



# Activity Report 2019

## Team PANAMA

### Sparsity and New Algorithms for Signal and Audio Processing

*Joint team with Inria Rennes – Bretagne Atlantique*

D5 – Digital Signals and Images, Robotics





## Table of contents

<b>1. Team, Visitors, External Collaborators</b> .....	<b>2</b>
<b>2. Overall Objectives</b> .....	<b>3</b>
2.1. Overall positioning	3
2.2. Scientific foundations	3
2.3. Applications	3
<b>3. Research Program</b> .....	<b>3</b>
3.1. Axis 1: Sparse Models and Representations	3
3.1.1. Efficient Sparse Models and Dictionary Design for Large-scale Data	3
3.1.2. Compressive Learning	4
3.2. Axis 2: Robust Acoustic Scene Analysis	4
3.2.1. Compressive Acquisition and Processing of Acoustic Scenes	4
3.2.2. Robust Audio Source Separation	4
3.2.3. Robust Audio Source Localization	5
3.3. Axis 3: Large-scale Audio Content Processing and Self-organization	5
3.3.1. Motif Discovery in Audio Data	5
3.3.2. Structure Modeling and Inference in Audio and Musical Contents	6
<b>4. Application Domains</b> .....	<b>6</b>
4.1. Acoustic Scene Capture	6
4.2. Audio Signal Separation in Reverberant Environments	6
4.3. Multimedia Indexing	7
4.4. Remote Sensing	7
<b>5. Highlights of the Year</b> .....	<b>7</b>
<b>6. New Software and Platforms</b> .....	<b>7</b>
6.1. FAuST	7
6.2. SPADE	8
6.3. FASST	9
6.4. Multi-channel BSS Locate Basic	9
6.5. VoiceHome-2	9
<b>7. New Results</b> .....	<b>10</b>
7.1. Sparse Representations, Inverse Problems, and Dimension Reduction	10
7.1.1. Computational Representation Learning: Algorithms and Theory	10
7.1.2. Generalized matrix inverses and the sparse pseudo-inverse	11
7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching	11
7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction	12
7.1.5. Low-rank approximations: fast constrained algorithms	13
7.1.6. Algorithmic Exploration of Sparse Representations for Neurofeedback	14
7.2. Emerging activities on high-dimensional learning with neural networks	14
7.3. Emerging activities on Nonlinear Inverse Problems	15
7.4. Source Localization and Separation	15
7.4.1. Towards Real-world Localization and Separation	16
7.4.2. Separation for Remixing Applications	16
7.5. Towards comprehensive audio scene analysis	17
7.5.1. Room Properties: Estimating or Learning Early Echoes	17
7.5.2. Multichannel Audio Event and Room Classification	18
7.6. Music Content Processing and Information Retrieval	18
7.6.1. Modeling music by Polytopic Graphs of Latent Relations (PGLR)	18
7.6.2. Exploring Structural Dependencies in Melodic Sequences using Neural Networks	19
7.6.3. Graph Signal Processing for Multiscale Representations of Music Similarity	19

<b>8. Partnerships and Cooperations</b> .....	<b>20</b>
8.1. National Initiatives	20
8.1.1. Labex Comin Labs projects	20
8.1.2. ANR INVATE project with IRT b-com, Rennes	22
8.1.3. ANR OATMIL project	22
8.1.4. Collaboration with 5th dimension – dynamic separation of localized sound sources	22
8.2. International Initiatives	22
8.3. International Research Visitors	23
<b>9. Dissemination</b> .....	<b>23</b>
9.1. Promoting Scientific Activities	23
9.1.1. Scientific Events: Organisation	23
9.1.2. Scientific Events: Selection	23
9.1.2.1. Chair of Conference Program Committees	23
9.1.2.2. Member of Conference Program Committees	23
9.1.3. Membership of Editorial Boards	23
9.1.4. Invited Talks	23
9.1.5. Scientific Expertise	24
9.1.6. Research Administration	24
9.2. Teaching	24
9.3. Popularization	24
9.3.1. Journée Science et Musique (JSM 2019)	24
9.3.2. Participation to Equinoxes#1	25
9.3.3. Internal or external responsibilities	25
<b>10. Bibliography</b> .....	<b>25</b>

## Project-Team PANAMA

*Creation of the Project-Team: 2013 January 01*

### **Keywords:**

#### **Computer Science and Digital Science:**

- A1.2.6. - Sensor networks
- A3.1.1. - Modeling, representation
- A3.3.3. - Big data analysis
- A3.4.1. - Supervised learning
- A3.4.2. - Unsupervised learning
- A3.4.4. - Optimization and learning
- A3.4.5. - Bayesian methods
- A3.4.6. - Neural networks
- A3.4.7. - Kernel methods
- A3.4.8. - Deep learning
- A3.5.1. - Analysis of large graphs
- A5.3.2. - Sparse modeling and image representation
- A5.7.1. - Sound
- A5.7.2. - Music
- A5.7.3. - Speech
- A5.7.4. - Analysis
- A5.9. - Signal processing
  - A5.9.1. - Sampling, acquisition
  - A5.9.2. - Estimation, modeling
  - A5.9.3. - Reconstruction, enhancement
  - A5.9.4. - Signal processing over graphs
  - A5.9.5. - Sparsity-aware processing
  - A5.9.6. - Optimization tools
- A5.10.2. - Perception
- A5.11.2. - Home/building control and interaction
- A6.1.4. - Multiscale modeling
- A6.2.5. - Numerical Linear Algebra
- A6.2.6. - Optimization
- A6.3.1. - Inverse problems
- A6.3.2. - Data assimilation
- A8.6. - Information theory
- A8.7. - Graph theory
- A9.2. - Machine learning
- A9.3. - Signal analysis

#### **Other Research Topics and Application Domains:**

- B2.6. - Biological and medical imaging
- B5.6. - Robotic systems

- B5.8. - Learning and training
- B6.3.3. - Network Management
- B8.1.2. - Sensor networks for smart buildings
- B8.4. - Security and personal assistance
- B9.1. - Education
- B9.2.1. - Music, sound
- B9.2.2. - Cinema, Television
- B9.2.3. - Video games
- B9.8. - Reproducibility
- B9.11.1. - Environmental risks

## 1. Team, Visitors, External Collaborators

### Research Scientists

- Frédéric Bimbot [Team leader, CNRS, Senior Researcher, HDR]
- Rémi Gribonval [Team leader, Inria, Senior Researcher, until Jul 2019, HDR]
- Nancy Bertin [CNRS, Researcher]
- Jeremy Cohen [CNRS, Researcher]

### Faculty Members

- Nicolas Courty [Univ de Bretagne sud, Professor, until Aug 2019, HDR]
- Lilian Besson [École Normale Supérieure Rennes, AGPR, from Nov 2019]

### Post-Doctoral Fellow

- Clément Elvira [Inria, Post-Doctoral Fellow, until Nov 2019]

### PhD Students

- Antoine Chatalic [Univ de Rennes I, PhD Student]
- Diego Di Carlo [Inria, PhD Student]
- Cassio Fraga Dantas [Inria, PhD Student]
- Clément Gaultier [Inria, PhD Student, until Jan 2019]
- Valentin Gillot [INSA Rennes, PhD Student]
- Mohammed Hafsati [Institut de recherche technologique B-com, PhD Student]
- Nathan Libermann [Inria, PhD Student]
- Axel Marmoret [Univ de Rennes I, PhD Student, from Oct 2019]
- Pierre Stock [Facebook, PhD Student, granted by CIFRE]

### Technical staff

- Ewen Camberlein [Inria, Engineer, until Aug 2019]
- Clément Gaultier [Inria, Engineer, from Feb 2019 until Aug 2019]
- Hakim Hadj Djilani [Inria, Engineer, until Apr 2019]
- Romain Lebarbenchon [Inria, Engineer, until Jul 2019]

### Interns and Apprentices

- Thomas Cusson [Inria, from Feb 2019 until Jun 2019]
- Axel Marmoret [Inria, from Feb 2019 until Jun 2019]
- Clement Merdrignac [Inria, from May 2019 until Jul 2019]
- Younes Zine [École Normale Supérieure de Rennes, until Jul 2019]

### Administrative Assistant

- Stéphanie Lemaile [Inria, Administrative Assistant]

### Visiting Scientist

- Andreas Loukas [EPFL, until Feb 2019]

## 2. Overall Objectives

### 2.1. Overall positioning

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

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

### 2.2. Scientific foundations

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

- The extension of the sparse representation paradigm towards that of “sparse modeling”, with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning.
- A focus on sophisticated probabilistic models and advanced statistical methods to account for complex dependencies between multi-layered variables (such as in audiovisual streams, musical contents, biomedical data, remote sensing ...).
- 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, audio and visual 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 and remote sensing. 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.

One of the objectives of PANAMA is to instantiate and validate specific instances of audio source separation approaches and to target them 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.

### 4.4. Remote Sensing

Remote sensing is an image acquisition paradigm in which various pieces of information from the surface of the earth are obtained by airborne or satellite imaging. In the recent years the quality and variety of sensors available on the market for such remote acquisitions has drastically improved, leaving a large number of unsolved or difficult data processing tasks on the hand of data scientists. In particular, members of the PANAMA project-team have focused on processing hyperspectral images, that are images collected at a large number of wavelengths. The tasks that have been addressed are mainly spectral unmixing, that aims at identifying the various materials on the earth surface with their relative abundances, and denoising, that aims at removing structured noise from remote acquisitions that can be polluted by e.g. sensor movement or clouds. The spectra measured at each pixel of an hyperspectral image are heavily dependent on surface features (material, slope, material mixture type) which makes these data extremely rich and interesting.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

- The **Premier Prix de Thèse de la Fondation Rennes 1** in the area of *Mathématiques, Sciences et Technologies de l'Information et de la Communication*, was awarded to **Himalaya Jain** for his Ph.D. [73] titled "Learning compact representations for large scale image search", conducted under the joint supervision of R. Gribonval and Patrick Perez, Technicolor R & I, Rennes.
- The **Prix Jeune Chercheur** from the *Journée Science et Musique 2019 (Rennes)* was awarded to **Corentin Louboutin** for a contribution titled "Modélisation multi-échelle et multi-dimensionnelle de la structure musicale", in relation to his PhD thesis [13].

## 6. New Software and Platforms

### 6.1. FAuST

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

SCIENTIFIC DESCRIPTION: FAuST allows to approximate a given dense matrix by a product of sparse matrices, with considerable potential gains in terms of storage and speedup for matrix-vector multiplications.

FUNCTIONAL DESCRIPTION: Faust 1.x contains Matlab routines to reproduce experiments of the PANAMA team on learned fast transforms.

Faust 2.x contains a C++ implementation with Matlab / Python wrappers (work in progress).

NEWS OF THE YEAR: A Software Development Initiative (ADT REVELATION) started in April 2018 for the maturation of FAuST. A first step achieved this year was to complete and robustify Matlab wrappers, to code Python wrappers with the same functionality, and to setup a continuous integration process. A second step was to simplify the parameterization of the main algorithms. The roadmap for next year includes showcasing examples and optimizing computational efficiency. – In 2017, new Matlab code for fast approximate Fourier Graph Transforms have been included. based on the approach described in the papers:

-Luc Le Magoarou, Rémi Gribonval, "Are There Approximate Fast Fourier Transforms On Graphs?", ICASSP 2016 .

-Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay, "Approximate fast graph Fourier transforms via multi-layer sparse approximations", IEEE Transactions on Signal and Information Processing over Networks,2017.

- Participants: Luc Le Magoarou, Nicolas Tremblay, Rémi Gribonval, Nicolas Bellot, Adrien Leman and Hakim Hadj-Djilani
- Contact: Rémi Gribonval
- Publications: [Approximate fast graph Fourier transforms via multi-layer sparse approximations](#) - [Analyzing the Approximation Error of the Fast Graph Fourier Transform](#) - [Flexible Multi-layer Sparse Approximations of Matrices and Applications](#) - [Are There Approximate Fast Fourier Transforms On Graphs?](#) - [Efficient matrices for signal processing and machine learning](#) - [FA \$\mu\$ ST: speeding up linear transforms for tractable inverse problems](#) - [Chasing butterflies: In search of efficient dictionaries](#) - [Multi-layer Sparse Matrix Factorization](#)
- URL: <http://faust.inria.fr/>

## 6.2. SPADE

*Sparse Audio Declipper*

KEYWORDS: Audio - Sparse regularization - Declipping

SCIENTIFIC DESCRIPTION: SPADE (the Sparse Audio Declipper) allows to reproduce audio declipping experiments from the papers:

- Srđan Kitić, Nancy Bertin, Remi Gribonval. Audio Declipping by Cosparsely Hard Thresholding. iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, Aug 2014, Namur, Belgium.

- Srđan Kitić, 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.

FUNCTIONAL DESCRIPTION: SPADE is a declipping algorithm developed by the PANAMA project-team. To the best of our knowledge SPADE achieves state-of-the-art audio declipping quality. Real-time processing of audio streams is possible.

The web site <http://spade.inria.fr> provides example audio files and allows users to test SPADE on their own files, either by downloading Matlab routines or using Inria's software demonstration platform, Allgo, to test it on the web.

NEWS OF THE YEAR: In 2018, the code has been robustified with the help of InriaTech for a potential industrial transfer. – In 2017, a web interface to demonstrate the potential of SPADE has been setup using the Allgo platform.

- Participants: Nancy Bertin, Clement Gaultier, Ewen Camberlein, Romain Lebarbenchon, Alexandre Sanchez, Rémi Gribonval and Srđan Kitić
- Contact: Rémi Gribonval
- Publications: [Audio Declipping by Cosparsely Hard Thresholding](#) - [Sparsity and cosparsity for audio declipping: a flexible non-convex approach](#)
- URL: <http://spade.inria.fr/>

## 6.3. FASST

*Flexible Audio Source Separation Toolbox*

KEYWORD: Audio signal processing

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

FASST 1.0 development was achieved by the METISS team in Rennes and is now deprecated.

FASST 2.1 development was jointly achieved by the PAROLE team in Nancy and the (former) TEXMEX team in Rennes through an Inria funded ADT (Action de Développement Technologique). PANAMA contributed to the development by coordinating and performing user tests, and to the dissemination in a Show-and-Tell ICASSP poster [58]. While the first implementation was in Matlab, the new implementation is in C++ (for core functions), with Matlab and Python user scripts. Version 2, including speedup and new features was released in 2014 and can be downloaded from <http://bass-db.gforge.inria.fr/fasst/>.

Version 2.2.2 (current version) has been released in May 2018. This version was developed in the PANAMA team through the Inria funded ADT "FFWD" (FASST For Wider Dissemination). A version 3.0 is currently under development and will be released in 2019.

FUNCTIONAL DESCRIPTION: FASST is a Flexible Audio Source Separation Toolbox designed to speed up the conception and automate the implementation of new model-based audio source separation algorithms. It is the only audio source separation software available to the public (QPL licence) which simultaneously exploits spatial and spectral cues on the sources to separate.

NEWS OF THE YEAR: Version 2.2.2 (current version) has been released in May 2018. This version was developed in the PANAMA team through the Inria funded ADT FFWD (FASST For Wider Dissemination). A version 3.0 is currently under development and will be released in 2019.

- Participants: Alexey Ozerov, Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Frédéric Bimbot and Yann Salaun
- Contact: Emmanuel Vincent
- URL: <http://bass-db.gforge.inria.fr/fasst/>

## 6.4. 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).

NEWS OF THE YEAR: In 2018, with the help of InriaTech, selected parts of Multi-channel BSS Locate were ported to C++ in the perspective of a transfer

- Authors: Charles Blandin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Alexey Ozerov and Nancy Bertin
- Contact: Emmanuel Vincent
- URL: [http://bass-db.gforge.inria.fr/bss\\_locate/](http://bass-db.gforge.inria.fr/bss_locate/)

## 6.5. VoiceHome-2

KEYWORDS: Speech processing - Audio signal processing - Source Separation - Source localization

SCIENTIFIC DESCRIPTION: New, extended version of the voiceHome corpus for distant-microphone speech processing in domestic environments. This 5-hour corpus includes short reverberated, noisy utterances (smart home commands) spoken in French by 12 native French talkers in diverse realistic acoustic conditions and recorded by an 8-microphone device at various angles and distances and in various noise conditions. Noise-only segments before and after each utterance are included in the recordings. Clean speech and spontaneous speech recorded in 12 real rooms distributed in 4 different homes are also available. All data have been fully annotated.

- Participants: Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Sunit Sivasankaran, Irina Illina and Frédéric Bimbot
- Contact: Nancy Bertin
- Publication: [VoiceHome-2, an extended corpus for multichannel speech processing in real homes](#)

## 7. New Results

### 7.1. Sparse Representations, Inverse Problems, and Dimension Reduction

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

The team’s activity ranges from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems and dimension reduction.

#### 7.1.1. Computational Representation Learning: Algorithms and Theory

**Participants:** Rémi Gribonval, Hakim Hadj Djilani, Cássio Fraga Dantas, Jeremy Cohen.

*Main collaborations:* Luc Le Magoarou (IRT b-com, Rennes), Nicolas Tremblay (GIPSA-Lab, Grenoble), R. R. Lopes and M. N. Da Costa (DSPCom, Univ. Campinas, Brazil)

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. In our work introduced below, by imposing structural constraints on the dictionary and pruning provably unused atoms, we could alleviate the curse of dimensionality.

**Multilayer sparse matrix products for faster computations.** Inspired by usual fast transforms, we proposed a general dictionary structure (called FA $\mu$ ST for Flexible Approximate Multilayer Sparse Transforms) that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation, with reduced sample complexity. A comprehensive journal paper was published in 2016 [75], and we further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms [76], empirically showing that  $\mathcal{O}(n \log n)$  approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opened the way to substantial accelerations for Fourier Transforms on large graphs. This year we focused on the development of the FA $\mu$ ST software library (see Section 6), providing transparent interfaces of FA $\mu$ ST data-structures with both Matlab and Python.

**Kronecker product structure for faster computations.** In parallel to the development of FA $\mu$ ST, we proposed another approach to structured dictionary learning that also aims at speeding up both sparse coding and dictionary learning. We used the fact that for tensor data, a natural set of linear operators are those that operate on each dimension separately, which correspond to rank-one multilinear operators. These rank-one operators may be cast as the Kronecker product of several small matrices. Such operators require less memory and are computationally attractive, in particular for performing efficient matrix-matrix and matrix-vector operations. In our proposed approach, dictionaries are constrained to belong to the set of low-rank multilinear operators, that consist of the sum of a few rank-one operators. The general approach, coined HOSUKRO for High Order Sum of Kronecker products, was shown last year to reduce empirically the sample complexity of dictionary learning, as well as theoretical complexity of both the learning and the sparse coding operations [67]. This year we demonstrated its potential for hyperspectral image denoising. A new efficient algorithm with lighter sample complexity requirements and computational burden was proposed and shown to be competitive with the state-of-the-art for hyperspectral image denoising with dedicated adjustments [50], [28], [27].

**Combining faster matrix-vector products with screening techniques.** We combined accelerated matrix-vector multiplications offered by FA $\mu$ ST / HOSUKRO matrix approximations with dynamic screening [59], that safely eliminates inactive variables to speedup iterative convex sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary [68]. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. This lead to a journal publication [21] that includes new techniques based on duality gaps to optimally switch from a coarse dictionary approximation to a finer one. Significant complexity reductions were obtained in comparison to screening rules alone.

### 7.1.2. Generalized matrix inverses and the sparse pseudo-inverse

**Participant:** Rémi Gribonval.

*Main collaboration: Ivan Dokmanic (University of Illinois at Urbana Champaign, USA)*

We studied linear generalized inverses that minimize matrix norms. Such generalized inverses are famously represented by the Moore-Penrose pseudoinverse (MPP) which happens to minimize the Frobenius norm. Freeing up the degrees of freedom associated with Frobenius optimality enables us to promote other interesting properties. In a first part of this work [64], we looked at the basic properties of norm-minimizing generalized inverses, especially in terms of uniqueness and relation to the MPP. We first showed that the MPP minimizes many norms beyond those unitarily invariant, thus further bolstering its role as a robust choice in many situations. We then concentrated on some norms which are generally not minimized by the MPP, but whose minimization is relevant for linear inverse problems and sparse representations. In particular, we looked at mixed norms and the induced  $\ell^p \rightarrow \ell^q$  norms.

An interesting representative is the sparse pseudoinverse which we studied in much more detail in a second part of this work published this year [19], motivated by the idea to replace the Moore-Penrose pseudoinverse by a sparser generalized inverse which is in some sense well-behaved. Sparsity implies that it is faster to apply the resulting matrix; well-behavedness would imply that we do not lose much in stability with respect to the least-squares performance of the MPP. We first addressed questions of uniqueness and non-zero count of (putative) sparse pseudoinverses. We showed that a sparse pseudoinverse is generically unique, and that it indeed reaches optimal sparsity for almost all matrices. We then turned to proving a stability result: finite-size concentration bounds for the Frobenius norm of  $p$ -minimal inverses for  $1 \leq p \leq 2$ . Our proof is based on tools from convex analysis and random matrix theory, in particular the recently developed convex Gaussian min-max theorem. Along the way we proved several results about sparse representations and convex programming that were known folklore, but of which we could find no proof.

### 7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching

**Participants:** Rémi Gribonval, Antoine Chatalic.

*Main collaborations this year: Nicolas Keriven (ENS Paris), Phil Schniter & Evan Byrne (Ohio State University, USA), Laurent Jacques & Vincent Schellekens (Univ Louvain, Belgium), Florimond Houssiau & Y.-A. de Montjoye (Imperial College London, UK)*

**Sketching for Large-Scale Learning.** When learning from voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [60] and Nicolas Keriven [74] an approach based on sketching. A low-dimensional sketch is computed by averaging (random) features over the training collection. The sketch can be seen as made of a collection of empirical generalized moments of the underlying probability distribution. Leveraging analogies with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough, and released an associated toolbox for reproducible research (see SketchMLBox, Section 6) with the so-called Compressive Learning Orthogonal Matching Pursuit (CL-OMP) algorithm which is inspired by Matching Pursuit. Three unsupervised learning settings have been addressed so far: Gaussian Mixture Modeling,  $k$ -means clustering, and principal component analysis. A survey conference paper on sketching for large-scale learning was published this year [25], and an extended journal version of this survey is in preparation.

**Efficient algorithms to learn for sketches** Last year, we showed that in the high-dimensional setting one can substantially speedup both the sketching stage and the learning stage with CL-OMP by replacing Gaussian random matrices with fast random linear transforms in the sketching procedure [63]. We studied an alternative to CL-OMP for cluster recovery from a sketch, which is based on simplified hybrid generalized approximate message passing (SHyGAMP). Numerical experiments suggest that this approach is more efficient than CL-OMP (in both computational and sample complexity) and more efficient than k-means++ in certain regimes [61]. During his first year of Ph.D., Antoine Chatalic visited the group of Phil Schniter to further investigate this topic, and a journal paper has been published as a result of this collaboration [15].

**Privacy-preserving sketches** Sketching provides a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum. We established theoretical privacy guarantees (with the *differential privacy* framework) and explored the utility / privacy tradeoffs of Compressive K-means [24]. A journal paper is in preparation where we extend these results to Gaussian mixture modeling and principal component analysis.

**Advances in optical-based random projections** Random projections are a key ingredient of sketching. Motivated by the recent development of dedicated optics-based hardware for rapid random projections, which leverages the propagation of light in random media, we tackled the problem of recovering the phase of complex linear measurements when only magnitude information is available and we control the input. A signal of interest  $\xi \in \mathbb{R}^N$  is mixed by a random scattering medium to compute the projection  $y = \mathbf{A}\xi$ , with  $\mathbf{A} \in \mathbb{C}^{M \times N}$  a realization of a standard complex Gaussian independent and identically distributed (iid) random matrix. Such optics-based matrix multiplications can be much faster and energy-efficient than their CPU or GPU counterparts, yet two difficulties must be resolved: only the intensity  $|y|^2$  can be recorded by the camera, and the transmission matrix  $\mathbf{A}$  is unknown. We showed that even without knowing  $\mathbf{A}$ , we can recover the unknown phase of  $y$  for some equivalent transmission matrix with the same distribution as  $\mathbf{A}$ . Our method is based on two observations: first, conjugating or changing the phase of any row of  $\mathbf{A}$  does not change its distribution; and second, since we control the input we can interfere  $\xi$  with arbitrary reference signals. We showed how to leverage these observations to cast the measurement phase retrieval problem as a Euclidean distance geometry problem. We demonstrated appealing properties of the proposed algorithm in both numerical simulations and real hardware experiments. Not only does our algorithm accurately recover the missing phase, but it mitigates the effects of quantization and the sensitivity threshold, thus improving the measured magnitudes [33].

#### 7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction

**Participants:** Rémi Gribonval, Clément Elvira, Jérémy Cohen.

*Main collaboration: Nicolas Keriven (ENS Paris), Gilles Blanchard (Univ Postdam, Germany), Cédric Herzet (SIMSMART project-team, IRMAR / Inria Rennes), Charles Soussen (Centrale Supélec, Gif-sur-Yvette), Mila Nikolova (CMLA, Cachan), Nicolas Gillis (UMONS)*

**Information preservation guarantees with low-dimensional sketches.** We established a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework provides theoretical grounds supporting the experimental success of our algorithmic approaches to compressive K-means, compressive Gaussian Mixture Modeling, as well as compressive Principal Component Analysis (PCA). A comprehensive preprint is being revised for a journal [71].

**Recovery guarantees for algorithms with continuous dictionaries.** We established theoretical guarantees on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous dictionaries [66], e.g. as appearing in the context of sparse spike deconvolution. Analyses based on discretized dictionary fail to be conclusive when the discretization step tends to zero, as the coherence goes to one. Instead, our analysis is directly conducted in the continuous setting and exploits specific properties of the positive definite kernel between atom parameters defined by the inner product between the corresponding atoms. For the Laplacian kernel in dimension one, we showed in the noise-free setting that OMP exactly recovers the atom parameters as well as their amplitudes, regardless of the number of distinct atoms [66]. A



preprint describing a full class of kernels for which such an analysis holds, in particular for higher dimensional parameters, has been released and submitted to a journal [30], [36], [31], [51].

**Identifiability of Complete Dictionary Learning** In the era of deep learning, dictionary learning has proven to remain an important and extensively-used data mining and processing tool. Having been studied and used for over twenty years, dictionary learning has well-understood properties. However there was a particular stone missing, which was understanding deterministic conditions for the parameters of dictionary learning to be uniquely retrieved from a training data set. We filled this gap partially by drastically improving on the previously best such conditions in the case of complete dictionaries [16]. Moreover, although algorithms with guaranties to compute the unique best solution do exist, they are seldom used in practice due to their high computational cost. In subsequent work, we showed that faster algorithms typically used to compute dictionary learning often failed at computing the unique solution (in cases where our previous result guaranties this uniqueness), opening the way to new algorithms that are both fast and guaranteed [26].

**On Bayesian estimation and proximity operators.** There are two major routes to address the ubiquitous family of inverse problems appearing in signal and image processing, such as denoising or deblurring. The first route is Bayesian modeling: prior probabilities are used to model both the distribution of the unknown variables and their statistical dependence with the observed data, and estimation is expressed as the minimization of an expected loss (e.g. minimum mean squared error, or MMSE). The other route is the variational approach, popularized with sparse regularization and compressive sensing. It consists in designing (often convex) optimization problems involving the sum of a data fidelity term and a penalty term promoting certain types of unknowns (e.g., sparsity, promoted through an L1 norm).

Well known relations between these two approaches have lead to some widely spread misconceptions. In particular, while the so-called Maximum A Posteriori (MAP) estimate with a Gaussian noise model does lead to an optimization problem with a quadratic data-fidelity term, we disprove through explicit examples the common belief that the converse would be true. In previous work we showed that for denoising in the presence of additive Gaussian noise, for any prior probability on the unknowns, the MMSE is the solution of a penalized least-squares problem, with all the apparent characteristics of a MAP estimation problem with Gaussian noise and a (generally) different prior on the unknowns [72]. In other words, the variational approach is rich enough to build any MMSE estimator associated to additive Gaussian noise via a well chosen penalty.

This year, we achieved generalizations of these results beyond Gaussian denoising and characterized noise models for which the same phenomenon occurs. In particular, we proved that with (a variant of) Poisson noise and any prior probability on the unknowns, MMSE estimation can again be expressed as the solution of a penalized least-squares optimization problem. For additive scalar denoising, the phenomenon holds if and only if the noise distribution is log-concave, resulting in the perhaps surprising fact that scalar Laplacian denoising can be expressed as the solution of a penalized least-squares problem [22] Somewhere in the proofs appears an apparently new characterization of proximity operators of (nonconvex) penalties as subdifferentials of convex potentials [54].

### 7.1.5. Low-rank approximations: fast constrained algorithms

**Participant:** Jeremy Cohen.

*Main collaborations: Nicolas Gillis (Univ. Mons, Belgium), Andersen Man Shun Ang (Univ. Mons, Belgium), Nicolas Nadisic (Univ. Mons, Belgium).*

Low-Rank Approximations (LRA) aim at expressing the content of a multiway array by a sum of simpler separable arrays. Understood as a powerful unsupervised machine learning technique, LRA are most and foremost modern avatars of sparsity that are still not fully understood. In particular, algorithms to compute the parameters of LRA demand a lot of computer resources and provide sub-optimal results. An important line of work over the last year has been to design efficient algorithms to compute constrained LRA, and in particular constrained low-rank tensor decompositions. This work has been carried out through a collaboration with the ERC project COLORAMAP of Nicolas Gillis (Univ. Mons, Belgium) and his PhD students Nicolas Nadisic (co-supervision) and Andersen Man Shun Ang.

**Extrapolated Block-coordinate algorithms for fast tensor decompositions** State-of-the-art algorithms for computing tensor decompositions are based on the idea that solving alternatively for smaller blocks of parameters is easier than solving the large problem at once. Despite showing nice convergence speeds, the obtained Block Coordinate Descent algorithms (BCD) are prone to being stuck near saddle points. We have shown in preliminary work, which is still ongoing, that BCD algorithms can be improved using Nesterov extrapolation in-between block updates. This improves empirical convergence speed in constrained and unconstrained tensor decompositions tremendously at almost no additional computation cost, and is therefore bound to have a large impact on the community [37].

**Exact sparse nonnegative least-squares solutions to least-squares problems** Another important LRA is Nonnegative Matrix factorization, which has found many diverse applications such as in remote sensing or automatic music transcription. Sometimes, imposing sparsity on parameters of NMF is crucial to be able to correctly process and interpret the output of NMF. However, sparse NMF has scarcely been studied, and its computation is challenging. In fact, even only a subproblem in a BCD approach, sparse nonnegative least-squares, is already NP-hard. We proposed to solve this sparse nonnegative least-squares problem exactly using a combinatorial algorithm. To reduce as much as possible the cost of solving this combinatorial problem, a Branch and Bound algorithm was proposed which, on average, reduces the computational complexity drastically. A next step will be to use this branch and bound algorithm as a brick for proposing an efficient algorithm for sparse NMF.

### 7.1.6. Algorithmic Exploration of Sparse Representations for Neurofeedback

**Participant:** Rémi Gribonval.

*Claire Cury, Pierre Maurel & Christian Barillot (EMPENN Inria project-team, Rennes)*

In the context of the HEMISFER (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation) Comin Labs project (see Section 1), in collaboration with the EMPENN team, we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [82]. We then focused on directly estimating neuro-feedback scores rather than brain activity. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) both allow measurement of brain activity for neuro-feedback (NF), respectively with high temporal resolution for EEG and high spatial resolution for fMRI. Using simultaneously fMRI and EEG for NF training is very promising to devise brain rehabilitation protocols, however performing NF-fMRI is costly, exhausting and time consuming, and cannot be repeated too many times for the same subject. We proposed a technique to predict NF scores from EEG recordings only, using a training phase where both EEG and fMRI NF are available [39]. A journal paper has been submitted.

## 7.2. Emerging activities on high-dimensional learning with neural networks

**Participants:** Rémi Gribonval, Himalaya Jain, Pierre Stock.

*Main collaborations: Patrick Perez (Technicolor R & I, Rennes), Gitta Kutyniok (TU Berlin, Germany), Morten Nielsen (Aalborg University, Denmark), Felix Voigtlaender (KU Eichstätt, Germany), Herve Jegou and Benjamin Graham (FAIR, Paris)*

dictionary learning, large-scale indexing, sparse deep networks, normalization, sinkhorn, regularization

Many of the data analysis and processing pipelines that have been carefully engineered by generations of mathematicians and practitioners can in fact be implemented as deep networks. Allowing the parameters of these networks to be automatically trained (or even randomized) allows to revisit certain classical constructions. Our team has started investigating the potential of such approaches both from an empirical perspective and from the point of view of approximation theory.

**Learning compact representations for large-scale image search.** The PhD thesis of Himalaya Jain [73], which received the Fondation Rennes 1 PhD prize this year, was dedicated to learning techniques for the design of new efficient methods for large-scale image search and indexing.

**Equi-normalization of Neural Networks.** Modern neural networks are over-parameterized. In particular, each rectified linear hidden unit can be modified by a multiplicative factor by adjusting input and output weights, without changing the rest of the network. Inspired by the Sinkhorn-Knopp algorithm, we introduced a fast iterative method for minimizing the  $l_2$  norm of the weights, equivalently the weight decay regularizer. It provably converges to a unique solution. Interleaving our algorithm with SGD during training improves the test accuracy. For small batches, our approach offers an alternative to batch- and group- normalization on CIFAR-10 and ImageNet with a ResNet-18. This work was presented at ICLR 2019 [41].

**Approximation theory with deep networks.** We study the expressivity of sparsely connected deep networks. Measuring a network's complexity by its number of connections with nonzero weights, or its number of neurons, we consider the class of functions which error of best approximation with networks of a given complexity decays at a certain rate. Using classical approximation theory, we showed that this class can be endowed with a norm that makes it a nice function space, called approximation space. We established that the presence of certain "skip connections" has no impact on the approximation space, and studied the role of the network's nonlinearity (also known as activation function) on the resulting spaces, as well as the benefits of depth. For the popular ReLU nonlinearity (as well as its powers), we related the newly identified spaces to classical Besov spaces, which have a long history as image models associated to sparse wavelet decompositions. The sharp embeddings that we established highlight how depth enables sparsely connected networks to approximate functions of increased "roughness" (decreased Besov smoothness) compared to shallow networks and wavelets. A preprint has been published and is under review for a journal [23].

## 7.3. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

### 7.3.1. Audio Inpainting and Denoising

**Participants:** Rémi Gribonval, Nancy Bertin, Clément Gaultier.

*Main collaborations: Srdan Kitic (Orange, Rennes)*

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 [57], we proposed over the last five years a series of algorithms leveraging the competitive cosparsity approach, which offers a very appealing trade-off between reconstruction performance and computational time, and its extensions to the incorporation of the so-called "social" into problems regularized by a cosparsity prior. We exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [69]; these results, together with listening tests results that were specified and prepared in 2019 and will be run soon, will be included in an ongoing journal paper, to be submitted in 2020. This year, following Clément Gaultier Ph.D. defense [12], we progressed towards industrial transfer of these results through informal interaction with a company commercializing audio plugins, in particular with new developments to alleviate some artifacts absent from simulation but arising in real-world use cases.

## 7.4. Source Localization and Separation

Source separation, sparse representations, probabilistic model, source localization

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, which is the task of retrieving the source signals underlying a multichannel mixture signal. Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [9], hereafter summarized as the "local Gaussian modeling", and 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. This model became the basis of a large number of works in the field, including our own. This accumulated progress led, in 2015, to two main

achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [84] 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 [11].

From there, our recent work divided into several tracks: maturity work on the concrete use of these tools and principles in real-world scenarios, in particular within the INVATE project and the collaboration with the startup 5<sup>th</sup> dimension (see Sections 8.1.2, 8.1.4), on the one hand; on the other hand, an emerging track on audio scene analysis with machine learning, evolved beyond the “localization and separation” paradigm, and is the subject of a more recent axis of research presented in Section 7.5.

#### 7.4.1. Towards Real-world Localization and Separation

**Participants:** Nancy Bertin, Frédéric Bimbot, Rémi Gribonval, Ewen Camberlein, Romain Lebarbenchon, Mohammed Hafsati.

*Main collaborations: Emmanuel Vincent (MULTISPEECH Inria project-team, Nancy)*

Based on the team’s accumulated expertise and tools for localization and separation using the local Gaussian model, two real-world applications were addressed in the past year, which in turn gave rise to new research tracks.

First, our work within the voiceHome project (2015-2017), an OSEO-FUI industrial collaboration <sup>1</sup> aiming at developing natural language dialog in home applications, such as control of domotic and multimedia devices, in realistic and challenging situations (very noisy and reverberant environments, distant microphones) found its conclusion with the publication of a journal paper in a special issue of Speech Communication [14].

Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique). This new development phase of the FASST software started in September 2017 and was achieved this year by the release of the third version of the toolbox, with significant progress towards efficient initialization, low latency and reduction of the computational burden.

In addition, evolutions of the MBSSLocate software initiated during this project led to a successful participation in the IEEE-AASP Challenge on Acoustic Source Localization and Tracking (LOCATA) [77], and served as a baseline for the publication of the for the IEEE Signal Processing Cup 2019 [21]. The SP Cup was also fueled by the publicly available DREGON dataset 5 recorded in PANAMA, including noiseless speech and on-flight ego-noise recordings, devoted to source localization from a drone [117].

Finally, these progress also led to a new industrial transfer with the start-up 5<sup>th</sup> dimension (see Section 8.1.4). During this collaboration aiming at equipping a pair of glasses with an array of microphones and “smart” speech enhancement functionalities, we particularly investigated the impact of obstacles between microphones in the localization and separation performance, the selection of the best subset of microphones in the array for side speakers hidden by the head shadow, and the importance of speaker enrolment (learning spectral dictionaries of target users voices) in this use case.

#### 7.4.2. Separation for Remixing Applications

**Participants:** Nancy Bertin, Rémi Gribonval, Mohammed Hafsati.

*Main collaborations: Nicolas Epain (IRT b<>com, Rennes)*

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b<>com with the INVATE project, see Section 8.1.2) started in November 2016, we investigated a new application of source separation to sound re-spatialization from Higher Order Ambisonics (HOA) signals [70], in the context of free navigation in 3D audiovisual contents. We studied the applicability conditions of the FASST framework to HOA signals and benchmarked localization and separation methods in this domain. Simulation results showed that separating sources in the HOA domain results in a 5 to 15 dB increase in signal-to-distortion ratio, compared to the microphone domain. These results were accepted for publication in the DAFx international conference [34]. We continued extending our methods following two tracks: hybrid acquisition scenarios, where the separation

<sup>1</sup>With partners: onMobile, Delta Dore, eSoftThings, Orange, Technicolor, LOUSTIC, Inria Nancy.

of HOA signals can be informed by complementary close-up microphonic signals, and the replacement of spectrogram NMF by neural networks for a better spectral adaptation of the models. Future work will include subjective evaluation of the developed workflows.

## 7.5. Towards comprehensive audio scene analysis

Source localization and separation, machine learning, room geometry, room properties, multichannel audio classification

By contrast to the previous lines of work and results on source localization and separation, which are mostly focused on the *sources*, the following emerging activities consider the audio scene and its analysis in a wider sense, including the environment around the sources, and in particular the *room* they are included in, and their properties. This inclusive vision of the audio scene allows in return to revisit classical audio processing tasks, such as localization, separation or classification.

### 7.5.1. Room Properties: Estimating or Learning Early Echoes

**Participants:** Nancy Bertin, Diego Di Carlo, Clément Elvira.

*Main collaborations:* Antoine Deleforge (Inria Nancy – Grand Est), Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Helena Peic-Tukuljac (EPFL, Switzerland).

In [85] we showed that the knowledge of early echoes improved sound source separation performances, which motivates the development of (blind) echo estimation techniques. Echoes are also known to potentially be a key to the room geometry problem [65]. In 2019, two different approaches to this problem were explored.

As a competitive, yet similar approach to our previous work in [83], we proposed a new analytical method for off-the-grid early echoes estimation, based on continuous dictionaries and extensions of sparse recovery methods in this setting. From the well-known *cross-relation* between room impulse responses and signals in a “one source - two microphones” settings, the echo estimation problem can be recast as a Beurling-LASSO problem and solved with algorithms of this kind. This enables near-exact blind and off-grid echo retrieval from discrete-time measurements, and can outperform conventional methods by several orders of magnitude in precision, in an ideal case where the room impulse response is limited to a few weighted Diracs. Future work will include alternative initialization schemes, extensions to sparse-spectrum signals and noisy measurements, and applications to dereverberation and audio-based room shape reconstruction. This work, mostly lead by Clément Elvira, was submitted for publication in *Icassp* 2020.

On the other hand, the PhD thesis of Diego Di Carlo aims at applying the “Virtual Acoustic Space Traveler” (VAST) framework to the blind estimation of acoustic echoes, or other room properties (such as reverberation time, acoustic properties at the boundaries, etc.) Last year, we focused on identifying promising couples of inputs and outputs for such an approach, especially by leveraging the notions of relative transfer functions between microphones, the room impulse responses, the time-difference-of-arrivals, the angular spectra, and all their mutual relationships. In a simple yet common scenario of 2 microphones close to a reflective surface and one source (which may occur, for instance, when the sensors are placed on a table such as in voice-based assistant devices), we introduced the concept of microphone array augmentation with echoes (MIRAGE) and showed how estimation of early-echo characteristics with a learning-based approach is not only possible but can in fact benefit source localization. In particular, it allows to retrieve 2D direction of arrivals from 2 microphones only, an impossible task in anechoic settings. These first results were published in *ICASSP* [29]. In 2019, we improved the involved DNN architecture in MIRAGE and worked towards experimental validation of this result, by designing and recording a data set with annotated echoes in different conditions of reverberation. Future work will include extension of this data set, extension to more realistic and more complex scenarios (including more microphones, sources and reflective surfaces) and the estimation of other room properties such as the acoustic absorption at the boundaries, or ultimately, the room geometry. Some of these tracks currently benefit from the visit of Diego di Carlo to Bar-Ilan University (thanks to a MathSTIC doctoral outgoing mobility grant.)

### 7.5.2. Multichannel Audio Event and Room Classification

**Participants:** Marie-Anne Lacroix, Nancy Bertin.

*Main collaborations: Pascal Scalart, Romuald Rocher (GRANIT Inria project-team, Lannion)*

Typically, audio event detection and classification is tackled as a “pure” single-channel signal processing task. By contrast, audio source localization is the perfect example of multi-channel task “by construction”. In parallel, the need to classify the type of scene or room has emerged, in particular from the rapid development of wearables, the “Internet of things” and their applications. The PhD of Marie-Anne Lacroix, started in September 2018, combines these ideas with the aim of developing multi-channel, room-aware or spatially-aware audio classification algorithms for embedded devices. The PhD topic includes low-complexity and low-energy stakes, which will be more specifically tackled thanks to the GRANIT members area of expertise. During the first year of the PhD, we gathered existing data and identified the need for new simulations or recordings, and combined ideas from existing single-channel classification techniques with traditional spatial features in order to design several baseline algorithms for multi-channel joint localization and classification of audio events. The impact of feature quantization on classification performance is also currently under investigation and a participation to the 2020 edition of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE) is envisioned.

## 7.6. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model, complexity

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 [58], in particular the System & Contrast model [1].

### 7.6.1. Modeling music by Polytopic Graphs of Latent Relations (PGLR)

**Participants:** Corentin Louboutin, Frédéric Bimbot.

The musical content observed at a given instant within a music segment obviously tends to share privileged relationships with its immediate past, hence the sequential perception of the music flow. But local music content also relates with distant events which have occurred in the longer term past, especially at instants which are metrically homologous (in previous bars, motifs, phrases, etc.) This is particularly evident in strongly “patterned” music, such as pop music, where recurrence and regularity play a central role in the design of cyclic musical repetitions, anticipations and surprises.

The web of musical elements can be described as a Polytopic Graph of Latent Relations (PGLR) which models relationships developing predominantly between homologous elements within the metrical grid.

For regular segments the PGLR lives on an  $n$ -dimensional cube(square, cube, tesseract, etc...),  $n$  being the number of scales considered simultaneously in the multiscale model. By extension, the PGLR can be generalized to a more or less regular  $n$ -dimensional polytopes.

Each vertex in the polytope corresponds to a low-scale musical element, each edge represents a relationship between two vertices and each face forms an elementary system of relationships.

The estimation of the PGLR structure of a musical segment can be obtained computationally as the joint estimation of the description of the polytope, the nesting configuration of the graph over the polytope (reflecting the flow of dependencies and interactions between the elements within the musical segment) and the set of relations between the nodes of the graph, with potentially multiple possibilities.

If musical elements are chords, relations can be inferred by minimal transport [79] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [81] where the PGLR approach is presented conceptually and algorithmically, together with an extensive evaluation on a large set of chord sequences from the RWC Pop corpus (100 pop songs).

Specific graph configurations, called Primer Preserving Permutations (PPP) are extensively studied in [80] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

In parallel, recent work has also been dedicated to modeling melodic and rhythmic motifs in order to extend the polytopic model to multiple musical dimensions.

Results obtained in this framework illustrate the efficiency of the proposed model in capturing structural information within musical data and support the view that musical content can be delinearised in order to better describe its structure. Extensive results are included in Corentin Louboutin's PhD [13], defended in March 2019 and which was awarded the Prix Jeune Chercheur Science et Musique, in October.

### 7.6.2. Exploring Structural Dependencies in Melodic Sequences using Neural Networks

**Participants:** Nathan Libermann, Frédéric Bimbot.

*This work is carried out in the framework of a PhD, co-directed by Emmanuel Vincent (Inria-Nancy).*

In order to be able to generate structured melodic phrases and section, we explore various schemes for modeling dependencies between notes within melodies, using deep learning frameworks.

As a first set of experiments, we have considered a GRU-based sequential learning model, studied under different learning scenarios in order to better understand the optimal architectures in this context that can achieve satisfactory results. By this means, we wish to explore different hypotheses relating to temporal non-invariance relationships between notes within a structural segment (motif, phrase, section).

We have defined three types of recursive architectures corresponding to different ways to exploit the local history of a musical note, in terms of information encoding and generalization capabilities.

Initially conducted on the Lakh MIDI dataset, experiments have switched to the Meertens Tune Collections data set (Dutch traditional melodies) and confirm the trends observed in [78], w.r.t. the utility of non-ergodic models for the generation of melodic segments.

Ongoing work is extending these findings to the design of specific NN architectures, which incorporate attention models, to account for this non-invariance of information across musical segments.

### 7.6.3. Graph Signal Processing for Multiscale Representations of Music Similarity

**Participants:** Valentin Gillot, Frédéric Bimbot.

“Music Similarity” is a multifaceted concept at the core of Music Information Retrieval (MIR). Among the wide range of possible definitions and approaches to this notion, a popular one is the computation of a so-called content-based similarity matrix ( $S$ ), in which each coefficient is a similarity measure between descriptors of short time frames at different instants within a music piece or a collection of pieces.

Matrix  $S$  can be seen as the adjacency matrix of an underlying graph, embodying the local and non-local similarities between parts of the music material. Considering the nodes of this graph as a new set of indices for the original music frames or pieces opens the door to a “delinearized” representation of music, emphasizing its structure and its semiotic content.

Graph Signal Processing (GSP) is an emerging topic devoted to extend usual signal processing tools (Fourier analysis, filtering, denoising, compression, ...) to signals “living” on graphs rather than on the time line, and to exploit mathematical and algorithmic tools on usual graphs, in order to better represent and manipulate these signals. Toy applications of GSP concepts on music content in music resequencing and music inpainting are illustrating this trend.

From exploratory experiments, first observations point towards the following hypotheses :

- local and non-local structures of a piece are highlighted in the adjacency matrix built from a simple time-frequency representation of the piece,
- the first eigenvectors of the graph Laplacian provide a rough structural segmentation of the piece,
- clusters of frames built from the eigenvectors contain similar, repetitive sound sequences.

The goal of Valentin Gillot's PhD is to consolidate these hypotheses and investigate further the topic of Graph Signal Processing for music, with more powerful conceptual tools and experiments at a larger scale.

The core of the work will consist in designing a methodology and implement an evaluation framework so as to (i) compare different descriptors and similarity measures and their capacity to capture relevant structural information in music pieces or collection of pieces, (ii) explore the structure of musical pieces by refining the frame clustering process, in particular with a multi-resolution approach, (iii) identify salient characteristics of graphs in relation to mid-level structure models and (iv) perform statistics on the typical properties of the similarity graphs on a large corpus of music in relation to music genres and/or composers.

By the end of the PhD, we expect the release of a specific toolbox for music composition, remixing and repurposing using the concepts and algorithms developed during the PhD. First results obtained this year in music recomposition have proven very conclusive [32].

## 8. Partnerships and Cooperations

### 8.1. National Initiatives

#### 8.1.1. Labex Comin Labs projects

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

- **HEMISFER (2014-2017) and HEMISFER-CLINICAL (2018-2019)**

**Participant:** Rémi Gribonval.

*Acronym: HYBRID (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation)*

*<http://hemisfer.cominlabs.u-bretagne-normandie.fr/>*

*Research axis: 3.1*

*CominLabs partners : EMPENN, HYBRID and PANAMA Inria project-teams;*

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

*Coordinator: Christian Barillot, EMPENN Inria project-team*

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

*Contribution of PANAMA: PANAMA, in close cooperation with the EMPENN 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.*

- **TEPN**



**Participant:** Rémi Gribonval.

*Acronym:* *TEPN (Toward Energy Proportional Networks)*

<http://tepn.cominlabs.u-bretagne.fr/>

*Research axis:* **3.1**

*CominLabs partners :* *IRISA OCIF - Telecom Bretagne; IETR SCN; IETR SCEE; PANAMA Inria project-team*

*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, 2016), focuses on the design of new waveforms for multi carrier systems with reduced Peak to Average Power Ratio (PAPR).*

- **FAWI (2019-2020)**

- **Fourier Adaptive Waveform Implementation**

- **Participant:** Rémi Gribonval.

- This project is a follow-up to TEPN. Its main goal is to implement a prototype demonstrating concretely the feasibility of the new Fourier Adaptive Waveform modulation which has been patented [62].
      - Contribution of PANAMA: to provide initial training to the recruited engineer in charge of the implementation.
      - *Partners:* *PANAMA, IETR.*
      - Funding: 18 months of engineer, hosted by IETR.

- **SPARSE (2019)**

- **Sparse representations in continuous dictionaries**

- **Participants:** Rémi Gribonval, Clément Elvira, Clément Merdrignac.

- This short exploratory action aims to explore the new paradigm of sparse representations in “continuous” dictionaries.
      - Contribution of PANAMA: to design algorithms for the sparse representation problem in continuous dictionaries with theoretical success guarantees.
      - *Partners:* *PANAMA, SIMSMART (Inria-Rennes), ENSTA Bretagne, IMT Atlantique.*
      - Funding: 5.6kEuros (internship + travel)

### 8.1.2. ANR INVATE project with IRT b-com, Rennes

**Participants:** Rémi Gribonval, Nancy Bertin, Mohammed 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.

### 8.1.3. ANR OATMIL project

**Participants:** Rémi Gribonval, Antoine Chatalic, Nicolas Courty.

*Duration: 4 years (2017-2021)*

*Acronym: OATMIL (Bringing Optimal Transport and Machine Learning Together)*

*<http://people.irisa.fr/Nicolas.Courty/OATMIL/>*

*Research Axis 3.1*

*Partners: Obelix team and PANAMA Inria project-team, IRISA; LITIS, Rouen; Lagrange Laboratory, Nice; Technicolor R&I France, Rennes.*

*Coordinator: Nicolas Courty (Obelix team)*

*Description: The OATMIL project will propose novel concepts, methodologies, and new tools for exploiting large data collections. This will result from a cross-fertilization of fundamental tools and ideas from optimal transport (OT) and machine learning (ML). The main objective of OATMIL is to develop new techniques for large-scale machine learning, encompassing adaptability, scalability, and robustness, by a cross-fertilization of ideas coming from OT and ML. This cross-fertilization leads to two complementary scientific challenges : bringing OT to ML and bringing ML to OT.*

*Contribution of PANAMA: PANAMA will explore the use of dimension-reduction with sketching strategies in the context compressive optimal transport.*

*Funding: ANR*

### 8.1.4. Collaboration with 5th dimension – dynamic separation of localized sound sources

**Participants:** Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon.

*Duration: 1 year (2018-2019)*

*Research axis: 3.2*

*Partner: 5<sup>th</sup> dimension (<https://5dim.com/>)*

*Funding: LABEX AMIEX (<https://www.agence-maths-entreprises.fr/a/>)*

After a first phase of this contract which involved porting in C++ a subset of our source localization library Multichannel BSS Locate (Oct.-Nov. 2018, in collaboration with InriaTech), a second phase was realized in 2019 with support from LABEX AMIES. We specified and recorded new data adapted to the partner's use case (microphones on glasses temples) and investigated the interplay between localization and separation, using the FASST library, on simulated and real data recorded with a prototype.

## 8.2. International Initiatives

### 8.2.1. Inria International Partners

#### 8.2.1.1. Informal International Partners

Nancy Bertin is “external collaborator” of the MERLIN project (project between the Acoustics Research Institute of the Austrian Academy of Sciences and the Signal Processing Laboratory at Brno University of Technology.)

## 8.3. International Research Visitors

### 8.3.1. Visits of International Scientists

#### 8.3.1.1. Internships

- Pavel Závíška and Ondřej Mokřý, visiting students from Brno University of Technology, in December 2018 (within the MERLIN collaboration).
- Andersen Man Shun Ang, visiting student from University of Mons, in February 2019.

## 9. Dissemination

### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific Events: Organisation

- Nancy Bertin, event coordinator of the 2019 Science and Music Day (Journée Science et Musique) organized by IRISA (with Claire Cury from EMPENN project-team.)
- Nancy Bertin, member of the LVA/ICA conference steering committee.

#### 9.1.2. Scientific Events: Selection

##### 9.1.2.1. Chair of Conference Program Committees

- Nancy Bertin, coordinator of the Science and Music Young Researcher Award (Prix Jeune Chercheur Science et Musique).
- Rémi Gribonval and Frédéric Bimbot, scientific coordinators of the Science and Music Day (Journée Science et Musique) organized by IRISA.

##### 9.1.2.2. Member of Conference Program Committees

- Rémi Gribonval, member of the program committee of the GRETSI.

#### 9.1.3. Membership of Editorial Boards

- Frédéric Bimbot, Editor-in-Chief of the international journal *Speech Communication*.
- Rémi Gribonval, Associate Editor of the international journal *Constructive Approximation*.
- Rémi Gribonval, Senior Area Editor of the *IEEE Transactions on Signal Processing*.
- Nancy Bertin, Guest Editor of a special issue in the international journal *IEEE Journal of Selected Topics in Signal Processing*.

#### 9.1.4. Invited Talks

- Jérémy Cohen, invited speaker at workshop "Low Rank Optimization and Applications" (Max Plank Institute for the Mathematics in the Sciences, Leipzig, Germany, April 2019)<sup>2</sup>
- Jérémy Cohen, invited speaker at workshop "AI and Tensor Factorizations" (Santa Fe, USA, September 2019)
- Rémi Gribonval, keynote at SampTA 2019 (Sampling Theory and Applications) in Bordeaux, France.
- Rémi Gribonval, tutorial at Summer School "Sparsity for Physics, Signal and Learning", Inria Paris, June 2019
- Rémi Gribonval, invited talk at 4TU meeting on "Mathematics of Deep Learning", Delft University, Nov 2019
- Rémi Gribonval, invited talk at Alan Turing Institute on "Mathematics of Data", London, May 2019
- Rémi Gribonval, invited talk at GDR ISIS meeting on "Theory of Deep Learning", Paris, Oct 2019
- Rémi Gribonval, panel session on "Signal processing and AI" at GRETSI 2019, Lille, Aug 2019

<sup>2</sup><https://www.mis.mpg.de/calendar/conferences/2019/lroa2019.html>

### 9.1.5. Scientific Expertise

- Rémi Gribonval, vice-president of the Scientific Advisory Board of the Acoustics Research Institute from the Austrian Academy of Sciences in Vienna.
- Frédéric Bimbot, member of the International Advisory Council of ISCA (International Speech Communication Association).
- Frédéric Bimbot, member of the ANR Expert Scientific Committee n°38, dedicated to the Digital Revolution and its relations to Knowledge and Culture.
- Rémi Gribonval, member of the EURASIP Special Area Team (SAT) on Signal and Data Analytics for Machine Learning (SiG-DML) since 2015.

### 9.1.6. Research Administration

- Frédéric Bimbot, board member of the GDR MADICS
- Rémi Gribonval, member of the organization committee of the 2019 **GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing**.
- Nancy Bertin, elected member of the “Comité de Centre” at Inria Rennes Bretagne Atlantique research center.
- Nancy Bertin, coordinator of IRISA “Arts, Heritage and Culture crosscutting axis” (*with Valérie Gouranton, HYBRID project-team.*)

## 9.2. Teaching

Master : N. Bertin, coordination of the VAI module "Vocal and Acoustic Interactions" within the SIF M2, 20 hours, Université de Rennes 1, France

Master : N. Bertin, "Vocal and Audio Interactions", 4 hours, M2, Université Rennes 1, France.

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

Master : N. Bertin, "Foundations of Smart Sensing", 9 hours, M2, Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI), 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, "High dimensional statistical learning", 12 hours, M2, Université Rennes 1, France.

Master: R. Gribonval, coordination of the HDL module "High dimensional statistical learning" within the SIF M2, 20 hours, Université Rennes 1, France.

Master : R. Gribonval, "Sparsity in Signal and Image Processing", 8 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.

Master : J. Cohen, "Parcimony", 9 hours, ENSAI / "Smart Sensing" course

Master : J. Cohen, "Sparsity in Signal and Image Processing", 19 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France

## 9.3. Popularization

### 9.3.1. Journée Science et Musique (JSM 2019)

**Participants:** Nancy Bertin, Rémi Gribonval, Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Stéphanie Gosselin-Lemaile, Corentin Louboutin, Antoine Chatalic, Clément Gaultier, Cássio Fraga Dantas, Diego Di Carlo.

with contributions and support from: Claire Cury, Elisabeth Lebret, Valérie Gouranton, Ronan Gaugne, Florian Nouviale, Evelyne Orain, Agnès Cottais, Catherine Jacques-Orban and many more.

PANAMA coordinated the organization of a public event called “Journée 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 2019 hosted approximately 300 visitors and was a partner of the “Festival des Sciences” and an official event of CNRS 80th birthday celebrations.

### 9.3.2. Participation to Equinoxes#1

Frédéric Bimbot gave a popularization conference at the **Equinoxes#1** event, organised by Artoutaï (13 October 2019), on "the mysterious spectrum of sound" (4 times during the day)

### 9.3.3. Internal or external responsibilities

Frédéric Bimbot is a member of a working group of the CNRS/DR17 on the topic "changer le regard sur le handicap au travail"

## 10. 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", June 2016, vol. 33, n<sup>o</sup> 5, pp. 631-661 [DOI : 10.1525/MP.2016.33.5.631], <https://hal.inria.fr/hal-01188244>
- [2] M. CHAFII. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, <https://hal.archives-ouvertes.fr/tel-01399509>
- [3] 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<sup>o</sup> 7, pp. 1830–1840 [DOI : 10.1109/TASL.2010.2050716], <https://hal.inria.fr/inria-00541865>
- [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<sup>o</sup> 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<sup>o</sup> 2, pp. 335-348 [DOI : 10.1109/TSP.2015.2480045], <https://hal.inria.fr/hal-01133087>
- [6] S. KITIĆ. *Cospars regularization of physics-driven inverse problems*, IRISA, Inria Rennes, November 2015, <https://hal.archives-ouvertes.fr/tel-01237323>
- [7] C. LOUBOUTIN, F. BIMBOT. *Modeling the multiscale structure of chord sequences using polytopic graphs*, in "18th International Society for Music Information Retrieval Conference", Suzhou, China, October 2017, <https://hal.archives-ouvertes.fr/hal-01653455>

- [8] S. NAM, M. E. DAVIES, M. ELAD, R. GRIBONVAL. *The Cospase Analysis Model and Algorithms*, in "Applied and Computational Harmonic Analysis", 2013, vol. 34, n<sup>o</sup> 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>
- [9] 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<sup>o</sup> 4, pp. 1118 - 1133, 16, <http://hal.inria.fr/hal-00626962>
- [10] H. PEIC TUKULJAC, A. DELEFORGE, R. GRIBONVAL. *MULAN: A Blind and Off-Grid Method for Multi-channel Echo Retrieval*, in "NIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, pp. 1-11, <https://arxiv.org/abs/1810.13338> , <https://hal.inria.fr/hal-01906385>
- [11] E. VINCENT, N. BERTIN, R. GRIBONVAL, F. BIMBOT. *From blind to guided audio source separation*, in "IEEE Signal Processing Magazine", December 2013, <http://hal.inria.fr/hal-00922378>

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

- [12] C. GAULTIER. *Design and evaluation of sparse models and algorithms for audio inverse problems*, Université Rennes 1, January 2019, <https://tel.archives-ouvertes.fr/tel-02148598>
- [13] C. LOUBOUTIN. *Multi-scale and multi-dimensional modelling of music structure using polytopic graphs*, Université Rennes 1, March 2019, <https://tel.archives-ouvertes.fr/tel-02149728>

### Articles in International Peer-Reviewed Journals

- [14] N. BERTIN, E. CAMBERLEIN, R. LEBARBENCHON, E. VINCENT, S. SIVASANKARAN, I. ILLINA, F. BIMBOT. *VoiceHome-2, an extended corpus for multichannel speech processing in real homes*, in "Speech Communication", January 2019, vol. 106, pp. 68-78 [DOI : 10.1016/J.SPECOM.2018.11.002], <https://hal.inria.fr/hal-01923108>
- [15] E. BYRNE, A. CHATALIC, R. GRIBONVAL, P. SCHNITER. *Sketched Clustering via Hybrid Approximate Message Passing*, in "IEEE Transactions on Signal Processing", September 2019, vol. 67, n<sup>o</sup> 17, pp. 4556-4569, <https://arxiv.org/abs/1712.02849> [DOI : 10.1109/TSP.2019.2924585], <https://hal.inria.fr/hal-01991231>
- [16] J. E. COHEN, N. GILLIS. *Identifiability of Complete Dictionary Learning*, in "SIAM Journal on Mathematics of Data Science", 2019, vol. 1, n<sup>o</sup> 3, pp. 518–536, <https://arxiv.org/abs/1808.08765> [DOI : 10.1137/18M1233339], <https://hal.archives-ouvertes.fr/hal-02183578>
- [17] C. CURY, P. MAUREL, R. GRIBONVAL, C. BARILLOT. *A sparse EEG-informed fMRI model for hybrid EEG-fMRI neurofeedback prediction*, in "Frontiers in Neuroscience", January 2020 [DOI : 10.3389/FNINS.2019.01451], <https://www.hal.inserm.fr/inserm-02090676>
- [18] A. DELEFORGE, D. DI CARLO, M. STRAUSS, R. SERIZEL, L. MARCENARO. *Audio-Based Search and Rescue with a Drone: Highlights from the IEEE Signal Processing Cup 2019 Student Competition*, in "IEEE Signal Processing Magazine", September 2019, vol. 36, n<sup>o</sup> 5, pp. 138-144, <https://arxiv.org/abs/1907.04655> [DOI : 10.1109/MSP.2019.2924687], <https://hal.archives-ouvertes.fr/hal-02161897>

- [19] I. DOKMANIĆ, R. GRIBONVAL. *Concentration of the Frobenius norm of generalized matrix inverses*, in "SIAM Journal on Matrix Analysis and Applications", 2019, vol. 40, n<sup>o</sup> 1, pp. 92–121, <https://arxiv.org/abs/1810.07921> - Revised/condensed/renamed version of preprint "Beyond Moore-Penrose Part II: The Sparse Pseudoinverse", forthcoming [DOI : 10.1137/17M1145409], <https://hal.inria.fr/hal-01897046>
- [20] S. FOUCART, R. GRIBONVAL, L. JACQUES, H. RAUHUT. *Jointly Low-Rank and Bisparse Recovery: Questions and Partial Answers*, in "Analysis and Applications", 2020, vol. 18, n<sup>o</sup> 01, pp. 25–48, <https://arxiv.org/abs/1902.04731> [DOI : 10.1142/S0219530519410094], <https://hal.inria.fr/hal-02062891>
- [21] C. FRAGA DANTAS, R. GRIBONVAL. *Stable safe screening and structured dictionaries for faster L1 regularization*, in "IEEE Transactions on Signal Processing", July 2019, vol. 67, n<sup>o</sup> 14, pp. 3756-3769, <https://arxiv.org/abs/1812.06635> [DOI : 10.1109/TSP.2019.2919404], <https://hal.inria.fr/hal-01954261>
- [22] R. GRIBONVAL, M. NIKOLOVA. *On bayesian estimation and proximity operators*, in "Applied and Computational Harmonic Analysis", 2019, pp. 1-25, <https://arxiv.org/abs/1807.04021> , forthcoming [DOI : 10.1016/J.ACHA.2019.07.002], <https://hal.inria.fr/hal-01835108>

### Invited Conferences

- [23] R. GRIBONVAL, G. KUTYNIOK, M. NIELSEN, F. VOIGTLAENDER. *Approximation spaces of deep neural networks*, in "SMAI 2019 - 9<sup>ème</sup> Biennale des Mathématiques Appliquées et Industrielles", Guidel, France, May 2019, 1 p. , <https://hal.inria.fr/hal-02127179>
- [24] V. SCHELLEKENS, A. CHATALIC, F. HOUSSIAU, Y.-A. DE MONTJOYE, L. JACQUES, R. GRIBONVAL. *Differentially Private Compressive k-Means*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 7933-7937 [DOI : 10.1109/ICASSP.2019.8682829], <https://hal.inria.fr/hal-02060208>

### International Conferences with Proceedings

- [25] A. CHATALIC, N. KERIVEN, R. GRIBONVAL. *Projections aléatoires pour l'apprentissage compressif*, in "GRETSI 2019 - XXVII<sup>ème</sup> Colloque francophone de traitement du signal et des images", Lille, France, August 2019, pp. 1-4, <https://hal.inria.fr/hal-02154803>
- [26] J. E. COHEN, N. GILLIS. *Nonnegative Low-rank Sparse Component Analysis*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 8226-8230 [DOI : 10.1109/ICASSP.2019.8682188], <https://hal.archives-ouvertes.fr/hal-02201471>
- [27] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL. *Hyperspectral Image Denoising using Dictionary Learning*, in "WHISPERS 2019 - 10th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing", Amsterdam, Netherlands, September 2019, pp. 1-5, <https://hal.inria.fr/hal-02175630>
- [28] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL. *Learning Tensor-structured Dictionaries with Application to Hyperspectral Image Denoising*, in "EUSIPCO 2019 - 27th European Signal Processing Conference", A Coruña, Spain, September 2019, pp. 1-5, <https://hal.inria.fr/hal-02126782>
- [29] D. DI CARLO, A. DELEFORGE, N. BERTIN. *Mirage: 2D Source Localization Using Microphone Pair Augmentation with Echoes*, in "ICASSP 2019 - IEEE International Conference on Acoustic, Speech Signal

Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 775-779, <https://arxiv.org/abs/1906.08968> [DOI : 10.1109/ICASSP.2019.8683534], <https://hal.archives-ouvertes.fr/hal-02160940>

- [30] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN. *Uniform  $k$ -step recovery with CMF dictionaries*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, pp. 1-2, <https://hal.inria.fr/hal-02157561>
- [31] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *OMP and continuous dictionaries: Is  $k$ -step recovery possible ?*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 1-5 [DOI : 10.1109/ICASSP.2019.8683617], <https://hal.archives-ouvertes.fr/hal-02049486>
- [32] V. GILLOT, F. BIMBOT. *Polytopic reconfiguration: a graph-based scheme for the multiscale transformation of music segments and its perceptual assessment*, in "SMC 2019 - 16th Sound & Music Computing Conference", Malaga, Spain, May 2019, pp. 1-8, <https://hal.archives-ouvertes.fr/hal-02132955>
- [33] S. GUPTA, R. GRIBONVAL, L. DAUDET, I. DOKMANIĆ. *Don't take it lightly: Phasing optical random projections with unknown operators*, in "NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, pp. 1-13, <https://arxiv.org/abs/1907.01703> , <https://hal.inria.fr/hal-02342280>
- [34] M. HAFSATI, N. EPAIN, R. GRIBONVAL, N. BERTIN. *Sound source separation in the higher order ambisonics domain*, in "DAFx 2019 - 22nd International Conference on Digital Audio Effects", Birmingham, United Kingdom, September 2019, pp. 1-7, <https://hal.inria.fr/hal-02161949>
- [35] V. SCHELLEKENS, A. CHATALIC, F. HOUSSIAU, Y.-A. DE MONTJOYE, L. JACQUES, R. GRIBONVAL. *Compressive  $k$ -Means with Differential Privacy*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, pp. 1-2, <https://hal.inria.fr/hal-02154820>

### National Conferences with Proceedings

- [36] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *Identification de supports en  $k$  étapes avec OMP pour les dictionnaires continus*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, pp. 1-4, <https://hal.inria.fr/hal-02157571>

### Conferences without Proceedings

- [37] A. ANG, J. E. COHEN, N. GILLIS. *Accelerating Approximate Nonnegative Canonical Polyadic Decomposition using Extrapolation*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, pp. 1-4, <https://hal.archives-ouvertes.fr/hal-02143969>
- [38] C. CURY, P. MAUREL, R. GRIBONVAL, C. BARILLOT. *Can we learn from coupling EEG-fMRI to enhance neuro-feedback in EEG only?*, in "OHBM 2019 - Annual Meeting Organization for Human Brain Mapping", Rome, Italy, June 2019, 1 p. , <https://www.hal.inserm.fr/inserm-02074623>
- [39] C. CURY, P. MAUREL, G. LIOI, R. GRIBONVAL, C. BARILLOT. *Learning bi-modal EEG-fMRI neurofeedback to improve neurofeedback in EEG only*, in "Real-Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, pp. 1-2 [DOI : 10.1101/599589], <https://www.hal.inserm.fr/inserm-02368720>



- [40] A. LORENTE MUR, M. OCHOA, J. E. COHEN, X. INTES, N. DUCROS. *Handling negative patterns for fast single-pixel lifetime imaging*, in "2019 - Molecular-Guided Surgery: Molecules, Devices, and Applications V", San Francisco, United States, SPIE, February 2019, vol. 10862, pp. 1-10 [DOI : 10.1117/12.2511123], <https://hal.archives-ouvertes.fr/hal-02017598>
- [41] P. STOCK, B. GRAHAM, R. GRIBONVAL, H. JÉGOU. *Equi-normalization of Neural Networks*, in "ICLR 2019 - Seventh International Conference on Learning Representations", New Orleans, United States, May 2019, pp. 1-20, <https://arxiv.org/abs/1902.10416> , <https://hal.archives-ouvertes.fr/hal-02050408>
- [42] P. STOCK, A. JOULIN, R. GRIBONVAL, B. GRAHAM, H. JÉGOU. *And the Bit Goes Down: Revisiting the Quantization of Neural Networks*, in "ICLR 2020 - Eighth International Conference on Learning Representations", Addis-Abeba, Ethiopia, February 2020, pp. 1-11, <https://hal.archives-ouvertes.fr/hal-02434572>
- [43] T. VAYER, R. FLAMARY, R. TAVENARD, L. CHAPEL, N. COURTY. *Sliced Gromov-Wasserstein*, in "NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, vol. 32, <https://arxiv.org/abs/1905.10124> , <https://hal.archives-ouvertes.fr/hal-02174309>

### Scientific Books (or Scientific Book chapters)

- [44] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. *The Spectral Graph Wavelet Transform: Fundamental Theory and Fast Computation*, in "Vertex-Frequency Analysis of Graph Signals", L. STANKOVIĆ, E. SEJDIĆ (editors), Signals and Communication Technology, Springer International Publishing, December 2019, pp. 141-175 [DOI : 10.1007/978-3-030-03574-7\_3], <https://hal.inria.fr/hal-01943589>

### Research Reports

- [45] B. CARAMIAUX, F. LOTTE, J. GEURTS, G. AMATO, M. BEHRMANN, F. BIMBOT, F. FALCHI, A. GARCIA, J. GIBERT, G. GRAVIER, H. HOLKEN, H. KOENITZ, S. LEFEBVRE, A. LIUTKUS, A. PERKIS, R. REDONDO, E. TURRIN, T. VIÉVILLE, E. VINCENT. *AI in the media and creative industries*, New European Media (NEM), April 2019, pp. 1-35, <https://arxiv.org/abs/1905.04175> , <https://hal.inria.fr/hal-02125504>

### Software

- [46] C. F. DANTAS, R. GRIBONVAL. *Stable Screening - Python code*, June 2019  
[SWH-ID : swh:1:dir:3ccb3fde105b05aee192367fb5e07e192e3b6774], Software, <https://hal.inria.fr/hal-02129219>
- [47] M. KOWALSKI, E. VINCENT, R. GRIBONVAL. *Underdetermined Reverberant Source Separation*, October 2019  
[SWH-ID : swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d], Software, <https://hal.archives-ouvertes.fr/hal-02309043>

### Other Publications

- [48] A. ACKAOUY, N. COURTY, E. VALLEE, O. COMMOWICK, C. BARILLOT, F. GALASSI. *Unsupervised Domain Adaptation with Optimal Transport in multi-site segmentation of Multiple Sclerosis lesions from MRI data : Preprint*, October 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02317028>
- [49] B. B. DAMODARAN, R. FLAMARY, V. SEGUY, N. COURTY. *An Entropic Optimal Transport Loss for Learning Deep Neural Networks under Label Noise in Remote Sensing Images*, July 2019, <https://arxiv.org/>

[abs/1810.01163](https://hal.archives-ouvertes.fr/hal-02174320) - Under Consideration at Computer Vision and Image Understanding [DOI : 10.01163], <https://hal.archives-ouvertes.fr/hal-02174320>

- [50] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL. *Tensor-structured Dictionaries for Hyperspectral Imaging*, July 2019, pp. 1-2, SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations, <https://hal.inria.fr/hal-02169405>
- [51] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *When does OMP achieve support recovery with continuous dictionaries?*, April 2019, working paper or preprint, <https://hal.inria.fr/hal-02099464>
- [52] C. ELVIRA, C. HERZET. *Safe Squeezing for Antispase Coding*, November 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02368134>
- [53] R. GRIBONVAL, G. KUTYNIOK, M. NIELSEN, F. VOIGTLAENDER. *Approximation spaces of deep neural networks*, June 2019, <https://arxiv.org/abs/1905.01208> - working paper or preprint, <https://hal.inria.fr/hal-02117139>
- [54] R. GRIBONVAL, M. NIKOLOVA. *A characterization of proximity operators*, November 2019, <https://arxiv.org/abs/1807.04014> - working paper or preprint, <https://hal.inria.fr/hal-01835101>
- [55] A. LORENTE MUR, J. E. COHEN, N. DUCROS. *Factorisation exacte pour la généralisation de motifs en imagerie monopixel*, March 2019, 1 p. , JIONC - 14ème Journées d'Imagerie Optique Non Conventionnelle, Poster, <https://hal.archives-ouvertes.fr/hal-02170148>
- [56] T. VAYER, L. CHAPEL, R. FLAMARY, R. TAVENARD, N. COURTY. *Fused Gromov-Wasserstein distance for structured objects: theoretical foundations and mathematical properties*, July 2019, <https://arxiv.org/abs/1811.02834> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02174316>

## References in notes

- [57] 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<sup>o</sup> 3, pp. 922 - 932 [DOI : 10.1109/TASL.2011.2168211], <http://hal.inria.fr/inria-00577079>
- [58] F. BIMBOT. *Towards an Information-Theoretic Framework for Music Structure*, 2016, pp. 167-168, Invited talk at Dagstuhl Seminar 16092 on Computational Music Structure Analysis (Ed. : M. Müller, E. Chew, J.P. Bello) [DOI : 10.4230/DAGREP.6.2.147], <https://hal.archives-ouvertes.fr/hal-01421013>
- [59] A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL. *Dynamic Screening: Accelerating First-Order Algorithms for the Lasso and Group-Lasso*, in "IEEE Transactions on Signal Processing", 2015, vol. 63, n<sup>o</sup> 19, 20 p. [DOI : 10.1109/TSP.2015.2447503], <https://hal.archives-ouvertes.fr/hal-01084986>
- [60] A. BOURRIER. *Compressed sensing and dimensionality reduction for unsupervised learning*, Université Rennes 1, May 2014, <https://tel.archives-ouvertes.fr/tel-01023030>
- [61] E. BYRNE, R. GRIBONVAL, P. SCHNITER. *Sketched Clustering via Hybrid Approximate Message Passing*, in "Asilomar Conference on Signals, Systems, and Computers", Pacific Grove, California, United States, October 2017, <https://hal.inria.fr/hal-01650160>

- [62] M. CHAFII, J. PALICOT, R. GRIBONVAL. *Dispositif de communication à modulation temps-fréquence adaptative*, July 2016, n<sup>o</sup> Numéro de demande : 1656806 ; Numéro de soumission : 1000356937, <https://hal.inria.fr/hal-01375661>
- [63] A. CHATALIC, R. GRIBONVAL, N. KERIVEN. *Large-Scale High-Dimensional Clustering with Fast Sketching*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 4714-4718 [DOI : 10.1109/ICASSP.2018.8461328], <https://hal.inria.fr/hal-01701121>
- [64] I. DOKMANIĆ, R. GRIBONVAL. *Beyond Moore-Penrose Part I: Generalized Inverses that Minimize Matrix Norms*, July 2017, working paper or preprint, <https://hal.inria.fr/hal-01547183>
- [65] I. DOKMANIĆ, R. PARHIZKAR, A. WALTHER, Y. M. LU, M. VETTERLI. *Acoustic echoes reveal room shape*, in "Proceedings of the National Academy of Sciences", 2013, vol. 110, n<sup>o</sup> 30, pp. 12186–12191, <http://dx.doi.org/10.1073/pnas.1221464110>
- [66] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN. *A case of exact recovery with OMP using continuous dictionaries*, in "CS 2018 - 9th International Conference on Curves and Surfaces", Arcachon, France, June 2018, <https://hal.inria.fr/hal-01937532>
- [67] C. FRAGA DANTAS, J. E. COHEN, R. GRIBONVAL. *Learning fast dictionaries for sparse representations using low-rank tensor decompositions*, in "LVA/ICA 2018 - 14th International Conference on Latent Variable Analysis and Signal Separation", Guildford, United Kingdom, LNCS, Springer, July 2018, vol. 10891, pp. 456-466 [DOI : 10.1007/978-3-319-93764-9\_42], <https://hal.inria.fr/hal-01709343>
- [68] C. FRAGA DANTAS, R. GRIBONVAL. *Dynamic Screening with Approximate Dictionaries*, in "XXVIème colloque GRETSI", Juan-les-Pins, France, September 2017, <https://hal.inria.fr/hal-01598021>
- [69] C. GAULTIER, N. BERTIN, S. KITIĆ, R. GRIBONVAL. *A modeling and algorithmic framework for (non)social (co)sparse audio restoration*, November 2017, working paper or preprint, <https://hal.inria.fr/hal-01649261>
- [70] M. A. GERZON. *Periphony: With-Height Sound Reproduction*, in "J. Audio Eng. Soc", 1973, vol. 21, n<sup>o</sup> 1, pp. 2–10, <http://www.aes.org/e-lib/browse.cfm?elib=2012>
- [71] R. GRIBONVAL, G. BLANCHARD, N. KERIVEN, Y. TRAONMILIN. *Compressive Statistical Learning with Random Feature Moments*, December 2017, Main novelties compared to version 1: improved concentration bounds, improved sketch sizes for compressive k-means and compressive GMM that now scale linearly with the ambient dimension, <https://hal.inria.fr/hal-01544609>
- [72] R. GRIBONVAL. *Should penalized least squares regression be interpreted as Maximum A Posteriori estimation?*, in "IEEE Transactions on Signal Processing", May 2011, vol. 59, n<sup>o</sup> 5, pp. 2405-2410 [DOI : 10.1109/TSP.2011.2107908], <http://hal.inria.fr/inria-00486840>
- [73] H. JAIN. *Learning compact representations for large scale image search*, Université Rennes 1, June 2018, <https://tel.archives-ouvertes.fr/tel-01889405>
- [74] N. KERIVEN. *Sketching for large-scale learning of mixture models*, Université Rennes 1, October 2017, <https://tel.archives-ouvertes.fr/tel-01620815>

- [75] 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>
- [76] L. LE MAGOAROU, R. GRIBONVAL, N. TREMBLAY. *Approximate fast graph Fourier transforms via multi-layer sparse approximations*, in "IEEE transactions on Signal and Information Processing over Networks", June 2018, vol. 4, n<sup>o</sup> 2, pp. 407–420, <https://arxiv.org/abs/1612.04542> [DOI : 10.1109/TSIPN.2017.2710619], <https://hal.inria.fr/hal-01416110>
- [77] R. LEBARBENCHON, E. CAMBERLEIN, D. DI CARLO, C. GAULTIER, A. DELEFORGE, N. BERTIN. *Evaluation of an Open-Source Implementation of the SRP-PHAT Algorithm within the 2018 Locata Challenge*, in "LOCATA Challenge Workshop, a satellite event of IWAENC 2018", Tokyo, Japan, September 2018, <https://hal.archives-ouvertes.fr/hal-02187964>
- [78] N. LIBERMANN, F. BIMBOT, E. VINCENT. *Exploration de dépendances structurelles mélodiques par réseaux de neurones récurrents*, in "JIM 2018 - Journées d'Informatique Musicale", Amiens, France, May 2018, pp. 81-86, <https://hal.archives-ouvertes.fr/hal-01791381>
- [79] 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>
- [80] C. LOUBOUTIN, F. BIMBOT. *Modeling the multiscale structure of chord sequences using polytopic graphs*, in "18th International Society for Music Information Retrieval Conference", Suzhou, China, October 2017, <https://hal.archives-ouvertes.fr/hal-01653455>
- [81] C. LOUBOUTIN, F. BIMBOT. *Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments*, in "6th International Conference on Mathematics and Computation in Music (MCM 2017)", Mexico City, Mexico, O. A. AGUSTÍN-AQUINO, E. LLUIS-PUEBLA, M. MONTIEL (editors), Lecture Notes in Computer Science book series, Springer, June 2017, vol. 10527, <https://hal.archives-ouvertes.fr/hal-01653445>
- [82] S. NOORZADEH, P. MAUREL, T. OBERLIN, R. GRIBONVAL, C. BARILLOT. *Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints*, in "MICCAI 2017 - 20th International Conference on Medical Image Computing and Computer Assisted Intervention", Quebec, Canada, September 2017 [DOI : 10.1007/978-3-319-66182-7\_51], <https://hal.inria.fr/hal-01586495>
- [83] H. PEIC TUKULJAC, A. DELEFORGE, R. GRIBONVAL. *MULAN: A Blind and Off-Grid Method for Multi-channel Echo Retrieval*, in "NIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, pp. 1-11, <https://arxiv.org/abs/1810.13338> , <https://hal.inria.fr/hal-01906385>
- [84] 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>
- [85] R. SCHEIBLER, D. DI CARLO, A. DELEFORGE, I. DOKMANIĆ. *Separake: Source Separation with a Little Help From Echoes*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, April 2018, <https://hal.inria.fr/hal-01909531>