Activity Report 2017

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

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
A5.9.1. - 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

Other Research Topics and Application Domains:

B1.2. - Neuroscience and cognitive science
B2.5.1. - Sensorimotor disabilities
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
B9.6. - Reproducibility
B9.9.1. - Environmental risks

1. Personnel

Research Scientists

Rémi Gribonval [Team leader, Inria, Senior Researcher, HDR]
Nancy Bertin [CNRS, Researcher]
Frédéric Bimbot [CNRS, Senior Researcher, HDR]
Antoine Deleforge [Inria, Researcher]
Pierre Vandergheynst [EPFL, Senior Researcher, Chaire Internationale Inria]

Post-Doctoral Fellows

Clément Elvira [Inria, from Dec 2017]
Yann Traonmilin [Inria, until Aug 2017]

PhD Students

Antoine Chatalic [Univ de Rennes I, from Sep 2017]
Diego Di Carlo [Inria, from Oct 2017]
Cássio Fraga Dantas [Inria]
Clément Gaultier [Inria]
Valentin Gillot [INSA Rennes, from Oct 2017]
Mohammed Hafsati [Institut de recherche technologique B-com]
Himalaya Jain [Technicolor]
Nicolas Keriven [Univ de Rennes I, until Sep 2017]
Nathan Libermann [Inria, joint with MULTISPEECH team; in PANAMA from Sep 2017]
Corentin Louboutin [Univ de Rennes I]

Technical staff

Nicolas Bellot [Inria, until Apr 2017]
Ewen Camberlein [CNRS, until Jul 2017; Inria, since Sep 2017]
Romain Lebarbenchon [CNRS, until Jul 2017; Inria, since Sep 2017]

Interns

Antoine Chatalic [Ecole normale supérieure de Rennes, from Feb 2017 until Jun 2017]
Valentin Gillot [CNRS, from Mar 2017 until Jul 2017]
Victor Miguet [Ecole normale supérieure de Rennes, from May 2017 until Jul 2017]
Victor Miguet [Ecole normale supérieure de Rennes, from Sep 2017]
Eric Zajler Grinstein [Inria, from Oct 2017]

Administrative Assistant

Stéphanie Lemaile [Inria]

Visiting Scientists

Flavio Castro Alves Teixeira [Univ Insbruck, from May 2017 until Jun 2017]
Andreas Loukas [EPFL, from Dec 2017]
Helena Peic Tukuljac [EPFL, from Oct 2017]

Martin Strauss [Univ Erlangen-Nuremberg, from Oct 2017]
Corentin Guichaoua [until Sep 2017]

2. Overall Objectives

2.1. Overall positioning

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

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

2.2. Scientific foundations

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

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

2.3. Applications

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

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

3. Research Program

3.1. Axis 1: Sparse Models and Representations

3.1.1. Efficient Sparse Models and Dictionary Design for Large-scale Data

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

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

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

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

3.1.2. Compressive Learning

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

3.2. Axis 2: Robust Acoustic Scene Analysis

3.2.1. Compressive Acquisition and Processing of Acoustic Scenes

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

3.2.2. Robust Audio Source Separation

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

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

In parallel to these efforts, expected progress in sparse modeling for inverse problems shall bring new approaches to source separation and modeling, as well as to source localization, which is often an important first step in a source separation workflow.

3.2.3. Robust Audio Source Localization

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

3.3. Axis 3: Large-scale Audio Content Processing and Self-organization

3.3.1. Motif Discovery in Audio Data

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

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

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

3.3.2. Structure Modeling and Inference in Audio and Musical Contents

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

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

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

4. Application Domains

4.1. Acoustic Scene Capture

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

4.2. Audio Signal Separation in Reverberant Environments

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

4.3. Multimedia Indexing

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

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

EDF and the French Academy of Technology awarded the 2017 *Paul Caseau Ph.D. prize to Luc Le Magoarou* for his work on "Efficient Matrices for signal processing and machine learning" defended in 2016 [84].

The thesis of Luc Le Magoarou [84] was also awarded a special mention of the AFRIF (the French Association for Shape Recognition and Interpretation) annual Ph.D. prize.

Nicolas Keriven has been awarded the Best Student Paper Award at the international workshop SPARS 2017 .

A 2017 EURASIP best paper award was awarded to the paper *Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques*, co-authored by Gilles Puy, Pierre Vandergheynst, Rémi Gribonval and Yves Wiaux, published in EURASIP Journal on Advances in Signal Processing in 2012 [93].

BEST PAPER AWARD:

[46]
N. KERIVEN, R. GRIBONVAL, G. BLANCHARD, Y. TRAONMILIN. *Random Moments for Sketched Mixture Learning*, June 2017, SPARS2017 - Signal Processing with Adaptive Sparse Structured Representations workshop, https://hal.inria.fr/hal-01494045

6. New Software and Platforms

6.1. VoiceHome Corpus

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

- Participants: Ewen Camberlein, Romain Lebarbenchon, Nancy Bertin and Frédéric Bimbot
- Contact: Nancy Bertin
- URL: http://voice-home.gforge.inria.fr/voiceHome_corpus.html

6.2. 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: 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 and Adrien Leman
- 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 - FAμST: 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/

6.3. SketchMLBox

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

- Authors: Nicolas Keriven, Nicolas Tremblay and Rémi Gribonval
- Partner: Université de Rennes 1
- Contact: Rémi Gribonval
- Publications: Sketching for Large-Scale Learning of Mixture Models - Compressive K-means - Spikes super-resolution with random Fourier sampling - Sketching for large-scale learning of mixture models - Blind Source Separation Using Mixtures of Alpha-Stable Distributions - Sketching for Large-Scale Learning of Mixture Models - Compressive Gaussian Mixture Estimation by Orthogonal Matching Pursuit with Replacement
- URL: http://sketchml.gforge.inria.fr

6.4. 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:

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

- Srdan Kitic, Nancy Bertin, Remi Gribonval. Sparsity and cosparsity for audio declipping: a flexible non-convex approach. LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation, Aug 2015, Liberec, Czech Republic.
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, A||go, to test it on the web.
NEWS OF THE YEAR: In 2017, a web interface to demonstrate the potential of SPADE has been setup using the A||go platform.

- Participants: Nancy Bertin, Clement Gaultier, Ewen Camberlein, Romain Lebarbenchon, Rémi Gribonval and Srdan Kitic
- Contact: Rémi Gribonval
- Publications: Audio Declipping by Cosparse Hard Thresholding - Sparsity and cosparsity for audio declipping: a flexible non-convex approach
- URL: http://spade.inria.fr/

6.5. 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 (current version) 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/ .

A new version is currently under development in the PANAMA team through the Inria funded ADT "FFWD" (FASST For Wider Dissemination) and will be released in 2018.
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.

- 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/

6.6. PHYSALIS

KEYWORDS: Source localization - Cosparsity
SCIENTIFIC DESCRIPTION: PHYSALIS (Physics-Driven Cosparse Analysis) gathers algorithms for (joint) source localization and estimation, expressed as inverse problems and addressed with co-sparse regularization. A particular emphasis is put on the acoustic and EEG settings.
FUNCTIONAL DESCRIPTION: PHYSALIS is distributed as a set of Matlab routines to reproduce experimental results from the Ph.D. thesis of Srdan Kitic.
NEWS OF THE YEAR: In 2017, the code of PHYSALIS has been packaged at the occasion of the writing of an overview chapter on co-sparse source localization.

- Participants: Laurent Albera, Nancy Bertin, Rémi Gribonval and Srdan Kitic
- Contact: Rémi Gribonval
- Publications: Physics-driven inverse problems made tractable with cosparse regularization - Cosparse regularization of physics-driven inverse problems - Versatile and scalable cosparse methods for physics-driven inverse problems - Hearing behind walls: localizing sources in the room next door with cosparsity - Sparse Acoustic Source Localization with Blind Calibration for Unknown Medium Characteristics - The best of both worlds: synthesis-based acceleration for physics-driven cosparse regularization
- URL: http://cosoloc.gforge.inria.fr/

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 has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems, and dimension reduction.

7.1.1. Algorithmic and Theoretical results on Computational Representation Learning

Participants: Rémi Gribonval, Nicolas Bellot, Cássio Fraga Dantas.

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. Inspired by usual fast transforms, we proposed a general dictionary structure (called FAμST for Flexible Approximate Multilayer Sparse Transforms) that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation.

The principle and its application to image denoising appeared at ICASSP 2015 [81] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [80]. A Matlab library has been released (see FAμST in Section 6.2) to reproduce the experiments from the comprehensive journal paper published in 2016 [83], which additionally includes theoretical results on the improved sample complexity of learning such dictionaries. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou on this topic [84].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms. A conference paper on this latter topic appeared in ICASSP 2016 [82], and a journal paper has been published this year [17] where we empirically show that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opens the way to substantial accelerations for Fourier Transforms on large graphs. The approximation error of such Fast Graph Fourier Transforms has been studied in a conference paper [31].

A C++ version of the FAμST software library has been developed (see Section 6) to release the resulting algorithms and interface them with both Matlab and Python (work in progress).

As a complement to the FAμST structure for matrix approximation, we proposed a learning algorithm that constrains the dictionary to be a sum of Kronecker products of smaller sub-dictionaries. A special case of the proposed structure is the widespread separable dictionary. This approach, named SuKro, was evaluated experimentally on an image denoising application [39].

We combined accelerated matrix-vector multiplications offered by FAuST matrix approximations with dynamic screening [53], that safely eliminates inactive variables to speedup iterative sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. Significant complexity reductions were obtained in comparison to screening rules alone [35].

7.1.2. Theoretical results on 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 [37], 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 $\ell^p \to \ell^q$ norms.

An interesting representative is the sparse pseudoinverse which we studied in much more detail in a second part of this work [38], 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, Nicolas Keriven, Antoine Chatalic, Antoine Deleforge.

Main collaborations: Patrick Perez (Technicolor R&I France, Rennes), Anthony Bourrier (formerly Technicolor R&I France, Rennes; then GIPSA-Lab, Grenoble), Antoine Liutkus (ZENITH Inria project-team, Montpellier), Nicolas Tremblay (GIPSA-Lab, Grenoble), Phil Schniter & Evan Byrne (Ohio State University, USA)

Sketching for Large-Scale Mixture Estimation. When fitting a probability model to voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [54], [57], [55], [56] a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. The proposed algorithm provided good reconstruction and scaled to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of voluminous datasets. It also provided a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum.

During the Ph.D. thesis of Nicolas Keriven [12], we consolidated our extensions to non-isotropic Gaussians, with a new algorithm called CL-OMP [73] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper appeared at ICASSP 2016 [72] and the journal version has been accepted this year [44], accompanied by a toolbox for reproducible research (see SketchMLBox, Section 6.3). Nicolas Keriven was awarded the SPARS 2017 Best Student Paper Award for this work [46].

Sketching for Compressive Clustering and beyond. Last year we started a new endeavor to extend the approach beyond the case of Gaussian Mixture Estimation.

First, we showed empirically that sketching can be adapted to compress a training collection while still allowing large-scale *clustering*. The approach, called "Compressive K-means", uses CL-OMP at the learning stage and is described in a paper published at ICASSP 2017 [23]. In the high-dimensional setting, it is also possible to substantially speedup both the sketching stage and the learning stage by replacing Gaussian random matrices with fast random matrices in the sketching procedure. This has been demonstrated by Antoine Chatalic during his internship and submitted for publication to a conference. An alternative algorithm for cluster recovery from a sketch was proposed this year, 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 [25].

Then, we leveraged the parallel between the mathematical expression of sketched clustering and super-resolution to explore the potential of sketching and CL-OMP for the stable recovery of signals made of few spikes (in the gridless setting) from few random weighted Fourier measurements [48]. We also demonstrated that sketching can be used in blind source localization and separation, by learning mixtures of alpha-stable distributions [45], see details in Section 7.4.2.

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

Participants: Rémi Gribonval, Yann Traonmilin.

Main collaboration: Mike Davies (University of Edinburgh, UK), Gilles Puy (Technicolor R&I France, Rennes),

Inverse problems and compressive sensing in Hilbert spaces.

Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the ℓ^1-norm or low-rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). Last year we published a comprehensive paper [97] establishing generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. We also described a generic technique to construct linear maps from a Hilbert space to $\mathbb{R}^m$ that satisfy the RIP [20]. These results have been surveyed in a book chapter completed this year[49].

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 has been completed and submitted to a journal [42]. Future work will include expliciting the impact of the proposed framework on a wider set of concrete learning problems.

7.1.5. Algorithmic Exploration of Sparse Representations in Virtual Reality and Neurofeedback

Participant: Rémi Gribonval.

Ferran Argelaget & Anatole Lecuyer (HYBRID Inria project-team, Rennes), Saman Noorzadeh, Pierre Maurel & Christian Barillot (VISAGES Inria project-team, Rennes)

In collaboration with the VISAGES team we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [34].

Our work in collaboration with the HYBRID team on sparse dictionary learning for spatial and rotation invariant gesture recognition has been published this year [24]. Our work on multi-modal

7.1.5.1. An Alternative Framework for Sparse Representations: Sparse "Analysis" Models

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Technicolor R & I France, Rennes), Laurent Albera and Siouar Bensaid (LTSI, Univ. Rennes)

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome.

Building on our pioneering work on the cosparse model [71], [90][8], successful applications of this approach to sound source localization, brain imaging and audio restoration have been developed in the team during the last years [74], [76], [75], [51]. Along this line, two main achievements were obtained this year. First, and following the publication in 2016 of a journal paper embedding in a unified fashion our results in source localization [5], we wrote a book chapter (currently in press) gathering our contributions in physics-driven cosparse regularization, including new results and algorithms demonstrating the versatility, robustness and computational efficiency of our methods in realistic, large scale scenarios in acoustics and EEG signal processing [47]. Second, we continued extending the cosparse framework on audio restoration problems: improvements on our released real-time declipping algorithm (A-SPADE - see Section 6), new results on the denoising task [41], [28], and the submission of a journal paper encompassing several denoising and declipping methods in a common framework [40].

7.2. Activities on Waveform Design for Telecommunications

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

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

Participant: Rémi Gribonval.

Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (SCEE team, CentraleSupelec, Rennes)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 9.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii [58], defended in October 2016 and co-supervised by R. Gribonval), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems.

A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, leading to a journal paper [62] and several conference communications [60], [61]. Our characterization of waveforms with optimum PAPR [63] has been published in a journal in 2016 [59]. Our work on the design of new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR which gave rise to a patent in 2016 [64] has been been submitted for publication as a journal paper. Our study of the tradeoffs between PAPR and Power Spectral Density properties of a wavelet modulation scheme has been published this year [14].

7.3. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Locally-Linear Inverse Regression

Participant: Antoine Deleforge.

Main collaborations: Florence Forbes (MISTIS Inria project-team, Grenoble), Emeline Perthame (HUB team, Institut Pasteur, Paris), Vincent Drouard, Radu Horaud, Sileye Ba and Georgios Evangelidis (PERCEPTION Inria project-team, Grenoble)

A general problem in machine learning and statistics is that of *high- to low-dimensional mapping*. In other words, given two spaces $\mathbb{R}^D$ and $\mathbb{R}^L$ with $D \gg L$, how to find a relation between these two spaces such that given a new observation vector $y \in \mathbb{R}^D$ its associated vector $x \in \mathbb{R}^L$ can be estimated? In *regression*, a set of training pairs $\{(y_n, x_n)\}_{n=1}^N$ is used to learn the relation. In *dimensionality reduction*, only vectors $\{y_n\}_{n=1}^N$ are observed, and an intrinsic low-dimensional representation $\{x_n\}_{n=1}^N$ is sought. In [67], we introduced a probabilistic framework unifying both tasks referred to as *Gaussian Locally Linear Mapping* (GLLiM). The key idea is to learn an easier other-way-around locally-linear relationship from x to y using a joint Gaussian Mixture model on x and y. This mapping is then easily reversed via Bayes' inversion. This framework was notably applied to hyperspectral imaging of Mars [65], head pose estimation in images [16], sound source separation and localization [66], and virtually-supervised acoustic space learning (see Section 7.4.4). This year, in [19], we introduced the *Student Locally Linear Mapping* (SLLiM) framework. The use of heavy-tailed Student's t-distributions instead of Gaussian ones leads to more robustness and better regression performance on several datasets.

7.3.2. Phase Estimation in Multichannel Mixtures

Participants: Antoine Deleforge, Yann Traonmilin.

Main collaboration: Angélique Drémeau (ENSTA Bretagne and Lab-STICC, Brest)

The problem of estimating source signals given an observed multichannel mixture is fundamentally ill-posed when the mixing matrix is unknown or when the number of sources is larger that the number of microphones. Hence, prior information on the desired source signals must be incorporated in order to tackle it. An important line of research in audio source separation over the past decade consists in using a model of the source signals' magnitudes in the short-time Fourier domain [9]. Such models can be inferred through, *e.g.*, non-negative matrix factorization [9] or deep neural networks [91]. Magnitudes estimates are often interpreted as instantaneous variances of Gaussian-process source signals, and are combined with Wiener filtering for source separation. In [26], we introduced a shift of this paradigm by considering the *Phase Unmixing* problem: how can one recover the instantaneous phases of complex mixed source signals when their magnitudes and mixing matrix are known? This problem was showed to be NP-hard, and three approaches were proposed to tackle it: a heuristic method, an alternate minimization method, and a convex relaxation into a semi-definite program. The last two approaches were showed to outperform the oracle multichannel Wiener filter in under-determined informed source separation tasks. The latter yielded best results, including the potential for exact source separation in under-determined settings. In [27] we applied this framework to the classical problem of *phase retrieval* with a novel multivariate Von Mises prior on phases. We showed that enforcing this prior yielded more accurate estimates than state-of-the art phase retrieval methods.

7.3.3. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Technicolor R&I France, 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 (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [50]), we proposed over the last two years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [75], [78] [6]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [78] resulted in a software release in 2016.

In 2017, this work was extended towards advanced (co)sparse decompositions, including several forms of structured sparsity in the time-frequency domain and across channels, and towards their application to the denoising task, in addition to the previously introduced declipping task, which we continued to improve. In particular, we investigated the incorporation of the so-called "social" structure constraint [79] into problems regularized by a cosparse prior [28], [41], and exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [40]. A new algorithm for joint declipping of multichannel audio was also derived (one submitted conference publication.)

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 lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [94] 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 [10].

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 voiceHome and INVATE projects (see Section 7.4.1) ; more exploratory work towards new approaches diverging away from local Gaussian modeling (Section 7.4.2) ; formulating and addressing a larger class of problems related to localization and separation, in the context of robotics (Section 7.4.3) and audio scene analysis with machine learning (Section 7.4.4).

7.4.1. Towards Real-world Separation and Remixing Applications

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

Main collaborations: Emmanuel Vincent (MULTISPEECH Inria project-team, Nancy), Nicolas Epain (IRT b<>com, Rennes)

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, we were part of the voiceHome project (2015-2017, see Section 9.1.4), an industrial collaboration 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). We benchmarked, improved and optimized existing localization and separation tools to the particular context of this application, worked on a better interface between source localization and source separations steps and on optimal initialization scenarios, and reduced the latency and computational burden of the previously available tools, highlighting operating conditions were real-time processing is achievable. Automatic selection of the best microphones subset in an array was investigated. A journal publication including new data (extending the voiceHome Corpus, see Section 6.1), baseline tools and results was submitted to a special issue of Speech Communication. Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique), which started in September 2017, for a new development phase of the FASST software (see Section 6.5).

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b<>com with the INVATE project, see Section 9.1.2) 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. We started extending our methods to hybrid acquisition scenarios, where the separation of HOA signals can be informed by the complementary close-up microphonic signals. Future work will include systematic experimental evaluation.

7.4.2. *Beyond the Local Complex Gaussian Model*

Participants: Antoine Deleforge, Nicolas Keriven.

Main collaboration: Antoine Liutkus (ZENITH Inria project-team, Montpellier)

The team has also recently investigated a number of alternative probabilistic models to the local complex Gaussian (LCG) model for audio source separation. An important limit of LCG is that most signals of interest such as speech or music do not exhibit Gaussian distributions but heavier-tailed ones due to their important dynamic [85]. In [45] we proposed a new sound source separation algorithm using heavy-tailed alpha stable priors for source signals. Experiments showed that it outperformed baseline Gaussian-based methods on under-determined speech or music mixtures. Another limitation of LCG is that it implies a zero-mean complex prior on source signals. This induces a bias towards low signal energies, in particular in under-determined settings. With the development of accurate magnitude spectrogram models for audio signals such as nonnegative matrix factorization [92][9] or more recently deep neural networks [91], it becomes desirable to use probabilistic models enforcing strong magnitude priors. In [26], we explored deterministic magnitude models (see section 7.3.2 for details). An approximate and tractable probabilistic version of this referred to as BEADS (Bayesian Expansion Approximating the Donut Shape) is currently under development. The source prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle.

7.4.3. *Applications to Robot Audition*

Participants: Nancy Bertin, Antoine Deleforge, Martin Strauss, Victor Miguet.

Main collaborations: Aly Magassouba, Pol Mordel and François Chaumette (LAGADIC Inria project-team, Rennes), Alexander Schmidt and Walter Kellermann (University of Erlangen-Nuremberg, Germany)

Implicit Localization through Audio-based Control. In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to vision. Yet audio sensing is a valid alternative or complement to vision in robotics, for instance in homing tasks. Most existing works are based on the relative localization of a defined system with respect to a sound source, and the control scheme is generally designed separately from the localization system. In contrast, the approach that we investigated in the context of Aly Magassouba's Ph.D. (defended in December 2016) focused on a sensor-based control approach. A journal paper encompassing and extending the results obtained before 2017 [89], [87], [88] has been submitted to IEEE Transactions on Robotics (accepted with minor revisions). In 2017, we obtained new results on the use of interaural level difference as the only input feature of the servo, with new experimental validation on humanoid robots. A publication about these last results has been submitted to IEEE Robotics and Automation Letters.

Ego-noise Reduction with Motor-Data-Guided Dictionary Learning. Ego-noise reduction is the problem of suppressing the noise a robot caused by its own motions. Such noise degrades the recorded microphone signal such that the robot's auditory capabilities suffer. To suppress it, it is intuitive to use also motor data, since it provides additional information about the robot's joints and thereby the noise sources. In [96], we incorporated motor data to a recently proposed multichannel dictionary algorithm [69]. We applied this to ego-noise reduction on the humanoid robot NAO. At training, a dictionary is learned that captures spatial and spectral characteristics of ego-noise. At testing, nonlinear classifiers are used to efficiently associate the current robot's motor state to relevant sets of entries in the learned dictionary. By this, computational load is reduced by one third in typical scenarios while achieving at least the same noise reduction performance. Moreover, we proposed to train dictionaries on different microphone array geometries and used them for ego-noise reduction while the head on which the microphones are mounted is moving. In such scenarios, the motor-data-guided approach resulted in significantly better performance values.

Sound Source Localization with a Drone. Flying robots or drones have undergone a massive development in recent years. Already broadly commercialized for entertainment purpose, they also underpin a number of exciting future applications such as mail delivery, smart agriculture, archaeology or search and rescue. An important technological challenge for these platforms is that of localizing sound sources in order to better analyse and understand their environment. For instance, how to localize a person crying for help in the context of a natural disaster? This challenge raises a number of difficult scientific questions. How to efficiently embed a microphone array on a drone? How to deal with the heavy ego-noise produced by the drone's motors? How to deal with moving microphones and distant sources? Victor Miguet and Martin Strauss tackled part of these challenges during their masters' internships. A light 3D-printed structure was designed to embed a USB sound card and a cubic 8-microphone array under a Mikrokopter drone that can carry up to 800 g of payload in flights. Noiseless speech and on-flights ego-noise datasets were recorded. The data were precisely annotated with the target source's position, the state of each drone's propellers and the drone's position and velocity. Baseline methods including multichannel Wiener filtering, GCC-PHAT and MUSIC were implemented in both C++ and Matlab and were tested on the dataset. Up to 5° speech localization accuracy in both azimuth and elevation was achieved under heavy-noise conditions (-5 dB signal-to-noise-ratio). We plan to make the datasets and code publicly available in 2018.

7.4.4. Virtually-Supervised Auditory Scene Analysis

Participants: Antoine Deleforge, Nancy Bertin, Diego Di Carlo, Clément Gaultier.

Main collaborations: Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA) and Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Saurabh Kataria (IIT Kanpur, India)

Classical audio signal processing methods strongly rely on a good knowledge of the *geometry* of the audio scene, *i.e.*, what are the positions of the sources, the sensors, and how does the sound propagate between them. The most commonly used *free field* geometrical model assumes that the microphone configuration is perfectly known and that the sound propagates as a single plane wave from each source to each sensor (no reflection or interference). This model is not valid in realistic scenarios where the environment may be unknown, cluttered, dynamic, and include multiple sources, diffuse sounds, noise and/or reverberations. Such difficulties critical hinders sound source separation and localization tasks. In some ongoing work, we showed that the knowledge of a few early acoustic echoes significantly improve sound source separation performance over the free-field model.

Recently, two directions for advanced audio geometry estimation have emerged and were investigated in our team. The first one is physics-driven [47]. This approach explicitly solves the wave propagation equation in a given simplified yet realistic environment assuming that only few sound sources are present, in order to recover the positions of sources, sensors, or even some of the wall absorption properties. Encouraging results were obtained in simulated settings, including "hearing behind walls" [77]. However, these methods rely on approximate models and on partial knowledge of the system (e.g. room dimensions), limiting their real-world applicability so far. The second direction is data-driven. It uses machine learning to bypass the use of a physical model by directly estimating a mapping from acoustic features to source positions, using training data obtained in a real room [66], [68]. These methods can in principle work in arbitrarily complex environments, but they require carefully annotated training datasets. Since obtaining such data is time consuming, the methods are usually working well for one specific room and setup, and are hard to generalize in practice.

We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually acoustic space travelling* (VAST) [22], [30]. The idea is to use a physics-based room-acoustic simulator to generate arbitrary large datasets of room-impulse responses corresponding to various acoustic environments, adapted to the physical audio system at hand. We demonstrated that mappings learned from these data could potentially be used to not only estimate the 3D position of a source but also some acoustical properties of the room [30]. We also showed that a virtually-learned mapping could robustly localize sound sources from real-world binaural input, which is the first result of this kind in audio source localization [22]. The starting PhD thesis of Diego Di Carlo aims at applying the VAST framework to the blind estimation of acoustic echoes. The ultimate goal is to use these estimates to recover partial acoustic properties of the scene and enhance audio signal processing methods.

7.5. 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 [52], in particular the System & Contrast model [1].

7.5.1. Tensor-based Representation of Sectional Units in Music

Participants: Corentin Guichaoua, Frédéric Bimbot.

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

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

This approach has been studied in the context of Corentin Guichaoua's PhD [11] where a novel method for the inference of musical structure based on the optimisation of a tensorial compression criterion has been designed and experimented.

This tensorial compression criterion exploits the redundancy resulting from repetitions, similarities, progressions and analogies within musical segments in order to pack musical information observed at different time-scales in a single n-dimensional object.

The proposed method has been introduced from a formal point of view and has been related to the System & Constrast Model [1] as a extension of that model to hypercubic tensorial patterns and their deformations.

From the experimental point of view, the method has been tested on 100 pop music pieces (RWC Pop database) represented as chord sequences, with the goal to locate the boundaries of structural segments on the basis of chord grouping by minimizing the complexity criterion. The results have clearly established the relevance of the tensorial compression approach, with F-measure scores reaching 70 %

7.5.2. Modeling music by polytopic graphs of latent relations

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 [86] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [33] 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 [32] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

These results illustrate the efficiency of the proposed model in capturing structural information within musical data and is currently being explored on melodic sequences and rythmic patterns.

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

Participant: Frédéric Bimbot.

Main collaborations Gabriel Sargent (LinkMedia Inria project-team, Rennes)

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

Former work carried out in our group [95] has illustrated the benefits that can be expected from a regularity constraint on the structural segmentation of popular music pieces : a constraint which favors structural segments of comparable size provides a better conditioning of the boundary estimation process.

As a further investigation, we have explored the benefits of the regularity constraint as an efficient way for combining the outputs of a selection of systems presented at MIREX between 2010 and 2015. These experiments have yielded a level of performance which is competitive to that of the state-of-the-art on the ”MIREX10" dataset (100 J-Pop songs from the RWC database) [21].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

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

Participants: Rémi Gribonval, Himalaya Jain.

Duration: 3 years (2015-2018)

Research axis: 3.1.2

Partners: Technicolor R&I France; Inria-Rennes

Funding: Technicolor R&I France; ANRT

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

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Labex Comin Labs projects

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

9.1.1.1. HEMISFER

Participant: Rémi Gribonval.

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

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

Research axis: 3.1

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

9.1.1.2. TEPN

Participant: Rémi Gribonval.

Acronym: TEPN (Toward Energy Proportional Networks)

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

Research axis: 3.1

CominLabs partners : IRISA OCIF - Telecom Bretagne; IETR SCN; IETR SCEE; 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).

9.1.2. 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.

9.1.3. ANR OATMIL project

Participants: Rémi Gribonval, Antoine Chatalic.

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

9.1.4. OSEO-FUI: voiceHome

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

Duration: 3 years (2015-2017)

Research axis: 3.2

Partners: voicebox (formerly known as onMobile), Delta Dore, eSoftThings, Orange, Technicolor R&I France, LOUSTIC, Inria Nancy

Coordinator: voicebox

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

Contributions of PANAMA are focused on audio source localization and separation with distant microphones in real environments. In both cases, the issue of energy frugality is central and strongly constrains the available resources. This cooperation, which reached its end in November 2017, allowed us to make progress towards operational low-resource audio source localization and separation schemes, to disseminate software, collected data and scientific results, and to identify new research and development perspectives in adaptive microphone array processing for fast and robust audio scene analysis.

9.2. International Initiatives

9.2.1. Inria International Partners

9.2.1.1. Informal International Partners

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

9.3. International Research Visitors

9.3.1. Visits of International Scientists

- Flavio Castro Alves Teixeira, in May-June 2017, Post-doc, University of Innsbruck, Austria
- Pierre Vandergheynst, in June-July 2017, Professor of Signal and Image Processing, EPFL (Chaire Internationale Inria), Lausanne, Switzerland
- Gilles Blanchard, in September 2017, Professor, University of Potsdam, Germany
- Mike Davies, in October 2017, Professor, University of Edimburgh, UK
- Jérémy Cohen, in November 2017, Post-doc, University of Mons, Belgium
- Andreas Loukas, in December 2017, Post-doc, EPFL, Lausanne, Switzerland

9.3.1.1. Internships

- Helena Peic Tukuljac, from October to December 2017, PhD Student at EPFL, Lausanne, Switzerland
- Martin Strauss, from October to December 2017, M1 student, Friedrich-Alexander University, Erlangen, Germany

10. Dissemination

10.1. Promoting Scientific Activities

Antoine Deleforge was elected member of the IEEEAudio and Acoustic Signal Processing Technical Committee.

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

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

Rémi Gribonval is a member of the EURASIP Special Area Team (SAT) on Signal and Data Analytics for Machine Learning (SiG-DML) since 2015.

Rémi Gribonval has been a member of the Steering Committee of the SPARS international workshop (chairman until 2013) until 2017.

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

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

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

Nancy Bertin and Frédéric Bimbot are the coordinators of the Science and Music Young Researcher Award (Prix Jeune Chercheur Science et Musique).

Antoine Deleforge organized and will co-chair with Ivan Dokmanic and Robin Schleibler a special session on "Geometry-Aware Auditory Scene Analysis" at ICASSP 2018, Calgary, Canada.

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

Antoine Deleforge is Area Chair in Bayesian Inference for the 14th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2018).

R. Gribonval was an invited speaker at the Learning Theory Workshop during the conference on Foundations of Computational Mathematics (Barcelona, July 2017) and the Les Houches Workshop on Statistical Physics (Les Houches, February 2017).

R. Gribonval has been a member of the jury of the GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing from 2014 to 2017.

R. Gribonval is the scientific organizer of the 13th Peyresq summer school in signal and image processing, jointly organized in July 2018 by GRETSI and GDR ISIS.

R. Gribonval has joined in 2017 the Scientific Advisory Board of the Acoustics Research Institute from the Austrian Academy of Sciences in Vienna.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

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

Master : N. Bertin, "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, "Sparse representations and compressive sensing", 15 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", 4 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

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

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

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

Master : A. Deleforge, "Vocal and Audio Interaction", 6 hours, M2, Université Rennes 1, France.

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

10.3. Popularization

10.3.1. Journée Science et Musique

Participants: Antoine Deleforge, Rémi Gribonval, Frédéric Bimbot, Romain Lebarbenchon, Clément Gaultier, Nancy Bertin, Ewen Camberlein, Stéphanie Lemaile, Corentin Louboutin, Corentin Guichaoua, Cássio Fraga Dantas, Valentin Gillot, Antoine Chatalic, Yann Traonmilin.

with contributions and support from: Valérie Gouranton, Ronan Gaugne, 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 2017 hosted more than 800 visitors and was a partner of the “Festival des Sciences” and “Festival Maintenant” in Rennes.

11. Bibliography

Major publications by the team in recent years

[1] F. BIMBOT, E. DERUTY, G. SARGENT, E. VINCENT. *System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces*, in "Music Perception", June 2016, vol. 33, n^{o} 5, pp. 631-661 [*DOI :* 10.1525/MP.2016.33.5.631], https://hal.inria.fr/hal-01188244

[2] N. DUONG, E. VINCENT, R. GRIBONVAL. *Under-determined reverberant audio source separation using a full-rank spatial covariance model*, in "IEEE Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n^{o} 7, pp. 1830–1840 [*DOI :* 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865

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

[4] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. *Wavelets on graphs via spectral graph theory*, in "Applied and Computational Harmonic Analysis", March 2011, vol. 30, n^{o} 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^{o} 2, pp. 335-348 [*DOI :* 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087

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

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

[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^{o} 4, pp. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962

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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] C. GUICHAOUA. *Compression models and complexity criteria for the description and the inference of music structure*, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01687054

[12] N. KERIVEN. *Sketching for large-scale learning of mixture models*, Université Rennes 1, October 2017, https://tel.archives-ouvertes.fr/tel-01620815

Articles in International Peer-Reviewed Journals

[13] H. BECKER, L. ALBERA, P. COMON, J.-C. NUNES, R. GRIBONVAL, J. FLEUREAU, P. GUILLOTEL, I. MERLET. *SISSY: an efficient and automatic algorithm for the analysis of EEG sources based on structured sparsity*, in "NeuroImage", August 2017, vol. 157, pp. 157-172 [*DOI :* 10.1016/J.NEUROIMAGE.2017.05.046], http://www.hal.inserm.fr/inserm-01617155

[14] M. CHAFII, J. PALICOT, R. GRIBONVAL. *Wavelet modulation: An alternative modulation with low energy consumption*, in "Comptes Rendus Physique", 2017, vol. 18, n^{o} 2, pp. 156-167 [*DOI :* 10.1016/J.CRHY.2016.11.010], https://hal.archives-ouvertes.fr/hal-01445465

[15] C. F. DANTAS, M. N. DA COSTA, R. DA ROCHA LOPES. *Learning Dictionaries as a sum of Kronecker products*, in "IEEE Signal Processing Letters", May 2017, vol. 24, n^{o} 5, pp. 559 - 563 [*DOI :* 10.1109/LSP.2017.2681159], https://hal.inria.fr/hal-01672349

[16] V. DROUARD, R. HORAUD, A. DELEFORGE, S. BA, G. EVANGELIDIS. *Robust Head-Pose Estimation Based on Partially-Latent Mixture of Linear Regressions*, in "IEEE Transactions on Image Processing", March 2017, vol. 26, n^{o} 3, pp. 1428 - 1440, https://arxiv.org/abs/1603.09732 [*DOI :* 10.1109/TIP.2017.2654165], https://hal.inria.fr/hal-01413406

[17] 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 2017, 15 p. , https://arxiv.org/abs/1612.04542 [*DOI :* 10.1109/TSIPN.2017.2710619], https://hal.inria.fr/hal-01416110

[18] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Aural servo: sensor-based control from robot audition*, in "IEEE Transactions on Robotics", 2018, pp. 1-14, https://hal.inria.fr/hal-01694366

[19] E. PERTHAME, F. FORBES, A. DELEFORGE. *Inverse regression approach to robust nonlinear high-to-low dimensional mapping*, in "Journal of Multivariate Analysis", January 2018, vol. 163, pp. 1 - 14 [*DOI :* 10.1016/J.JMVA.2017.09.009], https://hal.inria.fr/hal-01652011

[20] G. PUY, M. E. DAVIES, R. GRIBONVAL. *Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$*, in "IEEE Transactions on Information Theory", 2017, https://arxiv.org/abs/1509.06947 - Submitted in 2015 [*DOI :* 10.1109/TIT.2017.2664858], https://hal.inria.fr/hal-01203614

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

Invited Conferences

[22] C. GAULTIER, S. KATARIA, A. DELEFORGE. *VAST : The Virtual Acoustic Space Traveler Dataset*, in "International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)", Grenoble, France, February 2017, https://arxiv.org/abs/1612.06287 , https://hal.archives-ouvertes.fr/hal-01416508

[23] N. KERIVEN, N. TREMBLAY, Y. TRAONMILIN, R. GRIBONVAL. *Compressive K-means*, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://arxiv.org/abs/1610.08738 , https://hal.inria.fr/hal-01386077

International Conferences with Proceedings

[24] F. ARGELAGUET, M. DUCOFFE, A. LÉCUYER, R. GRIBONVAL. *Spatial and rotation invariant 3D gesture recognition based on sparse representation*, in "IEEE Symposium on 3D User Interfaces", Los Angeles, United States, March 2017, pp. 158 - 167 [*DOI :* 10.1109/3DUI.2017.7893333], https://hal.inria.fr/hal-01625128

[25] 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

[26] A. DELEFORGE, Y. TRAONMILIN. *Phase Unmixing : Multichannel Source Separation with Magnitude Constraints*, in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://arxiv.org/abs/1609.09744 , https://hal.inria.fr/hal-01372418

[27] A. DRÉMEAU, A. DELEFORGE. *Phase Retrieval with a Multivariate Von Mises Prior: From a Bayesian Formulation to a Lifting Solution*, in "ICASSP 2017 - 42nd IEEE International Conference on Acoustics, Speech and Signal Processing", New Orleans, United States, March 2017, pp. 1-5 [*DOI :* 10.1109/ICASSP.2017.7953027], https://hal.archives-ouvertes.fr/hal-01653732

[28] C. GAULTIER, S. KITIĆ, N. BERTIN, R. GRIBONVAL. *AUDASCITY: AUdio Denoising by Adaptive Social CosparsITY*, in "25th European Signal Processing Conference (EUSIPCO)", Kos, Greece, August 2017, https://hal.inria.fr/hal-01540945

[29] H. JAIN, J. ZEPEDA, P. PÉREZ, R. GRIBONVAL. *SUBIC: A supervised, structured binary code for image search*, in "The IEEE International Conference on Computer Vision (ICCV)", Venise, Italy, IEEE, October 2017, https://arxiv.org/abs/1708.02932 - Accepted at ICCV 2017 (Spotlight) [*DOI :* 10.1109/ICCV.2017.96], https://hal.inria.fr/hal-01683390

[30] S. KATARIA, C. GAULTIER, A. DELEFORGE. *Hearing in a shoe-box : binaural source position and wall absorption estimation using virtually supervised learning* , in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New-Orleans, United States, March 2017, https://arxiv.org/abs/1609.09747 , https://hal.inria.fr/hal-01372435

[31] L. LE MAGOAROU, N. TREMBLAY, R. GRIBONVAL. *Analyzing the Approximation Error of the Fast Graph Fourier Transform*, in "ASILOMAR conference on Signals, Systems, and Computers", Monterey, California, United States, October 2017, https://hal.archives-ouvertes.fr/hal-01627434

[32] 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

[33] 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

[34] 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

National Conferences with Proceedings

[35] 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

Scientific Popularization

[36] K. DÉGUERNEL, N. LIBERMANN, E. VINCENT. *La musique comme une langue*, March 2017, Commission française pour l'enseignement des mathématiques livret "Mathématiques et langages - Panorama du thème", https://hal.inria.fr/hal-01485209

Other Publications

[37] I. DOKMANIĆ, R. GRIBONVAL. *Beyond Moore-Penrose Part I: Generalized Inverses that Minimize Matrix Norms*, July 2017, https://arxiv.org/abs/1706.08349 - working paper or preprint, https://hal.inria.fr/hal-01547183

[38] I. DOKMANIĆ, R. GRIBONVAL. *Beyond Moore-Penrose Part II: The Sparse Pseudoinverse*, July 2017, https://arxiv.org/abs/1706.08701 - working paper or preprint, https://hal.inria.fr/hal-01547283

[39] C. FRAGA DANTAS, R. GRIBONVAL, R. R. LOPES, M. N. DA COSTA. *Learning Dictionaries as Sums of Kronecker Products*, June 2017, SPARS 2017 - Signal Processing with Adaptive Sparse Structured Representations workshop, https://hal.inria.fr/hal-01514044

[40] C. GAULTIER, N. BERTIN, S. KITIĆ, R. GRIBONVAL. *A modeling and algorithmic framework for (non)social (co)sparse audio restoration*, November 2017, https://arxiv.org/abs/1711.11259 - working paper or preprint, https://hal.inria.fr/hal-01649261

[41] C. GAULTIER, S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Cosparse Denoising: The Importance of Being Social*, June 2017, The Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop, https://hal.inria.fr/hal-01510710

[42] R. GRIBONVAL, G. BLANCHARD, N. KERIVEN, Y. TRAONMILIN. *Compressive Statistical Learning with Random Feature Moments*, December 2017, https://arxiv.org/abs/1706.07180 - 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

[43] R. GRIBONVAL, H. JAIN, J. ZEPEDA, P. PÉREZ. *Learning a Complete Image Indexing Pipeline*, January 2018, https://arxiv.org/abs/1712.04480 - working paper or preprint, https://hal.inria.fr/hal-01683385

[44] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Sketching for Large-Scale Learning of Mixture Models*, May 2017, https://arxiv.org/abs/1606.02838 - working paper or preprint, https://hal.inria.fr/hal-01329195

[45] N. KERIVEN, A. DELEFORGE, A. LIUTKUS. *Blind Source Separation Using Mixtures of Alpha-Stable Distributions*, November 2017, https://arxiv.org/abs/1711.04460 - working paper or preprint, https://hal.inria.fr/hal-01633215

[46] *Best Paper*
N. KERIVEN, R. GRIBONVAL, G. BLANCHARD, Y. TRAONMILIN. *Random Moments for Sketched Mixture Learning*, June 2017, SPARS2017 - Signal Processing with Adaptive Sparse Structured Representations workshop, https://hal.inria.fr/hal-01494045.

[47] S. KITIĆ, S. BENSAID, L. ALBERA, N. BERTIN, R. GRIBONVAL. *Versatile and scalable cosparse methods for physics-driven inverse problems*, March 2017, working paper or preprint, https://hal.inria.fr/hal-01496767

[48] Y. TRAONMILIN, N. KERIVEN, R. GRIBONVAL, G. BLANCHARD. *Spikes super-resolution with random Fourier sampling*, June 2017, SPARS workshop 2017, https://hal.archives-ouvertes.fr/hal-01509863

[49] Y. TRAONMILIN, G. PUY, R. GRIBONVAL, M. E. DAVIES. *Compressed sensing in Hilbert spaces*, July 2017, https://arxiv.org/abs/1702.04917 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01469134

References in notes

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

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

[52] 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

[53] 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^{o} 19, 20 p. [*DOI :* 10.1109/TSP.2015.2447503], https://hal.archives-ouvertes.fr/hal-01084986

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

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

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

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

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

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

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

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

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

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

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

[65] A. DELEFORGE, F. FORBES, S. BA, R. HORAUD. *Hyper-Spectral Image Analysis With Partially Latent Regression and Spatial Markov Dependencies*, in "IEEE Journal of Selected Topics in Signal Processing", 2015, vol. 9, n^{o} 6, pp. 1037–1048

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

[67] A. DELEFORGE, F. FORBES, R. HORAUD. *High-dimensional regression with gaussian mixtures and partially-latent response variables*, in "Statistics and Computing", 2015, vol. 25, n^{o} 5, pp. 893–911

[68] A. DELEFORGE, R. HORAUD, Y. Y. SCHECHNER, L. GIRIN. *Co-localization of audio sources in images using binaural features and locally-linear regression*, in "IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)", 2015, vol. 23, n^{o} 4, pp. 718–731

[69] A. DELEFORGE, W. KELLERMANN. *Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures*, in "Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on", IEEE, 2015, pp. 355–359

[70] M. A. GERZON. *Periphony: With-Height Sound Reproduction*, in "J. Audio Eng. Soc", 1973, vol. 21, n^o 1, pp. 2–10, http://www.aes.org/e-lib/browse.cfm?elib=2012

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

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

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

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

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

[76] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Hearing behind walls: localizing sources in the room next door with cosparsity*, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014 [*DOI :* 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779

[77] S. KITIĆ, N. BERTIN, R. GRIBONVAL. *Hearing behind walls: localizing sources in the room next door with cosparsity*, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014 [*DOI :* 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779

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

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

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

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

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

[83] 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

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

[85] A. LIUTKUS, R. BADEAU. *Generalized Wiener filtering with fractional power spectrograms*, in "Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on", IEEE, 2015, pp. 266–270

[86] 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

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

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

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

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

[91] A. A. NUGRAHA, A. LIUTKUS, E. VINCENT. *Multichannel audio source separation with deep neural networks*, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2016, vol. 24, n^{o} 9, pp. 1652–1664

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

[93] G. PUY, P. VANDERGHEYNST, R. GRIBONVAL, Y. WIAUX. *Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques*, in "EURASIP Journal on Advances in Signal Processing", 2012 [*DOI :* 10.1186/1687-6180-2012-6], http://hal.inria.fr/inria-00582432

[94] 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

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

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

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