Chapter 1

Audio, Acoustics and Waves (AAO)

1.1 Executive Summary (Appendix 1)

Team Leader G. Richard

Initial Staff 7 Professors; 1 Research Scientist;

Staff who Left 3 Permanent Staff (89 months) ; 16 PhD Students (450 months) ; 5 Postdocs (103 months) ; 1 Engineer (24 months).

Permanent Staff who Were Hired A. Gramfort (09/2012), (PostDoc, CEA)

Scientific Highlights

- *International Projects:* Participation in 5 European projects including 2 Networks of excellence: IST Kspace (*Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content*) and 3DLife (*Bringing the Media Internet to Life*)); Obtention of a 3 years Marie Curie Grant for a research fellow exchange between AAO and Columbia University (Prof. D. Ellis).
- *Publications:* 224 publications (56 journals, 141 conferences, 11 book chapters, 17 PhD thesis) for a global H-number of the group for this period of 25 (e.g. 25 papers published in [2008-2013] are cited at least 25 times). The ten most cited papers gather an average of 89 citations (source: Googlescholar).
- *Patents:* 3 new patents were filed and one previous patent was transfered to the SME Invoxia. (Patent transfer and scientific support for the development of a hands-free IP telephone, including microphone array and loudspeaker array).
- *Open source software:* Lead participation in Scikit-Learn (the corresponding journal paper published in 2011 is cited 265 times) and full development and distribution of YAAFE (*Yet Another Audio Features Extractor*) with a growing impact with over 2300 downloads since March 2010 (463 downloads in 01-03/2013) from 79 different countries.
- *Award:* PhD prize in 2010 (jointly awarded by EEA club, GRETSI and ISIS) (N. Bertin, who is now a permanent CNRS researcher);

Scientific Production 56 Journals; 11 Book chapters; 141 Articles in Proceedings; 17 PhD thesis ; 32 Talks

---

Major Publications

- G. Richard, S. Sundaram and S. Narayanan, "An overview on Perceptually Motivated Audio Indexing and Classification", Proceedings of the IEEE, Vol. 101, n°9, September 2013.
- A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, M. Kowalski, Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations, Neuroimage, 15;70:410-22, April 2013

- V. Emiya, R. Badeau and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle", IEEE Transactions on Audio, Speech and Language Processing, vol. 18, n° 6, pp. 1643 1654, 2010.
- S. Essid and C. Févotte. Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring. IEEE Transactions on Multimedia, 15(2):415–425, Mar. 2013.
- R. Badeau, N. Bertin et E. Vincent, "Stability analysis of multiplicative update algorithms and application to non negative matrix factorization", IEEE Transactions on Neural Networks, vol. 21, n° 12, pp. 1869 1881, 2010.

Major Documents

- S. Essid et al. A multi-modal dance corpus for research into interaction between humans in virtual environments. Journal on Multimodal User Interfaces, pages 1–14, Aug. 2012.;
- A. Gramfort is one of the lead developer of *scikit-learn* which is a widely used machine learning toolbox and of the *MNE-python* toolbox for M/EEG data analysis. These two projects are open to students in the framework of the Google Summer of Code program.
- Database production and public release which includes 3 databases for music separation and transcription (MAPS, ENST-Drums, QUASI), 1 for robot audition (ROMEO-HRTF) and 1 for multimodal scenes analysis (ACM Grand Challenge).

Impact and Attractivity

- *Editing activities:* Editor and Associate editors of journals (A. Gramfort for *Jour. FBMIM*; G. Richard for IEEE Trans. on ASLP (2007–2011) and R. Badeau for EURASIP Jour. on AMSP (since 2012)); Guest editors of special issues in journals (B. David, Lead Guest Editor - IEEE Trans. on ASLP 2010; G. Richard, Guest Editor - IEEE JSTSP 2011 ; G. Richard Lead Guest Editor of EURASIP JASP 2013).
- *Organization of International conferences and workshops :* 14th edition of the International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS) 2013 (S. Essid, G. Richard : General Co-chairs); IEEE MMSP (Y. Grenier: Technical Co-chair); CFA 2010 (B. David, Technical co-chair); Acoustics08 (B. David, Technical Co-chair)
- Participation to technical committees of Scientific bodies (G. Richard, IEEE AASP TC), Major conferences (A. Gramfort, PRNI; G. Richard, ICASSP, Interspeech; S. Essid ACM MM, ICME; Y. Grenier, IWAENC; R. Badeau, ISSPA) and International PhD committees (G. Richard in 7 European countries).
- *National and International collaborations:* 65 % of published journal papers are co-authored with external collaborators; 5 European projects ; Collaboration with other research groups of LTCI in projects(OSEO-Quaero, FP7-VERVE, FP7-REVERIE, FP7-3Dlife) and PhD thesis supervision (3 joint PhDs with STA including 2 ongoing; 1 starting with MM).
- Invitation of tutorial/Keynotes talks in conferences (A. Gramfort, PRNI'2013, S. Essid and G. Richard at WIAMIS'2012, G. Richard at ACM Multimedia'2011) and in major international research labs.

Interaction with Economic and Social Spheres

- Public scientific conference at "Espace Pierres Gilles de Genes, ESPCI" ("Does the computer has the sense of rythmn ?") by B. David and G. Richard.
- 6 CIFRE PhD theses with Orange (1), INA (2), Arkamys (1), Parrot (1), Audionamix (1)).
- A patent from AAO was transfered to the SME Invoxia (Patent transfer and scientific support for the development of a hands-free IP telephone, including microphone array and loudspeaker array).
- Serving as experts for funding agencies : ANR-CONTINT (G. Richard, member of Programme committee), OSEO (S. Essid), Dutch Technology Foundation STW (S. Essid), European Union (G. Richard)
- Technology transfer to instrument makers: for more than 10 years now, AAO regularly attends the JFIS workshop (ITEMM) with the goal to tackle applied science projects with the stringed instruments luthiers. Leads to the PAFI ANR-project (B. David) where a software and hardware platforms have been developed and used in the today practice of the craftsmen.

Contributions to Higher Education

- Participation in University Masters *Méthodologie en Imagerie Médicale-Paris Descartes* (A. Gramfort), *Informatique-UPMC* (A. Gramfort, Y. Grenier, G. Richard (resp. of 1 UE)), *MVA-Paris Descartes/ENS Cachan* (Y. Grenier, G. Richard (resp. of 1 UE)), *ATIAM-UPMC* (R. Badeau, B. David (both resp. of 1 UE), G. Richard).
- Introduction of a new course of Signal processing based on active learning (e.g. problem and project based learning) for the 1st year of Telecom ParisTech engineering studies (B. David).
- Leading role in the reshape of the 1st year of study at Telecom ParisTech and in the proposal and then coordination of the newly introduced 6 months collaborative project (PACT, coord. B. David, [87, 91]).
- PhD students coordination: 17 PhDs awarded in the period. Amongst the 23 Phds students awarded in [2005 - 2011], 10 are now permanently employed in Academia, 2 at the European Patent office, 6 have permanent position in industry and 1 has started his company.
- Sessions for the benefit of "classes préparatoires" teachers, aka LIESSE: two sessions of a 2-day course on Python (A. Gramfort, S. Essid), 1 session on High Resolution Method (Y. Grenier, B. David, M. Maazaoui)

1.2 People

Team leader G. Richard (FP)

Faculty R. Badeau (AP), B. David (AP), Y. Grenier (FP), C. Févotte (–12/09), N. Moreau (FP, –03/2010), S. Essid (AP), J. Prado (AP,-02/2011); A. Gramfort (AP,10/2012-).

PhD students M. Betser (–06/08), N. Bertin (10/05-10/09), J-L. Durrieu (01/07-05/10), M. Ramona (10/06-06/10), C. Joder (11/07-09/11), L. Oudre (10/07-11/10), F. Vallet (11/07-09/11), S. Gulluni (02/08-12/11), R. Hennequin (10/08-11/11), M. Maazaoui (01/09-), S. Fenêt (01/10-); B. Fuentes (10/09-03/12); R. Foucard (10/09-); M. Moussalam (10/09-12/12); F. Rigaud (10/10-); A. Liutkus (01/10-11/12);N. Lopez (05/11-) A. Masurelle (10/11-); X. Jaureguiberry (10/11-); C. Fox (10/10-); A-C Conneau (01/12-); H. Bai (10/12-); N. Seichepine (10/12-).

PostDocs, engineers and sabbaticals M. Lagrange (Postdoc,10/08-09/09), A. Ozerov (02/08-07/09), T. Fillon (Postdoc,10/08-04/13), B. Mathieu (Engineer,10/08-12/10), H.Takeugming (12/11-11/12), A. Dielmann (PostDoc, 11/10-03/11), A. Drémeau (PostDoc, 09/11-08/13), C. Damon (PostDoc, 01/12-), D. Mauro (Post Doc, 01/13-)

External collaborators L. Daudet (Institut Langevin, Paris), O. Derrien (LMA-Marseille), E. Vincent (INRIA Nancy), L. Devillers (LIMSI-CNRS, Orsay), L. Girin (GIPSA-Lab, Grenoble), R. Boyer (LSS, Orsay), A. Ozerov (Technicolor), S. Marchand (Univ. de Bretagne occidentale), N. Bertin (CNRS-IRISA, Rennes), F. Gautier (LAUM, Le Mans), X. Boutillon (LMS, Polytechnique), N. Evans (Eurecom), T. Sikora (Technical Univ. of Berlin), N. O'Connor (Dublin City University), E. Izquierdo (Queen Mary Univ., London), P. Daras (CERTH, Thessaloniki), B. Thirion et G. Varoquaux (INRIA-Saclay), M. Hamalaïnen (Harvard), M. Descoteaux (Sherbrooke Univ.), Y. Hua (Univ. of California), D. Ellis (Columbia Univ., New York).

1.3 Overview

The overall objective of this research group is to develop digital signal processing methods with applications to audio, music, multimodal and biomedical signals. Its activities range from theoretical work on machine learning for signal processing, signal models and sparse representations to computational optimization of algorithms.

An increased effort was in particular dedicated to adaptive methods for high resolution sinusoidal components tracking [4, 5, 3] and sparse signal representations with a specific interest on those based on Matching Pursuit (MP), Probabilistic Latent Component Analysis (PLCA) or Non-negative Matrix factorization (NMF), that allow to decompose a signal using a limited number of atoms or basis functions. Several very interesting results were for example obtained for NMF concerning the stability of multiplicative update algorithms [6, 79], or the description of beta-divergence as a subclass of Bregman divergence [29]. Several extensions of the NMF were also explored including the introduction of a new generalized model for High-Resolution NMF [72], the extension to multichannel [46], the presentation of a novel geometric algorithm based on single-class Support Vector Machines [104], and the proposal of a general formulation of under-determined source separation of Gaussian Processes [39]. The applicability of these methods to generic problems such as audio indexing in the (scalable) compressed domain [51], audio source separation or music signal indexing was demonstrated by introducing specific constraints deduced from the audio signal properties (use of harmonicity or temporal constraints for music transcription [7, 82, 56], use of source production or timbre models for source separation [13, 14], use of time-frequency activations to model non-stationary audio events [28],...). This methodological effort explores both deterministic and statistical approaches.

Source separation also appears to be at the heart of this research group with applications in nearly all the individual research themes.

Besides this methodological axis, the research tackled by the group can be organized in three main themes (which will be further discussed in section 1.4):

1. *Machine listening and audio source separation:* The objective of this theme is to improve the capability of machines to analyse and interpret complex audio situations by developing specific digital signal processing methods. This is the main research theme of the group.
2. *Audio and multimodal signal processing:* The objective of this theme is first to develop novel generic models and approaches for audio signal representation and compression and second to automatically process multimodal data streams (segmentation, structuring,...).
3. *Biomedical signal analysis:* is dedicated to the analysis of biomedical signals, in particular electroencephalographic (EEG) and magnetoencephalographic (MEG).

In terms of bibliometry (source GoogleScholar), the group's faculty members have co-authored over 200 papers including 56 journal papers, 141 conference papers and 11 books or book chapters. Considering that each faculty is at most half-time on research, the group staff is 3.25 Equivalent Full Time (EFT) researchers and which leads an average number of 11,5 papers per year per EFT researcher. The H-number of the group for this period is 25 (e.g. 25 papers published in the period 2008-2013 are cited at least 25 times) and the ten most cited papers gather an average of 89 citations (ranging from 56 to 280).

Besides publications, the group promotes whenever possible research reproducibility by distributing open source software or by participating to open source software initiatives (for example, one of the members of the group is one of the lead developer of the scikit-learn which is a widely used machine learning toolbox and of the MNE-python toolbox for M/EEG data analysis). Recently, the group has been preparing jointly with with ENS-Cachan, with the support of the LMH (Hadamard) Labex, the launch of a new journal for reproducible research currently named *Audio and Signal Processing Algorithms Reviews*. This new journal will follow the spirit of an existing journal for image processing (IPOL).

In terms of attractivity and impact, the team is particularly visible at the international level with the participation in 5 European projects and the participation to international campaigns such as MIREX. Besides all members of the group are particularly active in Editing and Reviewing activities for major journals and conferences (editing activities in some of the most prestigious journals in this field including IEEE Transactions on Audio Speech and Language Processing, IEEE Journal on Selected Topics in Signal Processing, EURASIP journal on Advances in Signal Processing, Journal Frontiers in Brain Imaging Methods,...). Some members of the group are also particularly active in scientific bodies (SFA, IEEE Audio and Acoustic signal Processing TC) and are regularly invited to international PhD committees, area chair in major conferences such as ICASSP for example. The group is also regularly receiving visits from researchers from all around the world and candidacies for sabbatical. As an example, following a previous 2 months visit of Dr. Mads Christensen, Professor Juan Pablo Bello (New York University) will come to the AAO group for a one year sabbatical starting in January 2014 on a Fullbright grant.

The group has also strong interactions with the socio-economic world. First the group is involved in a variety of collaborative projects with industry. Then, the group is also developping bilateral collaborations with industry in particular through CIFRE PhD thesis (2 with INA, 1 with Orange, 1 with Arkamys, 1 with Audionamix, 1 with Parrot). In parallel of an active publication activity, the AAO group has also filed three new patents in the period. One of the patents previously filed was also transferred to the SME Invoxia in this period along with a scientific support for an efficient technology transfer.

1.4 Research Themes

1.4.1 Machine listening and audio source separation

Faculty B. David, Y. Grenier, S. Essid, R. Badeau, G. Richard, C. Févotte (–12/09);

Highlights: Scientific Production [43] (collaboration with Fraunhofer, Columbia Univ. and Tampere University, cited 50 times), [19] (cited 75 times), [6] (collaboration with INRIA, theoretical results, cited 21 times), [44] (collaboration with STA), [31] (cited 23 times)

Highlights: Impact

- ANR projects: DReaM (*Active music listening*) and DESAM (*Audio object decomposition with application to music*).
- CapDigital-ROMEO (*a project within "pôle de compétitivité" CapDigital, led by Aldebaran Robotics and aiming at creating a humanoid robot*), ROMEO2 (*a PIA "Projet d'Investissement d'Avenir", a follow-up to ROMEO aimed at bringing learning skill to a humanoid robot*)
- PhD prize in 2010 (jointly awarded by EEA club, GRETSI and ISIS) (N. Bertin);
- Organisation of a special issue in Eurasip Journal on Advances in Signal Processing on Informed Acoustic source separation analysis (2013, collaboration with Tampere Univ. of Techn., Bogazici Univ., Dublin Institute of Technology and Technicolor).
- OSEO-QUAERO (*Towards multimedia and multilingual search engines for professional and general public applications*), a very large scale French/German project.

Highlights: Interactions with Society

- Technology transfer to SME Invoxia (Patent transfer and scientific support for the development of a hands-free IP telephone, including microphone array and loudspeaker array).
- 2 patents filed including 1 in 2013 on Audio Fingerprinting.
- The open-source software YAAFE (*Yet Another Audio Features Extractor*). Growing impact (with over 2300 Downloads since March 2010 and 463 downloads between 01/01/2013 and 01/04/2013 from 79 different countries (mainly France, United States, Germany and United Kingdom)).
- Public scientific conference at "Espace Pierres Gilles de Genes, ESPCI" ("Does the computer has the sense of rythmn ?").
- 3 CIFRE PhD theses (with INA on Electro-acoustic music segmentation and transcription, Arkamys on speech dereverberation and with Parrot on noise reduction).

The objective of this theme is to improve the capability of machines to analyse complex audio situations by developing specific digital signal processing methods. This research theme encompasses a variety of situations ranging from speech signal dereverberation using a single microphone to complex audio and music scene analysis using one or several sensors.

Music analysis and audio source separation

A topic of major interest to the group is *Music transcription and source separation*, which are two intricate problems. Indeed, efficient source separation facilitates the transcription of the resulting sources and vice-versa. In music signal transcription, the group is directly interested in the four main problems which are *multiple fundamental frequencies estimation* (e.g. detection of simultaneous notes in a polyphonic musical recording [19, 56, 122]), *rhythmical information tracking* (tempo and beat estimation [143], *harmonic information estimation* (recognition of the chords sequence [208, 44, 45]) and *timbre recognition* (musical instrument recognition [30, 149]). Whenever possible, the results obtained are submitted to national or international evaluation campaigns. In particular in 2011, our group has obtained the best results in several subtasks of the Quaero competitive internal evaluation campaigns. Further, source separation approaches were developed for specific music transcription tasks such as piano transcription [19] and main melody estimation (by the use of a NMF-based source-filter model for separating the singing voice from the musical accompaniment [14]) but also for specific audio rendering tasks such as stereo signal remastering [46].

Another topic of interest in the group, related to the recognition of musical audio events discussed above, is the so-called audio fingerprinting problem. The objective of audio fingerprinting is to identify a given audio excerpt (e.g. obtaining metada such as title and artist identification in the case of a musical song) using the sole audio signal. Our work in this domain has concentrated on the introduction of simple *audio fingerprints* which are highly robust to the major signal degradations observed in broadcast streams [111] and on its capacity to scale up to very large databases or dynamically growing databases [112]. More recently, a major extension was introduced which led to a versatile system capable of identifying not only identical excerpts but also "semantically similar" excerpts with large acoustical variations (such as re-recording, live/studio versions and in some cases cover versions recorded with complete different musicians). A patent was recently filled on this topic.

Another topic of interest to the group is *multimodal music classification* where the focus is on the incorporation of prior knowledge on the nature and structure of music data into discriminative classifiers, both at the signal level and the semantic level, using all the available data, including ancillary information possibly attached to the content (available meta-data, tags...) and/or user interaction (*relevance feedback*). As such, efforts have been dedicated to the alternative route to music transcription that consists in achieving *music-to-score-alignment*, given that musical scores have become widely available over the Internet, which has made the approach of using such scores for music transcription highly appealing. Our contributions along this line are mainly the introduction of an effective and scalable statistical framework using *Conditional Random Fields* [31, 32]. Further, user-interactive systems have been devised that rely on *active learning* techniques for the analysis of the structure of particular forms of music, namely *electro-acoustic music*, which cannot be envisaged without taking into account the viewpoint of a human analyst [127, 128, 129]. Finally, in view of music similarity analysis [35, 116], the task of music auto-tagging (that is predicting user-tags for musical pieces) has been addressed, where multi-level, especially multi-scale classification systems have been developed using boosting techniques [117, 118].

Robot audition and blind source separation

A strong focus in robot audition is on multiple microphone techniques: beamforming for microphone arrays and blind source separation, some of these techniques being also applied to single microphone source separation and dereverberation.

Current work addresses the difficult problem of humanoid robot audition which needs, using a limited number of sensors, to be robust to movements of the robot and to highly variable environments. This task is part of the Romeo project that aims at building an humanoid robot (Romeo) that can act as a comprehensive assistant for persons suffering from loss of autonomy. Our approach follows a two-stage blind source separation strategy. The first stage consists in a fixed beamforming preprocessing to reduce the reverberation and the environmental noise. Due to the highly constrained context of robot audition, pre-recorded Head Related Transfer Functions (HRTFs) are used to estimate the beamforming filters. The use of the HRTF to estimate the beamformers allows to capture the head and torso effect on the manifold of the microphone array. The second stage is a blind source separation algorithm based on a l_1 norm minimization sparsity criterion. The results obtained highlighted the merit of the fixed beamforming preprocessing for improving the separation performances [161, 41]. A recent extension was also proposed by using a modified l_p norm blind source separation criterion based on the source sparsity in the time-frequency domain. We followed a tempered approach where the sparsity constraint could be reinforced by varying the parameter p of the l_p to dynamically change from l_1 to l_0 norm. This variation is driven by a sigmoid function which allows to obtain smooth transition and to avoid the divergence of this tempered approach. The merits of this method were demonstrated and compared to more classical schemes [162].

Our cooperation with Invoxia has permitted to develop a combination of a microphone array and a loudspeaker array that allows hands-free communications with high quality of the captured speech, and a 3D restitution of various distant speakers in a local listening room.

The transfer of a patent (Y. Grenier inventor) to Invoxia was the conclusion of this study. Invoxia has already designed two products using our technology (they call it In Vivo Acoustic: http://www.invoxia.com/fr/technologies/invivoacoustic), and the first of these products NVX 610 received a Best Innovation Award in CES 2012 (Las Vegas).

Another axis in this domain relates to signal capture in reverberant environment using a single sensor and a dedicated collaboration with the company Arkamys has permitted to develop novel dereverberation algorithms, based upon an estimation of the reverberation time [159].

1.4.2 Audio and multimodal signal analysis

Faculty R. Badeau, S. Essid, G. Richard, N. Moreau (–03/2010);

Highlights: Scientific Production [50] (2008, cited 37 times; collaboration with Institut Langevin); [42] (collaboration with Institut Langevin); [55] (collaboration with INA), [20] (collaboration with 5 European partners); [47] (collaboration with Technicolor,).

Highlights: Impact

- 2 European Networks of Excellence: FP6 Network of Excellence (NoE) IST Kspace (*Knowledge Space of Semantic Inference for Automatic Annotation and Retrieval of Multimedia Content*) and FP7-ICT NoE 3DLife (*Bringing the Media Internet to Life*)
- 1 European Integrated project FP7-ICT REVERIE (*REal and Virtual Engagement in Realistic Immersive Environments*)
- ANR DReaM (*Active music listening*) with collaboration with INPG Grenoble, Institut Langevin, University of Brest, Iklax.
- ACM MM' Grand Challenges: organization and data production for the 2011-2012 3DLife/Huawei challenges on *Realistic Interaction in Online Virtual Environments*;
- Organisation of the 14th edition of the International Workshop on Image and Audio Analysis for Multimedia Interactive Services (WIAMIS) 2013 - technically co-sponsored by IEEE SPS (http://wiamis2013.telecom-paristech.fr).

Highlights: Interactions with Society

- 1 Patent jointly filled with INPG Grenoble on Informed source Separation.
- Collaboration with Technicolor on Informed source Separation.
- 1 CIFRE Phd Thesis with INA on Audiovisual document structuring

Sound source compression, Acoustics and 3D Audio

In audio compression, the work was mostly dedicated to low to medium bit rate parametric audio coding. For low bit rate music coding applications, parametric coders are an efficient alternative to transform coders. In particular, sinusoidal modeling is widely used in response to the fact that most real-world audio signals are dominated by tonal components. Less used, the exponentially damped sinusoidal model (EDS) combined with a variable-length time segmentation is however considered as more powerful, but at the cost of an increased number of parameters. Our work has shown, however, that it is possible to design a joint scalar quantizer for amplitude, damping and phase parameters and obtain increased coding capabilities compared to the more traditional sinusoidal model. Our model incorporates in particular a dynamic temporal segmentation and psychoacoustic modeling and an asymptotically optimal entropy-constrained quantization method for the four sinusoid parameters (e.g. including damping) [92, 12].

On the other hand, investigations were pursued to develop highly scalable transform coders which can seamlessly operate from very low bit rate up to transparency. To that aim, sparse overcomplete representations are used to decompose the audio signals over a redundant union of bases (such as Modified Discrete Cosine Transform bases at different scales)[50]. It was also shown that the high flexibility of the signal representations used in this coder allows to address various audio indexing tasks (such as beat tracking or musical genre recognition) directly in the transformed domain [51] or to perform a large variety of music similarity tasks or structural-based audio coding [174]. More recently, a novel Random Matching Pursuit algorithm was designed which allows to simulate a local search in a larger dictionary while operating at the cost of a search in a sub-sampled dictionary. The approach consists in using a non adaptive random sequence of subdictionaries in the decomposition process, thus parsing a large dictionary in a probabilistic fashion with no additional projection cost nor parameter estimation. Based upon a theoretical modeling exploiting order statistics and experimental evidences, it was shown that the novel algorithm can be efficiently used on sparse approximation problems and successfully applied to signal compression [42]. On a more transversal axis, a comparative study of sparse greedy algorithms that had been independently introduced in speech and audio research communities was conducted. It was in particularly shown that the Matching Pursuit (MP) family of algorithms (MP, OMP, and OOMP) are equivalent to multi-stage gain-shape vector quantization algorithms previously designed for speech signals coding. Following this unified view, a new family of algorithms was introduced based on cyclic minimization principles and on the recent Cyclic Matching Pursuit [15].

In parallel, our work on Informed source separation allowed us to propose a novel framework to close the gap between source separation and audio coding domains by exploiting source separation models and principles for multichannel audio coding [40]. This novel approach, called Coding-based ISS (CISS) encodes the individual sources using not only a model as in source coding but also the observation of the mixture. This approach has several advantages including state of the art performance for multi-source audio coding in terms of rate-distorsion using Nonnegative Tensor Factorization as a source model [184, 47].

The group is also pursuing its activity in Acoustics and especially in audio rendering (or Audio3D) and musical acoustics. The audio rendering activity also benefits from the two European projects 3DLife and REVERIE. The group is in particular interested in developing novel hybrid approaches between pure physics-based approaches and perception-based approaches. One of the current lines of research consists in extending radiance-based transfer method to be effective for both the early part of the reverberation (early echoes) and late reverberation for which it was initially designed for. The musical acoustics activity is particularly focused to applying subspace methods and enumeration methods to the modal analysis of musical instruments, where it allows to investigate successfully the mid-frequency range [18, 17, 16, 192, 102]. This activity benefited from the ANR PAFI project, a four years project in collaboration with French instrument makers.

Audio-visual content and human activity analysis

As far as multimedia content analysis is concerned, the group's efforts are mainly geared towards audio-visual document segmentation and structuring, where the focus has been mainly on radio and TV content analysis [188, 205, 69].

On the methodological level, a special interest has been directed to kernel-based methods (Support Vector Machines, probabilistic distances, kernel change detection...) [189, 49, 55] allowing us to develop original and effective architectures for tasks such as *audio diarization*, that is segmentation into broad classes of events (especially music/speech discrimination) and more specifically *speaker diarization* [55].

Another line of work, conducted in collaboration with the STA group, is concerned with the development of new matrix factorisation techniques, which turn out to be particularly useful for document structuring [105, 21]. More recently, the focus has been on methods allowing a meaningful joint decomposition of "temporally related" parallel streams of data, especially the audio and visual streams of a video content [199].

In parallel, the topic of *human activity analysis* has attracted a growing interest within the AAO group, especially as part of its involvement in the 3DLife, EMC^2 and REVERIE European projects.

The work is centered at the development of machine learning and signal processing techniques[1] amenable to the analysis of data recorded through multiple capturing devices of different natures (microphone and video-camera arrays, inertial measurement units and motion capture devices, depth sensors, physiological sensors...). In general, the originality of our approach lies in the adoption of methodologies whereby the useful information is hunted for by spotting regularities emerging jointly across the concurrent streams of observed data. From the applicative viewpoint the group's work revolves around multimodal action/gesture classification, especially dance gesture analysis, motivated by a use-case that has been promoted by the 3DLife/Huwaei Grand Challenge within ACM multimedia 2011-2013, that is a virtual dance class scenario [20, 125, 106]. Problems of interest include dance performance alignment [108, 93], representation [153] and recognition.

1.4.3 Biomedical signal analysis

Researchers J. Prado (-02/2011), S. Essid (30%), A. Gramfort (100%);

Highlights: Scientific Production

[27] (collobaration with INRIA/Neurospin, Harvard medical school, Ilmenau university, Supelec); [210] (conference acceptance rate $\leq$ 20%; collaboration with INRIA/Neurospin and Ecole Centrale); [88] (collaboration with ESPCI).

Highlights: Impact

- DGA-DGCIS project MEEGAPERF (*Monitoring EEG pour l'Anticipation des PERFormances*);
- European project FP7-VERVE (*Vanquishing fear and apathy through E-inclusion: personalized and populated Realistic Virtual Environments for clinical, home and mobile platforms*)
- Development of the MNE-Python (http://martinos.org/mne/) package supported by 2 Google Summer of Code student in 2013

The third research direction of the group is dedicated to the analysis of biomedical signals, in particular electroencephalographic (EEG) and magnetoencephalographic (MEG) which are respectively electrical and magnetic signals induced by the electrical activity of active neurons. M/EEG offer a unique opportunity to non-invasively measure the brain activity at a millisecond time scale with clinical applications (epilepsy, sleep disorders) as well as for cognitive neurosciences and brain computer interfaces (BCI).

The team has pursued its long-standing work on asleep subjects recorded using a single pair of EEG electrodes. The developed approach has two technological breakthroughs: an automated analysis pipeline and the use of a single EEG channel. The efficiency and robustness of the developed method have been quantified and experimentally validated in collaboration with a French company called Physip founded by a former PhD student. Another application of interest was the analysis of biomedical data about colonic transit time (CTT). In particular, a dedicated approach was designed to robustly estimate this colonic transit time even in situations where the patient omits to ingest the radiopaque markers for one or two days [9].

The effort of the group in the domain of biomedical signal processing (especially multichannel EEG analysis) has been strengthened with the acceptance of two research projects. The first project (MEEGAPERF), started in September 2009, is centered at EEG-analysis for the realtime detection of physical performance decrease, using portable and lightweight EEG devices. The most recent work has been on artifact rejection [88] with specific constraints: noisy experimental setups and limited number of electrodes. The second project (FP7-VERVE) aims at developing dedicated tools to support the treatment of people who are at risk of social exclusion due to fear and/or apathy associated with a disability. The group's work is focused on the analysis of a

[1] often related to the ones developed for multimedia content analysis

patient's emotional state as he/she is submitted to a serious game treatment, based on EEG and ECG recordings used to monitor him/her.

The recent arrival of a new associate professor in biomedical signal processing, A. Gramfort, will allow this research topic to be further developed. Current directions are on the use of time-frequency representations for brain source localization [27], as well as data-driven representation learning using sparse coding and dictionary learning techniques. In his research, A. Gramfort works on the development of statistical machine learning techniques for mining brain imaging data (MEG, EEG and functional MRI). A recent collaboration with Ecole Centrale Paris led to a paper at the IPMI conference [210], known for being very selective.

1.5 Achievements (Appendix 6)

1.5.1 Scientific Productions

Articles in Journals

[1] A. Aissa El Bey, K. Abed-Meraim, Y. Grenier, and Y. Hua. A general framework for second order blind separation of stationary colored sources. *Signal Processing*, 88(9):2123–2137, Sept. 2008.

[2] R. Badeau and R. Boyer. Fast multilinear singular value decomposition for structured tensors. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1008–1021, Sept. 2008.

[3] R. Badeau, B. David, and G. Richard. Cramér-Rao bounds for multiple poles and coefficients of quasipolynomials in colored noise. *IEEE Transactions on Signal Processing*, 56(8):3458–3467, Aug. 2008.

[4] R. Badeau, G. Richard, and B. David. Fast and stable yast algorithm for principal and minor subspace tracking. *IEEE Transactions on Signal Processing*, 56(8):3437–3446, Aug. 2008.

[5] 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.

[6] R. Badeau, N. Bertin, and E. Vincent. Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization. *IEEE Transactions on Neural Networks*, 21(12):1869–1881, Dec. 2010.

[7] N. Bertin, R. Badeau, and E. Vincent. Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3):538–549, Mar. 2010.

[8] 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.

[9] M. Bouchoucha, J. Prado, L. Chtourou, G. Devroede, C. Atanassiu, and R. Benamouzig. Non-compliance does not impair qualitative evaluation of colonic transit time. *Neurogastroenterology and Motility*, 23(1):103–108, Jan. 2011.

[10] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette. Fear-type emotion recognition for future audio-based. *Speech Communication*, 50(6):487–503, June 2008.

[11] O. Derrien and G. Richard. A new model-based algorithm for optimizing the mpeg-aac in ms-stereo. *IEEE Transactions on Audio, Speech and Language Processing*, 16(8):1373–1382, Nov. 2008.

[12] O. Derrien, R. Badeau, and G. Richard. Parametric audio coding with exponentially damped sinusoids. *IEEE Transactions on Audio, Speech and Language Processing*, 21(7):1489–1501, July 2013.

[13] J.-L. Durrieu, G. Richard, B. David, and C. Févotte. Source/filter model for unsupervised main melody extraction from polyphonic audio signals. *IEEE Transactions on Audio, Speech and Language Processing*, Mar. 2010.

[14] J.-L. Durrieu, B. David, and G. Richard. A musically motivated mid-level representation for pitch estimation and musical audio source separation. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1180–1191, Oct. 2011.

[15] P. Dymarski, N. Moreau, and G. Richard. Greedy sparse decompositions: A comparative study. *EURASIP Journal on Advances in Signal Processing*, Aug. 2011.

[16] K. Ege, X. Boutillon, and B. David. High-resolution modal analysis. *Journal of Sound and Vibration*, 325(4):852–869, May 2009.

[17] B. Elie, F. Gautier, and B. David. Macro parameters describing the mechanical behavior of classical guitars. *Journal of the Acoustical Society of America*, 132(6):4013–4024, Dec. 2012.

[18] B. Elie, F. Gautier, and B. David. Estimation of mechanical properties of panels based on modal density and mean mobility measurements. *Mechanical Systems and Signal Processing*, July 2013.

[19] V. Emiya, R. Badeau, and B. David. Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *IEEE Transactions on Audio, Speech and Language Processing*, 18 (6):1643–1654, Aug. 2010.

[20] S. Essid and et al. A multi-modal dance corpus for research into interaction between humans in virtual environments. *Journal on Multimodal User Interfaces*, pages 1–14, Aug. 2012.

[21] S. Essid and C. Févotte. Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring. *IEEE Transactions on Multimedia*, 15(2):415–425, Mar. 2013.

[22] S. Essid and G. Richard. Fusion of multimodal information in music content analysis. *Dagstuhl Follow-Ups: Multimodal Music Processing*, Jan. 2012.

[23] C. Févotte, B. Torrésani, 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*, 16(1):174–185, Jan. 2008.

[24] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3), Mar. 2009.

[25] B. Fuentes, R. Badeau, and G. Richard. Harmonic adaptive latent component analysis of audio and application to music transcription. *IEEE Transactions on Audio, Speech and Language Processing*, 21(9), Sept. 2013.

[26] 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.

[27] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, and M. Kowalski. Time-frequency mixed-norm estimates: Sparse m/eeg imaging with non-stationary source activations. *Neuroimage*, 70(15): 410–422, Apr. 2013.

[28] R. Hennequin, R. Badeau, and B. David. Nmf with time-frequency activations to model non stationary audio events. *IEEE Transactions on Audio, Speech and Language Processing*, 19(4):744–753, May 2011.

[29] R. Hennequin, B. David, and R. Badeau. Beta-divergence as a subclass of bregman divergence. *IEEE Signal Processing Letters*, 18(2):83–86, Feb. 2011.

[30] C. Joder, S. Essid, and G. Richard. Temporal integration for audio classification with application to musical instrument classification. *IEEE Transaction on Audio, Speech and Language Processing*, 17 (1):174–186, Jan. 2009.

[31] C. Joder, S. Essid, and G. Richard. A conditional random field framework for robust and scalable audio-to-score matching. *IEEE Transaction on Audio, Speech and Language Processing*, 19(8):2385–2397, Nov. 2011.

[32] C. Joder, S. Essid, and G. Richard. Learning optimal features for polyphonic audio-to-score alignment. *IEEE Transactions on Audio Speech and Language Processing*, 21(10):2118–2128, Oct. 2013.

[33] J.-R. King, F. Faugeras, A. Gramfort, A. Schurger, I. El Karoui, J. Sitt, B. Rohaut, C. Wacongne, E. Labyt, T. Bekinschtein, L. Cohen, L. Naccache, and S. Dehaene. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage*, July 2013.

[34] M. Lagrange and M. Raspaud. Spectral similarity metrics for sound source formation based on the common variation cue. *ACM Multimedia Tools and Applications Journal on Content-Based Multimedia Indexing*, 48(1):185–205, 2010.

[35] M. Lagrange, M. Raspaud, R. Badeau, and G. Richard. Explicit modeling of temporal dynamics within musical signals for acoustical unit formation and similarity. *Pattern Recognition Letters (PRNSA)*, 31 (12):1498–1506, Sept. 2010.

[36] M. Lagrange, R. Badeau, B. David, N. Bertin, O. Derrien, S. Marchand, and L. Daudet. Décompositions en éléments sonores et applications musicales. *Traitement du Signal*, 28(6):665–689, Oct. 2011.

[37] J.-L. Le Carrou, F. Gautier, and R. Badeau. Sympathetic string modes in the concert harp. *Acta Acustica united with Acustica*, 95(4):744–752, July 2009.

[38] 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*, Jan. 2008.

[39] A. Liutkus, R. Badeau, and G. Richard. Gaussian processes for underdetermined source separation. *IEEE Transactions on Signal Processing*, 59(7):3155–3167, July 2011.

[40] A. Liutkus, J. Pinel, R. Badeau, L. Girin, and G. Richard. Informed source separation through spectrogram coding and data embedding. *Signal Processing*, 92(8):1937–1949, Aug. 2012.

[41] M. Maazaoui, Y. Grenier, and K. Abed-Meraim. Blind source separation for robot audition using fixed hrtf beamforming. *EURASIP Journal on Advances in Signal Processing*, (58), Mar. 2012.

[42] M. Moussallam, L. Daudet, and G. Richard. Matching pursuits with random sequential subdictionaries. *Signal Processing*, (92):2532–2544, May 2012.

[43] M. Mueller, D. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1088–1110, Oct. 2011.

[44] L. Oudre, C. Févotte, and Y. Grenier. Probabilistic template-based chord recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 19(8):2249–2259, Nov. 2011.

[45] L. Oudre, Y. Grenier, and C. Févotte. Chord recognition by fitting rescaled chroma vectors to chord templates. *IEEE Transactions on Audio, Speech and Language Processing*, 19(7):2222–2233, Sept. 2011.

[46] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for

audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, 3(18), Mar. 2010.
[47] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard. Coding-based informed source separation: Non-negative tensor factorization approach. *IEEE Transactions on Audio, Speech and Language Processing*, 21(8):1699–1712, Aug. 2013.
[48] M. Ramona, S. Fenet, R. Blouet, H. Bredin, T. Fillon, and G. Peeters. A public audio identification evaluation framework for broadcast monitoring. *Applied Artificial Intelligence: An International Journal*, 26 (1-2):119–136, Feb. 2012.
[49] M. Ramona, G. Richard, and B. David. Multiclass feature selection with kernel gram-matrix-based criteria. *IEEE Transactions on Neural Networks and Learning Systems*, 23(10):1611–1623, Oct. 2012.
[50] E. Ravelli, G. Richard, and L. Daudet. Union of mdct bases for audio coding. *IEEE Transactions on Audio, Speech and Language Processing*, 16(8):1361–1372, Nov. 2008.
[51] E. Ravelli, G. Richard, and L. Daudet. Audio signal representations for indexing in the transform domain. *IEEE Transactions on Audio, Speech and Language Processing*, Mar. 2010.
[52] G. Richard, S. Sundaram, and S. Narayanan. An overview on perceptually motivated audio indexing and classification. *Proceedings of the IEEE*, 101(9), Sept. 2013.
[53] F. Rigaud, B. David, and L. Daudet. A parametric model and estimation techniques for the inharmonicity and tuning of the piano. *Journal of the Acoustical Society of America*, 133(5):3107–3118, May 2013.
[54] J. Salomon, E. Gomez, D. Ellis, and G. Richard. Melody extraction from polyphonic music signals: Approaches, applications and challenges. *IEEE Signal Processing magazine*, July 2013.
[55] F. Vallet, S. Essid, and J. Carrive. A multimodal approach to speaker diarization on tv talk-shows. *IEEE Transactions on Multimedia*, 15(3):509–20, Apr. 2013.
[56] E. Vincent, N. Bertin, and R. Badeau. Adaptive harmonic spectral decomposition for multiple pitch estimation. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3):528–537, Mar. 2010.

Books

[57] N. Moreau. *Outils pour la compression des signaux, applications aux signaux audio*. Hermès Lavoisier, Paris, 2009.
[58] N. Moreau. *Tools for Signal Compression: Applications to Speech and Audio Coding*. Wiley-ISTE, 2011.

Book Chapters

[59] G. Adda, G. Chollet, S. Essid, T. Fillon, M. Garnier-Rizet, C. Hory, and L. Zouari. *Sémantique et multimodalité en analyse de l'information*, chapter 4 : Traitement des modalités "audio" et "parole". Hermes/Lavoisier, 2011.
[60] C. Baras, N. Moreau, and T. Dutoit. *Applied Signal Processing*, chapter 7 : How could music contain hidden information, pages 223 – 264. Springer, 2009.
[61] R. Benmokhtar, B. Huet, G. Richard, T. Declerck, and S. Essid. *Multimedia Semantics: Metadata, Analysis and Interaction*, chapter 4 : Feature Extraction for Multimedia Analysis. Wiley, 2011.
[62] C. Clavel and G. Richard. *Systèmes d'Interaction Emotionnelle*, chapter 5 : Reconnaissance acoustique des émotions,. Hermès, 2010.
[63] C. Clavel and G. Richard. *Emotional Interaction System*, chapter Recognition of acoustic emotion. Wisley, 2011.
[64] T. Dutoit and N. Moreau. *Applied Signal Processing*, chapter 3 : How is sound processed in an MP3 player, pages 65–101. springer, 2009.
[65] T. Dutoit, N. Moreau, and P. Kroon. *Applied Signal Processing*, chapter 1 : How is speech processed in a cell phone conversation, pages 1–31. Springer, 2009.
[66] S. Essid, M. Campedel, G. Richard, T. Piatrik, R. Benmokhtar, and B. Huet. *Multimedia Semantics: Metadata, Analysis and Interaction*, chapter 5 : Machine Learning Techniques for Multimedia Analysis. Wiley, 2011.
[67] G. Richard. *Encyclopedia of Data Warehousing and Mining, Second Edition.*, chapter Audio Indexing. Information Science Reference - IGI Global, 2008.
[68] M. Toda, S. Maeda, and K. Honda. *Turbulent Sounds in Speech*, chapter Formant-cavity affiliation in sibilant fricatives, pages 1–33. Mouton de Gruyter, 2009.

[69] F. Vallet, S. Essid, J. Carrive, and G. Richard. *TV Content Analysis*, chapter High-Level TV talk show structuring centered on speakers' interventions". CRC Press, Taylor Francis LLC, 2012.

Articles in Conference Proceedings

[70] K. Apostolakis and et al. Blending real with virtual in 3dlife. In *Paris*, July 2013.
[71] S. Arberet, A. Ozerov, R. Gribonval, and F. Bimbot. Blind spectral-GMM estimation for underdetermined instantaneous audio source separation. In *International Conference on Independent Component Analysis and Blind Source Separation (ICA'09)*, Paraty, Brazil, Mar. 2009.
[72] R. Badeau. Gaussian modeling of mixtures of non-stationary signals in the time-frequency domain (hr-nmf). In *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 253–256, New Paltz, New York, USA, Oct. 2011. IEEE.
[73] R. Badeau and B. David. Weighted Maximum Likelihood Autoregressive and Moving Average Spectrum Modeling. In *ICASSP'08*, pages 3761–3764, Las Vegas, Nevada, USA, Apr. 2008.
[74] R. Badeau and A. Dremeau. Variational bayesian em algorithm for modeling mixtures of non-stationary signals in the time-frequency domain (hr-nmf). In *ICASSP*, pages 6171–6175, Vancouver, Canada, May 2013. IEEE.
[75] R. Badeau and A. Ozerov. Multiplicative updates for modeling mixtures of non-stationary signals in the time-frequency domain. In *EUSIPCO*, Marrakech, Morocco, Sept. 2013.
[76] R. Badeau and M. D. Plumbley. Probabilistic time-frequency source-filter decomposition of non-stationary signals. In *EUSIPCO*, Marrakech, Morocco, Sept. 2013.
[77] R. Badeau and M. D. Plumbley. Multichannel hr-nmf for modelling convolutive mixtures of non-stationary signals in the time-frequency domain. In *WASPAA*, New Paltz, New York, USA, Oct. 2013. IEEE.
[78] R. Badeau, V. Emiya, and B. David. Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra. In *ICASSP'09*, pages 3073–3076, Taipei, Taiwan, Apr. 2009.
[79] R. Badeau, N. Bertin, and E. Vincent. Stability analysis of multiplicative update algorithms for non-negative matrix factorization. In *36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011. IEEE.
[80] W. Bailer, E. Dumont, S. Essid, and B. Mérialdo. A collaborative approach to automatic rushes video summarization. In *IEEE ICIP Workshop on Multimedia Information Retrieval: New Trends and Challenges*, Oct. 2008.
[81] C. Berthomier, A. Muzet, P. Berthomier, J. Prado, and J. Mattout. Real-time automatic measurement of drowsiness based on a single eeg channel. In *European Sleep Research Society*, Glasgow Scotland, Sept. 2008.
[82] N. Bertin, R. Badeau, and E. Vincent. Fast bayesian NMF algorithms enforcing harmonicity and temporal continuity in polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 29–32, New Paltz, New York, USA, Oct. 2009.
[83] N. Bertin, C. Févotte, and R. Badeau. A tempering approach for Itakura-Saito non-negative matrix factorization. With application to music transcription. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, pages 1545–1548, Taipei, Taiwan, Apr. 2009.
[84] R. Blouet, G. Rapaport, I. Cohen, and C. Févotte. Evaluation of several strategies for single sensor speech/music separation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'08)*, Las Vegas, USA, Apr. 2008.
[85] R. Boyer, R. Badeau, and G. Favier. Fast orthogonal decomposition of volterra cubic kernels using oblique unfolding. In *36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011. IEEE.
[86] S. Bozonnet, F. Vallet, N. Evans, S. Essid, J. Carrive, and G. Richard. A multimodal approach to initialisation for top-down speaker diarization of television shows. In *Eusipco*, Aug. 2010.
[87] M. Campedel, B. David, S. Lemarchand, and P. Bellot. Construire un espace commun pour l'équipe pédagogique et les élèves ? In *TICE*, Lyon, Dec. 2012.
[88] C. Damon, A. Liutkus, A. Gramfort, and S. Essid. Non-negative matrix factorization for single-channel eeg artifact rejection. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.
[89] C. Damon, A. Liutkus, A. Gramfort, and S. Essid. Nonnegative tensor factorization for single-channel eeg artifact rejection. In *IEEE International Workshop on Machine Learning for Signal Processing*, Southampton, UK, Sept. 2013.

[90] B. David, R. Hennequin, J.-L. Durrieu, and R. Badeau. Including parametric models in spectrogram decomposition. In *2nd Pan-American/Iberian Meeting on Acoustics*, Cancun, Mexique, Nov. 2010.

[91] B. David, M. Campedel, S. Lemarchand, M. Grojnowski, and P. Bellot. Le projet pact : créer les conditions pour apprendre la collaboration. In *QPES, Question de pédagogie dans l'enseignement supérieur*, Sherbrooke, Canada, June 2013.

[92] O. Derrien, R. Badeau, and G. Richard. Entropy-constrained quantization of exponentially damped sinusoids parameters. In *36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011. IEEE.

[93] A. Dremeau and S. Essid. Probabilistic dance performance alignment by fusion of multimodal features. In *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

[94] E. Dumont, B. Merialdo, S. Essid, W. Bailer, D. Byrne, H. Bredin, N. E. O'Connor, G. J. F. Jones, M. Haller, A. Krutz, T. Sikora, and T. Piatrik. A collaborative approach to video summarization. In *SAMT 2008, 3rd International Conference on Semantic and Digital Media Technologies*, Koblenz, Germany, Dec. 2008.

[95] E. Dumont, B. Merialdo, S. Essid, W. Bailer, H. Rehatschek, D. Byrne, H. Bredin, N. E. O'Connor, G. J. F. Jones, A. F. Smeatonand, M. Haller, A. Krutz, T. Sikora, and T. Piatrik. Rushes video summarization using a collaborative approach. In *TRECVID 2008, ACM International Conference on Multimedia Information Retrieval 2008*, Vancouver, BC, Canada, Nov. 2008.

[96] E. Dupraz and G. Richard. Robust frequency-based audio fingerprinting. In *ICASSP*, Dallas, USA, Mar. 2010.

[97] J.-L. Durrieu, G. Richard, and B. David. Singer melody extraction in polyphonic signals using source separation methods. In *ICASSP'08*, Las Vegas, Nevada, USA, Apr. 2008.

[98] J.-L. Durrieu, A. Ozerov, C. Févotte, G. Richard, and B. David. Main instrument separation from stereophonic audio signals using a source/filter model. In *European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, Aug. 2009.

[99] J.-L. Durrieu, G. Richard, and B. David. An iterative approach to monaural musical mixture de-soloing. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, Apr. 2009.

[100] B. Elie, F. Gautier, and B. David. Catégorisation d'instruments à cordes basée sur des mesures de mobilité de table d'harmonie. In *Congrès Français d'Acoustique*, Lyon, Apr. 2010.

[101] B. Elie, F. Gautier, B. David, and M. Curtit. fulltext access analysis of bridge admittance of plucked string instruments in the high frequency range. In *Forum Acusticum 2011*, Aalborg, Danemark, June 2011.

[102] B. Elie, M. Curtit, B. David, and F. Gautier. fulltext access analysis of mechanical admittance of violins in the mid- frequency range. In *Acoustics 2012*, Nantes, France, Apr. 2012.

[103] V. Emiya, R. Badeau, and B. David. Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches. In *EUSIPCO 2008*, Lausanne, Switzerland, Aug. 2008.

[104] S. Essid. A single-class svm based algorithm for computing an identifiable nmf. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, Mar. 2012.

[105] S. Essid and C. Févotte. Decomposing the video editing structure of a talk-show using nonnegative matrix factorization. In *International Conference on Image Processing (ICIP)*, Orlando, FL, USA, Oct. 2012.

[106] S. Essid, Y. Grenier, M. Maazaoui, G. Richard, and R. Tournemenne. An audio-driven virtual dance-teaching assistant. In *ACM Multimedia*, Scottsdale, Arizona, USA, Nov. 2011.

[107] S. Essid, X. Lin, M. Gowing, G. Kordelas, A. Aksay, P. Kelly, T. Fillon, Q. Zhang, A. Dielmann, V. Kitanovski, R. Tournemenne, N. E. O'Connor, P. Daras, and G. Richard. A multimodal dance corpus for research into real-time interaction between humans in online virtual environments. In *ICMI WORKSHOP ON MULTIMODAL CORPORA FOR MACHINE LEARNING*, Alicante, Spain, Nov. 2011.

[108] S. Essid, D. Alexiadis, R. Tournemenne, M. Gowing, P. Kelly, D. Monhagan, P. Daras, A. Dremeau, and N. E. O'Connor. An advanced virtual dance performance evaluator. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, Mar. 2012.

[109] P. Fechteler, P. Eisert, A. Hilsmann, D. A. Mauro, S. Broeck, F. Kuijk, D. Monaghan, P. Cesar, P. Daras, J. Wall, T. Zahariadis, R. Mekuria, M. Sanna, D. Alexiadis, and C. Stevens. A framework for realistic 3d tele-immersion. In *MIRAGE 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications*, page 8, Berlin (Germany), June 2013.

[110] S. Fenet, Y. Grenier, and G. Richard. Une empreinte audio à base de cqt appliquée à la surveillance de flux radiophoniques. In *GRETSI*, page NA, Bordeaux, France, Sept. 2011.

[111] S. Fenet, G. Richard, and Y. Grenier. A scalable audio fingerprint method with robustness to pitch-shifting. In *ISMIR*, pages 121–126, Miami, USA, Oct. 2011.

[112] S. Fenet, M. Moussallam, Y. Grenier, G. Richard, and L. Daudet. A framework for fingerprint-based detection of repeating objects in multimedia streams. In *EUSIPCO*, pages 1464–1468, Bucharest, Romania, Aug. 2012.

[113] C. Févotte and A. T. Cemgil. Nonnegative matrix factorisations as probabilistic inference in composite models. In *17th European Signal Processing Conference (EUSIPCO'09)*, Glasgow, Scotland, Aug. 2009.

[114] T. Fillon and J. Prado. A flexible multi-resolution time-frequency analysis framework for audio signals. In *11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pages 1124–1129, Montreal (Canada), July 2012.

[115] T. Fillon, J. Prado, and R. Badeau. Outil d'analyse temps-fréquence multi-résolution appliqué aux signaux audio. In *Colloque GRETSI 2013*, Brest, France, Sept. 2013.

[116] R. Foucard, J.-L. Durrieu, M. Lagrange, and G. Richard. Multimodal similarity between musical streams for cover version detection. In *ICASSP*, Dallas, USA, Mar. 2010.

[117] R. Foucard, S. Essid, M. Lagrange, and G. Richard. Multi-scale temporal fusion by boosting for music classification. In *ISMIR*, pages 663–668, Miami, USA, Oct. 2011.

[118] R. Foucard, S. Essid, M. Lagrange, and G. Richard. A regressive boosting approach to automatic audio tagging based on soft annotator fusion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

[119] R. Foucard, S. Essid, G. Richard, and M. Lagrange. Exploring new features for music classification. In *WIAMIS*, Paris, France, July 2013.

[120] C. FOX, M. Charbit, R. Badeau, B. DAVID, and G. VITTE. A subband hybrid beamforming for in-car speech enhancement. In *EUSIPCO*, pages 11–15, Bucarest, Roumanie, Aug. 2012.

[121] B. Fuentes, R. Badeau, and G. Richard. Analyse des structures harmoniques dans les signaux audio : modéliser les variations de hauteur et d'enveloppe spectrale. In *XXIIIème Colloque GRETSI*, Bordeaux, France, Sept. 2011.

[122] B. Fuentes, R. Badeau, and G. Richard. Adaptive harmonic time-frequency decomposition of audio using shift-invariant plca. In *36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011. IEEE.

[123] B. Fuentes, R. Badeau, and G. Richard. Blind harmonic adaptive decomposition applied to supervised source separation. In *20th European Signal Processing Conference (EUSIPCO)*, pages 2654–2658, Bucharest, Romania, Aug. 2012. EURASIP.

[124] B. Fuentes, A. Liutkus, R. Badeau, and G. Richard. Probabilistic model for main melody extraction using constant-q transform. In *37th International Conference on Acoustics, Speech, and Signal Processing ICASSP'12*, pages 5357–5360, Kyoto, Japan, Mar. 2012. IEEE.

[125] M. Gowing, P. Kell, N. E. O'Connor, E. Izquierdo, V. Kitanovski, X. Lin, Q. Zhang, C. Concolato, S. Essid, J. Le Feuvre, and R. Tournemenne. Enhanced visualisation of dance performance from automatically synchronised multimodal recordings. In *ACM Multimedia*, Scottsdale, Arizona, USA, Nov. 2011.

[126] A. Gramfort, B. Thirion, and G. Varoquaux. Identifying predictive regions from fmri with tv-l1 prior. In *Pattern Recognition in Neuroimaging (PRNI)*, Philadelphia, USA, June 2013.

[127] S. Gulluni, S. Essid, O. Buisson, E. Favreau, and G. Richard. Interactive segmentation of electro-acoustic music. In *2nd International Workshop on Machine Learning and Music (MML - ECML - PKDD)*, Bled, Slovenia, Sept. 2009.

[128] S. Gulluni, S. Essid, O. Buisson, and G. Richard. Interactive classification of sound objects for polyphonic electro-acoustic music annotation. In *AES Conference*, Ilmenau, Allemagne, July 2011.

[129] S. Gulluni, S. Essid, O. Buisson, and G. Richard. An interactive system for electro-acoustic music analysis. In *ISMIR*, Miami, USA, Oct. 2011.

[130] Z. Harchaoui, F. Vallet, A. Lung-Yut-Fong, and O. Cappé. A regularized kernel-based approach to unsupervised audio segmentation. In *ICASSP 2009*, pages 1665–1668, Taiwan, Apr. 2009.

[131] R. Hennequin, R. Badeau, and B. David. Spectral similarity measure invariant to pitch shifting and amplitude scaling. In *10ème Congrès Français d'Acoustique (CFA)*, Lyon, France, Apr. 2010.

[132] R. Hennequin, R. Badeau, and B. David. Time-dependent parametric and harmonic templates in non-negative matrix factorization. In *13th International Conference on Digital Audio Effects (DAFx 2010)*, Graz, Austria, Sept. 2010.

[133] R. Hennequin, R. Badeau, and B. David. NMF with time-frequency activations to model non-stationary audio events. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*,

pages 445–448, Dallas, Texas, USA, Mar. 2010. IEEE.
[134] R. Hennequin, R. Badeau, and B. David. Scale-invariant probabilistic latent component analysis. In *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 129–132, New Paltz, New York, USA, Oct. 2011. IEEE.
[135] R. Hennequin, B. David, and R. Badeau. Score informed audio source separation using a parametric model of non-negative spectrogram. In *36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011. IEEE.
[136] X. Jaureguiberry, G. Richard, P. Leveau, R. Hennequin, and E. Vincent. Introducing a simple fusion framework for audio source separation. In *Machine Learning for Signal Processing (MLSP)*, Southampton, Royaume-Uni, Sept. 2013.
[137] C. Joder, S. Essid, and G. Richard. Alignment kernels for audio classification with application to music instrument recognition. In *EUSIPCO 2008*, Lausanne, Suisse, Aug. 2008.
[138] C. Joder, S. Essid, and G. Richard. étude des descripteurs acoustiques pour l'alignement temporel audio-sur-partition musicale. In *GRETSI*, Dijon, Sept. 2009.
[139] C. Joder, S. Essid, and G. Richard. A conditional random field viewpoint of symbolic audio-to-score matching. In *ACM Multimedia*, Florence, Italie, Oct. 2010.
[140] C. Joder, S. Essid, and G. Richard. Approche hiérarchique pour un alignement musique-sur-partition efficace. In *CORESA 2010*, Lyon, France, Apr. 2010.
[141] C. Joder, S. Essid, and G. Richard. A comparative study of tonal acoustic features for a symbolic level music-to-score alignment. In *ICASSP*, Dallas, TX, E-U, Mar. 2010.
[142] C. Joder, S. Essid, and G. Richard. An improved hierarchical approach for music-to-symbolic score alignment. In *ISMIR*, Utrecht, Holland, Aug. 2010.
[143] C. Joder, S. Essid, and G. Richard. Hidden discrete tempo model: a tempo-aware timing model for audio-to-score alignment. In *ICASSP*, Prague, Rep. Tchèque, May 2011.
[144] C. Joder, S. Essid, and G. Richard. Optimizing the mapping from a symbolic to an audio representation for music-to-score alignment. In *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, Oct. 2011.
[145] M. Khadkevich, T. Fillon, G. Richard, and M. Omologo. A probabilistic approach to simultaneous extraction of beats and downbeats. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 445–448, Kyoto Japan, Mar. 2012.
[146] M. Lagrange and M. Raspaud. Spectral similarity metrics for sound source formation based on the common variation cue. In *Seventh International Workshop on Content- Based Multimedia Indexing (CBMI 2009)*, Chania (Greece), Sept. 2009. IEEE.
[147] M. Lagrange, R. Badeau, B. David, N. Bertin, J. Echeveste, O. Derrien, S. Marchand, and L. Daudet. The DESAM toolbox: spectral analysis of musical audio. In *13th International Conference on Digital Audio Effects (DAFx 2010)*, Graz, Austria, Sept. 2010.
[148] M. Lagrange, R. Badeau, and G. Richard. Robust similarity metrics between audio signals based on asymmetrical spectral envelope matching. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 405–408, Dallas, Texas, USA, Mar. 2010. IEEE.
[149] M. Lardeur, S. Essid, G. Richard, M. Haller, and T. Sikora. Incorporating prior knowledge on the digital media creation process into audio classifiers. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, Apr. 2009.
[150] A. Lenoir, R. Landais, G. Peters, L. Oudre, and T. Fillon. Muma: A music search engine based on content analysis. In *IEEE International Conference on Multimedia and Expo*, Barcelona (Spain), July 2011.
[151] A. Liutkus, R. Badeau, and G. Richard. Informed source separation using latent components. In *Ninth International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, volume 6365, pages 498–505, Saint Malo, France, Sept. 2010. Springer.
[152] A. Liutkus, R. Badeau, and G. Richard. Multi-dimensional signal separation with gaussian processes. In *IEEE Workshop on Statistical Signal Processing (SSP2011)*, Nice, France, June 2011. IEEE.
[153] A. Liutkus, A. Dremeau, D. Alexiadis, S. Essid, and P. Daras. Analysis of dance movements using gaussian processes. In *ACM Multimedia*, Nara, Japan, Nov. 2012.
[154] A. Liutkus, S. Gorlow, N. Sturmel, S. Zhang, L. Girin, R. Badeau, L. Daudet, S. Marchand, and G. Richard. Informed audio source separation: A comparative study. In *20th European Signal Processing Conference (EUSIPCO)*, pages 2397–2401, Bucharest, Romania, Aug. 2012. EURASIP.
[155] A. Liutkus, A. Ozerov, R. Badeau, and G. Richard. Spatial coding-based informed source separation. In *20th European Signal Processing Conference (EUSIPCO)*, pages 2407–2411, Bucharest, Romania, Aug. 2012. EURASIP.

[156] A. Liutkus, Z. Rafii, R. Badeau, B. Pardo, and G. Richard. Adaptive filtering for music/voice separation exploiting the repeating musical structure. In *37th International Conference on Acoustics, Speech, and Signal Processing ICASSP'12*, pages 53–56, Kyoto, Japan, Mar. 2012. IEEE.

[157] A. Liutkus, R. Badeau, and G. Richard. Low bitrate informed source separation of realistic mixtures. In *ICASSP*, pages 66–70, Vancouver, Canada, May 2013. IEEE.

[158] A. Liutkus, J.-L. Durrieu, L. Daudet, and G. Richard. An overview of informed audio source separation. In *WIAMIS*, July 2013.

[159] N. López, Y. Grenier, G. Richard, and I. Bourmeyster. Low variance blind estimation of the reverberation time. In *13th International Workshop on Acoustic Signal Enhancement (IWAENC 2012)*, Aachen, Germany, Sept. 2012.

[160] N. López, M. Maazaoui, Y. Grenier, G. Richard, and I. Bourmeyster. Does dereverberation help multichannel blind source separation? a study case. In *European Signal Processing Conference (EUSIPCO)*, Marrakech, Maroc, Sept. 2013.

[161] M. Maazaoui, Y. Grenier, and K. Abed-Meraim. Frequency domain blind source separation for robot audition using a parameterized sparsity criterion. In *The European Signal Processing Conference (EUSIPCO-2011)*, pages 1869–1873, Barcelone, Espagne, Sept. 2011.

[162] M. Maazaoui, Y. Grenier, and K. Abed-Meraim. Blind source separation for robot audition using fixed beamforming with hrtfs. In *12th Annual Conference of the International Speech Communication Association (Interspeech-2011)*, Florence, Italie, Sept. 2011.

[163] M. Maazaoui, K. Abed-Meraim, and Y. Grenier. Adaptive blind source separation with hrtfs beamforming preprocessing and varying number of sources. In *The seventh IEEE Sensor Array and Multichannel Signal Processing Workshop*, New Jersey, USA, June 2012.

[164] M. Maazaoui, Y. Grenier, and K. Abed-Meraim. From binaural to multichannel blind source separation using fixed beamforming with hrtfs. In *The 19th International Conference on Systems, Signals and Image Processing, IWSSIP 2012*, Vienne, Autriche, Apr. 2012.

[165] S. Marchand, R. Badeau, C. Barras, L. Daudet, D. Fourer, L. Girin, S. Gorlow, A. Liutkus, J. Pinel, G. Richard, N. Sturmel, and S. Zhang. Dream: a novel system for joint source separation and multi-track coding. In *133rd AES Convention*, San Francisco, USA, Oct. 2012.

[166] A. Martelloni, D. A. Mauro, and A. Mancuso. Further evidences of the contribution of the ear canal to directional hearing: design of a compensation filter. In *ICA*, volume 19, page 5, Montreal (Canada), June 2013.

[167] A. Masurelle, S. Essid, and G. Richard. Multimodal classification of dance movements using body joint trajectories and step sounds. In *International Workshop on Image and Audio Analysis for Multimedia Interactive Services WIAMIS*, Paris, France, 2013.

[168] B. Mathieu, S. Essid, T. Fillon, J. Prado, and G. Richard. Yaafe, an easy to use and efficient audio feature extraction software. In *ISMIR*, Utrecht, Pays-bas, Aug. 2010.

[169] D. A. Mauro. Audio convolution on gpus: a follow-up. In *AIA-DAGA*, page 4, Meran (Italy), Mar. 2013.

[170] D. A. Mauro, R. Mekuria, and M. Sanna. Binaural spatialization for 3d immersive audio communication in a virtual world. In *AudioMostly*, page 8, Pitea (Sweden), Sept. 2013.

[171] M. Moussallam, T. Fillon, G. Richard, and L. Daudet. How sparsely can a signal be approximated while keeping its class identity? In *MML10 workshop, satellite to ACM MM 2010*, firenze, Italy, Oct. 2010.

[172] M. Moussallam, P. Leveau, and S.-M. Aziz Sbaï. Sound enhancement using sparse approximation with speclets. In *ICASSP*, pages 221–224, Dallas - US, Mar. 2010.

[173] M. Moussallam, G. Richard, and L. Daudet. How sparsely can a signal be approximated while keeping its class identity? In *ACM 2010 : Workshop MML*, Florence, Italie, Nov. 2010.

[174] M. Moussallam, L. Daudet, and G. Richard. Audio signal representations for factorization in the sparse domain. In *ICASSP*, pages 513–516, Prague, Czech, May 2011.

[175] M. Moussallam, L. Daudet, and G. Richard. Random time-frequency subdictionary design for sparse representation with greedy algorithms. In *ICASSP*, pages 3577–3580, Kyoto, Japon, Mar. 2012.

[176] M. Moussallam, G. Richard, and L. Daudet. Audio source separation informed by redundancy with greedy multiscale decompositions. In *European Signal Processing Conference*, pages 2644–2648, Bucarest, Roumanie, Aug. 2012.

[177] M. Moussallam, A. Gramfort, L. Daudet, and G. Richard. Débruitage aveugle par décompositions parcimonieuses et aléatoires. In *GRETSI*, Brest, France, Sept. 2013.

[178] L. Oudre, Y. Grenier, and C. Févotte. Template-based chord recognition : influence of the chord types. In *International Symposium on Music Information Retrieval (ISMIR)*, pages 153–158, Kobe, Japan, Oct. 2009.

[179] L. Oudre, Y. Grenier, and C. Févotte. Chord recognition using measures of fit, chord templates and filtering methods. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 9–12, New York, USA, Oct. 2009.

[180] L. Oudre, C. Févotte, and Y. Grenier. Probabilistic framework for template-based chord recognition. In *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pages 183–187, St Malo, France, Oct. 2010.

[181] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures. with application to blind audio source separation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, Taipei, Taiwan, 2009.

[182] A. Ozerov and W. B. Kleijn. Optimal parameter estimation for model-based quantization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, 2009.

[183] A. Ozerov, C. Févotte, and M. Charbit. Factorial scaled hidden markov model for polyphonic audio representation and source separation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'09)*, Mohonk, NY, Oct. 2009.

[184] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard. Informed source separation: source coding meets source separation. In *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 257–260, New Paltz, New York, USA, Oct. 2011. IEEE.

[185] I. Potamitis and A. Ozerov. Single channel source separation using static and dynamic features in the power domain. In *EUSIPCO, 16th European Signal Processing Conference*, Laussane, Switzerland, Aug. 2008.

[186] G. Presti and D. A. Mauro. Continuous brightness estimation (cobe): Implementation and its possible applications. In *CMMR*, page 8, Marseille (France), Oct. 2013.

[187] M. Ramona and G. Richard. Segmentation parole/musique par machines à vecteurs de support. In *Journées d'Etudes sur la Parole (JEP'08)*, Avignon, France, June 2008.

[188] M. Ramona and G. Richard. Comparison of different strategies for a svm-based audio segmentation. In *European Signal Processing Conference (EUSIPCO)*, Glasgow, UK, Sept. 2009.

[189] M. Ramona, G. Richard, and B. David. Vocal detection in music with support vector machines. In *ICASSP'08*, Las Vegas, USA, Apr. 2008.

[190] E. Ravelli, G. Richard, and L. Daudet. Matching pursuit in adaptive dictionaries for scalable audio coding. In *EUSCIPCO*, Lausanne, Suisse, Sept. 2008.

[191] E. Ravelli, G. Richard, and L. Daudet. Fast mir in a sparse transform domain. In *ISMIR*, Philadelphie, USA, Sept. 2008.

[192] H. Ricateau, B. Elie, M. Curtit, J. ch. Valière, F. Gautier, and B. DAVID. Analysis of dead tones of classical guitars. In *Acoustics 2012*, Nantes, France, Apr. 2012.

[193] F. Rigaud, B. David, and L. Daudet. A parametric model of piano tuning. In *Proc. of the 14th Conf. on Digital Audio Effects (DAFx-11)*, pages 393–399, Paris, France, Sept. 2011.

[194] F. Rigaud, B. David, and L. Daudet. Piano sound analysis using non-negative matrix factorization with inharmonicity constraint. In *European Signal Processing Conference*, pages 2462–2466, Bucharest, Romania, Aug. 2012.

[195] F. Rigaud, A. Dremeau, B. David, and L. Daudet. A probabilistic line spectrum model for musical instrument sounds and its application to piano tuning estimation. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz NY, USA, Oct. 2013.

[196] F. Rigaud, A. Falaize, B. David, and L. Daudet. Does inharmonicity improve an nmf-based piano transcription model? In *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pages 11–15, Vancouver, Canada, May 2013. IEEE.

[197] M. Robine, M. Lagrange, and P. Hanna. Meter class profile for music similarity and retrieval. In *Proceedings of the International Conference on Music Information Retrie- val (ISMIR)*, Sept. 2009.

[198] T. Rocher, M. Robine, P. Hanna, and L. Oudre. Concurrent estimation of chords and keys from audio. In *International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands, Nov. 2010.

[199] N. Seichepine, S. Essid, C. Févotte, and O. Cappé. Soft nonnegative matrix co-factorization with application to multimodal speaker diarization. In *ICASSP*, Vancouver, May 2013.

[200] N. Seichepine, S. Essid, C. Févotte, and O. Cappé. Co-factorisation douce en matrices non-négatives. application au regroupement multimodal de locuteurs. In *GRETSI*, Brest, France, Sept. 2013.

[201] N. Sturmel, A. Liutkus, J. Pinel, L. Girin, S. Marchand, G. Richard, R. Badeau, and L. Daudet. Linear mixing models for active listening of music productions in realistic studio conditions. In *132nd AES Convention*, Budapest, Hongrie, Apr. 2012.

[202] V. Y. F. Tan and C. Févotte. Automatic relevance determination in nonnegative matrix factorization.

In *Workshop on Signal Processing with Adaptative Sparse Structured Representations (SPARS'09)*, St-Malo, France, Apr. 2009.
[203] F. Vallet, G. Richard, S. Essid, and J. Carrive. Detecting artist performances in a tv show. In *Kspace PhD Jamboree*, Paris, France, July 2008.
[204] F. Vallet, S. Essid, J. Carrive, and G. Richard. Descripteurs visuels robustes pour l'identification de locuteurs dans des émissions télévisées de talk-shows. In *CORESA*, Oct. 2010.
[205] F. Vallet, S. Essid, J. Carrive, and G. Richard. Robust visual features for the multimodal identification of unregistered speakers in tv talk-shows. In *ICIP*, Sept. 2010.
[206] E. Vincent, N. Bertin, and R. Badeau. Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *ICASSP'08*, pages 109–112, Las Vegas, Nevada, USA, Apr. 2008.
[207] S. Wegener, M. Haller, J.-J. Burred, T. Sikora, S. Essid, and G. Richard. On the robustness of audio features for musical instrument classification. In *EUSCIPCO*, Lausanne, Suisse, Sept. 2008.
[208] J. Weil, J.-L. Durrieu, G. Richard, and T. Sikora. Automatic generation of lead sheets from polyphonic music signals. In *International Society for Music Information Retrieval Conference*, pages 603–608, Kobe, Japon, Oct. 2009.
[209] F. Weninger, J.-L. Durrieu, F. Eyben, G. Richard, and B. Schüller. Combining monaural source separation with long short-term memory for increased robustness in vocalist gender recognition. In *ICASSP 2011*, Prague, May 2011.
[210] W. Zaremba, K. Pawan, A. Gramfort, and M. Blaschko. Learning from m/eeg data with variable brain activation delays. In *International Conference on Information Processing in Medical Imaging (IPMI) 2013*, Asilomar, California, July 2013.

Invited Talks and Tutorial

[211] B. David and G. Richard. L'ordinateur a-t-il le sens du rythme ? In *Conférence Grand Public, Espace Pierre Gilles de Genes*, Paris, France, Dec. 2008.
[212] G. Richard. Multimedia music signal processing. In *Tutorial at ACM Multimedia Conference*, Phoenix, USA, Nov. 2011.
[213] G. Richard. Multimodal music processing. In *Dagstuhl Seminar on Multimodal music processing*, Dagstuhl, Germany, Jan. 2011.
[214] G. Richard. Indexation des signaux musicaux polyphoniques. In *Keynote aux Journées d'Informatique Musicale*, Rennes, France, May 2011.
[215] G. Richard. Audio processing research and technologies. In *International Korea university Workshop*, Seoul, Korea, Mar. 2012.
[216] G. Richard. Audio and multimedia music signal processing. In *Keynote at 13th International Workshop on Image Analysis for Multimedia Interactive Services*, Dublin, Ireland, May 2012.

Talks and Seminars

[217] R. Badeau. Analyse spectrale à haute résolution appliquée au traitement des signaux de musique. In *Seminar of SFA at TSI/AAO, Télécom ParisTech*, Paris, France, 2011.
[218] R. Badeau. Modèles probabilistes de représentations temps-fréquences. application au traitement des signaux de musique. In *Seminar at LIF/LATP, Aix-Marseille Université*, Marseille, France, May 2012.
[219] R. Badeau. High resolution spectral analysis and nonnegative decompositions applied to music signal processing. In *Seminar at C4DM, Queen Mary University of London*, London, UK, 2013.
[220] R. Badeau. Probabilistic modelling of time-frequency representations with application to music signals. In *Seminar at C4DM, Queen Mary University of London*, London, UK, 2013.
[221] R. Badeau. Probabilistic modelling of time-frequency representations with application to music signals. In *Seminar at MLG, City University London*, London, UK, 2013.
[222] R. Badeau. Probabilistic modelling of time-frequency representations with application to music signals. In *Seminar at SigProC, University of Cambridge*, Cambridge, UK, 2013.
[223] S. Essid. Classification automatique de signaux multimédia. In *Seminar at IRISA, INRIA*, Rennes, France, Nov. 2009.
[224] S. Essid. Music-to-score temporal alignment with discriminative graphical models. In *Seminar at the Music and Audio Research Laboratory, New York University (NYU)*, New York, USA, Oct. 2011.
[225] S. Essid. The 3dlife multimodal dance corpus and applications. In *Seminar at The University of Tokyo*, Tokyo, Japan, Mar. 2012.

[226] S. Essid. Audio-driven multimedia content analysis. In *Tutorial at MediaSense 2012: Summer School on Multi-modal Data Analytics*, Dublin, Ireland, May 2012.

[227] A. Gramfort. Supervised and unsupervised learning in brain imaging: from sparsity for the meg inverse problem to dictionary learning for fmri and dmri. In *Athena INRIA Team Group meeting*, Sophia-Antipolis, FR, 2012.

[228] A. Gramfort. Functional brain imaging: how to use meg and fmri to know the "where" and "when". In *Probabilistic structures of the brain*, Cergy, FR, 2012.

[229] A. Gramfort. Sparse methods for brain imaging. In *Workshop on Sparse Models and Machine Learning*, INRIA/IRISA, Rennes, FR, 2012.

[230] A. Gramfort. Decoding in source vs sensor space. In *Neurospin Decoding Symposium*, Gif sur Yvette, FR, 2013.

[231] A. Gramfort. An introduction to regularized risk minimization for predicting from neuroimaging data. In *Pattern Recognition in Neuroimaging (PRNI) conf.*, Philadelphia, USA, 2013.

[232] A. Gramfort. Analyse temps-fréquence et parcimonie pour la localisation de sources par eeg et meg. In *Société de mathématiques appliquées et industrielles (SMAI)*, Seignosse, FR, 2013.

[233] M. Moussallam. Greedy pursuits in random sequential sub-dictionaries. In *Seminar at Columbia University*, New York, USA, Oct. 2012.

[234] G. Richard. Beyond the bag-of frames approach for musical instrument recognition. In *Seminar at Aalborg University*, Aalborg, Denmark, 2009.

[235] G. Richard. Beyond the bag-of frames approach for musical instrument recognition. In *Seminar at Dublin Institute of Technology*, Dublin, Ireland, 2009.

[236] G. Richard. Beyond the bag-of frames approach for musical instrument recognition. In *Seminar at INESC, Porto*, Porto, Portugal, 2009.

[237] G. Richard. Automatic extraction of the main melody from polyphonic music signals: With application to transcription and separation. In *Seminar at Queen Mary University of London*, London, UK, 2010.

[238] G. Richard. Greedy pursuits algorithms for representing audio signals: with applications to compression, source separation and audio fingerprint. In *Seminar at ICSI, Berkeley University*, Berkeley, USA, 2012.

[239] G. Richard. Greedy pursuits algorithms for representing audio signals: with applications to compression, source separation and audio fingerprint. In *Seminar at Los Angeles University*, Los Angeles, USA, 2012.

[240] G. Richard. An overview of audio research at telecom paristech. In *Seminar at Technical University of Berlin*, Berlin, Germany, 2012.

[241] G. Richard. Some research in audio, music and multimodal signal processing. In *Seminar at Fraunhofer Institute*, Ilmenau, Germany, 2013.

1.5.2 Public Fundings

Period	Project details	Funding	Principal investigator	Annual total (k€)
2011-2014	REVERIE - Realistic and immersive 3D Environnements	Europe (IP)	S. Essid	323 k€
2011-2014	VERVE - Vanquishing fear and apathy through E-inclusion	Europe (STREP)	S. Essid	110 k€
2010-2013	3Dlife - Analysis/synthesis of 3D audiovisual content for 3D models animation, virtual humans and virtual environments creation	Europe (NoE)	G. Richard	272 k€
2008-2013	QUAERO - Automatic analysis, indexing of multimedia and multilingual documents	OSEO	G. Richard	803 k€
2009-2013	DREAM - Sound Separation, transformation and watermarking for active listening	ANR	G. Richard	128 k€
2008-2011	ROMEO - Sound capture by microphone arrays for Humanoid robots	Cap Digital	Y. Grenier	159 k€
2012-2016	ROMEO 2 - Sound scene capture for Humanoid robots	OSEO	Y. Grenier	179 k€
2008-2013	PAFI - Modular platform for music instruments	ANR	B. David	60 k€
2009-2013	ARTIS - Articulatory inversion of audiovisual speech for augmented speech	Europe	S. Maeda	159 k€
2012-2013	SPOL : Sound Processing On Line	Labex LMH 2012	G. Richard	5 k€
2011-2014	MeegaPerf - EEG Monitoring	Contract with Industry	S. Essid	131 k€
2011-2014	EMC2- Support action towards excellence in media computing and communication	Europe	G. Richard	25 k€
2013-2016	Marie Curie IOF Fellowship	Europe	G. Richard	280 k€
TOTAL				2,3 M€

1.5.3 Private Fundings

Period	Project details	Funding	Principal investigator	Annual total (k€)
2008-2009	GRANDE PAROISSE 6: sound source localisation	Contract with Industry	Y. Grenier	35 k€
2008-2011	INA - CIFRE PhD (S. Gulluni) on Audio segmentation	Contract with Industry	G. Richard	20 k€
2009	INA - Database collection	Contract with Industry	S. Essid	3 k€
2011	Egonocast - Algorithms	Contract with Industry	S. Essid	1,3 k€
2011-2014	Arkamys - CIFRE PhD on Speech dereverberation	Contract with Industry	G. Richard	48 k€
2011-2012	Audionamix - CIFRE PhD on source separation	Contract with Industry	G. Richard	16 k€
TOTAL				123 k€

1.5.4 Patents and software

Patents

- Sébastien Fenet, Yves Grenier and Gaël Richard (TSI), *Audiofingerprinting "Generation d'une signature d'un signal audio musical"*, Patent filled under N° FR 13/51752
- Antoine Liutkus, Laurent Girin, Roland Badeau and Gaël Richard, *Procédé et dispositif de représentation et de séparation/filtrage des composantes d'un signal mixé*, Patent filled under N° FR 10/58348
- Nicolas Lopez, Gaël Richard and Yves Grenier, *Procédé de suppression de la réverbération tardive*, Patent filled under N° 26875 FR.

Softwares

- Benoit Mathieu, Jacques Prado, YAAFE, (open source) software referenced under N° IDDN.FR.001.100013.000.S.P.2010.000.20000
- Benoit Mathieu, Jacques Prado, YAAFE extension, software referenced under N° IDDN.FR.001.100014.000.S.P.2010.000.20000
- Jacques Prado, Benoit Mathieu, SMARC, sotfware referenced under N° IDDN.FR.001.080018.000.S.P.2010.000.20000
- Jacques Prado, Benoit Mathieu, SMARC (Language C), sotfware referenced under N° IDDN.FR.001.080017.000.S.P.2010.000.20000

1.6 PhDs (Appendix 7)

1.6.1 Defended PhDs

[242] N. Bertin. *Les factorisations en matrices non-négatives. Approches contraintes et probabilistes, application à la transcription automatique de musique polyphonique.* PhD thesis, Télécom ParisTech, Oct. 2009.
[243] M. Betser. *Modélisation sinusoïdale et applications à l'indexation sonore*. PhD thesis, Telecom ParisTech, June 2008.
[244] J.-L. Durrieu. *Transcription et Séparation automatique de la mélodie principale dans les signaux de musique polyphoniques*. PhD thesis, Télécom ParisTech, May 2010.
[245] B. Elie. *Caractérisation vibratoire et acoustique des instruments à cordes - Application à l'aide à la facture instrumentale*. PhD thesis, LAUM, Nov. 2012.
[246] V. Emiya. *Transcription automatique de la musique de piano*. PhD thesis, TELECOM ParisTech, Oct. 2008.
[247] S. Fontana. *Déconvolution et applications à la technologie binaurale*. PhD thesis, TELECOM ParisTech, July 2008.
[248] B. Fuentes. *L'analyse probabiliste en composantes latentes et ses adaptations aux signaux musicaux. Application à la transcription automatique de musique et à la séparation de sources*. PhD thesis, Télécom ParisTech, Mar. 2013.
[249] S. Gulluni. *Un système interactif pour l'analyse des musiques électroacoustiques*. PhD thesis, Télécom ParisTech, Dec. 2011.
[250] R. Hennequin. *Décomposition de spectrogrammes musicaux informée par des modèles de synthèse spectrale : modélisation des variations temporelles dans les objets musicaux.* PhD thesis, Télécom ParisTech, Nov. 2011.
[251] C. Joder. *Alignement temporel musique-sur-partition par modèles graphiques discriminatifs*. PhD thesis, Telecom ParisTech, Sept. 2011.
[252] A. Liutkus. *Processus gaussiens pour la séparation de sources et le codage informé*. PhD thesis, Télécom ParisTech, Nov. 2012.
[253] M. Maazaoui. *Séparation de sources pour l'audition des robots*. PhD thesis, Télécom ParisTech, May 2012.
[254] M. Moussallam. *Représentations Redondantes et Hiérarchiques pour l'Archivage et la Compression de Scènes Sonores*. PhD thesis, Télécom ParisTech, Dec. 2012.
[255] V. S. Nguyen. *Etude de caractéristiques de la langue vietnamienne en vue de sa synthèse et de sa connaissance automatique. Aspects statiques et dynamiques*. PhD thesis, Télécom ParisTech, Dec. 2009.
[256] L. Oudre. *Reconnaissance d'accords à partir de signaux audio par l'utilisation de gabarits théoriques*. PhD thesis, TELECOM ParisTech, Nov. 2010.
[257] M. Ramona. *classification automatique de flux radiophoniques par machines à vecteurs de support*. PhD thesis, Télécom ParisTech, June 2010.
[258] F. Vallet. *Structuration automatique de talk shows télévisés*. PhD thesis, Télécom ParisTech, Sept. 2011.

1.6.2 Ongoing PhDs

[259] H. Bai. *Analyse automatique et synthese dynamique de secenes 3D audio*. PhD thesis, Télécom ParisTech.
[260] A.-C. Conneau. *Identification automatique et dynamique de l'état émotionnel par analyse de signaux biologiques hétérogènes*. PhD thesis, Télécom ParisTech.
[261] S. Fenet. *Identification audio par le contenu*. PhD thesis, Télécom ParisTech.
[262] R. Foucard. *Fusion multi-niveaux pour la recherche par similarité musicale*. PhD thesis, Télécom ParisTech.
[263] C. Fox. *Reduction de bruit acoustique en environnement automobile*. PhD thesis, Télécom ParisTech.
[264] X. Jaureguiberry. *Fusion et optimisation de modèles pour la séparation de sources audio*. PhD thesis, Télécom ParisTech.
[265] N. Lopez. *Méthodes partimonieuses pour la déréverbération des signaux audio*. PhD thesis, Télécom ParisTech.

[266] A. Masurelle. *Analyse automatique de scènes multimodales par approches discriminatives*. PhD thesis, Télécom ParisTech.

[267] F. Rigaud. *Apprentissage de modèles génératifs pour un instrument de musique sur des données enregistrées*. PhD thesis, Télécom ParisTech.

[268] N. Seichepine. *Factorisations multimodales pour la structuration non-supervisée des documents audiovisuels*. PhD thesis, Télécom ParisTech.