



IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2014

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

1. Members	1
2. Overall Objectives	2
2.1. Overall positioning	2
2.2. Scientific foundations	2
2.3. Applications	2
3. Research Program	2
3.1. Axis 1: sparse models and representations	2
3.1.1. Efficient sparse models and dictionary design for large-scale data	2
3.1.2. Compressive Learning	3
3.2. Axis 2: robust acoustic scene analysis	3
3.2.1. Compressive acquisition and processing of acoustic scenes	3
3.2.2. Robust audio source separation	3
3.3. Axis 3: large-scale audio content processing and self-organization	4
3.3.1. Motif discovery in audio data	4
3.3.2. Structure modeling and inference in audio and musical contents	5
4. Application Domains	5
4.1. Acoustic scene capture	5
4.2. Audio signal separation in reverberant environments	5
4.3. Multimedia indexing	5
4.4. Brain source imaging	6
5. New Software and Platforms	6
6. New Results	6
6.1. Highlights of the Year	6
6.2. Recent results on sparse representations	7
6.2.1. A new framework for sparse representations: analysis sparse models	7
6.2.2. Theoretical results on sparse representations	7
6.2.3. Algorithmic and theoretical results on dictionary learning	8
6.3. Activities on waveform design for telecommunications	9
6.4. Emerging activities on compressive sensing, learning and inverse problems	9
6.4.1. Audio inpainting	9
6.4.2. Blind Calibration of Compressive Sensing systems	10
6.4.3. Compressive Gaussian Mixture estimation	10
6.5. Recent results on tensor decompositions	10
6.6. Source separation and localization	12
6.6.1. A general framework for audio source separation	12
6.6.2. Towards real-world separation and remixing applications	13
6.6.3. Acoustic source localization	13
6.6.4. Brain source localization	14
6.6.5. Independent component analysis	16
6.6.6. Semi-nonnegative independent component analysis	16
6.7. Audio and speech content processing	17
6.7.1. Audio motif discovery	17
6.7.2. Mobile device for the assistance of users in potentially dangerous situations	17
6.8. Music Content Processing and Music Information Retrieval	18
7. Bilateral Contracts and Grants with Industry	18
7.1. Bilateral Contracts with Industry	18
7.2. Bilateral Grants with Industry	18
8. Partnerships and Cooperations	19
8.1. National Initiatives	19

8.1.1. Labex Comin Labs projets	19
8.1.1.1. HEMISFER	19
8.1.1.2. TEPN	19
8.1.2. OSEO-FUI: S-POD: “Assistance à personnes en danger potentiel”	20
8.1.3. Action de Développement Technologique: FASST	20
8.2. European Initiatives	21
8.3. International Initiatives	21
8.4. International Research Visitors	21
9. Dissemination	21
9.1. Promoting Scientific Activities	21
9.2. Teaching	22
9.3. Popularization	23
9.3.1. Journée science et musique	23
9.3.2. Radio and press	23
10. Bibliography	23

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 [Univ. Rennes I, Associate Professor, HdR]

Engineers

Ewen Camberlein [CNRS]
Jules Espiau de Lamaestre [Inria, until May 2014, granted by MAIA - Musiciens Artistes Interprètes Associés]
Romain Lebarbenchon [CNRS]
Jérémy Paret [Inria, until Sep 2014]

PhD Students

Anthony Bourrier [Technicolor R&I, CIFRE, until May 2014]
Corentin Guichaoua [Univ. Rennes I]
Srdan Kitic [Inria, granted by FP7 ERC PLEASE project]
Luc Le Magoarou [Inria, granted by FP7 ERC PLEASE project]
Nathan Souviraà -Labastie [Univ. Rennes I, granted by MAIA - Musiciens Artistes Interprètes Associés]
Nicolas Keriven [Univ. Rennes I, Contrat Doctoral Polytechnicien, from Sep 2014]

Post-Doctoral Fellows

Anthony Bourrier [Inria, from May 2014 until Aug 2014, granted by FP7 ERC PLEASE project]
Cagdas Bilen [Inria, until Aug 2014, granted by FP7 ERC PLEASE project]
Pierre Machart [Inria, granted by FP7 ERC PLEASE project]
Anaik Olivero [Inria, until Oct 2014, granted by Abondement Carnot then FP7 ERC PLEASE project]
Gilles Puy [Inria, from Feb 2014, granted by FP7 ERC PLEASE project]
Yann Traonmilin [Inria, from Sep 2014, granted by FP7 ERC PLEASE project]

Administrative Assistant

Stéphanie Lemaile [Inria]

Others

Thomas Aubert [Intern, from Apr 2014 until Jun 2014]
Theo Dabreteau [Intern, from Jun 2014 until Aug 2014]
Melanie Ducoffe [Intern, from Feb 2014 until Jun 2014]
Etienne Gerat [Intern, from Jul 2014 until Aug 2014]
Anh Tho Le [Intern, from Apr 2014 until Jun 2014]
Maxime Lecoq [Intern, from Apr 2014 until Jul 2014]
Nicolas Keriven [Intern, Univ. Rennes I, from Jul 2014 until Aug 2014]

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 scenarios, 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.

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. New Software and Platforms

5.1. 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 development was jointly achieved by the PAROLE team in Nancy and the TEXMEX team in Rennes through an Inria funded ADT (Action de Développement Technologique). PANAMA contributed to the development by coordinating and performing user tests, and to the dissemination in a Show-and-Tell ICASSP poster [58].

While the first implementation was in Matlab, the new implementation is in C++ (for core functions), with Matlab and Python user scripts. Version 2, including speedup and new features was released in 2014 and can be downloaded from <http://bass-db.gforge.inria.fr/fasst/>.

6. New Results

6.1. Highlights of the Year

The EUSIPCO 2014 Best Student Paper Award was awarded to our joint paper [32] on dynamic screening for sparse regularization. .

A review paper on audio source separation, rooted in METISS/PANAMA know-how and contributions to this topic over the years, was published in the IEEE Signal Processing Magazine [25].

A new version of the Flexible Audio Source Separation Toolbox (FASST) was released in January 2014 and downloaded 300 times.

BEST PAPER AWARD :

[32] **A Dynamic Screening Principle for the Lasso in European Signal Processing Conference EUSIPCO 2014.** A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL.

6.2. 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 8.2.1.1).

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

Participants: Rémi Gribonval, Nancy Bertin, Srđan Kitić, Cagdas Bilen, Laurent Albera.

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 [100]. 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 [101].

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 [97] and showed its efficiency in situations where usual methods fail ([37], see paragraph 6.6.3). It was also shown to be applicable to the hard declipping problem [49]. Application to EEG brain imaging was also investigated and a paper was published in MLSP14 [28] (see paragraph 6.6.4).

6.2.2. Theoretical results on sparse representations

Participants: Rémi Gribonval, Anthony Bourrier, Pierre Machart, Yann Traonmilin, Gilles Puy.

Main collaboration: Mike Davies (University of Edinburgh), Patrick Perez (Technicolor R&I France), Tomer Peleg (The Technion)

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 whose 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 [106], [105]. This year, noiseless and noise-robust settings were both considered in the journal paper [16]. 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. Current work explores the existence of convex decoders for general union of subspaces models under generalized RIP assumptions, as well as conditions ensuring that random low-dimensional projections ensure the RIP even when the projection is from an infinite-dimensional space to a finite dimensional one. Envisioned applications are in compressive learning (see Section 6.4).

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 $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$. In 2013, we extended this result to general inverse problems [88], [86], [87]. This year, we worked on the characterization of such relations beyond the Gaussian noise model, with the objective of understanding whether similar results hold when the quadratic data-fidelity term is replaced with other convex losses.

6.2.3. Algorithmic and theoretical results on dictionary learning

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen, Luc Le Magoarou, Melanie Ducoffe.

Main collaboration (theory for dictionary learning): Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleinsteuber, Matthias Seibert (TU-Munich),

Main collaboration (dictionary learning for gesture recognition): Anatole Lecuyer, Ferran Argelaguet (EPI HYBRID, Rennes)

Theoretical guarantees for 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 literature 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 [89], [109] [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 [54]. Moreover, we established new sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [55], [46], [38].

Learning computationally efficient dictionaries Classical dictionary learning is limited to small-scale problems. Inspired by usual fast transforms, we proposed a general dictionary structure that allows cheaper manipulation, and an algorithm to learn such dictionaries –and their fast implementation [50]. A preprint is available [56], a paper will appear at ICASSP 2015, and a journal paper is in preparation.

Operator learning for cospase representations : Besides standard dictionary learning, we also considered learning in the context of the cospase model. The overall problem is to learn a low-dimensional signal model from a collection of training samples. The mainstream approach is 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 cospase analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator.

In specific situations, when prior information is available on the operator, it is possible to express it in parametric form and learn this parameter. For instance, in the sound source localization problem, we showed that the unknown speed of sound can be learned jointly in the process of cospase recovery, under mild conditions. This work was presented at iTWist'14 workshop [48].

Dictionary learning for gesture modeling In collaboration with the HYBRID project-team (internship of Melanie Ducoffe), we explored the potential of dictionary learning in the context of motion tracking. Motion tracking technology, especially for commodity hardware, requires robust gesture recognition algorithms to fully exploit the benefits of natural user interfaces. We proposed a gesture recognition algorithm based on the sparse representation of motion data, with a learning phase consisting in learning a dictionary of basic gestures. A paper is in preparation.

6.3. Activities on waveform design for telecommunications

Participant: Rémi Gribonval.

Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (Equipe SCEE, Supelec, Rennes)

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 8.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii co-supervised by R. Gribonval), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems. A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, leading to a journal paper [17] and several conference communications [27], [33]. The work now concentrates on characterizing waveforms with optimum PAPR.

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

Compressive sensing, compressive learning, acoustic wavefields, audio inpainting,

6.4.1. Audio inpainting

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

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. A series of works about audio inpainting was initiated by the METISS team in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data).

Building upon our previous contributions (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [60]), new results were obtained this year to address the case of audio declipping with the competitive cosparsity approach. Its promising results, especially when the clipping level is low, were confirmed experimentally by the formulation and use of a new algorithm named Cosparsity Iterative Hard Thresholding, which is a counterpart of the sparse Consistent Iterative Hard Thresholding. These results were presented during the iTwist'14 workshop [49]. Additional experiments were performed (internship of Anh Tho Le) to confirm the results on a larger database and investigate optimal parameters (nature and redundancy of the dictionary, relaxation parameter for the cosparsity level).

Current and future works deal with developing advanced (co)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).

6.4.2. *Blind Calibration of Compressive Sensing systems*

Participants: Rémi Gribonval, Cagdas Bilen, Gilles Puy.

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

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists of 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]. We extended the framework to phase-only decalibration, using techniques revolving around low-rank matrix recovery [66], [65], [110], [64], and to joint phase and gain decalibration [15].

6.4.3. *Compressive Gaussian Mixture estimation*

Participants: Rémi Gribonval, Anthony Bourrier, Nicolas Keriven.

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 [71], [69], [70]. This year, extensions to non-isotropic Gaussians, with new algorithms and preliminary applications to speaker verification have been conducted.

6.5. 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 [96]. 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 [92] 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 exercises 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 [114]. 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 [91], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [74]. In honor to Hitchcock's pioneer 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 result in a slow convergence (or lack of convergence) of iterative algorithms such as the ALS one [98]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

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

Participant: Laurent Albera.

Main collaboration (Line search and trust region strategies): 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 $CG_{2\text{steps}}$ -like and LM-like algorithms. This work was published in the Elsevier Linear Algebra and Applications journal [19].

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 [108] and LM [112] algorithms without symmetry and nonnegativity constraints, and the ACDC algorithm [115] 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 published in the Elsevier Signal Processing journal [18].

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 and rational subproblems using i) a square change of variables to impose nonnegativity and ii) LU or QR 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 matrices, each one depending on only one parameter.

The proposed approach reduces the optimization problem to the computation of the two equal nonnegative loading matrices only. 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 published in a journal paper [26]. Numerical experiments on simulated matrices emphasize the advantages of the proposed algorithms over classical CP and JDC techniques, especially in the case of degeneracies.

6.6. Source separation and localization

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

6.6.1. A general framework for audio source separation

Participants: Frédéric Bimbot, Rémi Gribonval, Nancy Bertin.

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

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal.

About a decade ago, state-of-the-art approaches consisted of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. These approaches rely only on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [102], which generalizes a number of existing techniques including our former study on spectral GMMs [61]. We showed how it could be used to quickly design new models adapted to the data at hand and estimate its parameters via the EM algorithm., and it became the basis of a large number of works in the field, including our own. In the last years, improvements were obtained through the use of prior knowledge about the source spatial covariance matrices [81], [95], [94], knowledge on the source positions and room characteristics [82], or a better initialization of parameters thanks to specific source localization techniques [68].

This accumulated progress lead to two main achievements this year, which show the maturity of our work and which will leverage its impact. First, a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released. It will provide the community with an efficient and ergonomic software, making available the tools from past years' research [58]. Second, we published an overview paper on recent and going research along the path of *guided* separation, *i.e.*, techniques and models allowing to incorporate knowledge in the process towards efficient and robust solutions to the audio source separation problem, in a special issue of IEEE Signal Processing Magazine devoted to source separation and its applications [25].

6.6.2. Towards real-world separation and remixing applications

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

Emmanuel Vincent (EPI PAROLE, Inria Nancy)

While some challenges remain, work from previous years and our review paper on guided source separation [25] highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications, especially when some knowledge can be incorporated. This is exemplified by the contract with MAIA Studio, which reaches its end in December 2014 and showed in particular how user input or side information could raise source separation tools to efficient solutions in real-world applications.

In this context, new tools were developed this year. The introduction of manually-tuned parameters in the automated separation process, which modifies the Wiener filtering coefficients obtained from estimation of the mixtures covariance matrices, allows to find a better trade-off between artifacts and interferences. In order to ensure high audio quality for such applications, some user-guided corrections remain necessary even after an automatic pre-separation ; to this end, we developed an improved display (based on cepstrum and automatic contrast adaptation) and semi-automatic selection and suppression tools in the time-frequency domain. Those tools take as few inputs as possible from the user and their result can be ergonomically adjusted from the baseline output to a manually fine-tuned area, in a very small operating time. We also proposed tools to suppress a time-frequency area and replace it by content extracted from its context, reducing the perceptual impact of the suppression.

In some applicative contexts of source separation, several mixtures are available which contain similar instances of a given source. We have designed a general framework for audio source separation guided by multiple audio references, where each audio reference is a mixture which is supposed to contain at least one source similar to one of the target sources. Deformations between the sources of interest and their references are modeled in a general manner. A nonnegative *matrix co-factorization* algorithm is used which allows sharing of information between the considered mixtures. We have experimented our algorithm on music plus voice mixtures with music and/or voice references. Applied on movies and TV series data, the algorithm improves the signal-to-distortion ratio (SDR) of the sources of lowest intensity by 9 to 12 decibels with respect to original mixture [40]

6.6.3. Acoustic source localization

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

Main collaborations (audio-based control for robotics): Aly Magassouba and François Chaumette (Inria, EPI LAGADIC, France)

Acoustic source localization is, in general, the problem of determining the spatial coordinates of one or several sound sources based on microphone recordings. This problem arises in many different fields (speech and sound enhancement, speech recognition, acoustic tomography, robotics, aeroacoustics...) and its resolution, beyond an interest in itself, can also be the key preamble to efficient source separation. Common techniques, including beamforming, only provides the *direction of arrival* of the sound, estimated from the *Time Difference of Arrival (TDOA)* [68]. This year, we have particularly investigated alternative approaches, either where the explicit localization is not needed (audio-based control of a robot) or, on the contrary, where the exact location of the source is needed and/or TDOA is irrelevant (cosparsity modeling of the acoustic field).

Implicit localization through audio-based control for robotics

In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to vision. Yet audio sensing is a valid alternative or complement to vision in robotics, for instance in homing tasks. Most existing works are based on the relative localization of a defined system with respect to a sound source, and the control scheme is generally designed separately from the localization system.

In contrast, the approach that we started investigating this year focuses on a sensor-based control approach. We proposed a new line of work, by considering the hearing sense as a direct and real-time input of closed loop control scheme for a robotic task. Thus, and unlike most previous works, this approach does not necessitate any explicit source localization: instead of solving the localization problem, we focus on developing an innovative modeling based on sound features. To address this objective, we placed ourselves in the sensor-based control framework, especially visual servoing (VS) that has been widely studied in the past [76].

From now on, we have established an analytical model linking sound features and control input of the robot, defined and analyzed robotic homing tasks involving multiple sound sources, and validated the proposed approach by simulations. This work is mainly lead by Aly Magassouba, whose Ph.D. is co-supervised by Nancy Bertin and François Chaumette. A conference paper presenting these first results was submitted to ICRA 2015. Future work will include real-world experiments with the robot Romeo from Aldebaran Robotics.

Cosparse modeling of the acoustic field

Cosparse modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator, which is the case for the acoustic field, ruled by the wave equation. Unlike usual localization techniques such as beamforming or TDOA-based direction estimation, which generally consider reverberation as an adverse condition, the cosparse modeling of sound propagation has also the interest of considering reverberation as a source of additional information for the localization task. Eventually, it can provide a full coordinate localization of the sources, and not only their direction of arrival.

Building upon our previous results on cosparse modeling and recovery algorithms for the wave equation [97], we have obtained additional evidence of the interest of this approach. In particular, we have showed that recasting source localization as a cosparse inverse problem allows to scale up to 3-dimensional problems which were untractable with the counterpart sparse approach. Moreover, we have confirmed that our model takes indeed advantage of reverberation, by showing that localization remains possible when the sources and the microphones are partly separated by an acoustically opaque obstacle (a situation where TDOA would obviously fail). These two results were published and presented during ICASSP'14 [37]. Recent results also include algorithmic improvements (through the use of the Alternating Direction Method of Multipliers (ADMM) framework), and evidence that, in addition to its scaling capabilities, the sparse analysis computational cost can even *benefit* from an increase in the number of measurements. A journal paper including these new results and presenting them jointly with co-space modeling in the context of brain source localization (see Section 6.6.4) is under preparation.

6.6.4. Brain source localization

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

Main collaborations (tensor-based approaches): 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), Gwénaél Birot (FBML, Genève), Martin Haardt (TUI, Germany)

Main collaborations (from tensor to sparse models): Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France)

Main collaborations (a sparsity-based approach): Hanna Becker (Technicolor, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France)

Main collaborations (a multimodal sparsity-based approach): Thomas Oberlin, Pierre Maurel, Christian Barillot (EPI VISAGES, Rennes, 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 [104], cortical LORETA (cLORETA) [103], and 4-ExSo-MUSIC [67]. 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 published in the Elsevier NeuroImage journal [13].

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 presented at ICASSP'14 [30] while the whole part was submitted to IEEE Signal Processing Magazine.

A cosparsity-based approach

Cosparsity 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 (Cosparsity 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 partially presented at MLSP'14 [28].

A sparsity-based approach

Identifying the location and spatial extent of several highly correlated and simultaneously active brain sources from EEG recordings and extracting the corresponding brain signals is a challenging problem. In our comparison of source imaging techniques presented at ICASSP'14 [30], the VB-SCCD algorithm [79], which exploits the sparsity of the variational map of the sources, proved to be a promising approach. We proposed several ways to improve this method. In order to adjust the size of the estimated sources, we added a regularization term that imposes sparsity in the original source domain. Furthermore, we demonstrated the application of ADMM, which permitted to efficiently solve the optimization problem. Finally, we also considered the exploitation of the temporal structure of the data by employing L1,2-norm regularization. The performance of the resulting algorithm, called L1,2-SVB-SCCD, was evaluated based on realistic simulations in comparison to VB-SCCD and several state-of-the-art techniques for extended source localization. This work was partially presented at EUSIPCO'14 [29] and a journal paper is in preparation.

A multimodal sparsity-based approach

In the context of the HEMISFER Comin Labs project (see Section 8.1.1.1), in collaboration with the VISAGES team, we investigated brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduced a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The joint imaging problem is expressed as a linear inverse problem, which is addressed using sparse regularization. The sparsity prior naturally reflects the fact that only few areas of the brain are activated at a certain time, and it is easily implemented through proximal algorithms. At this stage, the significance of the method and its effectiveness have been demonstrated through numerical investigations on a simplified head model and simulated data on a realistic brain model. A conference paper has been submitted and a journal paper is in preparation.

6.6.5. Independent component analysis

Participant: Laurent Albera.

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

Independent Component Analysis (ICA) is a very useful tool to process biomedical signals including EEG data. We proposed a Jacobi-like Deflationary ICA algorithm, named JDICA. More particularly, while a projection-based deflation scheme inspired by Delfosse and Loubaton's ICA technique (DeL^R) [78] was used, a Jacobi-like optimization strategy was proposed in order to maximize a fourth order cumulant-based contrast built from whitened observations. Experimental results obtained from simulated epileptic data mixed with a real muscular activity and from the comparison in terms of performance and numerical complexity with the FastICA [93], RobustICA [116] and DeL^R algorithms, show that the proposed algorithm offers the best trade-off between performance and numerical complexity. This work was submitted for publication in the IEEE Signal Processing Letters journal.

6.6.6. Semi-nonnegative independent component analysis

Participant: Laurent Albera.

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

ICA plays also an important role in many other areas including speech and audio [62], [63], [75], [72], radiocommunications [77] and document restoration [113] to cite a few.

For instance in [113], 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 [80] 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 [107], 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 [111]. Moreover, using the JADE algorithm [73] 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 [72] 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 [80], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in

[113] 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 [90]; 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 [99] 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 6.5.1. Hence we use the new methods described in section 6.5.1 to perform semi-nonnegative ICA. Performance results in audio and biomedical engineering were given in the different papers cited in section 6.5.1.

6.7. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

6.7.1. Audio motif discovery

Participants: Frédéric Bimbot, Nathan Souviraà -Labastie.

This work was performed in close collaboration with Guillaume Gravier from the Limkmedia 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.

We have designed a system to create audio thumbnails of spoken content, i.e., short audio summaries representative of the entire content, without resorting to a lexical representation. As an alternative to searching for relevant words and phrases in a transcript, unsupervised motif discovery is here used to find short, word-like, repeating fragments at the signal level without acoustic models. The output of the word discovery algorithm is exploited via a maximum motif coverage criterion to generate a thumbnail in an extractive manner. A limited number of relevant segments are chosen within the data so as to include the maximum number of motifs while remaining short enough and intelligible.

Evaluation has been performed on broadcast news reports with a panel of human listeners judging the quality of the thumbnails. Results indicate that motif-based thumbnails stand between random thumbnails and ASR-based keywords, however still far behind thumbnails and keywords humanly authored [35].

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

Participants: Romain Lebarbenchon, Ewen Camberlein, 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.

This ongoing work is focused on the development of robust techniques for audio scene analysis, including statistical classification of audio segments into threat vs non-threat categories, and the use of spatial information to determine the location of the user with respect to the potential threat.

6.8. Music Content Processing and Music Information Retrieval

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

6.8.1. Music structure representation and decomposition

Participants: Frédéric Bimbot, Corentin Guichaoua, Anaik Olivero.

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

Interest has been steadily growing in semantic audio and music information retrieval for the description of music structure, i.e. the global organization of music pieces in terms of large-scale structural units. Our group has defined a detailed methodology for the semiotic description of music structure, based on concepts and criteria which are formulated as generically as possible, i.e. without resorting to intrinsic properties of the musical content, but rather on relationships between musical elements resulting in well-identifiable patterns. The essential principles and practices developed during an annotation effort deployed by our research group on audio data, in the context of the Quaero project, has led to the public release of over 380 annotations of pop songs from three different data sets (<http://musicdata.gforge.inria.fr/structureAnnotation.html>) documented by a technical report which includes a few case studies and a concise statistical overview of the annotated data [31]. From the algorithmic point of view, we are currently exploring tree-based representation of music structure where a sequence of musical elements and their relationships are modeled hierarchically as the derivation of a context-free grammar. Parsimony criteria and specific cost functions adapted to music patterns are used to learn the grammar rules and the possibility of distorting the rules is introduced to account for variability across different repetitions of musical segments.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. 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 transferring 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. New PANAMA tools were integrated in the software developed by Imaging Factory and delivered to MAIA in May 2014, and the final release will be delivered in December 2014.

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 was to explore, both numerically and theoretically, the potential of compressive sensing for the manipulation of large (audiovisual) databases. A particular objective was 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. Labex Comin Labs projets

CominLabs is a Laboratoire d'Excellence funded by the PIA (Programme Investissements d'Avenir) in the broad area of telecommunications.

8.1.1.1. HEMISFER

Participant: Rémi Gribonval.

<http://www.hemisfer.cominlabs.ueb.eu/>

Research axis: 3.1

CominLabs partners : EPI VISAGES; EPI HYBRID; EPI PANAMA

External partners : EA 4712 team from University of Rennes I; EPI ATHENA, Sophia-Antipolis;

Coordinator: Christian Barillot, EPI VISAGES

Description: The goal of HEMISFER is to make full use of neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain (stroke, attention-deficit disorder, language disorders, treatment-resistant mood disorders, ...).

Contribution of PANAMA: PANAMA, in close cooperation with the VISAGES team, contributes to a coupling model between EEG and fMRI considered as a joint inverse problem addressed with sparse regularization. By combining both modalities, one expects to achieve a good reconstruction both in time and space. This new imaging technique will then be used for improving neurofeedback paradigms in the context of rehabilitation and psychiatric disorders, which is the final purpose of the HEMISFER project.

Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation

8.1.1.2. TEPN

Participant: Rémi Gribonval.

<http://www.tepn.cominlabs.ueb.eu/>

Research axis: **3.1**

CominLabs partners : IRISA OCIF - Telecom Bretagne; IETR SCN; IETR SCEE; EPI PANAMA

Coordinator: Nicolas Montavont, IRISA OCIF - Telecom Bretagne

Description: As in almost all areas of engineering in the past several decades, the design of computer and network systems has been aimed at delivering maximal performance without regarding to the energy efficiency or the percentage of resource utilization. The only places where this tendency was questioned were battery-operated devices (such as laptops and smartphones) for which the users accept limited (but reasonable) performance in exchange for longer use periods. Even though the end users make such decisions on a daily basis by checking their own devices, they have no way of minimizing their energy footprint (or conversely, optimize the network resource usage) in the supporting infrastructure. Thus, the current way of dimensioning and operating the infrastructure supporting the user services, such as cellular networks and data centers, is to dimension for peak usage. The problem with this approach is that usage is rarely at its peak. The overprovisioned systems are also aimed at delivering maximal performance, with energy efficiency being considered as something desired, but non-essential. This project aims at making the network energy consumption proportional to the actual charge of this network (in terms of number of served users, or requested bandwidth). An energy proportional network can be designed by taking intelligent decisions (based on various constraints and metrics) into the network such as switching on and off network components in order to adapt the energy consumption to the user needs. This concept can be summarized under the general term of Green Cognitive Network Approach.

Contribution of PANAMA: PANAMA, in close cooperation with the SCEE team at IETR (thesis of Marwa Chafii), focuses on the design of new waveforms for multi carrier systems with reduced Peak to Average Power Ratio (PAPR).

Toward Energy Proportional Networks

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

Participants: Frédéric Bimbot, Romain Lebarbenchon, Ewen Camberlein.

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. The software was publicly released in January 2014.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. 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, Nicolas Keriven, Yann Traonmilin, Laurent Albera, Gilles Puy.

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/>

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

PANAMA has strong recurrent collaborations with the LTS2 lab at EPFL, the Center for Digital Music at Queen Mary University of London, the Institute for Digital Communications at the University of Edimburgh.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Mike Davies, in November, Professor of Signal and Image Processing, University of Edinburgh
- Pierre Vandergheynst, in November, Professor of Signal and Image Processing, EPFL
- Karin Schnass, in July, University of Innsbruck Department of Mathematics
- Gilles Blanchard, in May, Professor, University of Postdam
- Ivan Dokmanic, in January, Assistant Professor, EPFL, Lausanne

8.4.1.1. Internships

- Thomas Aubert, from April until June, University of Rennes 1
- Theo Dabreteau, from June until August, Insa of Rennes
- Melanie Ducoffe, from February until June, ENS Rennes
- Anh-tho Le, from April until June, University of Hanoi
- Maxime Lecoq, from April until July, University of Rennes 1

9. Dissemination

9.1. Promoting Scientific Activities

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 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).

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

Rémi Gribonval organized a mini-symposium on "Dictionary learning and applications" at the international conference Curves & Surfaces in Paris.

Rémi Gribonval organized a session "Mathematical for Image Processing" at the SMAI-MODE conference in Rennes.

Rémi Gribonval was co-organizer of a one-day seminar in Lyon on sparse representations and compressive sensing for medical imaging sponsored by the GDR ISIS (french research groups in signal and image processing) and GDR STIC-SANTE.

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

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

Frédéric Bimbot was a member of the Scientific Committee of Journées d'Analyse Musicale 2014 of the SFAM (Société Française d'Acoustique Musicale).

Rémi Gribonval and Frédéric Bimbot are the scientific coordinators of the Science and Music Day (Journée Science et Musique) organized by IRISA.

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

9.2. Teaching

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 : N. Bertin, "Fundamentals of Signal Processing", 24 hours, M1, Ecole Normale Supérieure de Bretagne, Rennes, 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 Représentation 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", 6 hours, L2, Université Rennes 1, France.

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

Licence: L. Albera, "Mathematics", 53 hours, L3, Ecole Supérieure d'Ingénieurs de Rennes, France

Master: L. Albera, "Cardiac source localization from ECG signals", project supervision (with Ansys), M2, Ecole Supérieure d'Ingénieurs de Rennes, France.

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

Master : L. Albera, "Inverse problems", 3 hours, M2, 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: Anaik Olivero, Jules Espiau de Lamaestre, Pierre Machart, Gilles Puy, Nancy Bertin, Frédéric Bimbot, Rémi Gribonval, Corentin Guichaoua, Srdan Kitic, Stéphanie Lemaile, Nicolas Keriven, Ewen Camberlein, Romain Lebarbenchon, Luc Le Magoarou, Nathan Souviraà -Labastie.

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 2014 hosted 260 visitors.

9.3.2. *Radio and press*

Nancy Bertin and Rémi Gribonval co-authored a popularization article about compressed acoustic holography (with L. Daudet and F. Ollivier). It was accepted by the journal "Pour la Science" for publication in 2015.

Nancy Bertin, Jules Espiau and Rémi Gribonval (with L. Daudet, A. Bordenave and Inria audiovisual department) participated in the realization of a pedagogical video presenting the ANR ECHANGE project results on acoustic holography <http://videotheque.inria.fr/videotheque/embed/25816>.

Jules Espiau and Pierre Machart were interviewed in a one-hour scientific radio show on Radio Prune (radio of the Nantes region) at the occasion of the "Journée Science et Musique".

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 Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n^o 7, pp. 1830–1840 [DOI : 10.1109/TASL.2010.2050716], <https://hal.inria.fr/inria-00541865>

- [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 Identification - Sparse Matrix-Factorisation via ℓ_1 -Minimisation*, in "IEEE Transactions on Information Theory", July 2010, vol. 56, n^o 7, pp. 3523-3539 [DOI : 10.1109/TIT.2010.2048466], <https://hal.archives-ouvertes.fr/hal-00541297>
- [7] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. *Wavelets on graphs via spectral graph theory*, in "Applied and Computational Harmonic Analysis", March 2011, vol. 30, n^o 2, pp. 129-150 [DOI : 10.1016/J.ACHA.2010.04.005], <https://hal.inria.fr/inria-00541855>
- [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 Cospase 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. BOURRIER. *Compressed sensing and dimensionality reduction for unsupervised learning*, Université Rennes 1, May 2014, <https://tel.archives-ouvertes.fr/tel-01023030>

Articles in International Peer-Reviewed Journals

- [13] H. BECKER, L. ALBERA, P. COMON, M. HAARDT, G. BIROT, F. WENDLING, M. GAVARET, C.-G. BÉNDAR, I. MERLET. *EEG extended source localization: Tensor-based vs. conventional methods*, in "NeuroImage", August 2014, vol. 96, pp. 143-57 [DOI : 10.1016/J.NEUROIMAGE.2014.03.043], <https://hal.archives-ouvertes.fr/hal-01011856>
- [14] A. BENICHOX, L. S. R. SIMON, E. VINCENT, R. GRIBONVAL. *Convex regularizations for the simultaneous recording of room impulse responses*, in "IEEE Transactions on Signal Processing", January 2014 [DOI : 10.1109/TSP.2014.2303431], <https://hal.inria.fr/hal-00934941>
- [15] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. *Convex Optimization Approaches for Blind Sensor Calibration using Sparsity*, in "IEEE Transactions on Signal Processing", September 2014, vol. 62, n^o 18, pp. 4847-4856 [DOI : 10.1109/TSP.2014.2342651], <https://hal.inria.fr/hal-00853225>

- [16] A. BOURRIER, M. E. DAVIES, T. PELEG, P. PÉREZ, R. GRIBONVAL.

Invited Conferences

- [27] M. CHAFII, J. PALICOT, R. GRIBONVAL. *A PAPR upper bound of generalized waveforms for multi-carrier modulation systems*, in "ISCCSP - 6th International Symposium on Communications, Control, and Signal Processing", Athènes, Greece, May 2014, pp. 461-464 [DOI : 10.1109/ISCCSP.2014.6877913], <https://hal-supelec.archives-ouvertes.fr/hal-01072519>

International Conferences with Proceedings

- [28] L. ALBERA, S. KITIĆ, N. BERTIN, G. PUY, R. GRIBONVAL. *Brain source localization using a physics-driven structured cospase representation of EEG signals*, in "2014 IEEE International Workshop on Machine Learning for Signal Processing", Reims, France, September 2014, 6 p. , <https://hal.archives-ouvertes.fr/hal-01027609>

- [29] R. GRIBONVAL .

European Conference of the International Federation for Medical and Biological Engineering", Dubrovnik, Croatia, September 2014, 4 p. , <https://hal.archives-ouvertes.fr/hal-01015798>

- [37] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Hearing behind walls: localizing sources in the room next door with cosparsity*, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014, <https://hal.inria.fr/hal-00904779>
- [38] M. SEIBERT, M. KLEINSTEUBER, R. GRIBONVAL, R. JENATTON, F. BACH. *On The Sample Complexity Of Sparse Dictionary Learning*, in "SSP 2014 - IEEE Workshop on Statistical Signal Processing", Jupiters, Gold Coast, Australia, IEEE, June 2014, This work was partially supported Cluster of Excellence CoTeSys funded by the German DFG. This work was partially supported by the EU FP7, SMALL project, FET-Open grant number 225913, <https://hal.inria.fr/hal-00990684>
- [39] M. SEIBERT, J. WÖRMANN, R. GRIBONVAL, M. KLEINSTEUBER. *Separable Cosparsity Analysis Operator Learning*, in "EUSIPCO 2014 - European Signal Processing Conference", Lisbonne, Portugal, September 2014, <https://hal.inria.fr/hal-01054124>
- [40] N. SOUVIRAÀ-LABASTIE, A. OLIVERO, E. VINCENT, F. BIMBOT. *Audio source separation using multiple deformed references*, in "Eusipco", Lisboa, Portugal, September 2014, <https://hal.inria.fr/hal-01017571>
- [41] N. SOUVIRAÀ-LABASTIE, E. VINCENT, F. BIMBOT. *Music separation guided by cover tracks: designing the joint NMF model*, in "ICASSP - 40th IEEE International Conference on Acoustics, Speech and Signal Processing", Brisbane, Australia, April 2015, forthcoming, <https://hal.archives-ouvertes.fr/hal-01108675>
- [42] M. TILLMANN, R. GRIBONVAL, E. PFETSCH. *Projection onto the Cosparsity Set is NP-Hard*, in "ICASSP14 - International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, IEEE, May 2014, <https://hal.inria.fr/hal-00802359>
- [43] L. WANG, L. ALBERA, A. KACHENOURA, H. SHU, L. SENHADJI. *CP decomposition of semi-nonnegative semisymmetric tensors based on QR matrix factorization*, in "The eighth IEEE Sensor Array and Multi-Channel Signal Processing Workshop", A Coruna, Spain, June 2014, 4 p. , <https://hal.archives-ouvertes.fr/hal-01012125>
- [44] L. WANG, A. KACHENOURA, L. ALBERA, A. KARFOUL, H. ZHONG SHU, L. SENHADJI. *Nonnegative Compression for Semi-Nonnegative Independent Component Analysis*, in "The eighth IEEE Sensor Array and Multi-Channel Signal Processing Workshop", A Coruna, Spain, June 2014, 4 p. , <https://hal.archives-ouvertes.fr/hal-01012136>

Conferences without Proceedings

- [45] V. EMIYA, A. BONNEFOY, L. DAUDET, R. GRIBONVAL. *Le problème d'échantillonnage compressif avec permutation des capteurs et une solution de type branch-and-bound*, in "Congrès ROADEF de la société Française de Recherche Opérationnelle et Aide à la Décision", Marseille, France, February 2015, <https://hal.inria.fr/hal-01100265>
- [46] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. *Apprentissage de dictionnaire pour les représentations parcimonieuses*, in "46e Journées de Statistique", Rennes, France, Société Française de Statistique, June 2014, This is an abstract of the full preprint available at <http://hal.inria.fr/hal-00918142/>, <https://hal.inria.fr/hal-01054122>

- [47] R. GRIBONVAL, P. MACHART. *Conciliation d'a priori sans préjugé*, in "46è Journées de Statistique", Rennes, France, June 2014, <https://hal.inria.fr/hal-01055789>

Scientific Popularization

- [48] C. BILEN, S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Sparse Acoustic Source Localization with Blind Calibration for Unknown Medium Characteristics*, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, <https://hal.inria.fr/hal-01060320>
- [49] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Audio Declipping by Cospase Hard Thresholding*, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, <https://hal.inria.fr/hal-00922497>
- [50] L. LE MAGOAROU, R. GRIBONVAL. *Strategies to learn computationally efficient and compact dictionaries*, August 2014, International Traveling Workshop on Interactions between Sparse models and Technology (iTWIST), <https://hal.inria.fr/hal-01010766>

Other Publications

- [51] C. BILEN, G. PUY, R. GRIBONVAL, L. DAUDET. *Balancing Sparsity and Rank Constraints in Quadratic Basis Pursuit*, March 2014, <https://hal.inria.fr/hal-00960272>
- [52] A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL. *Dynamic Screening: Accelerating First-Order Algorithms for the Lasso and Group-Lasso*, November 2014, <https://hal.archives-ouvertes.fr/hal-01084986>
- [53] A. BOURRIER, F. PERRONNIN, R. GRIBONVAL, P. PÉREZ, H. JÉGOU. *Explicit embeddings for nearest neighbor search with Mercer kernels*, January 2015, To appear in Journal of Mathematical Imaging and Vision, <https://hal.inria.fr/hal-00722635>
- [54] R. GRIBONVAL, R. JENATTON, F. BACH. *Sparse and spurious: dictionary learning with noise and outliers*, September 2014, This is a substantially revised version of a first draft that appeared as a preprint titled "Local stability and robustness of sparse dictionary learning in the presence of noise", <http://hal.inria.fr/hal-00737152>, <https://hal.archives-ouvertes.fr/hal-01025503>
- [55] R. GRIBONVAL, R. JENATTON, F. BACH, M. KLEINSTEUBER, M. SEIBERT. *Sample Complexity of Dictionary Learning and other Matrix Factorizations*, November 2014, forthcoming, <https://hal.inria.fr/hal-00918142>
- [56] L. LE MAGOAROU, R. GRIBONVAL. *Learning computationally efficient dictionaries and their implementation as fast transforms*, June 2014, <https://hal.inria.fr/hal-01010577>
- [57] L. LE MAGOAROU, R. GRIBONVAL. *Chasing butterflies: In search of efficient dictionaries*, January 2015, <https://hal.archives-ouvertes.fr/hal-01104696>
- [58] Y. SALAÜN, E. VINCENT, N. BERTIN, N. SOUVIRAÀ-LABASTIE, X. JAUREGUIBERRY, D. T. TRAN, F. BIMBOT. *The Flexible Audio Source Separation Toolbox Version 2.0*, May 2014, ICASSP, <https://hal.inria.fr/hal-00957412>

- [59] N. SOUVIRAÀ-LABASTIE, A. OLIVERO, E. VINCENT, F. BIMBOT. *Multi-channel audio source separation using multiple deformed references*, November 2014, <https://hal.inria.fr/hal-01070298>

References in notes

- [60] 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>
- [61] 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>
- [62] 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
- [63] D. BARRY, D. FITZ

- [71] 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>
- [72] 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° 5, pp. 2295-2306
- [73] J.-F. CARDOSO, A. SOULOUMIAC. *Blind beamforming for non-gaussian signals*, in "IEE Proceesings-F", December 1993, vol. 140, no. 6, pp. 362-370
- [74] J. CARROLL, J. CHANG. *Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition*, in "Psychometrika", 1970, n° 9, pp. 267-283
- [75] 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/>
- [76] F. CHAUMETTE, S. HUTCHINSON. *Visual servo control, Part I: Basic approaches*, in "IEEE Robotics and Automation Magazine", 2006, vol. 13, n° 4, pp. 82-90, <https://hal.inria.fr/inria-00350283>
- [77] P. CHEVALIER, A. CHEVREUIL. *Application to telecommunications*, in "Handbook of blind source separation", P. COMON, C. JUTTEN (editors), Academic Press, March 2010
- [78] N. DELFOSSE, P. LOUBATON. *Adaptive blind separation of independent sources: a deflation approach*, in "Signal Processing, Elsevier", 1995, vol. 45, pp. 59-83
- [79] L. DING. *Reconstructing cortical current density by exploring sparseness in the transform domain*, in "Physics in Medicine and Biology", 2009, vol. 54, pp. 2683-2697
- [80] 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
- [81] N. DUONG, E. VINCENT, R. GRIBONVAL. *Spatial location priors for Gaussian model-based reverberant audio source separation*, Inria, September 2012, n° RR-8057, <http://hal.inria.fr/hal-00727781>
- [82] 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>
- [83] R. GIRYES, S. NAM, M. ELAD, R. GRIBONVAL, M. E. DAVIES. *Greedy-Like Algorithms for the Cospase 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. *Reconciling "priors" and "priors" without prejudice ?*, in "international biomedical and astronomical signal processing (BASP) Frontiers workshop", Villars-sur-Ollon, Switzerland, 2013, international biomedical and astronomical signal processing (BASP) Frontiers workshop, <http://hal.inria.fr/hal-00751996>
- [87] R. GRIBONVAL, P. MACHART. *Reconciling "priors" & "priors" without prejudice? (research report)*, Inria, September 2013, n^o RR-8366, <http://hal.inria.fr/hal-00862925>
- [88] 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>
- [89] R. GRIBONVAL, K. SCHNASS. *Dictionary identifiability from few training samples*, in "Proc. European Conf. on Signal Processing - EUSIPCO", August 2008
- [90] 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
- [91] 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
- [92] 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
- [93] A. HYVARINEN, E. OJA. *A fast fixed-point algorithm for independent component analysis*, in "Neural Computation", 1997, vol. 9, n^o 7, pp. 1483-1492
- [94] N. ITO. *Robust microphone array signal processing against diffuse noise*, University of Tokyo, January 2012, <http://hal.inria.fr/tel-00691931>
- [95] 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>
- [96] C. JORDAN. *Mémoire sur les formes bilinéaires*, in "Journal de mathématiques pures et appliquées 2e série, tome 19", 1874, pp. 35-54
- [97] S. KITIC, N. BERTIN, R. GRIBONVAL. *A review of cosparse signal recovery methods applied to sound source localization*, in "Le XXIVe colloque Grets", Brest, France, September 2013, <http://hal.inria.fr/hal-00838080>
- [98] 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

- [99] P. MCCULLAGH. *Tensor Methods in Statistics*, Chapman and Hall, Monographs on Statistics and Applied Probability, 1987
- [100] S. NAM, M. E. DAVIES, M. ELAD, R. GRIBONVAL. *The Cospase 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>
- [101] S. NAM, R. GRIBONVAL. *Physics-driven structured cospase 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>
- [102] 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>
- [103] 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
- [104] R. D. PASCUAL-MARQUI. *Standardized low resolution brain electromagnetic tomography (sLORETA): technical details*, in "Methods and Findings in Experimental and Clinical Pharmacology", 2002
- [105] 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, to appear in EUSIPCO 2013, <http://hal.inria.fr/hal-00812858>
- [106] 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>
- [107] 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
- [108] M. RAJH, P. COMON, R. A. HARSHMAN. *Enhanced Line Search: A Novel Method to Accelerate PARAFAC*, in "SIAM Journal in Matrix Analysis and Applications", 2008
- [109] 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
- [110] 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>
- [111] 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

-
- [112] 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
- [113] A. TONAZZINI, L. BEDINI, E. SALERNO. *Independent component analysis for document restoration*, in "International Journal on Document Analysis and Recognition IJDAR", April 2004, n^o 7, pp. 17–27
- [114] L. TUCKER. *Some mathematical notes on three-mode factor analysis*, in "Psychometrika", September 1966, vol. 31, n^o 3, pp. 279-311
- [115] 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
- [116] V. ZARZOSO, P. COMON. *Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast With Algebraic Optimal Step Size*, in "IEEE Transactions on Neural Networks", February 2010, vol. 21, n^o 2, pp. 248-261