IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2013

Project-Team PANAMA

Parcimonie et Nouveaux Algorithmes pour le Signal et la Modélisation Audio

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Language, Speech and Audio

Table of contents

Project-Team PANAMA

Keywords: Audio, Signal Processing, Machine Learning, Inverse Problem, Sparse Representations, Source Separation, Music

Creation of the Project-Team: 2013 January 01.

1. Members

Research Scientists

Rémi Gribonval [Team leader, Inria, Senior Researcher, HdR]
Nancy Bertin [CNRS, Researcher]
Frédéric Bimbot [CNRS, Senior Researcher, HdR]

Faculty Member

Laurent Albera [On partial leave from Univ. Rennes I, Associate Professor, HdR]

Engineers

Laurence Catanese [Inria, QUAERO OSEO ANVAR, until May 2013]
Jules Espiau de Lamaestre [Inria, MAIA - Musiciens Artistes Interprètes Associés]
Romain Lebarbenchon [CNRS, from Oct 2013]
Dimitri Moreau [CNRS, until Sep 2013]
Jérémy Paret [Inria, MAIA - Musiciens Artistes Interprètes Associés, until Dec 2013]
Laurent Simon [Inria, I3DMusics projet EUROSTARS, until Oct 2013]

PhD Students

Alexis Benichoux [Univ. Rennes I, until Nov 2013]
Anthony Bourrier [Technicolor R&I France, CIFRE]
Corentin Guichaoua [Univ. Rennes I]
Srdan Kitic [Inria, FP7 ERC PLEASE project]
Luc Le Magoarou [Inria, FP7 ERC PLEASE project, from Oct 2013]
Gabriel Sargent [Inria, OSEO Anvar QUAERO Project, until Apr 2013]
Nathan Souviraà-Labastie [Univ. Rennes I, Regional grant & MAIA]
Stefan Ziegler [CNRS & Regional grant]

Post-Doctoral Fellows

Cagdas Bilen [Inria, FP7 ERC PLEASE project]
Pierre Machart [Inria, FP7 ERC PLEASE project, from Jan 2013]
Anaik Olivero [Inria, ANR INSTITUT CARNOT ABDT project, from Sep 2013]
Stanislaw Raczynski [Inria, until Mar 2013]
Joachim Thiemann [CNRS, until Jun 2013]

Administrative Assistant

Stéphanie Lemaile [Inria]

2. Overall Objectives

2.1. Overall positioning

At the interface between audio modeling and mathematical signal processing, the global objective of PANAMA is to develop mathematically founded and algorithmically efficient techniques to model, acquire and process high-dimensional signals, with a strong emphasis on acoustic data.

Applications fuel the proposed mathematical and statistical frameworks with practical scenarii, and the developed algorithms are extensively tested on targeted applications. PANAMA's methodology relies on a closed loop between theoretical investigations, algorithmic development and empirical studies.

2.2. Scientific foundations

The scientific foundations of PANAMA are focused on sparse representations and probabilistic modeling, and its scientific scope is extended in three major directions:

- The extension of the sparse representation paradigm towards that of "sparse modeling", with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning.
- A focus on sophisticated probabilistic models and advanced statistical methods to account for complex dependencies between multi-layered variables (such as in audio-visual streams, musical contents, biomedical data ...).
- The investigation of graph-based representations, processing and transforms, with the goal to describe, model and infer underlying structures within content streams or data sets.

2.3. Applications

The main industrial sectors in relation with the topics of the PANAMA research group are the telecommunication sector, the Internet and multimedia sector, the musical and audiovisual production sector and, marginally, the sector of education and entertainment. Source separation is one of PANAMA's major applicative focus generating increasing industrial transfers. The models, methods and algorithms developed in the team have many potential applications beyond audio processing and modeling – the central theme of the PANAMA project-team – in particular to biomedical signals. Such applications are primarily investigated in partnership with research groups with the relevant expertise (within or outside Inria).

On a regular basis, PANAMA is involved in bilateral or multilateral partnerships, within the framework of consortia, networks, thematic groups, national and European research projects, as well as industrial contracts with various local companies.

2.4. Highlights of the Year

R. Gribonval was elevated to the grade of IEEE Fellow *for contributions to the theory and applications of sparse signal processing*.

Frédéric Bimbot was General Chairman of the Interspeech 2013 Conference in Lyon which gathered around 1400 participants.

The IEEE 2012 SPS Young Author Best Paper Award has been awarded to Ngoc Duong [4], former Ph.D. student in the METISS team.

3. Research Program

3.1. Axis 1: sparse models and representations

3.1.1. Efficient sparse models and dictionary design for large-scale data

Sparse models are at the core of many research domains where the large amount and high-dimensionality of digital data requires concise data descriptions for efficient information processing. Recent breakthroughs have demonstrated the ability of these models to provide concise descriptions of complex data collections, together with algorithms of provable performance and bounded complexity.

A crucial prerequisite for the success of today's methods is the knowledge of a "dictionary" characterizing how to concisely describe the data of interest. Choosing a dictionary is currently something of an "art", relying on expert knowledge and heuristics.

Pre-chosen dictionaries such as wavelets, curvelets or Gabor dictionaries, are based upon stylized signal models and benefit from fast transform algorithms, but they fail to fully describe the content of natural signals and their variability. They do not address the huge diversity underlying modern data much beyond time series and images: data defined on graphs (social networks, internet routing, brain connectivity), vector valued data (diffusion tensor imaging of the brain), multichannel or multi-stream data (audiovisual streams, surveillance networks, multimodal biomedical monitoring).

The alternative to a pre-chosen dictionary is a trained dictionary learned from signal instances. While such representations exhibit good performance on small-scale problems, they are currently limited to low dimensional signal processing due to the necessary training data, memory requirements and computational complexity. Whether designed or learned from a training corpus, dictionary-based sparse models and the associated methodology fail to scale up to the volume and resolution of modern digital data, for they intrinsically involve difficult linear inverse problems. To overcome this bottleneck, a new generation of efficient sparse models is needed, beyond dictionaries, which will encompass the ability to provide sparse and structured data representations as well as computational efficiency. For example, while dictionaries describe low-dimensional signal models in terms of their "synthesis" using few elementary building blocks called atoms, in "analysis" alternatives the low-dimensional structure of the signal is rather "carved out" by a set of equations satisfied by the signal. Linear as well as nonlinear models can be envisioned.

3.1.2. Compressive Learning

A flagship emerging application of sparsity is the paradigm of compressive sensing, which exploits sparse models at the analog and digital levels for the acquisition, compression and transmission of data using limited resources (fewer/less expensive sensors, limited energy consumption and transmission bandwidth, etc.). Besides sparsity, a key pillar of compressive sensing is the use of random low-dimensional projections. Through compressive sensing, random projections have shown their potential to allow drastic dimension reduction with controlled information loss, provided that the projected signal vector admits a sparse representation in some transformed domain. A related scientific domain, where sparsity has been recognized as a key enabling factor, is Machine Learning, where the overall goal is to design statistically founded principles and efficient algorithms in order to infer general properties of large data collections through the observation of a limited number of representative examples. Marrying sparsity and random low-dimensional projections with machine learning shall allow the development of techniques able to efficiently capture and process the information content of large data collections. The expected outcome is a dramatic increase of the impact of sparse models in machine learning, as well as an integrated framework from the signal level (signals and their acquisition) to the semantic level (information and its manipulation), and applications to data sizes and volumes of collections that cannot be handled by current technologies.

3.2. Axis 2: robust acoustic scene analysis

3.2.1. Compressive acquisition and processing of acoustic scenes

Acoustic imaging and scene analysis involve acquiring the information content from acoustic fields with a limited number of acoustic sensors. A full 3D+t field at CD quality and Nyquist spatial sampling represents roughly 10^6 microphones/m^3. Dealing with such high-dimensional data requires to drastically reduce the data flow by positioning appropriate sensors, and selecting from all spatial locations the few spots where acoustic sources are active. The main goal is to develop a theoretical and practical understanding of the conditions under which compressive acoustic sensing is both feasible and robust to inaccurate modeling, noisy measures, and partially failing or uncalibrated sensing devices, in various acoustic sensing scenarii. This requires the development of adequate algorithmic tools, numerical simulations, and experimental data in simple settings where hardware prototypes can be implemented.

3.2.2. Robust audio source separation

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the combination of (i) complex source characteristics, (ii) sophisticated underlying mixing model and (iii) adverse recording environments. Moreover, as opposed to the “academic” blind source separation task, most applicative contexts and new interaction paradigms offer a variety of situations in which prior knowledge and adequate interfaces enable the design and the use of informed and/or manually assisted source separation methods.

The former METISS team has developed a generic and flexible probabilistic audio source separation framework that has the ability to combine various acoustic models such as spatial and spectral source models. A first objective is to instantiate and validate specific instances of this framework targeted to real-world industrial applications, such as 5.1 movie re-mastering, interactive music soloist control and outdoor speech enhancement. Extensions of the framework are needed to achieve real-time online processing, and advanced constraints or probabilistic priors for the sources at hand will be designed, while paying attention to computational scalability issues.

In parallel to these efforts, expected progress in sparse modeling for inverse problems shall bring new approaches to source separation and modeling, as well as to source localization, which is often an important first step in a source separation workflow. In particular, a research avenue consists in investigating physically motivated, lower-level source models, notably through sparse analysis of sound waves. This should be complementary with the modeling of non-point sources and sensors, and a widening of the notion of “source localization” to the case of extended sources (i.e., considering problems such as the identification of the directivity of the source as well as its spatial position), with a focus on boundary conditions identification. A general perspective is to investigate the relations between the physical structure of the source and the particular structures that can be discovered or enforced in the representations and models used for characterization, localization and separation.

3.3. Axis 3: large-scale audio content processing and self-organization

3.3.1. Motif discovery in audio data

Facing the ever-growing quantity of multimedia content, the topic of motif discovery and mining has become an emerging trend in multimedia data processing with the ultimate goal of developing weakly supervised paradigms for content-based analysis and indexing. In this context, speech, audio and music content, offers a particularly relevant information stream from which meaningful information can be extracted to create some form of “audio icons” (key-sounds, jingles, recurrent locutions, musical choruses, etc ...) without resorting to comprehensive inventories of expected patterns.

This challenge raises several fundamental questions that will be among our core preoccupations over the next few years. The first question is the deployment of motif discovery on a large scale, a task that requires extending audio motif discovery approaches to incorporate efficient time series pattern matching methods (fingerprinting, similarity search indexing algorithms, stochastic modeling, etc.). The second question is that of the use and interpretation of the motifs discovered. Linking motif discovery and symbolic learning techniques, exploiting motif discovery in machine learning are key research directions to enable the interpretation of recurring motifs.

On the application side, several use cases can be envisioned which will benefit from motif discovery deployed on a large scale. For example, in spoken content, word-like repeating fragments can be used for several spoken document-processing tasks such as language-independent topic segmentation or summarization. Recurring motifs can also be used for audio summarization of audio content. More fundamentally, motif discovery paves the way for a shift from supervised learning approaches for content description to unsupervised paradigms where concepts emerge from the data.

3.3.2. Structure modeling and inference in audio and musical contents

Structuring information is a key step for the efficient description and learning of all types of contents, and in particular audio and musical contents. Indeed, structure modeling and inference can be understood as the task of detecting dependencies (and thus establishing relationships) between different fragments, parts or sections of information content.

A stake of structure modeling is to enable more robust descriptions of the properties of the content and better model generalization abilities that can be inferred from a particular content, for instance via cache models, trigger models or more general graphical models designed to render the information gained from structural inference. Moreover, the structure itself can become a robust descriptor of the content, which is likely to be more resistant than surface information to a number of operations such as transmission, transduction, copyright infringement or illegal use.

In this context, information theory concepts will be investigated to provide criteria and paradigms for detecting and modeling structural properties of audio contents, covering potentially a wide range of application domains in speech content mining, music modeling or audio scene monitoring.

4. Application Domains

4.1. Acoustic scene capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

4.2. Audio signal separation in reverberant environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

4.3. Multimedia indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

4.4. Brain source imaging

Epilepsies constitute a common neurological disorder that affects about 1% of the world population. As the epileptic seizure is a dynamic phenomenon, imaging techniques showing static images of the brain (MRI, PET scan) are frequently not the best tools to identify the brain area of interest. Electroencephalography (EEG) is the technique most indicated to capture transient events directly related to the underlying epileptic pathology (like interictal spikes, in particular). EEG convey essential information regarding brain (patho-)physiological activity. In addition, recording techniques of surface signals have the major advantage of being noninvasive. For this reason, an increased use in the context of epilepsy surgery is most wanted. However, to reach this objective, we have to solve an electromagnetic inverse problem, that is to say to estimate the current generators underlying noisy EEG data. Theoretically, a specific electromagnetic field pattern may be generated by an infinite number of current distributions. The considered inverse problem, called "brain source imaging problem", is then said to be ill-posed.

5. Software and Platforms

5.1. MPTK: the Matching Pursuit Toolkit

Participants: Rémi Gribonval [contact person], Jules Espiau de Lamaestre.

The Matching Pursuit ToolKit (MPTK) is a fast and flexible implementation of the Matching Pursuit algorithm for sparse decomposition of monophonic as well as multichannel (audio) signals. MPTK is written in C++ and runs on Windows, MacOS and Unix platforms. It is distributed under a free software license model (GNU General Public License) and comprises a library, some standalone command line utilities and scripts to plot the results under Matlab. This software has been registered at the APP (Agence de Protection des Programmes).

http://mptk.gforge.inria.fr, http://mptk.irisa.fr

5.2. FASST: a Flexible Audio Source Separation Toolbox

Participants: Nancy Bertin, Frédéric Bimbot.

Emmanuel Vincent [contact person]

FASST is a Flexible Audio Source Separation Toolbox, designed to speed up the conception and automate the implementation of new model-based audio source separation algorithms.

FASST is currently being developed jointly with the PAROLE team in Nancy and the TEXMEX team in Rennes through an Inria funded ADT (Action de Développement Technologique). The first implementation is in Matlab. http://bass-db.gforge.inria.fr/fasst/

5.3. NACHOS: Nearfield Acoustic HOlography with Sparse regularization

Participants: Nancy Bertin [contact person], Rémi Gribonval.

The software and associated database were developed within the ANR ECHANGE project, with the participation of Gilles Chardon, Laurent Daudet, François Ollivier and Antoine Peillot.

NACHOS (Nearfield Acoustic HOlography with Sparse regularization) is a downloadable companion software for the journal paper [79], distributed to comply with the "reproducible research" principle. It performs the reconstruction of operational deflection shapes of a vibrating structure, from acoustic measurements of the generated sound field. The software consists in Matlab source code, and automatically downloads the needed database. It allows to reproduce all results and figures of the paper, and to experiment some additional settings. It is distributed under GPL 3.0 license. Inter Deposit Digital Numbers: IDDN.FR.001.420023.000.S.P.2013.000.31235 (NACHOSDB) % IDDN.FR.001.420023.000.S.P.2013.000.31235 (NACHOS).

http://echange.inria.fr/nah.

6. New Results

6.1. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, sample complexity

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the ERC project PLEASE (projections, Learning and Sparsity for Efficient Data Processing, see section 8.2.1).

6.1.1. A new framework for sparse representations: analysis sparse models

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen.

Main collaboration: Mike Davies, Mehrdad Yaghoobi (Univ. Edinburgh), Michael Elad (The Technion).

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. Within the SMALL FET-Open project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an ℓ^1-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented.

These results have been published in conferences and in a journal paper [19]. Other algorithms based on iterative cosparse projections [83] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements [100].

Successful applications of the cosparse approach to sound source localization, audio declipping and brain imaging have been developed this year. In particular, we compared the performance of several cosparse recovery algorithms in the context of sound source localization [39] and showed its efficiency in situations where usual methods fail [60]. It was also shown to be applicable to the hard declipping problem [61]. Application to EEG brain imaging was also investigated and a paper was submitted to ICASSP'14 (see below).

6.1.2. Theoretical results on sparse representations

Participants: Rémi Gribonval, Anthony Bourrier, Pierre Machart.

Main collaboration: Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)), Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (Inria - CEMAGREF, Rennes)), Mehrdad Yaghoobi, Mike Davies (University of Edinburgh), Patrick Perez (Technicolor R&I France), Tomer Peleg (The Technion)

Sparse recovery conditions for Orthogonal Least Squares : We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. We extended Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ^1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis showed that none of the considered algorithms is uniformly better than the other [21]. More recently, we obtained simpler coherence-based conditions [18] and pursued the analysis of unrecoverable subsets [43].

Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems: The primary challenge in linear inverse problems is to design stable and robust "decoders" to reconstruct high-dimensional vectors from a low-dimensional observation through a linear operator. Sparsity, low-rank, and related assumptions are typically exploited to design decoders which performance is then bounded based on some measure of deviation from the idealized model, typically using a norm. We characterized the fundamental performance limits that can be expected from an ideal decoder given a general model, ie, a general subset of "simple" vectors of interest. First, we extended the so-called notion of instance optimality of a decoder to settings where one only wishes to reconstruct some part of the original high dimensional vector from a low-dimensional observation. This covers practical settings such as medical imaging of a region of interest, or audio source separation when one is only interested in estimating the contribution of a specific instrument to a musical recording. We defined instance optimality relatively to a model much beyond the traditional framework of sparse recovery, and characterized the existence of an instance optimal decoder in terms of joint properties of the model and the considered linear operator [42], [33]. Noiseless and noise-robust settings were both considered [56]. We showed somewhat surprisingly that the existence of noise-aware instance optimal decoders for all noise levels implies the existence of a noise-blind decoder. A consequence of our results is that for models that are rich enough to contain an orthonormal basis, the existence of an L2/L2 instance optimal decoder is only possible when the linear operator is not substantially dimension-reducing. This covers well-known cases (sparse vectors, low-rank matrices) as well as a number of seemingly new situations (structured sparsity and sparse inverse covariance matrices for instance). We exhibit an operator-dependent norm which, under a model-specific generalization of the Restricted Isometry Property (RIP), always yields a feasible instance optimality and implies instance optimality with certain familiar atomic norms such as the ℓ^1 norm.

Connections between sparse approximation and Bayesian estimation: Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an ℓ^1 penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

In 2011 we obtained a result [85] highlighting the fact that, while this is *one* possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty $\phi(x)$ need not be interpreted as assuming a prior $C \cdot \exp(-\phi(x))$ and using the MAP estimator. In particular, we showed that for *any* prior P_X, the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty $\phi(x)$, which can be interpreted as the MAP estimator with the prior $C \cdot \exp(-\phi(x))$. Vice-versa, for *certain* penalties $\phi(x)$, the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior P_X. In general $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$. This year, we extended this result to general inverse problems [30], [58], [47].

6.1.3. Algorithmic and theoretical results on dictionary learning

Participants: Rémi Gribonval, Nancy Bertin, Cagdas Bilen, Srdan Kitic.

Main collaboration: Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleinsteuber, Matthias Seibert (TU-Munich), Mehrdad Yaghoobi, Mike Davies (University of Edinburgh),

Dictionary learning : An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the litterature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

Beyond our pioneering work [86], [110] [6] on this topic, which concentrated on the noiseless case for non-overcomplete dictionaries, this year we obtained new results showing the relevance of an ℓ^1 penalized cost function for the locally stable identification of overcomplete incoherent dictionaries, in the presence of noise and outliers. Moreover, we established new sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [59].

Analysis Operator Learning for Overcomplete Cosparse Representations : Besides standard dictionary learning, we also considered learning in the context of the cosparse model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator.

We considered several approaches to learn an analysis operator from a training corpus [102]. For one of them, which uses a constrained optimization program based on ℓ^1 optimization, we derived a practical learning algorithm, based on projected subgradients, and demonstrated its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition was derived, providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions [24]. Extensions to deal with noisy data have been obtained as well [119].

In more specific situations, when prior information is available on the operator, it is also possible to express the operator on a parametric form, and learn this parameter. For instance, in the sound source localization problem, we showed that unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions. This work was submitted to the iTwist'14 workshop.

6.2. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, acoustic wavefields, audio inpainting,

6.2.1. Audio inpainting (SMALL FET-Open project)

Participants: Rémi Gribonval, Nancy Bertin, Corentin Guichaoua, Srdan Kitic.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval was the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on ℓ^1 minimization);

- addressing the case of audio declipping (*i.e.* desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work appeared as a journal paper [63].
- addressing the case of audio declipping with the competitive cosparse approach, with promising result especially when the clipping level is low. A contribution was submitted to the iTwist'14 workshop [61].

Current and future works deal with developping advanced sparse decomposition for audio inpainting, including several forms of structured sparsity (*e.g.* temporal and multichannel joint-sparsity), dictionary learning for inpainting, and several applicative scenarios (declipping, time-frequency inpainting).

6.2.2. Blind Calibration of Compressive Sensing systems

Participants: Rémi Gribonval, Cagdas Bilen.

Main collaborations: Gilles Chardon, Laurent Daudet (Institut Langevin), Gilles Puy (EPFL)

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists in unknown gains on each measure. We focus on blind calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using ℓ^1 minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, when the gains are real valued and non-negative, we showed that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrated the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observed that the success/failure of the approach seems to obey sharp phase transitions [84]. This year, we focused on extending the framework to phase-only decalibration, using techniques revolving around low-rank matrix recovery [27], [26], [34], [52], and to joint phase and gain decalibration [54].

6.2.3. Compressive Gaussian Mixture estimation

Participants: Rémi Gribonval, Anthony Bourrier.

Main collaborations: Patrick Perez (Technicolor R&I France)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this paper, we propose a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally show that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. Our algorithm provides good reconstruction and scales to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of numerous data. It also provides a privacy-preserving data analysis tool, since the sketch does not disclose information about individual datum it is based on [38], [40], [29].

6.3. Recent results on tensor decompositions

Multi-linear algebra is defined as the algebra of q-way arrays ($q > 2$), that is, the arrays whose elements are addressed by more than two indices. The first works back as far as Jordan who was interested in simultaneously diagonalizing two matrices at a time [92]. It is noteworthy that such two matrices can be interpreted as both slices of a three-way array and their joint diagonalization can be viewed as Hitchcock's polyadic decomposition [89] of the associated three-way array. Other works followed discussing rank problems related to multi-way structures and properties of multi-way arrays. However, these exercices in multilinear algebra were not linked to real data analysis but stayed within the realm of mathematics. Studying three-way data really started with Tucker's seminal work, which gave birth to the three-mode factor analysis [115]. His model is now often referred to as the Tucker3 model. At the same moment, other authors focused on a particular case of the Tucker3 model, calling it PARAFAC for PARAllel FACtor analysis [88], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [77]. In honor to Hitchcock's pionneer work, we will call it the Canonical Polyadic (CP) decomposition.

Achieving a CP decomposition has been seen first as a mere non-linear least squares problem, with a simple objective criterion. In fact, the objective is a polynomial function of many variables, where some separate. One could think that this kind of objective is easy because smooth, and even infinitely differentiable. But it turns out that things are much more complicated than they may appear to be at first glance. Nevertheless, the Alternating Least Squares (ALS) algorithm has been mostly utilized to address this minimization problem, because of its programming simplicity. This should not hide the inherently complicated theory that lies behind the optimization problem. Moreover, in most of the applications, actual tensors may not exactly satisfy the expected model, so that the problem is eventually an approximation rather than an exact decomposition. This may results in a slow convergence (or lack of convergence) of iterative algorithms such as the ALS one [94]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

6.3.1. A novel direct algorithm for CP decompositions

Participant: Laurent Albera.

Main collaborations: Sepideh Hajipour (LTSI & BiSIPL), Isabelle Merlet (LTSI, France), Mohammad Bagher Shamsollahi (BiSIPL, Iran)

Nowadays several techniques are available to solve the CP problem. They can be classified in three main groups [113]: alternating algorithms, which update only a subset of the parameters at each step; derivative-based methods, seeking for an update of all the parameters simultaneously by successive approximations; and direct procedures. The latter algorithms compute the CP decomposition by solving an alternative algebra problem of lower dimensions, but they do not provide a solution in terms of least squares contrarily to the alternating and derivative-based techniques.

We proposed a new direct algorithm to compute the CP decomposition of complex-valued multi-way arrays. The proposed algorithm is based on the Simultaneous Schur Decomposition (SSD) of particular matrices derived from the array to process. We also proposed a new Jacobi-like algorithm to calculate the SSD of several complex-valued matrices. Besides, we analysed our SSD and SSD-based CP techniques in terms of i) identifiability, ii) computational complexity and iii) estimation accuracy through a large number of scenarios including synthetic and real data in the context of CP decomposition. Computer results showed the efficiency of the proposed SSD-based CP method of dealing with some well-known difficult scenarios with swamp-like degeneracies. We also showed that the proposed method outperformed the classical CP algorithms in processing of Paatero multi-way arrays. Finally, the robustness of the proposed algorithm with respect to overfactoring was highlighted. This work was briefly presented at ICASSP'13 [31] while a journal paper for submission to IEEE Transactions on Signal Processing is in preparation.

6.3.2. CP decomposition of semi-symmetric semi-nonnegative three-way arrays

Participant: Laurent Albera.

Main collaboration (line search methods): Julie Coloigner (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France)

Main collaborations (Jacobi-like approaches): Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)

We proposed new algorithms for the CP decomposition of semi-nonnegative semi-symmetric three-way tensors. In fact, it consists in fitting the CP model for which two of the three loading matrices are nonnegative and equal. Note that such a problem can also be interpreted as a nonnegative Joint Diagonalization by Congruence (JDC) problem.

Line search and trust region strategies

We first circumvented the nonnegativity constraint by means of changes of variable into squares, leading to a (polynomial) unconstrained optimization problem. Two optimization strategies, namely line search and trust region, were then studied. Regarding the former, a global plane search scheme was considered. It consists in computing, for a given direction, one or two optimal stepsizes, depending on whether the same stepsize is used in various updating rules. Moreover, we provided a compact matrix form for the derivatives of the objective function. This allows for a direct implementation of several iterative algorithms such as Conjugate Gradient (CG), Levenberg-Marquardt (LM) and Newton-like methods, in matrix programming environments like MATLAB. Note that the computational complexity issue was taken into account in the design phase of the algorithms, and was evaluated for each algorithm, allowing to fairly compare their performance.

Thus, various scenarios have been considered, aiming at testing the influence of i) an additive noise, which can stand for modeling errors, ii) the collinearity between factors, iii) the array rank and iv) the data size. The comparisons between our CG-like, Newton-like and LM-like methods (where semi-nonnegativity and semi-symmetry constraints are exploited), and classical CP algorithms (where no constraints are considered), showed that a better CP decomposition is obtained when these a priori are exploited, especially in the context of high dimensions and high collinearity. Finally, based on our numerical analysis, the algorithms that seem to yield the best tradeoff between accuracy and complexity are our $\mathrm{CG}_{\mathrm{2steps}}$-like and LM-like algorithms.

This work was accepted for publication with minor revisions to the Elsevier Linear Algebra and Applications journal.

Next, we considered an exponential change of variable leading to a different (non-polynomial) unconstrained optimization problem. Then we proposed novel algorithms based on line search strategy with an analytic global plane search procedure requiring new matrix derivations. Their performance was evaluated in terms of estimation accuracy and computational complexity. The classical ELS-ALS [109] and LM [113] algorithms without symmetry and nonnegativity constraints, and the ACDC algorithm [120] where only the semi-symmetry constraint is imposed, were tested as reference methods. Furthermore, the performance was also compared with our algorithms based on a square change of variable. The comparison studies showed that, among these approaches, the best accuracy/complexity trade off was achieved when an exponential change of variable was used through our ELS-ALS-like algorithm.

This work was submitted to the Elsevier Signal Processing journal.

Jacobi-like approaches

The line search (despite the use of global plane search procedures) and trust region strategies may be sensitive to initialization, and generally require a multi-initialization procedure. In order to circumvent this drawback, we considered in this work Jacobi-like approaches, which are known to be less sensitive to initialization. Note that our line search and trust region approaches can then be used to refine the solution obtained by the latter.

More particularly, we formulated the high-dimensional optimization problem into several sequential polynomial subproblems using i) a square change of variables to impose nonnegativity and ii) LU matrix factorization for parameterization. The two equal nonnegative loading matrices are actually written as the Hadamard product of two equal matrices which can be factorized as the product of elementary lower and upper triangular matrices, each one depending on only one parameter.

The first approach minimizes alternatively the classical least squares objective criterion with respect to each parameter of the two equal nonnegative loading matrices and each column of the third loading matrix. This work was published in the IEEE Signal Processing Letters journal [23]. The second technique reduces the previous optimization problem to the computation of the two equal nonnegative loading matrices. The third loading matrix is algebraically derived from the latter. This requires an appropriate parameterization of the set of matrices whose inverse is nonnegative. This work was briefly presented at EUSIPCO'13 [37] while a journal paper for submission to IEEE Transactions on Signal Processing is in preparation. Numerical experiments on simulated matrices emphasize the advantages of the proposed algorithms over classical CP and JDC techniques, especially in the case of degeneracies.

6.4. Source separation and localization

Source separation, sparse representations, tensor decompositions, semi-nonnegative independent component analysis, probabilistic model, source localization

6.4.1. A general framework for audio source separation

Participants: Frédéric Bimbot, Rémi Gribonval, Nancy Bertin.

Main collaboration: E. Vincent (EPI PAROLE, Inria Nancy); N.Q.K. Duong (Technicolor R&I France)

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Recently, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [103], which generalizes a number of existing techniques including our former study on spectral GMMs [66]. This framework makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to use some prior knowledge about the source spatial covariance matrices, either via probabilistic priors [82] or via deterministic subspace constraints [91]. The latter approach was the topic of the PhD thesis of Nobutaka Ito [90]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [74]. This year, we showed that the approach provides a statistically principled solution to the permutation problem in a semi-infomed scenario where the source positions and certain room characteristics are known [15].

6.4.2. Towards real-world separation and remixing applications

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Jérémy Paret, Laurent Simon, Nathan Souviraà-Labastie, Joachim Thiemann.

Shoko Araki, Jonathan Le Roux (NTT Communication Science Laboratories, JP), E. Vincent (EPI PAROLE, Inria Nancy)

Following our founding role in the organization of the Signal Separation Evaluation Campaigns (SiSEC) [65], [101], our invited paper summarized the outcomes of the three first editions of this campaign from 2007 to 2010 [116]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the ongoing i3DMusic project and the contracts with Canon Research Centre France and MAIA Studio.

Our involvement in evaluation campaigns and source separation community was reinforced by the recording and the public release of the DEMAND (Diverse Environments Multi-channel Acoustic Noise Database) database, which provides multichannel real-world indoor and outdoor environment noise [44] under Creative Commons licence.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time [111], how to adapt EM rules for faster computation in multichannel setting [35], how to reduce artifacts [96], how our techniques compare to beamforming in realistic conditions [36], and (in the context of our collaboration with MAIA studios) how best to exploit extra information or human input. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed some improvements to the perceptually motivated metrics introduced last year [117].

6.4.3. Exploiting filter sparsity for source localization and/or separation

Participants: Alexis Benichoux, Rémi Gribonval, Frédéric Bimbot.

E. Vincent (EPI PAROLE, Inria Nancy)

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff [72], [71]. This was a central part of the Ph.D. thesis of Alexis Benichoux [12] defended this year. A journal paper including extensive experiments with real data has been submitted [69].

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. On a more theoretical side, we studied the frequency permutation ambiguity traditionnally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been published in a conference [70] and as a journal paper [14].

Finally, we considered the problem of blind sparse deconvolution, which is common in both image and signal processing. To counter-balance the ill-posedness of the problem, many approaches are based on the minimization of a cost function. A well-known issue is a tendency to converge to an undesirable trivial solution. Besides domain specific explanations (such as the nature of the spectrum of the blurring filter in image processing) a widespread intuition behind this phenomenon is related to scaling issues and the nonconvexity of the optimized cost function. We proved that a fundamental issue lies in fact in the intrinsic properties of the cost function itself: for a large family of shift-invariant cost functions promoting the sparsity of either the filter or the source, the only global minima are trivial. We completed the analysis with an empirical method to verify the existence of more useful local minima [25].

6.4.4. Semi-nonnegative independent component analysis

Participant: Laurent Albera.

Main collaborations: Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)

Independent Component Analysis (ICA) plays an important role in many areas including biomedical engineering [93], [64], [95], [118], [106], [81], speech and audio [67], [68], [78], [75], radiocommunications [80] and document restoration [114] to cite a few.

For instance in [114], the authors use ICA to restore digital document images in order to improve the text legibility. Indeed, under the statistical independence assumption, authors succeed in separating foreground text and bleed-through/show-through in palimpsest images. Furthermore, authors in [81] use ICA to solve the ambiguity in X-ray images due to multi-object overlappings. They presented a novel object decomposition technique based on multi-energy plane radiographs. This technique selectively enhances an object that is characterized by a specific chemical composition ratio of basis materials while suppressing the other overlapping objects. Besides, in the context of classification of tissues and more particularly of brain tumors [106], ICA is very effective. In fact, it allows for feature extraction from Magnetic Resonance Spectroscopy (MRS) signals, representing them as a linear combination of tissue spectra, which are as independent as possible [112]. Moreover, using the JADE algorithm [76] applied to a mixture of sound waves computed by means of the constant-Q transform (Fourier transform with log-frequency) of a temporal waveform broken up into a set of time segments, the authors of [75] describe trills as a set of note pairs described by their spectra and corresponding time envelopes. In this case, pitch and timing of each note present in the trill can be easily deduced.

All the aforementioned applications show the high efficiency of the ICA and its robustness to the presence of noise. Despite this high efficiency in resolving the proposed applicative problems, authors did not fully exploit properties enjoyed by the mixing matrix such as its nonnegativity. For instance in [81], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in [114] for the background, the overwriting and the underwriting, which correspond to the mixing coefficients, are also nonnegative. Regarding tissue classification from MRS data, each observation is a linear combination of independent spectra with positive weights representing concentrations [87]; the mixing matrix is again nonnegative.

By imposing the nonnegativity of the mixing matrix within the ICA process, we shown through computer results that the extraction quality can be improved. Exploiting the nonnegativity property of the mixing matrix during the ICA process gives rise to what we call semi-nonnegative ICA. More particularly, we performed the latter by computing a constrained joint CP decomposition of cumulant arrays of different orders [98] having the nonnegative mixing matrix as loading matrices. After merging the entries of the cumulant arrays in the same third order array, the reformulated problem follows the semi-symmetric semi-nonnegative CP model defined in section 6.3.2. Hence we use the new methods described in section 6.3.2 to perform semi-nonnegative ICA. Performance results in audio and biomedical engineering were given in the different papers cited in section 6.3.2.

6.4.5. Brain source localization

Participants: Laurent Albera, Srdan Kitic, Nancy Bertin, Rémi Gribonval.

Main collaborations: Hanna Becker (GIPSA & LTSI, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France), Pierre Comon (GIPSA, France), Christian Benar (La Timone, Marseille), Martine Gavaret (La Timone, Marseille), Gwenaël Birot (FBML, Genève), Martin Haardt (TUI, Germany)

Main collaborations: Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France)

Tensor-based approaches

The localization of several simultaneously active brain regions having low signal-to-noise ratios is a difficult task. To do this, tensor-based preprocessing can be applied, which consists in constructing a Space-Time-Frequency (STF) or Space-Time-Wave-Vector (STWV) tensor and decomposing it using the CP decomposition. We proposed a new algorithm for the accurate localization of extended sources based on the results of the tensor decomposition. Furthermore, we conducted a detailed study of the tensor-based preprocessing methods, including an analysis of their theoretical foundation, their computational complexity, and their performance for realistic simulated data in comparison to three conventional source localization algorithms, namely sLORETA [105], cortical LORETA (cLORETA) [104], and 4-ExSo-MUSIC [73]. Our objective consisted, on the one hand, in demonstrating the gain in performance that can be achieved by tensor-based preprocessing, and, on the other hand, in pointing out the limits and drawbacks of this method. Finally, we validated the STF and STWV techniques on real epileptic measurements to demonstrate their usefulness for practical applications. This work was recently submitted to the Elesevier NeuroImage journal.

From tensor to sparse models

The brain source imaging problem has been widely studied during the last decades, giving rise to an impressive number of methods using different priors. Nevertheless, a thorough study of the latter, including especially sparse and tensor-based approaches, is still missing. Consequently, we proposed i) a taxonomy of the methods based on a priori assumptions, ii) a detailed description of representative algorithms, iii) a review of identifiability results and convergence properties of different techniques, and iv) a performance comparison of the selected methods on identical data sets. Our aim was to provide a reference study in the biomedical engineering domain which may also be of interest for other areas such as wireless communications, audio source localization, and image processing where ill-posed linear inverse problems are encountered and to identify promising directions for future research in this area. A part of this work was submitted to ICASSP'14 while the whole part was submitted to IEEE Signal Processing Magazine.

A cosparsity-based approach

Cosparse modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator. We showed how to derive a reduced non-singular analysis operator describing EEG signals from Poisson's equation, Kirchhoff's law and some other physical constraints. As a result, we proposed the CoRE (Cosparse Representation of EEG signals) method to solve the classical brain source imaging problem. Computer simulations demonstrated the numerical performance of the CoRE method in comparison to a dictionary-based sparse approach. This work was submitted to ICASSP'14.

6.5. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

6.5.1. Audio motif discovery

Participants: Frédéric Bimbot, Laurence Catanese.

This work was performed in close collaboration with Guillaume Gravier from the Texmex project-team.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. The developed algorithms allows discovering and collecting occurrences of repeating patterns in the absence of prior acoustic and linguistic knowledge, or training material.

Former work extended the principles of seeded discovery to near duplicate detection and spoken document retrieval from examples [99].

In 2012, the work achieved consisted in consolidating previously obtained results with the motif discovery algorithm and making implementation choices regardless of the structure and the code, in order to minimize the computation time. This has lead to the creation of a software prototype called MODIS.

After the code has been thoroughly optimised, further optimizations to improve the system performances was to change the method used for the search of similarities between patterns. A new functionality has been added to get rid of unrelevant patterns like silence in speech. New versions of dynamic time warping have been implemented, as well as the possibility to downsample the input sequence during the process, which allows a huge gain of computation time.

The principles of the MODIS software has been documented in details [48] and demonstrated during a Show & Tell session at the Interspeech 2013 conference [41].

This work has been carried out in the context of the Quaero Project.

6.5.2. Landmark-driven speech recognition

Participant: Stefan Ziegler.

This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.

Our previous studies indicate that acoustic-phonetic approaches to ASR, while they cannot achieve state-of-the-art ASR performance by themselves, can prevent HMM-based ASR from degrading, by integrating additional knowledge into the decoding.

In our previous framework we inserted knowledge into the decoding by detecting time frames (referred to as landmarks) which estimate the presence of the active broad phonetic class. This enables the use of a modified version of the viterbi decoding that favours states that are coherent with the detected phonetic knowledge [122].

In 2012 we focused on two major issues. First, we aimed at finding new ways to model and detect phonetic landmarks. Our second focus was on the extension of our landmark detector towards a full acoustic-phonetic framework, to model speech by a variety of articulatory features.

Our new approach for the classification and detection of speech units focuses on developping landmark-models that are different from existing frame-based approaches to landmark detection [121]. In our approach, we use segmentation to model any time-variable speech unit by a fixed-dimensional observation vector. After training any desired classifier, we can estimate the presence of a desired speech unit by searching for each time frame the corresponding segment, that provides the maximum classification score.

We used this segment-based landmark-detection inside a standalone acoustic-phonetic framework that models speech as a stream of articulatory features. In this framework we first search for relevant broad phonetic landmarks, before attaching each landmark with the full set of articulatory features.

Integrating these articulatory feature streams into a standard HMM-based speech recognizer by weighted linear combination improves speech recognition up to 1.5

Additionally, we explored the possibilities of using stressed syllables as an information to guide the viterbi decoding. This work was carried under the leaderhip of Bogdan Ludusan from the team TEXMEX at IRISA [97].

6.5.3. Mobile device for the assistance of users in potentially dangerous situations

Participants: Romain Lebarbenchon, Frédéric Bimbot.

The S-Pod project is a cooperative project between industry and academia aiming at the development of mobile systems for the detection of potentially dangerous situations in the immediate environment of a user, without requiring his/her active intervention.

In this context, the PANAMA research group is involved in the design of algorithms for the analysis and monitoring of the acoustic scene around the user, yielding information which can be fused with other sources of information (physiological, contextual, etc...) in order to trigger an alarm when needed and subsequent appropriate measures.

Currently in its initial phase, work has mainly focused on functional specifications and performance requirements.

6.6. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

6.6.1. Music language modeling

Participants: Frédéric Bimbot, Dimitri Moreau, Stanislaw Raczynski.

Main collaboration: S. Fukayama (University of Tokyo, JP), E. Vincent (EPI PAROLE, Inria Nancy), Intern: A. Aras

Music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively.

We pursued our pioneering work on music language modeling, with a particular focus on the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem [108]. We also applied the log-linear interpolation paradigm to the joint modeling of melody, key and chords, which evolve according to different timelines [107]. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.

The limited availability of multi-feature symbolic music data is currently an issue which prevents the training of the developed models on sufficient amounts of data for the unsupervised probabilistic approach to significantly outperform more conventional approaches based on musicological expertise. We outlined a procedure for the semi-automated collection of large-scale multifeature music corpora by exploiting the wealth of music data available on the web (audio, MIDI, leadsheets, lyrics, etc) together with algorithms for the automatic detection and alignment of matching data. Following this work, we started collecting pointers to data and developing such algorithms.

Effort was dedicated to the investigation of structural models for improving the modeling of chord sequence. Preliminary results obtained during Anwaya Aras' intersnship show that using a matricial structure of time dependencies between successive chords improves the predictability of chord sequences as compared to a purely sequential model.

6.6.2. Music structuring

Participants: Frédéric Bimbot, Anaik Olivero, Gabriel Sargent.

Main collaboration: E. Vincent (EPI PAROLE, Inria Nancy), Intern: E. Deruty

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

Last year, our methodology for the *semiotic* annotation of music pieces has developed and concretized into a set of principles, concepts and conventions for locating the boundaries and determining metaphoric labels of music segments. The method relies on a new concept for characterizing the inner organization of music segments called the System & Contrast (S&C) model [2]. The annotation of 383 music pieces has been finalized, documented [28] and released to the MIR scientific community: http://musicdata.gforge.inria.fr/structureAnnotation.html.

For what concerns algorithmic approaches to music structure description [13], we have formulated the segmentation process as the optimization of a cost function which is composed of two terms: the first one corresponds to the characterization of structural segments by means of audio criteria; the second one relies on the regularity of the target structure with respect to a "structural pulsation period". In this context, we have compared several regularity constraints and studied the combination of audio criteria through fusion. We also considered the estimation of structural labels as a probabilistic finite-state automaton selection process : in this scope, we have proposed an auto-adaptive criterion for model selection, applied to a description of the tonal content. We also proposed a labeling method derived from the system-contrast model. We have evaluated and compared several systems for structural segmentation of music based on these approaches in the context of national and international evaluation campaigns (Quaero, MIREX).

As a follow-up to this work on music structure description, we are currently designing new models and algorithms for segmenting and labeling music into structural units. In one approach (Corentin Guichaoua's PhD), music structure is described as a hierarchical tree estimated by a grammar inference process whereas a second approach (Anaik Olivero's Post-doc) addresses music structure description as the estimation of a graph of similarity relationships.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Contract with Canon Research Centre France SAS

Participants: Joachim Thiemann, Nancy Bertin, Frédéric Bimbot.

Duration: 1.5 years (2012–2013).

Research axis: 3.2.2

Partner: Canon Research Centre France SAS

This contract aims at transfering some of the research done within METISS/PANAMA to products developed by Canon Inc. Two patents were filed [50], [51]. Final internal report was delivered in October 2013.

7.1.2. Contract with Studio MAIA

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Jérémy Paret, Nathan Souviraà-Labastie.

Duration: 3 years (2012–2014).

Research axis: 3.2.2

Partners: Studio MAIA (Musiciens Artistes Interprètes Associés), Imaging Factory

This contract aims at transfering some of the research done within PANAMA towards new services provided by MAIA Studio.

More specifically, the main objective is to adapt source separations algorithms and some other advanced signal processing techniques elaborated by PANAMA in a user-informed context.

The objective is twofold:

- partial automation of some tasks which the user previously had to accomplish manually
- improved quality of separation and processing by exploiting user inputs and controls

The resulting semi-automated separation and processing will feed an integrated software used for the professional remastering of audiovisual pieces. A first version of PANAMA tools were integrated in the software developed by Imaging Factory and delivered to MAIA in December 2013.

7.2. Bilateral Grants with Industry

7.2.1. CIFRE contract with Technicolor R&I France on Compressive Sensing for the manipulation of large multimedia databases

Participants: Rémi Gribonval, Anthony Bourrier.

Duration: 3 years (2011-2014)

Research axis: 3.1.2

Partners: Technicolor R&I France, Inria-Rennes

Funding: Technicolor R&I France, ANRT

The objective of this thesis is to explore, both numerically and theoretically, the potential of compressive sensing for the manipulate of large (audiovisual) databases. A particular objective is to propose learning techniques that can work on strongly compressed versions of a large corpus of data while maintaining the ability to infer essential characteristics of the distribution of the items in the corpus.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. OSEO: QUAERO CTC and Corpus Projects

Participants: Frédéric Bimbot, Laurence Catanese, Gabriel Sargent.

Main academic partners : IRCAM, IRIT, LIMSI, Telecom ParisTech

Duration: 2008 -december 2013

Research axis: 3.3

Description: Quaero is a European research and development program with the goal of developing multimedia and multilingual indexing and management tools for professional and general public applications (such as search engines).

Partners: Other companies involved in the consortium are: France Télécom, Exalead, Bertin Technologies, Jouve, Grass Valley GmbH, Vecsys, LTU Technologies, Siemens A.G. and Synapse Développement. Many public research institutes are also involved, including LIMSI-CNRS, Inria, IRCAM, RWTH Aachen, University of Karlsruhe, IRIT, Clips/Imag, Telecom ParisTech, INRA, as well as other public organisations such as INA, BNF, LIPN and DGA.

Funding: This program is supported by OSEO.

Coordinator: The consortium is led by Technicolor.

Contribution of PANAMA:

PANAMA is involved in two technological domains : audio processing and music information retrieval (WP6). The research activities (CTC project) are focused on improving audio and music analysis, segmentation and description algorithms in terms of efficiency, robustness and scalability. Some effort is also dedicated on corpus design, collection and annotation (Corpus Project).

PANAMA also takes part to research and corpus activities in multimodal processing (WP10), in close collaboration with the TEXMEX project-team.

8.1.2. OSEO-FUI: S-POD: "Assistance à personnes en danger potentiel"

Participants: Frédéric Bimbot, Romain Lebarbenchon.

Duration: August 2012-November 2016

Research axis: 3.2

Partners: ERYMA, CAPT/FOTON, CASSIDIAN, KAPTALIA, KERLINK, le LOUSTIC and Telecom Bretagne

Coordinator: ERYMA

Description: S-POD gathers research teams and industrial partners to that aim at setting up a framework to process and fuse audio, physiological and contextual data. The goal is to design an embedded autonomous system able to detect situations of potential danger arising in the immediate environment of a person (military, police, CIT, fire, etc.)

Contribution of PANAMA: PANAMA is in charge of R&I activities related to the qualitative and quantitative analysis of information from the acoustic environment (intensity, direction of arrival, nature of noise sounds, properties of voices, etc.) as well as to the exploitation of these analyses. The need for real-time embedded processing induces specific constraints.

8.1.3. Action de Développement Technologique: FASST

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Nathan Souviraà-Labastie.

Duration: 2 years (2012–2014).

Research axis: 3.2.2

Partners: Inria Teams Parole (Nancy) and Texmex (Rennes)

Description: This Inria ADT aims to develop a new version of our FASST audio source separation toolbox in order to facilitate its large-scale dissemination in the source separation community and in the various application communities. A specific effort will be made towards the speech processing community by developing an interface with existing speech recognition software. A beta version was internally released and tested from July 2013. The first public release is planned for January 2014.

8.2. European Initiatives

8.2.1. ERC-StG: PLEASE (Projections, Learning, and Sparsity for Efficient Data Processing)

Participants: Rémi Gribonval, Srdan Kitic, Pierre Machart, Cagdas Bilen, Luc Le Magoarou, Nancy Bertin.

Duration: January 2012 - December 2016

Research axis: 3.1

Principal investigator: Rémi Gribonval

Program: ERC Starting Grant

Project acronym: PLEASE

Project title: Projections, Learning and Sparsity for Efficient data processing

Abstract: The Please ERC is focused on the extension of the sparse representation paradigm towards that of *sparse modeling*, with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning

Web site: https://team.inria.fr/panama/projects/please/

8.2.2. Eureka-Eurostars: i3DMusic

Participant: Laurent Simon.

Duration: October 2010 - September 2013

Research axis: 3.2.2

Partners: Audionamix (FR), Sonic Emotion (CH), École Polytechnique Fédérale de Lausanne (CH), PANAMA (FR)

Program: Eureka - Eurostars

Project acronym: i3DMusic

Project title: Real-time Interative 3D Rendering of Musical Recordings

Abstract:The i3DMusic project (Real-time Interative 3D Rendering of Musical Recordings) has been setup with the SMEs Audionamix and Sonic Emotion and the academic partner EPFL to provide a system enabling real-time interactive respatialization of mono or stereo music content. This will be achieved through the combination of source separation and 3D audio rendering techniques. Metiss is responsible for the source separation work package, more precisely for designing scalable online source separation algorithms and estimating advanced spatial parameters from the available mixture.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

- Mike Davies, from May until July, Professor of Signal and Image Processing, University of Edinburgh
- Anders Hansen, from April until April, Research Fellow Royal Society, Center for Mathematical Sciences, University of Cambridge
- Dan Stowell, from March until March, Postdoctoral research assistant, Center for Digital Music, Queen Mary University of London
- Bob Sturm, from March until March, Assistant Professor, Aalborg University Copenhagen
- Boris Mailhé, from March until March, Postdoctoral research assistant, Center for Digital Music, Queen Mary University of London
- Simon Foucart, from March until March, Assistant Professor, Drexel University

8.3.2. Internships

- Anwaya Aras, from July until December, Third year undergraduate, Department of Computer Science BITS-Pilani, India.
- Emmanuel Deruty, from April to September, PhD Preparatory year, Musicology Department, Catholic University Louvain, Belgium

9. Dissemination

9.1. Scientific Animation

Rémi Gribonval is in charge of the Action "Parcimonie" within the French GDR ISIS on Signal & Image Processing

Rémi Gribonval is a member of the International Steering Committee for the LVA conference series.

Rémi Gribonval is a member of the IEEE Technical Committee on Signal Processing Theory and Methods (2012–2014), and a member of the Awards sub-committee.

Rémi Gribonval was a member of the organizing committee of the GRETSI 2013 edition.

Rémi Gribonval is a member of the program committee of the GRETSI.

Rémi Gribonval is a member of the Steering Committee of the SPARS international workshop (chairman until 2013).

Frédéric Bimbot is the Head of the "Digital Signals and Images, Robotics" in IRISA (UMR 6074).

Frédéric Bimbot has been the General Chairman of the Interspeech 2013 Conference in Lyon (1400 participants) [46].

Frédéric Bimbot is a member of the International Advisory Council of ISCA (International Speech Communication Association)

Nancy Bertin is a member of the IEEE Technical Committee on Audio and Acoustic Signal Processing (2013–2015).

Nancy Bertin and Frédéric Bimbot co-organised a two-day workshop on "localization, separation and tracking of audio sources", in the context of IRISA D5 (Signal, Image & Robotics) department's scientific animation seminars.

Laurent Albera was a scientific council member at University of Rennes 1.

Laurent Albera was co-organizer with David Brie at Paris of a one-day seminar on tensor decompositions and applications sponsored by the GDR ISIS (french research groups in signal and image processing).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : N. Bertin, "Discovery of selected topics in audio signal processing research", 9 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France

Master : N. Bertin, "Audio rendering, coding and source separation", 9 hours, M2, Université Rennes 1, France

Master : N. Bertin, "Audio indexing and classification", 9 hours, M2, Université Rennes 1, France

Master : R. Gribonval, "Sparse representations for inverse problems in signal and image processing", 10 hours, M2, Université Rennes 1, France

Master : R. Gribonval, "Signal and image representations", 8 hours, M2, Université Rennes 1, France

Master: R. Gribonval, coordination of the ARD module "Acquisition et Représensation de Données", M2, Université Rennes 1, France

Laurent Albera gives lectures in Mathematics and in Signal Processing, and he supervises end of school year projects, mainly at the university of Rennes 1:

Licence: L. Albera, "Mathematics for electronics", 18 hours, L2, Université Rennes 1, France

Licence: L. Albera, "Mathematics for electronics", 21 hours, L3, Université Rennes 1, France

Licence: L. Albera, "Mathematics", 48 hours, L3, Ecole Supérieure d'Ingénieurs de Rennes, France

Master: L. Albera, "Multimodal analysis of biomedical signals", 12 hours, M1, Ecole Supérieure d'Ingénieurs de Rennes, France

Master: L. Albera, "Evaluation of EEG signals to measure brain activity during entertainment applications", project supervision (with Technicolor R&I France), M2, Ecole Supérieure d'Ingénieurs de Rennes, France

Master : L. Albera, "Blind equalization", 4.5 hours, M2, Université Rennes 1, France

Master: L. Albera, "Brain source imaging", project supervision, M2, Université Rennes 1, France

Doctorat : L. Albera, "Brain source imaging", 2.5 hours, EUROSAE formation continue, France

Doctorat : L. Albera, "Brain source imaging", 6 hours, CominLabs, Université Rennes 1, France

Laurent Albera is responsible of the "Signal Processing" branch of the SISEA (Signal, Images, Embedded Systems and Control) Master 2 of University of Rennes 1.

9.3. Popularization

9.3.1. Journée science et musique

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Rémi Gribonval, Corentin Guichaoua, Srdan Kitic, Stéphanie Lemaile, Pierre Machart, Laurent Simon, Nathan Souviraà-Labastie, Stefan Ziegler.

PANAMA coordinated the organization of a public event called "Journée Science et Musique" (Day of Music and Science). This yearly event organized by the METISS/ PANAMA Team since 2011 aims at sharing with the wide audience the latest innovations and research projects in music. The motivation for hosting this event is to explain and promote the technology behind audio-processing that people face in their daily lives. The event is free to everyone and people have the possibility to attend talks by selected speakers or meet numerous experts that demonstrate current projects in which people can interactively participate. Edition 2013 hosted 350 visitors.

9.3.2. Radio and press

Nancy Bertin was an invited speaker in the France Inter scientific radio show "La Tête au Carré".

10. Bibliography

Major publications by the team in recent years

[1] F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, P. Perrier. , *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech), 25-29 August 2013, Lyon (France)*, International Speech Communication Association (ISCA), August 2013, over 3500 p. , http://hal.archives-ouvertes.fr/hal-00931864

[2] F. Bimbot, E. Deruty, G. Sargent, E. Vincent. , *System & Contrast : a Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces (Original Extensive Version)*, December 2012, n^{o} IRISA PI-1999, 40 p. , http://hal.inria.fr/hal-00868398

[3] G. CHARDON, L. DAUDET, A. PEILLOT, F. OLLIVIER, N. BERTIN, R. GRIBONVAL. *Nearfield Acoustic Holography using sparsity and compressive sampling principles*, in "Journal of the Acoustical Society of America", 2012, Code & data for reproducing the main figures of this paper are available at the provided urls, http://hal.inria.fr/hal-00720129

[4] N. DUONG, E. VINCENT, R. GRIBONVAL. *Under-Determined Reverberant Audio Source Separation Using a Full-Rank Spatial Covariance Model*, in "IEEE Trans. Audio, Speech and Language Processing", 2010, vol. 18, n^o 7, pp. 1830–1840

[5] R. GRIBONVAL. *Should penalized least squares regression be interpreted as Maximum A Posteriori estimation?*, in "IEEE Transactions on Signal Processing", May 2011, vol. 59, n^o 5, pp. 2405-2410 [*DOI :* 10.1109/TSP.2011.2107908], http://hal.inria.fr/inria-00486840

[6] R. GRIBONVAL, K. SCHNASS. *Dictionary Identifiability - Sparse Matrix-Factorisation via ℓ_1 minimisation*, in "IEEE Trans. Information Theory", jul 2010, vol. 56, n^o 7, pp. 3523–3539

[7] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. *Wavelets on Graphs via Spectral Graph Theory*, in "Applied and Computational Harmonic Analysis", 2010

[8] A. MUSCARIELLO, F. BIMBOT, G. GRAVIER. *Unsupervised Motif Acquisition in Speech via Seeded Discovery and Template Matching Combination*, in "IEEE Transactions on Audio, Speech and Language Processing", September 2012, vol. 20, n^o 7, pp. 2031 - 2044 [*DOI :* 10.1109/TASL.2012.2194283], http://hal.inria.fr/hal-00740978

[9] S. NAM, M. E. DAVIES, M. ELAD, R. GRIBONVAL. *The Cosparse Analysis Model and Algorithms*, in "Applied and Computational Harmonic Analysis", 2013, vol. 34, n^o 1, pp. 30–56, Preprint available on arXiv since 24 Jun 2011 [*DOI :* 10.1016/J.ACHA.2012.03.006], http://hal.inria.fr/inria-00602205

[10] A. OZEROV, E. VINCENT, F. BIMBOT. *A General Flexible Framework for the Handling of Prior Information in Audio Source Separation*, in "IEEE Transactions on Audio, Speech and Language Processing", May 2012, vol. 20, n^o 4, pp. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962

[11] E. VINCENT, N. BERTIN, R. GRIBONVAL, F. BIMBOT. *From blind to guided audio source separation*, in "IEEE Signal Processing Magazine", December 2013, http://hal.inria.fr/hal-00922378

Publications of the year

Doctoral Dissertations and Habilitation Theses

[12] A. BENICHOUX. , *Fonctions de couˆt pour l’estimation des filtres acoustiques dans les melanges reverberants*, Université Rennes 1, October 2013, http://hal.inria.fr/tel-00910326

[13] G. SARGENT. , *Estimation de la structure de morceaux de musique par analyse multi-critères et contrainte de régularité*, Université Rennes 1 and Université européenne de Bretagne, February 2013, http://hal.inria.fr/tel-00853737

Articles in International Peer-Reviewed Journals

[14] A. BENICHOUX, P. SUDHAKAR, F. BIMBOT, R. GRIBONVAL. *Well-posedness of the permutation problem in sparse filter estimation with lp minimization*, in "Applied and Computational Harmonic Analysis", November 2013, vol. 35, n^{o} 3, pp. 394-406 [*DOI :* 10.1016/J.ACHA.2012.11.001], http://hal.inria.fr/hal-00767139

[15] N. DUONG, E. VINCENT, R. GRIBONVAL. *Spatial location priors for Gaussian model based reverberant audio source separation*, in "EURASIP Journal on Advances in Signal Processing", September 2013, 149 p. [*DOI :* 10.1186/1687-6180-2013-149], http://hal.inria.fr/hal-00865125

[16] R. GIRYES, S. NAM, M. ELAD, R. GRIBONVAL, M. E. DAVIES. *Greedy-Like Algorithms for the Cosparse Analysis Model*, in "Linear Algebra and its Applications", January 2014, vol. 441, pp. 22–60, partially funded by the ERC, PLEASE project, ERC-2011-StG-277906 [*DOI :* 10.1016/J.LAA.2013.03.004], http://hal.inria.fr/hal-00716593

[17] R. GRIBONVAL, M. NIELSEN. *The restricted isometry property meets nonlinear approximation with redundant frames*, in "Journal of Approximation Theory", January 2013, vol. 165, n^{o} 1, pp. 1–19 [*DOI :* 10.1016/J.JAT.2012.09.008], http://hal.inria.fr/inria-00567801

[18] C. HERZET, C. SOUSSEN, J. IDIER, R. GRIBONVAL. *Exact recovery conditions for sparse representations with partial support information*, in "IEEE Transactions on Information Theory", November 2013, vol. 59, n^{o} 11, pp. 7509-7524 [*DOI :* 10.1109/TIT.2013.2278179], http://hal.inria.fr/hal-00907646

[19] S. NAM, M. E. DAVIES, M. ELAD, R. GRIBONVAL. *The Cosparse Analysis Model and Algorithms*, in "Applied and Computational Harmonic Analysis", 2013, vol. 34, n^{o} 1, pp. 30–56, Preprint available on arXiv since 24 Jun 2011 [*DOI :* 10.1016/J.ACHA.2012.03.006], http://hal.inria.fr/inria-00602205

[20] M. RUBEAUX, J.-C. NUNES, L. ALBERA, M. GARREAU. *Medical image registration using Edgeworth-based approximation of Mutual Information*, in "IRBM", 2014, 9 p. , http://hal.inria.fr/hal-00914411

[21] C. SOUSSEN, R. GRIBONVAL, J. IDIER, C. HERZET. *Joint k-step analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares*, in "IEEE Transactions on Information Theory", May 2013, vol. 59, n^{o} 5, pp. 3158-3174 [*DOI :* 10.1109/TIT.2013.2238606], http://hal.inria.fr/hal-00637003

[22] E. VINCENT, N. BERTIN, R. GRIBONVAL, F. BIMBOT. *From blind to guided audio source separation*, in "IEEE Signal Processing Magazine", December 2013, http://hal.inria.fr/hal-00922378

[23] L. WANG, L. ALBERA, A. KACHENOURA, H. SHU, L. SENHADJI. *Nonnegative Joint Diagonalization by Congruence Based on LU Matrix Factorization*, in "IEEE Signal Processing Letters", August 2013, vol. 20, n^{o} 8, pp. 807 - 810 [*DOI :* 10.1109/LSP.2013.2267797], http://hal.inria.fr/inserm-00841128

[24] M. YAGHOOBI, S. NAM, R. GRIBONVAL, M. E. DAVIES. *Constrained Overcomplete Analysis Operator Learning for Cosparse Signal Modelling*, in "IEEE Transactions on Signal Processing", 2013, vol. 61, n^{o} 9, pp. 2341-2355 [*DOI :* 10.1109/TSP.2013.2250968], http://hal.inria.fr/hal-00805705

International Conferences with Proceedings

[25] A. BENICHOUX, E. VINCENT, R. GRIBONVAL. *A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors*, in "ICASSP - 38th International Conference on Acoustics, Speech, and Signal Processing - 2013", Vancouver, Canada, May 2013, http://hal.inria.fr/hal-00800770

[26] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. *Blind Phase Calibration in Sparse Recovery*, in "EUSIPCO - 21st European Signal Processing Conference - 2013", Marrakech, Morocco, September 2013, http://hal.inria.fr/hal-00813421

[27] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. *Blind Sensor Calibration in Sparse Recovery Using Convex Optimization*, in "SAMPTA - 10th International Conference on Sampling Theory and Applications - 2013", Bremen, Germany, 2013, http://hal.inria.fr/hal-00813409

[28] F. BIMBOT, G. SARGENT, E. DERUTY, C. GUICHAOUA, E. VINCENT. *Semiotic Description of Music Structure: an Introduction to the Quaero/Metiss Structural Annotations*, in "AES 53rd International Conference on Semantic Audio", London, United Kingdom, January 2014, pp. P1-1, 12, http://hal.inria.fr/hal-00931859

[29] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Compressive Gaussian Mixture Estimation*, in "ICASSP - 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, 2013, pp. 6024-6028, http://hal.inria.fr/hal-00799896

[30] R. GRIBONVAL, P. MACHART. *Reconciling "priors" and "priors" without prejudice?*, in "Advances in Neural Information Processing Systems", South Lake Tahoe, United States, June 2013, The companion technical report hal-00862925 contains additional worked examples, http://hal.inria.fr/hal-00918008

[31] S. HAJIPOUR, L. ALBERA, M. SHAMSOLLAHI, I. MERLET. *Canonical polyadic decomposition of complex-valued multi-way arrays based on simultaneous Schur decomposition*, in "2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Vancouver, Canada, May 2013, pp. 4178-4182, http://hal.inria.fr/hal-00921783

[32] S. KITIĆ, L. JACQUES, N. MADHU, M. P. HOPWOOD, A. SPRIET, C. DE VLEESCHAUWER. *Consistent Iterative Hard Thresholding for signal declipping*, in "ICASSP - The 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, May 2013, http://hal.inria.fr/hal-00922504

[33] T. PELEG, R. GRIBONVAL, M. E. DAVIES. *Compressed Sensing and Best Approximation from Unions of Subspaces: Beyond Dictionaries*, in "21st European Signal Processing Conference (EUSIPCO 2013)", Marrakech, Morocco, 2013, http://hal.inria.fr/hal-00812858

[34] H. SHEN, M. KLEINSTEUBER, C. BILEN, R. GRIBONVAL. *A Conjugate Gradient Algorithm for Blind Sensor Calibration in Sparse Recovery*, in "IEEE International Workshop on Machine Learning for Signal Processing - 2013", Southampton, United Kingdom, 2013, http://hal.inria.fr/hal-00871323

[35] J. THIEMANN, E. VINCENT. *A fast EM algorithm for Gaussian model-based source separation*, in "EUSIPCO - 21st European Signal Processing Conference - 2013", Marrakech, Morocco, September 2013, http://hal.inria.fr/hal-00840366

[36] J. THIEMANN, E. VINCENT. *An experimental comparison of source separation and beamforming techniques for microphone array signal enhancement*, in "MLSP - 23rd IEEE International Workshop on Machine Learning for Signal Processing - 2013", Southampton, United Kingdom, September 2013, http://hal.inria.fr/hal-00850173

[37] L. WANG, L. ALBERA, H. SHU, L. SENHADJI. *Jacobi-like nonnegative joint diagonalization by congruence*, in "XXI European Signal Processing Conference (EUSIPCO'13)", Marrakech, Morocco, September 2013, 5 p. , http://hal.inria.fr/hal-00921807

National Conferences with Proceedings

[38] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Estimation de mélange de Gaussiennes sur données compressées*, in "24ème Colloque Gretsi", France, September 2013, 191 p. , http://hal.inria.fr/hal-00839579

[39] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *A review of cosparse signal recovery methods applied to sound source localization*, in "Le XXIVe colloque Gretsi", Brest, France, September 2013, http://hal.inria.fr/hal-00838080

Conferences without Proceedings

[40] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Compressive Gaussian Mixture Estimation*, in "Signal Processing with Adaptive Sparse Structured Representations (SPARS) 2013", Switzerland, July 2013, http://hal.inria.fr/hal-00811819

[41] L. CATANESE, N. SOUVIRAÀ-LABASTIE, B. QU, S. CAMPION, G. GRAVIER, E. VINCENT, F. BIMBOT. *MODIS: an audio motif discovery software*, in "Show & Tell - Interspeech 2013", Lyon, France, August 2013, http://hal.inria.fr/hal-00931227

[42] T. PELEG, R. GRIBONVAL, M. E. DAVIES. *Generalized Null Space and Restricted Isometry Properties*, in "Signal Processing with Adaptive Sparse Structured Representations 2013 (2013)", Lausanne, Switzerland, EPFL, 2013, http://hal.inria.fr/hal-00811673

[43] C. SOUSSEN, C. HERZET, J. IDIER, R. GRIBONVAL. *Unrecoverable subsets by OMP and Basis Pursuit*, in "Signal Processing with Adaptive Sparse Structured Representations, SPARS 2013", Lausanne, Switzerland, July 2013, http://hal.inria.fr/hal-00811641

[44] J. THIEMANN, N. ITO, E. VINCENT. *The Diverse Environments Multi-channel Acoustic Noise Database (DEMAND): A database of multichannel environmental noise recordings*, in "21st International Congress on Acoustics", Montreal, Canada, Acoustical Society of America, June 2013, http://hal.inria.fr/hal-00796707

[45] A. TILLMANN, R. GRIBONVAL, M. PFETSCH. *Projection Onto The k-Cosparse Set is NP-Hard*, in "Signal Processing with Adaptive Sparse Structured Representations 2013 (2013)", Lausanne, Switzerland, 2013, http://hal.inria.fr/hal-00811671

Scientific Books (or Scientific Book chapters)

[46] , *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech), 25-29 August 2013, Lyon (France)*, International Speech Communication Association (ISCA), August 2013, over 3500 p. , http://hal.inria.fr/hal-00931864

Research Reports

[47] R. GRIBONVAL, P. MACHART. , *Reconciling "priors" & "priors" without prejudice? (research report)*, Inria, September 2013, n^o RR-8366, http://hal.inria.fr/hal-00862925

[48] N. SOUVIRAÀ-LABASTIE, L. CATANESE, G. GRAVIER, F. BIMBOT. , *The MODIS software for word like motif discovery and its use for zero resource audio summarization*, Inria, July 2013, n^o RT-0439, http://hal.inria.fr/hal-00848631

[49] L. YU, P. MAUREL, C. BARILLOT, R. GRIBONVAL. , *Quantifying CBF from Aterial Spin Labeling via Diverse-TI: sampling diversity or repetitions ?*, Inria, March 2013, n^{o} RR-8258, http://hal.inria.fr/hal-00799718

Patents and standards

[50] G. KERGOURLAY, J. CITÉRIN, E. NGUYEN, L. LE SCOLAN, J. THIEMANN, E. VINCENT, N. BERTIN, F. BIMBOT. , *Sound source separation method*, 2013, n^{o} 1313218.8, http://hal.inria.fr/hal-00923802

[51] G. KERGOURLAY, J. THIEMANN, E. VINCENT, N. BERTIN, F. BIMBOT. , *Method and apparatus for sound source separation based on a binary activation model*, 2013, n^{o} 1304774.1, http://hal.inria.fr/hal-00923803

Other Publications

[52] C. BILEN, R. GRIBONVAL, G. PUY, L. DAUDET. *Blind Calibration for Phase Shifts in Compressive Systems*, in "Signal Processing with Adaptive Sparse Structured Representations (SPARS 2013)", Lausanne, Switzerland, July 2013, http://hal.inria.fr/hal-00811848

[53] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. *Blind Sensor Calibration in Sparse Recovery*, in "international biomedical and astronomical signal processing (BASP) Frontiers workshop", Villars-sur-Ollon, Switzerland, January 2013, http://hal.inria.fr/hal-00751360

[54] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. , *Convex Optimization Approaches for Blind Sensor Calibration using Sparsity*, August 2013, submitted, http://hal.inria.fr/hal-00853225

[55] A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL. , *S-ISTA and Brothers : a Dynamic Screening test principle for the Lasso*, November 2013, submitted, http://hal.inria.fr/hal-00880787

[56] A. BOURRIER, M. E. DAVIES, T. PELEG, P. PÉREZ, R. GRIBONVAL. , *Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems*, November 2013, Submitted to IEEE Trans. Inf. Theory, http://hal.inria.fr/hal-00908358

[57] V. EMIYA, A. BONNEFOY, R. GRIBONVAL, L. DAUDET. , *Compressed sensing with unknown sensor permutation*, 2013, submitted, http://hal.inria.fr/hal-00881407

[58] R. GRIBONVAL. *Reconciling "priors" and "priors" without prejudice ?*, in "international biomedical and astronomical signal processing (BASP) Frontiers workshop", Villars-sur-Ollon, Switzerland, 2013, http://hal.inria.fr/hal-00751996

[59] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. , *Sample Complexity of Dictionary Learning and other Matrix Factorizations*, December 2013, submitted, http://hal.inria.fr/hal-00918142

[60] S. KITIĆ, N. BERTIN, R. GRIBONVAL. , *Hearing behind walls: localizing sources in the room next door with cosparsity*, December 2013, Submitted to ICASSP, http://hal.inria.fr/hal-00904779

[61] S. KITIĆ, N. BERTIN, R. GRIBONVAL. , *Audio Declipping by Cosparse Hard Thresholding*, January 2014, Submitted to iTwist 2014, http://hal.inria.fr/hal-00922497

[62] A. TILLMANN, R. GRIBONVAL, M. PFETSCH. , *Projection Onto The k-Cosparse Set is NP-Hard*, March 2013, submitted, http://hal.inria.fr/hal-00802359

References in notes

[63] A. ADLER, V. EMIYA, M. JAFARI, M. ELAD, R. GRIBONVAL, M. D. PLUMBLEY. *Audio Inpainting*, in "IEEE Transactions on Audio, Speech and Language Processing", March 2012, vol. 20, n^o 3, pp. 922 - 932 [*DOI :* 10.1109/TASL.2011.2168211], http://hal.inria.fr/inria-00577079

[64] L. ALBERA, P. COMON, L. PARRA, A. KARFOUL, A. KACHENOURA, L. SENHADJI. *Biomedical applications*, in "Handbook of blind source separation", P. COMON, C. JUTTEN (editors), Academic Press, March 2010

[65] S. ARAKI, F. NESTA, E. VINCENT, Z. KOLDOVSKY, G. NOLTE, A. ZIEHE, A. BENICHOUX. *The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio source separation -*, in "10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)", Tel Aviv, Israël, March 2012, pp. 414-422, http://hal.inria.fr/hal-00655394

[66] S. ARBERET, A. OZEROV, F. BIMBOT, R. GRIBONVAL. *A tractable framework for estimating and combining spectral source models for audio source separation*, in "Signal Processing", August 2012, vol. 92, n^o 8, pp. 1886-1901, http://hal.inria.fr/hal-00694071

[67] F. ASANO, S. IKEDA, M. OGAWA, H. ASOH, N. KITAWAKI. *Combined approach of array processing and independent component analysis for blind separation of acoustic signals*, in "IEEE Transactions On Speech and Audio Processing", May 2003, vol. 11, n^o 3, pp. 204-215

[68] D. BARRY, D. FITZ, E. COYLE, B. LAWLOR. *Single channel source separation using short-time independent component analysis*, in "Audio Engineering Society, 119's convention", 2005, pp. 1-6

[69] A. BENICHOUX, L. S. R. SIMON, E. VINCENT, R. GRIBONVAL. , *Convex regularizations for the simultaneous recording of room impulse responses*, Inria, November 2012, n^o RR-8130, http://hal.inria.fr/hal-00749585

[70] A. BENICHOUX, P. SUDHAKAR, F. BIMBOT, R. GRIBONVAL. *Some uniqueness results in sparse convolutive source separation*, in "International Conference on Latent Variable Analysis and Source Separation", Tel Aviv, Israël, Springer, March 2012, http://hal.inria.fr/hal-00659913

[71] A. BENICHOUX, E. VINCENT, R. GRIBONVAL. *A compressed sensing approach to the simultaneous recording of multiple room impulse responses*, in "WASPAA", United States, 2011, 1 p.

[72] A. BENICHOUX, E. VINCENT, R. GRIBONVAL. *Optimisation convexe pour l'estimation simultanée de réponses acoustiques*, in "GRETSI 2011", Bordeaux, France, 2011

[73] G. BIROT, L. ALBERA, F. WENDLING, I. MERLET. *Localisation of extended brain sources from EEG/MEG: the ExSo-MUSIC approach*, in "Elsevier Neuroimage", May 2011, vol. 56, n^o 1, pp. 102-113

[74] C. BLANDIN, A. OZEROV, E. VINCENT. *Multi-source TDOA estimation in reverberant audio using angular spectra and clustering*, in "Signal Processing", March 2012, vol. 92, pp. 1950-1960, Revised version including minor corrections in equations (17), (18) and Figure 1 compared to the version published by Elsevier [*DOI :* 10.1016/J.SIGPRO.2011.10.032], http://hal.inria.fr/inria-00630994

[75] J. C. BROWN, P. SMARAGDIS. *Independent component analysis for automatic note extraction from musical trills*, in "Acoustical Society of America", May 2004, vol. 115, n^o 5, pp. 2295-2306

[76] J.-F. CARDOSO, A. SOULOUMIAC. *Blind beamforming for non-gaussian signals*, in "IEE Proceesings-F", December 1993, vol. 140, no. 6, pp. 362-370

[77] J. CARROLL, J. CHANG. *Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition*, in "Psychometrika", 1970, n^o 9, pp. 267-283

[78] M. A. CASEY, A. WESTNER. , *Separation of mixed audio sources by independent component analysis*, Mistubishi electric research labs, cambridge, MA, USA, 2000, http://www.merl.com/publications/TR2001-031/

[79] G. CHARDON, L. DAUDET, A. PEILLOT, F. OLLIVIER, N. BERTIN, R. GRIBONVAL. *Nearfield Acoustic Holography using sparsity and compressive sampling principles*, in "Journal of the Acoustical Society of America", 2012, Code & data for reproducing the main figures of this paper are available at the provided urls, http://hal.inria.fr/hal-00720129

[80] P. CHEVALIER, A. CHEVREUIL. *Application to telecommunications*, in "Handbook of blind source separation", P. COMON, C. JUTTEN (editors), Academic Press, March 2010

[81] K. DONG-GOO, S. YOUNGHUM, K. SUNGSU, L. SEONGDEOK, K. CHANGYEONG. *Multiple object decomposition based on independent component analysis of multi-energy X-ray projections*, in "International Conference on Image Processing ICIP", 2010, pp. 4173-4176

[82] N. DUONG, E. VINCENT, R. GRIBONVAL. , *Spatial location priors for Gaussian model-based reverberant audio source separation*, Inria, September 2012, n^o RR-8057, http://hal.inria.fr/hal-00727781

[83] R. GIRYES, S. NAM, M. ELAD, R. GRIBONVAL, M. E. DAVIES. *Greedy-Like Algorithms for the Cosparse Analysis Model*, in "Linear Algebra and its Applications", January 2014, vol. 441, pp. 22–60, partially funded by the ERC, PLEASE project, ERC-2011-StG-277906 [*DOI :* 10.1016/J.LAA.2013.03.004], http://hal.inria.fr/hal-00716593

[84] R. GRIBONVAL, G. CHARDON, L. DAUDET. *Blind Calibration for Compressed Sensing by Convex Optimization*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing", Kyoto, Japon, 2012, http://hal.inria.fr/hal-00658579

[85] R. GRIBONVAL. *Should penalized least squares regression be interpreted as Maximum A Posteriori estimation?*, in "IEEE Transactions on Signal Processing", May 2011, vol. 59, n^o 5, pp. 2405-2410 [*DOI :* 10.1109/TSP.2011.2107908], http://hal.inria.fr/inria-00486840

[86] R. GRIBONVAL, K. SCHNASS. *Dictionary identifiability from few training samples*, in "Proc. European Conf. on Signal Processing - EUSIPCO", August 2008

[87] J. HAO, X. ZOU, P. M. WILSON, N. P. DAVIES, Y. SUN, C. A. PEET, T. N. ARVANITIS. *A comparative study of feature extraction and blind source separation of independent component analysis (ICA) on childhood brain tumour 1H magnetic resonance spectra*, in "NMR in Biomedicine", 2009, vol. 22, n^o 8, pp. 809-818

[88] R. A. HARSHMAN. *Foundation of PARAFAC procedure: Models and conditions for an 'Explanatory' multi-mode factor analysis*, in "UCLA working papers in Phonetics", 1970, n^{o} 16, pp. 1-84

[89] F. L. HITCHCOCK. *Multiple invariants and generalized rank of a p-way matrix or tensor*, in "J. Math. and Phys.", 1927, vol. 7, n^{o} 1, pp. 39-79

[90] N. ITO. , *Robust microphone array signal processing against diffuse noise*, University of Tokyo, January 2012, http://hal.inria.fr/tel-00691931

[91] N. ITO, E. VINCENT, N. ONO, S. SAGAYAMA. , *Robust estimation of directions-of-arrival in diffuse noise based on matrix-space sparsity*, Inria, October 2012, n^{o} RR-8120, http://hal.inria.fr/hal-00746271

[92] C. JORDAN. *Mémoire sur les formes bilinéaires*, in "Journal de mathématiques pures et appliquées 2e série, tome 19", 1874, pp. 35-54

[93] A. KACHENOURA, L. ALBERA, L. SENHADJI, P. COMON. *ICA: a potential tool for BCI systems*, in "IEEE Signal Processing Magazine, special issue on Brain-Computer Interfaces", January 2008, vol. 25, n^{o} 1, pp. 57-68

[94] J. B. KRUSKAL, R. A. HARSHMAN, M. E. LUNDY. *How 3-MFA data can cause degenerate Parafac solutions, among other relationships*, in "Multiway Data Analysis", 1989, pp. 115-121

[95] C. LADROUE, F. HOWE, J. GRIFFITHS, R. TATE. *Independent Component Analysis for Automated Decomposition of In Vivo Magnetic Resonance Spectra*, in "Magnetic Resonnance in Medicine", 2003, vol. 50, pp. 697-703

[96] J. LE ROUX, E. VINCENT. *Consistent Wiener filtering for audio source separation*, in "IEEE Signal Processing Letters", October 2012, http://hal.inria.fr/hal-00742687

[97] B. LUDUSAN, S. ZIEGLER, G. GRAVIER. *Integrating Stress Information in Large Vocabulary Continuous Speech Recognition*, in "Interspeech", Portland, U.S.A., 2012, http://hal.inria.fr/hal-00758622

[98] P. MCCULLAGH. , *Tensor Methods in Statistics*, Chapman and Hall, Monographs on Statistics and Applied Probability, 1987

[99] A. MUSCARIELLO, F. BIMBOT, G. GRAVIER. *Unsupervised Motif Acquisition in Speech via Seeded Discovery and Template Matching Combination*, in "IEEE Transactions on Audio, Speech and Language Processing", September 2012, vol. 20, n^{o} 7, pp. 2031 - 2044 [*DOI :* 10.1109/TASL.2012.2194283], http://hal.inria.fr/hal-00740978

[100] S. NAM, R. GRIBONVAL. *Physics-driven structured cosparse modeling for source localization*, in "Acoustics, Speech and Signal Processing, IEEE International Conference on (ICASSP 2012)", Kyoto, Japon, IEEE, 2012, http://hal.inria.fr/hal-00659405

[101] G. NOLTE, D. LUTTER, A. ZIEHE, F. NESTA, E. VINCENT, Z. KOLDOVSKY, A. BENICHOUX, S. ARAKI. *The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Biomedical data analysis -*, in "10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)", Tel Aviv, Israël, March 2012, pp. 423-429, http://hal.inria.fr/hal-00655801

[102] B. OPHIR, M. ELAD, N. BERTIN, M. D. PLUMBLEY. *Sequential Minimal Eigenvalues - An Approach to Analysis Dictionary Learning*, in "The 19th European Signal Processing Conference (EUSIPCO-2011)", Barcelona, Spain, 2011, http://hal.inria.fr/inria-00577231

[103] A. OZEROV, E. VINCENT, F. BIMBOT. *A General Flexible Framework for the Handling of Prior Information in Audio Source Separation*, in "IEEE Transactions on Audio, Speech and Language Processing", May 2012, vol. 20, n^{o} 4, pp. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962

[104] R. D. PASCUAL-MARQUI, M. ESSLEN, D. LEHMANN. *Functional imaging with low resolution brain electromagnetism (LORETA): review, new comparison, and new validations*, in "Japanese Journal of Clinical Neurophysiology", 2002, vol. 30, pp. 81-94

[105] R. D. PASCUAL-MARQUI. *Standardized low resolution brain electromagnetic tomography (sLORETA): technical details*, in "Methods and Findings in Experimental and Clinical Pharmacology", 2002

[106] J. PULKKINEN, A. M. HAKKINEN, N. LUNDBOM, A. PAETAU, R. A. KAUPPINEN, Y. HILTUNEN. *Independent component analysis to proton spectroscopic imaging data to human brain tumours*, in "European Journal of Radiology", November 2005, vol. 56, n^{o} 2, pp. 160-164

[107] S. RACZYNSKI, S. FUKAYAMA, E. VINCENT. , *Melody harmonisation with interpolated probabilistic models*, Inria, October 2012, n^{o} RR-8110, http://hal.inria.fr/hal-00742957

[108] S. RACZYNSKI, E. VINCENT, S. SAGAYAMA. , *Dynamic Bayesian networks for symbolic polyphonic pitch modeling*, Inria, September 2012, n^{o} RT-0430, http://hal.inria.fr/hal-00728771

[109] M. RAJIH, P. COMON, R. A. HARSHMAN. *Enhanced Line Search: A Novel Method to Accelerate PARAFAC*, in "SIAM Journal in Matrix Analysis and Applications", 2008

[110] K. SCHNASS, R. GRIBONVAL. *Basis Identification from Random Sparse Samples*, in "Proc. SPARS'09 (Signal Processing with Adaptive Sparse Structured Representations)", Saint-Malo, France, April 2009

[111] L. S. R. SIMON, E. VINCENT. *A general framework for online audio source separation*, in "International conference on Latent Variable Analysis and Signal Separation", Tel-Aviv, Israël, March 2012, http://hal.inria.fr/hal-00655398

[112] F. SZABO DE EDELENYI, A. SIMONETTI, G. POSTMA, R. HUO, L. BUYDENS. *Application of independent component analysis to 1H MR spectroscopic imaging exams of brain tumours*, in "Analytica Chimica Acta", July 2005, vol. 544, n^{o} 1-2, pp. 36-46

[113] G. TOMASI, R. BRO. *A comparison of algorithms for fitting the PARAFAC model*, in "Computational Statistics & Data Analysis", 2006, vol. 50, n^{o} 7, pp. 1700-1734

[114] A. TONAZZINI, L. BEDINI, E. SALERNO. *Independent component analysis for document restoration*, in "International Journal on Document Analysis and Recogniton IJDAR", April 2004, n^{o} 7, pp. 17–27

[115] L. TUCKER. *Some mathematical notes on three-mode factor analysis*, in "Psychometrika", September 1966, vol. 31, n^{o} 3, pp. 279-311

[116] E. VINCENT, S. ARAKI, F. J. THEIS, G. NOLTE, P. BOFILL, H. SAWADA, A. OZEROV, B. V. GOWREESUNKER, D. LUTTER, N. DUONG. *The Signal Separation Evaluation Campaign (2007-2010): Achievements and Remaining Challenges*, in "Signal Processing", March 2012, vol. 92, pp. 1928-1936 [*DOI :* 10.1016/J.SIGPRO.2011.10.007], http://hal.inria.fr/inria-00630985

[117] E. VINCENT. *Improved perceptual metrics for the evaluation of audio source separation*, in "10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA)", Tel Aviv, Israël, March 2012, pp. 430-437, http://hal.inria.fr/hal-00653196

[118] A. J. WRIGHT, G. FELLOW, T. J. BYRNES, K. S. OPSTAD, D. J. O. MCINTYRE, J. GRIFFITHS, B. A. BELL, C. A. CLARK, T. R. BARRICK, F. HOWE. *Pattern recognition of MRSI data shows regions of glioma growth that agree with DTI markers of brain tumor infiltration*, in "Magnetic Resonance in Medicine", 2009, vol. 62, pp. 1646-1651

[119] M. YAGHOOBI, S. NAM, R. GRIBONVAL, M. E. DAVIES. *Noise Aware Analysis Operator Learning for Approximately Cosparse Signals*, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing - 2012", Kyoto, Japon, IEEE, 2012, http://hal.inria.fr/hal-00661549

[120] A. YEREDOR. *Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation*, in "IEEE Transactions On Signal Processing", July 2002, vol. 50, n^{o} 7, pp. 1545-1553

[121] S. ZIEGLER, B. LUDUSAN, G. GRAVIER. *Towards a new speech event detection approach for landmark-based speech recognition*, in "IEEE Workshop on Spoken Language Technology", États-Unis, 2012, http://hal.inria.fr/hal-00758424

[122] S. ZIEGLER, B. LUDUSAN, G. GRAVIER. *Using Broad Phonetic Classes to Guide Search in Automatic Speech Recognition*, in "Interspeech", Portland, U.S.A., 2012, http://hal.inria.fr/hal-00758427