UNIVERSITÀ DEGLI STUDI
DI TRENTO

DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Real-time adaptation of stimulation protocols for neuroimaging studies

Elena Kalinina

Advisor

Prof. Dr. Paolo Avesani

Fondazione Bruno Kessler (FBK)

Università degli Studi di Trento

January 2018

While conjectural at this point in scientific history, it is possible that one could combine an understanding of neural codes, experimental manipulation via brain stimulation, and network control theory to design control strategies to achieve certain mental states. This defines the frontier of mind control.

JOHN D. MEDAGLIA, PERRY ZURN, WALTER SINNOTT-ARMSTRONG & DANIELLE S. BASSETT

Abstract

Neuroimaging techniques allow to acquire images of the brain involved in cognitive tasks. In traditional neuroimaging studies, the brain response to external stimulation is investigated. Stimulation categories, the order they are presented to the subject and the presentation duration are defined in the stimulation protocol. The protocol is fixed before the beginning of the study and does not change in the course of experiment.

Recently, there has been a major rise in the number of real-time neuroscientific experiments where the incoming brain data is analysed in an online mode. Real-time neuroimaging studies open an avenue for approaching a whole new broad range of questions, like, for instance, how the outcome of the cognitive task depends on the current brain state. Real-time experiments need a different protocol type that can be flexibly and interactively adjusted in line with the experimental scope, e.g. hypotheses testing or optimising design for individual subject's parameters. A plethora of methods is currently deployed for protocol adaptation: information theory, optimisation algorithms, genetic algorithms. What is lacking, however, is the paradigm for interacting with the subject's state, brain state in particular. I am addressing this problem in my research. I have concentrated on two types of real-time experiments: closed-loop stimulation experiments and brain-state dependent stimulation (BSDS). As the first contribution, I put forward a method for closed-loop stimulation adaptation and apply it in a real-time Galvanic Skin Response (GSR) experimental setting. The second contribution is an unsupervised method for brain state detection and a real-time functional Magnetic Resonance Imaging (rtfMRI) setup making use of this method.

In a neurofeedback setting the goal is for the subject to achieve a target state. Ideally, the stimulation protocol should be adapted to the subject to better guide them towards that state. One way to do this would be modelling the subject's activity in a way that we can evaluate the effect of various stimulation options and choose the optimised ones, maximising the reward or minimising the error. However, currently developing such models for neuroimaging neurofeedback experiments presents a number of challenges, namely: complex dynamics of a very noisy neural signal and non-trivial mapping of neural and cognitive processes. We designed a simpler experiment as a proof of concept using GSR signal. We showed that if it is possible to model the subject's state and the dynamics of

the system, it is also possible to steer the subject towards the desired state.

In BSDS, there is no target state, but the challenge lies in the most accurate identification of the subject state in any given moment. The reference, state-of-the-art method for determining the current brain state is the use of machine learning classifiers, or multivariate decoding. However, running supervised machine learning classifiers on neuroimaging data has a number of issues that might seriously limit their application, especially in real-time scenarios. For BSDS, we show how an unsupervised machine learning algorithm (clustering in real-time) can be employed with fMRI data to determine the onset of the activated brain state. We also developed a real-time fMRI setup for BSDS that uses this method.

In an initial attempt to base BSDS on brain decoding, we encountered a set of issues related to classifier use. These issues prompted us to developed a new set of methods based on statistical inference that help address fundamental neuroscientific questions. The methods are presented as the secondary contribution of the thesis.

Keywords

Closed-loop control, unsupervised Machine Learning, real-time functional Magnetic Resonance Imaging (rtfMRI), Galvanic Skin Response (GSR), Neurofeedback, Brain State Dependent Stimulation (BSDS), Top-Down Attention, Affective States.

Manuscripts under review

- E. Kalinina, F. Pedregosa, V. Iacovella, E. Olivetti, P.Avesani. *A Test for Shared Patterns in Cross-modal Brain Activation Analysis* Nature Scientific Reports.
- E. Kalinina, P. Avesani, L.K. Hansen. *Closed-loop control of electrodermal activity (EDA): real-time adaptive stimulation can reduce signal variability* Nature Scientific Reports.

Software

- CMPT Searchlight: a method for the analysis of shared patterns betweeen different cognitive modalities
  `https://github.com/elena-kalinina/CMPT_Searchlight`
- Setup for rtfMRI neurofeedback experiments based on supervised decoding
  `https://github.com/FBK-NILab/rtfMRIdecoding`
- Setup for real-time GSR experiments with closed loop protocol adaptation
  `https://github.com/elena-kalinina/GSR_RT_experiment`
- Setup for rtfMRI BSDS experiments based on unsupervised decoding
  `https://github.com/elena-kalinina/BSDS_GMM_rt/tree/master`

The research was funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012", project "Characterizing and improving brain mechanisms of attention - ATTEND".

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank here all the wonderful people that I crossed paths with in these 4 years of the doctoral program. First, I would like to say thank you to my advisor professor Paolo Avesani and all (ex-)colleagues from the lab: Emanuele Olivetti, Sandro Vega Pons, Diana Porro, Vittorio Iacovella. Thank you, fellow PhD students and lab interns Danilo Benozzo, Nusrat Sharmin, Giulia Berto, Seyed Mostafa Kia, Stefania Ebli and Simone Poli, our lunch breaks in Mattarello have always been great fun. I also thank the colleagues from the ATTEND project and especially Marius Peelen and Timo Stein who were an exceptional source of help and inspiration.

I am enormously grateful to all the colleagues from the Magnetic Resonance Imaging Lab (Centre for Mind and Brain Sciences, University of Trento) whose help and support were crucial for performing the experiments: Jorge Jovicich, Domenico Zacà, Claudio Buoninsegna, Manuela Orsini, Andrea Pellegrini, Giulia Fratti, Angela Serafin, and I owe very special thanks to Pietro Chiesa. I would also like to thank the present and former IT guys, whose assistance has always been indispensable, Davide Tabarelli e Stefano Tessari. I also express my gratitude to Valeria Nencini and Sabrina Adamoli for being so helpful with the administrative matters.

I would like to thank my collaborator Fabian Pedregosa, working with him I learned a lot. I also gained great insights from discussions with professor Ben Herbst, professor Jens Schwarzbach, Patrik Andersson and Gaëtan Sanchez who I would also like to thank here.

I thank my wonderful friends and colleagues who bravely said yes to my requests to participate in my experiments: Giulia Malfatti, Ben Timberlake, Folco Panizza, Romina Esposito, Doris Pischedda, Giorgio Marinato, Ivan Paschenko, Leysan Nurgalieva, Alesia Torgan, Diana Kucherbaeva, Pavel Kucherbaev, Giulia Onorati, Paolo Faccin, Linda Malvezzi, Ilaria Toscano, Sofia Pavan, Stefania Bennetti, Katia Abramova, Xenia Dmitrieva, Giacomo Ariani, Eleonora Visintin, Simona Monaco, Thibaud Griessinger, Anton Yeschenko, Sergio Esteban Pérez, Carlo Cisale, Valeria Occelli, Marco Barilari, Piermatteo Morucci, Philipp Schwedhelm, Flavio Ragni, Stephanie Cattoir, Joshua Zonca. I also say thank you to all participants of my experiments for being so commited and fun to work with. I thank all Mattarello and Rovereto PhD students and post-doc folks, for being such great company during CIMEC drinks, lunches, BBQs and other fun stuff.

Very special heartfelt thanks go to professor Lars Kai Hansen and his CogSys group at the Technical University of Denmark (DTU Compute). In particular, I would like to thank Sophie-Therese Hansen, Martin Axelsen, Albert Vilamala, Frans Zdyb, Benjamin Johansen, Georgios Arvanitidis, Tobias Andersen, Finn Årup Nielsen and Jens Madsen, for fun during journal clubs and for help with experiments. Also I say thank you to Wanja Anderson and Sine Ingemann for they contributed a lot to making my stay at DTU very "hyggeligt".

I would like to say thank you to the members of my Qualifying exam comission, Prof. Enrico Blanzieri, Prof. Dr. Andera Passerini and Prof. Dr. David Stoppa. Their valuable feedback to my Qualifying exam presentation helped a lot to shape the research project that resulted in the present thesis. I am very grateful to the reviewers of my thesis, professor Claude Frasson and professor Ramanathan Subramanian, for their kind words and for their suggestions on how to improve the dissertation. I say a huge thank you to professor Niculae Sebe and professor Paolo Giorgini. Last but not least, thanks to Francesca Belton and Andrea Stenico for always being there for us, PhD students.

Part I

Dissertation

Chapter 1

Introduction

1.1 Summary

Modern neuroimaging methods (functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG), functional near infrared spectroscopy (fNIRS)) capture the signal of the active brain and allow to study the fundamental question of neuroscience, that of how the brain processes natural stimuli and responds to them. Technical progress made it possible to acquire and analyze the brain data in an online mode, enabling an exponential growth of real-time neuroimaging. Real-time neuroimaging not only allows to study the brain in a more natural environment, constant activity mode, but it also introduced a new scientific question in the neuroscientific paradigm: how the current brain state affects the brain response to stimuli. This new scientific question calls forth for an update in the experimental research, namely the need for new experimental protocols that would allow to successfully investigate this question by being adapted in real-time based on the analysis of the subject's brain activity.

There are three main real-time neuroimaging experiment types that are used to investigate this question (and also in practical applications): Brain-Computer Interfaces (BCI), Brain-State Dependent Stimulation (BSDS) and Neurofeedback (NF). In neurofeedback experiments, the goal is to bring a subject into some desired state. While the subject is trying to reach that state, she is shown how well she is performing up- or downregulation of her brain activity based on the analysis of her brain state online. In brain-state dependent stimulation, the objective is to study different responses to a stimulus depending on the current brain state. So, the choice of stimulus is conditioned on the actual brain state. In brain-computer interfaces, some external device is set into action when the presence of a necessary brain state is established. The challenge for BSDS protocol design lies in

most accurate identification of the current brain state. That may involve the mapping between the measured signal and cognitive states. The main state-of-the art method used for this purpose, brain decoding, has several issues in real-time settings that I am going to expose. For BCI and NF, there is need for efficient design in a closed-loop manner that combines accurate estimation and prediction of the activity states with an optimal choice of the runtime policy at each step. In the first step of a closed loop experimental setup, like in BSDS, the current state of activity of the subject's brain has to be determined with high accuracy. It can be labeled or it can be assessed in terms of similarity to the target brain state, but most importantly, next state (or states) of the brain activity have to be estimated and predicted resulting from the stimulation input. In the next step, an appropriate optimal action should be carried out (showing stimulus, feedback or running the device) that maximizes the measure of success in achieving the experimental goal. On the one hand, we have methods for modelling dynamic systems. On the other hand, we have algorithms for runtime policy adaptation. However, the challenge lies in combining both to be able to exert control over the subject's brain so that we can also control their behaviour.

In the thesis, I am addressing the two problems described above. First, I am proposing a new method of determining the subject's state that is used in a real-time setup for fMRI BSDS experiments. This contribution was motivated by the need to investigate the neuroscientific question of top-down attention patterns and how they influence perception. This question is central for neuroscience as it is fundamental for understanding information processing in the brain and the interaction of bottom-up and top-down processes. As a secondary contribution, I propose a new method that also helps to understand interactions of top-down and bottom up processing. It is a statistical inference test that can be used as an alternative to brain decodings in addressing this question. As the second main contribution to address the second problem, I designed and ran an experiment where the subject's affective state (arousal) was controlled in a closed-loop manner based on the recorded Electrodermal Activity signal (EDA), or Galvanic Skin Response (GSR).

1.2 Background: real-time neuroscience

When humans perform cognitive tasks, there is neural activity going on in the brain: the neurons fire, that is, an electrical charge is sent from one neuron to another. The firing of neurons is accompanied by a number of metabolic processes as active neurons consume energy. Neuroimaging techniques are based on our ability to register these

physical changes in the brain and relate them to our cognitive activity. Each neuroimaging technique can capture a particular type of physical processes. EEG registers changes in the electrical fields that are created by the charges of neural firing. MEG is sensitive to changes in the magnetic field that is the byproduct of the current of neuronal firing. FMRI detects changes related to brain metabolism and oxygen consumption in the active areas of the brain via magnetic properties of oxygenated and deoxygenated blood flow. fNIRS tracks the same changes, but based on a different physical principle: different rates of light absorption by oxygenated and deoxygenated blood. All these techniques are non-invasive, which makes them suitable for experimental studies. However, each has its advatages and downsides. For instance, EEG and MEG have high temporal resolution, i.e. the changes in the registered signal occur with minimal latency with respect to changes in the brain. This is not the case with fMRI and fNIRS known for their slow signal: metabolic changes they track have a quite long delay with respect to neuronal firing underlying the activity and are not directly related to the timings of firing. On the other hand, fNIRS and fMRI have a high spatial resolution: they allow to locate the active, firing sites with precision that is unattainable for EEG and MEG. Finally, EEG and fNIRS are portable technologies and can be used outside of the lab environment to study the brain workings in the real world. fMRI and MEG, on the contrary, are bulky machines and can only be run in labs by qualified technicians. However, all these techniques provide means for investigating the active brain and help answer the fundamental questions of modern neuroscience about how the brain processes the external stimuli and reacts to them.

In classical neuroimaging experiments, the data is first collected and then analyzed offline. During the experiment, the researcher tests participants (healthy subjects or patients) to know how the brain responds to certain types of stimulation. Experimental procedures and protocols are defined by the purpose of investigation. For the researcher to successfully analyze the participants' data, it is crucial to establish the timings of the various stimuli to attribute the recorded activity data. Every neuroimaging study starts with designing a stimulation protocol. It defines the stimulus categories and the time that the stimulus is delivered to the subject (known as onset), so that this information could be subsequently used in data analysis. Stimulation is then organized into trials (when a single stimulus is presented) and runs (subblocks of the experimental session consisting of multiple trials). Stimulation protocol with all these characteristics (number of stimulus categories and number of stimuli in each category, onsets, durations, presentation order) is generally developed before the start of the neuroimaging study. It is fixed for the whole experiment and does not change in its course.

Technical developments made it possible to stream, register and analyze neuroimaging data on the fly, in an online mode as the imaging process unfolds. As this allows to have access to ongoing brain activity, real-time neuroimaging was discovered to have a huge potential for practical applications. In brain-computer interface applications, the signal from the subject's brain is used to control an external device in real time. In neurofeedback applications, the subjects get feedback about their brain activity to enable them to up- or down regulate it so that it could result in changes in behaviour. In brain state dependent stimulation, magnetic impulses or deep current are delivered to target areas of the brain when a certain state is detected, to modulate or enhance the activity. However, all these new real-time types of experiments apart from their practical importance allowed to pose new research questions in the neuroscientific paradigm, namely about how the response to external stimuli is conditioned by the current brain state. This question can hardly be answered with offline data analysis implying the use of standard experimental protocols.

New research questions call forth for new developments in experimental design. Unlike in the canonical neuroimaging, real-time experiments cannot have a fixed protocol. The stimulus presented to the subject depends on the outcome of signal analysis. Real-time neuroimaging has a huge potential for creating interactive stimulation paradigms that open new perspectives for neurological investigations. Only with this flexibility of design can we approach this whole new broad range of questions about how we can interfere with the subject's brain activity or how the performance in cognitive tasks depends on the current brain state.

1.2.1 Brain-State Dependent Stimulation (BSDS)

So, the question about how the current brain activity affects the subsequent reaction to stimulation is central to real-time neuroimaging experiments. In other words, real-time neuroimaging raises the possibility of manipulating cognitive processing by taking the online brain activity into account and delivering the stimulus in the moment when the subject is in a particular state [119]. This approach is known as Brain State Dependent Stimulation (BSDS). This manipulation can take physical form when Transcranial Magnetic Stimulation (TMS) or Transcranial Direct Current Stimulation (TDCS) impulses are delivered to the subject in real-time based on their brain state as a treatment for some clinical pathologies [178]. But most importantly, BSDS opens a new paradigm in neuroscientific research [268]. In the traditional neuroscientific studies, the brain state is manipulated as a dependent variable with the stimulation that is the independent variable. Real-time neuroimaging allows to study the effect of brain state on behavior - for

instance, in decision making and action (motion) research, the subject, depending on the state she is in, can take different decisions or engage in different types of motion. In this case, the relationship between the dependent and independent variables is swapped: the brain state becomes the independent variable, while the behavior becomes the dependent variable [268]. To allow for the modulation of stimuli in relation to brain states, the subject's current state has to be identified to be used as a control signal to adapt the stimulus [144]. Therefore, to provide valid conclusions about the impact of the brain state on behaviour, BSDS has to rely heavily on the possibility of identifying the subject's state in a reliable manner and labeling it appropriately.

An important use of the BSDS technique is the investigation into top-down attention patterns. Top-down and bottom-up are two types of attentional processes distinguished in neuroscience [87, 88]. Bottom-up attentional mechanisms are activated by the salient stimuli in the immediate environment, e.g. bright light or loud noise. Top-down attentional mechanisms, on the contrary, are those that support our readiness to perceive certain stimuli because of their inner significance for the person. BSDS provides a handy technique to study these mechanisms: if we detect this internal readiness in the brain and send the stimulus in that particular moment when the subject is ready to attend to it, we can spot behavioural differences (in terms of reaction times or performance metrics, for example) between these reactions and reactions when no top-down attentional state is detected.

1.2.2 Neurofeedback (NF)

In neurofeedback applications, the subject learns to upregulate or downregulate the activity in the particular region of interest (ROI) of the brain. The feedback that the subject gets is based on the signal coming from this area. The signal is recorded and processed according to the neuroimaging technique used. Usually, neurofeedback experiments proceed like this: the ROI for feedback is chosen that is known for its involvement in a certain cognitive activity, like, for instance, emotion or pain processing (e.g. [268]). The subject in the scanner performs some cognitive task engaging this area and gets feedback on the screen that takes into account registered changes in the activity in this area (typically, in the form of a temperature bar). The subject is instructed to regulate her brain activity in such a way, that the feedback screen produces a certain output - e.g.make the temperature bar go up or down. The cognitive task given to the subject in the scanner is supposed to facilitate achieveing the desired result. In this way, it is claimed that the subject can create lasting change in her brain activity pattern which in turn leads to

cognitive improvements: self-regulated neuronal activity can influence behavior.

Neurofeedback is being more and more widely used for clinical purposes for its perceived benefits. The technique has gained huge significance because of a great number of reported successful studies. It has been successfully used for a wide range of disorders to achieve improvements: motor rehabilitation and training (e.g., of Parkinson patients - [249]), pain management [211] and for emotional upregulation [158], to name just a few. The use of neurofeedback is very common in depression patients. The goal is to increase response in the areas processing positive emotions, because the patients perceive this increased response as improvement of mood. Examples of such studies are quite numerous - [279], [283], [159]: patients with major depression disorder were asked to upregulate the activity of areas responsive to positively valenced stimuli by recalling their happy memories. The authors claim that the patient groups showed clinical improvements as compaired with controls. Neurofeedback studies are the most common among all real-time neuroimaging studies, and the number of these studies continues to grow as more and more disorders are shown to benefit from neurofeedback.

However, in many studies it is noted that in neurofeedback settings the subject should be deliberately and more efficiently guided towards successful use of the feedback to retrain the brain. This is because many subjects are not able to figure out the strategy for the desired regulation. As has been pointed out in [22], patients and healthy subjects often show a large variability, or even inability, of brain self-regulation. This could be the reason why neurofeedback has not become an established therapeutical tool despite consistently reported benefits.

1.2.3 Brain-Computer Interfaces (BCI)

In BCI the subject uses her brain to control external devices, such as attention-based spellcheckers or robotic hands. [8], [222]. These are mainly meant to restore some missing capability for disabled people, but there are also BCIs for gaming that create an immersive virtual reality. For instance, in [182] the subjects learned to control an avatar with their brain activity, moving it around in a virtual space. Brain activity is analyzed, decoded and transformed into the control signal that is transmitted to the device. Neuroimaging tools typically used for BCI include EEG and fNIRS - the first one, for both its high temporal resolution and portability, the second mainly for reasons of portability. As has been said, fMRI measures a very indirect correlate of brain activity, Brain Oxygenation Level Dependent signal (known as BOLD), and has a very low temporal resolution. Besides, it is not a portable device. For this reason, it cannot be used directly in BCIs, but

it is claimed to be very practical for training potential BCI beneficiaries to use BCIs - especially because for many BCI applications, it is crucial, too, to capture a signal from a well-defined area in the brain. In [222] it is shown that still BCI control can be performed even with such seemingly slow tool as is real-time fMRI. Training subjects in the fMRI for BCI use could be regarded as a special case of neurofeedback experiments.

1.3 Motivation of the study: need for adaptive protocols

Neuroimaging techniques help address neuroscientific questions about the workings of the brain. However, the difference between classic neuroimaging and real-time neuroimaging does not just lie in the differences of data acquisition. These two methods answer two fundamentally different scientific questions. The first one is about how brain reacts to stimuli where the researcher manipulates the dependent variable (measured activity) with experimental stimuli (independent variable). In real-time experiments, as has been said, the question is about how brain state can condition reactions to stimuli, and the relationship between the variables is reversed: brain activity is the independent variable, while the stimulation becomes the dependent one. This change in the experimental paradigm requires a different experimental design, a design where the stimulation can be adapted to the changes in brain activity in real time based on the experimental goal.

1.3.1 Adaptive protocols for BSDS

In BSDS, the goal is to demonstrate how performance is different between stimulation sent when the subject is in the target state and the random stimulation or stimulation when the target state is not present. To be able to perform inference on this data, the researcher has to possess robust methods for the state identification, so that the behavioural differences in the aforementioned types of trials could be reliably attributed to different brain states. Identifying the brain state implies mapping a certain pattern of activity to some cognitive state label - e.g., emotional or attentional state. The state-of-the-art of establishing such mapping is Multivariate Pattern Analysis (MVPA), or brain decoding [106]. This analysis consists in using Machine Learning classifiers in a supervised manner: the classifier is trained on some portion of data labeled corresponding to the detectable states and then tested on the unseen portion of the data to predict the state label. An example of decoding to detect an attentional state in BSDS is [56]: the subjects were shown two overimposed images and were asked to attend to one of them in a sustained manner; in the meantime, a logistic regression classifier was run to monitor their attention state and

to detect attention lapses when the subject was found to attend to the wrong image.

1.3.2 Closed loop adaptive protocols

In BCI and neurofeedback, detecting the subject's current state is just the first step of the process. The real challenge lies in the fact that the subject's brain actively interacts with the stimulation system, and the stimulation has to be constantly interactively adapted to reach the experimentation goal. In the course of interaction either the subject controls the external device (BCI, neurofeedback), or we try to control the subject's brain to drive behavioural changes (neurofeedback). The latter interaction type can be regarded as a case of control in terms of control theory. Control means we are trying to introduce changes into the system from some external system. Control theory distinguishes between open loop control and closed loop control. They differ with respect to the use of adaptation. In the former, the control system performs control without getting feedback from the controlled plant. In the latter case, after the control input the plant receives the feedback from the controlled system and adjusts the next input respectively to achieve the desired control result. Closed loop control strategy comprises two elements. The first is system identification: the current state has to be estimated and the next state(s) resulting from control input predicted. This can be done, for instance, using state space models and Kalman filter [264], [265]. The second element is the choice of optimal input, or runtime policy. The choice of runtime policy for real-time protocol adaptation can be represented as the problem of optimal control [108], policy search algorithms (MDP & RL) or it can be framed as an optimization problem: choosing the optimal protocol with the scope of the experiment in mind. That of reinforcement learning [253] could, too, be a fruitful framework for neurofeedback experiments. In fact, in [151] the authors performed a neurofeedback study in the context of reinforcement learning. The participants learned to upregulate the signal from the anterior insula because it is involved in multiple clinical conditions related to emotional and affective dysfunctions. The researchers worked out a reward-related reinforcement learning model of neurofeedback with real-time functional Magnetic Resonance Imaging (rtfMRI). The starting point hypothesis was that in the target group, unlike in the controls, a positive linear increase of signal will be observed in the activation of the right anterior insula and no such linear increase was expected in the controls that were provided false feedback. The study confirmed the expected results. However, it is worth noting that this recent study is unique in its use of the reinforcement learning model for neurofeedback, despite its obvious relevance in the field.

The optimization problem can be restated as the problem of "finding an optimal de-

cision x in order to maximize a measure $goodness(x)$“ [19]:2. So, in this case we have to define our choice of options $x = \{x_1, \ldots, x_n\}$, work out a quantative assessment of the “goodness“ function and hand over the goodness model to an algorithm that will search for an optimal choice at each step of the algorithm. In case of neuroimaging experimental setups, x is a set of experimental design choices that we expect to have impact on the activity we study. This could be the list of stimulus categories (e.g. pictures of people, houses, food and instruments) or stimulus features (e.g. the degree of stimulus noisiness - 50, 25, 15, 5 %). This can be also the space of timing choices (e.g. the moment at which the stimulus is delivered).

Now, how can we measure the “goodness“? In [184]: 3 the problem of design optimization is formulated as "finding a design d^* that maximizes some utility function $U(d)$ that quantifies the quality of design d".

$$d^* = \arg\max_d U(d)$$

However, it can be noted that although a plethora of methods have been proposed for protocol adaptation, they do not aim at interacting with the subject in a closed loop manner, so that the utility function does not maximize the chances of the subject getting into a certain state. The goals of adaptation are different. For instance, in a series of works ([184], [224], [53], [40]) design optimization procedure has been worked out for online hypothesis testing: the idea behind this framework is finding a design choice that at each timepoint provides best discrimination between competing models of the subject's activity in response to stimuli. It offeres us a measure that quantifies the utility of each design choice from the viewpoint of model selection and thus for efficient sampling procedure: the utility function is defined in terms of information theory. In other studies, optimization algorithms (stochastic optimization, Bayesian opimization, genetic algorithms, swarm optimization) have been used to optimize the design for better quality of data or for ROI selection.

1.3.3 Generalization of protocol adaptation schemes in real-time neuroimaging

So, in this section, we will try to generalize the schemes of protocol adaptation in various real-time experiments. A principal distinction here is between interactive, closed-loop and non-interactive, open-loop protocols. BSDS is largely open-loop: we detect the state and send the stimulation. The effect of stimulation sent in real-time when the subject is or is not in the target state is often estimated offline with a separate inference method.

BCI and NF allow both interactive and non-interactive approach. If we consider them from the subject's viewpoint, both are inherently interactive and closed-loop: the subject learns to control feedback or external device. However, modelling the subject's behaviour in both types of interactions is a separate, however interesting, problem, which is outside of the scope of the present study. If we view BCI or NF experiments from the perspective of the second agent, namely the computer setup executing the stimulation, here both open and closed loop scenarios are possible. The non-interactive, open-loop scenario is similar to that of BSDS: we get the signal, detect the state and send the command to the external device in the case of BCI. In the case of NF, we get the signal and show it back to the subject in some other form, e.g., as a temperature bar. However, both BCI and NF allow an interactive scenario where the computer setup interacts with the subject - and this is the situation of the closed loop control. It is the computer now that gets feedback from the subject in the form of comparing the obtained measurement to the target state and choosing the input to the subject based on the error between the two. The presence of this feedback loop distinguishes interactive, closed-loop scenarios from non-interactive, open-loop ones. This difference is summarized in Figure 1.1.

The main challenge in the non-interactive scenario is brain state detection - decoding. In the interactive scenario, it is the design of such closed loop interactive protocols, that can actually execute control over the subject's state. For this control to be meaningful and beneficial, we need to have profound knowledge about the mapping of brain states and behaviour, and here the BSDS paradigm is fundamental, as it can largely inform the development of closed loop interactive protocols.

1.4 Problem statement

So, at the moment the challenge in real-time neuroscience is designing adaptive protocols that are based on the subject's state detection and can be updated interactively. In a BSDS experiment, the investigator wants to monitor the changes in the subject's brain state. In the experimental trials, the stimulus has to be shown to the subject when a certain state is identified. In control trials, the stimuli are shown at random. The effect of the BSDS shows in significant differences in behavioural performance between BSDS and control runs. The main challenge in these experiments lies in identifying the subject's current state in real-time with high accuracy so that a valid comparison could be made between those and control runs. The fact that the state identification occurs in real-time imposes certain requirements on the methods deployed for state identification: first, we

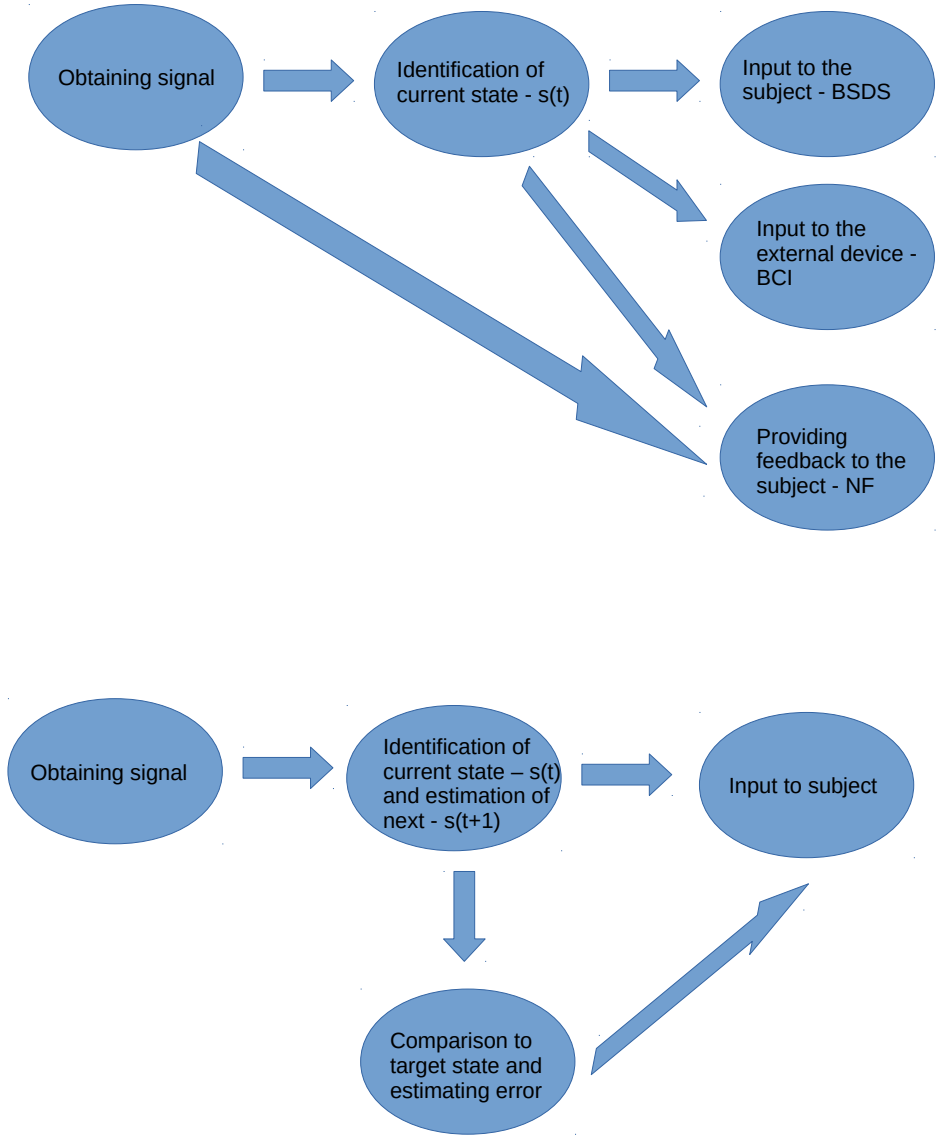


Figure 1.1: Different types of adaptive stimulation in real-time neuroimaging. Non-interactive scenarios (top) are based on most accurate detection and identification of the current brain state (this step is optional for neurofeedback). In the interactive scheme (bottom), an additional step is necessary to decide on the input to the subject, namely comparison of the current and next (estimated) state to some target state and then acting upon the error.

should be able to deal with a limited number of samples, and second, the learning method should be incremental, as it does not have access to all data of the session. Once the state is reliably identified, the runtime policy is trivial and consists in presenting the stimulus with minimal latency. However, at the moment the neuroscientific community possesses limited tools for this purpose. Brain decoding, or the use of supervised machine learning classifiers to label brain states, has several drawbacks, first and foremost, the curse of dimensionality in view of limited data usually available from experiments. In real-time settings, these drawbacks become even more pronounced. The development of robust methods for brain state identification to be used in real-time settings for online data processing can largely contribute to the promotion of the new, highly promising research paradigm of BSDS.

Control of a subject's state is the ultimate goal of many real-time neurofeedback experiments. However, in such experiments it is the subject who is trained to control her emotional state, up- or down-regulating the activity in some targeted brain area. In these experiments the control is performed by the subject and only indirectly by the external intervention. Exerting control over the subject's state in a closed loop manner could be required, especially in clinical context or when the subject is unable to learn to regulate the brain activity. Designing control strategies to achieve certain states of mind through experimental manipulation is thought to be a frontier of mind control, though possible [176]. To successfully implement a control strategy, a clearly defined target state is needed as the goal of the neural control. For the control intervention to be beneficial, it has to be defined in terms of cognitive representations ("mind"), but at the same time control strategy is executed at the physical level ("brain"), such as neural substrates and/or measured signal. For that we need an accurate mapping between neural (brain) states and states of mind. This is is seen at the moment rather as a challenge driving cognitive neuroscience [176] and a debated topic, not a solved issue. Another open challenge is estimating the physical system dynamics and determining its state in a way that we can evaluate the effect of various stimulation options and chose the optimized ones, maximizing the reward or minimizing the error. This is more achievable in an invasive manner, with the direct access to the neuronal measurements [228]. However, in view of limited applications of invasive control as well as general difficulties of measuring and modelling such noisy systems as human brain, we have to rely on other means of brain state definition, based on the analysis of non-invasive neuroimaging data. So, even though we have means for protocol adaptation, such as policy search, optimization algorithms and control theory approach, a meaningful definition of the subject state is at the core

of the protocol adaptation problem. That would allow dynamic system identification and the application of various models, e.g. state space models. Successful state definition requires collaboration with neuroscientists so that they could also formulate the experimental goal that could be formalized and modelled according to the chosen approach. Taking into account all this complexity, it might be preferable to strive for guiding the brain rather than controlling it in a compelling manner [100]. Even in the absense of an ideal control strategy, simplified control strategies built with limited access to the system and drawing upon the brain's natural dynamics could still provide benefits for the clinical scenarios [176].

1.5 Contributions

The contribution of the thesis consists in proposing new methods for the interaction with the subject's cognitive state in real-time. First, we propose a method for conducting brain-state dependent stimulation experiments. We show that the reference method for identifying brain states - brain decoding, or the use of supervised machine learning classifiers - has serious drawbacks which present particular limitations in the online mode. We propose to use an unsupervised clustering method - Gaussian Mixture Model (GMM) - to distinguish between activated and non-activated state. To enable real-time application of GMM, we use online clustering algorithm: only the model is kept in memory and updated with each incoming datapoint. We validate the method on a data from an fMRI experiment that shows that this method allows to present the stimulus on the majority of trials and thus fulfils the requirement of reliable brain state identification. We also present the real-time setup that implements the runtime policy: the stimulus is shown to the subject in the moment the state is identified with negligible latency. We discuss also some particularities of the use of GMM in real-time fMRI contexts.

Next, we designed and ran a proof-of-concept experiment to show how we can control the subject's affective state (arousal) in a closed-loop manner. In a visual task, the subjects were viewing sequences of negative and neutral images, while their level of arousal was measured with Galvanic Skin Response (GSR) sensors. GSR signal was used because it allows for a rather straighforward mapping of the signal level and affective state (arousal) and also for meaningful target state definition. A predictive model was applied to estimate the expected signal for each prospective sequence of stimulation images, which, in turn, allowed to compute the error between the predicted signal and the target signal. Then, the runtime policy consisted in choosing the sequence that minimized the error

between the predicted signal and the target signal. The error calculated for overall signal of the session with respect to the target signal was reduced significantly in the real-time experimental group as compared with the control group, although the mean number of negative pictures shown in a session at the group level was not significantly different for the groups. These findings suggest that it is possible to steer the subject's emotional state towards the desired target state.

Finally, as a secondary contribution, we propose a test for the analysis of shared patterns between different cognitive modalities. This test was the outcome of our investigation into supervised brain decoding for the purposes of BSDS.

1.6 Structure of the thesis

The thesis is divided into two main parts: Part I (Dissertation) and Part II. The former part summarizes the two main contributions of my PhD research and is structured as follows:

- Chapter 2 starts with an overview of the paradigm of real-time neuroscientific experiments mainly concentrating on real-time functional Magnetic Resonance Imaging. It outlines the main research question that is answered within this paradigm and presents three types of real-time experiments (Brain-State Dependent Stimulation, Brain-Computer Interfaces and Neurofeedback) run to answer this question. We show how the need for adaptive protocols arises in this context and what kind of adaptation is requested for each experiment type. Then we also show what methods and algorithms for protocol adaptation have been proposed in the literature.

- In Chapter 3 we present the problems that we set forth to tackle. We show where the main challenge lies in designing protocols for BSDS studies on the one hand and for neurofeedback and BCI on the other.

- Chapter 4 presents our solution for protocol adaptation in the context of BSDS. It is a real-time unsupervised clustering method that allows to detect the subject's activated state. This method is used to design a neuroscientific experiment aiming at studying top-down attentional patterns.

- Chapter 5 describes our second contribution, a closed-loop protocol that was also used in an experimental setting to control the subject's affective state (arousal). In this experiment, the subject's Galvanic Skin response signal was recorded while they were viewing a series of images. This sequence of images was adapted in a closed loop

manner in the experimental group, but followed a random presentation pattern in the control group. We show that this closed-loop intervention allows to reduce variability in the measured Electrodermal activity and conclude that this is a proof-of-concept of a feasible control strategy.

- Chapter 6 outlines the secondary contribution of the thesis that was the outcome of the work on the project "Characterizing and improving brain mechanisms of attention". It is a statistical hypothesis testing procedure meant to identify shared patterns of the activity between different cognitive modalities. This question is of fundamental importance in neuroscience.
- Chapter 7 concludes the thesis. It recalls the main results presented before and comments them by focusing on the related implications, limitations and future works.

The second part (Part II) contains references to the literature.

Chapter 2

Background

In this section, I am going to present the problem of protocol adaption in neuroscience and the computer science methods that have been used to address it. I will present various types of protocols in canonical neuroscientific experiments and the demand for protocol adaptation in real-time experiments. I will be addressing such types of real-time experiments, as neurofeedback, BCI and BSDS. I will also introduce the notions of decoding and closed loop real-time stimulation. I will overview technical methods that have been proposed in the literature for protocol adaptation.

2.1 Real-time neuroscience

When humans perform cognitive tasks, there is neural activity going on in the brain: neurons are firing, that is, sending an electrical charge from one neuron to another to excite or to inhibit the activity of the receiving neuron. Changes in the electric potential of the neural cell membrane are called post-synaptic potential and represent the exitatory and inhibitory input that these neurons receive. Neuroimaging techniques are based on our ability to register these physical changes in the brain and relate them to human cognitive activity. All these techniques are non-invasive, which makes them suitable for experimental studies. However, each has its advatages and downsides. Electroencephalography (EEG) and magnetoencephalography (MEG) directly track the electrical activity of brain cells by measuring its effects on the electrical and magnetic fields just outside the head, following a cognitive event, such as stimulus presentation. They allow researchers to record brain activity at a very high temporal resolution (milliseconds), close to the actual timescale of neural computations. These scalp potentials represent the bulk electrical activity of many pyramidal neurons depolarizing in synchrony (waves of neural activity)

measured in voltage/time (up to 2000 Hz), and they can be registered with electrodes placed all over the scalp. As the charges are quite weak, an amplifier is needed to distinguish the signal. Typically, between 64 and 256 electrodes are used. Higher number of electrodes allows to get more precise estimations of the signal source. EEG waves that accompany cognitive events can have different frequencies; most of the time our brain is emitting alpha waves, which are of a regular frequency (8-12/sec) and high amplitude. The response to external events measured by EEG is known as event-related potential. It is obtained by averaging the signal of many trials where the same stimulus was presented.

MEG measures the magnetic counterpart of that same electrical activity. As is known from physics, an electrical current always generates a magnetic field in the space around it. So, the current produced by firing neurons is also accompanied by a magnetic field, though quite weak. Changes in the cerebral magnetic field are measured through sensors named SQUIDs (superconducting quantum interference devices), which must operate at extremely low temperatures, close to absolute zero. A modern MEG scanner contains typically between 150 and 300 sensors, arranged as a helmet around the head and immersed in liquid helium. As both EEG and MEG signal is directly related to neuronal activity, they both have a very high temporal resoultion (up to a millisecond scale), that is, they allow the researcher to establish the timings of neural events with high precision. On the other hand, the disadvantage of both techniques is that they have quite low spatial resolution and do not allow to pinpoint the source of the signal with the same high precision. This is because the locations of activated brain regions must be estimated from the distribution of the signal outside the head. Estimation of the signal source is referred to as the inverse problem and it is mathematically ill-posed in the sense that there is, in principle, an infinite number of source configurations that can generate the observed signals. Therefore additional assumptions (e.g., diffuseness of sources) are necessary to estimate a unique solution. A mathematical model of the head is used to generate a forward model of how electrical activity at each point affects the signal distribution outside the head.

The firing of neurons is accompanied by a number of metabolic processes due to energy consumption. FMRI detects changes related to brain metabolism and oxygen consumption in the active areas of the brain via magnetic properties of oxygenated and deoxygenated blood flow (fMRI is described in more detail in a dedicated section below). Functional Near Infrared Spectroscopy (fNIRS) tracks the same changes, but based on a different physical principle: different rates of light absorption by oxygenated and deoxygenated blood. First, a near infrared light pulse is sent into the brain by an optical fiber - the

light can easily penetrate the skull. Because living organisms significantly scatter light, near infrared light introduced by optical fiber is scattered by various types of tissue. Then the light, after being absorbed and scattered at the cerebral cortex, is condensed by the optical fiber again from the surface of the head. However, at this point some portion of it has already been absorbed by the hemoglobin in the active parts of the cortex. Some of this scattered light reaches the optical fiber in the light receiving probe, and is then guided to a photomultiplier where it is converted to electrical signals so that we can know where most of the light had been absorbed. That indicates the location of the most responsive brain area. Unlike EEG and MEG, where the changes in the registered signal occur with minimal latency with respect to changes in the brain, fMRI and fNIRS are known for their slow signal and relatively low temporal resolution: metabolic changes they track have a quite long delay with respect to neuronal firing underlying the activity and are not directly related to the timings of firing. On the other hand, fNIRS and fMRI have a high spatial resolution: they allow to locate the active, firing sites with precision that is unattainable for EEG and MEG. Finally, EEG and fNIRS are portable technologies and can be used outside of the lab environment to study the brain workings in the real world. fMRI and MEG, on the contrary, are bulky and costly machines and can only be run in labs by qualified technicians. However, all these techniques provide means for investigating into the active brain and help answer the fundamental questions of modern neuroscience about how the brain processes the external stimuli and reacts to them.

Because my work was mainly focused on the fMRI data, the rest of this chapter will be largely dedicated to fMRI research with only a few examples from other neuroimaging methods.

2.2 Experimental design. Design of neuroscientific experiments

Data collection starts with a research question. As an answer to the research question, a hypothesis is formed which states how experimental manipulations are expected to impact the measurements in the data, and an experiment is set up to test the link between the hypothesis and the data in a controllable and controlled way. The hypothesis must predict the outcome of the experiment and it cannot be confirmed, but only disproved. The hypothesis to be disproved is called the null hypothesis, and it usually predicts a negative outcome of the experiment, i.e. it states that the experimental intervention does not evoke any stimulation related changes in the data. In a controlled experiment, some aspect is manipulated, and the data is collected before and after manipulation to see

if the manipulation had an effect. In neuroscientific experiments, the measurements of brain activity are taken as experimental data. The controlled manipulation is a temporally ordered collection of stimuli that are meant to evoke perceptual or cognitive activity in the brain. Usually some aspect of the stimulus is manipulated - categorical (different objects in pictures) or parametrical (easy vs difficult mathematical task) [113]. The way stimuli are chosen and organized in an ordered sequence is called experimental design. The changes in the measurements caused by stimuli are subject to statistical inference procedure meant to disprove the null hypothesis. The design is efficient if the null hypothesis can be disproved even when the experimental manipulations have a very small effect.

Manipulated aspect is the independent variable, measurement is the dependent variable. Conditions in the experimental design are different values of the independent variable, and they are usually categorical or parametrical. Another type of manipulation is between groups vs. within group manipulation. In the first case, experimental groups are constructed based on some group parameter (age, gender, healthy vs. patients) or manipulations parameter (medication vs. placebo, real feedback vs. sham feedback) [3].

So, by disproving the null hypothesis we intend to show that manipulating the IV will cause changes in the measurements. However, how can we tell slight fluctuations from changes that are big enough to claim that the manipulation worked? The null hypothesis is formulated as "Manipulation does not cause any changes in the measurement", and then we use statistical inference in the form of tests to disprove it. The outcome of tests is the value of the metrics - p-statistic or t-statistic - that is indicative of the effect of the manipulation. P-value is the probability to get the experimental measurements under the null hypothesis that manipulations do not have any effect. The purpose of design is to ensure that you can perform valid inference procedures on the data. An important characteristics of the design is its statistical power, i.e it should allow to reject the null hypothesis even when the effect is small.

Data acquisition has been studied in statistics under the topic of experimental design: the dataset must be constructed in such a way as to provide the possibility for statistical inference based on the data [83]. The dataset should contain as many subjects as possible for group inference, and as many data samples (trials) as possible for a single subject. Efficient design is the design where conditions are different in just one aspect, but it strongly allows to detect differences even with few data samples.

So, in the traditional design independent variables are stimulus categories, dependent variables are brain responses. The choice of stimulus type (factorial or parametrical) largely depends on the problem investigated. There are, however, more general choices

to make in the experimental design. These choices are motivated by the characteristics of the brain response signal, and their purpose is better capturing the properties of the response. One such choice is the choice between block or event-related design in fMRI experiments [195]. In blocked designs, stimuli are presented in blocks with short intervals between stimuli within the block and longer intervals between blocks. In event-related design, stimuli are presented as single events, but usually with varying intervals between events (duration is randomly distributed with the mean at 4 seconds, which is known in fMRI research as jittering). Each of these designs has its advantages with respect to capturing the signal. Block design has a good detection power to use with the fMRI BOLD signal: block duration of 12-20 seconds allows for the BOLD signal to reach certain strength. In event designs, the duration of stimulus presentation is less, but there is more variance in the responses to stimuli, which is an advantage for estimating the BOLD response shape. Jittered, i.e. varying-length intervals between stimuli are meant to better disentangle responses to different stimulus categories. Block or event related design with the stimulus presentation timings is the second characteristics of the experimental protocol design in fMRI. Time-locking of the stimulus presentation is important to be able to relate signal changes to the stimulus properties.

2.3 Functional Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) has become a powerful tool in neuroscience and a major workhorse in experimental research. This neuroimaging method is based on the magnetic properties of hydrogen protons that make up the majority of protons found in the brain. These protons spin around the axis that aligns with the direction of the strength of the scanner magnetic field - this phenomenon is known as precession. When a radio pulse is sent that resonates with the frequency of their spins, the energy transmitted gets the protons in the high energy state, and if no new impulses come, they start to come back slowly to the initial state, giving back energy: this is known as relaxation. Electromagnetic coils measure this energy that they give back - this is the signal that comes from the brain in the MRI scanner. In order to reconstruct 3D images, in the MRI scanner there are also gradient coils - transmitters that send impulses whose frequency varies in a constant way along one of the 3 axes - x, y or z. The signal is measured from the carefully timed sequence of gradient and radio pulses, and the sampling frequency in the MRI context is known as Repetition Time, or TR. In functional Magnetic Resonance Imaging (fMRI, see next paragpraph) TR needed to acquire the image of the whole brain

is usually around 2 seconds.

MRI allows us to get high quality anatomical images of the brain, and is widely used e.g. for diagnostic purposes. However, functional Magnetic Resonance Imaging (fMRI) is employed to get the images of the brain in action, i.e., while performing cognitive tasks. These tasks can involve processing of sensory information (visual, auditory, tactile etc.), motor control, memory formation, language, social interactions, emotion etc. FMRI is based on different magnetic properties of oxygenated and deoxygenated blood. When a certain area in the brain is active, neurons need energy for firing, and the cerebral blood flow rich in hemoglobin brings oxygen to the site of activity. This oxygen is then consumed, which results in higher concentration of deoxygenated blood in the active area. Deoxygenated blood is paramagnetic, and relaxation rate increases within the areas where it is concentrated. These differences in the magnetic field perturbation properties lead to signal variations that can be captured by the MRI coils. This signal is known as BOLD signal, or Blood Oxygenation Level Dependent signal. As we can localize the source of the signal with very good precision because of the gradient pulses, fMRI is known to have excellent spatial resolution. However, for the MRI signal to peak, a certain concentration of hemoglobin must be achieved, and this does not come before a lag of 6 seconds, as it takes time for the blood to be delivered to the working areas. This delay of 6 seconds is known as hemodynamic lag, and because of these particular properties of the BOLD signal, fMRI has rather poor temporal resolution. Another problem related to the BOLD sigal is its noisy nature: signal change in the activated area is very small, about 1-1.5 % [36] and subject to noise injected by head motion, physiological processes (respiration and heartbeat), thermal sources in electronics, baseline variations due to differences in vascular structures [97]. Signal losses and distortions can arise due to differences in magnetic susceptibility of various tissues, and indirect relationship between neural activity and hemodynamic response called neurovascular coupling can render spatial localization of activations less precise [252].

The introduction of functional MRI in the early 90s has had a huge impact on neuroscientific research. The number of papers using fMRI to investigate the functional organisation of the brain has been growing exponentially, so that we can state that it has become one of the fundamental tools of modern neuroscience. It is non-invasive, relatively easily available and relatively low-cost, but its main advantage is its high spatial resolution. In the course of the years technical issues related to the noisy signal, such as low contrast-to-noise ration, image distortions, signal dropouts have been successfully dealt with [97]. Functional MRI is also widely used for clinical purposes - such as pre-surgery mapping of

cognitive functions [97], [36], [44]. However, its clinical use is relatively small with respect to its significance in research: the method now is the mainstay in the experimental study of a wide range of cognitive capacities and their neural underpinnings [160].

In neuroimaging experiments the researcher tests participants (healthy subjects or patients) to know how the brain responds to certain types of stimulation. For fMRI it means looking into how the BOLD response reveals active areas in the brain as a result of experimental manipulation. Experimental procedures and protocols are defined by this purpose of investigation. For the researcher to successfully analyze the participants' data, it is crucial to establish the timings of the various stimuli to attribute the slow BOLD response [113]. Every neuroimaging study is designed with a stimulation protocol. It defines the stimulus categories and the time that the stimulus is delivered to the subject (known as onset), so that this information could be subsequently used in data analysis. An experimental session, i.e. the time that the subject passes in the scanner, typically lasts 60-90 minutes. For the data collection, tasks are organized in groups called runs. Each run lasts 7-10 minutes with short breaks between runs, so that the homogeneity of the data is not affected. Stimuli usually come in blocks (block design), or each stimulus can be regarded as a separate event (event-related design). Stimulus categories (e.g. images of faces vs. houses, positively or negatively charged emotional stimuli etc.) should be carefully balanced, that is, we should have equal quantitative representation of all categories in the protocol. The order of stimuli presentation is typically randomized (unless the design requires a certain presentation order). Stimulation protocol with all these characteristics (block design or event-related design, number of stimulus categories and number of stimuli in each category, onsets, randomization) is generally developed before the start of the neuroimaging study based on the research question. It is fixed for the whole experiment and does not change in its course.

The classical analysis of the experimentally obtained fMRI data is the univariate analysis. The time series of each voxel is analyzed separately under the null hypothesis that its time course is not affected by the stimulus presentation. To reveal task related effects, General Linear Model (GLM) approach is used to quantify linear correlations between the voxel time course and the reference model, where the box car function representing the experimental design is convolved with the hemodynamic response function [84], [169]. Each voxel is then assigned a probability to register this time course if the null hypothesis is true. Whole brain maps computed in this way with regard to the experimental manipulation are called statistical parametric maps [83]. Statistical inference can also be performed on contrasts between model parameters for two categories of stimuli, resulting

in the sets of voxels that selectively respond to a particular kind of stimulation [36]. This type of analysis allows to receive very precise answers to questions about the localization and mapping of various cognitive activities.

The introduction of the multivariate pattern analysis (MVPA) as an alternative to the traditional General Linear Model based univariate analysis in early 2000 allowed to obtain new insights into the functional organization of the cerebral cortex [106]. Unlike the univariate analysis performed voxelwise, MVPA takes into account the fine-grained pattern of activity of all voxels in a selected region. If the univariate analysis locates very specifically the source of activity in the brain, MVPA extracts information about the content of activity from the activity pattern in a particular area or in the whole brain - the content of activity can be both a stimulus category or a cognitive process. For this reason, the term "brain decoding" is often used as a synonim for MVPA. MVPA typically makes use of machine learning classifiers. The classifier (support vector machine (SVM), linear discriminant (LDA), k-nearest neighbour etc.) is trained on the samples of brain data along with labels, corresponding to the stimulus category. The samples of data are obtained based on the protocol timings and correspond to the duration of a block of stimuli or an event in event-related design. A 4D fMRI image reconstruction is a collection of time series coming from each brain voxel. Each time series represents a feature in the learning process. Next, the trained classifier is tested on brain activity data to predict the stimulus labels. Typically, for testing the cross-validation procedure known as "leave one run out" is used: the data of all runs but one is employed in training, and the classifier is tested on the remaining run, so that we consequently use one by one all runs of the session for testing. Whole brain decoding uses all time series as features. However, it is possible to effectively reduce the number of features by identifying most relevant regions of interest. At the group level, the mean cross-validation accuracy is used as a metrics that is tested for significance. To establish the subject-independent pattern at the group level cross-validation can be performed leaving out the data of one subject for testing. The procedure is performed on the whole dataset leaving out each subject once for testing, and sometimes is termed cross-subject decoding [126].

To localize areas containing information about the presented stimuli, Searchlight analysis is widely used to explore the spatial structure of the brain responses [140]. The Searchlight method is of a hybrid nature, combining features of both univariate and multivariate analyses. The outcome of the Searchlight analysis are maps, where each voxel is assigned some qualitative measure of information about the stimulus contained in the sphere centered around this voxel. The maps produced with Searchlight are of the same

nature as the maps obtained with the univariate GLM approach, but they are based on a more fine-grained pattern identification from multiple voxels and better reflect the spatial properties of the BOLD signal (that is, adjacent voxels have similar activation patterns). Most commonly used information measure to produce Searchlight maps is classification accuracy obtained with cross-validation. The maps are obtained in the following way. As a starting point of the analysis, a sphere is formed centering on a particular voxel containing all voxels around the center. Time series for all voxels in the sphere are extracted, and an SVM or sometimes an LDA classifier is trained on a part of this data. When it is tested on the new, unseen part of the data, the resulting cross-validation accuracy is assigned to the center of the sphere as an information measure. The procedure consists in probing all voxels, so that a whole brain map is produced as the output where each voxel is assigned a value. [2, 71]. Maps are first calculated individually for each subject and then pooled together to run group analysis. In the group analysis, the significance of the obtained values is tested at the group level. The most common method here is running t-tests voxelwise [244].

2.4 Real-time functional Magnetic Resonance Imaging

Functional magnetic resonance imaging in real-time was first introduced in 1995 [46]. The authors claimed that the method allowed to control for activations in real-time, thus assuring the quality of data acquisitions and choosing the best stimulation options [34], [270]. Real-time fMRI (rtfMRI) applications allow to access brain activations in the online mode by providing image analysis as the imaging process unfolds. The challenge of rtfMRI data processing consists in time constraints: all operations on the volume should be performed within the time that the next volume is acquired and reconstructed - usually around 2 seconds (1 TR). This is a significant limitation of rtfMRI as compared to offline analysis. However, in the past years, acquisition methods have improved significantly to permit this kind of access: higher-field MRI scanners were introduced, fast data acquisition sequences developed, methods of immediate image pre-processing have been proposed, statistical analysis techniques and activation visualization methods have been adapted to the real-time setting [261], [89], [240], [180]. All this made rtfMRI an important tool and also an important area of research in neuroscience and neuroimaging [241]. As a tool, real-time fMRI provides valuable insights on how the brain works by allowing to access and influence the ongoing brain activity [119].

Fig. 2.1 from [145] shows how real-time fMRI works. In rtfMRI acquisition process,

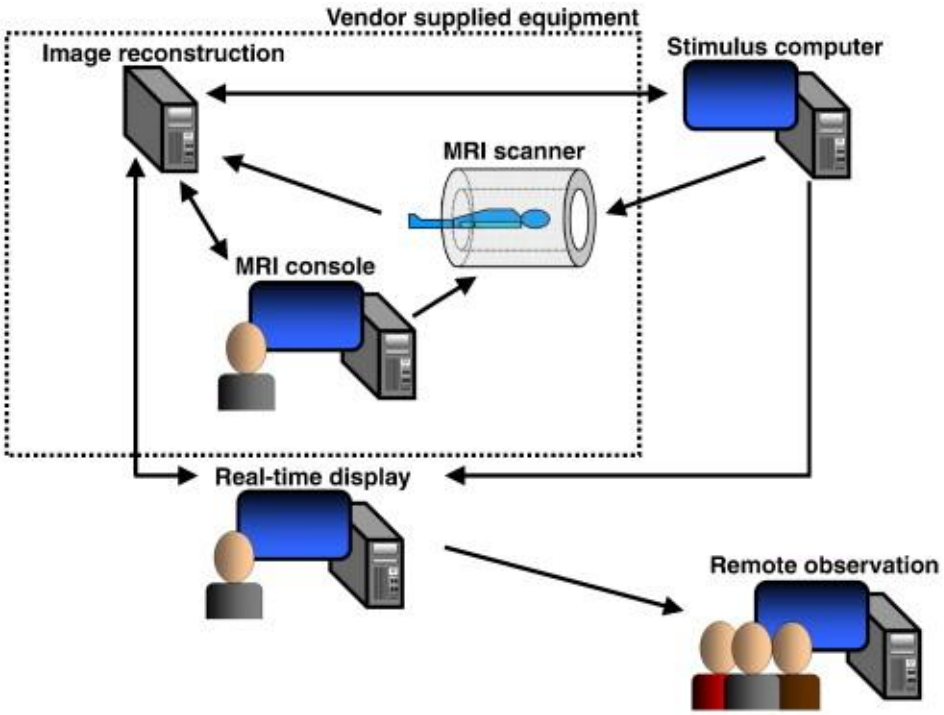


Figure 2.1: A scheme of a real-time fMRI setup. Stimuli are sent to the subject in the scanner, where images of their brain activity are acquired. The images are then reconstructed by the image reconstruction PC, and the process is managed by the MRI console. The reconstructed images are then sent to the real-time processing PC that performs preprocessing and analysis of the images. The results are shown on the real-time display.

the raw data from the MRI scanner are directly transfered to a real-time PC for further processing as soon as the volumes are acquired. Data processing needs to keep pace with the data acquisition, and now it is possible because of faster acquisition speed, special data transfer techniques, and computation capacities and algorithms [34]. Unlike in canonical fMRI studies, rtfMRI experiments do not have a fixed stimulation protocol. The stimulus presented to the subject depends on the outcome of signal analysis. When the rtfMRI was first introduced in 1995 [46], the authors saw its potential for creating interactive stimulation paradigms that open new perspectives for neurological investigations. This flexibility of design allows to approach a whole new broad range of questions about how we can interfere with the subject's brain activity or how the performance in cognitive tasks depends on the current brain state.

The fundamental challenge in real-time fMRI is related to the fact that at the moment of analysis the full data are not available as the acquisition is in progress [180]. To overcome this challenge, various incremental varieties of the traditional GLM method have been developed along with visualization techniques that allow to see gradually recomputed activation maps in real-time. In incremental GLM, model parameters are reestimated along with the arrival of new data [261], [109], [274], [57]. An alternative method consists in computing correlations between the measured signal and the reference model vector in a "sliding window" manner: the computations involve only a few latest datapoints [89].

MVPA is also used in rtfMRI: machine learning algorithms are applied to decode the subject's brain state volume by volume. However, the cross-validation procedure is not applicable in real-time settings unlike in offline analysis. There is a variety of scenarios of how training and testing can be organized in the decoding pipeline. In many studies, the classifier is trained offline on the data obtained in a pre-real time session and only tested on the data obtained in the real-time session. This is advantageous because it allows to fit the analysis into a 2 second interval, but can carry unwanted batch effects between sessions. An example of such as study is [225]: they present an FSL integrated BCI toolbox, where, however, there is only the option for offline classifier training. Another extreme scenario is to carry out both training and testing online. In this case, the experiment can be performed in a single session, but the classifier has to be retrained with each volume. There is the danger that computations will increase with each step. The methodology of online learning is meant to tackle this challenge. The studies of this type were carried out, for instance by [238]. The authors decoded the subjects' emotional states (happiness, anger and disgust) in real time using a procedure they developed for feature selection that they termed Effect mapping. The SVM classifier was also trained online.

In the mixed scenarios, the classifier can be trained offline after the pre-rt session and adapted online before carrying out predictions. In this way, the time-sensitive computations can be carried out properly and also batch effects can be neutralized. An example of a recent real-time decoding experiment is the study by [188]. In this study, the classifier was trained on the fMRI data from subjects, watching images of faces and places. Next, in the real-time it was adapted to decode the subject's attention either to the face image or to the place image in superimposed images. The mixed scenario at the moment is the most common one among real-time decoding experiments because of its advantages.

In real-time settings data processing is usually carried out on a single subject basis. However, a fascinating line of research is the use of multiple subject data in the real-time studies. That means, that the classifier can be trained on the data from X subjects and tested on X+1 subject data. That can be advantageous if the training data with X subjects can be collected offline and only tested in real time with subject number X+1. As an example I can cite the study by [210] where the authors propose a toolbox for subject independent classification. For example, when working with patients and neurofeedback, the toolbox allows you to avoid training the classifier on the abnormal patient's brain activity. Instead, it can be trained on healthy brains and used for real-time classification and feedback with patients. They successfully trained a classifier on a group of healthy volunteers to distinguish between happy and disgust emotional states, and then demonstrated its robust performance on participants' data who were not part of the training group.

As has been mentioned above, the principal advantage of fMRI is its high spatial resolution. As we can access brain activity in an online mode, targeting a certain area with high precision, rtfMRI has a huge potential for practical applications in spite of its slow signal. Examples of such applications include Brain State Dependent Stimulation (BSDS), Brain-Computer interfaces (BCI) and neurofeedback (NF). In BCI the subject controls external devices with her brain: attention-based spellcheckers, robotic hands, game avatars etc. [8], [222]. In neurofeedback applications, the subject learns to upregulate or downregulate the activity in a particular area of the brain. Usually, neurofeedback experiments proceed like this: there is a region of interest (ROI) that is known for its involvement in a certain cognitive activity, like, for instance, emotion or pain processing (e.g. [268]). The subject in the scanner performs some cognitive task, involving this area, and gets feedback on the screen based on the registered changes in the activity in the target area (for instance, in the form of a temperature bar). In this way, it is claimed that the subject can create lasting change in her brain activity pattern which in turn

leads to cognitive improvements. Neurofeedback is being more and more widely used for clinical purposes because of its preceived benefits. Notably numerous are studies on major depression patients that were claimed to show clinical improvements as a result of the neurofeedback intervention [279], [283].

Apart from clinical uses presented above, real-time fMRI is also a valuable tool in research. With rtfMRI we can ask questions that can hardly be answered with offline analysis which opens up new avenues of research. A very important type of questions we can answer with rtfMRI is about how the current brain activity affects the subsequent reaction to stimulation. RtfMRI creates a new paradigm in experimental studies [268]. In the majority of traditional neuroscientific studies, the brain state is manipulated as a dependent variable with the stimulation that is the independent variable. Real-time fMRI allows to study the effect of brain state on behavior - for instance, in decision making and action (motion) research the subject, depending on her state, can take different decisions or engage in different types of motion. In this case, the relationship between the dependent and independent variables is swapped: the brain state becomes an independent variable, while the behavior becomes a dependent variable (ibid.). This approach known as Brain State Dependent Stimulation (BSDS) and it consists in delivering the stimulus in the moment when a particular state is detected in the brain. In the clinical context, that means delivering a physical pulse (magnetic or electrical) to elicit desired changes in the brain activity [178].

So, we can see that in spite of the delay between the activity in the brain and the moment it is captured in the data, real-time fMRI is a very helpful tool for practical applications and also an important contribution to the research.

In the next sections I am going to give a more detailed overview of the main three applications of rtfMRI: BSDS, BCI and NF. However, each of these is a major direction of research, and could be worth writing a dedicated chapter, if not a book, on the topic. In my presentation, I will only highlight a few points, first, to understand the concept, and second, to put my research in a context.

2.4.1 Brain State Dependent Stimulation

The main principle of brain-state dependent stimulation consists in introducing stimuli in real-time as the function of the current brain state [119]. In the clinical context, certain brain states are recognized as beneficent for delivering stimulation that can actually elicit or enhance positive changes. In this case, stimulation does not consist of just cognitive stimuli, but takes the form of physical interventions into the brain activity, such

as Transcranial Magnetic Stimulation (TMS) or Transcranial Direct Current Stimulation (TDCS). In [178] several interactive real-time TMS protocols are reviewed for the treatment of major depression and schizophrenia. The patients perform tasks involving emotional response so that they can get into the state where a specific EEG pattern of alpha rhythm is recognized. It is in this moment that the TMS pulse has to be delivered for it to create the desired effect on the brain. In [167] several studies are listed that use TDCS in the same manner to treat alcohol and drug addictions. Another clinical context is neurorehabilitation: in [95] TMS was applied concurrently with motor-imagery to induce specific neural plasticity during motor training for neurorehabilitation of a patient lacking residual hand function. The authors significantly increased the excitability of the stimulated area by coupling TMS pulses to ipsilateral sensorimotor desynchronization during motor-imagery, while this effect was not observed in non-BSDS protocols.

In the research context, the idea is to investigate the functional role of brain states by delivering stimuli in real time, depending on the actual physiological state of the subject's brain. The differences in behavioural performance resulting from different brain states at the moment of stimulus presentation are attributed to the impact of the moment-to-moment changes in the brain activity. A series of neuroscientific studies on perception was driven by the interest in brain-state fluctuations and how the pre-stimulus state affects perception. These studies mostly concentrated on alpha oscillations recorded with EEG and reported their effect on perceiveing an induced phosphene - light dots perceived in the absense of light [219], [104]. In another series of studies, the effect of phase of alpha oscillations on perception was investigated: visual stimuli were flashed with a short delay, and were delivered to the subject in various phases of alpha oscillations, depending on which they were perceived either as two flashes or as a moving light [96]. Also response to auditory stimuli was shown to be determined by pre-stimulus alpha phase [143]. However, perception is affected not only by pre-stimulus fluctuations. It has been shown that the brain develops a predictive code for perceptual events that are most likely to occur [251]. This is due to a strong autocorrelation in the perceived signal, e.g. the scene you are seeing usually does not change fast enough to generate a completely new sensory input. The role of such a predictive signal would be to distribute the computational burden between pre-stimulus and post-stimulus processing, thereby limiting post-stimulus processing: once a stimulus has been presented, bottom-up sensory information is matched to a predictive code rather than being processed anew in feedforward succession. Predictive coding implies that the brain anticipates the forthcoming sensory environment, generating a template against which to match observed sensory evidence. In [250] a neural

representation of predicted perception was observed in the medial frontal cortex, while human subjects decided whether visual objects were faces or not. Moreover, perceptual decisions about faces were associated with an increase in top-down connectivity from the frontal cortex to face-sensitive visual areas, consistent with the hypothesis of matching of predicted and observed evidence. But this kind of predictive coding for the expected stimulus is just one of many other top-down processes in the brain that can be fruitfully investigated with BSDS; for instance, activity state has also been shown to have impact on higher cognitive processes, such as memory [274]. In this study, parahippocampal area of the subjects was monitored with real-time fMRI, and different states were recognized as "good" or "bad" for memory formation. The authors showed images of the natural scenes to the subjects both when they were in the "good" and in the "bad" state, and recall for scenes shown in the "good" states was significantly better.

Most important among these top-down processes is top-down attention. Top-down and bottom-up are two types of attentional processes distinguished in neuroscience [130]. Bottom-up attentional mechanisms are activated by the salient stimuli in the immediate environment, e.g. bright light or loud noise. Top-down attentional mechanisms, on the contrary, are those that support our readiness to perceive certain stimuli because of their inner significance for the person. Top-down attention is the key mechanism that prioritizes processing some stimuli over other by restructuring cortical information flow, thus sorting out behaviorally relevant over irrelevant and distracting information. The behavioral consequences of attentional restructuring of information flow are striking. Attended sensory inputs are processed more rapidly and accurately and with higher spatial resolution and sensitivity for fine changes, while nonattended information appears lower in contrast and is sometimes not perceived at all [88]. One of the most fundamental questions in neuroscience is about establishing neural correlates of these top-down attentional patterns because of their extermely important role in perception. BSDS is a powerful paradigm to investigate this question: if we formulate a hypothesis about these patterns, we can test the hypothesis by presenting the stimuli when the relevant state is detected; in this way, we can demonstrate behavioural differences between "in-state" trials and "out of state" trials, thus validating the hypothesis.

So, BSDS is a valuable technique not only for clinical applications, but also in neuroscientific research context. However, the success of this approach heavily relies on the answer to the following question: what reliable brain activity markers could be used to detect the relevant state? The state-of-the-art answer to this question is that of pattern analysis-based decoding of neuroimaging data, or supervised machine learning (e.g. clas-

sification and regression) and brain state prediction. Decoding fMRI data in real time allows for the modulation of stimuli in relation to brain states, that is, the decoded state is used as a control signal to adapt the stimulus [144]. An example of the use of decoding to adjust stimulation in real-time is [56]: the subjects were shown two overimposed images and were asked to attend to one of them in a sustained manner; in the meantime, a logistic regression classifier was run to monitor their attention state. When an attention lapse was detected (the subject was found to attend to the wrong image), a more difficult stimulus was shown for the same task, with less strong signal.

The set of studies mentioned above illustrates how BSDS has been applied in clinical context and in research in order to gain new insight into the functional role of ongoing brain activity. Given the growing interest in the online stimulation based on brain states, BSDS is likely to become a more frequently used tool in cognitive neuroscience. Most importantly, BSDS informs other types of real-time experimental studies - BCI and neurofeedback - that are going to be overviewed in the next two sections.

2.4.2 Brain Computer Interfaces

Brain-Computer Interface (BCI) is a system that allows for an external device to be controlled directly by the brain signal, without other mechanical or physical intervention. The primary role of BCIs is practical and clinical: augmenting the possibiltieis of interaction with the environment for people with disabilities [39]. BCIs can have either assistive or restorative function [94]. In the former case, they replace the function permanently lost to disability, for instance, the communication function. Assistive spellers belong to this category [1, 256]. In the latter, case, BCI is used as a neurorehabilitation device helping to restore the impaired or temporarily lost function, such as limb mobility. In [275] stroke patients with severe motor deficits practiced rehabilitation exercises with the assistance of a BCI and showed significant clinical improvements.

Apart from restorative and assistive BCIs for neuroprostethics and communication, there are also other types: gaming (providing augmented reality) or probing into consiousness of vegetative state patients. External devices that are controlled by the brain signal can be robotic arms, spellers, game avatars [182] or even robots [7]. EEG is the neuroimaging modality largely used for BCIs because of its high temporal resolution. Besides, fNIRS can also be used because it is portable [165]. There is a debate on whether it makes sense to use fMRI BCIs because of the slow fMRI signal and hemodynamic response delay [152, 237, 269]. However, fMRI still allows for some types of BCI and besides, turns out to be a valuable tool to train potential BCI users to successfully use BCIs [236], [222].

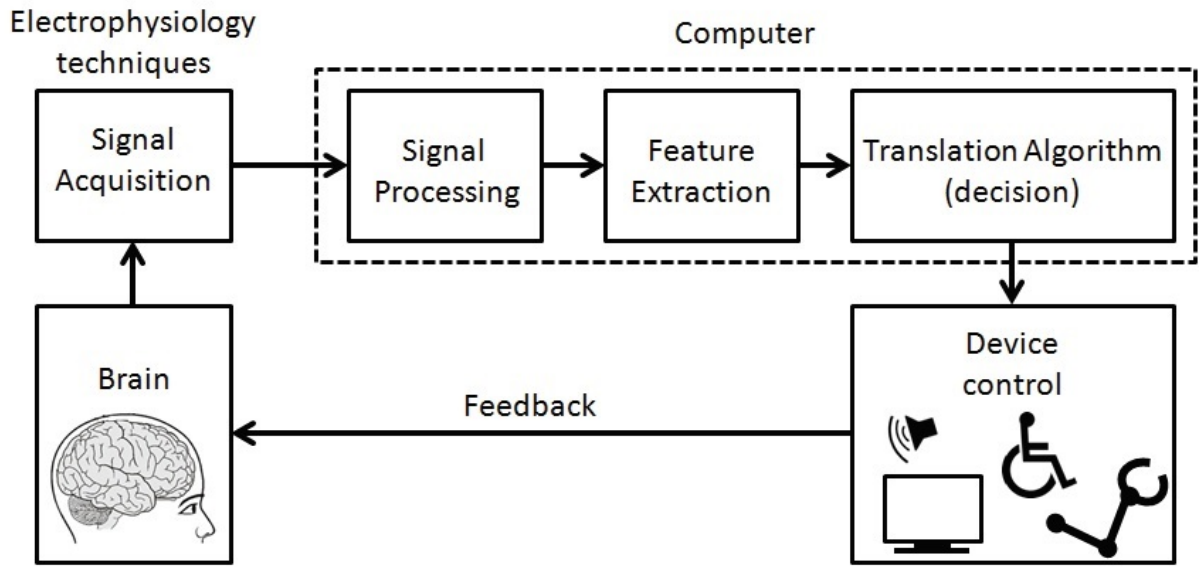


Figure 2.2: A typical BCI system translating input from brain signals into commands to an external device - from [223].

BCIs are designed as any communication or control system: they take input signal (i.e. electrophysiological activity) and transform it into output (i.e. feedbacks, device commands). Translating input into output involves such intermediate steps as signal acquisition, preprocessing (or signal enhancement), feature extraction, classification. The protocol that determines timing and decision-making on the interaction is part of the device's control interface [223]. A typical BCI system is shown in Fig. 2.2

BCI works because it recognizes a specific set of patterns in brain signals necessary to execute control, mostly they rely on electrophysiological brain signals. Decoding brain signals in real time is challenging: one has to be capable of making efficient and robust inference online based on very limited, complex and noisy observations. Large efforts have recently been put into developing and improving signal processing, feature selection and classification methods [60], [212], [133], as well as acquisition hardware techniques and dedicated software environments [114, 146, 282].

Methodological and technical progress has largely solved the challenge of processing these signals online. The main issue that remains, however, is the identification of a reliable mapping between electrophysiological (or other neuroimaging) measures and relevant states of mind. This is a common challenge for all real-time neuroimaging applications,

and this is why BCIs, as well as neurofeedback and brain-state dependent stimulation, are highly dependent upon advances in cognitive neuroscience and neuroimaging research. Actually, if BCI research has not yet produced many successful new therapies with a routine application, it is exactly because of this challenge of interpreting brain activation from the signal given the hidden neural code [223]. BCI can be an important platform for the new experiments on brain and behaviour that can inform neuroscientific research about this problem of mapping neural code to behaviour. On the one hand, BCIs need to decode a wide range of cognitive activities of high order [254] (speech segments, the desired object kinematics, motor planning, goals of eye movements, the internal representations of auditory and visual percepts, and decisions being made by the subject), on the other, they have to choose the strongest and most reliable signals for that [119]. This is a convenient testbed for neuroscientific hypotheses about neural underpinnings of these processes. Besides, certain areas of BCI research, such as motor imagery or covert attention [8], contribute to the neuroscientific understanding of these processes as much as to neurorehabilitation purposes, if not more [119].

Apart from this challenge, common for all real-time neuroimaging applications, BCIs present a very particular challenge arising due to the fact that there are two adaptive controllers, the computer system and the user's brain. BCI signal features will be affected by the device commands they are translated into: computer outputs will affect computer inputs [76]. Inappropriate action choice on the part of the computer could seriously impair performance of BCI protocols. Because the BCI interaction involved these two adaptive controllers, the users brain and the computer system, its design is among the most difficult problems confronting BCI research [223].

BCI context could significantcly modify the way that we will conduct future experimental studies, because it provides a dynamic, interactive environment for basic neuroscience that calls for computational models of psychophysiological functional mechanisms to ensure efficient investigation of neuroimaging markers involved [85].

2.4.3 Neurofeedback

Humans regulate their behaviour receiving feedback from the environment: this is the fundamental learning mechanism. When the feedback is positive (the individual receives an expected reward for her actions), the behaviour is reinforced. In case of negative feedback (the reward is not up to expectations) the behaviour is adjusted to get the expected reward. This learning mechanism is employed in biofeedback interventions when the person learns to regulate her physiological processes such as heartbeat. A case of

biofeedback I am going to analyze here is neurofeedback, when the physiological process a person learns to regulate is her brain activity [174]. The goal of neurofeedback is inducing behavioral changes that result from regulating the brain activity. These changes can be twofold. In clinical populations, neurofeedback is meant to revert some abnormal functioning by retraining the brain and bringing it towards the desired functioning. Examples of conditions that have been treated with neurofeedback include affective disorders [278], chronic pain [41], unhealthy eating patterns [257], addictions [47]. In healthy subjects, neurofeedback helps to improve performance on various cognitive tasks - e.g. working memory [281] or perceptual sensitivity [227, 232].

Neurofeedback keeps gaining interest both from clinical and research perspectives. In clinical perspective, neurofeedback studies are inspired by the reported success in treating such conditions as major depression, post-traumatic stress disorder (PTSD), autistic spectrum disorders and schizophrenia [91,197], so it could be an alternative to medication. From research viewpoint, neurofeedback is an exciting paradigm for studying how brain state affects behaviour [239, 266].

Although there is great variety in neuroimaging modalities used to provide neurofeedback (EEG, fNIRS, fMRI etc), signal processing methods and means of providing feedback (visual, auditory, tactile etc), the scheme of a neurofeedback experiment can be generalized. This generalized scheme is shown in Fig. 2.3 from [239]. First, a region of interest is chosen to target based on the knowledge about its functions. For instance, in the treatment of affective disorders, amygdala is often chosen as target ROI. In targeting cravings, the ROI selected can be the anterior cingulate cortex because it is known for its involvement. Next, the signal from this region is obtained by the neuroimaging setup - fMRI, EEG, fNIRS. The signal is then preprocessed by the state-of-the-art techniques of the chosen modality. In fMRI feedback, the signal amplitude is targeted as the performance measure. In EEG, the targeted signal parameter is most often a certain frequency in the frequency domain. Then the signal is analyzed and converted to some form that can be presented to the subject - e.g. changing temperature bar or auditory signal pitch. As an optional step, the signal can be compared against the desired performance or targeted state and only then converted to the feedback form. Visual feedback usually is shown as a temperature bar that goes up or down dependening on how well the subject is doing in reaching the desired state of brain regulation. Auditory feedback is typically presented as a sequence of sounds varying in pitch, so that higher pitch signals more success in brain regulation. Sometimes neurofeedback is given in the form of gaming environment or virtual reality [10].

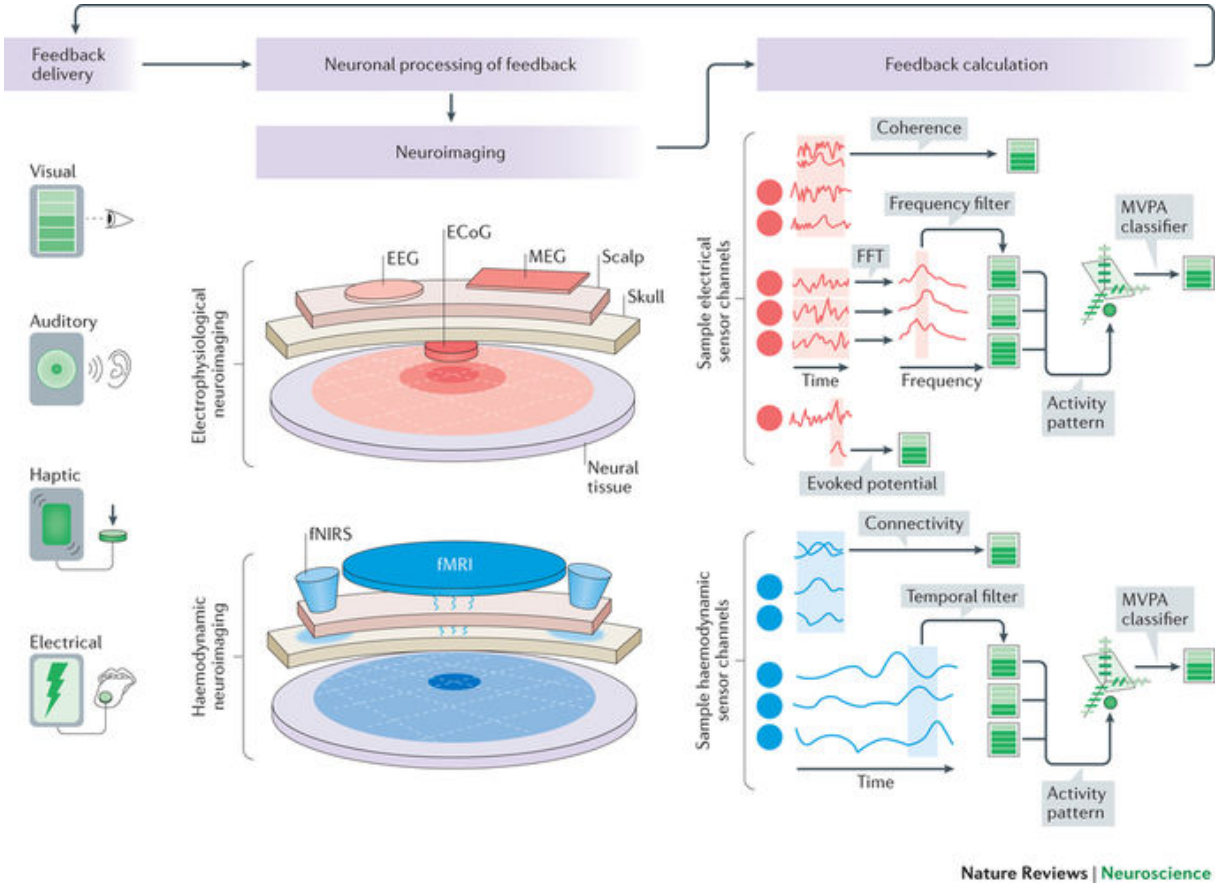


Figure 2.3: Generalized representation of a neurofeedback experiment (from [239]). First, neural activity is detected and observed with neuroimaging methods (EEG, MEG or invasive electrocorticography (ECoG); hemodynamic imaging methods include functional MRI (fMRI) and functional near-infrared spectroscopy (fNIRS)). The grids on canonical neural tissues provide a qualitative reference for the relative spatial resolutions of the various imaging technologies. Sample signals that are extracted from the sensor channels provide a qualitative representation of the difference in temporal resolution. Then signal is extracted; calculation of coherence or connectivity between two channels as a measure of functional connectivity is another common feedback method. Features from a set of sensors, such as the power at a frequency window or the level of activation, can be classified as multivariate patterns of activity (MVPAs). The calculated measure of signal quality is then presented to the individual via visual, auditory, haptic or electrical stimulation feedback, allowing the user to alter neural function and complete the loop.

Neurofeedback experiments can employ a plethora of methods for signal processing. In its simplest form, the raw signal is taken after some necessary preprocessing. To minimize the impact of random fluctuations, Kalman filter can be applied to the signal to reflect only those changes that correspond to the regulation effort [138]. At the next stage of complexity, multivariate decoding can be used to detect the target brain state [231]. Neurofeedback can be given based on the signal from a single ROI or it can reflect the changes in brain networks if correlations between ROIS or other measures of involvement of several ROIs are taken into account [136].

The choice of the neuroimaging setup depends on the neurofeedback scope. Usually, when a particular ROI is targeted, fMRI is chosen for its spatial precision and inspite of its slow signal (then it has to be accomodated in the experimental setup). For instance, nearly all affective disorder neurofeedback interventions are fMRI based because of the need to target amygdala (only a recent study found the way to do it with EEG [276]) [190,198,271]. EEG neurofeedback is preferable when a particular frequency has to be targeted [131,202], as, for instance, in experiments with PTSD [280].

Although there have been a lot of studies showing the effectivenes of neurofeedback interventions [197], it is still in the center of debate [255]. First, not all conditions can be efficiently treated with neurofeedback - a case in point is ADHD [285]. Second, there is a question about the lasting character of the changes induced by the neurofeedback [216]. Third, it is not quite clear what are the prerequisites for the neurofeedback treatment to succeed. For instance, a large percent of subjects cannot learn to regulate their brain activity on their own and need assistance because they cannot work out the strategy of the regulation when left to themselves [22]. However, little is known about what the "winning" strategy should really be like. Still, the number of studies demonstrating the benefits of neurofeedback is large enough, and it is expected that the research in this area will be quite intensive. The field is also moving in the direction of combining different neuroimaging modalities (like EEG and fMRI) to improve the quality of the feedback interventions [172, 206].

2.5 Motivation of the study: reasons for protocol adaptation

Neuroimaging techniques help address neuroscientific questions about the workings of the brain. However, the difference between classic neuroimaging and real-time neuroimaging does not just lie in the differences of data acquisition. These two methods answer two fundamentally different scientific questions. The first one is about how brain reacts to

stimuli where the researcher manipulates the dependent variable (measured activity) with experimental stimuli (independent variable). Standard experimental design is meant to study the brain response to external stimulation and provide data on which robust inference can be performed. The foundation of design in this case, as has been said, is that the stimulation categories are the values of the independent variable while measured brain response becomes the dependent variable. As real-time neuroimaging is closer to natural environment and is not restricted to controlled lab environment, real-time analysis of brain activity is strongly driven by the development of practical applications. However, real-time brain imaging is also a testbed for a new research paradygm of investigation into the relationship between brain activity and behaviour where brain activity is the independent variable and cognitive performance and/or behaviour are dependent variables [239]. In real-time experiments, the question is about how brain state conditions reactions to stimuli, and the relationship between the variables is reversed. This change in the experimental paradigm requires a different experimental design, a design where the stimulation can be adapted to the changes in brain activity in real time based on the experimental goal. Real-time experimentation methodologically is not meant for data collection with the purpose of conducting inference in the same way as offline analysis. For one, it is much more single-subject oriented [246]. Group inference also is mainly based on between group comparison. It should allow to investigate a different, fundamental research question about how current brain activity affects behaviour. To create experimentation investigating into that question we need new protocols that can be updated in the real-time mode based on the ongoing analysis of the brain activity. Protocol adaptation in real-time neuroimaging is the tool of interacting with the environment in a controlled and optimized fashion taking into account dynamic aspects of brain responses - such as learning, adaptation, plasticity etc.

2.5.1 Adaptive protocols for BSDS

In BSDS, the goal is to demonstrate how performance is different between stimulation sent when the subject is in the target state and the random stimulation or stimulation when the target state is not present. To be able to perform inference on this data, the researcher has to possess robust methods for the state identification, so that the behavioural differences in the aforementioned types of trials could be reliably attributed to different brain states. Identifying the brain state implies mapping a certain pattern of activity to some cognitive state label - e.g., emotional or attentional state. The state-of-the-art of establishing such mapping is Multivariate Pattern Analysis (MVPA), or brain decoding [106]. This analysis

consists in using Machine Learning classifiers in a supervised manner: the classifier is trained on some portion of data labeled corresponding to the detectable states and then tested on the unseen portion of the data to predict the state label. Brain volumes are the input to a trained model and the outputs are the predicted brain states. However, fMRI data are rather unique from the point of view of machine learning: not only is the number of features relatively large (current whole-brain studies can have 10 to 30,000+ voxels), but also the relative ratio of features to observations (ranging roughly from 100 to 1000 time samples) is atypical of many machine learning problems. This is known as the curse of dimensionality. To cope with this issue in MVPA, further advances are needed in visualizing and interpreting predictive fMRI models, evaluating preprocessing issues, maximizing information extraction, and characterizing signal-to-noise properties of fMRI data [144].

2.5.2 Closed loop adaptive protocols

In BCI and neurofeedback, detecting the current subject state is just the first step of the process. The real challenge lies in the fact that the subject's brain actively interacts with the stimulation system, and the stimulation has to be constantly updated to reach the experimentation goal. In the course of interaction either the subject adapts to the external device (BCI, neurofeedback), or we try to adapt to the subject's brain to drive behavioural changes (neurofeedback) [82]. We want the subject to alter her brain activity in order to affect the behaviour in some way. In standard neurofeedback situations, the subject is shown the feedback and is supposed to figure out a strategy to perform better. However, not all subjects are capable of learning to do it. In this case, we should exert some form of control in order for the subject to reach the state so that the behaviour changes in the desired way. Instead of the linear, reductionistic "stimulate and record response" approach, a more modern approach is taking hold: closed-loop neuroscience. It respects the inherent loopiness of neural circuits, and the fact that the nervous system is embodied, and embedded in an environment. By interposing our own technology in some of these loops, we can achieve unprecedented control over the system being studied and explore the functional consequences. In closed loop stimulation, the researcher gets the feedback from the subject by analyzing their activity and chooses the next stimulus for the subject to move towards the target state [209], [284]. In the next few sections we will analyze methods and techniques proposed in the literature for protocol adaptation to discuss which of them have been used in this particular kind of closed loop context that interacts with the subject's brain state.

In the majority of works discussed below, the problem of protocol adaptation can be regarded within the framework of optimization approach. The optimization problem in other words is the problem of "finding an optimal decision x in order to maximize a measure $goodness(x)$" [19]:2. So, in this case we have to define our choice of options $x = \{x_1, \ldots, x_n\}$, work out a quantative assessment of the "goodness" function and hand over the goodness model to am algorithm that will search for an optimal choice at each step. In case of neuroimaging experimental setups, x is a set of experimental design choices that we expect to have impact on the activity we study. This could be the list of stimulus categories (e.g. pictures of people, houses, food and instruments) or stimulus features (e.g. the degree of stimulus noisiness - 50, 25, 15, 5 per cent). This can be also the space of timing choices (e.g. the moment at which the stimulus is delivered).

Now, how can we measure the "goodness"? In [184]:3 the problem of design optimization is formulated as "finding a design d^* that maximizes some utility function $U(d)$ that quantifies the quality of design d".

$$d^* = \arg\max_d U(d)$$

The "goodness" is defined with regard to the experiment scope: it is the design choice that allows us to best attain the experimentation goal. Protocol adaptation in the studies discussed below varies with respect to optimization purposes (ROI selection, model selection, data quality) and methods (Information theory, Optimization algorithms, Genetic algorithms) used.

2.6 Adaptation based on Information Theory

Adaptive optimization procedure has been proposed recently in a series of works, mainly for EEG experiments [184], [224], [53], [40]. It has been termed Adaptive Design Optimization (ADO). The idea behind this framework is finding a design choice that at each timepoint provides best discrimination between competing models of the subject's activity in response to stimuli. It offeres us a measure that quantifies the utility of each design choice from the viewpoint of model selection and thus makes the sampling procedure more efficient. So, in [184]:3 the utility function is presented as "measure of model dissimilarity or distinguishability between two models" that each design choice provides. The utility function will be the larger, the more the models are dissimilar to each other. Generally, ADO works as follows: we have some evolution function that describes a sequence of hidden states, in our case, it is the dynamics of the subjects' behaviour. ADO chooses

the design options that have most statistical power, i.e. allow to identify the underlying model with minimal error.

The purpose of experimentation in this case is discriminating between candidate theories of mappings between physiological markers and mental states. As the starting point for the design optimization, each model under consideration should be formulated as a "statistical model defined as a parametric family of probability distributions, $p(y|theta, d)$, each of which specifies the probability of observing an experimental outcome (y) given a parameter value (theta) and a design value (d). That is, the data vector y is a sample drawn from the probability distribution (often called the sampling distribution) for given values of the parameter vector theta and of the design vector d" [184]: 6. The subsequent procedure is a sequential process consisting of a series of optimization-experimentation stages. First, the optimal design is chosen based on the current state of knowledge encoded in the prior. As ADO is essentially a Bayesian framework, it requires definition of priors for model parameters. Next, this design is used to obtain a sample of the data. Based on the observed outcomes, the priors are then updated to the posterior, which in turn becomes the prior on the next iteration of the optimization process. As the design is optimized at each step, it involves solving an optimization problem described in [184], where the optimal design parameter maximizes the utility function as follows:

$$d^* = \arg\max_d U(d)$$

However, in [224], [53] a revision of the design optimization procedure has been proposed. It involves calculating the Chernoff Laplace bound for each design choice as an approximation of error due to the wrong model choice (and consequently of the utility function as well). This procedure does not involve the optimization of the utility function. The Chernoff-Laplace bound is an information theoretical criterion that yields upper and lower bounds to the error due to selection of the wrong model. So if P_e is the probability of selecting a wrong model, it depends both on the data sampled (y) and the experimental design (u) used to sample the data. It can be written as

$$P_c = 1 - \max_M p(M|y, u)$$

The task of design optimization, according to [224]: is then "to reduce the effect of data sampling process upon the overall probability of selecting the wrong model". In other words, we want to optimize the design, minimizing the model selection error $E_{p(y|u)}$:

$$u^* = \arg\min_u E_{p(y|u)}[P_c]$$

The authors claim that the Chernoff bound they propose to minimize here can be efficiently computed in the online experiments.

So, the optimization procedure proposed in [184], [224], [53], can be generalized as follows:

1. Running the variational Bayes (VB) inference (or Bayesian logarithmic factor like in [137]) For each of the concurring models based on past observations and experimental design variables;

2. Updating the model priors to posteriors calculated with the obtained data samples;

3. Computing the Laplace-Chernoff bound for each possible value of the experimental design variable, u, to assess its efficiency in the design;

4. Selecting the optimal design for the next trial [224].

The ADO procedure and its differences from the standard protocol are summarized in Fig. 2.4 from [224]. ADO framework described in the previous section allows us to adapt stimlation also in an online manner to sample more efficiently for the purposes of hypotheses testing.

2.7 Optimization for design and data quality

So, as has been mentioned above, the general idea behind design adaptation is that we have a space of experimental design parameters and an algorithm that explores it in order to choose the optimal one with the experimental scope in mind. In the previous example, the goal was selecting the model that best fits the observed data, and the criterion for exploration of the design space was based on information theory: the design parameter was selected that was most informative about the choice of the model. In the next sections, we will examine protocol adaptation methods that aim at optimizing various constituents of the experimental design: statistical power, ROI selection, data quality (contrast). These methods are based on various optimization algorithms that I will present in the next sections. For each method, I will try to highlight the optimization objective, as well as input to the optimization algorithm, measure of fit and algorithm details. The objective is to demonstrate the range of available methods and to position our research among them.

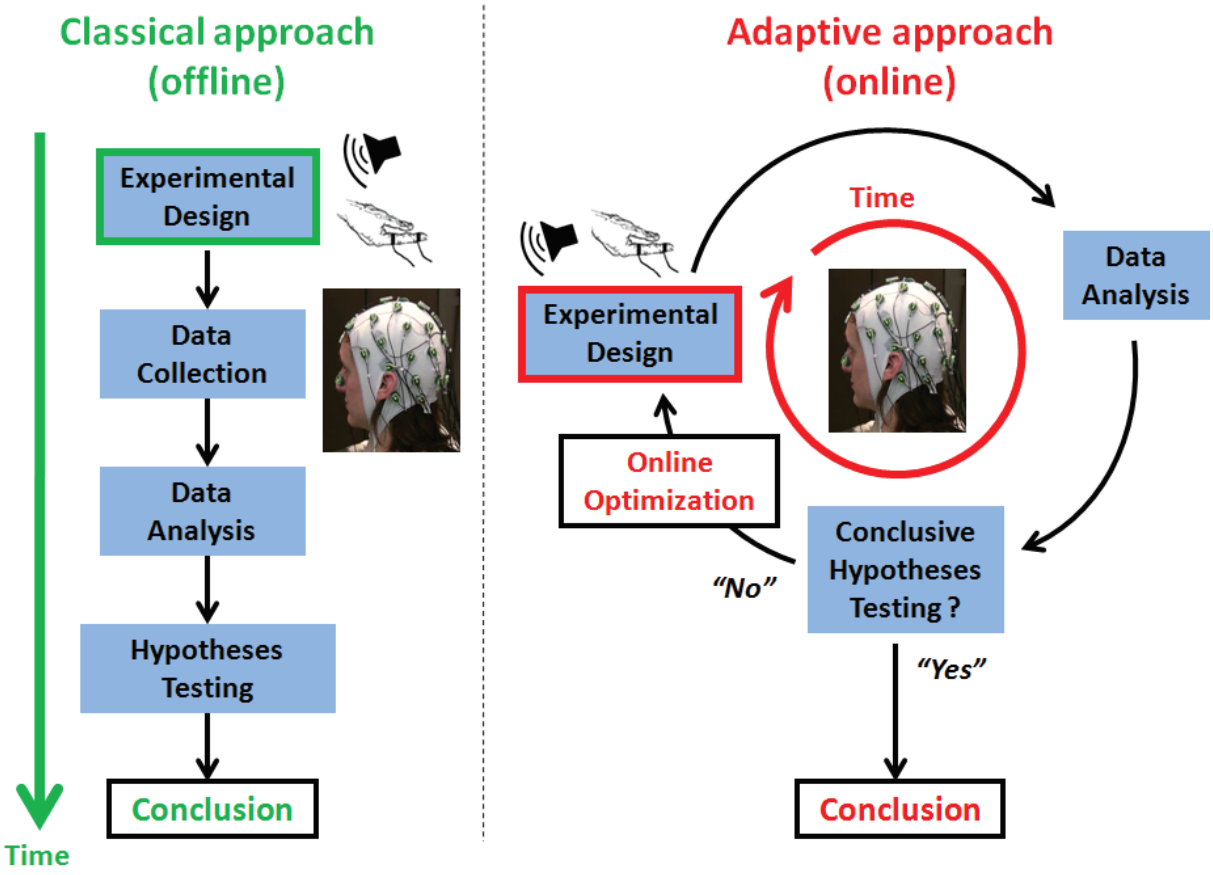


Figure 2.4: Left: sequential ordering of traditional steps of experimental data analysis. Right: adaptive approach with design optimization online for the purposes of hypothesis testing.

Genetic algorithms are a subset of a larger class of optimization algorithms, called evolutionary algorithms, which apply evolutionary principles in the search through high-dimensional problem spaces. Genetic algorithms code designs or candidate solutions to a problem as a digital chromosomea vector of numbers in which each number represents a dimension of the search space, and the value of the number represents the value of that parameter [111]. The elements of a design vector in a genetic algorithm are parameters specifying how to construct an object, analogous with genes of a biological chromosome. In [262] genetic algorithms were applied to find optimal sequences for event-related design in fMRI. Event-related design of fMRI experiments is powerful, but is somewhat mitigated by serious limitations posed by the shape and linearity of the BOLD response. The statistical power of effects in rapid event-related fMRI depends greatly on the timing parameters, the particular sequence of events chosen, and how these variables interact with signal nonlinearities, temporal noise autocorrelation, and psychological factors relating to the consistency and magnitude of the expected neural activity across trials. So, finding optimal event-related design could seriously improve the quality of the data for the inference purposes.

In fMRI design, the chromosome is a vector of N time bins, and the values in each slot represent the type of stimulus presented at that time (or the absense of it). Thus, the space of design parameter is created by temporal parameters: interstimulus intervals and jitter - varying interstimulus intervals, which is an important feature of event-related design. The design matrix is constructed based on the BOLD model taking into account signal autocorrelations for the contrast of stimuli.

The next feature of the genetic algorithms to be selected is a numerical measure of fitness on which the selection of design operates. Design vectors must be translatable into single numbers that represent the suitability of the design. In fMRI, there are several metrics that the researcher might want to optimize: e.g. maximize the efficiency of estimation of several contrasts and the efficiency of estimating the hemodynamic response for statistical reasons, minimizing at the same time stimulus predictability, so that the subject could not guess the stimulus to come.

The genetic algorithms operation consists of three processes that mimic the processes of the natural evolution: selection, crossover, and point mutation. The algorithm starts with a population of randomly generated design vectors, tests the fitness of those vectors, selects the best ones, and recombines the parameter values (i.e., exchanges some elements) of the best designs. The designs can then be tested again, and the process is iterated until some stopping criterion is reached. The process of exchanging elements among successful

designs is termed "crossover" because of its biological analog. Although there are many ways to exchange design elements, one of the simplest is for successful (i.e., high-fitness) vectors to be paired together, like biological chromosomes, exchanging the half of their parameters that lies above a randomly selected point along the vector. The third process is random reassignment of the values of a small number of design elements to other values, similar to biological mutations of DNA components. The functional value of this process in the algorithm is to ensure that the population of designs as a whole remains heterogeneous, because this heterogeneity determines how much of a search space the algorithm covers. Crossover relies on it as well. Random mutations alone are unlikely to generate high-quality designs, but random mutations in combination with crossover can be very effective.

Figure 2.5 from [262] illustrates the workings of a genetic algorithm as applied to the design of fMRI experiments.

In the authors' approach, the design vectors are design matrices constructed as explained above, and psychological factors (stimulus frequency and randomization) are taken into account in the form of hard and soft constraints. The resulting fitness measure is calculated as a weighted sum of efficiency, randomization measure and frequency measure. Next it is transformed with a sigmoid to produce a soft cutoff point. When the measure is above the mean, the design is selected for crossover.

Using the algorithms, the authors arrive at an optimal set of parameters for the event-related design to ensure quality data: interstimulus intervals should be 3 seconds or jittered around 6 seconds. The algorithm confirmed that for contrast efficiency block design is optimal, while for HRF estimation efficiency event-related design with 4 second intervals between same type events (or for estimation efficiency 6 secs) is preferred. It appears that it is difficult to simultaneously optimize response detection and response shape estimation, while counterbalancing the ordering of events in a stimulus sequence. However, GA optimization seems to produce designs that are much more effective in these aspects than random ordering of the event sequence.

However, this algorithm works in an offline mode. To optimize the design in an online mode for better activation detection a procedure was proposed based on a sliding window computation of correlation between the fMRI signal and the reference vector [89]. Based on the optimization of cost functions, the authors have developed a real-time algorithm that combines the correlation analysis in a computationally efficient manner with detrending, the sliding window technique, and with an optimization of the reference vector. The sliding-window technique provides constant sensitivity to changes in functional activation

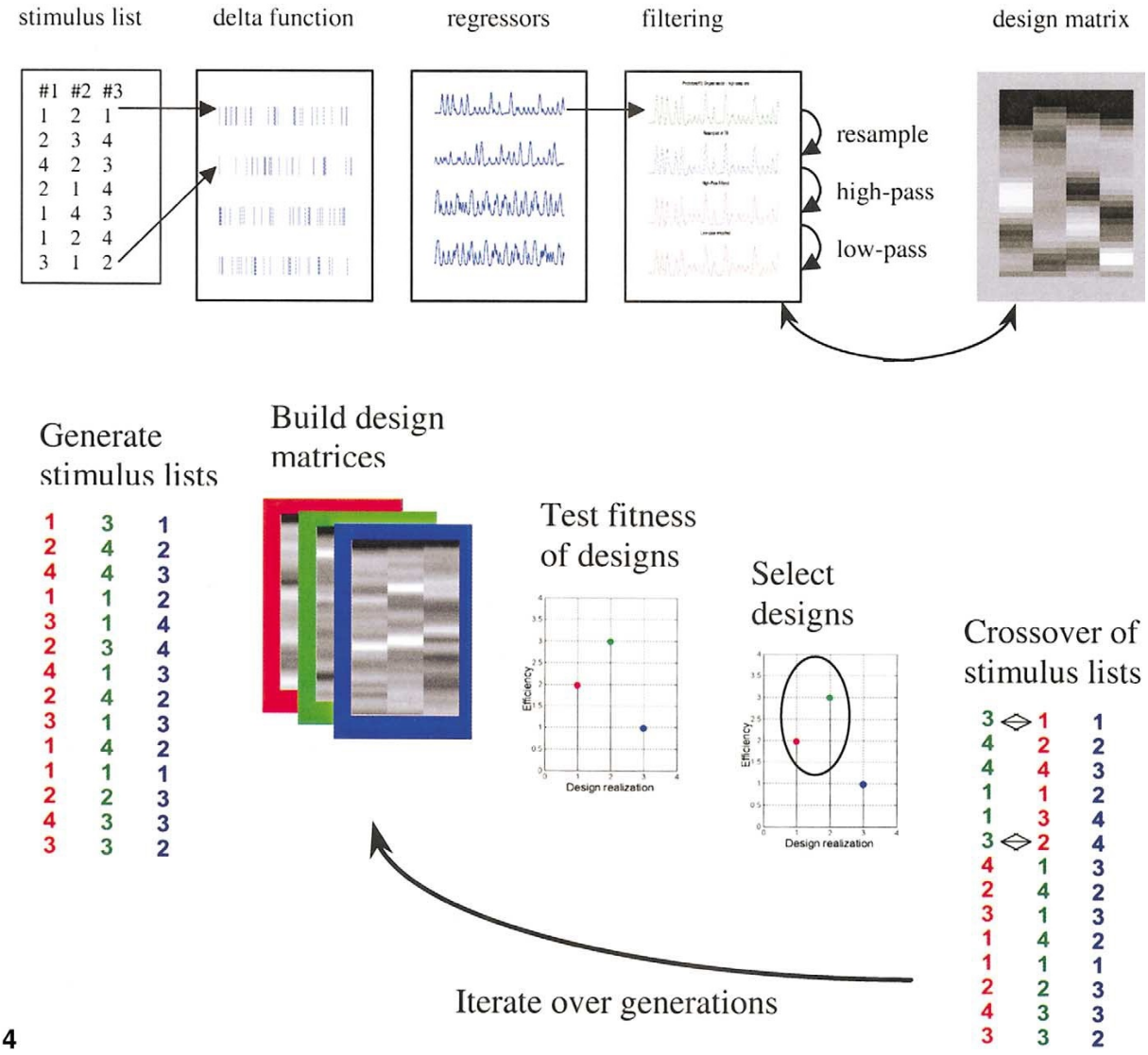


Figure 2.5: Scheme of GA in operation. Design vectors are transformed into model matrices, that are tested for fitness. Matrices were constructed by convolving a canonical HRF, normalized, with the stimulus delta function for each series. Selected designs undergo crossover, and the resulting children are retested iteratively until the algorithm is stopped.

during the entire scan. The reference-vector optimization facilitates online computation of physiological parameters, such as delay and dispersion of the hemodynamic response function, but, most importantly, also allows for the possibility of quickly changing elements of design (over a time period corresponding to the length of the sliding-window) during a measurement.

With the sliding-window technique, it is possible to restrict the correlation computation to only the most recent data samples, which is very important in real time. The reference vector depends on a set of parameters, such as the time of the gradual increase or decrease of blood flow. This signal is modeled by convolution of the design with the hemodynamic response function, which should take into account time delay and smoothing of the blood flow regulation. To increase the correlation with the measured data, either the hemodynamic response function or the reference vector can be optimized with respect to parameters that describe these effects. Since the reference vector can always be generated by convolution of the paradigm with a response function, the authors choose the former optimization option. The cost function to be optimized is as follows [89]:

$$E = |\overrightarrow{x_s} - \alpha \overrightarrow{r_s}(\omega_j)|^2$$

,

where ω is the set of parameters of the reference vector, $\overrightarrow{x_s}$ is the signal vector, and α is a linear parameter based on the correlation between signal vector and reference vector $\overrightarrow{r_s}$.

2.8 Optimization algorithms for ROI selection

ROI selection is an important part of neuroimaging experiments. It is about identifying an area in the brain where the signal is going to be analyzed. For many experiments the area is selected based on brain atlases that outline the functional organization of the brain. In the context of MVPA, we would try to further refine our selection and pin down most informative voxels within the area as features for classification analysis. MVPA is a high-dimensional pattern classification problem where classification (or prediction) model based on the fMRI BOLD signals is trained, and brain voxels are identified as features in response to different conditions as class labels. As has been mentioned, neuroscientific data is prone to the curse of dimensionality: the number of samples is significantly fewer than the number of voxels in the brain. Therefore, various advanced feature selection and sparse optimization techniques were proposed to enhance the classification efficacy

by selecting most informative voxels - e.g. [168]. It leads to a two-fold objective: (1) the number of voxels included in classification models has to be minimized while (2) the classification accuracy needs to be maximized.

In real-time experiments, the target area is typically selected beforehand based on relevant studies. However, in the context of all real-time experiments, it might be beneficial to perform ROI selection in real-time as well. An example of such study is [166]. To select a ROI for signal extraction in real-time, an extra step, such as localizer experiments and/or expert suggestions are required. The authors developed two unsupervised computational techniques based on the general linear model (GLM) and independent component analysis (ICA) of rt-fMRI data to automate this step. Decoding performances from the automatically defined ROIs were comparable to supervised methods.

In [161] optimization algorithms were used to create a protocol online, in which ROIs, that were most responsive to certain stimuli, were identified. In the first experiment, the goal was to traverse the two-dimensional parameter space for the algorithm to automatically learn the combination of audio-visual stimuli that best evokes a certain target brain state. Figure 2.6 (figures 2.6 - 2.9 all come from [161]) presents feature space in terms of image complexity, while in Fig. 2.7 features of the auditory stimuli are shown. Fig. contains 2.8 brain ROIs, for which most activating stimuli were being selected by the optimization algorith. To define the objective function, the authors took the difference of BOLD between two target regions in response to a complex audio-visual stimulus, sampled from the design space, and used a gradient approximation method - Simultaneous Perturbation Stochastic Approximations (SPSA) algorithm. The cycle of protocol adaptation is shown in Fig. 2.9. After two successive audio-visual stimuli, the BOLD was estimated for each and compared to an a-priori defined target brain state. The algorithm was regarded as converged when it was sampling the same combination of stimuli for three consecutive iterations.

In the second experiment, Bayesian optimization algorithm was applied with the same stimuli and ROIs. The goal was to optimize an unknown objective function (no analytical expression). The experiment proceeded according to an iterative scheme: subjects were presented with a stimulus, then BOLD was measured. BOLD activation was provided as feedback in real-time to learn a probabilistic model of the subjects target region activation as a function of both the auditory and visual stimuli. The model is employed to propose a new combination of auditory and visual stimuli, which is likely to increase activation in one of the target regions. The algorithm learns the distribution of the latent objective function, which is a relationship between the BOLD signal in the two regions for the grid space of

Visual complexity	Number of frames	Image size (pixel dimensions)	Image saturation (percentage)	2D Gaussian filter (*sd*)	Added Gaussian noise (variance)
1	-	-	same as baseline	same as baseline	same as baseline
2	1	65 x 116	1.7	8.9	8.9
3	3	125 x 221	2.8	7.8	7.8
4	6	184 x 327	4.6	6.7	6.7
5	11	243 x 432	7.7	5.6	5.6
6	21	303 x 538	12.8	4.5	4.5
7	39	362 x 644	21.2	3.4	3.4
8	72	422 x 749	35.3	2.3	2.3
9	133	481 x 855	58.7	1.2	1.2
10	250	540 x 960	original video	original video	original video

Figure 2.6: Design space of the visual stimuli used in the experiment.

design parameters (stimulus features). Prior distribution for the latent objective function was modelled as a Gaussian Process. The function to determine which combination of stimuli to play next was the expected improvement acquisition function. This function tries to maximize the expected improvement (the difference in BOLD between regions) over the current best. Together with GP, they are claimed to provide a closed form solution. In a later study Bayesian optimization was applied to dissociate not just single ROIs, but network regions [162].

2.9 Reinforcement learning

A fruitful framework for neurofeedback experiments would be that of reinforcement learning [253], because, unlike in the previous studies, here the protocol is adjusted to interact with the subject's state. In fact, in [151] the authors performed a neurofeedback study in the context of reinforcement learning. The participants learned to upregulate the signal from the anterior insula, as it is involved in multiple clinical conditions related to emotional and affective dysfunctions. They were instructed to only choose from a set of given strategies to upregulate the signal, and their process of decision making was modelled as an n-arm bandit problem, typical for reinforcement learning. The feedback they were given was temporally discounted reward, also standard in reinforcement learning problems, that correlated with the BOLD signal. The authors worked out a reward-related reinforcement learning model of neurofeedback in rtfMRI. The starting point hypothesis

Auditory complexity	Number of bands for noise-vocoded speech	Added Gaussian noise (*sd*)
1	same as baseline	same as baseline
2	1	.08
3	2	.05
4	3	.03
5	4	.01
6	5	.008
7	6	.005
8	10	.003
9	20	.001
10	original sentence	original sentence

Figure 2.7: Design space of the auditory stimuli used in the experiment.

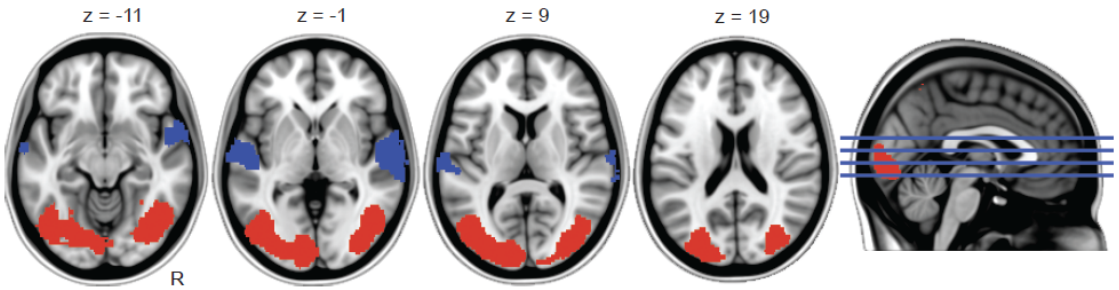


Figure 2.8: Targeted ROIs: visual and auditory cortex.

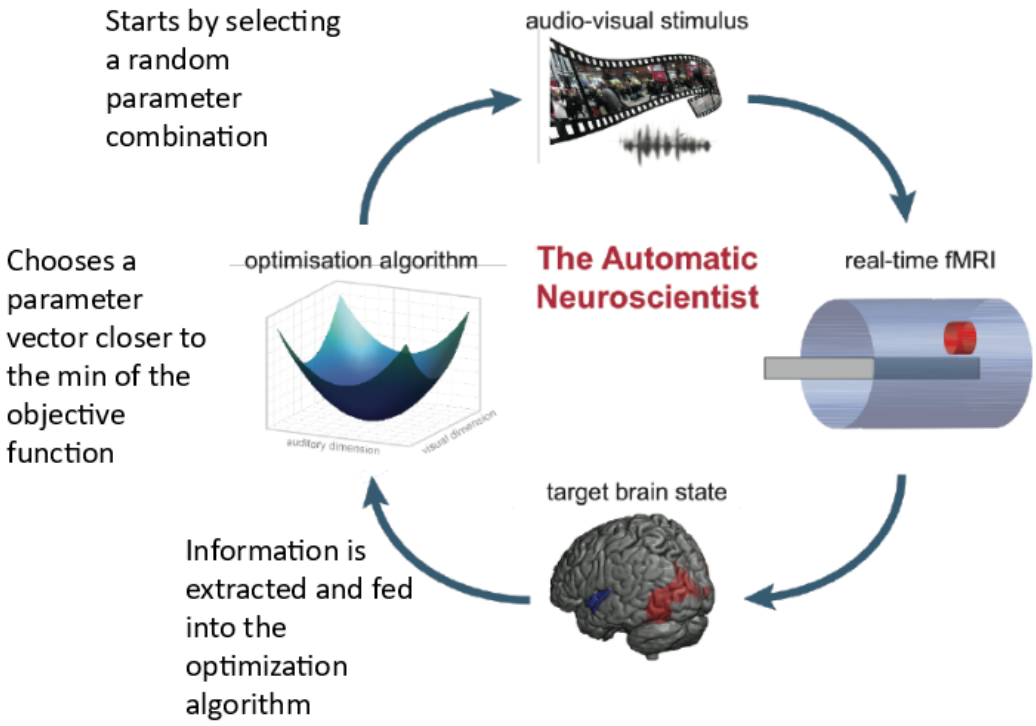


Figure 2.9: The cycle of protocol adaptation in the experiment.

was that in the target group, unlike in the controls, a positive linear increase will be observed in the activation of the right anterior insula, while no such linear interaction was expected in the controls that were provided false feedback. The study confirmed the expected results.

Another series of experiment used the reinforcement learning framework to adapt stimulation protocol for a different setting [186], [23], [20]. In these studies, the task difficulty for the participant was adjusted with the purpose of creating a restorative BCI for post-stroke rehabilitation. The subjects performed kinesthetic motor imagery of opening their left hand, while threshold-based classification of beta-band desynchronization resulted in proprioceptive robotic feedback. Electroencephalography was used to examine the cortical activation patterns during the training sessions. The participants rated their perceived mental effort nine times per block. In the adaptive block, the threshold was adjusted on the basis of these ratings, whereas adjustments were carried out at random in the other block. The perceived mental effort was correlated with the difficulty threshold of the neurofeedback training. The classifier used was a linear discriminant analysis classifier (LDA), and its threshold was adapted with a threshold adaptation strategy informed by a Bayesian reinforcement learning model. Adaptive threshold-setting reduced mental effort and increased also the classification accuracy.

So, we can see that in the context of BCI and Neurofeedback protocol adaptation should be based on interaction with the subject's state as inferred from the measured signal. However, it can be noted that although a plethora of methods have been proposed for protocol adaptation, the majority of them do not aim at interacting with the subject in a closed loop manner, so that the utility function does not maximize the chances of the subject getting into a certain state. We can conclude that technical solutions available to implement protocol adaptation for this kind of real-time neuroimaging are insufficient.

Chapter 3

Problem Statement and Proposed Contributions

In the previous chapter, I reviewed the modern paradigm of real-time neuroimaging and its three main experimentation types: BSDS, BCI and NF. I have also shown that this new paradigm needs new experimental protocols that allow adaptation in real-time in an interactive manner based on the subject's state. This is because the principal research question that this paradigm is answering is about how the brain state affects behaviour, and in this paradigm the brain state is the independent variable, while stimulation is a dependent variable. This is a reversed relationship with respect to traditional neuroimaging with offline data analisys, where the stimulation is an independent variable, and the brain measurement is the dependent variable. In this chapter, I will outline and discuss the challenges I am addressing in my research. The main challenge for BSDS stimulation protocols is finding reliable neural markers for distinguishing the different states of the brain and tracking their changes in real time. I will present this challenge in more detail in 4.1 using as a case study the neuroscientific problem of top-down attention. For BCI and NF, it is the development of real-time protocols that could be adapted in the course of interaction with the subject's brain state. In section 3.2.3 I will discuss in more detail how we can approach this challenge at the theoretical level and what are the prerequisits for practical implementations. As I am discussing in 3.2.3.4, for the successful development of interactive protocols for BCI and NF we have to rely on BSDS studies that address the mapping between neural signal and cognitive states and processes. By addressing these challenges, we give neuroscientists methodological tools to both investigate into the neuroscientific questions and to further advance the development of real-time neuroimaging protocols.

3.1 Study of top-down attention patterns with BSDS

3.1.1 Neuroscientific background of the study

Top-down attention patterns mediate attention biases during perception and affect behavioural performance in attention related tasks. In case of visual attention, these patterns can be revealed in visual search experiments via activity during preparatory delays. Visual search experiments are most common experiment used to study visual attention: the subject is cued to search for some object, then, after a certain delay period an image is shown, and the subject should respond with a button press of she detected the presence of the cued object. High-level top-down attention patterns have been actively studied in the category-related object selective areas [248] that process categorical information about the perceived natural objects. Several studies attempted at demonstrating the high-level nature of the preparatory patterns [200, 201, 248]: it is considered to reveal itself in the fact that these patterns are established in response to certain stimulus categories (object categories) independent of the sensory or cognitive process that operates with these categories (i.e., perception, imagery, visual search). Brain State Dependent Stimulation protocol is a perfect experimental tool to investigate into these patterns and to show that they do affect behavioural performance. That is, if we can send an attention engaging task, such as visual search stimulus, in the moment when the top-down attentional state is detected, we would expect that this would result in behavioural improvements, such as higher accuracy of detection or faster reaction times.

The state-of-the-art approach to demonstrate that a pattern is of high-level, modality-independent nature, is to cast the problem of identifying common activation patterns as a supervised learning or brain decoding problem [107, 126]. The data of one cognitive modality (e.g. perception) is used to train the classifier to distinguish category related patterns (e.g. people vs. cars) in the data coming from a different cognitive modality (e.g. imagery). In [187] it is pointed out that successful classification in this setting implies that neural patterns elicited by relevant cognitive factors in one modality generalize accross to the patterns in the other modality. A low misclassification error on a modality which is different than the one used to train the classifier provides empirical evidence that a given region of interest is related to the cognitive task encoded in both modalities. Such approach is referred to in the literature as cross-modal decoding analysis (CMDA). Statistical significance of this approach can be computed using a t-test on the accuracy obtained by the classifier [135, 170], or by a permutation test to compute the null distribution [70, 124, 259] of such statistic.

The patterns of brain activity that are shared between the processes of perception and imagery have been the subject of quite numerous studies. The question investigated was if we can arrive at abstract, top-down object representations [116, 213] containing distinguishing features [112, 218] that will have common neural substrate both for viewed and imagined object categories [218]. To test for the presence of shared patterns, many studies used cross-modal decoding - namely, multivariate pattern analysis with SVM classifiers [45, 153, 213]. Significant cross-modal classification accuracies were taken as the evidence in favour of shared activity patterns. What emerged from these studies was the view that, indeed, visual imagery activates the same areas that contain information about visually perceived stimuli [74, 218], and shared patterns for stimulus categories in these two processes can be established [5, 45, 153, 199, 213, 247]. The areas where these common representations were found include the ventral temporal pathway, lateral occipital cortex [45, 153, 213, 247] and extrastriate cortex [116, 153, 199, 247]. These conclusions influenced the selection of ROI in our study. The question about if these patterns are to be seen in attention tasks as top-down activity has been mainly investigated using cross-modal analysis with perception and visual search [200, 201, 248]. However, as object representations in the brain obtained during imagery tasks are thought to be closer to high-level top-down representations of objects in visual cortex [199, 218], the hypothesis naturally suggests itself that we should use imagery data to detect these high level patterns during visual search preparatory periods.

3.1.2 Offline experiment to test state identification

In order to be able to apply BSDS approach to the study of top-down attention, we need to a) define the marker of the relevant brain state; b) select tools that would allow us to detect this state marker in a real-time experimental setting and c) choose a runtime policy triggered by the detected markers. We hypothesized, based on the literature analysis, that the preparatory attentional state in visual search reveals itself as a high-level object category pattern in the object selective cortex that is independent of a cognitive modality. According to the literature, this should be closer to an object category pattern that shows up in imagery, i.e. when we imagine objects. We ran an experiment engaging all three cognitive processes (perception, imagery, visual search) that used the same object categories (whole-body images of people vs. cars). We performed empirical analysis of the data coming from this experiment with the goal of investigating the following questions: a) if cross-modal decoding analysis can reveal top-down modality independent object category pattern between such cognitive processes, as perception, imagery and visual

search; b) to test whether these patterns can be used as reliable markers of top-down state for the purposes of a brain-state dependent stimulation experiment; c) to verify whether we can perform real-time decoding of top-down states in a BSDS experiment using cross-modal decoding approach and whether this approach can be the basis of the runtime policy. fMRI data were collected to investigate object categorization during preparatory activity in a visual search experiment, designed in a similar manner to the one illustrated in [200]. The expriment is presented below.

3.1.2.0.1 Participants. 24 participants (8 male, mean age 27.1, st.dev. 4.3 years) were recruited and accessed the research facility. All participants, before starting the experiment, signed a form confirming their informed consent to participate in the experimental study. Each participant was instructed in advance about undergoing 2 experimental sessions (S1, S2) on two different days. The data on all three modalities in question (perception, imagery and visual search) were acquired in the same session, S1. Out of all 24, 22 completed a significant part (6/22) or the whole (16/22) of S1. During both S1 and S2, participants were also given other tasks, which we do not report here. Out of 16 participants that underwent the whole S1 only nine participants completed 4 runs of both perception/imagery task and visual search task (8 functional runs in total). Other participants failed to reach 4 runs at least in one of the task types. So, for the analysis we are using the data of the 9 participants that have the total of 8 functional runs each. The tasks are explained in the next section.

3.1.2.0.2 Stimuli. Two distinctive stimulus categories were presented to the participants throughout the tasks: people (whole body image) and cars. Participants were instructed to deal with these categories in three different ways: in *perception* modality, they had to attend to 8 presentations of 16 seconds long blocks of different instances of the same category (people, here depicted by whole body figures with no face, or cars), interspersed with 16 seconds long fixation periods. Participants were equipped with a two-button box. They were requested to perform a one-back task - i.e., to press a specific button whenever they detected the same image repeated twice in a row. In *imagery* modality participants were instructed to close their eyes and mentally visualize instances of the category, indicated by the letter cue shown at the beginning of the trial. They had to press the response button whenever they achieved a mental image that was sufficiently detailed, and then they had to switch to mental visualization of another instance of the same category. At the end of the 16 seconds block, an auditory cue told the participants to open their eyes and go on with the experiment. Perception and Imagery blocks were randomly

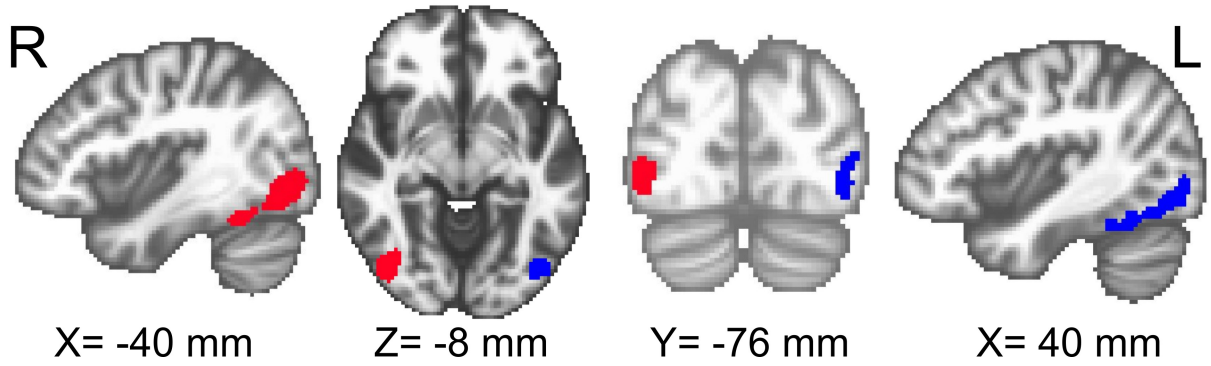


Figure 3.1: Bilateral Object Selective Cortex (OSC) group map in temporal-occipital cortex delineated based on the functional localizer (intact vs. scrambled objects). Its right part is shown in red, left in blue. X, Y and Z locate the coordinates of the slices.

presented within the same functional run. In *visual search* modality, participants were briefly (450 ms) shown images representing natural scenes (e.g.: crowded places, urban landscapes, etc ...). They were instructed through a visual cue (letter) to look for instances of one of the two categories within the scene. After scene presentation, they had 1.6 s to attend to the presentation of a mask and to give a positive or negative response by pressing a button. Visual search preparatory periods occurring between the presentation of the cue and the presentation of the scene had different lengths: 2, 4, 6, 8 or 10 seconds. Participants did not know in advance neither the lengths of preparatory period, nor their order of presentation throughout the task, which was random. Participants also had to perform a block-designed task, where we alternated presentation of images of intact and scrambled everyday objects, in order to functionally define an object selective region of interest (ROI) localized in the temporal - occipital cortex (Fig. 3.1). So, for perception and imagery the overall number of trials was 32 per modality (8 trials per 4 runs). Each visual search run consisted of 40 trials, where for each particular type of delay duration there were 8 trials. The total number of visual search trials is 160 (40 per 4 runs), while for each duration there are 32 trials (8 per 4 runs) in the dataset. All experimental procedures were approved by the Ethical Committee of the University of Trento.

3.1.2.0.3 Data acquisition. Images were acquired with a 4T Bruker (https://www.bruker.com/) scanner. For each participant, we started both experimental sessions by acquiring a structural scan using a 3D T1-weighted Magnetization Prepared RApid Gradient Echo (MPRAGE) sequence (TR/TE = 2700/4.18 ms, flip angle = 7°, voxel size = 1 mm isotropic, matrix = 256 × 224, 176 sagittal slices). Perception / Imagery and Visual

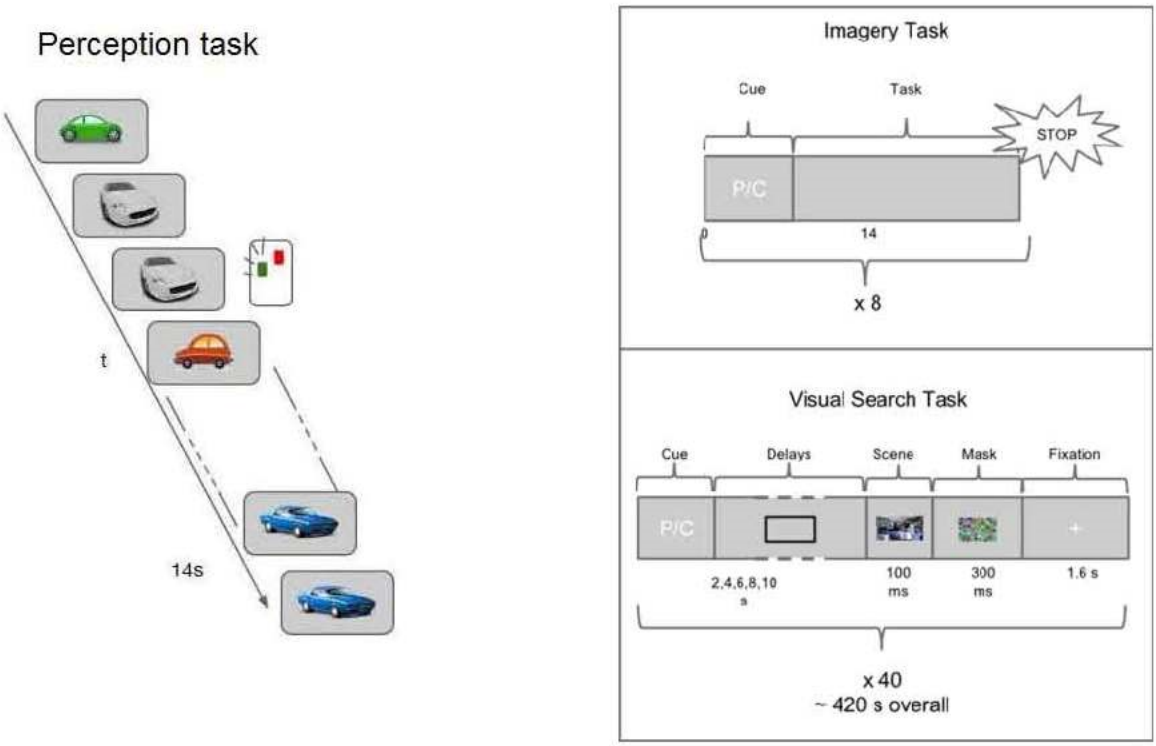


Figure 3.2: In the perception task (left), the participants were viewing blocked images of people's bodies and cars; in the imagery task (left top), they were visualizing the same categories when cued until they heard the sound signal. In visual search task, the participants detected the presence of the cued category (always people or cars) in images of cluttered scenes.

Search tasks were performed while acquiring, respectively, 177 and 195 functional scans with the following parameters: (TR/TE = 2000 / 33 ms, flip angle = 73°, voxel size = 3x3x3 mm, 1 mm slice spacing, matrix = 64x64, 34 axial slices covering the entire brain), during session 1. Same acquisition parameters were used for 165 scans acquired during Functional Localizer task in session 2.

3.1.2.0.4 Preprocessing. For data preprocessing FSL tools were used along with in-house built Python code. In all functional runs 5 initial volumes were discarded as dummy volumes. The skull was removed from both functional and structural images to extract the brain. Functional images were subsequently corrected for slice timing and motion artifacts. Transformation of the functional images to standard space was carried out in the following sequence. First structural scans were coregistered to the mean functional scan of each experimental run. Structural-in-functional-space images were then coregistered to standard (MNI) space to finally compute affine parameters. To extract task-related effects from functional localizer data, beta maps for both localizer conditions (Intact vs. Scrambled objects) were computed with linear regression and next fed into contrast analysis (Intact vs. Scrambled). This analysis resulted in a ROI located in bilateral temporal occipital cortex. We selected one cluster including 625 voxels from each hemisphere, ending up with a bilateral ROI with 1250 voxels overall. We applied the ROI mask to functional data coming from Perception / Imagery and Visual Search tasks and we obtained matrices containing time-series of the ROI voxels. For CMPT analysis, 1250x16 matrices of Perception / Imagery data and 1250x40 matrices of Visual Search data were considered per run.

3.1.2.0.5 Results The ROI chosen for the analysis was the Object Selective Cortex (OSC) map shown in Fig. 3.1. We analysed separately the performance of a cross-modal classifier for the left part of the ROI (denoted as L in Table 3.1, right part of the ROI (R in the same table) and the whole ROI (L+R). Encoding of the activation maps was of two types: raw BOLD and beta maps. For the raw BOLD encoding, the volumes were selected that corresponded to the peak of the hemodynamic response function (HRF, as rendered by SPM software - `www.fil.ion.ucl.ac.uk/spm/doc/`) convolved with the boxcar function that represented the experimental manipulation. One volume was selected per trial. For beta encoding, beta maps were calculated trialwise using linear regression.

Cross-modal decoding analysis (CMDA) was performed by training a logistic regression classifier with ℓ_2 regularization on the trials of one of the modalities. The regularization parameter was selected according to a nested cross-validation scheme (leave-one-run out).

Modalities	ROI	Encoding	CMDA
P vs I	L+R	Bold	0.12
P vs I	L+R	Beta	0.08
P vs I	L	Beta	**0.001**
P vs I	R	Beta	**0.005**
V vs I (delay 2 s)	L+R	Beta	0.272
V vs I (delay 4 s)	L+R	Beta	0.387
V vs I (delay 6 s)	L+R	Beta	0.165
V vs I (delay 8 s)	L+R	Beta	0.659
V vs I (delay 10 s)	L+R	Beta	0.249
P vs V (delay 2 s)	L+R	Beta	0.44
P vs V (delay 4 s)	L+R	Beta	0.44
P vs V (delay 6 s)	L+R	Beta	0.49
P vs V (delay 8 s)	L+R	Beta	0.08
P vs V (delay 10 s)	L+R	Beta	0.89
V vs I (all delays)	L+R	Beta	**0.012**

Table 3.1: Results of cross-modal decoding analysis (CMDA). The values refer to p-value for the discrimination task of *body* vs. *car* in each pair of the cognitive modalities listed in the extreme left column, where P stands for *perception*, I for *imagery* (I) and V for *visual search*.

The accuracy was then estimated on a test set from another modality. The training and test process was replicated for each subject. Then, the p-value was computed for the group using the permutation scheme described in [70]. The resulting p-values are reported in Table 3.1.

In Table 3.1 we report only one result related to raw BOLD volume encoding: the cross-modal analysis between Perception and Imagery. In this case decoding does not detect a meaningful shared activation pattern. Beta maps encoding on the other hand seems to be a more efficient representation. Chen and colleagues [42] demonstrate that using beta values is a way to get rid of intrinsic variabilities of BOLD signal throughout the brain and, specifically, within a single area. In our case, this means that beta values are more related to the effect size than raw BOLD signal changes during task-on periods. For this reason in the presentation of results we only focus on results obtained with data encoded with beta maps.

Table 3.1 illustrates the methodological limitations of decoding when applied to the problem of cross-modal pattern detection. Cross-modal analysis results for perception vs. imagery with beta maps encoding are significant when the two hemispheres are considered individually, namely the left and right OSC respectively. When the analysis is extended

to the joint ROI, the number of trials remain constant, while the number of voxels double. In this case the classifier is affected by the higher dimensionality of data, and CMDA does not succeed in rejecting the null hypothesis. If we compare the results for cross-modal decoding P vs V and I vs V, we can see that there are no significant results in P vs V decoding. On the other hands, p-value is significant for I vs V (all trials) which could be an argument in favour of the statement that high-level category patterns are indeed closer to those that are detected during imagery. However, we have another classifier issue here. If we consider the cumulative trials of visual search, irrespective of the delays, CMDA rejects the null hypothesis. When we restrict the cross-modal analysis to single delays of preparation period for visual search, the number of trials drop from 160 to 32. In this case CMDA fails to reject the null hypothesis.

The results shown in Table 3.1 demonstrate, that CMDA is sensitive to the curse of dimensionality and the number of samples used for classifier training. This is a serious hurdle for real-time mode, where only limited data is available, and subject independent decoding, that increases the number of samples available by using the data of other subjects, is still a matter of research and has its challenges, too [258]. So, supervised cross-modal decoding does not provide us with the means of top-down pattern identification that we could reliably use in the real-time mode. However, the question of cross-modal patterns in neurosicence is currently a very active area of research: the study of common representations between different cognitive processes and/or sensory modalities gains momentum now because it also sheds light on how sensory input interacts with categorical representations in human brain. This, in turn, is part of a broader, more fundamental neuroscientific question, namely interaction between bottom-up and top-down information processing in the brain in general. As the current reference method of investigation into these representations, brain decoding, has serious limitations, so as a secondary contribution we proposed a new method that has higher sensitivity. It allows to identify common activity patterns between various cognitive processes more robustly and thus achieves more reproducible results. The new method is meant to advance the quest for abstract, high-level neural representations independent of sensory modalities or cognitive processes. We illustrate the method with both synthetic data and experimental fMRI data. The method is presented in Chapter 6.

3.1.3 Unsupervised methods of state identification

So far the overwhelming majority of brain decoding studies have used supervised learning algorithms where the researcher provides stimuli category labels with the training data

so that the Machine Learning (ML) classifier can predict labels for the test data. At the same time, there is growing interest in the use of unsupervised learning methods for brain decoding studies - Principal Component / Independent Component Analysis (PCA/ICA, e.g. [6]), or different clustering methods (e.g. self-organizing maps [81]), representational dissimilarity matrices (e.g. [139]) etc, and also hidden Markov models (HMM, e.g. [73]). Most of the cited papers deal with the traditional offline decoding, but the efficiency of these methods for real time decoding has been demonstrated, for instance, in [6] and [242]. So, we also tried to approach brain state detection problem for a BSDS study of top-down attention employing an unsupervised algorithm: real-time application of Gaussian Mixture Models (GMM).

3.2 Development of closed-loop stimulation protocols

Human brain is a complex physical network system that supports mental function. Sometimes, in order to treat cognitive deficits or enhance mental abilities, an intervention is desirable that is targeted at modulating human brain activity to achieve beneficial behavioural changes. While empirically it was shown that such regulation is possible (e.g., cognitive behavioral therapy, deep brain stimulation, neurofeedback are all tools for such regulation), theoretical frameworks to guide these interventions and to optimize them for clinical use are fundamentally lacking. This, in turn, does not allow to offer practical solutions for the much needed case. Therefore, we consider it advantageous and well-timed to address the problem of development of neuroscientific interactive protocols that could make these interventions possible. As these interventions represent a form of control (one entity exerts physical influence over another), a promising approach lies in an engineering discipline known as control theory that investigates the question of how to guide complex systems from one state to another [123].

A very prominent part of neurofeedback and closed-loop studies are the studies on affective state regulation. Besides, affective states, such as, e.g., emotional states, are probably intuitively most readily understood as states, unlike, for instance, cognitive states. For these reasons, we are choosing to target affective states in the part of our research dedicated to closed-loop adaptive protocols. We want to make a meaningful contribution to the field by a) designing a real-time adaptive protocol in the framework of the control theory and b) implementing a proof-of-concept experiment. To do this, we are going to explore what has to be in place for such a protocol to be feasible. As I am going to show in the rest of the section, we have to be able to 1) meaningfully

define the target state as a mapping between the signal (brain level) and the subject's cognitive state (mind) and to 2) learn an optimal control sequence, namely a sequence of external inputs that will allow the subject to reach the desired state by interaction with her internal state dynamics. The first task requires deep neuroscientific insights that are not available to us at the moment, as will be discussed later in this chapter. So, we are going to investigate the second possibility, namely how to find the input sequence, at the theoretical level, assuming that the states and there dynamics are known. We also assume that the relationship between the two is modular, that is, we can "plug in" the necessary state definition(s) in the model of control once they are available from neuroscientific investigations. Still, we need to have condition 1 met if we want to design a proof-of-concept experiment. For the purpose of designing the experiment, we adopted an approach of "reduced complexity" in defining the subject's state(s). We chose the Galvanic Skin Respose (GSR) (also known as EDA - Electrodermal Activity) for two reasons: first, it is a unidimensional signal, unlike that of fMRI, and second, GSR signal is known to have a relatively straightforward linear correlation with the subject's affective state - arousal. Because of this correlation, we can avoid modelling separately the relationship between the signal and the state, as the level of the signal, high or low, corresponds nicely to the high or low arousal. In a somewhat simplified manner we can even say that by controlling the signal we are controlling the underlying state. The goal of the proposed experiment is to test empirically the feasibiulity of the optimal control approach for controlling the subjects' internal states.

3.2.1 Foundations of control theory

In the formal terms of control theory, control means applying some input to move a system from an initial state to a target state. Typically, control is exerted by a controller C via control input, u, to change the state of a system (usually called plant P in control theory). Fig. 3.3 from [177] illustrates two known control scenarios - so called open-loop (a) and closed-loop (b) control.

The control strategy is based on a cost function, i.e. it aims at minimizing the distance between the observed state of the system and the target state, often while also maximizing technical simplicity and minimizing input [128].

In the context of human cognition, the plant (system) is the physical brain or some part of it (e.g., a single neuron, an ensemble of neurons, a group of brain regions, or a single region of interest). As a physical system, the brain is characterized by specific states - patterns of neural activity. The control or guidance of the brain from one state

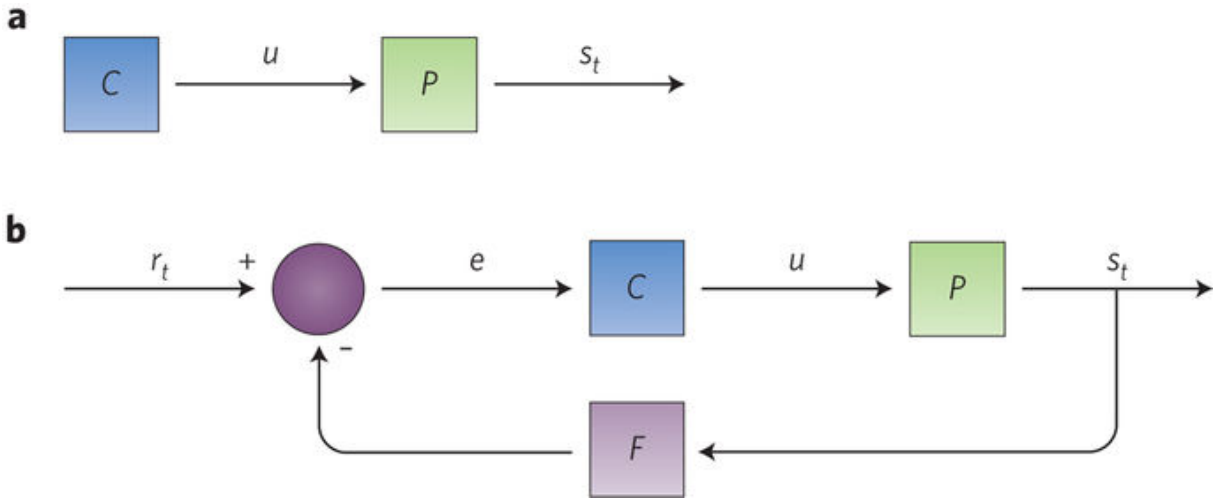


Figure 3.3: (a) A classic open-loop control scheme. The controller (e.g., a system in the environment or designed device) delivers an input u to the system under control (plant P) to influence the system state s_t. (b) A classic closed-loop feedback control scheme. The goal is to guide the system (P) to a reference value r_t. The system state s_t is fed back through a sensor measurement F to compare to the reference value r_t. The controller C then takes the error e (difference) between the reference and the output to change the control inputs u to the system P under control.

to another can either be intrinsic (the brain controls itself [100]) or extrinsic (the brain is guided externally, for example by brain stimulation [183]). The controller input is stimulation which can also take the physical form, such as applied current for deep brain stimulation - e.g. in Parkinson's disease. In this case, the neural reference state is often represented in the frequency domain, and the neural control goal is to achieve a target state of basal ganglia activity. But the control input can also be the sequence of non-invasive external stimuli that set the brain in action. By employing the framework of the control theory we address the challenge of designing strategies that interact naturally with a system's structure and dynamics. Here, the goal is to induce changes in neural states that support cognitive functions, i.e. to induce movement from one neural state to another, so that we can evoke the desired tangible behavioural changes. Control theory provides us with principled approaches to identifying strategies that influence states in the context of indirect interactions within the system and produce impact on cognitive function both in health [179] and disease [35].

3.2.2 Optimal control

The problem of optimal control consists in optimizing a sequence of actions to attain some future goal [80, 245]. It views an agent as an automaton that seeks to maximize expected reward (or minimize cost) over some future time period. If we consider an example of a hunter who throws a spear to kill an animal, the optimal control solution is a sequence of motor commands that results in killing the animal by throwing the spear with minimal physical effort [128]. The components of the optimal approach are: a) a dynamical system described by some mathematical model, e.g., differential equations; b) external control input that is able to influence the system's dynamics; c) some constraints on the system state or control input (e.g., maximal velocity or brake force of a car) and d) some objective function, or cost function to be minimized. Fig. 3.4 from M. Gerdts' lecture slides based on [90] illustrates how optimal control works. The state of the system at each time point is described by the state vector $x(t)$. The control vector $u(t)$ allows to influence the dynamic behavior of the state at time t. The differential equation

$$x'(t) = f(t, x(t), u(t))$$

defines the rate of change of the state vector as a function of the current time t, the current state vector $x(t)$, and the current control vector $u(t)$.

In an open-loop control problem, we are not receiving feedback from the plant. The solution of an optimal control problem is a time-dependent control $u(t)$ (open-loop control), which minimizes a given objective function (cost function) subject to constraints. The controlled dynamic system reads like the following [90]:

$$x'(t) = f(x(t); u(t))$$
$$x(0) = x_0$$

That is, the next state of the system is the function of its current state and the control input. Open-loop control does not adapt to perturbations, while closed-loop control receives feedback from the plant and is able to act on perturbations. Closed-loop control finds a feedback control law of type $u = U(t; x)$ for the control system, such that a given reference trajectory is tracked. The controlled dynamic system (closed system) is given by the following equation from [90]:

$$x'(t) = f(x(t); U(t; x(t)))$$
$$x(0) = x_0$$

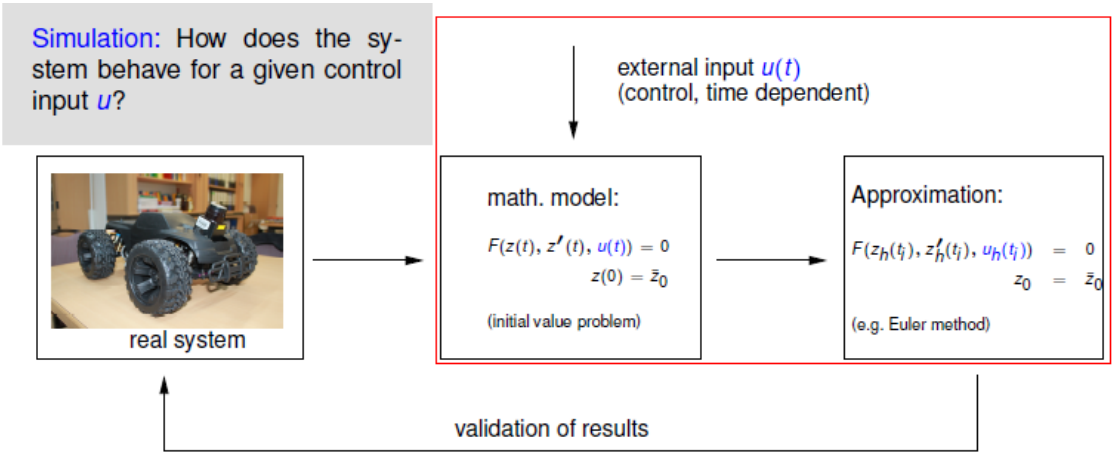


Figure 3.4: Example of how control task work. We have a physical system - a car. It can be described by its states, such as positions on a trajectory. Control is applied for the car to change its trajectory, and the model allows to estimate the effect of the input on the system state. Next, we can track the real trajectory of the physical car and act on the error.

In this case, the next state of the system is, again, the function of its current state and the control input, but now the control input is also the function of time and the current state of the system. Turning to our example of a hunter with a spear, we can say that if x denotes the state space (the positions and velocities of the muscles), the optimal control solution is a function $u(x,t)$ that depends both on the actual state of the system at each time and also depends explicitly on time [128]. Optimal control problem is usually solved by using a cost function. In our example, if a motor program is a sequence of actions, this sequence can be assigned a cost that consists generally of two terms: a path cost, that specifies the energy consumption to contract the muscles in order to execute the motor program, and an end cost, that specifies whether the spear will kill the animal, just hurt it, or misses it altogether.

To illustrate how the optimal control problem is solved, we take a simple example of a finite horizon discrete time deterministic control problem [128].

Consider the control of a discrete time dynamical system:

$$x_{t+1} = x_t + f(t, x_t, u_t)$$
$$t = 0, 1, \ldots, T-1$$

x_t is an n-dimensional vector describing the state of the system and u_t is an m-dimensional vector that specifies the control or action at time t (we are taking a deterministic example, and noise is not taken into account). If we specify x at $t = 0$ as x_0 and we specify a sequence of controls $u_{0:T-1}$ as $u_0, u_1, ..., u_{T-1}$, we can compute future states of the system $x_{1:T}$ recursively. Next, we define a cost function that assigns a cost to each sequence of controls:

$$C(x_0, u_{0:T-1}) = \phi(x_T) + \sum_{t=0}^{T-1} R(s, x_s, u_s)$$

$R(t,x,u)$ is the cost that is associated with taking action u at time t in state x, and ϕ(T) is the cost associated with ending up in state x_T at time T. The problem of optimal control is to find the sequence $u_{0:T-1}$ that minimizes $C(x_0, u_{0:T-1})$:

$$J(t, x_t) = \arg\min_{u_{t:T-1}} \phi(x_T) + \sum_{t=0}^{T-1} R(s, x_s, u_s)$$

The solution is usually found by dynamic programming [75].

Optimal control models are presently constructed by guessing the cost function, obtaining the corresponding optimal control law, and comparing its predictions to experimental

data. Ideally we would be able to do the opposite: record data, and automatically infer a cost function for which the observed behavior is optimal [108, 128].

Optimal control and optimal estimation are closely related mathematical problems - the problem of control, in fact, often poses itself as the problem of joint inference and control [128]. Learning becomes part of the control problem, when there is uncertainty about probability distributions over the hidden states, parameters or rewards. The optimal action policy has to comprise both actions that optimize the expected future reward given the current beliefs and actions that improve the accuracy of the beliefs [128]. In other words, the agent faces the problem of finding the right compromise between learning and control. This is known as dual adaptive control problem [78]. Examples of dual adaptive control are Simultaneous Localization and Mapping, SLAM, [33] and Simultaneous Localization and Planning - SLAM framework applied to modelling spatial cognition in the brain [203]. The approaches to this problem can be summarized as filter-based methods and optimization-based methos [33]. While the optimization methods deploy a variery of models and algorithms, the filtering methods are mainly based on Kalman filter to estimate the future state of the system: the new state estimate is deduced from the previous estimation by addition of a correction term proportional to the prediction error (or the innovation of the measured signal) - Kalman gain [125].

3.2.3 Protocol adaptation as an Optimal control problem

In sections 2.6 - 2.8 we presented several state-of-the-art methods used in real-time neuroscience for protocol adaptation. We also showed that they all could be generalized under the umbrella term of optimization approach beacuse they were meant to find an optimal design for some desired experimental charachteristic - e.g. data quality or ROI. We also mentioned that methods for finding an optimal policy for managing the subject's state are scarce. So, control theory and especially optimal control look like very promising theoretical framework to address the question of real-time protocol adaptation with the purpose of driving the subject in a particular state in the same line as other optimization-based methods.

To approach experimental protocol adaptation problem as an optimal control problem, we will have to define the state space of the subjects states, the target state and learn the optimal stimulation sequence that can get us into the target state. We should also be able to estimate the dynamics of the system and define the cost function that would allow us to learn the optimal action sequence, or policy. We will also have to address the issue of learning and inference. In the next few sections, we are going to show how we can tackle

the problem of control at the theoretical level and establish a theoretical framework to be able to approach the protocol adaptation problem as an optimal control problem. We do not claim to have a full-fledged solution here (this could be a whole separate research project in itself), the purpose of this section is to demonstrate at the theoretical level what is possible and what is missing in order to put that idea into action. We have already many key elements in place. For instance, for the estimation part, Kalman filter has already been used in the context of fMRI signal for system identification and to predict its state [264], [265]. Next, in 3.2.3.1 we are going to show that we can learn the sequence of (stimulation) inputs that produced a given sequence of (brain) states interacting with the system (neural) dynamics, and we can also demonstrate how to address the learning problem of joint inference and control . We can check for controllability of the system, namely, the brain, (see 3.2.3.2), and define cost functions that would allow to learn the control sequences (see 3.2.3.3). So, at the theoretical level, we can work out a framework for addressing the problem of interactive protocols as an optimal control problem. However, in the last section of this chapter, 3.2.3.4, I am also going to outline the major challenge that has to be overcome to be able to tackle the real-time protocol adaptation within the framework of optimal control in practice: meaningful definition of states and the target state in terms of mapping between neural signal and cognitive performance. I will also present our solution that we chose to enable setting up a proof-of-concept experiment on control of the subject's state(s).

3.2.3.1 Learning control from a sequence of states

In section 2.4.3 of chapter 2 on neurofeedback, I mentioned that neurofeedback experiments targeting affective states are very prominent and frequent in the field. So, to demonstrate how we can address the problem of learning the optimal stimulation sequence and also the problem of joint inference and control in theory, I will conjure a hypothetical example of such an experiment for merely illustrative purposes. Imagine an experiment in which we want to drive a major depression patient in a more stable positive state of mind by providing two types of stimulus inputs: neutral stimuli and positively charged stimuli. Neutral stimuli have to be there to counterbalance the adaptation effect that inevitably arrives if we stimulate with positive stimuli only. Putting several positive stimuli one after another can magnify their effect, but only up to a certain point, when the adaptation comes into play. That means that positive stimuli will have major effect depending on their place in the stimulation sequence, which, in turn, depends on the train of states in the subject's brain. So, we want to have such a sequence, that the effect will

be maximal - i.e. we want to optimize our stimulation to be able to drive the patient in the desired state.

In this experimental setting, we have to learn how the brain states transition as a function of the stimulation. In an interactive experimental design, to adapt the stimulation protocol in real-time, we have to learn both the optimal stimulation sequence and the model of the system under control. We are going to outline a tentative approach using simulated experiments with Hidden Markov Models (HMM).

Suppose we have a sequence of states S= s_1, s_2 ... s_n. The transition probabilities between states are contained in the transition matrix A_0. This is the model of the system dynamics.

At some timepoints on a timeline the experimental stimulus will be used to induce a certain state. We will refer to the sequence of stimuli as control signal. Here is an example of a control signal sequence X where 1 denotes the presence of stimulus (positively charged, in our example) and 0 means no stimulus (or neutral stimulus): $X = 0001001000...$

The sequence of states induced by the control signal can be regarded as a Markov chain with the underlying transition matrix A_1, containing the probabilities of transitions between states when the control signal is applied. Each transition probability is a function of the control signal: $A_{ij} = f(x)$.

The problem is posed in the following way: can we learn the model of state transitions (i.e., transition probabilities) given the state sequences produced by the control signal? That is, we have to learn two transition matrices, one for the system and one for control. In the next few sessions, we are going to show how to do that.

3.2.3.1.1 Approach outline So, we propose to view the state transition matrix A as the softmax function of the control sequence. We introduce new parameters - linear coefficients a and b.

$$A_{ij} = \frac{e^{a_{ij}x+b_{ij}}}{\sum_k e^{a_{ik}x+b_{ik}}}$$

However, the slope coefficient is only present in the case when the transition is induced by the control signal. When the states transition without the control being applied, the slope a is equal to 0: $a_{ij} = 0$. The function behind the transition matrix of the system states in the absense of control (i.e. describing the system dynamics) would be as follows in this case:

$$A_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

Let's denote the former matrix A_0 and the latter that describes transitions depending on controls as A_1.

The matrices A_0 and A_1 are constrained in the following way: 1) $A_{ij} \geqslant 0$, 2) $\sum_i A_i{=}1$.

Next, we propose learning the model parameters by maximizing the likelihood of the observed states chain given the data.

3.2.3.1.2 Implementation details In implementation of the parameter update we will follow [17]. The authors parametrize the transition probabilities with a softmax function in the following way:

$$t_{ij} = \frac{e^{\lambda w_{ij}}}{\sum_k e^{\lambda w_{ik}}}$$

As a starting point, we could follow the update they propose for the parameter w_{ij} (in the batch version of the algorithm), where O stands for an observation:

$$\Delta w_{ij} = \lambda \sum_O \frac{1}{L_O(M)} \{n_{ij}(O) - t_{ij} n_i(O)\}$$

Basically, at the core of this update is the difference between the number of transtions between s_i and s_j in the data $n_{ij}(O)$ that can be estimated, for instance, by the forward pass of the Viterbi algorithm, and the number of the transtitions from the state s_i predicted by the transtition model: $n_i(O)$ is the number of observed states s_i, and t_{ij} is the transition probability from this state defined in the equation above. $L_O(M)$ is the likelihood of the model given the data. The λ parameter is the learning rate. In the paper, it is claimed that it corresponds to the gradient descent on the negative log likelihood. So, to learn the parameters of transition probabilities, we will also take the difference between the number of observed transitions between states s_i and s_j and the number of transitions from states s_i and s_j predicted by the model.

Our method does not need the forward step because our states are fully observed, but we need to distinguish between states produced with or witout control signal. We will distinguish between $n(s_i)$ that denotes the occurences of state s_i in the absense of control and $n(x_i)$ that denotes the occurences of state s_i under the induction of control. Then, n_{ij} will denote the number of transitions between states s_i and s_j without control, and accordingly $n(x_{ij})$ stands for the count of transitions between states s_i and s_j in

the presence of control. In the same way, we will calculate separately the number of transtions predicted by the matrix A^0 and A^1 depending on the absense vs presence of the control signal in the transition. O_m will stand for an observation chain of length T where $1 \leqslant m \leqslant M$, M being the overall number of chains in the training data. We will use two different updates: one for the intercept that does not necessarily depend on the control signal x, and one for the slope. In the latter case, we should take into account only transitions that occured in the presence of control:

$$\Delta b_{ij} = \lambda \frac{\Delta b_{ij} + \sum_{T-1} n_{ij}(O_m) - (\sum_{T-1} nx_i(O_m) A^1_{ij} - \sum_{T-1} ns_i(O_m) A^0_{ij})}{M * T}$$

$$= \lambda \frac{\Delta b_{ij} + \sum_{T-1} n_{ij}(O_m) - \sum_{T-1} nx_i(O_m) A^1_{ij} + \sum_{T-1} ns_i(O_m) A^0_{ij}}{M * T} \tag{3.1}$$

To update the slope, we just take the difference between the count of a transition under control and the number of transitions under control predicted by the model:

$$\Delta a_{ij} = \lambda \frac{(\Delta a_{ij} + \sum_{T-1} nx_{ij}(O_m) - \sum_{T-1} A^1_{ij} nx_i(O_m))}{M * T} \tag{3.2}$$

3.2.3.1.3 Experiment 1: description and results In the learning experiment, we will first simulate the observation sequence where transtions are functions of the control signal. Suppose we have $K = 5$ states and a $5x5$ transition matrices. We set the intercept $b = 1$ and produce a $5x5$ matrix of random numbers for the slope coefficient a. Then we obtain the matrices A^0 and A^1 as the softmax values from our $5x5$ coefficient matrices where

$$A^0_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

$$A^1_{ij} = \frac{e^{a_{ij}x + b_{ij}}}{\sum_k e^{a_{ik}x + b_{ik}}}$$

Then, we generate a control sequence as $X = 0001001000...$ with the probability of control signal p_0.

Finally, from this matrices and the control sequence we generate M observation sequences of length T with K states where all states are equally probable as initial states (i.e. no particular initial state distribution). When there is no control, the transition happens with probabilites from A^0, and when the control is present, transition probabilities are those of A^1.

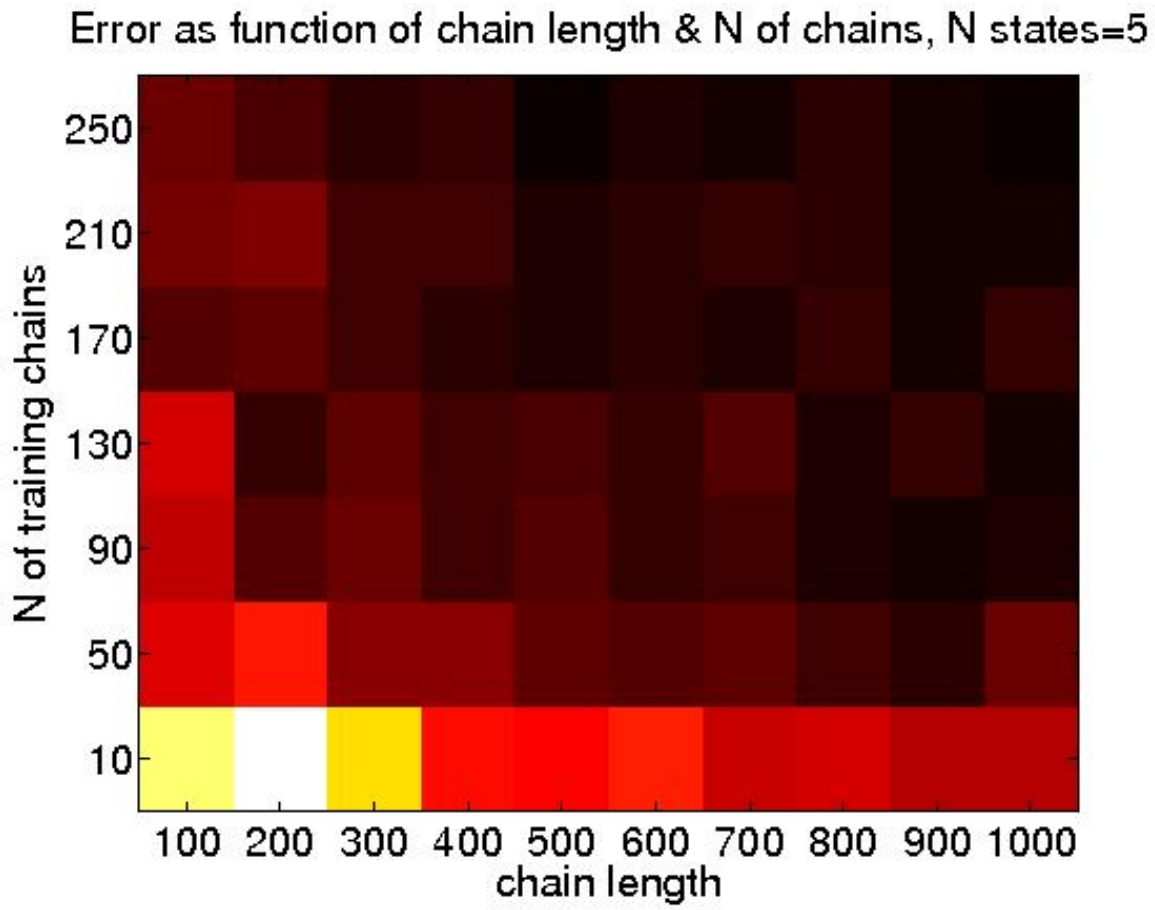


Figure 3.5: Error as function of the amount of training data. Training chain length is shown on the x axis, and number of the training chains on the y axis. We can see that, as expected, the error tends to be the lowest in the top right corner of the map (darker heat map), corresponding to more extensive training data - more chains and/or longer chains.

In the second part of the experiment, we use the generated sequences as the training data and try to learn the coefficient matrices A^0 and A^1. We start by a random guess and update the matrix according to Eq. 3.2 and Eq. 3.1 until convergence.

We tested the learning curve as function of different parameters: number of training elements (length of the training chain T and the number of chains M) and number of state in the transition matrix. The tests conformed with the expected outcomes. First, when more training data is available (more chains, longer chains), the error is reduced considerably - Figure 3.5. Second, the models with more states are more complex and harder to learn, and the error grows proportionally to the number of states in the model - Figure 3.6.

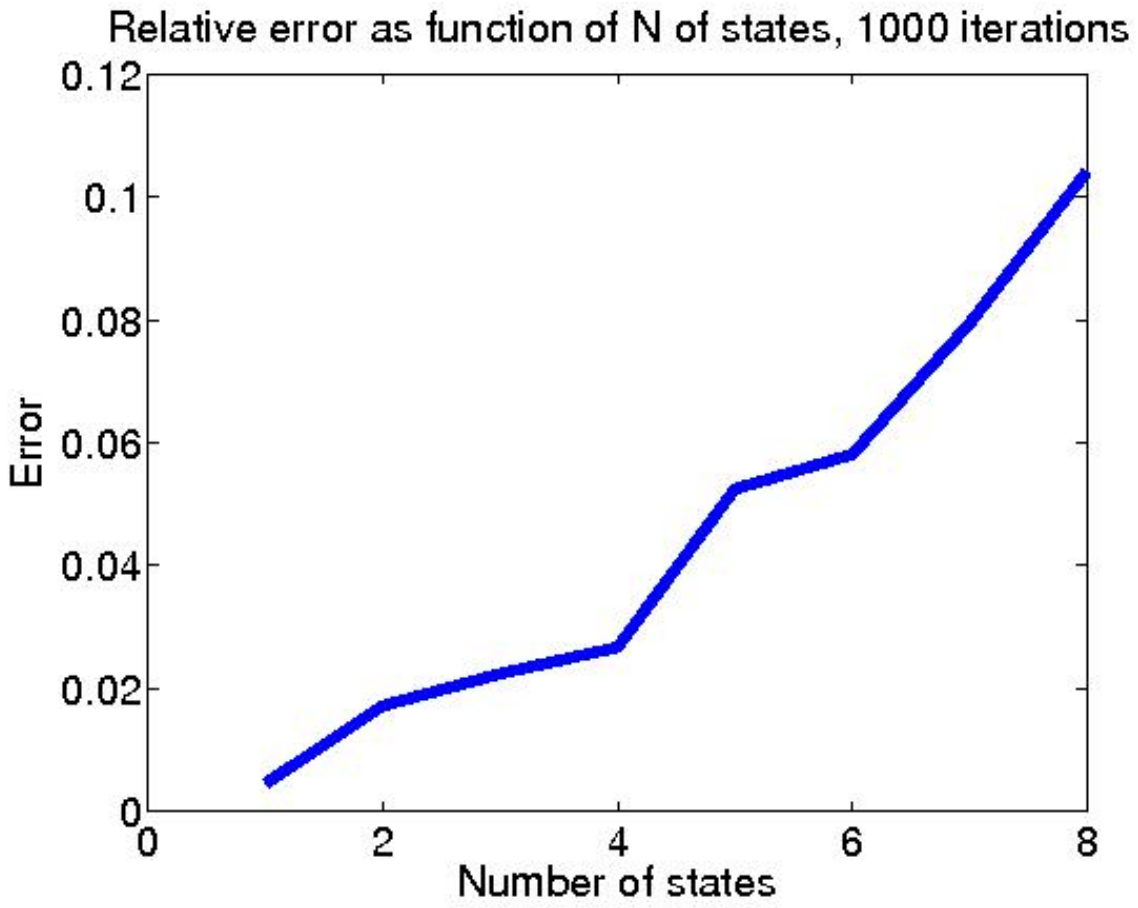


Figure 3.6: Error as function of the number of states. As expected, models with more states are learned with higher error.

3.2.3.1.4 Expectation Maximization (EM) algorithm: Experiment 2 In this part of the experiment, we will try to learn both the control sequence used to generate transition chains and the model - transition matrices. In this way we simulate an experimental situation, when we have to both perform inference about the system and search for an optimal stimulation strategy. We will do that by running an Expectation Maximization algorithm (EM). At the Expectation step, we will estimate the control sequence that generated the chains, and then at the Maximization step we will try to learn the transition matrices with the estimated control sequence. These transition matrices will then be used to estimate the control sequence at the next algorithm iteration, and the cycle will continue until convergence.

Let's introduce the notation x_{tj} where $j \in [0,\ 1]$: x_{t0} means position t where $x_t = 0$, and accordingly x_{t1} means position t where $x_t = 1$. Elements of the control sequence will be denoted as s.

$$p(\bar{x}_{1...T}, \bar{s}_{1...T}|A^0, A^1) =$$

$$= \Pi_{t=2}^{T} p(\bar{s}_t|\bar{s}_{t-1}, \bar{x}_{1...T}, A^0, A^1) =$$

$$= \Pi_{t=2}^{T} \Pi_{i,j}^{K} (A_{ij}^0)^{x_{t0} s_{tj} s s_{t-1i}} (A_{ij}^1)^{x_{t1} s_{tj} s s_{t-1i}}$$

We maximize the posteriors with respect to x, selecting between the two matrices the one that is active at step t.

$$p(x_{1:T}|p(s_{1:T}, A^0, A^1) = \frac{p(x_{1:T}, s_{1:T}|A^0, A^1)}{p(s_{1:T}|A^0, A^1)}$$

The denominator is just a constant, so for further we can ignore it. Let's calculate the log likelihood of the control sequence:

$$\log p(x_{1:T}, s_{1:T}|A^0, A^1) =$$

$$= \sum_{t=2}^{T} \sum_{ij} x_{t0} s_{tj} s_{t-1i} \log A_{ij}^0 + x_{t1} s_{tj} s_{t-1i} \log A_{ij}^1$$

So, we want to understand which of the matrices was most likely to produce the current state. Then, the matrix is the one that maximizes the following equation:

$$\hat{x_t} = \arg\max_q \{A_{s_{t-1}st}^q, A_{s_{t-1}st}^q\}, q \in \{0, 1\}$$

3.2.3.1.5 Experiment 2: description and results In this experiment, we will generate the data the way it was generated in Experiment 1 in 3.2.3.1.3. Next, we will start with random matrices and try to learn the control signal from them as the expectation step. With this control, we will enter the maximization step described in 3.2.3.1.4 and then repeat the Expectation step.

As it has beed said, data generation will be the same as in Experiment 1, but we will use particular matrices A^0, A^1 so that they match the most plausible conditions in real experiments. For instance, the system transition matrix without control A^0 is typically a diagonal matrix with some non-diagonal elements. As the matrices are softmax matrices, the sum of the rows should be equal to 1 (constraints described in 3.2.3.1.1). Let's choose values of transition probabilities in the A^0 matrix in the following way: $a = 0.99$, $b = 0.005$, so that one corresponds to the diagonal transition, and the other - to the non-diagonal transition. For the matrix to follow the softmax constraint, we constrain a and b in the following way:

$$a + b \approx 1$$

and the example matrix should look like the following:

Table 3.2: Example of a transition matrix

a	b	0	0	b
b	a	b	0	0
b	b	a	0	0
0	b	b	a	0
b	0	b	b	a

Figure 3.7 and figure 3.8 show that it is possible to learn the control signal by choosing which matrix was active at each step. A^0 shows state transition probabilities in the absense of control, while A^1 is the matrix of state transition probabilities under control. The comparison of the "ground truth" matrix A^1 (A^1 true in the figure) with the learned matrix A^1 (A^1 est in the figure) demonstrates that the model is learned quite well. The figure shows the matrices that allowed to achieve the best performance, those defined as mentioned above - A^0 strongly diagonal and A^1 somewhat noisy. Again, the comparison of 3.7 and 3.8 shows that more training data in the latter case (3.8) allows to learn the model faster.

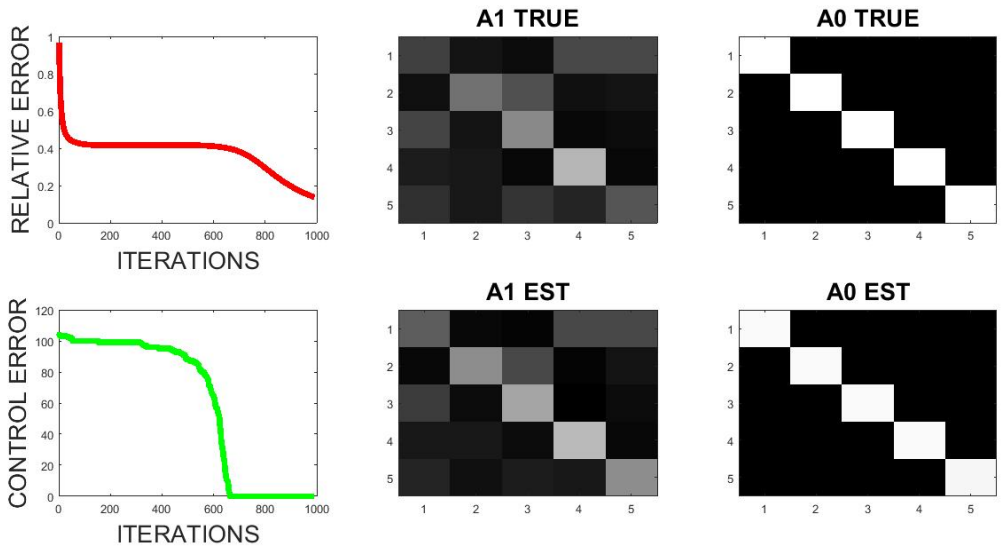


Figure 3.7: Learning curve of the control sequence, with 100 training chains of the length 1000. A^0 is the matrix of state transition probabilities in the absense of control, A^1 is the matrix of state transition probabilities under control. A^0 true and A^1 true are the ground truth matrices, A^0 est and A^1 est are the learned matrices. The red curve in the upper left corner shows the error dynamics while learning the model. The green curve in the bottom left corner is the error curve for learning the control sequence.

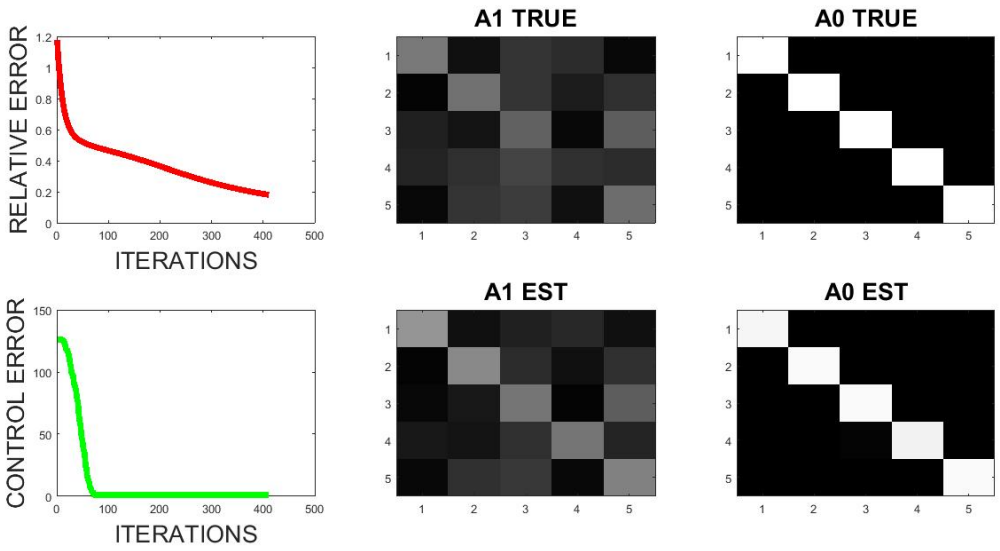


Figure 3.8: Learning curve of the control sequence, with 100 training chains of the length 1200. A^0 is the matrix of state transition probabilities in the absense of control, A^1 is the matrix of state transition probabilities under control. A^0 true and A^1 true are the ground truth matrices, A^0 est and A^1 est are the learned matrices. The red curve in the upper left corner shows the error dynamics while learning the model. The green curve in the bottom left corner is the error curve for learning the control sequence.

3.2.3.2 Controllability

In 3.2.3.1, we set out to show how it was possible to learn the model of the system under control and the control sequance that allows to produce a certain chain of states. However, before diving deeper into designing control systems for human brain, we might want to actually check if the system (the brain) is controllable at all. In this section, we are going to propose the apparatus for this purpose.

If we model the state dynamic in the subject with the application of control signal, we can describe it with a filter equation:

$$\bar{x}(t+1) = \bar{A}\bar{x}(t) + \bar{B}\bar{u}(t) + \bar{\epsilon}(t),$$

where x variable corresponds to the subject states, and u - to the control signal. For simplicity, we assume that the states are fully observable, so $\bar{x}(t) = x(t)$.

If we take the partial derivative of this equation with respect to x, we get a differential equation of the form:

$$\frac{\delta \bar{x}}{\delta t} = q\bar{x} + \bar{\epsilon} + \bar{B}\bar{u}(t)$$

where $q = \bar{A} - 1$.

If we pass to the time step form, we get:

$$\tau \frac{\delta \bar{x}}{\delta t} = qx_t + bu_t + \epsilon$$

If we then integrate both parts, we get:

$$x(t) = \int_{-\inf}^{\inf} k_t u_t d\epsilon$$

where k is a kernel matrix $k = e^{-qt}$.

We are interested in the properties of this matrix because it shows the relationship between the control signal and the subject's state. If it is a full rank matrix, we can say that the subject's states are controllable. If it is rank deficient, then controllability is an issue. So, we need to check for non-zero eigenvalues.

The matrix can be learned from the data of the subject's states and corresponding control sequences by applying the Expectation Maximization (EM) algorithm as described in [92], [194].

3.2.3.3 Cost Functions

As was explained in 3.2.1, if we want to learn the optimal sequence of control inputs, we have to define a cost function meant to minimize the error between the measured state and the target state. In the next two sections (3.2.3.3.1 - 3.2.3.3.2) we are going to contemplate what a cost function could be like in case of controlling human brain based on the neuroimaging signal.

3.2.3.3.1 The simpliest model In the simpliest case, we represent the current signal (corresponding to a certain state of the subject) as a function from the previous state x and the applied control signal a (which can be 0 or 1):

$$\hat{x}(t+1) = f(x(t), a(t))$$

Let's assume that in a hypothetic experiment described in 3.2.3.1, we have two types of stimuli: the control signal 1, evoking a response (e.g. an arousing positive stimulus) and the control signal 0, evoking a weak or no response (e.g. a neutral non-arousing stimulus) that makes the signal decay, so:

$$\hat{x}(t+1) < x(t)$$

if a(t) = 0, and

$$\hat{x}(t+1) > x(t)$$

if a(t) = 1.

We can see from the signal, that if no stimulation of type 1 (arousing positive) is administered, the signal level descends gradually over time. Because of that, we introduce a decay constant - let's set it to 0.1. Then, each next state will decay by 0.1 if there is no arousing control input and grow by 1 discounted for decay of 0.1 = 1 - 0.1=0.9 in the presence of positive arousing input:

$$x(t+1) = x(t) - (0.1 - a(t))$$

So, the expected signal at some point will be equal to the accumulated neutral stimulation minus accumulated arousing stimulation:

$$\mathbb{E} = \sum_t ((x^0(t) - x^1(t)))^2$$

As neutral stimulation can be expressed just in the decay constant, we can model the signal at any timepoint t as the difference between administered arousing stimulations minus decay constant multiplied by the number of timepoints as the expression of the non-arousing stimulation:

$$x(t) = (-0.1) * t + \sum_{t=1}^{t} a^1(t)$$

Based on that, we can formulate the error between actual and predicted signal as:

$$\mathbb{E} = \sum_{t}(x_\gamma - (-0.1 * t + \sum_{\tau=1}^{t} a_\tau))^2$$

And we have a cost function that can be minimized.

3.2.3.3.2 Control Theory and Filter Equation approach On the other hand, the signal can be modelled linearly with GLM. The canonical response function adopted for several types of human response signal ([13, 52, 83]) gets convolved with the vector of zeros and ones representing the design matrix, and model parameters are obtained.

$$x(t) = \beta^T h_\tau * a(t) + \epsilon(t)$$

In a standard GLM analysis, $h_\tau * a(t)$ is given, and we try to learn the parameters β. For our task, we have β and h_τ but we need to figure out the design matrix (i.e. the control signal) that will give us the desired level of the signal x(t). How can we do it?

One tentative way would be starting with the filter equation:

$$x(t+1) = Fx(t) + Gu(t+1),$$

$x(0) = x_0$.

This can be regarded as a difference equation. Let's consider the general form of a difference equation:

$$x_t = a_t x_{t-1} + b_t$$

The solution of the difference equation has a form of:

$$x_t = \prod_{\tau=1}^{t} a_\tau \times x_0 + \sum_{\tau=1}^{t} \prod_{\psi}^{\tau-1} \times b_\tau =$$

$$= e^{\sum_{\tau=1}^{t} \log a_\tau} \times x_0 + \sum_{\tau=1}^{t} e^{\sum_{\psi=1}^{t} \log a_\psi} \times b_\tau$$

If we introduce the transfer function H(s) [108] relating the input and the output, we can regard it as the solution of our difference equation:

$$x(t) = \sum_{\tau\prime=0}^{t} H(F, G, \tau)u(\tau) + \sum_{\tau=0}^{t} F^\tau x_0$$

At the same time,

$$H(F, G, \tau)u(\tau) = \beta^T h_\tau * a(t)$$

That is, we can estimate this part from the data with the GLM.

Then our reward or error (that we are going to minimize or maximize correspondingly) is going to be like:

$$R = \sum_{t=0}^{T} (x(t) - r(t))^2$$

where r is the reference level of the signal.

If we plug in the expression with the transfer function ,we get:

$$R = \sum_{t=0}^{T} (\sum_{\tau=0}^{t} H(F, G, \tau) * u(\tau) + \frac{1 - F^{t+1}}{1 - F} \times x_0 - r(t))^2$$

Now, we have the error to minimize:

$$[F, G, u(t)] = argmin(R)$$

In our experimental process, we can assume that F and G are known (they are model parameters), and so:

$$\hat{u}(t) = argmin(R)$$

,

and the error to minimize to find $u(t)$ would be ordinary least squares error:

$$E = \sum_{t} (x(t + 1) - (F(x(t)) + G(u(t))))^2$$

At this point, we can outline an expectation maximization algorithm: start with random control, estimate F and G and then iteratively reestimate control and consequently F and G.

3.2.3.4 The challenge of state definition

As we have seen in the previous sections, we have theoretical tools in place to frame protocol adaptation for BCI and NF as a closed-loop optimal control problem. We can learn the sequence of (stimulation) inputs that produced a given sequence of (brain) states interacting with the system (neural) dynamics, and we can also demonstrate how to address the learning problem of joint inference and control (section3.2.3.1). We can check for controllability of the brain system (section 3.2.3.2), and define cost functions that would allow to learn the control sequences (section 3.2.3.3). However, there is one huge missing piece: defining state space and the target state in terms of meaningful mapping between neural signal and its measurement on the one hand and cognitive and mental states on the other (examples of these mental states are the levels of arousal, fatigue, emotion, workload or other variables). Defining control goals for neural applications is challenging because the mapping from neural states and state transitions to cognition is presently incomplete. First, neural processes are extremely complex, relevant brain states are present at multiple time scales (i.e. circadian rhythm, alpha oscillation phase etc.), and a major challenge for any mind control goal is to identify which aspects of neuronal activity actually form a brain state. Next, in clinical contexts, we often know that we want to guide behavior away from a particular pattern, but we might not necessarily be able to define the target state we want to achieve [228]. Moreover, optimizing intervention protocols to individuals will require knowledge of neural representations within a single individual. Both researchers and clinicians must carefully evaluate our models of health, neuropathology, and neural repair in light of the problem of brain and mind control if we want to solve it [177].

The ability to detect states of the system is the critical component in the controller-observer (feedback sensor) cycle of closed-loop control. In case of human brain, this is easier to achieve invasively, when the sensors and controllers have direct access to neurons - e.g. with implanted microsensors or optogenetics. However, these have limited application, especially in experimental neuroscioence. Still, the problem of modelling brain states in terms of mappings between neural code and signal, on the one hand, and cognitive and behavioral labels, on the other, is at the core of control problem and has to be addressed. For instance, human behaviour in several studies has been modelled

using state space models [120, 204]. Some studies have defined states in the neuroimaging data (EEG or fMRI) - [191], [93], [243]. But for coupling behaviour and cognition with neural signal we need to make neuroscientists think in these terms so that we can apply the control approach. While cognitive neuroscience can inform the development of real-time applications such as BCIs, the reverse also holds. Actually, the development of closed loop protocols brings to us several new opportunities to investigate functional brain hypotheses in detail. Broadly speaking, it pushes neuroscientists to refine and identify clear hypotheses about the mapping between brain signal and neurophysiological markers, on the one hand, and cognitive processes and mental states on the other. In this light, the paradigm of BSDS is the fundamental tool that would inform immensly the development of closed-loop control protocols for human brain.

Given the difficulties in measuring neural states, modeling system dynamics, and estimating signals within noise, simplifying assumptions can be useful to facilitate insights into control of brain function. Encouragingly, neural control strategies can be designed with limited information about neural systems. Thus, while an ideal control strategy would include precise mapping between neural codes and the mind, numerous findings in brain stimulation research across neural scales suggest that the mind can be influenced, and not necessarily fully controled, in ways that provide interpretative and practical value. In other words, the brain can be guided rather than controled in its entirety, and it can still have positive impact [100]. Even in the absense of an ideal control strategy, "low-cost" control strategies could be designed for the clinical and experimental scenarios that could be highly beneficial [177]. As a proof-of-concept, we are putting forward an example of such a strategy with a reduced level of complexity. We are going to show how we can use closed-loop optimal control approach to adapt the stimulation protocol and drive the subject toward the target state given that a meaningful state definition is possible. We propose a closed loop experimental protocol using the signal that has a relatively straightforward mapping to a cognitive (affective) state, that of arousal: Electrodermal Activity (EDA) recordings, or Galvanic Skin Response (GSR) signal. We chose the GSR for two reasons: first, it is a unidimensional signal, unlike the multivariate signal you get in fMRI, which introduces an extra level of complexity; and second, GSR signal is known to linearly correlate with the subject's arousal. Because of this correlation, we can avoid modelling separately the relationship between the signal and the state and even say that by controlling the signal we are controlling the arousal. We use a predictive model to estimate the next state of the system under control based on its current state and to choose the stimulation sequence that minimizes the error between the target signal and

the estimated signal.

Chapter 4

Brain State Dependent Stimulation for the Study of Top-Down Attention

In this chapter, we are presenting our contribution to the development of BSDS protocols for the study of top-down attention. As has been mentioned in the previous chapter, neuroscientists distinguish between bottom-up and top-down attention. When attention is attracted by external events (e.g. loud noise or unpleasant smell), this is an example of how bottom-up attention works. Top-down attention implies internal state of readiness to attend to certain types of objects, e.g. when the person expects to see an object and is actively screening the field of view in search for it. The study of top-down attention would greatly benefit from efficient BSDS methods, when we can reliably detect the state of attentional readiness and adjust stimulation respectively.

With BSDS, we would like to test the hypothesis of the following type: if the stimulus is perceived in the moment that a top-down attention pattern is active, it is perceived differently from when it is not active. This is the same as to say that when you attend to something, it results in a more clear perception of this object. To investigate this hypothesis in a real-time experiment, we would need to send the stimulus in the moment when an attentional state is detected, and then we can compare those trials to trials with no attentional state detected. As a result, we expect to observe behavioural differences between the two types of trials. Before we can launch the development of the setup, we have to answer two questions: the first one is about how we can define these patterns to be detected and the second one about how we are actually going to detect them. Besides, detection has to take place in real time as we want to be able to send a stimulus in the moment the state is detected. So, the tools we are going to use should be suitable for a real-time runtime policy.

In the previous chapter, we described an experiment where top down attention patterns were hypothesized to be cross-modal brain activity patterns, i.e., common patterns between different cognitive modalities - e.g. perception vs. imagery. The task in which the subject was tested for these patterns was a visual search task: the subject had to search for an object category in a scene. Attention state was thought to be defined as a high-level category pattern that shows up in a cross modal fashion. Several studies investigated these shared patterns between perception and imagery and agreed that imagery results in a more general and high level pattern. Because of this we, too, expected the top-down attentional pattern to be similar to imagery signals while the subject was attending to a given stimulus category. To detect the category we planned to use cross-modal decoding: the classifier that was supposed to monitor the subject's state while they were preparing for visual search, was going to be trained on imagery data and tested on the preparatory activity in the visual search task. The attention state was expected to positively affect performance.

However, in the course of data analysis, it became clear that cross-modal decoding was not adapted to a real-time BSDS setting. This experiment allowed us to bring to light methodological limitations of decoding first with regard to cross-modal analysis and second with respect to online analysis. In an initial attempt to base BSDS on cross modal decoding we identified various drawbacks of using classifiers for this purpose: for one, high type II error. Cross-modal decoding is sensitive to ROI size (noise and the curse of dimensionality) and to the number of training samples. The question of shared activity patterns between modalities is of vital importance in neuroscience, because it contributes to the study of such fundamental neurocientific problem as the interaction of top-down and bottom-up cognitive processes in the brain. So, for the cross-modal analysis we proposed a different method - a statistical inference test, that is described in Chapter 6. For the online analysis of top-down attentional patterns, we also propose a different, unsupervised machine learning method and develop a real-time setup that allows to carry out real-time BSDS experiments for the study of top-down attention.

4.1 Unsupervised machine learning methods for neuroimaging data

So far the overwhelming majority of brain decoding studies have used supervised learning algorithms where the researcher provides stimulus category labels with the training data so that the Machine Learning (ML) algorithm can predict labels for the test data. At

the same time, unsupervised learning methods have also been employed for decoding, and there is growing interest in their application in neuroscience. Unsupervised methods are reference methods for the analysis of resting state fMRI - collecting neuroimaging data of subjects when they are not involved in any cognitive tasks and are just awake and engaged in mind wandering. But as we have been working on task-based fMRI, i.e., scanning subjects involved in various cognitive tasks, I will mention a few examples of the use of unsupervised methods from task-based fMRI. For instance, in [267], an unsupervised transfer method (spectral transfer using information geometry) was tested, which ranks and combines unlabeled predictions from an ensemble of information geometry classifiers built on data from individual training subjects. The method is meant to be used in BCI context for calibration in view of large intra- and inter-individual variability of brain-signals. In several studies, unsupervised pretraining of Deep Neural Networks to be used on fMRI data was performed with Restricted Bolzmann machines [118], [59,132]. This kind of unsupervised pre-training scheme is particularly well-suited when the supervised training data are limited, which is always the case in neuroimaging studies. Other examples include PCA/ICA (e.g. [6]), or different clustering methods: self-organizing maps [81], representational dissimilarity matrices (e.g. [139]), and also hidden Markov models (e.g. [73]). Most of the cited papers deal with the traditional offline decoding, but the efficiency of these methods for real time decoding has been demonstrated, for instance, in [6] where the authors used a combination of hybrid SVM-Markov model and an Individual Component (IC) dictionary to decode online the state of a nicotine addicted subject in terms of craving vs. craving resistance. In [242] the authors compared 14 different ICA algorithms from the viewpoint of their performance in real-time experiments and identified top four in performance. However, they highlight the importance of parameter choice when using ICA algorithms in real-time experiments. So, we also tried to tackle brain state detection problem for a BSDS study of top-down attention employing an unsupervised algorithm: real-time application of Gaussian Mixture models (GMM).

4.2 Gaussian Mixture Models applied to neuroimaging data

The Gaussian Mixture Model (GMM) is a semi-parametric method for high-dimensional density estimation. It has been widely used across different research fields (such as computer vision, object recognition, tracking, segmentation etc): its flexibility allows for the modelling of complex and nonlinear pattern variations, it is efficient in terms of memory load, and there are algorithms for model parameter estimation that are theoretically

guaranteed to converge [28]. The most common way of fitting a GMM is the Expectation-Maximization (EM) algorithm [58]. Starting from an estimate of model parameters, clustering of data data is computed (the Expectation step) which is then used to update the parameters in the maximal likelihood (ML) manner (the Maximization step). This two-step cycle is repeated until convergence, which is theoretically guaranteed. GMMs are most often initialized using the K-means clustering algorithm [9].

GMM have also been used with fMRI data in the context of various studies. For example, in [229] the authors reconstructed orthonumeric symbols that the subjects were viewing from fMRI data. In this study, GMM was used to expand their linear Gaussian framework for percept decoding to better represent the prior distribution of natural images: different mixture components correspond to different character categories. What is most important for the present study, GMMs have been fitted to detect activation in brain voxels, where the two components were corresponding to two voxel states: activated vs. deactivated. In [171] a generalization of a region based Bayesian detection/estimation approach was suggested that relied on a 2-class Gamma-Gaussian prior mixture modeling to classify the voxels of the brain region either as activated or unactivated. In [72], too, GMM was employed to extract the activation patterns from fMRI data and to model various activation patterns considering the strength, delay and duration of the epochs. In [217] a GMM based technique was proposed to estimate the onset and the duration of the voxel activation in situations where the nature, timing and duration of the psychological processes is not known (e.g. studies of drug uptake, emotional states or experiments with a sustained stimulus). The authors assume that the timing of a subject's activation onset and duration are random variables drawn from unknown population distributions. Then they fit a Gaussian mixture model with two components to each voxel time course, allowing that the shifts between active and inactive states occur at unknown times, and that between shifts observations are drawn from the same mixture component. The time between shifts is treated as missing data. The voxel shifts between states at a series of change points. This approach also allows for the possibility that some subjects may show no response which can certainly be the case in experimental studies.

So, following [217] and related line of research, we can model changes between brain states with GMM. Besides, in a BSDS setting we need to have the clustering work in real-time. However, real-time incremental learning of GMMs is a surprisingly difficult task. One of the main challenges here is the model complexity selection which is required to be dynamic by the very nature of the incremental learning framework. But at the same time if all information that is available at any time is the current GMM estimate, a

single novel point never carries enough information to cause an increase in the number of Gaussian components [9]. Still, incremental real-time techniques are available in robotics [37, 38]: in these studies, they were used to teach robots to imitate human gestures. To support brain state detection in real-time, we used a direct update method proposed in [9] and employed in [37, 38]. The method consists of a three-stage model update each time a new data point becomes available. At each time step: (i) model parameters are updated under the constraint of fixed complexity - i.e., the number of components; (ii) new Gaussian components are postulated by model splitting and (iii) components are merged to minimize the expected model description length. Only two GMMs are kept in memory, with no historical data: current GMM estimate, and Historical GMM - the oldest model of the same complexity after which no permanent new cluster creation took place

4.3 State identification: GMM to detect change in Percent Signal Change data stream

So, our goal is still to detect top-down attentional state in a visual search task, where there is a delay between the cue to search for an object and stimulus presentation when the subject has to respond if the object is present in the picture. The hypothesis we test states that during preparatory periods top-down attentional patterns are activated, and these attentional patterns correspond to the object category representation. In the previous experiment, these patterns, though specific for different categories, were found in the same ROI, OSC. However, they can also belong to different ROIs. For instance, it is well known that images of faces activate Fusiform Face area (FFA), while images of houses activate Parahippocampal place area (PPA, e.g. [250]). So, we could take the BOLD signal in each of these areas and try to detect change during the preparatory period. However, both areas also respond to the non-congruent object categories - such as houses for FFA and faces for PPA, but this amplitude is relatively low [251]. If we just take the absolute BOLD, it is prone to signal fluctuations that are not necessarily indicative of attending to the congruent category. That is why we should take a relative metrics of activity in the congruent area with respect to non-congruent area - such as Percent Signal Change (PSC). PSC will take into account only those changes in the signal that are higher than signal fluctuations in response to non-congruent category. So, if we detect change in PSC with respect to non-congruent area during the preparatory period, we are very likely to detect the expected state, that of attending to the congruent object category.

Following the approach outlined in [217], we propose to fit a GMM model on the PSC in one area with respect to the other, and PSC is calculated volume by volume during preparatory periods. We make no a priori assumptions about the nature of changes in the PSC timeseries. We model it as a two-state system with active and inactive states. PSC shifts between states at a series of change points. If the number of timepoints in the PSC vector of one session is N, its time profile is modeled as a sequence of normally distributed observations $y_i, i = 1 \dots N$ which may at an unknown time i undergo a shift in mean of unknown magnitude, representing a transition into an activated state. PSC may return to baseline at time $\tau_i + \omega_i$, where ω_i is also unknown. τ_i and ω_i represent the onset and duration, respectively, of the activated state (both are random variables drawn from some distribution, and in theory, could be estimated as well). We started by fitting a Gaussian mixture model with two components to each PSC time course, allowing that shifts between the activate and inactivate states occur at unknown times, and that between shifts observations are drawn from the same mixture component. This analysis was performed only during the preparatory period, volumes between stimulus onset and the onset of the next trial are treated as missing data. In the course of experiments, however, it was noted that the baseline could shift both within the same session and between sessions - due to low-frequency signal fluctuations and various effects on the subject's side - such as learning and adaptation. To accomodate this shift, we increased complexity of the models, trying various numbers of components between 3 and 5. In this case, even if the baseline fluctuates, the change from one component to another with a higher mean will still be detected.

An example of such fitting is shown in Fig. 4.1. We can see that baseline values were all clustered in Component 1, while values corresponding to preparatory activity and activated state are beling assigned to Component 2 and 3. Even when the baseline does not return to Component 1, the change between Component 2 and Component 3 can still be captured. Another observation we can make with respect to Fig. 4.1 concerns the fact that here we present a GMM model fitted on the whole PSC timecourse data, including both preparatory periods and stimulus presentation periods. We can see that the increase in PSC during preparatory periods is comparable to that during visual presentation of the stimulus. Usually the signal corresponding to some perceptual activity is quite strong even in higher cognitive areas, but at the same time in our experiment visual presentation was quite short, and that could explain relative parity of the signal strength between preparatory and presentation periods. The fact that both signals are on similar scale confirms our expectations to see preparatory activiy in the target areas and to duly detect

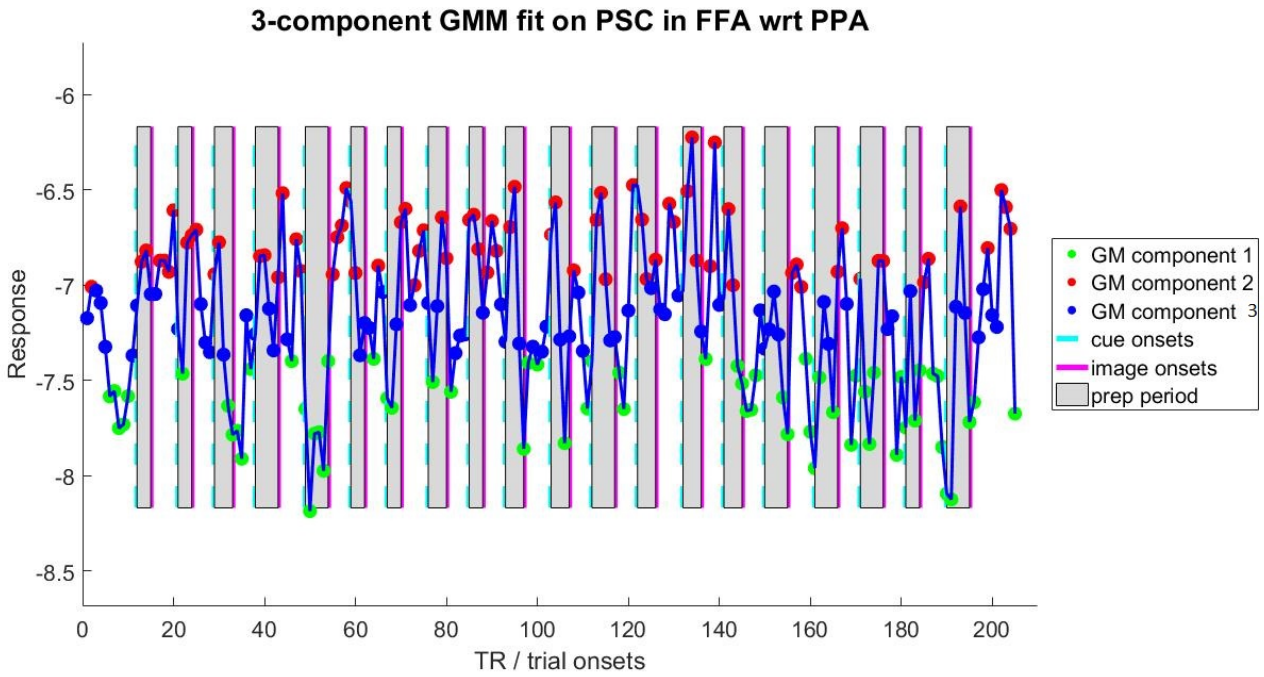


Figure 4.1: An example of fitting a 3-component Gaussian Mixture Model on the data coming from 1 experimental session of 1 subject. The detection task was face detection. PSC was taken from face responsive voxels in FFA with respect to house responsive voxels in PPA. Preparatory periods are shown as grey blocks, while stimulus presentation and response periods together with interstimulus intervals fall in white spaces. Cue onsets are marked with cyan dotted lines, and the end of the preparatory period corresponding also to stimulus presentation, is marked with magenta solid line.

it. Another remark concerns the fact that there are only 3 trials in 20 where the change during preparatory period could not be detected, which gives us a very good accuracy of about 80 %.

4.4 Runtime policy: GMM clustering in real time

Fig 4.1 was obtained in an offline mode, when the algorithm had all data available. In a real-time setting, only one data point will be available at each timepoint. We adopted implementing incremental online GMM clustering [9,37,38] where the GMM is initialized on a few data points with KNN, and then whenever a new data point arrives, Expectation Maximization (EM) is run to recompute model parameters; only the model is stored in the memory and not data points. This method detects change between components on 75-80

percent of trials, so it can be employed in an experimental setting to ensure that the state will be detected on the majority of trials, and consequently, the stimulus will be presented. A demo of this method is available (`https://github.com/elena-kalinina/BSDS_GMM_rt_demo/tree/master`), which demonstrates incremental fitting of a Gaussian Mixture Model (GMM) using a direct update method. In the demo, a PSC timecourse consisting of a number of data points is loaded, and these datapoints are presented one-by-one, in a real-time like manner, to update the GMM parameters by using an incremental version of the Expectation-Maximization (EM) algorithm in the direct update method (described in 4.2). The learning mechanism only uses the latest observation to update the model (no historical data is used).

4.5 Experiments

In this section, we will present two series of experiments carried out to develop our BSDS real-time setup. It is aimed at investigating the neuroscientific hypothesis about the role of top-down attention patterns in visual perception. The trials where stimulation is sent at the moment when preparatory activity is detected in the targeted area, are expected to demonstrate significant differences in behavioural performance (accuracy, reaction time, d'-criterion, ROC curve etc). So, the setup has to ensure that both types of trials - BSDS and control trials with random delay length between the cue and the stimulus - are collected and equally represented for statistical inference.

4.5.1 Offline experiment

In this experiment, we collected the data to develop and test methods for extracting signal and calculating PSC. We also tested GMM fitting on this data. The experiment had the scope of collecting FFA and PPA data in a visual search task and also to establich a behavioral baseline in terms of accuracy and reaction times. The stimulus categories used were noisy images of faces and/or houses, areas targeted FFA and/or PPA respectively. The subject was detecting faces or houses in noisy images, with a varying length delay period between the cue and the image (between 6 and 10 seconds) when the preparatory activity takes place. We took the PSC in the congruent area (FFA for faces) with respect to the non-congruent area (PPA for faces) volume by volume to distinguish between true activity and random fluctuations, and tested fitting a GMM on this data. In the BSDS design the stimulus is to be shown with the first activated volume after the cue, i.e. when the change from one component to another is detected.

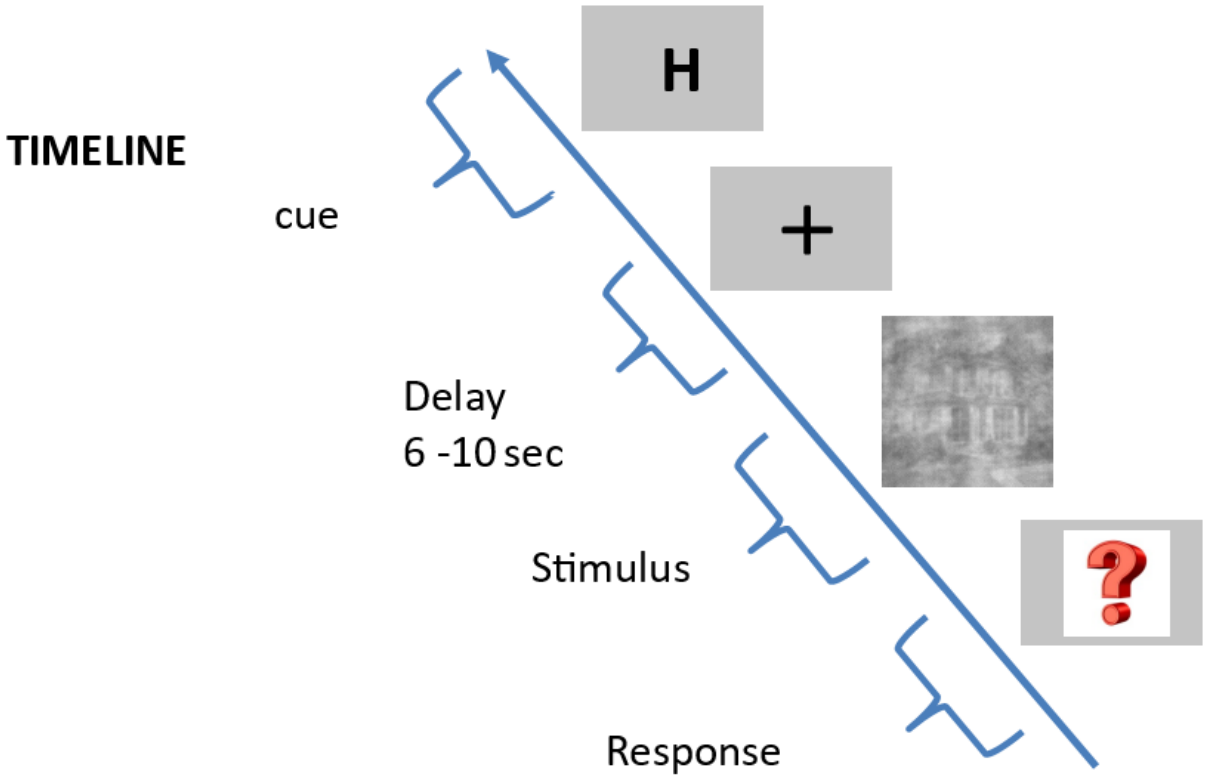


Figure 4.2: Trial scheme for the visual search experiment where the data was collected for offline analysis.

4.5.1.1 Trial scheme

The trial starts with a letter cue (to induce activity, H for houses or F for faces). Then comes a delay of varying duration when the screen goes gray (6-8-10 secs delay randomly varying on different trials). Next, the stimulus is shown for a brief period (300 ms), and the subject is given 2 seconds to respond if they have seen a face or a house in the noisy image. The task is a detection task, and the subject has to respond yes or no with a button press (right and left buttons of the respons box correspondingly). After each trial there is a fixation period of 6 seconds (when the subject only sees a grey screen and a fixation cross) to let the signal return to the baseline. The stimuli are pictures where the image of a face/house is overlaid with noise. On half of the trials, pictures are just pure noise. Other 50% are divided between different levels of contrast: 70% noise/30% signal, 45% noise/55% signal, 40% noise/60% signal (taking into account light and luminance in the scanner). In this way, we make sure we cover all sensitive range. The trial scheme is shown in Fig. 4.2.

4.5.1.2 Experimental setup

The experiment session starts with the functional localizer to identify voxels in PPA and FFA that are most responsive to houses and faces respectively. The voxels are identified through the following procedure. The subject sees blocked images of faces and houses, where each block of 20 images lasts for 14 seconds with 6 seconds fixation periods between the blocks. The run is broken into short series of blocks, each containing 4 blocks. After each series, there is a rest period (fixation) lasting 20 seconds. SPM contrast analysis is run on the data, and then functional masks of FFA and PPA derived by the method described in [121] are applied to select most responsive voxels within FFA and PPA. The masks are shown in Fig. 4.3. During the second run, the subject's high resolution anatomical image of the brain is acquired. Subsequent runs (3 - 9) follow the trial scheme described above with random delay timings (3-4-5 volumes = 6-8-10 secs), but face and house tasks alternate between runs: all odd runs are face task runs, while even runs are house task runs. There are 20 trials in a run.

4.5.1.3 Data and model

4 subjects took part in this experiment. Their data was used to test various complexity levels of the GMM model (2-5 components) and also various training scenarios (training data coming from the previous session vs. training data coming from the same session). Examples of ROI timecourses of an fMRI session are demonstrated in Fig. 4.4. In Fig. 4.5 we show an example of a PSC timeseries calculated on timecourses from Fig. 4.4. Figures 4.6 and 4.7 illustrate fitting GMM of different complexity (3 and 5 components) on the data of delay periods only (training data for initialization coming from the same session). We can see that higher complexity helps deal with the baseline shifts: with the number of components higher than 2 the change still can be detected. Fig. 4.8 illustrates how a 5-component model is robust when facing a baseline change: too high values are filtered out, and when the baseline goes down, we can still reliably detect 3 clusters in the data.

4.5.2 Real-time experiment

In this experiment, we developed and tested setup components pertaining to real-time model fitting and runtime policy. First, we tested various training scenarios. In the first scenario, previous session was used for initial clustering. However, although we are working with a relative measure, the baseline changes between sessions, as was explained

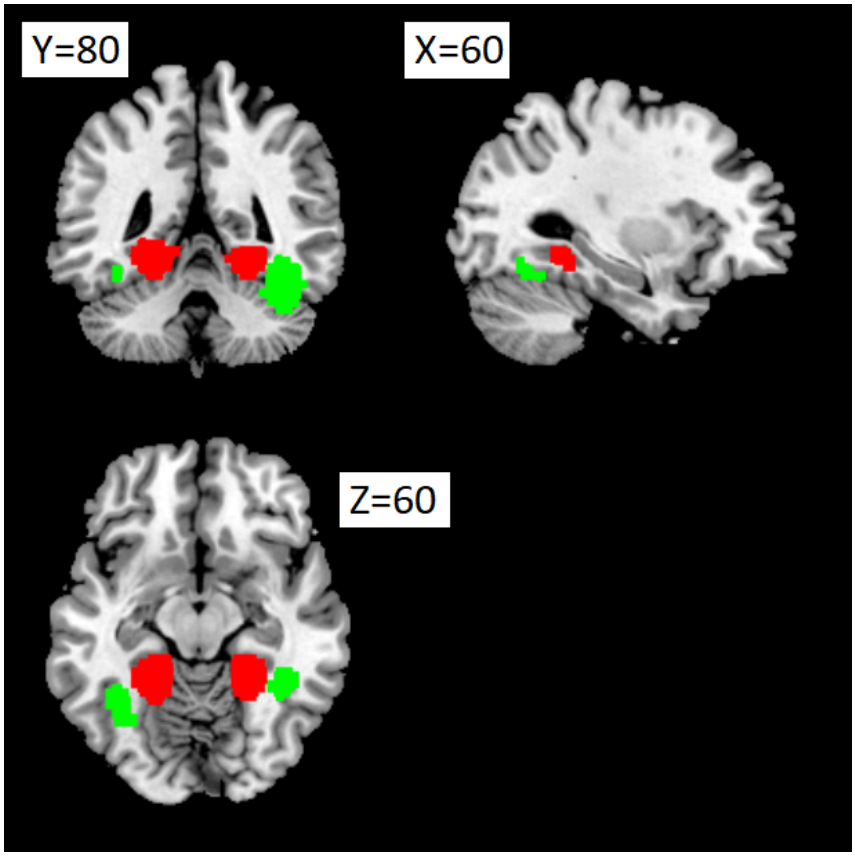


Figure 4.3: Functional masks used to extract responsive voxels: FFA is shown in red, while PPA in green.

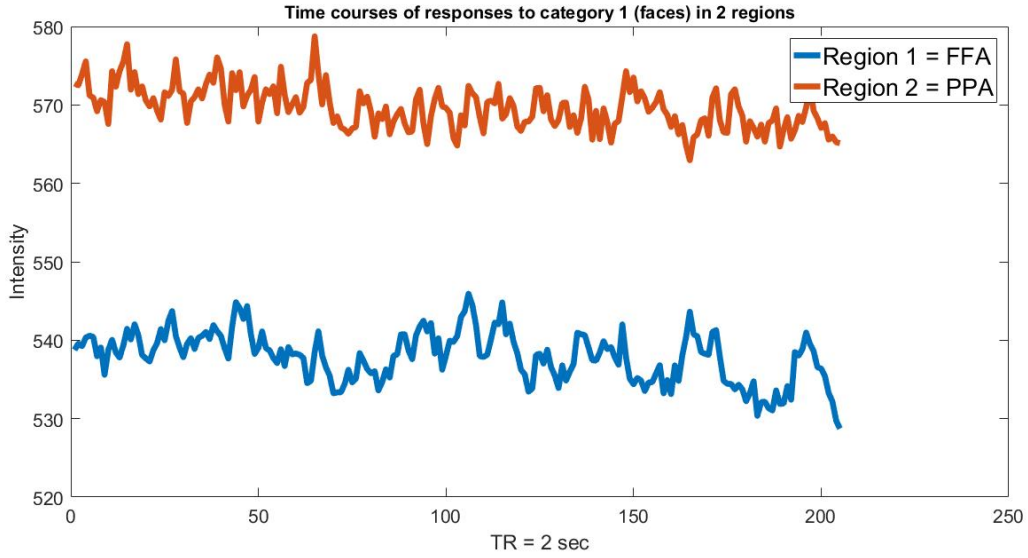


Figure 4.4: Full timecourses extracted from the ROIs obtained by the localizer analysis and functional masks: FFA voxels timecourse is shown in blue, while PPA voxels timecourse is shown in red.

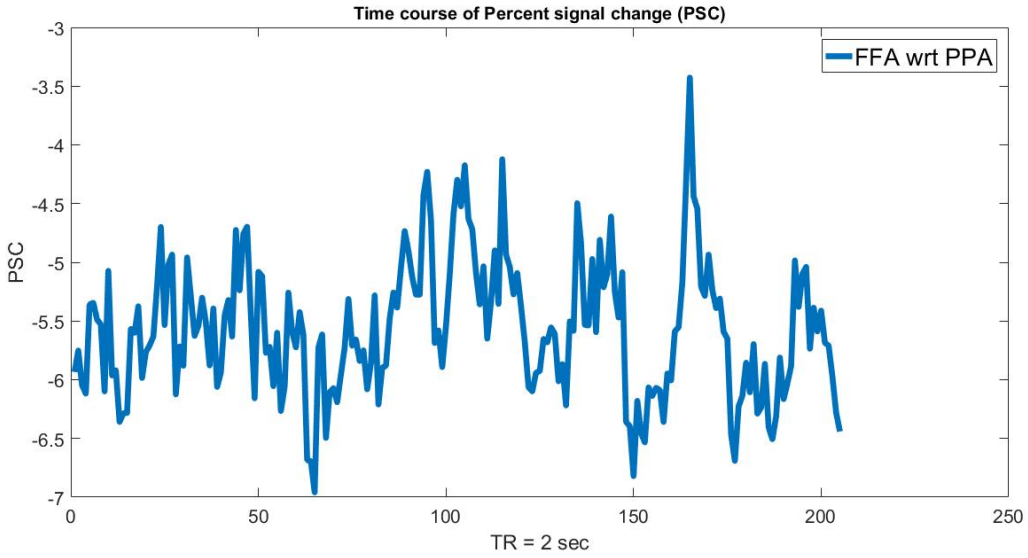


Figure 4.5: Percent signal change taken from FFA face responsive voxels with respect to the PPA house responsive voxels - full timecourse.

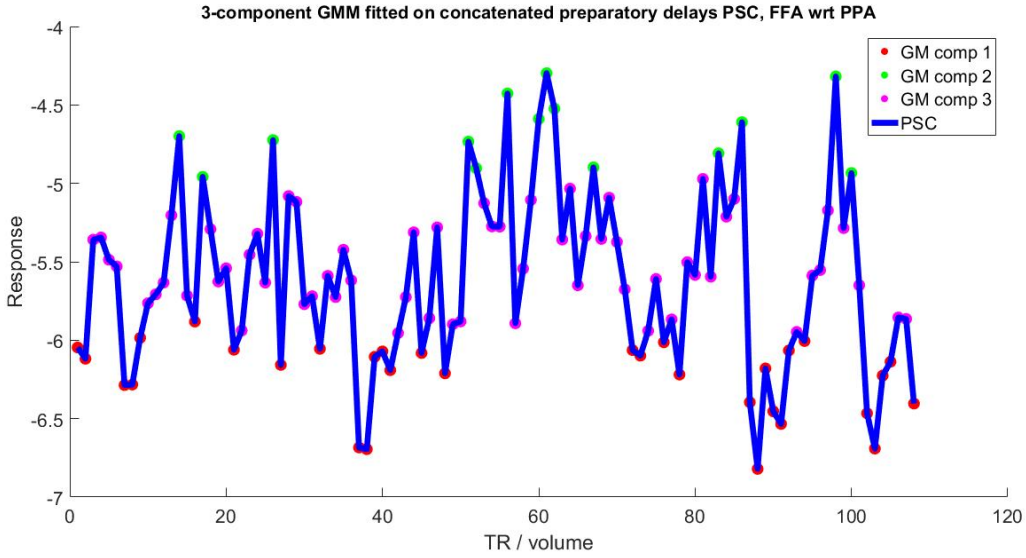


Figure 4.6: A 3-component GMM fitted to the PSC timecourse of delay period datapoints only (image presentation, response and fixation periods between the trials are not included).

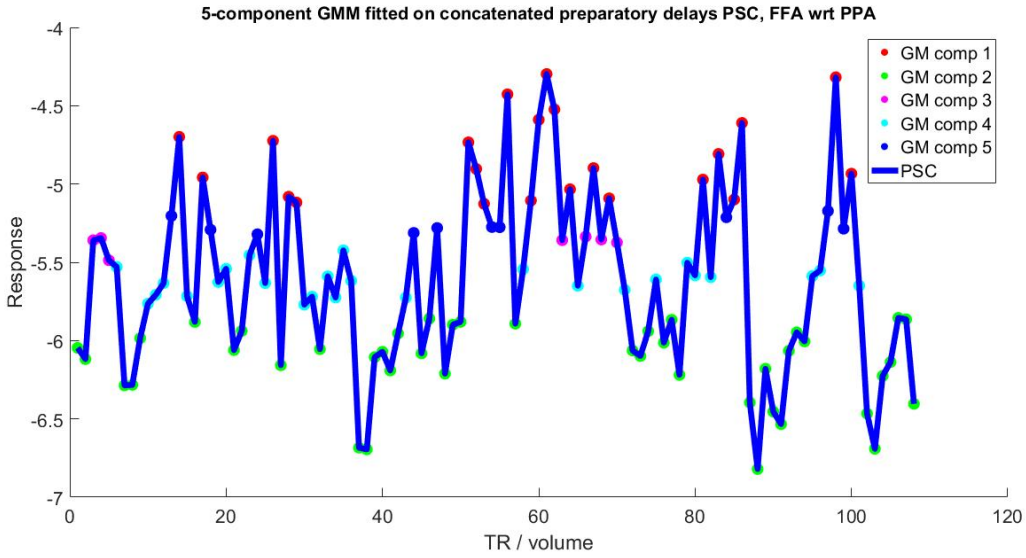


Figure 4.7: A 5-component GMM fitted to the PSC timecourse of delay period datapoints only (image presentation, response and fixation periods between the trials are not included).

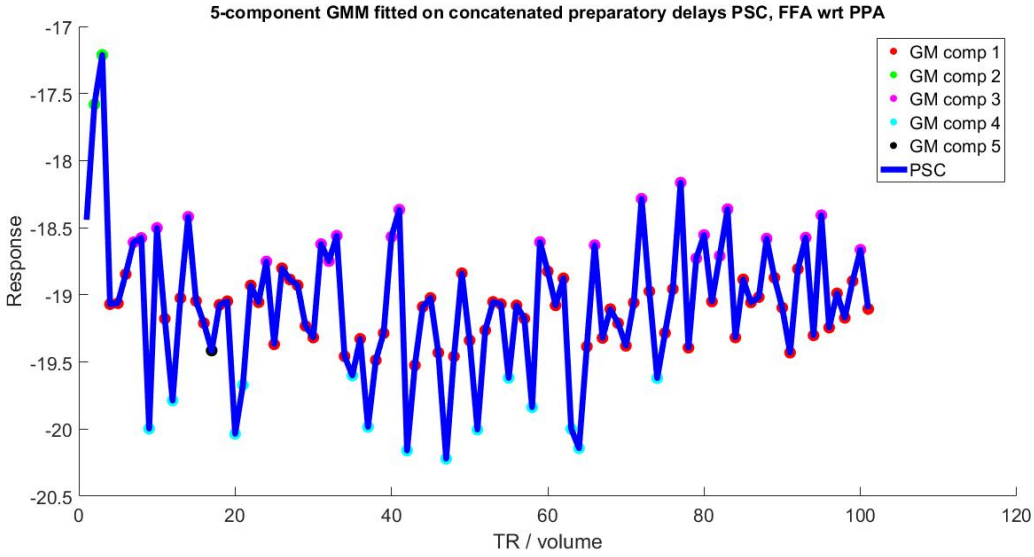


Figure 4.8: A 5-component GMM fitted to the PSC timecourse of delay period datapoints only (image presentation, response and fixation periods between the trials are not included). Although at the beginning there are some high values, they are "clustered out", and when the baseline goes dow, the data can still be clustered in 3 components corresponding to different activity levels.

in 4.3. Another scenario would be using half of the run for training and another half for BSDS. This is not desirable because we introduce a confound - a position in the run. BSDS trials will be systematically different by their position in the second half of the run, and we will not be able to attribute observed behavioural differences unambiguously to the experimental manipulation. So, in a compromise variant, we use 5 initial trials (of 10 second duration each, that is, 5 preparatory volumes each) for clustering and exclude 5 initial trials from the control runs from comparison with BSDS trials.

Second, we tested different levels of the model complexity, i.e., the numer of GMM components. Baseline changes within the run as well and not in a linear way (so that we cannot detrend): we have to deal with slow low-frequency fluctuations in the signal known to exist for fMRI signal on the one hand, and changes possibly due to various learning and adaptation effects in the subject on the other. We tested models with the number of components between 3 and 6, and 5 seems to be the optimal number that allows to still detect changes even with the changing baseline.

Finally, we tested the latencies. Most importantly, we had to make sure that in the moment the change is detected, the stimulus is shown with a negligible latency with respect to the change detection. Our setup conforms to this requirement. Another timing

issue was training time: initial clustering should not interfere with processing times. Clustering was performed during interstimulus interval, and in this way did not create any slowing down in the experiment course.

4.5.2.1 Trial scheme

In the experimental setup, the stimulation computer that shows images to the subject is connected to the computer performing data analysis in real time by means of the serial port. Overall trial scheme is similar to the one described in 4.5.1.1 for the offline experiment. The trial starts with a letter cue (to induce activity, H for houses or F for faces). Then comes a delay when the screen goes gray, and at this point the stimulation computer sends a signal to the processing computer to start processing volumes. PSC in the congruent area with respect to the non-congruent area is calculated, and the resulting value is assigned to a GMM component. If the value is assigned to a component with a higher mean that the previous volume, the processing PC sends a signal to the stimulation PC to show the stimulus to the subject and stops clustering until the next trial onset, when the signal from the stimulation PS is received again. The volume when the change was detected relatively to the cue onset (e.g., 1, 2, 3 etc.) is registered in the protocol record. These timings are then reshuffled and used in the control runs for stimulus presentation. In this way, we keep the same distribution of delay durations, but stimulus presentation is uncoupled from the activity state. The volumes between stimulus onset and next cue onset are treated as missing data. If no change is detected, the analysis goes on until the timeout of 10 seconds (5 volumes), and then the stimulus is shown. Usually when the change is detected, it happens earlier than 10 seconds, so when we have a 10 second delay in the protocol record, we know that it was a failed BSDS trial. The stimulus is shown for a brief period (300 ms), and the subject is given 2 seconds to respond with a button press if they have seen a face in the noisy image. The stimuli are pictures of faces overlaid with noise. On 50% of the trials, pictures are just pure noise. Other 50% are divided between different levels of contrast: 70% noise/30% signal, 45% noise/55% signal, 40% noise/60% signal (taking into account light and luminance in the scanner). The trial scheme is shown in Fig. 4.9.

4.5.2.2 Experimental setup

The real-time experiment, too, starts with the functional localizer in order to identify most responsive voxels in PPA and FFA, which is run in the same manner, as described in 4.5.1.2 for the offline experiment. The subject sees blocked images of faces and houses to

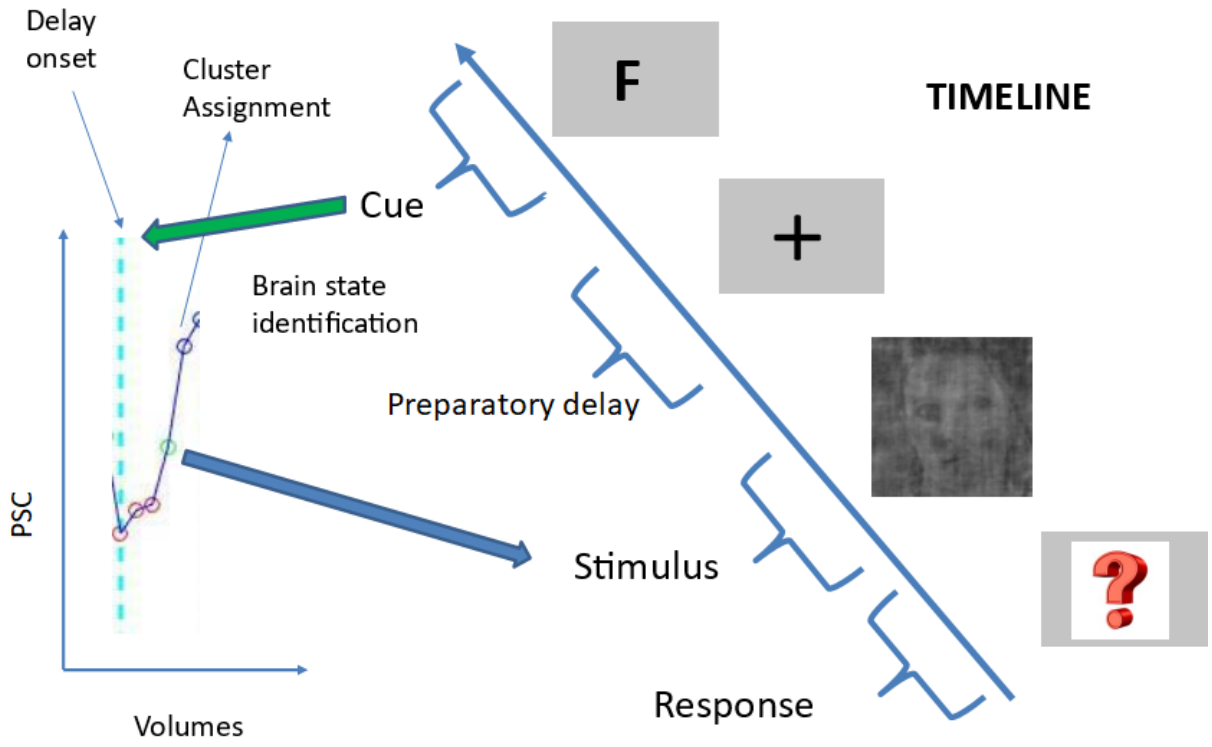


Figure 4.9: Trial scheme for the real-time BSDS experiment with the visual detection task.

identify voxels in FFA/PPA that respond to faces/houses accordingly. In the second run, a high-resolution anatomical image of the subject's brain is acquired; in the meantime, SPM contrast analysis is performed on the localizer data (1 -1 first for faces and then for houses), and the identified voxels are masked with the functional masks described above in 4.5.1.2 . The resulting ROIs are used throughout the subsequent runs to extract BOLD timecourses.

In runs 3 to 8 the task was limited to faces only. The first 5 trials all had delays of 10 seconds (5 volumes) and were used to perform k-means on the PSC in FFA face responsive voxels with respect to the house responsive voxels in PPA. In the real-time setup, BSDS runs were interleaved with control runs - in control runs, stimulus presentation timings were the shuffled timings from the previous BSDS runs (apart from the first 5 trials that will have the same delay duration of 10 seconds to control for a potential confound). The goal was comparing the trials where we send the stimulus at the onset of the brain activity (BSDS) with the trials where the stimulus timings are in no way related to the activity onset. If we use shuffled timings from the previous BSDS runs, on the one hand the timings and frequency on average will be the same as on BSDS runs but the correlation with the activity timings will be disrupted. In this way, we eliminate potential confounds consisting in introducing some very different timings on non-BSDS runs. However, we are assuming that there is enough meaningful variation in FFA/PPA activity in this 6-10 seconds period.

4.5.2.3 Data and model

4 subjects were scanned in this experiment. Each incoming volume was preprocessed performing slice timing correction and motion correction with respect to the first volume of the session as reference volume using a customized FieldTrip real-time fMRI preprocessing pipeline [196]. We took PSC in the congruent area (FFA for faces) with respect to non-congruent area (PPA for faces) during the preparatory period between the cue and the stimulus. We did it volume by volume on the absolute values (no z-scoring or other normalization). Next, we fitted a 5-component Gaussian mixture model to detect subpopulations in the data stream each having their own distribution parameters such as mean and variance. In this way, we were clustering all the incoming values into low, middle and high, so that we could say about each incoming data point which cluster it belongs to. Change, i.e. the rising signal, is detected when we pass from one subpopulation to another with a higher mean, i.e. if the previous value was low and the next one is middle, or if the previous was middle and the next is high. Initial model is obtained based

on the k-means clustering of delay period PSC in the first five trials. Starting with the cue onset, we assigned each value of the PSC to one of the model clusters, and whenever the next value was assigned to a higher cluster than the previous one, we showed the stimulus. The stimulus was sent on about 75-80 % of the trials.

These two experiments allowed us to put in place a fully functioning real-time setup for BSDS experiments for the study of top-down attention. However, for technical reasons we could not perform the neusoscientific validation of the method, that is, to test if the two types of trials, control and BSDS, result in any behavioural differences. As we piloted and adjusted various setup features in the course of real-time experiments, we do not have a consistent data set that would allow performing valid statistical inference to test the neuroscientific hypothesis. This is going to be the scope of subsequent experiments.

Chapter 5

Closed Loop Control of Affective States

Control of a subject's affective state is the ultimate goal of many real-time neurofeedback experiments. However, in such experiments it is the subject who is trained to control her emotional state, up- or down-regulating the activity in some targeted brain area. For instance, the aim can be to down-regulate the activity of the amygdala in order to weaken negative emotions [173] or, on the contrary, up-regulate the activity in the networks known to process positive emotions to enhance the subject's positive state of mind [157]. Or the subject could be suppressing craving responses to tobacco, alcohol or food [105, 115, 134]. In these experiments the control is performed by the subject and only indirectly by the external intervention. As has been pointed out in [22], patients and healthy subjects often show a large variability, or even inability, of brain self-regulation. Ideally, stimulation protocol in a neurofeedback setting should be adapted to the subject to better guide them towards successful regulation. One way to do this would be modelling the subject's activity in a way that we can evaluate the effect of various stimulation options and chose the optimized ones, maximizing the reward or minimizing the error. However, currently developing such models for neuroimaging experiments presents a number of challenges that make the task demanding and overcomplicated. The main obstacle is the complex nature of the relevant brain states so that it would be hard to define a target state and compare it to the current state in order to define the steps to achieving the target state by stimulation optimization. In view of these difficulties we chose to run an experiment where we could define some relatively (!) straighforward mapping between the state and the measurements: controlling arousal via recording Galvanic Skin Response (GSR) signal. This modality has such an advantage that we can define the target state

in meaningful terms, e.g., as "high arousal" or "low arousal".

5.1 Background of the study

Emotional control is not only a key concept for affective disorders such as major depression [277], but also in developmental (autism) [163] or mental disorders (schizophrenia) [141]. Furthermore, emotional control is no less important in healthy population - as, for instance, during emotional turmoil in adolescence [122]. Neurofeedback can train users to exert emotional control [239], with application in major depression [221], Post-traumatic stress disorder [214] or addictions [129]. However, in neurofeedback training the subject is assumed to learn a control strategy based on the *indirect supervision* of the feedback signal. While adapting neurofeedback stimuli to the subject leads to better results [21], further improvement would be conceivable if the learning process was more directly assisted [24, 86]. Such efforts should be considered within a context of general stimulus optimization, e.g., in [23, 62, 101].

Much experimental evidence related to emotion can be summarized within the so-called circumplex model involving the two main dimensions denoted arousal and valence [208]. The two dimensional circumplex representation may be seen as a convenient reduction of higher dimensional representations and linked to the hierarchy of neurotransmitters, see e.g., [18, 164], specifically arousal is hypothesized to be related to norepinephrine/noradrenaline [68, 103]. At the individual level arousal signals immediate or future relevance of given stimuli for survival and reproduction [102]. From a cognitive viewpoint, the arousal level is thought to guide attentional resource allocation during emotional stimulus processing: more arousing stimuli (e.g., threat or sexual attraction) have higher intrinsic motivational value and get more attentional resources as a result [193].

Autonomic reactions, such as increased respiratory rate, heart beat rate or skin conductance response (SCR) are all important aspects of the emotional response. Measured SCRs register fluctuations in the electrical conductivity of the skin caused by the sweat released from eccrine sweat glands. These are concentrated mainly on the palms of the hands and on the soles of the feet and reflect activation levels of the sympathetic branch of the autonomic nervous system [99]. SCR is found to be a reliable indicator of arousal [4,31] possibly mediated by norepinephrine/noradrenaline [26]. To study the basic dimensions of emotional response Lang et al. [150] collected a large set of photographic stimuli (International Affective Picture System, IAPS). These pictures were rated by young adults on the scale of valence (positive - neutral - negative) and arousal (high - medium - low

arousal). Such ratings allowed to systematically vary valence and arousal during experiments, so that covariation between dependent measures and emotional valence and arousal of picture stimuli can be analysed [147]. It was shown that significant correlation exists between the SCR magnitude and arousal rating [148, 149]. However, the causal relationship between the state of arousal and stimulation is complex. On the one hand, there is statistical relation at the population level: a viewer is more likely to be aroused when presented with a strongly negative image. On the other hand, at the individual level the causal relationship is subject specific [148, 149]. Arousal comes as the result of the interaction of the stimulus input and the current cognitive state.

SCR has become an important autonomic nervous system (ANS) measure in psychophysiological research [55], which is probably more widely used than any other ANS measures [175], viz., heart rate variability [226] or pupil size [32]. SCR is considered a direct and sensitive way to assess affective arousal and to gain insights about sympathetic nervous system reactions. Moreover, it has a number of advantages as a method: relative ease of setup, recording and processing, non-invasive nature, comfort for participants, low cost. Most importantly, it has high specificity in measuring arousal despite its simple waveform [99]. Besides, SCR recordings can provide information not only about autonomic nervous system activity, but also about brain states and emotional information processing: the activity of eccrine sweat glands is regulated by multiple brain structures that govern the autonomic nervous system, such as limbic regions, basal ganglia, and frontal cortex [65]. For example, measures from the autonomic nervous system can complement brain recordings in order to discriminate between cognitive processes of emotion evaluation on the one hand and experiencing emotions on the other hand [25].

Because of the observed link between the measured SCR signal and arousal, control over the level of the SCR measurements permits to gain control over the activity of the autonomic nervous system. This, in turn, could have significant consequences both for the study of emotion and also for clinical applications. To address this question, we carried out an experiment that aims at bringing the subject in the state of arousal with the desired level of arousal and maintaining the arousal level constant during the duration of experiment. While the subjects were viewing sequences of negative and neutral images, their skin conductance response (SCR) was measured with GSR sensors. When arousal decreases, that shows in low amplitude of the response to the stimulus and baseline downward shift. When the opposite is true, baseline rises, and so do response amplitudes. A predictive model was used to estimate the dynamics of the subject's arousal for the coming 30-40 seconds. Then the squared error was calculated between the signal predicted

for each of the possible stimulation sequence for the upcoming trials (3 or 4) and the target signal. We expect that negative images get the subject into the arousal state, while viewing a sequence of neutral images induces a less aroused state. The algorithm was attempting to construct a balanced sequence, where the subject's level of arousal was getting high, but was not supposed to get too high by overshooting the target level. Typically, less negative pictures were presented when the arousal was higher while if the arousal was decreasing, the subject was seeing more negative pictures.

As has been discussed in Chapter 2, adaptive stimulation in real-time comprises such objectives as optimization of the sequence for hypothesis testing [54], to obtain better neural responses [62], or for ROI localization [161]. In all these examples, input from the subject's data is used to decide on the next stimulus in the row in order to achieve the optimization goal. Among these, there are also prominent studies where such closed-loop intervention is used to interact with the subject's mental state - for attention training [56] or for improved BCI use [76]. As the subject's mental state results from distributed activity in connectivity networks, a promising line of experiments targets brain networks to exert control over - for instance, for deep brain stimulation [162, 176, 181]. Our study marks the first example of closed loop control over a subjects' affective state by optimizing stimulation with the goal of steering the subject towards that state.

5.2 Experiment

We carried out an experiment with two groups of participants, the real-time group and the control group. The experiment involved one session for each participant consisting of four experimental runs carried out on the same day with short breaks between the runs. Each run included a number of trials where one picture was shown to the participant on each trial, either negative or neutral. The first run was constructed in the same way for the real-time and control groups. The experimental sequence of stimulus pictures was following a fixed protocol, defined before the experiment start. In runs 2-4, one group was viewing sequences constructed in real time, while the sequences in the second group were defined offline before the beginning of the session in the same manner as for the first run. The subjects were only instructed to attend to the pictures and were unaware of the difference in the protocols. Each experimental run started with two minutes of baseline recording, during which the subjects were just viewing the fixation cross. The mean and the standard deviation of the baseline were used to set the target level for each participant. This method was chosen because GSR baseline largely varies within participants, and it

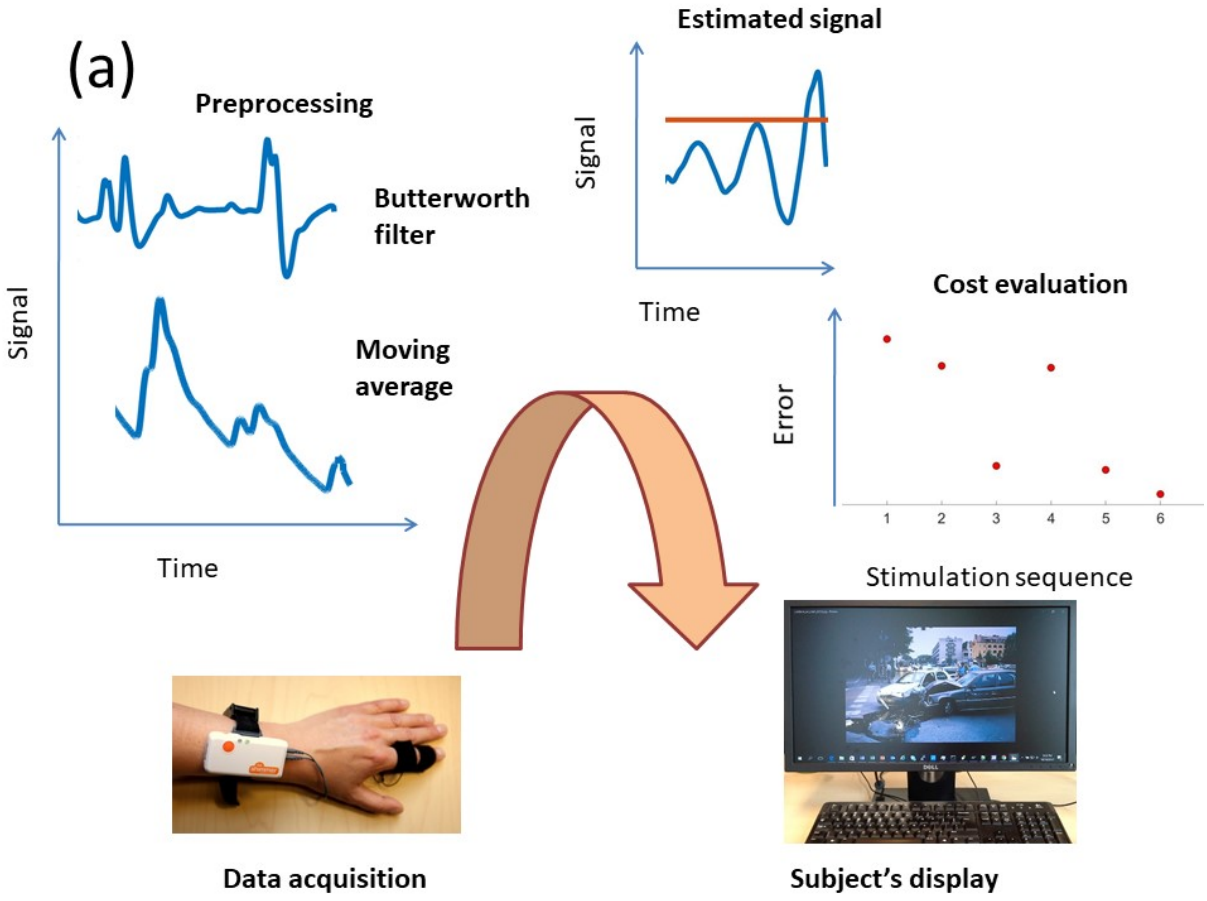
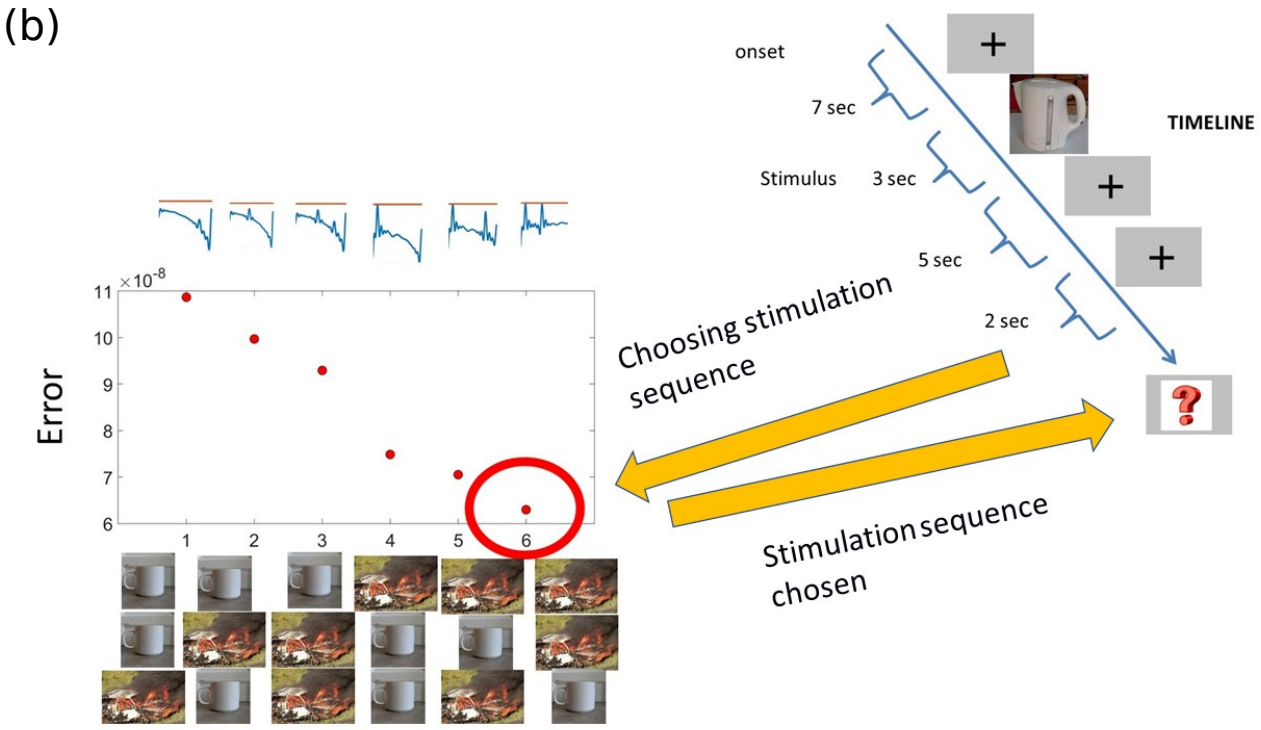


Figure 5.1: (a) Real-time pipeline. GSR sensors were placed on the subject's left hand. The resistance signal in Ohms was converted to conductance (Siemens) and preprocessed with Moving Average filter and then with Butterworth filter for online analysis. Based on the cost calculated for the upcoming trials, the sequence was chosen that produced the least error between the predicted and the target signal. Overall structure of the experimental real-time run is shown in Fig. 5.2. The image on the display comes from Wikimedia Commons `https://commons.wikimedia.org/wiki/File:Ljubljana_car_crash_2013.jpg` and is distributable under Creative Commons license. (b) Choice of stimulation sequence. All combinations of negative vs. neutral stimuli for the upcoming three or four trials were calculated. Only mixed combinations were considered, combinations of all negative or all neutral pictures were excluded. The predictive model allowed to estimate the signal for each of the possible sequences, and to compute the cost of each sequence. The protocol was updated during the fixation period, 2 seconds before the next stimulus onset. Signal estimations show that bringing arousal to the target level is not only the matter of the number of negative pictures, but mainly of the picture sequence. The image of the burning car was taken from freely available resources on the Internet `https://pixabay.com/it/auto-combustione-relitto-fuoco-872265/` and is distributable under Creative Commons license.



is impossible to choose a unified value as target. At the end of each run, the overall error - Sum Of Squared Errors (SSE) - was calculated between the target signal and the observed signal of the whole run. Then, for each participant the error values obtained in runs 2-4 were compared against the error value of the first run. We hypothesized that the application of the control strategy would reduce significantly the error in runs 2-4 as compared to the initial run in the real-time group. Conversely, we expected that no such reduction would be observed in the control group that viewed randomized image sequences unrelated to the state of the subject.

5.2.1 Participants

$N = 44$ healthy participants took part in the experiments, half of them (22 subjects) were assigned to the real-time group and the other half to the control group. They signed an informed consent form for participation and received either monetary reimbursement or ECT credits. One participant's dataset was discarded because of the problems during real-time recording. 2 subjects' data were removed because of too noisy signal. Subjects who demonstrated absolute error values more than 10 times higher than the absolute value of the first run in at least one of the runs had their data removed because these

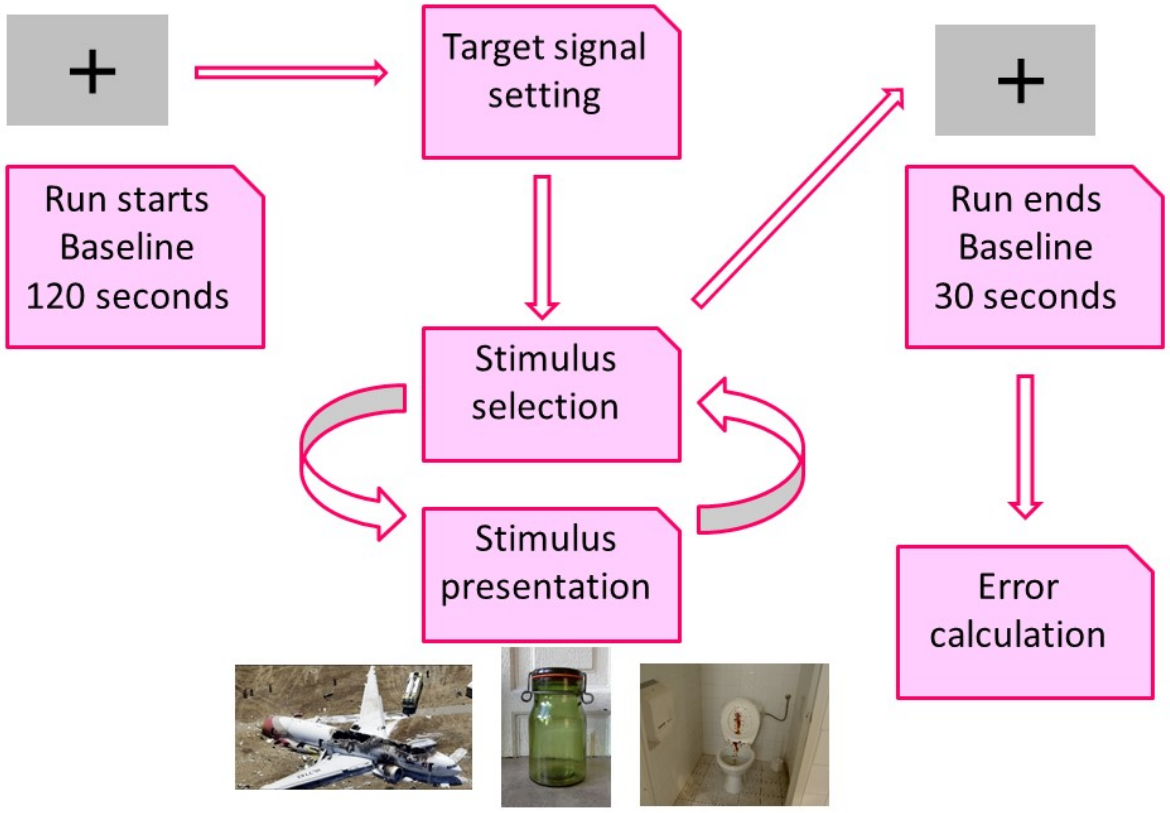


Figure 5.2: Structure of the experimental real-time run. The run starts with the baseline recording for 2 minutes. At the end of the baseline period, the target signal is computed as 2.8 standard deviations above the mean of the baseline signal. Then comes a sequence of stimulation protocol updates: the first sequence after the baseline is composed of 4 trials, and the subsequent 6 updates create sequences of 3 trials each. At the end of the run, there is another baseline recording period of 30 seconds. Finally, the sum of squared errors is calculated for the signal comprising the first 20 trials of the session. The leftmost and the rightmost images in the sequence were obtained from Wikimedia Commons `https://commons.wikimedia.org/wiki/File:Asiana_Airlines_Plane_Crash.png` and `https://commons.wikimedia.org/wiki/File:Pr%C5%AFjem.JPG` respectively. Both are distributable under Creative Commons license.

error values were largely due to artifacts resulting from deep breathing (6 subjects), but also cough artifacts (one subject), excessive motion (one subject) and artifacts from the loud noise in the environment (one subject) [30]. The groups were matched at the group level for main demographic parameters (age and gender) and handedness. Each group contains 17 subjects, of which 6 are males in each group. Each group has one left-handed participant. The mean age in the real-time group was 28 with range between 19 and 50, in the control group 27 with range between 20 and 40. All participants have lived in Europe for at least 3 years.

5.2.2 Stimuli

The stimuli were negative and neutral images from the International Affective Picture System (IAPS) set [150]. The images were chosen in the following way. First, a sample of images was drawn from the IAPS where negative arousing images were images with ratings for pleasure less that 2.0 (below the mean) and for arousal more than 6.5 (above the mean). The neutral non-arousing images were images that received pleasure ratings between 5.5 and 6.5 (around the mean) and arousal ratings below 3.25, i.e., below the mean. This sample of images was next tested in a pilot study, where the participants were also asked to provide feedback about the impact of the images, both neutral and negative. After the pilot, the final sample of images was manually corrected - the negative images that were evaluated by the pilot subjects as "cinematographic" and "fake" were replaced with images that the pilot participants judged more arousing. Large part of images containing people were removed from the neutral sample and replaced with images of objects as participants admitted that images of people were more arousing than images of objects.

In the initial run of both real-time and control groups the experimental sequence was following a fixed protocol, defined before the experiment start, where the negative image in each trial was chosen with the probability of $p = 0.6$. In the control group, all other runs were constructed in the same way, that is, in each trial the negative image appears with a certain probability. In runs 2-4 the probability was controlled using three levels of probability $p \in \{0.3, 0.45, 0.6\}$, so that the mean number of the negative pictures in a run for all participants was not too different from the mean number of negative images in the same run in the real-time group. After each real-time acquisition, the mean number of negative pictures per session was calculated in the experimental group. In the next control acquisition, the probabilities of negative images were adjusted, i.e. a higher or a lower probability was chosen for the run.

5.2.3 Procedure

All experiments were carried out in the same room with air conditioning, with a temperature of 20 ± 1.5 degrees. Natural differences in humidity seemed not to affect the variance in the data since participants recorded on the same day show similar variabilities as subjects acquired on different days, see e.g., [30]. All experimental procedures had been approved by the Ethical Committee of the University of Trento and were carried out in accordance with applicable guidelines and regulations on safety and ethics.

The experimental session consists of 4 runs (a calibration run and three experimental runs) of about 6.5-7 minutes' duration each. A run starts with 2 minutes of baseline recording, and the first run consists of 25-28 trials, where one image is shown on each trial. In runs 2-4 the number of trials was between 21 and 24 in the control group and 22 in the real-time group. Furthermore another baseline recording of 30 seconds was acquired at the end of each session. Images did not repeat between runs. As for visual stimulus presentation, in GSR experiments its duration can vary from 1 second in [12] to 8 seconds in [4] depending on the necessary amount of cognitive processing. In our experiment, images were shown for 3 seconds to allow for some cognitive processing followed by 7 seconds of fixation cross viewing, leaving enough time for the response to unfold in accordance with general guidelines of skin conductance experiments [77]. The experimental setup is visualized in Fig. 5.1a and b. The full structure of an experimental real-time run is further illustrated in Fig. 5.2.

We note that responsiveness may vary between subjects - for instance, there always are hypo- and hyperresponders among subjects, independent of the stimulation effects [215]. To account for differences in responsiveness, the mean and the standard deviation of the baseline were used in the process to set the target level for each participant at 2.8 standard deviations above the baseline mean. Ideally, the target state in the control scenario should be defined in cognitive or behavioural terms. As GRS signal measures the person's arousal, we should strive to define the target state as a certain level of arousal. However, in this case we encounter problems related to the elusive nature of such definition in the absense of strict behavioural metrics and cognitive criteria. For this reason, we set the target state as a certain signal level in relation to the baseline (2.8 standard deviations above the baseline, as mentioned above). As we know that GSR signal is very nicely correlated with arousal level, we can say that this level of signal linearly corresponds quite well to a certain degree of arousal. However, to further characterise this state of arousal in cognitive and behavioural terms (especially at the group level), we need additional experiments - I will talk about that later in Chapter 7.

5.2.4 Data acquisition

Data was acquired using Shimmer (www.shimmersensing.com) wearable GSR sensors. The two sensors were placed on two phalanges of the same finger, left index finger, one under another. During acquisition, the sampling rate was set to 51.2 Hz. The sensor was sending the signal via Bluetooth to the computer that carried out data preprocessing and analysis. The resistance signal in Ohms was converted to conductance (Siemens) and preprocessed with Moving Average filter and then with Butterworth filter for online analysis. During preprocessing, the data was down-sampled to 10 Hz. Then the data processing computer was sending instructions to the stimulation computer about the next sequence to show to the participant. Images were presented with Psychtoolbox-3 and the ASF plugin for Psychtoolbox [230]. Data analysis was carried out in Matlab (8.5.0, MathWorks, NatickMA, USA) using Shimmer software package for Matlab (Shimmer Matlab Instrument Driver v.2.6), customized for the experiment purposes, and additional custom-made scripts.

5.3 Results

To explore the effect of the protocol adaptation in real time, we analyzed the relative values of SSE calculated for each subject in runs 2-4 with respect to the values of the SSE obtained for the first run. To reduce skewness in their distribution, absolute error values were log-transformed - see Fig. 5.3. We compute the relative error as the ratio of the error in the first run with respect to error in each of the subsequent sessions to reduce non-stationarity. A relative error value less than 1 in runs 2-4 demonstrates that the error is reduced. Finally, the mean relative error was calculated for each group - Fig. 5.4b. Our hypothesis predicts, that the error will be reduced in sessions 2 - 4 with respect to the first session, against no such change in the control group. To test this hypothesis, the error in runs 2-4 are compared with the first run in each experimental group.

We found that in the real-time group, the mean relative error between the target arousal level and the observed signal was significantly reduced already in the second run with respect to the first run ($P = 0.003$, Fig. 5.4b) and stays significantly below the initial error also through sessions 3 and 4 ($P = 0.005$ and 0.0145 for session 3 and 4 respectively, see Fig. 5.4b). The relative error in the control group stays at the same level throughout the whole experiment ($P = 0.8$, 0.78 and 0.98 for sessions 2, 3 and 4 respectively, Fig. 5.4b).

The real-time and the control groups shared common clutural background and were

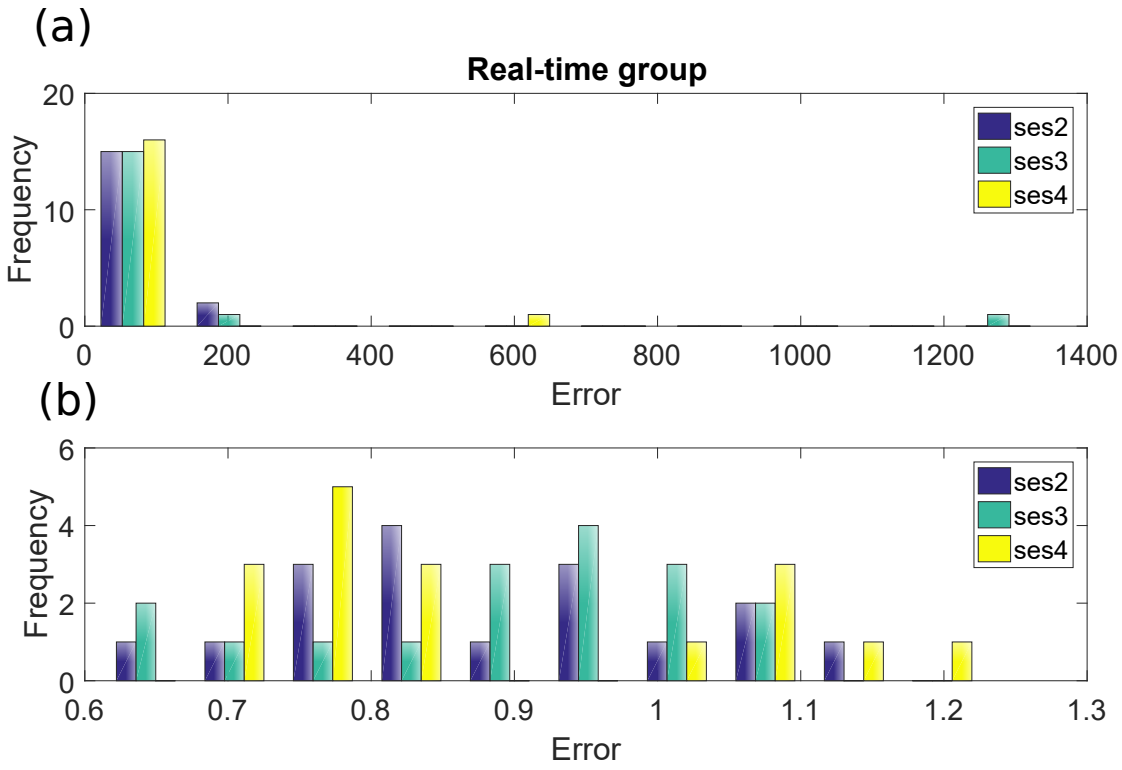
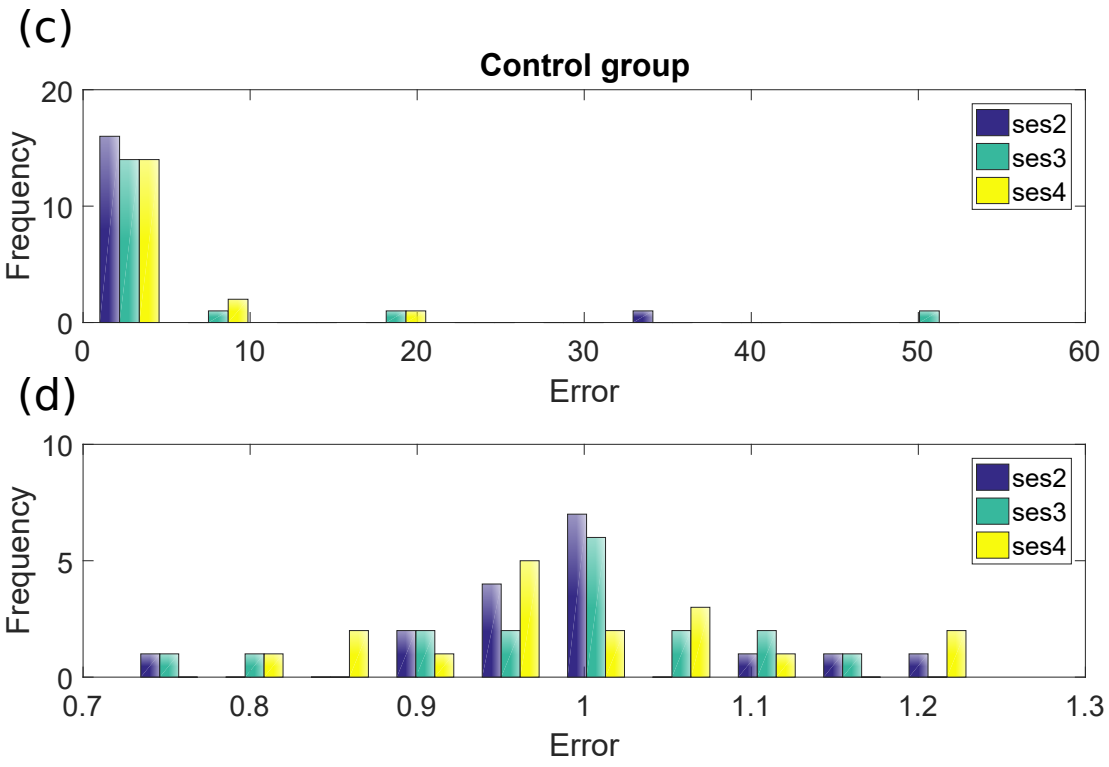


Figure 5.3: Relative error values before and after logarithmic transformation in the real-time group (a, b) and in the control group (c, d). Figures (a), (c) show the distribution of relative error values with respect to the first run calculated as the ratio run 1 / run 2,3,4. We can see that both distributions are skewed to the right. Figures (b), (d) show the relative error values after the logarithmic transformation. We took the natural logarithm of the absolute error values for each subject in each session and then calculated the relative error in the same way as it was calculated for the values before transformation. The distributions result more normal, though one can see differences between the group: in the control group, the mean is centered around 1 (d), while in the real-time group, the distribution is fatter with the median around 0.8 (b).


(c)
Control group
Frequency
Error
ses2
ses3
ses4
0
10
20
30
40
50
60
(d)
Frequency
Error
ses2
ses3
ses4
0.7
0.8
0.9
1
1.1
1.2
1.3

Factor	dF	F-stat	p-value
Within-subject (Time)	3	3.06	0.03
Between-subject	1	9.261	0.004
Interaction	3	2.614	0.0556

Table 5.1: Results of repeated-measures mixed effects ANOVA

not only matched by demographic parameters, such as age, gender and handedness: we also controlled for the mean number of negative pictures that a group would see in a session - Fig. 5.4a. As the first session was constructed in the same way for both groups of subjects, with negative pictures having a certain probability to be shown at each time point, that resulted in only slight differences in numbers between groups. There were no significant differences in the mean number of negative pictures shown in sessions 2-4 between groups (see also Fig. 5.5). In this way, the differences between groups can be attributed to the experimental manipulation and the protocol design.

If we look at the shape of the group error curve in Figure 5.4b, we can see that in neither of the groups there are significant differences between runs at the group level. For the real-time group that implies that the effect of adaptive stimulation is stable across runs. If there was an additional learning (or subject adaptation) effect, we would expect that the subject becomes "resistant" to the control, and the error would grow between the second and the last sessions. If there was a cumulative effect to expect, i.e. the more we control, the more "controllable" the subject becomes, then the error would go down from session 2 to session 4. At the group level we do not observe any of these. However, with a closer examination of the individual error curves in the real-time group (in the left part of Figure 5.4b), we can hypothesize, that with more data in the sample we should see at least three subpopulations in the group: those who show no effect of control, those who show the learning effect and those who show the cumulative effect. These subpopulations are described in more detail in Fig. 5.5b-c.

For between-group analysis, we conducted mixed effects (random-fixed) ANOVA with 1 within (time) and 1 between (protocol) subjects' factor for between group effect and interaction effect. We also ran a repeated measures ANOVA separately for each group to test for the effect of time.

Repeated measures ANOVA conducted separately in each group with 3 degrees of freedom shows that there is effect of time in the experimental group (F-stat 4.07, p-value 0.01), but no effect in the control group (F-stat 0.04, p-value 0.98). The results of mixed effects ANOVA are shown in Table 5.1.

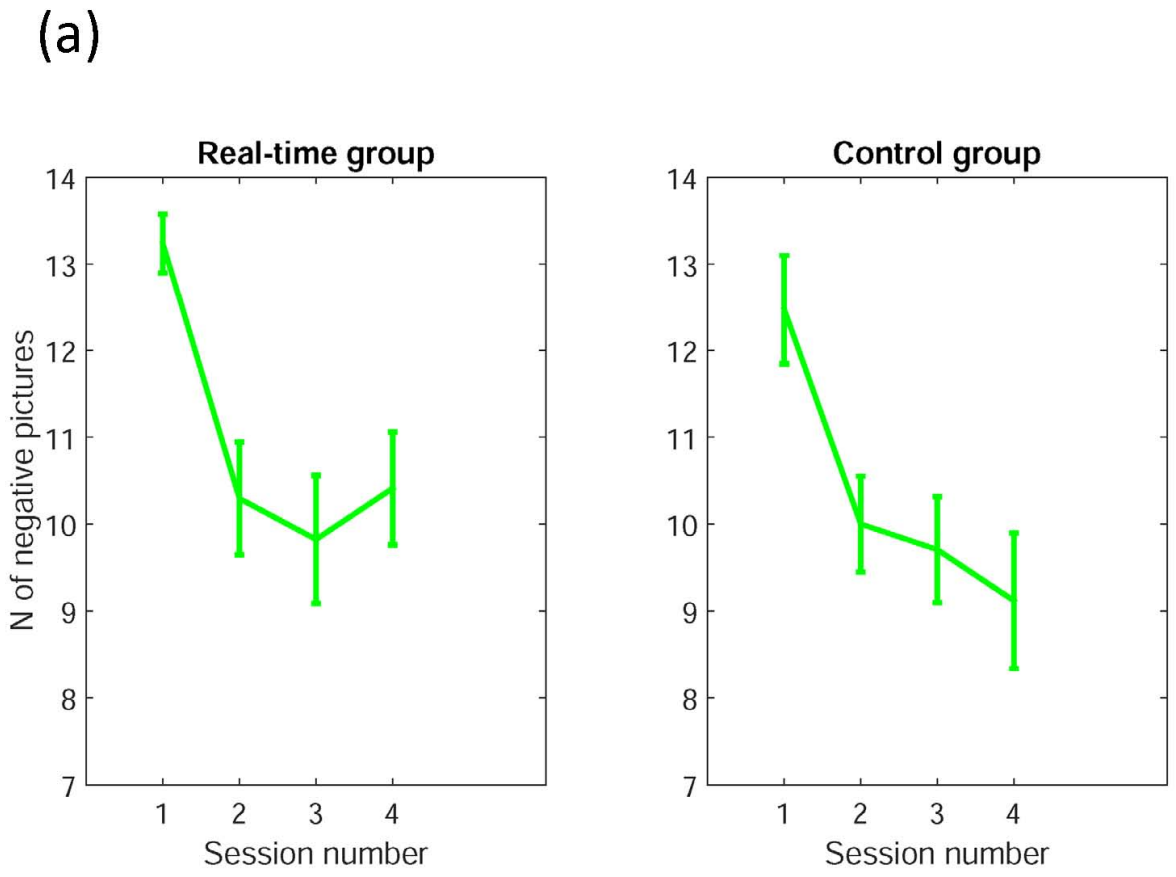
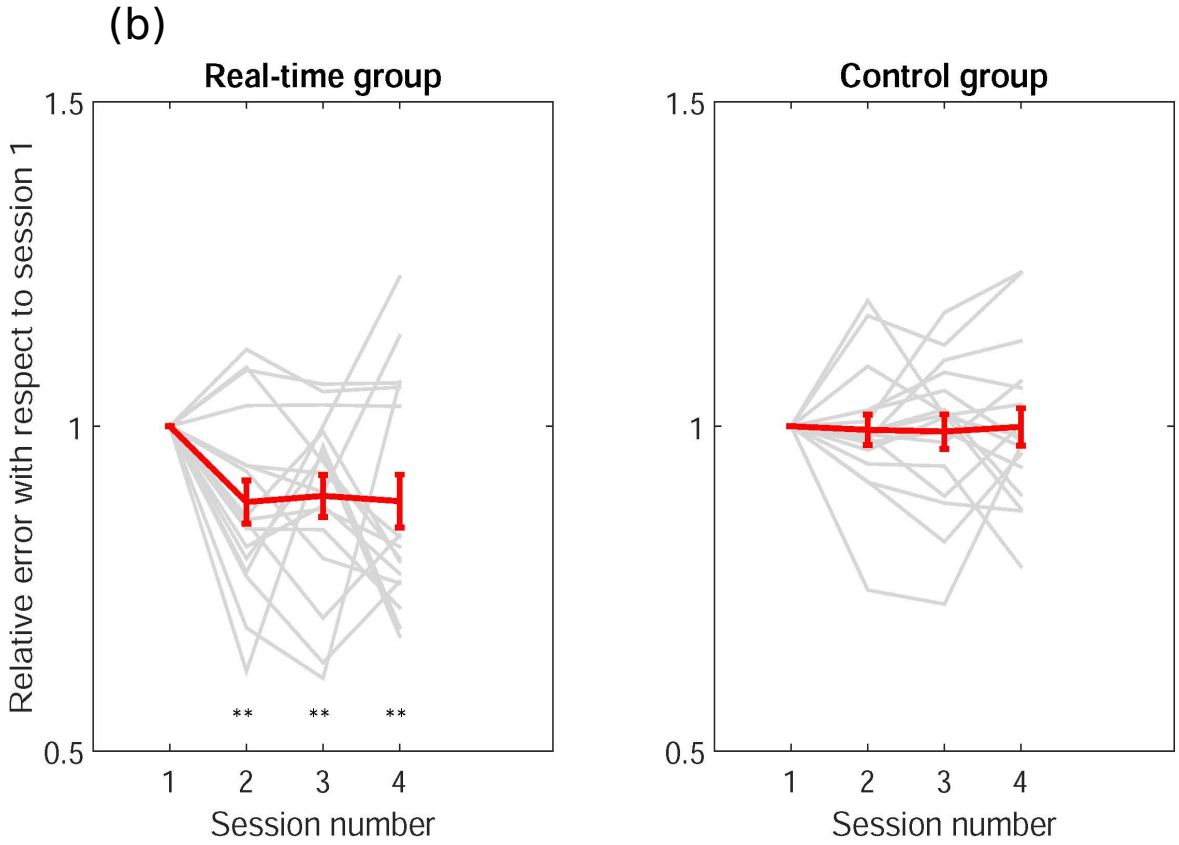


Figure 5.4: (a) Number of negative pictures per run in real-time and control groups. The mean number of negative pictures per run was balanced between groups. After each real-time acquisition, the mean number per run was calculated for both groups and the mean of the control group was adjusted accordingly in the next control acquisition. That is, the protocol of the coming control acquisition was constructed so, that the two means would not be significantly different. The proportion of negative pictures in that run would be decreased or increased to bring the means at the same level. Number of negative pictures for each subject per run are shown in Fig. 5.5a. (b) Relative error in runs 2-4 for real-time and control groups. Individual error values are rendered in gray, group means are presented in red. Asterisks show significative p-values. P-values in the control group: 0.824, 0.771 and 0.978 for runs 2,3, and 4 respectively. Repeated measures ANOVA with 3 degrees of freedom for testing the effect of time: F-stat 0.04, p-value 0.98. P-values in the real-time group: 0.003, 0.005 and 0.0145 for runs 2,3 and 4 respectively. Repeated measures ANOVA with 3 degrees of freedom for testing the effect of time: F-stat 4.07 p-value 0.01. Fig. 5.5b-c, apart from individual error values and group means, show also subject IDs which are the same as in Fig. 5.5a. Figures 5.5a-c illustrate how the number of negative pictures each subject saw per run can be related to their individual error dynamics.


(b)
Real-time group
Control group
Relative error with respect to session 1
1.5
1
0.5
1
2
3
4
Session number
**
**
**

So, we have p=0.004 for between group effect. It means that our groups are different in their responses. Next, we also see the effect of time. And for the interaction, the p-value is less than 0.06, which is on the verge of being significant. If it was significant, that would mean that temporal profiles in the groups were different, that is, we would have observed the effect of the session. However, as I explained previously, we don't really see a session effect in the experimental group (neither cumulative nor adaptation effect). Apart from the change between the first session and the rest, temporal profiles of the groups are not different. So, we would not expect a big effect of interaction here. But the between group effect is quite strong, as expected according to our hypothesis.

We also conducted a power analysis to know what the minimum effect is that we can test with sufficient (80%) power given the sample size, for this between-group comparison. As the basis for power analysis, we used Pocock's formula for sample size:

$$N = 2 * \sigma^2 \div (\mu_2 - \mu_1)^2 * f(\alpha, \beta)$$

where α is the desired significance level (0.05) and β is the desired power (0.8). $f(\alpha, \beta)$ is a table value, and for the given values of significance level and power it is equal to 7.9. Now let's denote effect size as x. If μ_1 and μ_2 are the two means, and σ is the standard deviation, effect size (Cohen's formula) is usually expressed as follows:

$$x = (\mu_2 - \mu_1) \div \sigma$$

Then the non-table part of the Pocock's formula can be expressed using effect size as follows:

$$N = 2 \div x^2 * f(\alpha, \beta)$$

Knowing that our sample $N = 34$ and plugging in the table value we can calculate the minimal effect that can be detected given the sample size, the power and the significance level:

$$34 = 2 \div x^2 * 7.9$$

$$x = \sqrt{}(15.8 \div 34) = \sqrt{}0.46 = 0.68$$

Then we used a calculator of effect size givent the F-value `https://www.psychometrica.de/effect_size.html#fvalue` to calculate effect size d for the between-group difference with 1 degree of freedom: $d = 1.076, d > 0.68$. So, we can say that our sample size gives us enough power to detect the between-group effect.

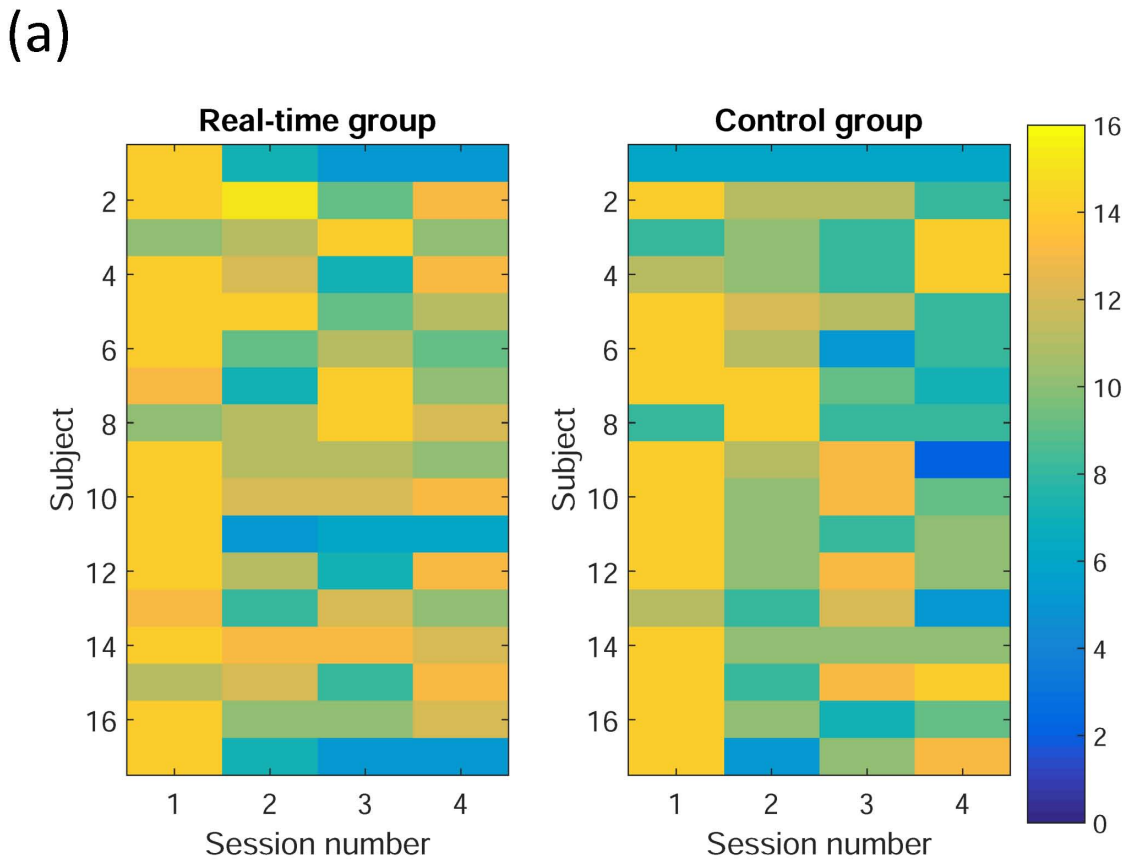
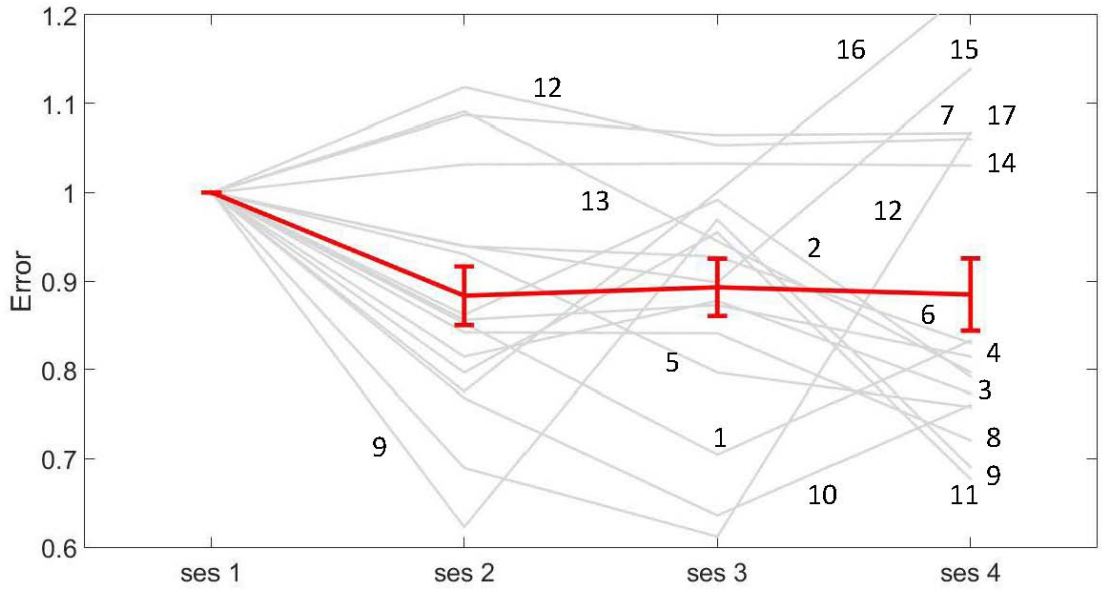
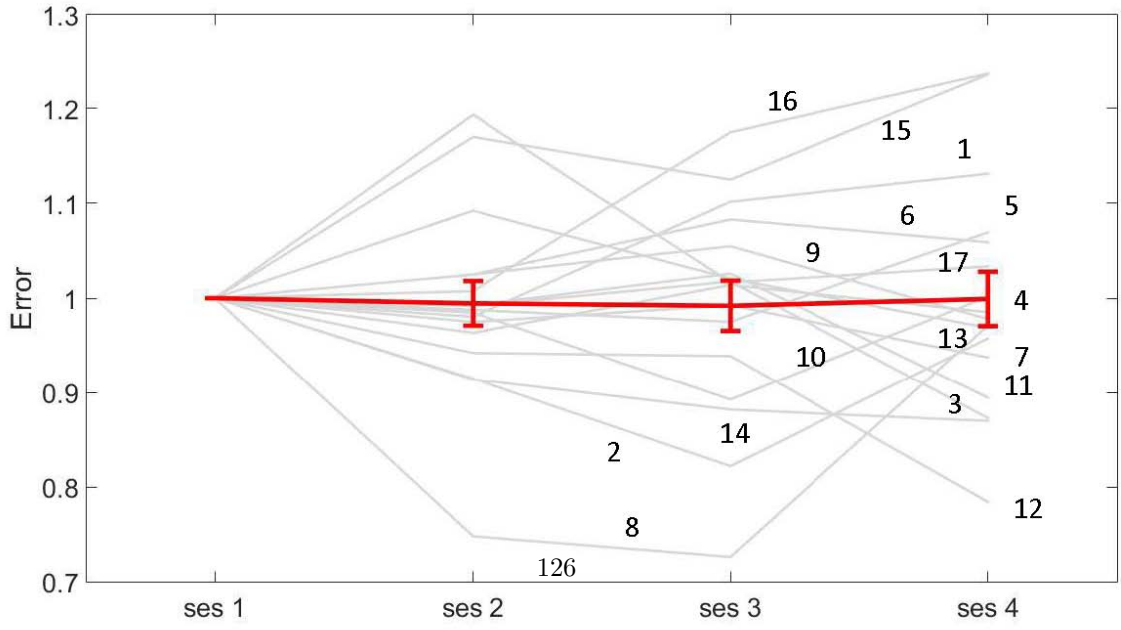


Figure 5.5: (a) The figure shows the number of negative pictures each subject was shown in each run. The number is color-coded. The subject IDs are the same throughout figures (a)-(c), so that the number of negative pictures can be related to the error dynamics of each subject. (b)-(c) In these figures, the error dynamics from run to run is shown for the real-time group (b) and for the control group (c). The subject ID is next to the error line plot showing the error for this subject. (b) Subjects 7, 12, and 14 potentially form a subpopulation "resistant to control": they show no effect throughout the sessions. Subjects 3,4,15,16, 17 could be regarded as the subgroup showing learning effects, as their error goes up in the last session. Among the rest of the subjects, we see a tendency for the cumulative effect: the error descends as the experiment progresses.

(b) Real-time group



(c) Control group



5.4 Methods

5.4.1 Model for online data analysis

Let's say we have a signal of length L, with the number of samples $\tau = 1, \dots L$. We also have a timeline of the control signal, applied with a certain periodicity: $t = 1, \dots T$. Assuming that the short term dynamics of the GSR signal $x(t)$ is linear autoregressive, with additive control $u(t) \in \{0, \theta\}$, we obtain

$$x(t) = \sum_{\tau=1}^{L} a_\tau x(t-\tau) + u(t) \tag{5.1}$$

where $t = 1, \dots T-1$. Let $\bar{x}$ be a vector of states of of length T: $\bar{x} = [x(0) \dots x(T)]^T$, where $x(0)$ is the initial state. If we rearrange 5.1 to express $u(t)$ which we are interested in, we get

$$x(t) - \sum_{\tau=1}^{L} a_\tau x(t-\tau) = u(t), \tag{5.2}$$

where $x(o) = const$, $x(t) = 0$ for $t < 0$. Equation 5.2 is equivalent to a linear matrix equation if we use polynomial notation:

$$\bar{\bar{G}}\bar{x} = \bar{u}, \tag{5.3}$$

with $\bar{u} = [x(0), u(1), \dots u(T)]$. From these equations we obtain a matrix G of the form

$$G = \begin{matrix} 1 & 0 & 0 & \dots & 0 \\ -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & -a_1 & 1 & \dots & 0 \\ -a_3 & -a_2 & -a_1 & 1 & 0 \\ 0 & \dots & -a_2 & -a_1 & 1 \end{matrix}$$

To predict the signal we must first estimate $\bar{\bar{G}}$ in a calibration window. We observe the linear system under random control and additive Gaussian noise:

$$\bar{\bar{G}}\bar{x} = \bar{u} + \bar{\epsilon}, \tag{5.4}$$

where $\epsilon(t) \sim N(0, \sigma^2)$. Here $\bar{u}$ is known, but $\bar{\bar{G}}$ is unknown and can be estimated using the least squares:

$$\hat{\bar{\bar{G}}} = \underset{\bar{\bar{G}}}{\operatorname{argmin}} \|\bar{\bar{G}}\bar{x} - \bar{u}\|^2. \tag{5.5}$$

Returning to the initial equation 5.1, we can express G as follows:

$$G_{tt't} = \delta_{tt't} - \sum_{\tau-1}^{L} a_\tau \delta_{t-t',\tau} \tag{5.6}$$

Basically, it is a polynomial corresponding to the autoregressive process. If we want to learn the parameters and take partial derivatve for a, we get:

$$\frac{\delta}{\delta a_\tau} \sum_{t=0}^{T-1} (\sum_{t'=0)}^{T-1} G_{tt'} x_{t'} - u_t)^2 = 2 \sum_{t=0}^{T-1} (\sum_{t'=0}^{T-1} G_{tt'} x_{t'} - u_t)^2 \sum_{t''=0}^{\tau-1} \frac{\delta G_{tt''}}{\delta a_\tau x_{t''}}, \tag{5.7}$$

where $\frac{\delta G_{tt''}}{\delta a_\tau} = -\delta_{t-t'',\tau}$ and $t - t'' = \tau, t'' = t - \tau$. Now we plug in the obtained expression to get 5.8:

$$= -2 \sum_{t=0}^{T-1} (\sum_{t'=0}^{T-1} G_{tt'} x_{t'} - u_t) \sum_{t''=0}^{T-1} x_{t''} \delta_{t-t'',\tau} = -2 \sum_{t=0}^{T-1} (\sum_{t'=0}^{T-1} G_{tt'} x_{t'} - u_t) x_{t-\tau} \tag{5.8}$$

As we know that the matrix G is a Toeplitz matrix, we can apply Levinson-Durbin recursion [64, 156] to solve for $a_{\tau'}$ as it is identical to the first column of G shifted by one element and with the opposite sign (linear system for $a_{\tau'}$):

$$\sum_{t=0}^{T-1} (\sum_{t'=0}^{T-1} G_{tt'} x_{t'} x_{t-\tau}) = \sum_{t=0}^{T-1} u_t x_{t-\tau} \tag{5.9}$$

$$\Rightarrow \sum_{t=0}^{T-1} \sum_{t'=0}^{T-1} (\delta_{tt'} - \sum_{\tau'=1}^{L} a_{\tau'} \delta_{t-t',\tau'}) x_{t'} x_{t-\tau} = r_\tau \tag{5.10}$$

Vector r holds the expression for what happens to the system under control.

$$r_\tau = \sum_{t=0}^{T-1} u_t x_{t-\tau} \tag{5.11}$$

Vector q holds just the temporal dynamics of the system.

$$q_\tau = \sum_{t=0}^{T-1} x_t x_{t-\tau} \tag{5.12}$$

$$q_\tau - r_\tau = \sum_{t=0}^{T-1} \sum_{t'=0}^{T-1} \sum_{\tau'=1}^{L} a_{\tau'} \delta_{t-t',\tau'} x_{t'} x_{t-\tau} \tag{5.13}$$

$$q_\tau - r_\tau = \sum_{\tau'=1}^{L} a_{\tau'} \sum_{t=0}^{T-1} x_{t-\tau'} x_{t-\tau} \tag{5.14}$$

$$\sum_{t=0}^{T-1} x_{t-\tau'} x_{t-\tau} = \tilde{q}_{\tau-\tau'} \tag{5.15}$$

So we have another linear system of order L' (= number of timesteps, corresponding to τ'),

$$\bar{q} - \bar{r} = \bar{\bar{Q}}\bar{a} \tag{5.16}$$

Here, the signal x is known, and we can get both the vector q and the matrix Q. The control signal u is known, too, so we can also get the vector r. Then we just can solve the matrix equation for parameters a and obtain the matrix G.

For online analysis we can predict the control for the upcoming trials after having learnt the system matrix. We know the estimation matrix $\bar{\bar{G}}$ until a certain timepoint $t = 0$. At $t = 0$, $x(t) = x_0$.

$$\bar{\bar{G}}\bar{x} = \bar{u} + \bar{\epsilon}, \tag{5.17}$$

where $t = 0, \ldots, H$. Loss (or cost) is given by the distance to a preferred set of the values that x could take. We will denote that set as y: $\bar{y} = [x_0, y_1, \ldots y_H]^T$. The expression for loss is $\|\bar{x} - \bar{y}\|^2$:

$$\mathbb{E}_\epsilon(Loss) = \mathbb{E}_\epsilon \|\bar{x} - \bar{y}\|^2 \tag{5.18}$$

with $\bar{\bar{G}}\bar{x} = \bar{u} + \bar{\epsilon}$. Turning to $\bar{u}$, $\bar{\bar{G}}\bar{x} = \bar{u} + \bar{\epsilon}$, so the optimal $\bar{u}$ can be expressed as:

$$\bar{u}^* = \underset{\bar{u}}{\operatorname{argmin}} \mathbb{E}_\epsilon \|\bar{x} - \bar{y}\|^2 \tag{5.19}$$

with $\bar{\bar{G}}\bar{x} = \bar{u} + \bar{\epsilon}$. From these we derive the optimization formula:

$$\bar{u}^* = \underset{\bar{u}}{\operatorname{argmin}} \|\bar{\bar{G}}^{-1}\bar{u} - \bar{y}\|^2 \tag{5.20}$$

If we optimize over discrete $u(t) = (0, \theta)$ (known θ), then for short horizons $H \leq 5$ direct, exact optimization is possible. For longer planning horizons we may resort to approximate inference, like, for instance importance sampling or mean field annealing [27, 220].

5.4.2 Simulated experiments

So, in this experiment we generate "ground truth" signal starting with some known control and known estimator matrix. Then we try to learn the estimator matrix from the "ground truth". It was done in the following way: a) we fixed the experiment parameters as number of timepoints = 30 corresponding to 3 trials of length 10 seconds. This is going to be the order L of the linear system; b) we fixed the control sequence as 0 1 1; c) we generated the coefficients as a decaying sinewave as in an autoregressive process - figure 5.6.

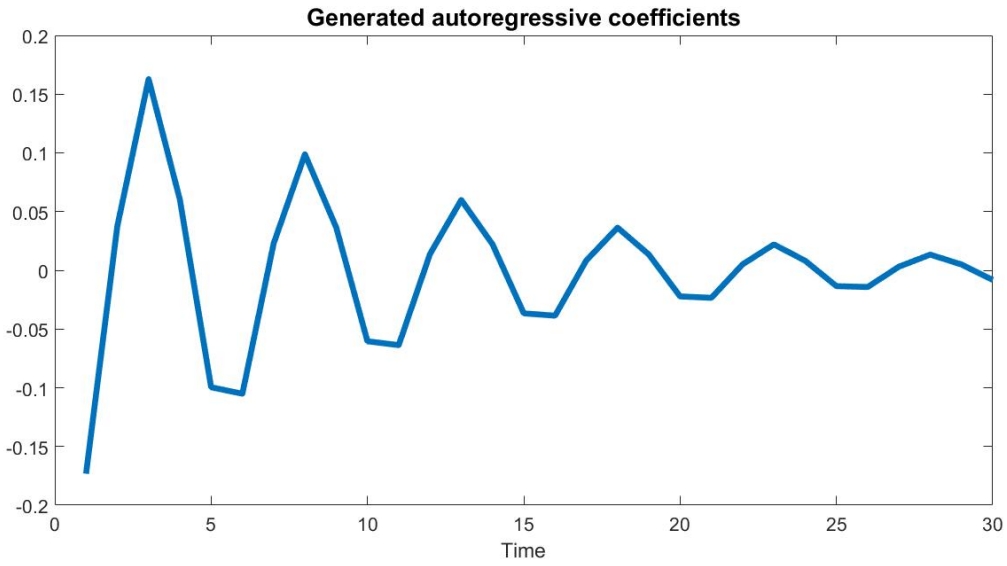


Figure 5.6: Decaying sine wave representing autoregressive coefficients used to generate the signal. The scale of coefficient values is represented by the values on the Y-axis, while the X-axis consists of consequent time points.

Next, we interpolated the signal to 150 timepoints simulating the sampling rate of 5 Hz (that is typically used for downsampling the experimental GSR data). The ground truth signal obtained in this way is shown in figure 5.7.

Finally, the estimation matrix was learned using the system identification equation of order L' (= number of timesteps) - Eq. 5.16:

$$\bar{q} - \bar{r} = \bar{\bar{Q}}\bar{a}$$

Then, new signal was generated with the learned coefficients and the same stimulation sequence. The signal is shown in figure 5.8. We can see that the model allows us to obtain a pretty good estimation of the ground truth signal. Root-mean-square error (RMSE) calculated as the standard deviation of residuals (i.e the difference between the predicted signal and the "ground truth") is equal to $4.6e-18$, that is, a very small value.

Please note that the amplitude of the response in the signal is equal to one. It is because we represent the control sequence as zeros and ones. In the eventual expetimental design 1 was replaced with the actual signal amplitude parameter θ (we can represent control as 1 multiplied by a constant θ equal to the amplitude).

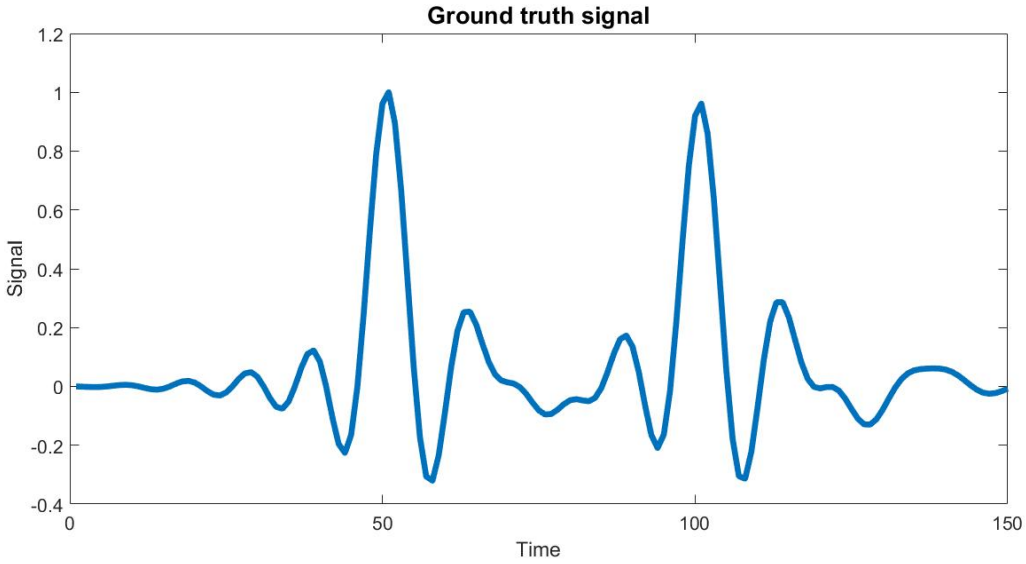


Figure 5.7: "Ground truth" signal generated with the coefficients shown above and the 0 1 1 control sequence

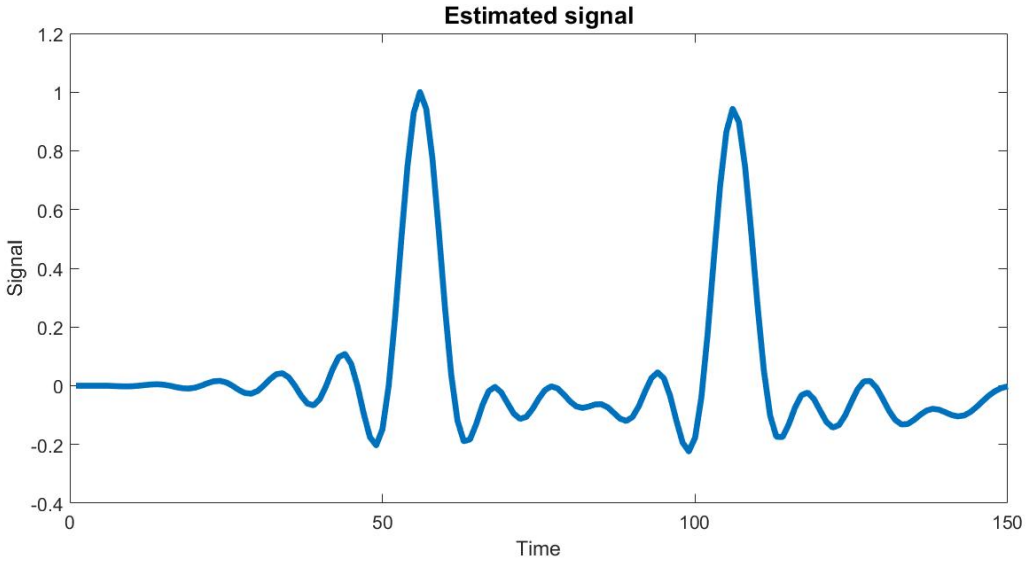


Figure 5.8: Signal generated with the estimator matrix learned. RMSE $= 4.6e-18$

5.4.3 Optimizing control with short horizon

In this experiment, we used the data from the pilot offline GSR experiment with IAPS pictures where the random stimulation sequence was fixed before the session. One session data was used as training data for learning the estimation matrix. The data of another session was used as test data: we predicted what optimal control could be like for this session given the signal, and then it was compared to the actual control of the session. The estimator matrix was learned with the data counting 10 trials backwards, and the control was predicted 5 trials forward. Next, predicted signal was generated using the design matrix of the predicted control for the whole session and response parameters obtained with the canonical response model described in [12]. Two examples of comparison between observed signal and estimated signal generated using the learned control sequence are shown in Fig. 5.10. We can see that the learned control corresponds to the one used to produce actual signal. Fig. 5.9 shows cumulative error for all possible binary sequences encoding controls, and we can see that there is a sequence with minimal error (it corresponds to the one used to produce predicted response). Fig. 5.11 plots the whole error matrix, where the timepoints contained in 5 trials (12 s each, timestep = 500 ms) are plotted on the x axis, and the error - on the y axis. Different lines correspond to 1 binary sequence each, there are 32 lines in total (2^5 because only the first position in the trial is independent and can be 0 or 1, control or no control, the rest is just padded with 0).

5.4.4 Optimizing control with long horizon

When the horizon comprizes the whole session (20 trials), we cannot just test all possible binary sequences to pick up one with the minimal error, so importance sampling was used to produce predicted controls. Again, one session data of the offline pilot experiment was used as training data for learning the estimation matrix and the data of another session was used as test data. First, a sample of 1000 possible binary sequences was produced, and the predicted signal was generated using the estimation matrix learned with the data of the calibration session. Next, the error between observed and predicted signal was calculated for each of positions in the sequence. Next, cumulated error in each position served as target function to produce the weight of each element in the position (in our case, 0 and 1). Then these weights were used to approximate the distribution while sampling, and a new binary sequence was generated based on the weights of 0 and 1 in each position. This sequence was considered predicted control. We also produced estimated signal for this sequence using parameter estimates and the canonical response

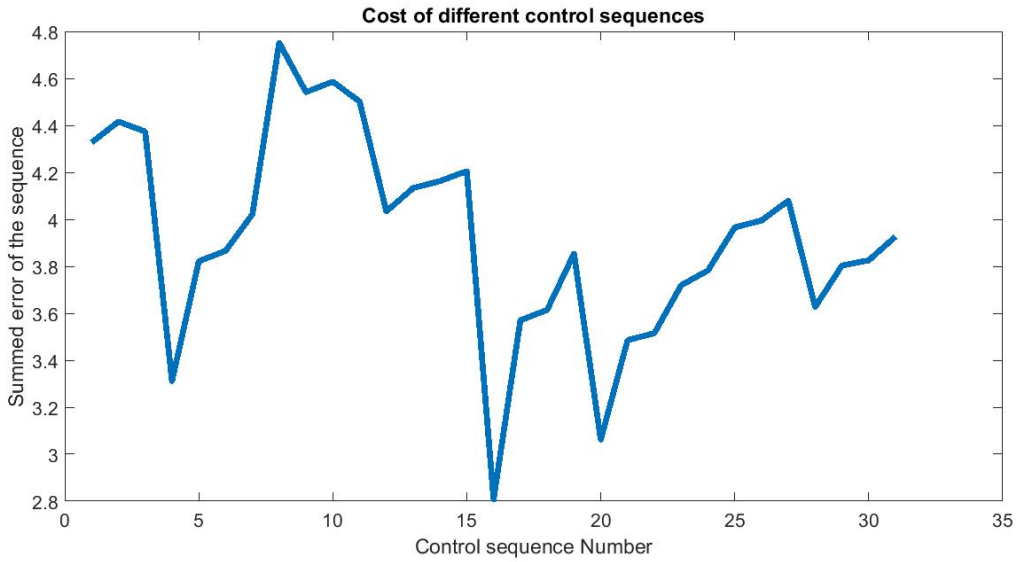


Figure 5.9: Cumulative error for each possible binary sequence representing controls.

model. The results are shown in Fig. 5.12. To compare the signal generated with the learned control with the ground truth, we looked into the temporal correlation between the two signals. We calculated correlation function between the observed response and predicted response in a session, which is presented in Fig. 5.13. We also checked for time lag of the point where the maximum correlation occured, the result is presented in Fig. 5.14. Correlation function and the lag plot show that there is a very good temporal correlation between the two signal types.

So, simulated experiments described in this session were performed to see how good are the estimates of the signal and control sequences produced by the model, and the results show that the estimation is quite good. However, it should be kept in mind that the canonical response function model from [12] that we used to genetrate the estimated signal for illustrative purposes is not a predictive model. It is a statistical model that was introduced for estimating responses from the data, and it is arguable whether it can be also used for the purposes of predictive modelling. We just used it in these experiments to represent schematically the signal that could be produced by the learned control sequences for mere illustrative purposes.

5.4.5 Model for amplitude parameters

Parameter θ is generally referred to as the response amplitude [30, 55]. As we treat responses to neutral images as defining the reference condition, θ quantifies the relative

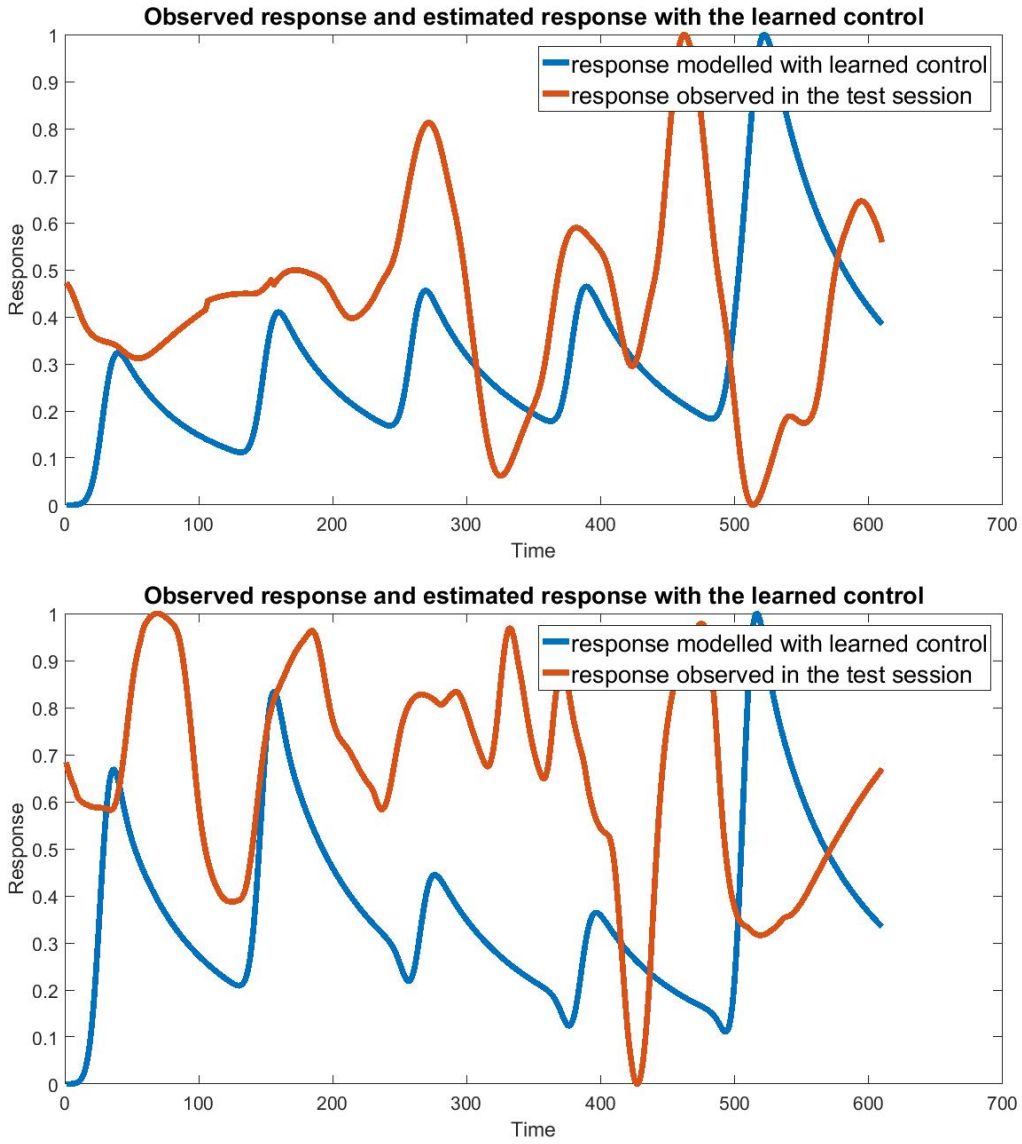


Figure 5.10: Predicted response with learned control and observed response. High peaks in the estimated signal correspond to the negative images, where high response is expected, while smaller ones - to neutral images.

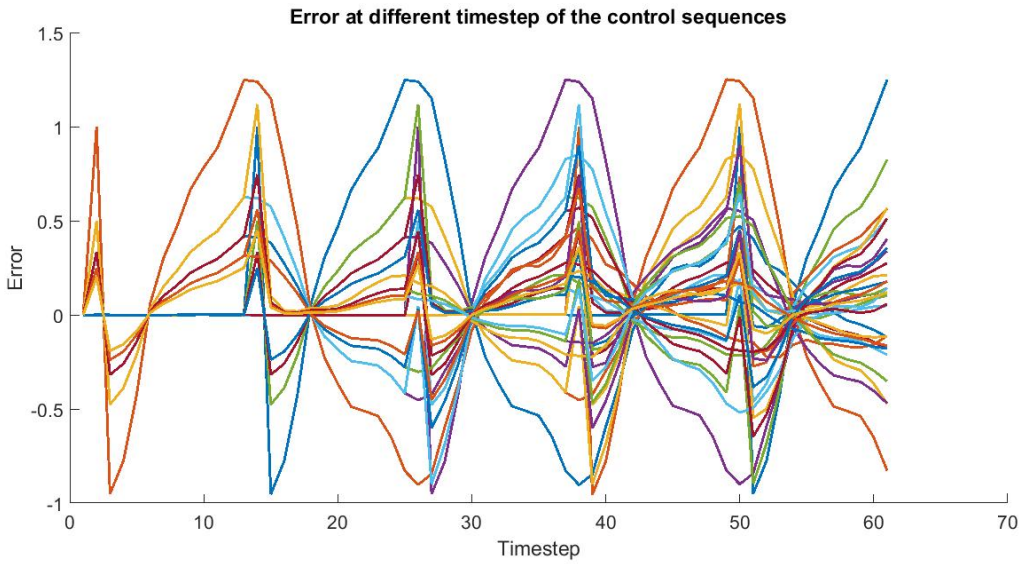


Figure 5.11: Error matrix for all binary sequences.

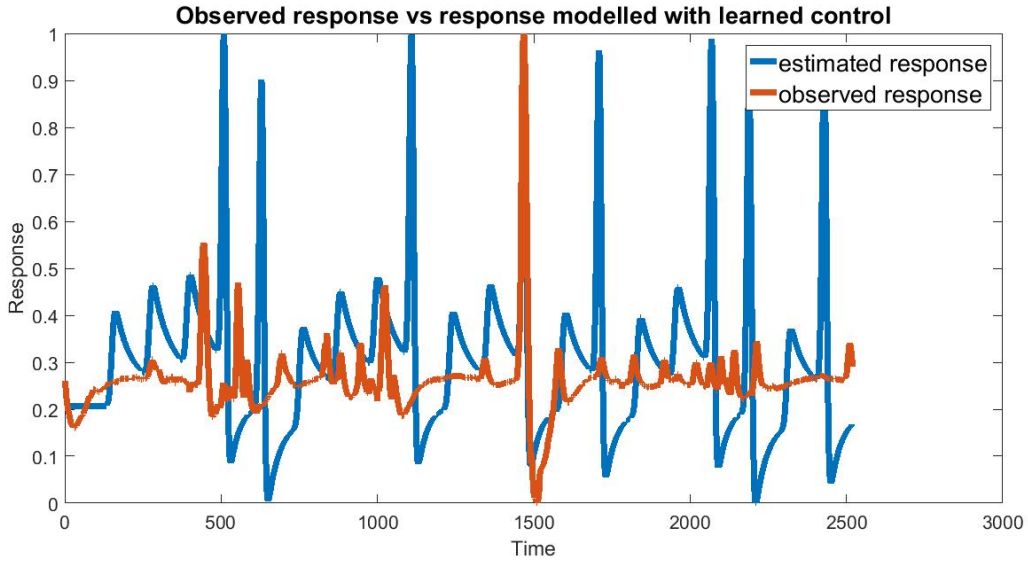


Figure 5.12: Predicted response with learned control and observed response: whole session. High peaks in the estimated signal correspond to the negative images, where high response is expected, while smaller ones - to neutral images.

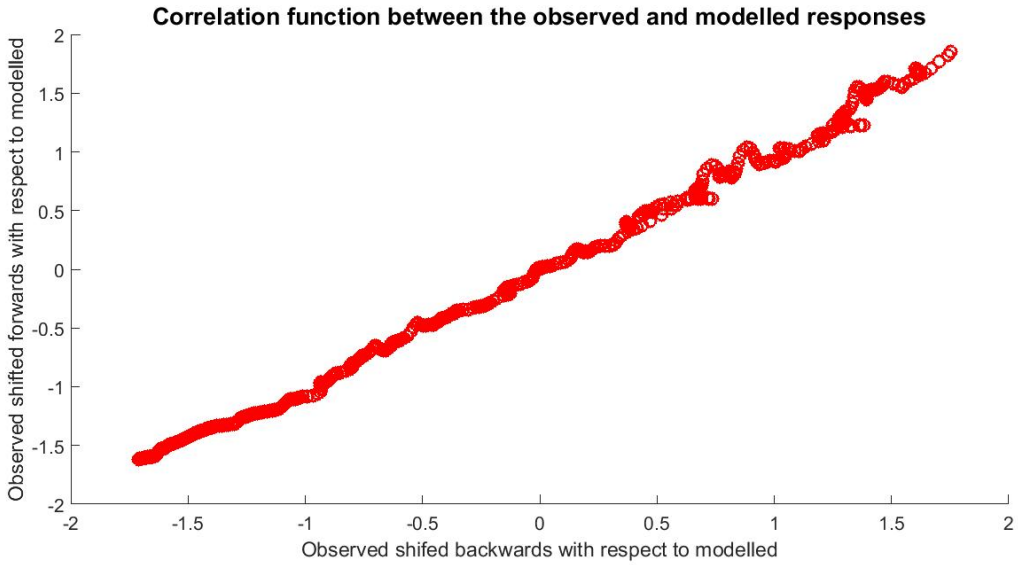


Figure 5.13: Correlation function between observed (x) and predicted (y) response

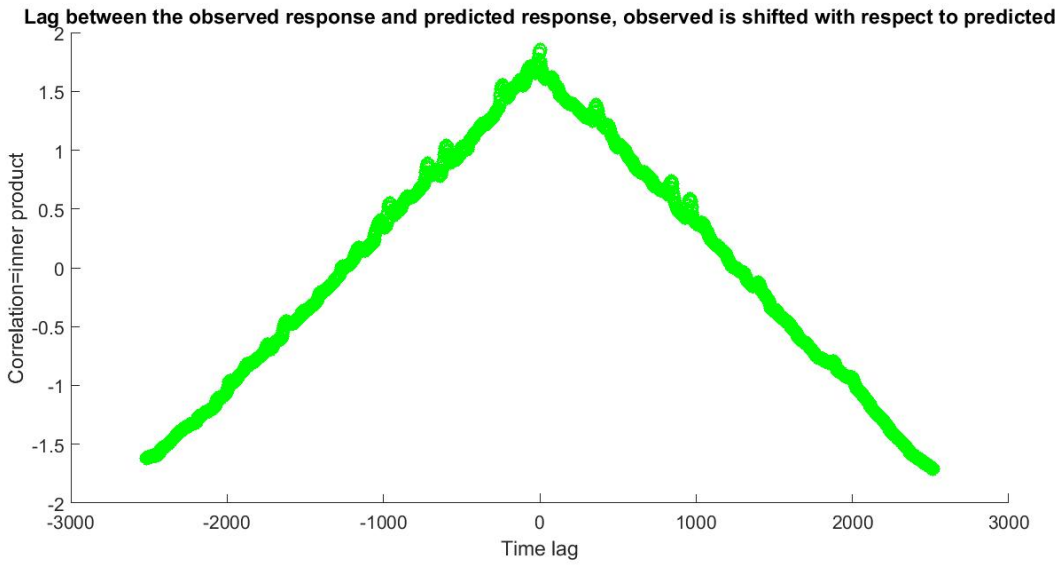


Figure 5.14: Time lag of the correlation function (in samples, sr=10Hz)

strength of a negative image over a neutral. The typical GSR response looks like a hill-shaped curve delayed with respect to the stimulus onset of 1-3 seconds [30, 55]. The amplitude is calculated as the difference between the minimum and the maximum of the signal taken in the response window (which is usually the time interval between 1 sec and 5-6 seconds after the stimulus onset) [55]. When updating the stimulation in real-time, we use the mean amplitude of response to negative images from the 3 preceding trials. However, for the first stimulation sequence of four trials after the baseline no data on amplitude are available. To obtain the parameter in the first four trials of a real-time session, we apply the framework of [14], so that we can proceed with error estimation and protocol selection. The authors propose a canonical response function for the skin conductance response, which was found to represent experimental data quite well. Model based GLM analysis for the GSR signal is implemented in the PsPM toolbox for Matlab (version 3.1, formerly SCRalyze) [11, 15, 16]. Running GLM on the training data from the previous sessions allows us to obtain parameters for each type of response (negative arousing picture versus neutral non-arousing picture) and reconstruct predicted responses. We used only one parameter, that of the canonical response function. As in the design matrix the stimulus was represented by zeros and ones, we can say that the obtained parameters represent the response amplitude. However, as already mentioned above, we only employ the negative response parameter to represent the amplitude of the response to negative pictures. Figure 5.15 shows an example of such reconstructed responses from the data.

As was already mentioned in the previous section, canonical response function model is a statistical model - it estimates parameters from data. To provide reliable estimates, this model, as any statistical model, needs enough data samples. This is the reason we only use this approach to estimate the amplitude of responses to negative pictures in the first four trials of a run: all trials of the previous session are available. This model cannot be employed in the same manner to predict response amplitudes for the upcoming trials in a session because at each time point, only a limited number of trials are available, which will not permit the model to provide good estimates. This is the reason we only use a rough estimate - mean amplitude of 3-4 previous trials. We also did not want to use the same estimate from the previous run throughout the real-time run because we assumed that response parameters could change between runs.

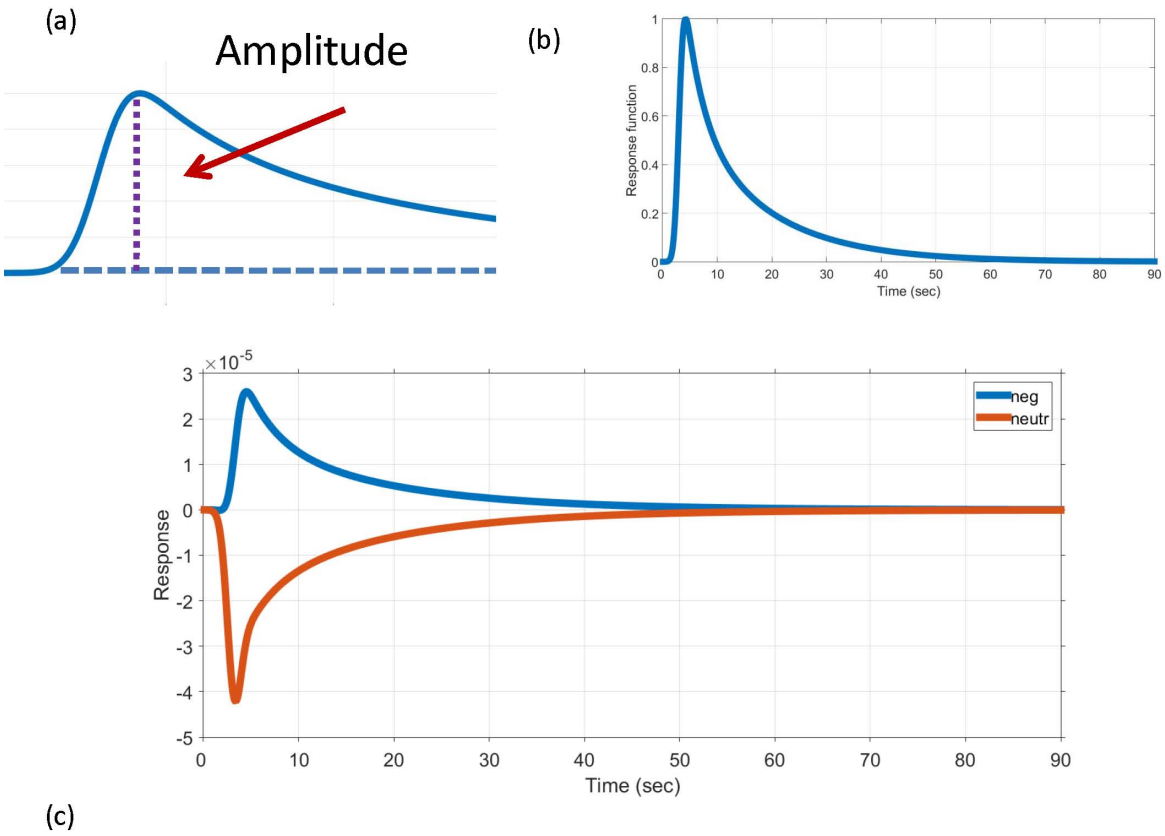


Figure 5.15: (a) GSR response and its amplitude. (b) Canonical response function for the GSR signal rendered by PsPM toolbox. (c) Responses - negative and neutral - estimated from the data with PsPM toolbox.

5.4.6 Session SSE computation

At the end of each run, the overall session error - Sum Of Squared Errors (SSE) - was calculated between the target signal and the observed signal of the experimental block:

$$\mathbb{E}_{\epsilon}(Loss) = \sum_{T}(\bar{y} - \bar{x})^2. \tag{5.21}$$

The real-time stimulation sequence had a fixed length of 22 trials: the initial update for the first 4 trials that was followed by 6 updates for 3 trials each (we chose to update the protocol for sequences of trials rather than for single trials). On the contrary, in the control group the number of trials was varying between 21 and 24 because of the probabilistic protocol creation: when generating a trial, the script was choosing the negative picture with a given probability, and it stopped generating trials when a certain number of negative pictures was reached within the run. To account for that difference in the number of trials between the groups, the performance measure (i.e., the sum of squared errors) calculation was restricted to the first 20 trials only in each session.

5.4.7 Statistics

Before running statistical tests on the error values, the distribution of values was tested for normality. As the distributions appeared strongly skewed, the log transformation was applied to make the data conform to the assumption of normality. Log-transformed values, indeed, yielded normal distributions of relative error values both in the experimental and the control group (Fig. 5.3). Significance testing of the relative error values was performed with one-sample one-tail t-test against 1 (the relative error in run 1 with respect to run 1) with 16 degrees of freedom (DoFs) as it was run separately for each group. Repeated measures ANOVA had 3 DoFs for within-subject effect (time) testing, 1 DoF for between-group effect testing and 3 DoFs for the interaction effect testing. Repeated-measures ANOVA was chosen because within each group our measurements are not independent and come from the same individuals. Both t-tests and ANOVA were performed with Matlab.

5.5 Data and code availability

A Matlab demo version showing how real-time protocol updating works is available from a public Github repository `https://github.com/elena-kalinina/GSR_demo_NHB` with instructions on how to run it. Full source code is contained in a private Github repository

https://github.com/elena-kalinina/GSR_RT_experiment and is available from the author upon a reasonable request. Raw data, preprocessed data and error files can be downloaded from Google drive usink the links https://drive.google.com/open?id=0BziIVITY-jdoRjdLcjFvdnhIODA and

https://drive.google.com/open?id=0BziIVITY-jdoeVU2YjB2MkFvQ0E.

Chapter 6

A test for cross-modal activation analysis as an alternative to decoding

Determining the extent to which different cognitive modalities (understood here as the set of cognitive processes underlying the elaboration of a stimulus by the brain) rely on overlapping neural representations is a fundamental issue in cognitive neuroscience. In the last decade, the identification of shared activity patterns has been mostly framed as a supervised learning problem. For instance, a classifier is trained to discriminate categories (e.g. faces vs. houses) in modality I (e.g. perception) and tested on the same categories in modality II (e.g. imagery). This type of analysis is often referred to as cross-modal decoding. We take a different approach and instead formulate the problem of assessing shared patterns across modalities within the framework of statistical hypothesis testing. We propose both an appropriate test statistic and a scheme based on permutation testing to compute the significance of this test while making only minimal distributional assumption. We denote this test cross-modal permutation test - Cmpt. We also provide empirical evidence on synthetic datasets that our approach has greater statistical power than the cross-modal decoding method while maintaining low Type I errors (rejecting a true null hypothesis). We compare both approaches on an fMRI dataset with three different cognitive modalities (perception, imagery, visual search). Finally, we show how Cmpt can be combined with Searchlight analysis to explore spatial distribution of shared activity patterns.

6.1 Neuroscientific background

Functional MRI recordings enable the investigation of activation patterns that characterize the working brain. The main goal is to detect whether the neural pattern of a region of interest correlates with a cognitive task, like, for example, object category identification. Such investigations are usually focused on a specific cognitive modality, i.e.: *visual perception*, *real auditory*, *visual imagery*, *auditory imagery*, etc. A qualitative discrimination task can be designed to extract relevant information from patterns activated by two (or more) stimuli categories (like *body* and *car*) in one of these cognitive modalities.

Identification of activation patterns that are shared across modalities has been the subject of numerous neurocognitive studies, with such modalities as mental calculations [135], sensory/motor stimulation [70] or words and picture-viewing [233], to name a few. The most common approach here is to cast the problem of identifying common activation patterns as a supervised learning, or brain decoding, problem [107,127]. Nastase and colleagues [187] point out that successful classification in this setting allows to conclude that neural patterns elicited by relevant cognitive factors in one modality generalize accross to the patterns in the other modality.

In other words, a low misclassification error on a modality which is different from the one used to train the classifier provides empirical evidence that a given region of interest is involved in the cognitive task encoded in both modalities. In the literature, such approach is referred to as cross-modal decoding analysis (Cmda). Statistical significance of its result can be assessed using a t-test on the accuracy obtained by the classifier [135,170], or by a permutation test based on computing the null distribution [70,124,259] of such statistic. However, Cmda suffers from a number of practical issues. As we are going to discuss in section 6.2, the accuracy of a decoding model is often low in a cross-modal setting, which most probably implies an exaggerated amount of Type II errors (failure to reject the null hypothesis).

Neuroscientific investigations into the activity patterns in the fMRI data studies can be formulated as "confirmatory" or "exploratory" analysis. Confirmatory analysis is centered on a pre-established region of interest (ROI). The Cmda method presented above is often used for the confirmatory approach, when it is run on the data coming from a predefined ROI (or a set of predefined ROIs). Exploratory analysis aims at localization of areas containing information about the presented stimuli. Here, Cmda is employed in conjunction with Searchlight technique to explore the spatial structure of cross-modal activations [187]. The outcome of the Searchlight procedure are maps, where each voxel is assigned some quantitative measure of information that the voxel contains about the

stimulus. The procedure of obtaining the maps consists in applying decoding sphere by sphere on time series extracted from the sphere voxels, and most commonly used information measure to produce Searchlight maps is classification accuracy. Maps are first calculated individually for each subject and then pooled together to run group analysis, where the significance of the obtained values is typically established running t-tests voxelwise with respect to chance level. Classification accuracy and t-tests have been subject to numerous criticisms with regard to their role in Searchlight analysis [71]. Besides, use of Searchlight for cross-modal analysis faces interpretation challenges introduced by the asymmetries both in accuracies and p-values when training and testing on data coming from two different cognitive modalities [187].

We developed a permutation test for the investigation of shared patterns across modalities that we denote *cross-modal permutation test* (Cmpt). This test builds on a long tradition of randomization inference in the statistics literature, which can be traced back to the first half of the 20th century [79, 154, 207]. Permutation tests have recently seen renewed interest in neuroimaging [66, 189, 272, 273] thanks to their minimal distributional assumptions and the availability of cheap computational resources. We provide empirical evidence on synthetic datasets that this method reduces Type II errors (failure to reject a false null hypothesis) while maintaining Type I errors comparable (incorrect rejection of the null hypothesis) with respect to Cmda. Our results highlight particular advantages of Cmpt in the small sample/high dimensional regime, a setting of practical importance in neuroimaging studies. Next, we compare Cmpt and Cmda on an fMRI study of three cognitive modalities: visual attention, imagery and perception. We conduct confirmatory analysis comparing the performance of Cmpt and Cmda when identifying the presence of shared patterns within a functionally defined Region of Interest (ROI). Finally, we present the results of an exploratory analysis with Searchlight making use of the proposed Cmpt test for the information based mapping to explore the presence of common patterns between different modalities at the whole brain level. The use of Cmpt allows to overcome major methodological drawbacks that had been pointed out for Searchlight in the literature [71].

6.2 Methods: Cross-modal permutation test (Cmpt)

In this section we describe a statistical test for cross-modal activation pattern analysis that we denote Cmpt. We formulate the problem of assessing cross-modal activation as a hypothesis testing problem and propose an inference procedure for this test based on a

permutation schema.

Setting. We assume that the experimental task consists of two modalities (e.g., auditory and visual perception) and each image in the dataset containing an activation pattern has an associated condition A or B (e.g., two stimuli categories like human body and car). In total, we observe n activation patterns for each modality, corresponding to the number of conditions in the experiment, where each activation image is a mean image (averaged by the number of trials in the experiment) representative of one condition, and the goal is to decide whether there is a common condition effect across the different modalities. Let us formalize this in the language of statistical hypothesis testing. Consider the set of pairs $Z(X,Y) = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ sampled iid from some unknown probability distribution P, where $X = \{X_1,\ldots,X_n\}$ (resp. $Y = \{Y_1,\ldots,Y_n\}$) are the activation patterns corresponding to the first (resp. second) modality, and where the experimental paradigm is designed such that the X_i and Y_i are associated with the same condition (but different modality). Since the image pairs belong to the same condition, as long as there is a condition effect shared across modality, the sequences X and Y cannot be independent. We hence formulate the null hypothesis (which we want to reject) that both sequences are independent and so their joint probability distribution P factorizes over their marginal:

$$H_0 : P = P_X \times P_Y \ , \quad \text{with } P_X,\ P_Y \text{ the marginal distribution of } P. \tag{6.1}$$

Test statistic. Given the set of image pairs $X = \{X_1,\ldots,X_n\}$ and $Y = \{Y_1,\ldots,Y_n\}$ described in the previous paragraph, let $\mathcal{A}$ (resp. $\mathcal{B}$) be the set of indices for the A (resp. B) category. We define $X_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} X_a$ (resp. $X_{\mathcal{B}} = \frac{1}{|\mathcal{B}|}\sum_{b\in\mathcal{B}} X_b$) as the average of activation patterns in X with index in $\mathcal{A}$ (resp. $\mathcal{B}$). $Y_{\mathcal{A}}$ (resp. $Y_{\mathcal{B}}$) are defined in similar way as the average of activation patterns in Y with index in $\mathcal{A}$ (resp. $\mathcal{B}$). Note that the index set is computed from images of X (and not Y) on both cases. This asymmetry will be useful when designing the permutation scheme. Consider also that we have access to a similarity measure between images that we denote by ρ. For simplicity we will initially suppose that this measure is the Pearson correlation coefficient, although we will see later that this can be generalized to any similarity measure between images.

We now have all necessary ingredients to present the test statistic that we propose to distinguish the null hypothesis from the alternative. This test statistic has values in $[-1,1]$ and is defined as

$$T(X,Y) = \frac{1}{4}\Big(\underbrace{\rho(X_{\mathcal{A}},Y_{\mathcal{A}}) + \rho(X_{\mathcal{B}},Y_{\mathcal{B}})}_{\text{within-condition similarity}} - \underbrace{(\rho(X_{\mathcal{A}},Y_{\mathcal{B}}) + \rho(X_{\mathcal{B}},Y_{\mathcal{A}}))}_{\text{between-condition similarity}} \Big) \quad . \tag{6.2}$$

At first, its form might seem strange. Let us give two intuitions on the form of this test statistic:

1. *As a difference of similarities.* The test statistic can be split as a difference of two terms. The first term is the sum of similarities for images from the same condition (and different modalities), while the second term is a sum of similarities for images of different conditions (and different modalities). Hence, large values of the test statistic are achieved whenever the within-condition similarity is larger than the between-condition similarity, bringing evidence for the existence of a condition-specific activation across modalities.

2. *As a singularity test.* If we compute all pairwise similarities between the images $X_{\mathcal{A}}$, $X_{\mathcal{B}}$, $Y_{\mathcal{A}}$ and $Y_{\mathcal{B}}$, we obtain 4 scalars that can be arranged in a 2-by-2 matrix as follows:

$$\begin{bmatrix} \rho(X_{\mathcal{A}}, Y_{\mathcal{A}}) & \rho(X_{\mathcal{A}}, Y_{\mathcal{B}}) \\ \rho(X_{\mathcal{B}}, Y_{\mathcal{A}}) & \rho(X_{\mathcal{B}}, Y_{\mathcal{B}}) \end{bmatrix}. \tag{6.3}$$

   Under the null hypothesis, the samples X and Y are independent and so $Y_{\mathcal{A}} \approx Y_{\mathcal{B}}$ (recall that the indexing was derived from images in X). Whenever $Y_{\mathcal{A}} = Y_{\mathcal{B}}$ the matrix above becomes colinear. A standard way to test for colinearity is through its determinant. Computing the determinant of the above equation we obtain our test statistic (modulo the normalizing factor $\frac{1}{4}$).

Statistical inference. We will estimate the distribution of this test statistic under the null hypothesis from the sample by repeatedly computing the test statistic over a permuted version of the initial sample, a technique often known as *permutation or randomization test*. For this to be valid, it is necessary to identify the quantities that we wish to permute and verify that under the null hypothesis, all permutations yield the same sample distribution [155].

Consider the sequence X_π which results from a random reordering of the activation images in X and the sequence of pairs $Z(X_\pi, Y)$. Under the null hypothesis, since the probability distribution factorizes over its marginal, the permuted sequence is distributed as $P'_X \times P_Y$, where P'_X is the distribution of X_π. Now, by the iid assumption made previously (which is commonplace in the context of permutation testing), this distribution is invariant to permutations, and so $P'_X = P_X$ and the condition is verified. After computing the permuted test statistic for a large number of random permutations (typically around 10000), the significance of this test, i.e., the probability of observing a test statistic equal

or as large as the one obtained, can be computed as

$$p = \frac{\text{number of times } \{T(X_\pi, Y) \geq T(X, Y)\}}{\text{number of permutations}} \quad . \tag{6.4}$$

Extensions. This test extends naturally to the setting of group analysis. In this case, the test statistic (6.2) can be taken as the sum over all subjects of the subject-specific test statistic. Ideally, the same permutation should be used across subjects to obtain each value of the permuted test statistic [69]. It is theoretically possible to perform a two-tailed test using this test statistic. A large negative value of the test statistic would also bring evidence to reject the null hypothesis of independence. However, since the neuroscientific interpretation of such negative values is not useful for our practical purpose, we will only use the one tailed test in this paper.

For simplicity, we have considered ρ the Pearson correlation coefficient as similarity measure, but the method remains valid using any other similarity measures. The Pearson correlation, being a measure of the linear correlation, works best when the effect is (close to) linear, but other more complex similarities can be used such as a (negative) Malahanobis [263] or Wasserstein [98] distance.

Relationship with cross-modal decoding analysis (CMDA). CMDA can be regarded within the same hypothesis testing framework outlined before, but with a different test statistic. In CMDA, the test statistic is the accuracy of a classifier on images from one modality when it was trained on images from the other modality.

Since both CMDA and CMPT follow the same permutation test approach to computing significance, both rely implicitly on a label exchangeability assumption behind the data-generating process. As we have seen in the previous subsection, a sufficient condition for this is to assume that the data we observe is sampled iid. Note that this iid assumption is on the pairs from different modalities (X_i, Y_i) and also on the experimental paradigm but not on the decoding train/test split, which divides the data by modality and is obviously not iid. This is a much weaker assumption than the distributional assumptions made by traditional parametric methods, it is important to keep in mind that permutation tests are *not* fully assumption-free methods and at the bare minimum require exchangeability of the observations.

A practical difference between both approaches is that the cross-modal permutation test is symmetric with respect to modalities while brain decoding is not. That is, CMPT would yield the same p-value regardless of the order in which the different modalities are labeled. This is not true for CMDA, where two possible tests can be performed (train on

A and test on B or train on B and test on A), and both can (and typically do) yield different p-values.

6.3 Datasets

First, we used real data of the neuroscientific experiment described in 3.1.1. Next, for simulated experiments we constructed a synthetic dataset according to a model in which the signal is a superposition of a modality-specific effect (M_X, X_Y), a condition-specific effect (C_i) and a Gaussian noise (ε_i):

$$X_i = \alpha \cdot C_i + \beta \cdot M_X + \varepsilon_i, Y_i = \alpha \cdot C_i + \beta \cdot M_Y + \varepsilon_i$$

where α, β are scalars that regulate the amount of modality-specific and condition-specific signal in the image, respectively. We then generated a total of 20 different images according to this model, considering two different modality-specific signals and two different condition-specific signals, all of them randomly generated from a Gaussian distribution. We generate 3 versions of this dataset, one with 10 voxels, one with 100 and another one with 1000 voxels.

6.4 Results

For the rest of the paper we will refer to the rejection of the null hypothesis with the traditional significance of 0.05 without explicitly mentioning this number.

6.4.1 Experiments on synthetic data

In Figure 6.1 we plot the resulting p-value after performing both CMPT and CMDA on the synthetic dataset described in section 6.2, for varying magnitudes of the condition-specific effect (α) and different image sizes. In the case of decoding, this p-value was computed as described in [192]. For $\alpha = 0$, the dataset has no condition-specific signal and so the test is not expected to produce a statistically significant result. Indeed, the p-value of CMPT is around 0.5. Note that because of the discreteness of the test statistic (test set accuracy), the average p-value need not converge towards 0.5 as α goes to zero in the case for CMDA. As the magnitude of the effect (α) increases, the method that yields a lower p-value has greater statistical power, because it is able to reject the null hypothesis with a greater probability. We can see in the figure, that in general CMDA p-values are

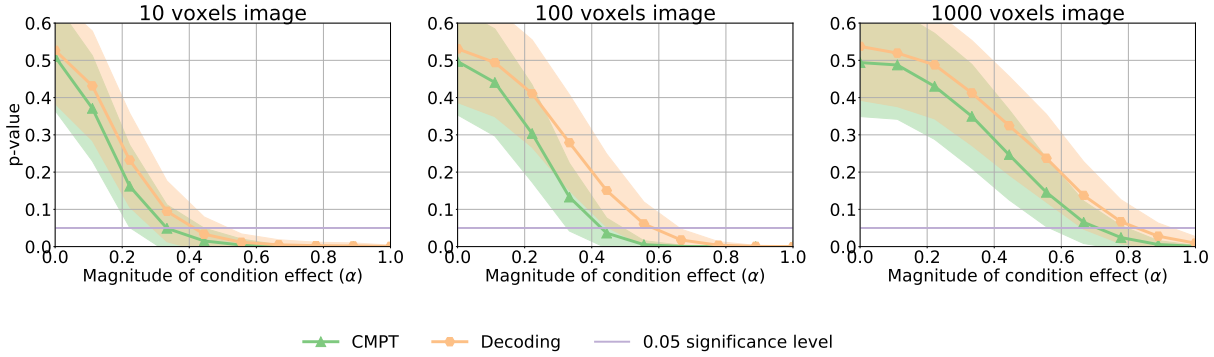


Figure 6.1: Statistical power as a function of the effect size. We compare the obtained p-value (y) for different magnitudes of the condition-specific effect (x) and different image sizes (10, 100 and 1000 voxels). Shaded areas correspond to the standard deviation. Across these different scenarios, CMPT yields a higher significance than the decoding-based approach.

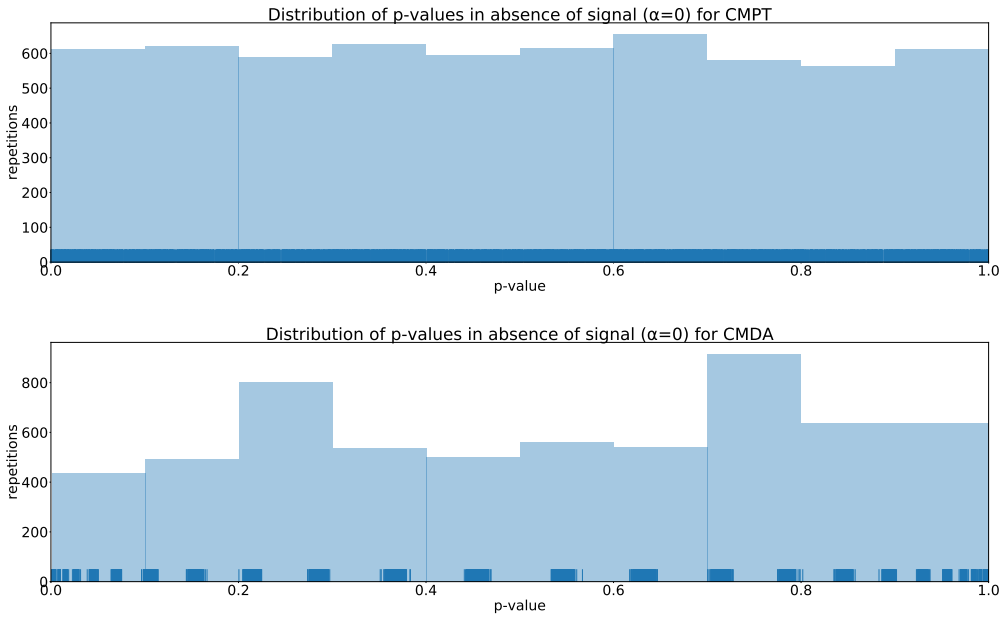


Figure 6.2: The distribution of p-values generated with CMPT (top) and CMDA (bottom) in the absense of condition-specific effect after 6000 repetitions of the experiment. p-values are on the x axis, while the frequencies are shown on the y axis. In case of CMPT (top) the distribution converges to the uniform. The false positive rate (Type I error) for a significance level of β is at the expected value of β. For CMDA (bottom), the distribution is not fully flat due to the discreteness of the scoring rule.

higher, which translates into a lower probability of rejecting the null hypothesis under this approach and hence higher Type II error.

In Figure 6.2 (top row) we can see that in the absence of signal ($\alpha = 0$), the distribution of p-values generated with CMPT(for 6000 repetitions) is relatively flat, showing that the false positive rate (Type I error) for a significance level of β is at the expected value of β. In the bottom row of that figure, we can see the same experiment for CMDA. In this case because of the discreteness of the test statistic, the distribution is not completely flat.

From the simulation results (Figure 6.1) we see that the average p-values yielded by CMPT are always below those of CMDA. This implies that smaller effects can be detected, and hence, that CMPT has a higher sensitivity than CMDA. Furthermore, this effect is replicated across images with different number of voxels, highlighting the benefits of CMPT in the high-dimensional setting, which is of great practical importance in neuroimaging.

6.4.2 Comparison of CMDA and CMPT on fMRI data

In this section we assess the agreement or disagreement in detecting shared activation patterns between CMDA and CMPT. The similarity of activation maps for the the discrimination of *body* vs. *car* categories was computed for the following pairs of modalities: perception and imagery, imagery and visual search. The visual search modality was investigated more in detail by first considering all durations of preparation periods put together and then analysing separately different delays (2, 4, 6, 8, and 10s). This analysis was meant to emphasize the issue of small sample size typical for neuroscientific data. The comparison between CMDA and CMPT took into account additional elements such as the choice of ROI and the type of encoding of activation maps.

The ROI chosen for the analysis was the Object Selective Cortex (OSC) map shown in Fig. 3.1. We analysed separately the performance of methods for the left part of the ROI, right part of the ROI and the whole ROI. Encoding of the activation maps was of two types: raw BOLD and beta maps. For the raw BOLD encoding, the volumes were selected that corresponded to the peak of the hemodynamic response function (HRF, as rendered by SPM software - `www.fil.ion.ucl.ac.uk/spm/doc/`) convolved with the boxcar function that represented the experimental manipulation. One volume was selected per trial, and for CMPT the volumes were averaged to produce a single representative volume per subject per condition per modality. For beta encoding, beta maps were calculated trialwise using linear regression; for CMPT the maps were averaged over trials, too, to produce a single beta map per subject per condition per modality.

CMDA was performed by training a logistic regression classifier with ℓ_2 regularization

Modalities	ROI	Encoding	CMDA	CMPT
P vs I	L+R	Bold	0.12	0.75
P vs I	L+R	Beta	0.08	**0.001**
P vs I	L	Beta	**0.001**	**0.001**
P vs I	R	Beta	**0.005**	**0.008**
V vs I (delay 2 s)	L+R	Beta	0.272	0.589
V vs I (delay 4 s)	L+R	Beta	0.387	0.141
V vs I (delay 6 s)	L+R	Beta	0.165	0.303
V vs I (delay 8 s)	L+R	Beta	0.659	**0.004**
V vs I (delay 10 s)	L+R	Beta	0.249	0.461
V vs I (all delays)	L+R	Beta	**0.012**	**0.001**

Table 6.1: Comparison of cross-modal decoding analysis (CMDA) and cross-modal permutation test (CMPT). The values refer to p-value for the discrimination task of *body* vs. *car* in each pair of the cognitive modalities listed in the extreme left column, where P stands for *perception*, I for *imagery* (I) and V for *visual search*.

on the trials of one of the modalities. The regularization parameter was selected according to a nested cross-validation scheme (leave-one-run out). The accuracy was then estimated on a test set from another modality. The training and test process was replicated for each subject. Then, the p-value was computed for the group using the permutation scheme described in [70]. The resulting p-values are reported in Table 1.

CMPT group analysis was carried out as described in section 6.2. The similarity distance between activation maps that we used is the Pearson correlation measure, both for raw BOLD volume and beta maps encoding. The significance of the proposed test statistics was computed by a permutation scheme with 10.000 iterations to estimate the null distribution. The resulting p-values are reported in Table 1.

In Table 1 we report only one result for the comparison between CMDA and CMPT related to raw BOLD volume encoding: the cross-modal analysis between Perception and Imagery. In this case none of the methods detect a meaningful shared activation pattern. Beta maps encoding on the other hand seems to be a more efficient representation. Chen and colleagues [42] demonstrate that using beta values is a way to get rid of intrinsic variabilities of BOLD signal throughout the brain and, specifically, within a single area. In our case, this means that beta values are more representative of the effect size than raw BOLD signal changes during task-on periods. For this reason in the presentation of results we only focus on results obtained with data encoded with beta maps.

The results in Table 1 confirm our expectations about the presence of common patterns

between modalities in Object Selective Cortex. At the same time, CMPT appears to have higher statistical power and sensitivity in revealing these patterns. The results reported in Table 1 illustrate two main scenarios: both CMDA and CMPT show significant p-values or only CMPT. In light of the simulation results we may argue, that since CMDA has a higher false error rate (Figure 6.1), in case of such disagreements the CMPT result is more reliable. This argument is further supported by the additional empirical evidence that the false positive rate or Type I error is similar for the two tests, limiting the risk of the disagreement being biased by a more optimistic rejection of the null hypothesis.

Cross-modal analysis results for perception vs. imagery with beta maps encoding are in agreement between CMDA and CMPT when the two hemispheres are considered individually, namely the left and right OSC respectively. When the analysis is extended to the joint ROI, the number of trials remains constant, while the number of voxels double. In this case the classifier is affected by the higher dimensionality of data, and CMDA does not succeed in rejecting the null hypothesis.

The empirical results also support the claim that CMPT is more robust not only in high-dimensional but also in small sample setting. Simulations show that the Type II error of CMPT is below that of CMDA in the small sample regime (Fig. 6.1). We may find analogous behaviour for the cross-modal analysis of visual search vs. imagery. If we consider the cumulative trials of visual search, irrespective of the delays, CMDA rejects the null hypothesis. When we restrict the cross-modal analysis to single delays of preparation period for visual search, the number of trials drop from 160 to 32. In this case CMDA fails to reject the null hypothesis while CMPT does not.

CMPT results are in line with the view that we should expect the presence of shared activity patterns between perceived and imagined object categories. CMPT analysis also confirms that the presence of these patterns can be expected in high-level visual areas processing information about object categories. We are going to further elaborate on this point in 6.5. On the other hand, CMDA results appear to be affected by the data sample size relative to the high dimensionality of data.

6.4.3 Exploratory data analysis with CMPT

We ran Searchlight analysis of the whole brain, using CMPT. First, we wanted to show if and how inserting CMPT as the elementary unit within the Searchlight framework could identify voxels that store information related to common activation patterns for two different cognitive modalities. Next, we intended to compare the spatial profiles of the exploratory analysis with the ROI individuated for the confirmatory analysis.

To construct group level maps, we referred to the procedure we illustrated in section Cross-modal Permutation Test (CMPT) at a ROI level. Here we consider the spheres centered on each voxel as ROIs: we first compute the "true" statistic and then we proceeded by using permutations. We started by computing single-participants' CMPT - Searchlight maps, where each voxel was considered as the center of a sphere (r= 8), and calculated the T-statistic. Then, we summed up single-participants' T-values and ended up with the true group-level statistic. Next, we created N=10000 permutations of the session labels, and we subsequently constructed 10000 averaged beta maps for each of the conditions based on the permuted labels - that is, two maps per subject per modality per permutation. In this way, we made sure that the data coming from different participants were tested against the same permutations in a uniform way. We then applied CMPT procedure on these permuted maps by first computing an individual T-statistic and then by summing up the group values - that is, we constructed an ad-hoc null distribution. Finally, we simply counted how many times the "true" group statistic was higher than the permuted group statistic, and we transformed the count in a fraction of the total number of permutations, obtaining a p-value. This value was assigned to the voxel in the center of the sphere. The procedure was repeated for each voxel within the gray matter mask. We ended up with a CMPT Searchlight map of p-values coming from a combination of permutation-like tests.

In Figure 6.3 we present results coming from the exploratory analysis. We took the maps for the pairs of cognitive processes where our confirmatory analysis yielded significant results, namely perception vs. imagery, visual search (delay 8s) vs. imagery, visual search (all delays) vs. imagery (see Table 6.1). For merely illustrative purposes, the maps were thresholded at the conventional significance level of 0.05 (as we are not aiming at significant cluster identification, no correction for multiple comparisons was carried out). In the left column of Fig. 6.3 we demonstrate the overlaps between the ROI identified in the course of the group analysis (Fig. 3.1) and the portions of the map that signal the presence of information about common patterns between two cognitive processes for a single slice (z=-16). The fact that the ROI identified contains a high portion of informative voxels is further illustrated by the histograms in the right column of the same figure. These are histograms of the p-values of the voxels within the ROI. We can see that all three histograms have a skewed shape, signalling the presence of a rather large number of voxels with p-values under 0.05 in the ROI.

For comparison, we also ran CMDA Searchlight with the same sphere size (r=8) for the same modality pairs: perception vs. imagery, visual search (all delays) vs. imagery, visual

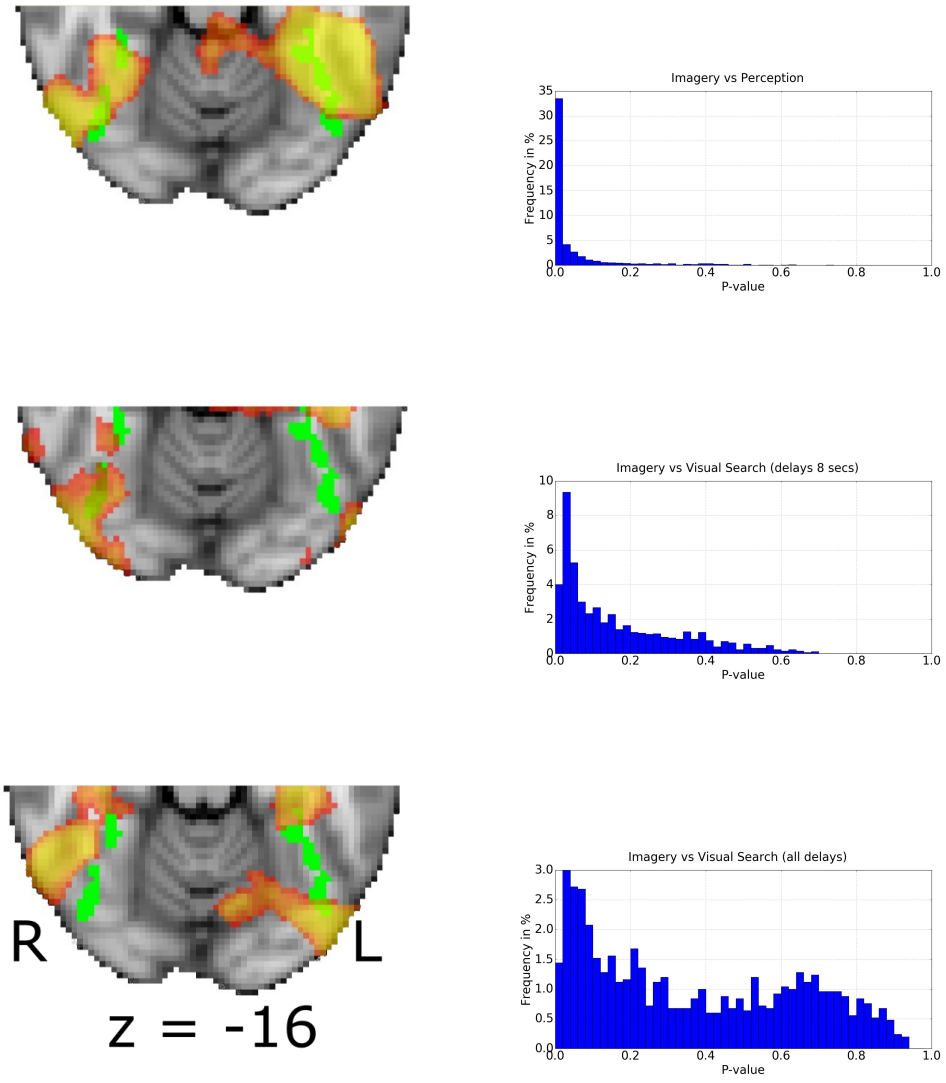


Figure 6.3: Overlap between the whole OSC ROI and informative voxels identified by Searchlight in the occipital-temporal cortex for the cross-modal pair of perception vs. imagery. The ROI (the same as in Figure fig:roi) is shown in blue. P-values in the Searchlight map are colorcoded. The maps were thresholded at the conventional significance level of 0.05, and the color code shows values between 0.95 (corresponding to the p-value of 0.05) in red and 1 (corresponding to the p-value of 0.00) in yellow. The leftmost column of the figure shows sagitarial (top) and coronal (bottom) slices (at x= 42 and y = -74 correspondingly) where the rectangle shaded in blue delimits the axial slices further presented in columns 2 - 4. These are six slices of the axial plane taken at steps of 8 between coordinates z = -24 and z = 16. We can see that in all these six slices, there are overlaps between the OSC ROI and areas where Searchlight identified voxels informative about shared activity patterns between perception and imagery.

search (delay 8) vs. imagery. The analysis was performed with Matlab 8.5.0, MathWorks, NatickMA, USA using in-house code and Libsvm library (https://www.csie.ntu.edu.tw/ cjlin/libsvm/). Classifier used for producing the maps was an SVM classifier with a linear kernel as implemented in the Libsvm library. For each subject, two Searchlight maps were obtained for each modality pair, one where the classifier was trained on the Imagery data and tested on the other modality data, and one where the assignment of train - test data was reversed. Then, these two maps were averaged as suggested in [187] yielding a single map per subject per modality pair. For the group analysis, a one-sample t-test against chance level (50 %) was performed using SPM software. The resulting group maps were thresholded at the significance level of 0.05 and cluster size of 10 voxels. Then we compared the group maps to the OSC ROI selected for the confirmatory analysis. The results are presented in Table 2. It shows percentages of voxels within the OSC ROI that were identified by the CMDA Searchlight as informative about shared patterns between two modalities. The numbers concerning the size of intersection between the Searchlight map and the ROI are given both as an absolute number of voxels within the ROI and in terms of percentages.

Modalities	ROI	Map Type	Intersection N Vox	Intersection %
P vs I	L	mean	4	0.64
P vs I	L	single P I	0	0
P vs I	L	single I P	239	38.24
V vs I	L	mean	5	0.8
V vs I	L	single V I	0	0
V vs I	L	single I V	12	1.92
V vs I (delay 8)	L	mean	0	0
V vs I (delay 8)	L	single I V	29	4.64
V vs I (delay 8)	L	single V I	0	0
P vs I	R	mean	45	7.2
P vs I	R	single P I	13	2.08
P vs I	R	single I P	193	30.88
V vs I	R	mean	0	0
V vs I	R	single V I	0	0
V vs I	R	single I V	1	0.16
V vs I (delay 8)	R	mean	0	0
V vs I (delay 8)	R	single I V	10	1.6
V vs I (delay 8)	L	single V I	0	0
P vs I	L+R	mean	49	3.92
P vs I	L+R	single P I	13	1.04
P vs I	L+R	single I P	432	34.56
V vs I	L+R	mean	5	0.4
V vs I	L+R	single V I	0	0
V vs I	L+R	single I V	13	1.04
V vs I (delay 8)	L+R	mean	0	0
V vs I (delay 8)	L+R	single I V	39	3.12
V vs I (delay 8)	L+R	single V I	0	0

.

Table 2. Percentage of voxels within the OSC ROI identified by the CMDA Searchlight as informative about shared patterns between two modalities. The cognitive modalities are listed in the extreme left column. P stands for *perception*, I for *imagery* and V for *visual search*. In the second column, it is shown which part of the OSC mask was used for calculations: left (L), right (R) or the whole mask (L+R). Number of voxels in L+R ROI is equal to 1250; each half of it (both L and R) has 650 voxels. The third column shows which type of map was used for calculations: an averaged map (denoted as mean),

or a single training - test pair map (denoted as single). In the names of the single maps, first comes the training modality, and second the test modality. In the last two columns you can find the numbers concerning the size of intersection between the map in column 3 and the ROI in column 2. In column 4, the intersection size is given as an absolute number of voxels within the ROI. In the last, extreme right column, it is given in terms of percentages.

The problem of asymmetry between classifier results when swapping train and test modalities in a cross-modal setting is well attested for the pair of perception vs. imagery. The accuracies obtained with the classifier trained on the imagery data have been shown to be consistently higher than after training on perception data [45,213]. To minimize the impact of this asymmetry in cross-modal investigations, it was suggested to average the maps resulting from different train-test combinations for a pair of modalities [187]. However, the authors of the paper showed that the divergence in accuracy numbers obtained with different train-test combinations in their data was insignificant. CMDA Searchlight results in our dataset seem to be rather seriously affected by the issues stemming from the accuracy asymmetries. First, for the pair of perception vs. imagery, CMDA Searchlight trained on imagery data identifies a high number of voxels within the OSC mask, both in right and left OSC. If trained on perception data, the Searchlight finds a much lower number of voxels within the same area, all of them in the right OSC. In the averaged mask, the number of the voxels that survive is nearly 10 times lower than that identified by the Searchlight trained on imagery data (49 against 432). This same kind of asymmetry is even more prominent for the pair of imagery vs. visual search: Searchlight, trained on imagery data, identifies voxels within the OSC ROI, while it does not identify any in the same area if trained with visual search data (neither all delays, nor delay 8). This result makes us pose a question about the extent to which the voxels idenitified belong actually to really shared patterns between modalities, or we should rather talk about voxels in one modality that are informative about the patterns in the other modality. Our overall conclusion about CMDA Searcchlight is that its use might be questionable in cases when notable asymmetry is expected, as is the case with perception vs. imagery. For some modality pairs asymmetry does not seem to be a big issue, as is the case with the data used in [187], and the use of CMDA Searchlight could be more justified for these data.

6.5 Discussion

The patterns of brain activity that are shared between the cognitive processes of perception and imagery have been the subject of quite numerous studies. The question investigated was if we can arrive at abstract, top-down object representations [116, 213] containing distinguishing features [112, 218] that will have common neural substrate both for viewed and imagined object categories [218]. To test for the presence of shared patterns, many studies used cross-modal decoding - namely, multivariate pattern analysis with SVM classifiers [45, 153, 213]. Significant cross-modal classification accuracies were taken as the evidence in favour of the presence of shared activity patterns. In some studies, correlation-based analysis was also performed to visualize and estimate similarity between these patterns in terms of distance [153, 213]. What emerged from these studies was the view that, indeed, visual imagery activates the same areas that contain information about visually perceived stimuli [74, 218], and shared patterns for stimulus categories in these two processes can be established [5, 45, 153, 199, 213, 247]. The areas where these common representations were found include the ventral temporal pathway, lateral occipital cortex [45, 153, 213, 247] and extrastriate cortex [116, 153, 199, 247]. The question of shared patterns in early visual areas, such as V1, remains controversial [153, 213]. Horikawa [112] showed that it depended on the feature type: lower visual features had similar representations for perception and imagery in lower visual areas, while the same was true for higher visual features in higher visual areas. Cichy [45] arrived at a similar conclusion about the subdivision of features: although they did not find significant accuracies for decoding object categories in lateral early visual cortex, they could identify shared representations of object locations in these areas.

Top-down attention patterns mediate attention biases during perception and affect behavioural performance in attention related tasks. In case of visual attention, these patterns can be revealed in visual search experiments via activity in the category-related object selective areas during preparatory delays [248]. Several studies attempted at demonstrating the high-level nature of the preparatory patterns through cross-modal analysis, mostly with visual perception as the other modality [200, 201, 248]. As object representations in the brain obtained during imagery tasks are thought to be closer to high-level top-down representations of objects in visual cortex [199, 218], the hypothesis naturally suggests itself that we can expect these patterns to show up also during visual search preparatory periods. We tried to shed light on this hypothesis using both CMDA and CMPT on visual imagery and visual search data. Besides, we ran cross-modal analysis separately for preparatory periods of varying length (between 2 and 10 seconds) to get insights into

preparatory dynamics. We were expecting that only certain dealys would result significant, conforming different hypotheses about this dynamics. For instance, if only shorter delays (2-4 s) had resulted significant, that could be evidence in favour of transitory and cue-related nature of the preparatory activity in the Object Selective Cortex. If, on the other hand, we had seen significant results in the longer delays, that could reveal the fact that it takes time for the activity to build up. First, we see that both methods confirm expectations about imagery patterns being more high level than perception. None of the methods yielded significant results in the pairs of visual search vs. perception. As for the presence of the shared patterns between visual imagery and visual search, we are faced again with limitations of CMDA as a method: its results can be significant and it can reveal the presence of shared patterns between preparatory periods and imagery, but this type of analysis needs a lot of data. On the other hand, CMPT can reveal shared patterns even with fewer data as is the case with 8 seconds delay. Further study is needed to uncover the temporal dynamics of the preparatory top-down patterns. We hypothesize that delays shorter that 8 s do not allow the preparatory activity to build up, while in case of 10 s the delay it is too long, and the subject might be loosing concentration after a certain period of time.

We placed CMPT side by side with other standard data analysis techniques in order to examine whether this approach could be as informative as others. We have shown that in confirmatory, top-down contexts CMPT can yield better results than CMDA. However, it is necessary to mention one limitation of the method. One of the overarching questions in the study of visual imagery is identifying neural representations of categorical features in the form of brain activation maps [199, 218]. CMPT method cannot provide insights into the location of the discriminative patterns at a ROI level. Despite being a more robust test for cross-modal analysis, CMPT is not appropriate to investigate the shape of shared pattern within a given ROI. In this case, CMPT doesn't support a sensitivity analysis at the voxel level needed to compute granular brain maps of activations that are common between modalities. On the other hand, CMDA (at least when linear classifiers are used) contains a vector of weights that can give some clues about the relevance of the input features. However, CMPT combined with Searchlight technique can be a helpful method to locate brain regions that contain information about common patterns between modalities.

We took advantage of one strong point of the Searchlight analysis, its "modular" nature: Searchlight might be thought of as a generic framework of data examination that can subsume various analysis techniques as elementary units. Searchlight is widely used in

neuroimaging, but it suffers from a number of issues. Conducting Searchlight analysis has several major advantages: first, it can be run on the whole brain, no prior ROI selection is required. Next, it avoids the "curse of dimensionality" of full brain classification, by reducing the number of features used at each point by the classifier. Finally, it has proven to be quite successful in identifying subject specific activation patterns [71]. The maps produced with Searchlight are of the same nature as the maps obtained with the univariate GLM approach, but they are based on a more fine-grained pattern identification from multiple voxels and better reflect the spatial properties of the BOLD signal (that is, adjacent voxels have similar activation patterns). However, major criticisms of the Searchlight approach regard the use of classification accuracy as the information measure and the t-test as the method to obtain group significances. As is pointed out in [71], SVM classifiers can correctly classify even with a few number of highly informative voxels and when weakly informative voxels are numerous enough. Both of these behaviours can cause distortions in a map: in the first case, all searchlights overlapping with one of a few informative voxels will be significant. In this way, the number of informative voxels is overestimated. In the second case, the cause of distortions is "discontinuous information detection": groups of weakly informative voxels will be missed out if their size is below a certain threshold, but can be judged significant if you just add a single voxel. That leads to underestimation of the number of significant clusters just because the number of weakly significant voxels does not reach a certain mass. Efficiency of using classifiers with Searchlight depends strongly on the classifier parameters and sphere size [2, 71]. In [2], the point is raised against interpretability of classifier accuracy with neuroscientific data: unlike distance measures, its value depends on the properties of the dataset (amount of training data and what kind of data is used as test data) and not only on the presence of a particular effect in the data. Besides, the authors point out that capturing interactions of several factors in a factorial experimental design cannot be cast as a classification task. So, addressing these methodological issues for Searchlight can significantly improve this valuable tool and make its result more scientifically rigorous.

Classification accuracy is not the only way to represent information content. In the original paper by Kriegeskorte [140] the metrics used was Mahalanobis distance between the distributions corresponding to stimulus categories. In [2] the authors build on the probabilistic model of the data proposing a cross-validated multivariate ANOVA (MANOVA) as the informational content measure. In [263] three various measures - classification accuracy, Euclidean/Mahalanobis distance, and Pearson correlation distance - are compared for reliability in the context of Searchlight analysis. In this paper, it was

shown that continuous crossvalidated distance estimators such as Euclidean/Mahalanobis distance or Pearson correlation should be preferred for Searchlight because they are more interpretable from the neuroscientific viewpoint.

Another bunch of critical remarks concerns the use of t-tests for assessing significance at the group level. Certain properties of neuroscientific data make the use of t-tests questionable for this purpose, particularly, the low number of observations and the non-gaussianity of the probability distribution of accuracy. As a consequence, several assumptions of the t-statistic are not met, rendering the procedure invalid from a theoretical point of view" [244]. However, it is not the only option here. In [244] a non-parametric test for group significance and cluster inference was proposed based on permutations and bootstrapping procedure. Nastase and colleagues [187] also opt for permutation tests in Searchlight context.

Methodologically, using CMPT in conjunction with the Searchlight technique for cross-modal pattern analysis has several advantages over the common Searchlight procedure because it does not rely neither on classification accuracy nor on the t-tests and hence avoids the common methodological pitfalls. At the same time, we are following the suggestions in the literature that are considered more appropriate for the Searchlight. First, the test statistic proposed in equation 6.2 that is used as the measure of information contained at each voxel is based on Pearson correlation and is interpretable in terms of similarity. Second, group significance is tested non-parametrically with permutation tests that do not make assumptions about the shape of the data distribution. We found that CMPT integrated into Searchlight has proven effective also to explore and, potentially, confirm what we observed using top-down, ROI-based analysis, which suggests both robustness and efficiency of the CMPT Searchlight in fMRI data analysis.

However, it is important to note that confirmatory and exploratory analyses report different p-values. While it is possible to qualitatively compare the outcomes of these two analyses, plainly putting their p-values side by side might be misleading. CMPT-ROI p-values come from an extended, functionally well-defined area including 625 or 1250 voxels. CMPT-SL analysis spans over the whole brain sphere by sphere, extracting results from spheres including about 200 voxels each. This means that p-values coming from confirmatory and exploratory analysis should not be compared on a purely quantitative level.

The question of shared patterns between various cognitive modalities is relevant not only for object categorization in visual processing. It is fundamental in the study of interactions between top-down and bottom-up processing streams in the human brain

in general. Further directions of study could include using the CMPT method and the CMPT Searchlight technique with a wider number of other cognitive modalities, such as auditory or linguistic [29, 234, 235]. Besides, we could investigate other areas that can share representations with imagery - for instance, working memory areas [199]. Finally, the CMPT method could be tried with other types of neuroimaging data - as, for example, EEG motor imagery data for Brain-Computer interfaces [43] or MEG data [61, 110].

6.6 Data and code availability.

Processed data (ROI beta-maps, BOLD volumes and Searchlight maps) are available from a public repository on Github `https://github.com/elena-kalinina/Code_CMPT_paper`. The code used to generate and analyze synthetic data is available from the same repository along with data analysis code and Searchlight implementation. Sharing the whole dataset involves the consent of third-parties who participated in the experiment design. If the consensus is reached, the data could be made publicly available.

Chapter 7

Conclusions and Discussion

The subject of our research was protocol adaptation in real-time neuroscientific studies. The need to adapt protocol interactively in an online mode stems from the paradigm shift that takes place in real-time experimental neuroimagning. Data acquisition methods in these studies are different from those in traditional experiments, because in the former the dependent variable is the measured brain activity, while stimulation is the independent variable. Based on these data, the researcher tests hypotheses about how external stimuli evoke responses from the brain, and the protocol has to be fixed before the study, to enable statistical inference. In real-time experimental manipulations, the question investigated is about how the current state of the subject conditions their response to external stimuli. Fixed stimulation protocols do not allow to study this question, because the relationship between variables is reversed: the signal becomes independent variable, while stimulation is adapted based on the signal as dependent variable. For this reason, real-time neuroimaging studies do not use fixed protocols (or use them very rarely) and need robust adaptive procedures for constructing stimulation sequences in real-time along with data analysis.

We analyzed major types of real-time neuroimaging experiments (BSDS, BCI, Neurofeedback) in Chapter 2. We also gave an overview of available protocol adaptation methods, where it was noted that real-time adaptation methods for interacting with the subject's states are very scarce. In Chapter 3 we outlined main challenges that each experimental type is facing for real-time protocol adaptation. In BSDS, it is detecting the subject's state reliably in real-time, when all samples are not yet available, and choosing a policy given severe temporal restrictions. BSDS as a method can produce valuable insights about mappings of the measured signal to subject states. These insights are critical for developing advanced closed loop stimulation protocols for BCIs and Neurofeedback

studies.

In our investigation of BSDS, we arrived at the conclusion that the reference method - supervised classification, or brain decoding - has several restrictions on its use in real-time settings, for a number of reasons, e.g. the curse of dimensionality. We had a very concrete state detection task in our case study of top-down attention, that involved not just supervised decoding, but cross-modal decoding, where training and test data were coming from two different cognitive modalities. This method was supposed to detect high-level common state patterns for these cognitive modalities, but was also failing, also because of limitations of supervised classification. So, for BSDS we proposed and evaluated an unsupervised method of state detection. We also came up with an alternative method for establishing common patterns between cognitive modalities, as this question is of fundamental importance in neuroscience. We framed detection of high-level common patterns as statistical hypothesis testing, and proposed an appropriate test statistic along with permutation tests schema for establishing significance. The method is described in Chapter 6.

In Chapter 4 we presented a real-time fMRI setup that allows to detect the subject's attentional state (top-down preparatory attention pattern) in a Brain-State-Dependent Stimulation setting. The method was based on an unsupervised model - GMM - for state detection, and an incremental variant of the model fitting algorithm was used to classify brain states in real-time. There are some issues that we would like to be pursued by future research. First, it is known that GMMs depend heavily on initialization. By trial and error we adopted the number of initial clusters as 5. In this way, we ensure that even though baseline shifts during the session, we still detect changes between components. Other directions for future studies could be learning the optimal number of components from the data - e.g. using Bayesian Information Criterion [38]. There are also dynamic models that create and continually adjust a Gaussian Mixture Model consistent to all sequentially presented data: in this way, they are able to create new components and "forget" other, unnecessary [67].

Second, as we were piloting an fMRI setup, its various components were tested, and in the end we do not possess a consistent dataset that would allow to perform good-quality statistical inference. In other words, we were not able to establish behavioural differences between two types of trials - BSDS trials and control trials. However, an additional validation method is needed anyway. If behavioural data would not not show differences between trials, the reason could be unclear: on the one hand, the method of state detection could have failed; on the other, the state detected in this way does not

necessarily have influence on perception. So we still need a separate validation method to confirm that indeed the desired states were detected.

In order to contribute to the develoment of closed loop protocols for BCIs and Neurofeedback, we designed and implemented an experiment with the scope of controlling a subject's affective state (arousal). As our second contribution, we presented this real-time experiment where the subject's skin conductance was measured, and stimulation was adjusted in a closed loop manner to maintain the signal at a target level. GSR measurements in real-time are mostly used for monitoring of arousal with the scope of detecting stress or other dysfunctional states [260]. We proposed and tested a novel method making use of real-time GSR signal processing for closed loop affective control. With the new approach we test the limits to controlling arousal by external measures. The results of the experiment, presented in Chapter 5, suggest it is quite possible to exert some degree of control over the subject's arousal. This work combines and extends several lines of research on the control of human mind. Designing control strategies to achieve certain states of mind through experimental manipulation is thought to be a frontier of mind control, though possible [176]. To successfully implement a control strategy, a clearly defined target state is needed as the goal of the neural control. For the control intervention to be beneficial, at least in the case of emotion control, it has to be defined in terms of cognitive representations, but at the same time control strategy is executed at the physical level, such as neural substrates and/or measured signal. For that we need a mapping between neural states and states of mind. This is is seen at the moment rather as a challenge driving cognitive neuroscience [176] than a solved issue. Another open challenge is estimating the physical system dynamics and determining its state. This is more achievable in an invasive manner, with the direct access to the neuronal measurements [228]. However, in view of limited applications of invasive control as well as general difficulties of measuring and modelling such noisy systems as human brain, it might be preferable to strive for guiding the brain rather than controlling it in a compelling manner [100]. Even in the absense of an ideal control strategy, simplified control strategies built with limited access to the system and drawing upon the brain's natural dynamics could still be beneficial for the clinical scenarios [176]. Control strategy based on GSR signal could be such a way of efficiently assisting affective state regulation in the brain.

With GSR measurements, we were able to define the target state and capture the system dynamics due to a very straightforward linear relationship between arousal and the signal. GSR is a very good correlate of arousal, but arousal is mediated and processed in the brain. If we can match activity in emotional networks in the brain with

GSR measurements, we can have access to control over these brain areas using GSR signal as an intermediary. Brain activity related to autonomic responses has been the subject of several studies aimed at identifying brain areas involved in autonomic control of emotional states. The principal approach here was experimentally perturbing the signal coming from autonomic nervous system with stimulation, so that increases and decreases in the autonomic signal could be correlated with the changes in the brain activity occurring concurrently [51]. Some of these studies used direct recordings of the sympathetic nerve activity [117]. However, more numerous studies employed non-invasive measurements of the autonomic response, such as heart rate, blood pressures and skin response [51]. This was done in a variety of tasks gambling task [48], the Stroop task [49] or biofeedback task, when the subject was asked to up- or downregulate arousal based on the feedback from the screen [50], [185]. The goal was establishing neural underpinnings of the dynamics of autonomic emotion control in order to shed some light on the dynamic integration of the visceral and cognitive processes [48]. In particular, that means disentangling brain mechanisms that generate changes in autonomic arousal during behavioral reactions to stimuli and those which process internal feedback signals to create representations and bring to awareness subjective feeling states. These neuroimaging studies have demonstrated that the constructs of emotion are consistently associated with topographically distinct and recurrent patterns of cerebral activity in humans [205]. The regions involved comprise orbitofrontal cortex [117, 185], insular cortex [50, 51, 117, 185, 205], thalamus [51, 117], ventromedial prefrontal cortex [185, 205]. Importantly, the amygdala as part of a more extensive supporting network including prefrontal cortex and striatum, was found to control arousal by modulating the neuroendocrine responses and the activity of noradrenergic nuclei [205]. These are regions evaluating the emotional value of stimuli, nearly all of them were found to be associated with viscerosensory processing. On the other hand, anterior/mid cingulate cortex [50, 51, 185, 205] has been identified as the region not only mapping visceral responses, but also controlling emotional arousal in a volitional manner. These brain areas can be regarded as the system forming an interface between cognitive/volitional behaviour and autonomic bodily responses. GSR signal can be a "window" into what happens in this system during emotional processing. That means, it allows to obtain necessary information about the systemic mechanisms involved in cognitive processing and volitional modulation of peripheral autonomic arousal states, represented by skin conductance, and it is possible that we can guide the system efficiently through that window.

The goal of our study was to test an adaptive stimulation protocol that can be applied

in a closed-loop manner to externally control affective states. This kind of approach could be beneficial for assisting clinical populations in need of emotion regulation, e.g., in autism or schizophrenia, as well as for healthy groups in situations where emotional control is required, e.g. under stress. It would also be interesting to manipulate arousal in order to control other cognitive states that arousal influences. For one, it would be interesting to see behavioural or other differences between groups that arise from the changes in arousal level - e.g. perceptual biases [142] or memory formation [63, 102].

We conclude that our closed loop real-time protocol allows to minimize the error signal consistently and significantly below that in the randomized sequence. Our results suggest that brain state control is feasible combining well-known stimuli and rather basic physiological measures.

Part II

Bibliography

Bibliography

[1] Laura Acqualagna and Benjamin Blankertz. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clinical neurophysiology*, 124(5):901–908, May 2013.

[2] Carsten Allefeld and John-Dylan Haynes. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage*, 89:345–357, April 2014.

[3] Edson Amaro and Gareth J. Barker. Study design in fMRI: basic principles. *Brain and cognition*, 60(3):220–232, April 2006.

[4] Christine Amrhein, Andreas Mühlberger, Paul Pauli, and Georg Wiedemann. Modulation of event-related brain potentials during affective picture processing: a complement to startle reflex and skin conductance response? *International journal of psychophysiology*, 54(3):231–240, November 2004.

[5] Andrew James J. Anderson, Elia Bruni, Alessandro Lopopolo, Massimo Poesio, and Marco Baroni. Reading visually embodied meaning from the brain: Visually grounded computational models decode visual-object mental imagery induced by written text. *NeuroImage*, 120:309–322, October 2015.

[6] Ariana Anderson, Jennifer Bramen, Pamela K. Douglas, Agatha Lenartowicz, Andrew Cho, Chris Culbertson, Arthur L. Brody, Alan L. Yuille, and Mark S. Cohen. Large Sample Group Independent Component Analysis of Functional Magnetic Resonance Imaging Using Anatomical Atlas-Based Reduction and Bootstrapped Clustering. *International journal of imaging systems and technology*, 21(2):223–231, June 2011.

[7] Patrik Andersson, Josien P. Pluim, Max A. Viergever, and Nick F. Ramsey. Navigation of a telepresence robot via covert visuospatial attention and real-time fMRI. *Brain topography*, 26(1):177–185, January 2013.

[8] Patrik Andersson, Nick F. Ramsey, Mathijs Raemaekers, Max A. Viergever, and Josien P. Pluim. Real-time decoding of the direction of covert visuospatial attention. *Journal of neural engineering*, 9(4):045004+, August 2012.

[9] O. D. Arandjelovic and R. Cipolla. Incremental learning of Temporally-Coherent gaussian mixture models. In *Procedings of the British Machine Vision Conference 2005*, page 59.1. British Machine Vision Association, 2005.

[10] Gabor Aranyi, Florian Pecune, Fred Charles, Catherine Pelachaud, and Marc Cavazza. Affective interaction with a virtual character through an fNIRS Brain-Computer interface. *Frontiers in computational neuroscience*, 10, 2016.

[11] Dominik R. Bach. A head-to-head comparison of SCRalyze and ledalab, two model-based methods for skin conductance analysis. *Biological psychology*, 103:63–68, December 2014.

[12] Dominik R. Bach, Guillaume Flandin, Karl J. Friston, and Raymond J. Dolan. Time-series analysis for rapid event-related skin conductance responses. *Journal of neuroscience methods*, 184(2):224–234, November 2009.

[13] Dominik R. Bach, Guillaume Flandin, Karl J. Friston, and Raymond J. Dolan. Modelling event-related skin conductance responses. *International journal of psychophysiology*, 75(3):349–356, March 2010.

[14] Dominik R. Bach, Guillaume Flandin, Karl J. Friston, and Raymond J. Dolan. Modelling event-related skin conductance responses. *International journal of psychophysiology*, 75(3):349–356, March 2010.

[15] Dominik R. Bach and Karl J. Friston. Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, 50(1):15–22, January 2013.

[16] Dominik R. Bach, Karl J. Friston, and Raymond J. Dolan. An improved algorithm for model-based analysis of evoked skin conductance responses. *Biological psychology*, 94(3):490–497, December 2013.

[17] Pierre Baldi and Yves Chauvin. Smooth On-Line learning algorithms for hidden markov models. *Neural Computation*, 6(2):307–318, March 1994.

[18] Christian Balkenius and Peter Gärdenfors. Spaces in the brain: From neurons to meanings. *Frontiers in psychology*, 7, 2016.

[19] Roberto Battiti and Mauro Brunato. *The LION way. Machine Learning* plus *Intelligent Optimization*. LIONlab, University of Trento, Italy, 2014.

[20] Robert Bauer, Meike Fels, Vladislav Royter, Valerio Raco, and Alireza Gharabaghi. Closed-loop adaptation of neurofeedback based on mental effort facilitates reinforcement learning of brain self-regulation. *Clinical neurophysiology*, 127(9):3156–3164, June 2016.

[21] Robert Bauer, Meike Fels, Vladislav Royter, Valerio Raco, and Alireza Gharabaghi. Closed-loop adaptation of neurofeedback based on mental effort facilitates reinforcement learning of brain self-regulation. *Clinical neurophysiology*, 127(9):3156–3164, June 2016.

[22] Robert Bauer and Alireza Gharabaghi. Estimating cognitive load during self-regulation of brain activity and neurofeedback with therapeutic brain-computer interfaces. *Frontiers in behavioral neuroscience*, 9, 2015.

[23] Robert Bauer and Alireza Gharabaghi. Reinforcement learning for adaptive threshold control of restorative brain-computer interfaces: a Bayesian simulation. *Frontiers in neuroscience*, 9, 2015.

[24] Robert Bauer and Alireza Gharabaghi. Constraints and Adaptation of Closed-Loop Neuroprosthetics for Functional Restoration. *Frontiers in neuroscience*, 11, 2017.

[25] Thomas Baumgartner, Michaela Esslen, and Lutz Jäncke. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1):34–43, April 2006.

[26] J Behari and DV Rai. Effect of some physiologically important drugs on the skin impedance. *Medical and Biological Engineering and Computing*, 19(2):244–246, 1981.

[27] G. L. Bilbro, W. E. Snyder, S. J. Garnier, and J. W. Gault. Mean field annealing: a formalism for constructing GNC-like algorithms. *IEEE Transactions on Neural Networks*, 3(1):131–138, January 1992.

[28] Christopher M. Bishop. *Neural networks for pattern recognition*. Clarendon Press ; Oxford University Press, 1 edition, January 1995.

[29] Valentina Borghesani, Fabian Pedregosa, Evelyn Eger, Marco Buiatti, and Manuela Piazza. A perceptual-to-conceptual gradient of word coding along the ventral path.

In *Pattern Recognition in Neuroimaging, 2014 International Workshop on*, pages 1–4. IEEE, June 2014.

[30] Wolfram Boucsein. *Electrodermal Activity*. Springer US, Boston, MA, 2012.

[31] Margaret M. Bradley, Bruce N. Cuthbert, and Peter J. Lang. Pictures as prepulse: Attention and emotion in startle modification. *Psychophysiology*, 30(5):541–545, September 1993.

[32] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, July 2008.

[33] Guillaume Bresson, Zayed Alsayed, Li Yu, and Sebastien Glaser. Simultaneous localization and mapping: A survey of current trends in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 2(3):194–220, September 2017.

[34] A. B. Bruhl. Making Sense of Real-Time Functional Magnetic Resonance Imaging (rtfMRI) and rtfMRI Neurofeedback. *International Journal of Neuropsychopharmacology*, 18(6):pyv020, February 2015.

[35] Andre R. Brunoni, Michael A. Nitsche, Nadia Bolognini, Marom Bikson, Tim Wagner, Lotfi Merabet, Dylan J. Edwards, Antoni Valero-Cabre, Alexander Rotenberg, Alvaro Pascual-Leone, Roberta Ferrucci, Alberto Priori, Paulo S. Boggio, and Felipe Fregni. Clinical research with transcranial direct current stimulation (tDCS): Challenges and future directions. *Brain Stimulation*, 5(3):175–195, July 2012.

[36] Bradley R. Buchbinder. Functional magnetic resonance imaging. *Handbook of clinical neurology*, 135:61–92, 2016.

[37] S. Calinon. *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press, 2009.

[38] S. Calinon and A. Billard. Incremental learning of gestures by imitation in a humanoid robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 255–262, March 2007.

[39] Laura Carelli, Federica Solca, Andrea Faini, Paolo Meriggi, Davide Sangalli, Pietro Cipresso, Giuseppe Riva, Nicola Ticozzi, Andrea Ciammola, Vincenzo Silani, and Barbara Poletti. Brain-Computer interface for clinical purposes: Cognitive assessment and rehabilitation. *BioMed research international*, 2017, 2017.

[40] Daniel R. Cavagnaro, Mark A. Pitt, and Jay I. Myung. Model discrimination through adaptive experimentation. *Psychonomic bulletin & review*, 18(1):204–210, February 2011.

[41] Heather Chapin, Epifanio Bagarinao, and Sean Mackey. Real-time fMRI applied to pain management. *Neuroscience letters*, 520(2):174–181, June 2012.

[42] Xu Chen, Francisco Pereira, Wayne Lee, Stephen Strother, and Tom Mitchell. Exploring predictive and reproducible modeling with the single-subject FIAC dataset. *Human brain mapping*, 27(5):452–461, May 2006.

[43] Kyuwan Choi. Electroencephalography (EEG)-based neurofeedback training for brain-computer interface (BCI). *Experimental brain research*, 231(3):351–365, November 2013.

[44] Maggie S. M. Chow, Sharon L. Wu, Sarah E. Webb, Katie Gluskin, and D. T. Yew. Functional magnetic resonance imaging and the brain: A brief review. *World Journal of Radiology*, 9(1):5+, 2017.

[45] Radoslaw M. Cichy, Jakob Heinzle, and John-Dylan Haynes. Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, 22(2):372–380, February 2012.

[46] R. W. Cox, A. Jesmanowicz, and J. S. Hyde. Real-time functional magnetic resonance imaging. *Magnetic resonance in medicine*, 33(2):230–236, February 1995.

[47] W. Miles Cox, Leena Subramanian, David E. Linden, Michael Lührs, Rachel McNamara, Rebecca Playle, Kerenza Hood, Gareth Watson, Joseph R. Whittaker, Raman Sakhuja, and Niklas Ihssen. Neurofeedback training for alcohol dependence versus treatment as usual: study protocol for a randomized controlled trial. *Trials*, 17(1), October 2016.

[48] Hugo D. Critchley. Neural mechanisms of autonomic, affective, and cognitive integration. *The Journal of comparative neurology*, 493(1):154–166, December 2005.

[49] Hugo D. Critchley. Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International Journal of Psychophysiology*, 73(2):88–94, August 2009.

[50] Hugo D. Critchley, Raphael N. Melmed, Eric Featherstone, Christopher J. Mathias, and Raymond J. Dolan. Volitional control of autonomic arousal: a functional magnetic resonance study. *NeuroImage*, 16(4):909–919, August 2002.

[51] Hugo D. Critchley, Yoko Nagai, Marcus A. Gray, and Christopher J. Mathias. Dissecting axes of autonomic control in humans: Insights from neuroimaging. *Autonomic neuroscience : basic & clinical*, 161(1-2):34–42, April 2011.

[52] Simone Cutini and Sabrina Brigadoi. Unleashing the future potential of functional near-infrared spectroscopy in brain sciences. *Journal of neuroscience methods*, 232C:152–156, May 2014.

[53] Jean Daunizeau, Vincent Adam, and Lionel Rigoux. VBA: A probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Comput Biol*, 10(1):e1003441+, January 2014.

[54] Jean Daunizeau, Kerstin Preuschoff, Karl Friston, and Klaas Stephan. Optimizing experimental design for comparing models of brain function. *PLOS Computational Biology*, 7(11):e1002280+, November 2011.

[55] Michael E. Dawson, Anne M. Schell, and Diane L. Filion. *The Electrodermal System*, chapter 10, pages 217–243. Cambridge University Press, Cambridge, 2017.

[56] Megan T. deBettencourt, Jonathan D. Cohen, Ray F. Lee, Kenneth A. Norman, and Nicholas B. Turk-Browne. Closed-loop training of attention with real-time brain imaging. *Nature neuroscience*, 18(3):470–475, March 2015.

[57] R. Christopher deCharms, Kalina Christoff, Gary H. Glover, John M. Pauly, Susan Whitfield, and John D. Gabrieli. Learned regulation of spatially localized brain activation using real-time fMRI. *NeuroImage*, 21(1):436–443, January 2004.

[58] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[59] Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(34):197–387, 2014.

[60] Gopikrishna Deshpande, D. Rangaprakash, Luke Oeding, Andrzej Cichocki, and Xiaoping P. Hu. A new generation of Brain-Computer interfaces driven by discovery of latent EEG-fMRI linkages using tensor decomposition. *Frontiers in neuroscience*, 11, 2017.

[61] Suzanne Dikker and Liina Pylkkänen. Predicting language: MEG evidence for lexical preactivation. *Brain and language*, 127(1):55–64, October 2013.

[62] Christopher DiMattina and Kechen Zhang. Adaptive stimulus optimization for sensory systems neuroscience. *Frontiers in Neural Circuits*, 7, 2013.

[63] Florin Dolcos and Roberto Cabeza. Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cognitive, affective & behavioral neuroscience*, 2(3):252–263, September 2002.

[64] J Durbin. The fitting of time series models. 28, 01 1960.

[65] R. Edelberg. Electrodermal recovery rate, goal-orientation, and aversion. *Psychophysiology*, 9(5):512–520, September 1972.

[66] Anders Eklund, Thomas E Nichols, and Hans Knutsson. Cluster failure: why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, page 201602413, 2016.

[67] Paulo Engel and Milton Heinen. Incremental learning of multivariate gaussian mixture models advances in artificial intelligence SBIA 2010. In Antônio da Rocha Costa, Rosa Vicari, and Flavio Tonidandel, editors, *Advances in Artificial Intelligence SBIA 2010*, volume 6404 of *Lecture Notes in Computer Science*, chapter 9, pages 82–91. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2011.

[68] Rodrigo A España, Brooke E Schmeichel, and Craig W Berridge. Norepinephrine at the nexus of arousal, motivation and relapse. *Brain research*, 1641:207–216, 2016.

[69] J. A. Etzel. MVPA permutation schemes: Permutation testing for the group level. In *Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on*, pages 65–68. IEEE, June 2015.

[70] Joset A. Etzel, Valeria Gazzola, and Christian Keysers. Testing Simulation Theory with Cross-Modal Multivariate Classification of fMRI Data. *PLoS ONE*, 3(11), November 2008.

[71] Joset A. Etzel, Jeffrey M. Zacks, and Todd S. Braver. Searchlight analysis: promise, pitfalls, and potential. *NeuroImage*, 78:261–269, September 2013.

[72] Brian S. Everitt and Edward T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 7(1):1–14, 1999.

[73] Sylvain Faisan, Laurent Thoraval, Jean-Paul P. Armspach, Marie-Noëlle N. Metz-Lutz, and Fabrice Heitz. Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models. *IEEE transactions on medical imaging*, 24(2):263–276, February 2005.

[74] M. J. Farah. The neural basis of mental imagery. *Trends in neurosciences*, 12(10):395–399, October 1989.

[75] Eugene A. Feinberg and Adam Shwartz. Handbook of markov decision processes : Methods and applications, 2002.

[76] Jacobo Fernandez-Vargas, Hanns U. Pfaff, Francisco B. Rodríguez, and Pablo Varona. Assisted closed-loop optimization of SSVEP-BCI efficiency. *Frontiers in neural circuits*, 7, 2013.

[77] B. Figner and R. O. Murphy. Using skin conductance in judgment and decision making research. *A handbook of process tracing methods for decision research: A critical review and user's guide*, pages 163–184, 2010.

[78] Nikolaj M. Filatov and Heinz Unbehauen. *Adaptive dual control : theory and applications*. Springer, 2004.

[79] Ronald A Fisher. The design of experiments. 1935. *Oliver and Boyd, Edinburgh*, 1935.

[80] Wendell H. Fleming and Halil M. Soner. *Controlled Markov Processes and Viscosity Solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, New York, 2006.

[81] A. Fournel and E. Reynaud. Exploring whole brain fMRI data with unsupervised artificial neural networks. In *Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on*, pages 53–56. IEEE, May 2011.

[82] Elisabeth V. Friedrich, Guilherme Wood, Reinhold Scherer, and Christa Neuper. Mind over brain, brain over mind: cognitive causes and consequences of controlling brain activity. *Frontiers in human neuroscience*, 8, 2014.

[83] K. J. Friston. Statistical parametric mapping : the analysis of functional brain images, 2011.

[84] K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Hum. Brain Mapp.*, 1(2):153–171, January 1994.

[85] Karl J. Friston and Raymond J. Dolan. Computational and dynamic models in neuroimaging. *NeuroImage*, 52(3):752–765, September 2010.

[86] A. Gaume, A. Vialatte, A. Mora-Sánchez, C. Ramdani, and F. B. Vialatte. A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. *Neuroscience and biobehavioral reviews*, June 2016.

[87] Michael S. Gazzaniga. Handbook of cognitive neuroscience, 1984.

[88] Michael S. Gazzaniga, George R. Mangun, and Sarah-Jayne Blakemore. *The cognitive neurosciences*. MIT Press, 2014.

[89] D. Gembris, J. G. Taylor, S. Schor, W. Frings, D. Suter, and S. Posse. Functional magnetic resonance imaging in real time (FIRE): sliding-window correlation analysis and reference-vector optimization. *Magnetic resonance in medicine*, 43(2):259–268, February 2000.

[90] Matthias Gerdts. Optimal control of ODEs and DAEs., 2012.

[91] Mattia I. Gerin, Harlan Fichtenholtz, Alicia Roy, Christopher J. Walsh, John H. Krystal, Steven Southwick, and Michelle Hampson. Real-Time fMRI neurofeedback with war veterans with chronic PTSD: A feasibility study. *Frontiers in psychiatry*, 7, 2016.

[92] Zoubin Ghahramani and Geoffrey E. Hinton. Parameter estimation for linear dynamical systems. Technical Report CRG-TR-96-2, Department of Computer Science. University of Toronto, FebFeb˜ 1996.

[93] Ramla Ghali, Sébastien Ouellet, and Claude Frasson. Using electrophysiological features in cognitive tasks: An empirical study. *International Journal of Information and Education Technology*, 6(8):584–590, 2016.

[94] Alireza Gharabaghi. What turns assistive into restorative Brain-Machine interfaces? *Frontiers in neuroscience*, 10, 2016.

[95] Alireza Gharabaghi, Dominic Kraus, Maria T. LeÃ£o, Martin SpÃler, Armin Walter, Martin Bogdan, Wolfgang Rosenstiel, Georgios Naros, and Ulf Ziemann. Coupling brain-machine interfaces with cortical stimulation for brain-state dependent stimulation: enhancing motor cortex excitability for neurorehabilitation. *Frontiers in Human Neuroscience*, 8, March 2014.

[96] M. Gho and F. J. Varela. A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal de physiologie*, 83(2):95–101, 1988.

[97] Gary H. Glover. Overview of Functional Magnetic Resonance Imaging. *Neurosurgery Clinics of North America*, 22(2):133–139, April 2011.

[98] Alexandre Gramfort, Gabriel Peyré, and Marco Cuturi. Fast optimal transport averaging of neuroimaging data, April 2015.

[99] Steven R. Green, Philip A. Kragel, Matthew E. Fecteau, and Kevin S. LaBar. Development and validation of an unsupervised scoring system (autonomate) for skin conductance response analysis. *International Journal of Psychophysiology*, 91(3):186–193, March 2014.

[100] Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K. Telesford, Alfred B. Yu, Ari E. Kahn, John D. Medaglia, Jean M. Vettel, Michael B. Miller, Scott T. Grafton, and Danielle S. Bassett. Controllability of structural brain networks. *Nature Communications*, 6:8414+, October 2015.

[101] Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K. Telesford, Alfred B. Yu, Ari E. Kahn, John D. Medaglia, Jean M. Vettel, Michael B. Miller, Scott T. Grafton, and Danielle S. Bassett. Controllability of structural brain networks. *Nature communications*, 6, October 2015.

[102] Stephan Hamann. Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, 5(9):394–400, September 2001.

[103] Carolyn W Harley. A role for norepinephrine in arousal, emotion and learning?: limbic modulation by norepinephrine and the kety hypothesis. *Progress in Neuropsychopharmacology and Biological Psychiatry*, 11(4):419–458, 1987.

[104] Thomas Hartmann, Hannah Schulz, and Nathan Weisz. Probing of brain states in Real-Time: Introducing the ConSole environment. *Frontiers in Psychology*, 2, 2011.

[105] Karen J. Hartwell, Colleen A. Hanlon, Xingbao Li, Jeffrey J. Borckardt, Melanie Canterberry, James J. Prisciandaro, Megan Moran M. Moran-Santa Maria, Todd LeMatty, Mark S. George, and Kathleen T. Brady. Individualized real-time fMRI neurofeedback to attenuate craving in nicotine-dependent smokers. *Journal of psychiatry & neuroscience : JPN*, 41(1), October 2015.

[106] James V. Haxby. Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage*, 62(2):852–855, August 2012.

[107] John-Dylan Haynes and Geraint Rees. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–534, July 2006.

[108] Ronald A. Hess. Control theory for humans: Quantitative approaches to modeling performance by r. j. jagacinski and j. m. flach, lawrence erlbaum associates, mahwah, NJ, 2003, 379 pp. IBSN 0-8058-2292-5. *Int. J. Robust Nonlinear Control*, 14(16):1371–1373, November 2004.

[109] Oliver Hinds, Satrajit Ghosh, Todd W. Thompson, Julie J. Yoo, Susan Whitfield-Gabrieli, Christina Triantafyllou, and John D. Gabrieli. Computing moment-to-moment BOLD activation for real-time neurofeedback. *NeuroImage*, 54(1):361–368, January 2011.

[110] Gerrit Hirschfeld, Pienie Zwitserlood, and Christian Dobel. Effects of language comprehension on visual processing - MEG dissociates early perceptual and late N400 effects. *Brain and language*, 116(2):91–96, February 2011.

[111] John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. The University of Michigan Press, 2008.

[112] Tomoyasu Horikawa and Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, 8:15037+, May 2017.

[113] Scott A. Huettel, Allen W. Song, and Gregory McCarthy. *Functional magnetic resonance imaging*. Sinauer, 2014.

[114] René J. Huster, Zacharais N. Mokom, Stefanie Enriquez-Geppert, and Christoph S. Herrmann. Brain-computer interfaces for EEG neurofeedback: peculiarities and solutions. *International journal of psychophysiology*, 91(1):36–45, January 2014.

[115] Niklas Ihssen, Moses O. Sokunbi, Andrew D. Lawrence, Natalia S. Lawrence, and David E. Linden. Neurofeedback of visual food cue reactivity: a potential avenue to alter incentive sensitization and craving. *Brain imaging and behavior*, May 2016.

[116] Alumit Ishai. Seeing faces and objects with the "mind's eye". *Archives italiennes de biologie*, 148(1):1–9, March 2010.

[117] Cheree James, Luke Henderson, and Vaughan G. Macefield. Real-time imaging of brain areas involved in the generation of spontaneous skin sympathetic nerve activity at rest. *NeuroImage*, 74:188–194, July 2013.

[118] Hojin Jang, Sergey M. Plis, Vince D. Calhoun, and Jong-Hwan H. Lee. Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *NeuroImage*, 145(Pt B):314–328, January 2017.

[119] Ole Jensen, Ali Bahramisharif, Robert Oostenveld, Stefan Klanke, Avgis Hadjipapas, Yuka O. Okazaki, and Marcel A. van Gerven. Using brain-computer interfaces and brain-state dependent stimulation as tools in cognitive neuroscience. *Frontiers in psychology*, 2, 2011.

[120] Imène Jraidi, Maher Chaouachi, and Claude Frasson. A hierarchical probabilistic framework for recognizing learners' interaction experience trends and emotions. *Advances in Human-Computer Interaction*, 2014:1–16, 2014.

[121] J. B. Julian, Evelina Fedorenko, Jason Webster, and Nancy Kanwisher. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4):2357–2364, May 2012.

[122] Kathrin Cohen C. Kadosh, Qiang Luo, Calem de Burca, Moses O. Sokunbi, Jianfeng Feng, David E. Linden, and Jennifer Y. Lau. Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. *NeuroImage*, 125:616–626, January 2016.

[123] Thomas Kailath. *Linear Systems*. Prentice Hall, January 1980.

[124] Daniel Kaiser, Damiano C. Azzalini, and Marius V. Peelen. Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, pages jn.01074.2015+, January 2016.

[125] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[126] Jonas T. Kaplan, Kingson Man, and Steven G. Greening. Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, 9, March 2015.

[127] Jonas T. Kaplan, Kingson Man, and Steven G. Greening. Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, 9:151, 2015.

[128] Hilbert J. Kappen. *Optimal control theory and the linear Bellman equation*, chapter 17, pages 363–387. Cambridge University Press, Cambridge, 2011.

[129] Susanne Karch, Daniel Keeser, Sebastian Hümmer, Marco Paolini, Valerie Kirsch, Temmuz Karali, Michael Kupka, Boris-Stephan Rauchmann, Agnieszka Chrobok, Janusch Blautzik, Gabi Koller, Birgit Ertl-Wagner, and Oliver Pogarell. Modulation of craving related brain responses using Real-Time fMRI in patients with alcohol use disorder. *PLoS ONE*, 10(7):e0133034+, July 2015.

[130] Fumi Katsuki and Christos Constantinidis. Bottom-up and top-down attention: different processes and overlapping neural systems. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 20(5):509–521, October 2014.

[131] André W. Keizer, Roland S. Verment, and Bernhard Hommel. Enhancing cognitive control through neurofeedback: a role of gamma-band activity in managing episodic retrieval. *NeuroImage*, 49(4):3404–3413, February 2010.

[132] Junghoe Kim, Vince D. Calhoun, Eunsoo Shim, and Jong-Hwan H. Lee. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage*, 124(Pt A):127–146, January 2016.

[133] Elsa Andrea A. Kirchner, Su Kyoung K. Kim, Sirko Straube, Anett Seeland, Hendrik Wöhrle, Mario Michael M. Krell, Marc Tabie, and Manfred Fahle. On the applicability of brain reading for predictive human-machine interfaces in robotics. *PloS one*, 8(12), 2013.

[134] Martina Kirsch, Isabella Gruber, Matthias Ruf, Falk Kiefer, and Peter Kirsch. Real-time functional magnetic resonance imaging neurofeedback can reduce striatal cue-reactivity to alcohol stimuli. *Addiction biology*, June 2015.

[135] André Knops, Bertrand Thirion, Edward M. Hubbard, Vincent Michel, and Stanislas Dehaene. Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. *Science*, 324(5934):1583–1585, June 2009.

[136] Yury Koush, Djalel-E E. Meskaldji, Swann Pichon, Gwladys Rey, Sebastian W. Rieger, David E. Linden, Dimitri Van De Ville, Patrik Vuilleumier, and Frank Scharnowski. Learning control over emotion networks through Connectivity-Based neurofeedback. *Cerebral cortex (New York, N.Y. : 1991)*, December 2015.

[137] Yury Koush, Maria Joao J. Rosa, Fabien Robineau, Klaartje Heinen, Sebastian W Rieger, Nikolaus Weiskopf, Patrik Vuilleumier, Dimitri Van De Ville, and Frank Scharnowski. Connectivity-based neurofeedback: dynamic causal modeling for real-time fMRI. *NeuroImage*, 81:422–430, November 2013.

[138] Yury Koush, Mikhail Zvyagintsev, Miriam Dyck, Krystyna A. Mathiak, and Klaus Mathiak. Signal quality and bayesian signal processing in neurofeedback based on real-time fMRI. *NeuroImage*, 59(1):478–489, January 2012.

[139] Nikolaus Kriegeskorte. Pattern-information analysis: from stimulus decoding to computational-model testing. *NeuroImage*, 56(2):411–421, May 2011.

[140] Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868, March 2006.

[141] Ann M. Kring and Janelle M. Caponigro. Emotion in schizophrenia. *Current Directions in Psychological Science*, 19(4):255–259, August 2010.

[142] Kamesh Krishnamurthy, Matthew R. Nassar, Shilpa Sarode, and Joshua I. Gold. Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour*, 1(6):0107+, May 2017.

[143] Stepan Y. Kruglikov and Steven J. Schiff. Interplay of electroencephalogram phase and auditory-evoked neural activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(31):10122–10127, November 2003.

[144] Stephen M. LaConte. Decoding fMRI brain states in real-time. *NeuroImage*, 56(2):440–454, May 2011.

[145] Stephen M. LaConte, Scott J. Peltier, and Xiaoping P. Hu. Real-time fMRI using brain-state classification. *Human brain mapping*, 28(10):1033–1044, October 2007.

[146] R. Gregory Lande and Miriam Pourzand. Brain computer interface technology: Usability and applications in psychiatry. *Technology and health care*, April 2015.

[147] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. Emotion, attention, and the startle reflex. *Psychological review*, 97(3):377–395, July 1990.

[148] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, May 1993.

[149] Peter J Lang. Emotions response patterns: The brain and the autonomic nervous system. *Emotion Review*, 6(2):93–99, 2014.

[150] Peter J. Lang, Margaret M. Bradley, and B. N. Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, 2008.

[151] Emma J. Lawrence, Li Su, Gareth J. Barker, Nick Medford, Jeffrey Dalton, Steve C. Williams, Niels Birbaumer, Ralf Veit, Sitaram Ranganatha, Jerzy Bodurka, Michael Brammer, Vincent Giampietro, and Anthony S. David. Self-regulation of the anterior insula: Reinforcement learning using real-time fMRI neurofeedback. *NeuroImage*, 88C:113–124, November 2013.

[152] Jong-Hwan H. Lee, Jeongwon Ryu, Ferenc A. Jolesz, Zang-Hee H. Cho, and Seung-Schik S. Yoo. Brain-machine interface via real-time fMRI: preliminary study on thought-controlled robotic arm. *Neuroscience letters*, 450(1):1–6, January 2009.

[153] Sue-Hyun Lee, Dwight J. Kravitz, and Chris I. Baker. Disentangling visual imagery and perception of real-world objects. *NeuroImage*, 59(4):4064–4073, February 2012.

[154] Eric L Lehmann and Charles Stein. On the theory of some non-parametric hypotheses. *The Annals of Mathematical Statistics*, pages 28–45, 1949.

[155] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2008.

[156] Norman Levinson. The wiener (root mean square) error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 25(1-4):261–278, April 1946.

[157] Zhonglin Li, Li Tong, Linyuan Wang, Yongli Li, Wenjie He, Min Guan, and Bin Yan. Self-regulating positive emotion networks by feedback of multiple emotional brain states using real-time fMRI. *Experimental brain research*, 234(12):3575–3586, December 2016.

[158] David E. Linden. Neurofeedback and networks of depression. *Dialogues in clinical neuroscience*, 16(1):103–112, March 2014.

[159] David E. Linden, Isabelle Habes, Stephen J. Johnston, Stefanie Linden, Ranjit Tatineni, Leena Subramanian, Bettina Sorger, David Healy, and Rainer Goebel. Real-time self-regulation of emotion networks in patients with depression. *PloS one*, 7(6), 2012.

[160] Nikos K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878, June 2008.

[161] Romy Lorenz, Ricardo P. Monti, Ines R. Violante, Christoforos Anagnostopoulos, Aldo A. Faisal, Giovanni Montana, and Robert Leech. The automatic neuroscientist: automated experimental design with real-time fMRI, June 2015.

[162] Romy Lorenz, Ines R. Violante, Ricardo P. Monti, Giovanni Montana, Adam Hampshire, and Robert Leech. Dissociating frontoparietal brain networks with neuroadaptive bayesian optimization. *bioRxiv*, pages 128678+, April 2017.

[163] Katherine A. Loveland. *Social-emotional impairment and self-regulation in autism spectrum disorders*, pages 365–382. Oxford University Press, December 2004.

[164] Hugo Lövheim. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2):341–348, 2012.

[165] Michael Lührs and Rainer Goebel. Turbo-Satori: a neurofeedback and brain-computer interface toolbox for real-time functional near-infrared spectroscopy. *Neurophotonics*, 4(4), October 2017.

[166] Michael Lührs, Bettina Sorger, Rainer Goebel, and Fabrizio Esposito. Automated selection of brain regions for real-time fMRI brain-computer interfaces. *Journal of neural engineering*, 14(1), November 2016.

[167] J. Luigjes, R. Breteler, S. Vanneste, and D. de Ridder. [neuromodulation as an intervention for addiction: overview and future prospects]. *Tijdschrift voor psychiatrie*, 55(11):841–852, 2013.

[168] Xinpei Ma, Chun-An A. Chou, Hiroki Sayama, and Wanpracha Art A. Chaovalitwongse. Brain response pattern identification of fMRI data using a particle swarm optimization-based approach. *Brain informatics*, 3(3):181–192, September 2016.

[169] Abdelhak Mahmoudi, Sylvain Takerkart, Fakhita Regragui, Driss Boussaoud, and Andrea Brovelli. Multivoxel Pattern Analysis for fMRI Data: A Review. *Computational and Mathematical Methods in Medicine*, 2012:1–14, 2012.

[170] Steve Majerus, Nelson Cowan, Frédéric Péters, Laurens Van Calster, Christophe Phillips, and Jessica Schrouff. Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. *Cerebral cortex (New York, N.Y. : 1991)*, 26(1):166–179, January 2016.

[171] S. Makni, P. Ciuciu, J. Idier, and J. B. Poline. Bayesian joint Detection-Estimation of brain activity using MCMC with a Gamma-Gaussian mixture prior model. In *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, pages V–1093–V–1096. IEEE, 2006.

[172] Marsel Mano, Anatole Lécuyer, Elise Bannier, Lorraine Perronnet, Saman Noorzadeh, and Christian Barillot. How to build a hybrid neurofeedback platform combining EEG and fMRI. *Frontiers in Neuroscience*, 11, March 2017.

[173] Michael Marxen, Mark J. Jacob, Dirk K. Müller, Stefan Posse, Elena Ackley, Lydia Hellrung, Philipp Riedel, Stephan Bender, Robert Epple, and Michael N. Smolka. Amygdala regulation following fMRI-neurofeedback without instructed strategies. *Frontiers in human neuroscience*, 10, 2016.

[174] Hengameh Marzbani, Hamid Reza R. Marateb, and Marjan Mansourian. Neurofeedback: A comprehensive review on system design, methodology and clinical applications. *Basic and clinical neuroscience*, 7(2):143–158, April 2016.

[175] Iris B. Mauss and Michael D. Robinson. Measures of emotion: A review. *Cognition and Emotion*, 23(2):209–237, February 2009.

[176] John D. Medaglia, Perry Zurn, Walter Sinnott-Armstrong, and Danielle S. Bassett. Mind control as a guide for the mind. *Nature Human Behaviour*, 1(6):0119+, June 2017.

[177] John D. Medaglia, Perry Zurn, Walter Sinnott-Armstrong, and Danielle S. Bassett. Mind control as a guide for the mind. *Nature Human Behaviour*, 1(6):0119+, June 2017.

[178] J-A A. Micoulaud-Franchi, R. Richieri, C. Lancon, and J. Vion-Dury. [interactive rTMS protocols in psychiatry]. *L'Encephale*, 39(6):426–431, December 2013.

[179] Carlo Miniussi, Justin A. Harris, and Manuela Ruzzoli. Modelling non-invasive brain stimulation in cognitive neuroscience. *Neuroscience & Biobehavioral Reviews*, 37(8):1702–1712, September 2013.

[180] Masaya Misaki, Nafise Barzigar, Vadim Zotev, Raquel Phillips, Samuel Cheng, and Jerzy Bodurka. Real-time fMRI processing with physiological noise correction - Comparison with off-line analysis. *Journal of neuroscience methods*, September 2015.

[181] Ricardo P. Monti, Romy Lorenz, Christoforos Anagnostopoulos, Robert Leech, and Giovanni Montana. Measuring the functional connectome "on-the-fly": towards a new control signal for fMRI-based brain-computer interfaces, February 2015.

[182] Charles Mueller, Michael Luehrs, Sebastian Baecke, Daniela Adolf, Ralf Luetzkendorf, Michael Luchtmann, and Johannes Bernarding. Building virtual reality fMRI paradigms: a framework for presenting immersive virtual environments. *Journal of neuroscience methods*, 209(2):290–298, August 2012.

[183] Sarah F. Muldoon, Fabio Pasqualetti, Shi Gu, Matthew Cieslak, Scott T. Grafton, Jean M. Vettel, and Danielle S. Bassett. Stimulation-Based control of dynamic brain networks. *PLOS Computational Biology*, 12(9):e1005076+, September 2016.

[184] Jay I. Myung, Daniel R. Cavagnaro, and Mark A. Pitt. A tutorial on adaptive design optimization. *Journal of mathematical psychology*, 57(3-4):53–67, June 2013.

[185] Y. Nagai, H. D. Critchley, E. Featherstone, M. R. Trimble, and R. J. Dolan. Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a default mode of brain function. *NeuroImage*, 22(1):243–251, May 2004.

[186] Georgios Naros and Alireza Gharabaghi. Reinforcement learning of self-regulated -oscillations for motor restoration in chronic stroke. *Frontiers in human neuroscience*, 9, 2015.

[187] S. A. Nastase, Y. O. Halchenko, B. Davis, and U. Hasson. Cross-modal searchlight classification: methodological challenges and recommended solutions. In *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4, June 2016.

[188] Adnan M. Niazi, Philip L. van den Broek, Stefan Klanke, Markus Barth, Mannes Poel, Peter Desain, and Marcel A. van Gerven. Online decoding of object-based

attention using real-time fMRI. *The European journal of neuroscience*, 39(2):319–329, January 2014.

[189] Thomas E Nichols and Andrew P Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

[190] Andrew A. Nicholson, Daniela Rabellino, Maria Densmore, Paul A. Frewen, Christian Paret, Rosemarie Kluetsch, Christian Schmahl, Jean Théberge, Richard W. J. Neufeld, Margaret C. McKinnon, Jim Reiss, Rakesh Jetly, and Ruth A. Lanius. The neurobiology of emotion regulation in posttraumatic stress disorder: Amygdala downregulation via real-time fMRI neurofeedback. *Hum. Brain Mapp.*, page n/a, September 2016.

[191] Søren F. V. Nielsen, Mikkel N. Schmidt, Kristoffer H. Madsen, and Morten Mørup. Predictive assessment of models for dynamic functional connectivity. *NeuroImage*, December 2017.

[192] Markus Ojala and Gemma C. Garriga. Permutation tests for studying classifier performance. 11:908–913, 11 2009.

[193] Jonas K. Olofsson, Steven Nordin, Henrique Sequeira, and John Polich. Affective picture processing: an integrative review of ERP findings. *Biological psychology*, 77(3):247–265, March 2008.

[194] Rasmus K. Olsson and Lars K. Hansen. Linear State-Space models for blind source separation. *J. Mach. Learn. Res.*, 7:2585–2602, December 2006.

[195] Hernando Ombao, Martin Lindquist, Wesley Thompson, and John Aston. Handbook of neuroimaging data analysis, 2017.

[196] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs M. Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:1–9, 2011.

[197] Franklin Orndorff-Plunkett, Fiza Singh, Oriana R. Aragón, and Jaime A. Pineda. Assessing the effectiveness of neurofeedback training in the context of clinical and social neuroscience. *Brain sciences*, 7(8), August 2017.

[198] Christian Paret, Rosemarie Kluetsch, Jenny Zaehringer, Matthias Ruf, Traute Demirakca, Martin Bohus, Gabriele Ende, and Christian Schmahl. Alterations of

amygdala-prefrontal connectivity with real-time fMRI neurofeedback in BPD patients. *Social cognitive and affective neuroscience*, February 2016.

[199] Joel Pearson, Thomas Naselaris, Emily A. Holmes, and Stephen M. Kosslyn. Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in cognitive sciences*, 19(10):590–602, October 2015.

[200] Marius V. Peelen, Li Fei-Fei, and Sabine Kastner. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, 460(7251):94–97, June 2009.

[201] Marius V. Peelen and Sabine Kastner. A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29):12125–12130, July 2011.

[202] Frenk Peeters, Jacco Ronner, Lonneke Bodar, Jim van Os, and Richel Lousberg. Validation of a neurofeedback paradigm: Manipulating frontal EEG alpha-activity and its impact on mood. *International Journal of Psychophysiology*, 93(1):116–120, July 2014.

[203] Will D. Penny, Peter Zeidman, and Neil Burgess. Forward and backward inference in spatial cognition. *PLOS Computational Biology*, 9(12):e1003383+, December 2013.

[204] Alex Pentland and Andrew Liu. Modeling and prediction of human behavior. *Neural Computation*, 11(1):229–242, January 1999.

[205] Martin Peper, Martin Herpers, Joachim Spreer, Jürgen Hennig, and Josef Zentner. Functional neuroimaging of emotional learning and autonomic reactions. *Journal of Physiology-Paris*, 99(4-6):342–354, June 2006.

[206] Lorraine Perronnet, Anatole Lécuyer, Marsel Mano, Elise Bannier, Fabien Lotte, Maureen Clerc, and Christian Barillot. Unimodal versus bimodal EEG-fMRI neurofeedback of a motor imagery task. *Frontiers in human neuroscience*, 11, 2017.

[207] Edwin JG Pitman. Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):119–130, 1937.

[208] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

[209] Steve M. Potter, Ahmed El Hady, and Eberhard E. Fetz. Closed-loop neuroscience and neuroengineering. *Frontiers in Neural Circuits*, 8:115, 2014.

[210] Mohit Rana, Nalin Gupta, Josue L. Dalboni Da Rocha, Sangkyun Lee, and Ranganatha Sitaram. A toolbox for real-time subject-independent and subject-dependent classification of brain states from fMRI signals. *Frontiers in neuroscience*, 7, 2013.

[211] Mariela Rance, Michaela Ruttorf, Frauke Nees, Lothar Rudi R. Schad, and Herta Flor. Real time fMRI feedback of the anterior cingulate and posterior insular cortex in the processing of pain. *Human brain mapping*, July 2014.

[212] Andreas M. Ray, Ranganatha Sitaram, Mohit Rana, Emanuele Pasqualotto, Korhan Buyukturkoglu, Cuntai Guan, Kai-Keng K. Ang, Cristián Tejos, Francisco Zamorano, Francisco Aboitiz, Niels Birbaumer, and Sergio Ruiz. A subject-independent pattern-based Brain-Computer interface. *Frontiers in behavioral neuroscience*, 9, 2015.

[213] Leila Reddy, Naotsugu Tsuchiya, and Thomas Serre. Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, 50(2):818–825, April 2010.

[214] Karen Reiter, Søren Bo B. Andersen, and Jessica Carlsson. Neurofeedback Treatment and Posttraumatic Stress Disorder: Effectiveness of Neurofeedback on Posttraumatic Stress Disorder and the Optimal Choice of Protocol. *The Journal of nervous and mental disease*, 204(2):69–77, February 2016.

[215] Stacey Reynolds and Shelly J. Lane. Diagnostic validity of sensory over-responsivity: a review of the literature and case reports. *Journal of autism and developmental disorders*, 38(3):516–529, March 2008.

[216] Fabien Robineau, Djalel E. Meskaldji, Yury Koush, Sebastian W. Rieger, Christophe Mermoud, Stephan Morgenthaler, Dimitri Van De Ville, Patrik Vuilleumier, and Frank Scharnowski. Maintenance of voluntary self-regulation learned through Real-Time fMRI neurofeedback. *Frontiers in human neuroscience*, 11, 2017.

[217] Lucy F. Robinson, Tor D. Wager, and Martin A. Lindquist. Change point estimation in multi-subject fMRI studies. *NeuroImage*, 49(2):1581–1592, January 2010.

[218] Stephanie M. Roldan. Object Recognition in Mental Representations: Directions for Exploring Diagnostic Features through Visual Mental Imagery. *Frontiers in Psychology*, 8, May 2017.

[219] Vincenzo Romei, Verena Brodbeck, Christoph Michel, Amir Amedi, Alvaro Pascual-Leone, and Gregor Thut. Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cerebral cortex (New York, N.Y. : 1991)*, 18(9):2010–2018, September 2008.

[220] Reuven Y. Rubinstein. *Simulation and the monte carlo method. (uden titel) 1 1.* John Wiley, 2008.

[221] Sergio Ruiz, Niels Birbaumer, and Ranganatha Sitaram. Editorial: Learned brain Self-Regulation for emotional processing and attentional modulation: From theory to clinical applications. *Frontiers in behavioral neuroscience*, 10, 2016.

[222] Sergio Ruiz, Korhan Buyukturkoglu, Mohit Rana, Niels Birbaumer, and Ranganatha Sitaram. Real-time fMRI brain computer interfaces: Self-regulation of single brain regions to networks. *Biological Psychology*, 95:4–20, January 2014.

[223] Gaëtan Sanchez. *Real-time electrophysiology in cognitive neuroscience : towards adaptive paradigms to study perceptual learning and decision making in humans.* Theses, Université Claude Bernard - Lyon I, June 2014.

[224] Gaëtan Sanchez, Jean Daunizeau, Emmanuel Maby, Olivier Bertrand, Aline Bompas, and Jérémie Mattout. Toward a new application of Real-Time electrophysiology: Online optimization of cognitive neurosciences hypothesis testing. *Brain Sciences*, 4(1):49–72, January 2014.

[225] João R. Sato, Rodrigo Basilio, Fernando F. Paiva, Griselda J. Garrido, Ivanei E. Bramati, Patricia Bado, Fernanda Tovar-Moll, Roland Zahn, and Jorge Moll. Real-time fMRI pattern decoding and neurofeedback using FRIEND: an FSL-integrated BCI toolbox. *PloS one*, 8(12), 2013.

[226] Kristina Schaaff and Marc T. P. Adam. Measuring emotional arousal for online applications: Evaluation of ultra-short term heart rate variability measures. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 362–368. IEEE, September 2013.

[227] Frank Scharnowski, Chloe Hutton, Oliver Josephs, Nikolaus Weiskopf, and Geraint Rees. Improving visual perception through neurofeedback. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(49):17830–17841, December 2012.

[228] Steven J. Schiff. *Neural control engineering the emerging intersection between control theory and neuroscience*. MIT Press, 2012.

[229] Sanne Schoenmakers, Umut Güçlü, Marcel van Gerven, and Tom Heskes. Gaussian mixture models and semantic gating improve reconstructions from human brain activity. *Frontiers in Computational Neuroscience*, 8, January 2015.

[230] Jens Schwarzbach. A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior research methods*, 43(4):1194–1201, December 2011.

[231] Aaron R. Seitz. Cognitive neuroscience: Targeting neuroplasticity with neural decoding and biofeedback. *Current Biology*, 23(5):R210–R212, March 2013.

[232] K. Shibata, T. Watanabe, Y. Sasaki, and M. Kawato. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*, 334(6061):1413–1415, December 2011.

[233] Svetlana V. Shinkareva, Vicente L. Malave, Robert A. Mason, Tom M. Mitchell, and Marcel A. Just. Commonality of neural representations of words and pictures. *NeuroImage*, 54(3):2418–2425, February 2011.

[234] Irina Simanova, Jolien C. Francken, Floris P. de Lange, and Harold Bekkering. Linguistic priors shape categorical perception. *Language, Cognition and Neuroscience*, 31(1):159–165, January 2016.

[235] Irina Simanova, Peter Hagoort, Robert Oostenveld, and Marcel A. van Gerven. Modality-independent decoding of semantic information from the human brain. *Cerebral cortex (New York, N.Y. : 1991)*, 24(2):426–434, February 2014.

[236] R. Sitaram, N. Weiskopf, A. Caria, R. Veit, M. Erb, and N. Birbaumer. fMRI Brain-Computer interfaces. *Signal Processing Magazine, IEEE*, 25(1):95–106, 2008.

[237] Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. FMRI brain-computer interface: a tool for neuroscientific research and treatment. *Computational intelligence and neuroscience*, 2007.

[238] Ranganatha Sitaram, Sangkyun Lee, Sergio Ruiz, and Niels Birbaumer. *Real-Time Regulation and Detection of Brain States from fMRI Signals*, pages 227–438. Elsevier, 2011.

[239] Ranganatha Sitaram, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, Maria Laura L. Blefari, Mohit Rana, Ethan Oblak, Niels Birbaumer, and James Sulzer. Closed-loop brain training: the science of neurofeedback. *Nature reviews. Neuroscience*, December 2016.

[240] C. Smyser, T. J. Grabowski, R. J. Frank, J. W. Haller, and L. Bolinger. Real-time multiple linear regression for fMRI supported by time-aware acquisition and processing. *Magnetic resonance in medicine*, 45(2):289–298, February 2001.

[241] Moses O. Sokunbi. Feedback of real-time fMRI signals: From concepts and principles to therapeutic interventions. *Magnetic Resonance Imaging*, 35:117–124, January 2017.

[242] Nicola Soldati, Vince D. Calhoun, Lorenzo Bruzzone, and Jorge Jovicich. ICA analysis of fMRI with real-time constraints: an evaluation of fast detection performance as function of algorithms, parameters and a priori conditions. *Frontiers in human neuroscience*, 7, 2013.

[243] Hojen P. Sorensen, L. Hansen, and C. Rasmussen. Bayesian modelling of fMRI time series, 1999.

[244] Johannes Stelzer, Yi Chen, and Robert Turner. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, 65:69–82, January 2013.

[245] Robert F. Stengel. Optimal control and estimation, 1994.

[246] K. E. Stephan, F. Schlagenhauf, Q. J. M. Huys, S. Raman, E. A. Aponte, K. H. Brodersen, L. Rigoux, R. J. Moran, J. Daunizeau, R. J. Dolan, K. J. Friston, and A. Heinz. Computational neuroimaging strategies for single patient predictions. *NeuroImage*, 145:180–199, January 2017.

[247] Mark Stokes, Russell Thompson, Rhodri Cusack, and John Duncan. Top-down activation of shape-specific population codes in visual cortex during mental imagery. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(5):1565–1572, February 2009.

[248] Mark Stokes, Russell Thompson, Anna C. Nobre, and John Duncan. Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences*, 106(46):19569–19574, November 2009.

[249] Leena Subramanian, John V. Hindle, Stephen Johnston, Mark V. Roberts, Masud Husain, Rainer Goebel, and David Linden. Real-time functional magnetic resonance imaging neurofeedback for treatment of parkinson's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(45):16309–16317, November 2011.

[250] Christopher Summerfield, Tobias Egner, Matthew Greene, Etienne Koechlin, Jennifer Mangels, and Joy Hirsch. Predictive codes for forthcoming perception in the frontal cortex. *Science (New York, N.Y.)*, 314(5803):1311–1314, November 2006.

[251] Christopher Summerfield, Tobias Egner, Jennifer Mangels, and Joy Hirsch. Mistaking a house for a face: neural correlates of misperception in healthy humans. *Cerebral cortex (New York, N.Y. : 1991)*, 16(4):500–508, April 2006.

[252] Bradley P. Sutton, Cheng Ouyang, Dimitrios C. Karampinos, and Gregory A. Miller. Current trends and challenges in MRI acquisitions to investigate brain function. *International journal of psychophysiology*, 73(1):33–42, July 2009.

[253] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning — The MIT Press*. A Bradford Book, March 1998.

[254] Ariel Tankus, Itzhak Fried, and Shy Shoham. Cognitive-motor brain-machine interfaces. *Journal of physiology, Paris*, 108(1):38–44, February 2014.

[255] Robert T. Thibault, Michael Lifshitz, and Amir Raz. The climate of neurofeedback: scientific rigour and the perils of ideology. *Brain : a journal of neurology*, December 2017.

[256] Matthias S. Treder and Benjamin Blankertz. (c)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions*, 6(1):28+, May 2010.

[257] D. Val-Laillet, E. Aarts, B. Weber, M. Ferrari, V. Quaresima, L. E. Stoeckel, M. Alonso-Alonso, M. Audette, C. H. Malbert, and E. Stice. Neuroimaging and neuromodulation approaches to study eating behavior and prevent and treat eating disorders and obesity. *NeuroImage: Clinical*, 8:1–31, 2015.

[258] Gaël Varoquaux, Pradeep R. Raamana, Denis A. Engemann, Andrés Hoyos-Idrobo, Yannick Schwartz, and Bertrand Thirion. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145:166–179, January 2017.

[259] Petra Vetter, Fraser W. Smith, and Lars Muckli. Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology*, 24(11):1256–1262, June 2014.

[260] María V. Villarejo, Begoña G. Zapirain, and Amaia M. Zorrilla. A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. *Sensors*, 12(5):6075–6101, May 2012.

[261] J. T. Voyvodic. Real-time fMRI paradigm control, physiology, and behavior combined with near real-time statistical analysis. *NeuroImage*, 10(2):91–106, August 1999.

[262] Tor D. Wager and Thomas E. Nichols. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage*, 18(2):293–309, February 2003.

[263] Alexander Walther, Hamed Nili, Naveed Ejaz, Arjen Alink, Nikolaus Kriegeskorte, and Jörn Diedrichsen. Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 137:188–200, August 2016.

[264] B. Douglas Ward, John Janik, Yousef Mazaheri, Yan Ma, and Edgar A. DeYoe. Adaptive kalman filtering for real-time mapping of the visual field. *NeuroImage*, 59(4):3533–3547, February 2012.

[265] B. Douglas Ward and Yousef Mazaheri. State-space estimation of the input stimulus function using the kalman filter: a communication system model for fMRI experiments. *Journal of neuroscience methods*, 158(2):271–278, December 2006.

[266] Takeo Watanabe, Yuka Sasaki, Kazuhisa Shibata, and Mitsuo Kawato. Advances in fMRI Real-Time neurofeedback. *Trends in cognitive sciences*, October 2017.

[267] Nicholas R. Waytowich, Vernon J. Lawhern, Addison W. Bohannon, Kenneth R. Ball, and Brent J. Lance. Spectral transfer learning using information geometry for a User-Independent Brain-Computer interface. *Frontiers in neuroscience*, 10, 2016.

[268] Nikolaus Weiskopf. Real-time fMRI and its application to neurofeedback. *NeuroImage*, 62(2):682–692, August 2012.

[269] Nikolaus Weiskopf, Klaus Mathiak, Simon W. Bock, Frank Scharnowski, Ralf Veit, Wolfgang Grodd, Rainer Goebel, and Niels Birbaumer. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE transactions on bio-medical engineering*, 51(6):966–970, June 2004.

[270] Nikolaus Weiskopf, Ranganatha Sitaram, Oliver Josephs, Ralf Veit, Frank Scharnowski, Rainer Goebel, Niels Birbaumer, Ralf Deichmann, and Klaus Mathiak. Real-time functional magnetic resonance imaging: methods and applications. *Magnetic resonance imaging*, 25(6):989–1003, July 2007.

[271] Leanne M. Williams. Amygdala-Guided neurofeedback for major depression. *The American journal of psychiatry*, 174(8):717–718, August 2017.

[272] Anderson M. Winkler, Gerard R. Ridgway, Gwenalle Douaud, Thomas E. Nichols, and Stephen M. Smith. Faster permutation inference in brain imaging. *NeuroImage*, 141:502 – 516, 2016.

[273] Mark W. Woolrich, Christian F. Beckmann, Thomas E. Nichols, and Stephen M. Smith. *Statistical Analysis of fMRI Data*, pages 183–239. Springer New York, 2016.

[274] Julie J. Yoo, Oliver Hinds, Noa Ofen, Todd W. Thompson, Susan Whitfield-Gabrieli, Christina Triantafyllou, and John D. Gabrieli. When the brain is prepared to learn: enhancing human learning using real-time fMRI. *NeuroImage*, 59(1):846–852, January 2012.

[275] Brittany M. Young, Zack Nigogosyan, Veena A. Nair, Léo M. Walton, Jie Song, Mitchell E. Tyler, Dorothy F. Edwards, Kristin Caldera, Justin A. Sattin, Justin C. Williams, and Vivek Prabhakaran. Case report: post-stroke interventional BCI rehabilitation in an individual with preexisting sensorineural disability. *Frontiers in neuroengineering*, 7, 2014.

[276] Kymberly Young, Greg Siegle, Vadim Zotev, Wayne Drevets, and Jerzy Bodurka. EEG correlates of Real-Time fMRI neurofeedback amygdala training in depression. *Biological Psychiatry*, 81(10):S379+, May 2017.

[277] Kymberly D. Young, Masaya Misaki, Catherine J. Harmer, Teresa Victor, Vadim Zotev, Raquel Phillips, Greg J. Siegle, Wayne C. Drevets, and Jerzy Bodurka. Real-Time functional magnetic resonance imaging amygdala neurofeedback changes positive information processing in major depressive disorder. *Biological psychiatry*, March 2017.

[278] Kymberly D. Young, Greg J. Siegle, Masaya Misaki, Vadim Zotev, Raquel Phillips, Wayne C. Drevets, and Jerzy Bodurka. Altered task-based and resting-state amygdala functional connectivity following real-time fMRI amygdala neurofeedback training in major depressive disorder. *NeuroImage. Clinical*, 17:691–703, 2018.

[279] Kymberly D. Young, Vadim Zotev, Raquel Phillips, Masaya Misaki, Han Yuan, Wayne C. Drevets, and Jerzy Bodurka. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PloS one*, 9(2):e88785+, February 2014.

[280] Raheel Zafar, Sarat C. Dass, and Aamir Saeed S. Malik. Electroencephalogram-based decoding cognitive states using convolutional neural network and likelihood ratio based score fusion. *PloS one*, 12(5), 2017.

[281] Gaoyan Zhang, Li Yao, Hang Zhang, Zhiying Long, and Xiaojie Zhao. Improved working memory performance through Self-Regulation of dorsal lateral prefrontal cortex activation using Real-Time fMRI. *PLoS ONE*, 8(8):e73735+, August 2013.

[282] Catharina Zich, Maarten De Vos, Cornelia Kranczioch, and Stefan Debener. Wireless EEG with individualized channel layout enables efficient motor imagery training. *Clinical neurophysiology*, July 2014.

[283] Vadim Zotev, Frank Krueger, Raquel Phillips, Ruben P. Alvarez, W. Kyle Simmons, Patrick Bellgowan, Wayne C. Drevets, and Jerzy Bodurka. Self-Regulation of Amygdala Activation Using Real-Time fMRI Neurofeedback. *PLoS ONE*, 6(9):e24522+, September 2011.

[284] Christoph Zrenner, Paolo Belardinelli, Florian Müller-Dahlhaus, and Ulf Ziemann. Closed-Loop Neuroscience and Non-Invasive Brain Stimulation: A Tale of Two Loops. *Frontiers in cellular neuroscience*, 10, 2016.

[285] Agnieszka Zuberer, Daniel Brandeis, and Renate Drechsler. Are treatment effects of neurofeedback training in children with ADHD related to the successful regulation of brain activity? a review on the learning of regulation of brain activity and a contribution to the discussion on specificity. *Frontiers in human neuroscience*, 9, 2015.