Expanding the analysis of functional Near-Infrared Spectroscopy (fNIRS) data with multivariate techniques: application to a children's literacy study

vorgelegt von
M.Sc.
Jessica Gemignani
ORCID: 0000-0002-7722-4489

von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (*Dr.-Ing*)
genehmigte Dissertation

Promotionausschuss:
Vorsitzender: Prof. Dr. Klaus Obermayer
Gutachter: Prof. Dr. Benjamin Blankertz
Gutachter: Prof. Dr. Judit Gervain
Gutachter: Prof. Dr. Isabell Wartenburger
Gutachter: Dr. Christoph H. Schmitz
Tag der wissenschaftlichen Aussprache: 25. September 2019

Berlin 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme, under the Marie Sklodowska-Curie grant agreement No. 641858, Innovative Training Network *PredictAble*, "Understanding and predicting developmental language abilities and disorders in multilingual Europe".

Acknowledgements

These years as a Ph.D. student in Berlin have been a true rollercoaster of experiences and emotions that I will never forget, and now that I finished writing this thesis I want to take a moment to thank all those who made it that way.

As someone who pursued her Ph.D at university, while being employed at a company, framed in a network with several other universities, and did a research visit overseas, I surely have come in touch with a lot of people. And it is indeed the case that when I began, my worst fear was that there was just "too many people" involved and that I wouldn't be able to meet everyone's expectations.

Little did I know that those "too many people" would actually be key to the success of this work and that they would all help me shape my project and my scientific growth in such a smooth way I could not possibly imagine at the beginning.

First of all, I want to thank Prof. Benjamin Blankertz, who guided me scientifically and from whom I learned so much - but I wish I had learnt so much more!

Then my gratitude to the whole NIRx family: you have been, indeed, like a family to me. In particular, I want to thank Dr. Christoph Schmitz and Prof. Randall Barbour for always helping and sharing their immense knowledge on the theoretical aspects of fNIRS, you definitely have been the ace in the hole that anyone in my position would have wanted; I want to express my gratitude to Dr. Eike Middell, former colleague of the Development division, for always jumping together with me into my data analysis problems and contributing to shape the project from the very start. I want also to thank the whole Support Team, in particular Ms. Lamija Pašalić, because by giving me constant exposure to the support cases they helped me understand so much better many practical aspects of various fNIRS applications - this was so precious for my learning. Finally, a big thank you to Mr. Richard Barbour, NIRx's CEO, for believing

in my abilities and making possible that I could be in contact with all these very different aspects of the company life.

I want to express my gratitude also to Prof. Kenneth Pugh, director of Haskins Laboratories, where I spent three wonderful and so productive months: I have been amazed at how available he has been to me during the whole visit and afterwards, and how open and helpful the Haskins members have all been to me; working there has been a real landmark moment of my Ph.D journey.

I am also very grateful I had the opportunity of being a Marie Skłodowska-Curie Early Stage Researcher (ESR): knowing to be part of a network has been very encouraging from the very beginning, and I want to thank in particular Prof. Isabell Wartenburger, for letting me in with her fNIRS data, guiding me through her data analysis, sharing tips, discussing with me shortcomings and problems that were to be addressed, in other words, for teaching me a lot!

That name of mine on this dissertation represents therefore not only my work but also the sum of the contributions of all of you, who helped me so much along my way.

Finally, nothing would have been possible without the support of my wonderful family: my husband, my siblings, my loving parents, always ready to jump on a plane at a moment's notice and come to Berlin just to see me for a few hours. I feel so lucky to have you in my life.

Abstract

The use of functional Near-Infrared Spectroscopy is experiencing a rapid growth in use in every area of neuroscientific research, but one of the most prolific areas of application is the field of language development in children.

If over the past decades there have been incredible advancements on the technology side, the same can not be said about data analysis techniques: for fNIRS, an *ad-hoc* standardized data analysis procedure is still lacking; there is no general consensus around the best set of pre-processing steps to be performed; there is no standard choice of which hemoglobin component should be used to infer functional brain activity (oxygenated or deoxygenated hemoglobin); the statistical analysis has been carried out for a long time with methods borrowed from the fMRI practice. Only recently, effort has been made in defining fNIRS-specific techniques, both at single-channel and multi-channel level.

The first contribution of this work was to introduce a single-channel classifier that combines features from both oxy- and deoxyhemoglobin and that employs Linear Discriminant Analysis (LDA) to classify channels as 'active' or 'not-active'. Its performances were compared to those achieved by the General Linear Model (GLM) and it was found that the LDA-based classifier not only yields higher classification accuracies but those accuracies are more stable across different subjects, i.e. the multivariate method was more robust with respect to intersubjective variability of the hemodynamic response.

The second contribution of this work was to investigate the impact of literacy on functional brain organization, with both single-channel (univariate) and multi-channel (multivariate) analysis approaches. Using a large fNIRS dataset including children of various ages and literacy levels, we were able to show that the univariate approach has a unique strength in localizing the effects under investigation. On the other hand,

the multi-channel analysis approach did not produce a statistically significant effect, most likely because the experimental design was not optimally suited for the trial-by-trial classification; nevertheless, it highlighted a trend that had eluded the univariate analysis.

We conclude that both types of analysis should be routinely employed, because their complementary strengths can answer to different questions, but also that the single-channel analysis should be rendered more robust by using both hemoglobin components and that a data-driven approach such as the one proposed may mitigate the shortcomings of the model-based single-channel analysis. Finally, in order to perform a multivariate pattern analysis, the experimental paradigm should be designed as to include enough trials for classification.

Zusammenfassung

Der Einsatz der funktionellen Nahinfrarotspektroskopie nimmt in allen Bereichen der neurowissenschaftlichen Forschung rasant zu, und eines der produktivsten Anwendungsgebiete ist das Feld der Sprachentwicklung bei Kindern.

Wenn es in den letzten Jahrzehnten technologisch gesehen unglaubliche Fortschritte gegeben hat, kann man das Gleiche nicht über Datenanalyseverfahren sagen: Für fNIRS fehlt noch ein ad-hoc standardisiertes Datenanalyseverfahren; es gibt keinen allgemeinen Konsens über die beste Reihe von durchzuführenden Vorverarbeitungsschritten; es gibt keine Standardauswahl, welche Hämoglobinkomponente verwendet werden sollte, um funktionelle Hirnaktivität abzuleiten (sauerstoffreiches oder sauerstoffarmes Hämoglobin); die statistische Analyse wird seit langem mit Methoden durchgeführt, die der fMRI-Praxis entnommen wurden. Erst kürzlich wurden Anstrengungen unternommen, um fNIRS-spezifische Techniken zu definieren, sowohl auf Einkanal- als auch auf Mehrkanalebene.

Der erste Beitrag dieser Arbeit war die Einführung eines Einkanalklassifizierers, der Merkmale von Oxy- und Desoxyhämoglobin kombiniert und der die Linear Discriminant Analysis einsetzt, um Kanäle als "aktiv" oder "nicht aktiv" zu klassifizieren. Seine Leistungen wurden mit denen des General Linear Model verglichen und es wurde festgestellt, dass der LDA-basierte Klassifikator nicht nur höhere Klassifikationsgenauigkeiten liefert, sondern dass diese Genauigkeiten bei verschiedenen Probanden stabiler sind, d.h. die multivariate Methode war robuster in Bezug auf intersubjektive Variabilität der hämodynamischen Reaktion.

Der zweite Beitrag dieser Arbeit war die Untersuchung der Auswirkungen der Alphabetisierung auf die funktionelle Gehirnorganisation, mit einkanaligen (univariaten)

und mehrkanaligen (multivariaten) Analysemethoden. Anhand eines großen fNIRS-Datensatzes mit Kindern unterschiedlichen Alters und Alphabetisierungsgrades konnten wir zeigen, dass der univariate Ansatz eine einzigartige Stärke bei der Lokalisierung der untersuchten Effekte hat. Andererseits zeigte der Multi-Channel-Analyse-Ansatz keinen statistisch signifikanten Effekt, wahrscheinlich weil das experimentelle Design nicht optimal für die Klassifizierung im Versuch geeignet war; dennoch zeigte er einen Trend auf, der sich der univariaten Analyse entzogen hatte.

Wir kommen zu dem Schluss, dass beide Arten der Analyse routinemäßig eingesetzt werden sollten, weil ihre komplementären Stärken auf unterschiedliche Fragen antworten können, aber auch, dass die Einkanalanalyse durch den Einsatz beider Hämoglobinkomponenten robuster gemacht werden sollte und dass ein datengetriebener Ansatz wie der vorgeschlagene die Mängel der modellbasierten Einkanalanalyse mildern kann.

Schließlich, um eine multivariate Musteranalyse durchzuführen, sollte das experimentelle Paradigma so gestaltet werden, dass es genügend Versuche für die Klassifizierung enthält.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The use of functional Near-Infrared Spectroscopy in language research

Functional Near Infrared Spectroscopy (fNIRS) is a relatively recent technique: it was only in 1977 that Franz Jöbsis reported in his seminal Science article "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters" [1] that biological materials are transparent enough in the near-infrared region of the spectrum so to allow to observe and record changes in tissue blood volume and in hemoglobin-oxyhemoglobin equilibrium for research and clinical purposes.

Since then, the technique has experienced a rapid growth in use, thanks to steady advancements in the development of the technology: while in the early 90s recordings were conducted mostly with bulky, single-channel devices, nowadays fNIRS devices are extremely portable, include hundreds of channels, multi-distance measurements to measure depth-dependent contributions, they are compatible for simultaneous acquisition of any other measurement and can record at a very high sampling frequency.

As more sophisticated instrumentation has been developed and put to use in cognitive experiments, fNIRS has made important contributions to the understanding of functional brain activity and higher cognitive functions in virtually every area of neuroscientific research. In particular, one of the most prolific areas of application for fNIRS has been the field of language development in children.

Several aspects are key to this success: the fact that it is completely portable means that children do not need to be physically isolated from others to participate to an experiment, but instead they can be setup in a comfortable environment where they feel at ease; the fact that there are virtually no operating costs makes it possible to run several sessions on the same child, and this may prove especially useful when the goal is closely following up the outcome of a treatment or a remediation program. The fact that it is completely silent allows to administer any kind of auditory stimuli in completely ecological conditions. Finally, it is completely safe and harmless.

And it is indeed the study of typical and atypical language development in children that laid the foundations of this doctoral project.

The work presented in this thesis is embedded in the framework of "PredictAble", a large multi-disciplinary Marie Skłodowska Curie Innovative Training Network (ITN) that includes, besides the Technische Universität Berlin and NIRx Medizintechnik GmbH, also University of Potsdam, University Paris Descartes, University of Barcelona Pompeu Fabra and University of Jyväskylä.

The work scope of PredictAble is to pioneer an interdisciplinary approach to enhance the understanding of the cognitive mechanisms that underlie developmental disorders of spoken and written language, two areas of study that have been traditionally considered separately, and to do so by making the best use of the recent technological developments, such as fNIRS.

In this context, this doctoral project was carried out at the Technische Universität Berlin and at NIRx Medizintechnik, one of the leading manufacturers of fNIRS devices.

The role of this work in the grant was to expand and optimize the tools available for the analysis of fNIRS data, in order to render them suitable for use with very young children, with the ultimate goal of developing diagnostic tools to detect early risks for language-related impairments.

In fact, despite the dramatic increase in the use and application of fNIRS, an ad-hoc standardized data analysis procedure is still lacking; there is no general consensus around the best set of pre-processing steps to be performed, or around the correct channel exclusion criteria [2]; furthermore, there is no standard choice of which hemoglobin

component should be used to infer functional brain activity: while in principle functional activation should manifest with a concurrent increase in oxygenated-hemoglobin and a decrease in deoxygenated-hemoglobin, it is oftentimes the case that significant changes in response to a certain task are only found in one chromophore and not the other; this can happen for many different reasons, but it is clear that, without an established guideline, a serious issue of reproducibility arises: the same study may draw different conclusions when considering only results found on oxy-, or deoxy-hemoglobin, or a combination of the two [3]. As for the statistical analysis, the last step of the analysis chain, methods reported in literature mostly involve time series averaging and/or the use of methods borrowed from the fMRI practice, like Statistical Parametric Mapping (SPM) [4], which proved suboptimal for the analysis of fNIRS data [5]. Recently, a more advanced and fNIRS-tailored General Linear Model has been proposed that overcomes some of the drawbacks of SPM [6]. Nevertheless, all of these methods are most commonly applied to one channel at a time and to one hemoglobin component (univariate analysis). Recently, effort has been made in the direction of defining more sophisticated methods that involve classification of multi-channel patterns of hemodynamic activity by means of machine learning algorithms [7][8].

This new direction could represent a turning point in the fNIRS data analysis, since it could help mitigate many of the limitations of the aforementioned methods [9]: concurrent variations of both hemoglobin components could be considered in a multi-variate fashion, with the potential of increasing the sensitivity of the analysis while removing the ambiguity associated with choosing one or the other chromophore; furthermore, in this framework, the natural inter- and intra-subject variability of the hemodynamic responses, that for any model-based analysis constitutes a problem, could be accounted for in what might be called a self-referencing scheme; finer effects that are distributed across multiple channels could be detected, even when they don't achieve statistical significance at single-channel level. For these reasons, multi-variate analysis approaches based on the use of machine learning have the potential of offering novel insights while overcoming many technical problems related to the current analysis practice, and represent the object of study of this doctoral project.

1.2 Motivations and outline of this thesis

This thesis reviews the options available for the analysis of fNIRS data and describes the special consideration that need to be taken into account - especially when working with children.

While reviewing the opportunities offered by the application of exquisitely univariate analysis approaches, this work wants to promote the appreciation that the fNIRS signal's richness of features may open the way for the discovery of subtle effects, by means of multivariate analysis approaches.

To pursue this goal, this work was developed in two steps: the first step includes the quantitative comparison, under controlled conditions, of a univariate and a multivariate analysis approach. This step was crucial in that it allowed to precisely quantify in terms of classification accuracy the advantage of using a multivariate approach over a univariate.

The second part of this work was carried out on experimental data collected at Haskins Laboratories (New Haven, CT, USA), a secondary partner of the PredictAble network; Haskins Laboratories is a leading institution in the study of typical and atypical development of spoken and written language in children.

Here, a large dataset of fNIRS measurements was collected on beginning and skilled readers; we applied both univariate and multivariate analyses on this data to explore the effect of literacy acquisition on the functional organization of the brain and to investigate the possibility of developing a "literacy predicting tool". This allowed to quantitively and qualitatively compare the results yielded by the two approaches.

The thesis is structured in the following way:

- **Chapter 2** introduces the fundamentals about the physiological origin of the fNIRS signal, describes the instrumentation for signal acquisition and how the methodology compares to other neuroimaging techniques like fMRI and EEG; lastly, it describes the most common steps that need to be taken for the preprocessing of the data and outlines the concepts behind univariate and multivariate approaches for the data analysis.

- **Chapter 3** describes the first step of this research project, namely the quantitative comparison between the univariate analysis, performed by applying a General Linear Model (GLM) and a multivariate approach involving the use of Linear Discriminant Analysis (LDA) to classify single channels as "active" or "not active". To this end, both synthetic and experimental data; synthetic data, namely data with known ground-truth, allowed to obtain a quantitative comparison; experimental data allowed us to further characterize the reliability of the proposed multivariate approach. This study was published in [9].
- **Chapter 4** introduces a literacy study in which participants of various ages and literacy abilities were presented with purely auditory stimuli and illustrates the application of a channel-wise General Linear Model to detect group differences between beginning Readers and skilled Readers but also to detect a differential activation to non-words depending on the literacy levels. Such univariate analysis failed to expose the latter effect, and therefore another analytical approach was required. This work is being prepared for publication [10]
- **Chapter 5** extends the work of Chapter 4 and presents the implementation of a within-subject neural decoder and its application for the classification of multivariate distributed patterns of hemodynamic activations in response to two different classes of stimuli: words and non-words; in doing so, it compares the performances of Linear Discriminant Analysis and Support Vector Machines; individual classification accuracies were correlated with reading scores, with the expectation that high accuracies would indicate an increased effort in processing non-words, and therefore would be associated to poor reading scores. Such a correlation was found but it was not statistically significant. Finally, the Chapter discusses possible reasons for the underwhelming results. Preliminary results of this project were published in [11] and final results are being prepared for a journal publication [10].
- **Chapter 6** concludes the thesis, provides an overview of the findings of each single step of this doctoral work, represented by the three research chapters, and discusses the remaining limitations and future directions of this work.

1.3 Publications

1.3.1 First-author publications

The work in this thesis has been published in peer-reviewed journals and conferences and follows the publications listed below:

Peer-Reviewed journals

1. **Gemignani, J.**, Middell, E., Barbour, R. L., Graber, H. L. and Blankertz, B. (2018) Improving the analysis of near-infrared spectroscopy data with multivariate classification of hemodynamic patterns: A theoretical formulation and validation, *Journal of Neural Engineering*, 15(4)
2. **Gemignani, J.**, Blankertz, B., Aslin, R. N. and Pugh, K. R. Observing the effects of literacy acquisition on the brain with fNIRS. (2019) *Neuroimage* (in preparation)

Peer-Reviewed Conferences

3. **Gemignani, J.**, Bayet, L., Kabdebon, C., Blankertz, B., Pugh, K. R. and Aslin, R. N. (2018). Classifying the mental representation of word meaning in children with Multivariate Pattern Analysis of fNIRS. *Engineering in Medicine and Biology Society (EMBC) 2018 40th Annual International Conference of the IEEE* (pp. 295-298). IEEE.

Abstracts

4. **Gemignani, J.**, Middell. E., Blankertz, B. Investigating The Impact Of Individual Differences On The Accuracy Reached By Univariate And Multivariate Analysis of fNIRS Measurements, *FENS (2018, July, Berlin)*

5. **Gemignani, J.**, Middell, E., Blankertz, B. Improving the analysis of NIRS data with multivariate classification of hemodynamic patterns, *Joint Italian-French Workshop on fNIRS, Politecnico di Milano (2018,June, Milan)*

6. **Gemignani, J.**, Middell, E., Blankertz, B. From Univariate to Multivariate Analysis of NIRS data. *Neuroadaptive Technology Meeting (2017, July, Berlin)*

1.3.2 Additional contributions

Additional publications in peer-reviewed journals and conferences that were co-authored and/or that are not strictly related to the topic of the dissertation are listed below:

Peer-Reviewed journals

7. Fisher, S. P., Cui, N., McKillop, L. E., **Gemignani, J.**, Bannerman, D. M., Oliver, P. L., Peirson, S. N. and Vyazovskiy, V. V. (2016). Stereotypic wheel running decreases cortical activity in mice. *Nature Communications*, 7, 13138.

Peer-Reviewed Conferences

8. Bartkowski, C. H., **Gemignani, J.**, Barbour, R. L., Lang, K. D., von Krshiwoblozki, M., Vieroth, R., Dils, C., Jung, E. and Schmitz, C. H. (2019) Variable Source Emitter, Smart Textile Headgear Design for Functional Near-Infrared Spectroscopy. *Engineering in Medicine and Biology Society (EMBC) 2019 41th Annual International Conference of the IEEE*

Abstracts

9. **Gemignani, J.**, Barbour, R. L., Middell, E. and Schmitz, C. H. Optode design for improved optical efficiency and fast setup time on hairy head regions, *The Society for Functional Near Infrared Spectroscopy, Biennial Meeting (2016, October, Paris)*

Chapter 2

Physical principles and applications of functional Near-Infrared Spectroscopy

2.1 Physiological and biological principles of fNIRS

2.1.1 Vasculature and hemodynamics of the brain

The regulation of the blood flow in the human brain is complex and builds upon several paradigms and components [12]. In particular, there are three main regulatory mechanisms to consider: cerebral autoregulation, which is the process whereby cerebral arterioles maintain a constant blood flow in response to changing perfusion pressure; flow-metabolism coupling, which refers to the capability of the brain to change the blood flow to match the metabolic demands; neurogenic regulation, which refers to the role of the extensive network of perivascular nerves in controlling the flow. Out of these three mechanisms, the most relevant from the neuroimaging point of view is definitely the flow-metabolism coupling.

The tight correlation between brain metabolism and cerebral blood flow (CBF) was suggested already more than a century ago [13]. The main processes that contribute to the high energy need of the brain are the maintenance and restoration of ion gradients dissipated by signaling processes, like postsynaptic and action potentials, as well as the use and recycling of neurotransmitters [14].

The main energy source in the brain is glucose, and its oxidation is the main ATP (adenosine-tri-phosphate) producing mechanism [1]. While there are small amounts of glucose stored in the brain, there are no stores of oxygen, that must therefore be carried to the site of neuronal activation on demand. This demand is met by the surrounding vasculature: neuronal activation leads to arteriolar dilation, so that CBF increases and ATP can be produced [15][16].

2.1.2 The hemodynamic response

The relationship between cerebral metabolic rate of oxygen consumption (CMR_{O2}) and CBF is at the basis of functional neuroimaging. As said, the metabolic demand induces a local increase in CBF. The quantity of oxygen that is in fact removed from the blood is the arteriovenous oxygen difference, and it can be termed as Ca – Cv, where Ca is the oxygen concentration of arterial (oxygenated) blood and Cv is the oxygen concentration of venous (deoxygenated) blood. The arteriovenous difference can be defined as [16]:

$$\frac{CMR_{O2}}{CBF} = Ca - Cv \tag{2.1}$$

In an activated region, CBF increases, so the difference Ca-Cv becomes smaller. Since the blood is almost fully oxygenated at the arterial end, Ca-Cv can become smaller only if venous oxygenation is increased. In other words, the metabolically driven increase of local CBF is disproportionally larger than the increase of CMR_{O2} and this results in a local hyperoxygenation that can be observed as an increase of oxygenated hemoglobin (HbO) accompanied by a corresponding decrease in deoxygenated hemoglobin (HbR) at the venous end of the capillary [17]. This is called hemodynamic response and it can be usually observed with a latency of approximately [4-6] seconds after the onset of the stimulus that induced the activation [18]. When the neural activation is terminated, CBF decreases and both HbO and HbR return to their baseline values.

These concentration changes can be observed with functional techniques like fMRI and fNIRS. While fMRI relies on the oxygen-dependent magnetic properties of hemoglobin (Section 2.2.2), fNIRS, in turn, relies on the oxygen-dependent optical properties of

[1] To oxidate one mole of glucose , 6 moles of oxygen are needed, according to the stechiometry of complete glucose oxidation: C6H12O6 + 6O2 → 6CO2 + 6H20

hemoglobin. In fact, HbO and HbR show differential absorption spectra in the near-infrared spectral range ($\sim 650 - 950nm$), and for this reason, their relative concentration changes can be separately detected by illuminating the brain with specific wavelengths within this range.

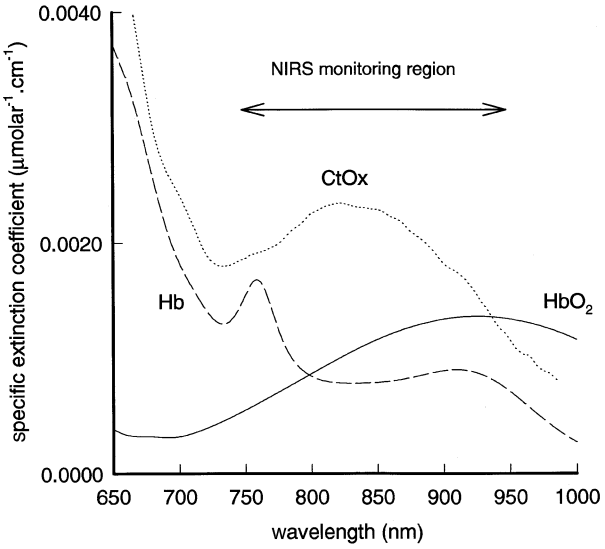


FIGURE 2.1: Extinction coefficients for HbO ($HbO2$), HbR (Hb) and cytochrome oxidase ($CtOx$). Figure adapted from Delpy and Cope, 1997 [19].

2.1.3 Instrumentation

From the technology point of view, fNIRS is most commonly implemented in one of three spectroscopic methods: Continuous Wave (CW-NIRS), Frequency Domain (FD-NIRS) and Time Domain (TD-NIRS).

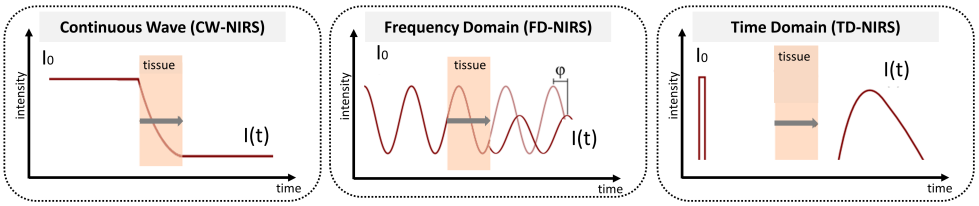


FIGURE 2.2: Different technological approaches for fNIRS

- **Continuous Wave (CW-NIRS)**

  The Continuous Wave (CW) method relies on the illumination of tissue with light of constant amplitude, and the detection of the transmitted light intensity, allowing therefore to assess the overall light attenuation inside the tissue. This method cannot differentiate between scattering and absorption effects, and can therefore only quantify concentration changes relatively to a baseline, rather than absolute values (see Section 2.1.6 for more details).

- **Frequency Domain (FD-NIRS)**

  In this technique the light source is intensity-modulated in the radio frequency range ($\sim 100MHz$). The propagating waves are then measured by the detector. This allows FD instruments to extract, besides the intensity attenuation, also two more independent quantities: a phase shift (ϕ) and the decay of modulation depth (ratio of AC to DC component). These quantities are affected differently by absorption and scattering of the tissue, therefore with FD systems it is possible to distinguish these parameters and compute absolution concentration levels.

- **Time Domain (TD-NIRS)**

  Time Domain measurements utilize picoseconds pulses of photons to irradiate the tissue and fast responding detectors to register the shape of the light pulse as it exits the tissue. The properties of the received photon distribution such as area under the curve, the time of maximum, and width allow assessment of the tissues absorption and scattering properties.

In terms of costs and complexity of the instrumentation, FD and TD systems are definitely more expensive and cumbersome than CW systems, which on the contrary can be miniaturized and made portable. Therefore, since quantification of absolute values is often not essential for most neuroscientific applications, the great majority of the fNIRS-based research utilizes CW systems.

2.1.4 Acquisition of the signal

To acquire the fNIRS signal, light sources and detectors are necessary. A source-detector pair constitutes a "channel", and the difference between the light emitted

through the source and that measured at the detector location represents the changes of the optical properties of the tissue underlying that channel.

Light in the near-infrared spectral range can propagate relatively deeply into the biological tissue, including the skull, because in this range light is only weakly absorbed by it (Figure 2.1). The selection of the precise wavelengths depends on several factors [20] but the pair 780 and 830 nm is most often used.

When light is emitted through the source, it diffuses in all directions through the tissues of the head (scalp, skull, cerebrospinal fluid); part of the light is scattered, part is absorbed by the chromophores. Although there are several chromophores that absorb light, the only ones that exhibit oxygenation-dependent absorption in the NIR region are HbO, HbR and cytochrome oxidase (CytOx), found in the cell mitochondrial membrane and involved in the oxidative metabolism of glucose.

On the other hand, the predominant factor of light transport in tissues is scattering, since the transport scattering coefficient μ_s of most tissues is much larger than their absorption coefficient μ_a, in the NIR region [19]. This introduces an unknown light loss but also results in a non-linear relationship between absorption changes and the resulting attenuation changes.

The fact that light is both absorbed and scattered coupled with the fact that the traveled tissues are inhomogeneous and anisotropic make it difficult to exactly determine the path followed by the light.

However, if we simplify the model and we assume the medium is homogenous, then we can model this path by only taking into account the phenomena of absorption and scattering. Several studies by Okada et al. were conducted on the propagation of light in simplified media (homogeneous or semi-homogeneous) and it was possible to formulate the Photon Measurement Density Function, or as it is most commonly called, the "banana-shaped" path [21, 22, 23]. This function is a three-dimensional model representing the probability that a photon travels through a certain optical path and allows therefore to determine the brain volume sampled by a given source-detector pair. The depth reached by the photons depends on the source-detector distance and it is usually approximately one half of it [24].

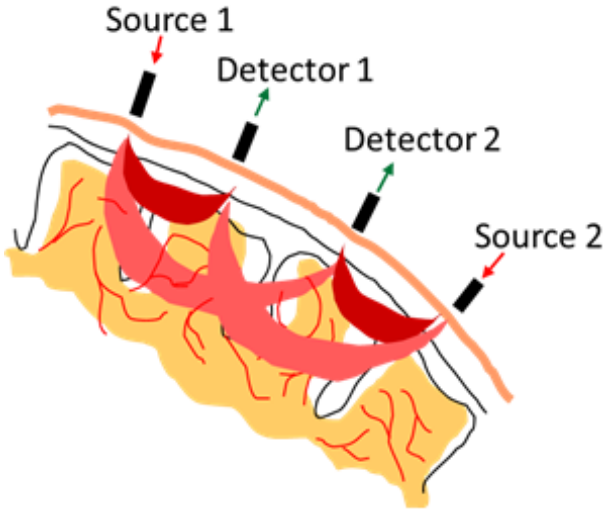


FIGURE 2.3: Each source-detector pair constitutes a NIRS "channel"; the light emitted from the source and measured by the detector follows the probability model given by the banana-shaped path. This path represents the probability of sampling a given optical path. In this case, the pairs source1-detector2 and source2-detector1 allow penetration of the light into the cortex, while the pairs source1-detector1 and source2-detector2 are close enough to only allow penetration through the shallow layers (*short-distance channels*). Figure is taken from www.nirx.net

2.1.5 The fNIRS signal: how the raw data looks like

The photon detectors detect light and convert it into a voltage signal, whose amplitude is proportional to the light intensity. The fNIRS signal can be looked at as a combination of several components, that can be distinguished according to the physical source of the signal, that can be cerebral or extracerebral, the relation of it with a task, namely it can be evoked by the task or independent on it, and its physiological origin, that can be neuronal or systemic, as exemplified in Figure 2.4. This model has been described by recent works of Tachsidis and Scholkmann ([20][25][26]), among others.

The components of the signal originated in the cerebral compartment can be either neuronal or systemic and can be either evoked by the task or spontaneous (not evoked). What the experimenter usually aims at is to unambiguously identify the cerebral neuronal component, i.e. the signal resulting from the neurovascular coupling, but the changes associated with it are relatively small compared to the overall variability of

the signal ($\sim 0.5\mu$M for HbO, $\sim 0.2\mu$M for HbR, compared to $\sim 1\mu$M for the whole signal for both HbO and HbR, [27]).

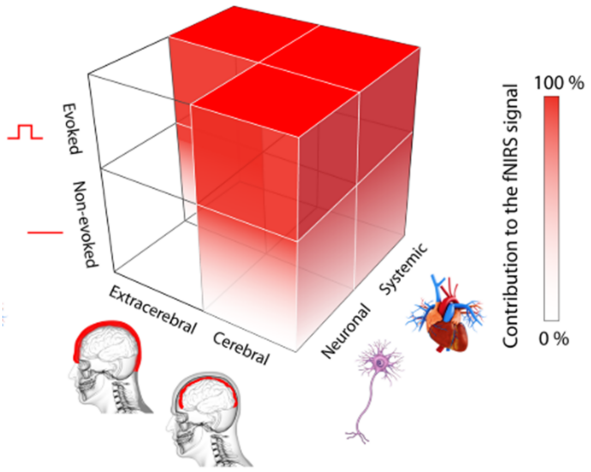


FIGURE 2.4: Composition of the fNIRS signal. Figure is taken from Tachsidis and Scholkmann, 2016 [26]

Besides the signal of neuronal origin, the cerebral component also contains a systemic element, that is sometimes referred to as "physiological noise" and that accounts for changes in blood pressure, cerebral blood flow, cerebral blood volume, heart rate and respiration. Although these elements are always physiologically present and spontaneous, and therefore non-evoked, it is also true that the execution of a task can significantly modulate their characteristics, like amplitude and frequency.

All these components are reflected and can be clearly observed in the rich frequency spectrum of the fNIRS signal, that shows clear contributions by the heartbeat ($\sim$1 Hz), the respiration ($\sim$0.3 Hz), the low-frequency oscillation also known as "Mayer wave" ($\sim$0.1 Hz) and very-low-frequency oscillations of different origin ($<$ 0.1 Hz) (Figure 2.5)

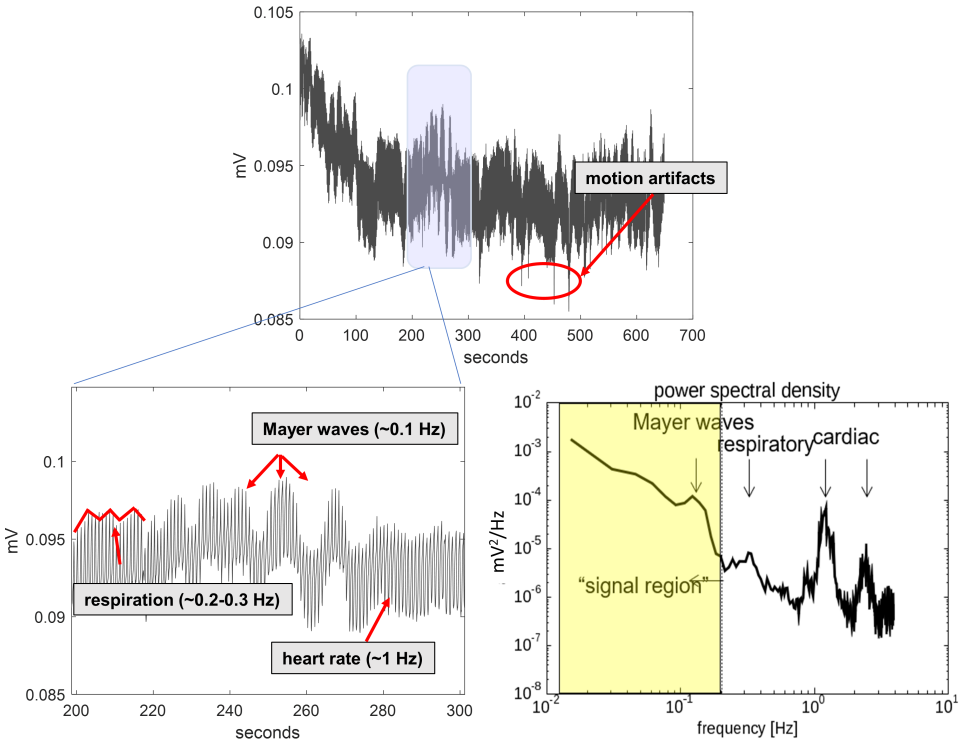


FIGURE 2.5: **(Top and bottom left)**: Example of the raw signal of light intensity. In this timetrace the different components (heartbeat, respiration, Mayer waves) are clearly recognizable, as well as a few spike artifacts, probably caused by the displacement between one of the optodes and the skin. **(Bottom right)** Typical frequency spectrum of the fNIRS signal

2.1.6 From light intensities to concentration changes: the modified Beer-Lambert Law

After the light has been measured by the photon detector and converted into a voltage signal, it is necessary to proceed and convert this raw signal into relative concentrations changes of the chromophores of interest. This is made possible by the *modified Beer-Lambert Law*.

Here we will focus only on the case where two wavelengths are used and the chromophores of interest are HbO and HbR, but of course, the same math applies when more wavelengths are used: in that case, it is possible to resolve the concentration changes of more chromophores, like that of the cytochrome oxidase.

The way light is transported through a homogeneous, non-scattering tissue, is modelled by the Beer-Lambert Law:

$$I(\lambda) = I_0(\lambda) \cdot e^{-\mu_a(\lambda) \cdot L} \tag{2.2}$$

where $I_0(\lambda)$ is the intensity of light emitted by the source, I(λ) is the intensity of light detected by the detector, $\mu_a(\lambda)$ is the absorption coefficient of the medium at the wavelength λ; L is the optical pathlength and it represents the distance travelled by the photons through the medium. The ratio between I and I_0 defines the *optical density of attenuation*:

$$OpticalDensity = -log(\frac{I}{I_0}) = \mu_a(\lambda) \cdot L \tag{2.3}$$

As said, the Beer-Lambert equation would hold true in case the medium was non-scattering, so when the light is only absorbed. But when light undergoes scattering, the distance travelled by the photons is no longer L, because the path followed by the photons becomes random, and photons will take more time to exit the tissue, and they will have more probability to be absorbed.

Furthermore, scattering introduces a certain amount of losses, which depends on the geometry and on the scattering coefficients of the involved tissues. The *modified Beer-Lambert Law* (mBLL) was formulated to take account scattering:

$$OpticalDensity = \mu_a(\lambda) \cdot d \cdot DPF(\lambda) + G(\lambda) \tag{2.4}$$

The optical pathlength, L, is replaced by $d \cdot DPF(\lambda)$, where d is the distance between source and detector and DPF is the Differential Pathlength Factor, that accounts for the increased path followed by the photons as a result of scattering. The term $d \cdot DPF(\lambda)$ corresponds to the mean light propagation distance in the medium; basically, the DPF is a scaling factor that indicates how many times more than d the detected light has travelled. $G(\lambda)$ accounts for the light losses due to scattering and, as the $DPF(\lambda)$, depends on the wavelength λ.

This equation can be used to calculate the chromophores concentrations, because the absorption coefficient of any chromophore, μ_a, can be written as the product of its

concentration, c, and its extinction coefficient, $\epsilon(\lambda)$:

$$\mu_a = c \cdot \epsilon(\lambda) \tag{2.5}$$

Therefore, the modified Beer-Lambert Law can be written as:

$$OpticalDensity = c \cdot \epsilon(\lambda) \cdot d \cdot DPF(\lambda) + G(\lambda) \tag{2.6}$$

where c is our unknown of interest: the chromophore concentration. In the real-world scenario, this equation is very difficult to be resolved exactly, because both $DPF(\lambda)$ and $G(\lambda)$ are not constant and depend on the amount of scattering. However, under the hypothesis of constant scattering, they can be considered constant. Concentration changes of the chromophores can therefore be calculated by measuring a change in optical density:

$$\Delta OD = OD_t - OD_{t_0} = \Delta c \cdot \epsilon(\lambda) \cdot d \cdot DPF(\lambda) \tag{2.7}$$

Where OD_t is the optical density measured at time t and OD_{t_0} is the optical density measured at time t_0. If we apply this equation to the two chromophores of our interest, HbO and HbR, the mBLL becomes:

$$\Delta OD = (\Delta[HbO] \cdot \epsilon_{HbO}(\lambda) + \Delta[HbR] \cdot \epsilon_{HbR}(\lambda)) \cdot d \cdot DPF(\lambda) \tag{2.8}$$

Where $\epsilon_{HbO}(\lambda)$ and $\epsilon_{HbR}(\lambda)$ are the extinction coefficients of HbO and HbR, respectively, at the wavelength of choice. By using at least two different wavelengths λ_1 and λ_2 it is possible to resolve the two unknowns, $\Delta[HbO]$ and $\Delta[HbR]$:

$$\Delta[HbO] = \frac{\epsilon_{HbO}(\lambda_1) \cdot \frac{\Delta OD(\lambda_2)}{DPF(\lambda_2)} - \epsilon_{HbR}(\lambda_2) \cdot \frac{\Delta OD(\lambda_1)}{DPF(\lambda_1)}}{(\epsilon_{HbR}(\lambda_1) \cdot \epsilon_{HbO}(\lambda_2) - \epsilon_{HbR}(\lambda_2) \cdot \epsilon_{HbO}(\lambda_1)) \cdot d} \tag{2.9}$$

$$\Delta[HbR] = \frac{\epsilon_{HbO}(\lambda_2) \cdot \frac{\Delta OD(\lambda_1)}{DPF(\lambda_1)} - \epsilon_{HbR}(\lambda_1) \cdot \frac{\Delta OD(\lambda_2)}{DPF(\lambda_2)}}{(\epsilon_{HbR}(\lambda_1) \cdot \epsilon_{HbO}(\lambda_2) - \epsilon_{HbR}(\lambda_2) \cdot \epsilon_{HbO}(\lambda_1)) \cdot d} \tag{2.10}$$

The extinction coefficients are represented in Figure 2.1. As for the DPFs, several works have been published over the years, attempting at tabulating values for different ages, tissues and wavelengths.

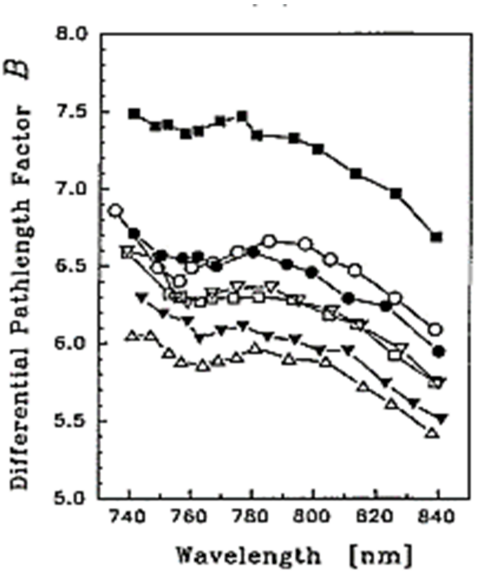


FIGURE 2.6: Tabulated DPF values from Essenpreis et al. 1993. The different timetraces correspond to different subjects (1 female, 6 males, median age 28 years old). It is clear that, despite the intersubject variability, the DPF exhibits a descending trend as the wavelength increases.

The DPF does not depend only on the wavelength but also on the absorption and scattering coefficients of the tissue; in particular, it increases with μs and decreases with μa. Furthermore, it depends on d, but this only for small d values: for $d \geq 2.5$ cm, the DPF is virtually independent on d [28]. In 1993, Essenpreis et al. [29] determined the DPFs on the adult head, forearm, calf and infant head, between 740 and 840 nm, using in vivo and post mortem measurements (Figure 2.6). In a recent paper, Scholkmann and Wolf [28] introduced a formula to calculate the dependency of the DPF on the age of the measured subject.

Finally, it is important to point out that CW-NIRS systems are not able to measure absolute concentration levels of the chromophores, but only concentration changes relative to a baseline. This is because, although the light emitted from the source

in principle could be determined, due to complex boundary conditions not all the light that is emitted actually enters the tissue. Since CW-NIRS systems are unable to quantify the optical properties of the tissue, in particular the scattering coefficient, it is not possible to know the incident light I_0 and use it in the mBLL. Instead, these systems use another reference level, a baseline, and express the chromophores' concentration changes relative to that.

2.2 Other neuroimaging techniques

fNIRS is only one of the many neuroimaging techniques available. Methods to investigate brain physiology include other techniques that rely on the hemodynamics of the brain, such as functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET). Also, brain physiology can be explored in terms of electrical activity, as it is the case with Electroencephalography (EEG), and magnetic activity, as it is the case for Magnetoencephalography (MEG). The strengths and limitations of these methods are listed in Table 2.1.

Property	EEG	MEG	fNIRS	fMRI	PET
Measurement	Electrical Activity	Electrical Activity	$\Delta[HbO]$, $\Delta[HbR]$, $\Delta[HbT]$	BOLD	Blood flow
Temporal Resolution	1 ms	1 ms	100 ms	2-5 s	1 s
Spatial Resolution	$> cm$	$< cm$	$< cm$	mm^3	mm^3
Cost	€/€€	€€€	€/€€	€€€	€€€
Required Immobility	low	high	low	high	high

Table 2.1: Characteristics of different neuroimaging modalities

In this paragraph, we will focus on the comparison between fNIRS and its closest – in terms of outcome – neuroimaging technique: fMRI, and its closest in terms of setup: EEG. Furthermore, a brief review of the research fields where fNIRS is currently employed will be presented, with particular attention to research studies including young subjects and the study of language development in children, the research topic of the grant under which this doctoral project was developed.

2.2.1 EEG and fNIRS

The most known neuroimaging technique is probably the Electroencephalography (EEG). EEG is a well-established modality that allows to non-invasively record electrical activity at scalp level, by using electrodes. The electrical activity is a measure of the flow of extracellular currents produced by the summation of the activity of thousands of neurons. The contribution of a single neuron cannot, therefore, be resolved at scalp level. Additionally, the skull is a semi-conductive medium and has a smearing effect on the signal: all these factors make it hard to determine the localization of the sources (*inverse problem*).

Nevertheless, EEG is very useful in that many cognitive processes and states of consciousness are reflected in electrical activity within specific frequency bands. Moreover, EEG can be used for the study of event related potentials (ERP), that are stereotypical electric responses resulting directly from a specific cognitive or motor stimulation. From the technical standpoint, EEG has a poor spatial resolution in comparison to fNIRS but an excellent temporal resolution: normally the sampling frequency is around 250 Hz but some modern systems can reach rates of 20 kHz, while for fNIRS that can be at most in the hundreds of Hz. They are both perfectly portable, allowing therefore to conduct a whole range of studies that require the subject to move freely, for example motor rehabilitation studies [30].

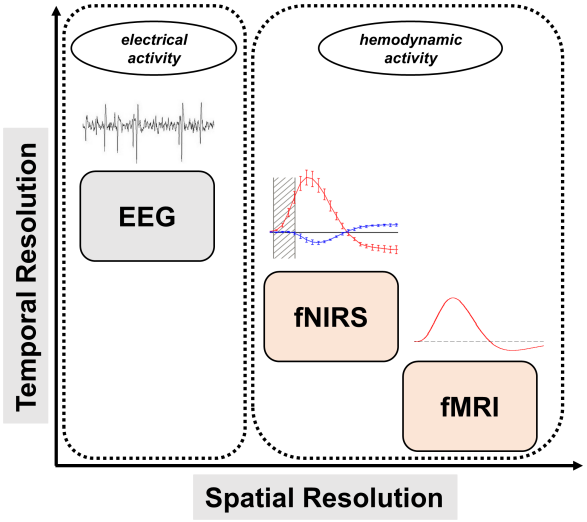


FIGURE 2.7: EEG, fNIRS and fMRI in terms of spatial and temporal resolution

Apart from these purely technical considerations, a meaningful comparison between the two modalities cannot really be drawn: they provide insights on different aspects of the neurophysiological processes under investigation, carrying complementary information that can become really meaningful when taken into account together [31][32]: in fact, it is the case that in recent years the integration of the two modalities has been gaining increasing interest and recognition, in several research fields. An example is the field of Brain-Computer Interfaces (BCIs): while until a few years ago BCIs were mostly EEG-based [33], recently Fazli and colleagues [34] demonstrated that a BCI based on the simultaneous acquisition of EEG and NIRS can achieve augmented classification performances, by taking advantage of each modality's individual strength, namely the fast responsiveness of the EEG and the information about the vascular response provided by the NIRS signal. In their study, they presented for the first time an EEG-fNIRS based Sensory Motor Rhythm (SMR) BCI and showed that the combination increased the classification of 5%, on average, with respect to using only EEG. Also in cognitive neuroscience simultaneous EEG-NIRS measurements are widely used, for example in the research areas of language processing [35][36] and epilepsy [37], just to name a few.

2.2.2 fMRI and fNIRS

fMRI is also a non-invasive technique, that builds on the MRI scanning technology. The MRI signal comes almost entirely from the protons of water of the tissues (water makes up for almost 65% of the body weight).

Protons have their own electric charge and have their own spin, i.e. they rotate about their own axes, resulting in a magnetic dipole that lies parallel to the axis of the nucleus. Normally, these dipoles are randomly oriented, but during an MRI scan a strong and static magnetic field B_0 (typically ≥ 3 Tesla) is applied, causing the dipoles to align with its direction. The summation of all the magnetic moments of the aligned dipoles is the magnetization vector M, and has the same direction of B_0.

When a brief radio-frequency pulse is applied, with a frequency equal to the resonating frequency of the rotating protons, then the phenomenon of magnetic resonance occurs: protons can change energy state and the magnetization vector M can rotate away from equilibrium. When the pulse terminates, protons go back to equilibrium by releasing the absorbed energy as a radio-frequency wave, with a process called *relaxation*, and producing a measurable signal.

With MRI it is possible to investigate the levels of oxygenation of the brain, thanks to the different magnetic properties showed by oxygenated and de-oxygenated hemoglobin; in fact, while oxy-hemoglobin is diamagnetic, and has therefore a little magnetic moment, deoxy-hemoglobin is paramagnetic, and produces a large magnetic moment. Large concentrations of deoxy-hemoglobin cause a reduction in the relaxation time of the tissue and therefore a decrease of the measured signal. Correspondingly, when a brain area activates in response to a certain stimulus, deoxy-hemoglobin decreases and the signal increases. This phenomenon is the Blood-Oxygen Level-Dependent (BOLD) effect, and this is the reason why the fMRI signal is also called BOLD signal[38].

It is fairly easy to recognize that fNIRS and fMRI are two different roads for the same destination: while fMRI relies on the oxygen-dependent magnetic properties of hemoglobin, fNIRS uses its oxygen-dependent optical properties. However, the two methods present some significant technical differences. First of all, the BOLD signal

reflects only the contribution of deoxy-hemoglobin, while fNIRS can resolve the concentration changes of oxy-hemoglobin too (and also of other chromophores, provided that the instrument can use more than two wavelengths).

The spatial resolution of fMRI is better than that of fNIRS, that is in the order of cm. However, the temporal resolution of fMRI is quite poor, and that is the cost for having such a good spatial resolution: due to the sequential acquisition of slices (30-100 ms each), it takes about 1-3 seconds (0.5 – 1.3 Hz) to achieve a whole-brain coverage at the spatial resolution of 3-5 mm [39], [40]. On the other hand, the fNIRS technical sampling frequency can be much higher, in the order of 100 Hz, depending on the technology used. However, it should be noted that in both cases of fMRI and fNIRS, the measured signal is the hemodynamic response, and this is naturally slow, showing a peak at around 4-6 seconds after stimulus and lasting for at least 20 seconds.

Aside from this technological comparison, practical considerations are really crucial when it comes to comparing fMRI and fNIRS: first of all, the fNIRS setup is relatively quick and easy, just like the EEG's, and an fNIRS experiment can be performed anywhere, while fMRI has some significant physical constraints: the scanner is cumbersome and can't be used outside its specific facility; both its purchase and maintenance are expensive; during an experiment the subject must lay very still inside to avoid the occurrence of motion artifacts; no one is allowed in the room where the scan takes place, and the scanner is also very noisy.

Especially the two latter features make fMRI inconvenient for babies and children, as well as for certain groups of adult subjects, depending on their pathologies or disabilities, and make fNIRS stand out: and fNIRS measurement is completely silent and allows a whole range of studies involving the administration of auditory stimuli, that could not be performed during an fMRI scan, at least not in ecological conditions. Also, because it is silent and can be performed anywhere, it is definitely more tolerable for children and babies, that can undergo an experiment while being around parents, for example, if the experiment allows so.

Another field where fNIRS is naturally preferred over fMRI is the field of motor and rehabilitation studies: whether the goal is measuring the functional effect of a rehabilitation task or the simple monitoring of a sustained movement, the portability of

the technique and the tolerance to motion artifacts clearly represent a key advantage [41][42].

Lastly, fNIRS techniques do not interfere with signals of other nature: this feature makes it ideal for multimodal integration with other modalities like EEG, as already mentioned, but also MRI: in this case, it would be possible to acquire structural information and localize anatomically the hemodynamic activity measured optically.

2.3 The use of fNIRS in language research on children

fNIRS is now used in a wide variety of fields in neuroscience, from neurology [43] to psychiatry [44], cognitive neuroscience [45], to basic research. Its versatility and relatively low-cost make it a useful technique for the investigation of many research questions. For the scope of this dissertation, we will focus on the use of fNIRS on children for language studies.

As already articulated, fNIRS proves to be particularly convenient for use on children: from a very practical point of view, children do not necessarily need to be isolated in order to be tested, as it is the case for fMRI. An fNIRS instrument is easily portable, therefore an experiment can be conducted in a comfortable environment where the child can feel at ease. The setup itself is easy and quick, and after the cap has been prepared the first time, it can be just taken off a subject and put on another subject with very little preparation needed, especially for children and infants, since they have little and thin hair.

Additionally, fNIRS is less sensitive than fMRI and EEG to motion artifacts and this allows to test freely behaving participants. Artifacts can certainly occur, especially when testing infants, as a result of bad coupling between the optodes and the skin following an abrupt movement of the head; the result is generally a spike artifact that is not very destructive of the underlying signal, and in many cases this can be recovered, provided that appropriate motion artifacts removal algorithms are in place. For a review and comparison of these algorithms, we refer the reader to [46].

Thanks to all these features, fNIRS can be used at any age and it is particularly employed on newborns; this has made possible to make giant progress in understanding the mechanisms by which the brain processes language and how this processing evolves throughout the whole span of the brain development, to the point that was even possible to infer, from fNIRS measurements at birth, the effects of prenatal experience on later language abilities and to document an innate hemispheric lateralization for speech processing [47, 48]. It also made possible to broadly localize the functional areas linked to the perception of other kinds of stimulation, like music, odours, particular visual patterns.

For a review on the use of fNIRS in developmental language research we recommend [49, 50, 51].

It is therefore clear that the silenceness of the technique proves extremely useful in language research, because allows the presentation of auditory stimuli in a non-intrusive environment.

Auditory stimuli are not used only in the context of developmental neuroscience of language, but also to the investigation of literacy abilities: in fact, findings suggest that the very way our brain processes auditory information is affected by the degree of literacy, since learning to read strengthens the neural pathways associated with phonological processing skills, that are otherwise weakened in illiterate subjects [52].

2.4 Analysis of fNIRS data

As already pointed out in a previous section, a raw fNIRS timetrace is a voltage signal representing the light intensity measured by each detector. This timetrace includes the contribution of several phenomena that are reflected in hemoglobin relative concentration changes, and the cognitive processes underlying the specific research question is only one among them. Furthermore, raw data may contain motion artifacts, in the form of shifts and spikes. All these considerations, related to quality and to physiological confounds, needs to be dealt with before starting the statistical analysis of the data. Therefore, we review in this section the most relevant state of the art pre-processing steps.

2.4.1 Reduction of motion artifacts

Motion artifacts are the result of a momentary loss of optical coupling between the scalp and the optodes, resulting in a brief and sharp wave superimposed to the signal; even when the coupling is restored, it is invariably different from before, and this results in a shift of the level of the signal reflective of a change of the baseline caused by the altered coupling.

The constant advances in hardware provide increasingly efficient ways to prevent this uncoupling from occurring; this is the case for example of spring-loaded fiber holders from NIRx (Figure 2.8).

FIGURE 2.8: Spring-loaded fiber holders. The spring keeps the tip of the optode at constant pressure, reducing the possibility of loss of contact between the tip and the scalp.

Nevertheless, motion artifacts still occur; many methods have been proposed to reduce and correct motion artifacts. Some of them require additional hardware, for example a short-distance fNIRS channel or an accelerometer, but the majority of the proposed methods seek to detect artifacts by relying only on changes of amplitude and frequency of the data corresponding to the occurrence of an artifact. Among these approaches, we will briefly describe the most popular ones. For a more complete review, we recommend the reader [46, 53].

Principle Component Analysis (PCA)

The correction of motion artifacts via PCA was first proposed by Zhang et al [54]. The idea behind it is that motion artifacts account for a large proportion of variance of the data, due to their large amplitude; therefore, when decomposing the data into uncorrelated principal components, they should be represented in the first M components, since these are organized in descending order of explained variance. Therefore, motion artifacts can be removed by removing the first M components and reconstructing the "clean" signal.

Wavelet Filtering

The motion artifact correction based on the use of wavelets was introduced by Molavi and Dumont [55]. The timetrace of each channel is decomposed into L levels by applying a discrete wavelet transform; for each level, a series of coefficients are obtained for *detail* and *approximation*. The proposed method relies on the assumption that the measured signal is a linear combination of the signal of interest plus motion artifacts and that the distribution of the detail coefficients accounting for the signal of interest is centered around zero and has low variance, while the detail coefficients corresponding to artifacts will appear as outliers in this distribution; therefore, a threshold of variance can be set on the probability distribution and coefficients exceeding the threshold are set to zero. The timetrace can be then reconstructed with the inverse discrete wavelet transform.

Spline Interpolation

The method based on spline interpolation was first proposed by Scholkmann et al [56]. In this procedure, a moving window is employed for the calculation of the standard deviation of the signal; a motion artifact is identified if, within the window, the standard deviation is greater than a certain threshold, predefined by the user. The timetrace is segmented and each segment containing an artifact is spline interpolated. Finally, the timetrace is reconstructed.

Transient Artifacts Reduction Algorithm (TARA)

TARA was introduced by Selesnick et al [57]. The method models the timeseries as a combination of a low-pass signal, an artifact of each type (spikes and shifts), and white noise. The estimation of the different components is formulated as an optimization problem. After shifts and spikes are estimated, they are removed from the signal.

2.4.2 Physiological confounds

As already described, the fNIRS signal of interest is significantly contaminated with global components arising from superficial layers of the scalp. A certain amount of these component is task-independent, and could therefore be modelled, in principle; however, the contribution to the signal arising in the superficial-layers is also modulated by the task, to some extent. This happens because, depending on the type of stimuli that are administered, they can produce significant changes in systemic variables, like heart-rate, blood pressure, respiration rate.

Many algorithms have been proposed to overcome this issue but one of the main limitations of any of them is that they rely uniquely on theoretical assumptions on the features of the confounding signal. For example, several studies proposed to decompose the signal via PCA and reconstruct the signal of interest by removing the first N components, under the assumption that they contain mainly the global contribution of the superficial layers [54, 58]. This kind of assumptions may yield adequate results when dealing with motion artifacts, since they are discrete events in the data, but they are at high risk of approximation when it comes to estimating the different sources of the signal and the extent of contribution given by each of them.

While this issue is definitely still an open research question, it is certainly true that the best way forward is to combine fine analysis tools with appropriate hardware means, for example multi-distance measurements.

Multi-distance channels allow to measure the signal at different depths, and therefore to obtain a signal from the superficial layers (Figure 2.9). The optimal short-distance separation was recently determined to be 8.4 mm for adults and 2.15 for term age infants [59].

Short-distance signals can then be regressed out from the whole signal. To this end, several methods have been introduced: one possibility is to employ them during the first-level statistical analysis as additional regressors of a General Linear Model (GLM) [60]; another approach is to use them in combination with Independent Component Analysis (ICA) [61].

In any case, the ideal amount of short-distance measurement to be removed from the corresponding long-distance measurement is still an open question. Saager and

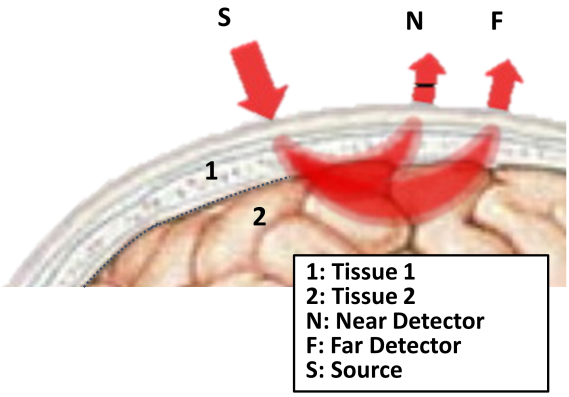


FIGURE 2.9: Multi-distance measurement; by placing a detector at a shorter distance, light measured by that detector will only come from the shallow layer

Berger [62] proposed a least-squares minimization procedure to fit the short-distance measurement to the long-distance measurement, and to use the result of the fit as a weighing coefficient.

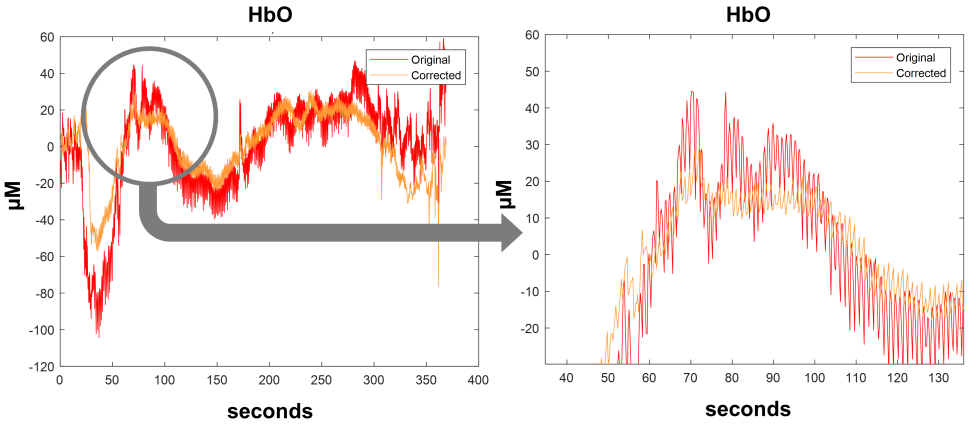


FIGURE 2.10: Example of application of the algorithm proposed by Saager and Berger [62]. On the right side of the figure, it is clearly visible that the ∼0.1 Hz fluctuations ("Mayer waves") are removed with the short-distance correction

2.4.3 Frequency filtering

Frequency filtering is the most straightforward method to eliminate components of the signal that are not of interest, for example those associated with the heart beat ($\sim$1 Hz) and the very-low frequency fluctuations (drifts $<$0.1 Hz). For this reason, the use of band-pass filters has been widely used in literature, usually with lower cut-off around 0.01 Hz and upper cut-off around 0.5 Hz; of course, the frequencies being removed should not include the "frequency of stimulation", i.e. the frequency at which hemodynamic activation is expected based on the stimulation timings: for example, with a task of duration 10 s, and a pause of 30 s, the frequency at which the task-specific activation is expected is 1/40 s = 0.025 Hz. A band-pass filter should be designed that does not exclude this frequency.

This is also the reason why band-pass filters can not generally be used to remove other kinds of unwanted features of the signal, such as Mayer waves or respiratory oscillations: their frequencies normally overlap with (or are very close to) the frequency of stimulation. Also, caution should be used when designing band-pass filters: the use of overly narrow bands could produce artifacts in the signal in the form of ripples, or could remove the hemodynamic responses that one is looking for.

Finally, also the filter choice requires special attention: Finite Impulse Response (FIR) filters are more stable than the Infinite Impulse Response (IIR) and the phase delay they introduce is constant across the whole frequency range, while IIRs introduce a frequency-dependent phase shift; on the other hand, IIRs are generally more computationally efficient.

In both cases, the phase-distortion can be corrected by shifting back in time the filtered signal of the delay amount (in case of FIRs) or by filtering the time series forward and backward (in case of IIRs). This latter option is named 'zero-phase filter' and it is widely used in fNIRS pre-processing [63]. Zero-phase filters are non-causal, in that they use samples not only from the past and present inputs but also from the future ones: as a consequence, they can be employed only as an offline operation, while real-time filtering has to be performed with causal filters.

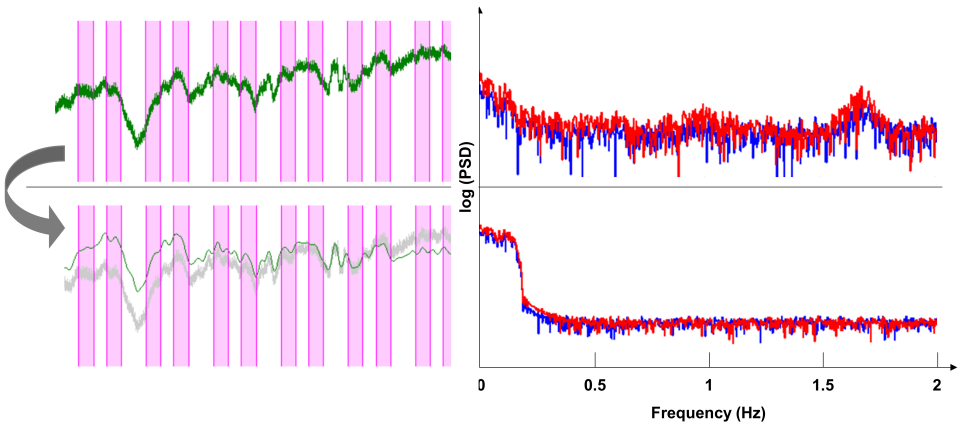


FIGURE 2.11: Effect of applying a band-pass filter [0.01-0.2 Hz]; the data has sampling frequency 4.4 Hz and the magenta parts represent the task periods (10 s, with a jittered pause of 19-26 s). On the PSD graphs, the red line is used for the PSD of the first wavelength and the blue one for the second wavelength (in this case 760 and 850 nm)

2.4.4 Computation of concentration changes

After the data has been pre-processed, relative concentration changes can be computed through the modified Beer-Lambert Law (mBLL). The formulation of the mBLL has been already described in detail in Section 2.1.6.

From a practical perspective, there is an important difference between the theory and how the algorithm is implemented in most of the available software tools. As illustrated in Section 2.1.3, CW-NIRS instruments – which represent the great majority of the fNIRS instruments employed in research - are unable to quantify the amount of incident light; therefore, some other reference signal is needed to compute the concentration changes: usually, the mean of the signal over the whole course of the study is used as a baseline, in place of the incident light. This is why CW instruments are not able to compute absolute values of concentration changes but changes relative to an arbitrary baseline.

As mentioned, the implementation of the mBLL, as well as of the other processing steps, is now offered by many available software solutions, like Homer2 [64], the NIRS Brain AnalyzIR Toolbox [65] and nirsLAB [66], just to name a few.

2.4.5 Statistical inference

Any analysis workflow is unique and depends on the specific research question that is being investigated. However, there is a general consensus around the necessary pre-processing steps that lead from the intensity measurements to having chromophores concentration changes.

The roadmap is not as well marked for the statistical inference. This is partially because the way data is statistically assessed depends on the question being investigated, but also because for fNIRS there is not yet a clear gold standard. Therefore, many different methods can be found in literature [67].

However, due to the similarity to the BOLD signal, the fNIRS signal has been oftentimes analysed for a long time with the same methods, namely by modelling the signal with a General Linear Model; the hypothesis of using a GLM for this purpose is that the signal can be seen as a superimposition of hemodynamic responses, elicited by the experimental conditions, and physiological oscillations.

Although the measured timeseries Y is really discrete, it is useful to treat it as a function of time, Y(t) [4, 68]; a GLM seeks to express the concentration timeseries Y(t) as:

$$Y(t) = X(t)\beta + \epsilon(t) \tag{2.11}$$

where X(t) is the design matrix, with regressors in each column; β is a vector containing the coefficients of the regressors, to be estimated; finally, $\epsilon(t)$ is the error term, having variance σ^2:

$$\epsilon(t) \sim N\left(0, \sigma^2\right) \tag{2.12}$$

If each regressor contains the timings of each experimental condition, the resulting β for that regressor represents therefore the impact of that condition on the measured data.

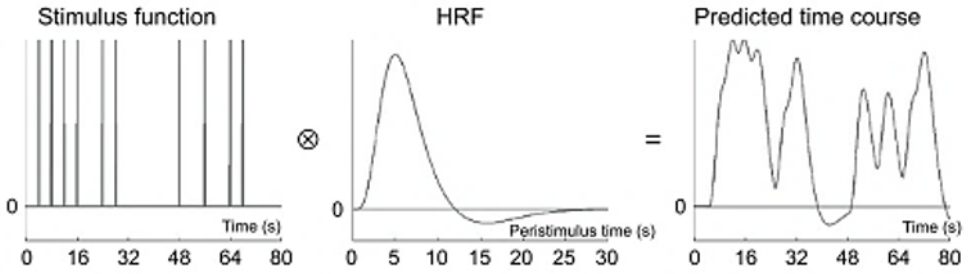


FIGURE 2.12: Each column of the design matrix X(t) is the result of the convolution of the stimulus vector of each experimental condition, u(t) (**left**), with a "basis function" representing the expected hemodynamic response elicited by that condition (**middle**). The result is a theoretical time course that is completely predicted by the administered stimuli (**right**), and this is the timetrace that is then fitted to the measured data. The more the condition actually had an impact on the data, the better will be the fit, in terms of amplitude of the corresponding ß. Figure is taken from [4]

The design matrix

Each column of the design matrix X encodes the timings of each experimental condition; in particular, a convolutional model is employed to express the corresponding column in X as the convolution of the stimulus vector u(t), usually a stick function containing the stimulation timings, with the expected hemodynamic response function (HRF) h(t) elicited by that stimulation:

$$X(t) = u(t) \otimes h(\tau) = \int_0^T u(t-\tau)h(\tau)d\tau \tag{2.13}$$

where τ indexes the peristimulus time(PST), over which the HRF is expressed. The validity of this model relies on the assumption that the hemodynamic response can be modelled as a linear time-invariant (LTI) process [69]: thus, it has a finite duration, it is independent of time and responses to successive inputs superimpose in a linear way.

The result of convolving a stimulus vector u(t) with a canonical HRF is shown in Figure 2.12.

The simplest and most commonly used basis function is the canonical HRF. The canonical HRF is a model characterized by two gamma functions, one modelling the peak and one modelling the undershoot. It is usually parametrized by a peak delay of 6 s and an undershot delay of 16 s, with a peak-undershoot amplitude ratio of 6.

However, the precise shape of the real HRF may vary over brain regions and over individuals; therefore, it is possible to choose a basis function different from the canonical HRF in order to accommodate this variability (Figure 2.13). Another possibility is to use the canonical HRF together with its partial derivatives. This set is known as *informed basis set* and it is used to capture variations of the peak delay and dispersion: the temporal derivative accounts for shifts in the latency of the response, while the dispersion derivative account for differences in the peak durations.

In a recent work, Santosa and colleagues investigated how the choice of the basis function affects the sensitivity and specificity of the fNIRS analysis, in presence of variability and/or systemic bias in the underlying evoked response, and especially in relation to the task duration [70]. They concluded that the sensitivity of the model depends heavily on the function choice for shorter task durations ($\sim$ 5 seconds), while for longer durations the difference in terms of outcome was less susbstantial.

Solving the model

Once the design matrix has been designed, the model is ready and it is typically solved via generalized least squares:

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^T Y \tag{2.14}$$

$$\mathrm{cov}(\beta) = \left(X^T X\right)^{-1} \tag{2.15}$$

In order for this solution to be the best unbiased estimate of the coefficients, the Gauss-Markov assumptions have to be fulfilled [71] , namely:

1. Each sample of the error term should be uncorrelated with each other: $cov\left(\epsilon_i, \epsilon_j\right) = 0 \ , \ \forall \ i \neq j$

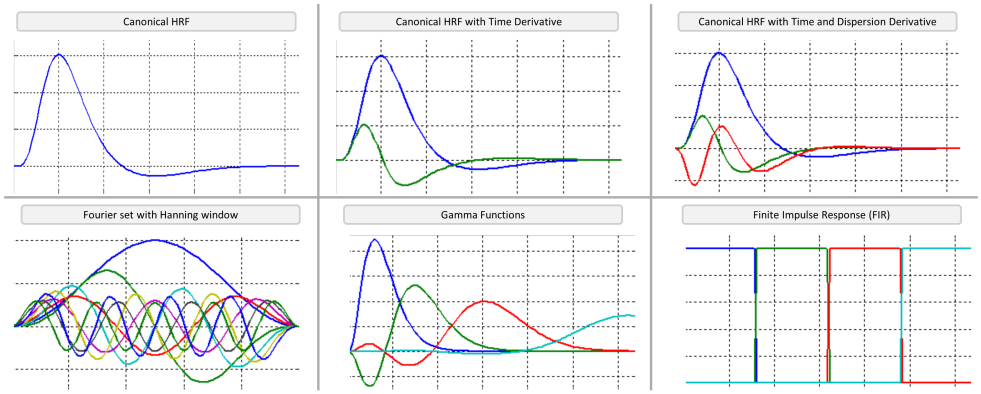


FIGURE 2.13: Basis functions that are most commonly used to model the expected HRF in a GLM. Figures are produced with the analysis software *nirsLAB* [66]. On the top row, the Canonical HRF is depicted (**top-left**); the Canonical HRF can be used in conjunction with its time derivative (**top-middle**) to capture temporal shifts in the peak latency, and also with the dispersion derivative (**top-right**) to capture differences in the peak duration. The (**bottom-left**) panel shows the Fourier set; the Fourier set consists of a constant + K sine functions + K cosine functions (tot: 2K +1) of harmonic period T, T/2, ... T/K. The (**bottom-middle**) panel shows K gamma functions (K=3); while the Finite Impulse Response (FIR) basis function (**bottom-right**) consists of K contiguous boxcar functions, each lasting T/K seconds, where T is the duration of the HRF. Linear combinations of the FIR or Fourier basis functions can capture any shape of response. For a more comprehensive description of these functions, we refer the reader to the book "Statistical Parametric Mapping: the analysis of functional brain images" [4]

2. Each sample of the error term has the same variance $\mathrm{var}(\epsilon_i) = \sigma^2\Sigma$, $\forall i$ (*homoscedasticity of the noise*)

Once the model is estimated, t-statistics can be estimated for the regressor(s) of interest:

$$t = \frac{\beta}{\sqrt{\mathrm{cov}(\beta)}} \tag{2.16}$$

If the goal is to statistically assess whether two or more regressors had a different impact on the data, a *contrast vector* c can be employed that encodes this difference; in practice, if one wants to test, for example, if *condition 1* produced a larger activation than *condition 2*, then the vector c= [1 -1 0 0 . . .] will be defined and the corresponding t-test will be:

$$t = \frac{c \cdot \beta}{\sqrt{c \cdot \mathrm{cov}(\beta) \cdot c^T}} \tag{2.17}$$

The whole procedure, from the definition of the model to the statistics on the regressors, is normally conducted channel-by-channel.

This way, it is possible to determine in which channels was the hemodynamic activity most affected by the experimental conditions, and therefore to have a clear localization of the effects under investigation.

The same analysis is usually performed on all the subjects. The possible options for proceeding with group level analysis are manifold and depend on the particular experiment. For example, it is possible to run a correlation analysis between each channel's regressor of a particular condition and a demographic variable, like a task score, or age of the subjects. Another possibility, when groups of subjects are specified, is to run a Linear Mixed Model with *group* as a between-subject factor, to test whether different experimental conditions had different effects on the groups.

The possibilities are numerous and it is not possible to describe a universal recipe that applies to all circumstances. In the following chapters several approaches that have been used in this doctoral work for group analysis will be presented.

2.4.6 Limitations of the univariate statistical analysis

fNIRS errors are non-spherical

The most important problem with using a GLM at subject-level for fNIRS analysis is that the requirements needed to have an unbiased estimate of the regressors are not met. The error term is nonspherical, namely is not uncorrelated and does not have uniform variance.

Let's start with the latter: the error term resulting from modelling fNIRS data does not have uniform variance. This is caused by the fact that fNIRS data may exhibit components that are much larger in amplitude than the physiological background, for example motion artifacts. These strong components will then appear as outliers in the noise distribution, making it not uniform.

Heteroscedasticity itself does not cause the coefficients estimates to be biased, but it can cause the estimates of their variance to be biased, possibly above or below the true variance. Therefore, while estimated β will still be unbiased, their variances will not, and wrong variances may lead to wrong computation of t-values, making the subsequent statistical inference possibly wrong.

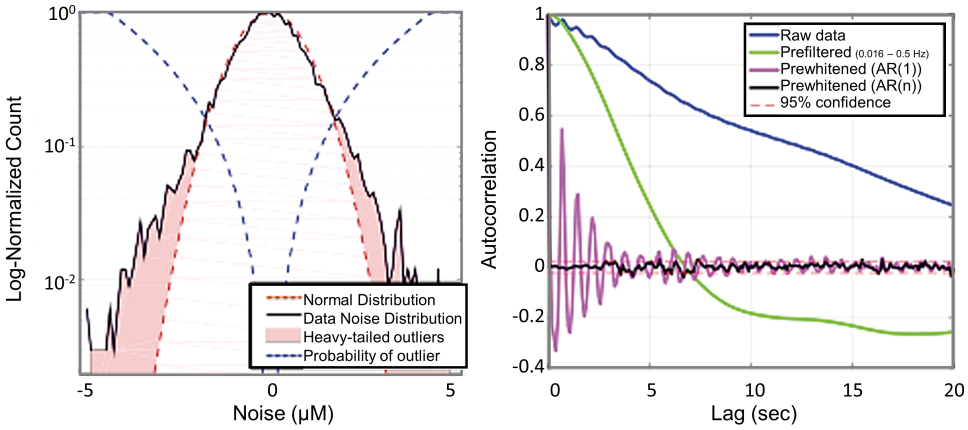


FIGURE 2.14: Figure taken and adapted from [5]. [**Left**] the distribution of the residuals is not Gaussian, but it's characterized by heavy-tails produced by large outliers, such as motion artifacts. [**Right**] The temporal autocorrelation of the raw data along with the autocorrelation of the data processed according to different algorithms; the chart shows that the *AR(n) prewhitening* yields the best results in terms of reduction of autocorrelation.

The other problem is that the samples of the error term are correlated. This is due to the fact that the physiological signal being measured is much slower than the sampling frequency of the acquisition instrument. For this reason, the acquired signal Y – and so also the error term ϵ - have high autocorrelations.

Ignoring these correlations when forming a t- or F-statistics leads to an inappropriate estimate of the degrees of freedom, that are effectively much less than the number of samples, and therefore to invalid tests.

This issue has been known for a long time to be true also for the fMRI analysis, but for the fNIRS analysis it is even more critical because of the much higher sampling rate.

To overcome this problem, a standard solution is to transform the noise structure before solving the model; in particular, the two main approaches that can be adopted are the noise *pre-whitening* and noise *pre-coloring*.

Pre-whitening refers to making the distribution of the noise uniform, thus white, by applying a whitening filter W, selected such that the new error term $W \cdot \epsilon$ is spherical:

$$W \cdot Y = W \cdot X \cdot \beta + W \cdot \epsilon \tag{2.18}$$

$$\widehat{\beta} = \left(X^T W^T W X\right)^{-1} X^T W^T W Y \tag{2.19}$$

The whitening matrix W can be obtained by iteratively fitting the GLM and examining the residuals: at the first iteration, the GLM is run without any whitening (W=I), and based on the residuals a W matrix is estimated. Usually, this is performed by fitting an autoregressive (AR) model to the residuals vector. The AR model will remove the noise structure related to the autocorrelation of the samples. The coefficients of the AR model are then used to build the W matrix.

A whitening matrix based on fitting an AR model to the residuals can definitely reduce their autocorrelations, but the extent of this effect depends on the model order, i.e. it depends on the modelled duration of the autocorrelation. In the software package SPM [4], that was designed for the analysis of fMRI data, the used model order is n=1. This means that correlation is assumed only in neighbouring data samples. While this may prove appropriate for fMRI data, it certainly is not for fNIRS data, since it has a much higher sampling frequency. Nevertheless, the SPM approach has been employed for some time also in the fNIRS data analysis [72].

The other option that has been proposed to solve this issue is noise *pre-coloring*. Pre-coloring refers to the idea of applying a known and specific color spectrum to the noise to mask its unknown structure. Basically the resulting noise structure is still coloured, only this time it has a known profile and can be accounted for by estimating a pre-coloring matrix C, to be applied to the GLM in a similar way of the pre-whitening matrix W.

Pre-whitening is generally preferred over pre-coloring [5], but the autocorrelation of the fNIRS data extends well beyond a model order of n=1. Therefore, research has progressed recently towards finding an fNIRS-specific version of this approach. In particular, an algorithm was recently proposed by Barker et al [6] that corrects both the autocorrelation of the residuals, by implementing a more powerful pre-whitening algorithm, and their heteroscedasticity.

In this method, the W matrix is iteratively estimated by fitting an AR(n) model to the residuals, with n determined by minimization of the Bayesian Information Criterion (BIC). This procedure is performed on each channel independently, resulting in different model orders across different channels. Additionally, at each iteration the weights of the least squares regression are updated to downweigh the timepoints corresponding to outliers, and enter the model via the diagonal weight matrix S. The solution to the model becomes:

$$\widehat{\beta} = \left(X^T W^T S W X\right)^{-1} X^T W^T S W Y \tag{2.20}$$

The authors found that for data acquired at 10 Hz, the optimal model order was n=33, ascertaining that indeed autocorrelation in fNIRS data do extend much beyond the n=1 assumed by SPM.

A-priori selection of the basis function

The choice of the right basis function is another item of complication posed by the use of a GLM. The basis function should represent as accurately as possible the real hemodynamic response evoked by the particular task that is carried on by the subject. In particular, complex tasks may engage temporally protracted processes that cannot be captured by the canonical set [73]. Moreover, even assuming that the canonical HRF is the appropriate choice, several studies have shown that there is a substantial variability in the shape of responses collected across subjects, and even across channels with the same single subject, all of which makes the definition of the parameters challenging [18, 74].

A strategy that has been proposed for addressing this variability is the incorporation of temporal and dispersion derivatives into the canonical set; however, the issue is surrounded by much debate [75, 76, 77, 78]: the main criticism regards the fact that although it is true that the additional terms do capture additional variance, the final test for statistical significance is still carried out on the amplitude of the nonderivative portion of the model, therefore the improvement made in model fitness does not lead to higher power estimate. Furthermore, the subsequent group-level analysis do not

typically account for the subject-level variances at all, thus it remains unaltered by the employment of the additional terms.

Lastly, it is worth noting that the hemodynamic response to the same stimulus can vary greatly across different channels' positions. For example, Zimeo Morais et al. [79] reported a consistent positive correlation between HbO and HbR in the anterior temporal region, that can be explained by extracerebral changes but more likely by oxygenation changes in the temporalis muscle. The contribution of muscle oxygenation changes has been described to play an important role in the signal measured with fNIRS [80, 81], as well as the contribution of small capillaries [82].

All these considerations explain why the GLM can hardly capture the natural inter- and intra-subject variability of the hemodynamic response: if the basis function is inaccurate, the design matrix is not a good model of the measured data, and actual task-induced neural activity may remain undetected.

Multiple comparisons

As the technological advances of the fNIRS hardware provide an increasingly wide coverage of the head, with many systems now offering caps with hundreds of channels, the issue of the multiple testing correction is as crucial as it has ever been, and can no longer be ignored.

The multiple comparisons problem occurs when several statistical tests are carried out simultaneously, as if they were independent: in any statistical test, there is a α% probability of rejecting the null hypothesis when it is true (Type I error, or false discovery), with α defined by the experimenter (usually $\alpha = 5\%$). If N tests are carried out simultaneously, the risk of Type I errors becomes (N x α)%. This is exactly what happens when a univariate statistics is employed on fNIRS data: each channel is statistically tested independently on each other, than a significance threshold α is applied and channels whose t-test resulted in $p \leq \alpha$ are considered *active*. Since channels are not really independent, there is an increased risk of Type I errors, namely, some of the identified active channels are actually false positives.

The simplest way of controlling Type I errors is using the Bonferroni correction: if there are k tests being performed, then the significance level α is replaced by α/k, for

each test. Although this method provides a strong control for errors, it is definitely too conservative for neuroimaging data, and can eliminate both false and true positives. For this reason, other methods have been proposed; for a thorough review, we recommend the reader the work of Nichols and Hayasaka [83].

One of the most advantageous for fNIRS seems to be the False Discovery Rate (FDR) procedure introduced by Benjamini and Hochberg [84, 85]. Briefly, FDR methods aim at controlling the proportion of errors not among the entire set of hypotheses tested, as it is the case for the Bonferroni correction, but only among those resulting in rejecting the null hypothesis, i.e. the declared active channels. This way, the result is less conservative and the power of this correction increases as the number of rejected null hypotheses increases.

Independent analysis of HbO and HbR

Lastly, a significant limitation with performing univariate statistics on fNIRS data is that the analysis is carried out on HbO and HbR independently, when in fact they are nonlinearly partially correlated variables.

The are several inconveniences to this. The first problem is a very theoretical one: the correlation between HbO and HbR is per se meaningful and should be taken into careful examination; in fact, hemodynamic activation should always involve a negative correlation between HbO and HbR (unless we are considering a particular group of subjects or infants [50]). However, this information is completely lost when keeping the two analyses separate. The examination of the two individual components alone can, therefore, lead to misinterpretation of the findings.

This leads to the other problem: for the same experimental condition, it is a very common situation that test statistics yielded by the HbO analysis do not mirror the test statistics yielded by the HbR analysis. In other words, one may find that channels considered *active* by looking at the HbO results, would be considered *not active* with the HbR results, or vice versa; this can happen because every subject has its own "hemodynamic signature", and while some exhibit stronger activation as an increase in the HbO signal, others may express activation more pronouncedly with a decrease in the HbR signal.

So, on one hand the correlation between the two components should always be negative for an activation to be considered as such, but on the other hand one may empirically observe that some subjects do not exhibit much activation with one or the other component [86].

This issue becomes especially problematic at group level: it is often the case indeed that group level results based on the statistical analysis of HbO values do not match results based on the analysis of the HbR values. To overcome this limitation, a univocal measure would be needed, that incorporates information on both components.

2.5 The potential role of Machine Learning

Machine Learning techniques have started being successfully employed in the analysis of fNIRS data only recently, with the research field of Brain-Computer Interfaces (BCI) leading this methodological renovation.

A BCI is a communication system that allows the use of brain activity to control external devices, by clustering these signals into two or more classes; this makes it an application where machine learning naturally belongs. It consists typically of five stages: brain-signal acquisition, pre-processing, feature extraction, classification and application interface. While the acquisition and the pre-processing steps have been already covered in previous sections, and the interface application is beyond the scope of this work, we will focus on the steps concerning feature extraction and classification.

The first study that demonstrated the feasibility of using fNIRS for a BCI was from Coyle et al [87]: in their work, participants had simply to imagine squeezing a rubber ball. The average HbO concentration level was computed within 1 second and used as a feature: if the feature exceeded a reference level (defined as the largest value of the signal within the first 10 seconds), then the event was classified as "task".

Since this study, many others followed and BCIs have become increasingly complex and articulated. Among the features that are mostly employed, the signal average amplitude and slope are the most common; others include signal variance, skewness, kurtosis and zero crossing. For a complete review we recommend the reader the work from Naseer and Hong [88]. As for the classification step, they found that the most

commonly used method in fNIRS-BCI is the Linear Discriminant Analysis (LDA). The reason for it can be found in the fact that LDA combines both good classification performances and low computational requirements. Other algorithms that have proven efficient for fNIRS-BCIs include Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Hidden Markov Models (HNN). A detailed description of LDA and SVM is provided in Sections 2.6 and 2.7.

The use of machine learning in the context of BCIs is then to classify brain signals into several classes. But it can also be used as an alternative to univariate statistics in order to investigate neurophysiological processes and cognitive functions.

This is the case of Multivariate Pattern Analysis (MVPA) methods. MVPA methods started being developed for the analysis of fMRI data, as an alternative to the univariate voxel-wise analysis.

The rationale behind this approach is that part of information is represented not only within each voxel but also across groups of voxel, as "spatial patterns" of activity; MVPA involves the use of pattern-classification algorithms to search for reproducible spatial patterns that differentiate across experimental conditions, and can therefore be considered as a supervised classification problem, where the chosen classifier attempts to capture the relationship between the patterns of activity and the experimental conditions [89].

This approach has several advantages over univariate methods: first of all, it can be more sensitive, simply because the voxel-wise response to the experimental manipulation might be not strong enough to be statistically significant, but could be discriminative of the condition if accumulated with the responses of neighbouring voxels and analysed with classification algorithms [90, 91]. This way, MVPA can extract a finer level of detail from the data.

Secondly, even if two regions do not individually bear information, they may do so when taken into account together.

For this reasons, this approach has been gaining increasing popularity, and has been applied to fMRI, EEG and, more recently, fNIRS [7, 8]. Emberson et al [7] described the first demonstration of neural decoding in infants using multivariate techniques on fNIRS data. In particular, they classified spatial patterns of hemodynamic activity arising

in infants in response to two conditions that were both audio-visual but of different nature (face+music vs fireworks+speech). They demonstrated that despite the absence of significant univariate results, the multivariate classification did manage to decode the two different states, with a classification accuracy averaged across subjects greater than 75%. This and other studies testify to the great potential that machine learning and model-free feature selection have to fill the gaps left by the univariate analysis.

2.6 Linear Discriminant Analysis (LDA)

The goal of classification is to take an input vector $\mathbf{x}$ and assign it to one of K discrete classes C_k (k=1, ….. K). The input space is divided into decision regions, and the boundaries of these regions are called decision surfaces [92].In this section, we will focus on the case of two-classes classification ($K = 2$).

One of the simplest approach to the classification problem is to use a decision surface that is a linear function of the input vector and it's defined by hyperplanes of dimension (D-1), with D being the number of dimensions of the input space. Such linear function is called *linear discriminant function* and it is defined as:

$$y(x) = \mathbf{w}^T\mathbf{x} + w_0 \tag{2.21}$$

where $\mathbf{w}$ is the weight vector, $\mathbf{x}$ is the input vector and w_0 is a *bias*.

An input vector $\mathbf{x}$ is assigned to class C1 if $y(\mathbf{x}) \geq 0$ and to class C2 otherwise. Therefore, the decision boundary is defined by the relation $\mathrm{y}(\mathbf{x}) = 0$, which corresponds to a (D-1)-dimensional hyperplane within the D-dimensional input space. The vector w determines the orientation of the decision surface, while the bias parameter w_0 determines the location of the decision surface (Figure 2.15). The bias can be chosen in different ways; often, it is chosen as the middle point between the projections of the classes' means on the normal vector $\mathbf{w}$.

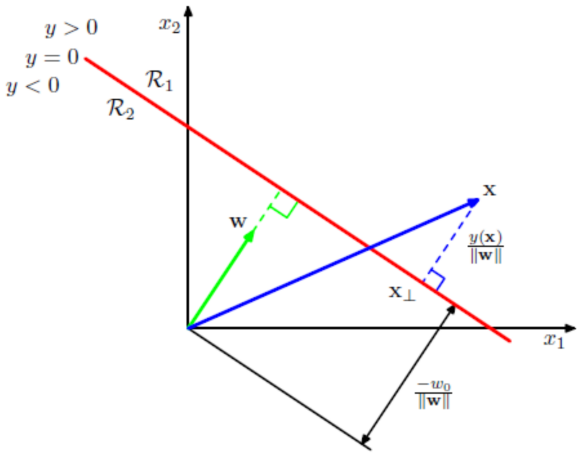


FIGURE 2.15: Geometry of the classification problem with Linear Discriminant Analysis: the decision surface (in red) is perpendicular to $\mathbf{w}$, and its intercept is defined by the bias w_0. Figure taken from Bishop (2006) [92]

The weight vector $\mathbf{w}$ is chosen such as to maximize the distance between the two classes' means (*between-class scatter*, S_B) and minimizes the interclass variances (*within-class scatter*, S_W). This criterion is known as Fisher criterion, and it can be formulated as follows. For class k, the within-class variance is defined as:

$$s_k^2 = \sum_{n \in C_k} (y_n - m_k)^2 \tag{2.22}$$

with $\mathrm{y}_n = \mathbf{w}^{\mathrm{T}}\mathbf{x}_n$ and m_k being the sample mean for the projected points of the class, defined as

$$m_k = \frac{1}{n_k} \sum_{n \in C_k} y_n = \frac{1}{n_k} \sum_{n \in C_k} \mathbf{w}^T x_n \tag{2.23}$$

Therefore, the objective function to maximize in LDA is

$$J(\boldsymbol{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \tag{2.24}$$

To find the optimum $\mathbf{w}$, J($\mathbf{w}$) must be expressed as a function of $\mathbf{w}$. It can be shown that the variance within the class k can be expressed as:

$$s_k^2 = \sum_{n \in C_k} \boldsymbol{w}^T \left(x_n - m_k\right) \left(x_n - m_k\right)^T \boldsymbol{w} \tag{2.25}$$

$$s_k^2 = \boldsymbol{w}^T S_k \boldsymbol{w} \tag{2.26}$$

Thus, the total within-class scatter is defined as:

$$s_1^2 + s_2^2 = \boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w} \tag{2.27}$$

Similarly, the difference between the projected means can be expressed in terms of the between-class scatter S_B:

$$\left(m_2 - m_1\right)^2 = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} \tag{2.28}$$

Therefore, we can express J($\mathbf{w}$) as:

$$J(\boldsymbol{w}) = \frac{\left(m_2 - m_1\right)^2}{s_1^2 + s_2^2} = \frac{\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}} \tag{2.29}$$

Now it's possible to find the optimum $\mathbf{w}$ by differentiating J($\mathbf{w}$); we find that J($\mathbf{w}$) is maximised when

$$\left(\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}\right) \boldsymbol{S}_W \boldsymbol{w} = \left(\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}\right) \boldsymbol{S}_B \boldsymbol{w} \tag{2.30}$$

By multiplying both sides of the equation by $\mathbf{S}_W{}^{-1}$ we obtain that the expression of the optimal weight vector is given by:

$$\boldsymbol{w} \propto \widehat{\boldsymbol{S}}_W^{-1} \left(\widehat{\boldsymbol{m}}_2 - \widehat{\boldsymbol{m}}_1\right) \tag{2.31}$$

It can be shown [93] that LDA is the optimal classifier when *a)* input features from all classes are normally distributed, *b)* the classes have the same covariance matrices and *c)* the true class distributions are known: this latter assumption is never met, therefore means, $\hat{m}_1$ and $\hat{m}_2$, and covariance matrices $\hat{\boldsymbol{S}}$ of the distributions have to be estimated from the data 2.8):

$$\widehat{\boldsymbol{S}} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \widehat{\boldsymbol{m}}) (\boldsymbol{x}_i - \widehat{\boldsymbol{m}})^T \tag{2.32}$$

The estimator of the covariance matrix is unbiased, but in the case of neuroimaging data the estimate may become imprecise: this is due to the fact the neuroimaging data is usually characterized by a limited number of data points in a high-dimensional feature space.

This curse of dimensionality leads to a systematically distorted estimation of the covariance matrix, given by an overestimation of the large eigenvalues and an underestimation of the small ones. The distortion eventually results in degrading the classification performances [94, 95]. The solution that has been proposed to mitigate this effect involves the introduction of a regularization term in the Equation 2.8, such that the new estimated covariance matrix is given by:

$$\tilde{S}(\gamma) = (1-\gamma)\widehat{S} + \gamma v I \tag{2.33}$$

Where $\gamma \in [0, 1]$ and v is defined as the average eigenvalue of $\hat{S}(v = trace(\hat{S})/d)$, with d being the dimensionality of the feature space. It can be shown [94] that $\hat{\mathbf{S}}$ and $\widetilde{\mathbf{S}}$ have the same eigenvectors, and that this approach, named *shrinkage* or *LDA regularization*, has indeed the effect to push extreme eigenvectors towards the average v. The regularization is largest when $\gamma = 1$ and absent when $\gamma = 0$. Several studies demonstrated that regularized LDA yields improved predictive accuracy in the context of ERP decoding [94], fMRI-based decoding [96] and also in fNIRS-based BCIs [97, 98].

2.7 Support Vector Machines (SVM)

While in LDA the separating hyperplane is obtained by seeking the projection that maximises the distance between the two classes' means and minimizes the interclass variances, Support Vector Machines are designed to maximise the distance between the separating hyperplane and the nearest training points, called *support vectors*.

Such distance M is called *margin* and maximising it means minimizing $\mathbf{w}^{\mathrm{T}}\mathbf{w}$. Additionally, $\mathbf{w}$ and w_0 should also ensure that datapoints are correctly classified, namely the input vectors $x_1, \ldots. \; x_N$ should be classified according to their target values t_1, $\ldots t_N$, with $t_n \in [-1, 1] : y(x_n) \geq 0$ for $t_n = +1$ and $y(x_n) \leq 0$ for $t_n = -1$. This criterion can be expressed by the following formula, named the *constraints*:

$$\mathrm{t_n}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x_n} + w_0\right) \geq 1 \qquad \text{for n} = 1, \ldots n \tag{2.34}$$

This is a quadratic problem; it can be shown that an optimal solution to this problem is found when the *Karush-Kuhn-Tucker (KKT) conditions* are satisfied:

$$\lambda_n^* \left(1 - t_n \left(\boldsymbol{w}^{*T}\boldsymbol{x}_n + w_0^*\right)\right) = 0 \tag{2.35}$$

$$\left(1 - t_n \left(\boldsymbol{w}^{*T}\boldsymbol{x}_n + w_0^*\right)\right) \leq 0 \tag{2.36}$$

$$\lambda_n^* \geq 0 \tag{2.37}$$

where λ_n are positive values known as *Lagrange multipliers*. By solving the quadratic problem, it can be shown that $\mathbf{w}^*$ and w_0* are given by:

$$w^* = \sum_{n=1}^{N} \lambda_n t_n x_n \tag{2.38}$$

$$\sum_{n=1}^{N} \lambda_n t_n = 0 \tag{2.39}$$

$$w_0^* = \frac{1}{N_s} \sum_{support\ vectors\ j} \left(t_j - \sum_{n=1}^{N} \lambda_n t_n x_n^T x_j \right) \tag{2.40}$$

And that a prediction on a new datapoint z can be made with the following formula:

$$w^{*T} z + w_0^* = \left(\sum_{n=1}^{N} \lambda_n t_n x_n \right)^T z + w_0^* \tag{2.41}$$

If the classes' distributions overlap in the feature space, we can still maximize the margin M, but we have to allow for some points to be on the wrong side of the margin, with a penalty that increases with the distance from that boundary.

This is achieved by introducing the *slack variables* $\xi_n = (\xi_1, \xi_2, ..., \xi_N)$, with N being the number of data points.

Data points that lie on or inside the correct margin boundary have $\xi_n = 0$, while for the other points $\xi_n = |t_n - y(x_n)|$. Points for which $o < \xi_n \leq 1$ lie inside the margin, but on the correct side of the decision boundary; points for which $\xi_n > 1$ like on the wrong side of the boundary and are misclassified.

Therefore, the constraint (Equation 2.34) must be adjusted to take into account slack variables. This is sometimes described as *relaxing* the hard margin:

$$\mathrm{t_n} \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}_{\mathrm{n}} + w_0 \right) \geq 1 - \xi_n \qquad \text{for n} = 1, \ldots . \mathrm{N} \tag{2.42}$$

with $\xi_n \geq 0$.

Slack variables allow to find an appropriate linear separation, even if the two classes are not completely linearly separable.

When it is impossible to find a linear separation between the classes, the solution adopted by SVMs is to transform the input features $\mathbf{x}$ with a *non-linear kernel*.

This doctoral work only employed Linear SVMs (see Chapter 5), therefore the choice of the kernel for non-Linear SVMs will not be described here; for a complete illustration of the subject, we refer the reader to Bishop (2006) [92] and Marsland (2014) [99].

2.8 Lessons learned

- Systemic oscillations and temporal correlation in the data can reduce the sensitivity of the General Linear Model.
- The independent univariate analysis of HbO and HbR can make difficult and mislead the interpretation of the results, especially at group level.
- The a-priori choice of the basis function impede accounting for inter- and intra-subject variability of the hemodynamic response.
- Machine learning methods can offer a viable alternative to uncover finer levels of detail, by classifying patterns without using any a-priori assumptions on the hemodynamic response; also, an univocal measure accounting for both HbO and HbR could be employed as a feature.

Chapter 3

From univariate to multivariate analysis of fNIRS data: a theoretical formulation and validation

3.1 Introduction

In the previous chapter, the limitations of the analysis based on GLMs were extensively described: the need for a-priori chosen model for the data, that restricts the possibility of accounting for subject- and channel-level shifts from the expected model; the problems yielded by the independent analysis of the HbO and HbR components; the systemic oscillations, that can considerably reduce the sensitivity of the GLM, that may therefore not produce significant results.

Despite the increasing use of machine learning-based algorithms for neuroimaging data, particularly driven by the field of BCIs, and the body of literature involving machine learning-based procedures for fMRI and EEG for the investigation of neurocognitive processes, to date fNIRS still lacks a comparable literature.

In particular, no systematic comparison has been performed between the GLM-based approach and an alternative analysis based on using machine learning, for fNIRS data.

An objective comparison between the two techniques is a critical factor in order to make a reasoned choice when it comes to data analysis.

For this reason, this effort was the very first step of this doctoral project and it is the subject addressed in this chapter, based on the following publication:

[9] Gemignani, J., Middell, E., Barbour, R. L., Graber, H. L. and Blankertz, B. (2018). Improving the analysis of near-infrared spectroscopy data with multivariate classification of hemodynamic patterns: a theoretical formulation and validation. Journal of Neural Engineering, 15(4), 045001. doi: 10.1088/1741-2552/aabb7c

The GLM approach was compared with a model-free approach based on the use of multivariate features, including both HbO and HbR, and on the use of LDA for the classification.

As described in Section 2.5, LDA is often used in the context of BCIs, thanks to its advantageous combination of low computational requirements and good performances. Here we want to assess if its characteristics make it also a convenient tool for offline statistical analysis, in quest of interpreting hemodynamic patterns with respect to the experimental conditions that elicited them. In particular, this work aims at comparing GLM and the LDA-based classifier in performing channel-wise classifications: *active channel* vs *non-active channel*.

The advantage of using a classifier for this purpose is that no assumptions on the structure of the noise are necessary, and that no prior knowledge of the shape of the expected hemodynamic response is needed. Furthermore, for the LDA-based classifier, the information regarding the simultaneous variations of HbO and HbR can be combined in a multivariate strategy. In this way, the analysis would yield a single metric for "activation" for each channel, and this would be easier to test than separately testing β coefficients from HbO and HbR, especially at group-level.

However, the benefits of using machine learning may go beyond "just" producing a larger classification accuracy: the comparison between the results yielded by the two analytic approaches may be informative in the sense that the classifier might identify unpredictable effects that elude the model-based analysis. For example, in a case where GLM analysis reports a channel as "not active", the availability of LDA results could facilitate the process of deciding whether activation truly was absent (i.e., because LDA

also classified the channel as "not active") or if the hemodynamic model used for GLM was not optimal (i.e., because LDA classified it as "active").

The present work comprises two steps: the first is a comparison of the proposed LDA-based method with the canonical GLM analysis. In order to do this, an extensive volume of simulated data is used to characterize the two algorithms in terms of receiver operator curves (ROC) when no systemic oscillation is present (i.e., simulated hemodynamic responses were added to simulated resting-state data) or when a considerable amount of systemic oscillation is present (i.e., simulated hemodynamic responses were added to experimental resting-state data); the real-data results also were used to characterize the impact of inter-subject variability on the outcomes of the classification analyses. Second, the two algorithms were used to analyse and classify the task-induced activations in a set of experimental data.

3.2 Methods

In order to compare the performances of the LDA-based and GLM-based methods under controlled conditions, sensitivity and specificity were quantified by recovering a known synthetic hemodynamic response added to either synthetic or real resting-state data. This approach has been used in several reports [5, 6, 100, 101] and it is particularly suited for studies that make use of ROC analysis, because it permits an exact quantification of true and false discovery rates.

In a second step, the two analysis pipelines were applied to real experimental data and channel-wise statistical assessments of each subject were compared. Figure 3.1 reports a summary description of the whole procedure followed in this work. Figure 3.2 shows how known hemodynamic responses were added over resting state time traces.

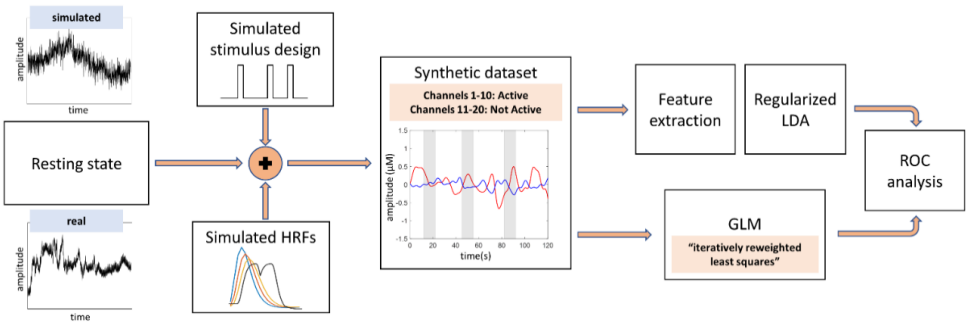


FIGURE 3.1: In each iteration, data are simulated, based on either synthetic or real resting-state data; HbO and HbR (red and blue time traces in the "Synthetic dataset" panel, respectively) were analyzed with GLM or LDA, and ROC analysis was performed to compare the classification accuracies. The simulated HRFs vary in shape and size, and 30% of them are characterized by a "double bump" as a simplified model of stimulus-locked Mayer waves.

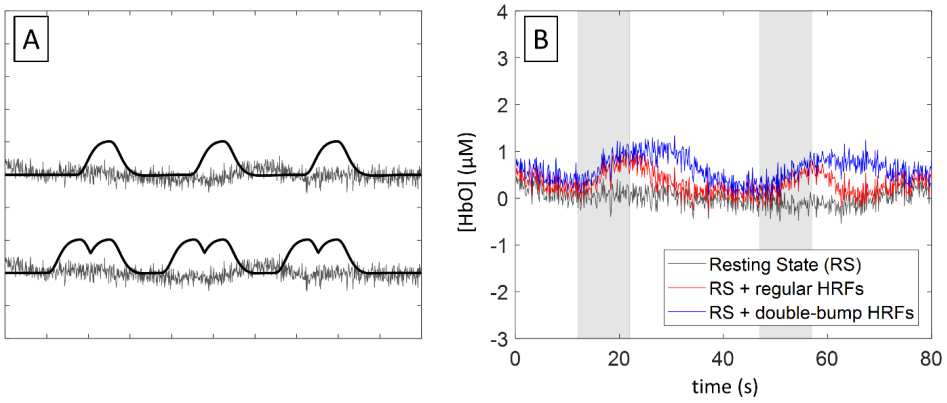


FIGURE 3.2: **(A)** Example of how simulated HRFs are created (black line) and added to a real resting-state time trace (dark grey line). The top trace is a simple HRF while the bottom trace contains a double bump. **(B)** The grey time trace represents the HbO signal before the hemodynamic activations are added; the red and blue ones are, respectively, the time traces after a simple HRF or a double-bump HRF have been added.

3.3 Theoretical formulation

3.3.1 Generation of the synthetic dataset

5000 datasets of NIRS data were iteratively simulated by combining temporally correlated ("colored") noise and synthetic hemodynamic response functions (HRFs). Baseline noise was produced by first generating white noise, then imposing temporal correlation on it by employing an autoregressive model of order 30, via tools in the NIRS Brain AnalyzIR toolbox [65]. Each dataset contained 20 channels, and synthetic HRFs were added to the resting state for half of them (i.e., Channels $1-10$). Channels that included synthetic HRFs in their HbO and HbR data were labelled as *"active"* and the others as *"not active"*. For the *active* channels, 'Start' and 'Stop' markers were created according to an experimental paradigm with 3 trials, each of 10 seconds duration. The position of the 'Start' markers was randomized, with the constraint that successive ones were separated in time by at least 35 seconds. The total duration of each time series was 4 minutes, at a sampling frequency of 7.81 Hz.

To model the inter- and intra-subject variability of real hemodynamic responses, synthetic evoked HRFs had variable size and shape across subjects and channels. While each HRF had the mathematical form of a canonical HRF [4], their peak amplitudes ranged from 0.01 to 0.1 µM [101], while, based on experience and existing literature on the variability of the hemodynamic response [18], the onset-to-peak times ranged from 2 to 8 seconds and the onset-to-undershoot times ranged from 14 to 18 seconds.

Positive-going synthetic HRFs (see Fig.3.2A) were added to the resting-state data for the HbO time series. The synthetic HRFs added to the corresponding HbR resting-state data had the same form as those for HbO, but were 50% reduced in magnitude and reversed in algebraic sign (i.e., they were negative-going). In addition, for 30% of the *active* channels in each dataset the synthetic HRF included a "double bump" (see Fig.3.2A), as an elementary model of systemic hemodynamic activity time-locked with the experimental condition. An example of this sort of additional activity that frequently is present in experimental data are the so-called "Mayer waves," which are systemic oscillations originating in the superficial layers of the tissues [102], and which occur, more or less prominently, at $\sim$0.1 Hz frequencies. Such oscillations are

particularly difficult to treat in a GLM-analysis framework, owing to their extensive spectral overlap with typical event-related activity (i.e., they cannot be eliminated via straightforward frequency filtering) [103].

The use of synthetic baseline noise has the clear benefit that a large volume of data can be created, and it allows to benchmark the proposed methodology against recent literature on the topic [5, 6].

However, synthetic data might not capture all the properties of real physiological data. Therefore we complemented the synthetic-data analysis by using experimental resting-state data as baseline noise. For this purpose, 15 young adults (mean age ± SD: 28.1 ± 4.0 years old; age range: 23-38; 11 women, 4 men) participated in the collection of 4 minutes of resting-state data. For a subset of the participants, this measurement was followed by the motor-task study that was used in a later stage of this analysis (see Section 2.2.1 for descriptions of the experimental setup and data collection).

Experimental resting-state recordings were used as a source of real physiological and correlated data, and were employed in the same manner as described above for the synthetic resting-state data, namely by performing 5000 randomizations of the positions of the 'Start' markers and adding simulated HRFs of variable shapes and amplitudes to only Channels 1-10 (left hemisphere), and labeling those channels as *active*.

3.3.2 Data analysis

Pre-processing

Both simulated and real data underwent the same pre-processing steps. Hemoglobin concentration changes were calculated using the modified Beer-Lambert law (Differential Pathlength Factor (DPF): 6 , absorption coefficients (μ_a, $cm^{-1}M^{-1}$) for HbO: μ_a(760 nm) = 1349 and μ_a(850 nm) = 2436, for HbR: μ_a(760 nm) = 3565 and μ_a(850 nm) = 1592).

Data was bandpass filtered in the range [0.01-0.2] Hz, with a zero-phase distortion digital FIR filter designed and implemented, respectively, with the MATLAB commands firls and filtfilt. For the subsequent statistical analysis, filtered data was used for the LDA analysis, in accordance with most fNIRS-based BCI literature [88], while

unfiltered data was used for the GLM computations because it has been reported that frequency filtering can produce biased estimates of the regressors [5]. In this way, both methods were used at their optimal settings.

Analysis with GLM

GLM was applied using the *autoregressive iteratively reweighted least squares* ("AR IRLS") algorithm available in the NIRS Brain AnalyzIR toolbox. This algorithm is reported to efficiently remove serial correlations from data, thereby achieving an acceptable false discovery rate [6]. HbO and HbR time traces were analyzed independently.

After the regressors β were estimated, the null hypothesis that there was no hemodynamic response (H0: $\beta = 0$) was tested by defining a contrast vector (c) and calculating the channel-wise t-statistic via the formula [67]:

$$t = \frac{c^T \beta}{\sqrt{c^T \operatorname{cov}(\beta) c}} \tag{3.1}$$

In this case, with only one experimental condition to be tested, the contrast vector would be [1 0], with the second column referring to the constant column added to the GLM design matrix. The p-values corresponding to the t-statistics from Eq. 3.1 were computed via two-tailed t-tests.

Analysis with LDA

For each HbO and HbR time series, trials were defined as the signal in the 15-second time interval following each 'Start' marker. Each trial was baseline-corrected by removing the mean value of the signal over the 3-seconds interval prior to stimulus onset. Channel-wise block averages were obtained by averaging across all trials within each channel.

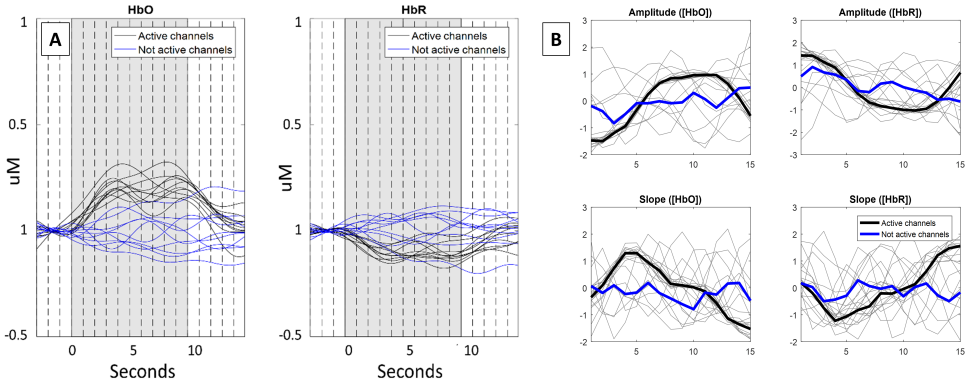


FIGURE 3.3: (**A**): Block averages of HbO (left) and HbR (right) signal used for feature extraction; dashed lines represent the 1s steps used for the moving-window computation of amplitude and slope. (**B**): Features vectors are obtained from the block averages by computing mean and slope of the signal over a sliding window of 3 s duration with 1 s steps, resulting in a 30-features vectors that were then normalized and concatenated to produce the 60-features multivariate (HbO + HbR) classifier. Grey lines represent individual trials, black lines highlight the mean value of the feature vectors corresponding to *"active"* channels, and blue lines highlight the mean value of the feature vectors corresponding to the *"not active"* channels. Values of the y axis are normalized values

Features were extracted from the channel-wise block averages (Figure 3.3 A). To do this, a 3-seconds-wide window was moved through the block-average time series in 1-second steps and the mean value and mean slope (computed as the change in signal amplitude over the time window divided by its size in number of samples) were computed within each window, yielding a 30-features vector (2 features $\times$ 15 windows) for each of HbO and HbR. Each feature vector was normalized to zero mean value and unit variance. Then the HbO and HbR feature vectors were concatenated, resulting in a 60-features vector that was used for the classification (Figure 3.3 B).

Channels were classified as *active* or *not active* with Regularized Linear Discriminant Analysis, via tools available in the Berlin Brain-Computer interfacing (BBCI) toolbox [104, 105]. Ten repetitions of 4-fold cross validation were performed: 20 trials (10 *active* channels and 10 *not active* channels) were separated into 4 folds, with three folds used for training and the remaining fold as the test dataset. The procedure was repeated 10 times.

Each feature vector $\mathbf{x} \in \mathbb{R}^N$ was assigned an output by the application of the formula of the separating hyperplane characterizing the LDA classifier [104]:

$$\boldsymbol{w}^T \boldsymbol{x} + w_0 = 0 \tag{3.2}$$

where $\mathbf{w}$ is the projection vector characterizing the classifier and w_0 is a bias term. For further detail, see Chapter 2, Section 2.6. In this implementation, the class *not active* was assigned to negative or zero outputs and class *active* was assigned to positive outputs.

Evaluation of performances

The performances of the GLM analysis and the LDA-based method were evaluated by computing receiver operating characteristic (ROC) curves.

ROC curves for the GLM results were computed by varying the significance threshold for the t-test p-values, from 0 to 1 in increments of 0.001, and computing the corresponding false positive rate and true positive rate for each threshold. ROC curves for the LDA results were computed by comparing the distributions of output values of

active and *not active* channels. We defined a significance threshold, varying from 0 to 1 in increments of 0.001, in the following manner: on the distribution of *not active* outputs, we defined a reference value as the percentile corresponding to the considered threshold. We defined as True Negatives (TN) the samples of the *not active* distribution that were smaller than the reference value, False Positive (FP) the samples of the *not active* distribution that were equal or greater than reference value, True Positives (TP) the samples of the *active* distribution that were equal or greater than the reference value and False Negative (FN) the samples of the *active* distribution that were smaller than the reference value. We repeated the procedure by sliding the reference value until 100% of the *not active* distribution was covered (i.e., significance threshold = 1). For example, by setting the threshold at 0.05, we computed the reference value on the distribution of *not active* outputs corresponding to its 5% percentile, and based on this reference value we computed TP, TN, FP, FN at p = 0.05.

For both GLM and LDA, classification accuracy was computed as the rate of correct classifications, (TP + TN)/(TP + TN + FP + FN), at p = 0.05. In order to investigate the impact of double-bump HRFs on classification accuracy, we conducted two separate analyses on the two subsets of data characterized by, respectively, only HRFs with no-double bumps and only HRFs coupled with double-bumps.

Analysis of the correlation between subjects' demographics and classification accuracy

Demographic information such as gender, age and chronobiology has been reported to play a role in the cerebral metabolism [106, 107, 108, 109], and therefore we tested for correlations between the individual-subject classification accuracies and each subject's characteristics. To do so, a linear mixed-effects (LME) model was fitted in Matlab 2017, with a random intercept for each participant and fixed effects for age, gender, hair color (two levels: Blond, Brown), and time-of-day of the measurement (three levels: 10 AM-1 PM, 1 PM-3 PM, 3 PM-6 PM):

$$Accuracy \sim Age + HairColor + Gender + TimeMeasurement + (1|Participant)$$

This analysis was carried out for LDA (HbO+HbR), GLM (HbO) and GLM (HbR) separately. Analysis of variance (ANOVA) was performed on each model to test the significance of the effects (error DF = 10 (15 observations minus 5 modelled effects)).

3.4 Application of the algorithm to experimental data

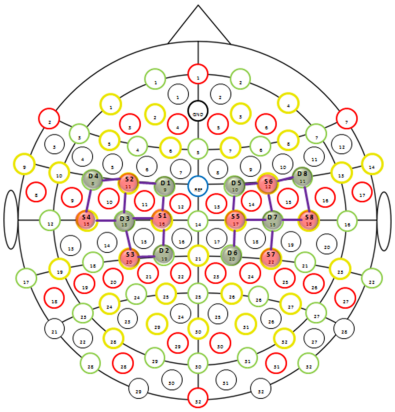
To provide a practical example of use of the proposed algorithm and compare it with the GLM analysis in the framework of a real experiment, a paradigm was chosen—finger tapping—that has a well-known effect on the motor cortex. In particular, it is expected to elicit a recognizable and significant response in the primary motor cortex (M1, Brodmann area 4, likely to underlie the C3/C4 positions of the EEG 10-20 system) and the premotor cortex (PMC, Brodmann area 6, likely to underlie the FC3/FC4 positions) [110].

3.4.1 Experimental setup and data collection

Seven healthy young adults (a subset of the 15 participants in the preceding study; mean± SD age 26.0± 2.3 yr, age range 23-30 yr; 5 female, 2 male) participated in this study. The experiment consisted of 16 trials of finger tapping (8 left, 8 right, alternating), each of 10 s duration, with 20 s rest periods between successive episodes. Before the experiment began, the subject was required to sit quietly for the collection of 4 minutes of resting-state data.

NIRS recordings were conducted with a NIRSport system (NIRx GmbH, Berlin, Germany), with sampling frequency 7.81 Hz, at wavelengths 760 nm and 850 nm, with 8 sources and 8 detectors. Sources and detectors were placed into a cap (EASYCAP, Herrsching, Germany), arranged according to the International 10-20 system. The source-detector distance was 2.5-3 cm, to form 20 channels evenly distributed between the hemispheres (Figure 3.4).

A



B

Channel	Source	Detector
Left Hemisphere		
1	1	1
2	1	2
3	1	3
4	2	1
5	2	3
6	2	4
7	3	2
8	3	3
9	4	3
10	4	4
Right Hemisphere		
11	5	5
12	5	6
13	5	7
14	6	5
15	6	7
16	6	8
17	7	6
18	7	7
19	8	7
20	8	8

FIGURE 3.4: (**A**) 8 sources × 8 detectors montage (covering the motor region); (**B**) Source-Detector pairs and corresponding channels

A spatial sensitivity profile was calculated based on the Monte Carlo photon migration modeling available in the AtlasViewer software [111], to prove that the probe design was selective for the regions relevant to the finger tapping task (underlying the 10-20 positions FC3/FC4 and C3/C4). The Monte Carlo modeling was performed with 10^6 photons. Figure 3.5 shows the probe arrangement and the resulting sensitivity profile.

3.4.2 Data analysis

The data analysis aims at identifying which channels are significantly activated by the motor task and can therefore be labeled *active*, as opposed to the *not active* channels that are not significantly activated by the task. For this reason, no distinction was made between left and right-hand finger tapping. The data was analyzed with the GLM analysis and the LDA-based method described in the previous section.

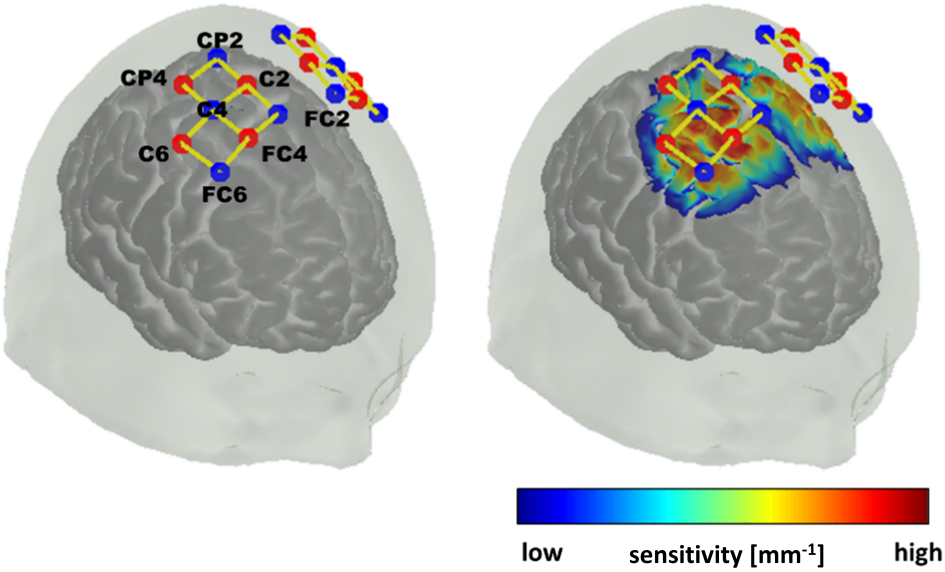


FIGURE 3.5: (**Left**) Probe setup. The arrangement of optodes follows the 10-20 standard and the placement is analogous in the other hemisphere. Red dots indicate the sources, blue dots indicate the detectors, and yellow lines indicate the formed channels. (**Right**) Sensitivity profile of the probe setup. The sensitivity profile has units of mm^{-1}. Multiplying the value of the sensitivity profile of a given measurement channel by an absorption change (mm^{-1}) and the area over which that change occurs (mm^2) will give an estimate of the optical density change that would be measured by that channel as a result of that absorption change [111]

Analysis with GLM

For the GLM analysis, the stimulus times of the task were convolved with a canonical hemodynamic response function (peakTime = 6s) to produce the single column ("Task" condition) of the design matrix. A GLM was applied using the autoregressive iteratively reweighted least squares algorithm available in the NIRS Brain AnalyzIR toolbox [65].

Analysis with LDA

For the LDA analysis, amplitudes and slopes were computed for each trial of finger tapping. For the "Rest" condition, an equal number (n = 16) of time intervals were produced by randomly sampling the initial 4 minutes of resting-state data of each measurement, and features were extracted. The sampling of "Rest" trials and the classification Task vs. Rest was iterated 2000 times for each channel and for each subject, to ensure robustness of the analysis. Ten repetitions of 4-fold cross validation were conducted and each of the 32 trials (16 task and 16 rest) was assigned a classifier output via Equation 3.2.

3.4.3 Statistical analysis

The results of the GLM analysis were statistically assessed by computing the channel-wise t-statistics (Equation 3.1) from the resulting β values, then testing them via two-tailed t-tests. The outputs of the LDA analysis were divided into "Task" and "Rest", then averaged over folds and over repetitions, and tested by comparing the two distributions (Task and Rest). The rationale of this procedure is that, if the task elicited a hemodynamic response and the classifier had a good discrimination between "Task" and "Rest", then the distributions of the outputs should be well separated and the channel will be labeled as *active*. If, on the contrary, the two distributions are not well separated, it means that for that channel the execution of the task did not elicit a response substantially different from the resting state, and the channel will be labeled as *not active*. As explained in Section 2.1.2, the class label *active* is assigned to positive outputs, while *not active* to the negative outputs. Therefore, the channel-wise p-value

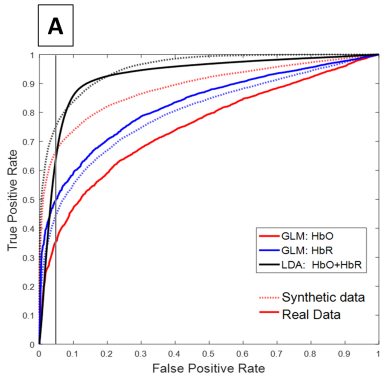
in this case was computed as the fraction of "Rest" outputs equal or greater than the mean value of the distribution of the "Task" outputs [112].

3.5 Results

3.5.1 Theoretical formulation

Performance of the algorithms: overall classification accuracies

Our first goal was to theoretically compare the two algorithms in terms of overall classification accuracy, both on synthetic and on real data. The other important objective was to evaluate whether, with real data, the achieved results are consistent across subjects, and to evaluate the impact of inter-subject variability on the performance of each algorithm.



B

Overall mean classification accuracy at p = 0.05		
	Synthetic resting state	Real resting state
LDA	85.25 %	78.76%
GLM(HbO)	79.08 %	65.76%
GLM(HbR)	67.75 %	70.29%

FIGURE 3.6: (**A**) ROC curves for GLM and LDA, using HbO, HbR and HbO+HbR features (only for LDA). The curves for the real resting-state data (solid lines) are obtained by averaging the individual curves across subjects, while dotted lines refer to the completely synthetic dataset. (**B**) The table reports the mean classification accuracies, over all iterations, of the three algorithms applied to synthetic and real resting-state data. The classification accuracy is computed from the ROC curves at the false positive rate of 0.05.

Figure 3.6 A shows the ROC curves obtained using synthetic and real resting-state data. In both cases the LDA classifier based on HbO+HbR features outperforms GLM

applied to either HbO or HbR, with results tabulated in Figure 3.6B. A difference between GLM results for synthetic and real resting-state data is also seen, in that GLM(HbO) is more accurate than GLM(HbR) in the former case, while GLM(HbR) is more accurate than GLM(HBO) in the latter.

We speculate that this difference indicates that the synthetic data does not entirely represent the properties of the real physiological data. For example, it certainly does not reflect the frequency structure of the resting-state signal, or its spatial dependence across the different channel positions. In addition, the temporal correlation in the synthetic data was imposed by using an autoregressive model of fixed order (N = 30) [5], which doesn't account for the variability that can be found in real data from different subjects.

To further investigate the performances of the three methods, we computed the classification accuracies for each subject individually (Figure 3.7 A). The barplots indicate the classification accuracy as computed from the individual subjects' ROC curves at p = 0.05, and the red line indicates the mean accuracy over all subjects, respectively (mean± SD) $78.76 \pm 5.1\%$ for LDA(HbO+HbR), $65.76 \pm 10.2\%$ for GLM(HbO) and $70.29 \pm 8.9\%$ for GLM(HbR), the standard deviation being computed across the 15 subjects. The individual errorbars represent the standard error of the mean across the 5000 repetitions.

Finally Figure 3.7 B shows the classification accuracies computed on two separate sets of data: data for all the channels that did not have Mayer waves modelled (i.e., no "double bumps" [Figure 3.2 A]) and data for all the channels that did have them. In this case, for the data without Mayer waves we found that LDA achieves an accuracy of $79.1 \pm 6\%$, GLM(HbO) $77.8 \pm 9.3\%$ and GLM(HbR) $82.4 \pm 8.2\%$, while for data with Mayer waves, the accuracy decreases to $77.0 \pm 11\%$ for LDA, $62.4 \pm 7.5\%$ for GLM(HbO)and $64.8 \pm 6.8\%$ for GLM(HbR).

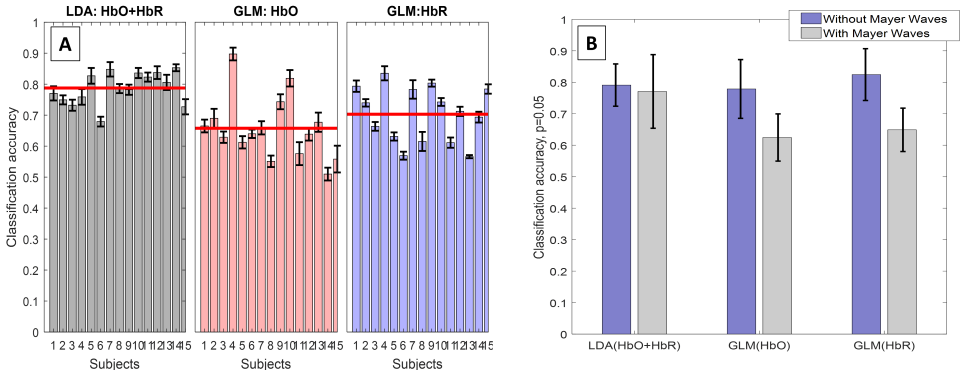


FIGURE 3.7: (**left**) Individual classification accuracies for the real-resting-state datasets, for LDA(HbO+HbR), GLM(HbO), and GLM(HbR) (left, middle, right). The red line indicates the mean accuracy reached by each algorithm over all the subjects. The errorbars represent the standard error of the mean for each individual subject, over all the iterations performed. The individual mean accuracies achieved by the LDA method is significantly higher than those achieved by the GLM(HbO) ($p = 0.0002$) and GLM(HbR) ($p = 0.01$), but no significant difference was found between GLM(HbO) and GLM(HbR) ($p = 0.24$, Repeated Measures ANOVA 1-way with Fixed Effect: "Analysis Method"). Also, the individual standard errors of the mean yielded by the LDA are significantly lower than those achieved by the GLM(HbO) and GLM(HbR) (p=0.021 and p=0.022, respectively), but no difference was found between those yielded by GLM(HbO) and GLM(HbR) (p=0.97). (**right**) Classification accuracies computed on two subsets of the real-resting-state datasets, one completely free from Mayer-wave oscillations and the other one with all the HRFs tainted by double-bumps. For the LDA, there is no significant difference between the accuracies reached in presence and absence of Mayer waves (paired t-test, $p = 0.44$), while for the GLM the difference was statistically significant (GLM(HbO), $p < 0.001$, GLM(HbR), $p < 0.001$) Note: the analysis, both with LDA and with GLM, was conducted on the two datasets (with and without Mayer waves) independently

Correlation between classification accuracies and individual measures

The goal of this analysis was to quantitatively assess the impact of individual characteristics (hair color, gender, age), and of the measurement time of day, on the individual classification accuracy. Table 3.1 shows the results of the LME analysis. The model shows a significant correlation between Hair Color and individual accuracies for

GLM(HbR), but not for any of the other fixed effects in the model, and there are no significant correlations for either LDA or GLM(HbO). Figure 3.8 reports distributions of individual accuracies grouped by hair color. More plots of accuracy distributions grouped by the other effects are provided in the Appendix (Figures A.1, A.2, A.3).

	LDA:HbO + HbR		GLM:HbO		GLM:HbR	
	β	p-value	β	p-value	β	p-value
Age	0.0042	0.1762	0.0010	0.8734	0.0068	0.1120
Hair Color	0.0361	0.1875	-0.0936	0.1020	-0.1408	0.0052
Gender	0.0038	0.1476	-0.0077	0.1552	-0.0026	0.4904
Time of Measurement	-0.0187	0.5109	0.0573	0.3308	-0.0072	0.7540

TABLE 3.1: Results of the linear model fitted to the individual classification accuracies, with fixed effects: Age, Hair Color, Gender and Time of Measurement

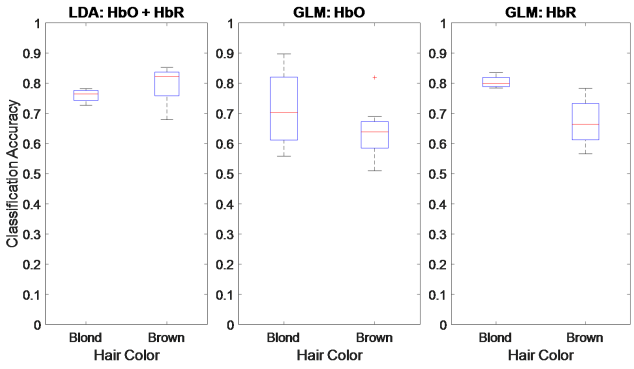


FIGURE 3.8: Distribution of individual classification accuracies within the two hair color classes (six blond and nine brown). The accuracies reached by the GLM(HbR) are significantly higher for blond-haired subjects than brown-haired. More distributions for the other modelled effects can be found in the Appendix. The central red marks represent the median values, the blue boxes extend from the 25th to the 75th percentiles, and the black whiskers extend to the most extreme data points not considered outliers (which are marked with red crosses).

3.5.2 Experimental results

Results from the experimental-data study are reported in Figure 3.9, as t-statistic values for GLM(HbO) and GLM(HbR), and classifier outputs for LDA, for $p \leq 0.05$

threshold. White cells indicate that the channel did not reach statistical significance.

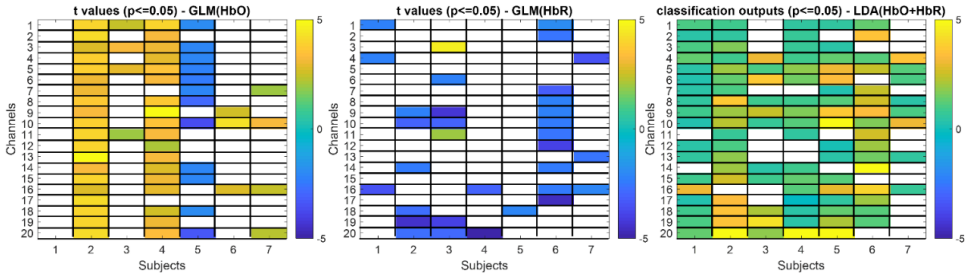


FIGURE 3.9: Classifier results for the finger-tapping experimental data, for the three different analyses. GLM t-statistic values and LDA classifier outputs (the latter derived from application of the separating hyperplane formula) are thresholded at $p \leq 0.05$. Blank cells indicate non-significant values (i.e., that the corresponding channel was classified as *not active*). The individual minimum value for statistical significance for the results of the LDA classifier varied across channels, ranging from 0.12 to 1.18. The numbers of channels classified as *active* by the three analyses are significantly different (p = 0.01, 1-way repeated measures ANOVA).

To better understand the source of the variability in the results, plots of the block-averaged trials for those channels, and topographic images of the channel-wise output values (LDA output values, and βs for GLM HbO/HbR), were produced for each subject. The images were produced via the visualization tool available in nirsLAB v2017.06 [66].

Plots for all subjects and corresponding output values are available in the Appendix, while here only the plots for subject 1 and subject 2 are reported.

Subject 1: Subject 1 results are non-significant at p = 0.05 for every channel according to the GLM(HbO) analysis, and significant for channels 1, 4 and 16 in the GLM(HbR) analysis, while the great majority of channels are classified as active (p<0.05) by the LDA classifier.

In the plots of Table 3.10, we observe that the HbO time traces are greatly affected by the double-bump typical of the 0.1 Hz systemic oscillation, and also that the first peak after stimulus occurs earlier than the onset-to-peak time of the theoretical model (6 sec).

An enlarged depiction of the block-average behaviour for Channel 16 is shown in Figure 3.11, together with the plots of the HRF model used in the GLM analysis and the block averages of the resting-state trials used by the LDA classifier. The resting-state HbO trace also includes a feature that is qualitatively similar to a hemodynamic response, but the task response is correctly discriminated from the resting-state time series nevertheless ($p<0.001$).

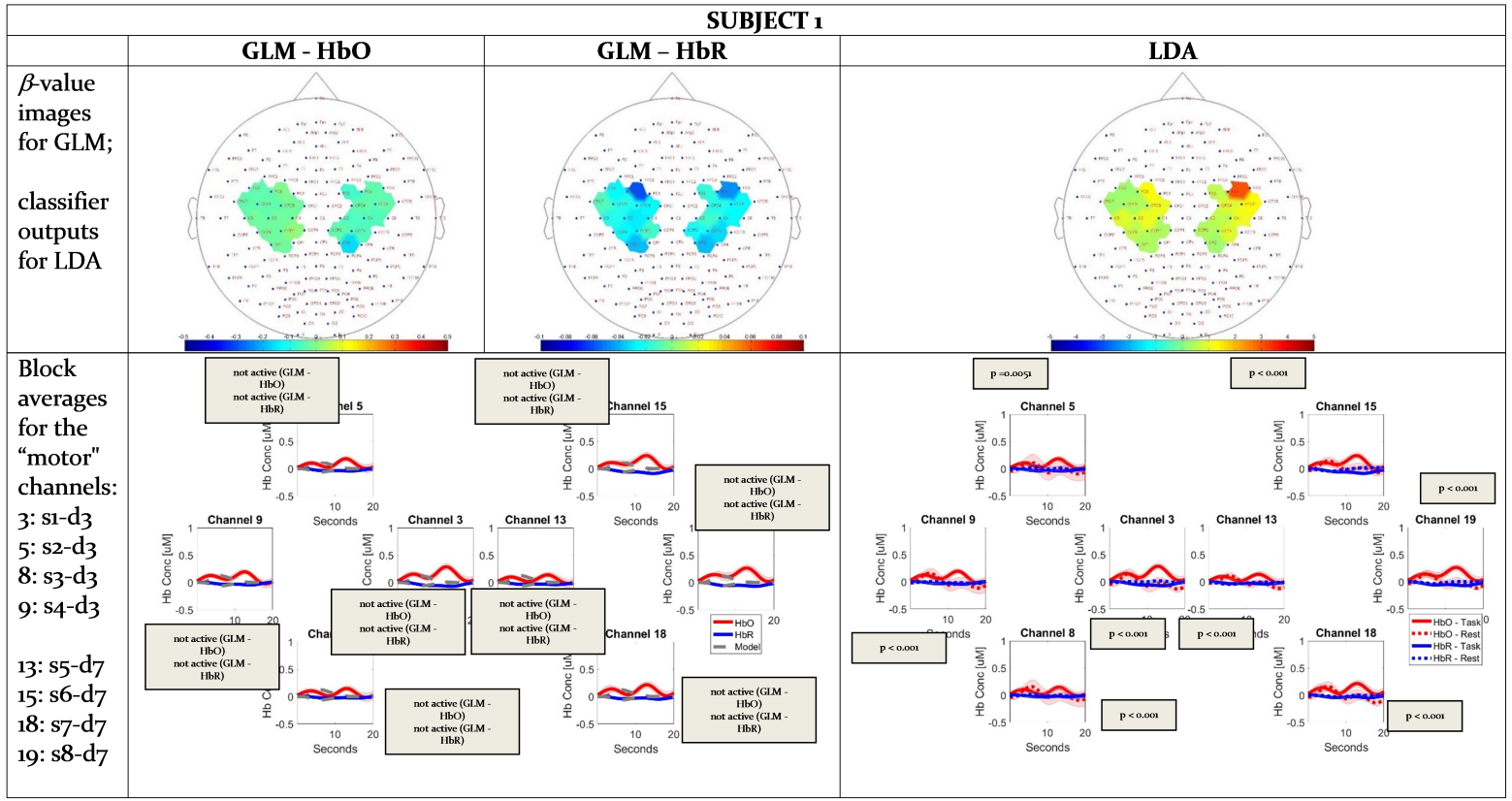


FIGURE 3.10: Topographic images and block averages for all the analyses on Subject 1

Subject 2: All channels of Subject 2 are classified as active ($p = 0.05$) in the GLM(HbO) analysis, while 6 of 20 are classified as active by the GLM(HbR) analysis, and 19 of 20 by the LDA classifier (3.13). A depiction of Channel 5 is presented in Figure 3.12).

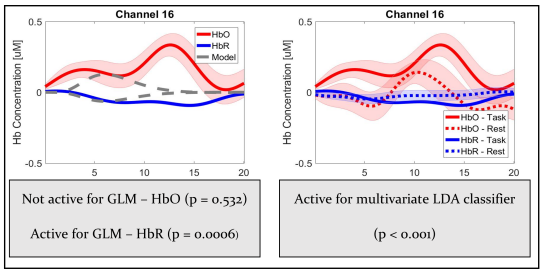


FIGURE 3.11: Block-average data for Subject 1, Channel 16. On the left the plot of the averaged signal is accompanied by the plot of the model used by the GLM analysis, namely a canonical HRF with peak time = 6s. On the right, the same plot is accompanied by the plot of an example of average of resting state trials against which the task trials are classified.

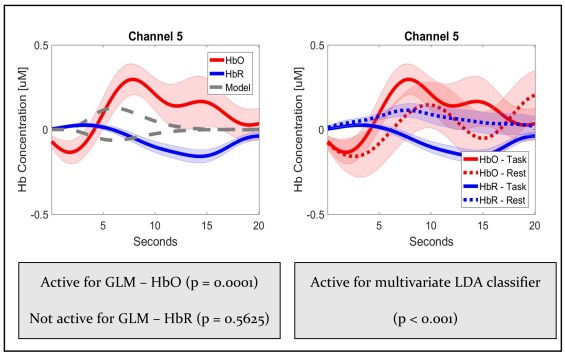


FIGURE 3.12: Block-average data for Subject 2, Channel 5.

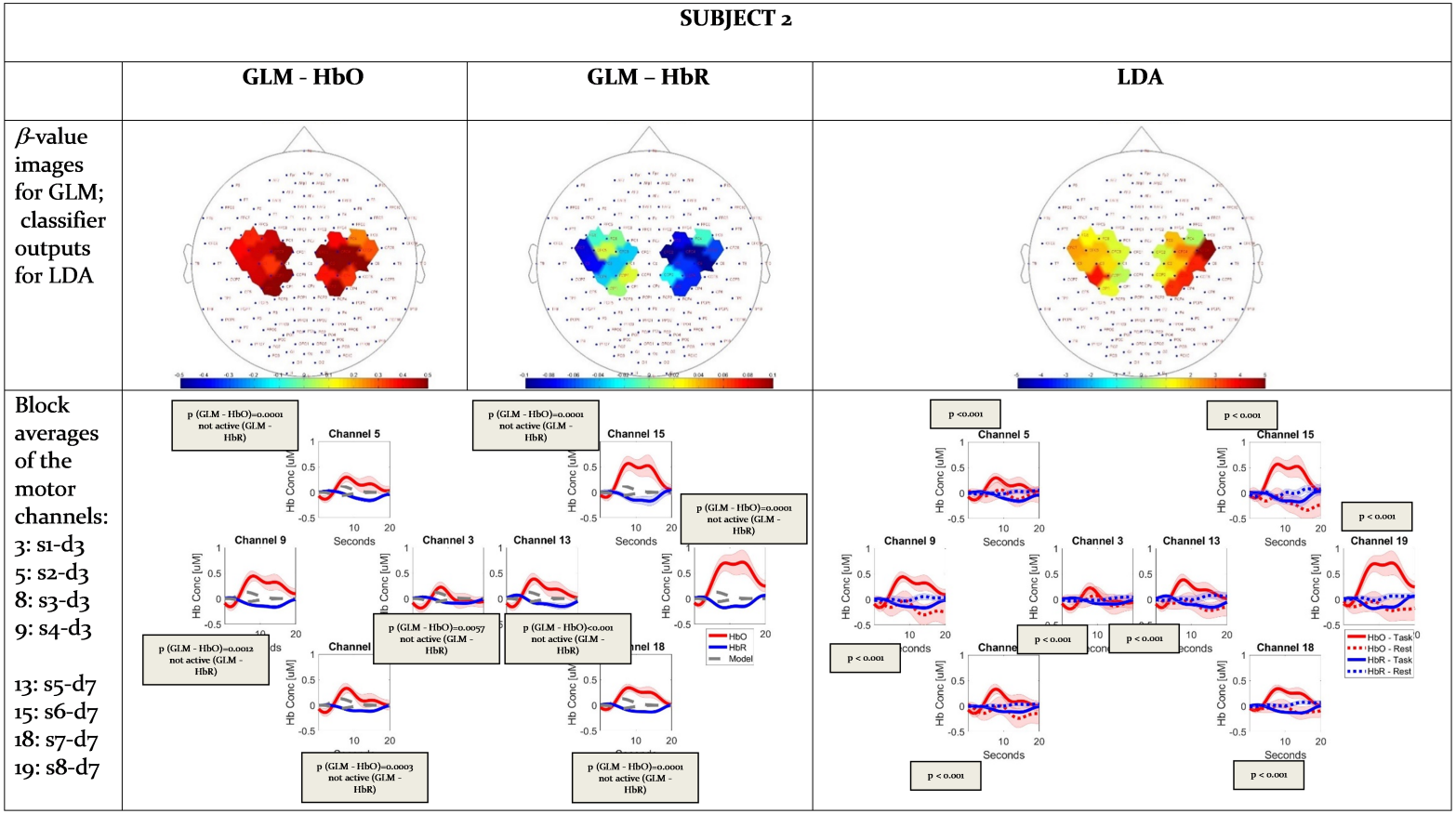


FIGURE 3.13: Topographic images and block averages for all the analyses on Subject 2

Corresponding results for all subjects can be found in the Appendix. For each subject, a table is reported with:

- Topographic images of channel-wise GLM β values and LDA classifier outputs. Large positive (or negative) β values indicate a good fit of the GLM model to the HbO (or HbR) data, and a correspondingly better chance of that channel having a statistically significant hemodynamic response. LDA outputs are negative if the channel is classified as *not active* and positive is the channel is classified as *active*. Therefore, a large positive classifier output value indicates a good chance that the channel is labeled as *active*.
- Block averages of the signal in response to the stimulus (read and blue curves for HbO and HbR, respectively). The shaded error bars indicate the standard error computed over the experimental trials. The GLM plots are accompanied by the canonical basis function used by the model; the LDA plots are accompanied by the block averages of the resting-state trials. The block averages are shown only for the channels covering the motor cortex.

3.6 Discussion

The statistical analysis of fNIRS data is often complicated by serial correlations, inter-subject variability of the hemodynamic response, and the presence of systemic oscillations and possibly motion artifacts. The study presented in this paper demonstrates that a data-driven approach (linear discriminant analysis, LDA) to data analysis is more robust than the most commonly employed model-based approach (general linear model, GLM) to many of these issues, and can therefore improve the detection of the hemodynamic activation.

Advantages of the proposed LDA approach are that no assumptions on the structure of the noise are necessary, and that no prior knowledge of the shape of the expected hemodynamic response is assumed. The LDA method compares data from different temporal segments of the same recording; namely, it compares, within the same subject, time intervals corresponding to the resting state and to execution of the task. Thus it constitutes a self-referencing approach, and in other fNIRS imaging contexts it has

been shown that this data-analysis strategy can enhance detectability of effects that are small in comparison to other sources of intra- and inter-subject variance [113]. As such, LDA can generate information potentially superior, or at least complementary, to the information yielded by a model-based approach. For example, if LDA recognizes activation where the GLM doesn't, it could mean that the GLM model does not accurately represent the real HRF, and it might be worth investigating why this is so.

An additional strength of the multivariate LDA classifier proposed in this study is that it combines features from the simultaneous variations in HbO and HbR time series, while the GLM approach analyzes them independently. This results in the former yielding a univariate channel-wise metric for "activation", while the latter yields separate beta coefficients for HbO and HbR. Performing statistical tests on a single metric is highly desirable, especially for group-level studies.

To quantify and compare the classification performances of the three methods, we made use of both synthetic and real resting-state data. The use of synthetic data, for which the ground truth is known with certainty, also allowed us to benchmark our methodology against recent literature regarding GLM classification accuracy [6].

The multivariate LDA classifier yielded greater classification accuracy than GLM, for both the synthetic and real resting-state data (78.7% for LDA, 65.76 for GLM(HbO) and 70.3% for GLM (HbR), in the real resting-state data case [Figure 3.6A]). Moreover, we demonstrated that the LDA had less inter-subject variability, as illustrated in Figure 3.7, where the standard deviation of individual results was 5.1% about the mean for LDA, as opposed to 10.2% for GLM(HbO) and 8.9% for GLM(HbR). In addition, the linear mixed model fit of individual-subject accuracies to predictors Age, Hair Color, Gender, and Time of Measurement revealed a significant effect only for Hair Color and only on the accuracy achieved with GLM(HbR) (blond-hair accuracy > brown-hair accuracy). The latter findings show that the observed differences between accuracies of the model-based and data-driven approaches is not simply accounted for by obvious (and easily absorbed into the classification model) demographic or physical characteristics of either the subject (e.g., gender) or the measurement (e.g., time of day).

The LDA results also are less sensitive than those for GLM to the "double bumps" that were used to approximate Mayer waves synchronized with hemodynamic task

responses (in the presence of double bumps, classification accuracy falls from $79.1 \pm 6\%$ to $77.0 \pm 11\%$ for LDA, from $77.8 \pm 9.3\%$ to $62.4 \pm 7.5\%$ for GLM(HbO), and from $82.4 \pm 8.2\%$ to $64.8 \pm 6.8\%$ for GLM(HbR) [Figure 3.7B]). These results confirm that the GLM, at least when used with a fixed basis function for all subjects, as was the case here, is less successful than LDA at picking up individual variability and atypical activation patterns, and is at risk of false negatives. These results suggest the possibility that the model used does not accurately represent the real hemodynamic responses and that therefore a different model would need to be designed. In this respect, the results of the one analysis can be used in support of interpreting the results of the other.

As a control study, additional simulations were performed to identify the classification performance of the LDA-based method when applied to data that did not actually contain any task-induced responses (either real or simulated) in the "Task" time intervals. The classifier performed at chance level in these cases (results not shown), suggesting that the possibility of false-positive results in the analyses of simulated and real task-response data is not an important concern.

To further understand and characterize the performances of the two pipelines in a real application, we used the two methods to analyze data from a motor experiment. The optode array covered the motor cortex and its vicinity on both hemispheres. Eight out of twenty channels were placed over the scalp positions most likely to cover the motor area. In this scenario we could verify that the LDA-based classifier is less susceptible to than GLM to 0.1 Hz systemic oscillations. This is illustrated for the subject considered in Figure 3.10: due to the systemic oscillations, the block-average HbO and HbR traces in the eight channels over the motor cortex differ from the hemodynamic response modeled in the GLM computations. Consequently, the GLM recognizes none of these channels as "active". Conversely, by contrasting "Task" temporal segments with "Rest" temporal segments, the LDA classifier finds significant differences in all of these channels, regardless of the presence of Mayer waves.

On the other hand, the LDA-based method generally classified more channels as *active* in response to the motor task than the canonical GLM analysis did. Because no established ground truth exists in the real data, this classification result warrants cautious interpretation. Especially for channels that extend beyond the center of the motor cortex, the activations found cannot be unambiguously attributed to neural activation

caused by the motor task. However, the resting-state data classification results, and inspection of the experimental hemoglobin time traces, show that these classification results also are not easily dismissed as false positives.

In fact, as shown in recent literature [26], the fNIRS signal is not composed exclusively of cerebral task-evoked signal but also includes cerebral non-evoked signal ("cerebral resting state"), extracerebral task-evoked signal ("extracerebral confound"), and extracerebral non-evoked signal. Quantitative characterization of the three latter components is still an open research question [25], which is why they could not be modeled separately in our simulations. However, they are likely to be present in the motor experiment data and offer a plausible explanation for the classification results: when a difference is found between "Task" and "Rest", not only the cerebral response to the task, but also all the systemic hemodynamic changes provoked in the extracerebral compartment by the execution the task (e.g., changes in heart rate, blood pressure, respiration rate), is discriminated from the resting state. These changes involve the whole extracerebral layer, and therefore their effect extends beyond the probes that specifically illuminate the motor cortex [26]. To exclusively associate the found activations with cerebral recruitment, a step that would be necessary, but is beyond the scope of the present work, would be to remove from the data the physiological component measured in the extracerebral layers before the analysis, for example with multi-distance NIRS measurements [59].

Finally, it is worthy of note that the experimental data, in agreement with the results of the theoretical simulations, reveal great inter-subject variability in the comparative sensitivities of GLM(HbO) and GLM(HbR). That is, some subjects' hemodynamic patterns are better interpreted, and hemodynamic task responses more detectable, using HbO data, while others' are better explained using HbR. This is a manifestation of the highly subject-specific hemodynamic fingerprint that has been reported [103]. A classifier, such as the one proposed, that takes into account simultaneous variations of both hemoglobin components has the potential to overcome this limitation and offer a more flexible analysis that adapts to the individual's own hemodynamic characteristics. In this respect, the proposed approach would be of especial application for populations, such as young children, that exhibit "atypical" patterns of hemodynamic responses, such as uncoupled HbO and HbR or inverted response direction [114, 115].

Nevertheless, in the current form the proposed approach only classifies "activations" vs "non-activations." As a future development, a non-parametric framework can be formulated in order to test more complex hypotheses on the distributions of classifier outputs, such as the comparison of amplitudes of the responses induced by different conditions, within subjects or between groups of subjects.

3.7 Lessons learned

- With the proposed LDA-based approach, no assumptions on the structure of the noise are required, and no prior knowledge of the shape of the HRF is expected.
- The proposed LDA-based classifier combines features from the simultaneous variation of HbO and HbR.
- LDA yielded classification accuracies that were, both overall and at individual level, higher than those yielded by the GLM analysis; also, they were more stable across subjects, i.e. less susceptible to individual variations, and less affected by the presence of systemic oscillations.
- When employed in the context of a motor experiment, LDA classified as *active* channels that extend beyond the motor cortex; the reason for this result could likely be found in the systemic hemodynamic changes provoked in the extracerebral compartment by the execution of the motor task, that are picked up by the classifier and used for the "task vs rest" classification.
- Future work should include the removal of the physiological confound from the data, by means of short-distance channels. This procedure would allow to exclusively associate the found activations with cerebral recruitment.

Chapter 4

Investigating the effect of literacy on the functional organization of the brain with *univariate* analysis of fNIRS data

4.1 Introduction

Reading is a complex cognitive task. It is usually acquired during childhood, it requires explicit effort and training, and it relies on a whole range of brain functions in order to develop successfully.

In particular, it has been shown that the acquisition of literacy skills is associated with the activation of a specific brain network that includes the temporo-parietal, occipito-temporal and inferior-frontal areas of the brain [116]; furthermore, fMRI studies demonstrated that as the development of these skills progresses, activation of these areas becomes less and less bilateral: while it is common in beginning readers to "employ" both left and right hemisphere (LH and RH, respectively), skilled reading involves mainly the LH only [117, 118].

This knowledge is very useful: if a "typical-development trajectory" can be clearly charted, then it is also possible to identify deviations from it and to characterize the

neurobiological features of reading disability, or to follow up a remediation program for reading skills improvement.

Studies have further shown that, in addition to brain changes, reading skills can also be predicted at behavioural level, by how well a child can read "pseudo-words", namely words that are phonologically but not lexically possible: pseudo-words – or "non-words" - are often used in conjunction to real words in experiments of developmental psychology because they allow testing separately phonological and lexical skills. It has been shown that while both skilled and less skilled readers can repeat real words, less skilled readers fail to repeat pseudo-words [52, 119].

The knowledge about the changes in brain organization associated with reading development and about the predictive power of behavioural measures lay the foundations of this project.

On one hand, neuroimaging has had a crucial role in the insights we have today about the brain organization changes, and this is especially true for fMRI; however, there would be several advantages of using fNIRS to follow these brain changes: its ease-of-use, low operating costs and portability would facilitate greatly the diagnosis of reading disability, that could be formulated in a more child-friendly setting and using ecological conditions; but especially it would make possible a tight-scheduled follow-up of the treatment, since there are no burdensome costs associated with each session, as it is the case for fMRI.

For these reasons, the first goal of this project was to ascertain whether it is possible to gain with fNIRS the same insights about brain organization changes associated to reading development that have gained with fMRI.

On the other hand, the knowledge about the predictive power of certain behavioural measures like pseudo-word repetition can be also be explored to see how this measure correlates with hemodynamic patterns at different levels of literacy skills.

The ultimate ambition of this project is establishing an "fNIRS-based predictor" of literacy: if it is possible to observe developmental changes in brain organization and also to verify that non-words elicit different brain patterns depending on reading skills, then we have a simple yet effective tool for facilitating the diagnosis of reading difficulties. This project is the object of this Chapter and of Chapter 5.

In particular, in this Chapter we apply a standard processing routine, that includes the use of a General Linear Model at subject-level, to respond to the two questions: can fNIRS see the same as fMRI does? And can fNIRS measure different hemodynamic patterns depending on reading skills? The first question could be successfully addressed: as it is illustrated in the Chapter, it was possible to observe a clear shift in the cerebral recruitment, from bilateral to left-hemispheric, as reading skills develop.

As for the second question, however, we will see that it was not possible, with the employed analysis approach, to highlight any differential activation to non-words depending on reading skills. For this reason, the focus of the next chapter will be to use an alternative procedure based on Multivariate Pattern Analysis to try and respond to the same question.

4.2 Methods

4.2.1 Participants

Eighty-three (83) healthy children (52 males, 31 females) between the ages of 3.7 and 7.9 years (M: 5.6, S: 1.05) participated in this study. Participants were recruited from New Haven and the surrounding areas of Connecticut. The experimental protocols were approved by the Yale Institutional Review Board. Participants were native English speakers. Children who had a formal diagnosis of a developmental disorder did not meet eligibility criteria for this study.

Measure	
n	83
Age (years, M=mean, S= standard deviation)	3.7÷7.8 (M: 5.6, S: 1.05)
Gender (Male:Female)	52:31
IQ	73÷134 (M:110.3, S:13.6)
Letter-Word Decoding (WJ)	4÷55 (M:25.8, S: 12.3)

Table 4.1: Participant's summary information

4.2.2 Behavioural assessment

Behavioural testing sessions assessed children's reading abilities. In particular, the assessment was performed using the standardized Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, and Mather, 2001).

WJ represents a comprehensive battery of psycho-educational tests aimed at measuring general intellectual abilities, specific skills, oral language and academic achievement across a wide age range. Out of the comprehensive collection of tests available, we focused on the "Letter-Word Identification" (WJ-LW): in this sub-test, participants are presented with written words of increasing difficulty, and they are required to read them out loud. The raw score represents the number of words that were read correctly. As simple as it is, the raw WJ-LW identification is an effective score of reading abilities. For this reason, it was used in this study as a measure of literacy skills of the participants and will be called from now on just "Reading Score".

Additionally, children's verbal and performance intelligence quotient (IQ) was measured using the Wechsler Abbreviated Scale of Intelligence (WASI-II or WPPSI-IV).

4.2.3 Experimental design

Participants were instructed to passively listen to 16 blocks of auditory stimuli, played though headphones while looking at a fixation cross that appeared on a monitor.

Each auditory condition consisted of 8 blocks, with each block consisting of the repeated sequence of one word or non-word. Non-words conformed to the phonological properties of English, but had no meaningful referent. The duration in each block was 7 s and contained 6 repetitions of the same word or non-word, with 100 ms of silence between each repetition. There was a 13 s rest period between each block and the order of blocks (word vs. non-word) was randomized.

Throughout the exposure to these stimuli, participants' fNIRS signals were measured.

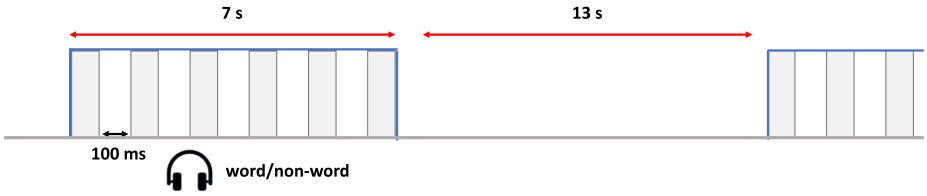


FIGURE 4.1: Structure of the experimental design

4.2.4 Data acquisition

4.2.5 Optode localization

A Patriot 3D Digitizer (Polhemus, Colchester, VT) was employed to measure for each participant the anatomical locations of optodes in relation to standard head landmarks. For each channel, the corresponding Montreal Neurological Institute (MNI) coordinates were obtained using functions available in the NFRI toolbox [120], and used to perform anatomical labelling, according to the Talairach atlas [121, 110].

The anatomical location of each channel can vary sensibly across different subjects, due to individual anatomical differences, and this can affect the group level results. To account for this effect, median MNI coordinates across participants were calculated for each channel (Table 4.2), and were later used as described in Section 4.2.7. With the MNI coordinates it was possible to determine the anatomical location of each channel, using the Talairach atlas [110].

Ch#	MNI coordinates			Brodmann Area	Pr(%)
	X	Y	Z		
1	-65.6	-52.33	-14.83	37 - Fusiform gyrus	64.8
2	-70	-28.33	-16.5	21 - Middle Temporal gyrus	97.5
3	-65.5	-8.33	-23.5	21 - Middle Temporal gyrus	58.7
4	-35.6	62.83	-7	10 - Frontopolar area	79.1
5	-12.3	72	-1.5	10 - Frontopolar area	98.8
6	17	72	-0.5	10 - Frontopolar area	100
7	39.3	63.33	-6.5	10 - Frontopolar area	87.3

8	68.17	-6.33	-21.33	21 - Middle Temporal gyrus	81.6
9	72	-26.5	-14.83	21 - Middle Temporal gyrus	98.7
10	67	-48.67	-15.17	37 - Fusiform gyrus	54.8
11	-62.3	-61	-5.5	37 - Fusiform gyrus	84.7
12	-70	-39.33	-2	21 - Middle Temporal gyrus	88.8
13	-70	-17	-6.33	21 - Middle Temporal gyrus	97.5
14	-62	2.3	-19	21 - Middle Temporal gyrus	100
15	-49.3	45.83	-5.33	47 - Inferior prefrontal gyrus	98.5
16	-25.3	67	9.34	10 - Frontopolar area	100
17	5	70	14.67	10 - Frontopolar area	100
18	28.83	67.67	10.17	10 - Frontopolar area	100
19	52	47.33	-2.83	47 - Inferior prefrontal gyrus	78.3
20	64	4.33	-14.83	21 - Middle Temporal gyrus	100
21	72	-14.5	-3.33	21 - Middle Temporal gyrus	83.5
22	72	-37.67	-0.83	21 - Middle Temporal gyrus	64.2
23	64	-58.5	-5.67	37 - Fusiform gyrus	93.1
24	-67.17	-50.33	9.67	22 - Superior Temporal Gyrus	65.8
25	-69	-26.33	13.5	42 - Primary and Auditory Association Cortex	98.7
26	-66.5	-5.33	7.83	22 - Superior Temporal Gyrus	93.2
27	-56.67	25.17	3.33	45 - pars triangularis Broca's area	56.8
28	-12	66.67	24.83	10 - Frontopolar area	100
29	15.67	67	25.83	10 - Frontopolar area	100
30	60	27.67	7.67	45 - pars triangularis Broca's area	96.3
31	68	-3.33	9	22 - Superior Temporal Gyrus	96.1
32	71.17	-25.33	13.17	42 - Primary and Auditory Association Cortex	97.5
33	69	-48.17	10.83	22 - Superior Temporal Gyrus	88.3
34	-63	-59.33	21.33	39 - Angular gyrus, part of Wernicke's area	50.7
35	-68	-36.5	30.67	40 - Supramarginal gyrus part of Wernicke's area	100
36	-67	-12.33	27.33	1, 3 - Primary Somatosensory Cortex	50.5
37	-61	11.33	14.17	44 - pars opercularis, part of Broca's area	100
38	-51	41	14.33	46 - Dorsolateral prefrontal cortex	98.6
39	-24.7	54.83	34.33	9 - Dorsolateral prefrontal cortex	90.6
40	1.67	57.17	39.67	9 - Dorsolateral prefrontal cortex	100

41	27.33	55	34.67	9 - Dorsolateral prefrontal cortex	98.2
42	53	40.33	18.67	46 - Dorsolateral prefrontal cortex	100
43	63	13.83	17	44 - pars opercularis, part of Broca's area	90.7
44	69	-9.33	28.5	6 - Pre-Motor and Supplementary Motor Cortex	47.4
45	70	-34.67	29.5	40 - Supramarginal gyrus part of Wernicke's area	100
46	64	-58.33	21	39 - Angular gyrus, part of Wernicke's area	44.7
47	-64	-46.5	39.67	40 - Supramarginal gyrus part of Wernicke's area	100
48	-65	-22.5	39.33	1,2,3 - Primary Somatosensory Cortex	98.8
49	-63	1.83	31.33	6 - Pre-Motor and Supplementary Motor Cortex	97.3
50	-53	30.33	26.67	46 - Dorsolateral prefrontal cortex	95.7
51	-37.67	44.17	36.33	9 - Dorsolateral prefrontal cortex	96.2
52	-12.33	47.67	49.67	8 - Includes Frontal eye fields	100
53	14.33	47.33	50.5	8 - Includes Frontal eye fields	100
54	38.67	44.67	38	9 - Dorsolateral prefrontal cortex	100
55	55	29.67	28.5	46 - Dorsolateral prefrontal cortex	98.6
56	65	3.33	33.33	6 - Pre-Motor and Supplementary Motor Cortex	87.7
57	67.5	-22.5	40.67	1 - Primary Somatosensory Cortex	50.6
58	65.17	-47	38.67	40 - Supramarginal gyrus part of Wernicke's area	100

TABLE 4.2: The X, Y, and Z columns represent median MNI coordinates across subjects. Anatomical labelling of the MNI coordinates was conducted using the Talairach atlas. The last column lists the atlas-based probability that the channel coordinates are within that anatomical location. In this table only the Brodmann area with the highest probability is displayed; the complete table is available in the Appendix.

4.2.6 Pre-processing

Data were pre-processed using custom scripts written in MATLAB^{TM}. In particular, raw data were converted into optical density and then concentration changes of oxy- and deoxyhemoglobin (HbO and HbR) using the modified Beer-Lambert Equation ($\Delta HbO = -1.488 \times \Delta A_{780} + 0.5970 \times \Delta A_{805} + 1.4847 \times \Delta A_{830}, \Delta HbR = 1.845 \times \Delta A_{780} - 0.2394 \times \Delta A_{805} - 1.0947 \times \Delta A_{830}$, with A780, A805 and A830 being the optical absorbances at 780, 805 and 830 nm, respectively).

After conversion into HbO and HbR changes, time traces were screened to detect and correct motion artifacts. This step was performed using the Wavelet-based algorithm described in [55] and available in Homer2 [64], and applied with parameter threshold =1.5 and order N=2 Daubechies wavelet. After the motion artifacts correction, the data quality of the channels was evaluated by computing the coefficient of variation (CV) of each channel, defined as the standard deviation of the timeseries divided by its mean value. Channels having CV $>$ 8% for either HbO or HbR were discarded from the analysis. This criterion was defined by visual inspection.

The next step was to remove from the data the contribution of global systemic effects originating in the superficial layers of scalp, dura, and peripheral vasculature and not in the cortex [26]. To this end, a spatial filter based on principal component decomposition [58] was applied to both HbO and HbR time traces.

Finally, a band-pass frequency filter was applied to the data in the range 0.01-0.2 Hz, to ensure that the time traces did not contain contributions from either the very low frequency systemic fluctuations, the respiration, or the heart rate oscillations. The filter was designed as a zero-phase digital FIR filter with the MATLAB function *filtfilt*.

4.2.7 Statistical analysis

The statistical analysis at Subject Level was performed with a General Linear Model (GLM), implemented via tools in the NIRS Brain AnalyzIR Toolbox [65]. In particular, the design matrix was obtained by convolving the stimuli times with a Canonical Hemodynamic Response Function (onset-to-peak time= 6 s, undershooting time= 16 s, duration= 32 s), and it was fit to the timeseries using the "*iteratively reweighted least squares autoregressive model*" presented in [6] and available in the toolbox. This implementation has been shown to have a good control of type-I-error rate [5].

This analysis resulted in two β coefficients for each experimental condition (Word, Non-Word).

To account for the fact that the anatomical location of each channel varies across subjects, beta values were projected into the MNI standard brain space, namely, they were normalized via linear interpolation according to the distance between each subject's

actual channel location and the median location of that channel across all subjects (Table 4.2), according to the procedure described in [122]. This step ensured beta values could be correctly compared at group-level.

For each channel independently, the Group Level Analysis was carried out by fitting a 2x2 Linear Mixed Model on the ß coefficients resulting from the Subject Level Analysis, with between-subject factor Group (Non Readers, Readers) and within-subject factor Stimulus Type (Word, Non Word), random intercept Participant and random slope Stimulus Type. The model was implemented with the MATLAB function *fitlme*.

On each model, an Analysis of Variance (ANOVA) was performed aimed at quantifying the main effects of Group and Stimulus Type and the interaction effect Group x Stimulus Type. Additional T-tests were conducted only on the channels that resulted significant ($p < 0.05$) from the previous step.

4.3 Results

4.3.1 Mixed-Effects Linear Model

The results of the channel-wise Mixed Effects Models are reported in Table 4.3, for both HbO and HbR. To avoid type I errors due to multiple comparisons, p-values were adjusted using the False Discovery Rate procedure by Benjamini and Hochberg [84].

Both HbO and HbR analyses showed a significant effect of Group in the Right Hemisphere, especially in those channels pertaining to the Middle Temporal Gyrus (BA: 21, channels 8, 9, 20, 21, 22), Fusiform Gyrus (BA: 37; channel 10, 23), Superior Temporal Gyrus (BA: 22; channels 31, 33, 46), Supramarginal and Angular Gyri ("Wernicke's area", BA:39, 40; channels 45, 46) and the Pars Opericularis and Triangularis parts of Broca's area (BA: 44, 45, channels: 30, 43), but only the analysis of HbR results for channels 32, 33, 44, 45, 46 kept statistical significance after FDR correction.

Only for these channels, additional two-sided t-tests were carried out to assess the direction of the effect and it was found that in all of them Pre-Readers exhibit a hemodynamic activation higher than Readers (Table 4.4).

Main Effect: Group (Non Readers vs Readers)								
	HbO				**HbR**			
Channel	**F**	**DF_1**	**p**	**p_{FDR}**	**F**	**DF_1**	**p**	**p_{FDR}**
8	7.22	203.4	0.0078	0.1130	0.75	243.0	0.3881	0.8658
9	4.69	231.0	0.0314	0.1861	2.30	231.0	0.1310	0.4468
10	0.94	219.0	0.3342	0.6057	6.11	219.0	0.0142	0.0634
20	7.94	243.0	0.0052	0.1011	1.10	243.0	0.2958	0.8658
21	8.07	243.0	0.0049	0.1011	5.17	243.0	0.0238	0.0987
22	4.79	234.0	0.0296	0.1861	6.37	201.9	0.0124	0.0634
23	4.08	213.0	0.0447	0.2159	7.89	187.5	0.0055	0.0510
31	8.07	246.0	0.0049	0.1011	6.72	199.9	0.0102	0.0594
32	6.77	246.0	0.0098	0.1140	8.82	203.4	0.0033	**0.0388**
33	4.62	234.0	0.0326	0.1861	8.89	191.6	0.0032	**0.0388**
43	5.80	237.0	0.0168	0.1621	6.21	170.9	0.0137	0.0634
44	4.48	237.0	0.0353	0.1861	10.11	175.1	0.0017	**0.0338**
45	3.65	243.0	0.0573	0.2555	10.88	191.8	0.0012	**0.0338**
46	4.59	228.0	0.0332	0.1861	10.19	186.1	0.0017	**0.0338**
56	2.44	234.0	0.1195	0.4348	7.43	159.5	0.0071	0.0510
57	1.30	225.0	0.2558	0.5506	7.21	180.3	0.0079	0.0510
58	2.34	231.0	0.1274	0.4348	7.36	178.3	0.0073	0.0510
Main Effect: Stimulus Type (Word vs NonWord)								
1	4.92	180.5	0.0278	0.4744	4.11	180.9	0.0440	0.3070
2	5.14	185.2	0.0246	0.4744	4.11	179.4	0.0442	0.3070
3	5.73	177.1	0.0178	0.4744	4.17	168.1	0.0427	0.3070
11	3.93	182.7	0.0491	0.4744	3.65	182.5	0.0577	0.3070
12	4.03	181.9	0.0462	0.4744	4.14	177.7	0.0435	0.3070
13	4.31	189.0	0.0393	0.4744	4.81	172.9	0.0297	0.3070
24	3.49	190.4	0.0632	0.5161	4.16	187.9	0.0427	0.3070
Interaction Effect: Stimulus Type x Group								
—								

TABLE 4.3: Channels showing a significant ($p \leq 0.05$) effect of Group, Stimulus Type or Stimulus Type x Group, before FDR correction. The formula employed for the model was 'beta $\sim$ stim*group + (stim|id)'; categorical variables were coded with the "*effects*" coding, that means that the sum of the dummy coefficients amounts to zero. This is also known as *deviation coding* and it is appropriate for testing main effects and interaction effects [123]

Channel	t	degrees of freedom	p
32	-2.62	244	0.0092
33	-2.82	232	0.0052
44	-3.23	235	0.0014
45	-3.37	241	0.0009
46	-3.14	226	0.0019

TABLE 4.4: Results of two-sided t-tests on HbR activation values between Pre-Readers and Readers in channels 32, 33, 44, 45 and 46. The negative t-values indicate that the mean of activation in Pre-Readers is significantly lower than the mean of activation in Readers, in each of the tested channels.

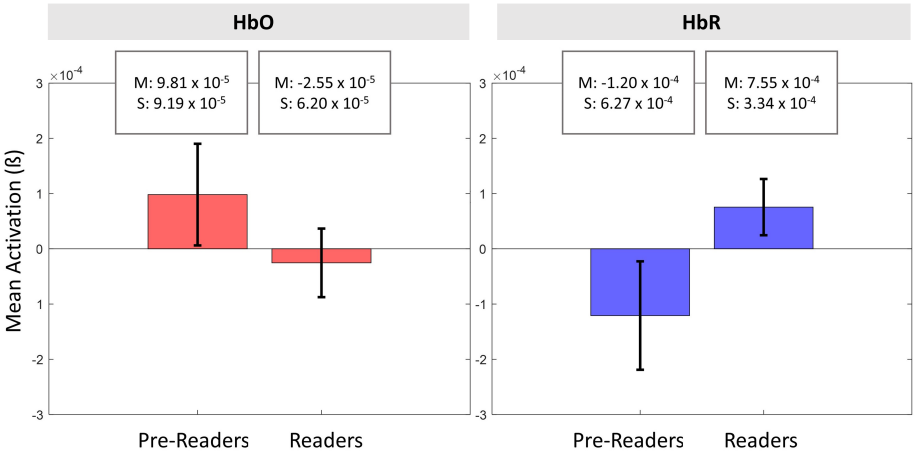


FIGURE 4.2: Bar plots of activation values averaged across conditions, channels (32, 33, 44, 45 ,46) and subjects within each group; bar plots are reported for both hemoglobin forms to show that, although the test did not prove statistically significant on HbO values, they nevertheless mirror the HbR results, in that they both indicate a higher hemodynamic response in these RH channels in Pre-Readers as compared to Readers.

As for the main effect of Stimulus Type, the Mixed-Effects models showed that a few channels, all located in the LH, mainly associated with the Middle Temporal Gyrus (BA: 37, channels 1, 2, 3, 11, 12, 13) are significantly associated with a differential activation between Words and Non Word, but this effect does not hold when the FDR correction is applied. Therefore, these channels were not followed up to. Finally, none of the channels resulted statistically significant for an interaction effect of Stimulus Type x Group, as opposed to what initially was hypothesized.

4.3.2 Block averages

To visualize the time course of activation, blocks were averaged from the onset of the presentation of the stimulus with a duration of 15 sec. A period of half a second before the onset was used for baseline correction.

To highlight the Group effect resulting from the statistical analyses, blocks were averaged over subjects within each group, irrespective of stimulus type, and these block averages are shown in Figure 4.3 along with the scalp map of F-values reported in Table 4.4. The shaded errorbars in the block averages' plots represent standard errors of the mean across subjects within each group. Scalp plots were produced with the NFRI toolbox [85, 120].

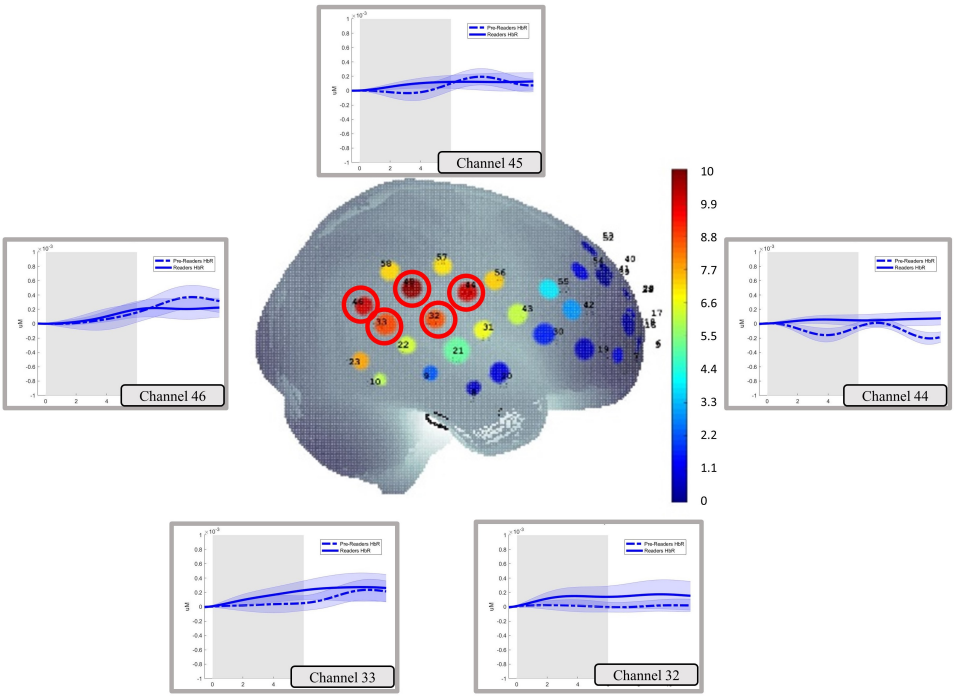


FIGURE 4.3: (**center**) Scalp plot of F-values resulting from the ANOVA test on the results of the Linear-Mixed Effects model on HbR beta values. From this analysis, a main effect of Group resulted statistically significant for channels 32, 33, 44, 45 and 46, whose block averages are reported around the scalp plot. The plots confirm that in these channels, hemodynamic activation is greater in Pre-Readers than in Readers.

4.4 Discussion

Cortical activation to auditory stimuli is known to be related to changes in brain organization as children develop their reading abilities [52]. In particular, we know from fMRI studies that, while beginning readers employ a wide network of brain areas distributed over both hemispheres, the advancement of reading skills is associated with a shift in the recruitment of brain regions towards the left hemisphere (LH). Also, we know that while skilled readers are able to repeat effectively both words and non-words, less skilled readers have a hard time with the latter ones.

With these two concepts in mind, the ultimate ambition of this project is to explore the feasibility of developing an "fNIRS-based predictor" of literacy: namely, we aim at investigating whether it is possible to assess someone's reading skills based solely on the analysis of their brain patterns in response to words and non-words. In order to do this, we had first of all to establish if it is possible to observe, with fNIRS, the changes in brain organization associated with reading development that have been observed with fMRI, namely, the shift from bilateral to left-hemispheric brain activation. Secondly, we wanted to see if participants' reading skills could be correlated to different patterns of activation in response to non-words.

To this end, we recorded the fNIRS signal of children of various ages while they were passively listening to words and non-words.

We divided participants into two groups based on their reading abilities: namely, the first group included children who were able to read up to the simplest 20 words, while the second group included children who could read more words. We called the first group "Pre-Readers" and the second group "Readers".

It is important to note here that groups were clearly not balanced with respect to the age factor, i.e., the mean age of the Pre-Readers group was much lower than that of Readers (M: 4.9, S: 0.7 for the Pre-Readers VS M: 6.1, S:0.8 for the Readers), but they were characterized by similar IQ (M:109.9, S: 13.6 for Pre-Readers, M:110.7, S: 13.9 for Readers) .

The fact that ages were very different in the two groups can be seen as a limitation of this analysis, but for the goals of this project this was not a problem: indeed,

we were not interested in distinguishing the effect of brain maturation from reading development, but instead in exposing any kind of group difference in the perception of the non-words, irrespectively of the source of this difference.

In the first steps of this project, described in this chapter, we employed a very standard analysis procedure: pre-processing included motion artifact correction, removal of global systemic component and frequency filtering; at subject-level, timetraces were modelled with a General Linear Model with the two experimental conditions, Words and Non-Words, as predictors. Coefficients resulting from subject-level statistics were then employed at group-level, to explore whether there was any significant main effect of Group (Pre-Readers vs Readers), or Stimulus Type (Word vs Non-Word), or, most interestingly, an interaction effect of Group x Stimulus Type.

This analysis showed an important main effect of Group in the brain areas of the RH associated with language processing: the Middle and Superior Temporal Gyri, the Fusiform Gyrus, the Pars Opericularis and the Pars Triangularis. In these areas, we found that Pre-Readers exhibit a significantly greater hemodynamic activation compared to Readers, irrespective of the type of stimulus. This finding is very important because it proves that changes in brain organization during reading development can indeed be observed also with fNIRS, and that has undeniable benefits: the possibility of observing the developmental trajectory opens up the way to using fNIRS for following up a reading remediation program, for example.

The second goal of this project, however, could not be pursued: the analysis did not show any significant interaction effect between Group and Stimulus type, namely, it was not possible to detect the differential activation pattern in response to Non-Words between the two groups that was initially hypothesized. Therefore, another analytical approach was explored that classifies distributed patterns of neural activity in response to the different stimuli, and it is the object of Chapter 5.

4.5 Lessons learned

- Pre-Readers exhibit a greater hemodynamic activation than Readers in certain channels of the Right Hemisphere, when listening to any auditory token – regardless of their meaning.
- The group effect was especially circumscribed to the Right Middle and Superior Temporal Gyri, the Fusiform Gyrus, the Pars Opericularis and the Pars Triangularis, all known to be associated to language processing.
- The group-level analysis did not show a differential activation to non-words depending on the literacy skills.

Chapter 5

Decoding the mental representation of word meaning in children to predict their literacy skills, with *multivariate* analysis of fNIRS data

5.1 Introduction

In the previous chapter, a literacy study was introduced where participants of various ages and literacy abilities were presented with purely auditory stimuli including only two conditions: words and non-words.

The goal of the study was twofold: on one hand, we aimed at confirming that learning to read is accompanied by an increasing lateralization towards the Left Hemisphere of the functions associated to language processing. On the other hand, we also wanted to test whether it was possible to determine, based solely on the fNIRS patterns elicited by the two conditions, the literacy abilities of the participants. Our hypothesis is motivated by the notion, confirmed by several fMRI and PET studies, that Non-Word reading and Non-Word verbal repetition is more difficult for non- and beginning readers [124]; a correct non-word repetition requires indeed a phonological processing capacity that is sculptured by learning to read and write [125]. For this reason, non-word repetition is known to be strongly predictive of reading outcomes [52].

Motivated by this understanding, our second goal was to test whether this increased processing effort - that at behavioral level is manifested in failure in non-word repetition – is encoded in the fNIRS signal. In other words, we wanted to test if it is possible to distinguish, without using the behavioral evidence, the mental representations of words and to non-words and, if such a dissimilarity exists, to use it for predicting reading outcomes.

The first goal was achieved in the previous Chapter: it was possible, indeed, to expose a more bilateral activation in the Pre-Readers, as opposed to the skilled Readers. This finding is completely consistent with the literature on the topic and it is the first time, to our knowledge, that such result has been revealed with fNIRS and by means of such a simple experimental setup.

The second point, however, was left unresolved: with the applied approach it was not possible to bring to light a differential activation for the two groups in response to the two classes of stimuli. Therefore, we followed-up these results with another analytical approach: Multi-Variate Pattern Analysis (MVPA).

MVPA is an approach aimed at decoding neural activity represented across a distributed group of channels, by means of powerful classification algorithms [90].

Several studies have shown that these methods can indeed uncover information that is not revealed by conventional univariate statistical analysis [126, 127, 128, 129], simply because distributed patterns of neural activity may encode significant information without producing a robust contrast when considered in a univariate fashion.

For this reason, the use of these techniques has become much widespread especially in the EEG field , and gradually to the analysis of other neuroimaging data like fMRI [91, 130], MEG [131] and also extracellular recordings [132]; these studies have confirmed that the distributed patterns of brain activity can decode many different classes of stimuli [7, 8].

The goal of this work was to use MVPA to classify patterns of hemodynamic responses to Words and to Non Words, both within-subject and within-group.

Consistently with the work presented in Chapter 3, a classifier was implemented that used features drawn from the simultaneous variation of both oxygenated and deoxygenated components of the hemodynamic response measured with fNIRS.

Furthermore, the classifier was developed based on Regularized Linear Discriminant Analysis (LDA), as it was done in Chapter 3, but also on Support Vector Machines (SVM), in order to compare the results yielded by the two methods.

As for the within-subject analysis, we explored whether the individual classification performances were representative of the lexical skills of the individuals, by means of correlation analysis and also by comparing the averages over groups of subjects.

The within-subject analysis was severely limited by the number of trials available for the trial-by-trial classification; therefore, we also applied the classifier to trials pooled across subjects, after they were grouped according to their literacy skills ("within-group" classification).

This chapter builds upon and extends the findings of the following publication:

[11] Gemignani J., Bayet L., Kabdebon C., Blankertz B., Pugh K.R, Aslin R.N. (2018), Classifying the mental representation of word meaning in children with Multivariate Pattern Analysis of fNIRS, 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)

5.2 Methods

5.2.1 Participants

Participants were the same as those described in Chapter 4. Out of 83 subjects, 11 were excluded because they completed less than 8 trials per condition.

In addition to this, 7 (seven) more subjects were discarded from the analysis because, based on the quality check of the timeseries described in Section 4.2.6, they had more than half channels with bad quality.

Therefore, the total number of participants used in this analysis was 65 (sixty-five). Table 5.1 summarizes the characteristics of the new subset:

Measure	
n	65
Age (years, M=mean, S= standard deviation)	3.7÷7.7 (M: 5.5, S: 1.1)
Gender (Male:Female)	35:30
IQ	74÷134 (M:110, S:13)
Letter-Word Decoding (WJ)	4÷55 (M:20.8, S: 13.7)

TABLE 5.1: Participant's summary information

Pre-processing

The pre-processing of the data was identical to the one described in Chapter 4. For the purpose of trials-classification, after the band-pass frequency filter, time traces were segmented into trials, with a time window of -0.5 s before each trigger to 15 s after; each trial was then baseline corrected by subtracting the mean value over the 0.5 seconds before the trigger. This ensured all trial to be aligned at time 0. The grand averages of the epoched time traces are depicted in Figure 5.1.

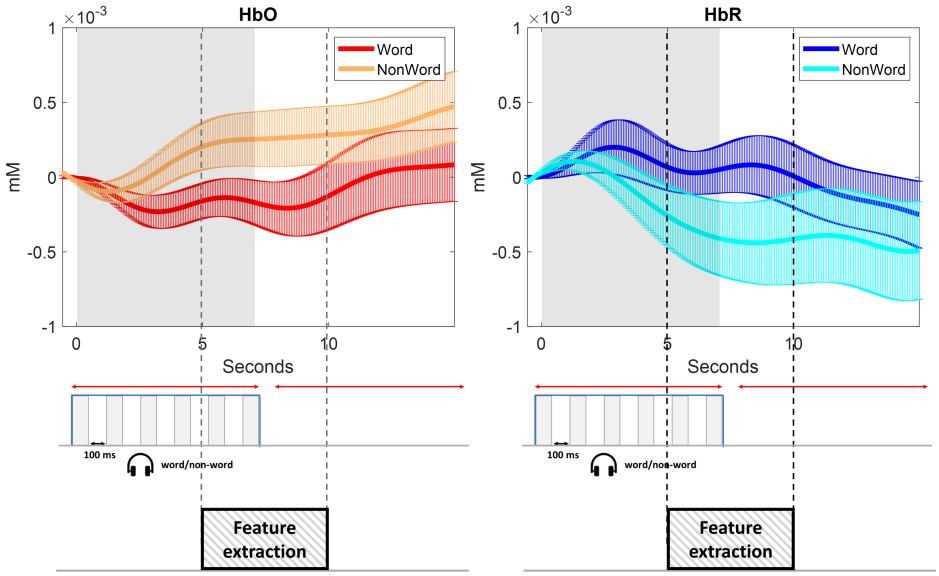


FIGURE 5.1: Grand averages of the epoched timeseries; timeseries were averaged across trials, channels and finally subjects. Errorbars represent the standard error of the mean across subjects. The shaded area of the plots represent the period of stimulation of 7 seconds, during which participants passively listened to the auditory stimuli (Words or Non Words).

5.2.2 Within-subject classifier

Feature extraction

Features were extracted for each block and for each channel from both HbO and HbR time traces, separately. The largest peak values of HbO and HbR were computed within the time window from 5 to 10 seconds after onset of the stimulus, since it's the time interval where the peak of the hemodynamic response is most likely to occur [34].

To ensure the meaningfulness of the features in representing the hemodynamic activation in response to the task, the peak value was computed as the largest positive value for HbO and the largest negative value of HbR, because HbR is expected to have a negative deflection in response to neural activation [19, 24].

Each trial's feature vector was then built by concatenating features from each of the selected channels. In order to account for the fact that the two hemispheres may contribute differently to lexical processing, and to explore the possible impact of this factor on the classification results, three different schemes were employed for channel selection and classification: in addition to using all channels (*whole-brain*), we also employed, for each hemisphere, a combination of several anatomically-defined regions of interest. Table 5.2 describes the locations of each of them.

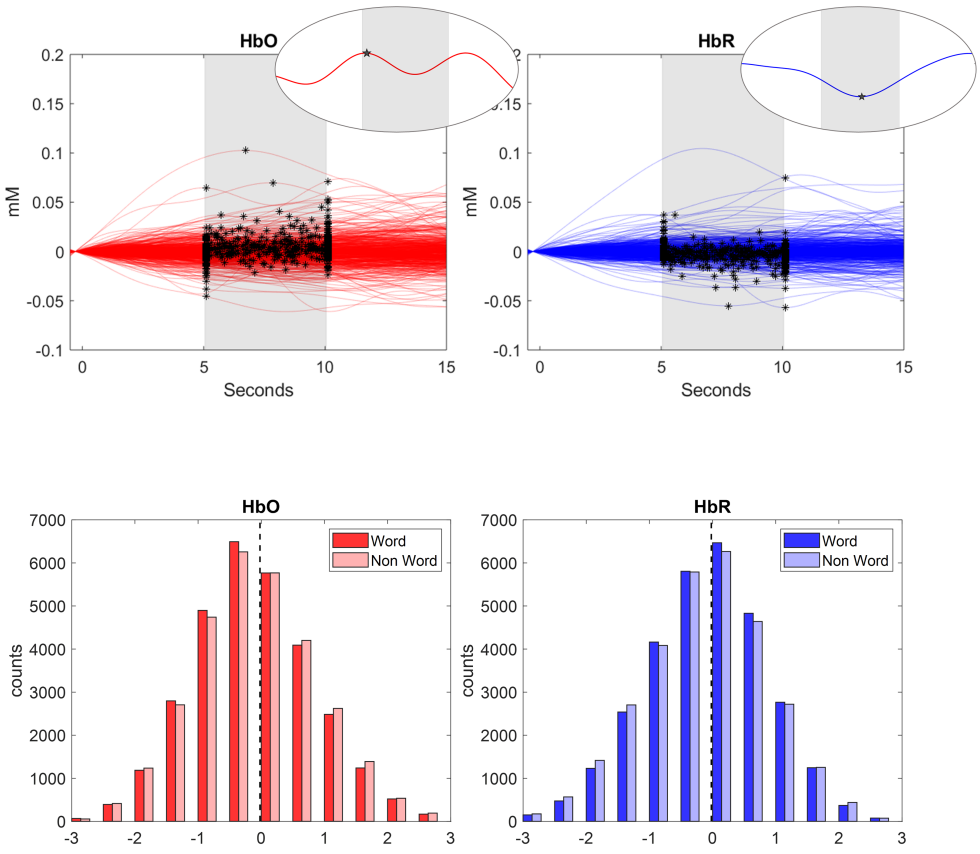


FIGURE 5.2: (**Top**) Features are extracted from each block of HbO and HbR timeseries as the positive and negative peak value, respectively, within the time window [5-10] seconds after the presentation of the stimulus. Features were normalized by removing their mean value and dividing by their standard deviation. (**Bottom**) Distribution of normalized features for each class, Word and Non Word.

Channel subset ID	**Description**	**Brodmann Areas**	**Selected channels**
1	Whole brain	–	1...58
2	(RH) Fusiform Gyrus, Middle Temporal Gyrus; Inferior prefrontal G., Superior Temporal G, Supramarginal G., Angular G., Pars Triangularis, Pars Opericularis	37, 21; 47, 22, 45, 39, 40, 44	1, 2, 3, 11, 12, 13, 14, 15, 24, 26, 27, 34, 35, 37, 47
3	(LH) Fusiform Gyrus, Middle Temporal Gyrus; Inferior prefrontal G., Superior Temporal G, Supramarginal G., Angular G., Pars Triangularis, Pars Opericularis	37, 21; 47, 22, 45, 39, 40, 44	8, 9, 10, 19, 20, 21, 22, 23, 30, 31, 33, 43, 45, 46, 58

TABLE 5.2: Classification was performed on brain patterns distributed over the whole brain, plus two hemispheric macro-regions, that were identified as groups of adjacent anatomically-defined regions of interest.

Features extracted from HbO and HbR timetraces were concatenated into a single feature vector; each feature was then normalized, by removing its mean value and dividing by its standard deviation. Figure 5.2 shows the procedure of features' extraction and their distributions after normalization.

Classification of Multivariate Patterns

The obtained trials were classified using both Linear Support Vector Machines as well as Regularized Linear Discriminant Analysis.

Linear SVMs were implemented in MATLAB 2017b with the libsvm-3.11 library [133], while Regularized LDA was implemented via tools available in the Berlin Brain-Computer Interfacing (BBCI) Toolbox [104].

In both cases, cross-validation was performed using 4 folds and 100 permutations: the sixteen blocks were randomly divided into 4 folds, with each fold containing 2 blocks for each class. At each repetition, three folds and the remaining fold were used for training and testing, respectively. At each repetition, the order of the trials was permuted. For each subject, classification accuracies were averaged across the permutations.

5.2.3 Analysis of individual classification accuracies

The achieved classification accuracies were evaluated both at subject-level and at group-level.

At subject-level, we wanted to test, for each participant, the statistical significance of the achieved accuracy; to this end, the empirical subject-level null distribution was estimated by classifying trials after classes' labels were randomly shuffled; to produce a robust distribution, the procedure was repeated 1000 times. After that, each participant's observed accuracy was tested by computing its p value as the proportion of the null distribution greater than, or equal to, the observed value.

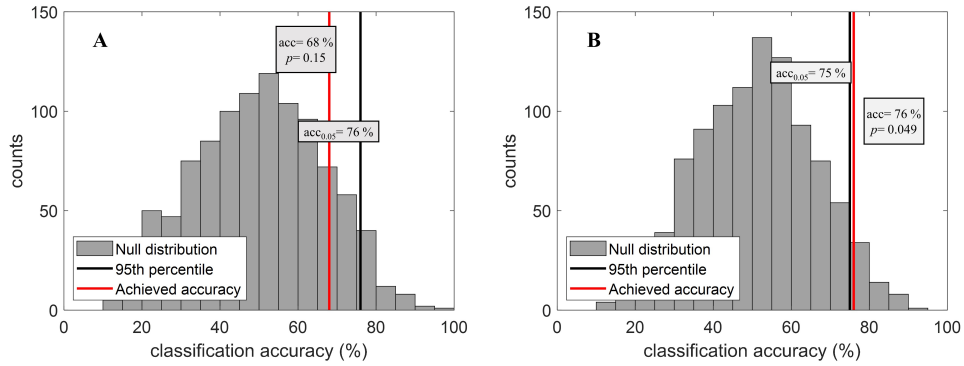


FIGURE 5.3: Examples of how individual classification accuracies were statistically assessed. In Panel **A** an example of individual performances that did not reach statistical significance, while in Panel **B** the individual classification accuracy of 76% was higher than the 95th percentile (75%), obtaining a p-value of 0.049.

At group-level, we were interested in exploring if and how classification accuracies were related to the reading abilities of the participants. Therefore, firstly, we employed the Reading Score as a continuous variable, by fitting a Mixed Effect Linear (LME) Model

with a random intercept for each participant, and Reading Score as fixed effect, for each subset of channels. The model was fitted in MATLAB 2017b.

Secondly, the Reading Score was used to divide participants into two groups, in the same fashion used in Chapter 4. For each area, an unpaired t-test was used to compare classification accuracies achieved by the two groups (Pre-Readers vs Readers).

5.2.4 Within-groups classifier

One of the major limitations of the within-subject classification is posed by the low number of available observations, especially considering the large number of features (116, in case of the *whole-brain* scheme). To mitigate this issue we also performed the same Word vs Non-Word classification on the trials pooled within the two groups (Pre-Readers and Readers). Within each group, each individual's trials were concatenated together; a feature matrix of 116 x 544 was obtained for the Pre-Readers and a 116 x 496 for the Readers.

The obtained trials were classified using LDA and SVM; in both cases, cross-validation was performed using 4 folds and 100 repetitions; at each repetition, trials were randomly permuted, and classification accuracies were averaged over folds and repetitions.

Additionally, for both groups, the null distribution was generated by repeatedly randomly shuffling class labels, and applying the same classification routine (1000 repetitions).

5.3 Results

5.3.1 Overall performances: SVM vs LDA

Across the whole pool of subjects and all brain areas, classification accuracies yielded by SVM and LDA did not result significantly different.

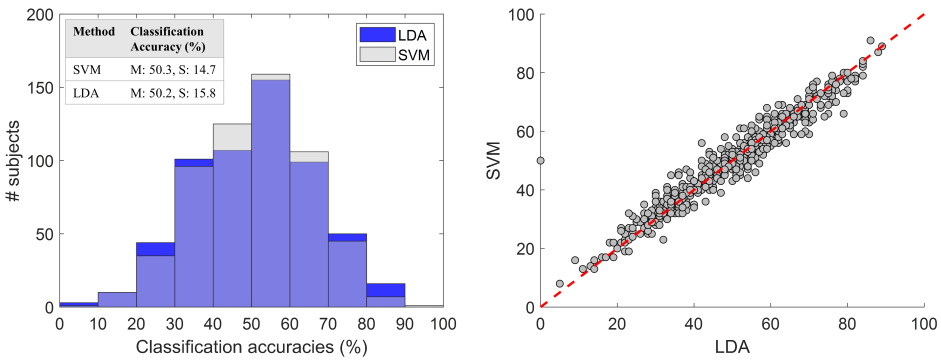


FIGURE 5.4: (**Left**) Distribution of classification performances across all subjects and all channel groups. (**Right**) The performances of the two algorithms did not significantly differ.

Even when considering the three employed channel subsets separately, the two methods performed similarly (Figure 5.5).

For this reason, in the subsequent sections only results yielded by LDA will be presented, but very similar analogous results were found by using SVM-based accuracies.

5.3.2 Individual performances

At individual level, statistically significant classification accuracies were achieved only for few participants. In particular, when whole-brain channels were employed, only Subject # 17 had a significant classification accuracy (79%, p= 0.038); when LH channels were employed, only Subjects # 11 and # 74 had significant accuracy (84%, 80% , respectively, and p= 0.007 and 0.033); finally, with RH channels, Subjects # 11, #74 and #90 had significantly high accuracies (76%, 77%, 79% with p= 0.040, 0.047, 0.045, respectively).

5.3.3 Correlation between individual decoding accuracies and reading abilities

The correlation analysis between achieved classification accuracies and individual reading scores was performed by means of Linear Mixed Effects (LME) models. In neither

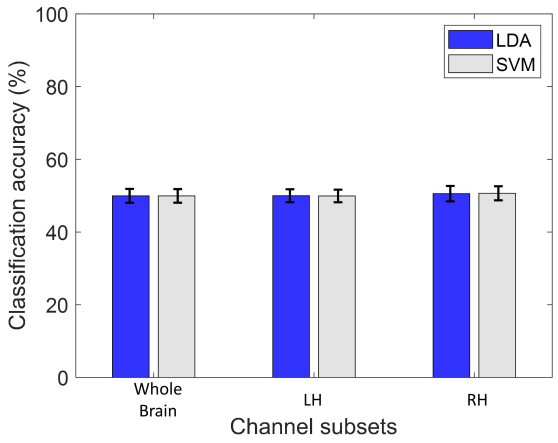


FIGURE 5.5: Classification accuracies achieved by LDA and SVM within the channel subsets defined in Table 5.2

of the three combinations of channels was the correlation between accuracies and scores significant at p= 0.05; correlations where found to be positive when using LH channels and negative when using RH channels:

Channel group	**Estimated regression coefficient**	**Standard Error**	***p***
Whole Brain	0.03	0.18	0.8548
Whole LH	0.26	0.16	0.1232
Whole RH	-0.10	0.20	0.6396

TABLE 5.3: Results of the LME models performed within each channel group ('$Accuracy \sim ReadingScores + (1|Participant)$')

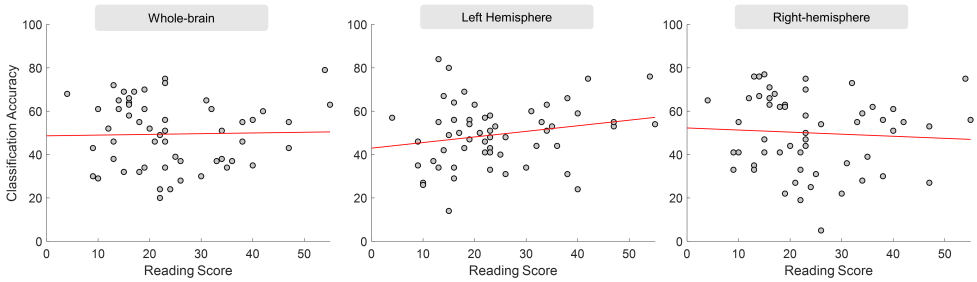


FIGURE 5.6: In none of the three employed channel groups was the correlation between Reading Scores and Classification Accuracies statistically significant at p=0.05; nevertheless, a pattern can be qualitatively detected of negative and positive correlations in the RH and LH, respectively.

In light of the fact that the Reading Score is highly correlated with the age of the subjects, an additional analysis was performed with the intent of removing the effect the age from the reading scores, by means of Partial Least Squares regression (Figure 5.7.

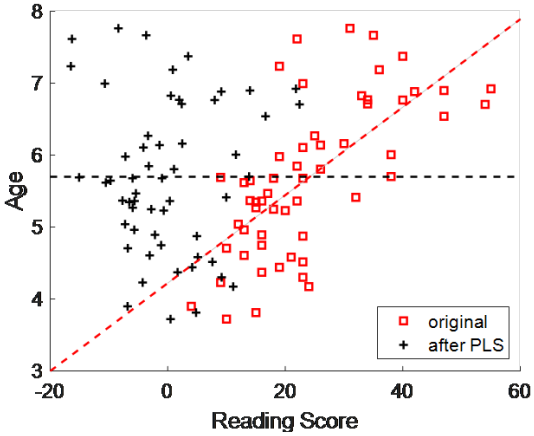


FIGURE 5.7: Effect of removing the effect of Age from the original Reading Scores, using Partial Least Squares (PLS) regression. The dashed lines represent the least squares correlation lines (r= 0.67 for the original scores, r= 0 for the "corrected" scores).

Then, the LME analysis with Classification Accuracies as the response variable was run again, this time using the residuals of the regression between age and reading scores as a predictor. The models yielded similar results, namely, negative and positive correlations, but not statistically significant, in the RH and LH, respectively.

Finally, we also employed the Reading Score to group participants in two groups Pre-Readers and Readers, to reproduce the conditions we employed in Chapter 4 for the univariate analysis. Then, classification accuracies of the two groups were compared using two-tailed unpaired t-tests.

Channel group	***t***	**95% Confidence Interval**		***p***
		Lower	**Upper**	
Whole Brain	1.665	-1.250	13.778	0.100
Whole LH	0.215	-6.399	7.943	0.830
Whole RH	2.248	1.031	17.490	0.026

TABLE 5.4: Results of unpaired t-tests between groups (Pre-Readers vs Readers), divided at ReadingScore=20 (degrees of freedom = 63).

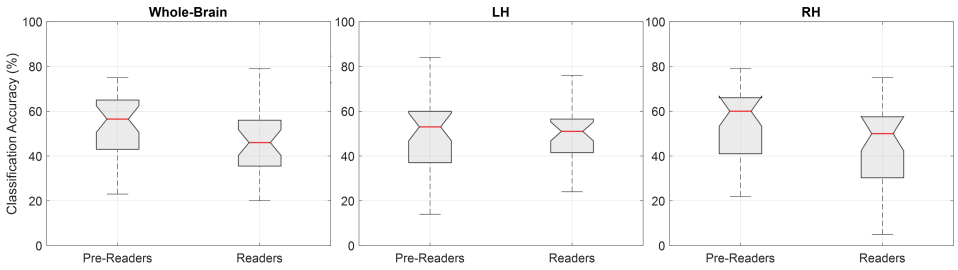


FIGURE 5.8: Boxplots of individual classification accuracies within the two groups Pre-Readers vs Readers; the central red line in each box indicates the median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent the whole range of observed values. In the RH, the unpaired t-test revealed a significant group difference, with Pre-Readers achieving better accuracies than Readers (60% vs 50%, respectively).

5.3.4 Within-groups classification

The results of the two within-groups classification support the findings already observed: the classification accuracy reached in the Pre-Readers group is slightly better than that reached in the Readers, but still not a good accuracy: the mean value is only 53.2%.

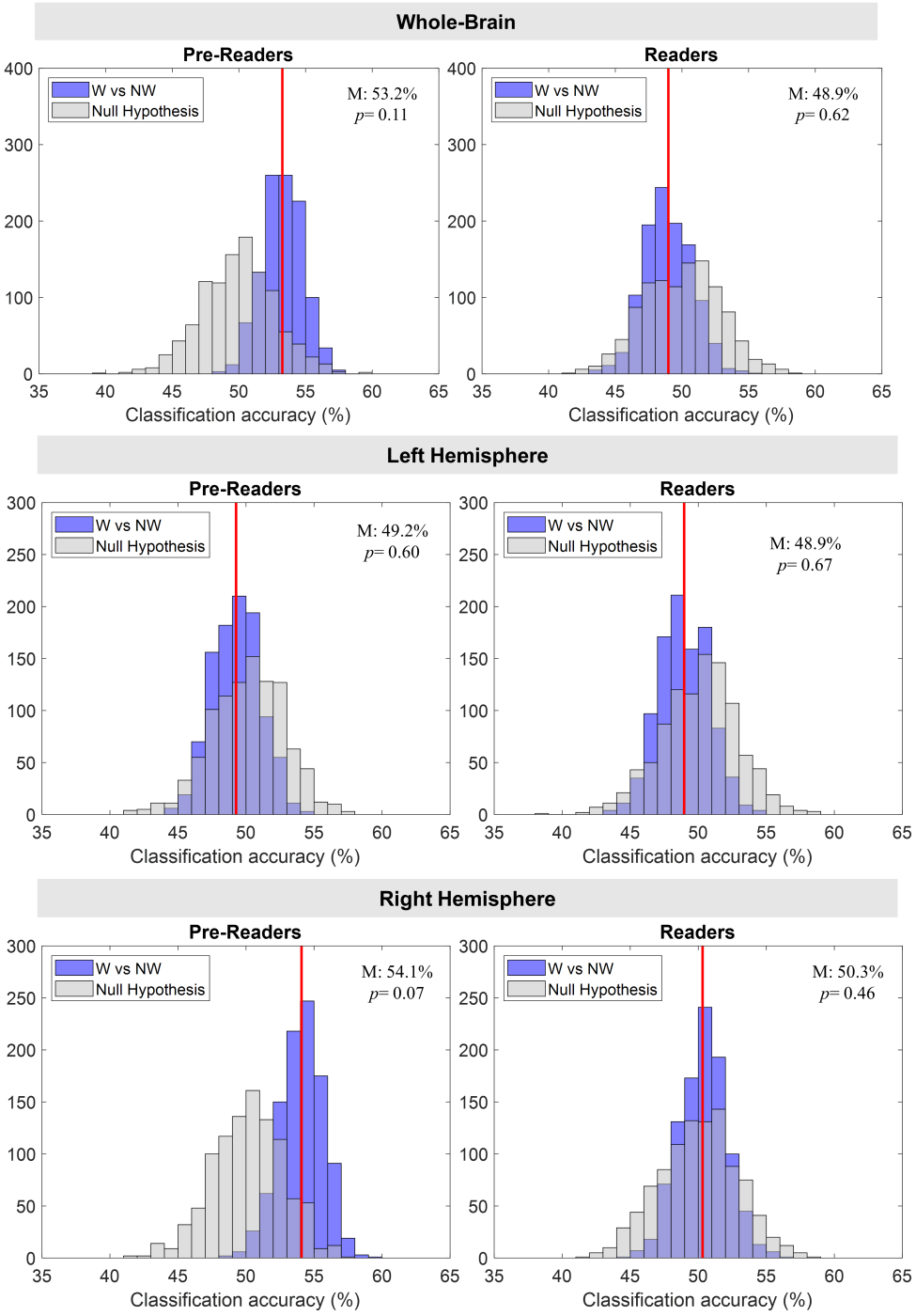


FIGURE 5.9: Results of within-groups classification. Classification accuracies obtained using trials from Readers are not statistically significant for any of the channel subsets, compared to their corresponding null distributions. For the Pre-Readers, accuracies are marginally significant when using *whole-brain* channels and more so when using RH channels, but still they are not high (53.2% and 54.1%, respectively).

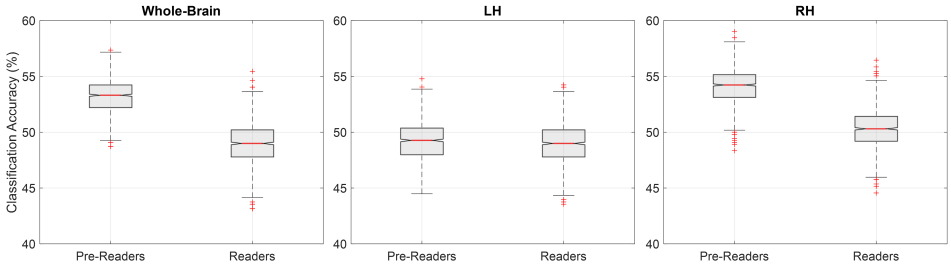


FIGURE 5.10: Boxplots of within-groups classification accuracies, for the three selections of channels; the central red line in each box indicates the median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent the whole range of observed values. The unpaired t-tests confirm a significant difference between the two groups when using channels from the RH and also Whole-Brain; however, the mean classification accuracy does not exceed 54.1%.

5.4 Discussion

Beginning readers make a greater effort, compared to skilled readers, when verbally repeating non-words; this is due to the fact that correct non-word repetition requires the emergence of a phonological processing capacity that is carved by the very process of learning to read.

In this project, we explored whether this differential processing capacity – that at behavioral level is manifested with a differential outcome in verbal repetition – is also reflected in the fNIRS signal measured when the subject is only listening to words and non-words.

To this end, we implemented a neural decoder, based on both LDA and SVM, and applied it to the classification of distributed patterns of hemodynamic activity elicited by word and non-words.

In accordance to the findings presented in Chapter 3, the classifier was trained to use features from both the HbO and the HbR component of the fNIRS signal. To our knowledge, this is the first attempt at using both hemoglobin components in the

context of a multi-channel analysis of fNIRS, thus it extends the work done so far in this field.

Also, to keep into account the possible spatial specificity of the processes under investigation, not only patterns distributed over the whole brain were examined but also, separately, patterns distributed only over the Left and over the Right Hemisphere were used for classification.

The classification was performed at two different levels: firstly, it was performed within-subject. This allowed to obtain a value of classification accuracy for each individual, that could be used to interrogate the relation between decoding and literacy skills.

The first interesting result is that the performances of LDA and SVM were basically equivalent; we were interested in this comparison because in our own previous work [11], following relevant literature, we had employed only SVM, but several other works reported that regularized LDA and linear SVMs tend to have both good and not significantly different performances [134, 135]. Our results supported this observation, and for this reason we focused on LDA for the remainder of the work.

Nonetheless, the within-subject classification yielded discouraging results: ranging from 20% to 79%, the average accuracy across all subjects did not exceed chance level; furthermore, when comparing each subject's observed accuracy with their respective null distributions, only one subject turned out to have a significant accuracy at $p<0.05$.

The large variability of the classification outcomes may suggest that the individual subjects' data quality was inconsistent, or also that the individuals' response patterns to the stimuli differed too much from one another.

Another possible explanations can be found in the limited number of trials available for classification: having only 8 trials per classes, the experimental design was probably not well suited for within-subject classification, even more so considering that, by concatenating HbO and HbR features, the vector included 116 features.

Furthermore, it becomes apparent from Figure 5.2 that using the peak value of each epoch as a single feature may not necessarily have been representative enough: rather, it would have been better to sample several values throughout the whole trial, as we did in Chapter 3 for example. While that would likely capture the characteristics of the trials much better, it would increase the dimensionality of the features far too much to

be feasible, considering the limited number of observations available and the number of channels.

Finally, the large variability could actually be representative of the question under investigation, namely, it could be related to individual differences in the mental representation of non-words and thus to individual differences in reading abilities. To test whether this was the case, we performed a correlation analysis between the achieved accuracies and the reading scores, but in neither of the three schemes for channel selection was the correlation found significant.

However, when looking at accuracies averaged over the groups of subjects, divided using reading scores, a significant group difference was found when using channels from the Right Hemisphere (p= 0.026), and a marginally significant difference was found for the "Whole-Brain" scheme (p= 0.1).

These results, however weak, show a tendency toward a greater word vs non-word difference in the Pre-Readers; when taken as a group, they exhibit indeed a higher decoding accuracy than Readers, especially in the Right Hemisphere (median: 60% vs 50%, respectively).

To mitigate the excessively unbalanced ratio between the number of features and the number of trials, classification was also performed within-groups: trials within each group, Pre-Readers and Readers, were pooled together so to achieve a higher number of observations. Also in this case, results were weak: only for Pre-Readers, the achieved accuracies were only marginally significant in the "Whole-Brain" scheme (p= 0.1) as well as in the "RH" scheme (p=0.07), with quite low nominal values (53.2% and 54.1%, respectively).

However low these accuracies are, a group difference does definitely exist: for the group Readers the distribution of accuracies across permutations overlaps completely with the null distribution (Figure 5.9) and it is significantly lower than that of Pre-Readers both when using channels from the Whole-Brain and from the RH (Figure 5.10); this observation supports the hypothesis that as children strengthen their reading abilities, the difference in fNIRS patterns between word and non-words lessen.

In this Chapter, we conclude that with multi-variate approaches it is indeed possible to explore finer-grained effects that escape the lens of the univariate analysis. In the

very case of this dataset, the experimental design was probably not optimally suited for this use; we did, however, manage to bring to light a trend, although only marginally statistically significant, that was not possible to observe in the channel-wise analysis.

Future applications of this method should include a careful planning of the experimental setup, so as to be suitable for trial-to-trial classification: especially it should have a sufficient number of trials in consideration of the number of channels to be used.

5.5 Lessons learned

- LDA and SVM performed very similarly.
- Subject-level classification accuracies were largely not significant.
- A statistically significant correlation between subject-level decoding accuracies and reading abilities could not be identified.
- When averaged across groups, Pre-Readers' accuracies were significantly higher than Readers when using patterns from the RH and marginally also when using whole-brain patterns.
- Within-groups classification accuracies differed substantially between the two groups, with Pre-Readers achieving a higher decoding accuracy than Readers, but their nominal value didn't exceed 54.1% at most.
- MVPA has a great potential to bring to light subtle effects that escape the channel-wise analysis, but for it to be really successful a careful planning of the experimental setup is crucial. In particular, a number of trials should be administered high enough to allow trial-by-trial classification within each subject.

Chapter 6

Summary and Conclusions

As the use and application of fNIRS is becoming increasingly widespread, the need for an *ad-hoc* standardized data analysis procedure is now more pressing than ever.

A long way of work and research is still required to establish guidelines and automated procedures. Nevertheless, much effort is being made in the direction of developing new analysis methods and systematically comparing them to traditional ones, in quest of defining a recipe for achieving the best and most accurate results. This doctoral thesis is framed within this attempt.

In particular, the goal of this work was to examine and tackle some of the most critical sources of concern of a typical data analysis workflow, in particular when the subjects are children.

The first one was the inter-subject variability of the hemodynamic response, which means that different subjects may not only respond differently to the same task but that, when responding to the task, their hemodynamic responses may themselves vary significantly in shape and characteristics. Furthermore, there is also a certain extent of intra-subject variability: due to changes in muscles' and small capillaries' configurations, different brain areas are sometimes characterized by hemodynamic responses of different shapes. These two issues together speak to the fact that the model-based type of analysis is oftentimes inaccurate; an a-priori fixed HRF model can hardly capture this natural variability and actual neural activity may remain undetected. This is especially relevant with children, since they often display a non-typical hemodynamic response, that is even more difficult to model.

Another source of complication is given by the choice of chromophore: by analyzing separately oxy- and deoxy-hemoglobin, it is rarely the case that the results on one chromophore do precisely mirror the results on the other. This issue is problematic, for instance, for the statistical test at group-level: since the hemodynamic fingerprint is so subject-specific, some individuals may naturally show stronger results with one chromophore while other individuals with the other one. In such a situation, two separate group-level analyses for the two chromophores may prove highly ineffective in uncovering the existing effects.

This doctoral work has been as a two-steps process: in the first part (Chapter 3) these problems have been tackled in a very theoretical fashion, while the second part (Chapters 4 and 5) provided a direct application of the newly proposed method to experimental data collected in the context of a literacy development study.

In the first part, we proposed to classify single channels as "active" or "not active", using Linear Discriminant Analysis. Features included amplitude and slope of both oxy- (HbO) and deoxy- hemoglobin (HbR). To precisely quantify the performances of this method, a synthetic dataset was generated in which hemodynamic activations of various sizes and shapes were added to physiological resting state signal. Moreover, part of the simulated hemodynamic responses were also artificially affected by Meyer waves; this allowed to compare the uni- and multivariate pipelines also in terms of robustness against this physiological confound.

The classification was conducted, within each subject, by comparing data from different temporal segments of the same recording ("resting state" vs "task"). As for the univariate approach, data was analyzed by fitting a General Linear Model (GLM) to HbO and HbR traces, separately. In terms of performances, the multivariate classifier yielded a significantly greater classification accuracy than GLM, both when applied to HbO and HbR.

Furthermore, the performances yielded by the (HbO+HbR) classifier also resulted less susceptible to individual variations and physiological confounds. When considering the correlation between achieved accuracy and demographic characteristics like Hair Color, Age, Gender and Time of Measurement, we found that none of these correlated significantly with the classification accuracies achieved by the multivariate classifier,

while accuracies achieved by GLM on HbR were significantly higher in blond-haired subjects compared to brown-haired.

This is a clear indication that a model-based analysis is an approximation that cannot capture the intrinsic variability that exists between and within subjects, while a data-driven strategy that works on the comparison "rest" vs "task" not only within the same subject but also within the same channel enhances the detectability of effects, also when they are small in comparison to other sources of variability.

In the second part of this project, the comparison between the two pipelines was put in practice by analyzing a large dataset of fNIRS data collected in the context of a literacy study. The goal of this project was twofold: on one hand, we aimed at highlighting a differential functional lateralization in response to auditory stimuli depending on literacy levels: since the visual system of a young reader specializes gradually, during the first stages of reading both hemispheres should functionally activate, even when only listening to speech, while with growing reading expertise, functional activation should be progressively more focused and should slowly converge to the left occipito-temporal areas.

The other goal of the project was to reveal a differential activation, depending on literacy levels, to non-words: this hypothesis comes from the notion that as reading increases phonological awareness, skilled readers can rely on the phonemic code they acquired and can easily process speech sounds, even when they are meaningless; on the other hand, beginning readers do not have yet this phonemic code, therefore they struggle in processing and repeating non-words. Therefore, we explored whether it was possible to distinguish various levels of literacy only by looking at hemodynamic responses to non-words.

To pursue these goals, both analysis approaches were employed: first, a General Linear Model was fit to the data at subject-level; these results were fed into a group-level model looking to predict the main effects of Group (Pre-Readers, Readers), Stimulus (Word, Non-Word) and the interaction effect Group x Stimulus. This step was successful in that it clearly highlighted that Pre-Readers exhibit a more bilateral functional activation than Readers.

This first result is a very important lesson of this project that should not go overlooked:

the channel-wise analysis has many limitations, that we extensively listed, but has the unique ability of providing a clear localization of the brain regions reflecting the cognitive process under investigation. With this respect, univariate methods have the advantage of being directly interpretable from a neuroscientific point of view.

However, in our experiment, it failed to expose a differential activation of one group to one type of stimulus. Therefore a multivariate approach was undertaken.

In doing so, the concept of the classifier proposed earlier was taken even further: while in the first part it was multivariate in that it combined features from the concurrent variations of HbO and HbR, here it was also multichannel; the goal was, indeed, to use HbO and HbR features from the whole pattern of channels to classify trials of Words and Non-Words, with the expectation that classification accuracies would be higher for lower levels of literacy skills. In addition, another extension we made to the previously developed classifier was that this time it was also implemented with Support Vector Machines (SVMs), for comparison with LDA.

The results of this analysis were admittedly underwhelming: when compared to their respective null distributions, individual classification accuracies were largely not significant. Furthermore, the correlation analysis between classification accuracies did not reveal a significant relation between different activations between Words and Non-Words and reading scores, when classifying patterns distributed across the whole brain.

However, when using patterns distributed over the right hemisphere and averaging accuracies across pre-defined groups, we did find that Pre-Readers' accuracies were significantly higher than the Readers'.

There are a few possible explanations for these mediocre results; one is that the starting hypothesis is flawed: beginning readers and illiterates do have a hard time processing, memorizing and repeating out loud pseudo-words, and this is a known fact that has been proven by several studies. However, this may not necessarily translate into a much different functional activation to non-words with respect to real words, and so into a high decoding accuracy, especially only at auditory level. This kind of neuroscientific considerations extend beyond the scope and the field of operation of this doctoral work.

Another, very likely, explanation has to do with the experimental design: with only 8 trials for each class, 58 channels and 2 features for each channel (HbO and HbR),

the feature matrix ended up having 16 observations and 116 features. This ratio was definitely excessively unbalanced and so, in order to mitigate the issue, an additional classification was performed by pooling participants' data within each group (Pre-Readers and Readers). This way, the feature matrices had a more balanced ratio between number of features and number of observations. In this case, classification accuracies achieved by the Pre-Readers group were marginally higher than those by Readers group, both using whole-brain patterns and RH patterns.

Therefore, MVPA did manage to uncover, if not really a significant correlation, at least a trend in the direction that we initially hypothesized.

This thesis has been an evolution: it identified some of the important challenging factors for fNIRS data analysis; it proposed a new multivariate method to tackle these challenges, and compared it against the most up to date state-of-the-art univariate method. Once the comparison proved successful, it moved onto the next level, by using it for the exploration of very fine-level questions that could not be answered by the "typical" analysis pipeline. This last step, with all its limitations, showed that multivariate and multichannel patterns do indeed have the potential of bringing to light effects that elude the channel-wise analysis.

This doctoral work constitutes a step forward in the scrutiny of the different fNIRS data analysis techniques, and it taught us many different lessons: first, that despite all its limitations, the channel-wise approach constitutes still a cornerstone of any analysis routine, since it's the only option we have to localize the cognitive processes of interest. Channel-wise doesn't necessarily mean GLM: the classifier presented in Chapter 3 also worked on single channels, and achieved great results by (1) using a data-driven procedure and (2) combining HbO and HbR features.

Second, we learned that by combining information from different channels we can decode brain states associated to the experiment, and that decoding models can answer a whole other range of fine-grained research questions. But for this approach to be really successful, it cannot just be applied to any dataset: a well-reasoned experimental design is crucial, especially in terms of number of trials and number of features, if the goal is trial-by-trial classification.

To conclude, advocating the exclusive use of one approach against the other is impossible: single-channel and multi-channel approaches respond to different questions and can actually greatly reinforce one another, when used in conjunction on the same dataset.

Appendix A

Appendix

A.1 Supplementary information of Chapter 3

A.1.1 Correlation between individual measures and classification accuracies

The linear model fitted to the individual classification accuracies revealed a significant effect of Hair Color on the classification accuracies reached by the GLM with HbR. The other predictors that were used were Gender, Time of Measurement, and Age. Reported below are the distributions of classification results for each of these effects.

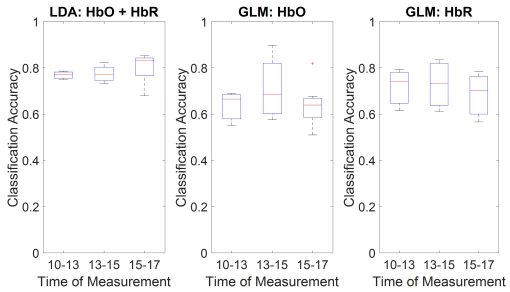


FIGURE A.1: Distributions of classification accuracies grouped by Time of Measurement. The central red marks represents the median values, the blue boxes extend from the 25th to the 75th percentiles, and the black whiskers extend to the most extreme data points not considered outliers (which are marked with red crosses)

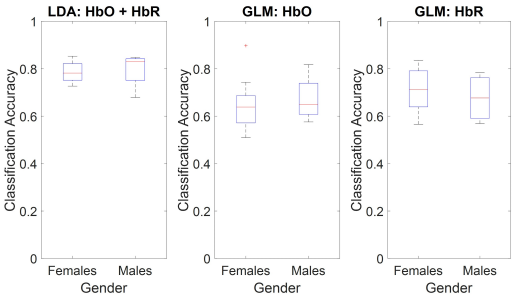


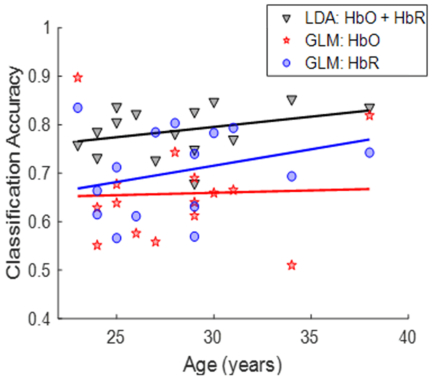
FIGURE A.2: Distributions of classification accuracies grouped by Gender

A.1.2 Experimental results

In this section, individual subjects' channel-wise classification results from the GLM-based and LDA-based analyses are reported.

For each subject, a table is presented, which includes:

- Topographic images of channel-wise output values, namely GLM β coefficient values and LDA classifier outputs. The images were produced with nirsLAB v2017.06 [1]. Large positive GLM(HbO) β values, or large negative GLM(HbR) β values, indicate a good fit of the model to the data, and so a better chance of that channel being statistically significant (i.e., of being classified as "active"). LDA outputs are negative if the channel is classified as "not active" and positive



Estimated coefficient for partial dependency of Classification Accuracy on Age

	ß	Lower Bound	Upper Bound	p value
LDA: HbO+HbR	0.0042	-0.002	0.010	0.176
GLM:HbO	0.0010	-0.012	0.014	0.873
GLM:HbR	0.0068	-0.001	0.015	0.112

FIGURE A.3: (**Left**) Effect of subject age on classification accuracy, according to Linear Mixed Effects analysis. Partial-dependency line for each method is superimposed on the individual-subject data points (**Right**) Coefficients estimated by the LME for Predictor: Age, Response Variable: Classification Accuracy. None of the methods has a statistically significant age dependence on accuracy.

is the channel is classified as "active". Therefore, a high classifier output indicates a good chance that the channel is labeled as "active".

- Block averages of the signal in response to the stimulus (red and blue curves for HbO and HbR, respectively). The shaded error bars indicate the standard error computed over the 16 episodes of alternating left- and right-hand finger tapping. The GLM plots are accompanied by the canonical basis function used by the model; the LDA plots are accompanied by the block averages of the resting-state trials. The block averages are shown for only the channels covering the motor cortex.

A second table reports the numerical output of each analysis, along with the corresponding p-values. Channels flagged as NaN were of poor quality and therefore discarded from the analysis.

	GLM – HbO							GLM – HbR							LDA						
SUB / CH	001	002	003	004	005	006	007	001	002	003	004	005	006	007	001	002	003	004	005	006	007
1	0.6123	0.0001	0.0173	0.0001	0.0465	0.5084	0.9314	0.0279	0.4403	0.0621	0.8131	0.3398	0.0470	0.1678	<0.0005	<0.001	0.1571	0.0019	<0.001	<0.001	0.3769
2	0.9445	<0.0001	0.2381	0.0017	0.1091	0.8705	0.9373	0.4607	0.5035	0.1794	0.4487	0.5527	0.0105	0.5467	0.0231	0.0013	0.6987	0.0090	0.0590	<0.001	0.7500
3	0.7284	0.0057	0.0015	0.0021	0.0338	0.5684	0.9946	0.1718	0.0953	<0.0001	0.3668	0.2898	0.4468	0.7848	<0.001	<0.001	0.8564	<0.001	0.0205	0.3237	0.0654
4	0.8388	0.0001	0.8818	0.0038	0.0290	0.1052	0.5185	0.0195	0.9807	0.0739	0.9854	0.9512	0.1521	0.0005	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
5	0.4510	0.0001	0.0030	0.0006	0.0285	0.1057	0.5144	0.3292	0.5625	0.9149	0.7486	0.4352	0.1053	0.4375	0.0051	<0.001	0.0013	<0.001	<0.001	<0.001	0.0410
6	0.5717	0.0021	0.6564	0.0014	0.0132		0.1921	0.2096	0.7014	0.0235	0.9017	0.6055		0.9605	0.0038	<0.001	<0.001	<0.001	<0.001		0.0019
7	0.7699	0.0022	0.8280	0.1023	0.0371	0.1033	0.0488	0.1496	0.9064	0.5426	0.9486	0.4538	0.0011	0.9745	0.0038	<0.001	0.6263	0.7673	<0.001	<0.001	0.2301
8	0.6594	0.0003	0.1407	0.0002	0.0044	0.1890	0.0724	0.1981	0.0770	0.1438	0.1867	0.2262	0.0231	0.7345	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.0359
9	0.6980	0.0012	0.1727	<0.0001	0.0900	0.0106	0.1108	0.3371	0.0209	<0.0001	0.2605	0.1352	0.0258	0.3021	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.0071
10	0.2241	0.0001	0.7121	0.0004	<0.0001	<0.0001	0.0012	0.2782	0.0021	0.0028	0.5742	0.0589	0.0188	0.1027	0.0359	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
11		0.0001	0.0425	0.0017	0.6511	0.7566	0.5920		0.0819	0.0449	0.6458	0.9495	0.0088	0.0777		0.0013	0.1096	0.0019	0.1737	<0.001	0.3622
12	0.4840	0.0001	0.8155	0.0419	0.9391	0.8723	0.5261	0.1372	0.4806	0.4507	0.7533	0.9715	0.0001	0.5056	0.0013	<0.001	0.5577	0.8654	0.0449	<0.001	0.9192
13	0.2990	<0.0001	0.7337	0.0012	0.3194	0.4382	0.6971	0.4318	0.0847	0.9526	0.9742	0.2530	0.0986	0.0094	<0.001	<0.001	0.3718	0.0583	<0.001	<0.001	0.0205
14	0.1854	0.0013	0.4129	0.0051	0.0388	0.8611	0.7868	0.5762	0.0188	0.9299	0.2778	0.5136	0.0135	0.1309	0.2667	<0.001	0.0276	<0.001	0.0987	<0.001	0.1635
15	0.5846	0.0001	0.7185	0.0022	0.0403	0.5710	0.6478	0.2743	0.0680	0.7669	0.5099	0.6586	0.2653	0.2499	<0.001	<0.001	0.0019	<0.001	<0.001	0.1840	0.7494
16	0.5326	0.0072	0.1422	0.0049	0.0813	0.0065	0.0150	0.0006	0.6436	0.7640	0.0026	0.4504	0.0167	0.0331	<0.001	0.1026	0.5821	<0.001	<0.001	<0.001	0.0115
17	0.1422	<0.0001	0.5986	0.0851	0.7729	0.9945	0.2683	0.2370	0.0928	0.0899	0.9955	0.2063	<0.0001	0.7517	<0.001	<0.001	0.2859	<0.001	0.0026	<0.001	0.9474
18	0.4069	0.0001	0.9827	0.0081	0.0426	0.6860	0.0539	0.0989	0.0072	0.0844	0.9628	0.0318	0.0840	0.1016	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.2622
19	0.4228	0.0001	0.0828	0.0021	0.0548	0.8409	0.2372	0.3372	<0.0001	0.0001	0.1160	0.7045	0.1513	0.2500	<0.001	<0.001	<0.001	0.0468	<0.001	0.0013	0.7263
20	0.5116	<0.0001	0.0952	0.0007	0.0006	0.0799	0.0204	0.1479	0.0074	0.0043	<0.0001	0.7523	0.3293	0.0674	<0.001	<0.001	<0.001	<0.001	<0.001	0.7885	0.1051

FIGURE A.4: Uncorrected p-values resulting from the 3 different analyses (GLM on HbO traces, GLM on HbR traces, LDA-based classification using combination of HbO and HbR features). Cells highlighted in red indicate the channels most likely covering the motor cortex. Green highlighting indicates channels that are classified as "active" at p=0.05. Empty (dark-gray highlighting) cells indicate channels with poor data quality, which were therefore discarded from the analysis.

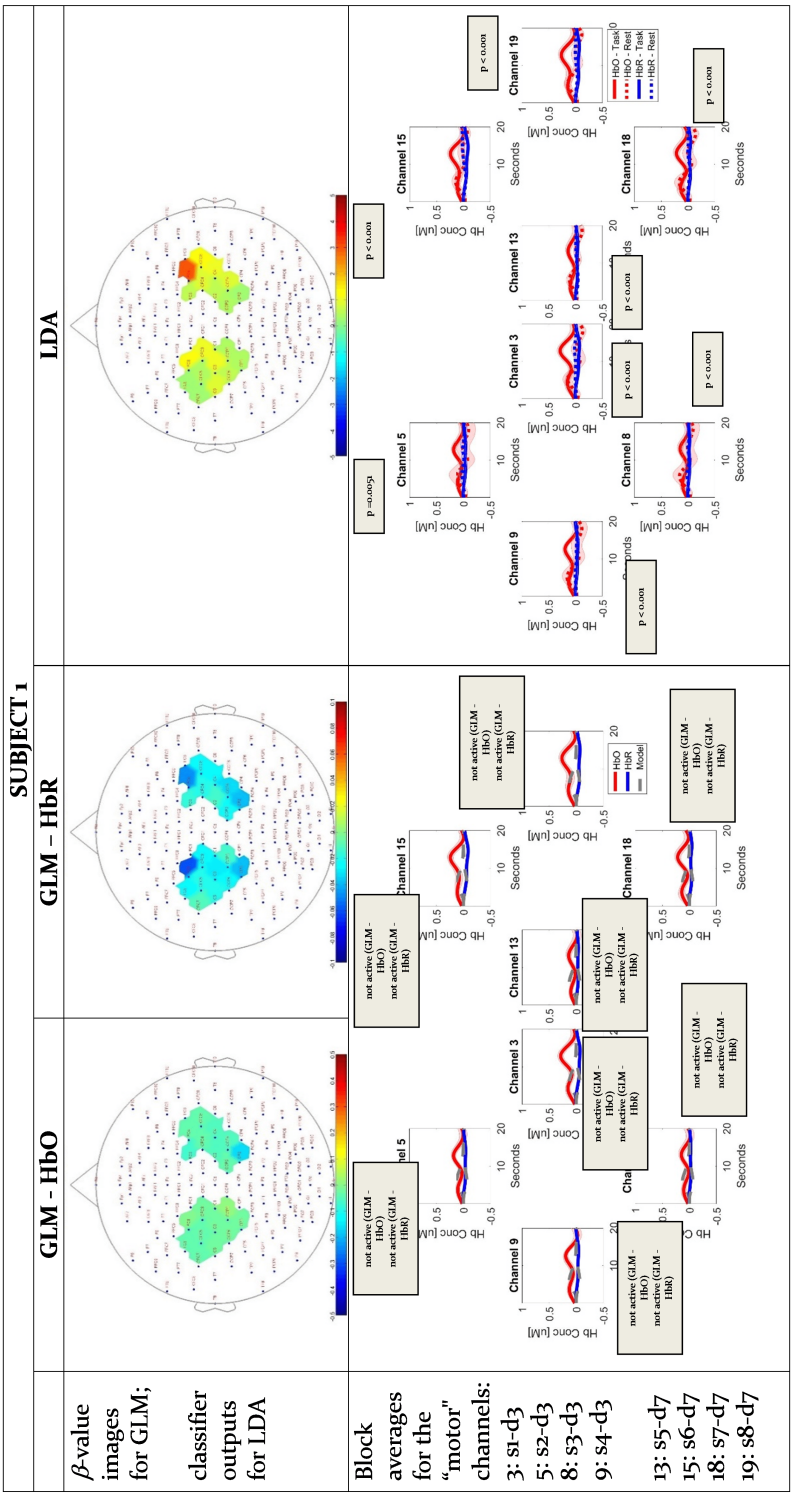


FIGURE A.5: Plots of GLM and LDA results- Subject 1

SUBJECT 1						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	-0.023	0.612	-0.027	0.028	1.013	<0.001
2	0.003	0.945	-0.013	0.461	0.467	0.023
3	-0.022	0.728	-0.029	0.172	0.973	<0.001
4	-0.013	0.839	-0.061	0.019	1.356	<0.001
5	-0.04	0.451	-0.021	0.329	0.383	0.005
6	-0.032	0.572	-0.026	0.21	0.218	0.004
7	-0.027	0.77	-0.04	0.15	0.361	0.004
8	-0.023	0.659	-0.026	0.198	0.507	<0.001
9	-0.021	0.698	-0.021	0.337	0.801	<0.001
10	-0.034	0.224	-0.012	0.278	0.214	0.036
11	NaN	NaN	NaN	NaN	NaN	NaN
12	-0.034	0.484	-0.023	0.137	0.436	0.001
13	-0.051	0.299	-0.015	0.432	0.53	<0.001
14	-0.063	0.185	-0.017	0.576	0.5	0.266
15	-0.031	0.585	-0.022	0.274	0.9	<0.001
16	-0.044	0.533	-0.049	0.001	3.032	<0.001
17	-0.156	0.142	-0.038	0.237	0.307	<0.001
18	-0.052	0.407	-0.031	0.099	0.855	<0.001
19	-0.058	0.423	-0.025	0.337	0.949	<0.001
20	-0.034	0.512	-0.026	0.148	1.37	<0.001

FIGURE A.6: Table of GLM and LDA results- Subject 1

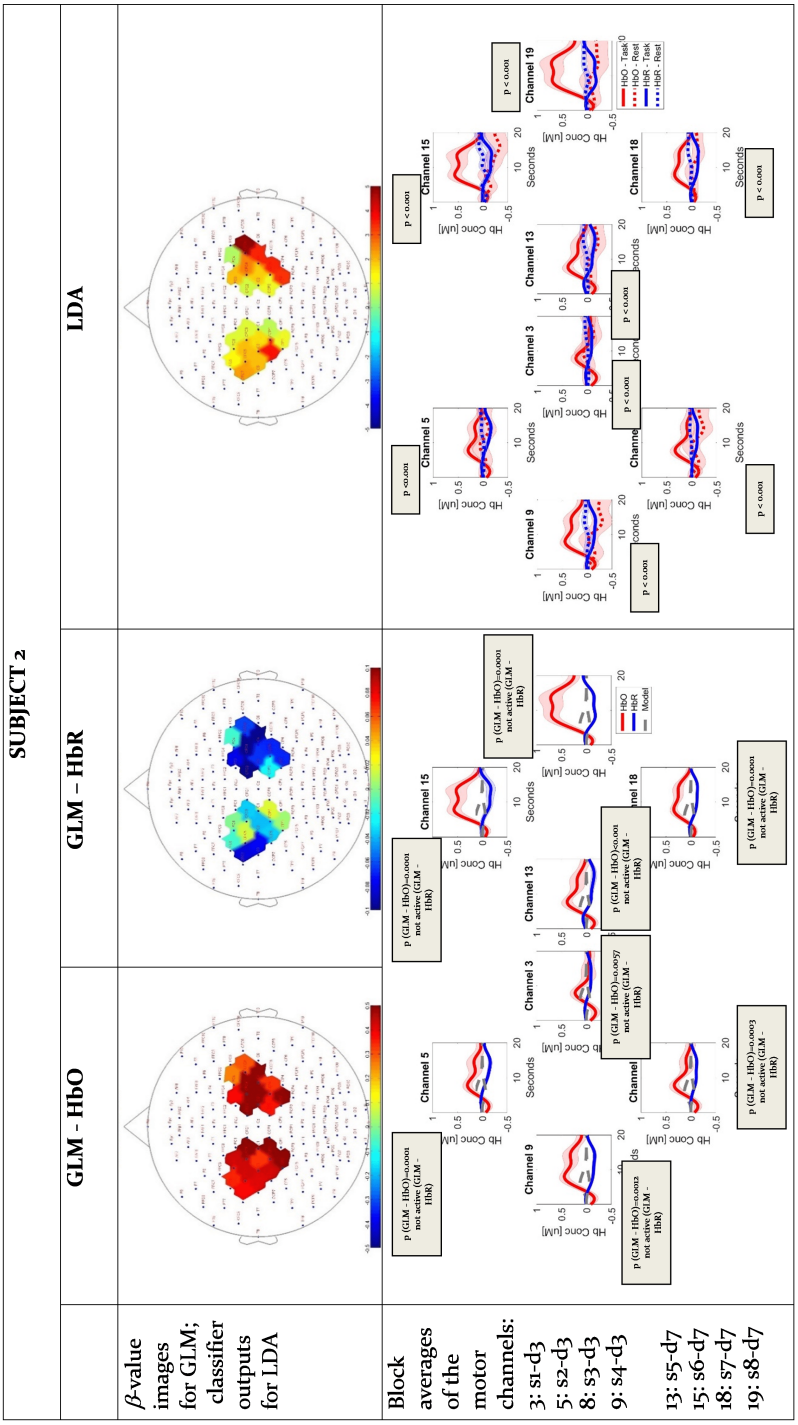


FIGURE A.7: Plots of GLM and LDA results- Subject 2

SUBJECT 2						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	0.531	<0.001	-0.036	0.44	0.928	<0.001
2	0.619	<0.001	0.017	0.503	0.611	0.001
3	0.332	0.006	-0.041	0.095	1.598	<0.001
4	0.431	<0.001	-0.001	0.981	0.93	<0.001
5	0.433	<0.001	0.015	0.563	1.892	<0.001
6	0.355	0.002	-0.014	0.701	1.384	<0.001
7	0.488	0.002	0.003	0.906	1.065	<0.001
8	0.424	<0.001	-0.037	0.077	3.623	<0.001
9	0.41	0.001	-0.076	0.021	1.885	<0.001
10	0.389	<0.001	-0.084	0.002	2.192	<0.001
11	0.492	<0.001	-0.15	0.082	0.678	0.001
12	0.373	<0.001	-0.022	0.481	0.969	<0.001
13	0.652	<0.001	-0.067	0.085	1.858	<0.001
14	0.375	0.001	-0.079	0.019	1.914	<0.001
15	0.576	<0.001	-0.128	0.068	2.537	<0.001
16	0.258	0.007	-0.013	0.644	0.42	0.103
17	0.494	<0.001	-0.07	0.093	3.42	<0.001
18	0.337	<0.001	-0.07	0.007	3.532	<0.001
19	0.529	<0.001	-0.179	0	3.508	<0.001
20	0.312	<0.001	-0.064	0.007	4.887	<0.001

FIGURE A.8: Table of GLM and LDA results- Subject 2

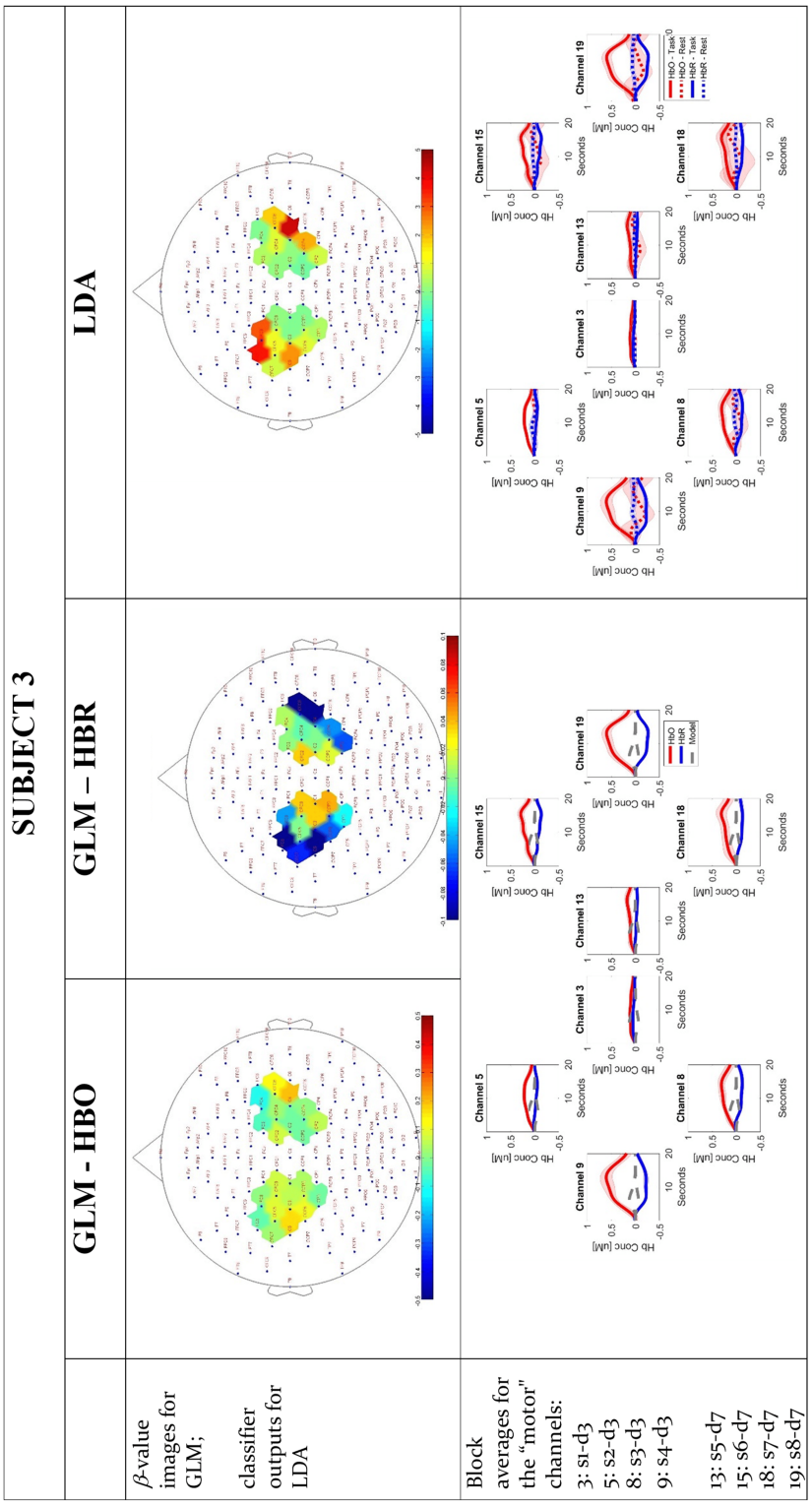


FIGURE A.9: Plots of GLM and LDA results- Subject 3

SUBJECT 3						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	0.073	0.017	0.037	0.062	0.029	0.157
2	0.045	0.238	0.043	0.179	-0.108	0.699
3	0.059	0.002	0.037	<0.001	-0.148	0.856
4	0.011	0.882	-0.047	0.074	3.02	<0.001
5	0.089	0.003	-0.002	0.915	1.087	0.001
6	-0.045	0.656	-0.099	0.024	3.666	<0.001
7	-0.026	0.828	-0.024	0.543	0.436	0.626
8	0.101	0.141	-0.057	0.144	0.84	<0.001
9	0.159	0.173	-0.166	<0.001	2.411	<0.001
10	0.023	0.712	-0.07	0.003	0.751	<0.001
11	0.067	0.043	0.035	0.045	0.027	0.11
12	-0.018	0.816	0.017	0.451	-0.184	0.558
13	-0.022	0.734	-0.001	0.953	0.129	0.372
14	-0.059	0.413	0.002	0.93	0.448	0.028
15	0.028	0.718	-0.009	0.767	0.87	0.002
16	-0.109	0.142	0.009	0.764	-0.025	0.582
17	0.056	0.599	-0.059	0.09	0.75	0.286
18	0.002	0.983	-0.047	0.084	2.073	<0.001
19	0.195	0.083	-0.155	<0.001	4.281	<0.001
20	0.117	0.095	-0.098	0.004	1.785	<0.001

FIGURE A.10: Table of GLM and LDA results- Subject 3

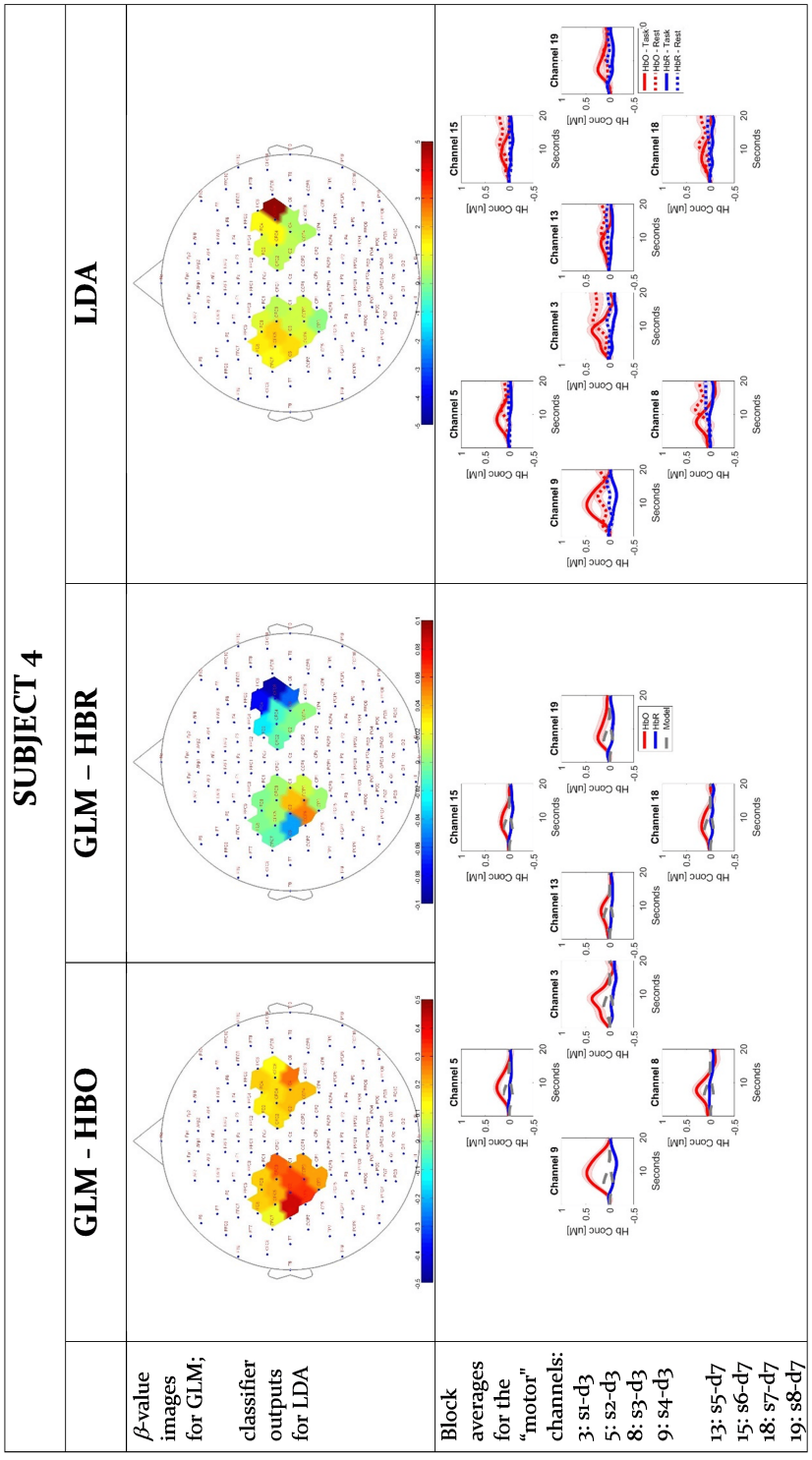


FIGURE A.11: Plots of GLM and LDA results- Subject 4

SUBJECT 4						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	0.282	<0.001	0.006	0.813	0.499	0.002
2	0.228	0.002	0.013	0.449	0.763	0.009
3	0.325	0.002	0.039	0.367	0.72	<0.001
4	0.18	0.004	0	0.985	0.868	<0.001
5	0.214	0.001	0.006	0.749	1.751	<0.001
6	0.18	0.001	-0.003	0.902	1.412	<0.001
7	0.23	0.102	0.005	0.949	0.136	0.767
8	0.336	<0.001	0.05	0.187	0.887	<0.001
9	0.434	<0.001	-0.044	0.26	1.675	<0.001
10	0.098	<0.001	-0.008	0.574	1.137	<0.001
11	0.161	0.002	-0.009	0.646	0.5	0.002
12	0.378	0.042	-0.03	0.753	-0.029	0.865
13	0.195	0.001	-0.001	0.974	0.605	0.058
14	0.149	0.005	-0.031	0.278	0.816	<0.001
15	0.162	0.002	-0.015	0.51	1.288	<0.001
16	0.176	0.005	-0.075	0.003	1.099	<0.001
17	0.476	0.085	0.001	0.995	-0.24	<0.001
18	0.193	0.008	0.002	0.963	0.393	<0.001
19	0.266	0.002	-0.057	0.116	0.316	0.047
20	0.137	0.001	-0.092	<0.001	8.254	<0.001

FIGURE A.12: Table of GLM and LDA results- Subject 4

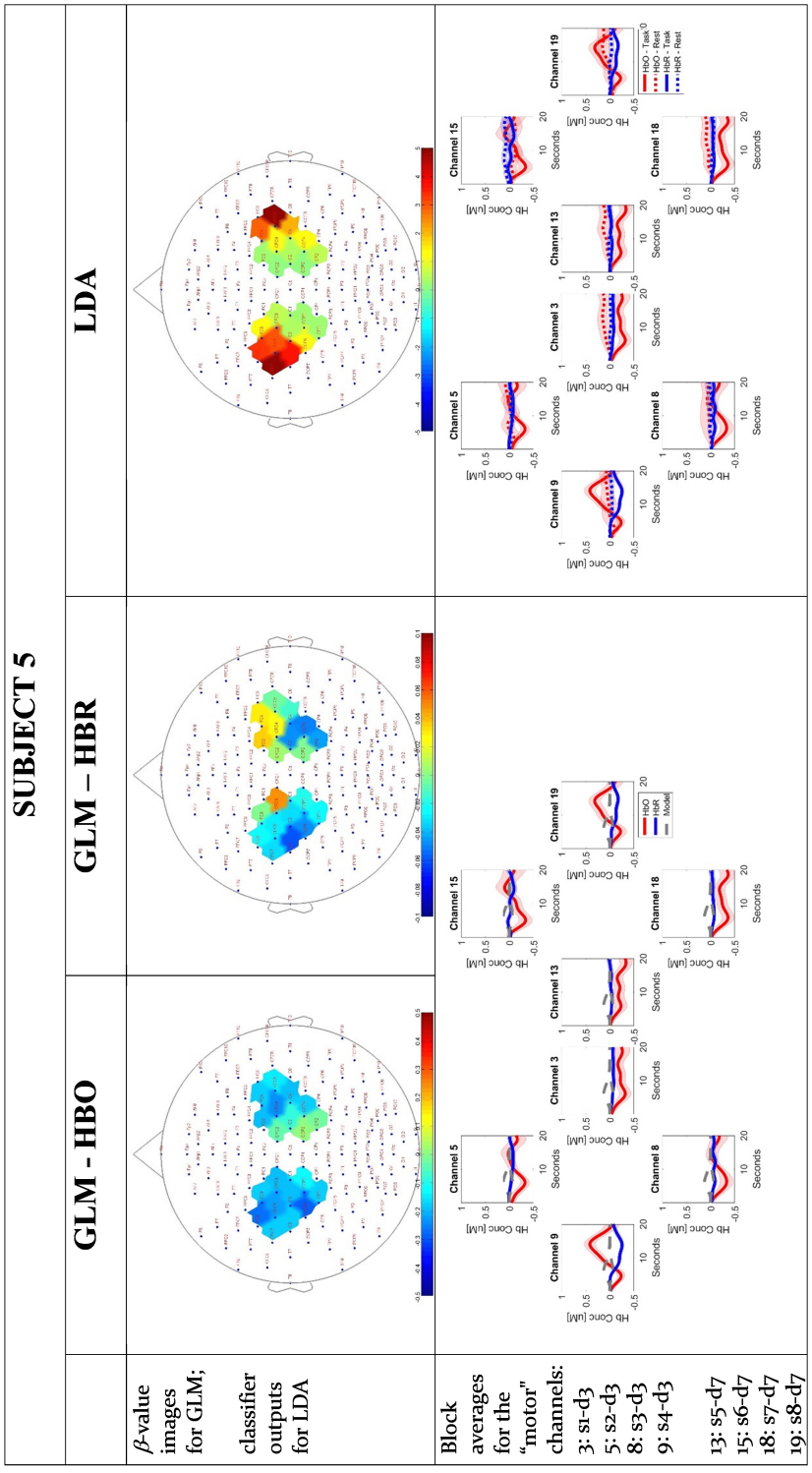


FIGURE A.13: Plots of GLM and LDA results- Subject 5

SUBJECT 5						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	-0.19	0.046	0.047	0.34	0.484	<0.001
2	-0.135	0.109	-0.021	0.553	0.233	0.059
3	-0.193	0.034	-0.031	0.29	0.448	0.021
4	-0.179	0.029	-0.002	0.951	1.298	<0.001
5	-0.226	0.028	-0.031	0.435	3.108	<0.001
6	-0.268	0.013	-0.023	0.606	3.311	<0.001
7	-0.174	0.037	-0.028	0.454	0.245	<0.001
8	-0.287	0.004	-0.049	0.226	1.155	<0.001
9	-0.183	0.09	-0.06	0.135	3.745	<0.001
10	-0.185	0	-0.033	0.059	5.692	<0.001
11	-0.041	0.651	0.005	0.95	0.018	0.174
12	0.004	0.939	-0.001	0.972	0.124	0.045
13	-0.082	0.319	-0.048	0.253	0.366	<0.001
14	-0.195	0.039	0.037	0.514	0.277	0.099
15	-0.247	0.04	0.023	0.659	1.21	<0.001
16	-0.182	0.081	0.029	0.45	2.885	<0.001
17	-0.024	0.773	-0.046	0.206	0.533	0.003
18	-0.177	0.043	-0.052	0.032	1.402	<0.001
19	-0.199	0.055	-0.015	0.704	2.035	<0.001
20	-0.149	0.001	-0.006	0.752	6.003	<0.001

FIGURE A.14: Table of GLM and LDA results- Subject 5

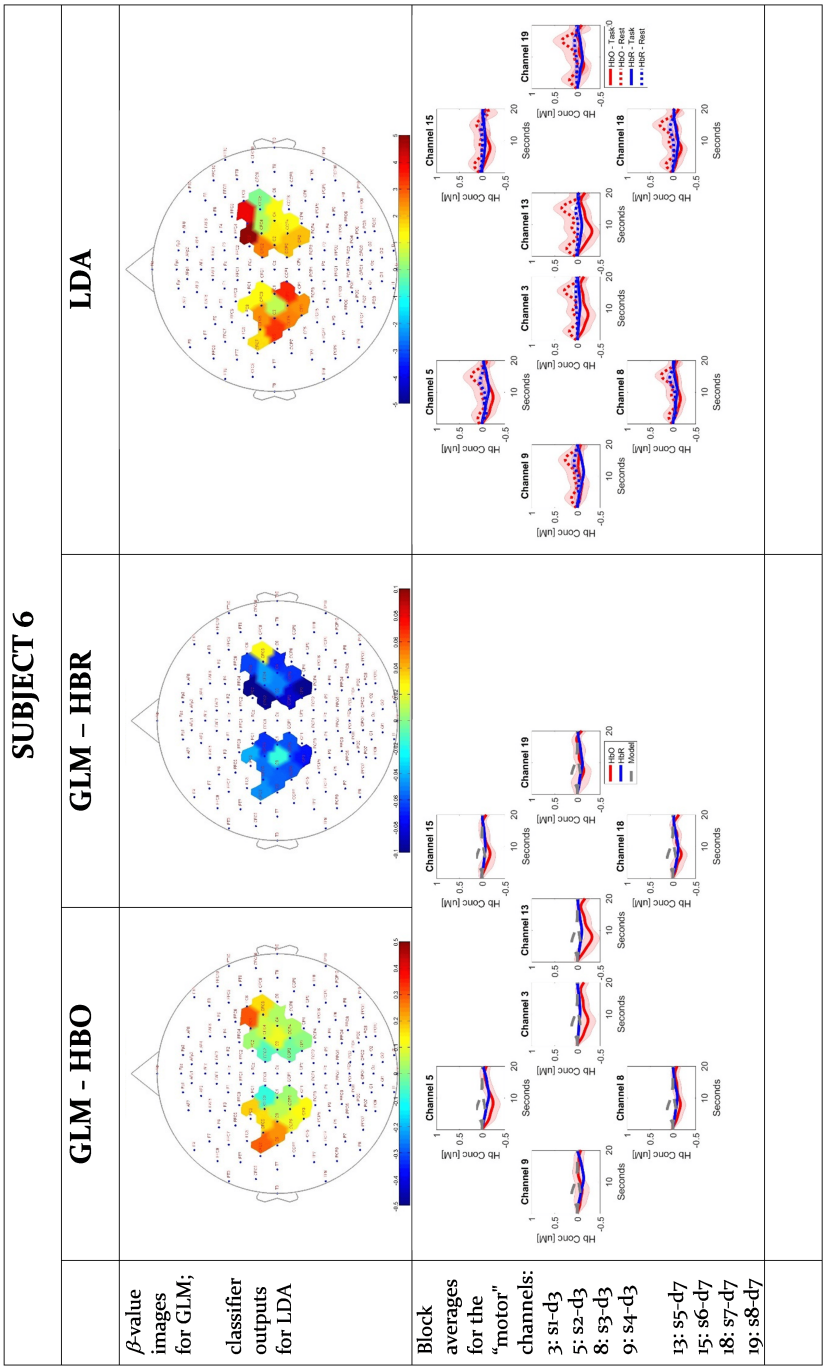


FIGURE A.15: Plots of GLM and LDA results- Subject 6

SUBJECT 6						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	-0.086	0.508	-0.066	0.047	1.307	<0.001
2	0.018	0.871	-0.062	0.011	3.578	<0.001
3	0.06	0.568	-0.018	0.447	0.528	0.324
4	0.174	0.105	-0.042	0.152	1.203	<0.001
5	0.163	0.106	-0.053	0.105	2.47	<0.001
6	NaN	NaN	NaN	NaN	NaN	NaN
7	0.138	0.103	-0.074	0.001	2.41	<0.001
8	0.1	0.189	-0.059	0.023	2.655	<0.001
9	0.242	0.011	-0.059	0.026	3.403	<0.001
10	0.283	<0.001	-0.047	0.019	1.675	<0.001
11	-0.051	0.757	-0.121	0.009	2.517	<0.001
12	-0.019	0.872	-0.085	<0.001	1.923	<0.001
13	0.104	0.438	-0.051	0.099	1.554	<0.001
14	0.023	0.861	-0.101	0.013	6.105	<0.001
15	0.069	0.571	-0.045	0.265	0.389	0.184
16	0.298	0.007	-0.066	0.017	3.987	<0.001
17	0.001	0.995	-0.113	<0.001	1.957	<0.001
18	0.041	0.686	-0.068	0.084	1.209	<0.001
19	0.031	0.841	-0.069	0.151	1.183	0.001
20	0.16	0.08	0.022	0.329	-0.012	0.788

FIGURE A.16: Table of GLM and LDA results- Subject 6

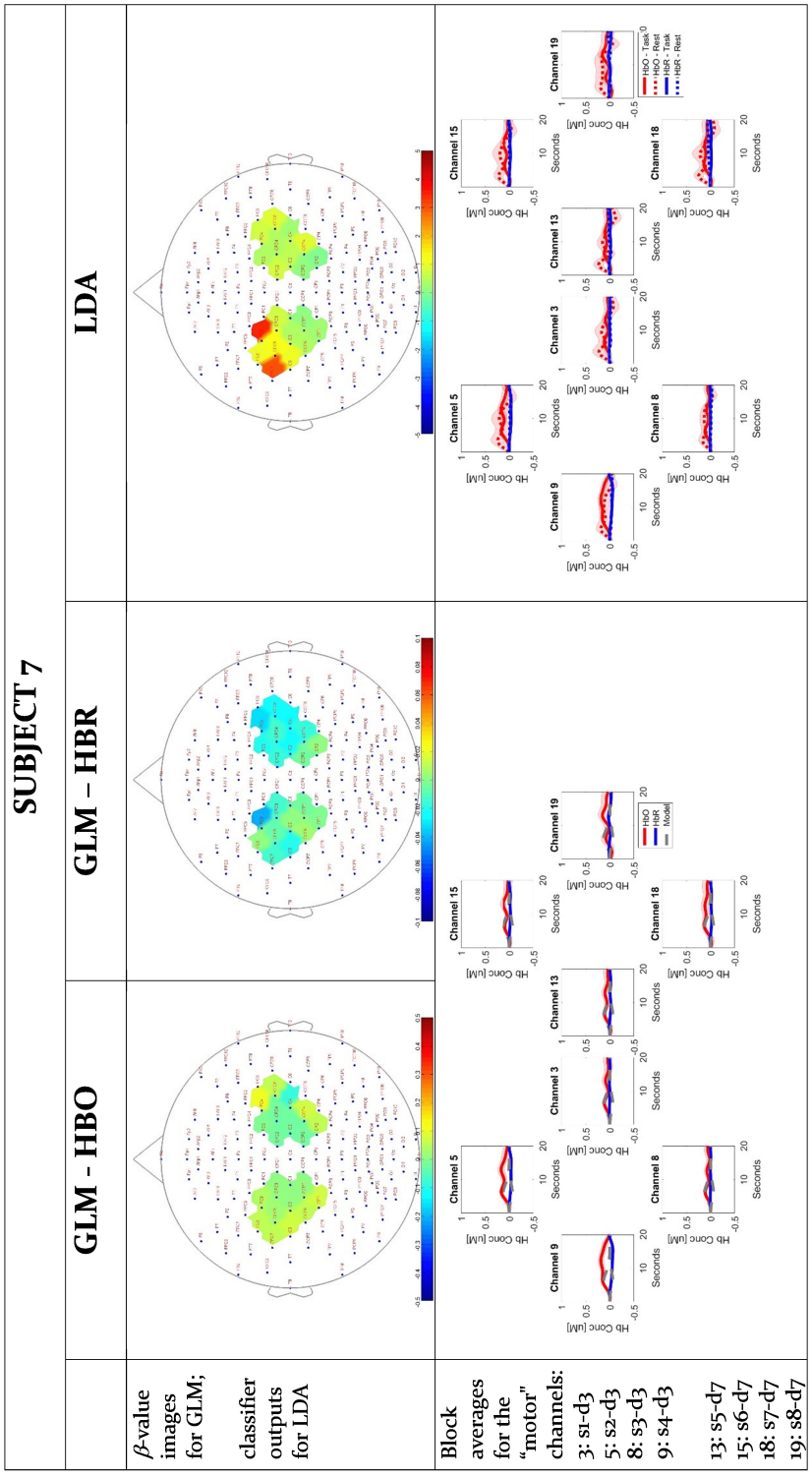


FIGURE A.17: Plots of GLM and LDA results- Subject 7

SUBJECT 7						
Channel	**Beta – GLM HbO**	**p – GLM HbO**	**Beta – GLM HbR**	**p – GLM HbR**	**Classif. Output**	**p – Classif.**
1	0.003	0.931	-0.014	0.168	0.743	0.377
2	0.003	0.937	-0.005	0.547	0.088	0.75
3	0	0.995	0.002	0.785	0.467	0.065
4	0.039	0.519	-0.041	<0.001	3.531	<0.001
5	0.026	0.514	-0.009	0.438	1.215	0.041
6	0.059	0.192	0.001	0.96	0.977	0.002
7	0.071	0.049	0	0.975	0.311	0.23
8	0.054	0.072	0.003	0.735	0.44	0.036
9	0.064	0.111	-0.021	0.302	1.067	0.007
10	0.062	0.001	-0.014	0.103	2.97	<0.001
11	-0.023	0.592	-0.016	0.078	0.662	0.362
12	-0.024	0.526	-0.004	0.506	0.076	0.919
13	-0.016	0.697	-0.024	0.009	0.526	0.021
14	-0.012	0.787	-0.02	0.131	0.384	0.163
15	-0.018	0.648	-0.013	0.25	0.376	0.749
16	0.094	0.015	-0.034	0.033	0.665	0.012
17	0.05	0.268	0.003	0.752	-0.05	0.947
18	0.075	0.054	-0.018	0.102	0.825	0.262
19	-0.079	0.237	-0.027	0.25	0.26	0.726
20	0.056	0.02	-0.024	0.067	0.834	0.105

FIGURE A.18: Table of GLM and LDA results- Subject 7

	GLM - HbO							GLM - HbR							LDA						
SUB / Ch	001	002	003	004	005	006	007	001	002	003	004	005	006	007	001	002	003	004	005	006	007
1	0.6123	0.0001	0.0173	0.0001	0.0465	0.5084	0.9314	0.0279	0.4493	0.0621	0.8131	0.3398	0.0470	0.1678	<0.0005	<0.001	0.1571	0.0019	<0.001	<0.001	0.3769
2	0.9445	<0.0001	0.2381	0.0017	0.1091	0.8705	0.9373	0.4607	0.5035	0.1794	0.4487	0.5527	0.0105	0.5467	0.0231	0.0013	0.6987	0.0090	0.0590	<0.001	0.7500
3	0.7284	0.0057	0.0015	0.0021	0.0338	0.5684	0.9946	0.1718	0.0953	<0.0001	0.3668	0.2898	0.4468	0.7848	<0.001	<0.001	0.8564	<0.001	0.0205	0.3237	0.0654
4	0.8388	0.0001	0.8818	0.0038	0.0290	0.1052	0.5185	0.0195	0.9807	0.0739	0.9854	0.9512	0.1521	0.0005	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
5	0.4510	0.0001	0.0030	0.0006	0.0285	0.1057	0.5144	0.3292	0.5625	0.9149	0.7486	0.4352	0.1053	0.4375	0.0051	<0.001	0.0013	<0.001	<0.001	<0.001	0.0410
6	0.5777	0.0021	0.6564	0.0014	0.0132		0.1921	0.2096	0.7014	0.0235	0.5017	0.6055		0.9605	0.0038	<0.001	<0.001	<0.001	<0.001		0.0019
7	0.7699	0.0022	0.8280	0.1023	0.0371	0.1033	0.0488	0.1496	0.9064	0.5426	0.9486	0.4538	0.0011	0.9745	0.0038	<0.001	0.6263	0.7673	<0.001	<0.001	0.2301
8	0.6594	0.0003	0.1407	0.0002	0.0044	0.1890	0.0724	0.1981	0.0770	0.1438	0.1867	0.2262	0.0231	0.7345	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.0359
9	0.6980	0.0012	0.1727	<0.0001	0.0900	0.0106	0.1108	0.3371	0.0209	<0.0001	0.2605	0.1352	0.0258	0.3021	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.0071
10	0.2241	0.0001	0.7121	0.0004	<0.0001	<0.0001	0.0012	0.2782	0.0021	0.0028	0.5742	0.0589	0.0188	0.1027	0.0359	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
11		0.0001	0.0425	0.0017	0.6531	0.7566	0.5920		0.0819	0.0449	0.6458	0.9495	0.0088	0.0777		0.0013	0.1096	0.0019	0.1737	<0.001	0.3622
12	0.4840	0.0001	0.8855	0.0419	0.9391	0.8723	0.5261	0.1372	0.4806	0.4597	0.7533	0.9715	0.0001	0.5956	0.0013	<0.001	0.5577	0.8654	0.0449	<0.001	0.9992
13	0.2990	<0.0001	0.7337	0.0012	0.3194	0.4382	0.6971	0.4318	0.0847	0.9526	0.9742	0.2530	0.0986	0.0094	<0.001	<0.001	0.3718	0.0583	<0.001	<0.001	0.0205
14	0.1854	0.0013	0.4129	0.0051	0.0388	0.8611	0.7868	0.5762	0.0188	0.9299	0.2778	0.5136	0.0135	0.1309	0.2667	<0.001	0.0276	<0.001	0.0987	<0.001	0.1635
15	0.5846	0.0001	0.7185	0.0022	0.0403	0.5710	0.6478	0.2743	0.0680	0.7669	0.5099	0.6586	0.2653	0.2499	<0.001	<0.001	0.0019	<0.001	<0.001	0.1840	0.7494
16	0.5326	0.0072	0.1422	0.0049	0.0813	0.0065	0.0150	0.0006	0.6436	0.7640	0.0026	0.4594	0.0167	0.0331	<0.001	0.1026	0.5821	<0.001	<0.001	<0.001	0.015
17	0.1422	<0.0001	0.5986	0.0851	0.7729	0.9945	0.2683	0.2370	0.0928	0.0899	0.9955	0.2063	<0.0001	0.7517	<0.001	<0.001	0.2859	<0.001	0.0026	<0.001	0.9474
18	0.4069	0.0001	0.9827	0.0081	0.0426	0.6860	0.0539	0.0989	0.0072	0.0844	0.9608	0.0318	0.0840	0.1016	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.2622
19	0.4228	0.0001	0.0828	0.0021	0.0548	0.8499	0.2372	0.3372	<0.0001	0.0001	0.1160	0.7045	0.1513	0.2500	<0.001	<0.001	<0.001	0.0468	<0.001	0.0013	0.7263
20	0.5116	<0.0001	0.0952	0.0007	0.0006	0.0799	0.0204	0.1479	0.0074	0.0043	<0.0001	0.7523	0.3293	0.0674	<0.001	<0.001	<0.001	<0.001	<0.001	0.7885	0.1051

FIGURE A.19: Table of all the experimental results. Uncorrected p-values resulting from the 3 different analyses (GLM on HbO traces, GLM on HbR traces, LDA-based classification using combination of HbO and HbR features). Highlighted cells (red) indicate the channels most likely covering the motor cortex. Highlighted cells (green) indicate $p<0.05$. Highlighted cells (grey) with missing values indicate channels with poor acquisition quality and therefore discarded from the analysis.

A.2 Supplementary information of Chapter 4

A.2.1 Localization of optodes

Ch#	MNI coordinates			Brodmann Area	Pr(%)
	X	Y	Z		
1	-65.6	-52.33	-14.83	20 - Inferior Temporal gyrus	28.2
				21 - Middle Temporal gyrus	7
				37 - Fusiform gyrus	64.8
2	-70	-28.33	-16.5	20 - Inferior Temporal gyrus	2.5
				21 - Middle Temporal gyrus	97.5
3	-65.5	-8.33	-23.5	20 - Inferior Temporal gyrus	41.3
				21 - Middle Temporal gyrus	58.7
4	-35.6	62.83	-7	10 - Frontopolar area	79.1
				11 - Orbitofrontal area	20.9
5	-12.3	72	-1.5	10 - Frontopolar area	98.8
				11 - Orbitofrontal area	1.2
6	17	72	-0.5	10 - Frontopolar area	100
7	39.3	63.33	-6.5	10 - Frontopolar area	87.3
				11 - Orbitofrontal area	12.7
8	68.17	-6.33	-21.33	20 - Inferior Temporal gyrus	18.4
				21 - Middle Temporal gyrus	81.6
9	72	-26.5	-14.83	20 - Inferior Temporal gyrus	1.3
				21 - Middle Temporal gyrus	98.7

10	67	-48.67	-15.17	20 - Inferior Temporal gyrus	38.4
				21 - Middle Temporal gyrus	6.8
				37 - Fusiform gyrus	54.8
11	-62.3	-61	-5.5	21 - Middle Temporal gyrus	15.3
				37 - Fusiform gyrus	84.7
12	-70	-39.33	-2	21 - Middle Temporal gyrus	88.8
				22 - Superior Temporal Gyrus	11.3
13	-70	-17	-6.33	21 - Middle Temporal gyrus	97.5
				22 - Superior Temporal Gyrus	2.5
14	-62	2.3	-19	21 - Middle Temporal gyrus	100
15	-49.3	45.83	-5.33	10 - Frontopolar area	1.5
				47 - Inferior prefrontal gyrus	98.5
16	-25.3	67	9.34	10 - Frontopolar area	100
17	5	70	14.67	10 - Frontopolar area	100
18	28.83	67.67	10.17	10 - Frontopolar area	100
19	52	47.33	-2.83	10 - Frontopolar area	21.7
				47 - Inferior prefrontal gyrus	78.3
20	64	4.33	-14.83	21 - Middle Temporal gyrus	100
21	72	-14.5	-3.33	21 - Middle Temporal gyrus	83.5
				22 - Superior Temporal Gyrus	16.5

22	72	-37.67	-0.83	21 - Middle Temporal gyrus	64.2
				22 - Superior Temporal Gyrus	35.8
23	64	-58.5	-5.67	21 - Middle Temporal gyrus	6.9
				37 - Fusiform gyrus	93.1
24	-67.17	-50.33	9.67	21 - Middle Temporal gyrus	34.2
				22 - Superior Temporal Gyrus	65.8
25	-69	-26.33	13.5	40 - Supramarginal gyrus part of Wernicke's area	1.3
				42 - Primary and Auditory Association Cortex	98.7
26	-66.5	-5.33	7.83	22 - Superior Temporal Gyrus	93.2
				42 - Primary and Auditory Association Cortex	6.8
27	-56.67	25.17	3.33	45 - pars triangularis Broca's area	56.8
				47 - Inferior prefrontal gyrus	43.2
28	-12	66.67	24.83	10 - Frontopolar area	100
29	15.67	67	25.83	10 - Frontopolar area	100
30	60	27.67	7.67	45 - pars triangularis Broca's area	96.3
				46 - Dorsolateral prefrontal cortex	3.7
31	68	-3.33	9	6 - Pre-Motor and Supplementary Motor Cortex	3.9
				22 - Superior Temporal Gyrus	96.1
32	71.17	-25.33	13.17	40 - Supramarginal gyrus part of Wernicke's area	2.5
				42 - Primary and Auditory Association Cortex	97.5
33	69	-48.17	10.83	21 - Middle Temporal gyrus	11.7
				22 - Superior Temporal Gyrus	88.3

34	-63	-59.33	21.33	22 - Superior Temporal Gyrus	45.2
				39 - Angular gyrus, part of Wernicke's area	50.7
				40 - Supramarginal gyrus part of Wernicke's area	4.1
35	-68	-36.5	30.67	40 - Supramarginal gyrus part of Wernicke's area	100
36	-67	-12.33	27.33	1 - Primary Somatosensory Cortex	32.5
				3 - Primary Somatosensory Cortex	18.8
				4 - Primary Motor Cortex	20
				6 - Pre-Motor and Supplementary Motor Cortex	7.5
				43 - Subcentral area	21.3
37	-61	11.33	14.17	44 - pars opercularis, part of Broca's area	100
38	-51	41	14.33	45 - pars triangularis Broca's area	1.4
				46 - Dorsolateral prefrontal cortex	98.6
39	-24.7	54.83	34.33	9 - Dorsolateral prefrontal cortex	90.6
				10 - Frontopolar area	9.4
40	1.67	57.17	39.67	9 - Dorsolateral prefrontal cortex	100
41	27.33	55	34.67	9 - Dorsolateral prefrontal cortex	98.2
				10 - Frontopolar area	1.8
42	53	40.33	18.67	46 - Dorsolateral prefrontal cortex	100
43	63	13.83	17	44 - pars opercularis, part of Broca's area	90.7
				45 - pars triangularis Broca's area	9.3
44	69	-9.33	28.5	3 - Primary Somatosensory Cortex	34.6

				4 - Primary Motor Cortex	10.3
				6 - Pre-Motor and Supplementary Motor Cortex	47.4
				43 - Subcentral area	7.7
45	70	-34.67	29.5	40 - Supramarginal gyrus part of Wernicke's area	100
46	64	-58.33	21	22 - Superior Temporal Gyrus	46.1
				39 - Angular gyrus, part of Wernicke's area	44.7
				40 - Supramarginal gyrus part of Wernicke's area	9.2
47	-64	-46.5	39.67	40 - Supramarginal gyrus part of Wernicke's area	100
48	-65	-22.5	39.33	1 - Primary Somatosensory Cortex	32.5
				2 - Primary Somatosensory Cortex	36.4
				3 - Primary Somatosensory Cortex	29.9
				40 - Supramarginal gyrus part of Wernicke's area	1.3
49	-63	1.83	31.33	6 - Pre-Motor and Supplementary Motor Cortex	97.3
				9 - Dorsolateral prefrontal cortex	2.7
50	-53	30.33	26.67	45 - pars triangularis Broca's area	4.3
				46 - Dorsolateral prefrontal cortex	95.7
51	-37.67	44.17	36.33	9 - Dorsolateral prefrontal cortex	96.2
				46 - Dorsolateral prefrontal cortex	3.8
52	-12.33	47.67	49.67	8 - Includes Frontal eye fields	100
53	14.33	47.33	50.5	8 - Includes Frontal eye fields	100
54	38.67	44.67	38	9 - Dorsolateral prefrontal cortex	100
55	55	29.67	28.5	45 - pars triangularis Broca's area	1.4

				46 - Dorsolateral prefrontal cortex	98.6
56	65	3.33	33.33	6 - Pre-Motor and Supplementary Motor Cortex	87.7
				9 - Dorsolateral prefrontal cortex	12.3
57	67.5	-22.5	40.67	1 - Primary Somatosensory Cortex	50.6
				2 - Primary Somatosensory Cortex	18.2
				3 - Primary Somatosensory Cortex	26
				4 - Primary Motor Cortex	5.2
58	65.17	-47	38.67	40 - Supramarginal gyrus part of Wernicke's area	100

TABLE A.1: The X, Y, and Z columns represent median MNI coordinates across subjects. Anatomical labelling of the MNI coordinates was conducted using the Talairach atlas. The last column lists the atlas-based probability that the channel coordinates are within that anatomical location. The Brodmann area with the highest probability is highlighted in bold.

Bibliography

[1] F. Jobsis. "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters". In: *Science* 198.4323 (1977), pp. 1264–1267. ISSN: 0036-8075. DOI: 10.1126/science.929199. arXiv: arXiv:1011.1669v3.

[2] Lia M. Hocke, Ibukunoluwa K. Oni, Chris C. Duszynski, Alex V. Corrigan, Blaise de B. Frederick, and Jeff F. Dunn. "Automated processing of fNIRS data-A visual guide to the pitfalls and consequences". In: *Algorithms* 11.5 (2018), pp. 1–25. ISSN: 19994893. DOI: 10.3390/a11050067.

[3] Paola Pinti, Ilias Tachtsidis, Antonia Hamilton, Joy Hirsch, Clarisse Aichelburg, Sam Gilbert, and Paul W. Burgess. "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience". In: *Annals of the New York Academy of Sciences* (2018), pp. 1–25. ISSN: 17496632. DOI: 10.1111/nyas.13948.

[4] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E Nichols. *Statistical parametric mapping: the analysis of functional brain images.* Academic press, 2011. ISBN: 9780124115118.

[5] Theodore J. Huppert. "Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy". In: *Neurophotonics* 3.1 (2016), p. 010401. ISSN: 2329-423X. DOI: 10.1117/1.NPh.3.1.010401. URL: http://neurophotonics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.NPh.3.1.010401.

[6] Jeffrey W Barker, Ardalan Aarabi, Theodore J Huppert, M Suzuki, I Miyai, T Ono, K Kubota, R J Cooper, J Selb, L Gagnon, D Phillip, H W Schytz, H K Iversen, M Ashina, J C Ye, S Tak, K E Jang, J Jung, and J Jang. "Autoregressive model based algorithm for correcting motion and serially correlated errors in

fNIRS". In: *Inst. Stat. Math. 22. A. M. Dale Hum. Brain Mapp* 4.8 (2013), pp. 35–54. ISSN: 2156-7085. DOI: 10.1364/BOE.4.001366.

[7] Lauren L. Emberson, Benjamin D. Zinszer, Rajeev D.S. Raizada, and Richard N. Aslin. "Decoding the infant mind: Multivariate pattern analysis (MVPA) usingfNIRS". In: *PLoS ONE* 12.4 (2017), pp. 1–23. ISSN: 19326203. DOI: 10.1371/journal.pone.0172500. URL: http://dx.doi.org/10.1371/journal.pone.0172500.

[8] Benjamin D. Zinszer, Laurie Bayet, Lauren L. Emberson, Rajeev D. S. Raizada, and Richard N. Aslin. "Decoding semantic representations from functional near-infrared spectroscopy signals". In: *Neurophotonics* 5.01 (2017), p. 1. ISSN: 2329-423X. DOI: 10.1117/1.NPh.5.1.011003. URL: https://www.spiedigitallibrary.org/journals/neurophotonics/volume-5/issue-01/011003/Decoding-semantic-representations-from-functional-near-infrared-spectroscopy-signals/10.1117/1.NPh.5.1.011003.full.

[9] Jessica Gemignani, Eike Middell, Randall L Barbour, Harry L Graber, and Benjamin Blankertz. "Improving the analysis of near-infrared spectroscopy data with multivariate classification of hemodynamic patterns : a theoretical formulation and validation". In: *Journal of Neural Engineering* 15.4 (2018), 045001 (15pp). ISSN: 1741-2560. DOI: 10.1088/1741-2552/aabb7c. URL: https://doi.org/10.1088/1741-2552/aabb7c.

[10] Jessica Gemignani, Benjamin Blankertz, Richard N. Aslin, and K. R. Pugh. "Observing the effects of literacy acquisition on the brain with fNIRS". In: *NeuroImage* in preparation (2019).

[11] Jessica Gemignani, Laurie Bayet, Claire Kabdebon, Benjamin Blankertz, Kenneth R. Pugh, and Richard N. Aslin. "Classifying the mental representation of word meaning in children with Multivariate Pattern Analysis of fNIRS". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. IEEE, 2018.

[12] Eric C Peterson, Zhengfeng Wang, and Gavin Britz. *Regulation of cerebral blood flow*. 2011. DOI: 10.1155/2011/823525.

[13] C.S. Roy and C.S Sherrington. "On the Regulation of the Blood-supply of the Brain". In: *Journal of Physiology* 11 (1890), pp. 85–158.

[14] Mireille Bélanger, Igor Allaman, and Pierre J. Magistretti. "Brain energy metabolism: Focus on Astrocyte-neuron metabolic cooperation". In: *Cell Metabolism* 14.6 (2011), pp. 724–738. ISSN: 15504131. DOI: 10.1016/j.cmet.2011.08.016. arXiv: arXiv:1011.1669v3.

[15] Helene Girouard and Costantino Iadecola. "Regulation of the Cerebral Circulation Neurovascular coupling in the normal brain and in hypertension , stroke , and Alzheimer disease". In: 10021 (2010), pp. 328–335. DOI: 10.1152/japplphysiol.00966.2005.

[16] R G Shulman and D L Rothman. "Brain Energetics and Neuronal Activity: Applications to fMRI and Medicine". In: *Medicine* (2004).

[17] Hellmuth Obrig and Arno Villringer. "Beyond the visible - Imaging the human brain with light". In: *Journal of Cerebral Blood Flow and Metabolism* 23.1 (2003), pp. 1–18. ISSN: 0271678X. DOI: 10.1097/01.WCB.0000043472.45775.29.

[18] GK Aguirre, E Zarahn, and M. D'Esposito. "The Variability of Human, BOLD Hemodynamic Responses • 1". In: *Neuroimage* 8.4 (1998), pp. 360–369. URL: http://www.sciencedirect.com/science/article/pii/S105381199890369X.

[19] D. T. Delpy and M. Cope. "Quantification in tissue near-infrared spectroscopy". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 352.1354 (1997), pp. 649–659. ISSN: 0962-8436. DOI: 10.1098/rstb.1997.0046.

[20] Felix Scholkmann, Stefan Kleiser, Andreas Jaakko Metz, Raphael Zimmermann, Juan Mata Pavia, Ursula Wolf, and Martin Wolf. "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology". In: *NeuroImage* 85 (2014), pp. 6–27. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.05.004. URL: http://dx.doi.org/10.1016/j.neuroimage.2013.05.004.

[21] Eiji Okada, Michael Firbank, Martin Schweiger, Simon R Arridge, Mark Cope, and David T Delpy. "Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head". In: *Applied optics* 36.1 (1997), pp. 21–31.

[22] Eiji Okada, Michael Firbank, and David T Delpy. "The effect of overlying tissue on the spatial sensitivity profile of near-infrared spectroscopy". In: *Physics in Medicine & Biology* 40.12 (1995), p. 2093.

[23] Eiji Okada and David T Delpy. "Near-infrared light propagation in an adult head model. II. Effect of superficial tissue thickness on the sensitivity of the near-infrared spectroscopy signal". In: *Applied optics* 42.16 (2003), pp. 2915–2921.

[24] Marco Ferrari and Valentina Quaresima. "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application". In: *NeuroImage* 63.2 (2012), pp. 921–935. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.03.049.

[25] Matthew Caldwell, Felix Scholkmann, Ursula Wolf, Martin Wolf, Clare Elwell, and Ilias Tachtsidis. "Modelling confounding effects from extracerebral contamination and systemic factors on functional near-infrared spectroscopy". In: *NeuroImage* 143 (2016), pp. 91–105. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2016.08.058. URL: http://dx.doi.org/10.1016/j.neuroimage.2016.08.058.

[26] Ilias Tachtsidis and Felix Scholkmann. "False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward". In: *Neurophotonics* 3.3 (2016), p. 031405.

[27] Martin Wolf, Ursula Wolf, Vlad Toronov, Antonios Michalos, L. Adelina Paunescu, Jee Hyun Choi, and Enrico Gratton. "Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: A near-infrared spectroscopy study". In: *NeuroImage* 16.3 I (2002), pp. 704–712. ISSN: 10538119. DOI: 10.1006/nimg.2002.1128.

[28] Felix Scholkmann and Martin Wolf. "General equation for the differential pathlength factor of the frontal human head depending on wavelength and age". In: *Journal of Biomedical Optics* (2013). ISSN: 1083-3668. DOI: 10.1117/1.JBO.18.10.105004.

[29] M. Essenpreis, C. E. Elwell, M. Cope, P. van der Zee, S. R. Arridge, and D. T. Delpy. "Spectral dependence of temporal point spread functions in human tissues". In: *Applied Optics* 32.4 (1993), p. 418. ISSN: 0003-6935. DOI: 10.1364/AO.32.000418. URL: https://www.osapublishing.org/abstract.cfm?URI=ao-32-4-418.

[30] Judith D. Schaechter. *Motor rehabilitation and brain plasticity after hemiparetic stroke*. 2004. DOI: 10.1016/j.pneurobio.2004.04.001.

[31] Felix Biessmann, Frank C. Meinecke, Arthur Gretton, Alexander Rauch, Gregor Rainer, Nikos K. Logothetis, and Klaus Robert Müller. "Temporal kernel CCA and its application in multimodal neuronal data analysis". In: *Machine Learning* 79.1-2 (2010), pp. 5–27. ISSN: 15730565. DOI: 10.1007/s10994-009-5153-3.

[32] Felix Biessmann, Sergey Plis, Frank C Meinecke, Tom Eichele, and Klaus-Robert Müller. "Analysis of Multimodal Neuroimaging Data". In: *IEEE Reviews in Biomedical Engineering* 4 (2011), pp. 26–58.

[33] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. "Brain-computer interfaces for communication and control." In: *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* 113.6 (2002), pp. 767–91. ISSN: 1388-2457. DOI: 10.1016/S1388-2457(02)00057-3. URL: http://www.ncbi.nlm.nih.gov/pubmed/12048038.

[34] Siamac Fazli, Jan Mehnert, Jens Steinbrink, Gabriel Curio, Arno Villringer, Klaus-Robert Müller, and Benjamin Blankertz. "Enhanced performance by a hybrid NIRS–EEG brain computer interface". In: *NeuroImage* 59.1 (2012), pp. 519–529. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.07.084. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811911008792.

[35] Silke Telkemeyer, Sonja Rossi, Stefan P Koch, Till Nierhaus, Jens Steinbrink, David Poeppel, Hellmuth Obrig, and Isabell Wartenburger. "Sensitivity of newborn auditory cortex to the temporal structure of sounds." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.47 (2009), pp. 14726–33. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1246-09.2009. URL: http://www.ncbi.nlm.nih.gov/pubmed/19940167.

[36] Sonja Rossi, Silke Telkemeyer, Isabell Wartenburger, and Hellmuth Obrig. "Shedding light on words and sentences: Near-infrared spectroscopy in language research". In: *Brain and Language* 121.2 (2012), pp. 152–163. ISSN: 0093934X. DOI: 10.1016/j.bandl.2011.03.008. URL: http://dx.doi.org/10.1016/j.bandl.2011.03.008.

[37] F. Wallois, A. Patil, C. Héberlé, and R. Grebe. "EEG-NIRS in epilepsy in children and neonates". In: *Neurophysiologie Clinique/Clinical Neurophysiology*

40.5-6 (2010), pp. 281–292. ISSN: 09877053. DOI: 10.1016/j.neucli.2010.08.004.

[38] John R Glover. "Overview of Functional Magnetic Resonance Imaging". In: *Neurosurg Clin N Am.* 22.2 (2011), pp. 133–139. DOI: 10.1016/j.nec.2010.11.001.Overview.

[39] Michael K. Stehling, Robert Turner, and Peter Mansfield. *Echo-planar imaging: Magnetic resonance imaging in a fraction of a second*. 1991. DOI: 10.1126/science.1925560.

[40] Fa Hsuan Lin, Kevin W.K. Tsai, Ying Hua Chu, Thomas Witzel, Aapo Nummenmaa, Tommi Raij, Jyrki Ahveninen, Wen Jui Kuo, and John W. Belliveau. *Ultrafast inverse imaging techniques for fMRI*. 2012. DOI: 10.1016/j.neuroimage.2012.01.072.

[41] Muyue Yang, Zhen Yang, Tifei Yuan, Wuwei Feng, and Pu Wang. "A systemic review of functional near-infrared spectroscopy for stroke: Current application and future directions". In: *Frontiers in Neurology* 10.58 (2019). ISSN: 16642295. DOI: 10.3389/fneur.2019.00058.

[42] Pei Yi Lin, Sang I. Lin, Trevor Penney, and Jia Jin Jason Chen. "Applications of near infrared spectroscopy and imaging for motor rehabilitation in stroke patients". In: *Journal of Medical and Biological Engineering* 29.5 (2009), pp. 210–221. ISSN: 16090985.

[43] Hellmuth Obrig. "NIRS in clinical neurology - a 'promising' tool?" In: *NeuroImage* 85.85 (2014), pp. 535–546. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.03.045. arXiv: NIHMS150003.

[44] Lena H. Ernst, Sabrina Schneider, Ann Christine Ehlis, and Andreas J. Fallgatter. *Functional near infrared spectroscopy in psychiatry: A critical review*. 2012. DOI: 10.1255/jnirs.970.

[45] Simone Cutini, Sara Basso Moro, and Silvia Bisconti. "Review: Functional near infrared optical imaging in cognitive neuroscience: An introductory review". In: *Journal of Near Infrared Spectroscopy* 20.1 (2012), pp. 75–92. ISSN: 09670335. DOI: 10.1255/jnirs.969.

[46] Sabrina Brigadoi, Lisa Ceccherini, Simone Cutini, Fabio Scarpa, Pietro Scatturin, Juliette Selb, Louis Gagnon, David A. Boas, and Robert J. Cooper. "Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data". In: *NeuroImage* 85.1 (2014), pp. 181–191. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.04.082. arXiv: NIHMS150003.

[47] Judit Gervain. "Plasticity in early language acquisition: the effects of prenatal and early childhood experience". In: *Current Opinion in Neurobiology* 35.JUNE (2015), pp. 13–20. ISSN: 09594388. DOI: 10.1016/j.conb.2015.05.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S095943881500094X.

[48] M. Pena, A. Maki, D. Kovacic, G. Dehaene-Lambertz, H. Koizumi, F. Bouquet, and J. Mehler. "Sounds and silence: An optical topography study of language recognition at birth". In: *Proceedings of the National Academy of Sciences* 100.20 (2003), pp. 11702–11705. ISSN: 0027-8424. DOI: 10.1073/pnas.1934290100.

[49] Richard N. Aslin. "Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy". In: *Cognitive Neuropsychology* 29.1-2 (2012), pp. 7–33. ISSN: 02643294. DOI: 10.1080/02643294.2012.654773.

[50] Judit Gervain, Jacques Mehler, Janet F. Werker, Charles A. Nelson, Gergely Csibra, Sarah Lloyd-Fox, Mohinish Shukla, and Richard N. Aslin. "Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium". In: *Developmental Cognitive Neuroscience* 1.1 (2011), pp. 22–46. ISSN: 18789293. DOI: 10.1016/j.dcn.2010.07.004. arXiv: arXiv:1011.1669v3.

[51] S. Lloyd-Fox, A. Blasi, and C. E. Elwell. "Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy". In: *Neuroscience and Biobehavioral Reviews* 34.3 (2010), pp. 269–284. ISSN: 01497634. DOI: 10.1016/j.neubiorev.2009.07.008.

[52] Kenneth R. Pugh, Nicole Landi, Jonathan L. Preston, W. Einar Mencl, Alison C. Austin, Daragh Sibley, Robert K. Fulbright, Mark S. Seidenberg, Elena L. Grigorenko, R. Todd Constable, Peter Molfese, and Stephen J. Frost. "The relationship between phonological and auditory processing and brain organization in beginning readers". In: *Brain and Language* 125.2 (2013), pp. 173–183. ISSN: 0093934X. DOI: 10.1016/j.bandl.2012.04.004. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.bandl.2012.04.004.

[53] Robert J. Cooper, Juliette Selb, Louis Gagnon, Dorte Phillip, Henrik W. Schytz, Helle K. Iversen, Messoud Ashina, and David A. Boas. "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy". In: *Frontiers in Neuroscience* 6.OCT (2012), pp. 1–10. ISSN: 16624548. DOI: 10.3389/fnins.2012.00147.

[54] Yiheng Zhang, Dana H. Brooks, Maria Angela Franceschini, and David A. Boas. "Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging". In: *Journal of Biomedical Optics* 10.1 (2005), p. 011014. ISSN: 10833668. DOI: 10.1117/1.1852552.

[55] Behnam Molavi and Guy A. Dumont. "Wavelet-based motion artifact removal for functional near-infrared spectroscopy". In: *Physiological Measurement* 33.2 (2012), pp. 259–270. ISSN: 09673334. DOI: 10.1088/0967-3334/33/2/259.

[56] F Scholkmann, S Spichtig, T Muehlemann, and M Wolf. "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation." In: *Physiological measurement* 31.5 (2010), pp. 649–662. ISSN: 0967-3334. DOI: 10.1088/0967-3334/31/5/004.

[57] Ivan W. Selesnick, Harry L Graber, Ying Ding, Tong Zhang, and Randall L Barbour. "Transient Artifact Reduction using Sparse Optimization". In: 62.1 (2014), pp. 5–7.

[58] Xian Zhang, Jack Adam Noah, and Joy Hirsch. "Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering". In: *Neurophotonics* 3.1 (2016), p. 015004. ISSN: 2329-423X. DOI: 10.1117/1.NPh.3.1.015004. URL: http://neurophotonics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.NPh.3.1.015004.

[59] Sabrina Brigadoi and Robert J Cooper. "How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy". In: *Neurophotonics* 2.2 (2015), p. 025005. ISSN: 2329-423X. DOI: 10.1117/1.nph.2.2.025005.

[60] Ilias Tachtsidis, Peck H. Koh, Charlotte Stubbs, and Clare E. Elwell. "Functional optical topography analysis using statistical parametric mapping (SPM) methodology with and without physiological confounds". In: *Advances in Experimental Medicine and Biology*. Vol. 662. 2010, pp. 237–243. ISBN: 9781441912398. DOI: 10.1007/978-1-4419-1241-1_34.

[61] Tsukasa Funane, Hirokazu Atsumori, Takusige Katura, Akiko N. Obata, Hiroki Sato, Yukari Tanikawa, Eiji Okada, and Masashi Kiguchi. "Quantitative evaluation of deep and shallow tissue layers' contribution to fNIRS signal using multi-distance optodes and independent component analysis". In: *NeuroImage* 85 (2014), pp. 150–165. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.02.026.

[62] Rolf B. Saager and Andrew J. Berger. "Direct characterization and removal of interfering absorption trends in two-layer turbid media". In: *Journal of the Optical Society of America A* 22.9 (2005), p. 1874. ISSN: 1084-7529. DOI: 10.1364/JOSAA.22.001874. URL: https://www.osapublishing.org/abstract.cfm?URI=josaa-22-9-1874.

[63] Paola Pinti, Felix Scholkmann, Antonia Hamilton, Paul Burgess, and Ilias Tachtsidis. "Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework". In: *Frontiers in Human Neuroscience* 12.January (2019), pp. 1–21. DOI: 10.3389/fnhum.2018.00505.

[64] Theodore J. Huppert, Solomon G. Diamond, Maria A. Franceschini, and David A Boas. "HomER: a review of time-series analysis methods for near- infrared spectroscopy of the brain". In: 15.10 (2009), pp. 1203–1214. ISSN: 08966273. DOI: 10.1016/j.drugalcdep.2008.02.002.A. arXiv: NIHMS150003.

[65] Hendrik Santosa, Xuetong Zhai, Frank Fishburn, and Theodore Huppert. "The NIRS Brain AnalyzIR Toolbox". In: (2018). DOI: 10.3390/a11050073.

[66] Yong Xu, Harry L. Graber, and Randall L. Barbour. "nirsLAB: A Computing Environment for fNIRS Neuroimaging Data Analysis". In: *Biomedical Optics 2014* (2014), BM3A.1. DOI: 10.1364/BIOMED.2014.BM3A.1. URL: https://www.osapublishing.org/abstract.cfm?uri=BIOMED-2014-BM3A.1.

[67] Sungho Tak and Jong Chul Ye. "Statistical analysis of fNIRS data: A comprehensive review". In: *NeuroImage* 85 (2014), pp. 72–91. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.06.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811913006538.

[68] Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ Frackowiak. "Statistical parametric maps in functional imaging: a general linear approach". In: *Human brain mapping* 2.4 (1994), pp. 189–210.

[69] Geoffrey M Boynton, Stephen A Engel, Gary H Glover, and David J Heeger. "Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1". In: *The Journal of Neuroscience* 16.13 (1996), pp. 4207–4221. ISSN: 0270-6474. URL: papers3://publication/uuid/B2E61BDB-C31A-4BC7-8119-A9F33C472D11.

[70] Hendrik Santosa, Theodore J Huppert, Hendrik Santosa, Frank Fishburn, Xuetong Zhai, and Theodore J Huppert. "Investigation of the sensitivity-specificity of canonical- and deconvolution-based linear models in evoked functional near-infrared spectroscopy". In: *Neurophotonics* 6.02 (2019), p. 1. ISSN: 2329-423X. DOI: 10.1117/1.NPh.6.2.025009. URL: https://www.spiedigitallibrary.org/journals/neurophotonics/volume-6/issue-02/025009/Investigation-of-the-sensitivity-specificity-of-canonical--and-deconvolution/10.1117/1.NPh.6.2.025009.full.

[71] A C Aitken. "IV.—On Least Squares and Linear Combination of Observations". In: *Proceedings of the Royal Society of Edinburgh* 55 (1936), pp. 42–48. DOI: 10.1017/S0370164600014346.

[72] J Ye, S Tak, K Jang, and J Jung. "NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy". In: *NeuroImage* 44.2 (2009), pp. 428–447. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.08.036. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811908009695.

[73] Richard Henson, Michael D Rugg, Karl J Friston, and Others. "The choice of basis functions in event-related fMRI". In: *NeuroImage* 13.6 (2001), p. 149.

[74] Minako Uga, Ippeita Dan, Toshifumi Sano, Haruka Dan, and Eiju Watanabe. "Optimizing the general linear model for functional near-infrared spectroscopy: an adaptive hemodynamic response function approach". In: *Neurophotonics* 1.1 (2014), p. 015004. ISSN: 2329-423X. DOI: 10.1117/1.nph.1.1.015004.

[75] V. D. Calhoun, M. C. Stevens, G. D. Pearlson, and K. A. Kiehl. "fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms". In: *NeuroImage* 22.1 (2004), pp. 252–257. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2003.12.029.

[76] M A Lindquist and T D Wager. "Validity and Power in Hemodynamic Response Modeling: A Comparison Study and a New Approach". In: *Hum Brain Mapp* 28.8 (2012), pp. 764–784. DOI: 10.1002/hbm.20310.Validity.

[77] C. H. Liao, K. J. Worsley, J. B. Poline, J. A.D. Aston, G. H. Duncan, and A. C. Evans. "Estimating the delay of the fMRI response". In: *NeuroImage* 16.3 I (2002), pp. 593–606. ISSN: 10538119. DOI: 10.1006/nimg.2002.1096.

[78] Valeria Della-Maggiore, Wilkin Chau, Pedro R. Peres-Neto, and Anthony R. McIntosh. "An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data". In: *NeuroImage* 17.1 (2002), pp. 19–28. ISSN: 10538119. DOI: 10.1006/nimg.2002.1113.

[79] Guilherne Augusto Zimeo Morais, Felix Scholkmann, Joana Bisol Balardin, Rogério Akira Furucho, Renan Costa Vieira de Paula, Claudinei Eduardo Biazoli, and João Ricardo Sato. "Non-neuronal evoked and spontaneous hemodynamic changes in the anterior temporal region of the human head may lead to misinterpretations of functional near-infrared spectroscopy signals". In: *Neurophotonics* 5.01 (2017), p. 1. ISSN: 2329-423X. DOI: 10.1117/1.NPh.5.1.011002. URL: https://www.spiedigitallibrary.org/journals/neurophotonics/volume-5/issue-01/011002/Non-neuronal-evoked-and-spontaneous-hemodynamic-changes-in-the-anterior/10.1117/1.NPh.5.1.011002.full.

[80] Nils Volkening, Anirudh Unni, Birte Sofie Löffler, Sebastian Fudickar, Jochem W. Rieger, and Andreas Hein. "Characterizing the Influence of Muscle Activity in fNIRS Brain Activation Measurements". In: *IFAC-PapersOnLine* 49.11 (2016), pp. 84–88. ISSN: 24058963. DOI: 10.1016/j.ifacol.2016.08.013.

[81] M. Ferrari, T. Binzoni, and V. Quaresima. "Oxidative metabolism in muscle". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 352.1354 (1997), pp. 677–683. ISSN: 09628436. DOI: 10.1098/rstb.1997.0049.

[82] Toru Yamamoto and Toshinori Kato. "Paradoxical correlation between signal in functional magnetic resonance imaging and deoxygenated haemoglobin content in capillaries: A new theoretical explanation". In: *Physics in Medicine and Biology* 47.7 (2002), pp. 1121–1141. ISSN: 00319155. DOI: 10.1088/0031-9155/47/7/309.

[83] Thomas Nichols and Satoru Hayasaka. "Controlling the familywise error rate in functional neuroimaging: A comparative review". In: *Statistical Methods in Medical Research* 12.5 (2003), pp. 419–446. ISSN: 09622802. DOI: 10.1191/0962280203sm341ra. arXiv: arXiv:1711.02231v1.

[84] Yoav Benjamini and Yosef Hochberg. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. 1995. DOI: 10.2307/3866483.

[85] Archana K Singh and Ippeita Dan. "Exploring the false discovery rate in multichannel NIRS". In: *Neuroimage* 33.2 (2006), pp. 542–549.

[86] Hellmuth Obrig, Sonja Rossi, Silke Telkemeyer, and Isabell Wartenburger. "From acoustic segmentation to language processing: evidence from optical imaging". In: *Frontiers in Neuroenergetics* 2.13 (2010). ISSN: 16626427. DOI: 10.3389/fnene.2010.00013. URL: http://journal.frontiersin.org/article/10.3389/fnene.2010.00013/abstract.

[87] Shirley Coyle, Tomás Ward, Charles Markham, and Gary McDarby. "On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces". In: *Physiological Measurement* 25.4 (2004), pp. 815–822. ISSN: 09673334. DOI: 10.1088/0967-3334/25/4/003.

[88] Noman Naseer and Keum-Shik Hong. "fNIRS-based brain-computer interfaces: a review". In: *Frontiers in Human Neuroscience* 9.January (2015), pp. 1–15. ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00003. arXiv: 0005074v1 [arXiv:astro-ph]. URL: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00003/abstract.

[89] Abdelhak Mahmoudi, Sylvain Takerkart, Fakhita Regragui, Driss Boussaoud, and Andrea Brovelli. "Multivoxel pattern analysis for fMRI data: A review". In: *Computational and Mathematical Methods in Medicine* 2012 (2012). ISSN: 1748670X. DOI: 10.1155/2012/961257.

[90] Kenneth A. Norman, Sean M. Polyn, Greg J. Detre, and James V. Haxby. "Beyond mind-reading: multi-voxel pattern analysis of fMRI data". In: *Trends in Cognitive Sciences* 10.9 (2006), pp. 424–430. ISSN: 13646613. DOI: 10.1016/j.tics.2006.07.005.

[91] John-Dylan Haynes and Geraint Rees. "Decoding mental states from brain activity in humans". In: *Nature Reviews Neuroscience* 7.7 (2006), pp. 523–534. ISSN: 1471-003X. DOI: 10.1038/nrn1931. URL: http://www.nature.com/doifinder/10.1038/nrn1931.

[92] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006, p. 758. ISBN: 9780387310732. DOI: 10.1021/jo01026a014. arXiv: arXiv:1011.1669v3.

[93] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. 2001. DOI: 10.1007/BF01237942. arXiv: 0-387-31073-8.

[94] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and K.-R. Müller. "Single-trial analysis and classification of {ERP} components–a tutorial". In: *NeuroImage* 56.2 (2011), pp. 814–825.

[95] Johannes Höhne, Daniel Bartz, Martin N. Hebart, Klaus-Robert Müller, and Benjamin Blankertz. "Analyzing neuroimaging data with subclasses: A shrinkage approach". In: *NeuroImage* 124 (2016), pp. 740–751. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2015.09.031. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811915008460.

[96] Maogeng Xia, Sutao Song, Li Yao, and Zhiying Long. "An empirical comparison of different LDA methods in fMRI-based brain states decoding". In: *Bio-Medical Materials and Engineering* (2015). ISSN: 18783619. DOI: 10.3233/BME-151415.

[97] Gunther Bauernfeind, David Steyrl, Clemens Brunner, and Gernot R. Muller-Putz. "Single trial classification of fNIRS-based brain-computer interface mental arithmetic data: A comparison between different classifiers". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*. 2014. ISBN: 9781424479290. DOI: 10.1109/EMBC.2014.6944008.

[98] Jaeyoung Shin, Klaus-R Müller, and Han-Jeong Hwang. "Near-infrared spectroscopy (NIRS)-based eyes-closed brain-computer interface (BCI) using prefrontal cortex activation due to mental arithmetic". In: *Scientific Reports* 6.October (2016), p. 36203. ISSN: 2045-2322. DOI: 10.1038/srep36203. URL: http://www.nature.com/articles/srep36203.

[99] Stephen Marsland. *Machine Learning: An Algorithmic Perspective*. Chapman and Hall/CRC, 2014, p. 406. ISBN: 1420067192. URL: http://books.google.com/books?id=n66O8a4SWGEC{\&}pgis=1.

[100] Louis Gagnon, Katherine Perdue, Douglas N. Greve, Daniel Goldenholz, Gayatri Kaskhedikar, and David A. Boas. "Improved recovery of the hemodynamic

response in diffuse optical imaging using short optode separations and state-space modeling". In: *NeuroImage* 56.3 (2011), pp. 1362–1371. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.03.001. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.neuroimage.2011.03.001.

[101] Jeffrey W. Barker. "Estimation of Cerebral Physiology and Hemodynamics via Near-Infrared Spectroscopy". PhD thesis. University of Pittsburgh, 2014. URL: http://d-scholarship.pitt.edu/.

[102] Claude Julien. "The enigma of Mayer waves: Facts and models". In: *Cardiovascular Research* 70.1 (2006), pp. 12–21. ISSN: 00086363. DOI: 10.1016/j.cardiores.2005.11.008.

[103] Meryem A. Yücel, Juliette Selb, Christopher M. Aasted, Pei-Yi Lin, David Borsook, Lino Becerra, and David A. Boas. "Mayer waves reduce the accuracy of estimated hemodynamic response functions in functional near-infrared spectroscopy". In: *Biomedical Optics Express* 7.8 (2016), p. 3078. ISSN: 2156-7085. DOI: 10.1364/BOE.7.003078. URL: https://www.osapublishing.org/abstract.cfm?URI=boe-7-8-3078.

[104] Benjamin Blankertz, Laura Acqualagna, Sven Dähne, Stefan Haufe, Matthias Schultze-Kraft, Irene Sturm, Marija Ušćumlic, Markus A Wenzel, Gabriel Curio, and Klaus-Robert Müller. "The Berlin Brain-Computer Interface: Progress Beyond Communication and Control". In: *Frontiers in Neuroscience* 10 (2016), p. 530. ISSN: 1662-453X. DOI: 10.3389/fnins.2016.00530. URL: https://www.frontiersin.org/article/10.3389/fnins.2016.00530.

[105] *BBCI toolbox*. URL: https://github.com/bbci/bbci{_}public (visited on 11/03/2016).

[106] Masaki Kameyama, Masato Fukuda, Toru Uehara, and Masahiko Mikuni. "Sex and age dependencies of cerebral blood volume changes during cognitive activation: A multichannel near-infrared spectroscopy study". In: *NeuroImage* 22.4 (2004), pp. 1715–1721. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2004.03.050.

[107] Hongyu Yang, Zhenyu Zhou, Yun Liu, Zongcai Ruan, Hui Gong, Qingming Luo, and Zuhong Lu. "Gender difference in hemodynamic responses of prefrontal area to emotional stress by near-infrared spectroscopy". In: *Behavioural Brain*

Research 178.1 (2007), pp. 172–176. ISSN: 01664328. DOI: 10.1016/j.bbr.2006.11.039.

[108] M. J. Herrmann, A. Walter, A. C. Ehlis, and A. J. Fallgatter. "Cerebral oxygenation changes in the prefrontal cortex: Effects of age and gender". In: *Neurobiology of Aging* 27.6 (2006), pp. 888–894. ISSN: 01974580. DOI: 10.1016/j.neurobiolaging.2005.04.013.

[109] J.A.E Anderson, K.L Campbell, T Amer, C.L Grady, and L Hasher. "Timing Is everything: Age differences in the cognitive control network are modulated by time of day." In: *Psychology and aging* 29.3 (2014), pp. 648–657. ISSN: 1939-1498. DOI: 10.1037/a0037243. URL: http://www.ncbi.nlm.nih.gov/pubmed/24999661.

[110] Jack L. Lancaster, Marty G. Woldorff, Lawrence M. Parsons, Mario Liotti, Catarina S. Freitas, Lacy Rainey, Peter V. Kochunov, Dan Nickerson, Shawn A. Mikiten, and Peter T. Fox. "Automated Talairach Atlas labels for functional brain mapping". In: *Human Brain Mapping* 10.3 (2000), pp. 120–131. ISSN: 10659471. DOI: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8.

[111] Christopher M. Aasted, Meryem A. Yücel, Robert J. Cooper, Jay Dubb, Daisuke Tsuzuki, Lino Becerra, Mike P. Petkov, David Borsook, Ippeita Dan, and David A. Boas. "Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial". In: *Neurophotonics* 2.2 (2015), p. 020801. ISSN: 2329-423X. DOI: 10.1117/1.NPh.2.2.020801. URL: http://neurophotonics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.NPh.2.2.020801.

[112] Mike X Cohen. *Analyzing neural time series data: theory and practice*. MIT Press. MIT Press, 2014.

[113] Harry L. Graber, Rabah Al abdi, Yong Xu, Armand P. Asarian, Peter J. Pappas, Lisa Dresner, Naresh Patel, Kuppuswamy Jagarlamundi, William B. Solomon, and Randall L. Barbour. "Enhanced resting-state dynamics of the hemoglobin signal as a novel biomarker for detection of breast cancer". In: *Medical Physics* 42.11 (2015), pp. 6406–6424. ISSN: 00942405. DOI: 10.1118/1.4932220. URL: http://dx.doi.org/10.1118/1.4932220.

[114] Saying Chen, Pan Ning, Kaoru Sakatani, Huancong Zuo, Wemara Lichty, and Shimin Zhao. "Auditory-evoked cerebral oxygenation changes in hypoxic-ischemic

encephalopathy of newborn infants monitored by near infrared spectroscopy". In: *Early Human Development* 67.1-2 (2002), pp. 113–121. ISSN: 03783782. DOI: 10.1016/S0378-3782(02)00004-X.

[115] K Sakatani, S Chen, W Lichty, H Zuo, and Y P Wang. "Cerebral blood oxygenation changes induced by auditory stimulation in newborn infants measured by near infrared spectroscopy." In: *Early human development* 55.3 (1999), pp. 229–236. ISSN: 0378-3782. DOI: S0378-3782(99)00019-5[pii].

[116] Karen E. Waldie, Charlotte E. Haigh, Gjurgjica Badzakova-Trajkov, Jude Buckley, and Ian J. Kirk. "Reading the wrong way with the right hemisphere". In: *Brain Sciences* 3.3 (2013), pp. 1060–1075. ISSN: 20763425. DOI: 10.3390/brainsci3031060.

[117] Donald Shankweiler, W. Einar Mencl, David Braze, Whitney Tabor, Kenneth R. Pugh, and Robert K. Fulbright. "Reading differences and brain: Cortical integration of speech and print in sentence processing varies with reader skill". In: *Developmental Neuropsychology* 33.6 (2008), pp. 745–775. ISSN: 87565641. DOI: 10.1080/87565640802418688.

[118] Karen E Waldie and James L Mosley. "Developmental trends in right hemispheric participation in reading". In: *Neuropsychologia* 38 (2000), pp. 462–474.

[119] Stanislas Dehaene. *Reading in the brain: The new science of how we read*. Penguin, 2009.

[120] Archana K. Singh, Masako Okamoto, Haruka Dan, Valer Jurcak, and Ippeita Dan. "Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI". In: *NeuroImage* 27.4 (2005), pp. 842–851. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2005.05.019.

[121] N Tzourio-Mazoyer, B Landeau, D Papathanassiou, F Crivello, O Etard, N Delcroix, B Mazoyer, and M Joliot. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain." In: *NeuroImage* 15.1 (2002), pp. 273–289. ISSN: 1053-8119. DOI: 10.1006/nimg.2001.0978.

[122] Xian Zhang, Jack Adam Noah, Swethasri Dravida, and Joy Hirsch. "Signal processing of functional NIRS data acquired during overt speaking". In: *Neurophotonics* 4.04 (2017), p. 1. ISSN: 2329-423X. DOI: 10.1117/1.NPh.4.4.041409. URL: https://www.spiedigitallibrary.org/journals/neurophotonics/

volume-4/issue-04/041409/Signal-processing-of-functional-NIRS-data-acquired-during-overt-speaking/10.1117/1.NPh.4.4.041409.full.

[123] Melissa Hardy and John Reynolds. "Incorporating Categorical Information into Regression Models: The Utility of Dummy Variables". In: *Handbook of Data Analysis*. 2012, pp. 209–237. DOI: 10.4135/9781848608184.n9.

[124] A. Castro-Caldas, K. M. Petersson, A. Reis, S. Stone-Elander, and M. Ingvar. "The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain". In: *Brain* 121.6 (1998), pp. 1053–1063. ISSN: 00068950. DOI: 10.1093/brain/121.6.1053.

[125] A Reis and A Castro-Caldas. "Illiteracy: a cause for biased cognitive development." In: *Journal of the International Neuropsychological Society : JINS* 3.5 (1997), pp. 444–50. ISSN: 1355-6177. URL: http://www.ncbi.nlm.nih.gov/pubmed/9322403.

[126] Federico De Martino, Giancarlo Valente, Noël Staeren, John Ashburner, Rainer Goebel, and Elia Formisano. "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns". In: *NeuroImage* 43.1 (2008), pp. 44–58. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.06.037.

[127] Christian Habeck and Y Stern. "Multivariate data analysis for neuroimaging data: overview and Application to Alzheimer's disease". In: *Cell Biochem Biophys* 58.2 (2010), pp. 53–67. DOI: 10.1007/s12013-010-9093-0.Multivariate. URL: http://link.springer.com/article/10.1007/s12013-010-9093-0.

[128] David D. Cox and Robert L. Savoy. "Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex". In: *NeuroImage* 19.2 (2003), pp. 261–270. ISSN: 10538119. DOI: 10.1016/S1053-8119(03)00049-1.

[129] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. "Distributed and overlapping representations of faces and objects in ventral temporal cortex". In: *Science* 293.September (2001), pp. 2425–2430. ISSN: 0036-8075. DOI: 10.1126/science.1063736.

[130] Irina Simanova, Marcel van Gerven, Robert Oostenveld, and Peter Hagoort. "Identifying object categories from event-related EEG: Toward decoding of

conceptual representations". In: *PLoS ONE* 5.12 (2010), e14465. DOI: https://doi.org/10.1371/journal.pone.0014465.

[131] Radoslaw Martin Cichy, Dimitrios Pantazis, and Aude Oliva. "Resolving human object recognition in space and time." In: *Nature neuroscience* (2014). ISSN: 1546-1726. DOI: 10.1038/nn.3635.

[132] Hesheng Liu, Yigal Agam, Joseph R. Madsen, and Gabriel Kreiman. "Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex". In: *Neuron* (2009). ISSN: 08966273. DOI: 10.1016/j.neuron.2009.02.025.

[133] Chih-chung Chang and Chih-jen Lin. "LIBSVM : A Library for Support Vector Machines". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 2 (2013), pp. 1–39. ISSN: 21576904. DOI: 10.1145/1961189.1961199. arXiv: 0-387-31073-8.

[134] Masaya Misaki, Youn Kim, Peter A. Bandettini, and Nikolaus Kriegeskorte. "Comparison of multivariate classifiers and response normalizations for pattern-information fMRI". In: *NeuroImage* 53.1 (2010), pp. 103–118. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.05.051. URL: http://dx.doi.org/10.1016/j.neuroimage.2010.05.051.

[135] Kelly Tai and Tom Chau. "Single-trial classification of NIRS signals during emotional induction tasks: Towards a corporeal machine interface". In: *Journal of NeuroEngineering and Rehabilitation* 6.1 (2009), pp. 1–14. ISSN: 17430003. DOI: 10.1186/1743-0003-6-39.