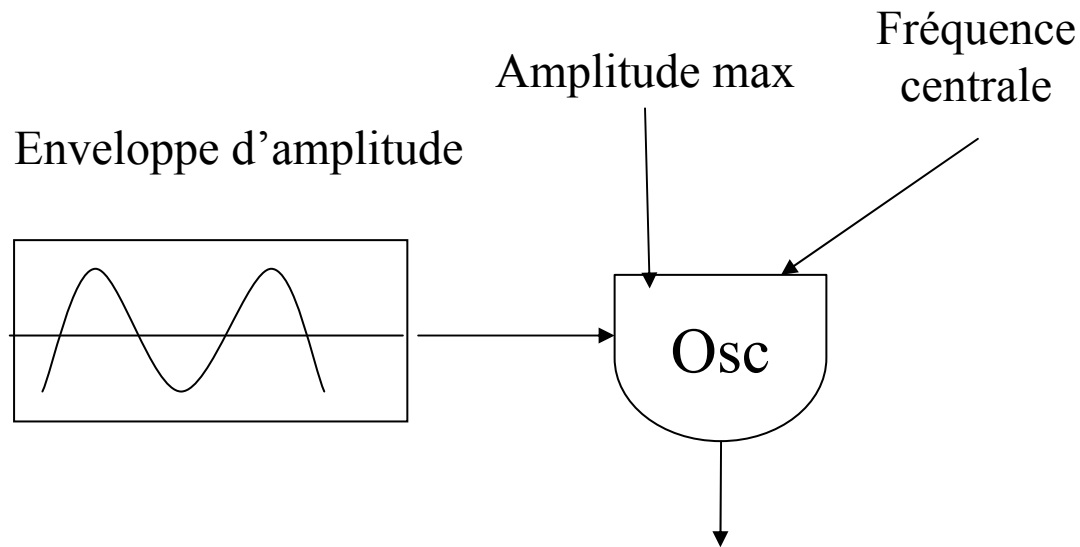
# Schéma de base

## Générateur de base

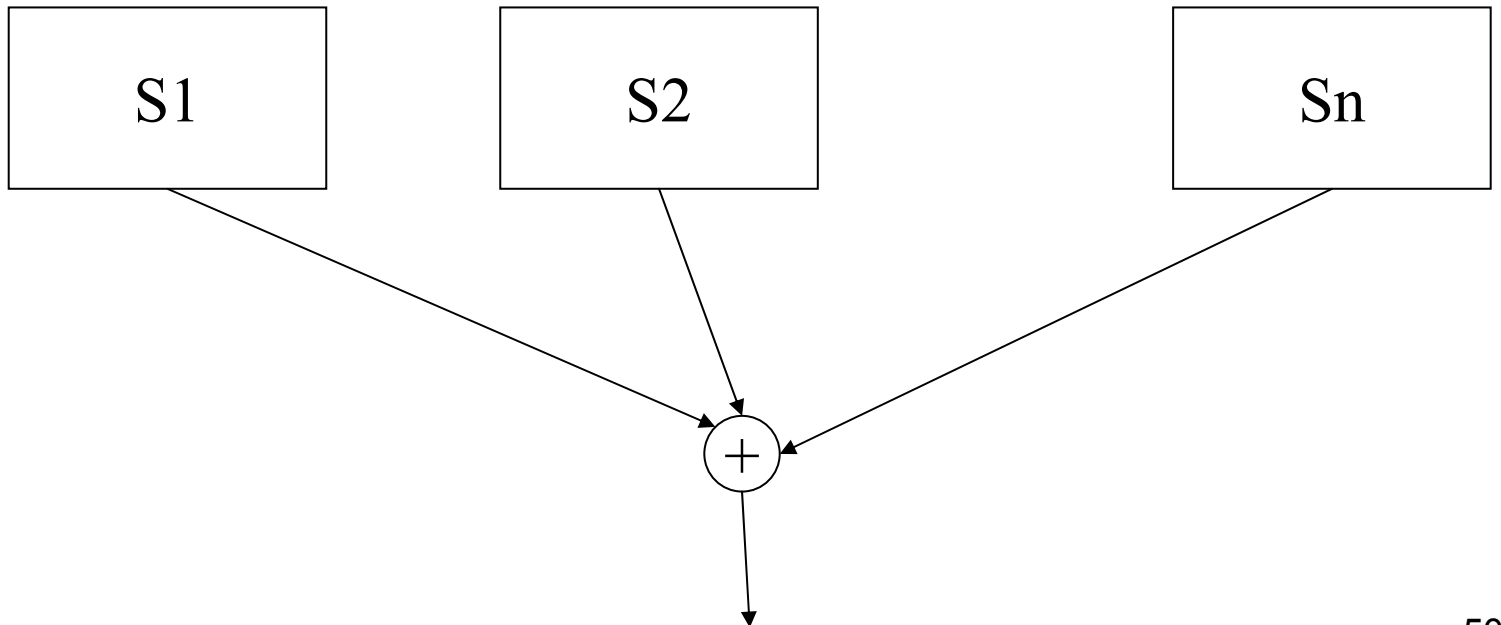




# Schéma d'une synthèse sinusoidale



# Synthèse complexe



# Synthèse par modulation

Principe:

Moduler un Signal sinuosidal par un autre signal

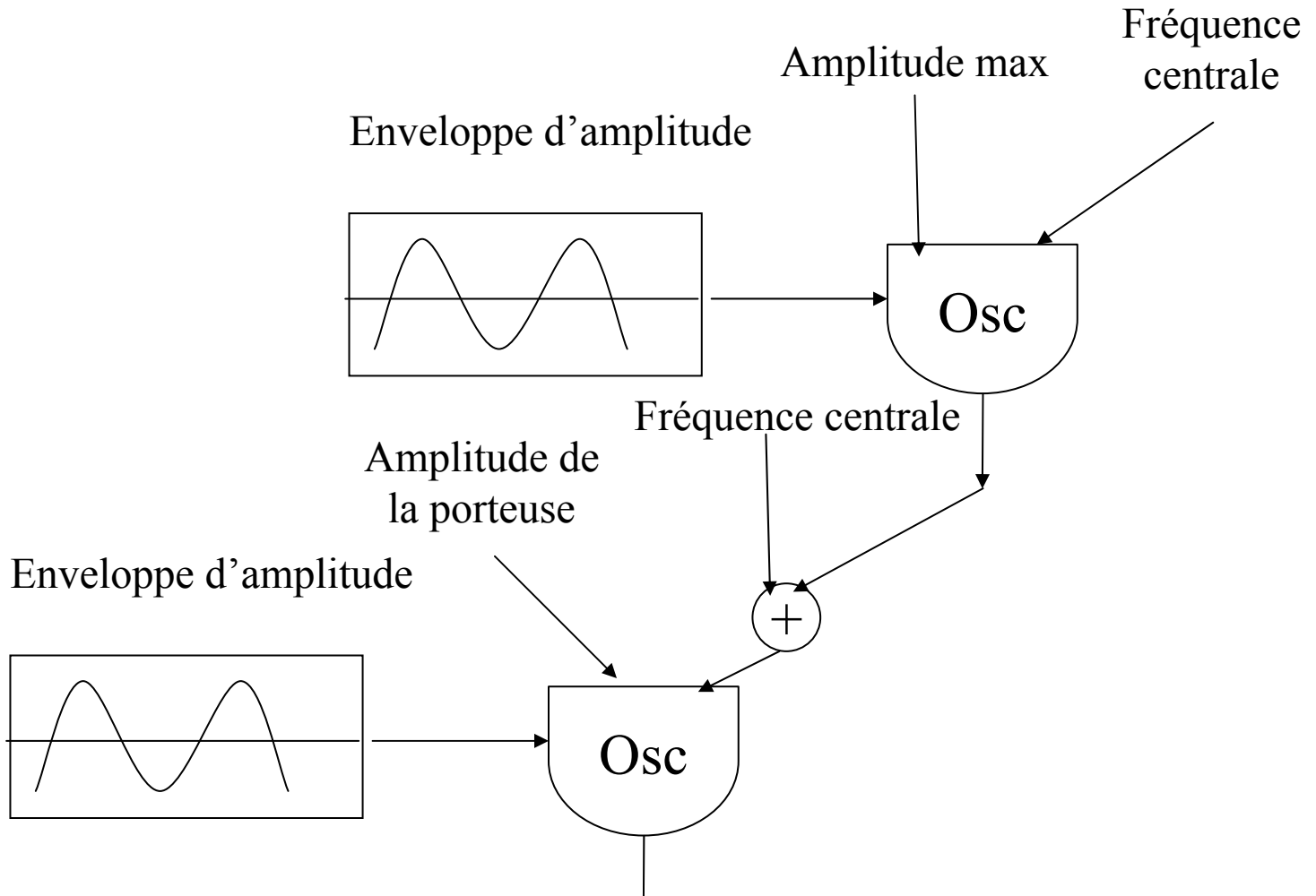
$S(t) = A(t) \cdot \sin(\omega \cdot t + \varphi)$  Amplitude

$S(t) = A \cdot \sin(\omega \cdot t + \varphi(t))$  Phase

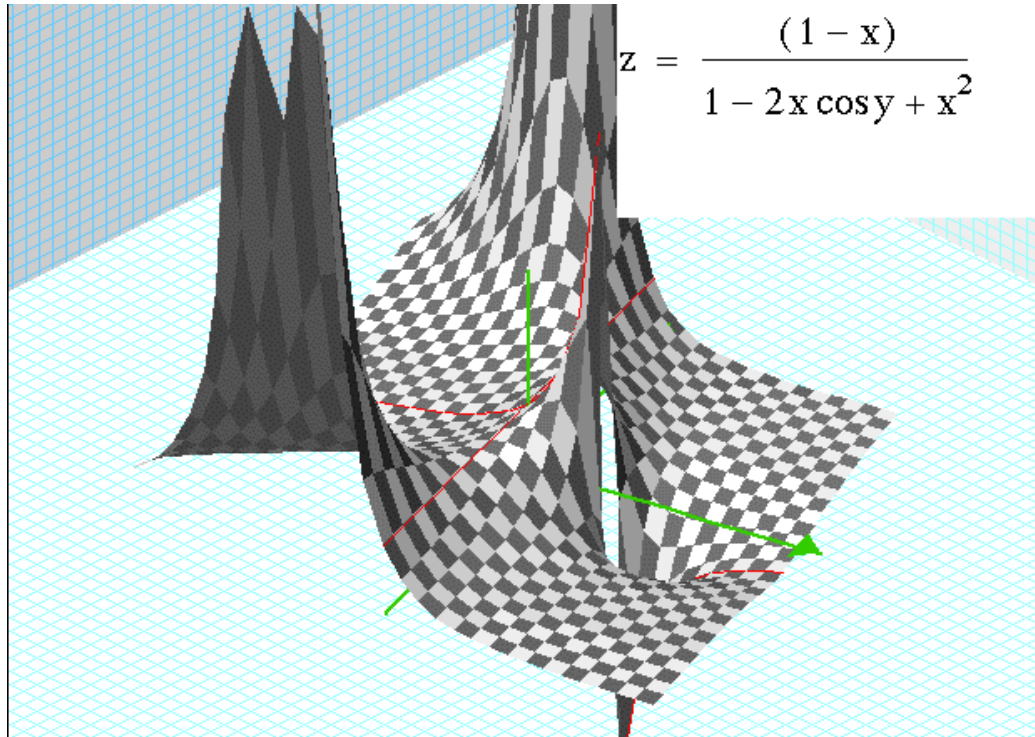
$S(t) = A \sin(\omega(t) \cdot t + \varphi)$  Fréquence

# Modulation de fréquences

## Schéma de base



# Utilisation de fonction complexes



$$\frac{1-x}{1-2x \cos(\theta) + x^2} = 1 + 2x \cos(\theta) + 2x^2 \cos(2\theta) + \dots$$

# Bibliographie

- L'audionumérique, Curtis Raod, Dunod 1998
- The technology of video and audio streaming, David Auterberry, Focal Press, 2002
- Le son musical, John Pierce, Pour la Science, Belin 1984