

ENSEA

Traitement du signal audio musical, descripteurs et
estimation

Geoffroy.Peeters@ircam.fr
UMR SMTS IRCAM CNRS UPMC

13/04/2015

1- Examen

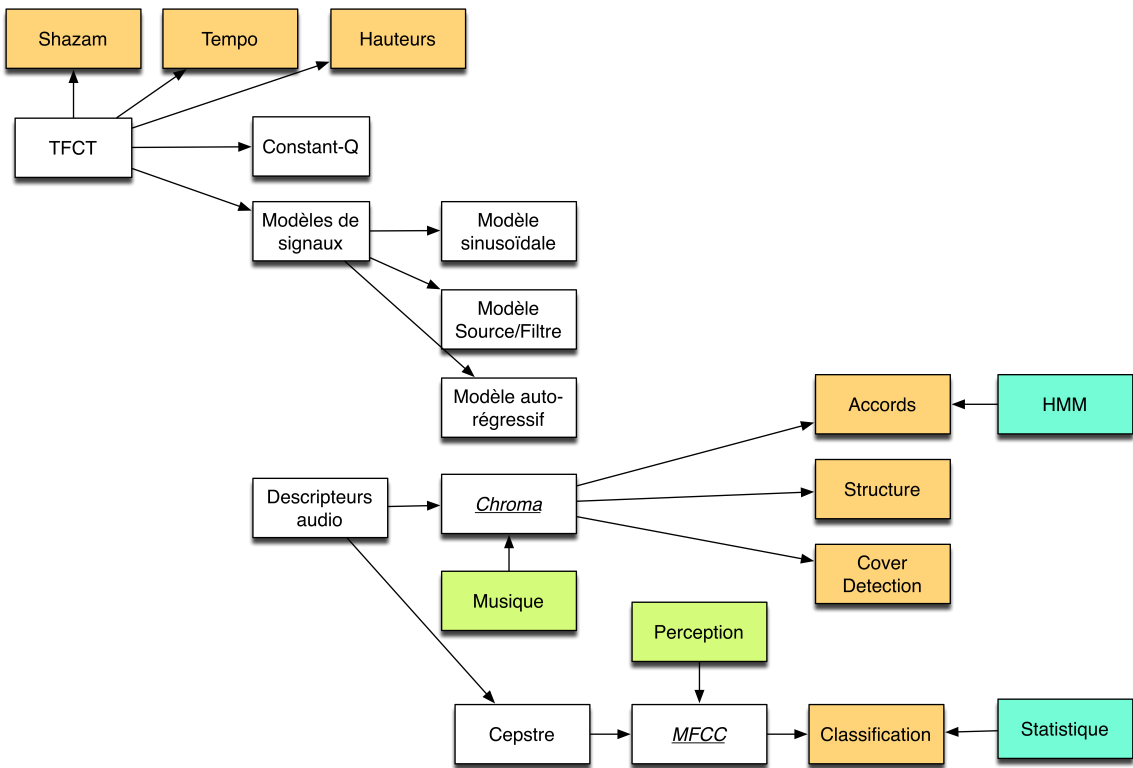
1.1- Option Audio-Parole-Musique 2nd semestre

Examen Option Audio-Parole-Musique 2nd semestre

- Quand : lundi 4 mai 2015 - 8h à 9h30 (1h30)
- Salle : 261
- sans document

Support de cours

Plan



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique



Enter a keyword, record a query or drag an example clip.

  [Search Audio](#) [Audio Preferences](#)
[Audio Help](#)



[Steve Jobs interview](#)
7 min 14 sec
Speech



[Metric - Raw Sugar](#)
3 min 47 sec
Music - Indie Pop



[Grenade explosion](#)
23 sec
Sound effect

[similarly random recordings >](#)

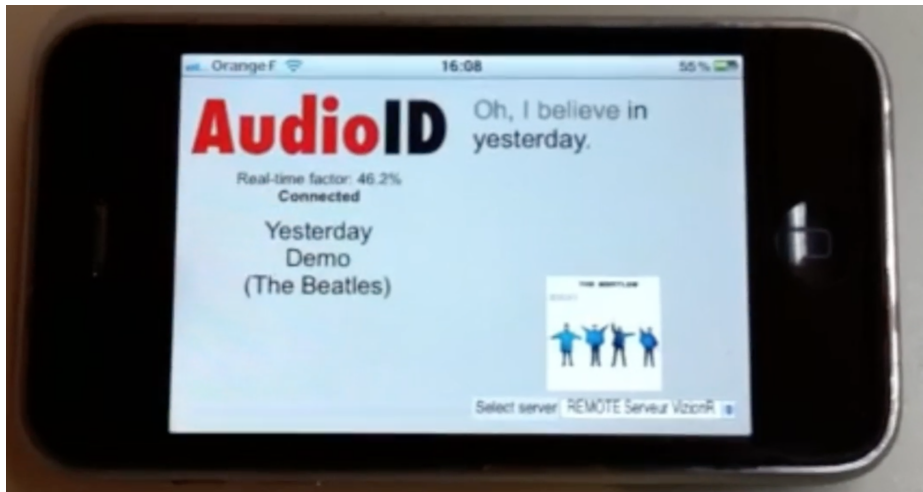
[Google Labs](#) - [Discuss](#) - [Terms of use](#) - [About Google Audio](#) - [Submit your recording](#)

source : Gaël Richard

3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

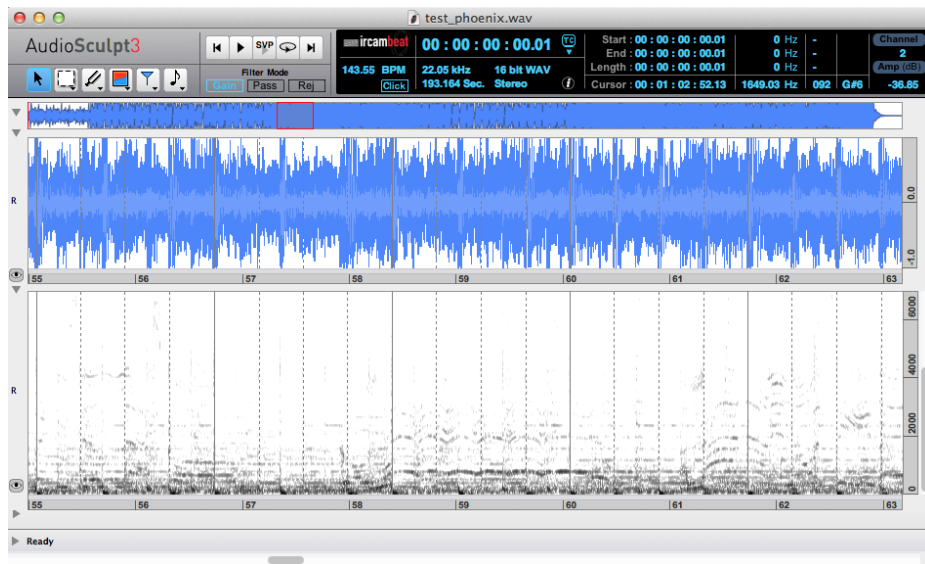
- Identification audio
 - recherche de doublons, gestion de copyright, attacher des méta données à une instance d'un morceau



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

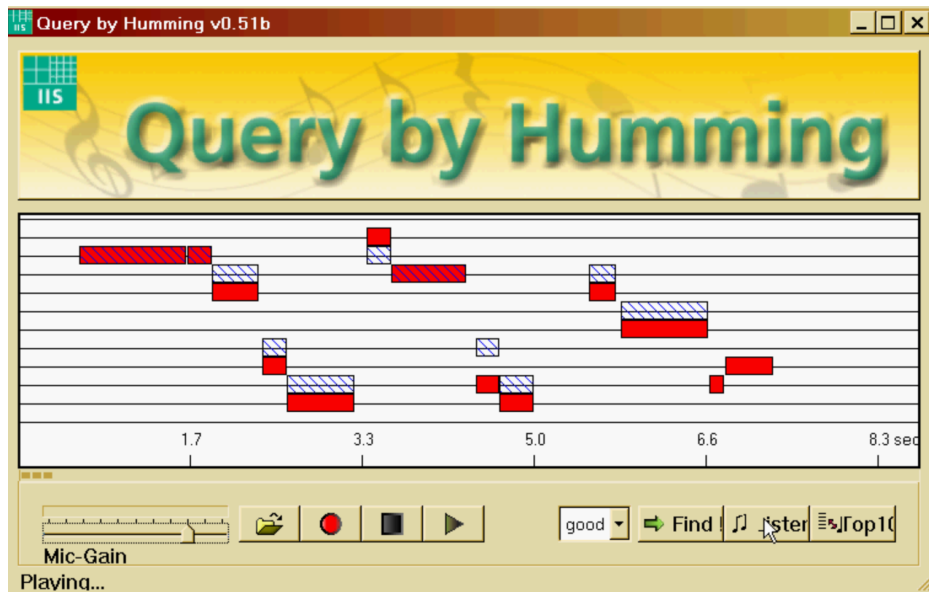
- Estimation du tempo, de la position des temps/ premier-temps
 - DJing, manipulation du contenu (add swing ...)



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

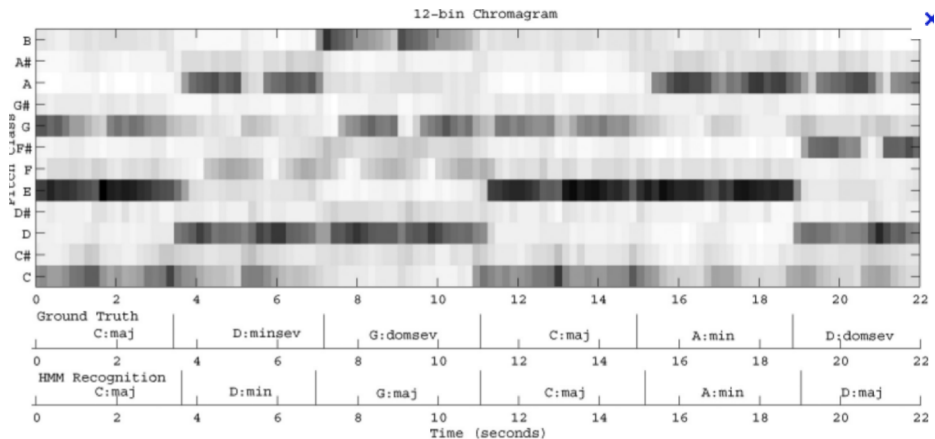
- Nouveaux modes de recherche :
 - par chantonnement/ sifflement



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

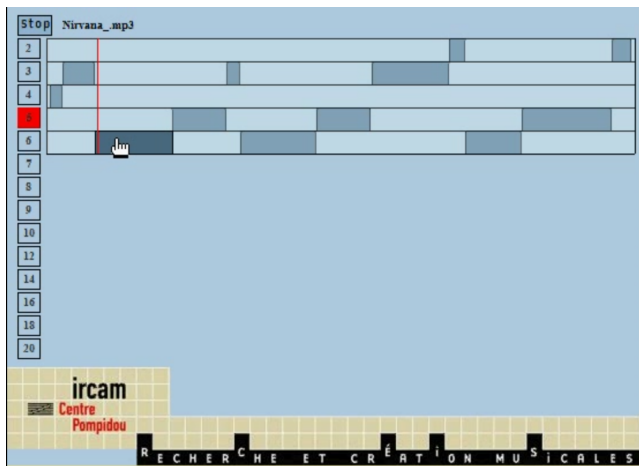
- Estimation des accords
 - obtenir des guitar-tab automatiquement



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

- Navigation à l'intérieur d'un morceau de musique par couplet/refrain
 - Génération automatique de résumé audio
- Dé-linéarisation d'un flux audio :
 - segmentation de flux radio, télé et étiquetage des parties



3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

- Détection des cover, reprises ou ... des plagias

Titre	Artiste	Album	D.	Pop.	
Let It Be	∨ The Beatles Recovered Band	30 Beatles Top Hits	03:50		<input type="checkbox"/>
Let It Be	∨ The Hit Co., The Tribute Co.	A Tribute to the Beatles: The Lat...	03:42		<input type="checkbox"/>
Let It Be	∨ Labrinth	Let It Be	03:05		<input type="checkbox"/>
Let It Be Me	∨ Ray LaMontagne	Gossip in The Grain	04:41		<input type="checkbox"/>
Let It Be - The Beatles Tribute	Let It Be	Let It Be - The Beatles Tribute	03:49		<input type="checkbox"/>
Let It Be	Lois	Let It Be - The Voice 2	03:15		<input type="checkbox"/>
Let It Be	The Yesteryears	A Tribute to #1 Beatles Hits - T...	03:48		<input type="checkbox"/>
Let It Be	∨ Aretha Franklin	This Girl's In Love Wlth You	03:33		<input type="checkbox"/>
Let It Be Sung	∨ Jack Johnson, Matt Costa, Zach Gill,...	If I Had Eyes	04:09		<input type="checkbox"/>
Let It Be	Vox Angeli	Gloria	03:26		<input type="checkbox"/>
Let It Be	∨ Paul McCartney	Good Evening New York City	03:54		<input type="checkbox"/>
Hey Jude	Let It Be	Hey Jude	03:55		<input type="checkbox"/>
Let It Be	Joan Baez	Greatest Hits And Others	03:51		<input type="checkbox"/>

3- Introduction

3.1- Applications des techniques d'indexation audio pour la musique

- Recherche d'un contenu audio dans une base de données
 - autrement que par "artistes", "titres" (Google musical)

Recherches proches : [maceo parker \(8\)](#) [all the king's men/maceo parker \(3\)](#)

Got to get you Dynamique - Soul/Funk - Batterie, Guitare électrique
Maceo Parker
Life on Planet Groove

ajouter à une playlist
chercher des musiques similaires

4:53 - 7:10 écouter le résumé | écouter l'intégral | afficher les passages chantés

RÉSULTATS (8)

enregistrement dans une playlist

Titre	Artiste	Album	durée
Got to get you	Maceo Parker	Life on Planet Groove	07:10
Pass the pass	Maceo Parker	Life on Planet Groove	11:28
Addictive Love	Maceo Parker	Life on Planet Groove	09:00
Shake everything you've got	Maceo Parker	Life on Planet Groove	16:41
Soul Power 52	Maceo Parker	Life on Planet Groove	14:13
Georgis on my mind	Maceo Parker	Life on Planet Groove	07:25
I got you (I Feel Good)	Maceo Parker	Life on Planet Groove	03:47
Children's World	Maceo Parker	Life on Planet Groove	06:23

Modèles de lecture : **enregistré** | **manipuler**

Get Similar Tracks

HUMEURS
JOYEUX CALME (1) DYNAMIQUE (5)
ROMANTIQUE TRISTE

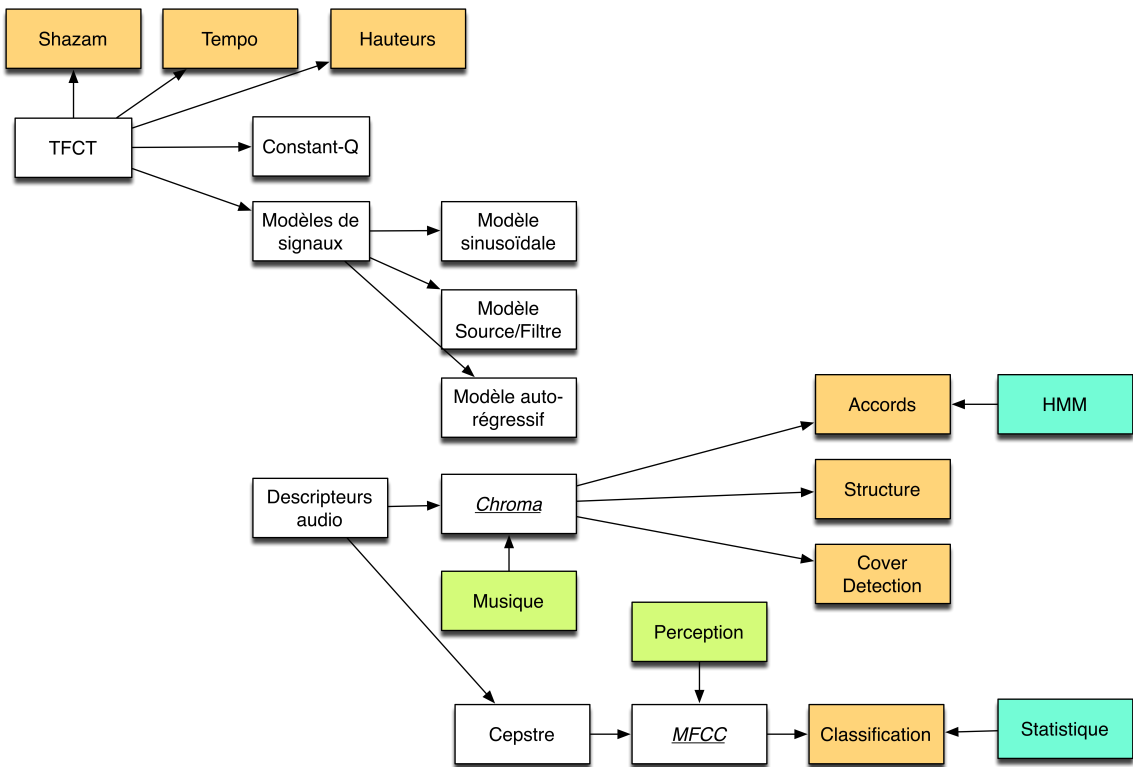
GENRES
POP/ROCK BLUES (2) ÉLECTRONIQUE
METAL/PUNK REGGAE CLASSIQUE JAZZ
RAP SOUL/FUNK (7) LATIN R&B

INSTRUMENTATIONS
GUITARE ÉLECTRIQUE (6) GUITARE ACOUSTIQUE
ÉLECTRONIQUE BATTERIE (8) CUVRES (2)
ORCHESTRE À CORDES PIANO ACCOUSTIQUE

ENREGISTREMENTS
STUDIO (1) LIVE (7)

MES PLAYLISTS
1 - Ismir (2)
[+ nouvelle playlist](#)

Plan

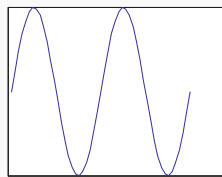


4- Théorie : Traitement du signal

4.1- Transformée de Fourier (temps et fréquences continus)

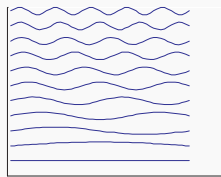
$$X(\omega) = \int_t x(t) \cdot \exp(j \cdot 2\pi f \cdot t) dt$$

- Variables :
 - t est le **temps**
 - $\omega = 2\pi f$ les **fréquences continues** exprimées en radian,
 - $\exp(j\omega t) = \cos(2\pi ft) + j \cdot \sin(2\pi ft)$.
- 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

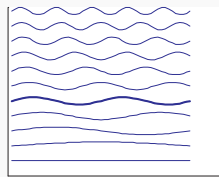


$x(t)$

X



$\sin(2\pi f t)$

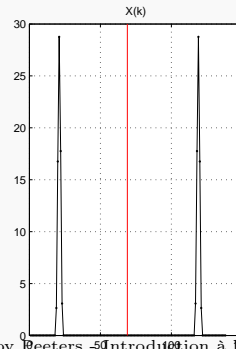
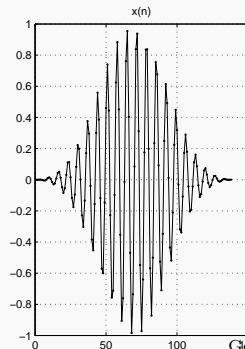


4- Théorie : Traitement du signal

4.2- Transformée de Fourier (temps et fréquences discrets)

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \exp(j \cdot 2\pi \frac{k}{N} \cdot n), \forall k \in [0, N]$$

- Variables :
 - n le numéro d'**échantillon**
 - k les **fréquences discrètes**
- Taux d'échantillonnage (sampling rate) : 44100 Hz
 - Distance entre deux échantillons : $\frac{1}{44100} = 0.000023s$.
 - **Fréquence de Nyquist** = taux d'échantillonnage
 - Repliement spectral (roue de voiture dans les films)



4- Théorie : Traitement du signal

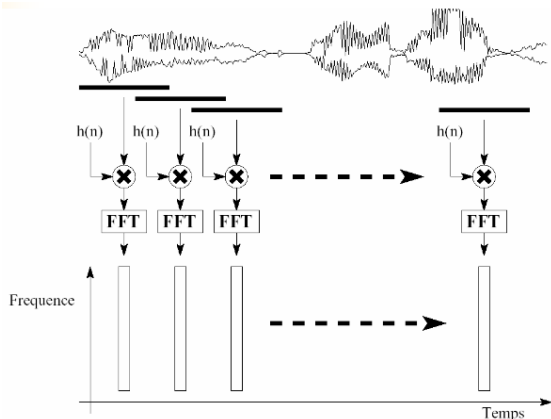
4.3- Transformée de Fourier (à Court Terme)

$$X(k, m) = \sum_{n=0}^{N-1} x(n) \cdot h(m - n) \cdot \exp(j \cdot 2\pi \frac{k}{N} \cdot (m - n)), \forall k \in [0, N]$$

- Application de la TFD à une portion du signal centrée autour de l'échantillon m

Pourquoi la Transformée de Fourier à Court Terme ?

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



source : Jean Laroche

4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

$$X(k, m) = \sum_{n=0}^{N-1} x(n) \cdot h(m - n) \cdot \exp(j \cdot 2\pi \frac{k}{N} \cdot (m - n)), \forall k \in [0, N]$$

Fenêtre de pondération $h(t)$

- $x(t) \cdot h(t) \Leftrightarrow X(\omega) * H(\omega)$
 - $h(t)$ est appelé "**fenêtre de pondération**"
 - $h(t)$ différents **types** de fenêtre
 - $h(t)$ 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 $-6dB_{20}$) : $Bw = \frac{Cw}{L}$
 - Hauteur des lobes secondaires

4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

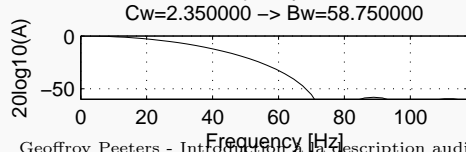
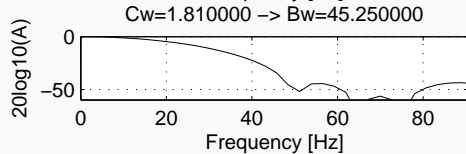
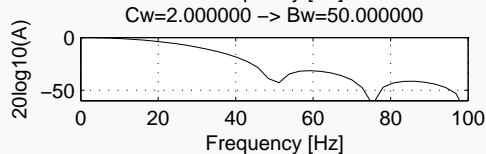
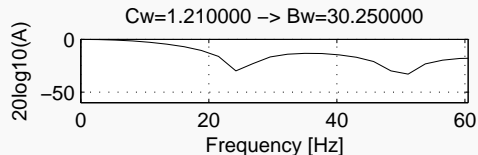
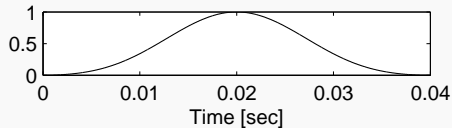
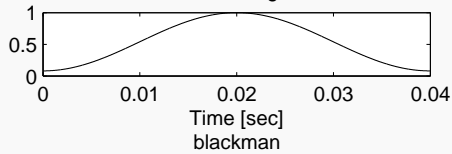
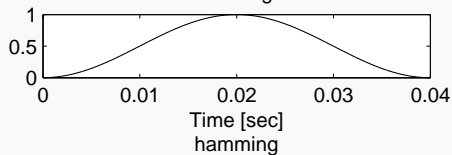
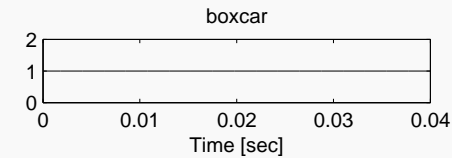
Choix du **type** de la fonction :

- rectangulaire $h(n) = 1 \rightarrow Bw = 1.81$
- hanning $h(n) = 0.5(1 - \cos(\frac{2\pi n}{N-1})) \rightarrow Bw = 2$
- hamming $h(n) = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1}) \rightarrow Bw = 1.81$
- blackman $h(n) = a_0 - a_1 \cos(\frac{2\pi n}{N-1}) + a_2 \cos(\frac{4\pi n}{N-1}) \rightarrow Bw = 2.35$

4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

Influence du **type** de la fonction



4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

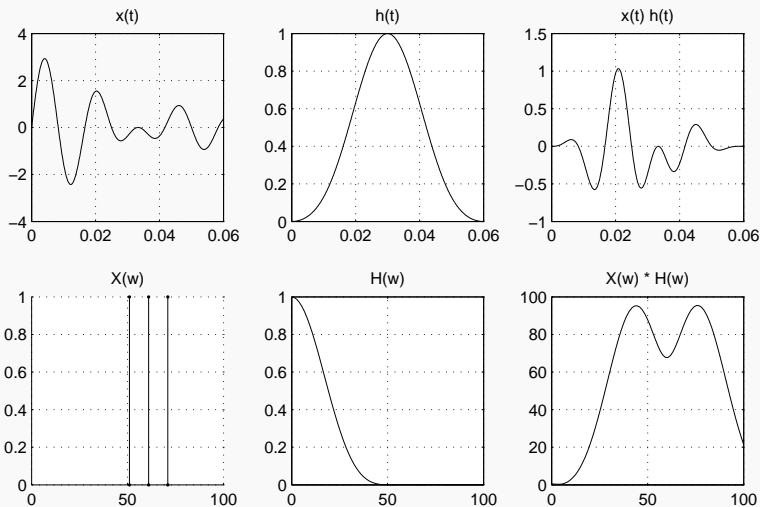
Choix de la **longueur temporelle** L :

- Au plus la fenêtre est courte,
 - au plus on observe précisément les temps.
- Au plus la fenêtre est longue,
 - au plus on observe précisément les fréquences.

4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

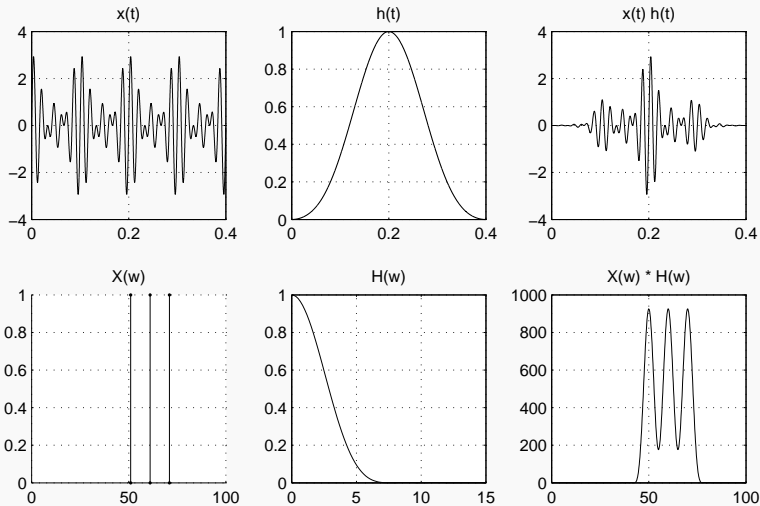
Influence de la **longueur temporelle** L ($L = 0.06$)



4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

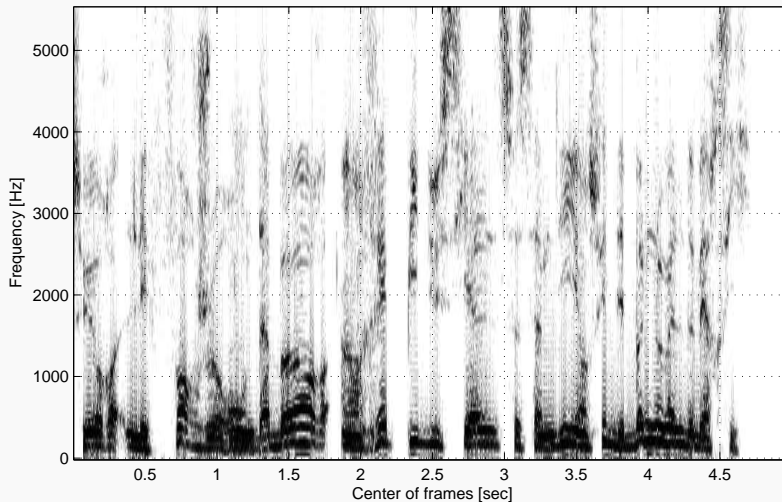
Influence de la **longueur temporelle** L ($L = 0.4$)



4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

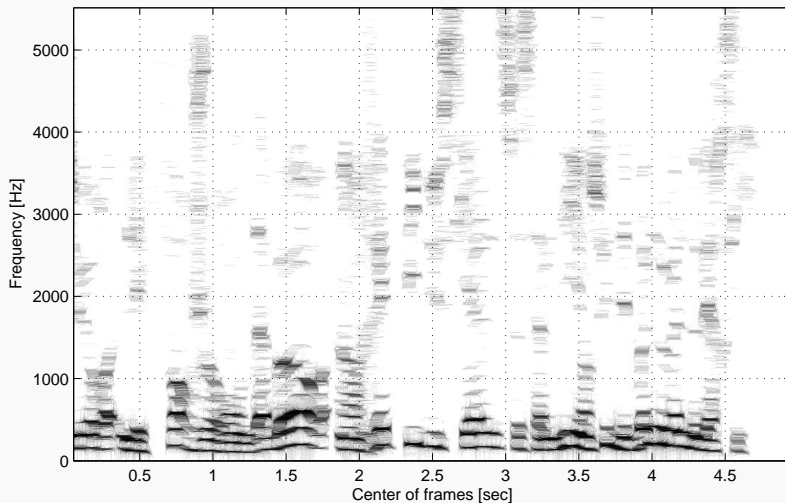
Influence de la **longueur temporelle** L ($L = 0.01$)



4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

Influence de la **longueur temporelle** L ($L = 0.1$)

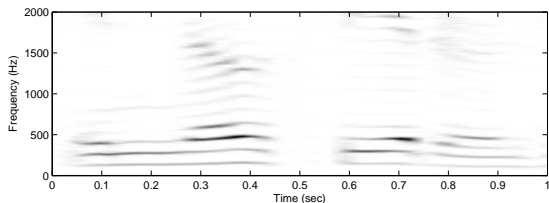
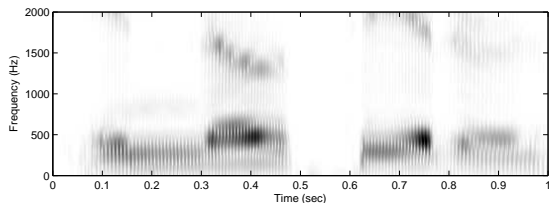
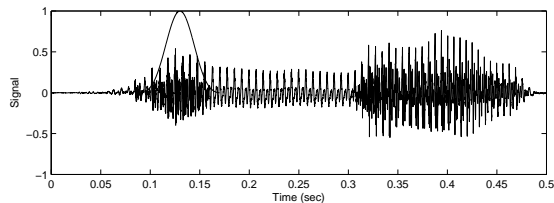
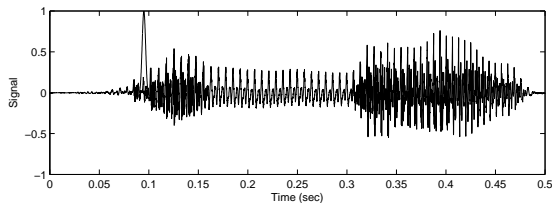


4- Théorie : Traitement du signal

4.3- Transformée de Fourier (à Court Terme)

Paradoxe temps/ fréquence

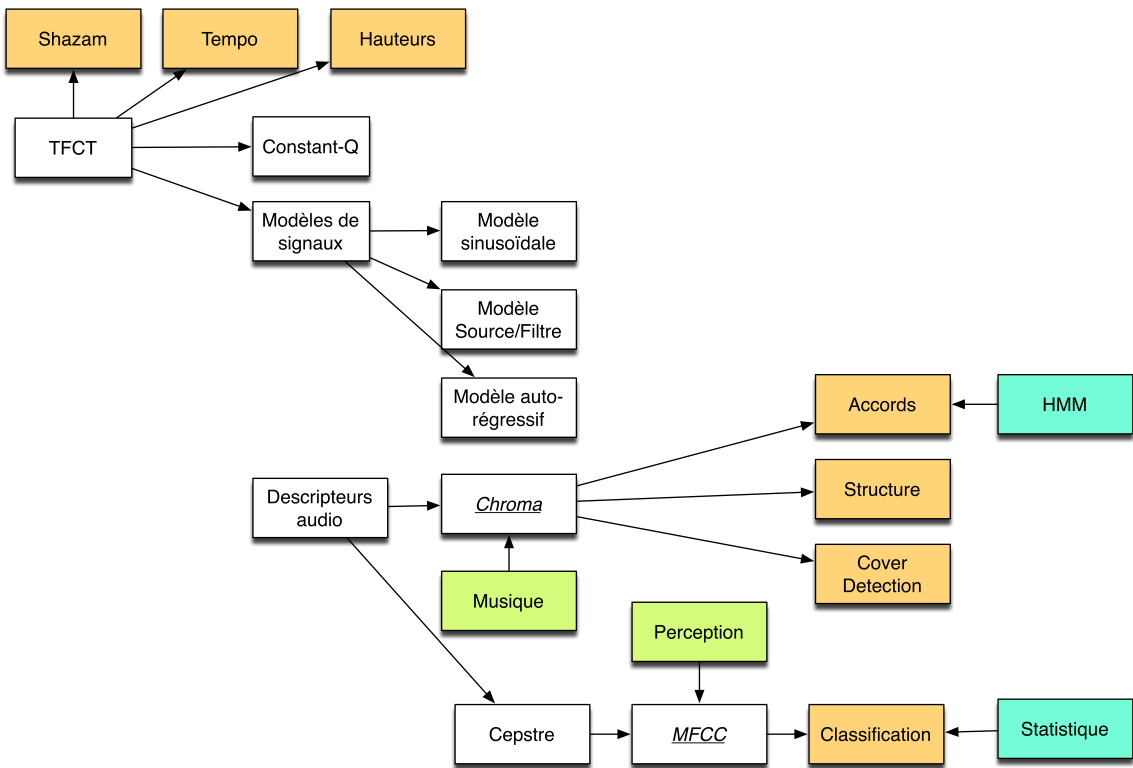
- Pas possible d'avoir simultanément une bonne localisation en temps et en fréquence !



Comme résoudre ce problème ?

- Utiliser d'autres transformées que celle de Fourier

Plan



5- Modèles de signaux

5.0-

Pourquoi des modèles de signaux ?

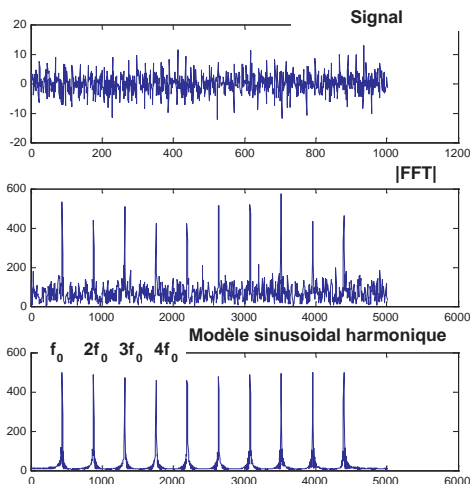
- On suppose que le signal a été produit par un certain **modèle**
- Permet de **réduire** le nombre de paramètres observés du signal
 - la TFCT contient beaucoup trop d'information
- Permet d'obtenir des paramètres plus facilement **interprétables** (indexation) et manipulables (transformation, synthèse)
 - la TFCT fournit des paramètres non directement exploitables
- Quels modèles ?
 - Modèle sinusoidal harmonique
 - Modèle source/ filtre
 - Modèle autorégressif

5- Modèles de signaux

5.1- Modèle sinusoidal harmonique

Modèle sinusoidal harmonique

- Hypothèse :
 - le signal $x(t)$ est un signal harmonique
 - Exemple : une note de musique, les parties voisées de la voix
- Modélisation :
 - le signal est représenté comme une somme de sinusoides dont les fréquences sont des multiples entiers de sa hauteur + un résiduel
 - Exemple : instrument de musique jouant un la-3 (A-4), $f_0 = 440$ Hz, modélisé comme la somme de ses harmoniques :
 $f_0 = 440$ Hz, $2f_0 = 880$ Hz,
 $3f_0 = 1320$ Hz, ...



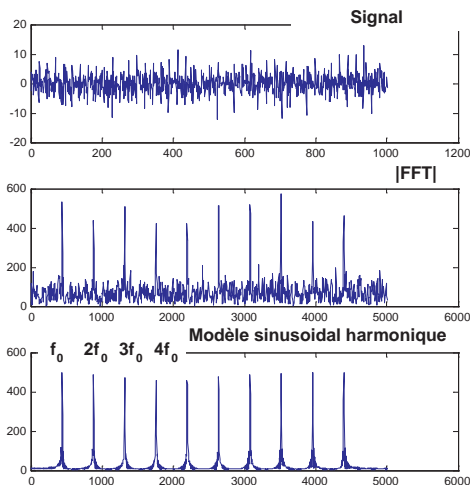
5- Modèles de signaux

5.1- Modèle sinusoidal harmonique

Modèle sinusoidal harmonique

$$X(f) = \sum_{h=1}^H A_h \sin(2\pi h f_0 + \phi_h) + \epsilon(f) \quad (1)$$

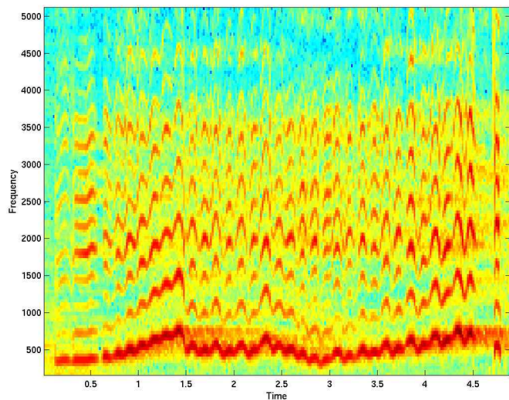
- ▶ A_h : amplitude de l'harmonique h
- ▶ $h f_0$: fréquence de l'harmonique h
- ▶ ϕ_h : phase initiale de l'harmonique h (représente la synchronisation temporelle)
- ▶ $\epsilon(f)$: résiduel, ce qui ne peut pas être représenté par le modèle



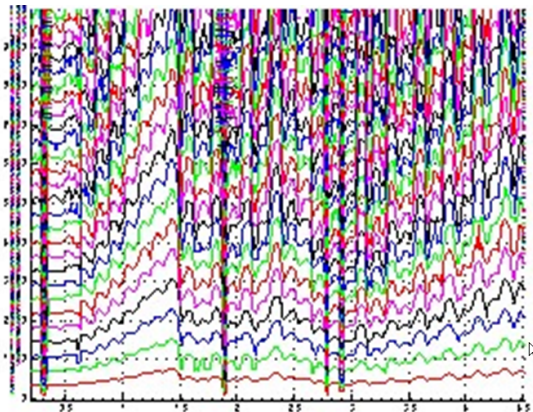
5- Modèles de signaux

5.1- Modèle sinusoidal harmonique

Les sinusoides harmoniques varient au cours du temps



TFCT



Modèle sinusoidal harmonique

5- Modèles de signaux

5.1- Modèle sinusoidal harmonique

Utilisation du modèle sinusoidal ?

- Synthèse, modification du signal
- Codage
 - paramétrisation du signal comme $f_0 +$ enveloppe spectrale $[A_1, A_2, \dots, A_H]$
- Extraction de paramètres pour l'indexation

Différentes estimations

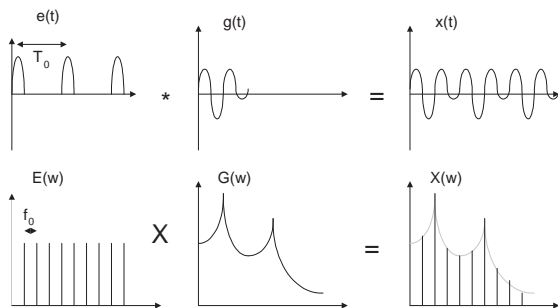
- de la fréquence fondamentale
 - **Temporel** : auto-corrélation, Average Mean Difference Function, Yin, Cepstre
 - **Fréquentiel** : Filtre en peigne harmonique, Maximum de vraisemblance, ...
- de l'enveloppe spectrale
 - LPC (Linear Predictive Coding)
 - Cepstre
 - MFCC

5- Modèles de signaux

5.2- Modèle source/ filtre

Modèle source/ filtre

- Hypothèse :
 - le signal $x(t)$ est le résultat du passage d'une excitation (un pulse, une série de pulse) dans un filtre (résonnant)
 - Exemples : le signal de parole, certains instruments de musique (trompette)
- Modélisation temporelle :
 - un signal d'excitation $e(t)$ passe (convolution) à travers un filtre $g(t)$:
 - $x(t) = e(t) * g(t)$
- Modélisation fréquentielle
 - la multiplication de la TF du signal d'excitation (source) par la TF du filtre.
 - $X(\omega) = E(\omega) \cdot G(\omega)$

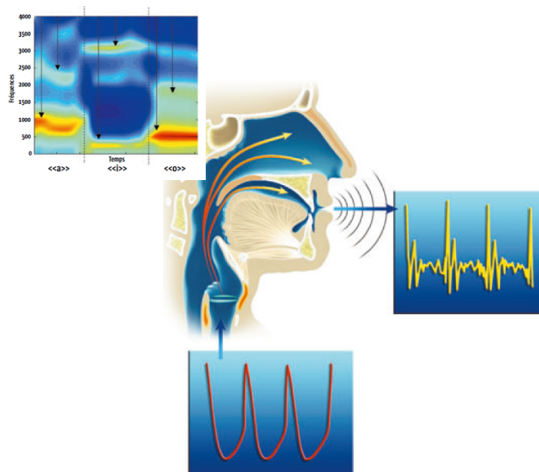


5- Modèles de signaux

5.2- Modèle source/ filtre

Modélisation du signal de parole :

- Le signal de parole (pour sa partie voisée) est créé par
 - les cordes vocales
 - une excitation régulière / périodique
 - le conduit bucco-nasal (bouches/ nez)
 - filtrage résonant/ anti-résonant

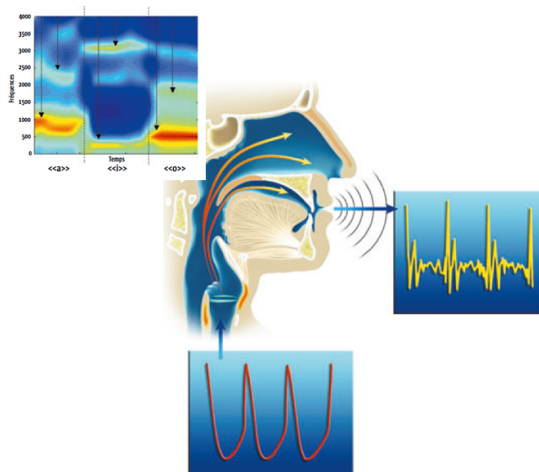


5- Modèles de signaux

5.2- Modèle source/ filtre

Modélisation du signal de parole :

- Ouverture/fermeture périodique des cordes vocales
 - détermine la hauteur
 - Hauteur de 100Hz ? pulses d'air sont espacés de
$$T_0 = 1/f_0 = 1/100 = 10ms.$$
 - Appelé signal d'excitation (ou signal source), $e(t)$.
- Conduit bucco-nasal
 - créer les différentes voyelles pour une hauteur donnée en renforçant (résonance) et retirant (anti-résonances) certains fréquences.
 - Filtre résonant (AR : Auto-Regressif) et anti-résonant (MA : Moving Average) : un filtre dit "ARMA".



5- Modèles de signaux

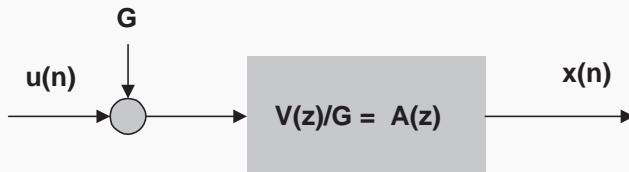
5.3- La prédiction linéaire

Modèle auto-régressif :

- Le signal à l'instant n peut être prédit à partir des instants précédents

$$\begin{aligned}x(n) &= a_1x(n-1) + a_2x(n-2) + a_3x(n-3)\dots + a_Px(n-P) \\ &= \sum_{p=1}^P a_p x(n-p) + G \cdot u(n)\end{aligned}\quad (2)$$

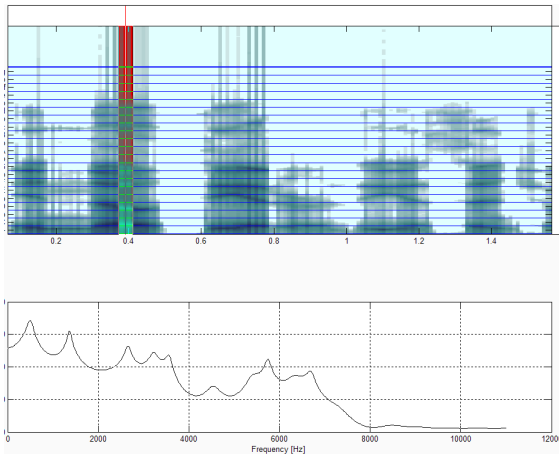
- Equivalent à passer le signal dans un filtre FIR tout-pôle : $V(z) = \frac{G}{1 + \sum_{p=1}^P a_p z^{-p}}$
- Objectif de la prédiction linéaire ?
 - déterminer le filtre $V(z)$ (donc les résonances ou les formants dans le cas de la voix) à partir du signal $x(n)$ (signal de pression micro)



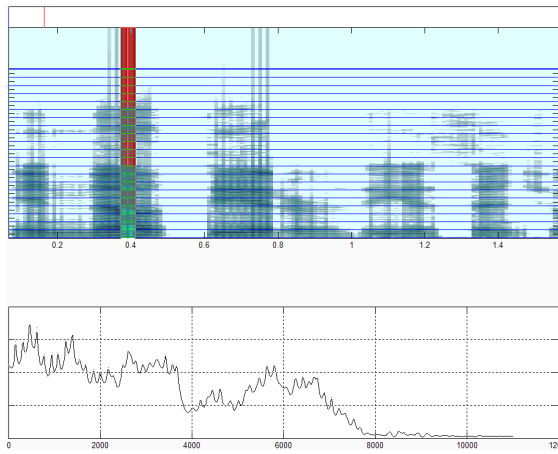
5- Modèles de signaux

5.3- La prédiction linéaire

Choix du nombre de pôle P



$P = 40$

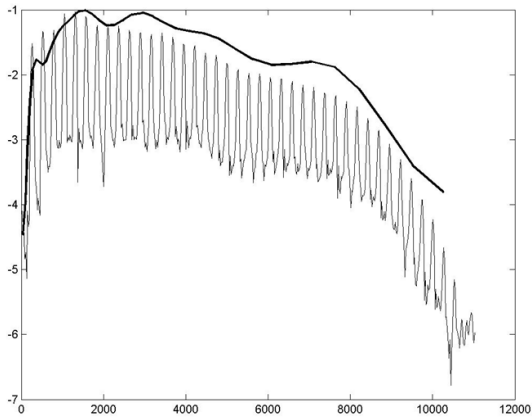


$P = 200$

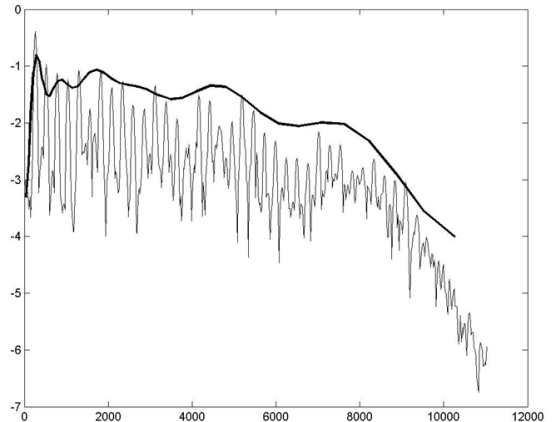
5- Modèles de signaux

5.3- La prédiction linéaire

Importance de la fréquence fondamentale et de l'enveloppe spectrale

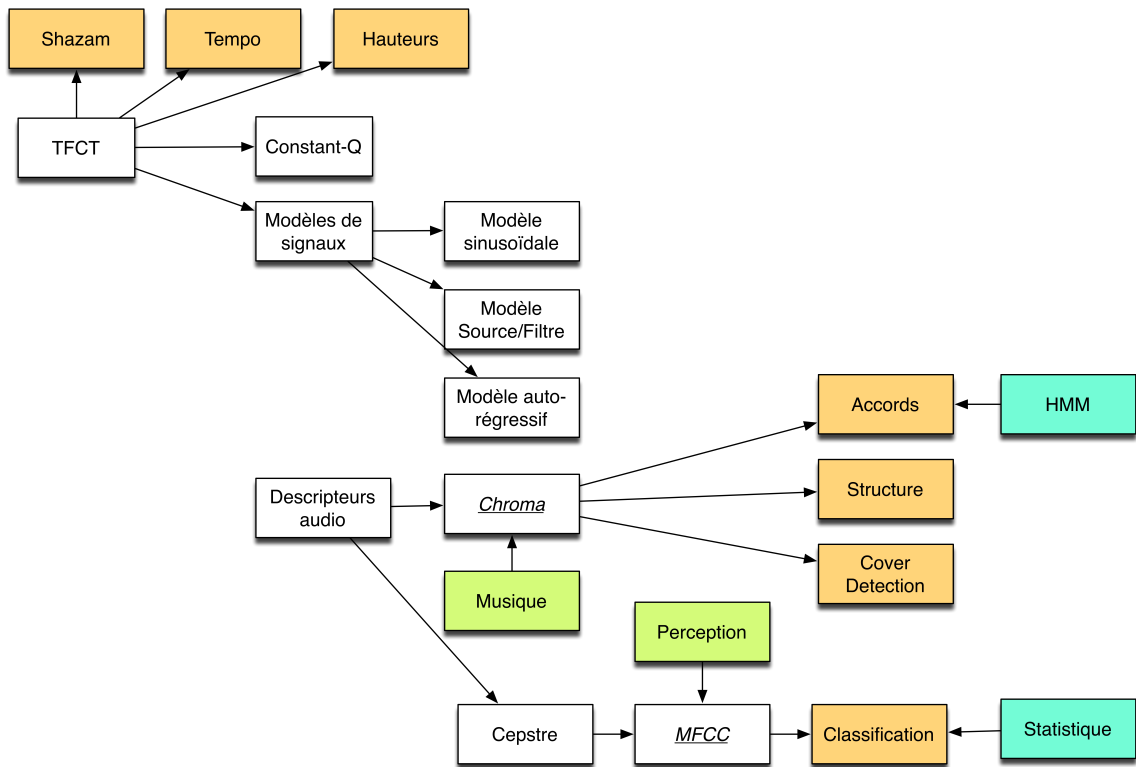


Enveloppe spectrale son de trompette



Enveloppe spectrale son de violon

Plan



Descripteurs audio ?

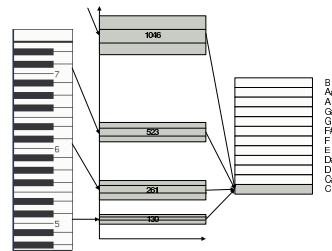
- Valeurs numériques extraites du signal audio dont le but est de représenter une propriété particulière de son contenu
 - Tout est dans la forme d'onde, dans la TFCT mais difficilement lisible et de trop grande dimension
- Forme ?
 - Descripteurs **scalaire**
 - Descripteur **vectorel**
- Extraction ?
 - **Opérateurs mathématique**
 - **Algorithme d'estimation**

6- Descripteurs audio

6.1- Chroma/ Class de Profil de Class

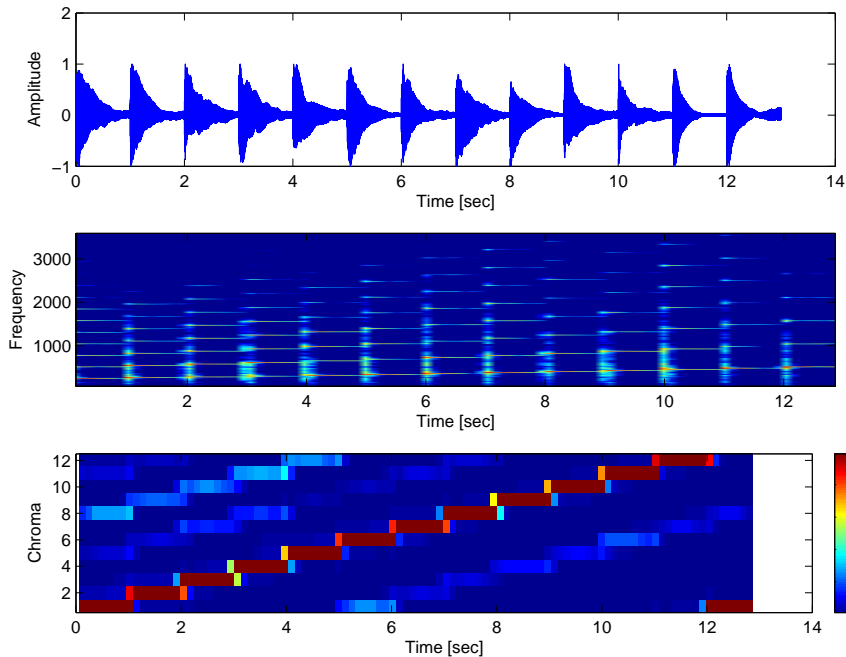
Chroma=Pitch Class Profile (PCP)

- Représenter le contenu harmonique du spectre à un instant donnée $X(k, m)$ sous forme d'un vecteur $C(n, m)$
 - Très utilisé dans les applications
 - reconnaissance de tonalité
 - reconnaissance de suite d'accords
 - identification de "cover versions"
- Shepard
 - représenter la hauteur comme une structure bi-dimensionnelles :
 - la hauteur tonale (numéro d'octave),
 - le chroma (classe de hauteur)
- Calcul ?
 - mapping entre les valeurs de la transformée de Fourier f_k et les 12 classes de hauteurs de demi-tons n
 - utilisation directe des valeurs de la Transformée à Q-Constant de classe de Chroma équivalente



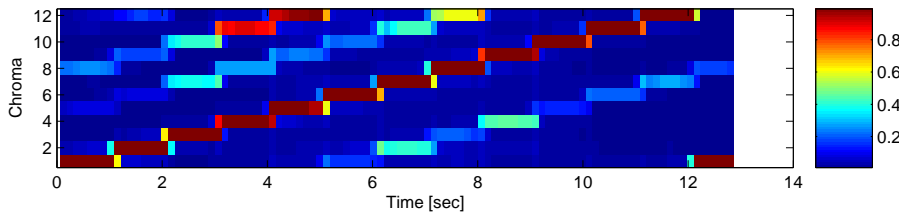
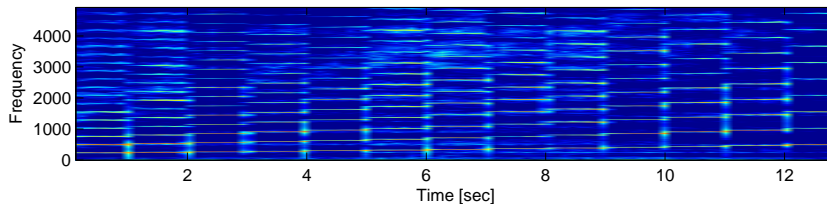
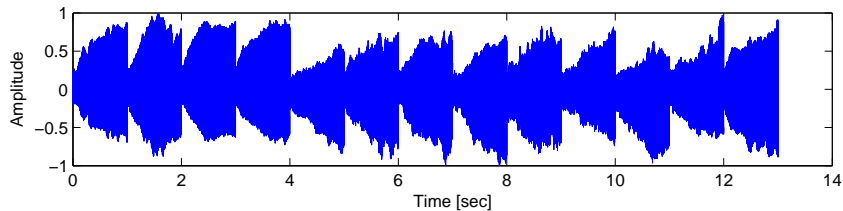
6- Descripteurs audio

6.1- Chroma/ Class de Profil de Class

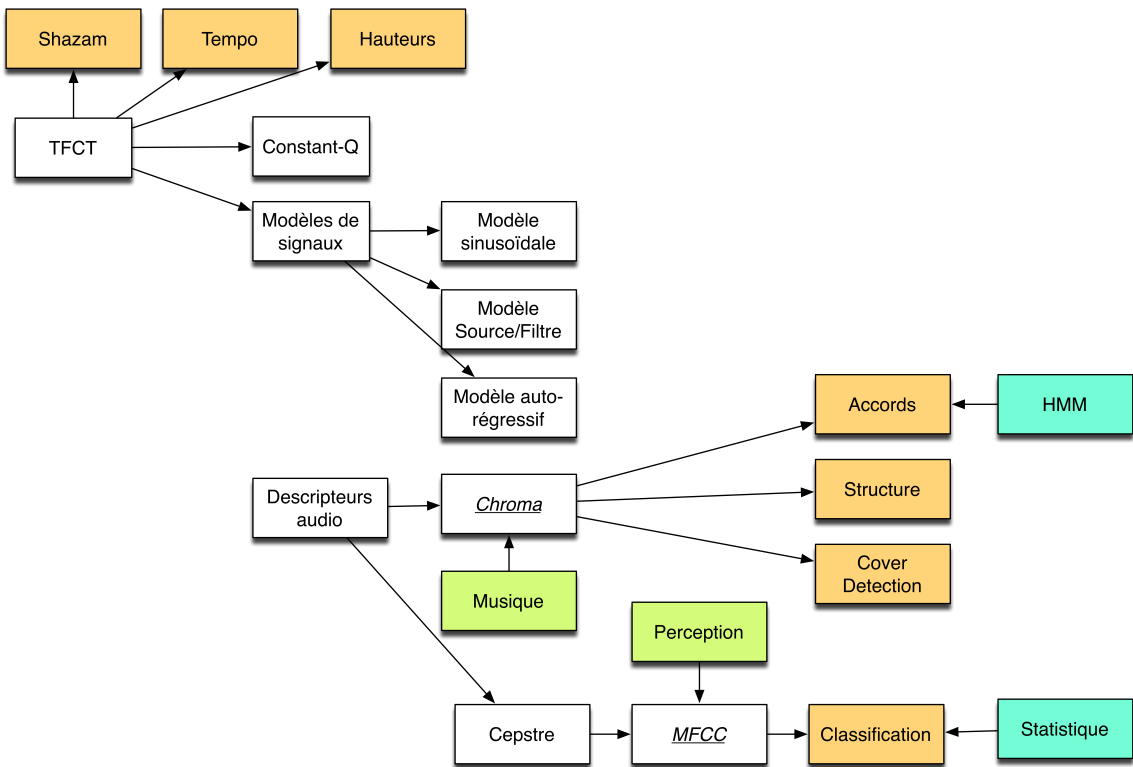


6- Descripteurs audio

6.1- Chroma/ Class de Profil de Class



Plan

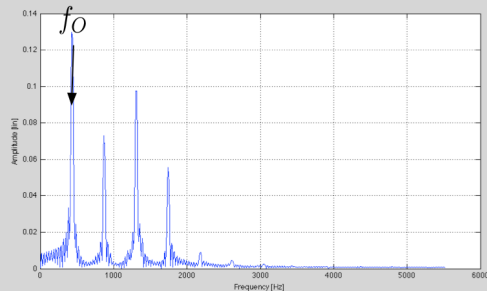
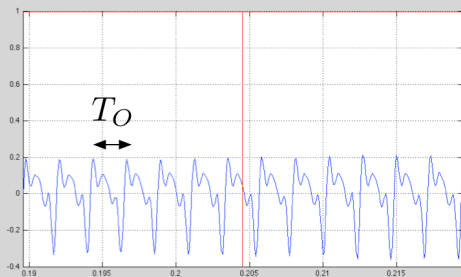


7- Estimation

7.1- Estimation de la hauteur d'une note

Période fondamentale T_0 ou fréquence fondamentale f_0

- f_0 : fréquence fondamentale en Hz
 - exemple La3/A4= 440Hz
- $T_0 = \frac{1}{f_0}$: période fondamentale en secondes
 - exemple La3/A4 = 0.0023s.

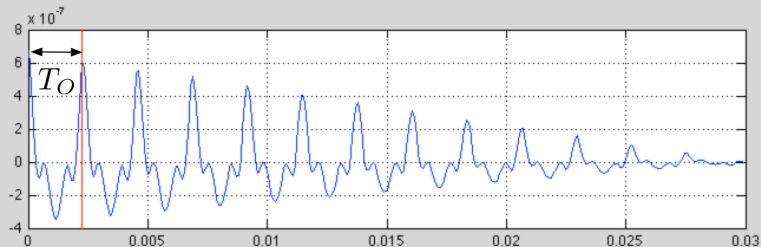
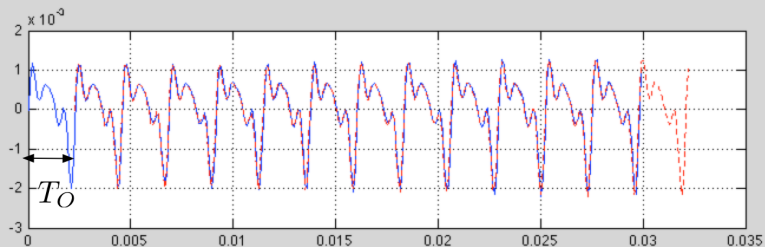


7- Estimation

7.2- Méthodes temporelles

Méthode de l'autocorrélation

- $r(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} x(n)x(n+m)$ si $m \geq 0$
- $r(m)$ est maximum si $m = T_0$

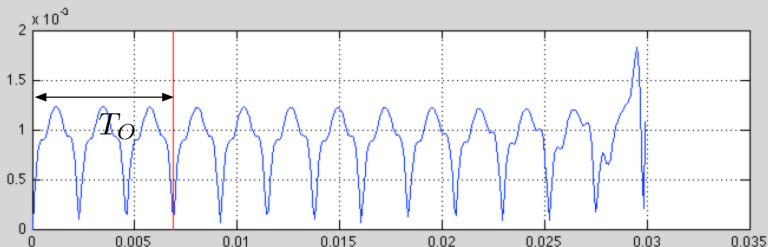
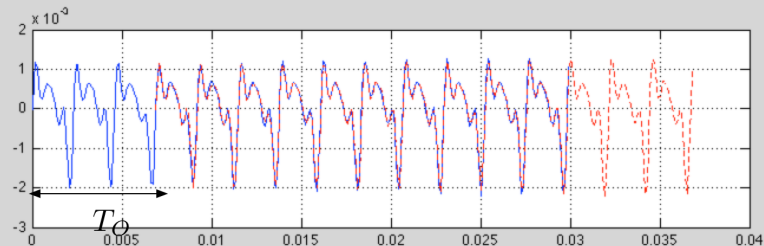


7- Estimation

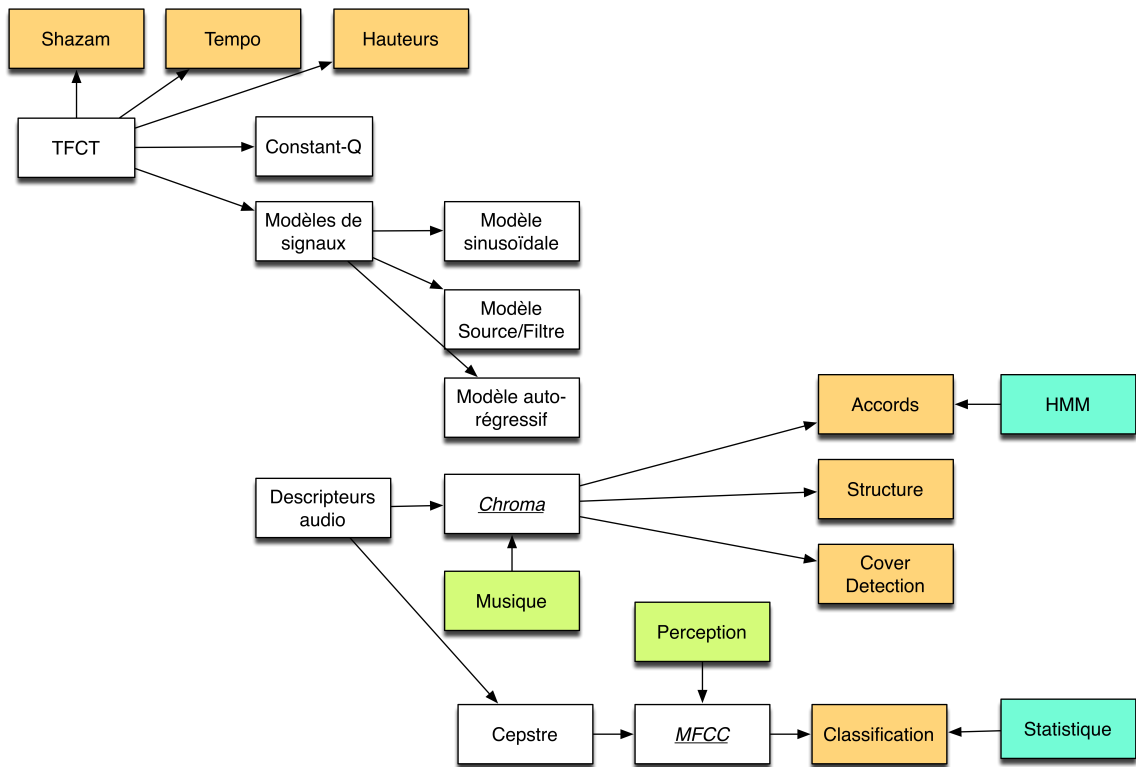
7.2- Méthodes temporelles

Méthode de l'Average Mean Difference Function (AMDF)

- $AMDF(m) = \frac{1}{N-m} \sum_{n=0}^{N-1-m} |x(n) - x(n+m)|$
- $AMFT(m) = 0$ si $m = T_0$



Plan



7- Estimation

7.3- Estimation du tempo

Rythme ?

- Tempo (beat)
 - indiquer sur une partition
 - "vitesse moyenne à laquelle les gens tapent du pied en écoutant la musique"
- Subdivision du rythme
 - mesure
 - entre deux barres, le groupement des noires
 - tactus
 - généralement la noire → le tempo
 - tatum
 - la vitesse la plus rapide
 - la subdivision de la noire en croches, triple-croches, double-croches

Andante grazioso (♩ = 120)

Beats Beats divided Beat subdivided

1 2 3 1 2 3 1 2 3

Beats Beats divided Beat subdivided

1 2 3 1 2 3 1 2 3

Beats Beats divided Beats subdivided

1 2 3 4 1 2 3 4 1 2 3 4

Beats Beats divided Beats subdivided

1 2 3 4 1 2 3 4 1 2 3 4

7- Estimation

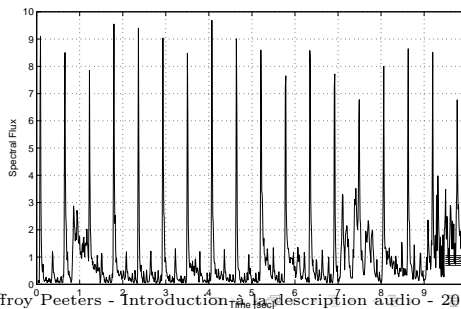
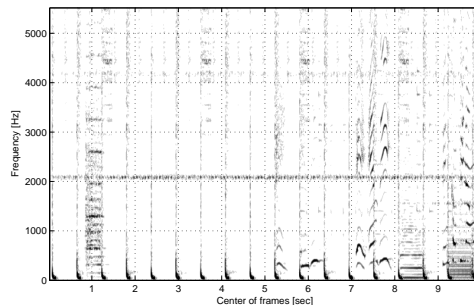
7.3- Estimation du tempo

Estimation du tempo ?

- Détecter la périodicité des évènements dans le signal audio

Détection des évènements ?

- Début des évènements = onsets
- Méthode 1
 - détecter les maxima locaux de la fonction d'énergie du signal
 - $ener(m) = \sum_k X(k, m)^2$
- Méthode2
 - détecter les maxima locaux du flux spectral :
 - $flux(m) = \sum_k hwr[X(k, m) - X(k, m - 1)]$
 - $hwr(x) = x$ if $x > 1$
 - $hwr(x) = 0$ otherwise



7- Estimation

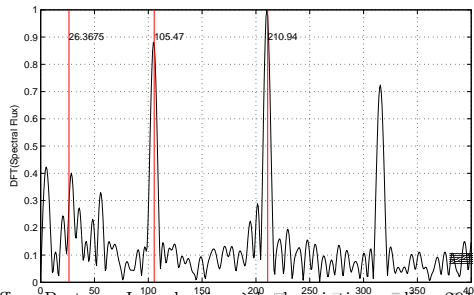
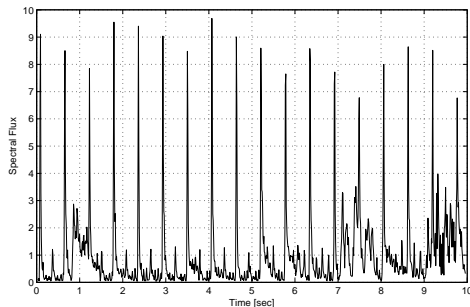
7.3- Estimation du tempo

Périodicité des évènements ?

- Calcul de la transformée de Fourier (DFT) du flux spectral :
 - $FLUX(k) = \sum_{n=0}^{N-1} flux(n) \cdot \exp(j2\pi \frac{k}{N}n), \forall k$
- Calcul de la fonction d'auto-corrélation du flux spectral

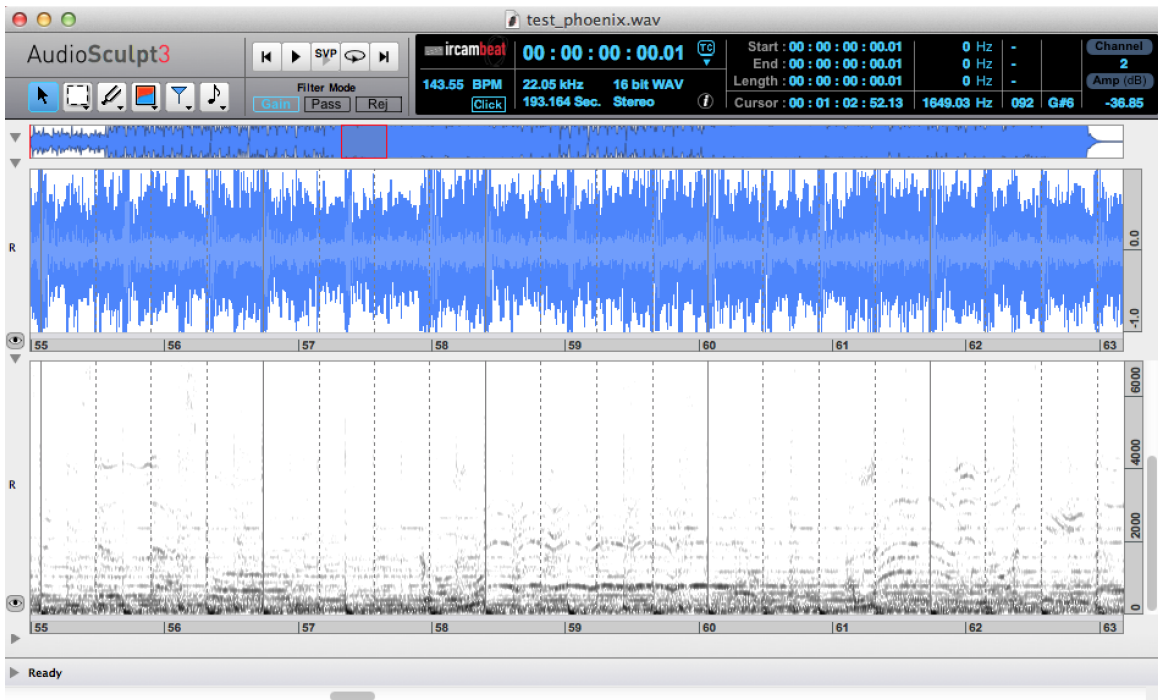
Estimation du tempo ?

- Détecter le peak (fréquence f_k de la DFT) correspondant au tempo
 - Tempo = $60f_k$ (BPM : Battement par Minute)
 - Peaks correspondant à la mesure, au tactus, au tatum

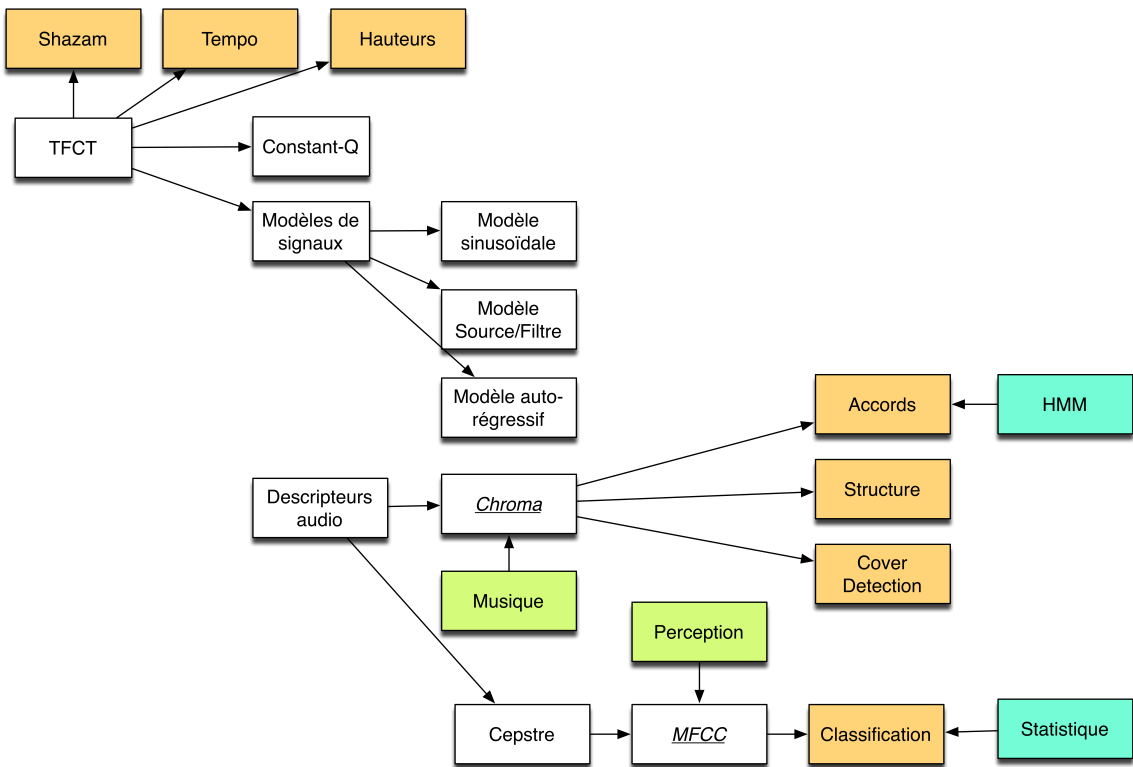


7- Estimation

7.3- Estimation du tempo



Plan

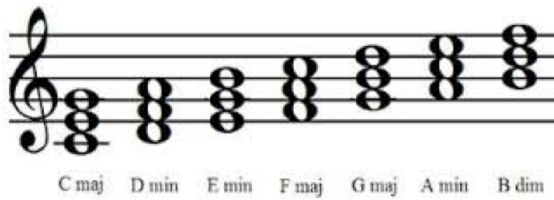


8- Applications

8.1- Estimation d'accords

Les accords ?

- réduction du contenu harmonique à un instant donné d'un morceau
 - ensemble de notes jouées (quasi) simultanément
- différents types d'accords



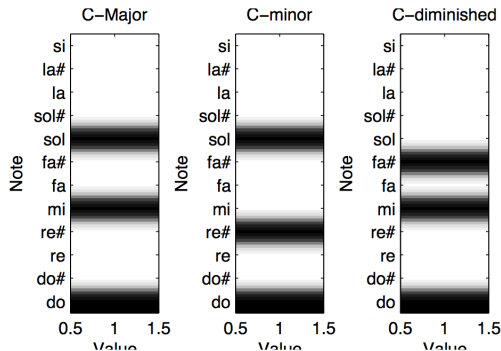
Root-note of the chord	Type of the chord
C, C#, D, D# ... B	Triads: major (CM: c,e,g) , minor (Cm: c, eb, g), suspended (Csus2: c, d, g / Csus4: c, f, g), augmented (Caug: c, e, g#), diminished (Cdim: c, eb, gb),
	Tetrads: major 7 (CM7: c, e, g, b), minor 7 (Cm7: c, eb, g, bb), dominant 7 (C7: c, e, g, sib), major 6 (CM6: c, e, g, a), minor 6 (Cm6: c, eb, g, a) ...
	Pentads: major 9 (CM9: c, e, g, b, d), dominant 9 (C9: c, e, g, bb, d) ...

8- Applications

8.1- Estimation d'accords

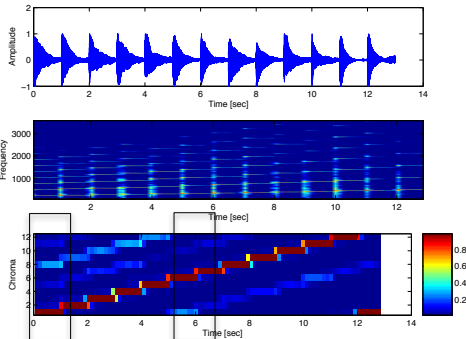
Estimation des accords ?

- Représentation des accords dans un ordinateur
 - Création d'un vecteur à 12 dimensions (les 12 notes)
 - valeur 1 si la note est présente dans l'accord
 - valeur 0 si la note est absente
 - Gabarits $G_a(n)$
 - a le nom de l'accord
 - n la note $n \in [1, 12]$
 - Vecteur de Chroma/PCP $C(n, m)$ à l'instant m
- Calcul d'une distance entre gabarit d'accord et vecteur de chroma
- Choix de la distance
 - $d(C(n, m), G_a(n))$
 - distance Euclidéenne
 - distance cosinusoidale



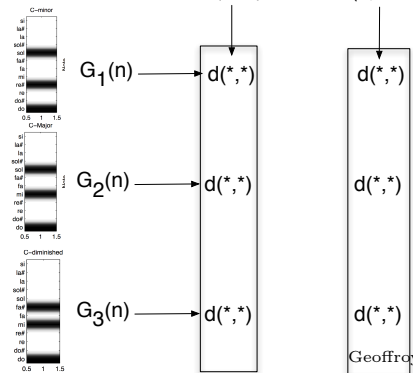
8- Applications

8.1- Estimation d'accords



$C(n,m)$

$C(n,m+1)$



8- Applications

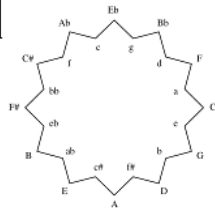
8.1- Estimation d'accords

Décodage de la suite d'accords

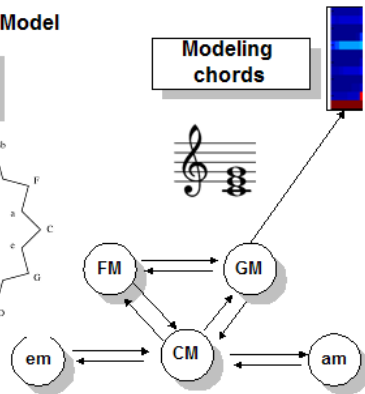
- 1. Les observations :
 - on extrait la séquence de descripteurs audio chroma/PCP
- 2. Les états :
 - les 24 accords qui sont cachés
- 3. Probabilités initiales :
- 4. Probabilités de transition entre accords :
 - suivent la théorie musicale (cercle des quintes, relatifs majeur-mineur) :
 - GMaj vers CMaj (consonance),
 - GMaj vers C#Maj (dissonance)
- 5. Probabilité d'émission des accords :
 - $p(S = CMaj|chroma)$,
 - $p(S = DbMaj|chroma)$, ...

Hidden Markov Model

Modeling transition between

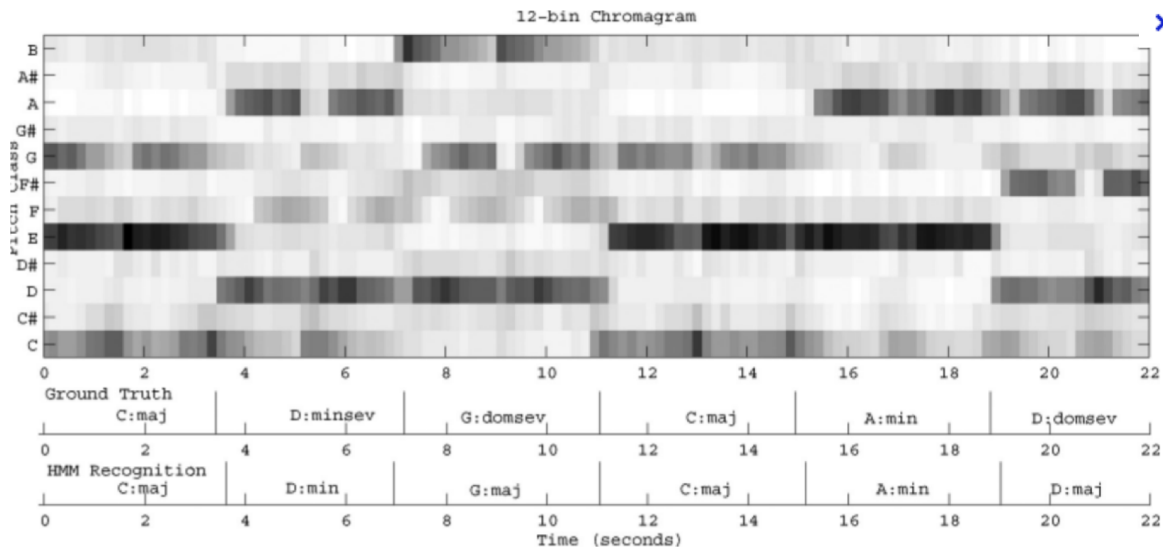


Modeling chords

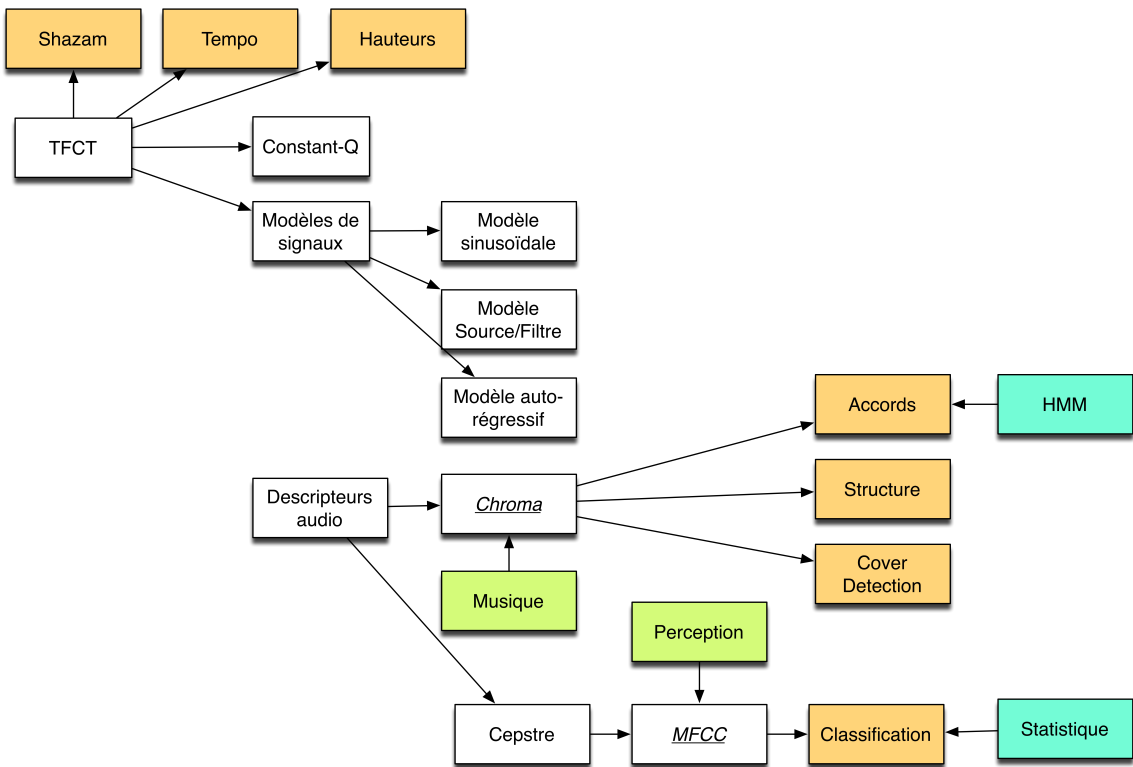


8- Applications

8.1- Estimation d'accords



Plan



8- Applications

8.2- Détection de cover-version

Cover-version ?

- Une reprise
- "Let it be" par The Beatles, Aretha Franklin Joan Baez, ...

Caractéristiques d'une cover-version :

- généralement la même suite harmonique
 - même suite d'accords,
 - même mélodie
- éventuellement transposée

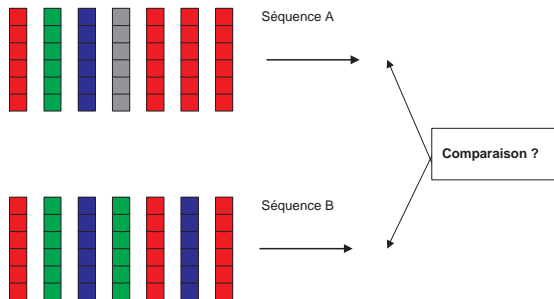
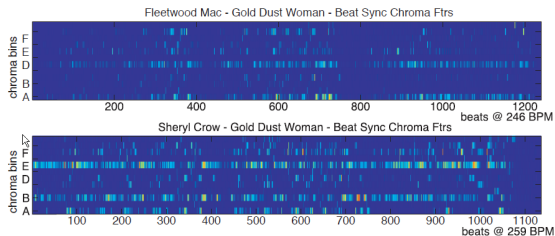
Titre	Artiste	Album	D.	Pop.	
Let It Be	▼ The Beatles Recovered Band	30 Beatles Top Hits	03:50		□
Let It Be	▼ The Hit Co., The Tribute Co.	A Tribute to the Beatles: The Lat...	03:42		□
Let It Be	▼ Labrinth	Let It Be	03:05		□
Let It Be Me	▼ Ray LaMontagne	Gossip in The Grain	04:41		□
Let It Be - The Beatles Tribute	Let It Be	Let It Be - The Beatles Tribute	03:49		□
Let It Be	Lois	Let It Be - The Voice 2	03:15		□
Let It Be	The Yesteryears	A Tribute to #1 Beatles Hits - T...	03:48		□
Let It Be	▼ Aretha Franklin	This Girl's In Love With You	03:33		□
Let It Be Sung	▼ Jack Johnson, Matt Costa, Zach Gill,...	If I Had Eyes	04:09		□
Let It Be	Vox Angeli	Gloria	03:26		□
Let It Be	▼ Paul McCartney	Good Evening New York City	03:54		□
Hey Jude	Let It Be	Hey Jude	03:55		□
Let It Be	Joan Baez	Greatest Hits And Others	03:51		□

8- Applications

8.2- Détection de cover-version

Méthode

- Chaque morceau est représenté par la séquence temporelle de ses chromas/PCP : $C(n, m)$
 - n est la pitch-class et
 - m est le temps
- Pour une collection de morceaux, on compare les morceaux deux à deux
 - Calcul du coût pour aligner $C_1(n, m)$ et $C_2(n, m)$
- Si le coût d'alignement est faible, il s'agit vraisemblablement d'une cover ou ... d'un plagia
- Technique utilisée :
 - Alignement Dynamique du Temps

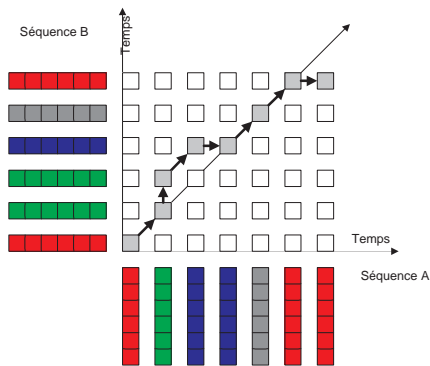


8- Applications

8.2- Détection de cover-version

Alignement de deux séquences temporelles A et B

- Hypothèse :
 - début et fin en correspondances
- Méthode
 - Parcours progressif de tous les points (a,b) de la matrice d'alignement
 - En un point (a,b) , recherche du meilleur chemin (de coût local minimal) pour y arriver parmi :
 - $(a-1, b-1) \rightarrow (a,b)$: chemin normal
 - $(a, b-1) \rightarrow (a,b)$: insertion
 - $(a-1, b) \rightarrow (a,b)$: deletion
 - Coût local =
 - coût du point précédent (cumul)
 - + coût défavorable si ins. ou del.



chemin normal

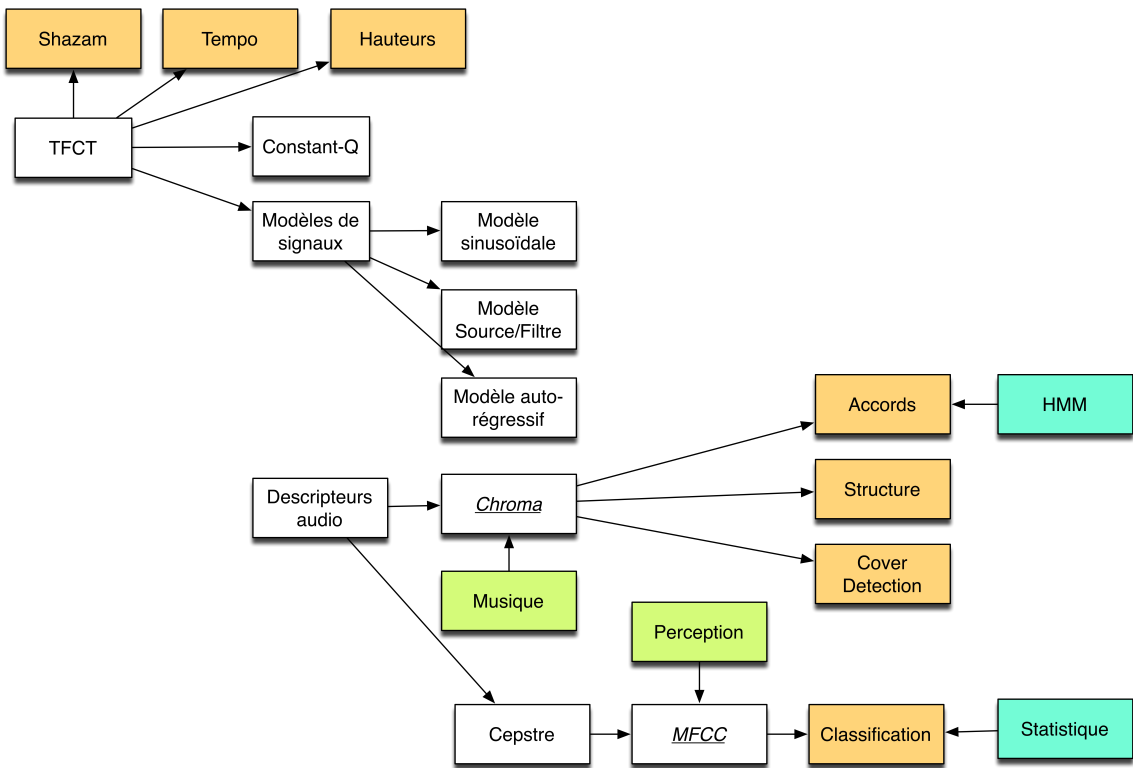


insertion



deletion

Plan



8- Applications

8.3- Identification audio

Identification audio ?

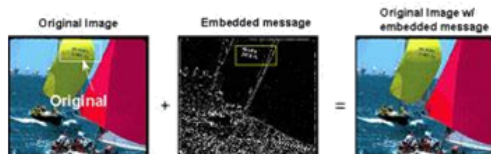
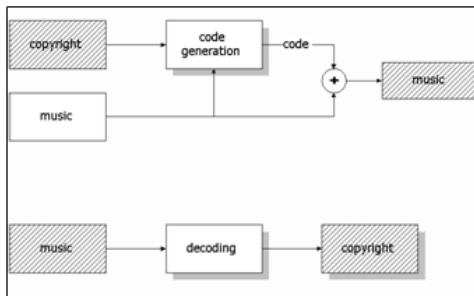
- Objectif :
 - Reconnaître un morceau diffusées sur radio, télé, Internet, bar, discothèque, ...
 - Identifier l'enregistrement (ISRC), pas l'oeuvre (ISWC)

8- Applications

8.3- Identification audio

Méthode 1 : Watermarking

- Codage :
 - introduction d'un code identifiant robuste mais inaudible dans le signal sonore
- Décodage :
 - pour un nouveau signal : extraction du code (si il est présent) et recherche de ce code dans une base de données

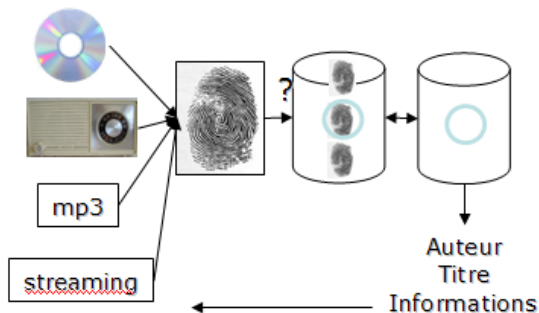


8- Applications

8.3- Identification audio

Méthode 2 : Fingerprint

- Shazam, Midomi, Philips, ...
- Codage :
 - prise d'empreinte du signal, stockage dans une base de données
- Décodage :
 - pour un nouveau signal, prise d'empreinte, comparaison avec les empreintes de la base de données
- Challenge :
 - déterminer un ensemble réduit de descripteurs audio extraits du signal sonore permettant d'identifier de manière unique un extrait musical

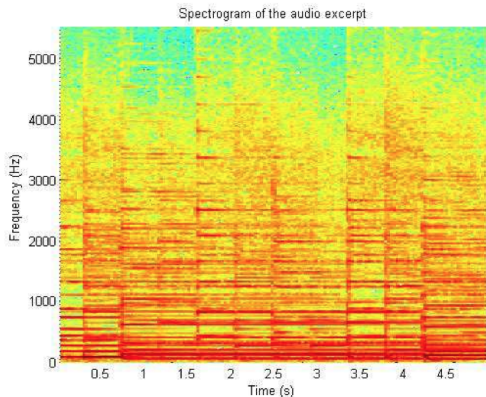


8- Applications

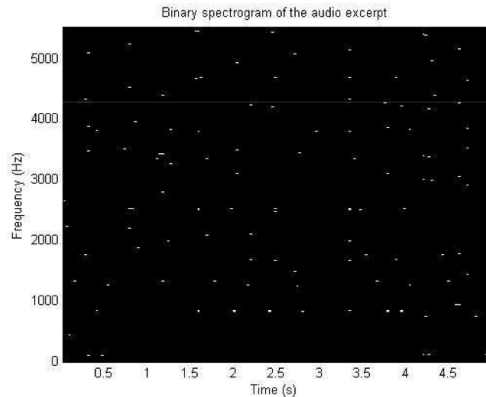
8.3- Identification audio

Fingerprint : algorithme de Shazam

- Extraction de points saillants dans le plan temps/fréquence
 - onsets de sinusoides, maxima locaux
 - "constellation points"



2D
peak
picking



8- Applications

8.3- Identification audio

Fingerprint : algorithme de Shazam

- Représentation des "constellation points" :
 - chaque point est pris comme un "anchor point" ayant une "target zone"
 - stockage :
 - fréquences des points f_1, f_2
 - le Δt du temps $t_2 - t_1$
 - (le temps de l'anchor t_1)

8- Applications

8.3- Identification audio

Fingerprint : algorithme de Shazam

- Matching :
 - pour chaque points correspondants dans la base de données on mesure le décalage temporel : $time_{db} - time_{sample} = constant$
 - puisque le décalage doit être constant, on calcul l'histogramme des Δt , il doit présenter un peak

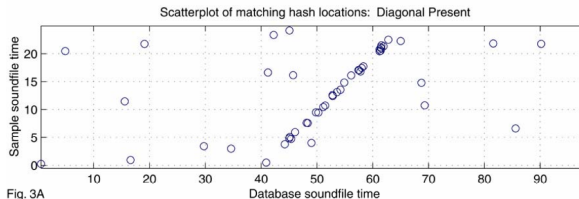


Fig. 3A

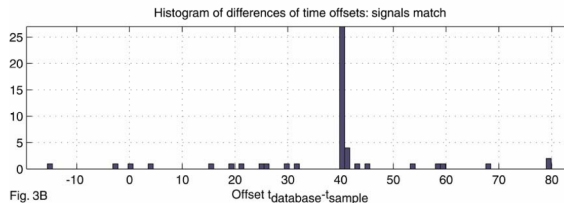


Fig. 3B

8- Applications

8.3- Identification audio

