

## 1er appel à participation ---

L'Ecole de Recherche Multimodale d'Information - Techniques & Sciences [ERMITES], édition VI, se tient les 27, 28 et 29 septembre 2011, sur Porquerolles la perle des Iles D'Or - Var

### Thème

**" Décomposition parcimonieuse, abstraction et structuration pour l'analyse de scène acoustique complexe "**

### Orateurs confirmés :

Adam O.; Bengio Y. ; Glotin H. ; Halkias X. ; Haton J.-P. ; Kowlaski S. ; MALLAT S. Razik J. ; Stylianou Y.

Organisée par le projet DYNI de l'UMR LSIS.

Soutenue par l'INRIA, CNRS LSIS, GDR13-ARIA, AFCP, USTV.

Sous le patronage de la Fédération de Recherche en Informatique et Interactions d'Aix-Marseille (FRIIAM) et du PEPS IN2SI DYNI "Indexation robuste à l'échelle".

Les participants peuvent soumettre un résumé de leurs travaux, qui si accepté sera présenté durant ERMITES et publié dans les actes.

### -- PROGRAMME --

Ouverture dès 10h le 27 septembre, centre IGESA Porquerolles

\*

**J.-P. Haton - Pr univ. Nancy - LORIA & IUF - <http://www.loria.fr/~jph/>**

**"Analyse de scène et reconnaissance stochastique de la parole"**

Les difficultés spécifiques pour la reconnaissance de la parole sont nombreuses: nombre variable de locuteurs connus ou inconnus, microphones distants, bruit ambiant, effets liés aux locaux (écho, réverbération).

Cet exposé rappelle les principes de la reconnaissance automatique de la parole et présentera les niveaux fonctionnels d'un système : prise de son, paramétrisation, analyses syntaxiques et sémantiques. Nous insisterons sur les modèles stochastiques actuellement les plus performants pour la reconnaissance de la parole.

\*

**Y. Stylianou – Pr univ. de Crète – ICS -**

**<http://www.ics.forth.gr/netlab/stylianou.html>**

**"Adaptive AM-FM decomposition of signals with applications to speech"**

I will present an adaptive AM-FM decomposition of signals and will comment its application to speech analysis, synthesis, and modifications.

As an introduction, I will present the classic Sinusoidal model showing its limitations in modeling non-stationary signals like speech.

Then the iterative adaptive AM-FM decomposition algorithm will be presented in the context of time-varying multi-component signals and robustness issues against wrong initial conditions/hypotheses and additive noise will be discussed on synthetic signals. Focusing in speech, it will be shown that this decomposition results in a high-resolution time frequency representation of speech. Comparisons with Gauss-Newton and Reassigned Spectrogram approaches will be presented.

Finally, possible applications on signals other than speech, like in vocalizations from marine mammals will be shortly discussed.

\*

**M. Kowlaski - MC univ. Paris XI - L2S -**

<http://web1.lss.supelec.fr/perso/kowalski/>

"Parcimonie et structures pour les décompositions des signaux dans des dictionnaires temps-fréquence"

Les décompositions parcimonieuses et structurées des signaux dans des dictionnaires de formes d'ondes peuvent s'obtenir par un modèle explicite dans le domaine des coefficients. On s'intéresse dans une première partie aux approches variationnelles, où les normes mixtes permettent d'introduire des couplages sur les coefficients préalablement organisés en groupes. L'optimisation des fonctionnelles s'effectue alors par des algorithmes de seuillage itératifs. En modifiant directement les opérateurs de seuillage, on peut introduire plus de flexibilité sur le couplage des coefficients, en particulier grâce à un Group-Lasso fenêtré.

Dans une seconde partie, on regardera des modèles de type Bernoulli-Gaussien, éventuellement multi-varié, pour la décomposition hybrides des signaux (aussi appelée décomposition en composantes morphologiques). Une telle approche à l'avantage de fournir un seuillage adaptatif des coefficients, et ainsi de s'affranchir des réglages parfois fastidieux d'hyper-paramètres intervenant dans les approches variationnelles.

\*

**Y. Bengio - Pr univ. Montréal - LISA -**

[http://www.iro.umontreal.ca/~bengioy/yoshua\\_en/](http://www.iro.umontreal.ca/~bengioy/yoshua_en/)

"Deep Unsupervised Learning of Representations"

Deep Unsupervised Learning of Representations et ferait un survol de découvertes récentes dans le domaine de l'apprentissage non-supervisé de représentations (surtout pour les images, transposable à l'acoustique).

Nous montrerons en particulier cette approche dans le but d'apprendre des abstractions de plus en plus haut niveau, et de découvrir les facteurs explicatifs des variations présentes dans les données.

\*

**S. Mallat - Ecole X, CMAP -** <http://www.cmap.polytechnique.fr/~mallat/>

"Scattering& Matching Pursuit for Acoustic Sources Separation"

Signal classes are usually invariant to groups of operators such as translations or scalings, and to larger Lie groups of deformations.

Invariants and dimensionality reduction play major roles for classification, as opposed to sparse complete representations.

Constructing invariants while keeping discriminant information deeply constrains the representation. We introduce a non-linear invariant representation, that reminds quantum scattering. It is implemented through the architecture of a deep neural network, but learning is needed at a single layer, and computed with  $O(N)$  operations. State of the art results are shown for object classification of deformed

patterns and random textures. Part of this work is described in

<http://arxiv.org/abs/1011.3023> .

\*

**J. Razik - MC univ. USTV LSIS -** <http://razik.univ-tln.fr/>

"Apprentissage de dictionnaire : de la parole humaine au chant de baleine"

L'utilisation des méthodes de dictionary learning non supervisé a montré dans le domaine de l'image être la méthode de l'état de l'art pour des tâches de classification. Dans le domaine de la parole, nous avons montré que ces méthodes permettent également de dépasser les classiques modèles

GMM en reconnaissance de phonèmes, tout en restant non supervisé. Nous avons également étudié l'apprentissage non supervisé par dictionary learning dans le cadre de l'analyse et la découverte d'un lexique associé au chant des baleines. L'objectif est d'obtenir des descripteurs plus génériques et plus efficaces que ceux habituellement extraits. Ceci permet également de déterminer des formes récurrentes dans la structure des chants entre les différents individus d'une population et sur plusieurs années. Ces travaux menés en collaboration avec S. Paris et H. Glotin sont illustrés de démonstrations.

\*

**X. Halkias - Columbia univ., Rosa Lab & post-doc DYNI Univ. Sud-Toulon Var**

### "Detection and Tracking of Dolphin Vocalizations"

Audio processing has widely affected the field of animal bioacoustics. Many audio engineers are now using their knowledge to advance our understanding of the world that surrounds us and especially that of animals. This work falls into that category, where the principles of signal processing, communication theory and machine learning are used to analyze the clandestine world of marine mammals and specifically dolphins.

In this talk, several algorithms are proposed for the successful detection of dolphin calls in long recordings. Starting with the simple and widely used thresholding detectors, several advancements are proposed, based on the data, leading to more intricate classifiers like Support Vector Machines (SVM) that are known for their robustness.

Lastly, two systems for pitch extraction are provided. The first system is based on a probabilistic framework and deals with the extraction of dolphin whistle calls while providing a first attempt on resolving simple overlaps. The second system assumes that the desired calls have already been detected and proceeds to identify the pitch for both whistle and burst calls using hierarchically driven Hidden

Markov Models (HMM).

\*

**O. Adam - Pr univ. Paris6 - LAM (Equipe Lutherie Acoustique Musicale)**

et CNPS (Centre de Neurosciences de Paris-Sud) -

<http://www.cb.upsud.fr/Olivier.htm>

### "Estimation de densité de population de baleine par analyse de leurs chants"

L'étude des cétacés passe par leurs observations. Plusieurs méthodes sont possibles et récemment le recours à l'acoustique passive a montré un grand intérêt, soit pour des observations d'espèces discrètes et/ou pour des observations à distance et/ou pour des observations continues.

Les techniques développées, relevant de l'acoustique sous-marine, l'instrumentation, du traitement du signal, de la reconnaissance des formes, ont pour but de détecter les sons émis par ces mammifères marins, de les classer, et de localiser lorsque cela est possible.

Nous présenterons une nouvelle activité dans ce domaine qui consiste à estimer la taille du groupe de cétacés à partir de leurs émissions sonores et des extrapolations statistiques. Plusieurs exemples sur différentes espèces seront présentés comme illustrations.

**H. Glotin - Pr univ. USTV - LSIS & IUF - <http://glotin.univ-tln.fr>**

### "Suivi de cachalot par acoustique passive"

Nous posons la problématique de l'analyse de scène acoustique au travers de la structuration de signaux bioacoustiques acquis dans des conditions variables. Nous nous intéresserons essentiellement au plus grand des odontocètes, le cachalot (45 tonnes, 18 mètres de long). Cet

extraordinaire plongeur, qui sonde plus d'une heure à moins 2000 mètres, possède le biosonar le plus puissant.

Avec plusieurs canaux (ici 4 hydrophones), nous montrons qu'il est possible de reconstruire la trajectoire de ses plongées écoutées à plusieurs kilomètres (voir démonstrations 3D <http://glotin.univ-tln.fr/PIMC/DEMO> et <http://lsis.univ-tln.fr/tcan/pimc/> ).

Nous montrons aussi que d'autres informations sont mesurables sur un seul canal, comme la taille de l'animal, son comportement. Des démonstrations sont faites sur des signaux que nous avons enregistrés au large de Porquerolles cette année, et sur des bases de référence de la communauté bioacoustique enregistrés au large des Bahamas par l'AUTEC

---

## Inscription :

Les pré-inscriptions sont ouvertes dès maintenant par simple email à [ermites@univ-tln.fr](mailto:ermites@univ-tln.fr) renseignant le type d'inscription, votre téléphone, email et affiliation.

Les places seront limitées aux trente premiers à avoir validé leur inscription par paiement à l'AFCP avant le 26 août (détails : <http://glotin.univ-tln.fr/ERMITES11> ).

---

## Tarifs :

A/ Formule pension complète (incluant 2 nuitées, 5 repas, 2 pts déj, pauses café / boissons, actes papier, en chambre studio 2 lits / salle de bain) :

A1/ Doctorant, Postdoc, Master = 280 euros,

A2/ Autre = 410 euros.

B/ Formule journée avec repas midi, actes, pauses café / boisson :

B1/ Doctorant, Postdoc, Master = 70 euros,

B2/ Autres = 100 euros.

---

## Comité de Programme :

H. Glotin (prés.), Pr univ. Sud-Toulon Var& IUF, LSIS DYNI

S. Paris, MC univ. Aix-Marseille, LSIS DYNI

J. Razik, MC univ. Sud-Toulon Var, LSIS DYNI

Y. Bengio, Pr univ. Montréal, LISA

S. Mallat, Ecole Polytechnique

O. Adam, Pr univ. Paris VI, LAM

Y. Stylianou, Pr univ. de Crête

J.-P. Haton, univ. Nancy& IUF, LORIA

## Comité d'Organisation :

S. Paris (prés.)

J. Razik

H. Glotin

F. Chamroukhi, MC USTV, LSIS DYNI

R. Abeille, doctorant USTV, LSIS DYNI

M. Chouchane, doctorant, LSIS DYNI

R. Delaye, ingénieur du son,

P. Machart, doctorant LSIS DYNI& LIF

F. Bénard, doctorant LSIS DYNI

Y. Doh, master, LSIS DYNI

---

## **Contexte général d'ERMITES :**

L'Ecole Recherche Multimodale d'Information : Technologies et Sciences fait intervenir, sur 3 jours, devant des chercheurs et ingénieurs en devenir, une dizaine de spécialistes qui travaillent sur les analyses conjointes de textes, images, sons et vidéos pour en extraire de l'information. La recherche d'information est de plus en plus complexe et hasardeuse compte tenu du volume sans cesse croissant des informations disponibles et de leur nature multimodale (images, sons, musique, parole, bio-acoustique, vidéos...). Nous avons décidé d'organiser ERMITES 2011 avec l'USTV, le support de l'UMR Laboratoire des Sciences de l'Information et des Systèmes (LSIS), du CNRS, de l'INRIA, du GDRI3-ARIA, du CNRS et LSIS, afin : - d'analyser les dernières avancées théoriques, et pratiques, des systèmes robustes de recherche d'information multimodale de grande dimension et sur des masses de données, - de montrer que ces nouveaux outils forment une base pour une domaine plus vaste que celui d'origine pour lequel ils ont été développés (plus particulièrement leur application au domaine audio et audio-vidéo),

- de sensibiliser les jeunes scientifiques au vaste champ scientifique mis en jeu par ces nouvelles avancées aux problèmes de fiabilité, robustesse, apprentissage, sous contrainte du passage à l'échelle.

Chaque spécialiste présentera un condensé de son savoir. L'esprit d'ERMITES étant de mettre l'accent sur les analyses jointes (transversales sur diverses modalités), ceci étant un défi majeur et encore peu pris en compte par la majorité des équipes qui restent plutôt dans le pré carré de leur spécialité de départ.

Site web de l'édition 2010 avec actes audiovisuels :

<http://glotin.univ-tln.fr/ERMITES10>

SITE WEB 2011 : <http://glotin.univ-tln.fr/ERMITES11>