

Analyse de scènes sonores pour le traitement des signaux musicaux

Mathieu Lagrange



October 16, 2009

Je me présente, je m'appelle...

- Nom : Mathieu Lagrange
- Équipe : Analyse / Synthèse
- Statut : Cr2 CNRS
- Formation : Informaticien

La technologie, un monde de défis

“**Quarante-deux** ! cria Loonquawl. Et c'est tout ce que t'as à nous montrer au bout de sept millions et demi d'années de boulot ?”

D. Adams

- Besoins :
 - capture
 - transmission
 - accès
- But :
 - bit précis
 - bit utile
 - bit pertinent
- Moyens :
 - biologie mécanique
 - psycho-acoustique
 - cognition

La technologie, un monde de défis

“**Quarante-deux** ! cria Loonquawl. Et c'est tout ce que t'as à nous montrer au bout de sept millions et demi d'années de boulot ?”

D. Adams

- Besoins :
 - capture
 - transmission
 - accès
- But :
 - bit précis
 - bit utile
 - bit pertinent
- Moyens :
 - biologie mécanique
 - psycho-acoustique
 - cognition

La technologie, un monde de défis

“**Quarante-deux** ! cria Loonquawl. Et c'est tout ce que t'as à nous montrer au bout de sept millions et demi d'années de boulot ?”

D. Adams

- Besoins :
 - capture
 - transmission
 - accès
- But :
 - bit précis
 - bit utile
 - bit pertinent
- Moyens :
 - biologie mécanique
 - psycho-acoustique
 - cognition

La musique, un monde d'enthousiastes

- La musique :
 - objet : arrangement de sons et de silences **au cours du temps**
 - fonction : forme d'expression
 - individuelle (sentiments)
 - collective (fête, chant, danse)
- Le Musicien :
 - est capable de produire de la musique
 - à partir de langages dédiés
 - et d'un dispositif souvent complexe

La musique, un monde d'enthousiastes

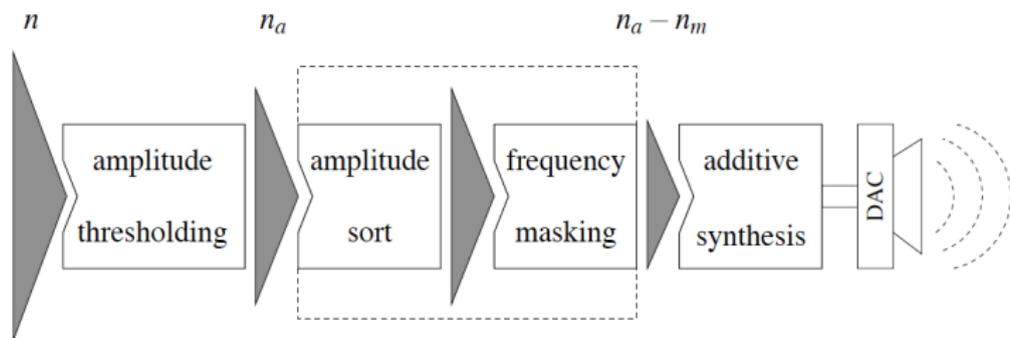
- La musique :
 - objet : arrangement de sons et de silences **au cours du temps**
 - fonction : forme d'expression
 - individuelle (sentiments)
 - collective (fête, chant, danse)
- Le Musicien :
 - est capable de produire de la musique
 - à partir de langages dédiés
 - et d'un dispositif souvent complexe

Travaux Antérieurs

Pourquoi se fatiguer alors qu'ils sont tous sourds ? (LaBRI)

Synthèse sonore à base de sinus [DAFx'01] :

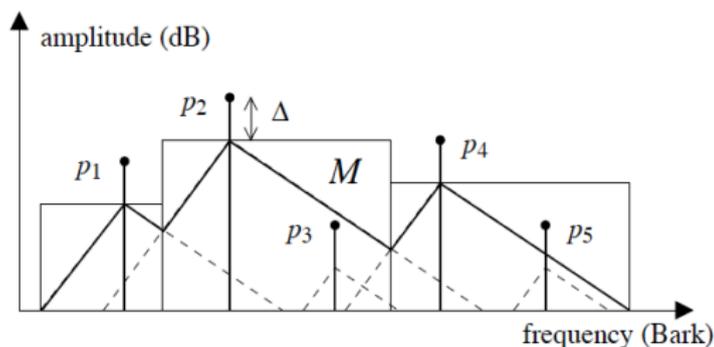
- Sans pré-traitement, la synthèse de larges scènes sonores a un coût de calcul qui est asymptotiquement linéaire en fonction du nombre de composantes sinusoïdales.
- Or, les capacités du système auditif humain sont limitées
 - Seuil d'audition
 - Masquage fréquentiel



Pourquoi se fatiguer alors qu'ils sont tous sourds ? (LaBRI)

Synthèse sonore à base de sinus [DAFx'01] :

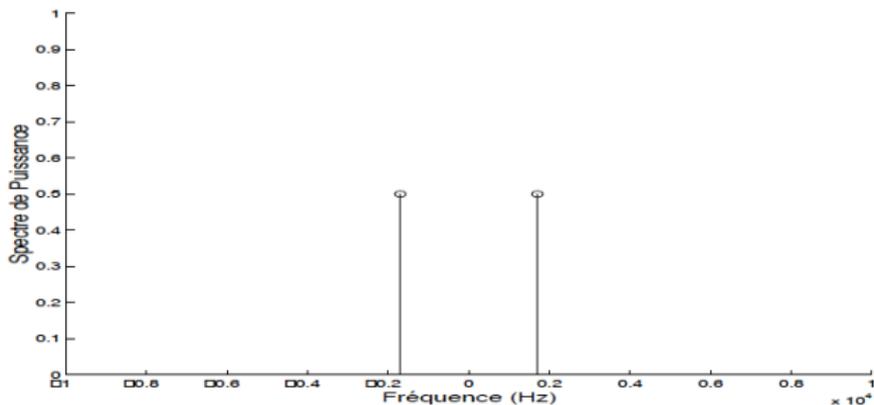
- Sans pré-traitement, la synthèse de larges scènes sonores a un coût de calcul qui est asymptotiquement linéaire en fonction du nombre de composantes sinusoïdales.
- Or, les capacités du système auditif humain sont limitées
 - Seuil d'audition
 - Masquage fréquentiel



De la théorie à la pratique

L'analyse est une affaire de compromis plus ou moins assumés :

- Observation partielle
- Violation de contraintes

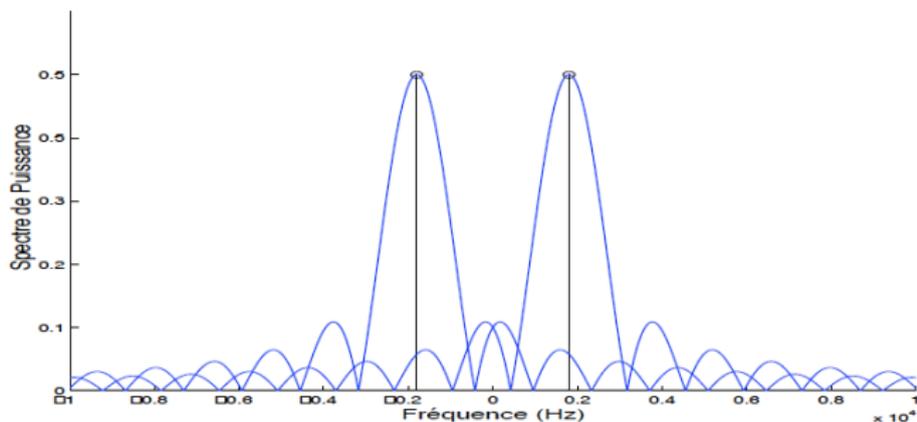


Une sinusoïde de fréquence constante (en temps continu)

De la théorie à la pratique

L'analyse est une affaire de compromis plus ou moins assumés :

- Observation partielle
- Violation de contraintes

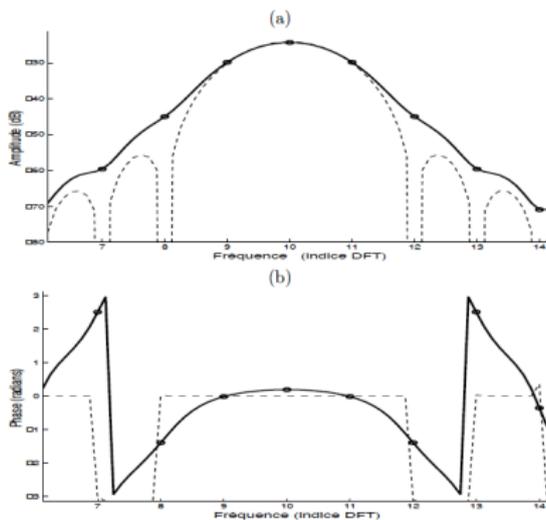


Une sinusoïde de fréquence constante (en temps discret)

De la théorie à la pratique

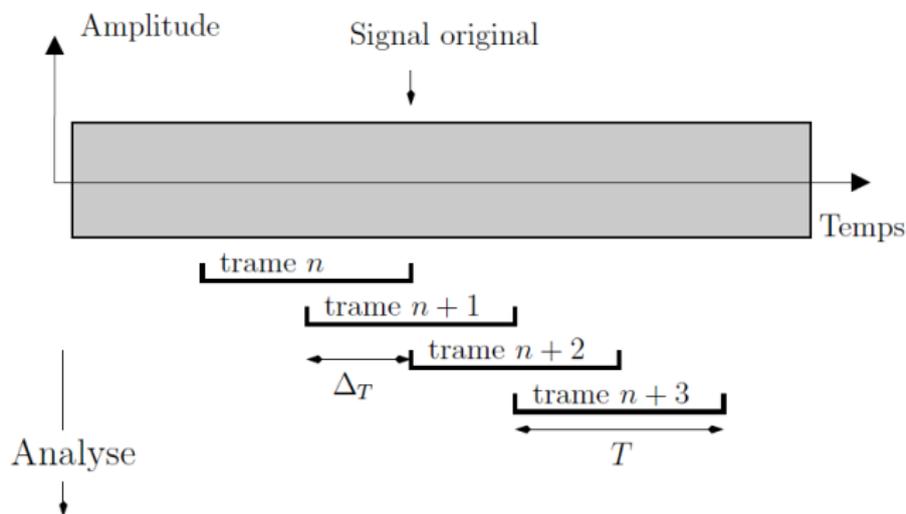
L'analyse est une affaire de compromis plus ou moins assumés :

- Observation partielle
- Violation de contraintes



Une sinusoïde dont la fréquence varie linéairement (cas non-stationnaire)

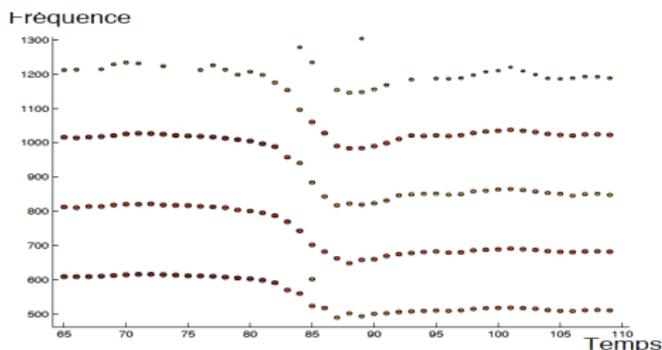
De la subtilité du hachoir...



Compromis usuellement considéré :

- Identification sur un horizon court
- Itération **avec recouvrement**
- Intégration temporelle par contrainte de régularité

De la subtilité du hachoir...

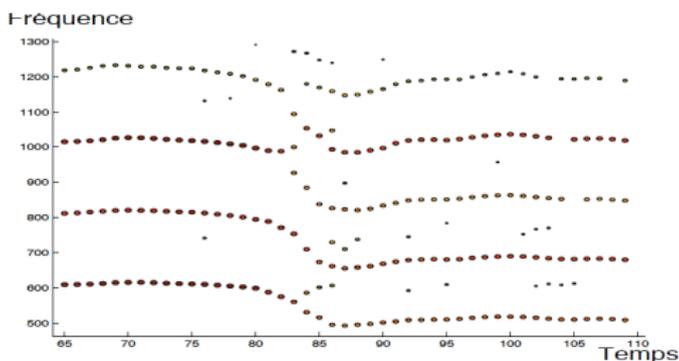


pas 12ms., fenêtre 24ms.

Compromis usuellement considéré :

- Identification sur un horizon court
- Itération **avec recouvrement**
- Intégration temporelle par contrainte de régularité

De la subtilité du hachoir...

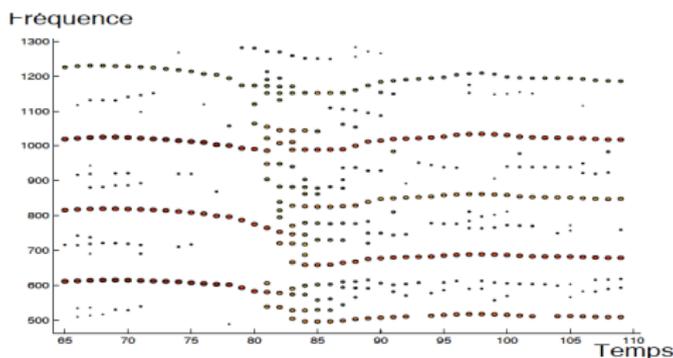


pas 12ms., fenêtre 48ms.

Compromis usuellement considéré :

- Identification sur un horizon court
- Itération **avec recouvrement**
- Intégration temporelle par contrainte de régularité

De la subtilité du hachoir...

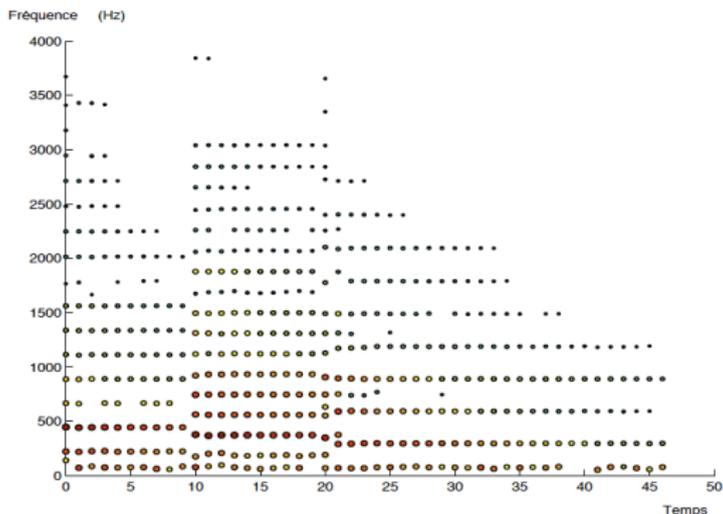


pas 12ms., fenêtre 96ms.

Compromis usuellement considéré :

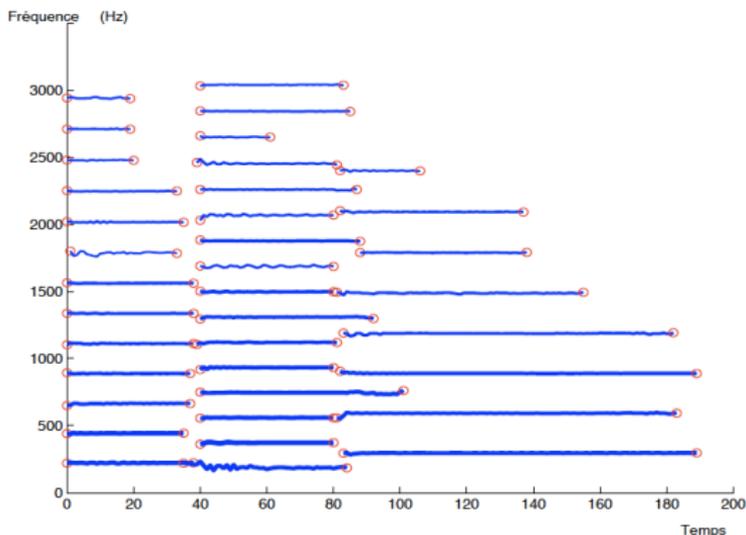
- Identification sur un horizon court
- Itération **avec recouvrement**
- Intégration temporelle par contrainte de régularité

Beaucoup on suivi avant nous...



- par heuristiques : MacAulay & al [Taslp'86], Serra [Phd'89]
- par HMM : Depalle & al [Icassp'93]
- par splines : Röbel [Tasslp'06]

Beaucoup on suivi avant nous...



- par heuristiques : MacAulay & al [Taslp'86], Serra [Phd'89]
- par HMM : Depalle & al [Icassp'93]
- par splines : Röbel [Tasslp'06]

De quelle manière les choses évolues t-elles ?

- GMM [Esling Ma'09]
- Polynomial + Sinusoïdes [Raspaud Phd'07]
- Modèle **Auto Régressif** [Lagrange Phd'04]

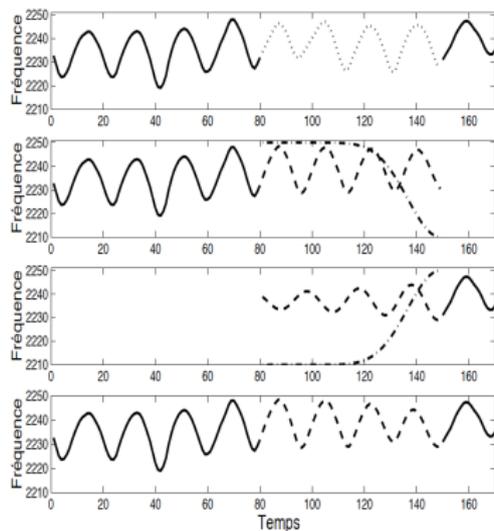
De quelle manière les choses évoluent-elles ?

- GMM [Esling Ma'09]
- Polynomial + Sinusoïdes [Raspaud Phd'07]
- Modèle **Auto Régressif** [Lagrange Phd'04]

$$x(n) = \sum_{i=1}^N a(i)x(n-i)$$

De quelle manière les choses évoluent-elles ?

- GMM [Esling Ma'09]
- Polynomial + Sinusoïdes [Raspaud Phd'07]
- Modèle **Auto Régressif** [Lagrange Phd'04]



De quelle manière les choses évolues t-elles ?

- GMM [Esling Ma'09]
- Polynomial + Sinusoïdes [Raspaud Phd'07]
- Modèle **Auto Régressif** [Lagrange Phd'04]

Original



Tronqué



AR temporel



Interpolation Linéaire



Interpolation AR



Dans le monde sinusoïdal : lentement !!

Une contrainte inhérente au modèle sinusoïdal est :

- l'évolution **lente** des paramètres
- utilisable pour le suivi de partiels (HFC) [Tasslp'07]

Dans le monde sinusoïdal : lentement !!

Une contrainte inhérente au modèle sinusoïdal est :

- l'évolution **lente** des paramètres
- utilisable pour le suivi de partiels (HFC) [Tasslp'07]

Dans le monde sinusoïdal : lentement !!

Une contrainte inhérente au modèle sinusoïdal est :

- l'évolution **lente** des paramètres
- utilisable pour le suivi de partiels (HFC) [Tasslp'07]

Dans le monde sinusoïdal : lentement !!

Une contrainte inhérente au modèle sinusoïdal est :

- l'évolution **lente** des paramètres
- utilisable pour le suivi de partiels (HFC) [Tasslp'07]

	Original	MAQ	HFC
Classique (Sc02)			
Pop (Sc03)			

Voir plus loin ...

- Au delà de la résolution de Fourier
- Ce qui est localement indécidable, ne l'est pas forcément en considérant des **à priori** structuraux :
 - Universels : un partiel n'est jamais seul
 - Contextuels

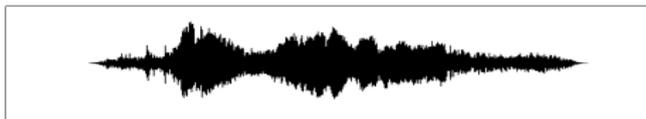
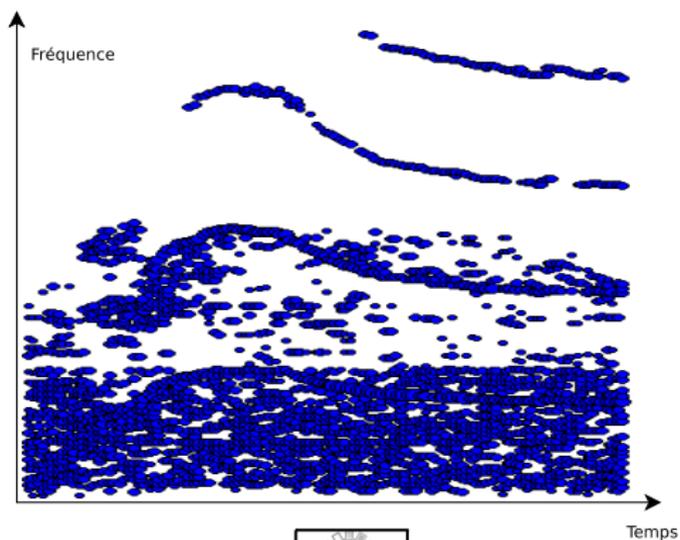
Lequel de l'œuf ou de la poule...

(U. of Vic.)

Modèle
sinusoïdal [S]



Signal



Lequel de l'œuf ou de la poule...

(U. of Vic.)

- Harmonicité et continuité temporelle doivent être considérées **conjointement**
- Approche par classification spectrale :
 - Représentation de la scène sonore sous forme de graphe
 - nœuds : atomes sinusoïdaux
 - arcs : pondérés par la probabilité des nœuds reliés d'appartenir à la même entité.
 - Formation d'entités sonores par coupures du graphe associé

Lequel de l'œuf ou de la poule...

(U. of Vic.)

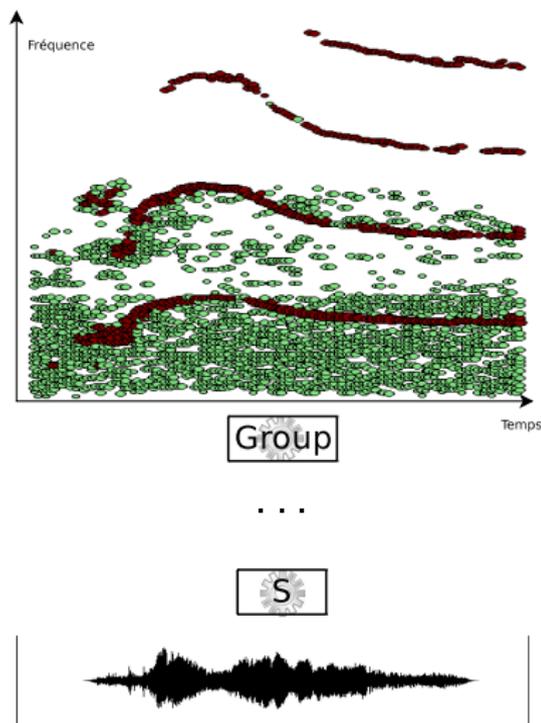
Structure
Instantanée



Modèle

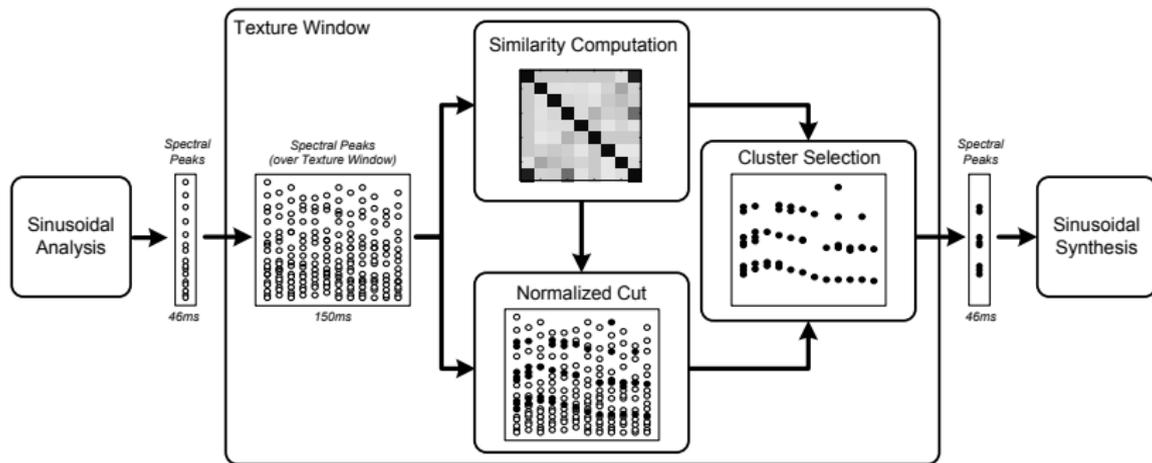


Signal



Lequel de l'œuf ou de la poule...

(U. of Vic.)



Un grand moment de solitude...

Différentes approches de séparation de la mélodie dominante :

- Modèle sinusoïdal + Classification Spectrale (SCS) [Lagrange & al Tasslp'08]
- Filtrage de Wiener + Factorisation de Matrices Non Négatives (WNMF) [Durrieu & al Icassp'08]

	Original	SCS	WNMF
Mirex song			
Pop song			
Celtic song			
U2 (Live)			
U2 (Live)			

Projet

L'Analyse de Scènes Auditives (ASA) comme cadre scientifique

- ASA : initié par des psychoacousticiens [Bregman '90]
- Aspect computationnel poursuivi par de nombreux laboratoires (Cambridge, Columbia, Tokyo, ...) [Wang & al '06]
- Domaine d'application principal : la parole
- Principe:
 - **décomposition** sous forme d'atomes
 - structuration **instantanée** : regroupement des atomes en entités (notes)
 - structuration **séquentielle** : regroupement des entités en séquences (voix)
- Traits courants de structuration instantanée :
 - Harmonicité
 - Continuité temporelle
 - ...

L'ASA appliquée aux signaux musicaux

Structure
Séquentielle



Structure
Instantanée



Modèles :

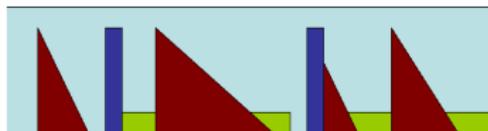
Sinusoïdal [S]

Transitoire [T]

Bruit [B]



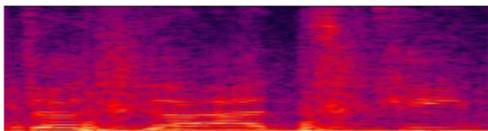
Signal



Group



Group



S

T

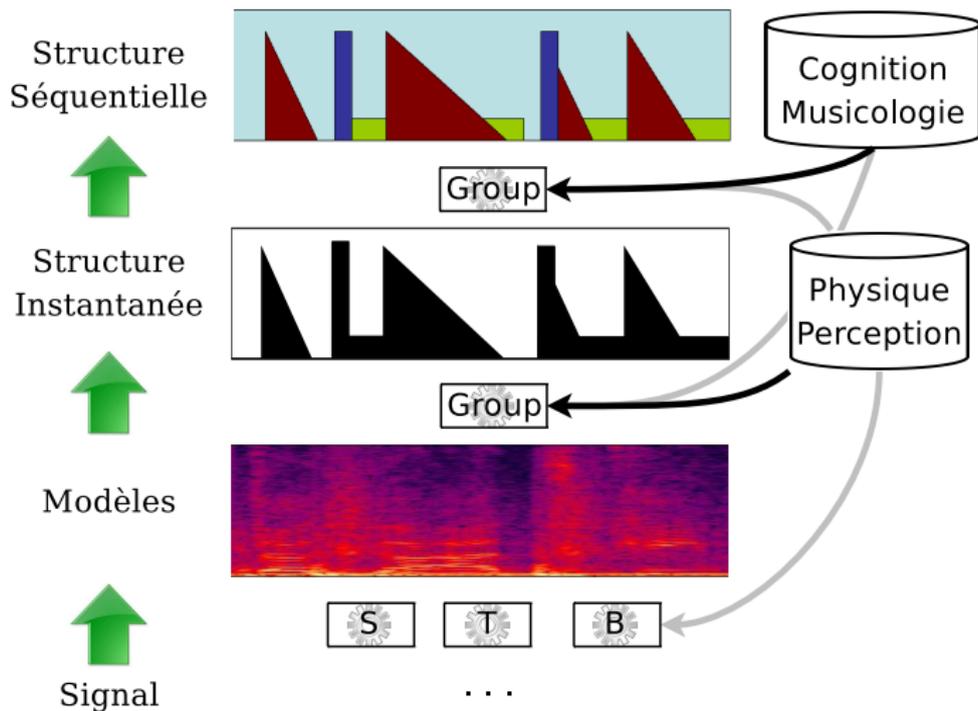
B

...

La structuration doit être multi-critères

- Critères de structuration universels :
 - Proximité fréquentielle, d'amplitude
 - Localité spatiale
- Critères de structuration spécifiques aux modèles :
 - Sinusoïdal : continuité temporelle, harmonicité, ...
 - Transitoires : régularité spectrale, synchronicité temporelle
 - Bruit : régularité spectro/temporelle
- Problématique :
 - Comment utiliser judicieusement ces critères ?
 - Prise en compte du **contexte** :
 - Type de signal
 - Type d'informations recherchées
 - ...

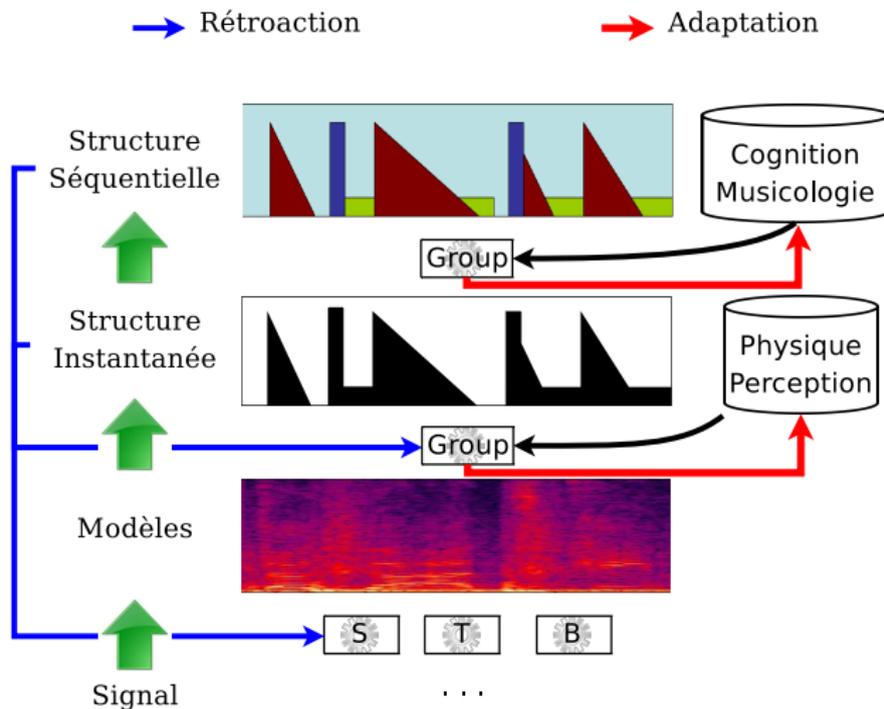
La structuration doit être simultanément “ascendante” et “descendante”



La structuration doit être simultanément “ascendante” et “descendante”

- Problématique complexe:
 - Formalisme commun entre les composants de traitements et les bases de connaissances
 - Pertinence des connaissances
- Premières approches :
 - Modèles de timbre
 - Modèles probabilistes
- Axe de recherche : approche hybride combinant **structuration** par coupures de graphes et **modélisation probabiliste** des connaissances

Rétroaction et adaptation



Rétroaction et adaptation : des outils fondamentaux pour une meilleure compréhension des signaux musicaux ?

- **Rétroaction :**

- Amélioration itérative de la qualité de description
- Challenge en terme d'architecture logicielle

- **Adaptation :**

- Problématique pluridisciplinaire
- Cognition : processus d'acculturation [Tillman '08]
 - apprentissage implicite par simple exposition
 - engendre connaissances et attentes musicales
- Valide une approche **phénoménologique** de la musique
- Problématique :
 - Quels sont les "patrons", les **invariants** structuraux ?
 - Comment **adapter** ces patrons aux phénomènes observés ?

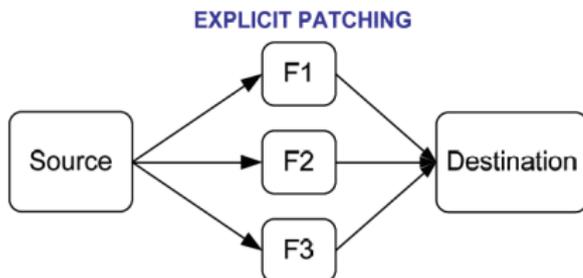
Travaux Annexes

“Wtf” ?

(U. of Vic.)

Architecture logicielle pour le traitement de signaux multimédias
[Acmm'08] :

- hétérogénéité des données
- flexibilité et expressivité :
 - des réseaux de traitements
 - des données traitées



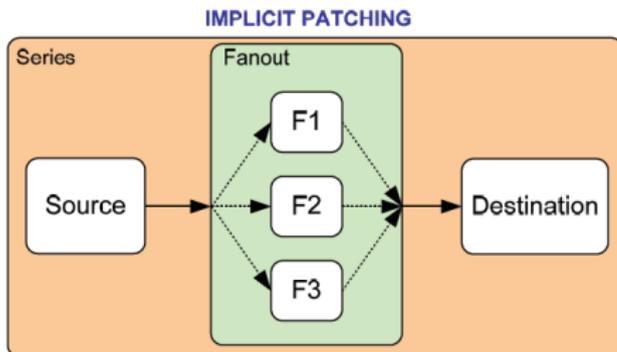
```
# EXPLICIT PATCHING
source, F1, F2, F3, destination
# connect the appropriate in/out ports
connect(source, F1);
connect(source, F2);
connect(source, F3);
connect(F1, destination);
connect(F2, destination);
connect(F3, destination);
```

“Wtf” ?

(U. of Vic.)

Architecture logicielle pour le traitement de signaux multimédias
[Acmm'08] :

- hétérogénéité des données
- flexibilité et expressivité :
 - des réseaux de traitements
 - des données traitées



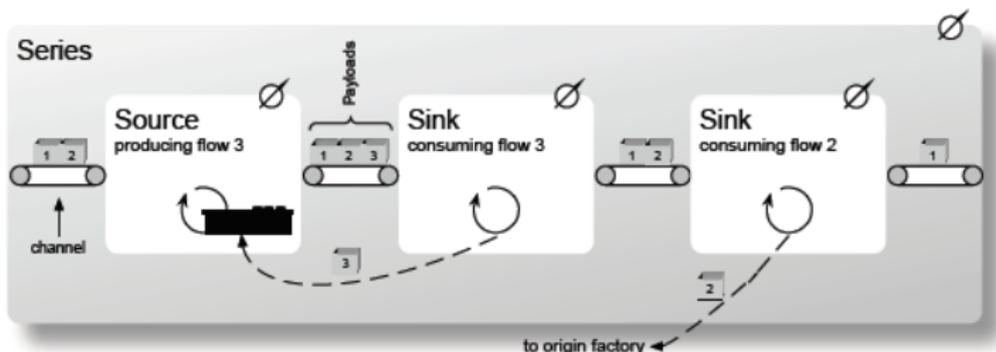
```
# IMPLICIT PATCHING
source, F1, F2, F3, destination
Fanout(F1, F2, F3)
Series(source, Fanout, destination);
```

“Wtf” ?

(U. of Vic.)

Architecture logicielle pour le traitement de signaux multimédias
[Acmm'08] :

- hétérogénéité des données
- flexibilité et expressivité :
 - des réseaux de traitements
 - des données traitées



Flots de données sous forme de "Payloads" (MarsyasX)

Similaire, vous avez dit similaire ? (Telecom ParisTech)

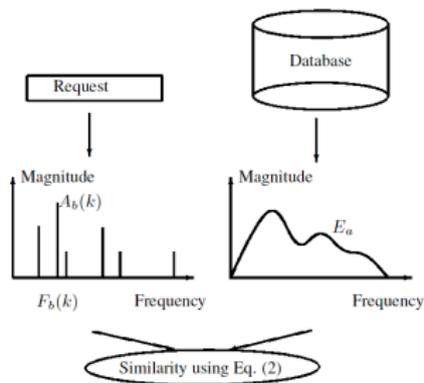
Décrire la musique implique de nombreuses dimensions :

- Rythme : descripteur compact de la métrique [Ismir'09]
- Timbre : modélisation robuste de l'enveloppe spectrale [sub Icassp'09]
- Hauteur : approche multimodale pour la détection de reprise [sub Icassp'09]

Similaire, vous avez dit similaire ? (Telecom ParisTech)

Décrire la musique implique de nombreuses dimensions :

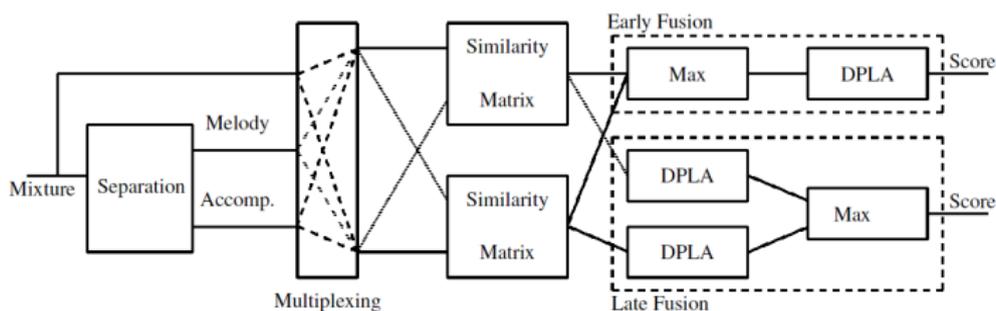
- Rythme : descripteur compact de la métrique [Ismir'09]
- Timbre : modélisation robuste de l'enveloppe spectrale [sub Icassp'09]
- Hauteur : approche multimodale pour la détection de reprise [sub Icassp'09]



Similaire, vous avez dit similaire ? (Telecom ParisTech)

Décrire la musique implique de nombreuses dimensions :

- Rythme : descripteur compact de la métrique [Ismir'09]
- Timbre : modélisation robuste de l'enveloppe spectrale [sub Icassp'09]
- Hauteur : approche multimodale pour la détection de reprise [sub Icassp'09]



Pierre qui roule...

(McGill)

La synthèse de sons de contact entretenus est un problème difficile, mais d'intérêt pour de nombreux domaines d'applications. Dans une optique temps réel, on considère :

- un modèle Source/Filtre à impulsions multiples
- estimation en boucle ouverte
- modélisation statistique des impulsions

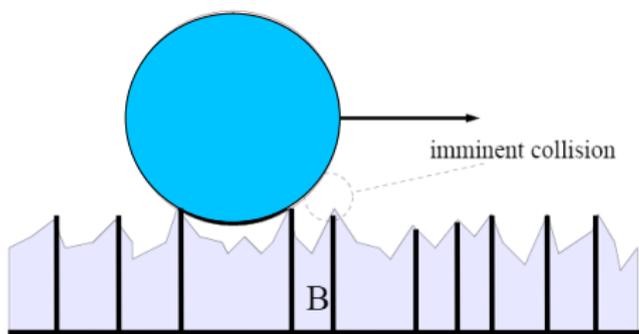


Pierre qui roule...

(McGill)

La synthèse de sons de contact entretenus est un problème difficile, mais d'intérêt pour de nombreux domaines d'applications. Dans une optique temps réel, on considère :

- un modèle Source/Filtre à impulsions multiples
- estimation en boucle ouverte
- modélisation statistique des impulsions

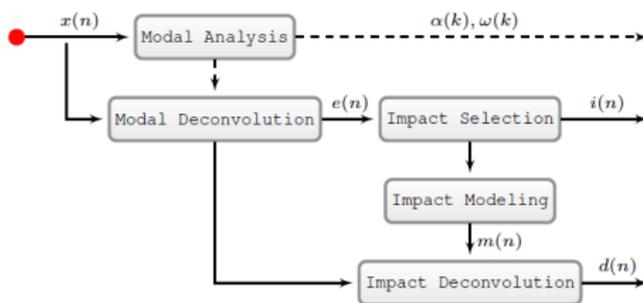


Pierre qui roule...

(McGill)

La synthèse de sons de contact entretenus est un problème difficile, mais d'intérêt pour de nombreux domaines d'applications. Dans une optique temps réel, on considère :

- un modèle Source/Filtre à impulsions multiples
- estimation en boucle ouverte
- modélisation statistique des impulsions

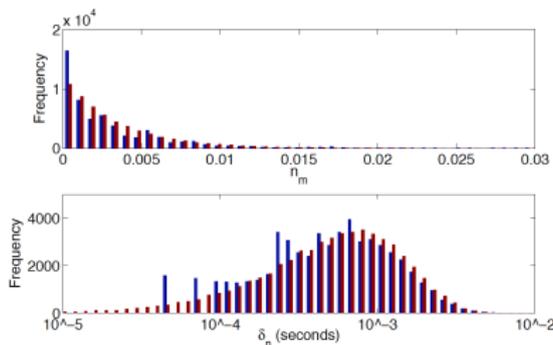


Pierre qui roule...

(McGill)

La synthèse de sons de contact entretenus est un problème difficile, mais d'intérêt pour de nombreux domaines d'applications. Dans une optique temps réel, on considère :

- un modèle Source/Filtre à impulsions multiples
- estimation en boucle ouverte
- modélisation statistique des impulsions



Souffle ici et je te dirais...

(LaBRI)

Étude sur le saxophone [Ismir'06] : l'évolution au cours du temps des paramètres spectraux portent de nombreuses informations sur l'instrumentiste :

- son expressivité
- son **niveau technique**

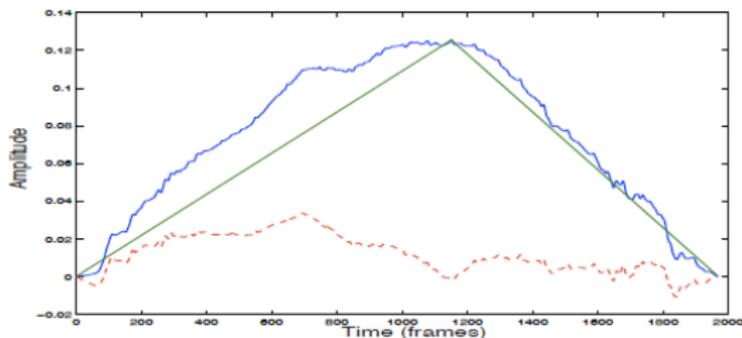


Souffle ici et je te dirais...

(LaBRI)

Étude sur le saxophone [Ismir'06] : l'évolution au cours du temps des paramètres spectraux portent de nombreuses informations sur l'instrumentiste :

- son expressivité
- son **niveau technique**

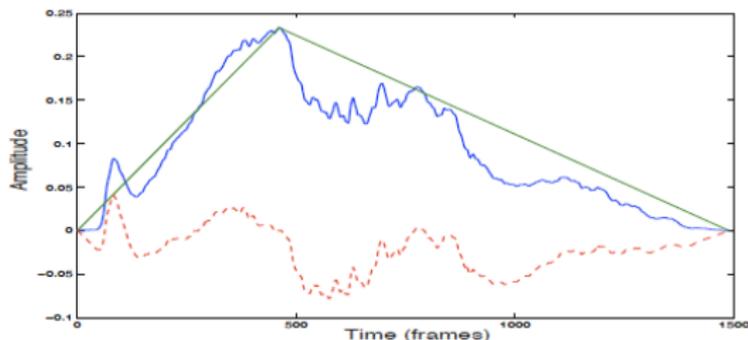


Souffle ici et je te dirais...

(LaBRI)

Étude sur le saxophone [Ismir'06] : l'évolution au cours du temps des paramètres spectraux portent de nombreuses informations sur l'instrumentiste :

- son expressivité
- son **niveau technique**



Souffle ici et je te dirais...

(LaBRI)

Étude sur le saxophone [Ismir'06] : l'évolution au cours du temps des paramètres spectraux portent de nombreuses informations sur l'instrumentiste :

- son expressivité
- son **niveau technique**

Amplitude results						
	α	p	mf	f	<>	vibrato
experts (3)	17	55 (8)	105 (40)	122 (67)	126 (24)	114 (150)
confirmed (7)	11	100 (22)	100 (33)	100 (10)	100 (28)	100 (98)
mid (6)	7	45 (12)	89 (22)	88 (36)	80 (21)	41 (42)
elementary (8)	4	47 (22)	65 (22)	76 (19)	44 (7)	100 (90)
beginners (6)	4	32 (16)	49 (22)	65 (39)	47 (17)	- (-)

Pour conclure...

A 204

48-71

lagrange@ircam.fr