Soutenance d'habilitation à diriger les recherches

« Modélisation Long-Terme de Signaux Sonores »

Mathieu Lagrange

Frédéric Bimbot	Directeur de Recherche CNRS, IRISA, Rennes
Alain de Cheveigné	Directeur de Recherche CNRS, ENS Paris
Béatrice Daille	Professeur, LS2N, Nantes
Patrick Flandrin	Directeur de Recherche CNRS, ENS Lyon
Stéphane Mallat	Professeur, Collège de France

14 Novembre 2019

Agenda

1. Curriculum Vitæ
2. Problématiques en traitement du signal audio-numérique
3. Analyse computationnelle de scènes auditives (CASA)
4. Expérimentation en traitement du signal audio-numérique
5. Perspectives de recherche

Curriculum Vitæ

Statut

Chargé de recherche CNRS classe normale

2009 : recrutement par la commission interdisciplinaire (CID) 44 : « Cognition, langage, traitement de l'information, systèmes »
2012 : rattachement à la section 07 : « Sciences de l'information »

Qualifications CNU

section 27 : « Informatique »
section 61 : « Génie informatique, automatique et traitement du signal »

Diplômes

Master Informatique (2001)

Intitulé : « Accélération de la synthèse sonore »
Encadrement : Sylvain Marchand, Robert Strandh
Lieu de soutenance : Université de Bordeaux 1

Doctorat Informatique (2004)

Intitulé : « Modélisation sinusoïdale des signaux polyphoniques »
Direction : Myriam Desainte-Catherine
Encadrement : Sylvain Marchand et Jean-Bernard Rault
Lieu de soutenance : Université de Bordeaux 1

Carrière

2001-4	**Doctorant Université Bordeaux 1** Ingénieur de Recherche à France Télécom R&D Rennes TECH/IRIS (équipe codage et multimédia)
2004-5	**Enseignant chercheur (ATER)** au LaBRI (U. Bx. 1)
2005-6	**Enseignant chercheur (ATER)** à l'Enseirb (U. Bx. 1)
2006-7	**Post-doctorant** au sein du département d'informatique Université de Victoria, BC, Canada
2007-8	**Post-doctorant** au sein du département "Music Technology" Université de McGill, QC, Canada
2008-9	**Post-doctorant** au sein de l'équipe « Acoustique Audio et Ondes », Télécom ParisTech
2009-13	**Chercheur CNRS** au sein de l'équipe Analyse / Synthèse Ircam (Umr 9912), Paris
2013- –	**Chercheur CNRS** au sein de l'équipe Signal, Images et Son (Sims) Ls2n (Umr 6004), Ecole Centrale de Nantes

Contributions

Indices bibliométriques

- 21 revues internationales à comité de lecture
- 62 conférences internationales à comité de lecture
- citations : 1784 (source Google Scholar, Oct. 2019)
- indice h : 19 (source Google Scholar, Oct. 2019)

Responsabilités

- relecteur pour 8 revues et 12 conférences du domaine
- adjoint à la direction de l'équipe SIMS
- membre du comité directeur de l'association sportive de l'École Centrale de Nantes

Enseignement (h eq. TD)

- Cursus Ingénieur
  - apprentissage automatique pour le traitement du signal audionumérique (24h)
  - musique numérique (15h)
- Master 2 : apprentissage automatique (9h)
- Formation doctorale : méthodologie de la recherche (18h)

Encadrement

- Rémi Foucard (2010 - 2013) : « Fusion multi-niveaux par boosting pour le tagging automatique »
- Grégoire Lafay (2013 - 2016) : « Simulation de scènes sonores environnementales : application à l'analyse sensorielle et à l'analyse automatique »
- Jean-Rémy Gloaguen (2015 - 2018) : « Estimation du niveau sonore de sources d'intérêt au sein de mélanges sonores urbains : application au trafic routier »
- Félix Gontier (2017 - –) : « Modélisation de signaux sonores par approches neuronales profondes »
- Tom Souaille (2019 - –) : « Conception interactive en design sonore »

Problématiques en traitement du signal audio-numérique

Traitement du signal audio-numérique

Besoins et problématiques associées

Transmission : Codage
Indexation : Recherche d'Information (IR)
Création : Synthèse sonore

Domaines d'application

- Musique
- Sons environnementaux

Besoins de compacité

Verrou

- une seconde de son :

$$x \in \mathbb{R}^{44100}$$

- besoin d'une représentation plus compacte

Types de compacité

Codage : compacité signal

Recherche d'information : compacité sémantique

Synthèse : compacité « signalo-sémantique »

Codage compressif par transformée

$$y = C(x) | \tilde{x} = C^{-1}(y), P_e(x) \simeq P_e(\tilde{x})$$

- C : Quantification adaptative d'un équivalent de la Transformée de Fourier à Court Terme (TFCT)
- P_e : Modélisation de la sensibilité aux déformations de la membrane basilaire

E. Zwicker et H. Fastl (2013). *Psychoacoustics : Facts and models.*

Transformée de Fourier à Court Terme (TFCT)

$$f[m,t] = \sum_{n=-\infty}^{\infty} x[n]w[n-t]e^{\frac{-2j\pi mn}{N}}$$

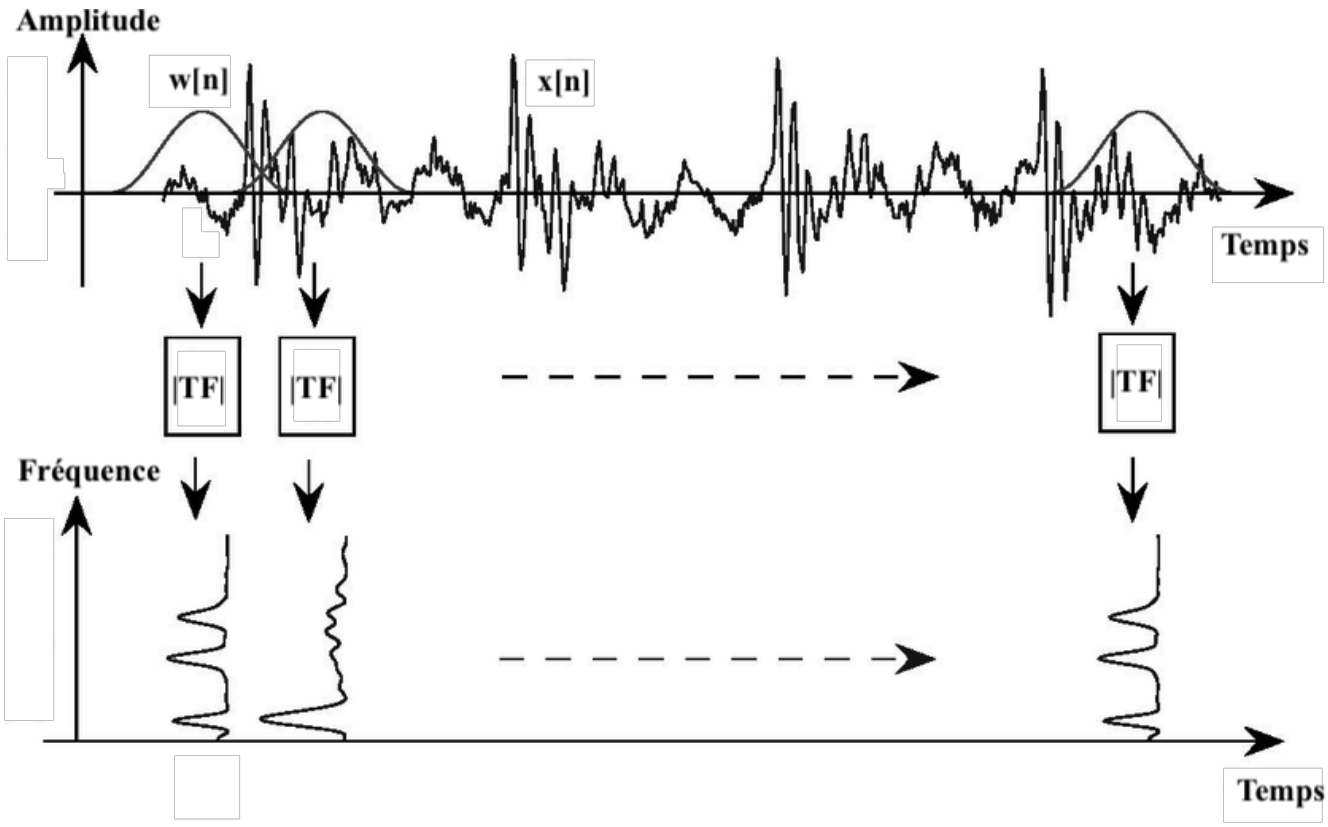
x : signal temporel

f : composantes fréquentielles (paniers, bins, ...)

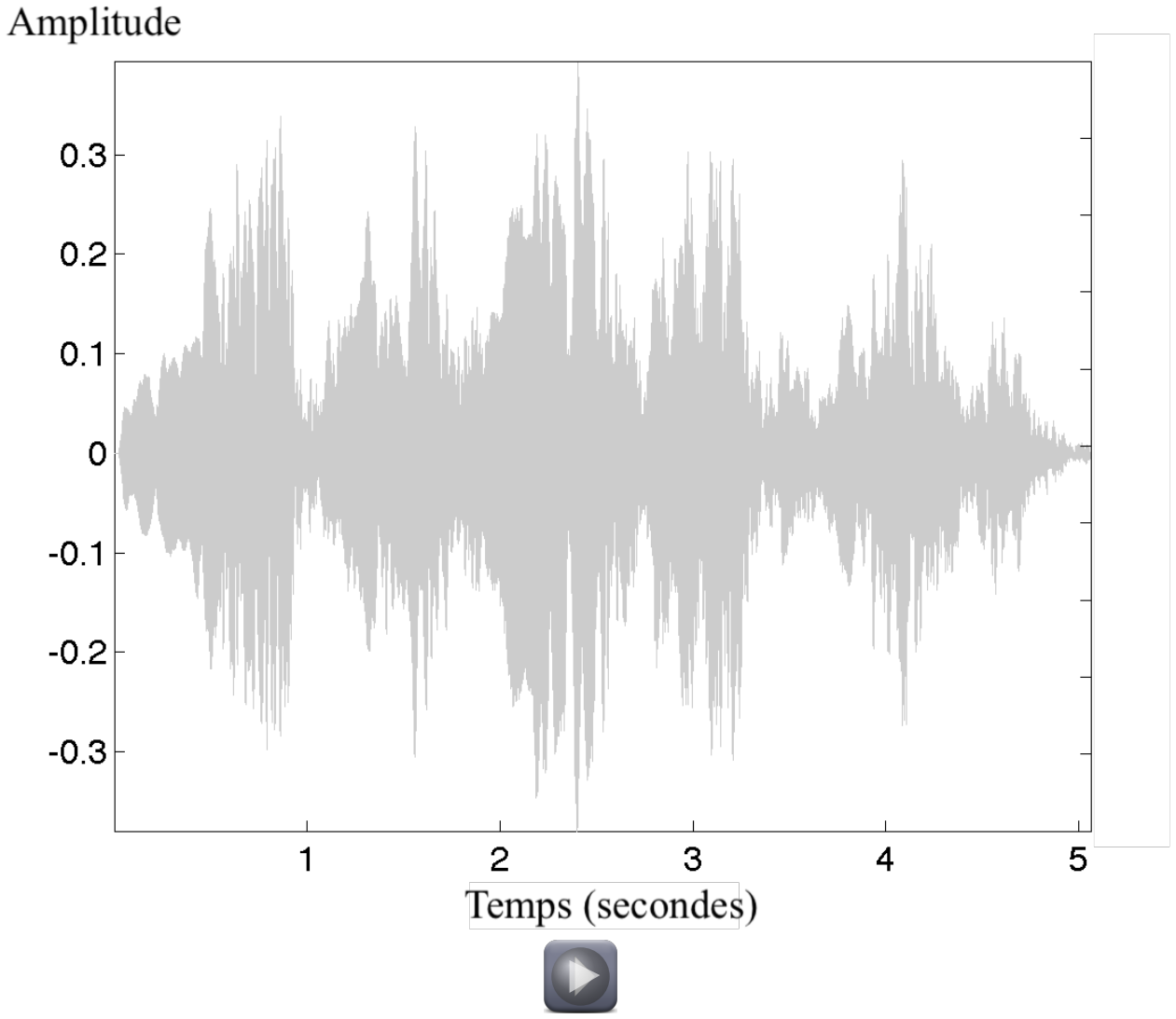
w : fenêtre

V. Lostanlen, J. Andén et M. Lagrange (2019). « Fourier at the heart of computer music : From harmonic sounds to texture ». In : *Comptes Rendus de Physique de l'Académie des Sciences*.

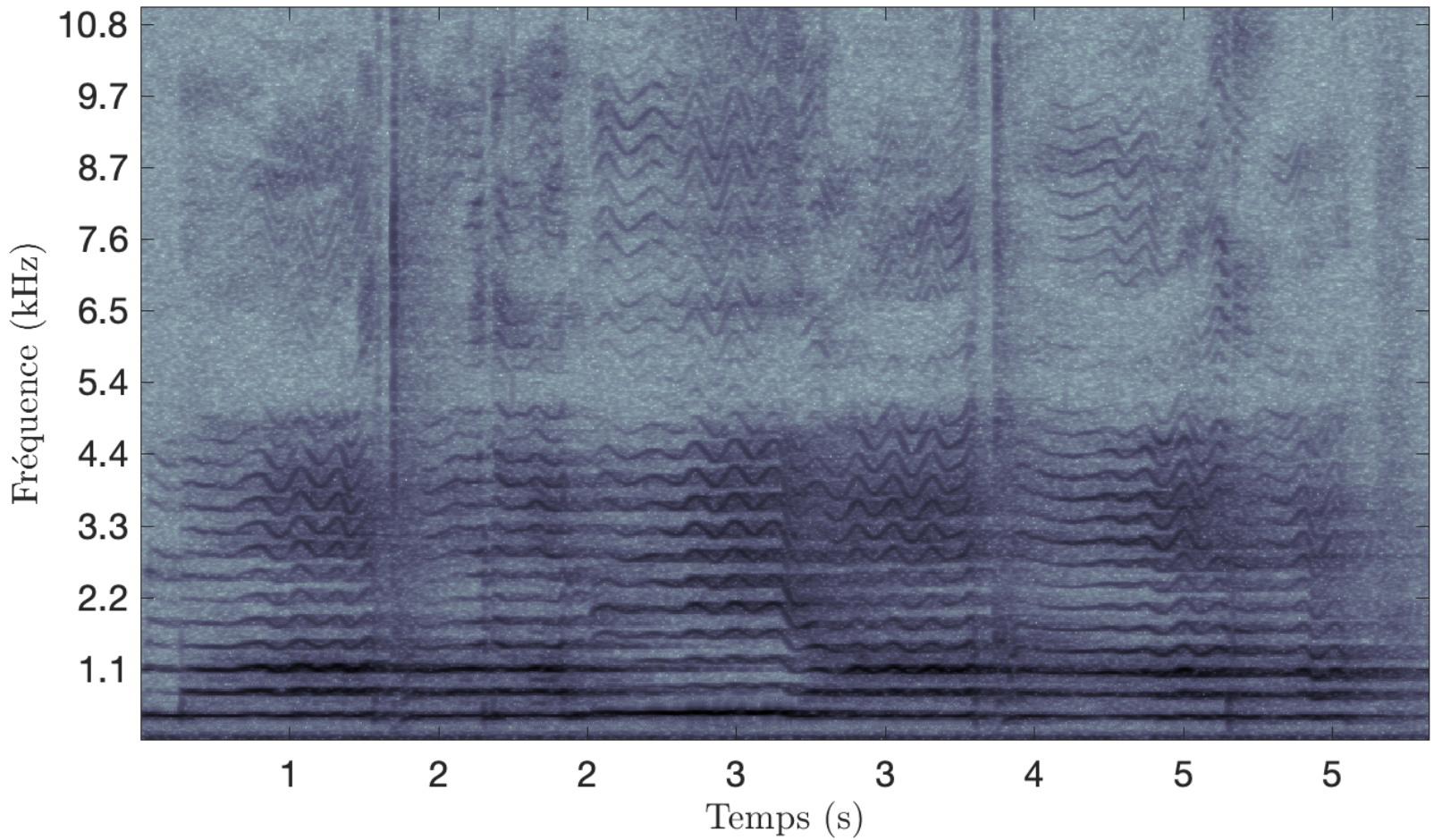
Spectrogramme : $|f[m,t]|$



Spectrogramme



Spectrogramme



Typologie des évènements sonores

	structure	
sons	horizontale	verticale
de parole	sons voisés	sons plosifs
	<a>, <o>	<pe>, <qe>
d'animaux	chants	clics
musicaux	chant lyrique	percussions
mécaniques	ventilation	marteau piqueur
environnementaux	vent	gouttes de pluie

Compromis temps/fréquence

$\hookrightarrow$ mitiger cette contrainte imposée par l'approche court-terme par l'utilisation d'*a priori* sur les sources d'intérêt.

Analyse computationnelle de scènes auditives (CASA)

Analyse de Scènes Auditives Computationnelle (CASA)

L'Analyse de Scène Auditives (ASA) étudie l'ensemble de traitements perceptifs permettant

- d'isoler les informations émanant d'entités sonores distinctes,
- de les organiser en un tout cohérent.
- à l'aide de processus « primitifs » et « séquentiels ».

L'approche CASA met en œuvre ces critères pour inférer automatiquement une organisation perceptuellement valide de la scène sonore.

A. S. Bregman (1994). *Auditory scene analysis : The perceptual organization of sound.*

D. Wang et G. J. Brown (2006). *Computational Auditory Scene Analysis : Principles, Algorithms, and Applications.*

Processus ASA « primitifs »

continuité : les propriétés d'un son isolé tendent à se modifier lentement et de façon continue

harmonicité : lorsqu'un corps sonore vibre à une période répétée, ses vibrations donnent naissance à un motif acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale ;

...

Processus ASA « primitifs »

continuité : les propriétés d'un son isolé tendent à se modifier lentement et de façon continue

harmonicité : lorsqu'un corps sonore vibre à une période répétée, ses vibrations donnent naissance à un motif acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale ;

...

Modèle sinusoïdal à long terme

$$x[n] = \sum_{l=1}^{L} a_l[n] \sin\left(\frac{2\pi}{F_s} f_l[n] \cdot n + \Phi_k\right)$$

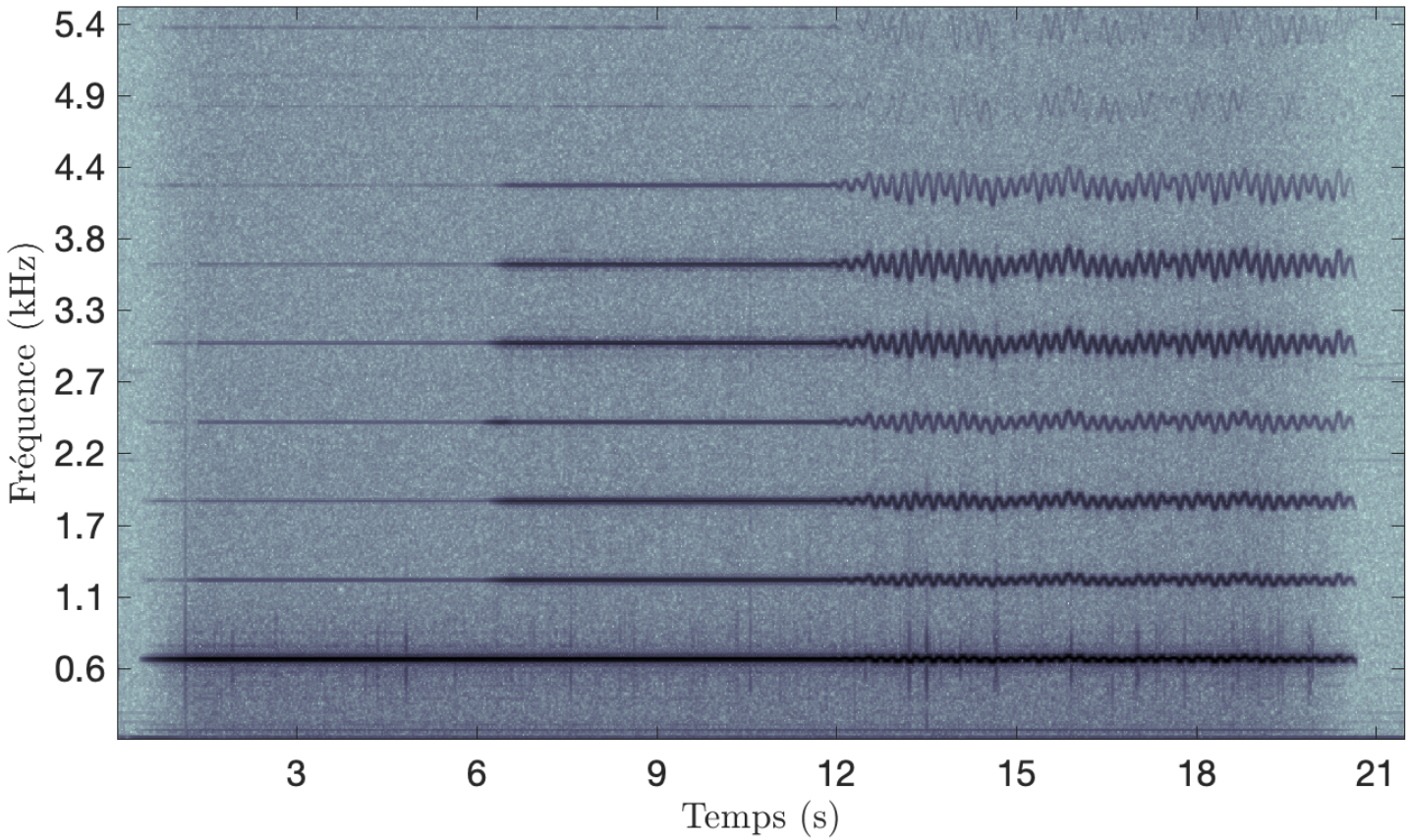
$a_l[n]$ et $f_l[n]$ sont des signaux basse fréquence contrôlant respectivement l'amplitude et la phase des L oscillateurs composant le modèle.

Modèle sinusoïdal à long terme

$$x[n] = \sum_{l=1}^{L} a_l[n] \sin\left(\frac{2\pi}{F_s} f_l[n] \cdot n + \Phi_k\right)$$

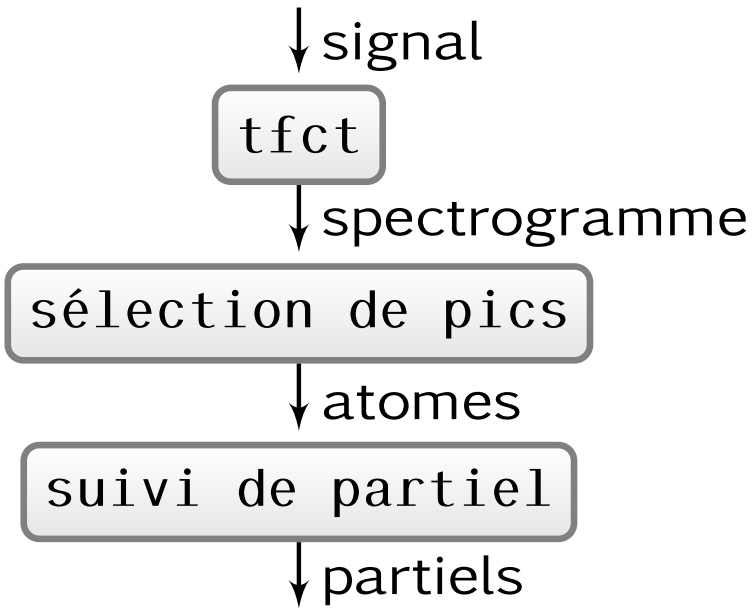
- $a_l[n] = 1$
- $f_l[n] = l * \left(f_0 + a_v \sin\left(\frac{2\pi * 5}{F_s} \cdot n\right) + FPB(\mathcal{N})\right)$
- f_0 : fréquence de fondamentale
- FBP : filtre passe bas

Modèle sinusoïdal à long terme

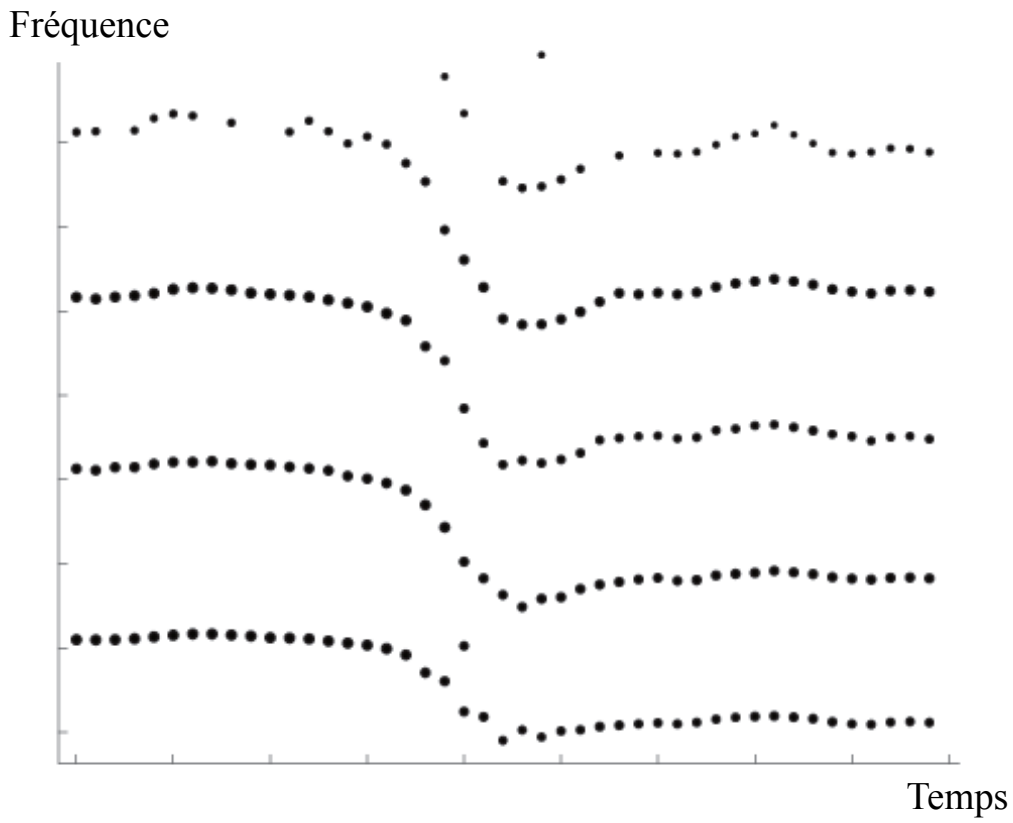


John Chowning (1970-80)

Procédé d'analyse de signaux sonores

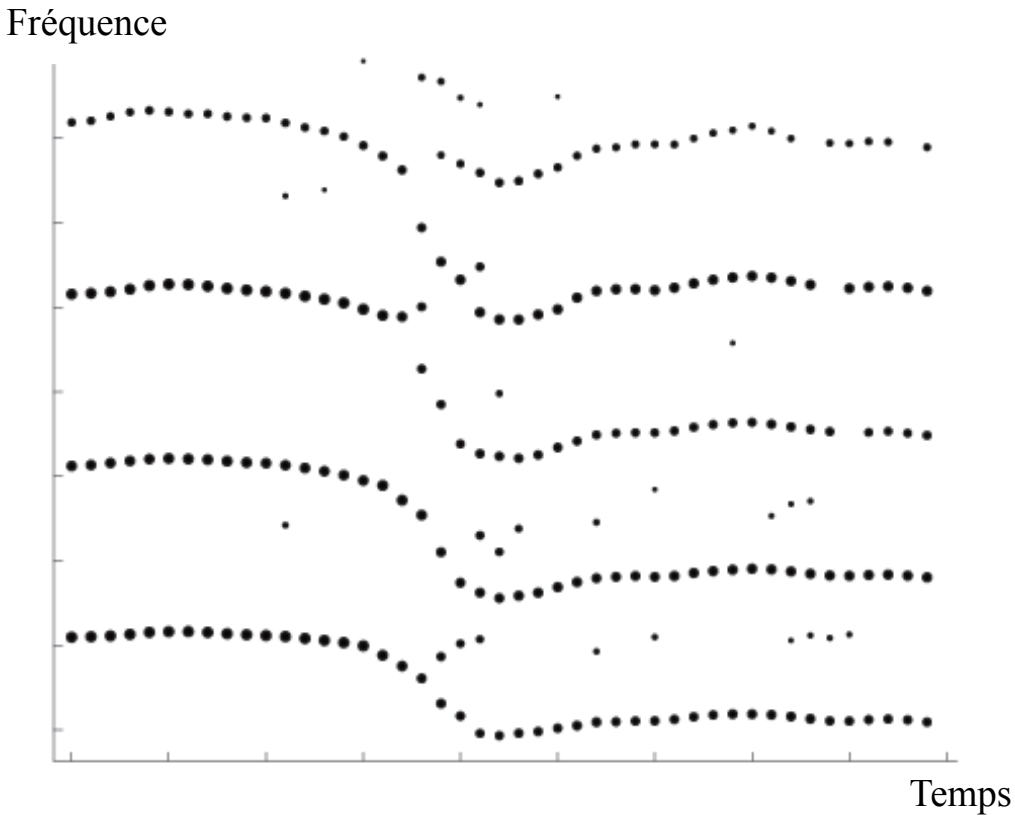


Atomes temps/fréquences



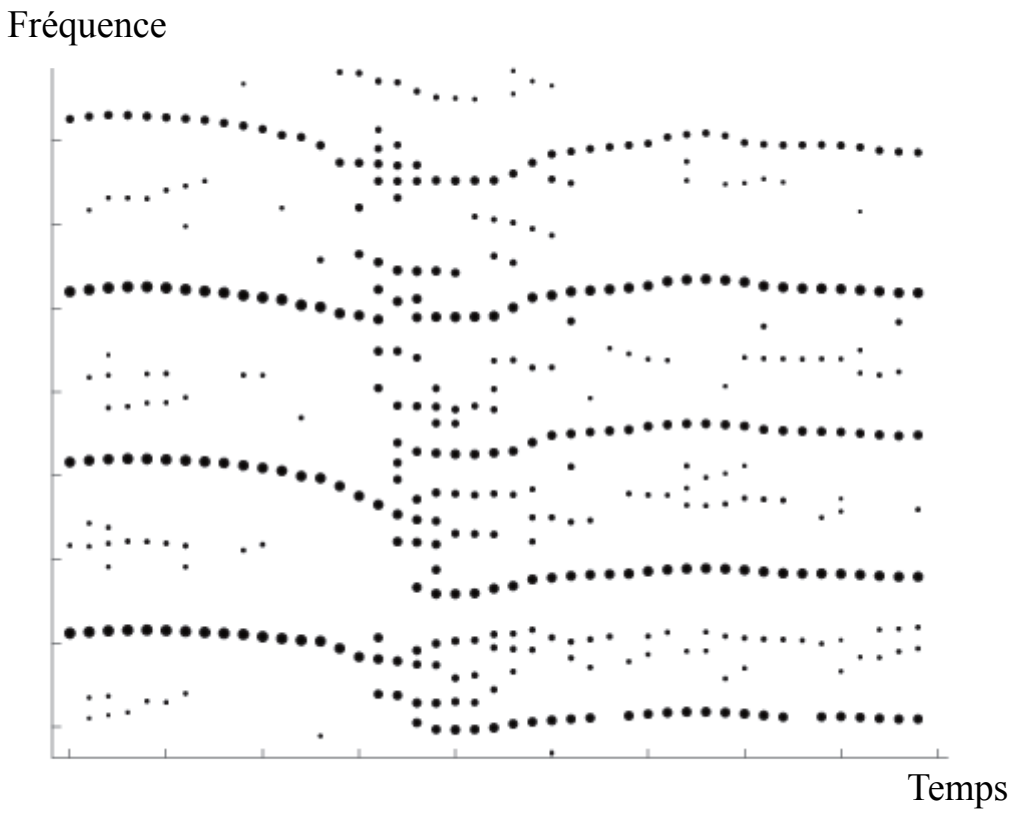
Taille de fenêtre : 23 ms, pas d'avancement 10 ms.

Atomes temps/fréquences



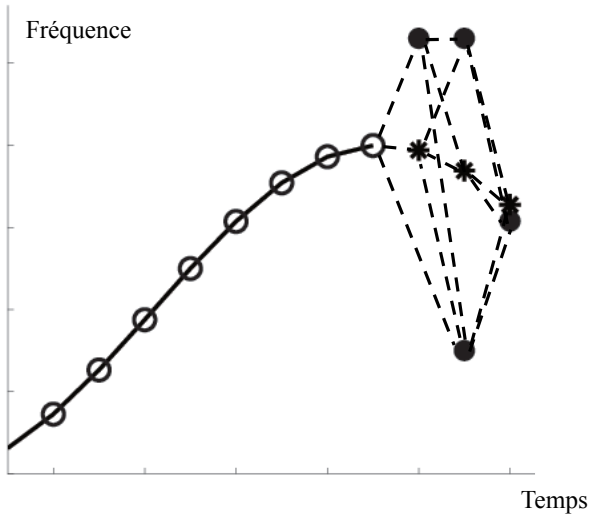
Taille de fenêtre : 46 ms, pas d'avancement 10 ms.

Atomes temps/fréquences



Taille de fenêtre : 92 ms, pas d'avancement 10 ms.

Algorithme de suivi amélioré



1. Prédiction auto-régressive de l'évolution des paramètres de fréquence et d'amplitude
2. Sélection des continuations engendrant le moins de hautes fréquences

---

M. Lagrange, S. Marchand et J. Rault (2007). « Enhancing the Tracking of Partials for the Sinusoidal Modeling of Polyphonic Sounds ». In : *IEEE Transactions on Acoustics, Speech, Signal and Language Processing*.

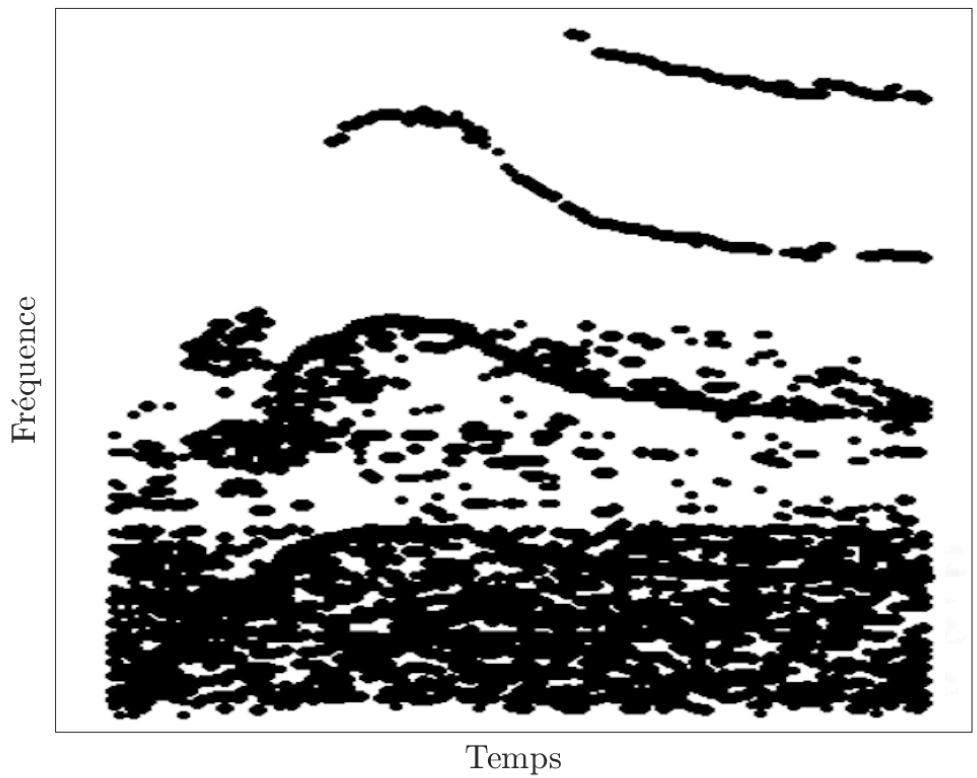
Processus ASA « primitifs »

continuité : les propriétés d'un son isolé tendent à se modifier lentement et de façon continue

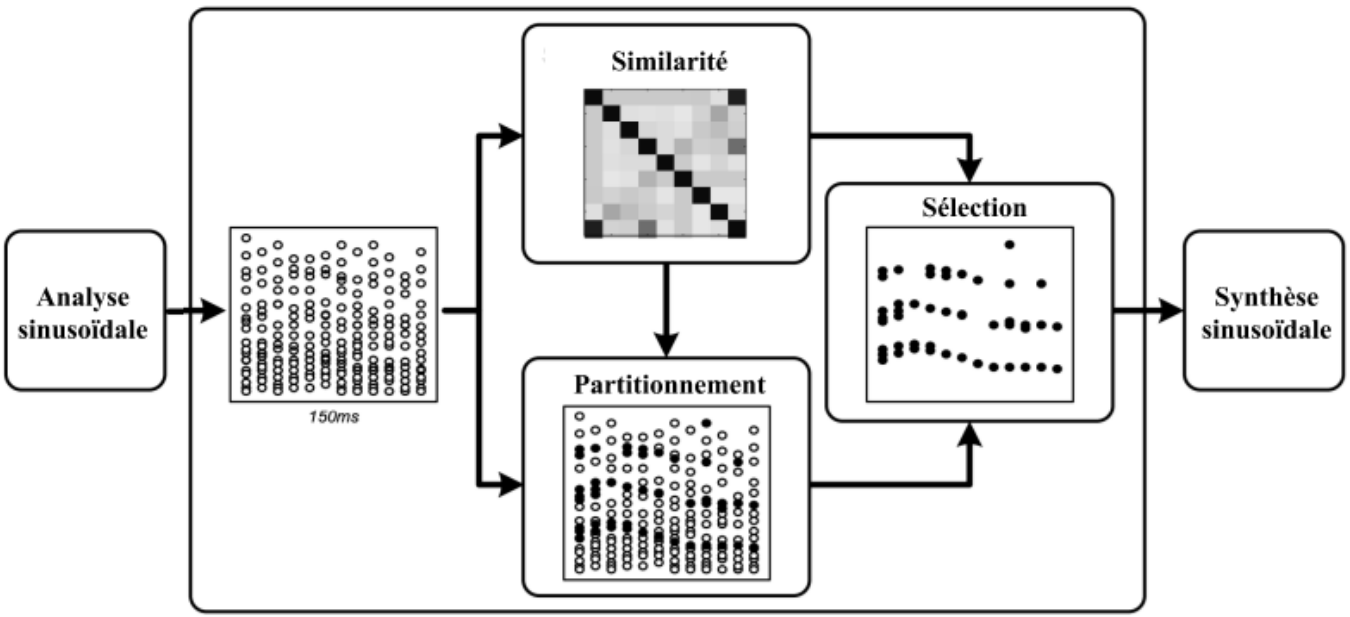
harmonicité : lorsqu'un corps sonore vibre à une période répétée, ses vibrations donnent naissance à un motif acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale ;

...

Algorithme par coupures normalisées de graphes

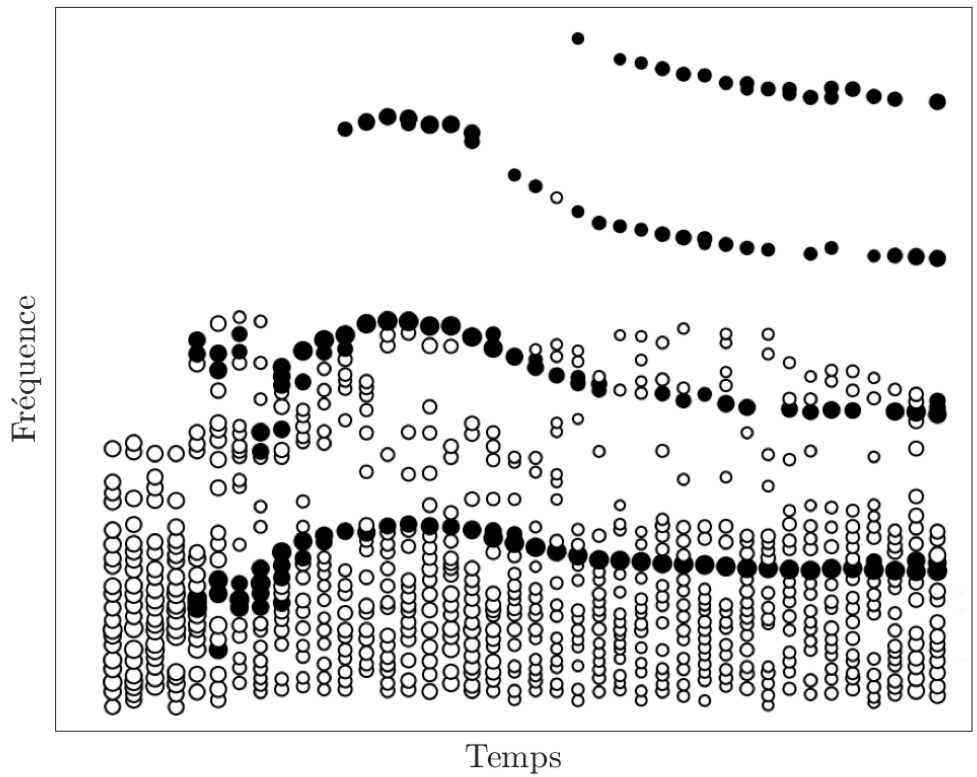


Algorithme par coupures normalisées de graphes



J. Shi et J. Malik (2000). « Normalized cuts and image segmentation ». In : *IEEE Transactions on pattern analysis and machine intelligence*.

Algorithme par coupures normalisées de graphes



M. Lagrange, L. G. Martins et al. (2008). « Normalized Cuts for Predominant Melodic Source Separation ». In : *IEEE Transactions on Acoustics, Speech, Signal and Language Processing*.

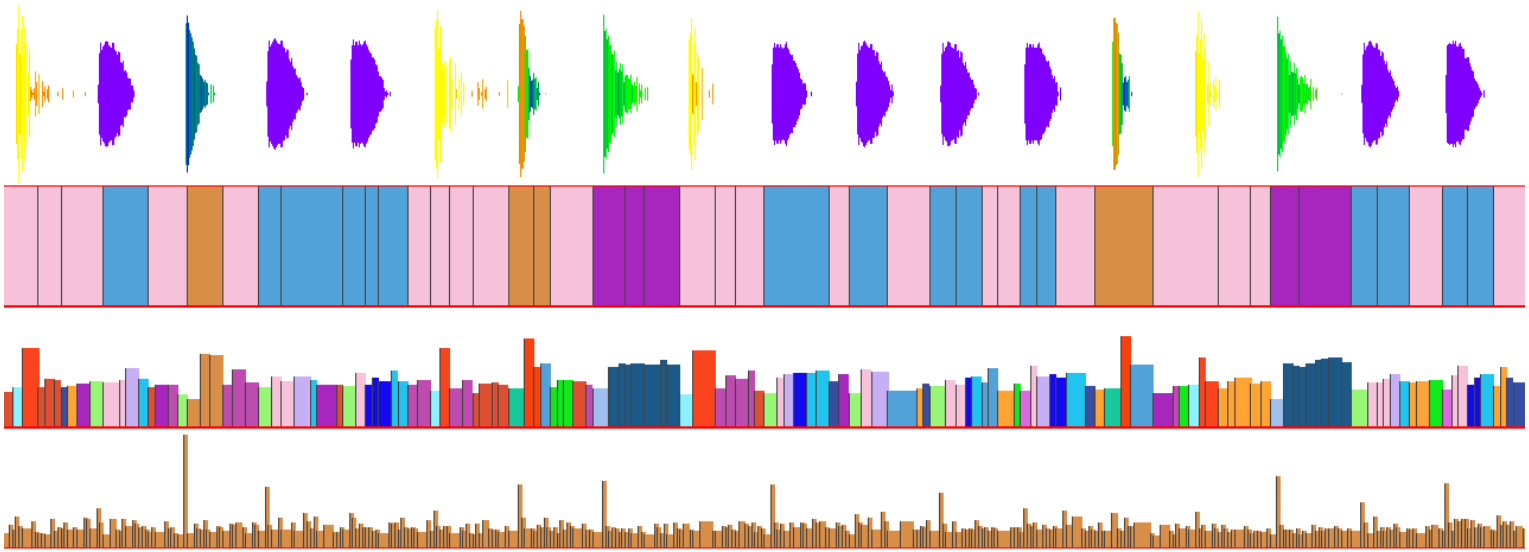
Processus ASA « séquentiels »

proximité : des éléments proches les uns des autres sur le plan temps/fréquence ont tendance à être groupés ensemble

similarité : des éléments qui se ressemblent ont tendance à être groupés ensemble (timbre).

...

Regroupement hiérarchique alterné (ANR JCJC Houle)



M. Rossignol, M. Lagrange et A. Cont (2018). « Efficient similarity-based data clustering by optimal object to cluster reallocation ». In : *PloS one*.

M. Rossignol, M. Lagrange, G. Lafay et al. (2015). « Alternate Level Clustering for Drum Transcription ». In : *European Conference on Signal Processing (EUSIPCO)*.

Approches « algorithmiques »

- Expression d'*a priori* sous forme d'heuristiques computationnelles
- Algorithmes de structuration non supervisés

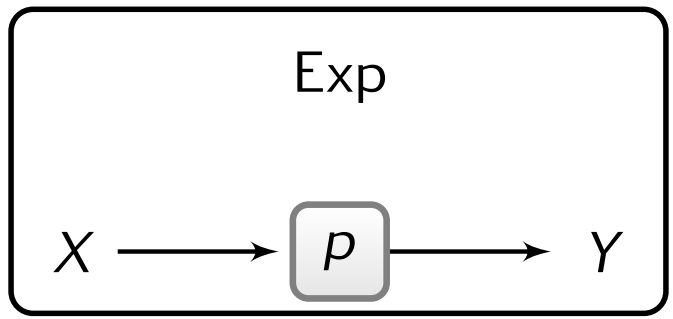
Bilan

+ approches long terme : exploitation d'intervalles d'observation long
- représentation court terme : problèmes de résolution
+ pas d'apprentissage : protocole expérimental simple, interprétabilité
- pas d'apprentissage : problèmes de tractabilité, problèmes d'efficience

Expérimentation en traitement du signal audio-numérique

Compétition

Expérimentation



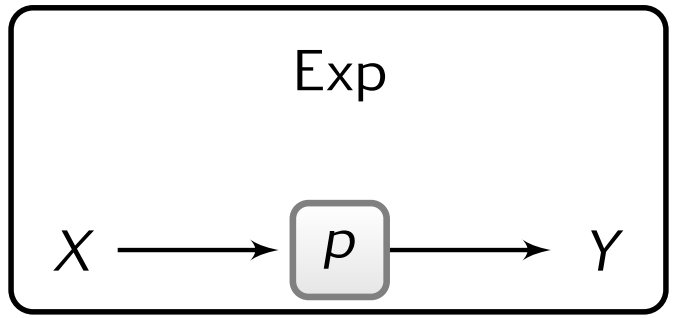
p : processus, traitement, prédicteur

X : entrée, signal, observation

Y : sortie, prédiction

Exp : protocole expérimental

Contributions



X : plus de contrôle

Y : plus de maîtrise

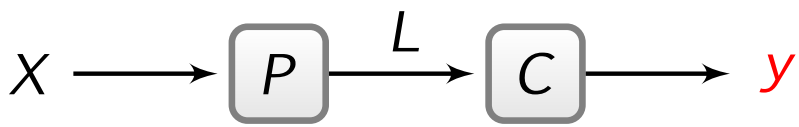
Exp : plus de formalisation

Tâche : recherche d'information (IR)

X : données observées (en grande dimension)

y : labels (en petite dimension)

Tâche : recherche d'information (IR)



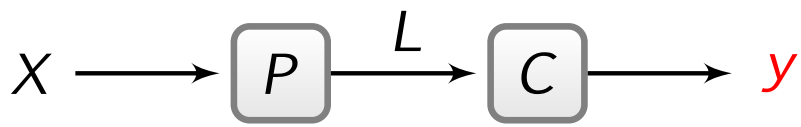
L données projetées dans un espace latent

but : $X_j = D(X_i), L_j = d(L_i)$ ssi $y_j = y_i$

D : grande déformation

d : petite déformation

Propriétés de L



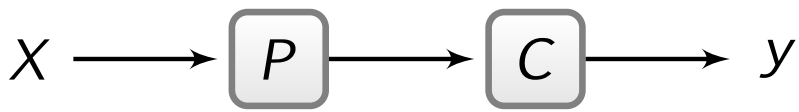
Guidé par les données X

- invariance : translation, ...
- stabilité : étirement, ...

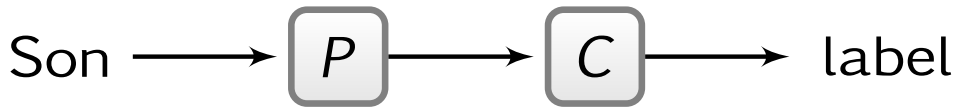
Guidé par la tâche y

$$|L_i - L_j| < |L_i - L_k| \text{ ssi } y_i = y_j \forall k | y_k \neq y_i$$

IR en audio



IR en audio



- P : module de « perception »
  - plusieurs TFCT à résolutions différentes
  - réseaux convolutionnels profonds
- C : module de « cognition »
  - réseaux neuronaux profonds totalement connectés
  - fusion

Communautés

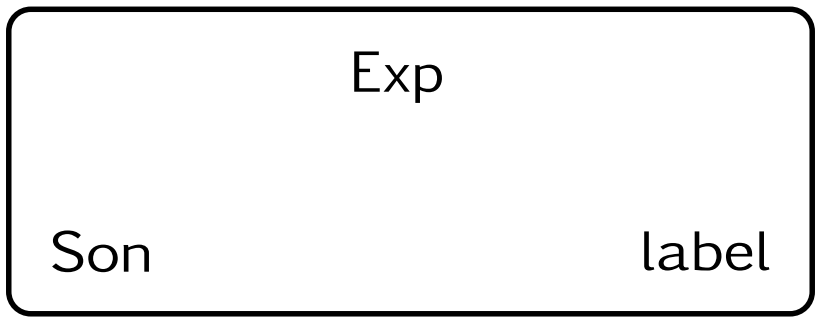
Music Information Retrieval (MIR)

- 2000 - –
- Compétition : 18 tâches (Mirex)
- Conference : 100 articles (Ismir)

Detection and Classification of Acoustic Scenes and Events (DCASE)

- 2013 - –
- Compétition : 7 tâches
- Workshop : 50 articles

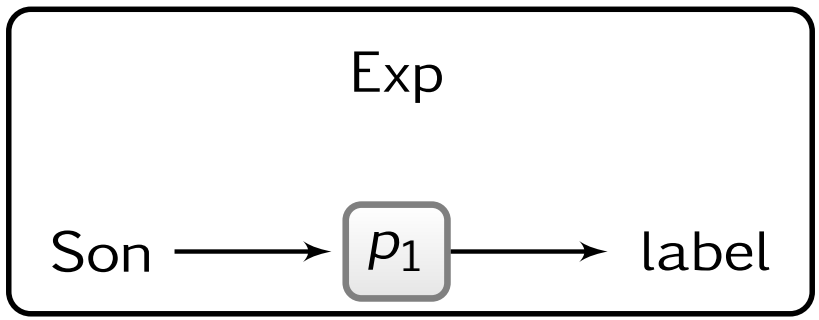
Compétitions en IR



L'organisateur de la compétition

- prépare les données ainsi qu'une description du problème
- fournit des méthodes d'évaluation
- ...

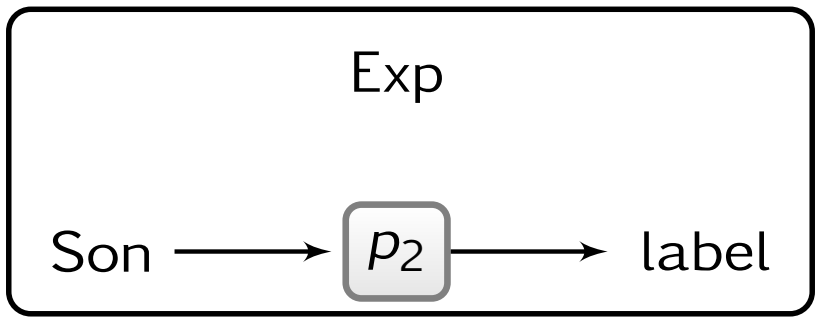
Compétitions en IR



L'organisateur de la compétition

- prépare les données ainsi qu'une description du problème
- fournit des méthodes d'évaluation
- ...

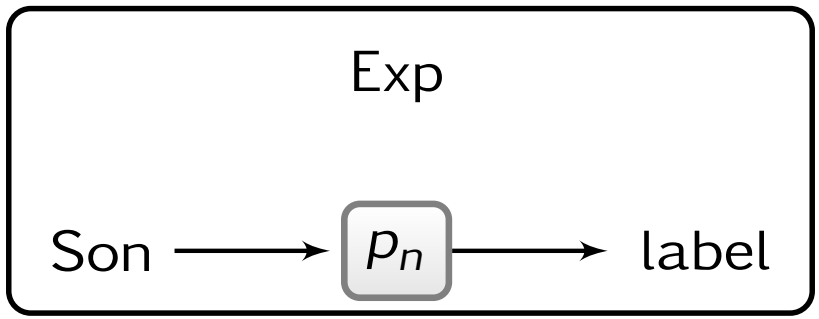
Compétitions en IR



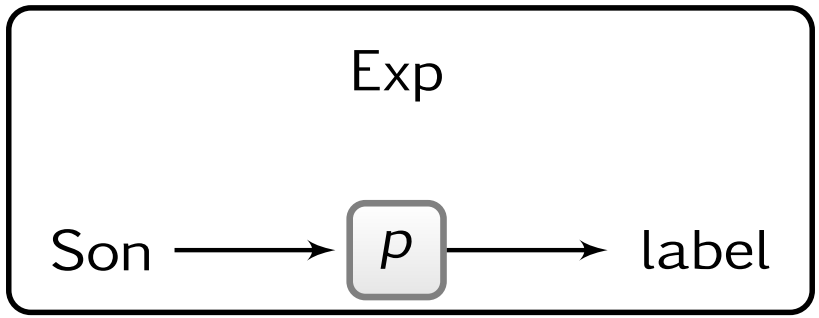
L'organisateur de la compétition

- prépare les données ainsi qu'une description du problème
- fournit des méthodes d'évaluation
- ...

Compétitions en IR

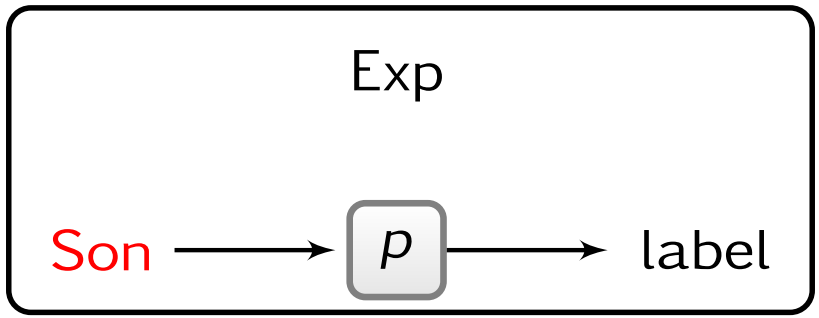


Compétitions en IR



- Alternative à l'approche « mon outil, mon jeu de données, ma métrique »
- Spécification de l'expérimentation « clés en main »
- biais de conception du protocole assumés collectivement

Compétitions en IR



- Alternative à l'approche « mon outil, mon jeu de données, ma métrique »
- Spécification de l'expérimentation « clés en main »
- biais de conception du protocole assumés collectivement

Compétitions en crise ?

- cauchemard des métriques
- « la fin justifie les moyens »

« Faites quelque chose d'intéressant !! »
« ... de scientifique !! »

panel DCASE 2018

↪ placer l'effort sur un questionnement plutôt que sur la démonstration d'un outil

Design de Compétition

Organisation d'une tâche DCASE (2013, 2016)

- Tâche de détection d'évènements
- corpus de scènes sonores simulées
- sons isolés enregistrés
- contrôle de haut niveau sur la composition de la scène

D. Stowell et al. (2015). « Detection and Classification of Acoustic Scenes and Events ». In : *IEEE Transactions on Multimedia*.

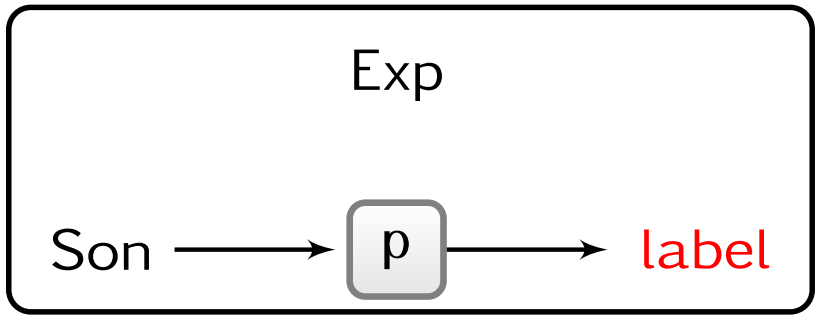
A. Mesaros et al. (2018). « Detection and Classification of Acoustic Scenes and Events : Outcome of the DCASE 2016 Challenge ». In : *IEEE/ACM Transactions on Audio, Speech and Language Processing*.

Approche « psychologie expérimentale »

- formulation d'une hypothèse : le degré de polyphonie impacte les algorithmes de détection d'évènement sonores
- production de corpus avec un degré variable de polyphonie
- choix d'un protocole expérimental adapté
- les algorithmes sont considérés comme des sujets et leurs concepteurs ne sont pas informés de la typologie du corpus
- analyse des résultats

G. Lafay et al. (2016). « A morphological model for simulating acoustic scenes and its application to sound event detection ». In : *IEEE/ACM Transactions on Audio, Speech and Language Processing*.

Que prédire ?



- interface riche avec d'autres communautés
- nécessité d'alignement des vocabulaires et des temporalités

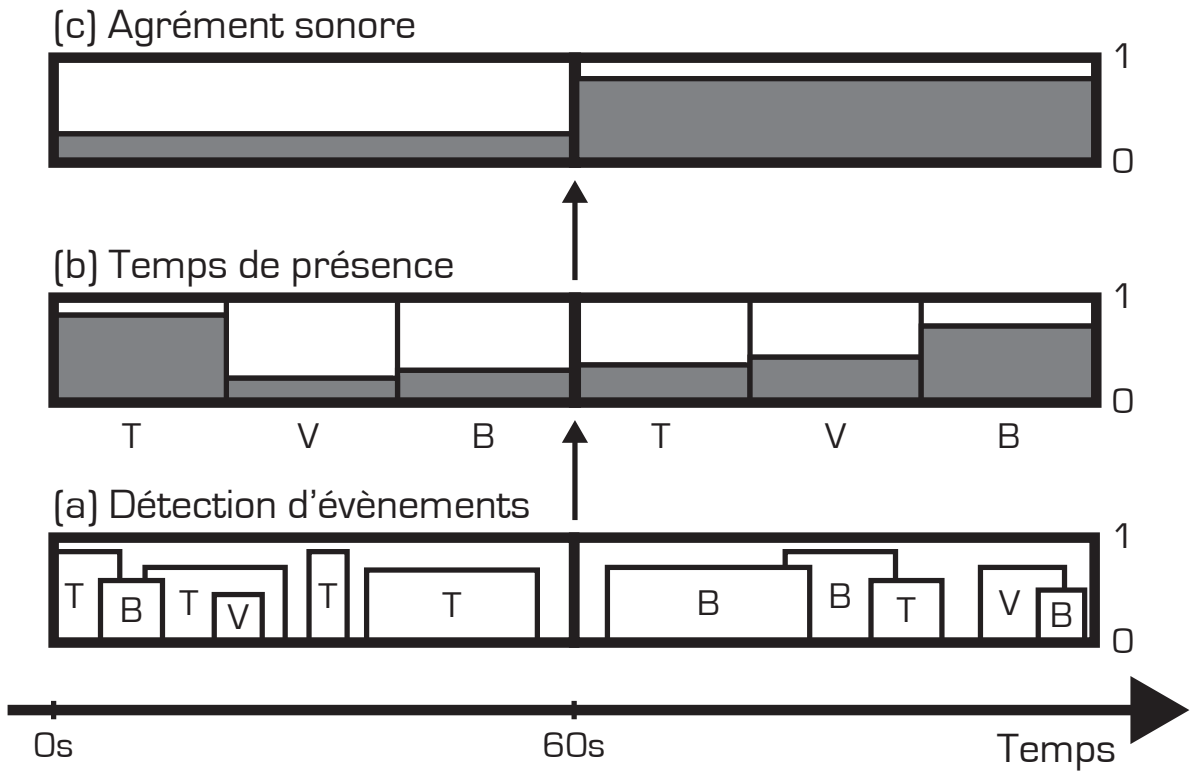
Caractérisation des environnements sonores urbains (ANR CENSE)

- les qualifiants perceptifs de haut niveau comme l'agrément sont corrélés au temps de présence perçu des sources
- prédiction de ces valeurs perceptives par des approches neuronales profondes

---

F. Gontier et al. (under revision). « Estimation of the perceived time of presence of sources in urban acoustic environments using deep learning techniques ». In : *Acta Acustica*.

Caractérisation des environnements sonores urbains (ANR CENSE)



F. Gontier et al. (under revision). « Estimation of the perceived time of presence of sources in urban acoustic environments using deep learning techniques ». In : *Acta Acustica*.

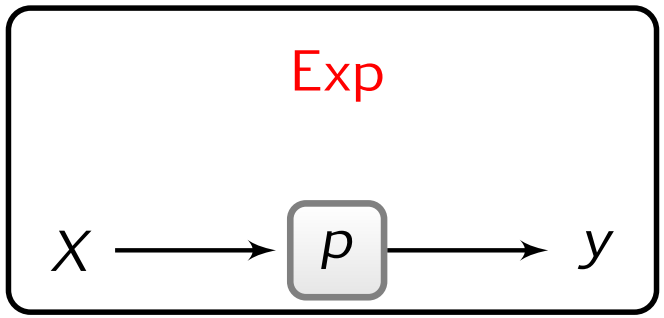
Projet de recherche avec le Conservatoire National de Musique et de Danse de Paris (CNSMDP)

- Recherche par similarité dans des corpus de modes de jeux étendus
- Confrontation d'un modèle computationnel de perception à des jugements experts
- L'opérateur de diffusion d'ondelettes apporte d'excellents résultats

joint (1s) + lmnn	-lmnn	(25 ms)	séparable	mfcc	
$96\% \pm 2$	$93\% \pm 3$	$91\% \pm 4$	$91\% \pm 4$	$82\% \pm 7$	($aP@5$)

V. Lostanlen, C. El-Hajj et al. (to be submitted). « Learning Auditory Similarities Between Instrumental Playing Techniques ». In : *EURASIP Journal on Audio, Speech, and Music Processing*.

ExpLanes



ExpLanes, un environnement logiciel qui facilite :

1. la gestion des calculs
2. le traitement des résultats
3. la reproducibilité

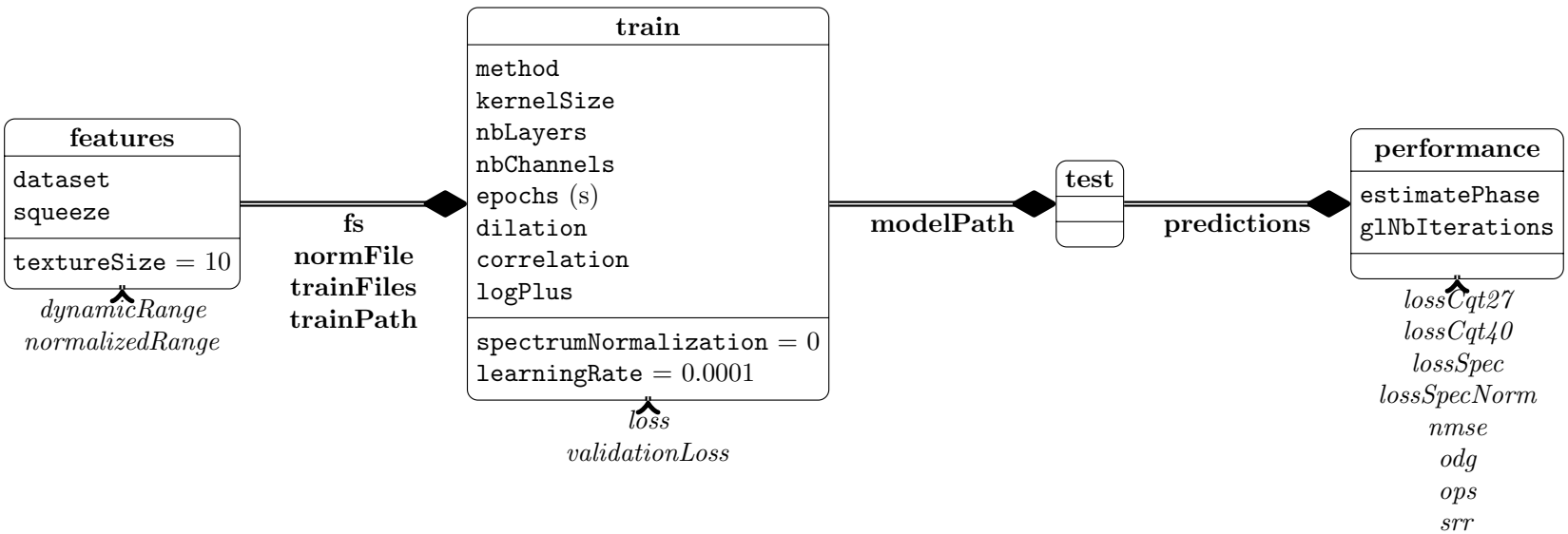
ExpLanes

`http://mathieulagrange.github.io/expLanes`

ExpLanes, un environnement logiciel qui facilite :

1. la gestion des calculs
2. le traitement des résultats
3. la reproducibilité

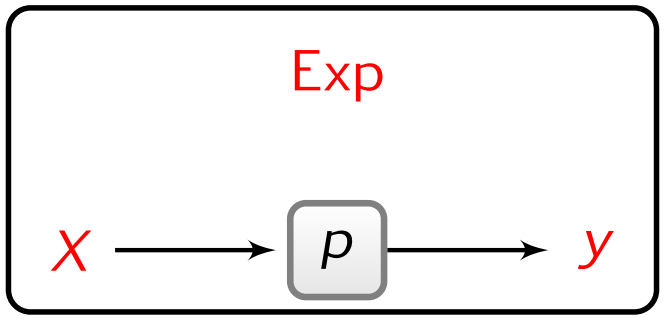
ExpLanes



https://mathieulagrange.github.io/paperBandwidthExtensionCnn/demo

M. Lagrange et F. Gontier (soumis). « Bandwidth extension of musical audio signals with no side information using dilated convolutional neural networks ». In : *ICASSP*.

Approches étudiées



- X plus de contrôle : données simulées
- y plus de maîtrise : collaboration avec les communautés expertes
- Exp plus de formalisation : développement d'explanes

Perspectives de recherche

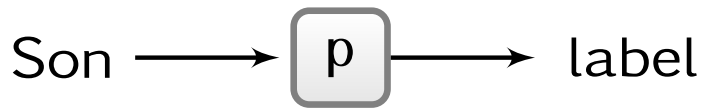
Problématiques en traitement du signal audionumérique

X-Y : codage, séparation de sources, extension de bande, inpainting, ...

X-y : recherche d'information

x-Y : synthèse

Problématiques en traitement du signal audionumérique

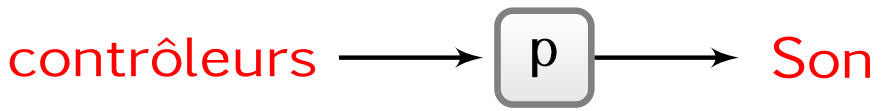


X-Y : codage, séparation de sources, extension de bande, inpainting, ...

X-y : recherche d'information

x-Y : synthèse

Problématiques en traitement du signal audionumérique

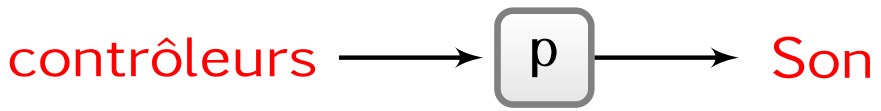


X-Y : codage, séparation de sources, extension de bande, inpainting, ...

X-y : recherche d'information

x-Y : synthèse

Problématiques en traitement du signal audionumérique



- attrait personnel pour l'inouï
- challenge
- en prise avec les avancées actuelles en apprentissage non supervisé

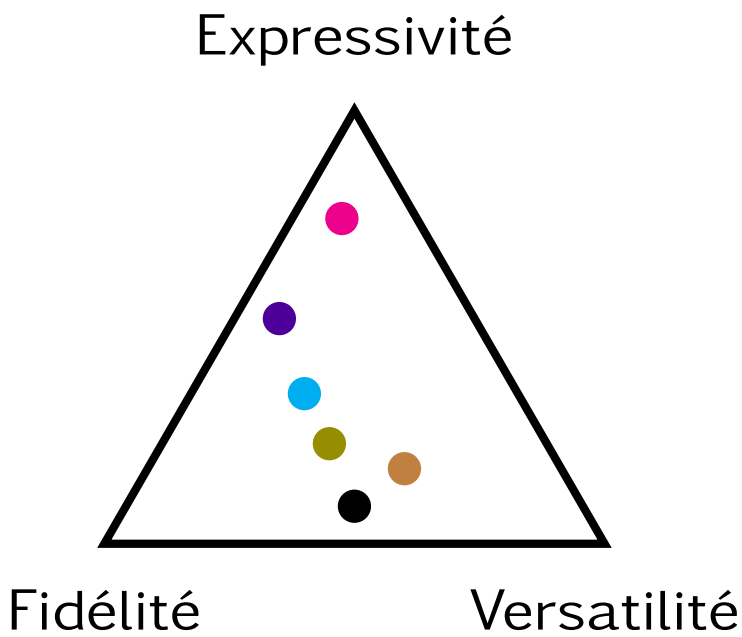
Requis

fidélité : ne pas produire d'artefacts audibles

expressivité : mécanismes de manipulation simples produisant une modification cohérente de la perception du signal résultant

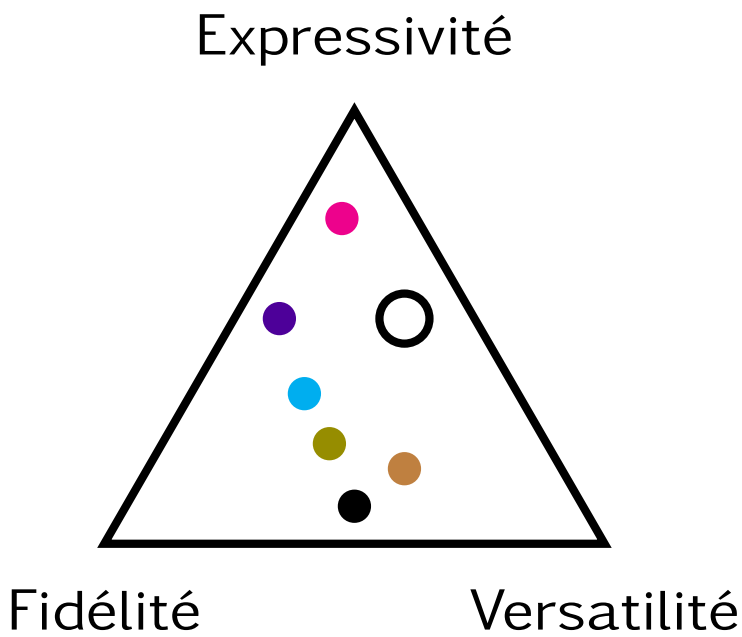
versatilité : les conditions de fidélité et d'expressivité sont remplies pour tout signal d'intérêt pour une tâche donnée

Existant



- Forme d'onde
- Spectrogramme
- Ondelettes
- Sinusoïdes à court terme
- Sinusoïdes à long terme
- Approches modales

Objectif



- Forme d'onde
- Spectrogramme
- Ondelettes
- Sinusoïdes à court terme
- Sinusoïdes à long terme
- Approches modales

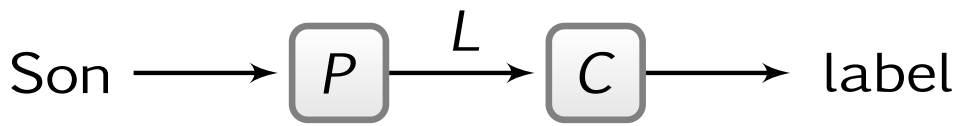
Challenges en traitement du signal

La synthèse en audio nécessite d'approcher les challenges suivants

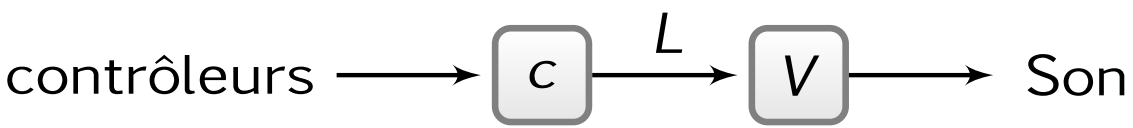
	fidélité	versatilité	expressivité
multirésolution	●	●	
causalité	●		●
non linéarité	●	●	
dimensionalité réduite			●
contrôle lent			●

Schéma fonctionnel

IR



Synthèse

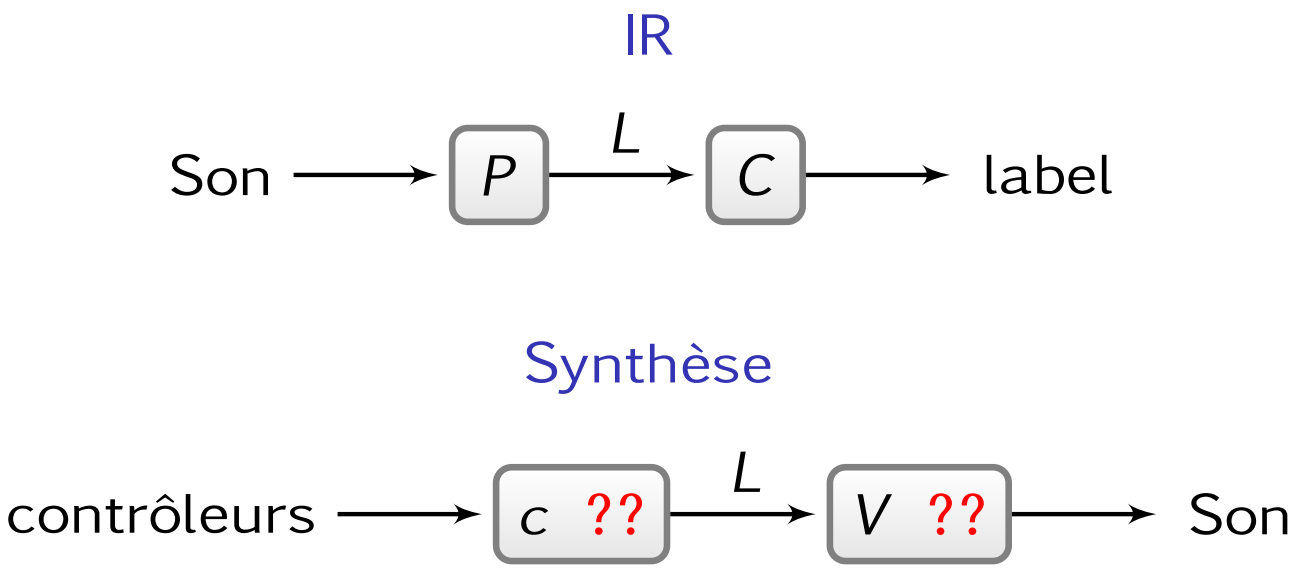


P : invariance / stabilité

c : conditionneur

V : vocodeur (P^{-1})

Schéma fonctionnel



P : invariance / stabilité

c : conditionneur

V : vocodeur (P^{-1})

Recherche en cours

Vocodeur : inversion de descripteurs pour la synthèse de scènes sonores respectueuses de la vie privée (Félix Gontier)

conditionneur : conception interactive en design sonore (Tom Souaille)

Problèmes ouverts

- Spécification des objectifs
  - mesure quantitatives de qualité perceptuelles
  - fonctions de coût perceptivement motivées
- Synthèse audio neuronale
  - modèles génératifs adversaires
  - approches basées échantillons
- Opérateur de diffusion d'ondelettes
  - pour le conditionnement
  - pour la synthèse par inversion

Réseau

Local

- Jean-François Petiot (LS2N)
- Arnaud Can, Judicaël Picaut, ... (UMRAE, IFSTTAR)

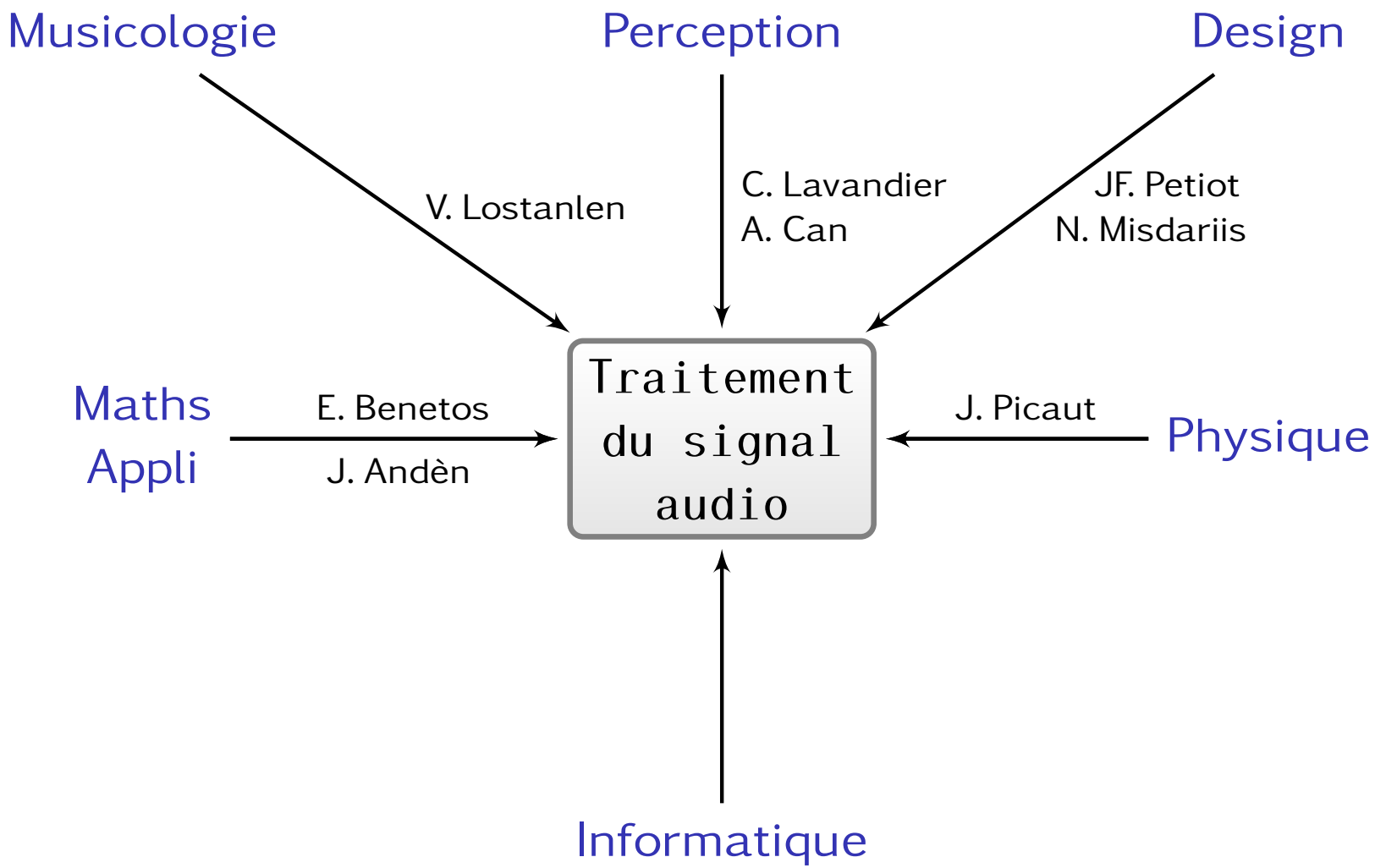
National

- Nicolas Misdariis (IRCAM)
- Catherine Lavandier (U. Cergy)

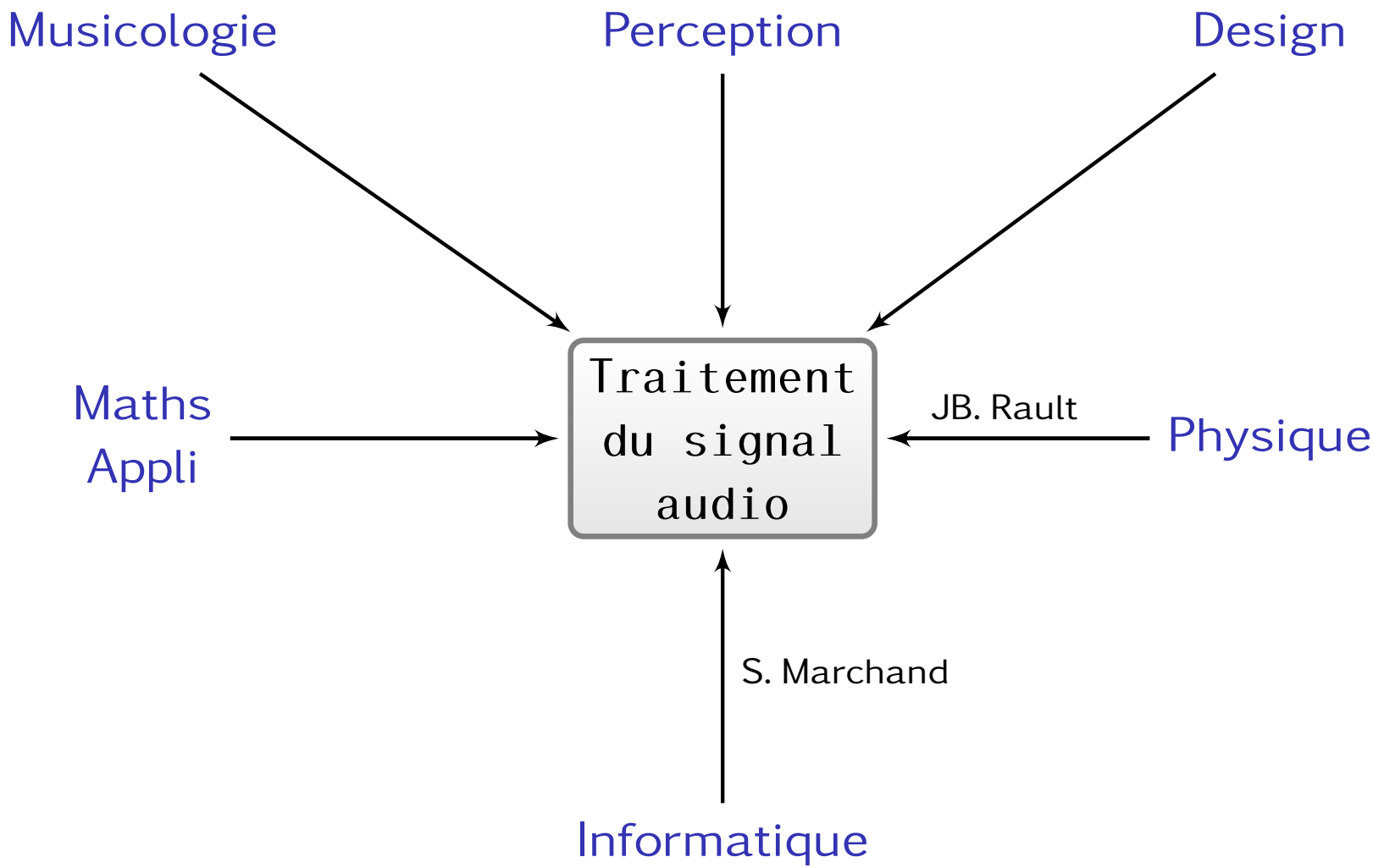
International

- Emmanouil Benetos (QMUL, UK)
- Vincent Lostanlen (NYU, US)
- Joakim Andèn (Flatiron Institute, US)

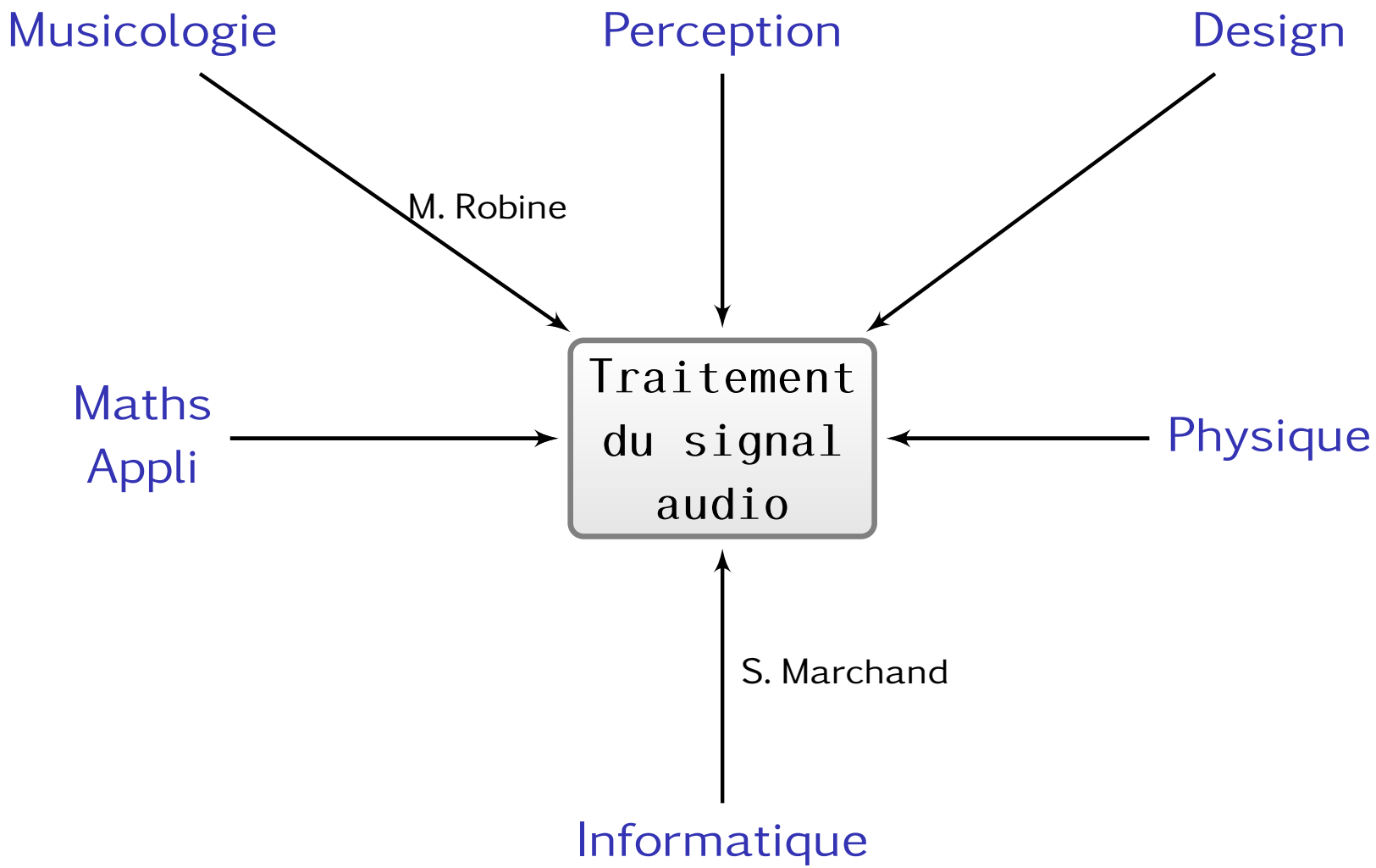
Carte thématique



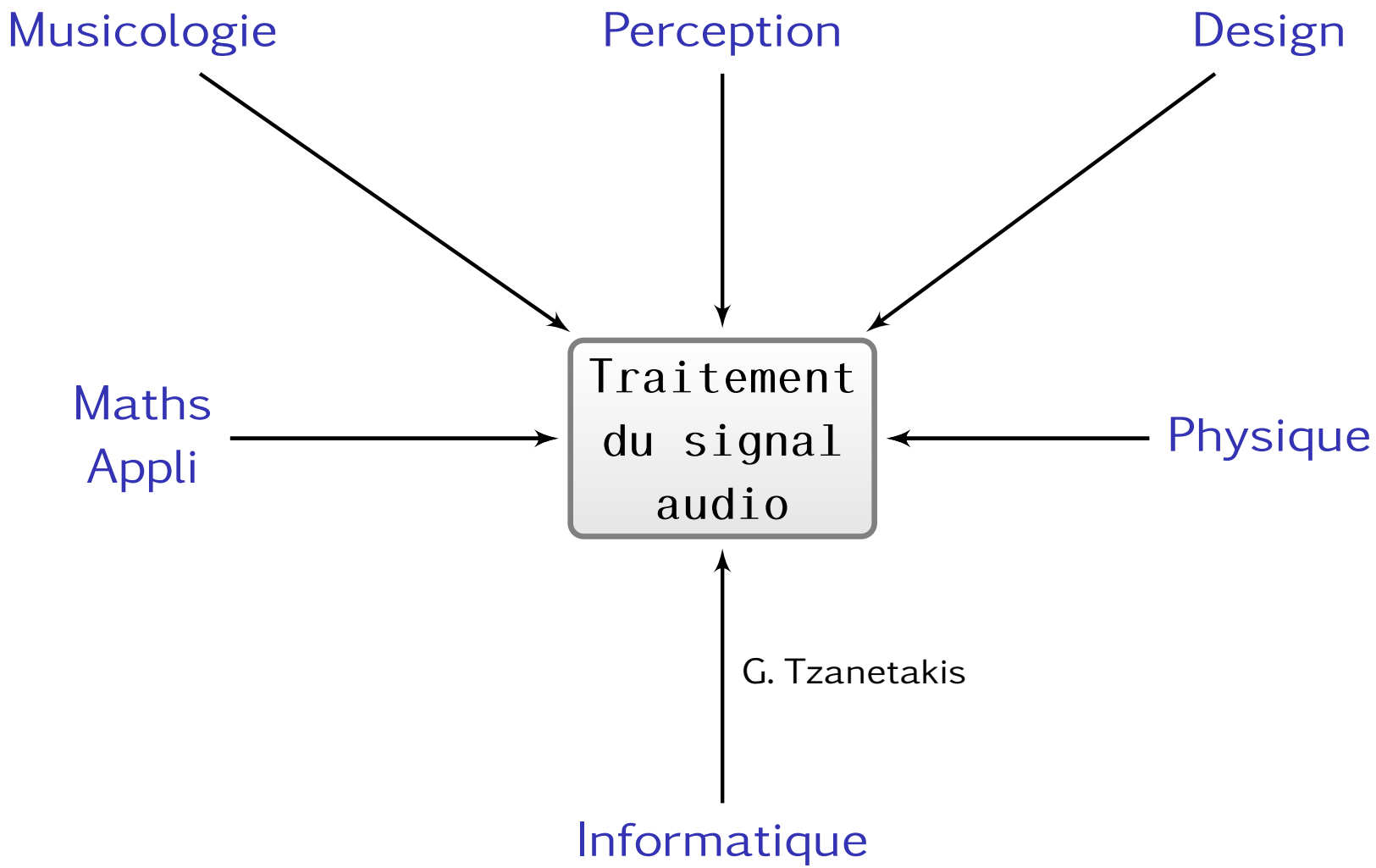
France Télécom R&D (2001-04)



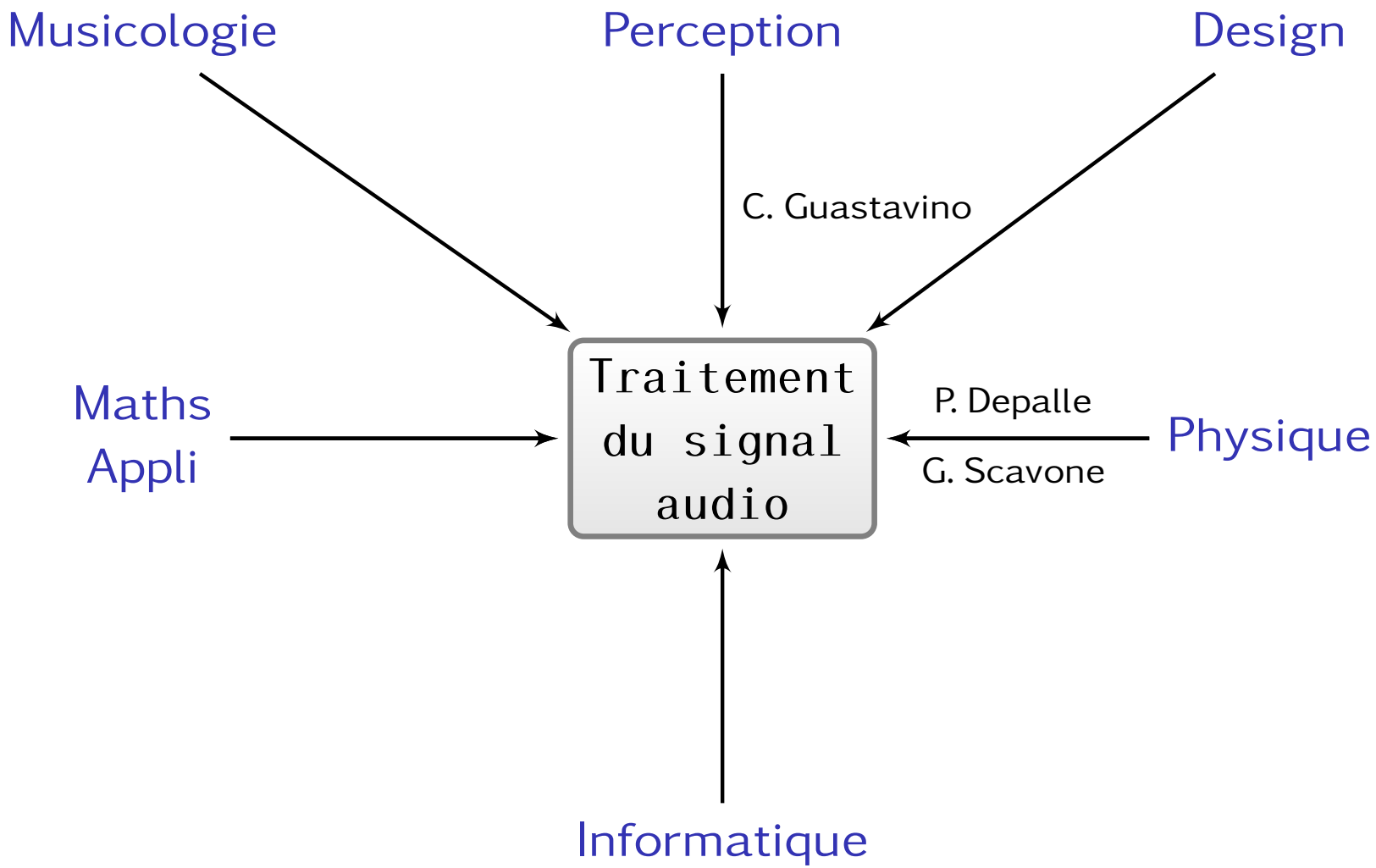
Bx (2004-06)


Musicologie
Perception
Design
M. Robine
Maths
Appli
Traitement
du signal
audio
Physique
S. Marchand
Informatique

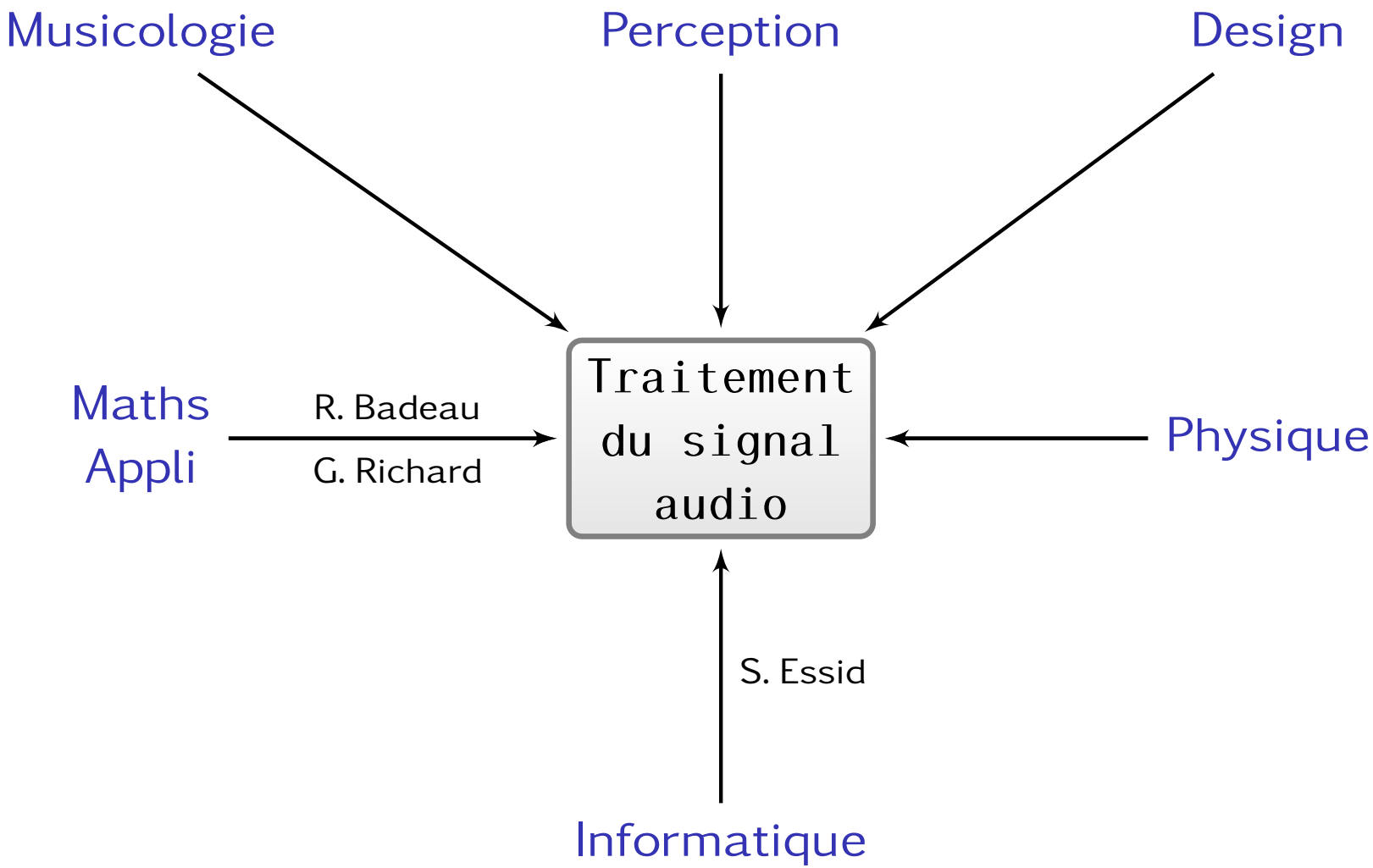
Uvic (2006-07)


Musicologie
Perception
Design
Maths
Appli
Traitement
du signal
audio
Physique
G. Tzanetakis
Informatique

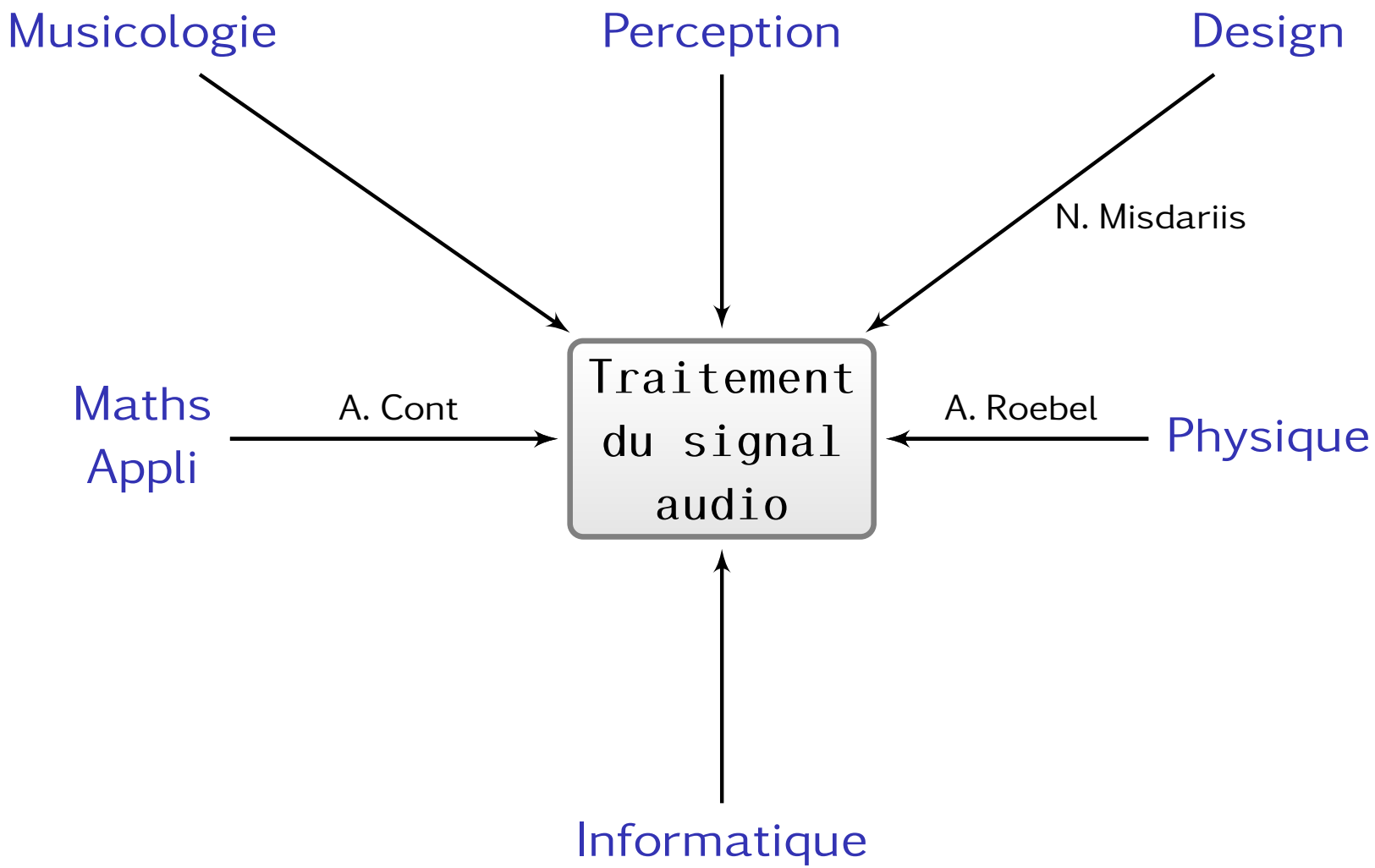
McGill (2007-08)



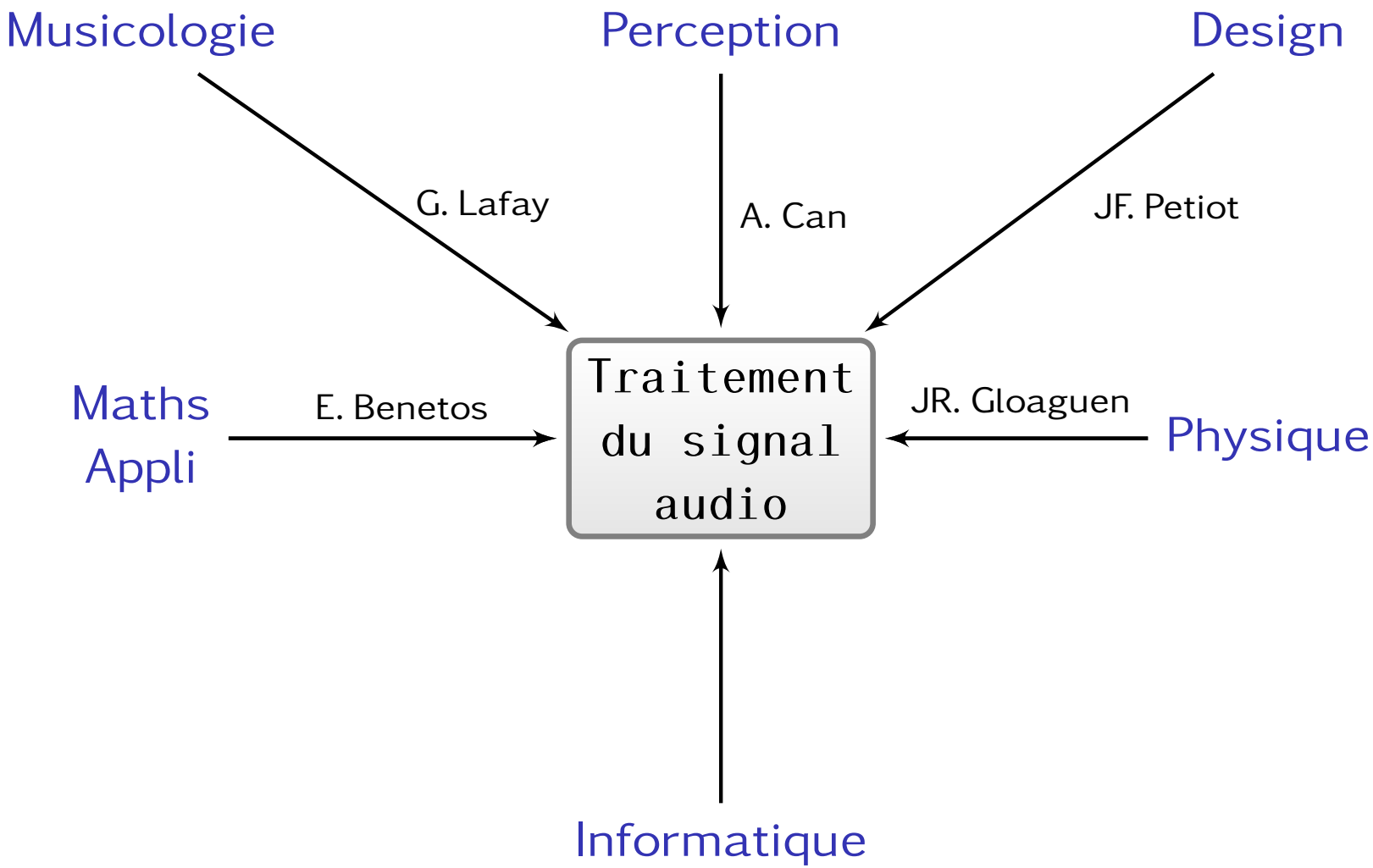
Télécom (2008-09)



Ircam (2009-13)



Ls2n (2013-19)


Musicologie
Perception
Design
G. Lafay
A. Can
JF. Petiot
Maths Appli
E. Benetos
Traitement du signal audio
JR. Gloaguen
Physique
Informatique

Un travail en inter-disciplinarité