Interdisciplinary approaches for Neurosciences, Artificial Intelligence and Sound

Nicolas Farrugia

Université de Brest Occidentale

Ecole Doctorale MATHSTIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

Habilitation à Diriger des Recherches

de l'Université de Brest Occidentale

Soutenue par

Nicolas FARRUGIA

Interdisciplinary approaches for Neurosciences, Artificial Intelligence and Sound

soutenue le 27 Juin 2022

Jury :

Rapporteurs :	Dimitri VAN DE VILLE	-	EPFL
	Peter VUUST	-	CFIN Aarhus (Denmark)
	Geoffroy PEETERS	-	Telecom Paris
Examinateurs :	Serge TIMSIT	-	Université de Brest Occidentale
	Mathieu PAQUIER	-	Université de Brest Occidentale
	Sophie ACHARD	-	Université de Grenoble Alpes

Acknowledgments

Brest, 1st of May, 2022.

The work described in this manuscript would not have been possible without the support of several people that I want to acknowledge here. First, as a graduate in electrical engineering, being able to embark on an endeavour in Auditory Cognitive Neurosciences would not have been possible without the confidence that Sonja Kotz, Simone Dalla Bella and Lauren Stewart have placed in my applications as a postdoctoral fellow in their respective teams in Leipzig, Warsaw and London, regardless of my background, and how they always supported me throughout the years. 2010 to 2013- the period of the EBRAMUS (Europe, Brain and Music) project, was without a doubt the most inspiring times in my scientific life, filled with countless human experiences, musical moments and scientific enlightenmnents. There would be too many people to acknowledge properly, but I would like to remember the support and friendship of Tom Fritz, Charles-Etienne Benoit, Marieke Van Der Steen, Eleanor Harding, Laura Verga, Richard Penninger, Sujin Hong, Nia Cason, Irene Fernandez, Laura Ferreri, Floris Van Vugt, Ahmad Kanaan, Ellen Schuessler, Georgia Floridou, Kelly Jakubowksi, the scientific inspiration from Michael Schwartze, Jonathan Smallwood, Daniel Margulies, and the motivation and drive of all the brilliant students and scientists met along the way. I also want to acknowledge the crucial role that BrainHack and the open science community played in my personal and scientific development, thanks to enriching and stimulating encounters during the hackathon I participated to, with people such as Gaël Varoquaux, Alexandre Granfort and Pierre Bellec, who later became a close collaborator.

In 2015 came another time of change, serendipity took me back to France, boosted by the confidence that Vincent Gripon and Michel Jezequel, and more formally IMT Atlantique, have put in my background. Being a former electrical engineer, now auditory cognitive neuroscientist, to be back as an associate professor in an engineering school came with a series of challenges, in particular maintaining consistency with my previous appointments. But the scientific environment provided by the team of the NEUCOD Advanced ERC Grant, led by Claude Berrou, has pushed me to develop original lines of research at the crossroads between artificial intelligence and neurosciences, with the continuous collaboration, mutual help and support of Vincent Gripon and Michel Jezequel to which I am deeply grateful. In this time of development, I also want to say thank you to my colleagues at IMT Atlantique for their support, and in particular to Carlos Lassance, Alan Aboudib, Jean-Charles Vialatte and Bastien Pasdeloup for their friendship.

As our research efforts were developing and our team was growing, in early 2020 came the COVID19 pandemic. I would like to acknowledge the ones who suffer through this crisis, everyone who helped in supporting each other in this difficult period, and the students for their patience and understanding. Despite the challenges of remote working and the lack of direct social interaction, our team managed to keep connected and caring, supported by our institution IMT Atlantique. In addition, being a scientist in the midst of a global pandemic also comes with a range of questions, on the sense and value that we provide to society. It is clearly thanks to friendship that I contributed to initiate the Silent Cities project (described in section 3.4.3), an idea of my friend with Samuel Challéat, and that we developed together. The genesis of Silent Cities was not to make just another science project, but was also finding something purposeful and inspiring, through continuous interactions with Samuel. I am grateful of this very special moment we shared when we poured all our scientific and personal enthusiasm in attempting a decentralized global data collection despite of the difficult context. With work conditions slowly coming back to "normal", our team developped quickly with the arrival of new permanent members Mathieu Leonardon and Giulia Lioi, that I want to thank personally for their friendship, trust and confidence, and for thriving to build together an excellent, safe and diverse team.

I would like to thank the reviewers of this manuscript for accepting to spend their precious time to give feedback. Each of them has a special role in giving purpose to this manuscript. I could not have wished a better scenario, as each of them represents different phases in the kind of contributions I am trying to make as a scientist.

Final words of acknowledgements will be for my family. I would like to dedicate this manuscript to my mother Lynda, who passed away in 2017. I want to thank my family, my father, my brothers and my partner Laetitia who support me unconditionally throughout all these years, despite many relocations around the world.

Contents

Chapter 1

Foreword

This manuscript presents a summary of my work from 2008 to 2022, and outlines research perspectives for the years to come. As always in such an exercise, there are choices and compromises on what to present. Here, I propose an humble attempt at a synthesis of past work and future perspectives, approached by the angle of Interdisciplinary approaches for neurosciences, artificial intelligence and sound. It is not by chance that sound can appear as a common denominator of most of the work presented here. Initially trained as a conservatory musician (until 1998), I moved on to scientific studies in electrical engineering up to the PhD level (until 2008), before moving on to Auditory Cognitive Neurosciences (until 2015), and finally joining IMT Atlantique as an assistant professor to contribute to a research effort in artificial intelligence and neurosciences (since 2016).

But why studying sound in the first place ? Sound affects us humans in various ways. Music, the art of placing sounds in space and time, is universal, affects our mood and emotions, shapes our identities and can even have powerful therapeutical effects. Being able to perceive and make music, and in particular being able to produces sounds in time, to synchronize to other sounds or to perceive a beat in music, are all very remarkable abilities of human beings, shared by very few other animal species. It is believed that such abilities to perceive regularities and synchronize with others have helped shape human societies, by fostering interpersonal coordination, boosting collaboration to achieve difficult tasks.

As a result, when awake, humans generally spent more time moving, thus making some sound, than standing still. Sound itself is produced by any movement, a breath, a heartbeat, or any vibration triggered not only by us, but also by what we have built. Such sounds produced by our cars, planes, trains or factories can be disruptive for human beings, as well as for other animal species. Sound can therefore be a pollution, in which case we tend to rather refer to it as "noise". But in those moments when we don't make or hear any sound, we might still be able to hear it in our heads. Those short loops of music or sounds that are sustained in our imagination are sometimes called "earworms", and are also very common. Earworms can be disruptive or helpful, can be elusive or stay for several days or weeks, and can be triggered by various reasons, sometimes unknown. Little is known about how the brain is able to self-generate auditory content without any actual vibration in the eardrum, but recent cognitive neuroscience research attempts at uncovering spontaneous brain processes, such as earworms or other self-generated thoughts happening when we daydream, by analyzing how neural activity is coupled when the brain is at rest.

The analysis of the brain's "resting-state" therefore consists in understanding the natural fluctuations of brain activity, and relate it to individual differences in phenotype. But the brain is a highly complex organ, composed of close to 170 billion cells including about 80 billion neurons, of which about 50 billion are actually in the cerebellum. Recent developments in imaging techniques now enable to measure brain structure and in-vivo brain activity with a spatial resolution of less than a few millimeters, and machine learning has helped to decompose spatial patterns of brain activity into a few hundred regions - this is called brain parcellation. Using these parcellations, it is possible to model the brain as a graph, with regions as graph vertices, and connections between regions as graph edges. Connections can be estimated either using imaging techniques that estimate white matter tracts (i.e. anatomical connectivity), or by computing statistical correlation. Once a graph vertices and edges are defined, graph theory and graph signal processing are powerful tools that can be exploited to decompose, learn and interpret patterns of brain activity, indvidual differences in brain connectivity as a function of phenotype, or help diagnosing disease.

Interdisciplinarity means being able to talk to several communities. Therefore, the first and most visible ambition of this manuscript is to be readable for members of distinct communities, namely cognitive neurosciences, machine learning for neuroimaging and audio signal processing (more specifically, sound event recognition, bioacoustics and ecoacoustics, and music information retrieval). In addition, the higher ambition of the work presented here is to contribute to an integration of several approaches, to open new possibilities in understanding facets of sound and its beneficial and detrimental effects on human beings. While we used the term "Artificial Intelligence" in the title, we will leave this expression aside for the rest of this manuscript ; as convenient as it can be for a general picture, we will prefer the terms Machine Learning and Deep Learning, as they are reflective of the modern methods in the field. In particular, some of my previous work includes a reflexion on the cost of Deep Learning in terms of computational and memory requirements, and I contribute to research that aims at understanding how to optimize such methods for efficiency at the algorithmic and architectural level. These optimizations can also be useful for interpretations of decisions taken by machine learning or deep learning approaches, therefore are also relevant for scientific discovery on the functionning of the human brain or for sound research.

Chapter 2 provides a summary of my past research activities. The chapter is divided in two parts, highlighting first my contributions in Auditory Cognitive Neurosciences, and second my work in Machine Learning and Graph Signal Processing. Each part itself begins with an overview of the main research questions and related work. Chapter 3 presents my research statement, which will include current projects, perspectives and future directions for my research. Some of the work included in this chapter was done by students who already finished, but have open perspectives that have their place here rather than in the previous work. Chapter 3 also contains some more long-term perspectives, on work that is only starting or planned. Finally, an extended CV is provided in appendix A. This appendix includes :

- a detailed list of acquired grants and funding in section A.3.
- a summary of my student supervision activities in section A.2
- a summary of my teaching activities in section A.4

Chapter 2

Overview of past research activities

Contents

2.1 Auditory Cognitive Neurosciences

2.1.1 Overview

Without specific effort, a large majority of healthy, adult human beings are able to entrain their movements to a regular rhythm, for example by tapping their foot or hands at a concert. This is the ability of sensorimotor synchronization (SMS), which enables beings to perform appropriately timed movements that are adapted to or synchronized with the environment. SMS is needed in daily life to engage in a conversation, to walk, or in other situations such as catching a ball or rowing, to name a few. In section 2.1.3, I describe my contributions to study of SMS in a child drummer prodigy, preschoolers, expert musicians, healthy non-trained musicians, and in patients with Parkinson's Disease (PD).

While the main function of SMS is to interact with the environment, several underlying, internal brain processes are at play. Recent theories suggest that the human brain has the ability to model the environment and predict future changes, and such an ability would naturally relate to SMS. Therefore, pathological situations such as PD, in which SMS is altered, can inform us on the neural underpinnings of SMS (section 2.1.5). In addition, it has been shown that specific rehabilitation strategies based on rhytmical entrainment can help alleviate motor symptoms in PD (section 2.1.4).

In the absence of sound, it is also possible to mentally recreate sensory content. This phenomenon, called musical imagery, can also be studied as a particular instance of SMS, for example if someone is tapping its hand to a music currently being imagined. Musical imagery can be deliberate and voluntary, but it can also be involuntary, uncontrolled and intrusive, in which case we refer to it as Involuntary Musical Imagery (INMI). Section 2.1.6 presents some contributions to understanding the neural correlates of INMI, as well as SMS to musical imagery.

2.1.2 Methods

We present in figure 2.1 an overview of the acquisition devices (panel A), paradigms (panel B) and analysis techniques (panel C) that we used to assess SMS in diverse situations. Electronic drums (e.g. Roland TD series) or drumpads (such as the AKAI MPD series) can record precise timing using the MIDI standard, while stimuli are presented to the subject using headphones or loudspeakers. For a more ecological setting, acoustic drums are used in combination with contact microphones based on piezoelectric pickups. In order to measure SMS while walking over music, we use motion capture (Mocap) with passive infrared markers on the lower limbs, while presenting the stimuli using wireless headphones. Finally, we measured SMS in subjects in their daily life, by capturing finger tapping with a wrist-worn watch that embeds an accelerometer. The proposed paradigms (self-paced, paced and synchronization continuation) enable to measure both the spontaneous rates, such as the spontaneous motor tempo or natural walking cadence, as well as synchronization with external stimuli.

Neuroimaging methods are presented in figure 2.2. We use Electroencephalography (EEG, Panel A) as a measure of time-resolved brain activity with a high temporal resolution (up to 500 Hz). Auditory stimuli are presented to subjects using loudspeakers in a soundproof booth (alternatively, a quiet small room). The precise timing of the onset of auditory stimuli is used to gather EEG responses time-locked to the stimuli. Event-related potentials (ERP) are estimated for each subject and condition, by averaging condition-wise stimuli repetitions. We use ERP to study the effect of temporal regularity in an auditory oddball experiment with healthy subjects and patients with PD (section 2.1.5).

Panels B of figure 2.2 presents the use of Magnetic Resonance Imaging (MRI) to study individual differences in brain structure. Recent progresses in MRI acquisition techniques enable to obtain high resolution 3D reconstructions of the brain, in order to measure the thickness of the cortical sheet, as well as estimates of local grey matter volume (also called grey matter density). We used these techniques to study the relationship between brain structure and individual differences in the experience of INMI, detailed in section 2.1.6.

Finally, Panel C describes the paradigm of resting-state fMRI, under which a subject is asked to relax and keep eyes open while fixating a cross on a screen, for a duration of at least six minutes. Resting-state fMRI measures spontaneous fluctuations in brain activity, and can be quantified using statistical correlation between signals, referred to as functional connectivity. This can be done by selecting a reference region, called a seed, and compute correlation across all other locations in the brain, thus obtaining a whole brain map of functional connectivity with the seed. A more comprehensive estimate of functional connectivity can be obtained by parcellating the brain into a set of regions of interest (ROI), and compute all pairwise correlations between ROIs, therefore yielding a functional connectivity matrix. We have measured resting-state functional connectivity in relationship with the experience of INMI (see section 2.1.6). In addition, we will show how to decompose, interpret and explore functional connectivity matrices

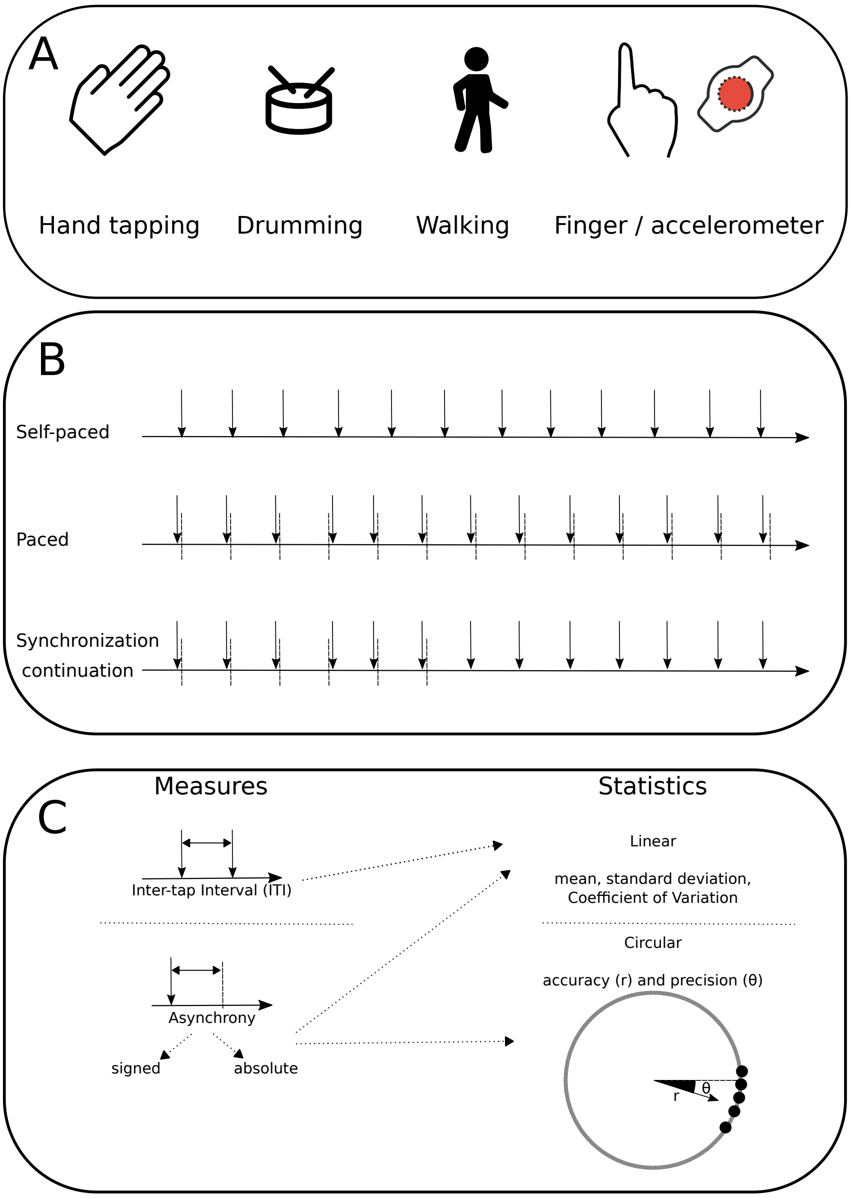


Figure 2.1: Methods for Sensorimotor Synchronization (SMS). Panel A : Acquisition techniques: tapping with hands, fingers on an worn wrist-band with accelerometers, or drumsticks, as well as walking using motion capture of lower limbs movement. Panel B: Experimental paradigms. Taps or steps performed by subjects are indicated with arrows, and external stimuli are dotted lines. Self-paced tapping or walking is performed to measure spontaneous rates, without any external reference. Paced paradigms involve an external stimulus on which the subject is asked to synchronize his movements. Synchronization Continuation paradigms involve a first synchronization phase, followed by a self-paced phase during which the subject is instructed to keep the previously paced rate. Panel C : Inter-tap interval (or interval between step onset for walking) is used to measure a global tendency of subjects to tap at a regular rate, in particular using the coefficient of variation, while asynchronies are used in paced paradigms in order to assess how close to the external stimulation subjects are able to synchronize. Linear statistics inform on average tendencies of subjects to anticipate or delay their taps / steps with respect to the external stimulus. In circular statistics, each tap (filled circles) is a vector on the unit circle, and the angle is defined relative to the periodic interval between stimuli. The average vector length r, also called accuracy, is indicative of the consistency of tap asynchronies, while the its angle θ is its precision relative to the external stimuli.

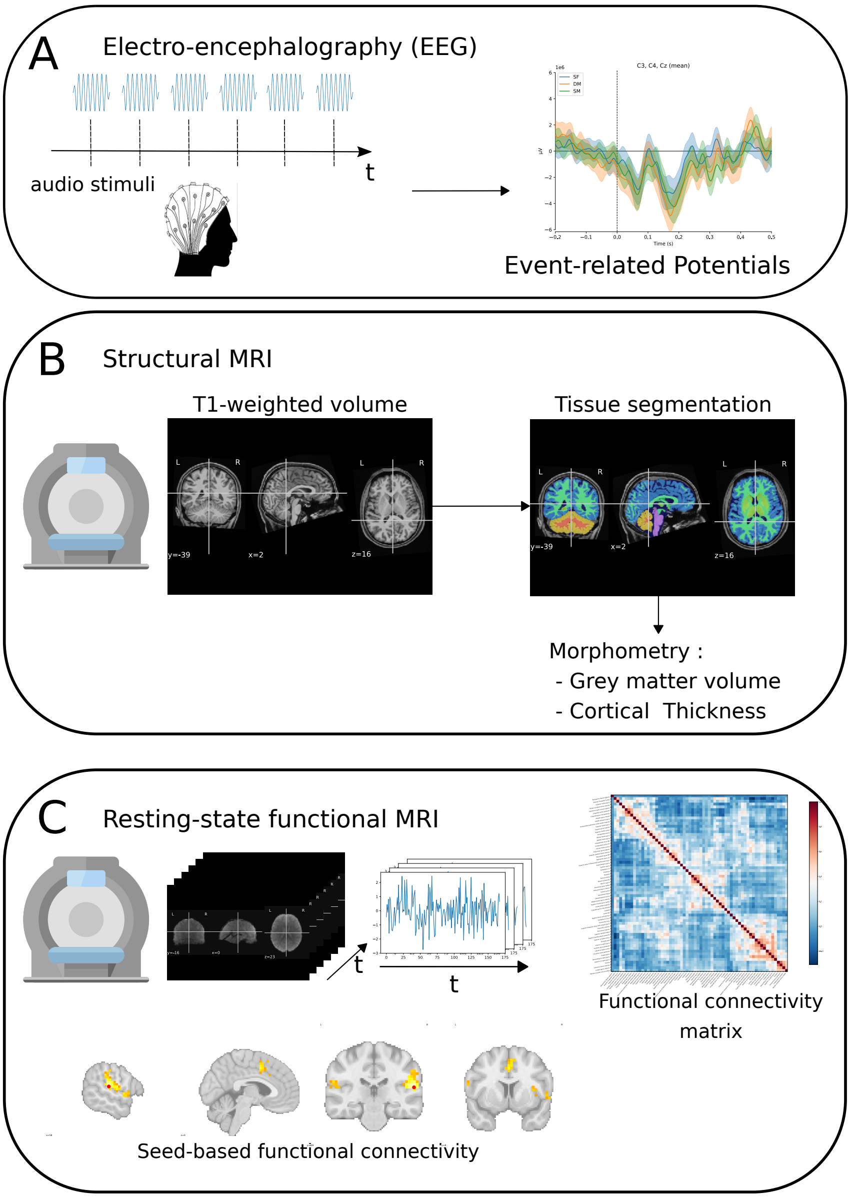


Figure 2.2: Neuroimaging methods. Panel A : Electroencephalography (EEG) consists in measuring the electrical current on the scalp, typically using a few dozen to a few hundred electrodes. Scalp-level electrical activity results partly from brain activity emerging from superficial layers of the cortex. Brain activity time-locked to controlled stimuli can be measured by averaging EEG signals corresponding to the same trial types, which is the study of Event Related Potentials (ERP). Panel B : Structural MRI. Specific acquisition sequences, for example T1-weighted sequences, are used to obtain images which can be segmented according to brain tissue types, such as gray matter and white matter. Cortical Thickness and Gray Matter Morphometry can subsequently be measured, yielding markers of individual differences in brain structure. Panel C : Functional MRI. Echo-planar Imaging (EPI) MRI sequences have a contrast that is sensitive to local changes in blood-oxygen concentration, which occur following neuronal activiation ; this is called the hemodynamic response. In the work summarized here, we measured resting-state brain activity (rsfMRI), by asking subject to relax while keeping their eyes open and fixating a cross on a screen, for at least 7 minutes. Statistical correlation is subsequently used to study how signals in regions of interest (ROI) are correlated with the rest of the brain. We show here an exemple using a seed region in the auditory cortex (red dot, bottom left panel), as well as a functional connectivity (FC) matrix depicting all pairwise correlations between several hundred ROIs.

with machine learning and graph signal processing approaches in the sections 2.2.4 and 2.2.5.

2.1.3 Sensorimotor synchronization and rhythmical entrainment

2.1.3.1 Rhythmical entrainment in preschoolers

This project was a collaboration between the team of Simone Dalla Bella (at that time, in Université de Montpellier 1, France and WSFiZ, Warsaw, Poland) and Fabia Franco (Middlesex University, London, UK), in the context of the PhD of Nina Politimou (Middlesex University, London, UK). I contributed with analysis techniques on synchronization tasks that were developed for the BAASTA, described in section 2.1.3.4. The context of this work is the study of a link between language and musical abilities. Previous studies have already shown the sensitivity of infants (less than one year old) to rhythmical regularity, as well as a capacity to entrain to regular rhythms from five year old, but an assesment of such abilities between two and four year old was missing. Critically, other studies suggest a link between language development and music. The goal of the two studies reported in [Politimou *et al.* 2019] was to reveal the relationship between language and musical skills in preschool children (40 subjects between 3 and 5 years), also as a function of informal musical experience at home. We found that the best predictors of phonological awareness were the variability of synchronization to a beat and rhythm perception. Those musical skills were better predictors than verbal memory and other cognitive skills such as non-verbal ability.

2.1.3.2 Rhythmical entrainment in a child drummer prodigy

As a postdoctoral fellow at the MPBLAB in Warsaw, in 2010-2011, I contributed to the analysis of previously collected data on SMS and rhythmical entrainement in a child drummer prodigy, Igor. Under the supervision of Simone Dalla Bella, this project was the PhD project of Jakub Sowinski, and another PhD student, Magdalena Berkowska, also contributed to this project. I performed the analysis of two experiments performed on Igor. The two paragraphs below are summarized and adapted from a book chapter I contributed to [Dalla Bella *et al.* 2016].

The first experiment was targeted at testing the learning of rhytmical patterns with varying level of metrical structure. Participants learned to play strongly metrical (SM), weakly metrical (WM) and non isochronous (NI) patterns on a percussion pad, which respectively contained none (SM), one (WM), or two (NI) missing strong beats in the sequence. The analysis of learning performance confirmed that Igor was more accurate than adults in playing the note durations in SM and WM patterns. However, his performance declined with NI sequences, while adult participants were able to learn those sequences successfully. With such sequences lacking an isochronous beat, Igor performed worse than the majority of adult nonmusicians. These findings suggest that Igor's rhythmic skills, which afford exceptional synchronization to the beat at his young age, are associated with outstanding skills in extracting the beat inherent in a rhythmic pattern. His sensitivity to the beat gives him an advantage when learning novel rhythmic patterns, provided that they possess a beat structure.

In the second experiment [Dalla Bella *et al.* 2011, Dalla Bella *et al.* 2016], the benefits of beat perception on Igor's performance were examined in a motion capture study. The effect of beat isochrony on the performance at the drums was studied by asking Igor, when he was 7 years old, to play a short isochronous rhythmic pattern at the beat indicated by a metronome. In one condition, the metronome was isochronous, while in another condition, the metronome was not always isochronous, but still predictable. It was isochronous while the pattern was played, but the interval between metronome tones corresponding to the break in between the repetitions was greater than the duration of the pattern (nonisochronous condition). The analysis of synchronization accuracy and movement amplitude (using the position of the tip of the drumstick) showed Igor's high sensitivity to beat structure while playing a simple rhythmic pattern along with a metronome (i.e. a predictable auditory stimulus). When the reproduction of a rhythmic pattern was timed by an isochronous beat, Igor exhibited greater anticipatory movement (defined as the time of the peak movement amplitude) prior to stimulus production (i.e. with higher amplitude and greater movement anticipation time) than children who received 1 to 2.5 years of drum training. This advantage disappeared when Igor was asked to play the same pattern with an auditory stimulus lacking isochrony. This is visible in Igor's sometimes insurmountable difficulty in executing the task and his reduced sensitivity to the metrical structure of the rhythmic pattern, as shown by the reduced movement amplitude in correspondence with the beat. These findings are in keeping with the results obtained in the first experiment, where Igor learned to play simple rhythmic patterns better than adult nonmusicians provided that the pattern conveyed a clear isochronous beat.

2.1.3.3 Anticipatory movements and expert performance in professional drummers

I summarize here an original study imagined, designed and run under the supervision of Simone Dalla Bella, while I was a visiting postdoctoral fellow at EUROMOV M2H in Montpellier, France, in the first six months of 2012. I was assisted by Charlotte Roy (at that time, master student) to run the experiments. This study was presented at an oral presentation at the ICMPC conference in 2014 [Farrugia & Dalla Bella 2014].

Skilled musicians are used to play with the beat at different tempos. The way their movements adjust to tempo for achieving high spatial and temporal accuracy is likely to reflect how audiomotor coupling shapes experts' movements. Indeed, motor strategies are built through training, and are commonly used as pedagogical aims (e.g. keeping fingers close to the keyboard to perform fast melodies). In particular, effects of tempo on anticipatory movements are reported in previous studies on piano and clarinet performance, e.g. larger movement amplitude with increasing tempo. In [Farrugia & Dalla Bella 2014], we assessed whether anticipatory movements in drumming vary as a function of musical training in synchronization. We were interested in (1) studying the effect of tempo on anticipatory movement and synchronization accuracy and variability, (2) relating anticipatory movement to temporal accuracy, and finally (3) explaining observed individual differences in movement patterns and synchronization by the amount of musical training. Eight professional drummers with varying degrees of musical training participated in this study. Participants synchronized to an isochronous sequence of tones presented at different tempi, spanning from 60 to 200 BPM. The movement of their right arm and stick was captured by a Vicon system during performance using passive reflective markers. The stick trajectory was analyzed in terms of maximum distance to the pad (movement amplitude), and time when this maximum is reached (movement anticipation). Results show that tempo modulated both synchronization accuracy and movement kinematics. Participants were less accurate and more variable at the fast tempi than at the slower tempi. They raised less the arm and anticipated more the stroke at fast than at slow tempi. Interestingly, both amplitude and anticipation time predicted synchronization accuracy at all tempi, while variability was mostly related to anticipation time. Finally, several measures of musical training could explain observed individual differences in synchronization as well as in movement patterns. Both timing and amplitude of anticipatory movements in drumming contribute to the optimization of sensorimotor performance. The underlying audiomotor coupling mechanism is developed progressively via training.

2.1.3.4 BAASTA : Battery for the Assessment of Auditory and Sensorimotor Abilities

Table 2.1 presents an overview of the perceptual and SMS tasks in the Battery for the Assessment of Auditory and Sensorimotor Abilities (BAASTA), introduced with preliminary data in [Farrugia *et al.* 2012] and in two subsequent validation studies in [Dalla Bella *et al.* 2017b]. The experimental design of some of the tasks was done in previous studies from Simone Dalla Bella or other authors [Dalla Bella *et al.* 2013, Sowiński & Dalla Bella 2013, Hyde & Peretz 2004, Iversen & Patel 2008, Schwartze *et al.* 2011a]. I contributed in finalizing the task design, implemented experimental tasks and performed analysis for the data presented in the following publications: [Benoit *et al.* 2014, Dalla Bella *et al.* 2017b, Farrugia *et al.* 2012, Dalla Bella *et al.* 2015].

BAASTA was designed in order to provide a comprehensive measurement of perceptual timing abilities; duration discrimination consists in judging whether two sounds are of equal duration or not, testing for "interval timing". Anisochrony detection tasks are related to beat perception, by asking subjects to detect a slight temporal shift in the last beat of a sequence. Finally, the Beat Alignement Test [Iversen & Patel 2008] measures the ability of subject to test whether a superimposed beat is in phase or at the right speed with a musical piece.

The production tasks of the BAASTA were designed to test abilities SMS using tapping tasks, as previously reviewed [Repp 2005]. Unpaced tapping is used to measure spontaneous, fastest and slowest rates of tapping. Paced tapping on metronome and music is employed to measure general synchronization abilities. Next, two more tasks are introduced to measure adaptation abilities. The synchronization-continuation task (figure 2.1, panel B) consists in a paced tapping phase, followed by a continuation phase during which the subject has to continue tapping at the same rate [Wing & Kristofferson 1973]. The adaptive tapping task is a synchronization-continuation task, during which a tempo change can occur during the synchronization phase, and subjects have to detect the change and tap at the new tempo in the continuation phase [Schwartze *et al.* 2011a]. In the following paragraphs, we give an overview of the results on a validation study [Dalla Bella *et al.* 2017b] performed on 24 subjects (12 females and 12 males, average age 23.9).

Results obtained for the perceptual tasks show that detection thresholds were higher in duration discrimination than anisochrony detection, while detecting a deviation from isochrony was easier in a musical sequence than in a sequence of tones. The results from the Beat Alignement Test showed that subjects could easily detect when a metronome was not aligned with the beat.

Results obtained for production tasks are given in figures 2.3 and 2.4. The spontaneous tapping rate was around 600 ms, confirmed results from previous litterature on the spontaneous motor tempo (see for example [McAuley *et al.* 2006]), while the variability of unpaced tapping did not vary as a function of condition. In the paced tapping task (figure 2.4, bottom panel), no differences were found according to the rate of stimulus, and participants were slightly more accurate on musical stimuli than on metronomes (difference found when averaging all metronome rates and musical stimuli).

Tests	Duration (min.)	Outcome measures
Perceptual tasks	60	
Duration discrimination	6	Duration discrimination threshold
Anisochrony detection with tones	20	Anisochrony detection threshold
Anisochrony detection with music	8	Anisochrony detection threshold
Beat Alignement Test (BAT)	15	Performance in detection of aligned beats
Production tasks	60	
Unpaced tapping	5	Spontaneous tapping rate and variability
Paced tapping with metronome and music	8	Synchronization accuracy and variability
Synchronization continuation	6	Tapping rate and variability in the continuation phase
Adaptive tapping	25	Measures of adaptation in tapping performance, and accuracy in the detection of tempo changes

Table 2.1: Overview of the BAASTA. Adapted from [Dalla Bella *et al.* 2017b]

In the synchronization-continuation task (figure 2.4, top panel), participants successfully continued tapping at the rate indicated in the synchronization task, while variability was larger for higher rates. Similarly, when subjects were presented with tempo changes in the adaptive task (figure 2.4, bottom panel), they were able to adapt to the new tempo similarly to faster and slower tempi. Adaptation indices (slope coefficient of a linear fit between tapped and stimulus tempo) were slightly above 1, indicating a tendency to overcorrect. Finally, when asked whether the stimulus was faster or slower, participants moslty detected the largest tempo changes (bottom panel, b).

Figure 2.5 shows significant Pearson correlations across the outcomes of several perceptual (detection thresholds for Duration Discrimination, hits minus false alarms for BAT averaged across tempi) and production (tapping accuracy, variability and adaptation indices, as well as detection of tempo change). From this correlation matrix, it appears clear that the outcomes of Duration Discrimination, and adaptation to slower or faster tempi are isolated from other outcomes, suggesting a distinct underlying mechanism. However, the variability of tapping was largely correlated across paced, unpaced and continuation tasks. Interestingly, some of the reported correlations suggest a link between perceptual abilities and tapping variability. First, the perceptual threshold for anisochrony detection with tones was positively correlated with tapping accuracy and variability. In addition, the ability to detect tempo changes in the adaptive task was negatively correlated with unpaced tapping variability, and with paced tapping accuracy and variability.

Taken together, these results reproduce classic results from the SMS litterature [Repp 2005, McAuley *et al.* 2006], and the variety of measures on the same subjects suggest enough sensitivity to detect interesting individual differences, as well as extreme cases of poor synchronizers

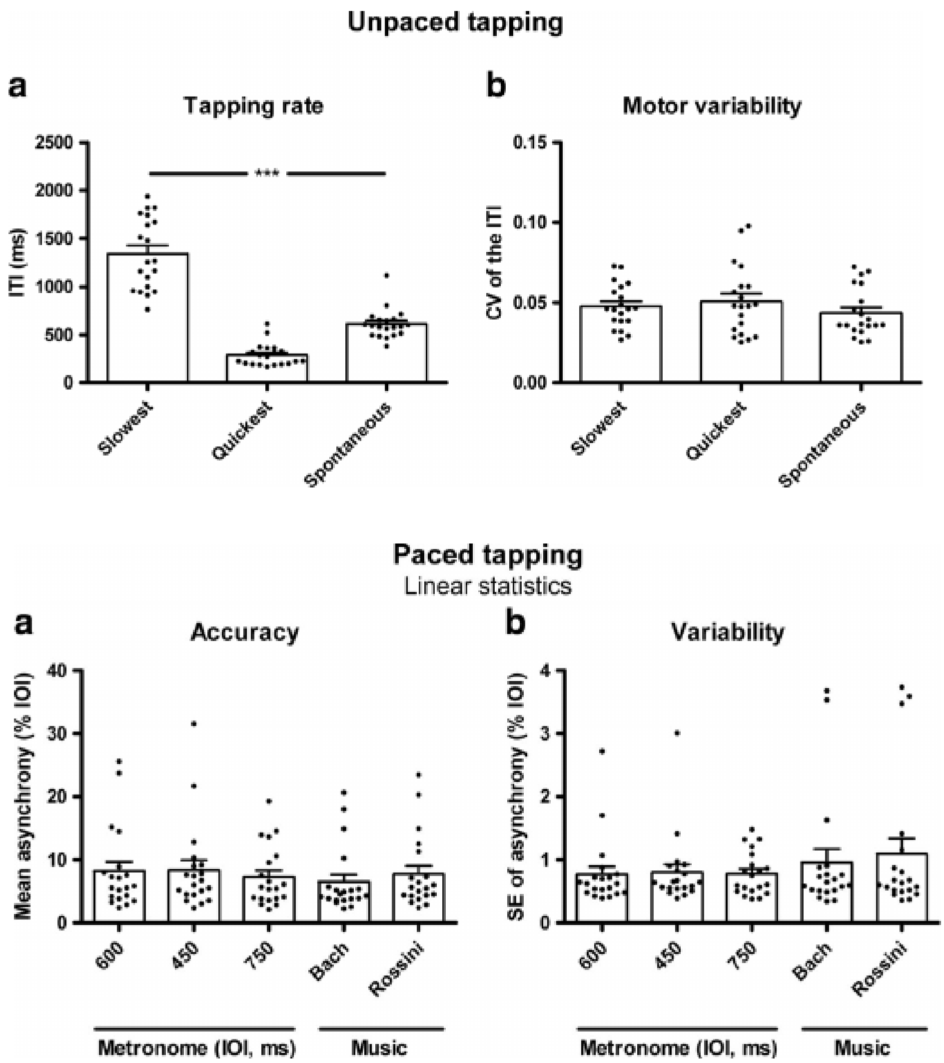


Figure 2.3: BAASTA Tapping results. Adapted from [Dalla Bella *et al.* 2017b]

/ beat deaf participants; more discussion on beat deafness can be found in the original paper [Dalla Bella *et al.* 2017b] as well as in subsequent studies such as [Bégel *et al.* 2017].

2.1.4 Music-based gait rehabilitation in Parkinson's Disease

2.1.4.1 Context

As reviewed previously in a meta-analysis of many studies [Lim *et al.* 2005], rehabilitation based on rhythmical auditory cues can improve gait in patients with PD. With the help of such a therapy, patients typically walk faster and exhibit greater stride length. However, this effect is highly variable among patients, with some exhibiting little or no response to the intervention. These individual differences may depend on patients' ability to synchronize their movements to a beat. I joined the European project EBRAMUS (Europe Brain and Music) as a postdoctoral fellow in 2010 and together with PhD student Charles-Etienne Benoit (WSFiz, Warsaw, Poland and Max Planck Institute for human cognitive and brain sciences, Leipzig, Germany), under the

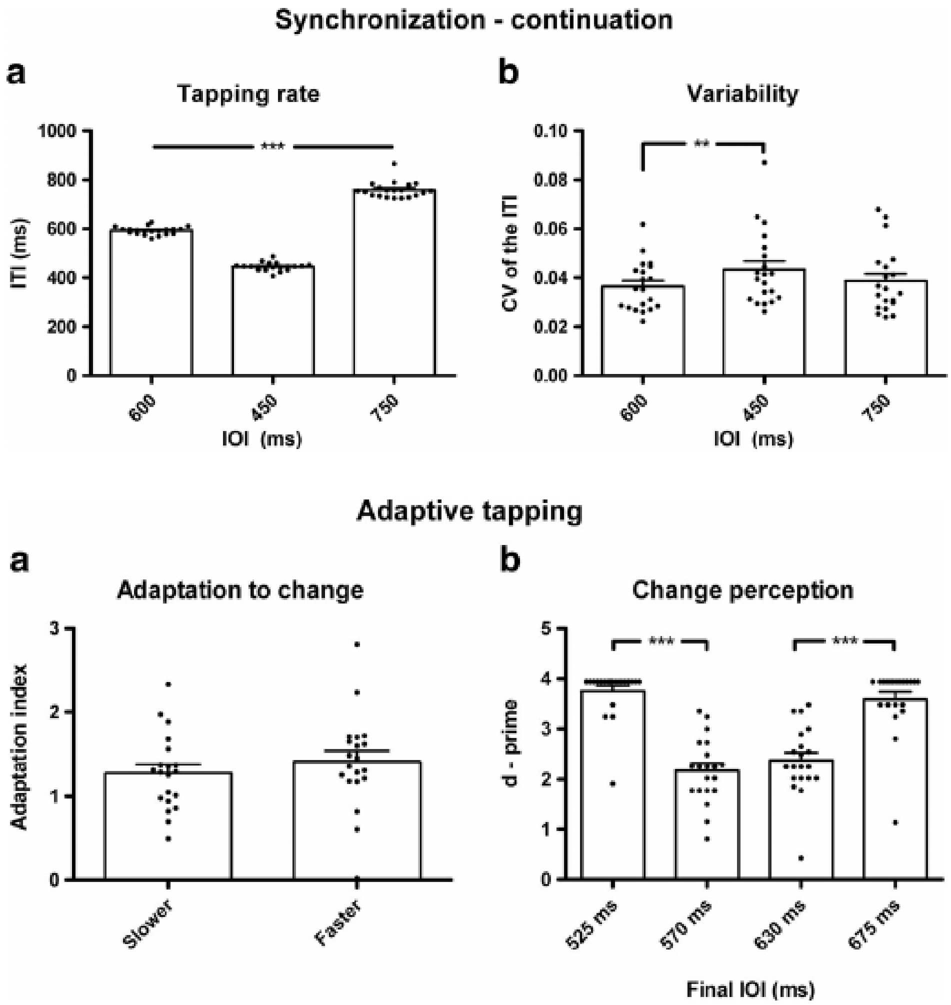


Figure 2.4: BAASTA Adaptive tasks results. Adapted from [Dalla Bella *et al.* 2017b]

supervision of Simone Dalla Bella and Sonja Kotz (Max Planck Institute for human cognitive and brain sciences, Leipzig, Germany), we implemented a project to test potential links between gait rehabilitation with music in PD and perceptual and sensorimotor timing skills. In particular, the BAASTA (see section 2.1.3.4) was designed to be used in this project.

2.1.4.2 Methods

Fourteen patients with PD were submitted to a Musically-Cued Gait Training (MCGT) program. Patients were all right-handed, non-demented, and showed moderate symptoms of PD, with an average Hoen and Yahr stage [Hoehn *et al.* 1998] of 2 (SD = 0.7; range = 0.5–3) and United Parkinson's Disease Rating Scale scores [Fahn *et al.* 1987] of 36.8 (SD = 19.1; range = 3–52). Other inclusion criteria were mild to moderate motor symptoms, normal hearing, no depression and no severe additional neurological or psychiatric illness. All participants had little musical training (below 1 year), except one patient with extensive amateur practice, but her performance in gait and behavioral tasks did not differ. Twenty age-matched non demented adults formed the control group (see [Dalla Bella *et al.* 2017a] for more details). PD patients were submitted

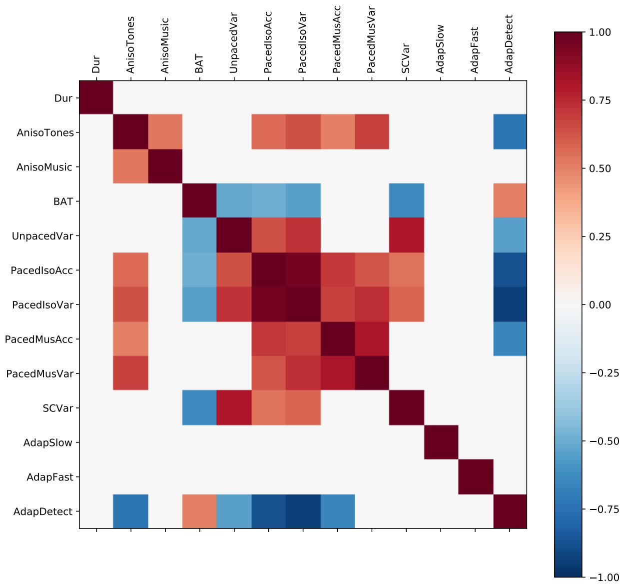


Figure 2.5: Significant Pearson correlations between outcomes from BAASTA tasks. Missing cells correspond to non significant correlation. Recreated using values from [Dalla Bella *et al.* 2017b]

to MCGT sessions three times a week for four weeks, in which they walked to music with an embedded metronome. Before (Pre), after the training (Post), and at a follow-up session one month after the end of training (Follow-up), patients' timing abilities were assessed using the BAASTA (see section 2.1.3.4).

During MCGT sessions, patients were asked to walk along with a familiar German folk song. The patients did not receive any explicit instructions to synchronize their footsteps to the beat of the music. The song was presented without the lyrics, and the beat of the song was emphasized with a superimposed salient high-pitch bell sound. Each training session lasted 30 minutes. Optimal beat frequency was determined at the beginning of the first testing session. The beat frequency (slower or faster), which led to the longest stride length was chosen and implemented in the MCGT for the entire duration of the training. Seven patients were trained with a beat frequency that was 10% faster than their preferred cadence, and the other seven patients were trained with a beat frequency that was 10% slower.

At each testing sessions (Pre, Post and Follow up), walking abilities were assessed using motion capture with passive reflective markers attached to the lower limbs, and patients were tested for both spontaneous walking (no cue), and walking with the same musical stimuli that are used during the MCGT sessions.

2.1.4.3 Results

First, motion capture revelead that patients with PD had smaller steps while walking at comfortable speed in the absence of an external cue (lower stride length as compared to controls).

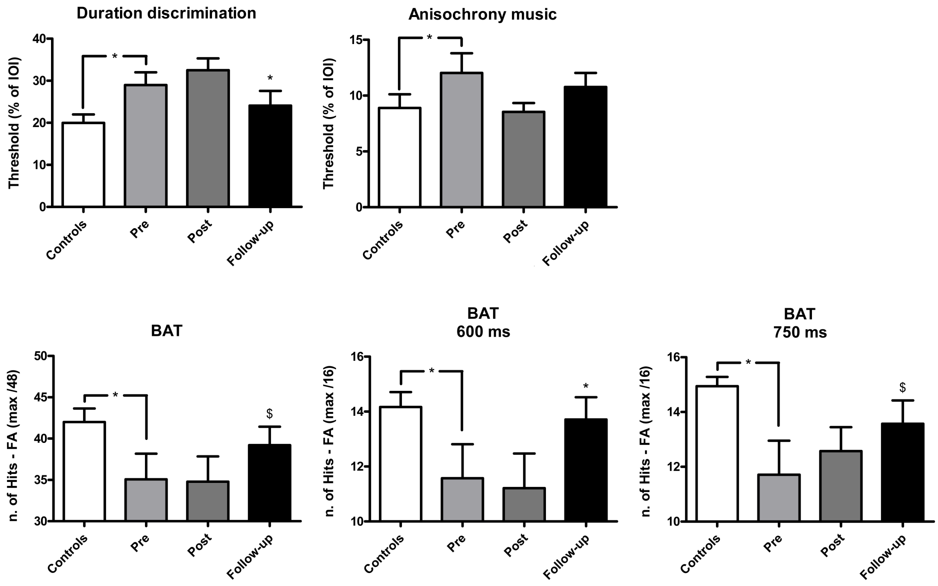


Figure 2.6: Behavioral test results of BAASTA with PD patients and the effect of auditory cueing rehabilitation. Tasks where patients differed from controls before the cueing training are selectively reported. Error bars indicate the standard error of the mean (SEM). Note: $*p < 0.05$; \$ marginally significant difference. Adapted from [Benoit *et al.* 2014]

MCGT increased stride length of patients on average, and this benefit was maintained at follow-up.

Figures 2.6 and 2.7 present the outcomes of BAASTA tasks on PD patients in the three testing sessions as a result of MCGT. Patients showed higher thresholds than controls in the duration discrimination task before the training, and improved following the training, an effect that was visible only at follow-up. Patients also displayed worse detection of anisochronies in musical stimuli than controls; yet, training did not improve their performance in this task. Finally, in the BAT, patients had more difficulties in detecting misaligned beats before intervention when compared to controls. This difference was present for slower tempi only, and the detection of misaligned beats generally improved when tested at follow-up.

In the unpaced tapping tasks patients did not differ from controls before the training, in terms of accuracy and variability. However, patients tended to be less accurate than controls when synchronizing with an isochronous sequence. The effect of the training was mostly visible in the follow-up session. Training led to increased synchronization accuracy with the isochronous sequences at the fastest tempo. Patients' performance in the follow-up evaluation did not statistically differ from that of controls.

In the synchronization–continuation task, patients tested prior to training were less accurate than controls only at the fastest tempo. However, the two groups did not differ in terms of variability across all tempi. Training had no effect on this task. Similar results were obtained in the adaptive tapping task. Moreover, patients performed worse in detecting tempo changes at 600 ms. Both groups displayed similar tapping variability, and comparable adaptation indexes. Training was effective only in improving the detection of tempo changes, an effect visible when comparing pre-intervention to follow-up. Patients' perception in the follow-up session did not significantly differ from the performance of healthy controls. More details,

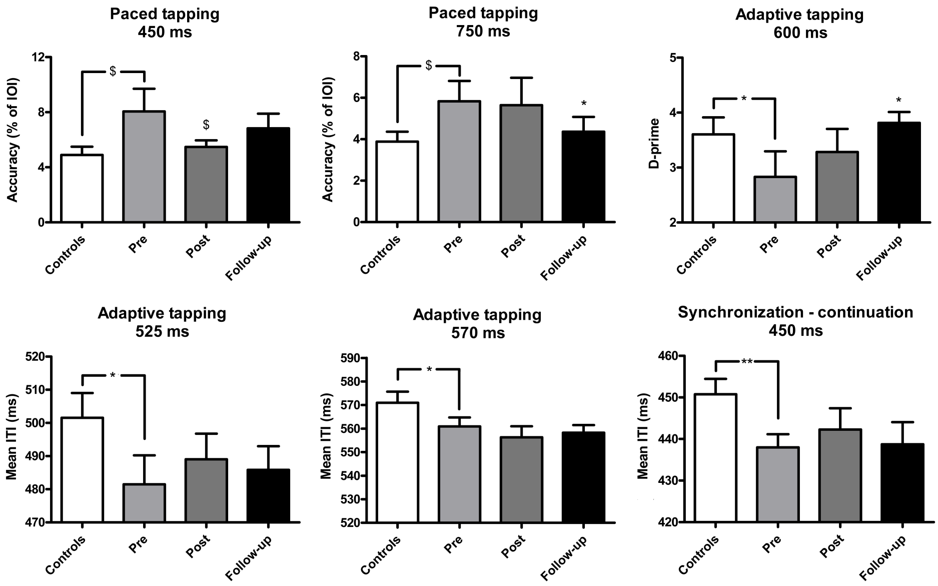


Figure 2.7: Motor tests results of BAASTA with PD patients and th effect of auditory cueing rehabilitation. Tasks where patients differed from controls before the cueing training are selectively reported. Error bars indicate the standard error of the mean (SEM). *p $<$ 0.05; **p $<$ 0.01; $ marginally significant difference. Adapted from [Benoit *et al.* 2014]

including an anlysis of individual differences in patients' profiles, can be found in the related publication [Benoit *et al.* 2014].

On the other hand, using motion capture analysis techniques, we were able to highlight links between the kinematics of the patients' walking, the synchronization of their steps with the music, and the effect of the rehabilitation [Dalla Bella *et al.* 2017a]. Patients increased gait speed and stride length in non-cued gait after training. However, individual differences were apparent as some patients showed a positive response to MCGT and others, either no response, or a negative response. These individual differences were modeled by entering patients' performance on various tapping and gait sensorimotor tasks into a logistic regression model aimed at predicting the probability of a positive response to MCGT. The model shows that patients who are most impaired in terms of their gait (i.e., those exhibiting the slowest gait speed), and who show the poorest synchronization accuracy are likely to benefit from RAS. This finding is not totally unexpected. Indeed, the most impaired patients are those who are likely to show the greatest improvement relative to patients who show less impaired gait (i.e., the latter may already perform at ceiling at the pre-test) or better synchronization accuracy. In addition, patients who are the least variable in a synchronized tapping task and the most responsive in adapting their movement timing (tapping) to a change in a pacing sequence (i.e., an acceleration) are those who benefit the most from RAS. An increased tendency to correct movement timing in responseto stimulus acceleration (adaptation index $>$1, overcorrection), accompanied by greater phase correction, favor a positive response to RAS. Thus patients' flexibility in adapting their walking behavior to a beat frequency departing from their natural cadence may contribute to the success of RAS. More details, including statistical results and modeling of individual differences using logistic regression can be found in the related publication [Dalla Bella *et al.* 2017a].

2.1.4.4 Discussion

The main goal of the current study was to examine the effects of a 1-month MCGT program. While we were able to replicate the general efficacy of the program, by showing improvements in gait parameters (longer step and stride length, faster cadence) on average in patients, our goal was to understand this type of rehabilitation, and individual differences, using two different point of views. First, we looked at the effect of MCGT on perceptual and motor timing abilities in PD patients (reported in [Benoit *et al.* 2014]). Performance was assessed with the BAASTA battery. Prior to the intervention, patients exhibited impaired perceptual timing across all BAASTA tasks except for anisochrony detection in isochronous sequences. On the contrary, motor timing was relatively spared, except lower accuracy in continuing tapping at a given rate, and in tapping along with an isochronous sequence. These findings are in line with previous evidence that PD is associated with a malfunctioning timing system.

Next, we investigated individual differences in patients' response to MCGT by relating their gait patterns, and more specifically the way patients synchronized their steps to music, with the amout of improvement after therapy. A positive response to MCGT was predicted by the synchronization performance in hand tapping and gait synchronization, using logistic regression modeling. More severe gait impairment, low synchronization variability, and a prompt response to a stimulation change foster a positive response to MCGT training. Thus, sensorimotor timing skills underpinning the synchronization of steps to an auditory cue may allow predicting the success of MCGT in PD.

What does such a pattern of results tell us on the brain basis of music-based rehabilitation of PD ? Benefits of auditory cueing on gait kinematics are likely to be mediated by a cerebello- thalamo-cortical network, which is also involved in timing (for a reviewa, see [Kotz & Schwartze 2010]). More specifically, projections from the supplementary motor area to the primary motor cortex may support motor output and modulate or stabilize gait kinematics over time. Compensation of a dysfunctional basal ganglia timing system via rhythmic auditory cues may be afforded by this compensatory cerebello-thalamo-cortical network. For example, evidence of hypermetabolism in the cerebellum of PD patients, as a result of cueing training (del Olmo et al., 2006), provides preliminary support for this hypothesis. This circuitry plays a key role in domain-general timing, and may underlie perceptual timing and coupling movement to an external pacing stimulus (Wing, 2002). Functional and/or structural changes in this compensatory network due to auditory cueing may affect both gait kinematics as well as perceptual and motor timing.

What mechanisms are responsible for these positive effects of MCGT as a function of individual differences in sensorimotor timing skills? In general, the core process underlying positive effects of MCGT may be the stimulus-driven allocation of attention to relevant aspects of gait kinematics, which augments temporal prediction, and thereby facilitates movement planning and initiation. Patients showing somehow spared sensorimotor timing skills, namely those who are quite consistent in a sensorimotor task (e.g., showing low variability in tapping to a beat, in spite of being inaccurate), and who can promptly react to a change in stimulation may benefit from the presentation of an external rhythmic cue. As a consequence, relatively spared extraction of a beat from the acoustic signal in some of the patients may create optimal conditions for the success of rehabilitation. MCGT could thus provide a regular temporal scaffolding for movement coordination, and facilitate the compensation of patients' impaired internal timing. In the next section, we are interested in uncovering the neural correlates of such internal timing processes, by tapping into the temporal processing system using EEG.

2.1.5 Cognitive neurosciences of temporal processing

2.1.5.1 Oddball paradigm and temporal processing

During my postdoctoral work at the Max Planck Institute in Leipzig in the team of Sonja Kotz, I collaborated with another postdoctoral fellow, Michael Schwartze, on the analysis of Electroencephalography (EEG) data in relation to temporal prediction phenomena in the brain. An oddball paradigm with manipulations of temporal regularity was proposed by Michael Schwartze in [Schwartze *et al.* 2011b], which reported effects of temporal regularity on late ERP such as the P300. This data was reanalyzed for ERP in the 50 to 100 ms range, in a study on which I am coauthor [Schwartze *et al.* 2013], and our results suggest that two early ERPs, namely the P50 and N100, have larger amplitudes when averaged over less predictable stimuli, i.e. in the irregular condition compared to regular, and on deviants compared to standard. This result suggests that brain responses are adapted according to predictability of the incoming stimuli, and these effects were stronger with temporal predictability. We argue that such an effect may hint at an underlying mechanism that optimizes the processing of environmental stimuli according to their occurence predictability.

2.1.5.2 Average ERP response of PD patients to deviance detection during temporally regular and irregular contexts

As reviewed in section 2.1.4, pathological populations such as PD are known for deficits in temporal processing and SMS. Therefore, we hypothesized that the putative predictive mechanism for processing temporal regularity is affected in PD, and that an effect of such disruption could be observed in the EEG, in the ERP as well as in background brain oscillations. We also hypothesized that the effect of auditory cueing therapy might be linked to the same underlying brain mechanism.

To test this hypothesis, we adapted the oddball paradigm from [Schwartze *et al.* 2013] with a longer interstimulus interval for the study on patients with PD, in order to analyse effect of background brain oscillations. The adapted paradigm is shown in figure 2.8-A. In the regular condition, pure tones with a duration of 300 ms are presented with a fixed inter-stimulus onset interval of 600 ms, while this interval is randomized between 200 and 1000 ms in the irregular condition. In figure 2.8-B, we show the effect of experimental manipulations with nineteen patients with PD and twenty age-matched healthy controls. The latter subjects had a similar pattern than young healthy adults from the previous study [Schwartze *et al.* 2013], with a larger amplitude for deviants than standard in both the irregular and regular condition, and more importantly, overall larger amplitudes for the irregular condition compared to the regular one. However, patients with PD showed a large amplitude of deviants compared to standards, irrespective of temporal regularity. These results suggest that the mechanism that would modulate brain responses may be impaired in PD, possibly due to basal ganglia disfunction.

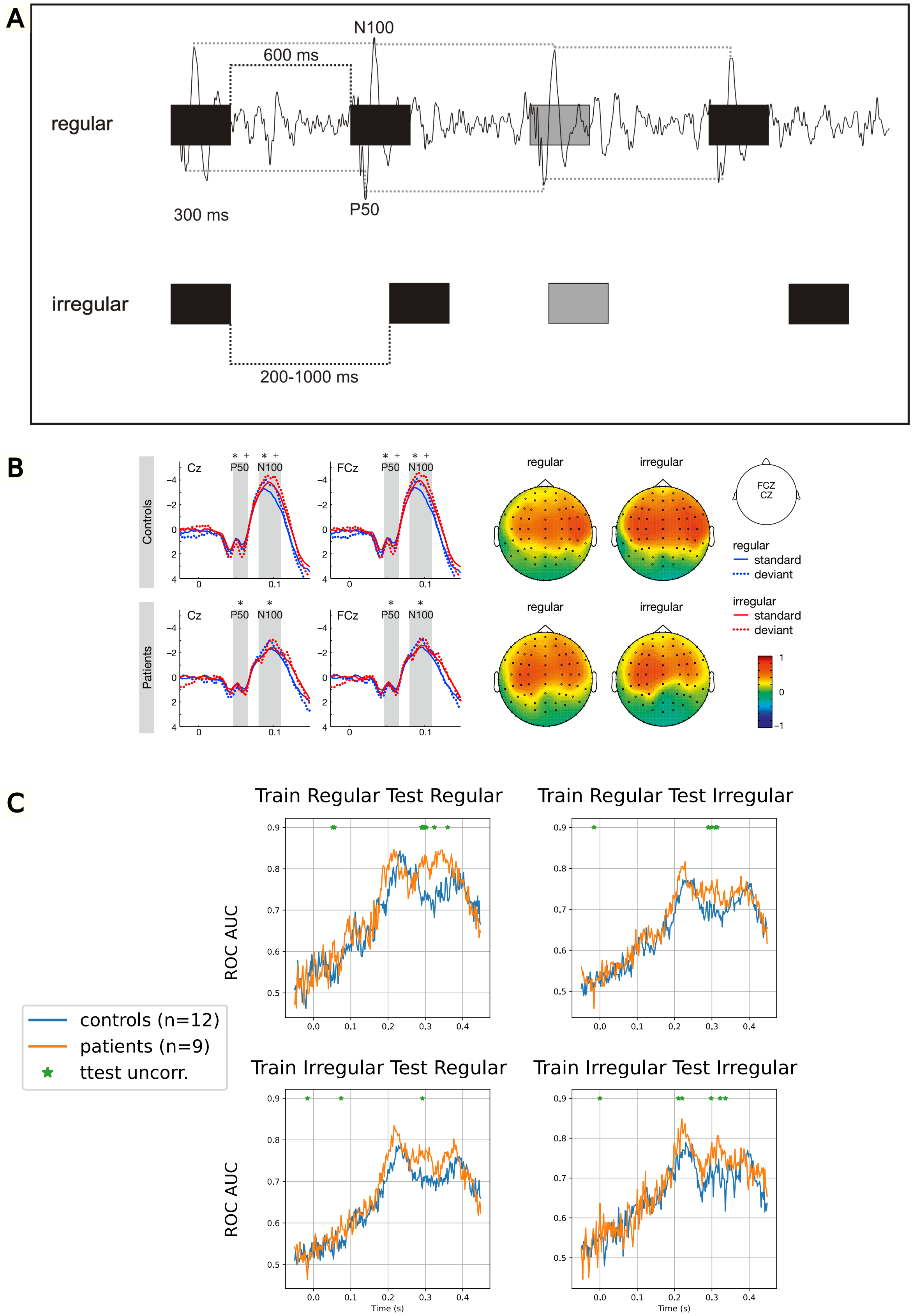


Figure 2.8: Neural correlates of temporal regularity in oddball detection in PD patients. Panel A : Oddball paradigm with manipulation of temporal regularity. Adapted from [Schwartze *et al.* 2013]. Panel B: Effect of temporal regularity in healthy controls and patients with PD (unpublished data). Panel C : Decoding single trials and generalization to different condition of temporal regularity, with subset of subjects with High signal to noise ratio (Unpublished data).

2.1.5.3 Exploratory analysis of individual differences using cross-condition decoding models

We intented to study the link between ongoing brain oscillations and subsequent ERPs. This type of analysis showed some potential with a few subjects, but with a large amount of individual differences ; some patients (as well as some controls) did not show any clear pattern, or incoherent patterns, with opposite effects, were found across subjects. Therefore, we decided to move on to individual subjects analysis in order to uncover potential sources of individual differences.

Single trial, single subject decoding models of deviance detection were defined, and tested for cross-decoding between Regular and Irregular conditions. The motivation behind such models is to investigate whether standards and deviants can be decoded from the scalp EEG at a specific time step, and to test whether the trained model can generalize across temporal regularity conditions. Failure in such generalization at the same time step would indicate that signals of deviance detection are of different scalp topographies (amplitude and location) between the regular and irregular conditions. In that respect, we hypothesize that deviance decoding would fail to generalize in controls, indicating that amplitudes and locations of EEG single trial responses to deviance detection are specific to the temporal regularity condition (which is what the average ERP study suggest, figure 2.8-B). On the contrary, if cross-decoding models generalize across temporal regularity and irregularity in PD patients, this would indicate that patients electrophysiological responses are "immune" to manipulations of temporal regularity.

Our decoding models employ supervised learning (see section 2.2), and are trained to classify standard and deviant tones, based on scalp topographies at a specific time point. Models are subsequently tested for generalization on other trials at the same time point using cross validation (five folds). Two metrics are analyzed to assess model performance :

- Probabilities for each class are computed, for each trial, enabling statistical testing of the overall confidence of the model at each time point, using Mann-Whitney U-tests corrected for multiple comparisons using False Discovery Rate [Benjamini & Hochberg 1995],
- As the ratio between examples of standard and deviant tones is imbalanced, we use a detection metric to measure the performance of trained model : the Receiver operator Characteristic, Area Under the Curve (ROC-AUC). A ROC-AUC of 0.5 means chance level, while a value of 1.0 is indicative of perfect detection of all deviants, and no confusion with standards.

Out of twenty control subjects and nineteen patients, decoded models yielded probabilities that were above than chance for only twelve controls and nine patients (FDR corrected); we will refer to this subset of subjects as having a high signal to noise ratio (SNR). The decoding models for the other subjects could not reliably distinguish between standard and deviants, in any condition. We present in figure 2.8-C average ROC-AUC values for high SNR subjects as a function of time, with cross-condition decoding ; training in either the regular or irregular condition, and testing (=generalization) for regular or irregular. Average ROC-AUC peaked at values of 0.85 around 200 to 250ms, corresponding to the time window of the N200 ERP. We tested for group differences in ROC-AUC in the four cross-decoding scenarios, and show uncorrected significance (green stars) as an exploratory analysis. The pattern of results suggest that patients may have more cross-condition generalization (Train Regular, Test Irregular and Train Irregular, Test Regular) than controls around 300 ms, as well as within-condition generalization.

While it is difficult to conclude with such as small sample of patients, the current results suggest that the brain responses of patients can be decoded similarly around 300ms irrespective of differences in temporal regularity of the stimulus, while it was not the case for controls. Future studies should attempt at replicating such cross-condition decoding with a large number of subjects, in order to better understand such individual differences. If confirmed at a larger scale, this result could be interpreted as follows : brain dysfunctions linked with PD are associated with an indifferent processing of temporal regularity, possibly linked to deficiencies found in behavioral and sensorimotor timing tasks (see section 2.1.4).

2.1.6 Musical Imagery

Musical mental imagery consists in "imagining" hearing music without any acoustic vibrations reaching the eardrums. Thus, involuntary musical imagery (INMI) is the cognitive phenomenon of having music in your head without conscious control. INMIs are very common; it is estimated that more than 90% of the population experiences an INMI more than once a week, and more than 30% experience them at least once a day, according to a meta-analysis of cognitive psychology research on this topic co-authored by my former collaborator at Goldsmiths Kelly Jakubowski (now at University of Durham) [Liikkanen & Jakubowski 2020]. The work presented in this section was done in the Music Mind and Brain group in the Department of Psychology in Goldsmiths, University of London, in which I was a postdoctoral fellow between 2013 and 2015. Some of the work was done in collaboration with Jonathan Smallwood, who was a Reader at the university of York at that time, as well as with the group of Daniel Margulies at the Max Planck Institute in Leipzig.

2.1.6.1 Tapping to voluntary and involuntary musical imagery

I contributed to several studies conducted by Kelly Jakubowski, then a PhD student under the supervision of Lauren Stewart (Goldsmiths, University of London). First, we developed a method to estimate the tempo of an INMI, by providing accelerometer wristwatches to participants, who kept them on for several days in order to "capture" INMI in daily life [Jakubowski *et al.* 2014, Jakubowski *et al.* 2015]. Participants were instructed to synchronize with their INMI by tapping the tempo of the music currently being imagined, by tapping with their finger on the wristwatch. We therefore analyzed the INMI episodes using methods I developed previously for the study of SMS (see section 2.1.3).

Figure 2.9-A shows the accelerometer data (averaged in the three X,Y and Z directions) for an INMI episode. A peak detection algorithm was applied to detect maximum acceleration peaks. Detected peak onsets clearly suggest a regularly tapped sequence. Panel B shows the Inter Tap Interval (ITI), corresponding to the derivative of the time series of detected peaks of acceleration. This ITI series fluctuates around 978 ms in this case. Panel C is a zoomed in version of the accelerometer data, along with the detected peaks. The two bottom histograms show the distributions of the tapped tempo of all captured INMI, as well as the ratio between the tapped and the veridical tempo of the original song, in cases we were able to match it with participants records. A ratio of 1 indicates that the INMI was tapped at a tempo that is close to the veridical tempo, and higher (resp. lower) values correspond to slower tapped tempi (resp. faster tempi). Most interestingly, we were also able to show that the tempo of the INMIs was correlated with self-reported emotional states : INMI reported during a high arousal

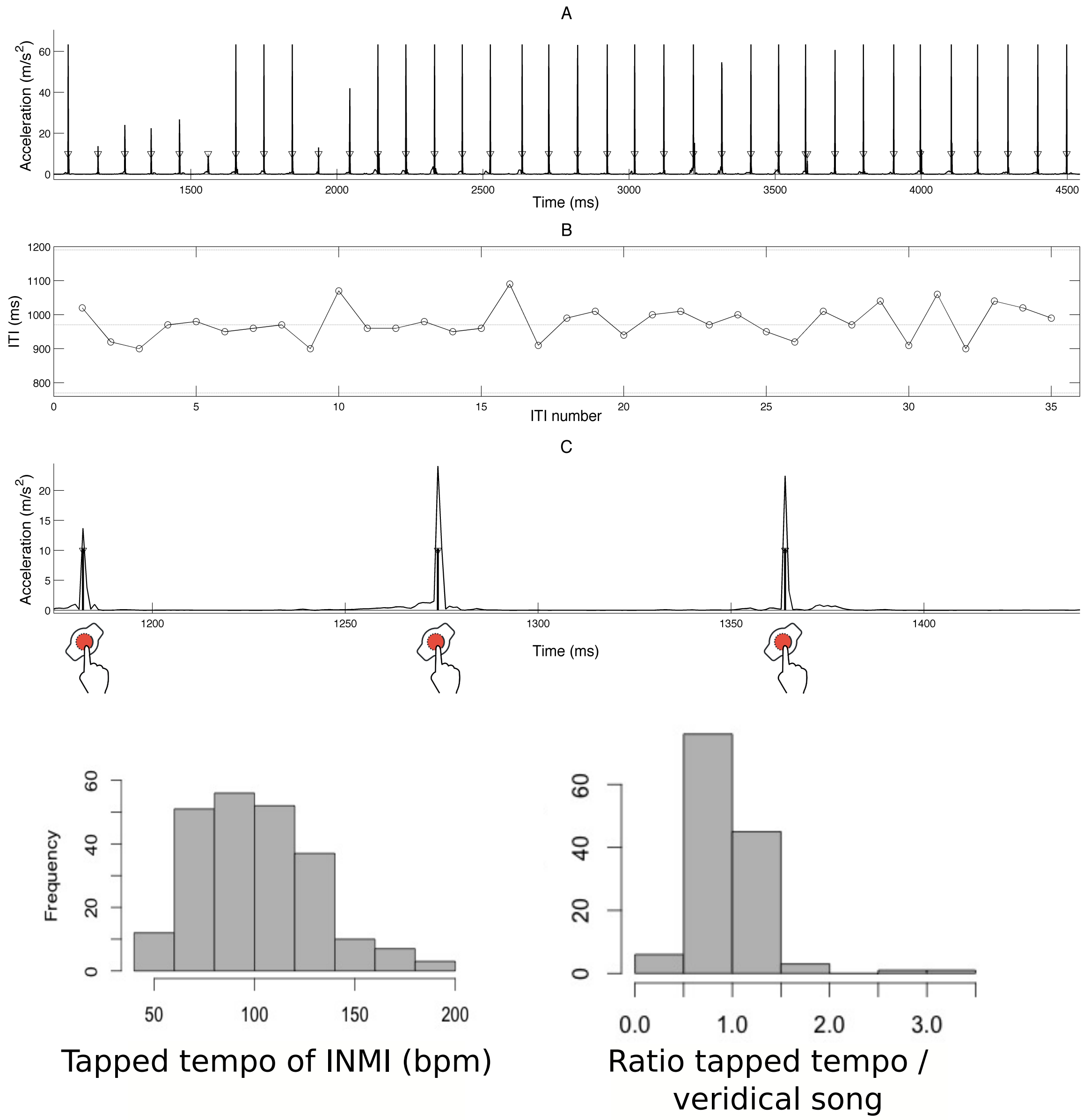


Figure 2.9: Capturing INMI tempo in everyday life. Adapted from [Jakubowski *et al.* 2015]

state (according to a short self-report emotional state scale filled by subjects at each episode) was positively correlated with the tempo of INMI [Jakubowski *et al.* 2015]. Said otherwise, INMI tempo were faster while people were in a more "aroused" (i.e. more excited) state. This interesting finding suggests that imagined mental content such as INMI might be related to emotional state, and to the sensorimotor coupling system. This results suggest new hypothesis regarding the putative neural basis of INMI, which could include brain networks related to emotions, as well as the sensorimotor system, in addition to auditory areas of the brain.

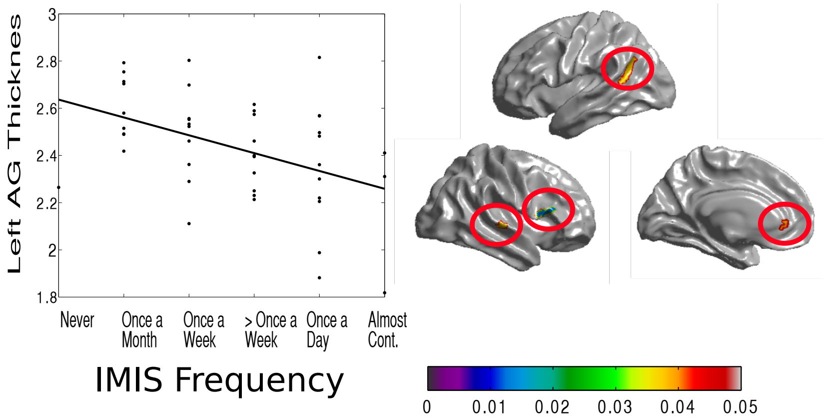
Additionally, I also contributed to two other studies : one on the tempo of voluntary mental imagery [Jakubowski *et al.* 2016], assessed using tapping, and the other one on the link between voluntary and involuntary recall of musical memories [Jakubowski *et al.* 2018], using the same acceleromter wristwatches than in the first study.

2.1.6.2 Neural correlates of individual differences in involuntary musical imagery

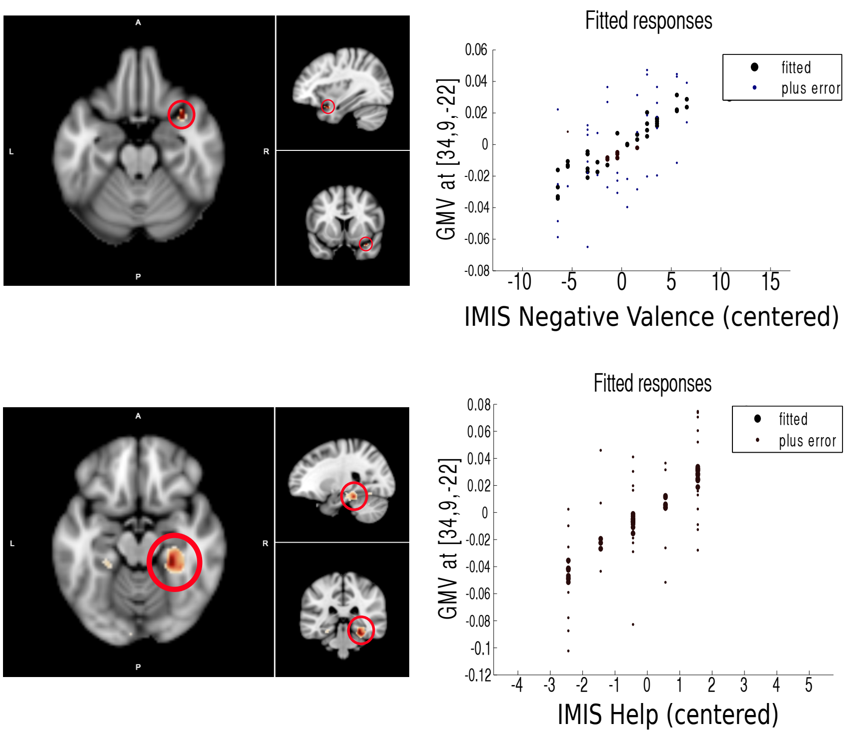
In order to investigate the neural correlates of the experience of INMI, we used a retrospective self-report scale to measure individual differences, the Involuntary Musical Imagery Scale (IMIS) [Floridou *et al.* 2015], developed in the Music Mind and Brain group by another PhD student, Georgia Floridou. The IMIS is composed of a set of questions to test for everyday experience of INMI. The IMIS consists of four factors that assess (1) the extent to which people negatively evaluate the INMI experience (Negative Valence factor), (2) the extent to which people move in time to their INMI (Movement factor), (3) the degree to which people are helped in their everyday activities by INMI (Help factor), and (4) the extent to which INMI reflects the content of personal matters, worries, or concerns (Personal Reflections factor). The IMIS also includes three additional questions regarding the overall frequency of INMI occurrence, the typical length of the section of music that is experienced as INMI (subsequently referred to as INMI section length), and the average length of an INMI episode (subsequently referred to as INMI episode length).

I led a first study [Farrugia *et al.* 2015b] that investigated links between brain morphometry (see figure 2.2) and individual differences in the IMIS. This study was a collaboration between the Music Mind and Brain group, and the Cognitive and Brain sciences Unit in Cambridge, with Rhodri Cusack. We analyzed structural brain data collected on fourty-four subjects, and we sent a questionnaire to subjects to measure individual differences in IMIS. Figure 2.10 shows some of the main results we obtained. First (top panel), the self-reported frequency of INMI episodes was related to cortical thickness of the right superior temporal gyrus, left angular gyrus, right inferior frontal gyrus and left ventral anterior cingulate cortex. Next, we found that two IMIS factors related to emotional responses to INMI were related to local gray matter volume in two ventral brain areas; increased values of the "Help" factor was linked to increased gray matter volume in the right parahippocampal cortex (PHC). In addition, the negative evaluation of INMI ("Negative Valence" factor) was linked with increased gray matter volume in the right temporal pole.

We interpreted these results in light of previous litterature in music neurosciences, as well as previous work on internal, self-generated thoughts (also referred to as "mind wandering"). Corresponding references are included in the discussion of our manuscript, which we adapted here [Farrugia *et al.* 2015b]. The right superior temporal gyrus has been strongly implicated both in auditory perception and voluntary musical imagery, while the right inferior frontal gyrus is held to have a role in pitch memory and is activated in both auditory perception and



(a)



(b)

Figure 2.10: Brain structure and involuntary musical imagery (INMI). (a) Cortical Thickness and Frequency of INMI. The left part shows the model fitting for the left angular gyrus (AG) thickness as a function of INMI Frequency. We found that INMI Frequency was also related to cortical thickness in the right auditory cortex, the anterior cingulate cortex and the right inferior frontal gyrus. (b) Gray Matter Volume (GMV) in the right temporal pole was linked to negative valence of INMI, and right PHC GMV was increasing with subjects that reported a beneficial effect of INMI (Help factor). Adapted from [Farrugia *et al.* 2015b]

imagery. Thus, our results suggest that the involvement of fronto-temporal areas also extends to non-volitional forms of musical imagery. In addition, the inferior frontal gyrus may also play a role in suppressing unwanted INMI episodes due to its role in inhibitory mechanisms.

Going beyond an interpretation only relating to auditory / musical related brain mechanisms, we found areas of covariation with aspects of the INMI experience that parallels with regions implicated in studies of self-generated thoughts (SGT). Specifically, we found a relationship between the frequency of INMI episodes and cortical thickness of a cluster in the left Angular Gyrus, notably, the ventral part, which is functionally connected to the PHC and medial prefrontal cortex, and lies within the default mode network, a set of regions known to be actived at rest during mind-wandering and self-generated thoughts.

We also found that local gray matter volume in regions of the right medial temporal lobe, namely the right PHC region (comprising the parahippocampal area and the temporal pole), correlated with factors of the IMIS representing both the negative evaluation of INMI and the extent to which participants find INMI episodes helpful. Specifically, gray matter volume in right temporal pole was greater for people who reported higher disturbance by their INMI episodes. The temporal pole has been suggested as part of a network for affective processing, and projects white matter tracts to the Orbito Frontal Gyrus and other emotion-related areas. As a consequence, a link between the structure of the right temporal pole and the disturbance of INMI episodes is in line with the fact that INMI might share neural resources with other forms of SGT. These results suggest that the tendency to be disturbed by INMI may be linked to the inability to inhibit negative emotions associated with the experience of INMI.

Finally, the IMIS Help factor measures how much participants feel their INMI episodes help them to get things done and to focus on their current task [Floridou *et al.* 2015]. Participants scoring highly on the IMIS Help factor had greater gray matter volume in the right PHC. Both regions are important for memory retrieval, and recent fMRI studies have reported that the PHC is activated in both the voluntary and involuntary recall of episodic memories. There is strong evidence from the fMRI literature highlighting an important role of the PHC in music-evoked emotions, which has been interpreted in the more general context of emotional processing and social functions of music. In particular, PHC volume is increased in individuals with a higher tendency to experience positive emotions related to music listening, which is congruent with our findings of increased GMV in the PHC in subjects with high scores on the Help factor.

In a second study in collaboration with Jonathan Smallwood in York [Farrugia *et al.* 2015a], we measured individual differences in INMI scores in relationship with resting-state functional connectivity. 40 subjects aged 18 to 31 years old (23 +/- 3.2 years) were recruited for this study. We examined connectivity between auditory areas, the right PHC, and the rest of the brain, by defining four seed regions : one seed in bilateral primary auditory auditory cortex (Heschl's gyrus), a seed in bilateral non-primary auditory cortex (npAC), a seed in bilateral middle temporal gyrus (MTG), and finally a seed in the right PHC taken from the peak coordinates of our previous study [Farrugia *et al.* 2015b].

Results are shown in figure 2.11. Two clusters showed significant negative correlation of the activity in npAC with respectively the precuneus and the anteromedial PFC (amPFC), modulated by the frequency of INMI episodes (Figure 2.11-A). Conversely, the opposite pattern was observed with the functional connectivity between npAC and amPFC, with high INMI frequencies related to negative correlation, while subjects with infrequent INMI had slightly positive correlation (Figure 2-A). Furthermore, more frequent INMI were related with decreasing correlation between the activity in the MTG and a cluster in the posterior cingulate cortex (PCC),

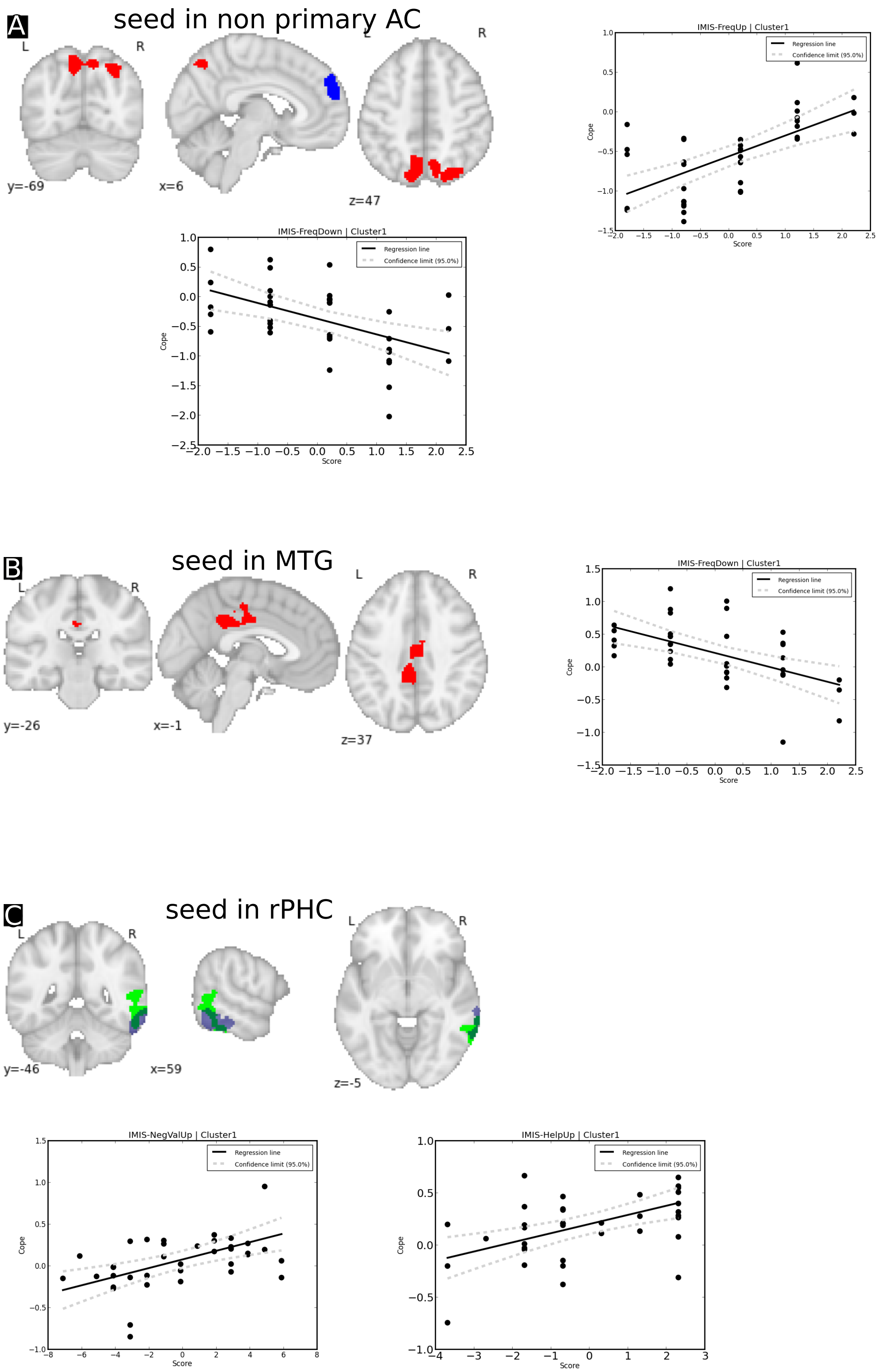


Figure 2.11: Resting-state functional connectivity and involuntary musical imagery (INMI). Panel A - Connectivity with the non primary auditory cortex was related with the frequency of INMI, in the precuneus and anteromedial prefrontal cortex. In panel B, we report a link between INMI frequency and the connectivity between the middle temporal gyrus and the posterior cingulate cortex. Finally, panel C shows that affective aspects of earworms were related with connectivity between the right PHC and the lateral temporal cortex.

extended to the supplementary motor area (Figure 2.11-B). Affective aspects of INMI were linked with the connectivity between the right PHC and the lateral temporal cortex (Figure 2.11-C). Specifically, subjects with high scores in the Help factor had higher connectivity with the anterior inferior temporal gyrus, while high scores in Negative Valence were related to higher connectivity with posterior parts of the inferior temporal gyrus, extending to the posterior MTG. These two clusters overlapped in the posterior inferior temporal gyrus.

Overall, we found that INMI frequency was related to functional connectivity between auditory areas on one side, and medial parietal, posterior cingulate and medial frontal areas on the other side, while affective aspects were linked with the connectivity between the right PHC and the lateral temporal cortex. These results are overall consistent with our previous structural MRI data that showed links between INMI facets and brain regions similar to the present study [Farrugia *et al.* 2015b].

INMI frequency was modulated by connectivity seeded in non-primary auditory areas (Figure 2.11-A and B). This confirms the role of connectivity with auditory areas, and is in line with previous literature on musical imagery that consistently reported a role of non-primary auditory areas. We also found that intrinsic connectivity between the right PHC and the lateral temporal cortex modulated the affective evaluation of INMI episodes (Figure 2.11-C). A pattern of co-activation between the lateral temporal cortex and the right PHC is similar to networks involved in involuntary episodic memory retrieval, as well as semantic retrieval. Similarly, functional connectivity between networks including the right PHC and lateral temporal cortex has previously been shown to underlie aspects of SGT such as autobiographical planning, mental simulation of future events or, more crucially regarding our results, affective evaluation and self-regulation of emotions. Memory associations is a commonly reported trigger of the INMI experience, and the nature of the association might influence the subsequent pleasantness of the INMI episode. In addition, music-evoked autobiographical memories have previously been associated with activations in the rPHC and this region is consistently recruited by music-evoked emotions, while its structure was linked to positive outcomes of INMI [Farrugia *et al.* 2015b]. As a consequence, increased PHC-lateral temporal cortex connectivity might be linked with a tendency to attach stronger affective value to INMI episodes.

Finally, taking together the results from both our studies suggest that the Default Mode Network might be a central integrative hub that serves a critical role in the INMI experiential state, as both the structure of areas of the Default Mode Network, and their intrinsic connectivity, influenced the frequency and the evaluation of INMI. Such an integrative role of the Default Mode Network has been proposed for more general cases of SGT. This may have important implication for the understanding of the idiosyncrasy of the INMI experience, as for instance subjects typically don't experience the same songs as INMI. Unique self-related processes originating in the Default Mode Network might either amplify or dampen a range of self-related thoughts such as involuntary memories or salient internal thoughts and emotions, resulting in varying occurrence and evaluation of INMI across subjects.

In order to replicate these results on a larger sample, we collaborated with Daniel Margulies, at that time groupe leader in the Max Planck Institute in Leipzig, to incorporate the IMIS in a large phenotyping dataset (194 subjects) using functional MRI, which resulted in a "Data Paper" of which I am co-author [Mendes *et al.* 2019]. The analysis of the IMIS data in relationship with the resting-state and structural data is still in progress.

2.2 Machine learning and Graph Signal Processing

2.2.1 Overview

This section describes my methodological contributions in Machine Learning and Graph Signal Processing, in three steps. First, I contributed to a research effort at IMT Atlantique on efficient deep learning (i.e., learning and using deep learning with less computational ressources, see section 2.2.3), related to previous work from my PhD as well as to my initial training as an electrical engineer. Second, I initiated a new research effort on the analysis of resting fMRI data using machine learning, described in section 2.2.4, which also contributed to cognitive neurosciences by shedding new light on impulsivity and addiction disorders, as well as musical training. Finally, I developed a line of work on the use of graph signal processing to analyze brain activity, by combining spectral graph extractors and machine learning 2.2.5.

2.2.2 Methods

2.2.2.1 Machine Learning and Deep Learning

We give here informal definitions of the concepts that will be used in the subsequent sections. Machine Learning (ML) refers to a set of algorithms that use data to solve a problem. Algorithms based on ML are sometimes qualified as data-driven, and ML is considered as a part of the larger fields of Artificial Intelligence and Data Science [1]. A ML algorithm uses data during a training set in order to tune its internal parameters. Once trained, the algorithm is tested for its generalization, by assessing a performance metric on new, unseen data that was not part of the training phase, called the test set. As the data considered is most commonly of a very high dimensionality (e.g., each sample of the training set is a parcellated brain image with 360 ROIs), ML algorithms typically constrain the search space of their solution by exploiting various forms of priors on the data, or on the form that must take the solution. The data on which ML algorithms are trained typically contains a few hundred to several / many thousands of examples, each example being a multidimensional vector whose number of dimensions is also called number of features.

By Feature Extraction, we refer to a function that uses "raw", input vectors (such as an audio waveform) with a large number of features, and uses a predetermined process to transform those vectors into another one that is more suited for further processing. For example, Fourier Transform, Wavelets or Parcellation can be considered as feature extraction. In this chapter, we will explore different ways to use Graph Theory and Graph Signal Processing in order to perform feature extraction.

Supervised learning is a class of ML algorithms whose goal is to infer a function f using examples of input vectors, usually in matrix form $\mathbf{X}$ (rows: samples and columns : features), and outputs $\mathbf{y}$. If $\mathbf{y}$ is a categorical variable (e.g. a finite integer set), such a problem is called a classification problem. Examples of classification problems are object recognition in images, speech recognition, automatic diagnosis of pathological conditions. In classification problems, the most common performance metric is the accuracy score, which simply measures the proportion of examples that were correctly classified. Other scores can be defined, to better

[1] Research fields and definitions used largely overlap, and refer more to scientific communities, specific conferences, and scientists or engineers with various backgrounds. There is no hard boundary between Artificial Intelligence, ML or Data Science.

deal with possible dataset biases and imbalances, such as the recall score or the sensitivity. If $\mathbf{y}$ is not categorical, for example if a real number, such a problem is called a regression problem. Example of regression problems are time series forecasting, age prediction using structural MRI, or reaction time prediction using brain activity. Performance metrics for regression problems can be the Mean Absolute Error (MAE), the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), or sometimes the Pearson correlation coefficient ρ.

Unsupervised learning is a class of ML algorithms that attempt at uncovering the structure of a dataset $\mathbf{X}$, without any other information. The overarching goal of unsupervised learning techniques is to reduce the dimensionality of the training set, by either finding a partition using a clustering approach, or approximate the training set using a decomposition or manifold learning. Decomposition methods try to find linear mapping between the training set, and a set of smaller dimension, and is usually formulated as a matrix factorization problem. Manifold learning attempts at finding non-linear projections to reduce dimensionality.

The high dimensionality of ML problems on real world data makes it very challenging to find appropriate solutions to a practical problem. Unsupervised learning, by attempting at reducing the dimensionality, seems a natural preliminary step before Supervised Learning, by using unsupervised learning for feature extraction. This kind of approach, combined with deterministic, expert defined feature extractors using domain-specific expertise (e.g. time frequency representations based on auditory perception, such as the mel-spectrogram), dominated many application fields of ML until 2012, when the first very large scale neural network for image recognition was trained ; this was the birth of Deep Learning [LeCun *et al.* 2015].

Deep Learning, or the learning Deep Neural Networks, is a particular class of ML algorithm that distinguishes itself by using a combination of three ideas:

- End-to-End : a Deep Learning approach attempts at learning a cascade of representations from raw signals until final decision, without expert-defined feature extraction.
- Compositional : a Deep Learning approach learns many (thousands / millions) elementary functions that are combined in a computational graph. These functions are usually a matrix-vector product, followed by a non-linear function[2].
- Differentiable : all functions can be tuned together by using variants of the stochastic gradient descent.

Training Deep Learning models is similar to other ML approaches, but differs by two important steps. First, Deep Learning algorithms typically need several tens to several hundreds of iterations, called Epochs, on the whole training set. Each epoch is further decomposed in smaller batches of data, and the optimization and update of parameters is done at each batch. A validation set is used during training to monitor generalization performance, and after training, performance is assessed on the test set. Second, an important concept in Deep Learning is Data Augmentation, which consists in randomly tranforming input vectors at the beginning of each epoch, in order to increase generalization performance. Example of data augmenation techniques are cropping or resizing, adding random noise, masking features by replacing them with zeros.

[2]The combination of a matrix-vector product with a non linearity (in particular one that includes a form of thresholding) bears an analogy with biological neurons, as initially proposed by McCulloch and Pitts, and other pioneering work since the 1940's. Nowadays, the term Deep Neural Networks is still popular, and relevant, also because the cascade of representations learning by deep learning can be sometimes seen in analogy with what is known about hierarchical levels of sensory representations in the human brain.

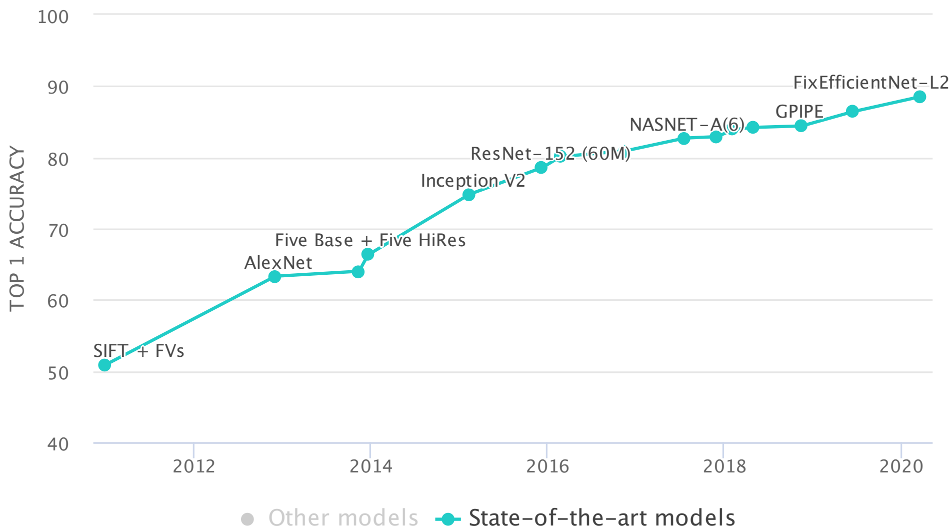


Figure 2.12: Top-1 accuracy for models trained on ImageNet ILSRVC2012, a large benchmark containing millions of images in one thousand different categories. All models except SIFT+FVs are based on Deep Learning and CNN. We only show the pareto frontier according to the corresponding axes, and only models trained without extra data are shown. Data from https://paperswithcode.com/sota/image-classification-on-imagenet

Data Augmentation makes learning more challenging at the level of individual epochs, but is a robust way to increase generalization performance.

Intermediate representations in deep learning are themselves feature vectors, and are also called feature maps. In computer vision, Convolutional Neural Networks (CNN) are an efficient class of deep learning architectures (Figure 2.12). CNN learn spatial filters using trainable 2D-convolution kernels, that are optimized during the training phase. The parameters of a deep learning architecture are also called weights, and the outputs of elementary functions (and in particular, of the non linearities) are called activations.

One remarkable property of representations learnt using deep learning is that they can be reused for another purpose, such as another deep neural network, or a simpler ML approach. This is called Transfer Learning, in the sense that representations for a deep neural net are transferred to another learning approach.

2.2.2.2 Graphs and Graph Signal Processing

The goal of Graph Signal Processing (GSP) is to generalize Fourier's approach, i.e. signal processing of one-dimensional signals such as regularly sampled time series, or two-dimensional signals such as images, to signals evolving on irregular structures by providing an adapted spectral space to decompose them in meaningful frequencies [Shuman *et al.* 2013, Ortega *et al.* 2018].

In the GSP framework, we consider a weighted and undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, W \rangle$ with vertices $\mathcal{V} = \{v_1, \dots, v_N\}$ of cardinal $|\mathcal{V}| = N$, edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, and a weighting function $W : \mathcal{E} \mapsto \mathbb{R}$. Such a graph can be equivalently represented by its weights matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ such that $\mathbf{W}[i,j] = W(\{v_i, v_j\})$ if $\{v_i, v_j\} \in \mathcal{E}$ and 0 otherwise. Additionally, we note $\mathbf{D} \in \mathbb{R}^{N \times N}$ the degrees matrix of $\mathcal{G}$, such that $\mathbf{D}[i,j] = \sum_{k=1}^{N} \mathbf{W}[i,k]$ if $i = j$ and 0 otherwise. From these two matrices, we can compute the normalized Laplacian matrix $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ of $\mathcal{G}$, where $\mathbf{I}_N$ is the identity matrix of dimension N. Since $\mathbf{L}$ is real and symmetric, it can be diagonalized as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U}$ is a matrix of orthonormal vectors associated with eigenvalues forming the diagonal matrix $\mathbf{\Lambda}$, sorted in increasing order. These eigenvalues are analogous to frequencies in Fourier analysis, and are called *graph frequencies*. A signal $\mathbf{x} \in \mathbb{R}^N$ on $\mathcal{G}$ is an observation on each of its vertices. Its Graph Fourier Transform $\hat{\mathbf{x}} = \text{GFT}(\mathbf{x}) = \mathbf{U}^\top\mathbf{x}$ shows the various contributions of eigenvalues of $\mathbf{L}$ in $\mathbf{x}$. Its inverse $\mathbf{x} = \text{GFT}^{-1}(\hat{\mathbf{x}}) = \mathbf{U}\hat{\mathbf{x}}$ transforms a graph spectrum into a graph signal.

In our case, we are interesting in defined graphs using brain regions and connectivity, and use brain activity as signals. This opens the possibility to use a GSP analysis to understand brain activity, for example by using GFT for feature extraction, which is what we will explore in the next sections. The application of graph theory and Graph GSP to neuroimaging is summarized in figure 2.13 (from [Lioi *et al.* 2021]), taken from an overview and perspectives paper that we coauthored with Giulia Lioi (Associate professor, IMT Atlantique).

2.2.3 Efficient Deep Learning

2.2.3.1 Context

As explained in section 2.2.2.1, Deep Learning approaches obtain impressive performances in many fields, and improvements keep increasing, but this has a cost. In figure 2.14, we show the computational capacity needed to train deep learning systems as a function of year, expressed in days spent calculating at one petaflop per second. A two year doubling can be observed from the 1960's until 2012, and is followed by a 3.4 month doubling. This figure also does not show Large Language Models (LLM, figure 2.15), as since 2018 state of the art LLM have a 10x increase in total number of parameter counts every year.

In Figure 2.16, we show the top-1 accuracy as a function of parameters (a) and computations (b), for state of the art models trained on ImageNet. This plot clearly shows that recent progress in increasing the accuracy (about 85% to 86%) have a very high cost on the model size (from about 20M to more than 400M). A similar observation can be made on audio datasets such as AudioSet (Figure 2.17) : more accurate models often come at a cost of increasing dramatically the model size and computational power.

In order to tackle this issue, many different techniques and ideas have been proposed in the litterature:

- Parameter pruning, which consists in permanently removing some parts of a network
- Quantization of weights or activations
- Rethinking operators so that they are less costly to implement
- Training a network (student) using the output probabilities or features from another trained network (teacher); this is called knowledge distillation / feature distillation.
- Using unsupervised learning (e.g. clustering) in order to decompose or quantize the possible values taken by network parameters.

In the following subsections, we present contributions to Efficient Deep Learning using pruning, quantization, and operators. Many of the contributions described in this section were done

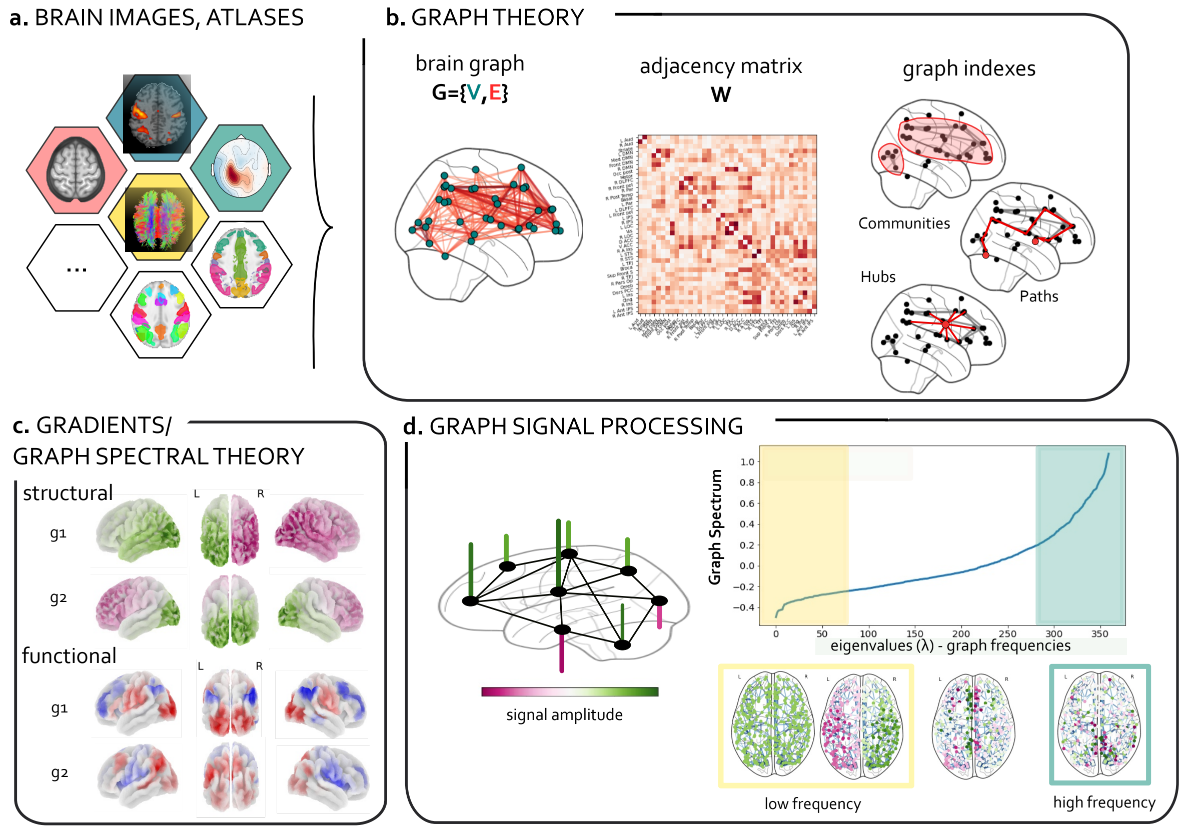


Figure 2.13: From graph theory to graph signal processing in brain imaging. **a.** Different areas of the brain can be represented as nodes and structural and functional relationships between them as edges of a complex large-scale network, also known as the *connectome*. Various approaches exist to identify the nodes of the connectome (atlas and anatomic based, data-driven, etc.). Similarly, edges of a brain network can be assessed with a range of neuroimaging techniques (DWI, EEG, fMRI, MEG, PET) and methods (structural, functional or effective connectivity) **b.** Graph theory allows us to describe salient properties of network topology with matrices (i.e. adjacency, Laplacian, Degree matrices, etc.) and graph indexes (i.e. efficiency, clustering, centrality) **c.** Graph spectral analysis (e.g. Laplacian eigenvectors) is used to extract low dimensional representations of brain networks known as brain gradients. **d.** *Graph signal processing (GSP)* takes a step forward as it associates a signal with an underlying graph. It extends classical analysis methods from regular domains (discrete time signals) to non-regular graphs. GSP allows us to analyze brain activity taking into account the underlying topology of brain networks. GSP also allows for a spectral decomposition of brain activity based on the underlying graph Laplacian eigenvectors (Graph Fourier Transform). In the figure, a brain signal (whose amplitude is encoded in the height and color of the vertical bars) "lives" on a brain network (black) and can be decomposed in low (high) graph frequency harmonics corresponding to small (high) graph Laplacian eigenvalues. In this example the graph spectrum and corresponding Laplacian eigenvectors were obtained from the spectral analysis of an averaged structural graph from the Human Connectome Project. Adapted from [Lioi *et al.* 2021].

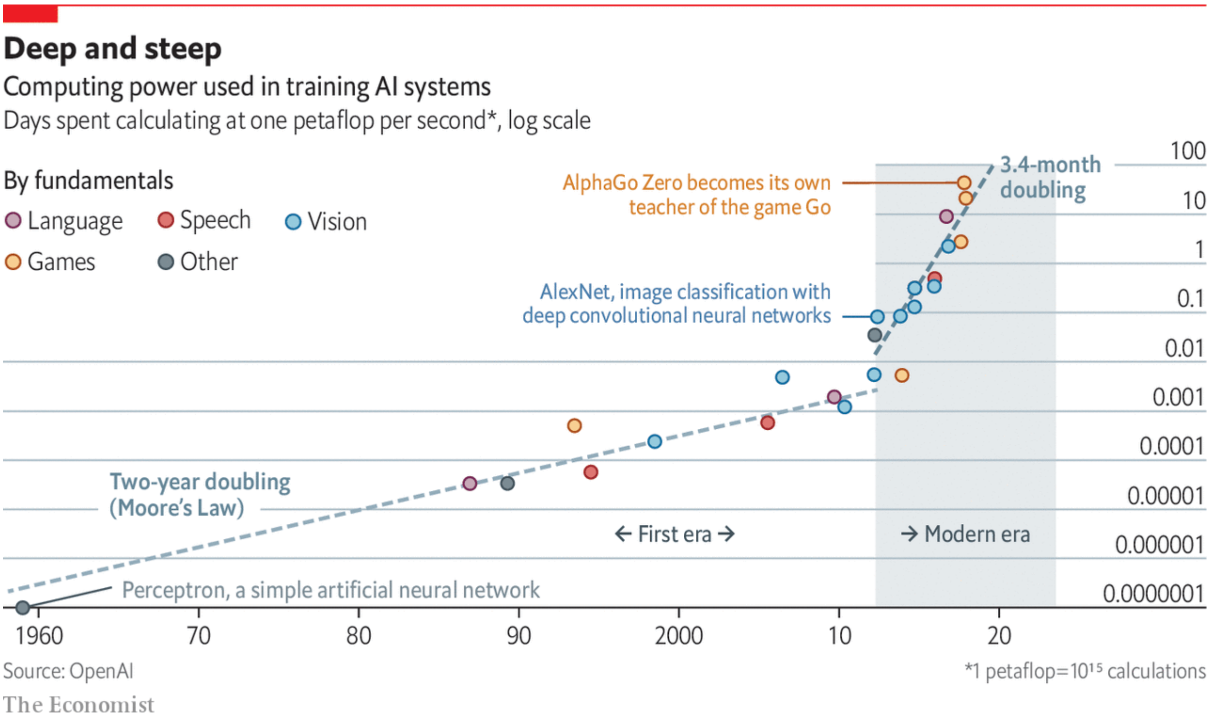


Figure 2.14: Compute capacity needed to train state of the art Deep Learning systems. Figure from The Economist, data from Open AI.

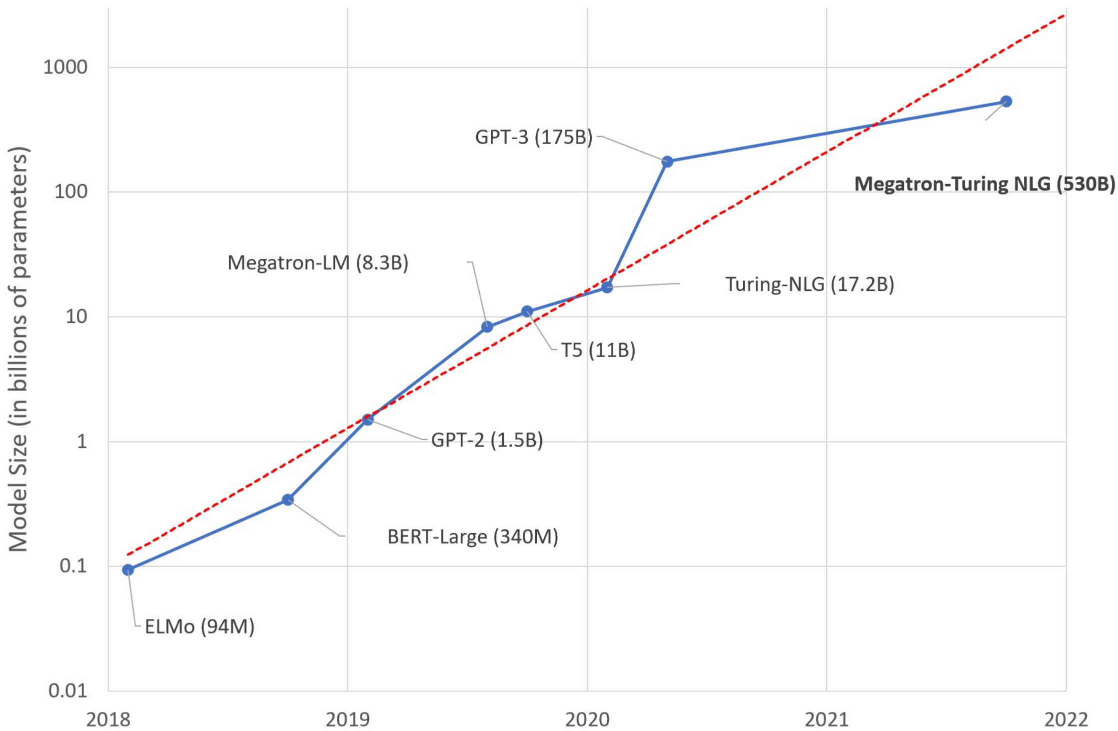
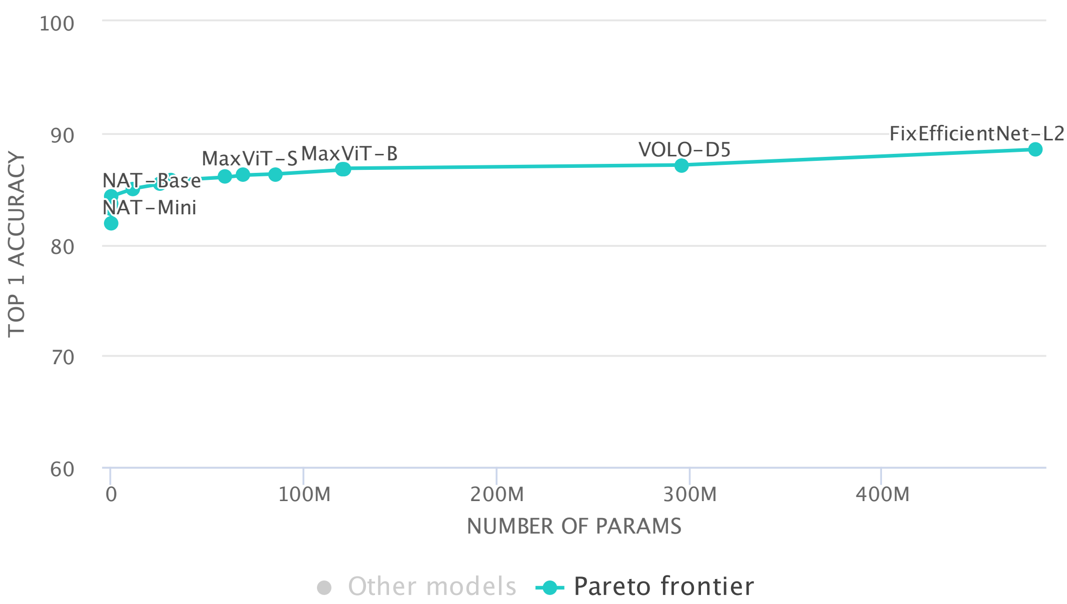
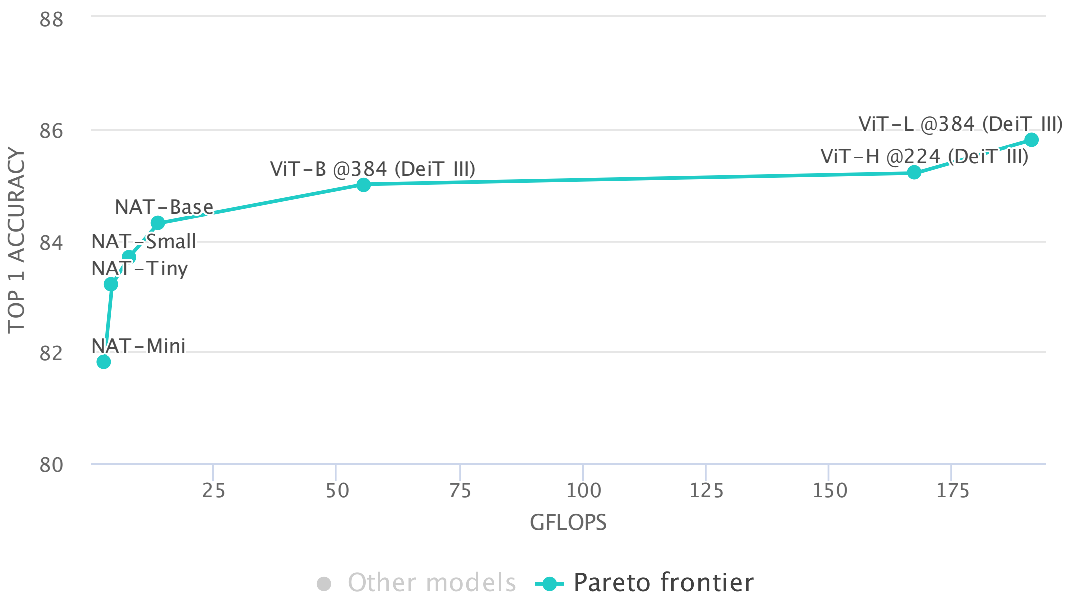


Figure 2.15: Number of parameters in modern large language models (LLM). From https://huggingface.co/blog/large-language-models

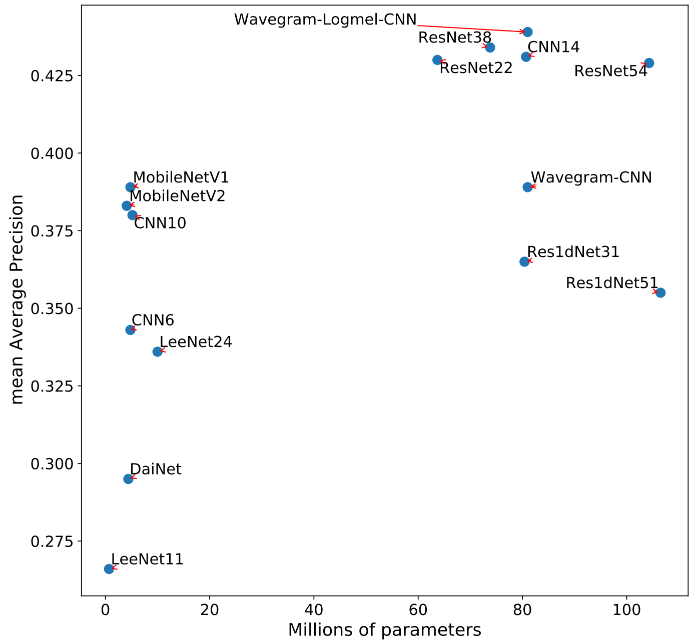


(a)

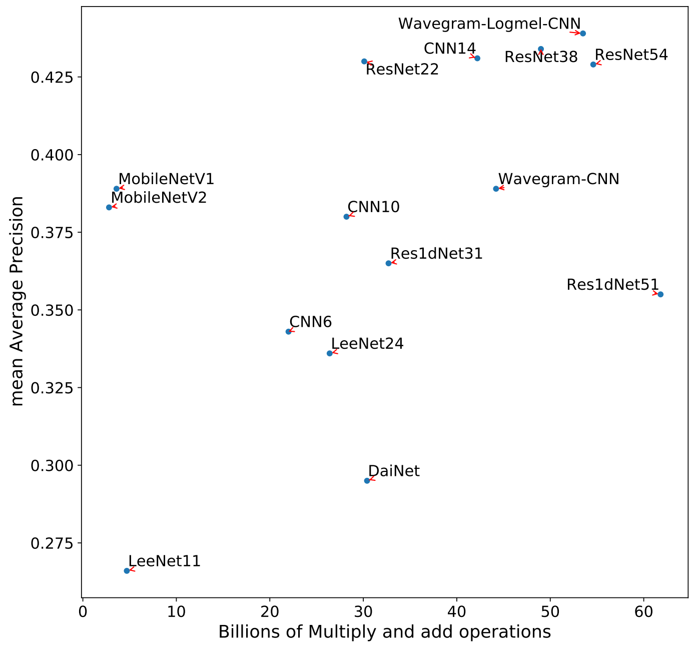


(b)

Figure 2.16: Top-1 accuracy as a function of the number of parameters (a) and computations in Giga floating-point operations (b) in State-of-the-art models for ImageNet ILSRVC2012. We only show the pareto frontier according to the corresponding axes, and only models trained without extra data are shown. Generated from https://paperswithcode.com/sota/image-classification-on-imagenet



(a)



(b)

Figure 2.17: Mean Average Precision on the AudioSet dataset, as a function of the number of parameters (a) and computations in Giga Multiply and Add operations (b), on models trained on Audioset presented in [Kong *et al.* 2020] (figure recreated using tables in the original paper).

in the context of the PhD project of Ghouthi Boukli Hacene, for which I was one of the advisors together with Vincent Gripon (at that time research fellow, now research director, IMT Atlantique) and Matthieu Arzel (Associate professor, IMT Atlantique), and the PhD was supervised by Michel Jezequel (Full professor, IMT Atlantique). Ghouthi proposed a hardware implementation of binary associative memories [Boukli *et al.* 2017], worked on incremental learning with an efficient algorithm [Boukli *et al.* 2019a, Boukli *et al.* 2019b, Boukli *et al.* 2018c] and a related hardware implementation [Boukli *et al.* 2019c].

2.2.3.2 Implementation of parallel architectures for convolutional neural networks

My PhD thesis (2005-2008) was directed by Michel Paindavoine and Fan Yang (both full professors, Université de Bourgogne), and advised by Franck Mamalet and Sébastien Roux (associate researchers, Orange Labs, Meylan). My research consisted in the implementation of parallel architectures for face analysis on Field-Programmable Gate Arrays (FPGA) [Farrugia 2008]. Face analysis is an important research domain with many applications in image indexing, security, videosurveillance and communications. The goal of this work was to efficiently implement a hardware for face analysis algorithms developped by Christophe Garcia (Orange Labs, Rennes). This work, initiated in 2005, was based on CNN, which have since become the standard technique for learning deep neural networks on images and speech data [LeCun *et al.* 2015]. We aimed at transforming, optimizing and proposing a hardware architecture which meets embedded systems' constraints. We set up an exploration methodology using an automated partiotionning tool (SynDEx), which allowed us to study high-level parallelism of the algorithm and to define a theoretically efficient architecture composed of Processor Elements (PE). We then explored and design several PE alternatives using a High Level Synthesis tool called UGH (User-Guided High level synthesis) and validate these PE with an implementation on FPGA. We then use these PE to build an optimized parallel architecture for face detection, which is scalable in terms of input image size, and describe the design of an effective FPGA demonstration board of our system. We have shown that it was possible to define an architecture exploiting mainly data parallelism [Farrugia *et al.* 2007, Farrugia *et al.* 2009], then we have defined elementary processors exploiting locally and globally this parallelism using C to VHDL synthesis techniques [Farrugia *et al.* 2008].

2.2.3.3 Hardware implementation of quantized Deep neural networks

Ghouthi Boukli proposed an original method that combined algorithmic level optimization using pruning, and a quantization scheme that effectively reduce the complexity and memory usage of convolutional layers of CNNs, by replacing the complex convolutional operation by a low-cost multiplexer [Boukli *et al.* 2018a, Boukli *et al.* 2020]. We performed experiments on CIFAR10, CIFAR100 and SVHN datasets and showed that the proposed method achieves almost state-of-the-art accuracy, while drastically reducing the computational and memory footprints compared to the baselines. We also proposed an efficient hardware architecture, implemented on Field Programmable Gate Arrays (FPGAs), to accelerate inference, which works as a pipeline and accommodates multiple layers working at the same time to speed up the inference process. In contrast with most proposed approaches which have used external memory or software defined memory controllers, our work is based on algorithmic optimization and full-hardware design, enabling a direct, on-chip memory implementation of a DNN while keeping close to state of the art accuracy.

2.2.3.4 Incremental learning and hardware implementation

Incremental learning is the problem of learning classes one by one without forgetting previously learned ones. Ghouthi Boukli introduced a new method for incremental learning, called Transfer Incremental Learning using Data Augmentation (TILDA) [Boukli *et al.* 2018c]. TILDA is based on pre-trained DNNs as feature extractors, robust selection of feature vectors in subspaces using a nearest-class-mean based technique, majority votes and data augmentation at both the training and the prediction stages. Experiments on challenging vision datasets demonstrate the ability of the proposed method for low complexity incremental learning, while achieving significantly better accuracy than existing incremental counterparts. A hardware implementation of this work was done in the following publications, using FPGA for prototyping [Boukli *et al.* 2019b, Boukli *et al.* 2019a].

2.2.3.5 Lightweight deep networks for auditory scene classification

Following the contributions of Ghouthi Boukli, in particular his work on quantization and pruning, our research team began to invest new efforts in efficient deep learning (several new PhD have been started, supervised by Vincent Gripon and Mathieu Leonardon, IMT Atlantique). In particular, I started a new research axis on applying similar ideas to audio deep learning.

In [Pajusco *et al.* 2020], we investigated the feasibility of training low complexity convolutional neural networks directly from waveforms. This study was done in the context of our participation to the 2020 DCASE (Detection and Classification of Auditory Scenes and Events) challenge, task 1B on low-complexity acoustic scene classification. The paper [Pajusco *et al.* 2020] was accepted for a presentation to the subsequent DCASE 2020 workshop. We were interested in testing the hypothesis that using a fixed feature extractor (such as spectograms) is detrimental for computational complexity, for two reasons. First, considering a spectrogram (or equivalent) as an image-like input may tend to overparametrize the downstream network, as the effort in training for classification becomes a two-dimensional problem. Second, a spectrogram only considers the power in frequency bands, ignoring the phase. In particular, when considering two channels as input, the phase difference between the channels could be informative. As a consequence, our goal was to show the feasibility to train low complexity networks directly on audio waveforms, with significantly less parameters than with a 2D CNN on spectrograms.

We demonstrated the feasibility of this approach on the TAU Urban Acoustic Scenes 2020 3class dataset, which consists in binaural recordings of urban soundscapes in three classes (indoor, outdoor, transportation). First, we performed resampling after a careful examination of the dataset. Next, we used both input channels to train 1D-CNNs, i.e. with one-dimensional convolution kernels, coupled with stride and/or max-pooling to reduce the size of internal feature maps. Third, we used various strategies for data augmentation, in order to challenge the network to learn relevant audio features with degraded or masked versions of the waveforms. We used five forms of DA: temporal masking, filtering, noise addition, Mixup and CutMix. We trained one dimensional Convolutional Neural Networks (1D-CNN) on raw, subsampled binaural audio waveforms, thus exploiting phase information within and across the two input channels.

In order to further reduce the parameter count, we applied iterative structured parameter pruning to remove the least important convolutional kernels, and perform weight quantization in floating point half precision. For all models, the training protocol is perfomed in the following sequence:

DA	Baseline (Retrained)	Baseline (binaural)	Proposed Model (mono)	Proposed Model (binaural)
None	87.9 ±0.6%	89.5 ±0.5%	84.3 ±0.5%	89.2 ±0.3%
TM, FILT, Noise	87.8 ±0.6%	89.7 ±0.4%	85.7 ±0.4	90.9 ±0.4
CutMix	87.0 ±0.9%	90.2 ±0.5%	86.3 ±0.2%	91.1 ±0.2%
Mixup	87.2 ±0.7%	90.0 ±0.6%	85.1 ±0.3%	89.6 ±0.4%
2D random crop	88.2 ±0.6%	89.6 ±0.7%	-	-
2D cutmix	88.0 ±0.7%	90.3 ±0.7%	-	-
2D mixup	86.7 ±0.6%	90.2 ±0.3%	-	-

Table 2.2: Ablation study comparing data augmentation strategies to train audio deep neural networks, mono versus binaural, between the baseline and our model (5 repetitions, average + 95% confidence interval). Our best result (binaural + cutmix) was obtained after the challenge deadline, thus the difference with the result in the main text. Adapted from [Pajusco *et al.* 2020]

- Training (using data augmentation) until early stopping as indicated by validation set performance,
- Iterative structured pruning on parameters (removing convolution kernels with smallest L2 norm) and fine tuning,
- Quantization to floating point half precision, and final evaluation on the test set.

We report here only the results from our best model : a simple CNN with 5 convolutional layers. The first convolutional layers has kernels of size 64, while the others have a kernel size of 4. Number of output feature maps of each convolutional layer are respectively 16, 16, 32, 64, 64, totalling in about 30 thousand trainable parameters. This model achieved an accuracy of 90.9% on the development set, and 90.6% on the evaluation set (challenge results), which resulted in the team being ranked 14 over 25 teams. However, we focused on reducing the parameter count, and not on optimizing the accuracy level ; in this respect, when taking all submissions with an accuracy higher than 90% (39 submissions over a total of 64 submissions), our model is the second smallest one, first one being of the models of the winning team.

Table 2.2 presents the result of an ablation study on data augmentation, using the proposed model as well as the DCASE baseline. Data Augmentation strategies did not provide clear gains in accuracy for the baseline model. However, having both channels as inputs leads to significant accuracy gains, both when using the raw waveform (+ 5.9%) or mel-features in the baseline model (+1.6%). The larger gain obtained for our model may be explained by the lack of instantaneous phase information in log-mel features. For 2D input, we tested 2D DA such as cutmix, mixup and random crop. When using raw waveform as inputs, the most efficient DA are CutMix (+2%) and temporal masking, filtering and noise together (+1.6%).

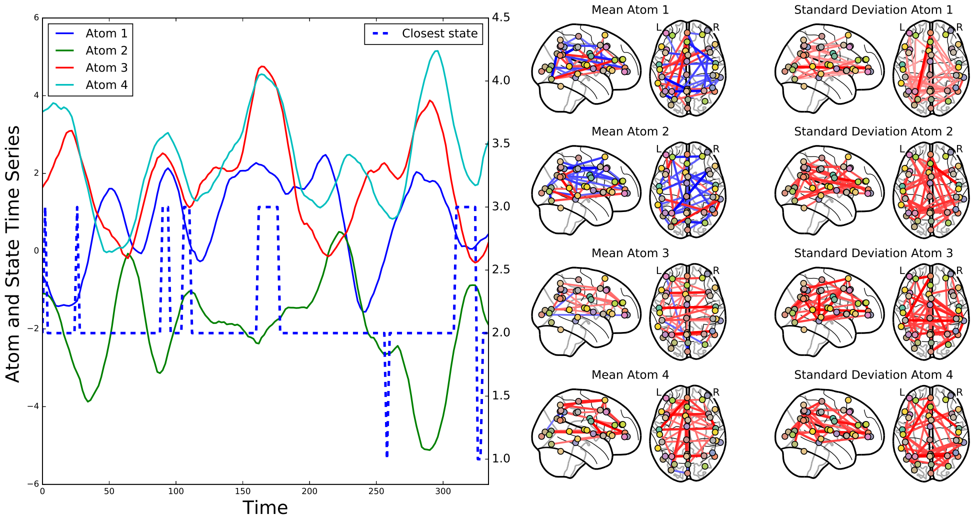


Figure 2.18: Decomposing transition dynamics of resting-state using dictionnary learning on sequences of functional connectivity matrices. Left panel - Time series of atoms (solid lines) for one subject, overlaid with current state obtained with k-means (dashed line). Right panel - Temporal averages and standard deviations of atom connectivity patterns. Blue colored edges (resp. red) depict ones with a negative (resp. positive) influence of the connectivity pattern.

Taken together, these results suggest that using both input channels on the raw waveform makes it possible to train a one-dimensional convolutional neural network with few parameters. We also showed that data augmention can significantly boost the accuracy of such a small model. In future work, we will investigate more complex operators, such as dilated convolutions, as well as attention mechanims, or other data augmentation strategies that better exploit the richness of raw waveforms, in order to explore the design space of functions trained on raw waveforms.

2.2.4 Machine Learning on spontaneous brain activity in the resting-state

2.2.4.1 Decomposition of spatiotemporal patterns of spontaneous cerebral and cerebellar activity

The study of temporal variations in the *resting state fMRI* is subject to many methodological issues and controversies [Lurie *et al.* 2020], and I have attempted to address two main difficulties:

- The temporal variations of the *resting state* are probably not characterized by a finite set of states,
- The temporal sequence is important to characterize the variations, as opposed to an approach considering simple independent sliding windows

I have introduced a method that relies on sparse dictionary learning [Mairal *et al.* 2009] to decompose sequences of FC matrices of brain activity (figure 2.2-C). The result of this sparse decomposition approach is a set of atoms that are themselves temporal sequences of changes in covariance.

Figure 2.18 depicts the result of this method [Farrugia *et al.* 2016]. The left panel of figure 2.18 shows the time series of contributions of all atoms for the same example subject. For each time point, the current connectivity spatio-temporal pattern can be described by the summed contribution of each atom. In comparison, the overlaid time series of the closest cluster shows the 3 possible clusters obtained using k-means clustering for this subject. The right panel of figure 2 describes the functional connectivity (FC) patterns for each atom by estimating their temporal average and standard deviation (over 15 time points). As was performed for clusters, to display these patterns the FC matrices corresponding to atoms were thresholded to keep only the 10% strongest connections. Connectivity patterns of atoms include expected features of spontaneous brain activity, such as strong inter-hemispheric connectivity between bilateral superior temporal regions, most probably auditory cortices in mean atoms 3 and 4, as well as posterior-anterior connectivity of medial regions in mean of atoms 1 and 2. Interestingly, the standard deviation of atoms can show temporal variability in connectivity between nodes that may not have a strong average, revealing transient or oscillating aspects of connectivity.

In figure 2.18, I show that the proposed method yields atom time series that are not linearly dependent, and thus not completely redundant. Additionally, sparsity constraints ensures that for all atoms, there exists a time step for which it reaches a null contribution. This leads to an analysis where only part of the states are overlapping at some time steps. Furthermore, the use of successive FC matrices smoothes the atom's time series, as opposed to the k-means approach. Note that results obtained using the k-means approach and our method are related; for example, state 3 occurs when atoms 2 and 3 have the largest contributions. Finally, the atoms are now temporal patterns reflecting dynamical behavior of the signal, without referring to the time series of atom's contributions anymore.

This work allowed me to start a new collaboration, which led to the supervision of Majd Abdallah's thesis, directed by Sandra Chanraud at the University of Bordeaux. Majd was also interested in the temporal variations of spontaneous brain activity, but sought to extend this research to the study of the cerebellum. Due to the many limits of methods based on sliding-windows [Lurie *et al.* 2020], we decided to use another technique to directly decompose spatiotemporal dynamics of fMRI directly from the timeseries, using the Hidden Markov Model-Multivariate AutoRegressive (HMM-MAR) method, based on variational bayesian inference of hidden Markov models [Vidaurre *et al.* 2017] (figure 2.19). Majd has adapted this technique as well as the preprocessing methods to be able to obtain a sufficient signal-to-noise ratio to study the cerebellum. In his first paper, Majd examined the potential associations between behavioral measures of impulsivity, and cerebello-cortical connectivity [Abdallah *et al.* 2020]. We employed robust data-driven techniques to identify cerebral and cerebellar resting-state networks and extract descriptive summary measures of static and dynamic cerebro-cerebellar FC. We observed evidence linking trait measures of impulsivity, sensation seeking, and lack of premeditation to the total strength and temporal variability of functional connectivity within networks connecting regions of the prefrontal cortex, precuneus, posterior cingulate cortex, basal ganglia, and thalamus with the cerebellum.

In a second paper, Majd examined dynamics of cerebello-cortical connectivity in patients with alcohol use disorder (AUD) [Abdallah *et al.* 2021]. We aimed at exploring cerebro-cerebellar FC dynamics in AUD patients ($N = 18$) and matched controls ($N = 18$). In particular, we quantified group-level differences in the temporal variability of FC, here using FC individual matrices computed with sliding windows because of the smaller sample size. FC matrices were estimated between the posterior cerebellum and large-scale cognitive systems, and we investi-

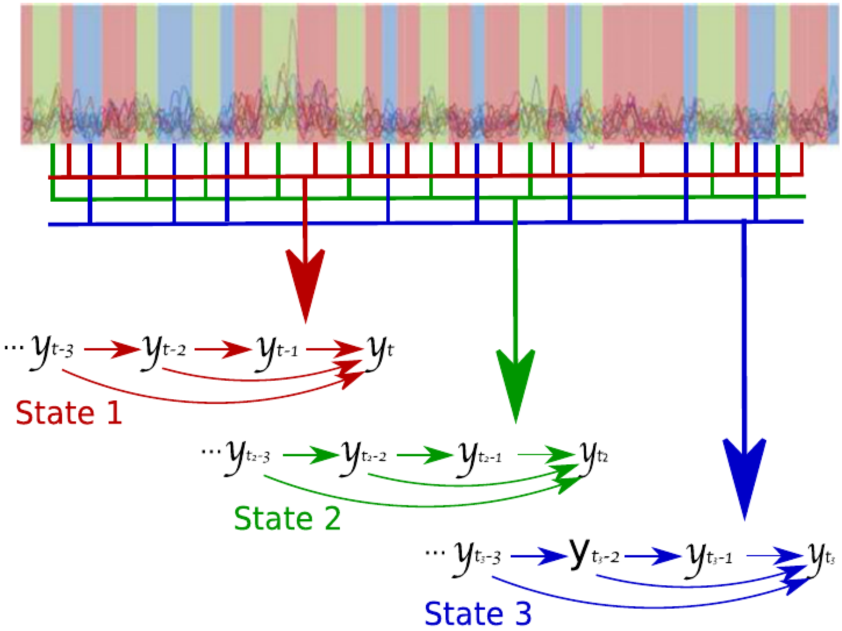


Figure 2.19: Graphical representation of the HMM-MAR model from [Vidaurre *et al.* 2017]. The time series (background) is partitioned into three states denoted by the blue, red and green slabs. Each state is characterised by a different set of dynamics, determined by the linear historical interactions between data points (small arrows). Figure from [Vidaurre *et al.* 2017]

gated the role of the cerebellum in large-scale brain dynamics in terms of the temporal flexibility and integration of its regions. We found that, relative to controls, the AUD group exhibited significantly greater FC variability between the cerebellum and both the frontoparietal executive control and ventral attention networks. Moreover, the AUD group exhibited significantly less flexibility and greater integration in the cerebellum. Finally, in an exploratory analysis, we found distributed changes in the dynamics of canonical large-scale networks in AUD. Overall, this study brings evidence of AUD-related alterations in dynamic FC within major cerebro-cerebellar networks. This pattern has implications for explaining the development and maintenance of this disorder and improving our understating of the cerebellum's involvement in addiction.

2.2.4.2 Phenotype prediction from resting-state graphs using CNN with edge to edge convolution

In another contribution (internship of Amine Echraibi), I have also shown that it is possible to train CNN with convolutions that exploit the weights of the graph edges in a phenotypic prediction task [Farrugia & Echraibi 2018]. Weight matrices $\mathbf{W}$ are obtained by estimating functional connectivity under the resting state, yielding a weight matrix for each subject. Weight matrices are then fed to a CNN, using a similar architecture than BrainNetCNN [Kawahara *et al.* 2017], composed of successive layers of 1D convolutions (conv1D) and fully connected layers. Edge to Edge (E2E) convolutional filters are instantiated in the first two layers. For each vertex $w_{i,j}$ of a weight matrix $\mathbf{W}$, an E2E convolutional filter is defined as a set of trainable weights r_k and

c_k that compute a new adjacency matrix $\mathbf{Z}$, with vertices $z_{i,j}$ defined as follows :

$$z_{i,j} = \sum_{k=1}^{N} [r_k w_{i,k} + c_k w_{k,j}] \tag{2.1}$$

Note than an E2E filter corresponds to a 1D convolution filter applied on connection weights of $w_{i,j}$. We instantiated an architecture starting with two E2E layers with 32 convolutional filters each, followed by a layer with 64 1D-convolutions, and two fully connected layers of size 258 and 30.

We used this architecture to investigate the functional connectivity profiles underlying musical training, on a subset of the Max Planck Institute Leipzing Mind Body Brain Dataset (MPILMBB) on which I contributed [Mendes *et al.* 2019]. We used 77 subjects (mean age 22.5, std 3.2). Musical training and engagement with musical activities was measured using the Golsmidths Musical Sophistication Index (Gold-MSI). Four resting-state runs of 15.5 min each were acquired on a Siemens 3T Tim Trio MRI. Participants were measured at rest with eyes-open. We used the minimally preprocessed version of the released data.

We trained this network on 1000 epochs, using a learning rate lr $= 0.01$ and mini-batches of size 14. A third of the dataset was kept for the validation. To visualize the connectivity patterns corresponding to maximizing the output scores, we used activation maximization (figure 2.20), regularised using L1 norm to enforce sparsity, to which we added an additional regularization term enforcing ternary outputs (1, -1 or 0). The trained network was able to predict both Gold-MSI scores accurately, yielding a mean Absolute error for both scores of 1.32% for Active engagement and 1.57% for Training). Predicted scores were highly correlated with true scores ($\rho = 0.52$ for Active Engagement, and $\rho = 0.69$ for Training, both $p < 0.001$).

2.2.5 Graph Signal Processing for Neuroimaging

In this section, we use the formalism introduced in section 2.2.2.2, and summarize a line of work that is based on a main principle : using Graph Signal Processing to perform feature extraction, followed by supervised learning on downstream tasks. Regarding the graphs, we will use either resting-state functional connectivity graphs, or anatomical connectivity graphs estimated using diffusion weighted imaging.

2.2.5.1 Graph Fourier Transform and dimensionality reduction for decoding brain activity

In [Ménoret *et al.* 2017] (work of Mathilde Ménoret, postdoctoral fellow), we evaluated whether GSP can lead to more accurate supervised classification of fMRI data (i.e. decoding brain activity), as well as whether GSP can be used for dimensionality reduction. We studied the influence of different types of graphs on decoding brain activity, taking into account either geometrical or statistical dependencies between voxels, or both. To do so, we used GFT to decompose brain activity into spectral components, compared several methods for dimensionality reduction of the decomposed signals (namely, graph sampling or statistical selection), and finally we evaluated the performance of these methods against state-of-the-art reduction techniques such as PCA and ICA. We performed our experiments on two datasets, a simulated fMRI dataset and a real open source fMRI dataset (Haxby dataset, a study with subjects viewing different types of images while being measured with fMRI). We simulated fMRI data in two conditions,

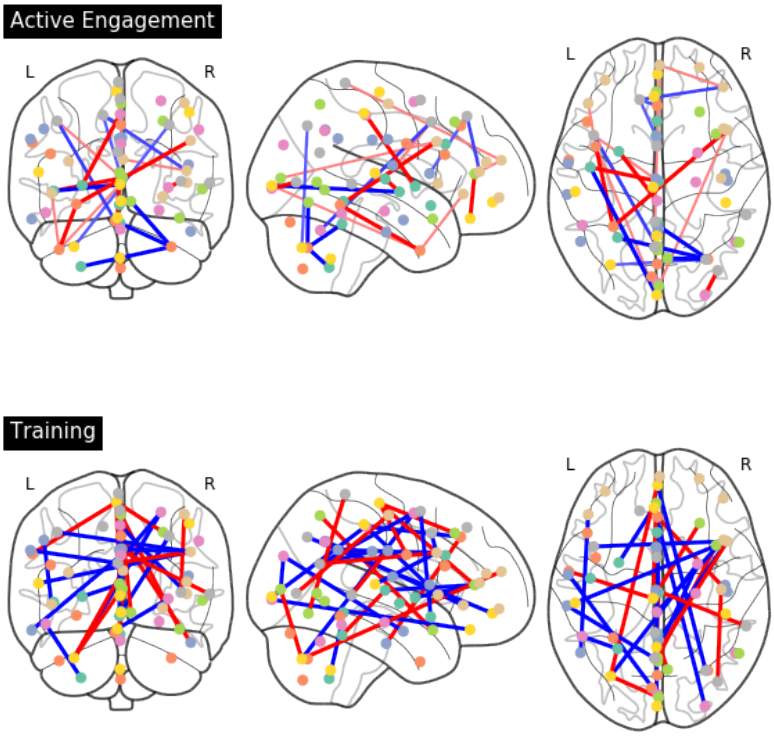


Figure 2.20: Prediction of Musical Training and Musical Engagement using Edge-to-edge convolutions. Most important edges are shown for visualisation. Adapted from [Farrugia & Echraibi 2018]

with 6 ROIs involved in visual processing. We obtained simulated subjects with varying signal to noise ratio, resulting in a group of "Easy" subjects and a group of "Difficult" subjects. More details on the simulation setup can be found in our paper [Ménoret *et al.* 2017].

In our experiments, we considered different methods of dimensionality reduction. In particular we compare graph sampling (GS) to other graph-free methods: PCA, ICA, and selection of the K best components using analysis of variance (ANOVA). GS is a method to select the vertices where the signal energy is the most concentrated. To do so, we compute the graph weighted coherence for a frequency band of interest delimited by indices (f_{min}, f_{max}), and extract the K vertices achieving maximum scores. The graph weighted coherence for a vertex v_i is defined as $\sum_{k=f_{min}}^{f_{max}} u_{ik}^2$. We restrict our analysis to either only *low frequencies* (LF) (below $N/2$) or *high frequencies* (HF) (above $N/2$).

Classification is performed to disentangle brain signals originating from different conditions using linear Support Vector Machine (SVM). To avoid excessive over-fitting and given the block design, cross-validation is performed across different sessions, leaving two sessions out: 16% of the data is used as test data, the remaining as training. Classification is performed on the signal projected in the graph Fourier domain using GFT, after dimensionality reduction using either PCA, ICA, ANOVA or GS.

Classification of the simulated data revealed that high frequencies are more relevant for the classification than low frequencies. Moreover, using a graph that exploits both geometrical

Method	Simulation		Haxby	
	Easy	Difficult	Face-House	Cat-Face
PCA	88.8%	65.5%	82.7%	63.6%
ICA	90.2%	65.3%	84.4%	67.0%
ANOVA	**92.1%**	67.3%	85.5%	65.5%
Graph sampling	90.9%	**72.5%**	**88.2%**	**69.0%**

Table 2.3: Comparison of Graph Sampling (*Semilocal* graph), PCA, ICA and ANOVA. Classification accuracy with 50 components for the Simulated and Haxby datasets. Adapted from [Ménoret *et al.* 2017]

information and covariance calculated from resting-state (semilocal graph), classification stands out from the other graph types and reaches better scores in both groups (72.5%, 90.9%). The *Semilocal* graph was selected for further analysis. When comparing the optimal number of dimensions, GS yields the best accuracy when selecting 30 components for the *Difficult* group and 50 components for the *Easy* group.

The performance of GS relative to state-of-the-art reduction techniques such as PCA, ICA and ANOVA is then compared for simulated and real fMRI data. Table 2.3 presents the results. For the simulated fMRI data, the classification with GS is significantly more accurate in the *Difficult* group than PCA ($Z = 5.9$, $p < 0.001$), ICA ($Z = 5.9$, $p < 0.001$) and ANOVA ($Z = 5.9$, $p < 0.001$). In the *Easy* group, classification is significantly more accurate for the ANOVA (PCA: $Z = 4.4$, $p < 0.001$; ICA: $Z = 5.5$, $p < 0.001$; and GS: $Z = 3.7$, $p < 0.001$). Classification with GS is significantly more accurate than PCA ($Z = 5.2$, $p < 0.001$), but not than ICA ($Z = 2.0$, $p = 0.05$). For the Haxby dataset, the classification with GS produces the most accurate results for both conditions. However, those differences do not reach statistical significance (PCA: $Z = 2.3$, $p = 0.022$ uncorrected, ICA: $Z = 1.5$, $p = 0.126$ uncorrected, ANOVA: $Z = 1.5$, $p = 0.126$ uncorrected). In figure 2.21, we show the spatial distribution of the classifier SVM trained on signals with the 50 best ROIs (i.e. graph vertices) sampled with GS, and compare this distribution with a simple statistical contrast between the two conditions (top panel). We can see that the ROIs selected by GS correspond to peak differences in the statistical contrast, confirming the relevance of using GS.

To summarize, in this work we tested the contribution of Graph Signal Processing to brain signal analysis. We constructed graphs that model the geometric and/or the functional dependencies of brain activity on simulated and real fMRI data, and compared classification accuracy for difference choices of graphs and dimensionality reduction techniques. We showed that applying graph sampling to a semilocal graph could select meaningful vertices for classification, without any prior hypothesis on the categories to distinguish, and led to a significant improvement in classification accuracy compared to PCA, ICA and ANOVA when categories are difficult to distinguish. The semilocal graph best fits the data structure by taking into account both the geometric structure of the data and functional connectivity between brain areas at rest, and improves classification and dimensionality reduction of neuroimaging data.

We also tested other graph types, such as the Kalofolias graph, which is a method that infers the graph by using the prior that signals are smooth on the graph, i.e. that signals are mostly concentrated in the low frequencies. We indeed observed that low frequency features

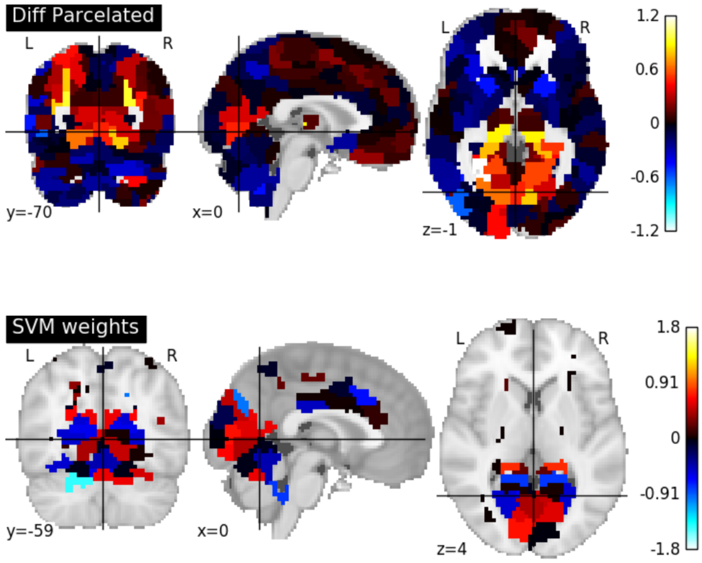


Figure 2.21: Results of fMRI classification of Face vs House in the Haxby dataset using Graph Sampling. Top panel : Statistical contrast of Face vs House using a general linear model. Bottom panel : SVM weights of a classification model on Graph Sampled signals.

are better for the Kalofolias graph. For the semilocal graph on the opposite, while low graph frequencies might correspond to more stable activity spread across the brain for both categories (e.g. the gradient that develops from occipital to more frontal areas during visual processing), the discriminative features between considered conditions are localized in the brain, and thus carried out by high frequencies of geometric graphs. We believe this observation is particularly interesting for anyone interested in applying a similar methodology to perform classification: the components of signals to be kept are highly dependent on the method used to build the graph, as well as on the type of expected discriminative features between conditions.

2.2.5.2 Spectral Graph Wavelet Transform as a feature extractor for fMRI decoding

In [Pilavci & Farrugia 2019], we investigated the benefits of using Spectral Graph Wavelet Transform (SGWT, [Hammond *et al.* 2011]) as a feature extractor for decoding brain activity. SGWT is defined by a wavelet and a scale function, both continuous functions defined in the spectral domain and evaluated only at the eigenvalues $\boldsymbol{\lambda}$ (diagonal elements of $\boldsymbol{\Lambda}$). We used a definition of wavelet functions, called warped translates [Shuman *et al.* 2015], using a warping function that modifies the kernel's behaviour on the spectrum. In [Shuman *et al.* 2015], the superior discriminatory power of warped translates is clearly demonstrated, because different parts of the evaluated spectrum are perfectly covered and segmented by different kernels. We could observe this effect in practice ; in figure 2.22, we show the different warped translate wavelet kernels obtained when calculated for a K-Nearest Neighbors graph estimated using resting state (KNN-Correlation Graph).

In order to test the relevance of using SGWT as a feature extractor, we first considered a synthetic regression problem to optimize the spectrum coverage using different wavelet shapes. Next, we evaluated the benefits obtained by SGWT on an open dataset of fMRI measurements on sub-

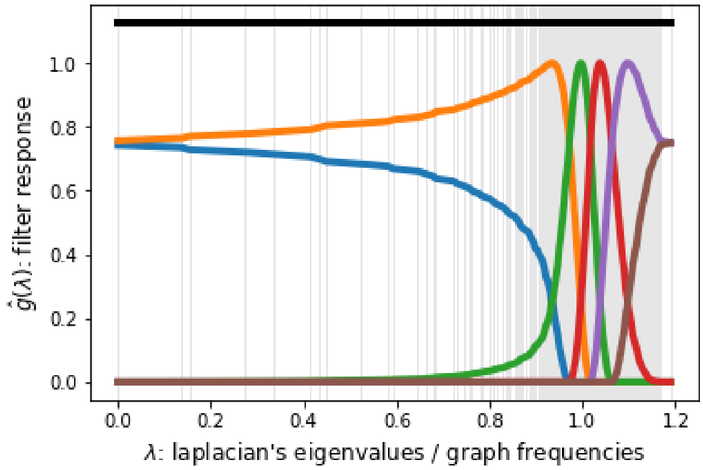


Figure 2.22: Warped Translate Wavelet Kernels on Spectral Domain, in the case of the KNN-Correlation Graph. Vertical lines depict placement of eigenvalues. Adapted from [Pilavci & Farrugia 2019].

jects who rated pictures with emotional content [Chang *et al.* 2015]. We fetched statistical maps of whole brain activity during single trials for each subject, from neurovault (collection number 1964). The supervised learning task consists in predicting the rating given by the subject from brain maps. We reproduced the decoding pipeline from the original paper [Chang *et al.* 2015] to be able to compare results. As this dataset didn't include connectivity data that could be used to compute subject-specific graphs, we estimated average brain graphs from resting state fMRI data (i.e. spontaneous fluctuations of the brain at rest) from 158 subjects of the MPILMBB, as in other studies reported in this manuscript [Mendes *et al.* 2019].

Results showed that SGWT provides significant performance gains (best model obtained a $RMSE = 0.983$, $\rho = 0.725$) when compared to using only parcellated ROI signals ($RMSE = 1.022$,$\rho = 0.701$), or original signals with no parcellation ($RMSE = 1.036$, $\rho = 0.693$) [Chang *et al.* 2015]. In particular, warped translate have a better potential for generalization to the test dataset. We suggest that SGWT enables an efficient exploitation of underlying multivariate dependencies, using spectrum-adapted wavelet kernels on a brain graph. In order to provide some spatial interpretation of our results, we depict in figure 2.23 the different scale and spatial localization of the largest regression coefficients

In this contribution, we have demonstrated the potential of combining SGWT and machine learning using synthetic data and fMRI data from open datasets. A key point of the proposed approach is to rely on Warped Translated kernels for wavelet definitions, which optimizes spectral coverage of SGWT. We showed that using features from SGWT can boost performance in a challenging regression task on neuroimaging data.

2.2.5.3 Graph Fourier Transform to extract relevant latent space for spatiotemporal sequence modeling

Building on the contributions presented in the previous paragraphs, we hypothesized that GFT could be useful to model spatiotemporal signals. The underlying idea was to rely on GFT to model spatial dependencies, and subsequently use Deep Learning models adapted to sequence

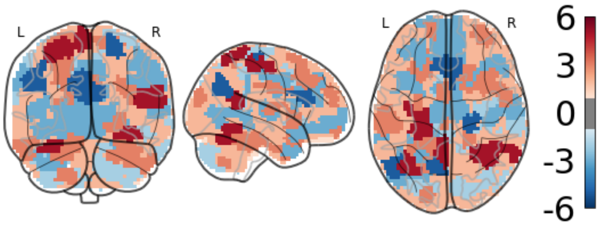


Figure 2.23: Significant Scale-Localization map on brain. Positive and negative weights are denoted with ± sign. Adapted from [Pilavci & Farrugia 2019].

modeling, such as LSTM (Long and Short-term Memory) models, using GFT features as input. This work [Bontonou *et al.* 2019] was performed in the beginning of Myriam Bontonou's thesis for which I was one of the advisor.

First, to model spatial dependencies, Myriam compared GFT, with autoencoders (AE). The fMRI dataset is a subset of 62 subjects from the MPILMBB resting-state dataset [Mendes *et al.* 2019]. Here, we only use the first measurement for each subject. The denoised data was z-scored and normalized to MNI space and parcellated on 523 non-overlapping ROIs from the finest scale of BASC atlas (444 networks). Two subjects were rejected because of corrupted data for one subject (the time series was shorter than expected), and because of excessive head motion for the other subject. 60 subjects remain, and we use 48 subjects for training and the 12 remaining subjects for testing.

We evaluated the ability of GFT applied on a correlation graph (GFT) and a linear autoencoder (AE), i.e. without activation function (the AE just corresponds to a learnable linear projetion). We compute the MSE between raw data and their reconstruction in function of the number of latent dimensions for 523 regions of interest (ROI) and for 171 ROI (table 2.4). The data for 171 ROI has been obtained after a hierarchical clustering of the 523 ROI. Interestingly, the AE obtains a better MSE, regardless of the latent dimension. Subsequently, when trying to model temporal dependencies using either the AE or GFT, all models performed equally than just predicting the average of the time series. Therefore, we did not pursue with this approach for modeling fMRI time-series.

	Loss	
Latent dimension	86	43
GFT	0.126	0.246
AE	0.065	0.134

(a) 171 ROI

	Loss	
Latent dimension	262	131
GFT	0.174	0.317
AE	0.102	0.189

(b) 523 ROI

Table 2.4: MSE loss on the fMRI test set as a function of the number of ROI and the number of latent dimensions.

2.2.5.4 Graph Fourier Transform as a feature extractor for benchmarking biomarkers using resting-state

The work presented in this paragraph takes a radical turn from the one described in section 2.2.5.3 ; while the idea is still to use GFT as a feature extractor, instead of using all time points in the time series, we attempted to summarize resting-state fMRI time series using simple statistical descriptors.

This work was done by a postdoctoral fellow that I supervised, Abdelbasset Brahim [Brahim & Farrugia 2020]. The application that was chosen to benchmark this idea is the ABIDE dataset, which consists in patients suffering from Autism Spectrum Disorder (ASD), measured in resting state fMRI in many different hospitals around the world.

We began by parcellating the fMRI data using the HCP-MMP (Human Connetome Project Multi Modal parcellation, Glasser et al. 2016), which includes 360 regions of interest (ROI) that we consider as graph vertices. We used the resting state time series to compute average, standard deviation and variance over time, for each ROI separetely, which yielded a single value per ROI per subject. Those ROI-wise signals were considered as signals on graph for the GSP framework, and we used an anatomical graph obtained from diffusion weighted imaging, obtained by an average over 56 subjects of the Human Connetome Project [Brahim & Farrugia 2020], and we defined the GFT using the anatomical graph. Therefore, we obtained several features based on temporal statistics of the resting state, and apply GFT, in order to test this combination as a feature extractor which combines resting-state and anatomical connectivity.

In order to compare our approach with other graph-based metrics, including ones that are commonly used in the litterature on the same dataset, we estimated a FC matrix for each subject, that we used to estimate several ROI-wise graph metrics: clustering coefficient, eigenvector centrality, and node strength (defined as the sum of the weights for the considered ROI). We also used the upper diagonal of the FC matrix as a feature extractor, as this is the most commonly used feature for this task. These steps yielded a second series of feature extractors, this time solely based on the resting state time series.

The same pipeline for classifying ASD was applied to all extracted features (GFT based, graph-metrics, and FC matrix) : we performed supervised univariate feature selection using ANOVA (K-Best method), followed by a simple classifier - here, we present results for a support vector classifier with Radial basis function kernel. We use accuracy, sensitivity and specificity as performance metrics in order to have a complete view on classifier biases.

A comparison of the different feature extractors is given in table 2.5, when selecting models that performed better than chance level classification (see [Brahim & Farrugia 2020] for details on the boostrapping procedure to estimate chance level). Such models have a different number of selected features after K-Best, and our best model selected 100 features. Applying GFT to the standard deviation over time of resting-state time series led to the best model in terms of accuracy and sensitivity in detecting ASD.

The proposed rs-fMRI analysis method was compared with several approaches for ASD diagnosis from the literature, as shown in Tables 2.6 and 2.7. The comparison is based on three criteria, i.e. the number of subjects, the number of features used for the classification and the resulting accuracy values. Taking into account more than 870 subjects from the ABIDE database, the proposed approach might be comparable with the other studies reported, although, we did not use the optimal pipeline for feature selection and classification. These differences in feature selection/classification algorithms might account for the differences in classification

Approaches	Acc (%)	Sen (%)	Spe (%)
Std Dev	59.09±0.034	44.69±0.054	76.54±0.055
Std Dev+GFT	60.89±0.039	53.46±0.051	69.47±0.052
Variance	57.26±0.020	33.02±0.088	81.33±0.078
Variance + GFT	58.51±0.036	45.86±0.066	72.71±0.070
Kurtosis	54.77±0.019	13.14±0.082	93.29±0.043
FC matrix	60.4±0.030	52.04±0.045	68.35±0.052
Eigenvector Centrality	56.43±0.033	50.37±0.04	70.05±0.063
Node Strength	56.31±0.032	46.04±0.069	70.95±0.051
Clustering Coefficient	57.54±0.029	42.16±0.064	71.64±0.060

Table 2.5: Maximum classification rates of the different approaches for ABIDE database using an SVC classifier with RBF kernel (max ± SD). Adapted from [Brahim & Farrugia 2020]

Approach	N. of subjects	N. of features	Accuracy (%)
Std Dev+GFT with SVC RBF kernel (Ours)	871	90	60.89
Nielsen et al. 2013	964	7266	60
Iidaka et al. 2015	640	2728	90
Ira Ktena et al. 2017	871	7260	62.9
Abraham et al. 2017	871	3486	66.9

Table 2.6: Comparison of different approaches for classification of the ABIDE database. Adapted from [Brahim & Farrugia 2020]

Approach	N. of subjects	N. of features	Accuracy (%)
Std Dev+GFT with SVC (Ours)	172	100	70.36
Nielsen et al. 2013	179	7266	65
Dodero et al. 2015	79	34,716	60.76
Wee et al. 2016	92	NA	71
Heinsfeld et al. 2018	175	19,900	66

Table 2.7: Comparison of different approaches for classification of a subset of the ABIDE dataset. Adapted from [Brahim & Farrugia 2020] with some data also from [Brahim *et al.* 2019]

accuracy between the studies. However, in the studies from Iidaka, only subjects under 20 years of age were included in their study and their model is age-dependent.
Moreover, when using the biggest subset from the full ABIDE database (we report this analysis also in [Brahim *et al.* 2019]), the proposed approach outperforms current methods. Specifically, our method is able to outperform recent methods based on deep learning applied on hundreds of subjects. The slight difference in the accuracy value with Wee et al. 2016, may be explained by the difference in the number of subjects for the same NYU dataset.

Taken together, these results reveal the reproducibility and generalizability of the proposed approach, which may work on small and even large databases, as exemplified by statistically robust gains in the classification metrics. However, from a methodological point of view, our main aim in this study is to present a novel modelling time series approach applied on rs-fMRI brain imaging, rather than the identification of biomarkers for ASD using intrinsic functional brain connectivity. The analytic procedure employed in the present study represents an entirely hypothesis-free, GSP-based approach, and we provide our analysis code for replicability (see `https://github.com/AbdelbassetBrahim/GSP-applied-on-ASD-classification`).

Chapter 3

Research perspectives

Contents

3.1 Takeaways from previous work

In chapter 2, I have described previous research in auditory cognitive neurosciences, machine learning and graph signal processing. The main takeaways can be summarized as follows:

- The abilities to perceive temporal regularity and to synchronize movements is widespread, still subject to important individual differences. Exceptional abilities occur, such as poor synchronizers that could be detected with the BAASTA (section 2.1.3.4), as well as prodigies (section 2.1.3.2). Spatial features of such movements (e.g. the length of steps in gait, or the amplitude of a drumstick movement) are crucially related to their temporal accuracy (section 2.1.3.2 and 2.1.3.3).

- Parkinson's Disease disrupts sensorimotor synchronization, and stimulating patients with a music cueing program can help rehabilitate gait ; however the outcomes of therapy might be linked to residual abilities in perceiving temporal regularity and synchronizing (section 2.1.4).

- Neural responses to temporal regularity can be studied with EEG using a modified oddball paradigm, and we could find attenuated differences as a function of temporal regularity in patients with PD, suggesting that temporal regularity in auditory information might not be processed optimally in the brain of patients (section 2.1.5).

- Individual differences in brain structure and brain connectivity could be linked to the frequency and affective evaluation of the music in our heads. In addition, the speed of our mental soundtrack may be related to our emotional states (section 2.1.6).
- Using Machine learning and models on Graphs, we could show that spontaneous brain activity (i.e. the brain's resting-state) could be linked with the spontaneous activity of the cerebellum, and that the strength and temporal dynamics of this cortico-cerebellar connectivity was linked to impulsivity traits and alcohol addiction. We also showed some first results relating connectivity patterns and musical training in a large cohort. (section 2.2.4).
- We have shown that Graph Signal Processing is a promising framework to integrate prior knowledge about brain connectivity (either anatomical or functional) to extract useful features for machine learning tasks, such as decoding mental states in fMRI or diagnosing autism spectrum disorder (section 2.2.5).
- Finally, we have shown how deep learning algorithms can be revisited in order to reduce their computational cost, by rethinking operators used, by applying parameter pruning or efficient reuse of transferred features (section 2.2.3).

In this final section, I will outline current work, middle term and long term research perspectives. Some of the work presented here is currently being investigated by members of my team, or with collaborators. Section 3.2 presents our current and future work planned on the theme of leveraging graph methods for neuroimaging, both by combining brain graphs in a deep learning framework, or by investigating new tools based on graph signal processing for multimodal neuroimaging. In section 3.3, I describe recent work that attempts at combining neuroimaging studies with computational modeling, by exploiting the richness of ecological conditions. Finally, I give some outlooks in section 3.4 on future directions that could integrate interdisciplinary perspectives to better understand sound using (artificial) audio intelligence, by rethinking how audio can be efficiently dealt with in deep learning, as well as investigating better representations. These perspectives carry the ambition to be applied to various dimensions of bioacoustics, ecoacoustics or geoacoustics.

3.2 Leveraging methods on graphs for Neuroimaging

This theme is the result of the research detailed in the section 2.2.5, which opens many perspectives in signal processing, machine learning, and neuroimaging. We have secured funding for several short and middle term projects, described in the next paragraphs, in order to develop this line of work. We are also setting up new collaborations with the Brest university hospital (Prof. MD. Serge Timsit, Faculty of Neurology) in order to investigate how clinical applications can gain insights and explanatory power from the methodological contributions descibed in this manuscript.

3.2.1 Towards a Graph Fourier Basis of Brain Activity

In parallel with the development of signal processing and deep learning techniques on graphs, the field of cognitive neuroscience and network neuroscience is now experiencing a revival with the proposal of connectivity gradients, which are based on the application of spectral graph theory to the analysis of brain connectivity. Work on these gradients has shown that they can be used to interpret brain organization in functional terms, that they also correspond to the anatomical and cellular organization of the brain. We proposed in a Perspectives article to exploit this new understanding of the role of gradients, by considering them as a Fourier basis of brain activity [Lioi *et al.* 2021]. Such a view opens a new interpretational framework for cognitive neuroscience, by integrating the brain organisational principles embedded in connectivity information into the analysis of brain activity.

3.2.2 Graph signal processing and Multimodality

In the continuity of our perspectives article [Lioi *et al.* 2021], we will try to exploit graph-based signal processing methods to combine multiple modalities in neuroimaging. The PhD project of Venkatesh Subramani started in January 2022, and is a joint project with the Computational Consciousness Lab (CoCoLab) directed by Karim Jerbi at the University of Montreal. Karim Jerbi is an expert in magnetoencephalography (MEG), and the goal of Venkatesh's thesis will be to combine the very good spatial resolution of functional MRI with the temporal resolution of MEG, using graph-based signal processing and deep learning methods, in order to study cognitive neuroscience questions related to sensory perception and Consciousness. Another PhD project will start in September 2022, in the context of the PEPERONI CominLabs project (with Pierre Maurel and Julie Coloigner, Empenn Team, INRIA, Rennes), in which we will investigate methods based on multimodal GSP to analyze a simultaneous EEG-fMRI neurofeedback dataset [Lioi *et al.* 2020]. The goal of this project is to integrate multimodal measures using GSP for personalized neurofeedback using cohorts of healthy subjects and patients (stroke and depression) that underwent the simultaenous EEG-fMRI training paradigm.

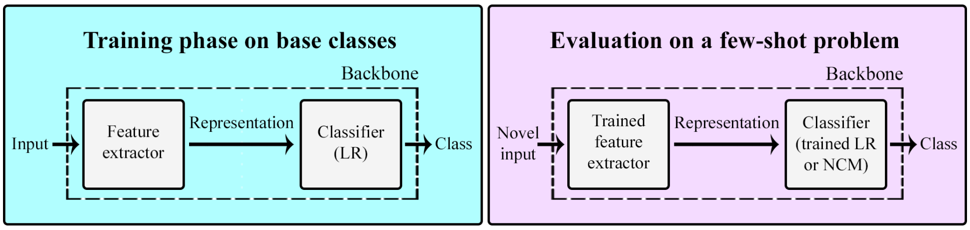
3.2.3 Deep Learning and Graphs for neuroimaging

3.2.3.1 Few-shot Learning for Neuroimaging

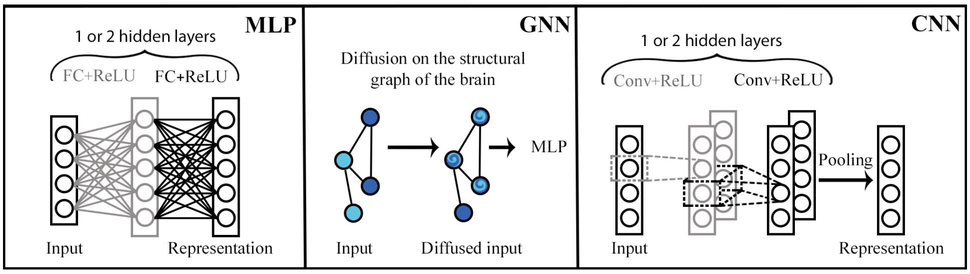
Learning with few examples is a recent theme in the activities of the team, and I contribute to it as a co-supervisor of Myriam Bontonou's thesis, who defended in December 2021.

Few-shot learning addresses problems for which a limited number of training examples are available. So far, the field has been mostly driven by applications in computer vision. Myriam was interested in adapting recently introduced few-shot methods to solve problems dealing with neuroimaging data, a promising application field. To this end, Myriam created a neuroimaging benchmark dataset for few-shot learning and compared multiple learning paradigms, including meta-learning, as well as various backbone networks. The general framework of few shot learning is shown in figure 3.1-(a) and the different backbones that were tested are shown in figure 3.1-(b)

Myriam has defined a benchmark to test learning with few examples in Neuroimaging [Bontonou *et al.* 2020], based on the Individual Brain Charting data set. Myriam showed that the sparse learning techniques developed in computer vision were usable to decode brain activity with only 1 to 5 learning examples. Myriam also tested GNN models that perform a single diffusion using a consensus anatomical graph (same one than we used in [Brahim & Farrugia 2020]), but the experiments with GNN yielded worse performances than simple MLP architectures. In another contribution, Myriam proposed a supervised learning



(a)



(b)

Figure 3.1: (a) During the training phase, the backbone learns rich representations from the abundant examples of base classes. During the evaluation phase, the backbone is adapted to new classes using only a few examples. (b) Architectures of the feature extractors. FC stands for fully connected layer and Conv for convolutional layer with 1×1 kernels. Adapted from [Bontonou *et al.* 2020]

method inspired by Linear Discriminant Analysis, but using the same graph to model the dependencies between dimensions, replacing the covariance estimation. This method was particularly efficient in cases where few examples are available for learning, in which case using the structure of the graph as a prior instead of covariance led to significant boosts for brain activity decoding [Bontonou *et al.* 2021].

Work in my team is currently ongoing on few shot learning for neuroimaging ; in particular, the PhD project of Yassine El Ouahidi, a PhD student who started in October 2021 and that I contribute to advise, together with Vincent Gripon, Giulia Lioi and Bastien Pasdeloup. Yassine will investigate few shot learning for Brain Computer Interfaces based on EEG.

3.2.3.2 Graph Convolutions using ChebNet

I am still developing the theme of combining graph knowledge with deep learning as an efficient way to integrate temporal and spatial dynamics for functional MRI decoding models, such as the work in collaboration with Yu Zhang [Zhang *et al.* 2021] in Hangzhou, China, and Pierre Bellec at the Université de Montréal. Figure 3.2 presents the proposed decoding model using ChebNet. ChebNet layers approximate the calculation of a graph convolution using high-order Chebyshev polynomials, therefore integrating signals from direct and indirect neighbours of each node. Our results revealed that between-network integration significantly boosted the decoding of high-order cognition such as visual working memory tasks. The representations learned through graph convolutions not only improved the functional alignment of brain responses across subjects, but also preserved inter-subject variability in brain organization during task, indicating shared genetic variance with in-scanner behavioral performance.

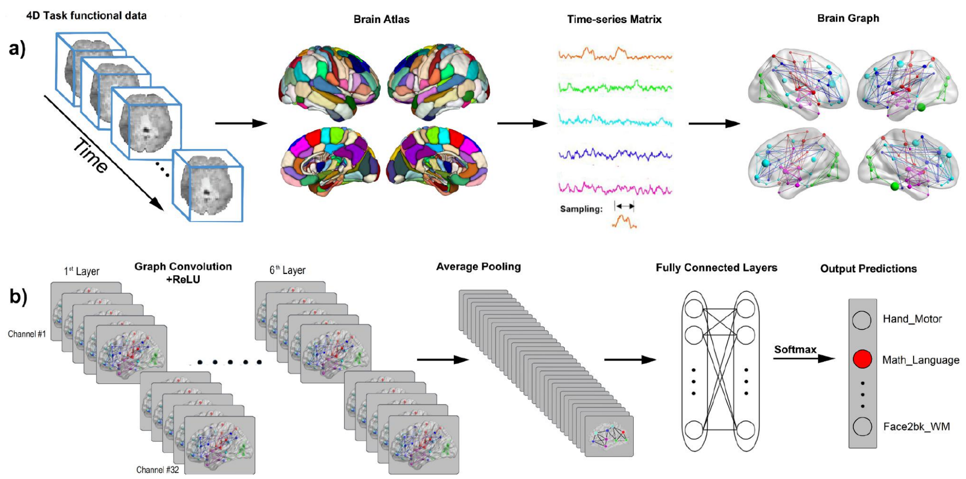


Figure 3.2: The decoding model consists of six ChebNet graph convolutional layers with 32 graph filters at each layer, followed by a flatten layer and 2 fully connected layers. Specifically, for a short series of fMRI volumes, the measured brain activity was first mapped onto a predefined brain atlas consisting of hundreds of brain regions. A functional graph was then constructed by calculating group-averaged resting-state functional connectivity for each pair of brain regions. Next, a new representation of task-evoked neural activity was generated through a multi-layer graph convolutional network, taking into account the segregation of localized brain activity and information integration among multiple brain networks. These representations were then used to predict the corresponding cognitive state associated with the short time window. The implementation of the ChebNet graph convolution was based on PyTorch 1.1.0, and was made publicly available in the following repository: `https://github.com/zhangyu2ustc/gcn_tutorial_test.git`. Adapted from [Zhang *et al.* 2021]

This work shows that graphs can be efficiently exploited to define efficient decoders of brain activity using the temporal dimension. In future work, we will investigate how this kind of approach can be extended to build efficient feature extractors for generalization in few-shot learning tasks, by using ChebNets with temporal signals to train a backbone for few-shot learning, similarly as the work from Myriam Bontonou described in section 3.2.3.1.

3.2.3.3 Selecting meaningful graph frequencies for cognitive decoding by Pruning Spectral Residual Deep Networks

Another way to define graph convolutions is to compute them directly as multiplications in the spectral domain. This idea has been tested by Yassine Ouahidi in his first paper [Ouahidi *et al.* 2022]. Following the usual setting of Fourier Analysis, convolution can be defined as entrywise multiplication $\odot$ in the spectral space. In more details, denote $\mathbf{h} \in \mathbb{R}^N$ a filter, then convolution is obtained as:

$$\mathbf{x} * \mathbf{h} = \mathrm{GFT}^{-1}\left(\mathrm{GFT}(\mathbf{x}) \odot \mathrm{GFT}(\mathbf{h})\right) \ . \tag{3.1}$$

Figure 3.3-(a) shows the architecture of the Spectral ResNet model that uses the proposed GSPConv blocks. Each GSPConv block contains a input and b output channels, resulting in ab trainable convolutional filters. Yassine compared this architecture with simple Multi Layer

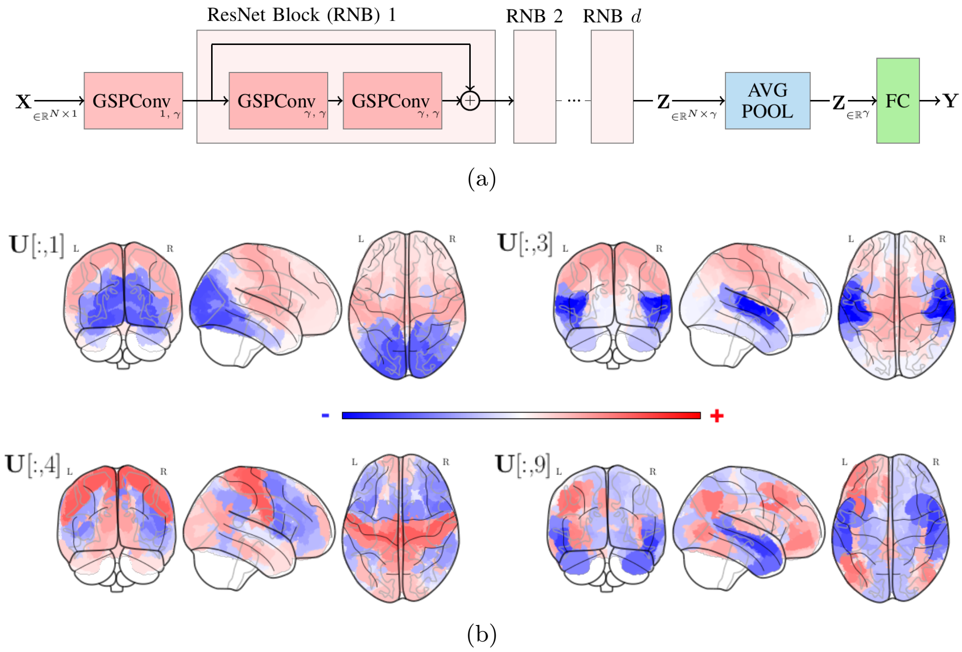


Figure 3.3: (a) Illustration of the Spectral ResNet model. The input is seen as a graph signal with 1 channel, it is first embedded as a graph signal with γ channels using a GSPConv layer. Then it goes through d ResNet blocks consisting of the sum between a shortcut path and a sequential path containing 2 GSPConv layers. The output is seen as a graph signal with γ channels. The values are averaged over the graph before being fed to a classical logistic regression. (b) The four most important graph frequencies selected by pruning, projected back on the brain for interpretation. Adapted from [Ouahidi *et al.* 2022]

Perceptrons (MLP), on two large benchmarks for fMRI decoding : activation maps from the Human Connectome Project (HCP, more than 800 subjects but with few tasks), and the Individual Brain Charting project, previously used by Myriam Bontonou, which contains many different tasks and multiple sessions on a sample of 13 subjects.

The main idea tested by Yassine is a combination of many different aspects that I developed in this manuscript : Brain Graphs, Deep Learning, and pruning deep networks. However, here pruning was used not to reduce the parameter count, but to provide an interpretation on which frequencies are the most important for cognitive decoding, shown in Figure 3.3-(b). Spatial distributions of largest positive and negative values of $\mathbf{U}[:,1]$ and $\mathbf{U}[:,3]$ correspond respectively to occipital and superior temporal brain areas, suggesting contributions of primary visual and auditory systems. $\mathbf{U}[:,4]$ most positive values are confined to the motor cortex, while $\mathbf{U}[:,9]$ includes medial contributions of the default mode network [Smallwood *et al.* 2021]. These results show how the proposed methodology is able to identify meaningful graph frequencies (interpreted here as spatial patterns) for brain activity decoding. As we consider the graph frequencies common to all the tasks combined, general purpose brain patterns were found, consistent with the literature [Mensch *et al.* 2021].

3.3 Sound and the brain in ecological contexts

3.3.1 Ecological paradigms and naturalistic stimuli

Almost all research in auditory cognitive neuroscience is based on the use of controlled behavioral paradigms, tested under laboratory conditions, in which human subjects selected on often limiting criteria participate. This is due to an adherence to a fundamentally reductionist postulate, which attempts to isolate a small number of factors influencing the object of study. Therefore, the main challenge of this research axis is to break free from these usual limitations of cognitive neuroscience research, by introducing the use of ecological paradigms, of naturalistic stimuli, as well as by proposing out-of-lab experiments.

A behavioral paradigm whose goal is to approximate as closely as possible the natural conditions that a subject might encounter will be called an ecological paradigm. Such paradigms are often based on natural stimuli, i.e. stimuli that are close to those that can be perceived in everyday life. The opposition between the "classical" experiment with controlled stimuli and paradigm, and the "ecological" experiment is illustrated in the table 3.1.

Cognitive function	Controlled experiment	Ecological experiment
Spatial navigation	2-way alternative forced choice between a limited amount of directions and locations	Free navigation a videogame-like environment
Music Perception	Short computer generated musical excerpts	Studio recorded music by professional musicians

Table 3.1: Examples of controlled and ecological experiments / stimuli

The advantage of classical paradigms (which is also their disadvantage!) lies in the fact that their analysis is essentially based on statistical inference tests, which make it possible to validate or invalidate the hypotheses posed by the controlled paradigm. The difficulty with ecological paradigms lies in the double complexity introduced by the variability of the stimuli and the behavior. In this research project, I propose to build on recent advances in deep learning to model this double complexity, and then to test new hypotheses on brain function and human behavior. In this regard, I recently co-authored a commentary article [Herholz *et al.* 2021] arguing the need to combine ecological paradigms, naturalistic stimuli and machine learning with the Neuromod project, described in the next paragraph.

3.3.2 Computational models of auditory neural representations

I collaborate with Pierre Bellec's team at the University of Montreal, who is in charge of the Courtois Neuromod project of brain activity modeling. Since the end of 2019, I am co-supervising the PhD project of Maëlle Freteault, whose objective is twofold. First, (1) Maëlle studies auditory perception and auditory imagery using fMRI with (long) naturalistic stimuli, extending and replicating previous work, and (2) Maëlle investigates whether brain activity can be used to finetune artificial neural networks, and how such finetuning can lead to better (i.e. more generalizable) representations. In order to address those questions, Maëlle has joined the Neuromod project. The Neuromod project measures brain activity in fMRI during video viewing (films, series) and video game playing during fMRI, on a limited number of subjects (6 human subjects) but with large quantities of acquisitions (to date, several hundred hours per subject, measured in weekly fMRI sessions). Maëlle builds on our preliminary work on transfer learning to model fMRI activity [Farrugia *et al.* 2019], and has replicated these results on Neuromod videos [Freteault *et al.* 2020].

Figure 3.4 presents the approach and first results obtained. This work shows that it is possible to model the brain activity associated with the soundtrack of about 40 hours of movie / series from a deep neural network, as shown by the good predictions made by the deep network, achieving a R^2 score (on a left-out test set of about 10 hours) around 0.4 with peak values in voxels close to the primary auditory cortex (figure 3.4, bottom panel, left subpanel). We also show that fine-tuning SoundNet can lead to increased performances in prediction, as demonstrated by higher peak values (maximum gain in R^2 : 0.05) and more voxels close to the maximum value (figure 3.4, bottom panel, right subpanel). Maëlle is currently confirming these results on all six subjects of the neuromod project, and will study the individual differences in spatial distribution of fine-tuning gains. The next step is to interpret internal representations of the initial (SoundNet) and finetuned versions. Our goal is to investigate how finetuning an artificial neural network with brain activity during naturalistic stimulation can reshape (i.e. regularize) its internal representations to increase generalization. We will benchmark this increase in generalization using the HEAR benchmark, described in section 3.4.2. Finally, Maëlle will run experiments on musical imagery during video game playing, by inserting silences during the game. The rationale taken is to test whether the trained neural encoding model can predict brain activity during silences. Using this paradigm, we directly test the hypothesis whether the brain actively simulates auditory perception during auditory imagery.

Two other PhD projects in Auditory Cognitive Neuroscience have recently started, on which I am a supervisor. The PhD project of Emma Ducos started in September 2021 and is directed by Luc Arnal at the Institut de l'Audition in Paris. Emma aims at establishing accurate models of neuronal responses to complex auditory stimuli (click trains), in healthy populations as well as patients. Tom Colas started his PhD project in September 2021, and is directed by Mathieu Paquier (Université de Brest) and co-supervised by Etienne Hendrickx (Université de Brest) and myself. Tom works on the perception of binaural sound, and in particular the phenomenon of externalization, i.e. the illusion of perceiving a distant source while the sound is presented on headphones. Tom uses EEG and machine learning to try to characterize the externalization phenomenon when natural stimuli are presented.

3.3.3 Studying the subjective and neural state during musical improvisation

As argued in section 3.3.1, recent advances in ML make it possible to consider ecological paradigms and tackle the increased complexity of experimental setups and measurements. In the context of a science and arts collaboration, I initiated what became an informal collaboration with the Ensemble Nautilis, an improvised free music ensemble based in Brest. The leader of Nautilis, a clarinet player named Christophe Rocher, was initially trained as an electrical engineer. We developed this collaboration since late 2018, which culminated in several public engagement events in France and Canada. We also setup a scientific environment to study musical improvisation under ecological conditions, and published a first paper [Farrugia *et al.* 2021].

Figure 3.5 presents the proposed ecological and controlled paradigms to study musical improvisation from the point of view of an improviser. This contribution is a single subject study attempting at a better understanding of the subjective mental state during musical improvisa-

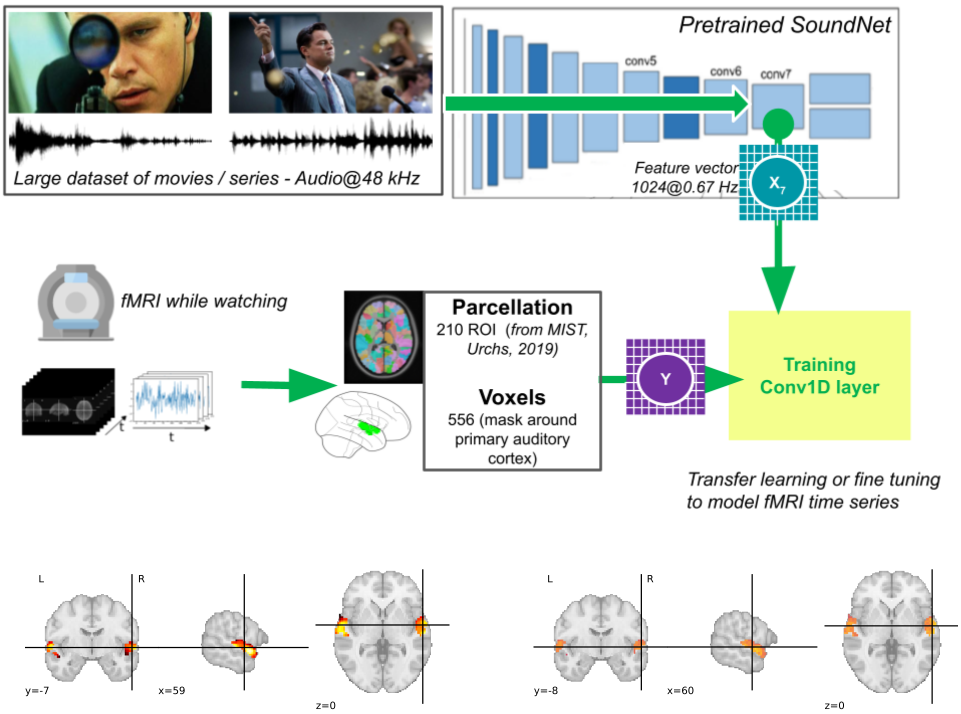


Figure 3.4: Transfer learning and fine-tuning deep neural networks to model brain activity during naturalistic movie watching. Top Panel : principle of the approach. Subjects' brain activity is measured using fMRI while they watch movies, spanned within several months of weekly fMRI sessions (about 45 hours of watching in total). Audio waveform is extracted from the stimuli, and fed into a pretrained 1D-convolutional neural network, SoundNet. Preprocessed fMRI data is parcellated either using 210 ROI, or using a mask of 556 voxels in bilateral auditory cortex. A single layer of 1D convolutions is trained to predict the fMRI signal, in two settings : transfer learning, in which only the final 1D-conv layer, and fine tuning, where the soundnet layers are also optimized. Bottom panel : Spatial distribution of performance between predicted and measured fMRI signals. Left subpanel corresponds to the R^2 scores in the transfer learning scenario, while right panel corresponds to the gain obtained by fine-tuning. (Manuscript under preparation, Freteault, Bellec and Farrugia)

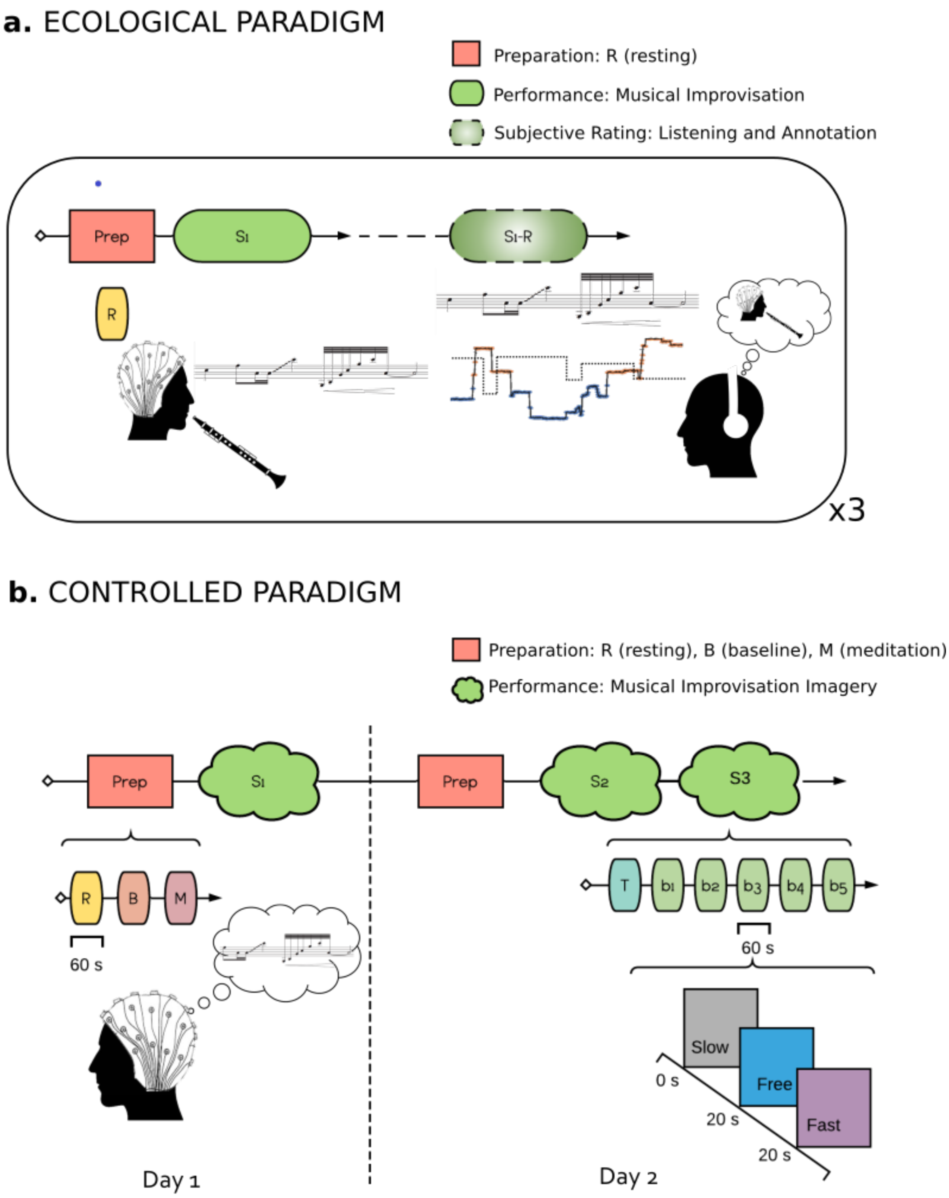


Figure 3.5: a. Ecological Paradigm to study subjective states of a musical improviser. The experiment included two parts. In the first, EEG was recorded while the subject performed musical improvisation. In the second part, the subject listened to his own performance and performed a retrospective rating, measuring for subjective Flow State, and subjective temporal resolution (STR). b. Controlled Paradigm. The experiment was carried out in two days. In the first, the subject underwent a Preparation session where he performed 60 s of Resting (Eyes Opened), 60 s of Baseline and 60 s of Meditation. He then performed a musical improvisation imagery task with a Slow, Fast or Free conditions. The second part (two days later) was as the first, with the exception that two training sessions were performed. Adapted from [Farrugia *et al.* 2021]

tion. In a first experiment, we setup an ecological paradigm measuring EEG on a musician in free improvised concerts with an audience, followed by retrospective rating of the mental state of the improviser. We defined Subjective Temporal Resolution (STR), a retrospective rating assessing the instantaneous quantization of subjective timing of the improviser. We identified high and low STR states using Hidden Markov Models in two performances, and were able to decode those states using supervised learning on instantaneous EEG power spectrum, showing increases in theta and alpha power with high STR values. In a second experiment, we found an increase of theta and beta power when experimentally manipulating STR in a musical improvisation imagery experiment. These results are interpreted with respect to previous research on flow state in creativity, as well as with the temporal processing literature. We suggest that a component of the subjective state of musical improvisation may be reflected in an underlying mechanism related to the subjective quantization of time. We also demonstrated the feasibility of single case studies of musical improvisation using brain activity measurements and retrospective reports, by obtaining consistent results across multiple sessions.

In future work, I plan to apply similar ideas (retrospective ratings, first person reports, together with other objective measures) in order to study mental states of improvisers as well as audience members. We have run a first pilot study on audience members listening to a public rehearsal of Nautilis while being measures with EEG, and used a tapping paradigm to study beat perception in polyrhythmic contexts played by the ensemble. More generally, the overarching goal of this line of work is to propose ecological ways to study interactions and idiosyncracies in subjective time, music perception and music imagery, and self-generated thoughts.

3.4 New paradigms and applications in audio intelligence

3.4.1 Rethinking deep learning for efficient audio processing

In section 2.2.3.5, I have presented a starting point in audio deep learning research in which I argue about using raw waveform to train efficient deep networks with few parameters [Pajusco *et al.* 2020]. The results obtained are promising, and call to be extended with more difficult benchmarks, larger datasets, to be compared with other competitive approaches in the field. In particular, I am particularly interested in exploring data augmentation or regularization techniques that are specific to the auditory domain, by exploiting properties of acoustic signals as well as inductive biases from previous work in auditory perception and psychoacoustics. Another currently ongoing work is the PhD thesis of Matteo Zambra, in collaboration with Ronan Fablet (Full professor, IMT Atlantique, PhD director) and Dorian Cazau (Associate Professor, ENSTA Bretagne, Brest), in the field of remote sensing and geoacoustics. Matteo uses deep learning models to predict wind speed using underwater passive acoustics. We use a variational model that combines assimilation methods with automatic differentiation to learn how to inverse a geophysical system [Fablet *et al.* 2021], and Matteo has proposed new architectures to learn temporal sequences of wind speed on a large amount of data. This contribution shows how to directly attempt at inversing the underlying physical model using deep learning in order to answer a question in geoacoustical modeling.

Another interesting new perspective for deep learning in general, and audio deep learning in particular, is Self Supervised Learning (SSL). The idea underlying SSL approaches is to use the data to create supervised tasks, for example by applying data augmentation to two vectors coming from either the same data point or different ones, and train a network to produce representations that can detect when the two inputs are different. It has been shown that SSL can be used to train representations that generalize well to supervised tasks; see [Niizumi *et al.* 2021] for an example of a recent SSL approach for audio. The PhD project of Ilyass Moummad started in December 2021, and is targeted towards this idea of exploring SSL for audio, with an application in diagnosis-aid for detecting respiratory pathologies using audio, in collaboration with the company OSO-AI. In particular, one possibility would be to explore how new data augmentation techniques can be used and combined to imagine innovative SSL paradigms.

3.4.2 Holistic evaluation of auditory representations (HEAR)

The HEAR benchmark was introduced at the NeurIPS 2021 conference in the context of a challenge. From the official website (`www.neuralaudio.ai`), "What audio embedding approach generalizes best to a wide range of downstream tasks across a variety of everyday domains without fine-tuning? The aim of the HEAR 2021 NeurIPS challenge is to develop a general-purpose audio representation that provides a strong basis for learning in a wide variety of tasks and scenarios. HEAR 2021 evaluates audio representations using a benchmark suite across a variety of domains, including speech, environmental sound, and music.". I will contribute to this research effort, by testing two lines of ideas :

- Can we use brain activity to train deep neural networks with internal representations that generalize well to everyday domains ? This idea will be tested in the context of the PhD of Maëlle Freteault, described earlier in this chapter.
- Can we define inductive biases from Auditory Cognitive Neurosciences to define better representations ? More precisely, is it possible to push representations to mimic neural mechanisms of auditory imagery, or predictive / temporal processing ? This is a new line of work that will necessitate to go back and forth between new cognitive neuroscience experiments, and computational modeling using methods described in this manuscript. This idea is more long-term, and some of its starting points will be explored in the context of the PhD of Ilyass Moummad.

3.4.3 Documenting complex socio-ecosystems using eco-acoustics

Since 2018, I have been collaborating with Samuel Challéat, researcher in environmental geography at the GEODE laboratory (UMR CNRS) in the framework of the Observatoire Hommes Milieux (OHM) Pyrénées [Challéat *et al.* 2018]. We use deep learning for sound recognition to characterize socio-ecosystems, i.e. the interaction between human activity and other animal species. In 2020, we launched the Silent Cities [Challéat *et al.* 2020] program, whose goal is to document the exceptional state of cities during COVID19's containment measures. We engaged the international community of eco-acoustic researchers, who had programmable recording devices, to record sound in urban environments according to a standardized protocol. In total, we collected more than two million minutes of recordings, 24 hours a day in 50 different countries on about 400 separate recorders (figure 3.6). In collaboration with eco-acoustic researchers (Amandine Gasc, IRD Aix en Provence and Jeremy Froideveaux, CESCO, MHN, Concarneau), as well as a master student in acoustics (Nicolas Pajusco, LAUM, Université du Mans), we will use audio signal processing and pretrained deep networks [Kong *et al.* 2020] in order to address research questions in ecology such as :

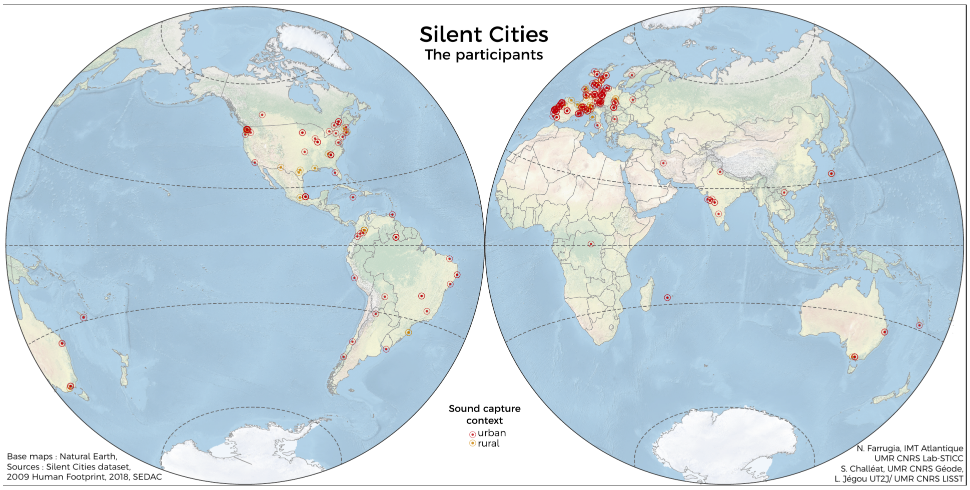


Figure 3.6: Contributors of the Silent Cities project

- Studies of biodiversity during and after 2020 lockdown in urban areas or during and after specific local political decisions,
- Lockdown effects along urbanization gradient: urban, semi-urban and rural areas,
- Impact of the type of environment (tropical, temperate etc..) on the lockdown effect: how is the soundscape affected around the world?

These measurements also open interesting perspectives for defining datasets for machine learning with specific properties that differ from traditional academic datasets ; as the measurements are continuous, they are neither identically distributed nor independent, can be highly correlated and confused between similar times and geographical sites, while they also have remarkable individual differences due to the variety of recorder placements, climates around the world or time period of measurement. As the measurements were done during the application of lockdown measures, there is a sometimes gradual, sometimes very brutal effect on soundscapes, and this effect may sometimes only be seen in long term temporal dependencies (days / weeks).

This project is contribution to Open Science ; we are currently writing a journal article (the submission of a data paper is planned for late 2022) describing the Silent Cities measurements, as well as code, preprocessed data with a wide range of acoustic descriptors, and results (outputs and internal representations) from inference using a pretrained deep network [Kong *et al.* 2020], in order to accelerate further research in the aformentionned questions for the global academic community.

Appendices

Appendix A

Extended CV

Contents

A.1 Resume

A.1.1 Academic training

PhD	Université de Bourgogne	Computer science	2008
M.Sc.	Université de Cergy Pontoise	Signal and Image Processing	2005
M. Engineering	ENSEA, Cergy Pontoise	Electrical Engineering	2005

A.1.2 Professional appointments

2016 – today	Associate Professor, IMT Atlantique, Brest, France.
2015 – 2016	Postdoctoral fellow, Telecom Bretagne, Brest, France.
2013 – 2015	Postdoctoral fellow, Golsmiths, Londres, Angleterre.
2012 – 2013	Postdoctoral fellow, Max Planck Institute, Leipzig, Allemagne
2010 – 2012	Postdoctoral fellow, WsFIz Varsovie, Pologne.
2008 – 2010	R&D engineer, CSSI, Paris, France.
2005 – 2008	PhD student, Orange Labs, Meylan, France.
2005 – 2008	PhD student, Université de Bourgogne, Dijon, France.
2005	Research Intern (Msc thesis), LE2I, Dijon, France.

A.2 Student supervision

A.2.1 PhD students

Previous students :

- Ghouthi Boukli, defended in Octoner 2019 [Boukli *et al.* 2018b, Boukli *et al.* 2018c, Boukli *et al.* 2019c, Boukli *et al.* 2019a, Boukli *et al.* 2019b, Boukli *et al.* 2020]
- Majd Abdallah, defended in September 2020 [Abdallah *et al.* 2021, Abdallah *et al.* 2020]
- Myriam Bontonou, defended in December 2021 [Bontonou *et al.* 2019, Bontonou *et al.* 2020, Bontonou *et al.* 2021]

Current students :

- Maëlle Freteault, started November 2019 [Freteault *et al.* 2020]
- Matteo Zambra, started November 2020
- Emma Ducos, started September 2021
- Tom Colas, started October 2021
- Yassine El Ouahidi, started November 2021 [Ouahidi *et al.* 2022]
- Ilyass Moummad, started December 2021
- Venkatesh Subramani, started January 2022

A.2.2 Postdotoral fellows

- Mathilde Ménoret, September 2016 to October 2017 [Ménoret *et al.* 2017].
- Abdelbasset Brahim, May 2018 to April 2020 [Brahim & Farrugia 2020, Brahim *et al.* 2019].
- Giulia Lioi, May 2020 to December 2020 [Farrugia *et al.* 2021, Bontonou *et al.* 2020, Lioi *et al.* 2021].
- Ghaith Bouallegue, since September 2021.

A.2.3 Master thesis supervision

- Venkatesh Subramani, June to November 2021
- Ludwig Segalen, January to June 2021
- Imene Djellad, March to August 2017
- Xueyao Ji, March to August 2022

A.2.4 Research internships

- Jules Bouvet, January to March 2021 [Farrugia *et al.* 2021]
- Gilles Schneider, March to June 2020
- Alix Lamouroux, June to July 2020 [Farrugia *et al.* 2021]
- Min Tri Truong, June to July 2020
- Imad El Kahoui, June to July 2020
- Nicolas Pajusco, January to March 2020 [Pajusco *et al.* 2020]
- Richard Huang, February to May 2020 [Pajusco *et al.* 2020]
- Victor Nepveu, August to October 2019 [Farrugia *et al.* 2019]
- Camila Deyci Vilamil, June to July 2019 [Farrugia *et al.* 2019]
- Thomas Dambricourt, June to July 2018
- Charlène Petit, June to July 2018
- Mariem Khlifi, July to August 2018
- Aymane Abdali, June to July 2018
- Nicolas Hérault, June to July 2018
- Matthieu Bachelot, July to August 2018
- Yusuf Yigit Pilavci, July to October 2018 [Pilavci & Farrugia 2019].
- Lison Blondeau-Pâtissier, July to August 2017
- Amine Echraibi, July to August 2017 [Farrugia & Echraibi 2018]
- Tristan Stérin, July to August 2016 [Stérin *et al.* 2017]
- Martin Dornier, July to August 2016
- Martin Guy, July to August 2016

A.3 Research Funding

A.3.1 Public funding

Table A.1 provides an exhaustive list of acquired public funding. For all these projects, I was the sole principal investigator. These projects contributed to funding postdoctoral fellows as well as co-funding for doctoral students. Funding for two other PhD students not mentionned in this table (Ghouthi Boukli Hacene, Majd Abdallah) was acquired by respective PhD directors (Michel Jezequel and Sandra Chanraud).

A.3.2 Funding from private institutions

Table A.2 is an exhaustive list of funding acquired from private instituions. For some of these projects, the responsability was shared with Vincent Gripon (Research Director, IMT Atlantique) as a co-principal investigator.

Année	Titre	Organisme	Montant (keuro)
2018	Graph Signal Processing	Britanny Region	86
2019	Neuro-inspired artificial intelligence	Britanny region	86
2019	Funding (half) for PhD (cotutelle with Université de Montréal) of Maelle Freteault - Deep learning models of auditory system (as advisor, director Michel Jezequel)	Britanny	56
2020	Multimodal Graph Signal Processing (MultiGSP)	Britanny and Finistère	76
2021	Funding (half) for PhD (cotutelle with Université de Montréal) of Venkatesh Subramani - Multimodal Graph Signal Processing and Deep Learning	IMT Atlantique	56
2021	Funding for PhD project of Ilyass Moumad - Deep learning for diagnosis aid of respiratory pathologies using audio	ANR	58

Table A.1: List of public funding obtained

Année	Titre	Organisme	Rôle	Montant (keuro)
2017	Automatic auditory scene summary with machine learning	Orange Labs	PI	30
2018	Novelty detection using associative memories	Thalès	co-PI	150
2019	Few shot learning	Thalès	co-PI	150
2020	Deep learning for diagnosis aid of respiratory pathologies using audio.	OSO-AI	co-PI	5
2021	Deep learning for diagnosis aid of respiratory pathologies using audio	OSO-AI	PI	58

Table A.2: Private funding acquired

A.4 Teaching

In this section I give an overview of my teaching activities. I have created and am responsible for two courses in artificial intelligence, described in sections A.4.1 and A.4.2. These two courses are taught by myself and Mathieu Leonardon, Giulia Lioi, Bastien Pasdeloup, Pierre-Henri Conze and Lucas Drumetz (all Associate Professors, IMT Atlantique). I have also created a short overview to cognitive neuroscience using fMRI targeted for students in speech therapy (sections A.4.4 and A.4.5) as well as a similar course for clinical neuroscience students (section A.4.6). I have also put together a short introduction to electrophysiological signal processing for engineers (A.4.3), as part of a larger course in healthcare engineering (led by Pierre-Henri Conze).

I also describe a few courses to which I contribute as a teacher while not being actively involved in designing the pedagogical content. Section A.4.7 gives a short summary of the Pyrat course, created by Bastien Pasdeloup and Vincent Gripon, and to which I contribute as a teacher, as I have helped in creating the online version of the course. Finally, I contribute to teaching digital electronics (section A.4.8), led by Charlotte Langlais (Full Professor, IMT Atlantique).

A.4.1 Introduction to Artificial Intelligence

Motivations Artificial Intelligence is a field that is quickly developing due to the recent impressive progress in Deep Learning. Applications are booming in every field of human society. Therefore, following this course will help a future engineer in becoming familiar with an engineering technique that is likely to become ubiquitous.

Keywords artificial intelligence, machine learning, reinforcement learning, supervised learning, unsupervised learning, game theory

Content The goal of this course is to give a general overview of the field of Artificial Intelligence. We thrive in giving students a clear sense of the difficulties in the field, by giving introductory lectures on the broad topics :

- Supervised Learning
- Unsupervised Learning
- Practical Ethics in AI
- Combinatorial Game Theory
- Reinforcement Learning

For each topic, we will focus on delivering the bigger picture, instead of doing a catalog of all possible methods. Students work in pairs during mini-projects where each pair focus on one particular method, and will present it to the whole class. The entire course is also structured around a Final Challenge, during which students can apply everything they have learnt during the course.

Course length 35 hours, in seven sessions.

Prerequisites Python programming, basics of linear algebra. This is an introductory course, the goal is that every student can take it, whatever their background.

Learning outcomes Students will have a general understanding of the field of modern artificial intelligence, and will be able to analyze a problem, and decide which machine learning solution should be best adapted to adress this problem.

Rather than going through a catalogue of AI methods, students are asked to each pick a different method, and prepare a presentation on this method. Each student is working on the same problem and the same dataset. The goal is to have everyone be familiar with the most common methods, while keeping a very general view on the methods.

Evaluation Short written evaluations at the beginning of each session. Oral presentations (4 in total), and technical evaluation of the projects.

A.4.2 Efficient Deep Learning

Motivations The applications of artificial intelligence on objects constrained in computing resources and energy are becoming more and more numerous, and will increase sharply in the years to come. In addition to SmartPhones, we can also mention Internet access boxes, network equipment, connected glasses or virtual reality helmets. Thus, the skills developed in this course are actively sought after in the industry.

Keywords artificial intelligence, deep learning, compression, quantization, distillation, factorization, embedded systems.

Content The objective of this course is to study and implement different techniques to reduce the complexity of deep neural networks. The course is organized around a long project (participation in a deep learning challenge), each session aims to study a family of techniques. The students present the results of their experiments to the other groups in lectures, and a final defense is made about the solution found to meet the challenge of the competition. This course is highly technical with a lot of time given to the projects.

Prerequisites Previous experience in deep learning (through one of the other UE Deep Learning, Introduction to AI, through one of the Core course in Machine Learning or a significant internship experience). This is expected from students taking this course:

- I know the main principles of machine learning; cross validation (training and test datasets), bias variance trade-off, dimension curse
- I know how to define a neural network architecture with a standard framework (preferably pytorch)
- I am fluent with python programming

Course materials available here: https://github.com/brain-bzh/ai-optim

Course length 35 hours, in seven sessions.

Learning outcomes By the end of the course, engineering students are able to dimension a deep learning / artificial intelligence solution by considering the constraints of the systems, and to validate the performances of such a solution.

Course organization The course is structured around a long project (the MicroNet Challenge on CIFAR100) during which students have to design a deep neural network with the least number of parameters and memory while keeping a performance better than a baseline model. Different optimization techniques are introduced each week : quantization, pruning, factorization, distillation. A lot of time is devoted to the project.

Evaluation Short written evaluations at each session, oral presentations and technical evaluation of the proposed solution.

A.4.3 Biomedical Signal Processing

This course is part of a larger course (Digital medicine, 35 hours long), led by Pierre-Henri Conze (Associate Professor, IMT Atlantique). I give a short overview of acquisition and analysis techniques for electrophysiology. I cover electrocardiography, electromyography, and electroencephalography. Students are taught with common analysis techniques, from preprocessing up to an introduction to advanced analysis techniques using modern signal processing and machine learning.

Course Materials Available here: `https://github.com/brain-bzh/health_exg`

Course Length 5 hours, in three sessions.

A.4.4 Overview of functional MRI

This course is part of a larger course (Neurosciences 2, 20+ hours long), led by Romuald Seizeur (Université de Brest, Faculty of Neurosurgery), target for students in speech therapy. The course is focused on informal explanations of concepts and applications.

The goal of this course is to provide a short overview of functional MRI acquisition and analysis techniques, focusing on applications in cognitive neuroscience. After having presented the main characteristics of fMRI, I present the classical analysis technique using the General Linear Model. In the second session, I provide an informal introduction to the study of functional connectivity (resting state) and the applications of machine learning to cognitive neuroimaging.

Course Length 3 hours, in two sessions.

A.4.5 Brain plasticity

This course is part of a larger course (Neurosciences 2, 20+ hours long), led by Romuald Seizeur, target for students in speech therapy. The course is focused on informal explanations of concepts and applications.

This course is taught together with François Rousseau (Full Professor, IMT Atlantique). The goal of the course is to introduce the concept of brain plasticity, and explain how it can be measured using anatomical MRI and functional MRI, using morphometry studies and resting-state. After giving a short overview in two lectures, students gather by groups (4 to 5 students per group) and have to present a research paper on brain plasticity.

Course Length 3 hours, in one session.

A.4.6 Functional MRI in clinical studies

This course is part of a one-day long course on Translational Research, led by Serge Timsit (Full Professor, MD, Université de Brest, Faculty of Neurology). This course is similar to the one described in section A.4.5, but focuses on clinical applications of functional connectivity studies as well as machine learning for precision medicine.

Course Length 2 hours, in one session.

A.4.7 Advanced algorithmics and Graph Theory

Pyrat is an algorithmics and graph theory course designed for first year engineering students at IMT Atlantique. It was designed by Vincent Gripon and Bastien Pasdeloup, and I joined the pedagogical team from 2016. PyRat is structured around a maze game, in which students program a character (a rat or a python) to go grab pieces of cheese faster than an opponent. Students are guided to achieve this difficult goal, by progressively reaching smaller objectives of gradually increasing difficulty. The course is designed to introduce concepts of graph theory, graph traversal, heuristic algorithms and computational complexity.

In 2018, we have designed an online version of this course and have shot a series of videos. I am one of the teachers that contributed to this online version and feature in the videos, together with Vincent Gripon and Patrick Meyer (Full professor, IMT Atlantique).

Course Materials Available here: https://formations.imt-atlantique.fr/pyrat/, and youtube videos here https://www.youtube.com/watch?v=hoDLCKqEDR4&list=PLjXls-kqM6JAbFBmFPI4UlIRN6ToZLibk

Course Length 35 hours, 7 sessions.

A.4.8 Digital electronics

The objective of this course (led by Charlotte Langlais) is to provide fundamentals in digital electronics, to students in first and second year IMT Atlantique. Students are a particular cohort who are part-time employees in private companies (so called "apprentices"), but eventually graduate with the same diploma as the "initial training" engineering students. This course is divided into two levels, for first and second year students respectively. First year students are taught with an overview of digital number representations, combinatorial logic, boolean operations, Karnaugh tables, canonical forms and arithmic operators. Second year students are taught with sequential logic design : latches and flip flops, synchronous logic and finite state machine design.

Course Length 42 hours in 7 sessions, per year.

Figure A.1: Inaugural event of national Brain Awareness Week, Grenoble, March 14th 2022, with the Brain Songs Project (see section 3.3.3)

A.5 Other synergestic activities

A.5.1 Public Engagement activities

Here I describe my various activities for public engagement and scientific outreach. Artificial Intelligence and Neuroscience are two topics that are particularly suited for such activities, because at the same time they attract a lot of fascination and curiosity, and also raise important ethical concerns.

2022

- Brain Songs, Conference / Concert on the neuroscience of musical improvisation. Estran, Guidel, France. April 2022.
- Main speaker for the inaugural event of the national Brain Awareness Week. Brain Songs, Conference / Concert on the neuroscience of musical improvisation. Grenoble, France. March 14 2022. Available online: https://youtu.be/Lpy_6qJgR1Y. Photo in Figure A.1.

2021

- I co-organised with Vincent Gripon a seminar on the ethical and societal challenges on Artificial Intelligence. This seminar was held online on March 11th 2021, because of the restrictions due to Covid19. This seminar consisted in a general presentation of the current state of Artificial Intelligence by Nicolas Courty (Université Bretagne Sud), followed by a roundtable with invited speakers from the private sector (Olivier Menut, OSO AI and Emilie Sirvent-Hient, Orange Labs), a former professor and former city council elected member (Michel Briand, emeritus professor IMT Atlantique, Brest) and a researcher in work sociology (Yann Ferguson, ICAM, Toulouse). The seminar is now visible online (in French) : https://www.youtube.com/watch?v=5esBl-51YX0

- Webinar on neurosciences and musical improvisation, organized by CESARE, Reims, April 2021. Available online : https://vimeo.com/542113705
- Soundscapes during Covid 19, the Silent Cities Project. "Sonference", April 2021. Available online : https://www.youtube.com/watch?v=9zJimdzEyNM&list=PLy_o1CrFrp0W3LhhBhqStsp07gakUBcK-&index=5&t=510s

2020

- Public conference on Music and Brain in a cider factory, Aerolithes, Tregor, July 2020.
- "Brain Songs, un projet art et sciences sur les états mentaux de l'improvisation musicale", Athénor, St Nazaire, juin 2020

2019

- Presentation on Involuntary Musical Imagery at the "Journeés Science et Musique", Rennes, November 2019. Available online : https://www.youtube.com/watch?v=sEmnH_iulG4
- Live demo of real-time object recognition at the "Nuit Des Chercheurs" September 2019, Oceanopolis, Brest.
- Brain Songs, Conference / Concert on the neuroscience of musical improvisation. The Neuro, McGill, Montreal, June 2019. Available online : https://youtu.be/ILhaZYtW8fs
- Brain Songs, Conference / Concert on the neuroscience of musical improvisation. Brest, March 2019.
- Public conference on Artificial Intelligence for "Université du Temps Libre", Brest, January 2019.

2018

- Real-time musical neurofeedback installation, "Nuit des chercheurs 2018", Oceanopolis, Brest.

2017

- Cinema - debate with the audience. Brain Awareness Week 2017, Brest, France.

2016

- "Musique Intérieure", solo concert and scientific conference. Brain Awareness Week Brest, March 14th 2016, Brest conservatory. Available online: https://www.youtube.com/watch?v=0odwd7aZI9g

Figure A.2: "Musique Intérieure", solo concert and scientific conference for Brain Awareness Week. March 14th 2016, Brest conservatory.

A.5.2 Scientific conferences and symposiums

- Co-organizer of the NEUROSTIC'2017 Conference in Brest, for the Research Group Neurostic, with Vincent Gripon. http://recherche.imt-atlantique.fr/brain/2017/10/26/french-journees-neurostic/
- Symposium on Involuntary Musical Imagery in Goldsmiths, University of London, 2015. Main organizer. https://www.musicmindbrain.com/earworm-symposium

Bibliography

[Abdallah *et al.* 2020] Majd Abdallah, Nicolas Farrugia, Valentine Chirokoff and Sandra Chanraud. *Static and dynamic aspects of cerebro-cerebellar functional connectivity are associated with self-reported measures of impulsivity: A resting-state fMRI study.* Network Neuroscience, vol. 4, no. 3, pages 891–909, 2020. (Cited on pages 39 and 68.)

[Abdallah *et al.* 2021] Majd Abdallah, Natalie M Zahr, Manojkumar Saranathan, Nicolas Honnorat, Nicolas Farrugia, Adolf Pfefferbaum, Edith V Sullivan and Sandra Chanraud. *Altered Cerebro-Cerebellar Dynamic Functional Connectivity in Alcohol Use Disorder: a Resting-State fMRI Study.* The Cerebellum, pages 1–13, 2021. (Cited on pages 39 and 68.)

[Bégel *et al.* 2017] Valentin Bégel, Charles-Etienne Benoit, Angel Correa, Diana Cutanda, Sonja A Kotz and Simone Dalla Bella. *"Lost in time" but still moving to the beat.* Neuropsychologia, vol. 94, pages 129–138, 2017. (Cited on page 11.)

[Benjamini & Hochberg 1995] Yoav Benjamini and Yosef Hochberg. *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal statistical society: series B (Methodological), vol. 57, no. 1, pages 289–300, 1995. (Cited on page 19.)

[Benoit *et al.* 2014] Charles-Etienne Benoit, Simone Dalla Bella, Nicolas Farrugia, Hellmuth Obrig, Stefan Mainka and Sonja A Kotz. *Musically cued gait-training improves both perceptual and motor timing in Parkinson's disease.* Frontiers in human neuroscience, vol. 8, page 494, 2014. (Cited on pages 9, 14, 15 and 16.)

[Bontonou *et al.* 2019] Myriam Bontonou, Carlos Lassance, Vincent Gripon and Nicolas Farrugia. *Comparing linear structure-based and data-driven latent spatial representations for sequence prediction.* In Wavelets and Sparsity XVIII, volume 11138, page 111380Z. International Society for Optics and Photonics, 2019. (Cited on pages 46 and 68.)

[Bontonou *et al.* 2020] Myriam Bontonou, Giulia Lioi, Nicolas Farrugia and Vincent Gripon. *Few-shot Learning for Decoding Brain Signals.* In arXiv preprint arXiv:2010.12500, 2020. (Cited on pages 53, 54 and 68.)

[Bontonou *et al.* 2021] Myriam Bontonou, Nicolas Farrugia and Vincent Gripon. *Graph-LDA: Graph Structure Priors to Improve the Accuracy in Few-Shot Classification.* arXiv preprint arXiv:2108.10427, 2021. (Cited on pages 54 and 68.)

[Boukli *et al.* 2017] Ghouthi Boukli, Vincent Gripon, Nicolas Farrugia, Matthieu Arzel and Michel Jezequel. *Finding all matches in a database using binary neural networks.* In Proc. COGNITIVE'17, pages 59–64, 2017. (Cited on page 35.)

[Boukli *et al.* 2018a] Ghouthi Boukli, Vincent Gripon, Matthieu Arzel, Nicolas Farrugia and Yoshua Bengio. *Quantized Guided Pruning for Efficient Hardware Implementations of Convolutional Neural Networks.* In arXiv preprint arXiv:1812.11337, 2018. (Cited on page 35.)

[Boukli *et al.* 2018b] Ghouthi Boukli, Vincent Gripon, Matthieu Arzel, Nicolas Farrugia and Yoshua Bengio. *Quantized Guided Pruning for Efficient Hardware Implementations of Convolutional Neural Networks.* In arXiv e-prints, pages arXiv–1812, 2018. (Cited on page 68.)

[Boukli *et al.* 2018c] Ghouthi Boukli, Vincent Gripon, Nicolas Farrugia, Matthieu Arzel and Michel Jezequel. *Transfer Incremental Learning using Data Augmentation.* Applied Sciences, vol. 8, no. 12, page 2512, 2018. (Cited on pages 35, 36 and 68.)

[Boukli *et al.* 2019a] Ghouthi Boukli, Vincent Gripon, Nicolas Farrugia, Matthieu Arzel and Michel Jezequel. *Budget restricted incremental learning with pre-trained convolutional neural networks and binary associative memories.* Journal of Signal Processing Systems, vol. 91, no. 9, pages 1063–1073, 2019. (Cited on pages 35, 36 and 68.)

[Boukli *et al.* 2019b] Ghouthi Boukli, Vincent Gripon, Nicolas Farrugia, Matthieu Arzel and Michel Jezequel. *Efficient Hardware Implementation of Incremental Learning and Inference on Chip.* In 2019 17th IEEE International New Circuits and Systems Conference (NEWCAS), pages 1–4. IEEE, 2019. (Cited on pages 35, 36 and 68.)

[Boukli *et al.* 2019c] Ghouthi Boukli, Vincent Gripon, Nicolas Farrugia, Matthieu Arzel and Michel Jezequel. *Efficient Hardware Implementation of Incremental Learning and Inference on Chip.* In arXiv e-prints, pages arXiv–1911, 2019. (Cited on pages 35 and 68.)

[Boukli *et al.* 2020] Ghouthi Boukli, Vincent Gripon, Matthieu Arzel, Nicolas Farrugia and Yoshua Bengio. *Quantized Guided Pruning for Efficient Hardware Implementations of Deep Neural Networks.* In 2020 18th IEEE International New Circuits and Systems Conference (NEWCAS), pages 206–209. IEEE, 2020. (Cited on pages 35 and 68.)

[Brahim & Farrugia 2020] Abdelbasset Brahim and Nicolas Farrugia. *Graph Fourier Transform of fMRI temporal signals based on an averaged structural connectome for the classification of neuroimaging.* Artificial Intelligence in Medicine, vol. 106, page 101870, 2020. (Cited on pages 47, 48, 53 and 68.)

[Brahim *et al.* 2019] Abdelbasset Brahim, Mehdi Hajjam El Hassani and Nicolas Farrugia. *Classification of Autism Spectrum Disorder Through the Graph Fourier Transform of fMRI Temporal Signals Projected on Structural Connectome.* In International Conference on Computer Analysis of Images and Patterns, pages 45–55. Springer, 2019. (Cited on pages 48, 49 and 68.)

[Challéat *et al.* 2018] Samuel Challéat, Alexis Laforge, Rémi Bénos, Nicolas Farrugia, Clélia Sirami, Georges Zissis, Laurent Canale and Sebastien Vauclair. *Sons et lumières du Haut Vicdessos. La nuit au prisme d'une écologie du son et de la lumière artificielle.* In Séminaire annuel du LabEx DRIIHM (Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieux), 2018. (Cited on page 62.)

[Challéat *et al.* 2020] Samuel Challéat, Nicolas Farrugia, Amandine Gasc, Jeremy Froidevaux, Jennifer Hatlauf, Frank Dziock, Adrien CHARBONNEAU, Juliette Linossier, Christopher Watson and Paul A Ullrich. *Silent· Cities.* In OSF, 2020. (Cited on page 62.)

[Chang *et al.* 2015] Luke J Chang, Peter J Gianaros, Stephen B Manuck, Anjali Krishnan and Tor D Wager. *A sensitive and specific neural signature for picture-induced negative affect.* PLoS biology, vol. 13, no. 6, page e1002180, 2015. (Cited on page 45.)

[Dalla Bella *et al.* 2011] Simone Dalla Bella, Jakub Sowinski and Nicolas Farrugia. *Lack of beat isochrony disrupts motor performance in a child drummer prodigy.* In BIO Web of Conferences, volume 1, page 00019. EDP Sciences, 2011. (Cited on page 7.)

[Dalla Bella *et al.* 2013] Simone Dalla Bella, Anita Białuńska and Jakub Sowiński. *Why movement is captured by music, but less by speech: Role of temporal regularity.* PloS one, vol. 8, no. 8, page e71945, 2013. (Cited on page 9.)

[Dalla Bella *et al.* 2015] Simone Dalla Bella, Charles-Etienne Benoit, Nicolas Farrugia, Michael Schwartze and Sonja A Kotz. *Effects of musically cued gait training in Parkinson's disease: beyond a motor benefit.* Annals of the New York Academy of Sciences, vol. 1337, no. 1, pages 77–85, 2015. (Cited on page 9.)

[Dalla Bella *et al.* 2016] S Dalla Bella, J Sowi, N Farrugia and M Berkowska. *Igor: a case study of a child drummer prodigy igor: a case study of a child drummer prodigy.* Musical Prodigies: Interpretations from Psychology, Education, Musicology, and Ethnomusicology, pages 391–408, 2016. (Cited on page 7.)

[Dalla Bella *et al.* 2017a] Simone Dalla Bella, Charles-Etienne Benoit, Nicolas Farrugia, Peter E Keller, Hellmuth Obrig, Stefan Mainka and Sonja A Kotz. *Gait improvement via rhythmic stimulation in Parkinson's disease is linked to rhythmic skills.* Scientific reports, vol. 7, no. 1, pages 1–11, 2017. (Cited on pages 12 and 15.)

[Dalla Bella *et al.* 2017b] Simone Dalla Bella, Nicolas Farrugia, Charles-Etienne Benoit, Valentin Begel, Laura Verga, Eleanor Harding and Sonja A Kotz. *BAASTA: Battery for the assessment of auditory sensorimotor and timing abilities.* Behavior research methods, vol. 49, no. 3, pages 1128–1145, 2017. (Cited on pages 9, 10, 11, 12 and 13.)

[Fablet *et al.* 2021] Ronan Fablet, Bertrand Chapron, Lucas Drumetz, Etienne Mémin, Olivier Pannekoucke and François Rousseau. *Learning variational data assimilation models and solvers.* Journal of Advances in Modeling Earth Systems, vol. 13, no. 10, page e2021MS002572, 2021. (Cited on page 61.)

[Fahn *et al.* 1987] SRLE Fahn, RL Elton, UPDRS Development Committee*et al.* *Unified Parkinson's Disease Rating Scale in Recent Developments.* Parkinson's Disease. Florham Park, NJ: Macmillan Health Care Information, pages 153–63, 1987. (Cited on page 12.)

[Farrugia & Dalla Bella 2014] Nicolas Farrugia and Simone Dalla Bella. *Tempo effects on movement kinematics: a marker of expertise in professional drummers.* In Joint Conference of the 13th International Conference on Music Perception and Cognition (ICMPC), Symposium on Musical movement: effects of cueing and feedback, 2014. (Cited on page 8.)

[Farrugia & Echraibi 2018] Nicolas Farrugia and Amine Echraibi. *Functional connectivity profiles underlying musical training and engagement.* In Annual Meeting of the Organization for Human Brain Mapping, pages 104–105. Organization for Human Brain Mapping, 2018. (Cited on pages 40, 42 and 69.)

[Farrugia *et al.* 2007] Nicolas Farrugia, Franck Mamalet, Sebastien Roux, Fan Yang and Michel Paindavoine. *A parallel face detection system implemented on FPGA*. In 2007 IEEE International Symposium on Circuits and Systems, pages 3704–3707. IEEE, 2007. (Cited on page 35.)

[Farrugia *et al.* 2008] Nicolas Farrugia, Franck Mamalet, Sébastien Roux, Fan Yang and Michel Paindavoine. *Design of a real-time face detection parallel architecture using high-level synthesis*. EURASIP Journal on Embedded Systems, vol. 2008, pages 1–9, 2008. (Cited on page 35.)

[Farrugia *et al.* 2009] Nicolas Farrugia, Franck Mamalet, Sébastien Roux, Fan Yang and Michel Paindavoine. *Fast and robust face detection on a parallel optimized architecture implemented on FPGA*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 4, pages 597–602, 2009. (Cited on page 35.)

[Farrugia *et al.* 2012] Nicolas Farrugia, Charles-Etienne Benoit, Eleanor Harding, Sonja Kotz and Simone Dalla Bella. *BAASTA: battery for the assessment of auditory sensorimotor and timing abilities*. In Joint Conference of the 12th International Conference on Music Perception and Cognition (ICMPC), 2012. (Cited on page 9.)

[Farrugia *et al.* 2015a] N Farrugia, J Smallwood and Lauren Stewart. *Frequency and affective evaluation of involuntary musical imagery correlate with resting state networks of spontaneous cognition*. In Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music, 2015. (Cited on page 24.)

[Farrugia *et al.* 2015b] Nicolas Farrugia, Kelly Jakubowski, Rhodri Cusack and Lauren Stewart. *Tunes stuck in your brain: The frequency and affective evaluation of involuntary musical imagery correlate with cortical structure*. Consciousness and Cognition, vol. 35, pages 66–77, 2015. (Cited on pages 22, 23, 24 and 26.)

[Farrugia *et al.* 2016] Nicolas Farrugia, Julia M Huntenburg, Daniel S Margulies and Vincent Gripon. *Identifying spatiotemporal patterns of functional connectivity using dictionary learning*. In 22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM), 2016. (Cited on page 39.)

[Farrugia *et al.* 2019] Nicolas Farrugia, Victor Nepveu and Deycy Camila Arias Villamil. *Estimating encoding models of cortical auditory processing using naturalistic stimuli and transfer learning*. In NeurIPS workshop on Hidden Neurons and Real Units, 2019. (Cited on pages 58 and 69.)

[Farrugia *et al.* 2021] Nicolas Farrugia, Alix Lamouroux, Christophe Rocher, Jules Bouvet and Giulia Lioi. *Beta and Theta Oscillations Correlate With Subjective Time During Musical Improvisation in Ecological and Controlled Settings: A Single Subject Study*. Frontiers in Neuroscience, vol. 15, page 676, 2021. (Cited on pages 58, 60, 68 and 69.)

[Farrugia 2008] Nicolas Farrugia. *Architectures parallèles pour l'analyse de visages embarqués*. PhD thesis, Dijon, 2008. (Cited on page 35.)

[Floridou *et al.* 2015] Georgia A Floridou, Victoria J Williamson, Lauren Stewart and Daniel Müllensiefen. *The involuntary musical imagery scale (IMIS)*. Psychomusicology: Music, Mind, and Brain, vol. 25, no. 1, page 28, 2015. (Cited on pages 22 and 24.)

[Freteault *et al.* 2020] Maëlle Freteault, Michel Jézéquel, Pierre Bellec and Nicolas Farrugia. *Encoding models of auditory features in full-length movies estimated using transfer learning.* In Annual Meeting of the Organization for Human Brain Mapping, pages 104–105. Organization for Human Brain Mapping, 2020. (Cited on pages 58 and 68.)

[Hammond *et al.* 2011] David K Hammond, Pierre Vandergheynst and Rémi Gribonval. *Wavelets on graphs via spectral graph theory.* Applied and Computational Harmonic Analysis, vol. 30, no. 2, pages 129–150, 2011. (Cited on page 44.)

[Herholz *et al.* 2021] Peer Herholz, Eddy Fortier, Mariya Toneva, Nicolas Farrugia, Leila Wehbe and Valentina Borghesani. *A roadmap to reverse engineering real-world generalization by combining naturalistic paradigms, deep sampling, and predictive computational models.* In arXiv preprint arXiv:2108.10231, 2021. (Cited on page 57.)

[Hoehn *et al.* 1998] Margaret M Hoehn, Melvin D Yahr*et al.* *Parkinsonism: onset, progression, and mortality.* Neurology, vol. 50, no. 2, pages 318–318, 1998. (Cited on page 12.)

[Hyde & Peretz 2004] Krista L Hyde and Isabelle Peretz. *Brains that are out of tune but in time.* Psychological science, vol. 15, no. 5, pages 356–360, 2004. (Cited on page 9.)

[Iversen & Patel 2008] John R Iversen and Aniruddh D Patel. *The Beat Alignment Test (BAT).* In International Conference for Music Perception and Cognition, Sapporo, Japan. Citeseer, 2008. (Cited on page 9.)

[Jakubowski *et al.* 2014] Kelly Jakubowski, Nicolas Farrugia and Lauren Stewart. *Capturing the speed of music in our heads: Developing methods for measuring the tempo of musical imagery.* In International Conference of Students of Systematic Musicology, 2014. (Cited on page 20.)

[Jakubowski *et al.* 2015] Kelly Jakubowski, Nicolas Farrugia, Andrea R Halpern, Sathish K Sankarpandi and Lauren Stewart. *The speed of our mental soundtracks: Tracking the tempo of involuntary musical imagery in everyday life.* Memory & Cognition, vol. 43, no. 8, pages 1229–1242, 2015. (Cited on pages 20, 21 and 22.)

[Jakubowski *et al.* 2016] Kelly Jakubowski, Nicolas Farrugia and Lauren Stewart. *Probing imagined tempo for music: Effects of motor engagement and musical experience.* Psychology of Music, vol. 44, no. 6, pages 1274–1288, 2016. (Cited on page 22.)

[Jakubowski *et al.* 2018] Kelly Jakubowski, Zaariyah Bashir, Nicolas Farrugia and Lauren Stewart. *Involuntary and voluntary recall of musical memories: A comparison of temporal accuracy and emotional responses.* Memory & cognition, vol. 46, no. 5, pages 741–756, 2018. (Cited on page 22.)

[Kawahara *et al.* 2017] Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker and Ghassan Hamarneh. *BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment.* NeuroImage, vol. 146, pages 1038–1049, 2017. (Cited on page 40.)

[Kong *et al.* 2020] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang and Mark D Plumbley. *Panns: Large-scale pretrained audio neural networks for audio pattern recognition.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pages 2880–2894, 2020. (Cited on pages 34, 62 and 63.)

[Kotz & Schwartze 2010] Sonja A Kotz and Michael Schwartze. *Cortical speech processing unplugged: a timely subcortico-cortical framework.* Trends in cognitive sciences, vol. 14, no. 9, pages 392–399, 2010. (Cited on page 16.)

[LeCun *et al.* 2015] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. *Deep learning.* nature, vol. 521, no. 7553, pages 436–444, 2015. (Cited on pages 28 and 35.)

[Liikkanen & Jakubowski 2020] Lassi A Liikkanen and Kelly Jakubowski. *Involuntary musical imagery as a component of ordinary music cognition: A review of empirical evidence.* Psychonomic bulletin & review, pages 1–23, 2020. (Cited on page 20.)

[Lim *et al.* 2005] Inge Lim, Erwin van Wegen, Cees de Goede, M Deutekom, Alice Nieuwboer, An Willems, Diana Jones, Lynn Rochester and Gert Kwakkel. *Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review.* Clinical rehabilitation, vol. 19, no. 7, pages 695–713, 2005. (Cited on page 11.)

[Lioi *et al.* 2020] Giulia Lioi, Claire Cury, Lorraine Perronnet, Marsel Mano, Elise Bannier, Anatole Lécuyer and Christian Barillot. *Simultaneous EEG-fMRI during a neurofeedback task, a brain imaging dataset for multimodal data integration.* Scientific data, vol. 7, no. 1, pages 1–15, 2020. (Cited on page 53.)

[Lioi *et al.* 2021] Giulia Lioi, Vincent Gripon, Abdelbasset Brahim, François Rousseau and Nicolas Farrugia. *Gradients of connectivity as graph fourier bases of brain activity.* Network Neuroscience, vol. 5, no. 2, pages 322–336, 2021. (Cited on pages 30, 31, 53 and 68.)

[Lurie *et al.* 2020] Daniel J Lurie, Daniel Kessler, Danielle S Bassett, Richard F Betzel, Michael Breakspear, Shella Kheilholz, Aaron Kucyi, Raphaël Liégeois, Martin A Lindquist, Anthony Randal McIntosh*et al. Questions and controversies in the study of time-varying functional connectivity in resting fMRI.* Network Neuroscience, vol. 4, no. 1, pages 30–69, 2020. (Cited on pages 38 and 39.)

[Mairal *et al.* 2009] Julien Mairal, Francis Bach, Jean Ponce and Guillermo Sapiro. *Online dictionary learning for sparse coding.* In Proceedings of the 26th annual international conference on machine learning, pages 689–696, 2009. (Cited on page 38.)

[McAuley *et al.* 2006] J Devin McAuley, Mari Riess Jones, Shayla Holub, Heather M Johnston and Nathaniel S Miller. *The time of our lives: life span development of timing and event tracking.* Journal of Experimental Psychology: General, vol. 135, no. 3, page 348, 2006. (Cited on pages 9 and 10.)

[Mendes *et al.* 2019] Natacha Mendes, Sabine Oligschläger, Mark E Lauckner, Johannes Golchert, Julia M Huntenburg, Marcel Falkiewicz, Melissa Ellamil, Sarah Krause, Blazej M Baczkowski, Roberto Cozatl*et al. A functional connectome phenotyping dataset including cognitive state and personality measures.* Scientific data, vol. 6, no. 1, pages 1–19, 2019. (Cited on pages 26, 41, 45 and 46.)

[Ménoret *et al.* 2017] Mathilde Ménoret, Nicolas Farrugia, Bastien Pasdeloup and Vincent Gripon. *Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction.* In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pages 618–622. IEEE, 2017. (Cited on pages 41, 42, 43 and 68.)

[Mensch *et al.* 2021] Arthur Mensch, Julien Mairal and Bertrand Thirion. *Extracting representations of cognition across neuroimaging studies improves brain decoding.* PLoS Computational Biology, pages 1–20, 2021. (Cited on page 56.)

[Niizumi *et al.* 2021] Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, Noboru Harada and Kunio Kashino. *Byol for audio: Self-supervised learning for general-purpose audio representation.* In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE, 2021. (Cited on page 62.)

[Ortega *et al.* 2018] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura and Pierre Vandergheynst. *Graph signal processing: Overview, challenges, and applications.* Proceedings of the IEEE, vol. 106, no. 5, pages 808–828, 2018. (Cited on page 29.)

[Ouahidi *et al.* 2022] Yassine El Ouahidi, Hugo Tessier, Giulia Lioi, Nicolas Farrugia, Bastien Pasdeloup and Vincent Gripon. *Pruning Graph Convolutional Networks to select meaningful graph frequencies for fMRI decoding.* arXiv preprint arXiv:2203.04455, 2022. (Cited on pages 55, 56 and 68.)

[Pajusco *et al.* 2020] Nicolas Pajusco, Richard Huang and Nicolas Farrugia. *Lightweight convolutional neural networks on binaural waveforms for low complexity acoustic scene classification.* In Detection and Classification of Auditory Scenes and Events, pages 197–212, 2020. (Cited on pages 36, 37, 61 and 69.)

[Pilavci & Farrugia 2019] Yusuf Yigit Pilavci and Nicolas Farrugia. *Spectral Graph Wavelet Transform as Feature Extractor for Machine Learning in Neuroimaging.* In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1140–1144. IEEE, 2019. (Cited on pages 44, 45, 46 and 69.)

[Politimou *et al.* 2019] Nina Politimou, Simone Dalla Bella, Nicolas Farrugia and Fabia Franco. *Born to speak and sing: Musical predictors of language development in pre-schoolers.* Frontiers in Psychology, vol. 10, page 948, 2019. (Cited on page 7.)

[Repp 2005] Bruno H Repp. *Sensorimotor synchronization: a review of the tapping literature.* Psychonomic bulletin & review, vol. 12, no. 6, pages 969–992, 2005. (Cited on pages 9 and 10.)

[Schwartze *et al.* 2011a] Michael Schwartze, Peter E Keller, Aniruddh D Patel and Sonja A Kotz. *The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes.* Behavioural brain research, vol. 216, no. 2, pages 685–691, 2011. (Cited on page 9.)

[Schwartze *et al.* 2011b] Michael Schwartze, Kathrin Rothermich, Maren Schmidt-Kassow and Sonja A Kotz. *Temporal regularity effects on pre-attentive and attentive processing of deviance.* Biological psychology, vol. 87, no. 1, pages 146–151, 2011. (Cited on page 17.)

[Schwartze *et al.* 2013] Michael Schwartze, Nicolas Farrugia and Sonja A Kotz. *Dissociation of formal and temporal predictability in early auditory evoked potentials.* Neuropsychologia, vol. 51, no. 2, pages 320–325, 2013. (Cited on pages 17 and 18.)

[Shuman *et al.* 2013] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega and Pierre Vandergheynst. *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains.* IEEE signal processing magazine, vol. 30, no. 3, pages 83–98, 2013. (Cited on page 29.)

[Shuman *et al.* 2015] David I. Shuman, Christoph Wiesmeyr, Nicki Holighaus and Pierre Vandergheynst. *Spectrum-Adapted Tight Graph Wavelet and Vertex-Frequency Frames.* IEEE Transactions on Signal Processing, vol. 63, pages 4223–4235, 2015. (Cited on page 44.)

[Smallwood *et al.* 2021] Jonathan Smallwood, Boris C. Bernhardt, Robert Leech, Danilo Bzdok, Elizabeth Jefferies and Daniel S. Margulies. *The default mode network in cognition: a topographical perspective.* Nature Reviews Neuroscience, vol. 22, no. 8, pages 503–513, Aug 2021. (Cited on page 56.)

[Sowiński & Dalla Bella 2013] Jakub Sowiński and Simone Dalla Bella. *Poor synchronization to the beat may result from deficient auditory-motor mapping.* Neuropsychologia, vol. 51, no. 10, pages 1952–1963, 2013. (Cited on page 9.)

[Stérin *et al.* 2017] Tristan Stérin, Nicolas Farrugia and Vincent Gripon. *An intrinsic difference between vanilla rnns and gru models.* In Proc. COGNITIVE'17, page 84, 2017. (Cited on page 69.)

[Vidaurre *et al.* 2017] Diego Vidaurre, Stephen M Smith and Mark W Woolrich. *Brain network dynamics are hierarchically organized in time.* Proceedings of the National Academy of Sciences, vol. 114, no. 48, pages 12827–12832, 2017. (Cited on pages 39 and 40.)

[Wing & Kristofferson 1973] Alan M Wing and Alfred B Kristofferson. *Response delays and the timing of discrete motor responses.* Perception & Psychophysics, vol. 14, no. 1, pages 5–12, 1973. (Cited on page 9.)

[Zhang *et al.* 2021] Yu Zhang, Nicolas Farrugia, Alain Dagher and Pierre Bellec. *Representational learning of brain responses in executive function and higher-order cognition using deep graph convolutions.* In bioRxiv. Cold Spring Harbor Laboratory, 2021. (Cited on pages 54 and 55.)

Interdisciplinary approaches for neurosciences, artificial intelligence and sound

Abstract: This manuscript summarizes my contributions to the fields of Auditory Cognitive Neurosciences, signal processing on graphs and machine learning. The first axis reported in this manuscript addresses the study of how humans are able to perceive temporal regularity and synchronize their movements with the environment, and how such regularities may be exploited for gait rehabilitation in Parkinson's disease. The second axis investigated individual differences in brain structure and connectivity related to the phenomenon of having music stuck in one's head in the absence of sound stimulation, called Involuntary Musical Imagery. This phenonenon is an endogeneous, spontaneous neural mechanism that may be the result of complex coupling between distinct brain systems of auditory processing and self-generated cognition. The third axis explores how machine learning and graphs can be combined in order to open new possibilities and explanatory power for neuroimaging, in particular by developing Graph Signal Processing using brain connectivity. I also report recent work in deep learning, aiming at reducing computational complexity and memory requirements using pruning techniques and efficient feature learning learning. The manuscript closes with possibilities to combine these different interdisciplinary directions into a set of research perspectives investigating innovative approaches to study the brain using machine learning and graphs, and better understand auditory cognition as well as the effect of sound on humans, on the environment and on societies.

Résumé: Ce manuscrit résume mes contributions aux domaines des neurosciences cognitives de l'audition, du traitement du signal sur graphes et de l'apprentissage automatique. Le premier axe rapporté dans ce manuscrit porte sur l'étude de la manière dont les humains sont capables de percevoir la régularité temporelle et de synchroniser leurs mouvements avec l'environnement, et comment de telles régularités peuvent être exploitées pour la rééducation de la marche dans la maladie de Parkinson. Le deuxième axe étudie les différences individuelles dans la structure et la connectivité du cerveau liées au phénomène d'avoir de la musique dans la tête en l'absence de stimulation sonore, appelé imagerie mentale musicale involontaire. Ce phénomène est un mécanisme neuronal endogène et spontané qui pourrait être le résultat d'un couplage complexe entre des systèmes cérébraux distincts de traitement auditif et de cognition auto-générée. Le troisième axe explore la manière dont l'apprentissage automatique et les graphes peuvent être combinés afin d'ouvrir de nouvelles possibilités et un pouvoir explicatif pour la neuroimagerie, notamment en développant le traitement du signal sur graphe en utilisant la connectivité cérébrale. Je fais également état de travaux récents en apprentissage profond, visant à réduire la complexité de calcul et les besoins en mémoire à l'aide de techniques d'élagage et d'un apprentissage efficace des caractéristiques. Le manuscrit se termine par les possibilités de combiner ces différentes directions interdisciplinaires en un ensemble de perspectives de recherche examinant des approches innovantes pour étudier le cerveau en utilisant l'apprentissage automatique et les graphes, et mieux comprendre la cognition auditive ainsi que l'effet du son sur les humains, sur l'environnement et les sociétés.

Keywords: Cognitive neuroscience, auditory, graph signal processing, machine learning, deep learning