Contexte industriel autour du TSA
Transducteurs électroacoustiques et fonctions analogiques
Psychoacoustique et application au codage avec perte
Traitements du signal fréquentiel
Traitements audionumériques
Architectures matérielles pour le TSA

#### Audio

#### S. Reynal

ENSEA
Laboratoire ETIS, UMR8051 CNRS/UCP/ENSEA
+ Institut ACTES, Univ. Paris 1 Panthéon-Sorbonne

2021-2022

# Plan du cours (20h CM + 16h TP)

- Transducteurs électroacoustiques et fonctions analogiques (3h)
- Psychoacoustique et application au codage MP3 (2h)
- Outils du traitement du signal fréquentiel : TF pour le signal musical, Vocoder, reconstruction (2h)
- Traitements audionumériques : égalisation, dynamique, spatialisation (3h)
- Architectures matérielles : CODEC, I2S, ALSA, Jack,... (2h)
- Deep Learning pour l'audio : Séparation drum / mélodie / harmonie, beat tracking... (4h)
- Indexation, descripteurs audio, classification et DL pour la MIR (4h)
- TP : audio temps réel sur STM32 Discovery ("mini-projet" 16h)

Contexte industriel autour du TSA
Transducteurs électroacoustiques et fonctions analogiques
Psychoacoustique et application au codage avec perte
Traitements du signal fréquentiel
Traitements audionumériques
Architectures matérielles pour le TSA

#### Plan

- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
  - Microphones, lignes et préamplification
  - Haut-parleurs et étages de puissance
- 3 Psychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application: time stretching par phase vocoder
- Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



Contexte industriel autour du TSA
Transducteurs électroacoustiques et fonctions analogiques
Psychoacoustique et application au codage avec perte
Traitements du signal fréquentiel
Traitements audionumériques
Architectures matérielles pour le TSA

# TSA pour les télécommunications, l'automobile, l'aéronautique

• UE: Thalès, Orange Labs (Rennes), SAGEM, ST, Nokia

• USA/JP : Motorola, Samsung

Automobile : Valeo, Kurtosis

Aéronautique : Dassault (Istres)

Architectures matérielles pour le TSA

## Editeurs et studio jeu video

- Ubisoft (UE, Canada)
- Unity (USA)
- Rockstar North : GTA (UK)
- Electronic Arts (USA)
- Activision : Guitar Hero+Call of duty (US)
- Nitendo (JP), Netease (CHINE)
- Réalité virtuelle/augmentée : perception sonore 3D (Oculus, USA)

#### Production musicale, post-prod cinéma, broadcast

#### Logiciel:

- Ableton (Berlin): Live, MaxMSP
- Steinberg (All, UE): cubase, nuendo
- Apple/Logic
- Avid/Protools
- Adobe Premiere (USA)

#### Matériel :

- Behringer, Seenheiser, Neumann, AKG (UE)
- Mackie, XLogic, Presonus, Alesis (USA)
- Audio Technica (JP)



# Hi-Fi, microphones, monitoring

- Deviallet (UE) : amplis
- Cabasse (UE): amplis, enceintes actives
- Edison (UE) : amplis, enceintes
- Philips (UE): casques noiseless, composants
- Genelec (Finlande, UE) : enceintes
- JBL (USA) : enceintes, sono
- Adam (All, UE) : enceintes, monitoring
- Dynaudio (Dan, UE): micros
- AKG: micros



# TSA et MIR pour la musique, la performance, etc

- Public: IRCAM, CIANT (UE), Berkeley (USA)
- Privé: Arturia (UE), Moog, Dave Smith Instruments (USA), Korg, Boss, Roland, Yamaha (JP)
- Fondeurs de CODEC, Analog ICs et DSP : Coolaudio, Curtis, Analog Device, Texas Instruments, ST Microelectronics,...
- Apps et IOT : enormément de startups à Paris-St-Lazare (ex : Lick / Mixfacer DJIT)

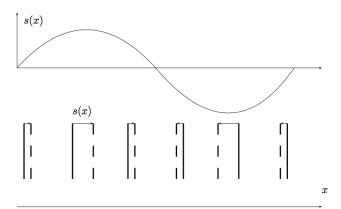
Rappels d'acoustique physique : rayonnement, propagatior Microphones, lignes et préamplification Haut-parleurs et étages de puissance

#### Plan

- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
  - Microphones, lignes et préamplification
  - Haut-parleurs et étages de puissance
- Sychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application: time stretching par phase vocoder
- Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



#### Le son



#### Rappels d'acoustique physique : rayonnement, propagation Microphones, lignes et préamplification Haut-parleurs et étages de puissance

$$v(x,t) = \frac{\partial s(x,t)}{\partial t}$$

$$P(x,t) = P_0 + p_{ac}(x,t)$$

$$N_{dB,SPL} = 20 \log \frac{p_{ac}}{p_0}$$

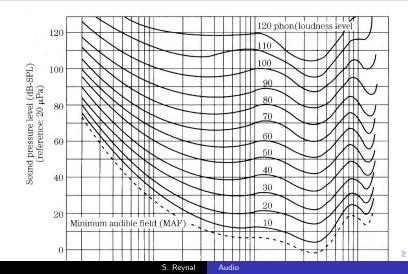
Seuil d'audition :

$$p_0 = 20 \mu Pa$$

Seuil de douleur :

$$p = 20Pa$$

#### Perception de l'intensité : courbe de Robison-Dadson



#### Formules essentielles

Equation d'Euler :

$$\rho_0 \frac{d\vec{v}}{dt} = -\vec{\mathsf{grad}} p$$

Impédance acoustique :

$$\underline{Z_a} = \frac{\underline{p}}{\underline{v}}$$

Rayonnement du monopole :

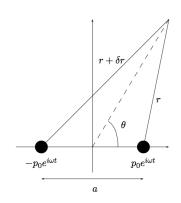
$$p(r,t) = p_0 \frac{e^{j(\omega t - kr)}}{kr}$$

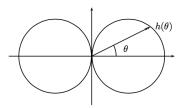
Puissance acoustique rayonné par le monopole :

$$\mathcal{P}_{ac} \approx 33k^2q_a^2(W)$$

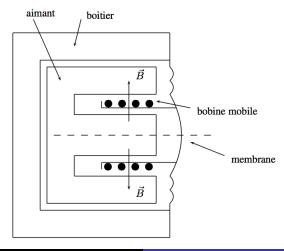


## Rayonnement et directivité du dipôle

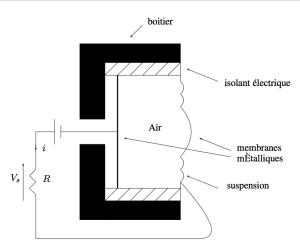




## Microphone électrodynamique

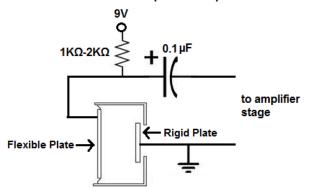


#### Microphone à condensateur

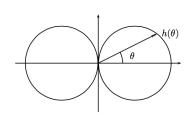


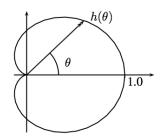
#### Microphone à condensateur : circuit de polarisation

#### Condenser Microphone Setup

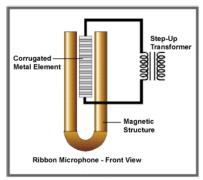


## Diagramme cardioïde

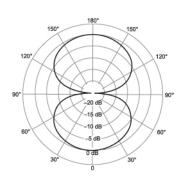




#### Microphone à ruban

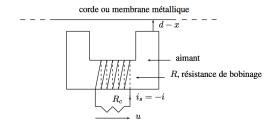


How a ribbon microphone works

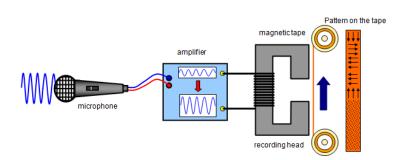


Bi-directional polar or pickup pattern

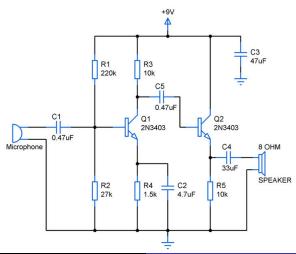
## Microphone à reluctance variable



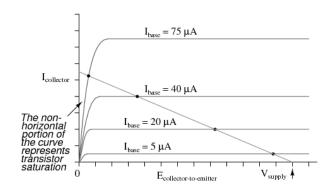
#### Tape recorder



## "Preamp stage" à transistor

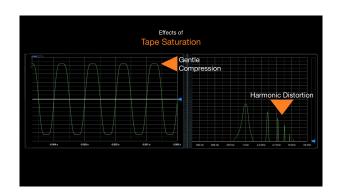


#### Pourquoi ça sature ?



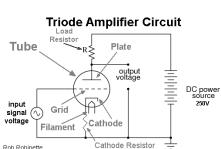
#### Conséquence sur le spectre harmonique

Architectures matérielles pour le TSA

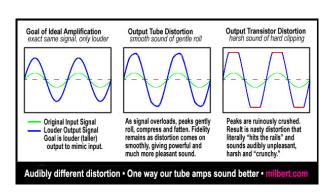


#### Preamp à tubes

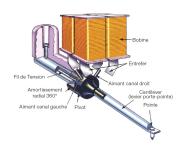


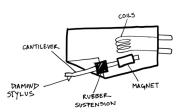


#### Comparaison saturation transistor vs tube



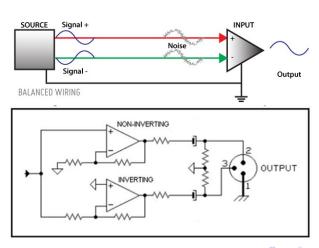
#### Cellule phono



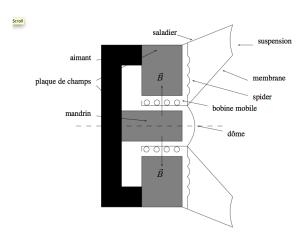


a

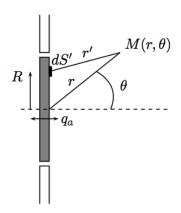
## Lignes symétrisées



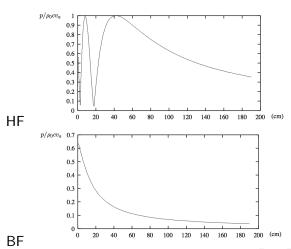
#### Haut-parleur électrodynamique



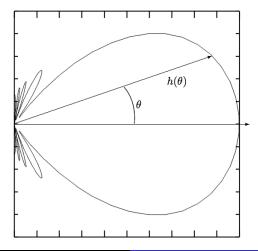
#### Rayonnement du haut-parleur



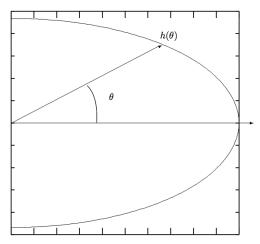
#### Pression sur l'axe



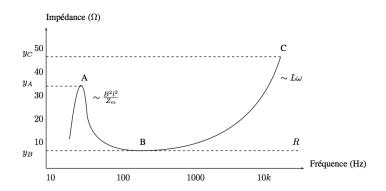
#### Diagramme de rayonnement à 1m en HF



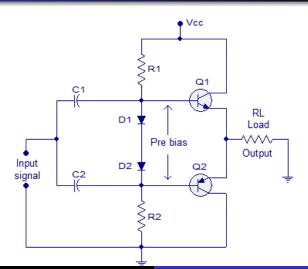
## Diagramme de rayonnement à 1m en BF



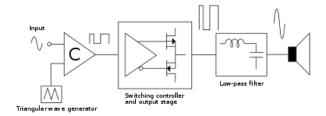
#### Impédance électrique du haut-parleur électrodynamique



# Etage de puissance classe B ("push-pull")



## Etage de puissance classe D



#### Plan

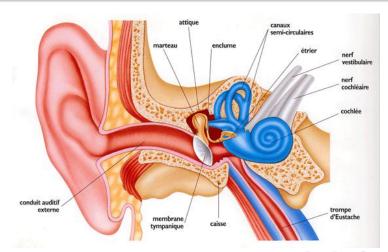
- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
    - Microphones, lignes et préamplification
    - Haut-parleurs et étages de puissance
- Sychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application: time stretching par phase vocoder
- Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



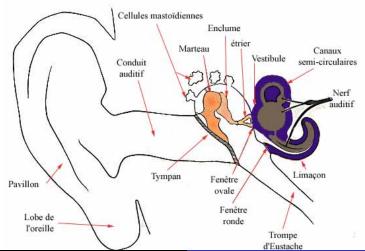
## Objectifs du codage psychoacoustique

- Encodage physique vs psychoacoustique
- Identifier le contenu imperceptible du signal
- Réduire le nbre de bits / optimiser l'allocation de ressources
- Cacher le bruit de quantification sous le seuil d'écoute
- Q : combien peut-on ajouter de bruit sans que cela s'entende ?
- PCM : gaspillage de bits...

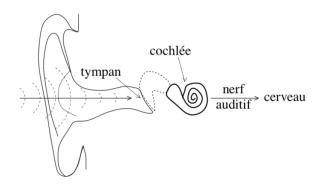
### Anatomie simplifiée de l'oreille



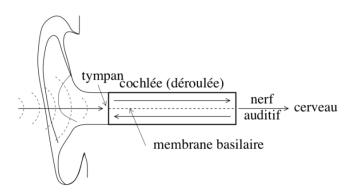
#### Anatomie simplifiée de l'oreille : détail de la cochlée



#### Anatomie simplifiée de l'oreille : nerf auditif



#### Anatomie simplifiée de l'oreille : membrane basilaire



#### Anatomie de l'oreille

- Pavillon : localisation verticale de la source
- Canal externe : résonance à 3 kHz (parole)
- Tympan : excursion maxi à 120 dB SPL
- 3 os : marteau enclume étrier → adaptation d'impédance
- Canal vestibulaire → sens de l'équilibre
- Cochlée = membrane basilaire enroulée en colimaçon  $\rightarrow$  mesure amplitude/fréquence.

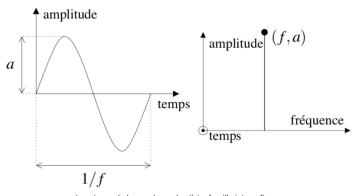
#### Cochlée

#### Estimation du pitch !!!

- Vibrations internes (fluide basilaire) transmises au nerf auditif (paquet de nerfs/neurones)
- 30000 cils le long de la membrane basilaire
- Chaque point de la membrane = une fréquence
- Position du maximum des ondes stationnaires dans la cochlée

   → fréquence (méthode 1)
- Stimulation de groupes de neurones → fréquence (méthode 2)

#### Cochlée



(au niveau de la membrane basilaire [oreille interne])

#### Echelles perceptives

#### Paramètres du son :

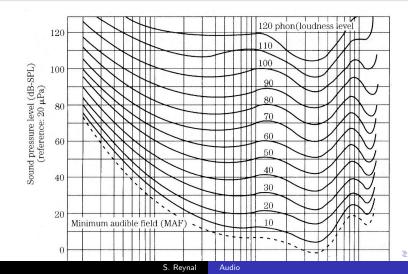
- amplitude et fréquence sont des grandeurs physiques, mesurables
- les paramètres perceptifs correspondants sont le volume et la hauteur (pitch)
- pour les psychoacousticiens : sonie et tonie

Loi de Fechner (s'applique à TOUT organe sensoriel) : la sensation est proportionnelle au logarithme de l'excitation.

## Perception de l'intensité

- Logarithmique (dB<sub>SPL</sub>)
- Sensibilité max @ 1-5kHz
- Equal-loudness curves (Robison-Dadson)
- Hardware : attention, il faut des potentiomètres étalonnés en log ! (aka "pot B") ; idem compresseurs/expandeurs

#### Perception de l'intensité : courbe de Robison-Dadson



# Perception fréquentielle - "pitch"

- Perception fréquentielle logarithmique ! (aka  $f_2/f_1$  plutôt que  $f_2-f_1$ )
- Importance des octaves  $(f_2/f_1 = 2)$  et quintes (3/2)
- Son graves : ça dépend aussi de l'amplitude !
- Influence : construction de toute la musique via le système de gammes !
- Attention : Fréquence ≠ hauteur tonale (pitch)

### Perception spatiale

- Comment localise-t-on ? Différence de phase G/D
- Localisation en azimuth : plus compliqué (filtrage créé par le pavillon auditif dépend de l'azimuth de la tête) ; cf. casques de VR
- Exemple d'illusion sonore : 2 mélodies asc./descendante jouées sur canaux G & D

### Bandes critiques

- Basse fréquence : signaux séparables si f2-f1 vaut quelques Hz
- Haute fréquence : quelques centaines de Hz
- Les cellules cilées répondent au plus fort signal de leur région : bande critique.
- 100Hz de largeur entre 20-400 Hz
- 1/5 octave pour 1-7kHz

# Echelle de Bark(hausen)

- les fréquences audibles vont approx. de 20Hz à 22kHz
- échelle beaucoup plus proche de la perception que l'échelle linéaire en Hz
- 1 Bark = 100 mels

$$B(F) = F/100 \text{ si } F < 500 Hz$$
  
 $B(F) = 9 + 4 \log_2(F/1000) \text{si si } F > 500 Hz$ 

#### Modèle de Zwicker

#### Largeur de bande exactement de 1 bark.

| numéro | début | centre | fin  |
|--------|-------|--------|------|
| 1      | 20    | 50     | 100  |
| 2      | 100   | 150    | 200  |
| 3      | 200   | 250    | 300  |
| 4      | 300   | 350    | 400  |
| 5      | 400   | 450    | 510  |
| 6      | 510   | 570    | 630  |
| 7      | 630   | 700    | 770  |
| 8      | 770   | 840    | 920  |
| 9      | 920   | 1000   | 1080 |
| 10     | 1080  | 1170   | 1270 |
| 11     | 1270  | 1370   | 1480 |
| 12     | 1480  | 1600   | 1720 |

| numéro | début | centre | fin   |
|--------|-------|--------|-------|
| 13     | 1720  | 1850   | 2000  |
| 14     | 2000  | 2150   | 2320  |
| 15     | 2320  | 2500   | 2700  |
| 16     | 2700  | 2900   | 3150  |
| 17     | 3150  | 3400   | 3700  |
| 18     | 3700  | 4000   | 4400  |
| 19     | 4400  | 4800   | 5300  |
| 20     | 5300  | 5800   | 6400  |
| 21     | 6400  | 7000   | 7700  |
| 22     | 7700  | 8500   | 9500  |
| 23     | 9500  | 10500  | 12000 |
| 24     | 12000 | 13500  | 15500 |

(valeurs des fréquences en Hz)

### Seuillage et masquage

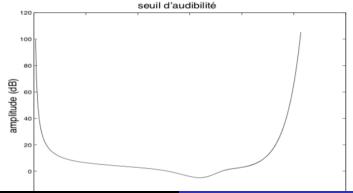
Deux phénomènes fondamentaux dans l'écoute :

- seuil d'écoute (0 sone) → dépend de la bande de fréquence
- masquage par amplitude → modification dynamique et temporaire du seuil par un son masquant (de la même bande critique ou d'une autre bande)

# Seuillage d'audibilité

- les volumes audibles vont approximativement de 0dB à 120dB
- le seuil de sensibilité  $s_a(f)$  dépend de la fréquence

$$s_a(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4$$



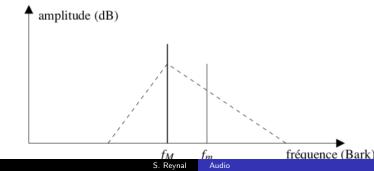
## Phénomène de masquage

- physiquement l'addition de signaux identiques est régie par une loi d'addition linéaire qui donne 6dB (amplitude doublée)
- perceptuellement, le seuil de perception est modifié pour un son m (masqué) quand il est joué en même temps qu'un son plus fort M (masquant)

## Masquage fréquentiel (simultané)

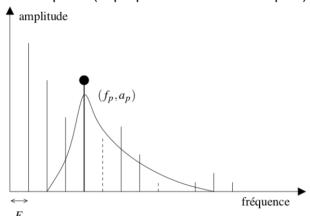
Cas où M et m sont deux sinusoïdes (sons purs/simples):

- de fréquences respectives  $f_M$  et  $f_m$
- d'amplitudes respectives  $a_M$  et  $a_m$   $(a_M > a_m)$
- si  $f_m$  est proche de  $f_M$ , le son m est masqué par le son M et peu devenir inaudible.



# Masquage fréquentiel (cas d'un son harmonique complexe)

#### son complexe (superposition de sons simples)



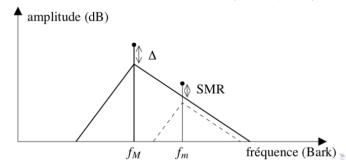
### Non-linéarités de l'oreille liées à la fréquence

- ullet des battements peuvent être perçus quand  $f_mpprox nf_M$
- les deux sont purs peuvent interagir pour créer perceptuellement des fréquences  $if_m \pm jf_M$ , produisant en quelque sorte des harmoniques "virtuelles" pour une sinusoïde pure
- conséquence importante : la fondamentale virtuelle est perçue même si elle est physiquement absente. Exemple : partiels à 200Hz, 300Hz et 400Hz. Fondamentale perçue à 100Hz.

## Modèle du triangle de masquage

Evaluation du rapport signal/masque (SMR) de chaque partiel. Le seuil de masquage est quasiment un triangle en échelles Bark-dB :

- différence  $\Delta$  entre vol. partiel masquant et masque (-10dB)
- demi-triangle de masquage vers graves (27 dB/bark)
- demi-triangle de masquage vers aigus ( -15dB/bark)



### Masquages temporels

#### Post-masquage:

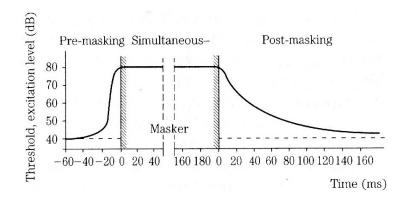
- survient quand le son masquant disparaît
- l'effet du masquage fréquentiel persiste alors, en s'estompant
- pendant plusieurs ms après cette disparition

#### Pré-masquage:

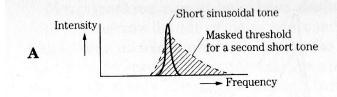
- plus surprenant...
- l'effet de masquage est actif quelques ms AVANT que le son n'apparaisse vraiment
- phénomène toutefois beaucoup moins prononcé

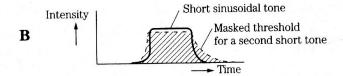


#### Masquage temporel

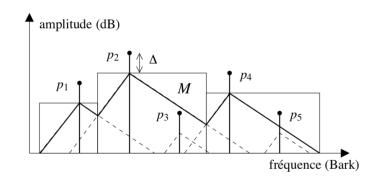


#### Masquage temporel





## Exemple de masque fréquentiel



## Exemple de masque fréquentiel

5 partiels et le masque associé M (en gras) :

- $p_1$ ,  $p_2$  et  $p_4$  sont des partiels masquants et contribuent au masque M
- p<sub>5</sub> n'est ni masquant, ni masqué
- $p_3$  est masqué (par  $p_2$ )

#### Algorithme de construction du masque

Initialement, masque mis à zéro.

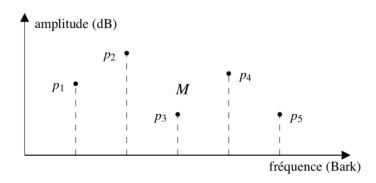
Puis parcourir partiels par amplitudes décroissantes.

Pour chaque partiel p de fréquence  $f_p$  et d'amplitude  $a_p$  (volume  $V(a_p)$  en dB) :

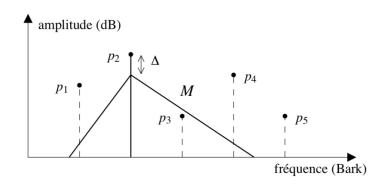
- si  $M(f_p) + \Delta < V(a_p)$  alors p est un partiel masquant et M doit être mis à jour avec sa contribution
- si  $M(f_p) < V(a_p) < M(f_p) + \Delta$  alors p n'est ni masquant ni masqué
- si  $V(a_p) < M(f_p)$  alors p est simplement masqué

On met à jour M itérativement, en conservant la plus grande valeur entre le triangle de masquage associé au partiel courant p et l'ancienne valeur de M.

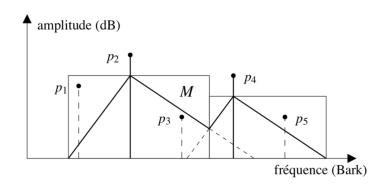
## Exemple de construction du masque (0/5)



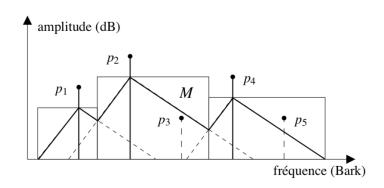
## Exemple de construction du masque (1/5)



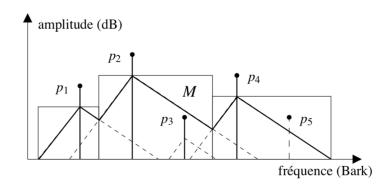
## Exemple de construction du masque (2/5)



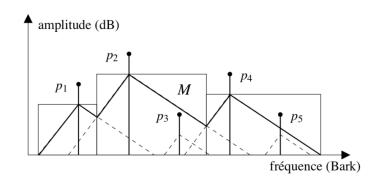
# Exemple de construction du masque (3/5)



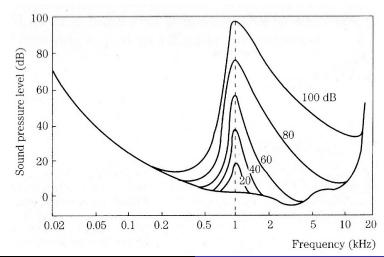
## Exemple de construction du masque (4/5)



# Exemple de construction du masque (5/5)



## Modèle de masquage fréquentiel : modèle avancé



# Codeur psycho-acoustique : vision "théorie de l'information"

- L'oreille ne perçoit qu'une fraction de l'information audiophonique : entropie !
- Signaux à entropie variable
- Supprimer les fractions non-pertinentes du signal = à faible entropie
- Encore mieux : VBR vs CBR !

## Débit des flux

| Bits/sample | Comp ratio | 48kHz    | 44.1kHz   |
|-------------|------------|----------|-----------|
| 16          | 1:1        | 768kbps  | 705.6kbps |
| 8           | 2:1        | 384kbps  | 352.8kbps |
| 4           | 4:1        | 192kbps  | 176kbps   |
| 2.67        | 6:1        | 128kbps  | 117.7kbps |
| 2           | 8:1        | 96kbps   | 88.2kbps  |
| 1.45        | 11:1       | 69.6kbps | 64kbps    |

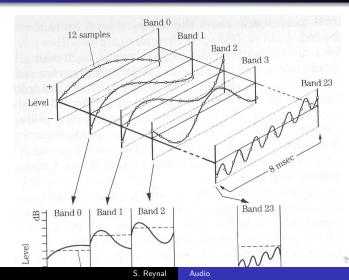
#### Allocation de ressources

- Allocation "forward": à l'encodage (allocation fine et précise, modèle psycho dans l'encodeur, décodeur léger, mais une partie du stream utilisée pour les informations d'allocation)
- Allocation "backward": au décodage (à partir du flux, complexe)

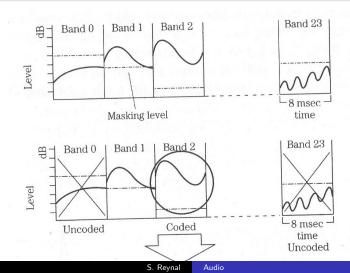
#### Tolérance aux erreurs

- PCM : erreur = click audible (bruit large bande)
- MPEG : erreur limité à une bande critique donnée (erreur = impulsion bande étroite)
- Correction d'erreur ciblée : pianissimo vs forte
- Attention à la mise en cascade !

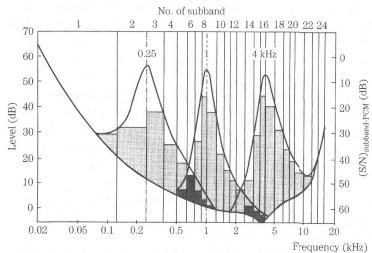
## Codage/décodage par bande



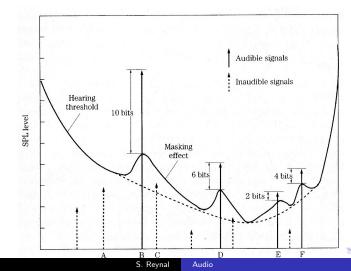
## Codage/décodage par bande



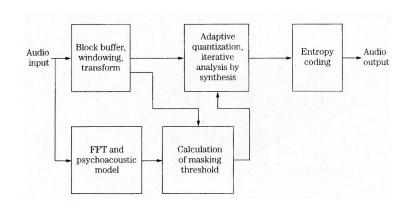
#### Allocation de bits



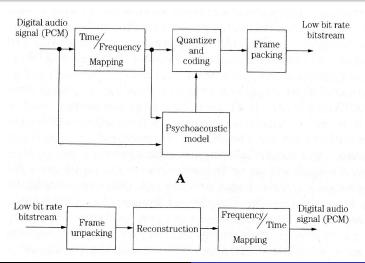
#### Allocation de bits

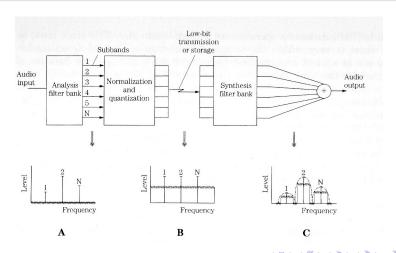


## Codage par transformée



#### Architecture du codeur



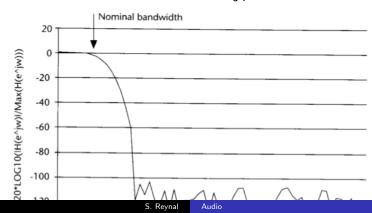


- Trame de 512 échantillons : x[n], n = 0...511
- $oldsymbol{t} t 
  ightarrow t+1$  : on fait "entrer" 32 nouveaux échantillons dans la trame
- Banc de filtre = 32 filtres indexés par i = 0...31
- Calcul des échantillons en sortie pour le filtre *i* :

$$s_t[i] = \sum_{n=0}^{511} x[t-n]H_i[n]$$

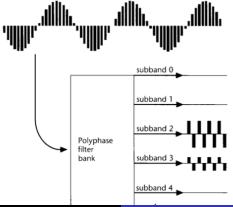
Réponse impulsionnelle du filtre i :

$$H_i[n] = h[n] \cos \frac{(2i+1)(n-16)\pi}{64}$$



#### Chevauchement!

Input audio: 1,500-Hz sine wave sampled at 32 kHz, 64 of 256 samples shown



Audio

Algorithme optimisé :

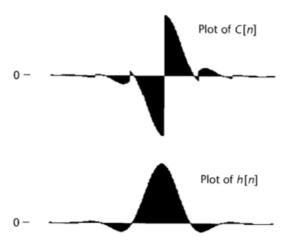
$$s_t[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] (C[k+64j]x[k+64j])$$

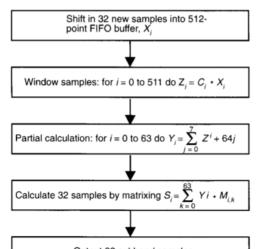
soit encore

$$s_t[i] = \sum_{k=0}^{63} M[i][k]Y[k]$$

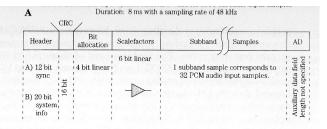
avec

$$Y[k] = \sum_{i=0}^{7} (C[k + 64j]x[k + 64j])$$

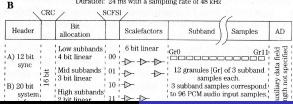




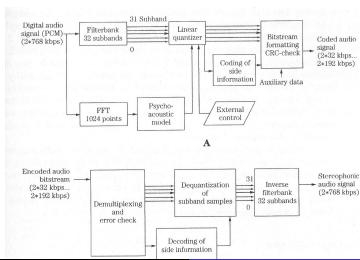
#### Norme MPEG1



ISO/MPEG/AUDIO layer II frame structure: valid for 1152 PCM audio input samples Duration: 24 ms with a sampling rate of 48 kHz



## Architecture MPEG1



#### Plan

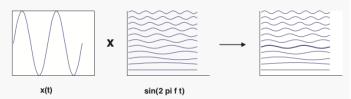
- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
  - Microphones, lignes et préamplification
  - Haut-parleurs et étages de puissance
- Sychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application : time stretching par phase vocoder
- Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



#### Transformée de Fourier continue

$$X(\omega) = \int_{t=-\inf}^{+\inf} x(t)e^{-j\omega t}dt \quad X(t) = \int_{t=-\inf}^{+\inf} \exp(-j2\pi ft)dt$$

- Variables :
  - t est le temps
  - $\omega = 2\pi f$  les fréquences continues exprimées en radian,
  - $\exp(j2\pi ft) = \cos(2\pi ft) + j \cdot \sin(2\pi ft)$ .
- Pourquoi la Transformée de Fourier?
  - Difficile d'extraire des observations directement à partir de la forme d'onde x(t)
  - Reproduire la décomposition en fréquences de l'oreille humaine



# Propriétés de la TF

| Propriétés  | x(t)                     | X(f)                            |
|-------------|--------------------------|---------------------------------|
| Similitude  | x(at)                    | $\frac{1}{ a }X(\frac{f}{ a })$ |
| Linéarité   | ax(t) + by(t)            | aX(f) + bY(f)                   |
| Translation | $x(t-t_0)$               | $X(f) \exp(-j2\pi f t_0)$       |
| Modulation  | $x(t) \exp(j2\pi f_0 t)$ | X(f-f0)                         |
| Convolution | $x(t) \circledast y(t)$  | X(f)Y(f)                        |
| Produit     | x(t)y(t)                 | $X(f) \circledast Y(f)$         |
| Parité      | réelle paire             | réelle paire                    |
|             | réelle impaire           | imaginaire paire                |
|             | imaginaire paire         | imaginaire paire                |
|             | imaginaire impaire       | réelle impaire                  |
|             | complexe paire           | complexe paire                  |
|             | complexe impaire         | complexe impaire                |
|             | réelle                   | $X(f) = X^{\star}(-f)$          |
|             |                          | $\Re(X(f))$ est paire           |
|             |                          | $\Im(X(f))$ est impaire         |
|             | $x^{\star}(t)$           | $X^{\star}(f)$                  |

#### Transformée de Fourier discrète

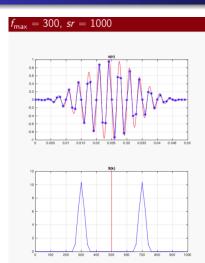
$$X(k) = \sum_{m=0}^{N-1} x(m) e^{-j2\pi \frac{k}{N}m} \quad \forall k \in [0, N]$$

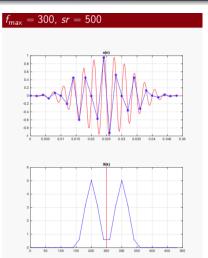
- Variables :
  - m le numéro d'échantillon
  - k les fréquences discrètes
- Fréquence d'échantillonnage (sampling rate) sr
  - sr définit à quelle fréquence le signal temporel va être échantillonné
  - Exemple :
    - Compact Disc sr = 44100 Hz
    - La distance temporelle entre deux échantillons (le pas d'échantillonnage) est de  $\Delta t = \frac{1}{44100} = 0.000023 \ s.$
- sr doit être > à deux fois la  $f_{max}$  présente dans le signal
  - Sinon : repliement spectral
    - exemple : captation d'une roue d'une voiture accélérant dans les films
  - Fréquence de Nyquist :  $f_{Nyquist} = \frac{sr}{2} > f_{max}$



Transformées de Fourier pour l'audio Application : time stretching par phase vocoder

## Transformée de Fourier discrète : repliement





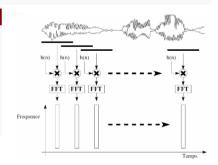
## Transformée de Fourier à Court Terme (TFCT)

## $X(k,n) = \sum_{m=0}^{N-1} x(m) w(n-m) e^{-j2\pi \frac{k}{N}m} \quad \forall k \in [0, N]$

• Application de la TFD à une portion du signal centrée autour de l'échantillon n

#### Pourquoi la TFCT?

- Signal audio = non-stationnaire
  - ses propriétés varient au cours du temps
- Stationnaires "localement" (en temps)
  - sur une durée de  $\pm$  40ms
- TFCT = suite d'analyses de Fourier sur des durées de + 40ms
  - = analyse à Court Terme ("trames/frames" en vidéo)



## Transformée de Fourier à Court Terme (TFCT)

$$X(k,n) = \sum_{m=0}^{N-1} x(m)w(n-m)e^{-j2\pi \frac{k}{N}m} \quad \forall k \in [0,N]$$

#### Fenêtre de pondération w(t)

- $x(t) \cdot w(t) \rightleftharpoons X(f) \circledast W(f)$ 
  - w(t) est appelé "fenêtre de pondération"
  - w(t) différents types de fenêtre
  - w(t) définie sur un horizon fini (longueur temporelle) [0, L].
  - Choix du type et de la longueur détermine les caractéristiques spectrales
    - Largeur de bande fréquentielle (à -6 $dB_{20}$ ) :  $Bw = \frac{Cw}{I}$
    - Hauteur des lobes secondaires

# TFCT : types de fenêtre

- rectangulaire
  - w(n) = 1
  - Bw = 1.21
- hanning

• 
$$w(n) = 0.5(1 - \cos(\frac{2\pi n}{N-1}))$$

• 
$$Bw = 2$$

hamming

• 
$$w(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1})$$

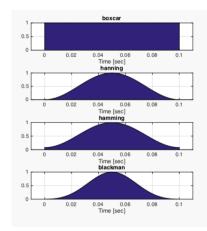
• 
$$Bw = 1.81$$

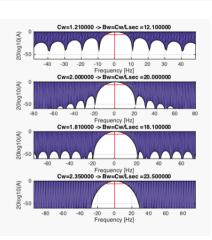
blackman

• 
$$w(n) = a_0 - a_1 \cos(\frac{2\pi n}{N-1}) + a_2 \cos(\frac{2\pi n}{N-1})$$

• 
$$Bw = 2.35$$

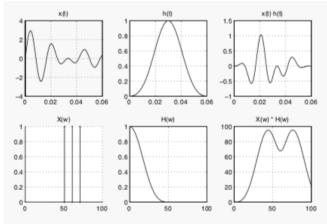
#### TFCT : influence de la fenêtre



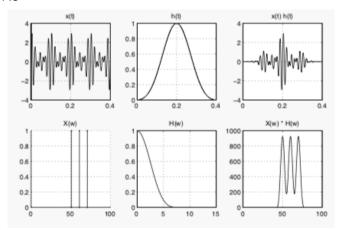


- Fenêtre courte : bonne précision temporelle
- Fenêtre longue : bonne précision fréquentielle

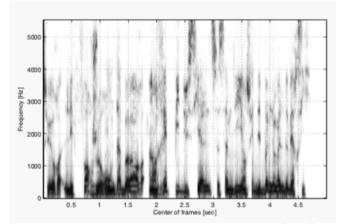
L = 0.06s



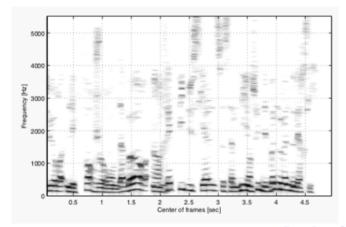
L = 0.4s



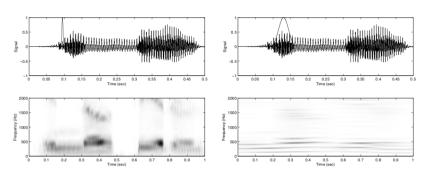
L = 0.01s



L = 0.1s



Impossible d'obtenir simultanément une bonne localisation temporelle et fréquentielle !



Solution: utiliser d'autres TF!



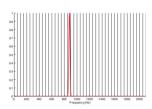
## TFCT: exercice

**Exercice**: on réalise une TF sur N = 1024 points à la fréquence d'échantillonnage  $s_r = 48000 \text{ kHz}$  avec une fenêtre de Hanning.

- Calculer la précision fréquentielle (=pas d'échantillonnage du spectre)
- Calculer la résolution fréquentielle (=pouvoir de séparation de deux fréquences proches)
- Ces deux paramètres sont-ils constants à travers les fréquences ?

#### Pourquoi la TFQ?

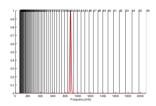
- En audio musical
  - les fréquences sont logarithimiquement espacées
    - pour passer des fréquences aux hauteurs de notes :  $m_k = 12 \cdot log_2 \frac{f_k}{dAO} + 69$
    - pour passer des hauteurs de notes aux fréquences :  $f = 440 \cdot 2^{\frac{m-69}{12}}$
  - les hauteurs de notes sont plus rapprochées en basses fréquences, plus espacées en hautes fréquences



Espacement linéaire de la DFT

#### Pourquoi la TFQ?

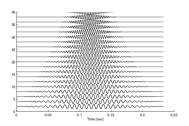
- La résolution fréquentielle de la DFT
  - n'est pas suffisante pour résoudre les hauteurs de notes adjacentes en basses fréquences,
  - est trop importante en hautes fréquences



Espacement logarithmique des hauteurs de notes

#### Principe:

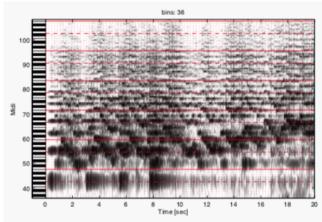
- Solution?
  - Changer la résolution fréquentielle en fonction des fréquences considérées
- Comment?
  - En changeant la longueur temporelle de la fenêtre pour chaque fréquence considérée
  - Le facteur  $Q = \frac{f_k}{f_{k+1} f_k}$  doit rester constant en fréquence



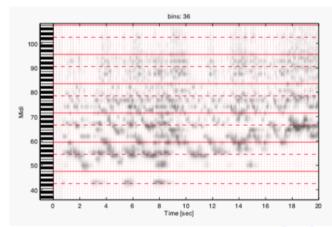
#### Exercice:

- Déterminer la relation entre  $L_k$ , Q,  $C_w$  et  $f_k$  pour que Q soit constant
- En déduire l'algorithme permettant de calculer la TF discrète à Q constant.

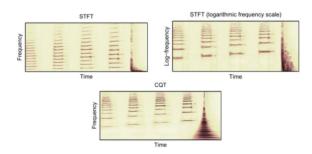
Exemple: DFT



Exemple: CQT



Sur une CQT, une différence de pitch correspond à une translation verticale!



# Comment changer le tempo de morceau?

#### Technique de DJ pour changer le tempo

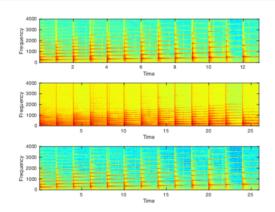
- Ralentir/accélérer la vitesse de lecture (du vynil, de la bande magnétique)
- $x(\alpha t) \rightleftharpoons \frac{1}{\alpha} X\left(\frac{f}{|\alpha|}\right)$
- ullet Ralentir le temps : lpha < 1
  - mais contracte aussi les fréquences (on abaisse les hauteurs)
- Accélérer le temps :  $\alpha > 1$ 
  - mais étend aussi les fréquences (on augmente les hauteurs)

#### Objectif

 Changer le temps et les hauteurs de manière indépendante

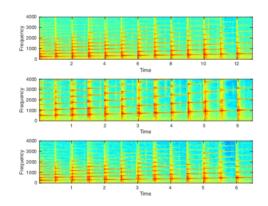


### Comment changer le tempo de morceau ?



[haut] : signal original, [milieu]  $\alpha<$  1 par ré-échantillonnage, [bas] :  $\alpha<$  1 par vocodeur de phase

### Comment changer le tempo de morceau ?



[haut] : signal original, [milieu]  $\alpha>$  1 par ré-échantillonnage, [bas] :  $\alpha>$  1 par vocodeur de phase

#### Phase vocoder

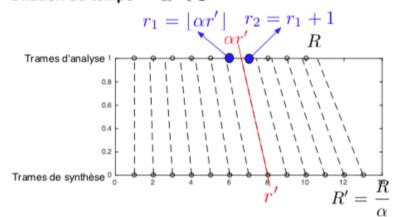
#### Le vocodeur de phase

- Méthode pour rallonger/raccourcir le signal :
  - changer le nombre de trames utilisées pour la resynthèse par TFCT<sup>-1</sup>
- Soit R :
  - le nombre de trames d'analyse de la TFCT
- Soit  $R' = \frac{R}{R}$ :
  - le nombre de trames de synthèse (utilisées pour la resynthèse par TFCT<sup>-1</sup>)

- $\alpha < 1 \rightarrow$  on dilate (ralentit) le temps
- $\bullet \ \alpha > 1 \rightarrow {\rm on~compresse}$  (accélère) le temps
- Le contenu d'une trame de synthèse  $r' \in [1,R'=\frac{R}{\alpha}]$  est obtenu en recherchant les trames d'analyse r correspondantes les plus proches
  - $r_1 = \lfloor \alpha r' \rfloor$  et
  - $r_2 = r_1 + 1$

#### Phase vocoder

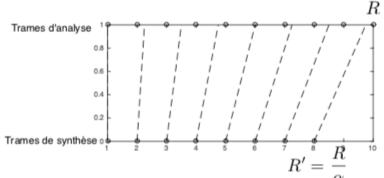
#### Dilation du temps $\alpha < 1$



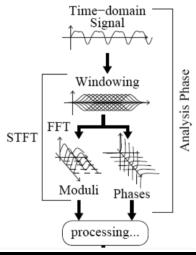


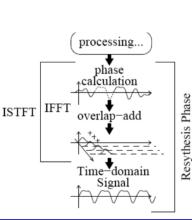
#### Phase vocoder

#### Compression du temps $\alpha > 1$



# Phase vocoder : analyse / synthèse





#### Phase vocoder: reconstruction

### 1) Spectre d'amplitude

- Le spectre d'amplitude à la trame r', est obtenu par interpolation linéaire des spectres d'amplitude en r<sub>1</sub> et r<sub>2</sub> = r<sub>1</sub> + 1 :
  - $A(k, r') = (1 \Delta)A(k, r_1) + \Delta A(k, r_2)$
  - avec  $\Delta = \alpha r' r_1$

### 2) Spectre de phase

c'est plus compliqué!!!



# Phase vocoder : phase et fréquence instantanée

- Considérons un signal formé d'une sinusoide à la fréquence f<sub>0</sub> :
  - $x(t) = \sin(2\pi f_0 t) = \sin(\phi(t))$
- Entre les instants  $t_1$  et  $t_2$ , sa phase a "tournée" de  $\phi(t_1)$  à  $\phi(t_2)$
- Puisqu'il s'agit d'une sinusoide pure, elle a tournée de

• 
$$\phi(t_2) = \phi(t_1) + 2\pi f_0(t_2 - t_1)$$

• On peut donc estimer  $f_0$  à partir de la différence de phase

• 
$$f_0 = \frac{\phi(t_2) - \phi(t_1)}{2\pi(t_2 - t_1)}$$

#### Problème :

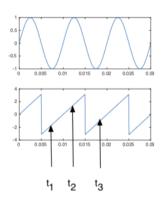
- la phase est uniquement définie dans l'intervalle  $[-\pi,\pi]$
- $\phi(t) \in [-\pi, \pi]$
- ullet donc en pratique le  $\hat{\phi}(t_2)$  qu'on observe n'est pas  $\phi(t_2)$  mais

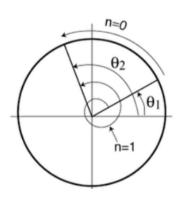
• 
$$\hat{\phi}(t_2) = \phi(t_2) + n2\pi = \phi(t_1) + 2\pi f_0(t_2 - t_1) + n2\pi$$

- avec  $n \in \mathbb{N}$  indéterminé
- pour estimer  $f_0$  il faut donc déterminer n

• 
$$f_0 = \frac{\phi(t_2) + n2\pi - \phi(t_1)}{2\pi(t_2 - t_1)}$$

### Phase vocoder : phase et fréquence instantanée





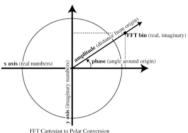
# Phase vocoder: phase dans la TFCT

• Pour chaque trame *n* et fréquence *k* la TFTC est un nombre complexe

• 
$$X(k,n) = \sum_{m} x(m)w(n-m)e^{-j2\pi \frac{k}{N}m}$$

 Il peut se décomposer en amplitude (module) et phase :

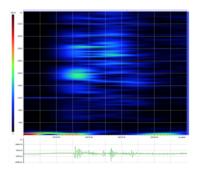
• 
$$X(k, n) = A(k, n)e^{j\phi(k, n)}$$



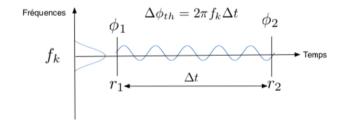
FF I Cartesian to Polar Conversion

# Phase vocoder: phase dans la TFCT

- On a donc une valeur d'amplitude et de phase pour chaque (k, n)
- Spectrogramme
  - d'amplitude A(k, n)
  - de phase  $\phi(k, n)$
- La phase indique la position de la cosinusoide.
- La variation temporelle de phase indique la fréquence instantanée
  - On peut donc calculer une fréquence instantanée pour chaque fréquence k et chaque couple de trames successives (n − 1) → n.



 A la trame r' le spectre de phase dans le filtre k de la TFCT est obtenu en propageant la phase à partir de la fréquence contenu dans ce filtre



- 1) Solution simplifiée :
  - On suppose qu'à travers le filtre k de la TFCT on peut observer uniquement une sinusoide à la fréquence f<sub>k</sub>
  - On propage l'évolution de la phase au cours du temps en utilisant la prédiction théorique de la phase : Δφ<sub>th</sub> = 2πf<sub>k</sub>Δt
  - Donc

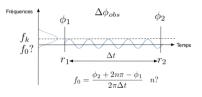
$$\phi(k, r') = \phi(k, r' - 1) + \Delta \phi_{th}$$
$$= \phi(k, r' - 1) + 2\pi f_k \Delta t$$

avec comme phase initiale :

$$\phi(k, r'=1) = \phi(k, r=1)$$

- 2) Solution correcte:
  - En pratique, à travers le filtre k de la TFCT, on peut observer des sinusoides à des fréquences proches mais différentes de f<sub>k</sub>
    - ceci est du à largeur du lobe principale, aux lobes secondaires
  - Il faut estimer cette fréquence f<sub>0</sub> que l'on observe à travers le filtre f<sub>k</sub> pour ensuite appliquer la propagation de phase

• 
$$\phi(k,r') = \phi(k,r'-1) + 2\pi f_0 \Delta t$$



- Estimer cette fréquence fo?
  - En utilisant la **fréquence instantanée** :

• 
$$f_0(n) = \frac{\phi_2 + n2\pi - \phi_1}{2\pi\Delta t}$$
• Déterminer  $n$ ?

• on cherche n tel que  $f_0 \simeq f_k$ 

$$\begin{split} n \text{ tel que } \min_n |f_0 - f_k| \\ \min_n |\frac{\phi_2 + n2\pi - \phi_1}{2\pi\Delta t} - f_k| \\ \min_n |\phi_2 + n2\pi - \phi_1 - 2\pi\Delta t f_k| \\ \min_n |\phi_2 + n2\pi - \phi_1 - \Delta\phi_{th}| \end{split}$$

- ce qui revient à
  - trouver la détermination principale (la valeur dans l'intervalle  $[-\pi,\pi]$ ) de

• n = 
$$[(\phi_2 - \phi_1 - \Delta\phi_{th})/(2\pi)]$$

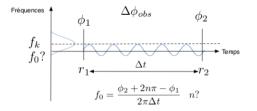
• il s'agit de la différence de phase non-expliquée par le modèle théorique  $\Delta\phi_{th}$ 

- 2) Solution correcte:
  - Finalement la phase est incrémentée de

$$\phi(k, r') = \phi(k, r' - 1) + 2\pi f_0 \Delta t$$
  
=  $\phi(k, r' - 1) + \phi_2 + n2\pi - \phi_1$ 

avec comme phase initiale :

$$\phi(k,r'=1) = \phi(k,r=1)$$



### Phase vocoder: pitch-shifter

• Ré-échantillonnage du signal pour correction de la longueur par phase-vocoder

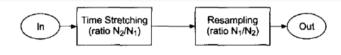


Figure 8.24 Resampling of a time stretching algorithm.

#### Plan

- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
  - Microphones, lignes et préamplification
  - Haut-parleurs et étages de puissance
- 3 Psychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application: time stretching par phase vocoder
- 5 Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



# Dynamique : représentations des échantillons

La plupart des CODEC audio fournissent des signaux quantifiés sur 16, 20 ou 24 bits, soit une dynamique de respectivement 96dB, 120dB et 144dB (rq : provient de  $20\log_{10}2^{Nbits}$ ) Trois représentations des nombres sont utilisées en traitement du signal audio :

- Nombres entiers
- Nombres à virgule fixe
- Nombres à virgule flottante

Calculs typiques : 
$$y[n] = a_0x[n] + a_1x[n-1] - b_0y[n] - b_1y[n-1]$$



### Dynamique : représentations des échantillons

Calculs sur des entiers de 32 bits (aka int, cf STM32 par ex) :

- $N = [x_{31} \dots x_0]$
- x<sub>31</sub> est le bit de signe (complément à 2)
- Range :  $-2^{31}$  à  $2^{31}$
- Dynamique :  $32x6.02 \approx 192dB$

#### Inconvénients:

- problème des étapes intermédiaires de calcul
- une multiplication de deux entiers sur 32 bits donne un entier sur 64 bits: il faut donc tronquer les 32 LSB (ou de manière équivalente, renormaliser en divisant par 2<sup>32</sup>)
- si seuls les 16 LSB des deux nombres sont non-nuls, le résultat fait 0
- une addition de deux nombres séparés de 192dB fait disparaître le plus faible dans le bruit de quantification

# Dynamique : représentations des nombres

#### Nombres à virgule fixe :

- $N = \pm 0.[x_{30} \dots x_0]$
- $x_{31} = \pm$  est le bit de signe (complément à 2)
- Range: -1 à 1
- Une multiplication de deux nombres donne toujours un nombre inférieur à 1
- Même dynamique que les "int" mais renormalisation automatique.
- Par contre, pb pas résolu : si seuls les 16 LSB des deux nombres sont non-nuls, le résultat fait toujours 0



# Dynamique : représentations des nombres

#### Nombres à virgule flottante:

- Représentation :  $N = S [E_7 \dots E_0] [M_{22} \dots M_0]$
- Valeur :  $N = (-1)^S 2^{E-127} (1 + M)$
- Range :  $-1.8 \ 10^{-38} \ \text{à} \ 3.4 \ 10^{38}$
- Dynamique : en théorie 1638dB (multiplication)
- En pratique, si on additionne deux nombres, ils ne peuvent différer de plus de 144*dB*. Exemple :

$$0.1\ 10^9 + 0.8\ 10^3 = 0.1\ 10^9 + 0.0000008\ 10^9 = 0.1000008\ 10^9$$

Risque de troncature dans la mantisse.



# Dithering

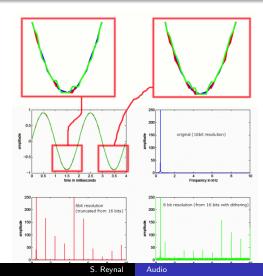
- Problématique : erreurs d'arrondi lorsqu'on passe de "floating point" à "integer".
- Exemple :  $[1\ 2\ 3\ 4\ 5]*0.8 = [0.8\ 1.6\ 2.4\ 3.2\ 4] \rightarrow [1\ 1\ 2\ 3\ 4]$
- Distorsion corrélée avec le signal : cela fait apparaître des harmoniques dans le spectre.
- Stratégie : ajouter un bruit uniforme avant conversion, pour décorréler ("quantization error randomization")
- Ainsi, on "étale" l'erreur sur l'ensemble du spectre ; on profite aussi du masquage



Contexte industriel autour du TSA Transducteurs électroacoustiques et fonctions analogiques Psychoacoustique et application au codage avec perte Traitements du signal fréquentiel **Traitements audionumériques** Architectures matérielles pour le TSA

Mixers virtuels Spectre et égalisation Dynamique et compression Spatialisation

# Dithering



# Dithering: exemples audio

https://www.dropbox.com/sh/ij1y1eeqj9xu1ci/AAByQjqtsyniIz2k9m6mI8eQa?dl=0

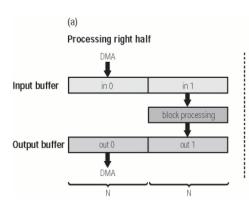
- Tom + reverbération numérique
- 1) Reverb Decay 16 bits
- 2) Reverb Tail (Boosted)
- 3) Reverb Decay + dithering
- 4) Reverb Tail + dithering (no disto, but noise)
- 5) Rev Decay 16bits (boost + limiter)
- 6) Rev Decay 24bits (idem)

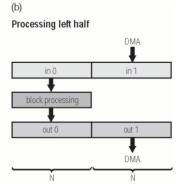


### Latence et temps réel

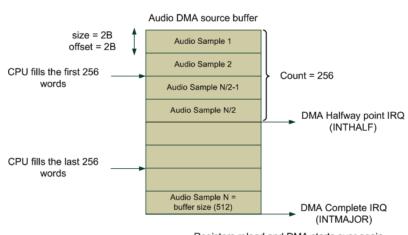
- Influence de la taille des buffers
- Double buffering, ring buffer, buffer circulaire
- Low latency mode
- Plugin delay compensation
- Problèmes liés à la phase (filtres en peigne)

### Double buffering





# Double buffering

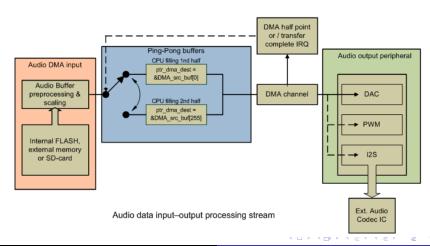


Registers reload and DMA starts over again

Contexte industriel autour du TSA Transducteurs électroacoustiques et fonctions analogiques Psychoacoustique et application au codage avec en Traitements du signal fréquentiel Traitements audionumériques Architectures matérielles pour le TSA

Mixers virtuels
Spectre et égalisation
Dynamique et compression
Spatialisation

# Double buffering



### Egalisation: pourquoi?

- Mixer = répartir les instruments de manière harmonieuse dans le spectre audible
- Notion d'équilibre tonal
- Altération du timbre : "thin vs fat", "big vs small", "clean vs dirty", "elegant vs rude"
- Séparation des sources (overlap, masquage)
- Définition (reconnaître un instrument)
- Mise à distance (cutoff LPF)

### Les quatre bandes standards

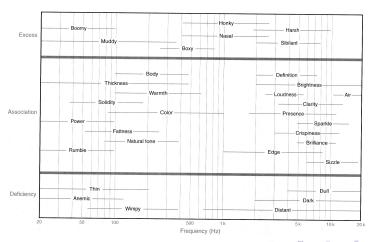
Lows: 20-300 Hz

Low-mids: 200-2kHz

• High-mids : 1k – 7kHz

• Highs: 6 kHz

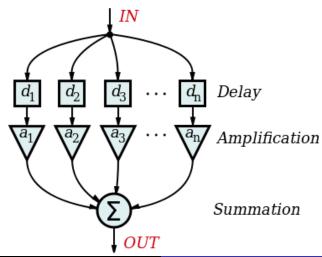
### Qualificateurs subjectifs



### Familles de filtres

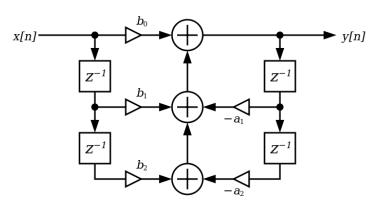
- HPF & I PF
- Résonance
- Shelving filters
- Filtres paramétriques
- EQ graphiques

## FIR Filter design



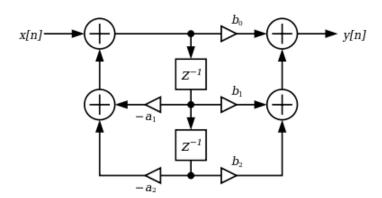
### IIR Filter design type 1

Direct form I (instable si N grand, 2N délais)

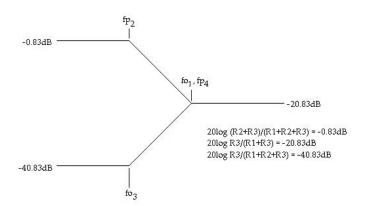


## IIR Filter design type 2

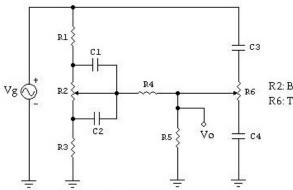
Direct form II: N délais / pb overflow.



### Filtres "shelf" dans les graves



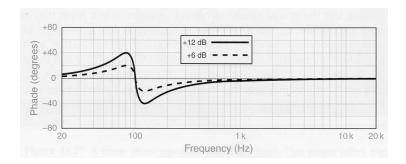
## Filtres "shelf" dans les aigus



R2: Bass control R6: Treble control

\_\_\_\_\_

### Phase artifacts



## Implémentations : FIR vs IIR

#### FIR

- + Phase linéaire (cf image stéréo)
- + Contrôle aisé de la réponse en fréq.
- Coûteux

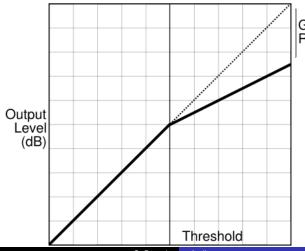
#### IIR

- + Réponse en fréquence de type "brickwall"
- + Peu gourmand en temps de calcul
- Artefacts de phase

Contexte industriel autour du TSA Transducteurs électroacoustiques et fonctions analogiques Psychoacoustique et application au codage avec en Traitements du signal fréquentiel Traitements audionumériques Architectures matérielles pour le TSA

Mixers virtuels Spectre et égalisation **Dynamique et compression** Spatialisation

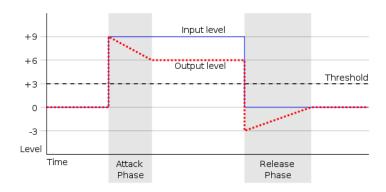
### Compresseur



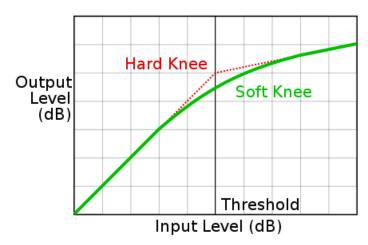
Gain Reduction

₹ 900

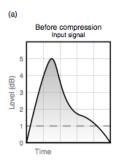
### Compresseur

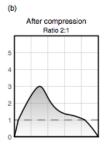


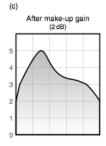
### Compresseur



### Compresseur



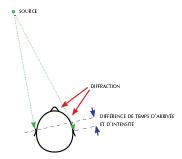




## Perception stéréophonique

La stéréophonie nous permet de localiser une source sonore dans l'espace :

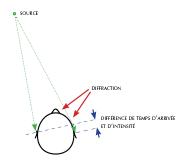
- ullet différence de marche :  $\delta\phi$  entre les deux signaux
- différence d'intensité



# Perception stéréophonique

### Exemple:

- ullet différence de marche :  $\delta\phi$  entre les deux signaux
- différence d'intensité



### **Panning**

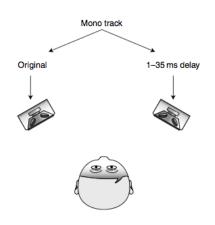
### Enjeu esthétique :

- Nous localisons chaque source sonore grâce à notre captation "stéréophonique" : c
- Localiser chaque source sonore dans l'espace stéréophonique.
- Réduire le masquage, donner davantage d'espace à chaque source en répartissant sur les deux oreilles

#### Outils et limitations :

- Mixage stereo : chaque source est "mixée" sur les deux canaux
- Effet Haas (< 35ms, < 10dB): crée un champ stéréophonique plus complet et large.
- Attention au combfiltering !
- On peut aussi inverser la phase sur un canal.

## Panning et effet Haas



### Acoustique des salles

- Distribution des modes :  $f^2$  (random freq. distrib. ok)
- Idem densité d'écho (en  $t^2$ )
- ullet Loi de Sabine :  $RT_{60}=rac{4V\log 10^6}{c\, Sa}pprox 0.1611 m^{-1}rac{V}{Sa}$

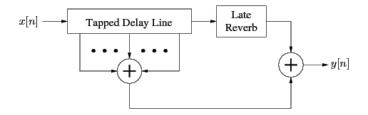
### Emulation réverbération

- Early reflection: en-dessous de 30ms, impression de volume de la salle (effet Haas)
- Masquage de la tail dépendant de la position auditeur vs source
- Réalisations : HP-Corridor / Spring / Plate / Digital.
- Dans les graves : distribution sparse de modes résonants
- Dans les aigus : haute densité de modes
- Idéalement : contrôle de RT60 dans 3 bandes de fréquence
- Clarity: ratio early/late reverb.



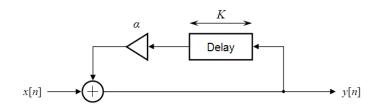
## Emulation réverbération : lignes à retard

- Emulation early reflexions: ligne à retard
- Eearly reflexions : Prise en compte de la géométrie
- Diffuse scattering (pas de ray tracing)



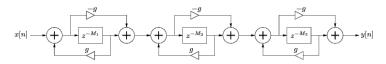
# Réverbération à delays

- Filtre en peigne ("Comb filter")
- Effet spectral si retard inférieur à 10ms
- Effet "echo" si retard supérieur à 10ms
- Feedback influence RT<sub>60</sub>



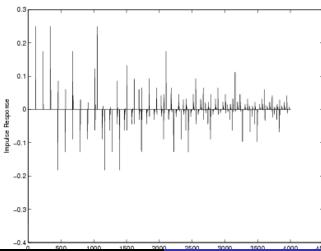
# Réverbération à delays

- Emulation : late reverb (densité d'écho)
- Algorithme de Schroeder : réverb par blocs
- Filtres allpass / retards premiers entre eux
- $G \approx 0.7$  et  $retard \approx 5 100 ms$



Spatialisation

### Emulation réverbération : convolution



### Réverbération

Pour éloigner une source, on peut :

- Baisser son amplitude
- Appliquer un filtre LP
- Ajouter de l'écho
- Ajouter de la réverbération

### Plan

- Contexte industriel autour du TSA
- 2 Transducteurs électroacoustiques et fonctions analogiques
  - Rappels d'acoustique physique : rayonnement, propagation
  - Microphones, lignes et préamplification
  - Haut-parleurs et étages de puissance
- Sychoacoustique et application au codage avec perte
  - Psychoacoustique, anatomie et perception
  - Encodeur MPEG layer I
- 4 Traitements du signal fréquentiel
  - Transformées de Fourier pour l'audio
  - Application: time stretching par phase vocoder
- Traitements audionumériques
  - Mixers virtuels
  - Spectre et égalisation



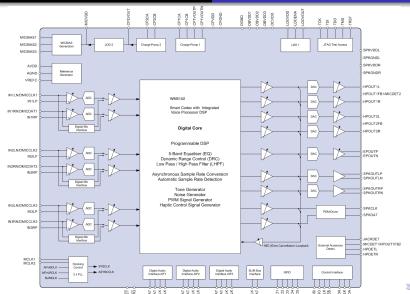
### Les CODEC audio

- System On Chip
- Embarque DAC, ADC, Mixer, AOp, DSP, ...
- Interfaces uP: I2S, TDM, MSB, LSB, SAI, ...
- Entrées line, micro (+ bias)
- Sorties line et puissance (HP, Casques)
- Interfaces numériques : SPDIF, AES, USB
- Applications: systèmes multimedia, smartphones, "cartes son", enregistreurs portatifs, consoles de mixage, caméra, TV, etc
- Fabriquants : Analog Device, Wolfson (Cirrus Logic), Texas Instrument



Contexte industriel autour du TSA
Transducteurs électroacoustiques et fonctions analogiques
Psychoacoustique et application au codage avec perte
Traitements du signal fréquentiel
Traitements audionumériques

Architectures matérielles pour le TSA



### Wolfson WM5102 Specifications

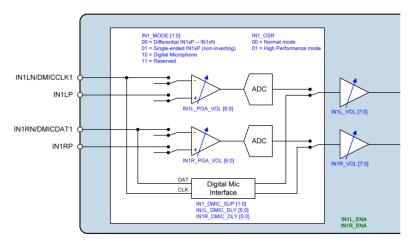
- DACs 7
- ADCs 6
- DAC THD+N (dB) -85
- ADC THD+N (dB) -86
- Sample Rate (kHz) 4-192
- Analog Power Supply (V) 1.8
- Digital Power Supply (V) 1.2 ou 1.8
- Power Consumption (mW) 7.6-1606

### Wolfson WM5102 Features

- Dynamic Range Control, Fully parametric EQs
- programmable filters
- Tone, Noise, PWM, Haptic control signal generators
- Multi-channel asynchronous sample rate conversion
- Integrated 6/7 channel 24-bit hi-fi audio hub codec
- 6 ADCs, 96 dB SNR microphone input (48kHz)
- 7 DACs, 113 dB SNR headphone playback (48kHz)
- Audio inputs : Single-ended or differential mic/line inputs
- Audio outputs : 29mW into 32 $\Omega$  load at 0.1% THD+N
- 2 x 2W stereo Class D speaker output drivers
- Pop suppression functions
- Flexible clocking

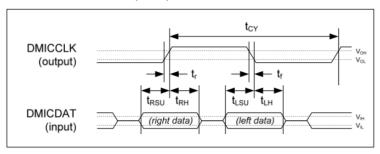


## Input Signal Path



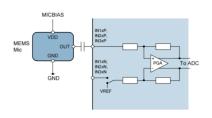
## Protocoles d'interfaçage capture/restitution

#### DIGITAL MICROPHONE (DMIC) INTERFACE TIMING

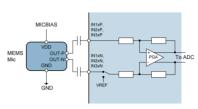


Ordre de grandeur : 10ns !

## Mems (Micro Electro Mechanical Systems) input



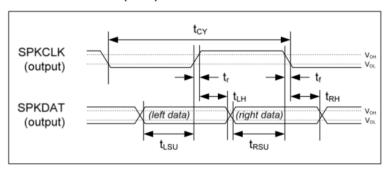
Single-Ended MEMS Input



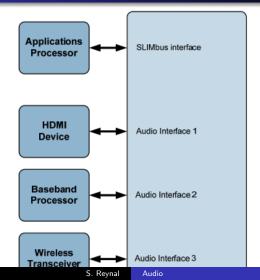
**Differential MEMS Input** 

# Protocoles d'interfaçage capture/restitution

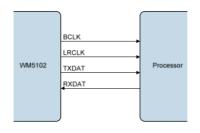
#### **DIGITAL SPEAKER (PDM) INTERFACE TIMING**

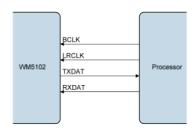


#### Digital Audio Interface (AIF) + SLIMbus interface

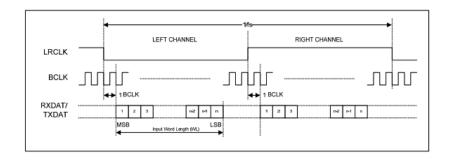


## Digital Audio Interface (AIF): I2S data format

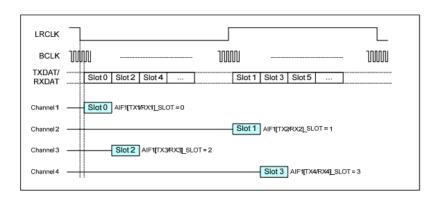




#### Digital Audio Interface (AIF): I2S data format

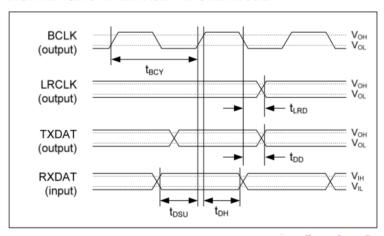


# Digital Audio Interface (AIF): I2S data format en multichannel

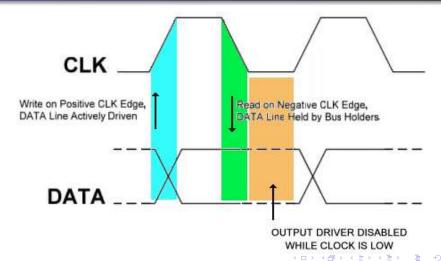


#### Digital Audio Interface (AIF): I2S bus timing

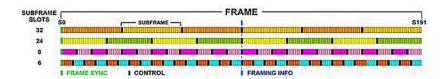
#### **DIGITAL AUDIO INTERFACE - MASTER MODE**



#### Control Interface : SLIMbus (couche PHY)

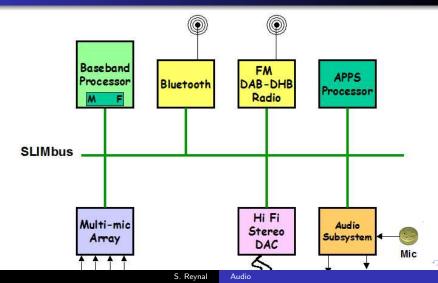


#### Control Interface : SLIMbus



Contexte industriel autour du TSA
Transducteurs électroacoustiques et fonctions analogiques
Psychoacoustique et application au codage avec perte
Traitements du signal fréquentiel
Traitements audionumériques
Architectures matérielles pour le TSA

#### Control Interface: SLIMbus



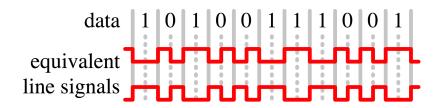
| REGISTER<br>ADDRESS | BIT | LABEL         | DEFAULT | DESCRIPTION                  |
|---------------------|-----|---------------|---------|------------------------------|
| R1280               | 4:0 | AIF1_BCLK_FRE | 01100   | AIF1BCLK Rate                |
| (0500h)             |     | Q [4:0]       |         | 00000 = Reserved             |
| AIF1                |     |               |         | 00001 = Reserved             |
| BCLK Ctrl           |     |               |         | 00010 = 64kHz (58.8kHz)      |
|                     |     |               |         | 00011 = 96kHz (88.2kHz)      |
|                     |     |               |         | 00100 = 128kHz (117.6kHz)    |
|                     |     |               |         | 00101 = 192kHz (176.4kHz)    |
|                     |     |               |         | 00110 = 256kHz (235.2kHz)    |
|                     |     |               |         | 00111 = 384kHz (352.8kHz)    |
|                     |     |               |         | 01000 = 512kHz (470.4kHz)    |
|                     |     |               |         | 01001 = 768kHz (705.6kHz)    |
|                     |     |               |         | 01010 = 1.024MHz (940.8kHz)  |
|                     |     |               |         | 01011 = 1.536MHz (1.4112MHz) |
|                     |     |               |         | 01100 = 2.048MHz (1.8816MHz) |
|                     |     |               |         | 01101 = 3.072MHz (2.8824MHz) |
|                     |     |               |         | 01110 = 4.096MHz (3.7632MHz) |
|                     |     |               |         | 01111 = 6.144MHz (5.6448MHz) |
|                     |     |               |         | 40000 - 9 400MU- (7 5064MU-) |

#### Exemple avec la bibliothèque HAL (STM32):

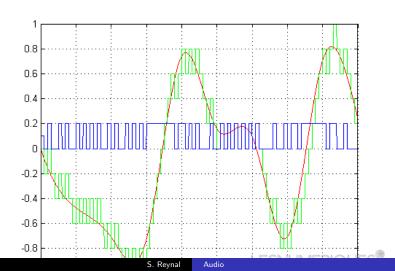
| REGISTER<br>ADDRESS | BIT | LABEL           | DEFAULT | DESCRIPTION   |
|---------------------|-----|-----------------|---------|---|
| ADIRES              | 7:0 | OUT1L_VOL [7:0] | 80h     | Output Path 1 (Left) Digital Volume -64dB to +31.5dB in 0.5dB steps 00h = -64dB 01h = -63.5dB (0.5dB steps) 80h = 0dB |
|                     |     |                 |         | (0.5dB steps) BFh = +31.5dB C0h to FFh = Reserved (See Table 65 for volume range)                                     |

| REGISTER<br>ADDRESS | BIT | LABEL                  | DEFAULT | DESCRIPTION  |
|---------------------|-----|------------------------|---------|--|
|                     | 9:8 | DRC1_QR_THR<br>[1:0]   | 00      | DRC1 Quick-release threshold<br>(crest factor in dB)<br>00 = 12dB<br>01 = 18dB |
|                     |     |                        |         | 10 = 24dB<br>11 = 30dB   |
|                     | 7:6 | DRC1_QR_DCY<br>[1:0]   | 00      | DRC1 Quick-release decay rate (seconds/6dB)                                    |
|                     |     |                        |         | 00 = 0.725ms   |
|                     |     |                        |         | 01 = 1.45ms  |
|                     |     |                        |         | 10 = 5.8ms   |
|                     |     |                        |         | 11 = Reserved  |
|                     | 5:3 | DRC1_HI_COM<br>P [2:0] | 011     | DRC1 Compressor slope (upper region)   |
|                     |     |                        |         | 000 = 1 (no compression)   |
|                     |     |                        |         | 001 = 1/2  |
|                     |     |                        |         | 010 = 1/4  |
|                     |     |                        |         | 011 = 1/8  |

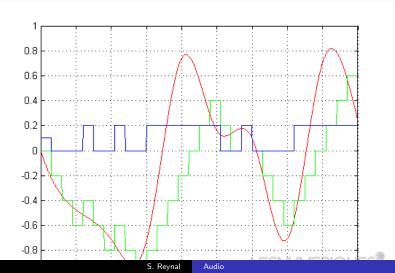
#### SPDIF: reconstruction d'horloge



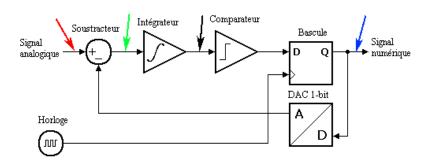
#### Convertisseurs sigma-delta: delta simple



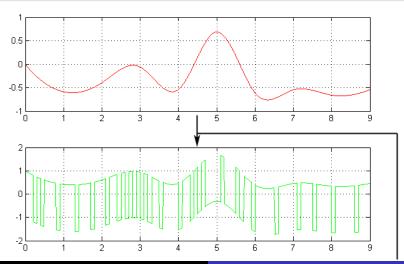
#### Convertisseurs sigma-delta : delta simple sous-échantilloné



#### Convertisseurs sigma-delta

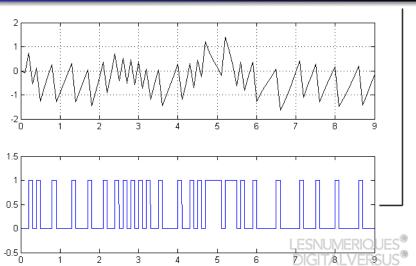


#### Convertisseurs sigma-delta: signal d'erreur





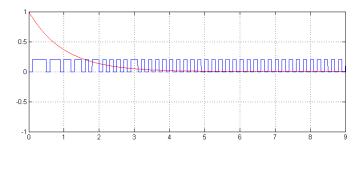
#### Convertisseurs sigma-delta : signal intégré et comparateur

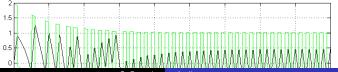


Audio

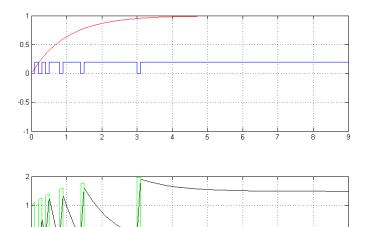
S. Reynal

#### Convertisseurs sigma-delta: exemple 1





#### Convertisseurs sigma-delta: exemple 2



#### Convertisseurs sigma-delta: exemple 3

