

# Groupe 1

## Audio, Acoustique et Ondes (AAO)

Le groupe de recherche AAO (**A**udio, **A**coustique et **O**ndes) accueille des activités de traitement du signal s'appuyant fortement sur les phénomènes physiques de formation des signaux, qu'ils soient acoustiques ou optiques. Il est organisé autour de deux projets structurants :

- Le projet Traitement du Signal Audio (*AudioSig*),
- Le projet Traitement Optique du Signal (*TOS*)

**Responsable** G. Richard (P)

**Permanents** R. Badeau (MC), B. David (MC), S. Essid (IE, depuis 10/2006), C. Févotte (CR2-CNRS, depuis 10/2007), R. Frey (P, 40%), Y. Grenier (P), S. Maeda (DR CNRS), A. Maruani (P), D. Maignon (MC), N. Moreau (P), J. Prado (MC), I. Zaquine (MC, 80 %),

**Doctorants** A. Aissa El Bey (10/04 - 06/07), S. Fontana, M. Guillaume (10/03 - 11/06), N. Bertin (10/05 - ), M. Betser (10/04 - ), C. Clavel (11/03 - 03/07), V. Emiya (10/03 - ), J-L. Durrieu (01/07 - ), O. Gillet (12/03 - 06/07), P. Leveau (11/04 - 11/07), M. Ramona (10/06 - ), M. Alonso (10/02 - 11/ 06), C. Baras (10/02 - 06/06), A. Moreau (10/03 - 09/06), Q. He (11/05 - ), J-L. Smirr (01/07 - ), E. Ravelli (10/05 - ), C. Joder (11/07 - ), F. Vallet (11/07 - )

**Non-permanents** C. Hory (Postdoc 16 mois), C. Févotte (Postdoc 9 mois), M. Christensen (Aalborg Univ. (DK) Sabb. 1 mois)

Permanents [ <i>Institut ; CNRS</i> ] ; post-doctorants (ETP en mois)	128 [101; 27] ; 29
Doctorants (ETP en mois)	238
Thèses soutenues	8
HDR soutenues	2
Articles de revues [publiés]	25
Articles de conférences	63
C.A. 2006–2007 contrats [Europe+ANR+ACI ; CIFRE ; Autres] (k€)	[324 ; 39 ; 319]

### 1.1 Projet structurant Traitement du Signal Audio (AudioSig)

#### 1.1.1 Objectifs

L'objectif de ce projet est de développer des méthodes de traitement numérique du signal audio, afin de proposer des solutions aux principales problématiques centrées sur le son (parole, mu-

sique, . . .) dans les applications multimédia. Nous nous intéressons à toute la chaîne du signal audio, depuis sa capture, sa transmission, et jusqu'à sa restitution. Les travaux sont menés aussi bien sur le développement de nouvelles méthodes et modèles de représentation du signal audio et en particulier du signal musical (méthodes adaptatives d'estimation et suivi de sinusoides, représentations parcimonieuses, décompositions en matrices non-négatives, . . .) que sur leur application à des problèmes concrets (tatouage, compression, analyse des signaux EEG, indexation). Le thème de l'indexation sonore est probablement le plus important de ce projet et intègre aussi bien des travaux sur la segmentation de flux audio radiophonique en classes d'événements sonores (parole/musique/chant, . . .) que des études sur les différentes composantes structurantes d'un signal musical (séparation de sources d'un signal audio polyphonique, estimation de rythme, extraction de fréquences fondamentales multiples, extraction de mélodie principale, . . .). Par ailleurs, une nouvelle orientation, transversale, prend actuellement de l'essor notamment avec l'arrivée d'un nouveau CR2-CNRS sur ce thème, et vise à développer de nouvelles approches statistiques pour le traitement du signal audio.

Le projet entretient des collaborations étroites avec d'autres partenaires aussi bien académiques (LAM-Paris 6, IRCAM, INRIA-IRISA) qu'industriels (FT R&D, RTL, INA, . . .).

## 1.1.2 Résultats

### Prise et restitution des sons, spatialisation

**Chercheurs** K.Abed-Meraim, B. David, Y. Grenier, J. Prado, G. Richard ;

**Faits marquants** Thèse en co-tutelle avec l'Université de Parme (Italie) ; contrat d'étude avec France Télécom sur la séparation de sources audio dans un contexte automobile.

L'objectif de ce thème d'étude est d'améliorer l'analyse et la synthèse des champs sonores par des techniques numériques de traitement du signal. Dans le domaine de la restitution binaurale des sons, nous avons développé un système rapide de mesures des réponses de tête (HRTF) et de personnalisation pour un nouvel auditeur [?]. Nous avons mené, en collaboration avec l'université de Parme, des tests auditifs de diverses techniques de spatialisation des sons [?]

En prise de sons, nous avons développé une technique d'analyse du champ sonore à partir d'un réseau de microphones [?]. Celle-ci s'apparente aux techniques de formation de voie multi-microphones ainsi qu'aux techniques d'estimation spectrale paramétriques. La composante du champ dans chaque direction s'obtient par un filtrage optimal maximisant la résolution spatiale autour de la direction visée. Ce filtrage s'exprime sous forme de fonctions sphéroïdales aplaties.

Nous avons aussi développé une approche originale pour la séparation aveugle de sources audio à partir d'un réseau de microphones, dans le cas sous-déterminé (plus de sources que de capteurs). Cette approche combine une analyse temps-fréquence ou en ondelettes de chaque signal avec une extraction par classification automatique des vecteurs représentant les positions de chaque source [?, ?, ?].

### Tatouage, compression de sources sonores

**Chercheurs** N. Moreau, G. Richard

**Faits marquants** : Projet Media Puppet, Collaborations académiques (Université de Toulon, INPG Grenoble, LAM/Paris 6)

Le but poursuivi dans le domaine du tatouage audio a d'abord été de chercher à augmenter les performances (en terme de débit/taux d'erreurs binaires) du système de tatouage en proposant de nouvelles méthodes basées sur le fait qu'un système de tatouage pouvait être vu comme un canal de transmission avec information adjacente [?]. Récemment, l'accent a été mis sur la nécessité, pour un système de tatouage ayant vocation à satisfaire les besoins d'utilisateurs industriels (médiamétrie par exemple), d'être capable de détecter un tatouage au travers d'une liaison haut-parleurs/microphone séparés d'au moins 1m50. C'est une contrainte très forte que les techniques envisagées (égalisation par suivi adaptatif pour compenser l'effet de salle) n'ont permis de

satisfaire que partiellement. D'une façon plus générale, l'objectif a été d'être capable de détecter un tatouage après une prise de son très dégradante (microphone éloigné des haut-parleurs, microphone de qualité très ordinaire).

L'activité compression des signaux s'est également poursuivie autour de la compression des signaux numériques multicanaux à très faible débit et autour de nouvelles solutions pour les codeurs par transformée (utilisation de dictionnaires redondants [?], utilisation de modèles statistiques pour la quantification [?]).

### **Analyse de scènes sonores et Indexation**

**Chercheurs** R. Badeau, B. David, S. Essid, C. Févotte, Y. Grenier, J. Prado, G. Richard ;

**Faits marquants** Collaborations industrielles (FT R&D, Thales, RTL) et académiques (LAM-Paris 6, IRISA, Ircam)

Projets : Réseau d'excellence NoE Kspace (*Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content*), ACI Musicdiscover (*Indexation et recherche dans des bases de données audio*), ANR-Desam (*Décompositions en Eléments Sonores et Applications Musicales*), ANR Sarah (*Standardisation du Remastering Audio Haute-définition*) ;

Prix de thèse ParisTech 2006 (R. Badeau)

Cette activité poursuit plusieurs axes de recherches. Sur le premier axe qui vise à développer de nouvelles méthodes et modèles de représentation du signal audio, plusieurs résultats très intéressants ont été obtenus notamment sur l'estimation et le suivi adaptatif des composantes sinusoïdales d'un signal ([?, ?]). Le second axe est consacré aux différents aspects de l'indexation des signaux musicaux et s'attaque en particulier aux problèmes d'estimation de fréquences fondamentales multiples (détection de notes simultanées dans un signal musical [?]) d'estimation de l'information rythmique d'un segment musical (estimation du tempo ([?, ?], extraction automatique de la piste de batterie [?]) et à l'estimation de mélodie principale (par exemple par séparation de la voix chantée et de l'accompagnement musical). Le troisième axe s'intéresse à la segmentation d'un flux audio en événements ou classes sonores avec application aux flux radiophoniques ou audiovisuels (segmentation parole/musique [?], reconnaissance automatique des émotions [?]) et aux flux musicaux (reconnaissance des instruments de musique [?, ?], alignement audio/vidéo [?]). Une nouvelle orientation, transversale aux axes précédents, prend actuellement de l'essor et vise à développer de nouvelles approches statistiques pour le traitement du signal audio.

Ces thèmes de recherche sont en partie réalisés dans le cadre de projets collaboratifs nationaux (ANR-Desam, ACI-Musicdiscover) ou internationaux (Réseau d'excellence K-space).

### **Contrôle actif de bruit et analyse de signaux bio-médicaux**

**Chercheur** J. Prado, Y. Grenier ;

**Résultats principaux et faits marquants** Collaborations externes, ACI Abrupt

Dans le cadre de l'ACI ABRUPT notre activité s'est concentrée sur le développement d'algorithmes pour la réduction par contrôle actif du bruit de fond dans les centrales d'appel téléphonique. En raison de ses bonnes performances, notamment en termes de largeur de bande, nous nous sommes basés sur l'algorithme GMDF<sub>a</sub> (Generalised Multi-Delay Filter) que nous avons modifié afin de supprimer la reconstruction du signal par addition-recouvrement, cette technique ne pouvant pas amener à un fonctionnement temps réel. La suppression de l'addition-recouvrement provoque une légère baisse de performance dans la réduction de bruit, mais reste plus efficace que les algorithmes temporels (de type FXLMS) notamment en termes de largeur de bande du signal traité.

Une autre part de notre activité concerne l'analyse automatique du signal d'EEG de sommeil à l'aide d'une seule paire de capteurs. Des améliorations sensibles ont été apportées à la méthode, notamment par la mise au point d'une procédure d'adaptation automatique des largeurs de bandes

des filtres d'analyse aux fréquences caractéristiques du dormeur. Les résultats obtenus, en comparaison avec l'analyse visuelle, ont fait l'objet de publications et montrent tout l'intérêt de la méthode [?, ?].

## 1.2 Projet structurant Traitement Optique du Signal

**Chercheurs** R. Frey , A. Maruani , I. Zaquine ;

**Résultats principaux et faits marquants** Contrat incitatif de l'institut TELECOM : *Fonctions Réseaux pour l'Information Quantique*

Contrat de type CNANO de la Région, en collaboration avec l'Institut d'Optique Graduate School. 2 HDR soutenues (I. Zaquine, D. Matignon)

### Objectifs

La conception et caractérisation de dispositifs toujours plus performants pour le traitement optique de l'information, avec en particulier le concept de diffraction intracavité, est poursuivie dans des configurations plus audacieuses : faisceaux gaussiens, milieux à gain, réseaux minces.

Un nouveau type de structure, dit cristal photonique à 2 dimensions, est abordé dans le but d'allier les avantages de l'intégration, propres aux nanotechnologies, aux performances des meilleurs dispositifs macroscopiques.

Dans le cadre des traitements quantiques de l'information, une nouvelle activité de l'équipe, nous souhaitons étudier les fonctions d'un futur réseau quantique dans lequel le premier problème qui se pose est de passer d'une longueur d'onde « télécom » pour la propagation des signaux sur de longues distances, à une longueur d'onde à laquelle on serait capable de stocker l'information dans les futures mémoires quantiques, en préservant la cohérence quantique.

### Résultats

Les résultats expérimentaux obtenus sur un micro-laser YAG confirment les prévisions théoriques sur l'intérêt du milieu à gain intracavité [?]. L'efficacité de diffraction du réseau est augmentée d'un facteur 5000 et la sélectivité angulaire d'un facteur 20 environ. Les modélisations développées permettent de décrire plusieurs configurations allant du réseau infiniment mince [?] au réseau épais remplissant entièrement la cavité résonnante qui a été mis en œuvre expérimentalement.

Pour les cristaux photoniques les modèles ont été affinés pour prendre en compte les ordres supérieurs de diffraction. A l'aide de la double source paramétrique accordable en régime picoseconde mise au point au laboratoire on a pu mettre en évidence un régime de diffraction de Bragg avec un accroissement énorme de l'efficacité de diffraction dans ces cristaux à deux dimensions, ce malgré leur épaisseur micronique.

Les caractéristiques d'une interface quantique de changement de longueur d'onde pour permettre le stockage d'un photon "télécom" dans une mémoire quantique permettant de préserver la polarisation d'un photon ont été déterminées et le matériel a été acheté pour commencer les expériences [?].

### Publications les plus significatives

- [1] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier. Blind separation of underdetermined convolutive mixtures using their time-frequency representation. *IEEE Transactions on Audio, Speech & Language Processing*, 15(5):1540–1550, jul 2007.
- [2] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier. Underdetermined blind audio source separation using modal decomposition. *EURASIP Journal on Audio, Speech & Music Processing*, 2007:1–15, mar 2007.

- [3] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier. Underdetermined blind separation of nondisjoint sources in the time-frequency domain. *IEEE Transactions on Signal Processing*, 55(3):897–907, mar 2007.
- [4] M. Alonso, G. Richard, and B. David. Accurate tempo estimation based on harmonic+noise decomposition. *Eurasip Journal on Applied Signal Processing*, jan 2007.
- [5] M. Alonso, G. Richard, and B. David. Tempo estimation for audio recordings. *Journal of New Music Research*, 36(1):17–26, mar 2007.
- [6] R. Badeau and R. Boyer. Fast multilinear singular value decomposition for structured tensors. *SIAM Journal on Matrix Analysis and Applications*, mar 2007.
- [7] R. Badeau, B. David, and G. Richard. High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials. *IEEE Transactions on Signal Processing*, 54(4):1341–1350, apr 2006.
- [8] R. Badeau, B. David, and G. Richard. A new perturbation analysis for signal enumeration in rotational invariance techniques. *IEEE Transactions on Signal Processing*, 54(2):450–458, feb 2006.
- [9] R. Badeau, G. Richard, and B. David. Performance of ESPRIT for estimating mixtures of complex exponentials modulated by polynomials. *IEEE Transactions on Signal Processing*, 56(2):492–504, feb 2008.
- [10] C. Baras, N. Moreau, and P. Dymarki. Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1772–1782, sep 2006.
- [11] C. Berthomier, X. Drouot, M. H.-S. M, J. Prado, J. Mattout, and M. d’Ortho. Automatic analysis of single-channel sleep eeg: Validation in healthy individuals. In *SLEEP 2007*, pages 1587–1595, 2007.
- [12] C. Berthomier, X. Drouot, M. H.-S. M, J. Prado, J. Mattout, and M. d’Ortho. Real-time automatic measurement of recorded sleep time. In *American College of Chest Physicians (ACCP) congress : CHEST*, pages 1587–1595, Chicago, US, oct 2007.
- [13] M. Betser, P. Collen, G. Richard, and B. David. Estimation of frequency for am/fm models using the phase vocoder framework. *IEEE Transactions on Signal Processing*, 56(2):505 – 517, feb 2008.
- [14] D. Bitauld, I. Zaquine, A. Maruani, and R. Frey. Numerical analysis of a high resolution fast tunable filter based on an intracavity bragg grating. *Applied Optics*, 46(21):4728–4735, jul 2007.
- [15] A. Capra, S. Fontana, F. Adriaensen, A. Farina, and Y. Grenier. Listening tests of the localization performance of stereodipole and ambisonic systems. In *123rd Convention of the Audio Engineering Society*, New York, USA, oct 2007.
- [16] C. Clavel, I. Vasilescu, G. Richard, and L. Devillers. Du corpus émotionnel au système de détection : le point de vue applicatif de la surveillance dans les lieux publics. *Revue en Intelligence Artificielle (RIA)*, 20(4-5):529–551, sep 2006.
- [17] O. Derrien, P. Duhamel, M. Charbit, and G. Richard. A new quantization optimization algorithm for the mpeg advanced audio coder using a statistical sub-band model of the quantization noise. *IEEE Transactions on Audio, Speech and Language Processing*, 14(4):1328–1339, jul 2006.
- [18] V. Emiya, R. Badeau, and B. David. Multipitch estimation of inharmonic sounds in colored noise. In *10th Int. Conf. on Digital Audio Effects (DAFx-07)*, pages 93–98, Bordeaux, France, sep 2007.
- [19] S. Essid, G. Richard, and B. David. Instrument recognition in polyphonic music based on automatic taxonomies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):68–80, jan 2006.
- [20] S. Essid, G. Richard, and B. David. Musical instrument recognition by pairwise classification strategies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1401–1412, jul 2006.
- [21] C. Févotte, B. Torrèsani, L. Daudet, and S. J. Godsill. Sparse linear regression with structured priors and application to denoising of musical audio. *IEEE Trans. Audio, Speech and Language Processing*, 2007.
- [22] S. Fontana, Y. Grenier, and A. Farina. A system for head related impulse responses rapid measurement and direct customization. In *120th Convention AES*, Paris, France, oct 2006.
- [23] O. Gillet, S. Essid, and G. Richard. On the correlation of automatic audio and visual segmentations of music videos. *IEEE Trans. on Circuit and Systems for Video Technology*, mar 2007.
- [24] O. Gillet and G. Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech and Language Processing*, 16(3):529–540, Mar 2008.
- [25] M. Guillaume and Y. Grenier. Sound field analysis with a two-dimensional microphone array. In *ICASSP*, volume V, pages 321–324, Toulouse, France, may 2006.

- [26] M. Guillaume and Y. Grenier. Sound field analysis based on analytical beamforming. *EURASIP Journal on Advances in Signal Processing*, 2007, aug 2007.
- [27] Q. He, I. Zaquine, A. Maruani, and R. Frey. Band edge induced bragg diffraction in 2d photonic crystals. *Optics Letters*, 31(9):1184–1186, may 2006.
- [28] T. Hélie and D. Matignon. Diffusive representations for the analysis and simulation of flared acoustic pipes with visco-thermal losses. *Mathematical Models and Methods in Applied Sciences (M3AS)*, 16(4):503–536, apr 2006.
- [29] T. Hélie and D. Matignon. Representations with poles and cuts for the time-domain simulation of fractional systems and irrational transfer functions. *Signal Processing (SP)*, 86(10):2516–2528, oct 2006.
- [30] T. Hélie, D. Matignon, and R. Mignot. Criterion design for optimizing low-cost approximations of infinite-dimensional systems: towards efficient real-time simulation. *International Journal of Tomography and Statistics*, 7(7):13–18, sep 2007.
- [31] J. Kergomard, V. Debut, and D. Matignon. Resonance modes in a 1d medium with two purely absorbing boundaries: calculation methods, orthogonality and completeness. *Journal of the Acoustical Society of America (JASA)*, 119(3):1356–1367, mar 2006.
- [32] P. Leveau, E. Vincent, G. Richard, and L. Daudet. Instrument-specific harmonic atoms for mid-level music representation. *IEEE Transactions on Audio, Speech and Language Processing*, 16(1):116–128, jan 2008.
- [33] D. Matignon and H. Zwart. Standard diffusive systems as well-posed linear systems. *SIAM Journal on Control and Optimization*, page 24, jun 2007.
- [34] Y. Menesguen, J.-L. Smirr, G. Pillet, C. Bourdarias, R. Alléaume, A. Maruani, I. Zaquine, and R. Frey. Interface de changement de longueur d'onde conservant l'intrication en polarisation. In *COLOQ*, Grenoble, jul 2007.
- [35] A. Moreau, Q. He, I. Zaquine, A. Maruani, and R. Frey. Intracavity gain gratings. *Optics Letters*, 32(3):208–210, feb 2007.
- [36] A. Moreau, I. Zaquine, A. Maruani, and R. Frey. Réseaux minces de diffraction en régime de bragg. *Journal de Physique IV*, 135:239, oct 2006.
- [37] E. Ravelli, G. Richard, and L. Daudet. Extending transform coding to very low bitrates using overcomplete dictionaries. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, US, oct 2007.
- [38] G. Richard, M. Ramona, and S. Essid. Combined supervised and unsupervised approaches for automatic segmentation of radiophonic audio streams. In *ICASSP'07*, Honolulu, Hawai, apr 2007.