IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2016

Project-Team PANAMA

Parcimonie et Nouveaux Algorithmes pour le Signal et la Modélisation Audio

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Language, Speech and Audio

Table of contents

Project-Team PANAMA

Creation of the Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

1.2.6. - Sensor networks
3.1.1. - Modeling, representation
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.4.5. - Bayesian methods
3.4.6. - Neural networks
3.4.7. - Kernel methods
3.4.8. - Deep learning
3.5.1. - Analysis of large graphs
5.3.2. - Sparse modeling and image representation
5.7.1. - Sound
5.7.2. - Music
5.7.3. - Speech
5.7.4. - Analysis
5.9.1. - Sampling, acquisition
5.9.2. - Estimation, modeling
5.9.3. - Reconstruction, enhancement
5.9.4. - Signal processing over graphs
5.9.5. - Sparsity-aware processing
5.9.6. - Optimization tools
5.10.2. - Perception
5.11.2. - Home/building control and interaction
6.1.4. - Multiscale modeling
6.2.5. - Numerical Linear Algebra
6.2.6. - Optimization
6.3.1. - Inverse problems
6.3.2. - Data assimilation
7.8. - Information theory
7.9. - Graph theory

Other Research Topics and Application Domains:

1.3. - Neuroscience and cognitive science
2.5.1. - Sensorimotor disabilities
2.6. - Biological and medical imaging
5.6. - Robotic systems
5.8. - Learning and training

6.3.3. - Network Management
8.1.2. - Sensor networks for smart buildings
8.4. - Security and personal assistance
9.1. - Education
9.2.1. - Music, sound
9.2.2. - Cinema, Television
9.2.3. - Video games
9.6. - Reproducibility
9.9.1. - Environmental risks

1. Members

Research Scientists

Rémi Gribonval [Team leader, Senior Researcher, Inria, HDR]
Nancy Bertin [Researcher, CNRS]
Frédéric Bimbot [Senior Researcher, CNRS, HDR]
Antoine Deleforge [Researcher, Inria]

Engineers

Ewen Camberlein [CNRS]
Romain Lebarbenchon [CNRS]
Nicolas Bellot [Inria, granted by FP7 ERC PLEASE project]
Adrien Leman [Inria, from Feb 2016, granted by FP7 ERC PLEASE project]

PhD Students

Corentin Guichaoua [Univ. Rennes I]
Luc Le Magoarou [Inria, granted by FP7 ERC PLEASE project]
Nicolas Keriven [Univ. Rennes I, Contrat Doctoral Polytechnicien]
Himalaya Jain [Technicolor, granted by CIFRE Technicolor]
Clément Gaultier [Inria, granted by FP7 ERC PLEASE project and Conseil Régional Bretagne]
Corentin Louboutin [Univ. Rennes I (ENS Rennes)]
Nathan Libermann [Inria, from Oct 2016, granted by ANR DEEPMUS project and Conseil Régional Bretagne]
Cassio Fraga Dantas [Inria, from Oct 2016]
Mohamed Hafsati [IRT b<>com, granted by ANR INVATE project, from Dec 2016]

Post-Doctoral Fellows

Gilles Puy [Inria, until Jan 2016, granted by FP7 ERC PLEASE project]
Nicolas Tremblay [Inria, until Aug 2016]
Yann Traonmilin [Inria, granted by FP7 ERC PLEASE project]
Srdan Kitic [Inria, until Mar 2016, granted by FP7 ERC PLEASE project]

Visiting Scientist

Pierre Vandergheynst [EPFL, Chaire Internationale Inria position, Jul 2015]

Administrative Assistants

Stéphanie Lemaile [Inria]
Armelle Mozziconacci [CNRS, from May 2016]

Others

Saurabh Kataria [intern, from May 2016 until Jul 2016]
Igal Rozenberg [intern, from Apr 2016 until Jul 2016]
Guillaume Versal [intern, from Jun 2016 until Aug 2016]

2. Overall Objectives

2.1. Overall positioning

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

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

2.2. Scientific foundations

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

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

2.3. Applications

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

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

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, encompassing 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. Building on this existing framework, a first objective of PANAMA 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 need to 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.

3.2.3. Robust Audio Source Localization

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. Knowing the geometry of an audio scene is often a pre-requisite to perform higher-level tasks such as speaker identification and tracking, speech enhancement and recognition or audio source separation. It can be decomposed into two sub-tasks : (i) compute spatial auditory features from raw audio input and (ii) map these features to the desired spatial information. Robustly addressing both these aspects with a limited number of microphones, in the presence of noise, reverberation, multiple and possibly moving sources remains a key challenge in audio signal processing. The first aspect will be tackled by both advanced statistical and acoustical modeling of spatial auditory features. The second one will be addressed by two complementary approaches. *Physics-driven* approaches cast sound source localization as an inverse problem given the known physics of sound propagation within the considered system. *Data-driven* approaches aim at learning the desired feature-to-source-position mapping using real-world or synthetic training datasets adapted to the problem at hand. Combining these approaches should allow a widening of the notion of source localization, considering problems such as the identification of the directivity or diffuseness of the source as well as some of the boundary conditions of the room. 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 need to 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.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Antoine Deleforge (new PANAMA team member), Florence Forbes (MISTIS team) and Radu Horaud (PERCEPTION team) received the 2016 Hojjat Adeli Award for Outstanding Contributions in Neural Systems for their paper [70].

A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound-source separation and localization on binaural manifolds," *International journal of neural systems*, vol. 25, no. 01, 2015, https://hal.archives-ouvertes.fr/hal-00960796

The Award for Outstanding Contributions in Neural Systems established by World Scientific Publishing Co. in 2010, is awarded annually to the most innovative paper published in the previous volume/year of the International Journal of Neural Systems.

6. New Software and Platforms

6.1. Audio Activity Detector

KEYWORD: Audio activity estimation

- Authors: Frédéric Bimbot, Ewen Camberlein, Romain Lebarbenchon and Vincent Soupe
- Contact: Frédéric Bimbot

6.2. Audio Breath Rhythm Estimator

KEYWORD: Breath rhythm estimation

- Authors: Frédéric Bimbot, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.3. Audio GMM Classifier

- Authors: Frédéric Bimbot, Vincent Soupe, Jérémy Paret, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.4. CSCbox

Compressive Spectral Clustering Toolbox
KEYWORD: Clustering
SCIENTIFIC DESCRIPTION

The Compressive Spectral Clustering Toolbox is a Matlab toolbox implementing routines to reproduce experiments from the paper "Compressive Spectral Clustering", by N. Tremblay, G. Puy, P. Vandergheynst and R. Gribonval.
FUNCTIONAL DESCRIPTION

Matlab toolbox implementing routines to reproduce experiments from the paper "Compressive Spectral Clustering"

- Authors: Nicolas Tremblay, Gilles Puy, Pierre Vandergheynst and Rémi Gribonval
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Rémi Gribonval
- URL: http://www.irisa.fr/panama/software

6.5. FASST2

Flexible Audio Source Separation Toolbox
KEYWORDS: Audio - Source Separation
SCIENTIFIC DESCRIPTION

Only source separation software publicly availble allowing to use both spacial and spectral source properties with a generalised EM algorithm (expectation - maximisation). Fast specification of each use case by the choice of suitable constraints in constraint libraries.
FUNCTIONAL DESCRIPTION

Toolbox for the fast design of audio source separation adapted to any use case.

- Participants: Emmanuel Vincent and Yann Salaun
- Contact: Emmanuel Vincent
- URL: http://fasst.gforge.inria.fr

6.6. FAuST

KEYWORDS: Learning - Sparsity - Fast transform - Multilayer sparse factorisation
FUNCTIONAL DESCRIPTION

C++ toolbox, designed to decompose a given dense matrix into a product of sparse matrices in order to reduce its computational complexity (both for storage and manipulation).

- Authors: Luc Le Magoarou, Rémi Gribonval, Adrien Leman, Nicolas Bellot and Thomas Gautrais
- Contact: Rémi Gribonval
- URL: http://faust.gforge.inria.fr/

6.7. Multi-channel BSS Locate Basic

KEYWORDS: Audio - Localization - Signal processing - Multichannel signal

SCIENTIFIC DESCRIPTION

Multi-Channel BSS Locate is a Matlab toolbox to estimate Direction Of Arrival (expressed both in azimuth and elevation) of multiple sources in a multi-channel audio signal recorded by an array of microphones. This toolbox implements the previous 8 angular spectrum methods presented in BSS Locate (GCC-PHAT, GCC-NONLIN, MUSIC and several SNR-based spectra).

- Authors: Emmanuel Vincent, Charles Blandin, Alexey Ozerov, Ewen Camberlein, Romain Lebarbenchon, Frédéric Bimbot and Nancy Bertin
- Contact: Emmanuel Vincent
- URL: http://bass-db.gforge.inria.fr/bss_locate/

6.8. SPADE

Sparse Audio Declipper
KEYWORDS: Audio - Sparse regularization - Declipping
SCIENTIFIC DESCRIPTION

Matlab routines to reproduce audio declipping experiments from the paper: - Srdan Kitic, Nancy Bertin, Remi Gribonval. Audio Declipping by Cosparse Hard Thresholding. iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, Aug 2014, Namur, Belgium. <hal-00922497v3> - Srdan Kitic, Nancy Bertin, Remi Gribonval. Sparsity and cosparsity for audio declipping: a flexible non-convex approach. LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation, Aug 2015, Liberec, Czech Republic. pp.8. <hal-01159700v2>

- Participants: Srdan Kitic, Nancy Bertin and Rémi Gribonval
- Contact: Rémi Gribonval
- URL: http://xspaad.gforge.inria.fr/

6.9. SPOD Audio

KEYWORDS: Audio source classification - Speaker verification - Breath rhythm estimation - Audio activity estimation

- Authors: Frédéric Bimbot, Vincent Soupe, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot
- URL: http://www.kerlink.fr/en/

6.10. SPOD Model Generation

KEYWORDS: Machine learning - Audio source classification - Statistical modeling - Speaker verification

- Authors: Frédéric Bimbot, Vincent Soupe, Jérémy Paret, Ewen Camberlein and Romain Lebarbenchon
- Contact: Frédéric Bimbot

6.11. SRP-PHAT

KEYWORD: Source localization

- Authors: Frédéric Bimbot, Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Charles Blandin and Alexey Ozerov
- Contact: Frédéric Bimbot

6.12. SketchMLBox

KEYWORD: Clustering

SCIENTIFIC DESCRIPTION

The SketchMLbox is a Matlab toolbox for fitting mixture models to large collections of training vectors using sketching techniques. The collection is first compressed into a vector called sketch, then a mixture model (e.g. a Gaussian Mixture Model) is estimated from this sketch using greedy algorithms typical of sparse recovery. The size of the sketch does not depend on the number of elements in the collection, but rather on the complexity of the problem at hand [2,3]. Its computation can be massively parallelized and distributed over several units. It can also be maintained in an online setting at low cost. Mixtures of Diracs ("K-means") and Gaussian Mixture Models with diagonal covariance are currently available, the toolbox is structured so that new mixture models can be easily implemented.

FUNCTIONAL DESCRIPTION

Matlab toolbox for fitting mixture models to large collections of feature vectors using sketching techniques.

- Authors: Nicolas Keriven, Rémi Gribonval and Nicolas Tremblay
- Partner: Université de Rennes 1
- Contact: Rémi Gribonval
- URL: http://sketchml.gforge.inria.fr

6.13. VoiceHome Corpus

KEYWORDS: Audio - Source Separation

FUNCTIONAL DESCRIPTION

This corpus includes reverberated, noisy speech signals spoken by native French talkers in a lounge and recorded by an 8-microphone device at various angles and distances and in various noise conditions. Room impulse responses and noise-only signals recorded in various real rooms and homes and baseline speaker localization and enhancement software are also provided.

- Contact: Nancy Bertin
- URL: http://voice-home.gforge.inria.fr/voiceHome_corpus.html

6.14. graphsamplingbox

- Authors: Nicolas Tremblay, Gilles Puy, Pierre Vandergheynst and Rémi Gribonval
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Rémi Gribonval
- URL: http://www.irisa.fr/panama/software

7. New Results

7.1. Recent results on Sparse Representations, Inverse Problems, and Dimension Reduction

Sparsity, low-rank, dimension-reduction, inverse problem, sparse recovery, scalability, compressive sensing

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems, and dimension reduction, which is at the core of the ERC project PLEASE (Projections, Learning and Sparsity for Efficient Data Processing, see Section 9.2.1.1).

7.1.1. Theoretical results on Sparse Representations, Graph Signal Processing, and Dimension Reduction

Participants: Rémi Gribonval, Yann Traonmilin, Gilles Puy, Nicolas Tremblay, Pierre Vandergheynst.

Main collaboration: Mike Davies (University of Edinburgh), Pierre Borgnat (ENS Lyon), and members of the LTS2 lab of Pierre Vandergheynst at EPFL

Stable recovery of low-dimensional cones in Hilbert spaces: Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the ℓ_1-norm or low rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). We established generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. These guarantees were illustrated on selected examples. For block structured sparsity in the infinite dimensional setting, we used the guarantees for a family of regularizers which efficiency in terms of RIP constant can be controlled, leading to stronger and sharper guarantees than the state of the art. This has been published in a journal paper [21].

Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$: We considered the problem of constructing a linear map from a Hilbert space (possibly infinite dimensional) to $\mathbb{R}^m$ that satisfies a restricted isometry property (RIP) on an arbitrary signal model set. We obtained a generic framework that handles a large class of low-dimensional subsets but also *unstructured* and *structured* linear maps. We provided a simple recipe to prove that a random linear map satisfies a general RIP on the model set with high probability. We also described a generic technique to construct linear maps that satisfy the RIP. Finally, we detailed how to use our results in several examples, which allow us to recover and extend many known compressive sampling results. This has been presented at the conference EUSIPCO 2015 [90], and a journal paper is under revision [91].

Signal processing on graphs: from filtering to random sampling and robust PCA: Graph signal processing is an emerging field aiming at extending classical tools from signal processing (1D time series) and image processing (2D pixel grids, 3D voxel grids) to more loosely structured numerical data: collections of numerical values each associated to a vertex of a graph, where the graph encodes the underlying “topology” of proximities and distances. Since our pioneering contributions on this topic [4], the team regularly works on various aspects of graph signal processing, in collaboration with the LTS2 lab of Pierre Vandergheynst at EPFL. This year, we studied the problem of sampling k-bandlimited signals on graphs. We proposed two sampling strategies that consist in selecting a small subset of nodes at random. The first strategy is non-adaptive, i.e., independent of the graph structure, and its performance depends on a parameter called the graph coherence. On the contrary, the second strategy is adaptive but yields optimal results. Indeed, no more than $O(klog(k))$ measurements are sufficient to ensure an accurate and stable recovery of all k-bandlimited signals. This second strategy is based on a careful choice of the sampling distribution, which can be estimated quickly. Then, we proposed a computationally efficient decoder to reconstruct k-bandlimited signals from their samples. We proved that it yields accurate reconstructions and that it is also stable to noise. Finally, we conducted several experiments to test these techniques. A journal paper has been published [17] accompanied by a toolbox for reproducible research (see Section 6.14). Other contributions from this year on the topic of graph signal processing include new subgraph-based filterbanks for graph signals [22], and new accelerated and robustified techniques for PCA on graphs [19], [20] (see also below our contributions in terms of new algorithms to obtain approximate Fast Graph Fourier Transforms [32], [53]).

Accelerated spectral clustering: We leveraged the proposed random sampling technique to propose a faster spectral clustering algorithm. Indeed, classical spectral clustering is based on the computation of the first k eigenvectors of the similarity matrix' Laplacian, whose computation cost, even for sparse matrices, becomes prohibitive for large datasets. We showed that we can estimate the spectral clustering distance matrix without computing these eigenvectors: by graph filtering random signals. Also, we took advantage of the stochasticity of these random vectors to estimate the number of clusters k. We compared our method to classical spectral clustering on synthetic data, and showed that it reaches equal performance while being faster by a factor at least two for large datasets of real data. Two conference papers have been presented, at ICASSP 2016 [39] and ICML 2016 [40] and a toolbox for reproducible research has been released (see Section 6.4).

7.1.2. An Alternative Framework for Sparse Representations: Sparse “Analysis” Models

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Clément Gaultier.

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.

Building on our pioneering work on the cosparse model [7] [73], [87] successful applications of this approach to sound source localization, audio declipping and brain imaging have been developed in 2015 and 2016. In addition, new applications to audio denoising were also introduced this year.

Versatile cosparse regularization: Digging the groove of previous years' results (comparison of the performance of several cosparse recovery algorithms in the context of sound source localization [77], demonstration of its efficiency in situations where usual methods fail ([79], see paragraph 7.4.2), applicability to the hard declipping problem [78], application to EEG brain imaging [56]), a journal paper embedding the latest algorithms and results in sound source localization and brain source localization in a unified fashion was published this year [5]. This framework was also exploited to extend results on audio inpainting (see Section 7.3.2).

New results include experimental confirmation of robustness and versatility of the proposed scheme, and of its computational merits (convergence speed increasing with the amount of data). In a work presented in a workshop [44], we also proposed a multiscale strategy that aims at exploiting computational advantages of both sparse and cosparse regularization approaches, thanks to the simple yet effective all-zero initialization which the synthesis-based optimization can benefit from, while retaining the computational properties of the analysis-based approach for huge scale optimization problems arising in physics-driven settings.

Parametric operator learning for cosparse calibration: In many inverse problems, a key challenge is to cope with unknown physical parameters of the problem such as the speed of sound or the boundary impedance. In the sound source localization problem, we previously showed that the unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions [58], [81]. This year, we extended the formulation to the case of unknown boundary impedance, and showing that a similar biconvex formulation and optimization could solve this new problem efficiently (conference paper published in ICASSP 2016 [29], see also Section 7.3.3).

7.1.3. Algorithmic and Theoretical results on Computational Representation Learning

Participants: Rémi Gribonval, Luc Le Magoarou, Nicolas Bellot, Adrien Leman, Cassio Fraga Dantas, Igal Rozenberg.

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, 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 together with their fast implementation. The principle and its application to image denoising appeared at ICASSP 2015 [84] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [83]. A Matlab library has been released (see Section 6.6) to reproduce the experiments from the comprehensive journal paper published this year [16], which additionally includes theoretical results on the improved sample complexity of learning such dictionaries. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou on this topic [85].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms. A conference paper on this latter topic appeared in ICASSP 2016 [32], and a journal paper has been submitted [53] where we empirically show that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opens the way to substantial accelerations for Fourier Transforms on large graphs.

A C++ software library has been developed (see Section 6.6) to release the resulting algorithms.

7.2. Activities on Waveform Design for Telecommunications

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

7.2.1. Characterizing and designing multi-carrier waveform systems with optimum PAPR

Participant: Rémi Gribonval.

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

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 9.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii [64], defended in October this year and 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 [67] and several conference communications [65], [66]. Our characterization of waveforms with optimum PAPR [68] has been published in a journal this year [14]. The work this year has concentrated on designing new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR. This has given rise to a patent [49] and a journal paper is in preparation.

7.3. Emerging activities on Compressive Learning and Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Phase Estimation in Multichannel Mixtures

Participants: Antoine Deleforge, Yann Traonmilin.

The problem of estimating source signals given an observed multichannel mixture is fundamentally ill-posed when the mixing matrix is unknown or when the number of sources is larger that the number of microphones. Hence, prior information on the desired source signals must be incorporated in order to tackle it. An important line of research in audio source separation over the past decade consists in using a model of the source signals' magnitudes in the short-time Fourier domain [8]. Such models can be inferred through, *e.g.*, non-negative matrix factorization [89] or deep neural networks [88]. Magnitudes estimates are often interpreted as instantaneous variances of Gaussian-process source signals, and are combined with Wiener filtering for source separation. In [50], we introduced a shift of this paradigm by considering the *Phase Unmixing* problem: how can one recover the instantaneous phases of complex mixed source signals when their magnitudes and mixing matrix are known? This problem was showed to be NP-hard, and three approaches were proposed to tackle it: a heuristic method, an alternate minimization method, and a convex relaxation into a semi-definite program. The last two approaches were showed to outperform the oracle multichannel Wiener filter in under-determined informed source separation tasks. The latter yielded best results, including the potential for exact source separation in under-determined settings.

7.3.2. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. Building upon our previous pioneering 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 [55]), we proposed over the last two years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [78], [80], [81]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [80], together with the associated toolbox for reproducible research (see Section 6.8) draw the attention of a world leading company in professional audio signal processing, with which some transfer has been negotiated. In 2016, real-time implementation of the A-SPADE algorithm was obtained and demonstrated at various events (HCERES evaluation, TechnofÃ©rence # 18 « Nouvelles expériences son et vidéo », ...).

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, denoising, time-frequency inpainting, joint source separation and declipping). In particular, we investigated the incorporation of the so-called "social" structure constraint [82] into problems regularized by a cosparse prior, including declipping and denoising. Publication of this work is currently under preparation.

7.3.3. Blind Calibration of Impedance and Geometry

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Main collaborations: Laurent Daudet, Thibault Nowakowski, Julien de Rosny (Institut Langevin)

Last year, we also investigated extended inverse problem scenarios where a "lack of calibration" may occur, i.e., when some physical parameters are needed for reconstruction but apriori unknown: speed of sound, impedance at the boundaries of the domain where the studied phenomenon propagates, or even the shape of these boundaries. In a first approach, based on our physics-driven cosparse regularization of the sound source localization problem [5] (see section 7.1.2), we managed to preserve the sound source localization performance when the speed of sound is unknown, or, equally, when the impedance is unknown, provided the shape is and under some smoothness assumptions. Unlike the previous case (gain calibration), the arising problems are not convex but biconvex, and can be solved with proper biconvex formulation of ADMM algorithm. In a second approach based on eigenmode decomposition (limited to a 2D membrane), we showed that impedance learning with known shape, or shape learning with known impedance can be expressed as two facets of the same problem, and solved by the same approach, from a small number of measurements. Two papers presenting these two sets of results appeared at ICASSP 2016 [29], [37].

7.3.4. Sketching for Large-Scale Mixture Estimation

Participants: Rémi Gribonval, Nicolas Keriven.

Main collaborations: Patrick Perez (Technicolor R&I France) Anthony Bourrier (formerly Technicolor R&I France, then GIPSA-Lab)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [60] 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 showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. The proposed algorithm provided good reconstruction and scaled to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of voluminous datasets. It also provided a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum it is based on [63], [61], [62]. Last year, we consolidated our extensions to non-isotropic Gaussians, with new algorithms [76] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper appeared at ICASSP 2016 [31] and a journal version has been submitted [52], accompanied by a toolbox for reproducible research (see Section 6.12).

This year the work concentrated on extending the approach beyond the case of Gaussian Mixture Estimation. First, we showed empirically that the algorithm can be adapted to sketch a training collection while still allowing to compute clusters. The approach, called "Compressive K-means", is described in a paper accepted at ICASSP 2017 [27]. Then, we expressed a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework already covers compressive K-means as well as compressive Principal Component Analysis (PCA), and a conference paper has been submitted. A comprehensive journal paper is under preparation, and future work will include expliciting the impact of the proposed framework on a wider set of concrete learning problems.

7.4. Source Separation and Localization

Source separation, sparse representations, probabilistic model, source localization

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 [8], which generalizes a number of existing techniques including our former study on spectral GMMs [57]. 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 [71], [75], [74], knowledge on the source positions and room characteristics [72], or a better initialization of parameters thanks to specific source localization techniques [59].

This accumulated progress lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [92] and we published an overview paper on recent and going research along the path of *guided* separation in a special issue of IEEE Signal Processing Magazine devoted to source separation and its applications [10]. This two achievements formed the basis of our work in 2016, exploring intensively the concrete use of these tools and principles in real-world scenarios, in particular within the voiceHome project (see Section 6.13).

7.4.1. Towards Real-world Separation and Remixing Applications

Participants: Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Romain Lebarbenchon.

In 2015, we began a new industrial collaboration, in the context of the VoiceHome project, aiming at another challenging real-world application: natural language dialog in home applications, such as control of domotic and multimedia devices. As a very noisy and reverberant environment, home is a particularly challenging target for source separation, used here as a pre-processing for speech recognition (and possibly with stronger interactions with voice activity detection or speaker identification tasks as well). In 2016, we publicly released a realistic corpus of room impulse responses and utterances recorded in real homes, and presented it during the Interspeech conference [28]. We also continued benchmarking and adapting existing localization and separation tools to the particular context of this application, worked on a better interface between source localization and source separations steps, and investigated new means to reduce the latency and computational burden of the currently available tools (low-resolution source separation preserving speech recognition improvement, automatic selection of the best microphones, joint localization and multichannel speech / non speech classification prior to any separation).

In november 2016, we started investigating a new application of source separation to sound respatialization from Higher Order Ambisonics (HOA) signals, in the context of free navigation in 3D audiovisual contents. This work is conducted in a collaboration with the IRT b<>Com, through the Ph.D. of Mohammed Hafsati (co-supervised by Nancy Bertin, RÃ©mi Gribonval).

7.4.2. Implicit Localization through Audio-based Control for Robotics

Participant: Nancy Bertin.

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*) [59]. 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 (cosparse modeling of the acoustic field, see Section 7.1.2).

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 investigate over the last three years 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 a 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 [69].

Last year, we established an analytical model linking the Interaural Time Difference (ITD) 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 and experiments with an actual robot [86]. This year, we consolidated these results and extended the range of applicative tasks [36] and obtained similar results (including theoretical and experimental) for the Interaural Level Difference (ILD), in combination with the absolute energy level [34]. Another set of experiments, presented during the IROS workshop [35] was successfully carried with a humanoid robot, notably without any measurement nor modeling of the robot's Head Relative Transfer Functions (HRTF). This work was mainly lead by Aly Magassouba, who defended his Ph.D. (co-supervised by Nancy Bertin and François Chaumette) in December 2016.

7.4.3. Emerging activities on Virtually-Supervised Sound Localization

Participants: Antoine Deleforge, Clément Gaultier, Saurabh Kataria.

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. It can be decomposed into two sub-tasks : (i) computing spatial auditory features from raw audio input and (ii) mapping these features to the desired spatial information.

Extracting spatial features from raw audio input: The most commonly used features in binaural (two microphones) sound source localization are frequency-dependent phase and level differences between the two microphones. To handle the presence of noise, several sources, or reverberation, most existing methods rely on some kind of aggregation of these features in the time-frequency plane, often in a heuristic way. In [25], we introduced the rectified binaural ratio as a new spatial feature. We showed that for Gaussian point-source signals corrupted by stationary Gaussian noise, this ratio follows a complex t-distribution with explicit parameters. This new formulation provides a principled, statistically sound and efficient method to aggregate these features in the presence of noise. Experiments notably showed the higher robustness of these features compared to traditional ones, in the task of localizing heavily corrupted speech signals.

Mapping features to spatial information: Existing methods to map auditory features to spatial properties divide into two categories. *Physics-driven* methods attempt to estimate an explicit mapping based on an approximate physical model of sound propagation in the considered system. *Data-driven* methods bypass the use of a physical model by learning the mapping from a training set, obtained by manually annotating features extracted from real data. We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually-supervised acoustic space mapping* [26], [51]. The idea is to use a physics-based room-acoustic simulator to generate arbitrary large datasets of room-impulse responses corresponding to various acoustic environments, adapted to the physical audio system at hand. We demonstrated that mappings learned from these data could potentially be used to not only estimate the 3D position of a source but also some acoustical properties of the room [51]. We also showed that a virtually-learned mapping could robustly localize sound sources from real-world binaural input, which is the first result of this kind in audio source localization [26].

7.5. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model

Current work developed in our research group in the domain of music content processing and information retrieval explore various information-theoretic frameworks for music structure analysis and description [24], in particular the System & Contrast model [1].

7.5.1. Tensor-based Representation of Sectional Units in Music

Participants: Corentin Guichaoua, Frédéric Bimbot.

Following Kolmogorov'ÂÂs complexity paradigm, modeling the structure of a musical segment can be addressed by searching for the compression program that describes as economically as possible the musical content of that segment, within a given family of compression schemes.

In this general framework, packing the musical data in a tensor-derived representation enables to decompose the structure into two components : (i) the shape of the tensor which characterizes the way in which the musical elements are arranged in an n-dimensional space and (ii) the values within the tensor which reflect the content of the musical segment and minimize the complexity of the relations between its elements.

This approach is currently developed and tested for the grouping of chord sequences into sectional units for pop music songs, with very encouraging segmentation results on pop songs.

7.5.2. Minimal Transport Graphs for the Modeling of Chord Progressions

Participants: Corentin Louboutin, Frédéric Bimbot.

In this work, we model relations between chords by minimal transport and we investigate different types of dependencies within chord sequences [33]. For this purpose we use the ÂÂSystem & Contrast (S&C) model [1], designed for the description of music sectional units, to infer non-sequential structures called chord progression graphs (CPG).

Minimal transport is defined as the shortest displacement of notes, in semitones, between a pair of chords. The paper [33] present three algorithms to find CPGs for chords sequences: one is sequential, and two others are based on the S& C model. The three methods are compared using the perplexity as an efficiency measure.

The experiments on a corpus of 45 segments taken from songs of multiple genres indicate that optimization processes based on the S&C model outperform the sequential model with a decrease in perplexity over 1.0.

7.5.3. Regularity Constraints for the Fusion of Music Structure Segmentation System

Participant: Frédéric Bimbot.

Main collaborations Gabriel Sargent (EPI LinkMedia, Rennes, France)

Music structure estimation has recently emerged as a central topic within the field of Music Information Retrieval. Indeed, as music is a highly structured information stream, knowledge of how a music piece is organized represents a key challenge to enhance the management and exploitation of large music collections.

Former work carried out in our group [9] has illustrated the benefits that can be expected from a regularity constraint on the structural segmentation of popular music pieces : a constraint which favors structural segments of comparable size provides a better conditioning of the boundary estimation process.

As a further investigation, we have explored the benefits of the regularity constraint as an efficient way for combining the outputs of a selection of systems presented at MIREX between 2010 and 2015. These experiments have yielded a level of performance which is competitive to that of the state-of-the-art on the "MIREX10" dataset (100 J-Pop songs from the RWC database) [18].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Licensing agreement contract with Cedar Audio Limited

Participants: Nancy Bertin, Srdan Kitic, Rémi Gribonval.

This contract aimed at licensing an audio desaturation (declipping) software developed in the team.

8.2. Bilateral Grants with Industry

8.2.1. CIFRE contract with Technicolor R&I France on Very large scale visual comparison

Participants: Rémi Gribonval, Himalaya Jain.

Duration: 3 years (2015-2018)
Research axis: 3.1.2
Partners: Technicolor R&I France, Inria-Rennes
Funding: Technicolor R&I France, ANRT

The grand goal of this thesis is to design, analyze and test new tools to allow large-scale comparison of high-dimensional visual signatures. Leveraging state of the art visual descriptors, the objective is to obtain new compact codes for visual representations, exploiting sparsity and learning, so that they can be stored and compared in an efficient, yet meaningful, way.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Labex Comin Labs projects

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

9.1.1.1. HEMISFER

Participant: Rémi Gribonval.

Acronym: HYBRID (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation)
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, Ã¢ÂÂ¦).
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.

9.1.1.2. TEPN

Participant: Rémi Gribonval.

Acronym: TEPN (Toward Energy Proportional Networks)

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).

9.1.2. ANR INVATE project with IRT b<>com France

Participants: Rémi Gribonval, Nancy Bertin, Mohamed Hafsati.

Thesis on 3D audio scene decomposition for interactive navigation

Duration: 3 years (2016-2019)

Research axis: 3.2.2

Partners: IRT b<>com, Inria-Rennes, IRISA

Funding: ANR INVATE project (PIA)

The objective of this thesis is to develop tools to analyze audio scenes in order to identify, locate, and extract the sources present in the scene to re-spatialize them according to the user head orientation and the movement of the user in the targeted virtual scene.

9.1.3. OSEO-FUI: voiceHome

Participants: Nancy Bertin, Frédéric Bimbot, Romain Lebarbenchon, Ewen Camberlein.

Duration: 3 years (2015-2017)

Research axis: 3.2

Partners: onMobile, Delta Dore, eSoftThings, Orange, Technicolor, LOUSTIC, Inria Nancy

Coordinator: onMobile

Description: The goal of the project is to design and implement a multi-channel voice interface for smart home and multimedia (set-top-box) appliances.

Contributions of PANAMA are focused on (i) audio activity monitoring and wake-up word detection and (ii) audio source localization and separation. In both cases, the issue of energy frugality is central and strongly constrains the available resources. We expect from this cooperation to make progress towards operational low-resource audio source separation schemes and we intend to investigate compressive sensing for the characterization of audio and voice activity.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

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

Participants: Rémi Gribonval, Srdan Kitic, Luc Le Magoarou, Nancy Bertin, Nicolas Keriven, Yann Traonmilin, Gilles Puy, Adrien Leman, Nicolas Bellot.

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/

9.3. International Initiatives

9.3.1. Inria International Partners

9.3.1.1. 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, and the Institute for Mathematics of the Postdam University.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Pierre Vandergheynst, in June-July, Professor of Signal and Image Processing, EPFL (Chaire Internationale Inria)
- Gilles Blanchard, in September, Professor, University of Potsdam
- Laurent Jacques, in October, Professor, Catholic University of Louvain
- Mike Davies, in November, Professor, University of Edimburgh

10. Dissemination

10.1. Promoting Scientific Activities

Rémi Gribonval is a member of the IEEE Technical Committee on Signal Processing Theory and Methods (2012–2017), 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).

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

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

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

Antoine Deleforge organized and co-chaired with Prof. Sharon Gannot (Bar-Ilan University) a special session on “Learning-Based Sound Source Localization and Spatial Information Retrieval” at ICASSP 2016, Shanghai, China.

N. Bertin, A. Deleforge, and R. Gribonval co-organized a special session entitled “From Source Position to Room Properties : Learning Methods for Audio Scene Geometry Estimation”, at the 2017 LVA/ICA Conference, Grenoble, February 2017

R. Gribonval was the organizer of a special session on “Multiscale Factorizations and Learning”, at the 2016 IEEE Information Theory Workshop, Cambridge, UK, September 2016

R. Gribonval was guest editor of a special issue of IEEE Journal on Selected Topics in Signal Processing, on Structured Matrices in Signal and Data Processing.

F. Bimbot was an invited speaker at the Dagstuhl Seminar 16092 on Computational Music Structure Analysis (February 2016, Dagstuhl, Germany)

R. Gribonval was an invited speaker at the London Workshop on Sparse Signal Processing (Imperial College, September 2016), the DALI (Data Learning and Inference) Workshop on Learning Theory (Sestri Levante, March 2016), the Workshop on Computational and Statistical Trade-offs in Learning (IHES, Bures-sur-Yvette, March 2016), and the Workshop on Low Complexity Models in Signal Processing (Hausdorff Institute, Bonn, February 2016).

R. Gribonval is a member of the jury of the GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing since 2014.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Bachelor : N. Bertin, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : N. Bertin, "Audio rendering, coding and source separation", 6 hours, M2, Université Rennes 1, France.

Master : N. Bertin, "Audio indexing and classification", 6 hours, M2, Université Rennes 1, France.

Master : N. Bertin, "Fundamentals of Signal Processing", 24 hours, M1, Ecole Normale Supérieure de Bretagne, Rennes, France.

Master : N. Bertin, "Sparse representations and compressive sensing", 15 hours, M2, Ecole Nationale de la Statistique et de l'Analyse de l'Information, Rennes, France.

Master : N. Bertin, "Sparsity in Signal and Image Processing", 10 hours, M1, Ecole Normale Supérieure de Bretagne, Rennes, France.

Master : N. Bertin, "Sparsity in Signal and Image Processing”, 12 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

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

Master: R. Gribonval, coordination of the ARD module "Acquisition et Représensation de Données", 20hours, M2, Université Rennes 1, France.

Master : R. Gribonval, "Sparsity in Signal and Image Processing”, 6 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : R. Gribonval, coordination of the module "Sparsity in Signal and Image Processing”, 48 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Bachelor : A. Deleforge, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : A. Deleforge, "Sparsity in Signal and Image Processing", 4 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

10.3. Popularization

10.3.1. Journée Science et Musique

Participants: Antoine Deleforge, Rémi Gribonval, Frédéric Bimbot, Romain Lebarbenchon, Adrien Leman, Nicolas Bellot, Clément Gaultier, Nancy Bertin, Nicolas Keriven, Ewen Camberlein, Luc Le Magoarou, Armelle Mozziconacci, Stéphanie Lemaile, Corentin Louboutin, Nicolas Tremblay, Corentin Guichaoua, Srdan Kitic.

with contributions and support from: Valérie Gouranton, Laurent Perraudeau, Nathalie Denis, Evelyne Orain, Agnès Cottais, and many more.

PANAMA coordinated the organization of a public event called "JournÃ©e Science et Musique" ("Music and Science Day"). 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 2016 hosted more than 500 visitors and was the official opening event of the "Festival des Sciences" week in Rennes.

11. Bibliography

Major publications by the team in recent years

[1] F. Bimbot, E. Deruty, G. Sargent, E. Vincent. *System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces*, in "Music Perception", 2016, 41 p. , To appear - http://mp.ucpress.edu/, https://hal.inria.fr/hal-01188244

[2] 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° 7, pp. 1830–1840 [*DOI :* 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865

[3] R. Gribonval, K. Schnass. *Dictionary Identification - Sparse Matrix-Factorisation via ℓ_1-Minimisation*, in "IEEE Transactions on Information Theory", July 2010, vol. 56, n° 7, pp. 3523–3539 [*DOI :* 10.1109/TIT.2010.2048466], https://hal.archives-ouvertes.fr/hal-00541297

[4] 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° 2, pp. 129–150 [*DOI :* 10.1016/J.ACHA.2010.04.005], https://hal.inria.fr/inria-00541855

[5] S. Kitić, L. Albera, N. Bertin, R. Gribonval. *Physics-driven inverse problems made tractable with cosparse regularization*, in "IEEE Transactions on Signal Processing", January 2016, vol. 64, n° 2, pp. 335-348 [*DOI :* 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087

[6] S. Kitić. *Cosparse regularization of physics-driven inverse problems*, IRISA, Inria Rennes, November 2015, https://hal.archives-ouvertes.fr/tel-01237323

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

[8] 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

[9] G. SARGENT. *Music structure estimation using multi-criteria analysis and regularity constraints*, Université Rennes 1, February 2013, https://tel.archives-ouvertes.fr/tel-00853737

[10] 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

[11] M. CHAFII. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509

Articles in International Peer-Reviewed Journals

[12] H. BECKER, L. ALBERA, P. COMON, R. GRIBONVAL, F. WENDLING, I. MERLET. *Localization of Distributed EEG Sources in the Context of Epilepsy: A Simulation Study*, in "IRBM", 2016 [*DOI :* 10.1016/J.IRBM.2016.04.001], https://hal.archives-ouvertes.fr/hal-01359237

[13] F. BIMBOT, E. DERUTY, G. SARGENT, E. VINCENT. *System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces*, in "Music Perception", 2016, 41 p. , https://hal.inria.fr/hal-01188244

[14] M. CHAFII, J. PALICOT, R. GRIBONVAL, F. BADER. *A Necessary Condition for Waveforms with Better PAPR than OFDM*, in "IEEE Transactions on Communications", 2016 [*DOI :* 10.1109/TCOMM.2016.2584068], https://hal.inria.fr/hal-01128714

[15] S. KITIĆ, L. ALBERA, N. BERTIN, R. GRIBONVAL. *Physics-driven inverse problems made tractable with cosparse regularization*, in "IEEE Transactions on Signal Processing", January 2016, vol. 64, n^o 2, pp. 335-348 [*DOI :* 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087

[16] L. LE MAGOAROU, R. GRIBONVAL. *Flexible Multi-layer Sparse Approximations of Matrices and Applications*, in "IEEE Journal of Selected Topics in Signal Processing", June 2016 [*DOI :* 10.1109/JSTSP.2016.2543461], https://hal.inria.fr/hal-01167948

[17] G. PUY, N. TREMBLAY, R. GRIBONVAL, P. VANDERGHEYNST. *Random sampling of bandlimited signals on graphs*, in "Applied and Computational Harmonic Analysis", June 2016 [*DOI :* 10.1016/J.ACHA.2016.05.005], https://hal.inria.fr/hal-01229578

[18] G. SARGENT, F. BIMBOT, E. VINCENT. *Estimating the structural segmentation of popular music pieces under regularity constraints*, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, https://hal.inria.fr/hal-01403210

[19] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, P. Vandergheynst. *Fast Robust PCA on Graphs*, in "IEEE Journal of Selected Topics in Signal Processing", April 2016, vol. 10, n^{o} 4, pp. 740 - 756 [*DOI :* 10.1109/JSTSP.2016.2555239], https://hal.inria.fr/hal-01277624

[20] N. Shahid, N. Perraudin, G. Puy, P. Vandergheynst. *Compressive PCA for Low-Rank Matrices on Graphs*, in "IEEE transactions on Signal and Information Processing over Networks", 2016, 17 p. , Titled changed from initial preprint "Compressive PCA on graphs" [*DOI :* 10.1109/TSIPN.2016.2631890], https://hal.inria.fr/hal-01277625

[21] Y. Traonmilin, R. Gribonval. *Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all*, in "Applied and Computational Harmonic Analysis", September 2016, https://hal.inria.fr/hal-01207987

[22] N. Tremblay, P. Borgnat. *Subgraph-based filterbanks for graph signals*, in "IEEE Transactions on Signal Processing", March 2016, https://hal.archives-ouvertes.fr/hal-01243889

[23] M. B. Wakin, R. Gribonval, V. Koivunen, J. Romberg, J. Wright. *Introduction to the Issue on Structured Matrices in Signal and Data Processing*, in "IEEE Journal of Selected Topics in Signal Processing", May 2016, vol. 10, n^{o} 4, pp. 605–607 [*DOI :* 10.1109/JSTSP.2016.2553398], https://hal.inria.fr/hal-01421797

Invited Conferences

[24] F. Bimbot. *Towards an Information-Theoretic Framework for Music Structure*, in "Dagstuhl Seminar on Computational Music Structure Analysis", Dagstuhk, Germany, February 2016, pp. 167-168 [*DOI :* 10.4230/DagRep.6.2.147], https://hal.archives-ouvertes.fr/hal-01421013

[25] A. Deleforge, F. Forbes. *Rectified binaural ratio: A complex T-distributed feature for robust sound localization*, in "European Signal Processing Conference", Budapest, Hungary, August 2016, pp. 1257-1261, https://hal.inria.fr/hal-01372337

[26] C. Gaultier, S. Kataria, A. Deleforge. *VAST : The Virtual Acoustic Space Traveler Dataset*, in "International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)", Grenoble, France, February 2017, https://hal.archives-ouvertes.fr/hal-01416508

[27] N. Keriven, N. Tremblay, Y. Traonmilin, R. Gribonval. *Compressive K-means*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://hal.inria.fr/hal-01386077

International Conferences with Proceedings

[28] N. Bertin, E. Camberlein, E. Vincent, R. Lebarbenchon, S. Peillon, É. Lamandé, S. Sivasankaran, F. Bimbot, I. Illina, A. Tom, S. Fleury, E. Jamet. *A French corpus for distant-microphone speech processing in real homes*, in "Interspeech 2016", San Francisco, United States, September 2016, https://hal.inria.fr/hal-01343060

[29] N. Bertin, S. Kitić, R. Gribonval. *Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization*, in "ICASSP - 41st IEEE International Conference on Acoustics, Speech and Signal Processing", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01247227

[30] H. JAIN, P. PÉREZ, R. GRIBONVAL, J. ZEPEDA, H. JÉGOU. *Approximate search with quantized sparse representations*, in "14th European Conference on Computer Vision (ECCV)", Amsterdam, Netherlands, October 2016, https://hal.archives-ouvertes.fr/hal-01361953

[31] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Sketching for Large-Scale Learning of Mixture Models*, in "2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.inria.fr/hal-01208027

[32] L. LE MAGOAROU, R. GRIBONVAL. *Are There Approximate Fast Fourier Transforms On Graphs?*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01254108

[33] C. LOUBOUTIN, F. BIMBOT. *Description of Chord Progressions by Minimal Transport Graphs Using the System & Contrast Model*, in "ICMC 2016 - 42nd International Computer Music Conference", Utrecht, Netherlands, September 2016, https://hal.archives-ouvertes.fr/hal-01421023

[34] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Audio-based robot controlfrom interchannel level difference and absolute sound energy*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, pp. 1992-1999, https://hal.inria.fr/hal-01355394

[35] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Binaural auditory interaction without HRTF for humanoid robots: A sensor-based control approach*, in "Workshop on Multimodal Sensor-based Control for HRI and soft manipulation, IROS'2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01408422

[36] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *First applications of sound-based control on a mobile robot equipped with two microphones*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01277589

[37] T. NOWAKOWSKI, N. BERTIN, R. GRIBONVAL, J. DE ROSNY, L. DAUDET. *Membrane Shape And Boundary Conditions Estimation Using Eigenmode Decomposition*, in "ICASSP 2016 – 41st IEEE International Conference on Acoustics, Speech and Signal Processing", Shanghai, China, March 2016, https://hal.inria.fr/hal-01254882

[38] A. SCHMIDT, A. DELEFORGE, W. KELLERMANN. *Ego-Noise Reduction Using a Motor Data-Guided Multichannel Dictionary*, in "International Conference on Intelligent Robots and Systems (IROS), 2016", Daejon, South Korea, IEEE/RSJ, October 2016, pp. 1281-1286, https://hal.inria.fr/hal-01415723

[39] N. TREMBLAY, G. PUY, P. BORGNAT, R. GRIBONVAL, P. VANDERGHEYNST. *Accelerated spectral clustering using graph filtering of random signals*, in "41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01243682

[40] N. TREMBLAY, G. PUY, R. GRIBONVAL, P. VANDERGHEYNST. *Compressive Spectral Clustering*, in "33rd International Conference on Machine Learning", New York, United States, June 2016, https://hal.archives-ouvertes.fr/hal-01320214

National Conferences with Proceedings

[41] M. CHAFII, J. PALICOT, R. GRIBONVAL. *La modulation en ondelettes: une modulation alternative à faible consommation d'énergie*, in "URSI Énergie et radiosciences", Rennes, France, Journées Scientifiques URSI "Énergie et radiosciences", March 2016, n^{o} 143, https://hal.archives-ouvertes.fr/hal-01401604

[42] N. MONTAVONT, D. SHEHADEH, J. PALICOT, X. LAGRANGE, A. BLANC, R. GRIBONVAL, P. MARY, J.-Y. BAUDAIS, J.-F. HÉLARD, M. CRUSSIÈRE, Y. LOUËT, C. MOY. *Toward Energy Proportional Networks*, in "Journées scientifiques, URSI 2016", Rennes, France, March 2016, https://hal.archives-ouvertes.fr/hal-01297894

Conferences without Proceedings

[43] M. CHAFII, M. LAMARANA DIALLO, J. PALICOT, F. BADER, R. GRIBONVAL. *Adaptive Tone Reservation for better BER Performance in a Frequency Selective Fading Channel*, in "IEEE VTC2016-Spring", Nanjing, China, May 2016, https://hal-centralesupelec.archives-ouvertes.fr/hal-01290918

[44] S. KITIC, N. BERTIN, R. GRIBONVAL. *The best of both worlds: synthesis-based acceleration for physics-driven cosparse regularization*, in "iTwist 2016 - International Traveling Workshop on Interactions Between Sparse Models and Technology", Aalborg, Denmark, August 2016, https://hal.archives-ouvertes.fr/hal-01329051

[45] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE. *Non linear robust regression in high dimension*, in "The XXVIIIth International Biometric Conference", Victoria, Canada, July 2016, https://hal.archives-ouvertes.fr/hal-01423622

[46] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE. *Regression non lineaire robuste en grande dimension*, in "48èmes Journées de Statistique organisées par la Société Française de Statistique", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01423630

Research Reports

[47] C. LOUBOUTIN, F. BIMBOT. *Tensorial Description of Chord Progressions: Complementary Scientific Material*, IRISA, équipe PANAMA, May 2016, https://hal.archives-ouvertes.fr/hal-01314493

[48] G. SARGENT, F. BIMBOT, E. VINCENT. *Supplementary material to the article: Estimating the structural segmentation of popular music pieces under regularity constraints*, IRISA-Inria, Campus de Beaulieu, 35042 Rennes cedex ; Inria Nancy, équipe Multispeech, September 2016, https://hal.inria.fr/hal-01368683

Patents and standards

[49] M. CHAFII, J. PALICOT, R. GRIBONVAL. *Dispositif de communication à modulation temps-fréquence adaptative*, July 2016, n^{o} Numéro de demande : 1656806 ; Numéro de soumission : 1000356937, https://hal.inria.fr/hal-01375661

Other Publications

[50] A. DELEFORGE, Y. TRAONMILIN. *Phase Unmixing : Multichannel Source Separation with Magnitude Constraints*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01372418

[51] S. KATARIA, C. GAULTIER, A. DELEFORGE. *Hearing in a shoe-box : binaural source position and wall absorption estimation using virtually supervised learning* , September 2016, working paper or preprint, https://hal.inria.fr/hal-01372435

[52] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Sketching for Large-Scale Learning of Mixture Models*, June 2016, working paper or preprint, https://hal.inria.fr/hal-01329195

[53] L. LE MAGOAROU, R. GRIBONVAL, N. TREMBLAY. *Approximate Fast Fourier Transforms on graphs via multi-layer sparse approximations*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01416110

[54] E. PERTHAME, F. FORBES, A. DELEFORGE. *Inverse regression approach to robust non-linear high-to-low dimensional mapping*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347455

References in notes

[55] A. ADLER, V. EMIYA, M. G. 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

[56] L. ALBERA, S. KITIĆ, N. BERTIN, G. PUY, R. GRIBONVAL. *Brain source localization using a physics-driven structured cosparse 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

[57] 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

[58] 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

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

[60] A. BOURRIER. *Compressed sensing and dimensionality reduction for unsupervised learning*, Université Rennes 1, May 2014, https://tel.archives-ouvertes.fr/tel-01023030

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

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

[63] 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

[64] M. CHAFII. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509

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

[66] M. CHAFII, J. PALICOT, R. GRIBONVAL. *Closed-form Approximations of the PAPR Distribution for Multi-Carrier Modulation Systems*, in "EUSIPCO 2014 - European Signal Processing Conference", Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01054126

[67] M. CHAFII, J. PALICOT, R. GRIBONVAL. *Closed-form approximations of the peak-to-average power ratio distribution for multi-carrier modulation and their applications*, in "EURASIP Journal on Advances in Signal Processing", 2014, vol. 2014, n^{o} 1, 121 p. [*DOI :* 10.1186/1687-6180-2014-121], https://hal.inria.fr/hal-01056153

[68] M. CHAFII, J. PALICOT, R. GRIBONVAL. *L'optimalité de l'OFDM en termes de performance en PAPR*, in "25ème Colloque Gretsi 2015", Lyon, France, September 2015, https://hal-supelec.archives-ouvertes.fr/hal-01165509

[69] F. CHAUMETTE, S. HUTCHINSON. *Visual servo control, Part I: Basic approaches*, in "IEEE Robotics and Automation Magazine", 2006, vol. 13, n^{o} 4, pp. 82-90, https://hal.inria.fr/inria-00350283

[70] A. DELEFORGE, F. FORBES, R. HORAUD. *Acoustic space learning for sound-source separation and localization on binaural manifolds*, in "International journal of neural systems", 2015, vol. 25, n^{o} 01, 1440003 p. , https://hal.archives-ouvertes.fr/hal-00960796

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

[72] 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

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

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

[75] 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

[76] N. KERIVEN, R. GRIBONVAL. *Compressive Gaussian Mixture Estimation by Orthogonal Matching Pursuit with Replacement*, July 2015, SPARS 2015, https://hal.inria.fr/hal-01165984

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

[78] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Audio Declipping by Cosparse Hard Thresholding*, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, https://hal.inria.fr/hal-00922497

[79] 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 [*DOI :* 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779

[80] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Sparsity and cosparsity for audio declipping: a flexible non-convex approach*, in "LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation", Liberec, Czech Republic, August 2015, 8 p. , https://hal.inria.fr/hal-01159700

[81] S. KITIĆ. *Cosparse regularization of physics-driven inverse problems*, Université Rennes 1 ; Inria Rennes Bretagne Atlantique, November 2015, https://hal.archives-ouvertes.fr/tel-01237323

[82] M. KOWALSKI, K. SIEDENBURG, M. DÖRFLER. *Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators*, in "IEEE Transactions on Signal Processing", May 2013, vol. 61, n^{o} 10, pp. 2498 - 2511 [*DOI :* 10.1109/TSP.2013.2250967], https://hal.archives-ouvertes.fr/hal-00691774

[83] L. LE MAGOAROU, R. GRIBONVAL, A. GRAMFORT. *FAμST: speeding up linear transforms for tractable inverse problems*, in "European Signal Processing Conference (EUSIPCO)", Nice, France, August 2015, https://hal.archives-ouvertes.fr/hal-01156478

[84] L. LE MAGOAROU, R. GRIBONVAL. *Chasing butterflies: In search of efficient dictionaries*, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Brisbane, Australia, April 2015 [*DOI :* 10.1109/ICASSP.2015.7178579], https://hal.archives-ouvertes.fr/hal-01104696

[85] L. LE MAGOAROU. *Efficient matrices for signal processing and machine learning*, INSA de Rennes, November 2016, https://tel.archives-ouvertes.fr/tel-01412558

[86] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Sound-based control with two microphones*, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'15", Hamburg, Germany, September 2015, pp. 5568-5573, https://hal.inria.fr/hal-01185841

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

[88] A. A. NUGRAHA, A. LIUTKUS, E. VINCENT. *Multichannel music separation with deep neural networks*, in "24th European Signal Processing Conference (EUSIPCO 2016)", 2016, pp. 1748–1752

[89] A. OZEROV, C. FÉVOTTE. *Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation*, in "IEEE Transactions on Audio, Speech, and Language Processing", 2010, vol. 18, n^{o} 3, pp. 550–563

[90] G. PUY, M. E. DAVIES, R. GRIBONVAL. *Linear embeddings of low-dimensional subsets of a Hilbert space to* $\mathbb{R}^m$, in "EUSIPCO - 23rd European Signal Processing Conference", Nice, France, August 2015, https://hal.inria.fr/hal-01116153

[91] G. PUY, M. E. DAVIES, R. GRIBONVAL. *Recipes for stable linear embeddings from Hilbert spaces to* $\mathbb{R}^m$, September 2015, Submitted to IEEE Transactions on Information Theory, https://hal.inria.fr/hal-01203614

[92] Y. SALAÜN, E. VINCENT, N. BERTIN, N. SOUVIRAÀ-LABASTIE, X. JAUREGUIBERRY, D. TRAN, F. BIMBOT. *The Flexible Audio Source Separation Toolbox Version 2.0*, May 2014, ICASSP, https://hal.inria.fr/hal-00957412