



Indexation des signaux musicaux polyphoniques



G. Richard
Professeur
Télécom ParisTech



« Licence de droits d'usage »
http://formation.enst.fr/licences/pedago_sans.html

Contenu

- **Indexation audio : quels intérêts ?**
- **Quelques travaux:**
 - Découvrir une caractéristique du signal: Application à la reconnaissance des instruments de musique
 - Découvrir une information tonale: Détection de hauteur
 - Découvrir une information rythmique: Tempo
- **Mais que faire avec les signaux polyphoniques !**
 - Découvrir l'information « globalement »
 - ...ou utiliser une décomposition ou séparation du son
- **Conclusion**



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Indexation audio : intérêts

- **Nouveaux challenges pour la société de l'information:**
 - Volume considérable de données numériques multimedia disponibles
 - L'accroissement rapide et continu de ces données numériques (qu'elles se trouvent sur le réseau Internet ou dans des bases personnelles)
 - Généralisation de leur utilisation pour de nombreuses applications

Diminution de « l'accessibilité » des données

Un fort besoin pour de nouvelles méthodes efficaces d'indexation, de classification et d'accès par le contenu.

- **L'indexation automatique vise ainsi à extraire du flux numérique des descripteurs de haut niveau permettant de réaliser par la suite une classification ou un accès à l'information par son contenu.**



Gaël RICHARD

JIM 2010, Rennes



Droits d'usage autorisé

Pourquoi analyser le signal musical ?

- **Rechercher par le contenu**
 - À partir d'un morceau ...
 - À partir d'un chantonnement...
 - De nouveaux morceaux à partir de ce que j'aime....
 - Une nouvelle version d'un air connu ..
 - Une vidéo qui « va bien » avec l'audio
 - ...

- **Nouvelles applications**
 - Playlist « sémantiques » (jouer des morceaux de plus en plus rapide...)
 - Karoké « intelligent » (l'accompagnement suit le chanteur...)
 - Prédire le potentiel succès d'un titre
 - Aide au mixage, Djing,
 - Ecoute active,...



Google



Recherche à la voix



Jogging musical






Gaël RICHARD

JIM



Droits d'usage autorisé



TELECOM
PARIS
école nationale
supérieure des
télécommunications

Comment retrouver des musiques par leur rythme ?



poum ta Poum Poum ta

Exprimez dans le microphone
la séquence rythmique
que vous souhaitez...



RECONNAISSANCE
VOCALE

ENREGISTREMENT
D'UNE NOUVELLE BOUCLE

BASE DE BOUCLES
DE BATTERIE

Reconnaissance automatique
(cymbales, grosse caisse...)
Extraction du tempo

... Le système trouve,
dans sa base de données
de boucles de batterie,
toutes celles qui
correspondent à votre choix.



BOUCLE DE BATTERIE
TROUVÉE

CONTACT : Olivier Gillet, Gaël Richard
POUR EN SAVOIR PLUS : www.telecom-paris.fr



CNRS
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE



Analyse des signaux monophoniques

Un seul instrument joue à la fois...et une seule note



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Systemes de classification

- **Plusieurs problèmes, une même approche**
 - Reconnaissance automatique du genre musical.
 - Reconnaissance des instruments utilisés dans un morceau.
 - Classification d'échantillons sonores.
 - Etiquetage d'une bande son (dialogues, scènes d'action, musique, effets spéciaux).
 - Organisation d'une collection musicale selon les habitudes d'écoute.
 - Détection de "hits" potentiels.

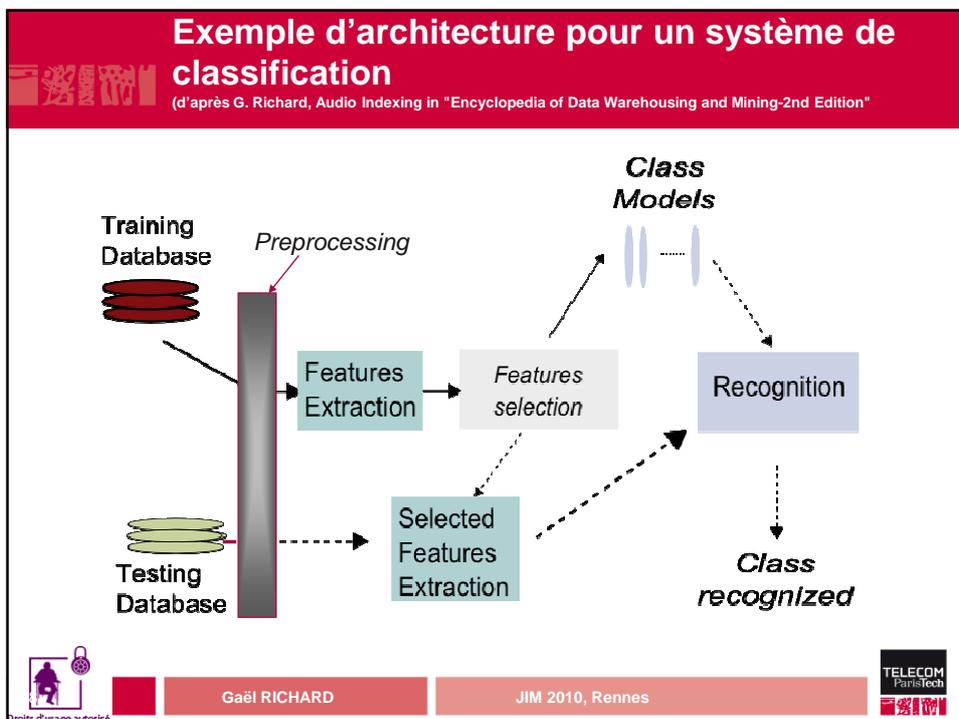


Gaël RICHARD

JIM 2010, Rennes



Droits d'usage autorisé



Le timbre des instruments de musique

- **Le timbre des instruments de musique**
 - Une définition possible: « l'attribut de la sensation auditive qui permet de différencier 2 sons de même hauteur et de même intensité »
- « Espace » de timbre

Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes

Le timbre des instruments de musique

- Espace 3D (Krumhansl, 1989, McAdams, 1992)

- BSN = basson
- CAN = cor anglais
- CNT = clarinet
- GTR = guitar
- HRN = cor
- HRP = harpe
- TPT = trompette
- PNO = piano
- VBS = vibraphone

Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Découvrir l'information rythmique....



Droits d'usage autorisé

Gaël RICHARD

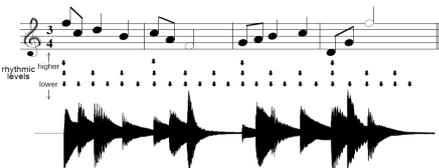
JIM 2010, Rennes





Extraction du rythme ou du Tempo

- **Le rythme: concept musical intuitivement simple à comprendre mais difficile à définir !!**



- **Handel (1989): « *The experience of rhythm involves movement regularity, grouping and yet accentuation and differentiation* »**
- **le rythme d'un signal écouté n'a pas nécessairement une interprétation unique !!**
- **On définit fréquemment la pulsation (beat en anglais)**



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Extraction du rythme ou du Tempo

Signal + Onsets

« Fonction de détection »

Autocorrélation

Suivi du tempo

Tempo « à la noire »

Gaël RICHARD

JIM 2010, Rennes

Droits d'usage autorisé

Extraction du rythme: un indice intéressant

- Recherche de musique par similarité (ou classification par genre) genre classification)
- Recherche de contenus audio adaptés à la vidéo

Demo

- Reference

O. Gillet and G. Richard, « On the Correlation of Automatic Audio and Visual Segmentations of Music Videos » | IEEE Trans. on CSVT, 2007

Gaël RICHARD

JIM 2010, Rennes

Droits d'usage autorisé



Découvrir une information tonale

détection de hauteur



Gaël RICHARD

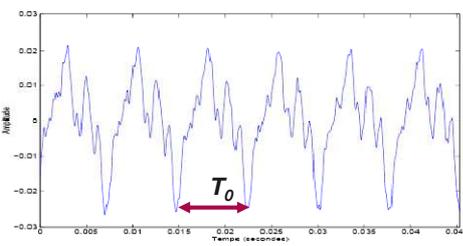
JIM 2010, Rennes



Droits d'usage autorisé



Un son quasi-périodique

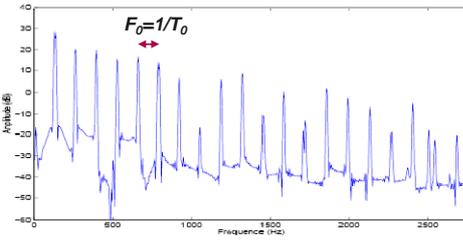


Amplitude

Temps (secondes)

T_0

Son de piano (C3)



Amplitude

Fréquence (Hz)

$F_0 = 1/T_0$

Spéctre du son de piano




Droits d'usage autorisé

Découvrir une information tonale

- Plusieurs notes : Comment faire pour extraire l'ensemble des notes d'un enregistrement polyphonique ?

Spectre d'amplitude de l'accord Do - Fa# (hautbois)

Spectre d'Amplitude (dB)

Fréquences (Hz)



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Vers l'analyse des signaux polyphoniques

Plusieurs notes, Plusieurs rythmes, Plusieurs instruments à la fois....

Vers l'analyse des signaux polyphoniques

Plusieurs notes, Plusieurs rythmes, Plusieurs instruments à la fois....



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Vers l'analyse des signaux polyphoniques

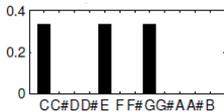
- **2 directions envisageables:**
 - « Reconnaître » globalement (accords, quartet, etc...)
 - « Reconnaître » après décomposition du signal ou séparation de sources


Droits d'usage autorisé
Gaël RICHARD
JIM 2010, Rennes


Découvrir une information tonale...

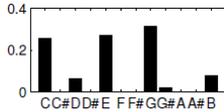
- **Reconnaître globalement**
 - Reconnaissance des accords
 - Utilisation de templates (Exemples de chromagrammes théoriques)

Do Majeur (1 harmonique)



Note	Amplitude
C	0.35
D	0.35
E	0.35
F	0
G	0
A	0
B	0

Do Majeur (6 harmoniques)



Note	Amplitude
C	0.25
D	0.1
E	0.25
F	0.3
G	0.05
A	0.1
B	0


Droits d'usage autorisé
Gaël RICHARD
JIM 2010, Rennes


Découvrir une information timbrale

■ Reconnaître globalement

- Utiliser des classes mixtes (par exemple classe « violon+violoncelle »,...)
- Pour plus d'instruments, on peut suivre une approche hiérarchique.....
- Mais aussi apprendre automatiquement les classes intermédiaires...

Droits d'usage autorisé

Gaël RICHARD JIM 2010, Rennes

TELECOM ParisTech

Découvrir une information timbrale

■ Exemples de classes pour quartet de Jazz

- Bs= Basse ; Dr = Drums ; Pn = Piano ; Eg= Electric Guitar; Tr = Trompette; V= Voix ;

S. Essid, G. Richard, B. David. *Instrument recognition in polyphonic music based on automatic taxonomies*. IEEE Trans. on Audio, Speech, and Language Proc. 14 (2006), no. 1

Droits d'usage autorisé

Gaël RICHARD JIM 2010, Rennes

TELECOM ParisTech

« Reconnaître » après décomposition du signal ou séparation de sources



Gaël RICHARD JIM 2010, Rennes

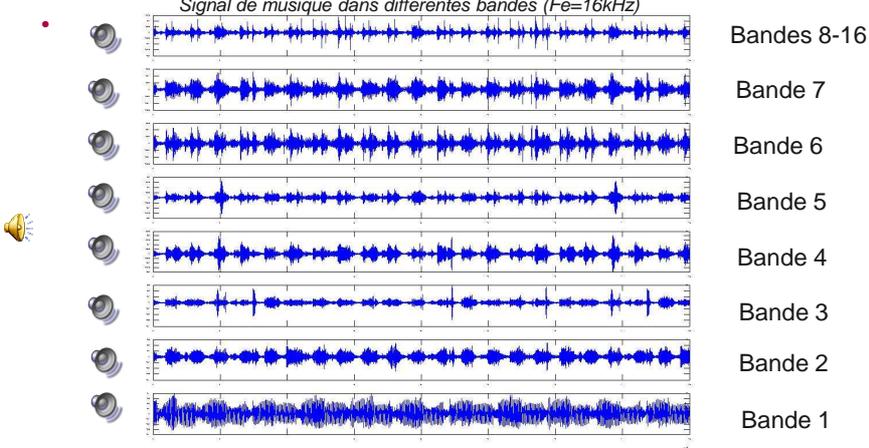


Droits d'usage autorisés

Découvrir l'information rythmique

■ Décomposer le signal en bandes de fréquences...

Signal de musique dans différentes bandes (Fe=16kHz)



Bandes 8-16
Bande 7
Bande 6
Bande 5
Bande 4
Bande 3
Bande 2
Bande 1



Gaël RICHARD JIM 2010, Rennes

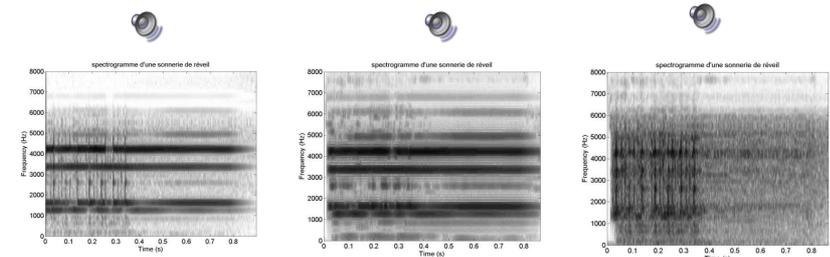


Droits d'usage autorisés

Découvrir l'information rythmique

- Décomposition « harmoniques » + bruit

• Original = Somme de sinusoïdes + Bruit



Exemples fournis par R. Badeau



Gaël RICHARD

JIM 2010, Rennes



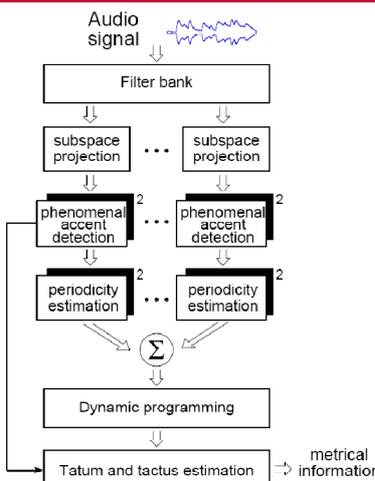
Droits d'usage autorisés

Découvrir l'information rythmique

- Un exemple pour l'extraction robuste du tempo
 - ➔ Utilisation d'un banc de filtres
 - ➔ Utilisation d'une décomposition harmoniques / bruit
 - ➔ Utilisation d'un suivi dynamique du tempo (programmation dynamique)



video_tracking.mp4



M. Alonso, G. Richard, B. David, "Accurate tempo estimation based on harmonic+noise de composition", *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 82795, 14 pages, 2007



Gaël RICHARD

JIM 2010, Rennes



Droits d'usage autorisés

Découvrir l'information tonale

- **Reconnaître toutes les notes itérativement...**
 - Reconnaître la note la plus forte ...
 - La soustraire de l'accord
 - Reconnaître la note suivante
 - La soustraire de l'accord
 - Etc... tant qu'il y a des notes dans l'accord



Droits d'usage autorisé

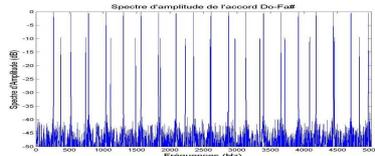
Gaël RICHARD

JIM 2010, Rennes



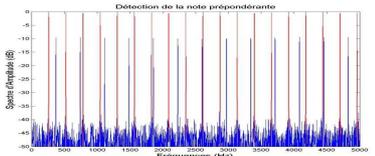
Découvrir l'information tonale

Accord de deux notes de synthèse Do – Fa#



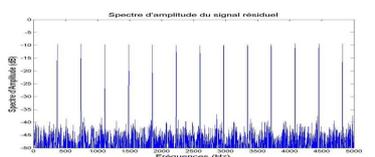
Spectre d'amplitude (dB)
Fréquences (Hz)

Détecter la note prépondérante (en rouge)



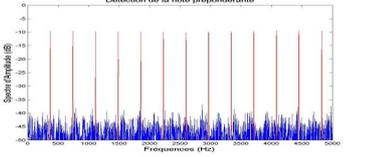
Spectre d'amplitude (dB)
Fréquences (Hz)

Soustraire la note détectée



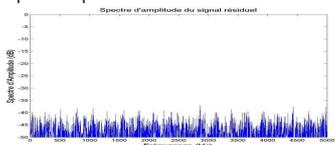
Spectre d'amplitude (dB)
Fréquences (Hz)

Détecter la note suivante....



Spectre d'amplitude (dB)
Fréquences (Hz)

Il n'y a plus de notes...l'accord de Fa# a été reconnu



Spectre d'amplitude (dB)
Fréquences (Hz)



Droits d'usage autorisé

Gaël RIC



Découvrir l'information tonale / timbrale

- **Utilisation de méthodes de décomposition parcimonieuse (ou séparation de sources)**
 - Décomposition atomique des signaux polyphoniques
 - On représente le signal comme une combinaison linéaire d'atomes $h_\lambda(t)$ pris dans un dictionnaire fixe:

$$x(t) = \sum_{\lambda \in \Lambda_x} \alpha_\lambda h_\lambda(t)$$



Droits d'usage autorisé

Gaël RICHARD

JIM 2010, Rennes



Découvrir l'information tonale / timbrale

- **La décomposition atomique est obtenue par exemple par "matching pursuit" :**
 - L'atome le plus proéminent (*i.e. le plus corrélé avec le signal*) est extrait et soustrait du signal original.
 - Itérer la procédure tant qu'un nombre prédéfini d'atomes n'a pas été extrait (*ou qu'un rapport signal à bruit prédéfini a été atteint*)

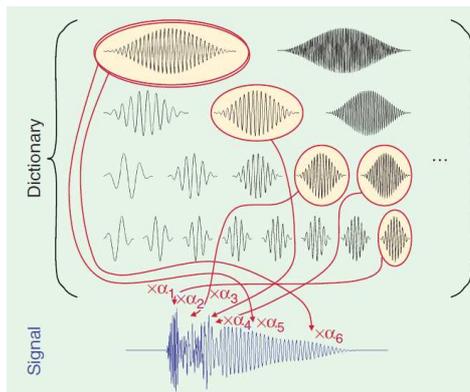


Figure d'après L. Daudet: *Audio Sparse Decompositions in Parallel*, IEEE Signal Processing Magazine, 2010



Droits d'usage autorisé

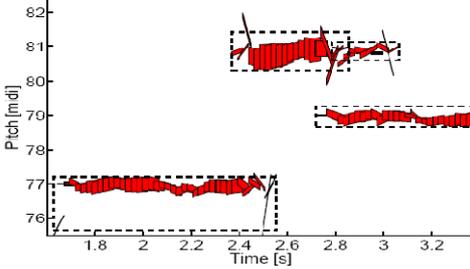
Gaël RICHARD

JIM 2010, Rennes



Découvrir l'information tonale / timbrale

- Utiliser des atomes informés: e.g. caractéristiques d'une hauteur tonale et d'un instrument
- Introduction de contraintes de continuité pour construire des "molécules" (note jouée par un instrument)



Pitch [mjd]

Time [s]

Demo from P. Leveau



Testfcl.mov

P. Leveau, E. Vincent, G. Richard and L. Daudet, « Instrument-Specific Harmonic Atoms for Mid-Level Musical Audio Representation » IEEE Trans. on ASLP, Volume 16, N°1 Jan. 2008 Page(s):116 - 128


Droits d'usage autorisés

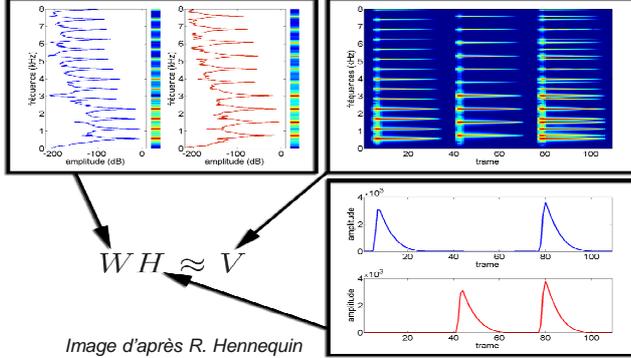
Gaël RICHARD

JIM 2010, Rennes



Découvrir l'information tonale / timbrale

- Utilisation de méthodes de décomposition non supervisées (par exemple par factorisation en matrices non-négatives : NMF)
- Principe de la NMF :



$WH \approx V$

Image d'après R. Hennequin


Droits d'usage autorisés

Gaël RICHARD

JIM 2010, Rennes



Découvrir l'information tonale / timbrale

- **Extraction de la mélodie principale:**
 - Utilisation de contraintes dans la décomposition :

$$X = V + M$$
 - Contrainte pour la voix : elle respecte un modèle source/filtre
 - Contraintes de continuité de la courbe mélodique et "timbrale"

```

graph LR
    A["(Glottal) Source"] --> B["(Vocal Tract) Filter"]
    B --> C["Voice Signal (V)"]
  
```

J-L Durrieu, G. Richard, B. David, C. Févotte, *Source/Filter Model for Unsupervised Main Melody Extraction From Polyphonic Audio Signals*, IEEE Transactions on ASLP, March 2010.

Gaël RICHARD
JIM 2010, Rennes

Droits d'usage autorisé

Conclusion

- **Complexité de la tâche d'indexation pour les signaux polyphoniques.**
 - Mais on a déjà des résultats encourageants...
- **Intérêt d'utiliser des méthodes de décomposition parcimonieuse et de la séparation de sources ...**
 - La séparation aide la transcription et l'analyse en général...
 - Mais la séparation marche bien quand on a des connaissances sur les sources...
 - Connaissances a priori, bien sur..
 - Mais aussi connaissances extraites du signal...l'indexation aide la séparation...

Gaël RICHARD
JIM 2010, Rennes

Droits d'usage autorisé



■ **Merci pour votre attention**

■ **Pour plus de détails, voir par exemple:**

- <http://www.tsi.telecom-paristech.fr/aao/>
 - Travaux du groupe AAO de Telecom ParisTech
 - Logiciels et bases de données distribuées
 - Liens vers les membres du groupe
- www.telecom-paristech.fr/~grichard/
 - Publications téléchargeables



Droits d'usage autorisés

Gaël RICHARD

JIM 2010, Rennes

