Activity Report 2018

Project-Team PANAMA

Parcimonie et Nouveaux Algorithmes pour le Signal et la Modélisation Audio

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. - Signal processing
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.9.3. - Reconstruction, enhancement
A5.9.4. - Signal processing over graphs
A5.9.5. - Sparsity-aware processing
A5.9.6. - Optimization tools
A5.10.2. - Perception
A5.11.2. - Home/building control and interaction
A6.1.4. - Multiscale modeling
A6.2.5. - Numerical Linear Algebra
A6.2.6. - Optimization
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A8.6. - Information theory
A8.7. - Graph theory
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

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

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

1. Team, Visitors, External Collaborators

Research Scientists
Rémi Gribonval [Team leader, Inria, Senior Researcher, HDR]
Nancy Bertin [CNRS, Researcher]
Frédéric Bimbot [CNRS, Senior Researcher, HDR]
Jeremy Cohen [CNRS, Researcher, from Oct 2018]
Antoine Deleforge [Inria, Researcher, until Mar 2018]

Faculty Member
Nicolas Courty [Univ de Bretagne sud, Professor, from Sep 2018, HDR]

Post-Doctoral Fellow
Clément Elvira [Inria]

PhD Students
Antoine Chatalic [Univ de Rennes I]
Diego Di Carlo [Inria]
Cássio Fraga Dantas [Inria]
Clément Gaultier [Inria]
Valentin Gillot [INSA Rennes]
Mohammed Hafsati [Institut de recherche technologique B-com]
Himalaya Jain [Technicolor, until Jun 2018]
Marie-Anne Lacroix [Univ de Rennes I, from Sep 2018]
Nathan Libermann [Inria]
Corentin Louboutin [Univ de Rennes I, until Aug 2018]
Corentin Louboutin [Inria, from Sep 2018 until Nov 2018]
Pierre Stock [Facebook, from Mar 2018]

Technical staff
Ewen Camberlein [Inria]
Hakim Hadj Djilani [Inria, from Apr 2018]
Romain Lebarbenchon [Inria]

Interns
Tudor Vlad Beleuta [Inria, from Mar 2018 until Jun 2018]
Antoine Gonon [Inria, from May 2018 until Jul 2018]
Xavier Rolland [Inria, from Feb 2018 until Jun 2018]
Eric Zajler Grinstein [Inria, until Jan 2018]
Younes Zine [Ecole normale supérieure de Rennes, from Nov 2018]

Administrative Assistant

Stéphanie Lemaile [Inria]
Visiting Scientists
Jeremy Cohen [CNRS, Sep 2018]
Roilhi Frajo Ibarra Hernandez [CICESE, Mexique, from Feb 2018 until Aug 2018]
Andreas Loukas [EPFL, from Dec 2018]
Andreas Loukas [EPFL, until Feb 2018]
Pierre Vandergheynst [Inria International Chair,Inria]

2. Overall Objectives

2.1. Overall positioning

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

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

2.2. Scientific foundations

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

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

2.3. Applications

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

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

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*

The 2018 **prix de thèse Signal, Image et Vision** was jointly awarded by the Club EEA, the GRETSI and the GdR ISIS to **Mrs Marwa Chafii** for her thesis entitled : *Étude d'une nouvelle forme d'onde multiporteuses à PAPR réduit*. This thesis was conducted within the IETR Lab at CentraleSupélec on the campus of Rennes, under the supervision of Jacques Palicot, Professeur, CentraleSupélec, Rennes and Rémi Gribonval, Directeur de recherche, Inria, Rennes.

5.1.2. *Other highlights*

Frédéric Bimbot is the new Editor-in-Chief of the journal "Speech Communication".

BEST PAPER AWARD:

[2]
M. CHAFII. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509

6. New Software and Platforms

6.1. FAuST

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

SCIENTIFIC DESCRIPTION: FAuST allows to approximate a given dense matrix by a product of sparse matrices, with considerable potential gains in terms of storage and speedup for matrix-vector multiplications.
FUNCTIONAL DESCRIPTION: Faust 1.x contains Matlab routines to reproduce experiments of the PANAMA team on learned fast transforms.

Faust 2.x contains a C++ implementation with Matlab / Python wrappers (work in progress).
NEWS OF THE YEAR: A Software Development Initiative (ADT REVELATION) started in April 2018 for the maturation of FAuST. A first step achieved this year was to complete and robustify Matlab wrappers, to code Python wrappers with the same functionality, and to setup a continuous integration process. A second step was to simplify the parameterization of the main algorithms. The roadmap for next year includes showcasing examples and optimizing computational efficiency. – In 2017, new Matlab code for fast approximate Fourier Graph Transforms have been included. based on the approach described in the papers:

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

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

- Participants: Luc Le Magoarou, Nicolas Tremblay, Rémi Gribonval, Nicolas Bellot, Adrien Leman and Hakim Hadj-Djilani

- Contact: Rémi Gribonval

- Publications: Approximate fast graph Fourier transforms via multi-layer sparse approximations - Analyzing the Approximation Error of the Fast Graph Fourier Transform - Flexible Multi-layer Sparse Approximations of Matrices and Applications - Are There Approximate Fast Fourier Transforms On Graphs? - Efficient matrices for signal processing and machine learning - 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/

6.2. SPADE

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

- 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 2018, the code has been robustified with the help of InriaTech for a potential industrial transfer. – In 2017, a web interface to demonstrate the potential of SPADE has been setup using the A∥go platform.

- Participants: Nancy Bertin, Clement Gaultier, Ewen Camberlein, Romain Lebarbenchon, Alexandre Sanchez, 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/

6.3. FASST

Flexible Audio Source Separation Toolbox

KEYWORD: Audio signal processing

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

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

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

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

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

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

- Participants: Alexey Ozerov, Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Frédéric Bimbot and Yann Salaun

- Contact: Emmanuel Vincent

- URL: http://bass-db.gforge.inria.fr/fasst/

6.4. Multi-channel BSS Locate Basic

KEYWORDS: Audio - Localization - Signal processing - Multichannel signal

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

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

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

6.5. VoiceHome-2

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

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

7. New Results

7.1. Sparse Representations, Inverse Problems, and Dimension Reduction

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

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

7.1.1. *Computational Representation Learning: Algorithms and Theory*

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

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

An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, classical dictionary learning is limited to small-scale problems.

Multilayer sparse matrix products for faster computations. 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, with reduced sample complexity. Besides the principle and its application to image denoising [105], we demonstrated the potential of the approach to speedup linear inverse problems [104], and a comprehensive journal paper was published in 2016 [107]. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou [108].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms [106], and studied their approximation error [109]. In a journal paper published this year [16] 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 FAµST software library (see Section 6) was first released as Matlab code primarily for reproducibility of the experiments of [107]. A C++ version is being developed to provide transparent interfaces of FAµST data-structures with both Matlab and Python.

Kronecker product structure for faster computations. In parallel to the development of FAuST, we have proposed another approach to structured dictionary learning that also aims at speeding up both sparse coding and dictionary learning. We used the fact that for tensor data, a natural set of linear operators are those that operate on each dimension separately, which correspond to rank-one multilinear operators. These rank-one operators may be cast as the Kronecker product of several small matrices. Such operators require less memory and are computationally attractive, in particular for performing efficient matrix-matrix and matrix-vector operations. In our proposed approach, dictionaries are constrained to belong to the set of low-rank multilinear operators, that consist of the sum of a few rank-one operators. A special case of the proposed structure is the widespread separable dictionary, named SuKro, which was evaluated experimentally last year on an image denoising application [81]. The general approach, coined HOSUKRO for High Order Sum of Kronecker products, has been shown this year to reduce empirically the sample complexity of dictionary learning, as well as theoretical complexity of both the learning and the sparse coding operations [27].

Combining faster matrix-vector products with screening techniques. We combined accelerated matrix-vector multiplications offered by FAµST / HOSUKRO matrix approximations with dynamic screening [57], that safely eliminates inactive variables to speedup iterative convex sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary [80]. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. This year we completed a comprehensive preprint submitted for publication in a journal [49] that includes new techniques based on duality gaps to optimally switch from a coarse dictionary approximation to a finer one. Significant complexity reductions were obtained in comparison to screening rules alone [28].

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

Participant: Rémi Gribonval.

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

We studied linear generalized inverses that minimize matrix norms. Such generalized inverses are famously represented by the Moore-Penrose pseudoinverse (MPP) which happens to minimize the Frobenius norm. Freeing up the degrees of freedom associated with Frobenius optimality enables us to promote other interesting properties. In a first part of this work [76], 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 [77], 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 $1 \le p \le 2$. 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. This year, a condensed version of these results has been prepared which is now accepted for publication [14].

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

Participants: Rémi Gribonval, 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 Keriven (ENS Paris), Nicolas Tremblay (GIPSA-Lab, Grenoble), Phil Schniter & Evan Byrne (Ohio State University, USA), Laurent Jacques & Vincent Schellekens (Univ Louvain, Belgium), Florimond Houssiau & Y.-A. de Montjoye (Imperial College London, UK)

Sketching for Large-Scale 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 [58], [61], [59], [60] to fit a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical generalized 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 Ph.D. thesis of Nicolas Keriven [97] consolidated extensions to non-isotropic Gaussians, with a new algorithm called CL-OMP [96] and large-scale experiments demonstrating its potential for speaker verification [95]. A journal paper was published this year [15], with an associated toolbox for reproducible research (see SketchMLBox, Section 6).

Sketching for Compressive Clustering and beyond. In 2016 we started a new endeavor to extend the sketched learning approach beyond Gaussian Mixture Estimation.

First, we showed empirically that sketching can be adapted to compress a training collection while allowing large-scale *clustering*. The approach, called "Compressive K-means", uses CL-OMP at the learning stage [98]. This year, we showed that in the high-dimensional setting one can substantially speedup both the sketching stage and the learning stage by replacing Gaussian random matrices with fast random linear transforms in the sketching procedure [23].

An alternative to CL-OMP for cluster recovery from a sketch is based on simplified hybrid generalized approximate message passing (SHyGAMP). Numerical experiments suggest that this approach is more efficient than CL-OMP (in both computational and sample complexity) and more efficient than k-means++ in certain regimes [62]. During his first year of Ph.D., Antoine Chatalic visited the group of Phil Schiter to further investigate this topic, and a journal paper is in preparation.

We also demonstrated that sketching can be used in blind source localization and separation, by learning mixtures of alpha-stable distributions [32], see details in Section 7.5.3.

Finally, sketching provides a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum. A conference paper establishing theoretical privacy guarantees (with the *differential privacy* framework) and exploring the utility / privacy tradeoffs of Compressive K-means has been submitted for publication.

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

Participants: Rémi Gribonval, Clément Elvira.

Main collaboration: Mike Davies (University of Edinburgh, UK), Gilles Puy (Technicolor R&I France, Rennes), Yann Traonmilin (Institut Mathématique de Bordeaux), Nicolas Keriven (ENS Paris), Gilles Blanchard (Univ Postdam, Germany), Cédric Herzet (SIMSMART project-team, IRMAR / Inria Rennes), Charles Soussen (Centrale Supelec, Gif-sur-Yvette), Mila Nikolova (CMLA, Cachan)

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). We published a comprehensive paper [20] 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 [121]. These results have been surveyed in a book chapter published this year [46]. In the context of nonlinear inverse problems, we showed that the notion of RIP is still relevant with proper adaptation [42].

Optimal convex regularizers for linear inverse problems. The ℓ^1-norm is a good convex regularization for the recovery of sparse vectors from under-determined linear measurements. No other convex regularization seems to surpass its sparse recovery performance. We explored possible explanations for this phenomenon by defining several notions of " best " (convex) regularization in the context of general low-dimensional recovery and showed that indeed the ℓ^1-norm is an optimal convex sparse regularization within this framework [43]. A journal paper is in preparation with extensions concerning nuclear norm regularization for low-rank matrix recovery and further structured low-dimensional models.

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 is under revision for a journal [88].

Recovery guarantees for algorithms with continuous dictionaries. We established theoretical guarantees on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous dictionaries [40], e.g. as appearing in the context of sparse spike deconvolution. Analyses based on discretized dictionary fail to be conclusive when the discretization step tends to zero, as the coherence goes to one. Instead, our analysis is directly conducted in the continuous setting amd exploits specific properties of the positive definite kernel between atom parameters defined by the inner product between the corresponding atoms. For the Laplacian kernel in dimension one, we showed in the noise-free setting that OMP exactly recovers the atom parameters as well as their amplitudes, regardless of the number of distinct atoms [40]. A journal paper is in preparation describing a full class of kernels for which such an analysis holds, in particular for higher dimensional parameters.

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

Well known relations between these two approaches have lead to some widely spread misconceptions. In particular, while the so-called Maximum A Posterori (MAP) estimate with a Gaussian noise model does lead to an optimization problem with a quadratic data-fidelity term, we disprove through explicit examples the common belief that the converse would be true. In previous work we showed that for denoising in the presence of additive Gaussian noise, for any prior probability on the unknowns, the MMSE is the solution of a penalized least squares problem, with all the apparent characteristics of a MAP estimation problem with Gaussian noise and a (generally) different prior on the unknowns [89]. In other words, the variational approach is rich enough to build any MMSE estimator associated to additive Gaussian noise via a well chosen penalty.

This year, we achieved generalizations of these results beyond Gaussian denoising and characterized noise models for which the same phenomenon occurs. In particular, we proved that with (a variant of) Poisson noise and any prior probability on the unknowns, MMSE estimation can again be expressed as the solution of a penalized least squares optimization problem. For additive scalar denoising, the phenomenon holds if and only if the noise distribution is log-concave, resulting in the perhaps surprising fact that scalar Laplacian denoising can be expressed as the solution of a penalized least squares problem. [51] Somewhere in the proofs appears an apparently new characterization of proximity operators of (nonconvex) penalties as subdifferentials of convex potentials [50].

7.1.5. *Algorithmic Exploration of Sparse Representations for Neurofeedback*
Participant: Rémi Gribonval.

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

In the context of the HEMISFER (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation) Comin Labs project (see Section 9.1.1.1), 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 [118]. This year we focused on directly estimating neuro-feedback scores rather than brain activity. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) both allow measurement of brain activity for neuro-feedback (NF), respectively with high temporal resolution for EEG and high spatial resolution for fMRI. Using simultaneously fMRI and EEG for NF training is very promising to devise brain rehabilitation protocols, however performing NF-fMRI is costly, exhausting and time consuming, and cannot be repeated too many times for the same subject. We proposed a technique to predict NF scores from EEG recordings only, using a training phase where both EEG and fMRI NF are available. A conference paper has been submitted.

7.1.6. *Sparse Representations as Features for Heart Sounds Classification*
Participant: Nancy Bertin.

Main collaborations: Roilhi Frajo Ibarra Hernandez, Miguel Alonso Arevalo (CICESE, Ensenada, Mexico)

A heart sound signal or phonocardiogram (PCG) is the most simple, economical and non-invasive tool to detect cardiovascular diseases (CVD), the main cause of death worldwide. During the visit of Roilhi Ibarra, we proposed a pipeline and benchmark for binary heart sounds classification, based on his previous work on a sparse decomposition of the PCG [91]. We improved the feature extraction architecture, by combining features derived from the Gabor atoms selected at the sparse representation stage, with Linear Predictive Coding coefficients of the residual. We compared seven classifiers with two different approaches in presence of multiple hearts beats in the recordings: feature averaging (proposed by us) and cycle averaging (state-of-the-art). The feature sets were also tested when using an oversampling method for balancing. The benchmark identified systems showing a satisfying performance in terms of accuracy, sensitivity, and Matthews correlation coefficient, with best results achieved when using the new feature averaging strategy together with oversampling. This work was accepted for publication in an international conference [30].

7.1.7. *An Alternative Framework for Sparse Representations: Sparse "Analysis" Models*
Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Orange, 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 [87], [117] [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 [99], [101], [100], [55]. 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],

a book chapter 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, was published this year [45]. Second, we continued extending the cosparse framework on audio restoration problems [85], [84], [82], especially improvements on our released real-time declipping algorithm (A-SPADE - see Section 6.2) and extension to multichannel data [29].

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. *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 [63], defended in October 2016 and co-supervised by R. Gribonval, and awarded with the GDR ISIS/GRETSI/EEA thesis prize, see Section 5.1.1), 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, [67], [65], [66]. We characterized waveforms with optimum PAPR [68], [64] as well as the tradeoffs between PAPR and Power Spectral Density properties of a wavelet modulation scheme [70]. Our design of new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR gave rise to a patent [69] and was published this year [22], [13].

7.3. Emerging activities on high-dimensional learning with neural networks

Participants: Rémi Gribonval, Himalaya Jain, Pierre Stock.

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

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

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

Learning compact representations for large-scale image search. The PhD thesis of Himalaya Jain [11] was dedicated to learning techniques for the design of new efficient methods for large-scale image search and indexing. A first step was to propose techniques for approximate nearest neighbor search exploiting quantized sparse representations in learned dictionaries [92]. The thesis then explored structured binary codes, computed through supervised learning with convolutional neural networks [93]. This year, we integrated these two components in a unified end-to-end learning framework where both the representation and the index are learnt [31]. These results have led to a patent application.

Equi-normalization of Neural Networks. Modern neural networks are over-parameterized. In particular, each rectified linear hidden unit can be modified by a multiplicative factor by adjusting input and out- put weights, without changing the rest of the network. Inspired by the Sinkhorn-Knopp algorithm, we introduced a fast iterative method for minimizing the l2 norm of the weights, equivalently the weight decay regularizer. It provably converges to a unique solution. Interleaving our algorithm with SGD during training improves the test accuracy. For small batches, our approach offers an alternative to batch- and group- normalization on CIFAR-10 and ImageNet with a ResNet-18. This work has been submitted for publication.

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

7.4. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.4.1. *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 [73], 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 [71], head pose estimation in images [79], sound source separation and localization [72], and virtually-supervised acoustic space learning (see Section 7.6.1). 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.

7.4.2. *Audio Inpainting and Denoising*

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Orange, Rennes)

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. Building upon our previous pioneering contributions [54]), we proposed over the last three years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [100], [102] [6]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [102] resulted in a software release in 2016. In 2017, this work was extended towards advanced (co)sparse decompositions, including several forms of structured sparsityand towards their application to the denoising task.In particular, we investigated the incorporation of the so-called "social" structure constraint [103] into problems regularized by a cosparse prior [84], [85], and exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [82].

In 2018, a new algorithm for joint declipping of multichannel audio was derived and published [29]. Extensive experimental benchmarks were conducted, questioning the previous state-of-the-art habits in degradation levels (usually moderate to inaudible) and evaluation (small datasets, SNR-based performance criteria) and setting up new standards for the task (large and diverse datasets, severe saturation, perceptual quality evaluation) as well as guidelines for the choice of the best variant (sparse or cosparse, with or without structural time-frequency constraints...) depending on the data and operational conditions. These new results will be included in an ongoing journal paper, to be submitted in 2019.

7.5. 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 [122] 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 Sections 7.5.1, 7.5.2); more exploratory work towards new approaches diverging away from local Gaussian modeling (Section 7.5.3) ; formulating and addressing a larger class of problems related to localization and separation, in the contexts of robotics (Section 7.5.4) and virtual reality (Section 7.5.2). Eventually, one of these new tracks, audio scene analysis with machine learning, evolved beyond the "localization and separation" paradigm, and is the subject of a new axis of research presented in Section 7.6.

7.5.1. *Towards Real-world Localization and Separation*

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

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

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

First, 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, submitted to a special issue of Speech Communication, was published this year [12].

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). In addition, evolutions of the MBSSLocate software initiated during this project led to a successful participation in the IEEE-AASP Challenge on Acoustic Source Localization and Tracking (LOCATA), and to industrial transfer (see Section 8.1.1).

7.5.2. *Separation for Remixing Applications*
Participants: Nancy Bertin, Rémi Gribonval, Mohammed Hafsati.

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

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b<>com with the INVATE project, see Section 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 [86], in the context of free navigation in 3D audiovisual contents. We studied the applicability conditions of the FASST framework to HOA signals and benchmarked localization and separation methods in this domain. Simulation results showed that separating sources in the HOA domain results in a 5 to 15 dB increase in signal-to-distortion ratio, compared to the microphone domain. These results led to a conference paper submission in 2018. We continued extending our methods to hybrid acquisition scenarios, where the separation of HOA signals can be informed by complementary close-up microphonic signals. Future work will include subjective evaluation of the developed workflows.

7.5.3. *Beyond the Local Complex Gaussian Model*
Participant: Antoine Deleforge.

Main collaboration: Nicolas Keriven (ENS Paris), 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 [110]. We provided a theoretical analysis of some limitations of the classical LCG-based multichannel Wiener filter in [21]. In [32] 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 [120][9] or more recently deep neural networks [119], it becomes desirable to use probabilistic models enforcing strong magnitude priors. In [75], we explored deterministic magnitude models. An approximate and tractable probabilistic version of this referred to as BEADS (Bayesian Expansion Approximating the Donut Shape) was presented this year [33]. The source prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle.

7.5.4. *Applications to Robot Audition*
Participants: Nancy Bertin, Antoine Deleforge.

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. Results obtained in the previous years [116], [114], [115] were encompassed and extended in two journal papers published this year [17], [18]. In particular, we obtained new results on the use of interaural level difference as the only input feature of the servo, a counter-intuitive result outside the robotic context. We also showed the robustness, low-complexity and independence to Head Related Transfer Function (HRTF) of the approach on humanoid robots.

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). The dataset was made publicly available at dregon.inria.fr and was presented together with the results in [37].

7.6. Towards comprehensive audio scene analysis

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

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

7.6.1. *Virtually-Supervised Auditory Scene Analysis*

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

Main collaborations: Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), 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.

Recently, two directions for advanced audio geometry estimation have emerged and were investigated in our team. The first one is physics-driven [45]. This approach implicitly 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. However, it relies 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 [72], [74]. 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) [83], [94]. 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 [94]. 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 [83]. The VAST datasets and approaches made the bed of several new works in 2018, including real-world source localization on a wider range of settings (LOCATA test data on various microphone arrays) and echo estimation (see below).

7.6.2. *Room Properties: Estimating or Learning Early Echoes*

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

Main collaborations: Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Helena Peic-Tukuljac (EPFL, Switzerland).

In [35] we showed that the knowledge of early echoes improved sound source separation performances, which motivates the development of (blind) echo estimation techniques. Echoes are also known to potentially be a key to the room geometry problem [78]. In 2018, two different approaches to this problem were explored.

In [34] we proposed an analytical method for early echoes estimation. This method builds on the framework of finite-rate-of-innovation sampling. The approach operates directly in the parameter-space of echo locations and weights, and enables near-exact blind and off-grid echo retrieval from discrete-time measurements. It is shown to outperform conventional methods by several orders of magnitude in precision, in an ideal case where the room impulse response is limited to a few weighted Diracs. Future work will include alternative initialization schemes and convex relaxations, extensions to sparse-spectrum signals and noisy measurements, and applications to dereverberation and audio-based room shape reconstruction.

As a concurrent approach exploration, the PhD thesis of Diego Di Carlo aims at applying the VAST framework to the blind estimation of acoustic echoes, or other room properties (such as reverberation time, acoustic properties at the boundaries, etc.) This year, we focused on identifying promising couples of inputs and outputs for such an approach, especially by leveraging the notions of relative transfer functions between microphones, the room impulse responses, the time-difference-of-arrivals, the angular spectra, and all their mutual relationships. In a simple yet common scenario of 2 microphones close to a reflective surface and one source (which may occur, for instance, when the sensors are placed on a table such as in voice-based assistant devices), we introduced the concept of microphone array augmentation with echoes (MIRAGE) and showed how estimation of early-echo characteristics with a learning-based approach is not only possible but can in fact benefit source localization. In particular, it allows to retrieve 2D direction of arrivals from 2 microphones only, an impossible task in anechoic settings. These first results were submitted to an international conference. Future work will consider extension to more realistic and more complex scenarios (including more microphones, sources and reflective surfaces) and the estimation of other room properties such as the acoustic absorption at the boundaries, or ultimately, the room geometry.

7.6.3. *Multichannel Audio Event and Room Classification*
Participants: Marie-Anne Lacroix, Nancy Bertin.

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

Typically, audio event detection and classification is tackled as a "pure" single-channel signal processing task. By constrast, audio source localization is the perfect example of multi-channel task "by construction". In parallel, the need to classify the type of scene or room has emerged, in particular from the rapid development of wearables, the "Internet of things" and their applications. The PhD of Marie-Anne Lacroix, started in September 2018, combines these ideas with the aim of developing multi-channel, room-aware or spatially-aware audio classification algorithms for embedded devices. The PhD topic includes low-complexity and low-energy stakes, which will be more specifically tackled thanks to the GRANIT members area of expertise. During the first months of the PhD, we gathered existing data and identified the need for new simulations or recordings, and combined ideas from existing single-channel classification techniques with traditional spatial features in order to design a baseline algorithm for multi-channel joint localization and classification of audio events, currently under development.

7.7. 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 [56], in particular the System & Contrast model [1].

7.7.1. *Tensor-based Representation of Sectional Units in Music*
Participant: Frédéric Bimbot.

This work was primarily carried out by Corentin Guichaoua, former PhD student with Panama, now with IRMA (CNRS UMR 7501, Strasbourg).

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 [90] 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 & Contrast 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 % on that task [41]

7.7.2. *Modeling music by Polytopic Graphs of Latent Relations (PGLR)*
Participants: Corentin Louboutin, Frédéric Bimbot.

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

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

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

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

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

If musical elements are chords, relations can be inferred by minimal transport [111] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [113] 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 [112] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

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

Results obtained in this framework illustrate the efficiency of the proposed model in capturing structural information within musical data and support the view that musical content can be delinearised in order to better describe its structure. Extensive results will be included in Corentin Louboutin's PhD, which is planned to be defended early 2019.

7.7.3. *Exploring Structural Dependencies in Melodic Sequences using Neural Networks*
Participants: Nathan Libermann, Frédéric Bimbot.

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

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

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

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

These experiments have been conducted on the Lakh MIDI dataset and more particularly on a subset of 8308 monophonic 16-bar melodic segments. The obtained results indicate a non-uniform distribution of modeling capabilities prediction of recurrent networks, suggesting the utility of non-ergodic models for the generation of melodic segments [38].

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

7.7.4. Graph Signal Processing for Multiscale Representations of Music Similarity
Participants: Valentin Gillot, Frédéric Bimbot.

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

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

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

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

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

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

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

By the end of the PhD, we expect the release of a specific toolbox for music composition, remixing and repurposing using the concepts and algorithms developed during the PhD.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Contract with 5th dimension on "dynamic separation of localized sound sources"

A first phase of this contract, in collaboration with InriaTech, involved porting in C++ a subset of our source localization library Multichannel BSS Locate. A second phase will involve further investigations on the interplay between localization and separation, using the FASST library, with support from LABEX AMIES.

8.1.2. Contract with Honda on "multichannel speech and audio processing"

This is a follow-up contract, which targets collaborative research on multichannel speech and audio processing and eventual software licensing in order to enable voice-based communication in challenging noisy and reverberant conditions in which current hands-free voice-based interfaces perform poorly.

8.2. Bilateral Grants with Industry

8.2.1. CIFRE contract with Technicolor R&I France on Very large scale visual comparison
Participants: Rémi Gribonval, Himalaya Jain.

Duration: 3 years (2015-2018)

Research axis: 3.1.2

Partners: Technicolor R&I France; Inria-Rennes

Funding: Technicolor R&I France; ANRT

The grand goal of this thesis was 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.

8.2.2. CIFRE contract with Facebook Artificial Intelligence Research, Paris on Deep neural networks for large scale learning
Participants: Rémi Gribonval, Pierre Stock.

Duration: 3 years (2018-2021)

Research axis: 3.1.2

Partners: Facebook Artificial Intelligence Research, Paris; Inria-Rennes

Funding: Facebook Artificial Intelligence Research, Paris; ANRT

The overall objective of this thesis is to design, analyze and test large scale machine learning algorithms with applications to computer vision and natural language processing. A major challenge is to design compression techniques able to replace complex and deep neural networks with much more compact ones while preserving the capacity of the initial network to achieve the targeted task. An avenue primarily envisioned to achieve this goal is to rely on structured linear layers.

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://hemisfer.cominlabs.u-bretagneloire.fr/

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://tepn.cominlabs.u-bretagneloire.fr/

Research axis: 3.1

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

Coordinator: Nicolas Montavont, IRISA OCIF - Telecom Bretagne

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

Contribution of PANAMA: PANAMA, in close cooperation with the SCEE team at IETR (thesis of Marwa Chafii, 2016), focuses on the design of new waveforms for multi carrier systems with reduced Peak to Average Power Ratio (PAPR).

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, Nicolas Courty.

Duration: 4 years (2017-2021)

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

http://people.irisa.fr/Nicolas.Courty/OATMIL/

Research Axis 3.1

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

Coordinator: Nicolas Courty (Obelix team)

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

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

Funding: ANR

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 was 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. 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 published in 2018 in a journal paper [12], 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 Institute for Digital Communications at the University of Edimburgh, and the Institute for Mathematics of the Postdam University.

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

9.3. International Research Visitors

9.3.1. Visits of International Scientists

- Gilles Blanchard, in Spring 2018, Professor, University of Potsdam, Germany
- Andreas Loukas, in January 2018 and December 2018, Post-doc, EPFL, Lausanne, Switzerland

9.3.1.1. Internships

- Roilhi Frajo Ibarra Hernandez, from March to August 2018, PhD Student at CICESE, Ensenada, Mexico

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

- Rémi Gribonval, the scientific organizer of the 13th Peyresq summer school in signal and image processing, jointly organized in July 2018 by GRETSI and GDR ISIS.
- Rémi Gribonval, scientific and organization board of the Conference on Deep Learning: from Theory to Applications organized in the framework of the LABEX Centre Henri Lebesgue, Rennes, September 4-6 2018.
- Valentin Gillot and Romain Lebarbenchon, event coordinators of the 2018 Science and Music Day (Journée Science et Musique) organized by IRISA.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

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

10.1.2.2. Member of the Conference Program Committees

- Antoine Deleforge, Area Chair in Bayesian Inference for the 14th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2018).
- Nancy, Area Chair in Sparsity-related Methods for the 14th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2018).
- Nancy, member of the program committee of the Journées d'Informatique Musicale (JIM 2018, Amiens, France).
- Rémi Gribonval, member of the program committee of the GRETSI.

10.1.2.3. Organization of special sessions

- Antoine Deleforge, organizer and co-chair with Ivan Dokmanic and Robin Schleibler of a special session on "Geometry-Aware Auditory Scene Analysis" at ICASSP 2018, Calgary, Canada.
- Antoine Deleforge, organizer and co-chair with Angélique Drémeau of a special session on "Advances in Phase Retrieval and Applications" at LVA/ICA 2018, Guildford, UK.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Frédéric Bimbot appointed as Editor-in-Chief of the international journal *Speech Communication*.
- Rémi Gribonval, Associate Editor of the international journal *Constructive Approximation*.

10.1.4. Invited Talks

- Frédéric Bimbot, JIM 2018 (Journées d'Informatique Musicale) in Amiens, France.
- Frédéric Bimbot, JEP 2018 (Journées d'Etudes sur la Parole), in Aix-en-Provence, France.
- Rémi Gribonval, Oberwolfach workshop on Applied Harmonic Analysis and Data Processing, March 25-31 2018, in Oberwolfach, Germany.
- Rémi Gribonval, CM+X workshop on Inverse Problems and Machine Learning, Feb 8-11 2018, in Caltech, Los Angeles, USA.
- Nancy Bertin, JJCAAS (Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio), June 6-8 2018, in Brest, France.

10.1.5. Scientific Expertise

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

10.1.6. Research Administration

- Frédéric Bimbot, Head of the "Digital Signals and Images, Robotics" department in IRISA (UMR 6074) until June 2018.
- Rémi Gribonval, member of the organization committee of the 2019 GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing.

10.2. Teaching - Supervision - Juries

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

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

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

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

Master : R. Gribonval, "High dimensional statistical learning", 12 hours, M2, Université Rennes 1, France.

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

Master : R. Gribonval, "Sparsity in Signal and Image Processing", 8 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

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

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 (JSM 2018)*

Participants: Valentin Gillot, Romain Lebarbenchon, Rémi Gribonval, Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Stéphanie Lemaile, Corentin Louboutin, Antoine Chatalic, Clément Gaultier, Cássio Fraga Dantas, Diego Di Carlo.

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

PANAMA coordinated the organization of a public event called "Journée Science et Musique" ("Music and Science Day"). This yearly event organized by the METISS/ PANAMA Team since 2011 aims at sharing with the wide audience the latest innovations and research projects in music. The motivation for hosting this event is to explain and promote the technology behind audio-processing that people face in their daily lives. The event is free to everyone and people have the possibility to attend talks by selected speakers or meet numerous experts that demonstrate current projects in which people can interactively participate. Edition 2018 hosted approximately 500 visitors and was a partner of the "Festival des Sciences" and "Festival Maintenant" in Rennes.

10.3.2. *250th Anniversary of Joseph Fourier's birth*

Participants: Nancy Bertin, Frédéric Bimbot, Antoine Deleforge, Rémi Gribonval.

PANAMA has been involved in the *Comité Fourier 250* by participating to, and triggering interest for the 250th Anniversary of Joseph Fourier's birth, in the framework of *Commémorations Nationales* for the year 2018. In particular, Frédéric Bimbot has been invited to give a talk at the JEP 2018 on Fourier's heritage in Speech Processing, and the *Journée Science et Musique 2018* (see above) has included a conference where Fourier's Transform was introduced and called for to explain concepts relating to music perception.

10.3.3. *Articles and contents*

Rémi Gribonval wrote an article published online on Interstices [47]

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, n⁰ 5, pp. 631-661 [*DOI :* 10.1525/MP.2016.33.5.631], https://hal.inria.fr/hal-01188244

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

[3] N. DUONG, E. VINCENT, R. GRIBONVAL. *Under-determined reverberant audio source separation using a full-rank spatial covariance model*, in "IEEE Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n⁰ 7, pp. 1830–1840 [*DOI :* 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865

[4] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. *Wavelets on graphs via spectral graph theory*, in "Applied and Computational Harmonic Analysis", March 2011, vol. 30, n⁰ 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, n⁰ 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, n⁰ 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, n⁰ 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] H. JAIN. *Learning compact representations for large scale image search*, Université Rennes 1, June 2018, https://tel.archives-ouvertes.fr/tel-01889405

Articles in International Peer-Reviewed Journals

[12] N. BERTIN, E. CAMBERLEIN, R. LEBARBENCHON, E. VINCENT, S. SIVASANKARAN, I. ILLINA, F. BIMBOT. *VoiceHome-2, an extended corpus for multichannel speech processing in real homes*, in "Speech Communication", 2018, https://hal.inria.fr/hal-01923108

[13] M. CHAFII, J. PALICOT, R. GRIBONVAL, F. BADER. *Adaptive Wavelet Packet Modulation*, in "IEEE Transactions on Communications", July 2018, vol. 66, n° 7, pp. 2947-2957 [*DOI :* 10.1109/TCOMM.2018.2809586], https://hal-centralesupelec.archives-ouvertes.fr/hal-01713821

[14] I. DOKMANIĆ, R. GRIBONVAL. *Concentration of the Frobenius norm of generalized matrix inverses*, in "SIAM Journal on Matrix Analysis and Applications", 2018, https://arxiv.org/abs/1810.07921 - Revised/condensed/renamed version of preprint "Beyond Moore-Penrose Part II: The Sparse Pseudoinverse", https://hal.inria.fr/hal-01897046

[15] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Sketching for Large-Scale Learning of Mixture Models*, in "Information and Inference", September 2018, vol. 7, n° 3, pp. 447-508, https://arxiv.org/abs/1606.02838 - to appear in Information and Inference, a journal of the IMA (available online since December 2017) [*DOI :* 10.1093/IMAIAI/IAX015], https://hal.inria.fr/hal-01329195

[16] L. LE MAGOAROU, R. GRIBONVAL, N. TREMBLAY. *Approximate fast graph Fourier transforms via multi-layer sparse approximations*, in "IEEE transactions on Signal and Information Processing over Networks", June 2018, vol. 4, n° 2, pp. 407–420, https://arxiv.org/abs/1612.04542 [*DOI :* 10.1109/TSIPN.2017.2710619], https://hal.inria.fr/hal-01416110

[17] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Aural servo: sensor-based control from robot audition*, in "IEEE Transactions on Robotics", June 2018, vol. 34, n° 3, pp. 572-585 [*DOI :* 10.1109/TRO.2018.2805310], https://hal.inria.fr/hal-01694366

[18] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Exploiting the distance information of the interaural level difference for binaural robot motion control*, in "IEEE Robotics and Automation Letters", July 2018, vol. 3, n° 3, pp. 2048-2055 [*DOI :* 10.1109/LRA.2018.2806560], https://hal.inria.fr/hal-01712981

[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-01347455

[20] 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", July 2018, vol. 45, n° 1, pp. 170–205, https://arxiv.org/abs/1510.00504 [*DOI :* 10.1016/J.ACHA.2016.08.004], https://hal.inria.fr/hal-01207987

[21] C. ZHENG, A. DELEFORGE, X. LI, W. KELLERMANN. *Statistical Analysis of the Multichannel Wiener Filter Using a Bivariate Normal Distribution for Sample Covariance Matrices*, in "IEEE/ACM Transactions on Audio, Speech and Language Processing", May 2018, vol. 26, n° 5, pp. 951 - 966 [*DOI :* 10.1109/TASLP.2018.2800283], https://hal.inria.fr/hal-01909612

International Conferences with Proceedings

[22] M. CHAFII, J. PALICOT, R. GRIBONVAL, F. BADER. *Fourier Based Adaptive Waveform*, in "ICT 2018 - 25th International Conference on Telecommunications", Saint-Malo, France, IEEE, June 2018, pp. 37-41 [*DOI :* 10.1109/ICT.2018.8464864], https://hal.archives-ouvertes.fr/hal-01779764

[23] A. CHATALIC, R. GRIBONVAL, N. KERIVEN. *Large-Scale High-Dimensional Clustering with Fast Sketching*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 4714-4718 [*DOI :* 10.1109/ICASSP.2018.8461328], https://hal.inria.fr/hal-01701121

[24] N. COURTY, R. FLAMARY, M. DUCOFFE. *Learning Wasserstein Embeddings*, in "ICLR 2018 - 6th International Conference on Learning Representations", Vancouver, Canada, April 2018, pp. 1-13, https://hal.inria.fr/hal-01956306

[25] B. B. DAMODARAN, B. KELLENBERGER, R. FLAMARY, D. TUIA, N. COURTY. *DeepJDOT: Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation*, in "ECCV 2018 - 15th European Conference on Computer Vision", Munich, Germany, LNCS, Springer, September 2018, vol. 11208, pp. 467-483, https://arxiv.org/abs/1803.10081 - European Conference on Computer Vision 2018 (ECCV-2018) [*DOI :* 10.1007/978-3-030-01225-0_28], https://hal.inria.fr/hal-01956356

[26] D. DI CARLO, A. LIUTKUS, K. DÉGUERNEL. *Interference reduction on full-length live recordings*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 736-740 [*DOI :* 10.1109/ICASSP.2018.8462621], https://hal.inria.fr/hal-01713889

[27] C. FRAGA DANTAS, J. E. COHEN, R. GRIBONVAL. *Learning fast dictionaries for sparse representations using low-rank tensor decompositions*, in "LVA/ICA 2018 - 14th International Conference on Latent Variable Analysis and Signal Separation", Guildford, United Kingdom, LNCS, Springer, July 2018, vol. 10891, pp. 456-466 [*DOI :* 10.1007/978-3-319-93764-9_42], https://hal.inria.fr/hal-01709343

[28] C. FRAGA DANTAS, R. GRIBONVAL. *Faster and still safe: combining screening techniques and structured dictionaries to accelerate the Lasso*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, AB, Canada, IEEE, April 2018, pp. 4069-4073 [*DOI :* 10.1109/ICASSP.2018.8461514], https://hal.inria.fr/hal-01706392

[29] C. GAULTIER, N. BERTIN, R. GRIBONVAL. *CASCADE : Channel-Aware Structured Cosparse Audio DEclipper*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 571-575 [*DOI :* 10.1109/ICASSP.2018.8461694], https://hal.inria.fr/hal-01714667

[30] R. F. IBARRA-HERNÁNDEZ, N. BERTIN, M. A. ALONSO-ARÉVALO, H. A. GUILLÉN-RAMÍREZ. *A benchmark of heart sound classification systems based on sparse decompositions*, in "SIPAIM 2018 - 14th International Symposium on Medical Information Processing and Analysis", Mazatlán, Mexico, October 2018, pp. 1-14, https://hal.inria.fr/hal-01935058

[31] H. JAIN, J. ZEPEDA, P. PÉREZ, R. GRIBONVAL. *Learning a Complete Image Indexing Pipeline*, in "CVPR 2018 - IEEE Conference on Computer Vision and Pattern Recognition", Salt Lake City, United States, IEEE, June 2018, https://arxiv.org/abs/1712.04480 , https://hal.inria.fr/hal-01683385

[32] N. KERIVEN, A. DELEFORGE, A. LIUTKUS. *Blind Source Separation Using Mixtures of Alpha-Stable Distributions*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 771-775, https://arxiv.org/abs/1711.04460 [*DOI :* 10.1109/ICASSP.2018.8462095], https://hal.inria.fr/hal-01633215

[33] A. LIUTKUS, C. ROHLFING, A. DELEFORGE. *Audio source separation with magnitude priors: the BEADS model*, in "ICASSP 2018 – IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, Signal Processing and Artificial Intelligence: Changing the World, April 2018, pp. 1-5 [*DOI :* 10.1109/ICASSP.2018.8462515], https://hal.inria.fr/hal-01713886

[34] H. PEIC TUKULJAC, A. DELEFORGE, R. GRIBONVAL. *MULAN: A Blind and Off-Grid Method for Multichannel Echo Retrieval*, in "NeurIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, pp. 1-11, https://arxiv.org/abs/1810.13338 , https://hal.inria.fr/hal-01906385

[35] R. SCHEIBLER, D. DI CARLO, A. DELEFORGE, I. DOKMANIĆ. *Separake: Source Separation with a Little Help From Echoes*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, April 2018, https://hal.inria.fr/hal-01909531

[36] V. SEGUY, B. B. DAMODARAN, R. FLAMARY, N. COURTY, A. ROLET, M. BLONDEL. *Large-Scale Optimal Transport and Mapping Estimation*, in "ICLR 2018 - International Conference on Learning Representations", Vancouver, Canada, April 2018, pp. 1-15, https://arxiv.org/abs/1711.02283 - 15 pages, 4 figures. To appear in the Proceedings of the International Conference on Learning Representations (ICLR) 2018, https://hal.inria.fr/hal-01956354

[37] M. STRAUSS, P. MORDEL, V. MIGUET, A. DELEFORGE. *DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization*, in "IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)", Madrid, Spain, October 2018, https://hal.inria.fr/hal-01854878

National Conferences with Proceedings

[38] N. LIBERMANN, F. BIMBOT, E. VINCENT. *Exploration de dépendances structurelles mélodiques par réseaux de neurones récurrents*, in "JIM 2018 - Journées d'Informatique Musicale", Amiens, France, May 2018, pp. 81-86, https://hal.archives-ouvertes.fr/hal-01791381

Conferences without Proceedings

[39] C. ELVIRA, H.-P. DANG, P. CHAINAIS. *Small variance asymptotics and bayesian nonparametrics for dictionary learning*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Rome, Italy, September 2018, pp. 1607-1611 [*DOI :* 10.23919/EUSIPCO.2018.8553142], https://hal.archives-ouvertes.fr/hal-01961852

[40] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN. *A case of exact recovery with OMP using continuous dictionaries*, in "CS 2018 - 9th International Conference on Curves and Surfaces", Arcachon, France, June 2018, https://hal.inria.fr/hal-01937532

[41] C. GUICHAOUA, F. BIMBOT. *Inférence de segmentation structurelle par compression via des relations multi-échelles dans les séquences d'accords*, in "JIM 2018 - Journées d'Informatique Musicale", Amiens, France, May 2018, pp. 71-79, https://hal.archives-ouvertes.fr/hal-01791367

[42] N. KERIVEN, R. GRIBONVAL. *Instance Optimal Decoding and the Restricted Isometry Property*, in "International Conference on New Computational Methods for Inverse Problems (NCMIP)", Cachan, France, May 2018, https://arxiv.org/abs/1802.09905 , https://hal.inria.fr/hal-01718411

[43] Y. TRAONMILIN, S. VAITER, R. GRIBONVAL. *Is the 1-norm the best convex sparse regularization?*, in "iTWIST'18", Marseille, France, Proceedings of iTWIST'18, November 2018, https://arxiv.org/abs/1806.08690 , https://hal.archives-ouvertes.fr/hal-01819219

Scientific Books (or Scientific Book chapters)

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

[45] S. KITIĆ, S. BENSAID, L. ALBERA, N. BERTIN, R. GRIBONVAL. *Versatile and scalable cosparse methods for physics-driven inverse problems*, in "Compressed Sensing and its Applications – Second International MATHEON Conference 2015", H. BOCHE, G. CAIRE, R. CALDERBANK, M. MÄRZ, G. KUTYNIOK, R. MATHA (editors), Series: Applied and Numerical Harmonic Analysis, Birkhaüser Basel, March 2018, pp. 291-332 [*DOI :* 10.1007/978-3-319-69802-1_10], https://hal.inria.fr/hal-01496767

[46] Y. TRAONMILIN, G. PUY, R. GRIBONVAL, M. E. DAVIES. *Compressed sensing in Hilbert spaces*, in "Compressed Sensing and its Applications – Second International MATHEON Conference 2015", H. BOCHE, G. CAIRE, R. CALDERBANK, M. MÄRZ, G. KUTYNIOK, R. MATHA (editors), Series: Applied and Numerical Harmonic Analysis, Birkhaüser Basel, March 2018, pp. 359–384, https://arxiv.org/abs/1702.04917 [*DOI :* 10.1007/978-3-319-69802-1_12], https://hal.archives-ouvertes.fr/hal-01469134

Scientific Popularization

[47] R. GRIBONVAL. *D'APB à Parcoursup : quelles méthodes d'affectation post-bac ?*, in "Interstices", February 2018, https://hal.inria.fr/hal-01827604

Other Publications

[48] E. BYRNE, A. CHATALIC, R. GRIBONVAL, P. SCHNITER. *Sketched Clustering via Hybrid Approximate Message Passing*, January 2019, https://arxiv.org/abs/1712.02849 - working paper or preprint, https://hal.inria.fr/hal-01991231

[49] C. FRAGA DANTAS, R. GRIBONVAL. *Stable safe screening and structured dictionaries for faster ℓ_1 regularization*, December 2018, https://arxiv.org/abs/1812.06635 - working paper or preprint, https://hal.inria.fr/hal-01954261

[50] R. GRIBONVAL, M. NIKOLOVA. *A characterization of proximity operators*, July 2018, https://arxiv.org/abs/1807.04014 - This work and the companion paper [17] are dedicated to the memory of Mila Nikolova, who passed away prematurely in June 2018. Mila dedicated much of her energy to bring the technical content to completion during the spring of 2018. The first author did his best to finalize the papers as Mila would have wished. He should be held responsible for any possible imperfection in the final manuscript, https://hal.inria.fr/hal-01835101

[51] R. GRIBONVAL, M. NIKOLOVA. *On bayesian estimation and proximity operators*, July 2018, https://arxiv.org/abs/1807.04021 - This work and the companion paper [10] are dedicated to the memory of Mila Nikolova, who passed away prematurely in June 2018. Mila dedicated much of her energy to bring the technical content to completion during the spring of 2018. The first author did his best to finalize the papers as Mila would have wished. He should be held responsible for any possible imperfection in the final manuscript, https://hal.inria.fr/hal-01835108

[52] T. LE NAOUR, N. COURTY, S. GIBET. *Kinematic driven by distances*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01838784

[53] A. RAKOTOMAMONJY, A. TRAORÉ, M. BERAR, R. FLAMARY, N. COURTY. *Distance Measure Machines*, November 2018, https://arxiv.org/abs/1803.00250 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01717940

References in notes

[54] 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° 3, pp. 922 - 932 [*DOI :* 10.1109/TASL.2011.2168211], http://hal.inria.fr/inria-00577079

[55] 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

[56] 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

[57] 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° 19, 20 p. [*DOI :* 10.1109/TSP.2015.2447503], https://hal.archives-ouvertes.fr/hal-01084986

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

[59] 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

[60] 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

[61] 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

[62] 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

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

[64] 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

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

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

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

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

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

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

[71] 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$^{\text{o}}$ 6, pp. 1037–1048

[72] 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$^{\text{o}}$ 01, 1440003 p., https://hal.archives-ouvertes.fr/hal-00960796

[73] 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$^{\text{o}}$ 5, pp. 893–911

[74] 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, nᵒ 4, pp. 718–731

[75] 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://hal.inria.fr/hal-01372418

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

[77] I. DOKMANIĆ, R. GRIBONVAL. *Beyond Moore-Penrose Part II: The Sparse Pseudoinverse*, July 2017, See https://hal.inria.fr/hal-01897046 for a revised/condensed/renamed version of this preprint, https://hal.inria.fr/hal-01547283

[78] I. DOKMANIĆ, R. PARHIZKAR, A. WALTHER, Y. M. LU, M. VETTERLI. *Acoustic echoes reveal room shape*, in "Proceedings of the National Academy of Sciences", 2013, vol. 110, nᵒ 30, pp. 12186–12191, http://dx.doi.org/10.1073/pnas.1221464110

[79] 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, nᵒ 3, pp. 1428 - 1440 [*DOI :* 10.1109/TIP.2017.2654165], https://hal.inria.fr/hal-01413406

[80] 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

[81] 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

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

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

[84] 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

[85] 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

[86] M. A. GERZON. *Periphony: With-Height Sound Reproduction*, in "J. Audio Eng. Soc", 1973, vol. 21, nᵒ 1, pp. 2–10, http://www.aes.org/e-lib/browse.cfm?elib=2012

[87] 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

[88] R. GRIBONVAL, G. BLANCHARD, N. KERIVEN, Y. TRAONMILIN. *Compressive Statistical Learning with Random Feature Moments*, December 2017, Main novelties compared to version 1: improved concentration bounds, improved sketch sizes for compressive k-means and compressive GMM that now scale linearly with the ambient dimension, https://hal.inria.fr/hal-01544609

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

[90] 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

[91] R. F. IBARRA-HERNÁNDEZ, M. A. ALONSO-ARÉVALO, S. VILLARREAL, C. I. NIEBLAS. *A parametric model for heart sounds*, in "Signals, Systems and Computers, 2015 49th Asilomar Conference on", IEEE, 2015, pp. 765–769

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

[93] 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, Accepted at ICCV 2017 (Spotlight) [*DOI :* 10.1109/ICCV.2017.96], https://hal.inria.fr/hal-01683390

[94] 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://hal.inria.fr/hal-01372435

[95] 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

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

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

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

[99] 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

[100] 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

[101] 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

[102] 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

[103] 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° 10, pp. 2498 - 2511 [*DOI :* 10.1109/TSP.2013.2250967], https://hal.archives-ouvertes.fr/hal-00691774

[104] 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

[105] 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

[106] 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

[107] 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

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

[109] L. LE MAGOAROU, N. TREMBLAY, R. GRIBONVAL. *Analyzing the Approximation Error of the Fast Graph Fourier Transform*, in "51st Asilomar Conference on Signals, Systems, and Computers (ACSSC 2017)", Monterey, California, United States, October 2017, https://hal.archives-ouvertes.fr/hal-01627434

[110] 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

[111] 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

[112] 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

[113] 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

[114] 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

[115] 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

[116] 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

[117] 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

[118] 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

[119] 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° 9, pp. 1652–1664

[120] 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° 3, pp. 550–563

[121] 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, Submitted in 2015 [*DOI : 10.1109/TIT.2017.2664858*], https://hal.inria.fr/hal-01203614

[122] 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