

Identifying and Decoding the Multivariate Brain Activity for Emotional States in Self – and Empathic Perception

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Biomedical Engineering

eingereicht von

Lukas Traxler

Matrikelnummer 1028999

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: a.o.Univ.-Prof. Dipl.-Ing. Dr. Robert Sablatnig
Mitwirkung: Ass.Prof. Dipl.-Ing. Dr. Georg Langs – CIR Lab, Medical University of Vienna
Univ.-Prof. Dr. Claus Lamm – SCAN Unit, University of Vienna

Wien, 28. Januar 2014

(Unterschrift Verfasser)

(Unterschrift Betreuung)

Identifying and Decoding the Multivariate Brain Activity for Emotional States in Self – and Empathic Perception

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Biomedical Engineering

by

Lukas Traxler

Registration Number 1028999

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: a.o.Univ.-Prof. Dipl.-Ing. Dr. Robert Sablatnig

Assistance: Ass.Prof. Dipl.-Ing. Dr. Georg Langs – CIR Lab, Medical University of Vienna
Univ.-Prof. Dr. Claus Lamm – SCAN Unit, University of Vienna

Vienna, 28. Januar 2014

(Signature of Author)

(Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Lukas Traxler
Schirnes 3, 3842 Thaya

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)

(Unterschrift Verfasser)

Acknowledgements

I would like to express my special appreciation and thanks to Ass. Prof. Dipl.-Ing. Dr. Georg Langs for providing this interesting thesis topic, guidance through the work, support, advice and taking the time for many discussions on the topic. I also want to thank Univ.-Prof. Dr. Claus Lamm for providing the fMRI–data and his expertise regarding the psychological context of the experiments. Thus I could profit from an interdisciplinary supervision which was very helpful during the course of this project. In addition I want to thank a.o. Univ.-Prof. Dipl.-Ing. Dr. Robert Sablatnig for the supervision of my thesis. I also extend my thanks to my family and friends who gave me their warmth and responsiveness during the time I carried out my work.

Abstract

Functional Magnetic Resonance Imaging (fMRI) is a widely used tool in neurosciences. It allows to measure metabolic properties related to neuronal activity in the working human brain. With that, processes in the human brain can be studied on a neuronal level. One field of current research in psychological neurosciences is the investigation of the neuronal basis for empathy. Empathy allows the human being to share affective states of others and is fundamental in social interaction.

A huge variety of methods for analyzing fMRI measurements exist. Currently methods based on univariate statistics are predominantly used in empathy research. One factor of interest is to identify shared representations between self- and empathic-perception. Methods capturing multivariate properties have successfully been used to decode participant's mental states from fMRI measurements and to increase sensitivity in detecting neuronal activity involved in the neuronal process. The aim of this thesis is to describe and demonstrate a novel method based on a multivariate Random Forest classifier on real data. The method successfully decodes the mental state induced by visuo- tactile stimuli. Further it is able to detect multivariate cortical activation patterns distributed across the entire brain, which are shared between self- and empathic perception. The thesis also deals with the visualization of the multivariate coding pattern, as well as with visualization of long ranged conjunctions which are informative in decoding the mental state. The results from this method will give support for new insight into distributed patterns of neuronal processes of empathy.

Kurzfassung

Funktionelle Magnetresonanztomographie (fMRT) ist eine weit verbreitete Methode für neurowissenschaftliche Fragestellungen. Dabei werden Veränderungen durch metabolische Aktivität, die durch neuronalen Prozesse hervorgerufen wird, gemessen. Dies ermöglicht die Untersuchung dieser neuronalen Prozesse im menschlichen Gehirn, während dieses arbeitet. Die Untersuchung der neuronalen Basis für Empathie ist ein aktuelles Forschungsgebiet in den psychologischen Neurowissenschaften. Empathie ist eine Fähigkeit des Menschen, welche es ermöglicht, Zustände anderer Personen nachzuempfinden und ist damit ein fundamentaler Baustein für soziale Interaktion.

Für die fMRT Datenanalyse stehen eine Vielzahl von Methoden zur Verfügung, wobei in der Empathieforschung derzeit vorwiegend Methoden, die auf univariater Statistik basieren, verwendet werden. Ein interessanter Aspekt dieser Untersuchungen ist die Identifikation von neuronalen Mechanismen die sowohl in der Eigenwahrnehmung als auch in der empathischen Wahrnehmung aktiv sind. Methoden, welche auf multivariater Statistik basieren, wurden erfolgreich eingesetzt, einerseits um aus den gemessenen Hirnaktivitäten den Mentalzustand der untersuchten Person abzulesen, andererseits zeigen diese Methoden eine erhöhte Sensitivität in der Identifikation von involvierten neuronalen Mechanismen. Zweck dieser Arbeit ist die Beschreibung und Demonstration einer neuen Methode, welche auf dem multivariaten Random Forest Klassifikator basiert. Die Methode ist in der Lage, die Emotion eines visuell-taktilen Reizes zu dekodieren. Des Weiteren werden die codierenden Muster in der Cortex-Aktivität, welche sowohl für Eigenwahrnehmung als auch in der empathischen Wahrnehmung relevant sind, identifiziert. Die Arbeit befasst sich darüber hinaus auch mit der Visualisierung dieser codierenden Muster sowie der Darstellung von informativen Zusammenhängen von örtlich weit entfernten Aktivierungen. Diese Resultate bieten neuen Einblick örtlich verteilten neuronalen Muster für empathische Wahrnehmung.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Motivation and Aim of the Work | 1 |
| 1.2 | Problem Statement | 2 |
| 1.3 | Methodological Approach | 3 |
| 1.4 | Mathematical Notation | 3 |
| 1.5 | Structure of the Thesis | 4 |
| 2 | Functional Magnetic Resonance Imaging as Tool for Neuroscience | 5 |
| 2.1 | Signal Origin in fMRI | 5 |
| 2.1.1 | Physical Principles of MRI | 5 |
| 2.1.2 | The Blood Oxygenation Level Dependent Contrast and fMRI | 8 |
| 2.1.3 | Limitations of fMRI | 8 |
| 2.2 | Principles on Study Design | 9 |
| 2.3 | Summary | 10 |
| 3 | Basics in Functional Magnetic Resonance Imaging Data Analysis | 11 |
| 3.1 | The General Linear Model | 11 |
| 3.2 | Multivariate Pattern Analysis | 13 |
| 3.2.1 | Preparing Data for MVPA | 13 |
| 3.2.2 | Estimating the Activity Pattern | 14 |
| 3.2.3 | Selecting Voxels and Searchlight MVPA | 14 |
| 3.2.4 | Classifiers in MVPA | 15 |
| 3.2.5 | Testing and Validation of the Classifier–Model | 15 |
| 3.3 | Activation – vs. Information – Based Analysis | 16 |
| 3.3.1 | What Questions are Answered by the Different Analysis Methods? | 16 |
| 3.3.2 | Comparison of Univariate Analysis and MVPA | 16 |
| 3.3.3 | Another Point of View: Forward- and Backward–Models | 17 |
| 3.4 | Random Forest Classifier | 18 |
| 3.4.1 | Functioning of the RF | 18 |
| 3.4.2 | Variable Importance Measures | 21 |
| 3.4.3 | Special Issues for RF Application on fMRI–Data | 23 |
| 3.5 | Summary and Limitations of the State-of-the-Art fMRI–Analysis Methods | 23 |

| | | |
|----------|---|-----------|
| 4 | The Empathic Brain | 25 |
| 4.1 | Definition and Factors of Empathy | 25 |
| 4.2 | Study of Principal Neuronal Mechanisms for Empathy | 26 |
| 4.2.1 | Empathy of Pain | 26 |
| 4.2.2 | Empathy in non Pain Experiments | 27 |
| 4.3 | Empathy Related Studies of Interest and MVPA | 27 |
| 4.4 | Map of the Empathic Brain | 28 |
| 4.5 | Summary | 28 |
| 5 | Methods for Detecting Shared Distributed Patterns of Neural Activity | 31 |
| 5.1 | Contribution of the Proposed Method | 32 |
| 5.1.1 | Contribution 1: Identification of Distributed Shared Patterns | 32 |
| 5.1.2 | Contribution 2: Quantification of Informativeness of Combined Features | 33 |
| 5.2 | Study Design: A Paradigm for Self – Perception and Empathy | 33 |
| 5.2.1 | Experimental Setup and Stimuli | 33 |
| 5.2.2 | Technical Details on the fMRI Data – Acquisition and Preprocessing | 35 |
| 5.3 | Data Preparation for Classification | 35 |
| 5.3.1 | Cortical Alignment of the fMRI–Data | 35 |
| 5.3.2 | Preparing Labels for the Classifier | 37 |
| 5.4 | Classifier Training for Decoding the Emotional State from fMRI–Data | 39 |
| 5.5 | Measuring the Variable Importance of Voxels Within a Stimulus Category | 39 |
| 5.6 | Measuring the Variable Importance of Voxels Shared Across Stimulus Category | 40 |
| 5.7 | Measuring the Variable Importance of Cliques of Voxels | 40 |
| 5.7.1 | Step 1: Joint Gini Importance | 41 |
| 5.7.2 | Step 2: Correction for Unconditioned Importance | 41 |
| 5.8 | Summary | 44 |
| 6 | Experiments and Results | 45 |
| 6.1 | Classical GLM Analysis | 46 |
| 6.2 | Classification Performance | 50 |
| 6.3 | Comparison of Variable Importance Measures | 52 |
| 6.3.1 | Importance of Voxels Within a Stimulus Category | 52 |
| 6.3.2 | Importance of Voxels Shared Across Stimulus Category | 56 |
| 6.4 | Visualization of Detected Informative Voxels | 58 |
| 6.4.1 | Voxels Informative Within a Stimulus Category | 59 |
| 6.4.2 | Voxels Informative Shared Across Stimulus Category | 62 |
| 6.5 | Identifying Informative Cliques of Voxles | 65 |
| 6.6 | Summary of Main Results | 68 |
| 7 | Discussion | 69 |
| 7.1 | Validity of the Classifier Models | 69 |
| 7.2 | Discussion on the Variable Importance Measures | 71 |
| 7.3 | RF Variable Importance Compared to GLM | 73 |
| 7.3.1 | Self Perception | 73 |

| | | |
|----------|---|------------|
| 7.3.2 | Empathic Perception | 74 |
| 7.3.3 | Shared Networks | 74 |
| 7.4 | Importance of Joint Voxels and the Importance of Long Range Connections . . | 76 |
| 7.5 | Summary | 77 |
| 8 | Conclusion and Future Work | 79 |
| 8.1 | Major Findings of this Thesis | 79 |
| 8.1.1 | Benefit of RF in fMRI Data Analysis | 79 |
| 8.1.2 | Insight into Empathic Mechanisms | 79 |
| 8.1.3 | Benefits Compared to State-of-the-Art Whole-Brain MVPA | 80 |
| 8.2 | Further Extensions to the Proposed Method | 80 |
| A | Alternative Data Preparation for Classification and Model Trainings | 81 |
| A.1 | Alternative Data Preparation Strategy | 81 |
| A.1.1 | Preparing Labels for Classification with a GLM Block Design | 81 |
| A.1.2 | Decoding Each Point in Time vs. Decoding Experimental Blocks | 83 |
| A.2 | Training Single Subject Models | 84 |
| B | Decoding Performance | 85 |
| B.1 | Single Subject Classification Accuracy | 85 |
| B.1.1 | Self-Model decoding Self-Session | 86 |
| B.1.2 | Self-Model decoding Other-Session | 88 |
| B.2 | Group Classification Accuracy | 90 |
| B.2.1 | Self-Model decoding Self-Session | 91 |
| B.2.2 | Self-Model decoding Other-Session | 97 |
| B.2.3 | Empathy-Model decoding Other-Session | 103 |
| B.2.4 | Other-Model decoding Self-Session | 106 |
| C | Schemaballs | 109 |
| D | List of Acronyms | 117 |
| E | List of Math Symbols | 119 |
| | Bibliography | 123 |

Introduction

This chapter describes the motivation to identify and decode the multivariate brain activity of emotional states in self- and empathic perception from a neuroscientific point of view. In the problem statement the technical difficulties to be solved are defined. The methodological approach is briefly outlined. The chapter closes with an explanation of the mathematical notation used in this thesis and a general description of the structure.

1.1 Motivation and Aim of the Work

In a recent Nature article “Mind reading” Smith [103] gives an overview on a growing field in neuroscience which tries to decode the brain’s activation pattern to read what people are seeing, hearing or feeling. One of the first papers demonstrating brain decoding was published in 2001 by Haxby et al. [51]. The popular term “mind reading”, is slightly misleading, since what the scientific community wants to do is to learn about the function of the brain itself, rather than reading someone’s mind [103].

Typically, the method for brain pattern decoding is referred to as MultiVariate Pattern Analysis (MVPA) in literature. According to Smith [103] decoders are generally built on individual brains, except for simple stimuli categories e.g. a binary choice if someone is looking at a picture A or B. Recent and frequently cited reviews on MVPA, like [87], [53] and [79], focus mainly on the typical procedure of decoding brain activity in single subjects at a high resolution.

Since functional Magnetic Resonance Imaging (fMRI) voxels typically serve as variables, and the points in time at which they are acquired as examples, the number of variables far exceeds the number of training examples [45]. This commonly deteriorates the classification rate, but the problem can be overcome by using preselected Region of Interest (ROI)s or restricting the analysis to local patterns [53, 79, 87].

While preselecting ROIs biases the result to previously expected regions the *searchlight* approach, introduced by Kriegeskorte et al. [69], can be applied to the whole brain without any prior knowledge. Searching distributed patterns outside predefined ROIs or small searchlight-volumes is not used [68]. Work on inter-subject multivariate pattern classification, done e.g. by Conroy et al. [17], focus on visual stimuli and use functional features for alignment rather than purely anatomical features as it is common in group alignment in the widely used analysis using the General Linear Model (GLM).

These points ask for a multivariate pattern identification and decoding method on a **whole brain** level, *without predefined ROIs*, at a *computational affordable expense* and which can be generalized to *group-level*.

The contribution of this thesis is a method to identify the cortical representation of the complete network involved in self-, as well as empathic experience of emotions induced by tactile and visual stimuli. For studying the neuronal basis of empathy it is further of interest to identify the neuronal patterns which are shared across self- and empathic perception [115]. Extensive work on the neuronal basis of empathy in pain has been done before [73]. While a local searchlight-approach using MVPA has been done in one study by Corradi et al. [18], in general MVPA has rarely been used in the field of empathy. In a recent review on the neural basis of empathy by Bernhardt and Singer [7] they mention “pattern classification” as a method of interest. Thus the data used for demonstration are a scientifically interesting application.

1.2 Problem Statement

The following problems arise with the development of such an analysis method as motivated by the “wish list” and application given above. The present thesis addresses these questions and offers solutions together with a discussion of related alternative approaches.

- **Capturing long-range interactions:** *Which classifier is robust in identifying the patterns of interest without limiting the features to specific ROIs? How to identify important features? How to measure the contribution of long-range connections? Which networks are shared between self and empathic emotions?*
- **Transferring global patterns across stimuli:** *Demonstration is performed on the self-, empathy-perception data-set. How to validate the output? How to visualize the data to be interpretable for psychological investigation and comparison with other experiments' results.*
- **Group-generalization:** *Due to subject to subject variability decoding cannot be performed on a high resolution level in group analysis [67]. Which common atlas can be used? Which resolution is appropriate?*

1.3 Methodological Approach

Experimental Setup

The data for devising these methods have been acquired in the context of a psychological experiment in emotional social judgment similar to experiments designed by Silani et al. [101] but with three emotional states namely *neutral*, *pleasant* and *unpleasant*. Data were recorded on 30 subjects with two trails (self and empathy) each. One trail contains 10 blocks of each stimulus. The data is provided by Professor Claus Lamm (Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna).

Group Alignment of Functional Data

It is known that the dimensionality reduction of cortical surface based alignment reduces the variability of activated regions, furthermore the cortical surface provides better features for inter-subject registration [110]. Thus in this work a cortical surface based approach will be used with the cost of losing information of deep brain regions.

Random Forest Classifier to Identify Informative fMRI Patterns

Langs et al. [77] showed that the Random Forest classifier [10] is stable in detecting distributed patterns. In the data for this work the identified pattern obtained by training on the self-session is used to identify the emphatic perceived emotions without any modifications on the classifier, this reveals the *cross condition* patterns that are shared between self- and empathic stimuli.

Importance Measures of Individual Voxels and Pairs of Voxles

Various importance measures in Random Forests exist [45]. These will be compared on the dataset. The measures are adapted first for defining the shared patterns between self- and empathic-perception and second to point out the contribution of distributed patterns, which should reveal the differences to the popularly used local searchlight-approach.

Validation of Classification and Comparison with other Methods

The decoding performance can either be evaluated by cross validation (leave one subject out) or by applying the self-session model to the empathy-session and vice versa. The findings also have to be validated against the classical GLM analysis. Further the results can be compared with results from related studies.

1.4 Mathematical Notation

For simplicity a coherent nomenclature is used in this thesis:

- **scalar values, constants, etc.** are noted with the default math font (e.g. a , A)
- **time-continuous functions** are explicitly noted as functions of time (e.g. $a(t)$)
- **vectors** are noted in lower case bold letters (e.g. \mathbf{a})
- **matrices** are noted in upper case bold letters (e.g. \mathbf{A})

A full list of all used mathematic symbols can be found in Appendix E

1.5 Structure of the Thesis

In Chapter 2 the usage of fMRI as a tool in neuroscience and signal origin of the measurements is explained. With that knowledge the *State of the Art* methods for fMRI-analysis can be outlined in Chapter 3. Limitations of the standard-methods are pointed out, which give rise to a new approach. Chapter 4 gives explanations to the neuroscientific research on empathy which is crucial for understanding the application of the proposed method. With the information given on the technical State of the Art and the necessary psychological background, the proposed method for detecting shared distributed patterns is described in detail in Chapter 5. Chapter 6 describes all conducted experiments and their results. They are discussed in detail in Chapter 7. Finally Chapter 8 summarizes the major findings and gives insight into future work.

Functional Magnetic Resonance Imaging as Tool for Neuroscience

fMRI is a method of detecting dynamic activity in the working human brain. To get an insight on the origin of the fMRI signals, the first section of this chapter focuses on the physical basis of Magnetic Resonance Imaging (MRI) and the physiological mechanism which enables the measurement of functional brain activity. The second section will give a brief overview on fMRI study design. This knowledge is necessary to understand how the measured brain activity can be linked to cognitive states and how it can be analyzed further.

2.1 Signal Origin in fMRI

The linkage between neuronal activity and the measured fMRI signals can be split in physical measurement process and the physiological connection between neuronal activity and the measurable local blood oxygenation.

2.1.1 Physical Principles of MRI

The fundamental principle of nuclear magnetic resonance was independently described by Bloch [8] and Purcell et al. [96] in 1946.

The human body is largely composed of water, each water molecule contains two hydrogen nuclei ^1H , which is a proton [14, 71]. A proton has a nonzero spin and thus a magnetic moment. Normally the nuclear moments are randomly orientated, but under a strong external magnetic field B_0 they weakly tend to align in the direction of this field [14]. Commonly the direction of B_0 is defined as the z -direction. The alignment causes a small magnetization M_0 which is in the order of $4 \cdot 10^{-9}$ times the applied external field and thus far too weak to be measured. [12]

For making this weak magnetization measurable the technique of nuclear magnetic resonance, described by the Bloch-equations, can be used. The key idea is to measure the weak magnetic moment while it oscillates in a plane perpendicular to the static field. For performing

this measurement the magnetic moment has to be tipped away from the z -direction. In the xy -plane the magnetic moment feels a torque proportional to B_0 , this causes a precession movement around the z -direction in a frequency proportional to B_0 . This precession movement produces a time-varying flux which can be measured in an induction coil antenna. [12] The tipping is done by a radio frequency pulse. The degree of tipping is defined as the *flip angle* and depends on the magnitude and duration of the radio frequency pulse. A 90° flip angle flips the magnetization from the z -direction in the xy -plane. After such a flip one can observe two relaxation processes: First the transversal magnetization in the xy -plane decays exponentially with a time constant T_2 , whereas the magnetization in z -direction recovers exponentially towards equilibrium with a time constant of T_1 . These values depend on the type of tissue e.g. at $B_0 = 3$ T the time constants for the gray matter in the human brain are approximately $T_1 = 1$ s and $T_2 = 0.1$ s [14]. These two relaxation processes are now explained in more detail:

Spin-lattice relaxation T_1 describes the longitudinal relaxation of the equilibrium magnetization M_0 in z -direction. This relaxation processes with energy exchange between lattice and spins. The recovery of the magnetic field in z -direction is given in Figure 2.1.

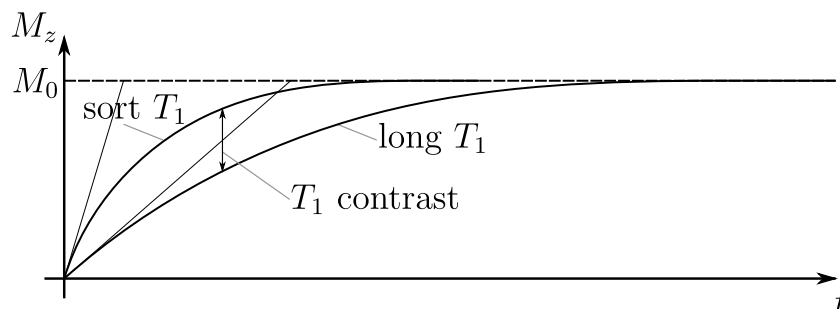


Figure 2.1: T_1 relaxation to equilibrium magnetization M_0 ; upper trace depicts the relaxation of a tissue with shorter T_1 than the tissue in the lower trace. This can be used as contrast in a T_1 -weighted image.

Spin-spin relaxation T_2 : This process is independent from the T_1 relaxation. In the T_2 -relaxation there is no energy exchange with the surrounding. It describes the exponential decay of the M_{xy} magnetization, which performs the precession movement around the z -axis. The precession frequency depends on the strength of the external magnetic field [12]. Random fluctuations of the local magnetic field cause a variation in precession frequencies and thus a dephasing of the spins. This results in a decay in the induced voltage in the induction coil. In an experiment one observes a faster decay T_2^* than the expected time constant T_2 explainable by random spin-spin relaxation effects [14]. This comes from local field inhomogeneities which cause an additional variance in precession frequency and a faster dephasing. In an experiment T_2 - and T_2^* - effects can be measured separately. This is based on the fact that non-random field inhomogeneities are continuous in time, and thus dephasing can be reversed by the following measurement procedure: Using a 180° flip pulse, the precession-movement flips in reverse direction and thus the spins will come in phase again. Being in phase again, the ensemble of spins

produce a measurable signal again. This signal is called *spin echo* [14]. Intuitively accessible the spin–echo can be explained with runners on a run–track: If runners with different speed start from the same position, they will fall out of “phase”. To a certain point in time all runners get a signal to turn around (180° – flip) and run back. The slower runners traveled a shorter distance, and thus also have shorter way back to the starting point. This means runners regardless their running speed will return to the starting point at exactly the same time (=spin echo). Figure 2.2 depicts a simple spin–echo experiment. The reduction in the echo’s amplitude is due to non reversible random speed differences, which is according to the “true” T_2 spin–spin relaxation process. In the shown example T_2^* –decay is much faster than the T_2 decay.

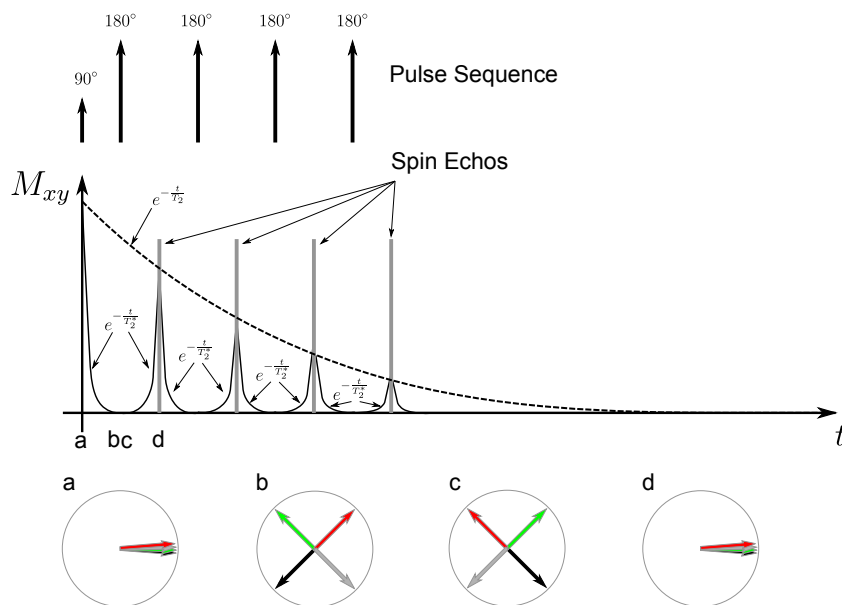


Figure 2.2: T_2 relaxation: a) After an initial 90° flip magnetization is flipped in xy direction. b) field inhomogeneities lead to fast dephasing of the spins. Destructive interference destroys the spin–ensemble’s xy –magnetization. c) A 180° spin flip reverses all phase–values. d) local static field inhomogeneities are the same as from “a” to “b” for each location, thus spins will run back in phase again. The spin–echo amplitudes decrease due to random dephasing, which is not reversible with the spin–echo sequence. The envelope function of the echo peak amplitudes describes the spin–spin relaxation decay with the time constant T_2 . The upper part of the Figure is adapted from [13].

This very short glance on the physical principles gives an overview on how the different decay functions of T_1 , T_2 and T_2^* can be measured. T_1 and T_2 vary across different types of tissues [9]. Thus these properties can be used to characterize the tissue and to produce an image contrast in anatomical images. Depending on the contrast used the images are called T_1 –weighted or T_2 –weighted image. Measuring these properties in specific locations within the brain is done by using different sequences of gradient fields and pulses as e.g. given in [58],

detailed explanations of this procedures are beyond the scope of this introduction to the physical signal origin in MRI. The usage of T_2^* for functional imaging is depicted in the following section.

2.1.2 The Blood Oxygenation Level Dependent Contrast and fMRI

To measure functional brain activity one can make usage of the Blood Oxygenation Level Dependent (BOLD)-effect. The pathway of this effect is depicted in Figure 2.3: The pathway shows that the neuronal activity is related to an increase in metabolism. A complex network of physiological effects are involved in changing blood-flow, blood-volume and blood-oxygenation. They aim to ensure nutrition and oxygen supply. Blood has different magnetic properties dependent on its oxygenation: oxygenated hemoglobin has diamagnetic- and deoxygenated hemoglobin has paramagnetic properties [86]. Beside other effects the change in blood oxygenation is the dominant effect [14, 86]. As depicted in Figure 2.3 the increased blood oxygenation increases the field uniformity and thus increases the T_2^* decay time, this is an measurable effect.

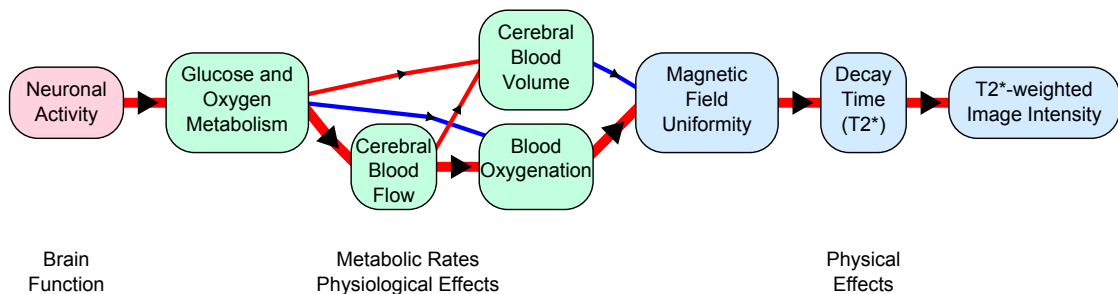


Figure 2.3: Pathway of the BOLD-effect. Red connections are positive correlated effects, blue connections are negative correlating effects. Bold lines indicate the dominant effects. Adapted from [86]

Figure 2.3 does not depict the temporal behavior of the effect and the physiological effects are complex and not known in all detail [2, 14]. But regardless of this fact one can measure the BOLD-signal response from a point like neuronal activity. This was e.g. done by Glover [46]. The resulting function is known as Hemodynamic Response Function (HRF) and describes the time dependent measurable¹ amplitude of the BOLD-effect. The function varies across subjects and areas in the brain [47]. A standard-HRF from a statistical toolbox for fMRI-analysis [114] is depicted in Figure 2.4. Mathematically the response to an arbitrary neuronal activity can be calculated by convolving the neuronal activity function with the HRF. This process can be interpreted as the temporal filtering of the neuronal activity with the HRF as filter-kernel.

2.1.3 Limitations of fMRI

The number of publications using fMRI show a tremendous increase since the beginning in 1992, this comes from the unprecedented possibility for noninvasively and safely measuring

¹measurable by MRI due to T_2^* -weighted imaging

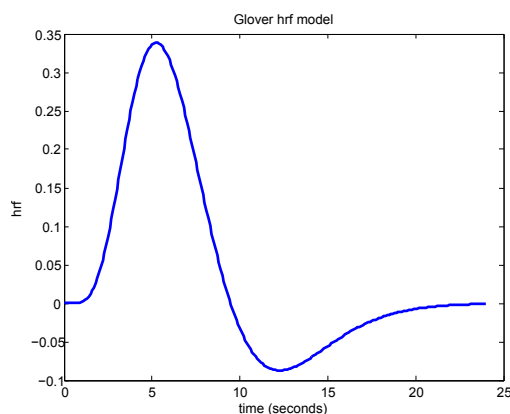


Figure 2.4: Glover HRF–model: measurable BOLD–effect from a single point in time neuronal activity. [46, 114]

brain activity [93]. Nevertheless this modality is limited by its spatial and temporal resolution which always have to be considered in any study:

fMRI is an indirect measurement of the neuronal activity, the quite complex BOLD–effect is between neuronal activity, which we want to observe, and the image contrast, which we can observe by a physical measurement. As it can be seen in Figure 2.4, the HRF has a Full Width Half Maximum (FWHM) of some seconds, conditionally on that the reduction of the repetition time T_R (the time in between two measurements) is of limited use. For whole brain studies a temporal resolution of 3 s is a typical value [14].

With a cortical neuronal density of approximately $40,000 \frac{\text{neurons}}{\text{mm}^3}$ [78] and a typical voxel size of $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ [14] one has to be aware that the smallest measured unit contains an ensemble of about one million neurons. A BOLD fMRI–measurement is a measurement of metabolism from a mass population of neurons.

The values depicted here are typical values found in current studies, today’s techniques are approaching 1 mm^3 spatial resolution and 1 s temporal resolution [14].

2.2 Principles on Study Design

The aim of the *Study Design* is to devise a setup which makes it possible to relate the measurements to the cognitive states one wants to observe. When designing a fMRI–study the limitations of fMRI as given in section 2.1.3 have to be considered. Besides one has to consider the psychological aspects: Amaro and Barker [2] describe what thousands of volunteers in fMRI–studies have experienced: “*Imagine yourself lying down inside a 60 cm wide, 120 cm long tube, exposed to 120dB acoustic noise (with mechanical vibration), trying not to move (or possibly restrained) while trying to perform a cognitive task.*” This clearly points out the complexity one has to face when designing a good experiment. One important aspect of each study design is to choose an appropriate way to present the stimuli, which induce the neuronal activity, to the subjects. The way of the presentation of the stimuli influences the neuronal perception and thus has to be

known for the appropriate analysis. The *paradigm* is defined as the construction and temporal organization of the cognitive task executed by the subject [2]. The data used for this thesis use a **block design** paradigm which is given described here:

The most basic block design is the so called “AB block” where two conditions are alternating. One cycle contains two epochs of each condition. Each condition has a length of several seconds. E.g. showing a certain picture to the person in the scanner for 10 s followed by a consecutive rest block reveals which areas in the brain are involved/activated by perceiving the image compared to the rest condition. The block–design has a relatively large change in the BOLD–signal compared to baseline and thus a high statistical power [2]. The block design is the first one used in neuroimaging and is still commonly used today [20]. In [20] several further aspects in the exact design are discussed: Stimulus or task should be as continuous as possible. Blocks altering every 16 to 20 s are maximally powerful and robust. When using high–pass filter to remove noise from the measurement one has to carefully choose the block design frequency and filtering frequency.

2.3 Summary

Neuronal activity causes a change in the metabolism. This metabolic process causes a change in magnetic properties. By measuring the T_2^* –decay, which is linked to field inhomogeneities by physical mechanisms, the neuroscientist has a tool for measuring neuronal processes in the living human brain. To link the indirect measurements to the neuronal processes the study design has to be known. For this work a block design is used. Many other aspects on study design exist, see [2] for an introductory review on this topic. The given information on paradigm used is sufficient for the further understanding of the fMRI analysis methods described in this thesis.

Basics in Functional Magnetic Resonance Imaging Data Analysis

In this chapter the established standard method of fMRI data analysis, the GLM, is explained in detail. Haxby et al. [51] introduced a way of analyzing fMRI data in 2001. This method subsequently came to be called MVPA [52]. An introduction to MVPA is given in Section 3.2. With this knowledge the terms *activation*- and *information-based* analysis are explained in this chapter. A big variety of classifiers is used for MVPA [77, 80, 90], in this work the Random Forest (RF)-classifier is used for the multivariate pattern detection. This classifier is explained in detail in Section 3.4, special emphasis is given on pointing out the advantages of RF for fMRI data analysis.

3.1 The General Linear Model

The GLM is a widely used standard method of fMRI data analysis, it is currently the most popular statistical approach [94, 104] and around with fMRI-data analysis since its early days [41]. The measurable BOLD-response of a neuronal activity is given by the HRF. For each observed voxel a BOLD-response $z(t)$ can be measured. Each neuron in the voxel can be involved in different neuronal activities. In the GLM it is assumed that individual neuronal activities sum up to the measured response in a linear way. Therefore the observed signal can be written as:

$$z(t) = m_1(t)\beta_1 + m_2(t)\beta_2 + \dots + m_q(t)\beta_q + e(t) \quad (3.1)$$

where $m_1(t) \dots m_q(t)$ are q predictor-functions of the model, with the corresponding weighting factors $\beta_1 \dots \beta_q$ and an error term $e(t)$ (noise). A model predictor $m(t)$ typically reflects the predicted BOLD-response of a neuronal activity. Thus it can be obtained by convolution of the neuronal activity with the HRF. It is also possible to include other additional predictor models which do not represent neuronal activity [98, 109]. The standard assumption of $e(t)$ is being normally distributed with zero mean [93]. If $z(t)$ is measured on n_{obs} discrete points,

in time it can be written as the vector \mathbf{z} of length n_{obs} . This notation can also be applied to the predictors $\mathbf{m}_1 \cdots \mathbf{m}_q$. We define $\boldsymbol{\beta}$ as a weighting vector $(\beta_1 \ \beta_2 \ \dots \ \beta_q)^T$ and $\mathbf{M} = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q)$ as a $n_{\text{obs}} \times q$ model predictor matrix, or design matrix¹, equation (3.1) can be written in matrix notation as:

$$\mathbf{z} = \mathbf{M}\boldsymbol{\beta} + \mathbf{e} \quad (3.2)$$

If the measurement is modeled as a GLM, we are trying to find an estimate for $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1 \ \hat{\beta}_2 \ \dots \ \hat{\beta}_q)^T$ so that the estimated BOLD–response $\hat{\mathbf{z}} = \mathbf{M}\hat{\boldsymbol{\beta}}$ fits best the measurement \mathbf{z} . To find $\hat{\boldsymbol{\beta}}$ the method of *least squares*, which minimizes the squares of the difference $\mathbf{e} = \mathbf{z} - \hat{\mathbf{z}}$, is used [93]. It can be shown that $\hat{\boldsymbol{\beta}}$ for minimizing the squares of differences $\mathbf{e}^T \mathbf{e}$ can be found with:

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{z} \quad (3.3)$$

To identify activated voxels from the GLM model one can use a t-test. For this purpose a null hypothesis has to be performed. It is formulated in the form of:

$$\text{H}_0 : \mathbf{c}^T \boldsymbol{\beta} = 0 \quad (3.4)$$

With the alternative hypothesis:

$$\text{H}_a : \mathbf{c}^T \boldsymbol{\beta} \neq 0 \quad (3.5)$$

In Equation (3.4) and (3.5) $\mathbf{c} = (c_1 \ c_2 \ \dots \ c_q)$ is the so called contrast vector to test for the correct null hypothesis according to Equation (3.4). For illustration two exemplary contrast vectors and the corresponding hypothesis are given:

1. $\mathbf{c}^T = (1 \ 0 \ 0)$ tests for $\text{H}_0 : \beta_1 = 0$ *Voxel shows no activation of the predicted neuronal response \mathbf{m}_1 .*
2. $\mathbf{c}^T = (1 \ 0 \ -1)$ tests for $\text{H}_0 : \beta_1 - \beta_3 = 0$ ($\text{H}_0 : \beta_1 = \beta_3$) *Voxel is equally activated by the predicted neuronal response \mathbf{m}_1 and neuronal response \mathbf{m}_3 .*

With a given contrast vector the t–statistic can be calculated [4]:

$$t = \frac{\mathbf{c}^T \hat{\boldsymbol{\beta}}}{\sqrt{\text{var}(\mathbf{e}) \mathbf{c}^T (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{c}}} \quad (3.6)$$

This t–value can then be compared to a specified significance level. One has to keep in mind if e.g. the test is $\text{H}_0 : \beta_1 = 0$ and the significance level is set to $p = 0.01 = 1\%$, just by chance for 50 out of 5000 voxels the null hypothesis has to be rejected and will falsely be considered as “activated”. This issue is known as multiple testing problem. Solutions to this problem can be found in many textbooks e.g. [93]

¹The convention in most fMRI textbooks is to denote the design matrix as \mathbf{X} . This is in conflict with most machine learning literature where the measurements/observations are denote with \mathbf{X} . Since main parts of this thesis deal with machine learning this nomenclature is used and the design matrix is renamed as \mathbf{M} ; furthermore in fMRI textbooks \mathbf{y} is used for the measurement instead of \mathbf{z} . In this thesis \mathbf{y} is used as response vector or label vector for the data \mathbf{X} , as usual in machine learning.

Detailed information on the GLM can be found in textbooks on fMRI data analysis see [4,13,58,93]. Poline and Brett [94] give an excellent review on the GLM and a critical reflection on the usage of this tool.

3.2 Multivariate Pattern Analysis

The GLM estimates how much each single voxel contributes to the expected neuronal activity. This completely ignores the interaction between voxels. Intuitively this seems not to be useful since neurons interact with each other to perform a specific task. Using multivariate statistics and methods from machine learning this multidimensional interaction can be analyzed. This type of analyzing fMRI-data has been used for more than 10 years when Haxby et al. [51] demonstrated a method for investigating the functional architecture for face and object recognition in ventral temporal cortex. Later this method came to be called MVPA and a diverse set of methods to analyze patterns of fMRI activity came up [52]. It has been proven that MVPA is more sensitive and informative for fMRI-data analysis than the GLM [52].

The following overview on MVPA is structured in five steps of pattern information analysis as given in a review by Mur et al. [85]:

1. Preprocessing and splitting the data in training and test data
2. Estimate the activity pattern
3. Select voxels for classifier training
4. Training of the classifier
5. Testing of the classifier

To gather the variety of the methods each of the steps is discussed in detail in the following.

3.2.1 Preparing Data for MVPA

This step aims to prepare data for training and testing the classifier. Cross validation is an effective way of using the data [85] e.g. train the model on all but one subjects and test the model on the left out subject. This can be repeated until all subjects are used once for testing. Preprocessing should, if possible, be done separately on training- and test-set to avoid dependencies. Preprocessing includes slice-time-correction, motion correction and trend removal. Spatial smoothing is in contrast to preprocessing for the GLM not done before MVPA. [60, 69, 79, 85] Spatial smoothing is discussed in literature quite controversial e.g. [89] recommends spatial smoothing. Kriegeskorte and Bandettini [67] give an overview on spatial resolutions in fMRI analysis. They report group alignment in Talairach space can be mismatched many millimeters between subjects. A cortex based alignment can provide a more precise alignment, where the authors give a value of approximately 2 to 7mm mismatch in a cortical alignment between subjects. To preserve fine grained patterns and because of doubts about the functional correspondence across subject, MVPA is often done on a single subject level [85].

Volume based as well as surface based approaches are used in conjunction with MVPA [15, 88]. These approaches require different preprocessing.

3.2.2 Estimating the Activity Pattern

Various methods are found in literature to gather the training data. The simplest way is to stay close to the raw data. This can be done for slow block designs where the BOLD–response of different conditions do not temporally overlap [85]. This is e.g. done in [77, 95].

Estimating the activation of each voxel is also possible in the univariate GLM–fashion. This method can be found in literature frequently and is particularly useful in rapid designs where BOLD–response of different conditions overlap in time [85]. In this preprocessing each experimental event or block is modeled with its own predictor–function $m(t)$ in the GLM. This approach is e.g. used in [70]. Using this method one gets a β –value for each single block for each voxel. The experimental condition of each block serves as training label for the β –value pattern across all voxels. This means the number of observations for training equals the number of blocks in the experiment.

3.2.3 Selecting Voxels and Searchlight MVPA

It is known from classical machine–learning that including too many (noisy) features in a classifier reduces the generalisability of the model which is known as over–fitting [55]. This is particularly a problem in fMRI since conventionally the number of voxels in the brain far exceeds the number of observations or experimental blocks². This deteriorates the classification performance and thus a dimensionality reduction has to be made [53]. Various methods exist:

- Classical methods of dimensionality reduction as Principal Component Analysis (PCA) as e.g. demonstrated in [82] can be used.
- Down-Sampling was used for lie detection by Davatzikos et al. [19], there they used averages of relatively large boxes (16mm \times 16mm \times 16mm) as features for the classifier.
- By using the training data or other data independent from test set, informative voxels for discriminating the conditions of interest can be identified [81, 85]
- Any univariate statistic can be used to identify voxels individually informative to a specific condition. Here the main concern is that even with a liberal threshold voxels are discarded which are uninformative alone but provide information in combination with others. [87]
- Also anatomical data or other knowledge to select a ROI in which a multivariate pattern is searched can be used [79, 85, 87]. This restricts the analysis to a volume which relies on prior assumptions where an informative activity pattern is expected. A typical example is to restrict the area of interest to the visual cortex when interested in the visual perception [61].
- A widely used method was introduced by Kriegeskorte et al. [69], it is referred to as *searchlight*–approach. In this approach a small control volume is shifted across the brain. These e.g. 15 mm \times 15 mm \times 15 mm volumes only contain a small number of 125 voxels when using a 3 mm scanning resolution. The activation pattern of the volume in all locations is compared across all experimental conditions to investigate whether the local activation pattern is informative for decoding the experimental condition. The GLM

²In standard fMRI the voxel–size is in the order of 3mm \times 3mm \times 3mm when assuming a brain volume of 1,200 cm³ more than 44,000 voxels could be included in the model. Using a block–length of 20 s more than 200 hours of scanning time would be necessary to obtain as much experimental block as voxels

can be seen as a special case of the MVPA which is only sensitive in the case where all voxels in a certain location change in the *same direction* for each experimental condition [85]. The searchlight method is illustrated in Figure 3.1. Beside the popularity of this approach Etzel et al. [30] outline pitfalls with the searchlight–approach which can lead to interpretational errors. Furthermore this method is not able to detect multivariate relations across distant voxels, it just looks at the local pattern.

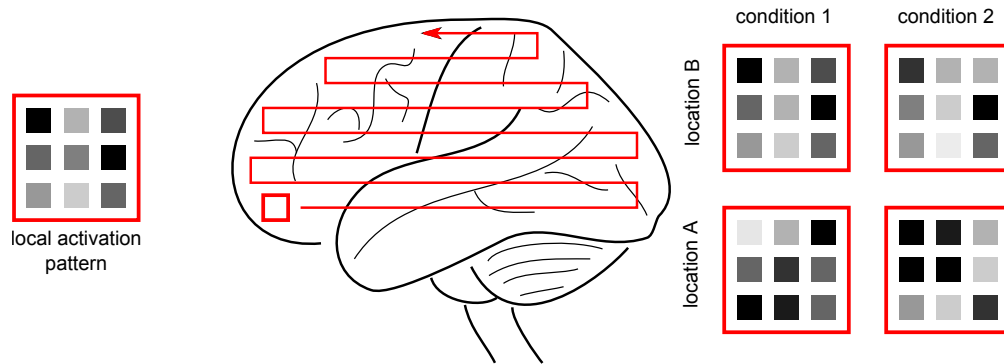


Figure 3.1: Simplified procedure of the searchlight method: A small control volume is shifted across the whole brain. In each location the patterns are compared across all experimental conditions. In the shown example the responses in location “B” are similar for both conditions, this location does not carry information to distinguish the two experimental conditions 1 and 2. Location “A” shows different response patterns for both experimental conditions 1 vs. condition 2, thus the pattern is informative in respect to identify the experimental condition. The average mean across these control volumes would be nearly the same across both conditions for each location. Thus there would not be a detectable contrast in the GLM between both conditions in the location “A” and “B”.

3.2.4 Classifiers in MVPA

At step 4 of MVPA a classifier is trained on the data provided by the previous steps. A big variety of classifiers is used for MVPA; overviews and comparison can be found in [77, 80, 90, 92]. Most interestingly for this work is the RF classifier which is proven to be useful for decoding fMRI and detecting distributed patterns [44, 45, 77]. As this classifier serves as the basis for the method developed, this classifier is discussed in Section 3.4.

3.2.5 Testing and Validation of the Classifier–Model

This last step tests if the model is valid and usable on data the classifier has never seen before. If the model generalizes well it will successfully decode the conditions in the test data set. If the model predominantly “explains” the noise in the training–set, it will fail to decode the unknown data where the random noise will behave in a different way. By performing a statistical test with the null hypothesis H_0 : *The classifier performs at chance level*, the model can be validated [85].

This test can be assessed on a permutation based approach [30]. For this test the labels are permuted, which corrupts the dependency between labels and the measurement. The classifier is then trained and evaluated in the same way as on the unpermuted data. To obtain a minimal p -value of $p = 0.001$ 1000 permutations have to be made for testing [29]. This can cause a significant amount of computation time as reported in [59].

For further detailed information on MVPA see the reviews [52,53,79,85,87,90]. For a very recent cursory overview on brain decoding and “mind reading” see [103]. [49] is an excellent textbook on statistical learning.

3.3 Activation – vs. Information – Based Analysis

The previous two sections gave an overview on two frequently used analysis tools. It has to be discussed when to use which type of analysis. In literature a distinction between activation-based analysis (this is mainly done with the GLM) and information-based analysis (this is done with MVPA) can be found [68]. Kriegeskorte and Bandettini [68] state that these methods are complementary tools. To point out the exact scientific question by these tools the terms activation- and information-based fMRI data analysis have to be defined:

In [67] *activation* is defined as specially averaged activity of a functional region, with distinguishing it from *activity* as the unsmoothed local activity from one voxel. Thus activation-based analysis reveals the activation related to a brain process. If the target is to identify activation of whole brain regions, an activation-based approach is very successful [67]. Information-based analysis determines if a region’s activity pattern carries information about the experimental condition.

3.3.1 What Questions are Answered by the Different Analysis Methods?

Kriegeskorte, Bandettini and Kleinschmidt [67] [65] and [68] discuss in partly controversial opinions on activation and information based analysis. Kriegeskorte and Bandettini [68] try to explain and clarify the goals and problems of different analysis methods, see Table 3.1.

3.3.2 Comparison of Univariate Analysis and MVPA

Jimura and Poldrack [59] directly compare a univariate analysis and a searchlight-MVPA on data of the same experiment. A spacial smoothing with a 5mm FWHM Gaussian kernel for the GLM-analysis and no spatial smoothing is used in the MVPA, the searchlight volume of a 3 voxel radius is used. The article reveals the following issue: MVPA and univariate analysis exhibit substantial differences in the identified regions. The authors state that the univariate analysis may reflect core processes whereas MVPA reflects subprocesses. MVPA shows increased sensitivity in frontal and parietal regions, in contrast the GLM is more sensitive in subcortical areas. This can be either explained with the different functional organisation of cortical and subcortical regions, also the partial inclusion of ventricle-voxels in the searchlight volume could overwhelm the MVPA with noise. They conclude that both approaches should be used and state that there is a need to better integrate these methods with each other.

| Area | Goal | Method | Problems | Promise |
|-------------|---------------------------------------|---|---|--|
| whole brain | finding activated regions | univariate on smoothed data | insensitive to fine-grained pattern | localization of extended homogeneous activation |
| whole brain | finding informative regions | MVPA searchlight | less sensitive to extended activation | localization of regions with informative pattern |
| ROI | characterization of ROI activation | univariate on ROI average | insensitive to fine-grained pattern | characterization of overall involvement |
| ROI | visualizing fine-grained ROI activity | univariate on unsmoothed data | fMRI accurate image of neuronal pattern? sensitivity taxing multiple comparison: ROI → OK; whole brain → problem with high resolution | visualizing of fine-grained pattern characterization, columnar spatial organization |
| ROI | characterization of ROI information | multivariate on unsmoothed data: e.g. classification analysis | sensitivity taxing curse of dimensionality: ROI → OK; whole brain → problem with high resolution | characterization of the overall information content |

Table 3.1: Overview on fMRI analysis methods; adapted from [68]

3.3.3 Another Point of View: Forward- and Backward-Models

Another way to think about the different objectives in the analysis are forward- and backward models. A *forward model* explains how the measured data are generated from the neuronal source [50]. This is for example done by the GLM which, as described above, models the measured data as a linear combination of neuronal activities. A model which extracts the neuronal information / the mental state from the measured data is called *reverse model* [50]. This corresponds to the MVPA as described above: the classifier reverses the generation process and determines the mental state from the measured data.

Haufe et al. [50] state that only forward models are interpretable with respect to a neural process. Further they state that backward models are accurate in decoding, and that there is no reason to believe that the decoding models should be interpretable. Thus they propose that a backward model first has to be transformed into a forward model to be interpretable and show how the weight vectors of a linear multivariate model can be transformed to an interpretable forward model.

3.4 Random Forest Classifier

The RF-classifier was chosen as the basis machine learning algorithm in this thesis, therefore it is described here in detail.

3.4.1 Functioning of the RF

Methods called *bootstrap aggregation* or *bagging* can be used to improve stability and accuracy of an estimated prediction function [49]. The *Random Forest* algorithm introduced by Breiman [10] is a substantial modification of bagging. In the RF-classifier, an ensemble of weak learners is created. Decision trees serve as the elementary base learners.

Individual Decision Trees

The functionality of a decision tree is illustrated in a simple example in Figure 3.2: The popular Fisher iris data [37] is used for demonstration. Three classes have to be identified by two variables. To find the decision for the correct class an observation belongs to, this observation is passed down the tree. Each *node* performs a binary *split* by evaluation if a certain feature exceeds a defined value. For each branch further splits are allowed. Nodes with no further splits are called *leaf nodes*. In the geometry of the feature space each split is represented by a partition of the feature space along a hyperplane perpendicular to the split feature. Each leaf node corresponds to one partition of the split feature space and is assigned to one final decision for a class. A 2D representation is given in Figure 3.2, hyperplanes degenerate to straight lines in this example.

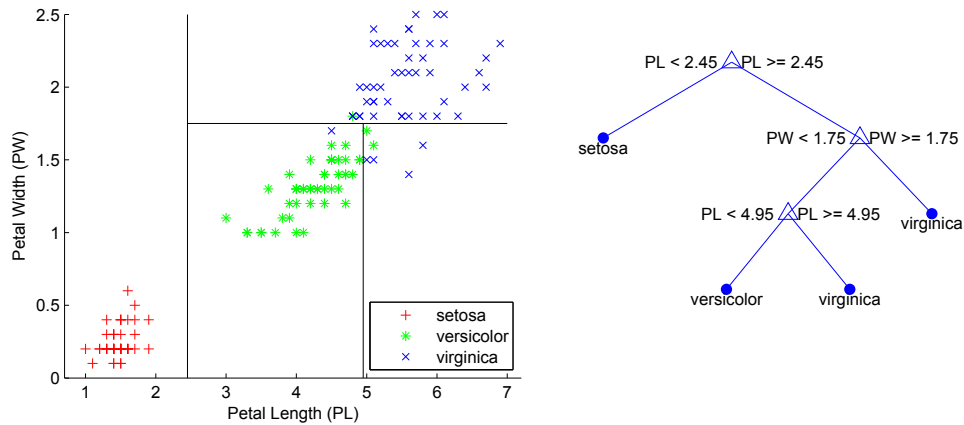


Figure 3.2: Example for a decision tree on the Fisher iris data [37]. Scatter plot shows the split-values of the tree which divide the feature space in maximally pure subdivisions. Each of the four partitions is assigned to one of the leaf nodes in the tree.

input : matrix \mathbf{X}
input : target vector \mathbf{y} containing a label for each observation
output: an ensemble of trees (Random Forest)

```

1 for  $i \leftarrow 1$  to  $n_{trees}$  do
2   Draw a bootstrap sample (random subset with replacement)  $\mathbf{X}^*$  of the size  $n_{obs}^*$ ;
3   recursively repeat
4     Randomly select  $m$  out of  $V$  features;
5     Find the feature  $v^*$  and split-value  $\eta^*$  that best splits the sample  $\mathbf{X}^*$  in pure
     classes;
6     Recursively proceed with both children;
7   until The tree is fully grown;
8 end

```

Algorithm 3.1: RF–algorithm: For each tree in the forest a classification tree is build on a random subset of all observations. At building the tree only a random selection of all features is used to find the best split.

Learning Ensembles of Decision Trees

The RF is built by training individual trees on randomly chosen subsamples from the whole dataset. For decorrelating the trees also the features are randomly subsampled for each node. For classification a majority-vote of all trees is taken as final decision. For further explanation the data matrix is denoted as \mathbf{X} . It has the size $V \times n_{obs}$, where V is the number of features and n_{obs} . In the example of fMRI data analysis this matrix could e.g. be equivalent to the BOLD measurements in each voxel v . For classifier training a target vector \mathbf{y} of length n_{obs} exists, which contains a label for each observation. The algorithm is outlined in Algorithm 3.1.

Mathematical Explanations of the Parameters

Each individual tree serves as a weak learner and delivers a random variable with variance σ^2 . An average over all n_{trees} for the case of independent identically distributed (i.i.d.) variables is $\frac{\sigma^2}{n_{trees}}$. As shown in [49] the average variance for the identically distributed (i.d.) case is

$$\sigma_{av}^2 = \rho\sigma^2 + (1 - \rho)\frac{\sigma^2}{n_{trees}} \quad (3.7)$$

where ρ is the pairwise correlation. The second term in Equation (3.7) does not contribute to the average variance for high n_{trees} . The remaining first term shows that the average variance can only be reduced for decorrelated trees. The reduction of correlation is done by the selection of m out of V variables, where reducing m reduces the correlation. m can be as low as 1. [49] Advised default value for m is \sqrt{V} for classification [49].

The Gini Split Criteria

The Algorithm 3.1 contains the task of finding the “best” split–variable and split–feature. Different split–criteria for tree classifiers could be used. In RF the *decrease in Gini Impurity* is used.

As the *Gini impurity* and *decrease in impurity* are not only used in growing the tree but also extensively used in analyzing the trees these terms are explained here in detail:

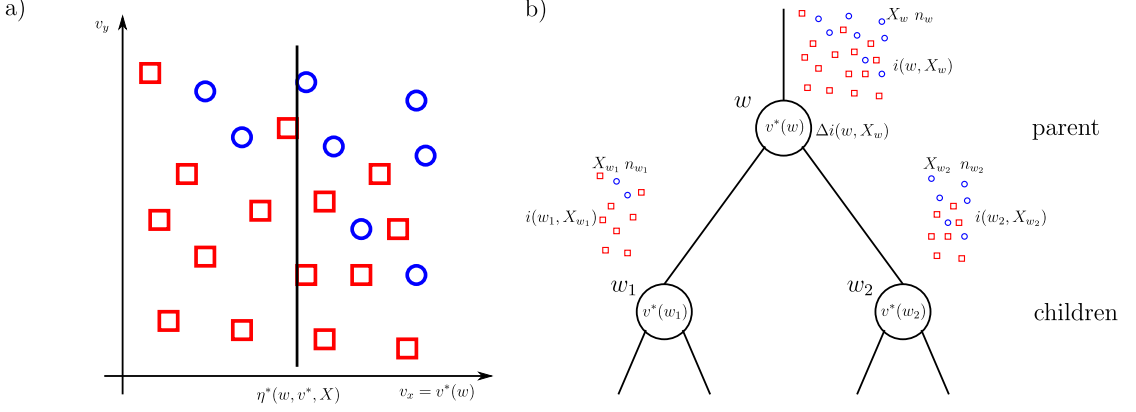


Figure 3.3: Exemplary illustration for a node w in a tree for the definition of the Gini-split criteria: a) Data \mathbf{X}_w at node w with the random subsample of features v_x and v_y . $v^*(w)$ is the split variable and $\eta^*(w, v^*, \mathbf{X}_w)$ the split-value for node w . b) Tree structure: parent node w splits the data into two children subsets \mathbf{X}_{w_1} and \mathbf{X}_{w_2} assigned to the children nodes w_1 and w_2 . i is the Gini impurity and $\Delta i(w, \mathbf{X}_w)$ is the decrease in impurity at parent node w .

Figure 3.3 shows a hypothetical example for any node w in a tree. We now want to find the best split-variable and split-value for node w . For simplicity it is assumed that the random subsample of features just contains the two features (namely v_x and v_y). The data reaching node w are denoted as \mathbf{X}_w . In addition a label l is known for each element of \mathbf{X}_w whereas L is the number of different labels.

The probability of a datapoint in \mathbf{X}_w belonging to a specific label l is defined as p_w^l . The Gini impurity of the data \mathbf{X}_w is given by:

$$i(\mathbf{X}_w) = \sum_{\text{all } l} p_w^l (1 - p_w^l) \quad (3.8)$$

The Gini impurity $i(\mathbf{X}_w)$ measures the probability that two independent draws from the data \mathbf{X}_w are from different classes. The Gini impurity thus can range from 0, which means pure data all with the same labels where it is impossible to draw data from different labels, to $1 - \frac{1}{L}$ for equal amount of data from each label.

Each possible split of the data \mathbf{X}_w produces two children \mathbf{X}_{w_1} and \mathbf{X}_{w_2} . This decreases the impurity by $\Delta i(w)$ which can be assessed by:

$$\Delta i(w) = i(\mathbf{X}_w) - \frac{n_{w_1}}{n_w} i(\mathbf{X}_{w_1}) - \frac{n_{w_2}}{n_w} i(\mathbf{X}_{w_2}) \quad (3.9)$$

n_w is the total number of examples at node w , n_{w_1} number of examples at node w_1 and n_{w_2} is total number of examples at node w_2 . The fractions $\frac{n_{w_1}}{n_w}$ and $\frac{n_{w_2}}{n_w}$ are necessary to normalize for

the amount of data in the children–nodes; the child’s impurity has to be weighted stronger if the child node contains more data. $\Delta i(w)$ never can take negative values, each split decreases the impurity or keeps it constant. With the Gini split–criteria the feature v^* with the corresponding split–value η^* is chosen among all possible variables and values, where the decrease in impurity $\Delta i(w, \mathbf{X}_w)$ by the split in node w is maximal.

RF Implementations

Implementations of the algorithm exist in many software packages. The original code by Breiman and Cutler is available in FORTRAN [11]. A MATLAB port by Jaiantilal [57] was used in this thesis.

3.4.2 Variable Importance Measures

Since the RF–classifier is applied on the whole dataset, which implies that it uses all features distributed across the entire brain. If it is proven that the classifier successfully decodes the mental state it is evident that the classifier “knows” which measured neuronal activity are related to specific mental states of the brain. The general approach to identify these features (=measurements of activity in certain spots) is to use variable or feature importance measures, where the variable importance tells us how important a certain feature is for decoding the correct mental state. It is shown by Langs et al. [77] that such an importance measure can be successfully used as functional imaging contrast in fMRI.

Gini Importance of Subsamples chosen by the RF–Algorithm

In each split the best split–variable v^* and a best split–value η^* is selected which provides the highest decrease in Gini–Impurity $\Delta i(w, \mathbf{X}_w)$. The higher this value, the more important is this split to purify the data. For an over–all objective measure of the importance it is also necessary to take the amount of data n_w split by this node into account. It is more useful if a split with a high $\Delta i(w, \mathbf{X}_w)$ is performed on a big dataset, thus we define:

$$\Delta I(w) = \Delta i(w, \mathbf{X}_w) \cdot n_w \quad (3.10)$$

as the weighted decrease in impurity. With Equation (3.9) we get:

$$\Delta I(w) = n_w \cdot i(\mathbf{X}_w) - n_{w_1} \cdot i(\mathbf{X}_{w_1}) - n_{w_2} \cdot i(\mathbf{X}_{w_2}) \quad (3.11)$$

If this decrease in weighted impurity is summed up separately for each variable v across all trees in the forest one gets the *Gini importance* of variable v

$$I_G(v) = \sum_{\text{all trees}} \left(\sum_{w:v^*(w)=v} \Delta I(w) \right) \quad (3.12)$$

Since every RF–implementation has to determine the best split–feature and split–value, the Gini–importance does not cause extra computational effort. It is available in RF–implementations

[107]. It has to be mentioned that this increase in Gini-impurity just relies on the drawn random subsamples \mathbf{X}^* of observations and random sub selection of m features in each node. It just accounts for the amount of decrease in impurity among all classes and thus this measure does not provide class specific importances.

In an experiment with simulation data Strobl et al. [107] demonstrate that the Gini importance is severely biased. This importance measure shows strong preference for variables which have more categories than other variables and also prefers continuous variables. Also in [105] a bias in Gini-based importance measures is reported. Strobl et al. [107] state that this derives from the fact that the RF-algorithm selects the best split according the Gini-split criterion, obviously a continuous variable provides more possible cut-points. Strobl et al. [107] conclude from that fact that it is just by chance more likely that such a variable is selected as split variable. Since in fMRI-data all variables are continuous variables and approximately i.d. this issue does not seem to be a problem in the application on fMRI-data.

Permutation Importance

A more advanced importance measure than the Gini importance is the *permutation importance* [107].

Since during training for each tree only a subsample of the dataset is taken, there is a set of datapoints not included in the bootstrap sample to construct the tree b . These samples are called Out Of Bag (OOB)-samples. These samples can be used to estimate the prediction strength during training since for each tree the OOB-samples, which were not included in the training, can be passed down the current tree [49]. To assess the importance of feature v in the OOB-sample of tree b the values of feature v are randomly permuted. This breaks the dependency between the values of feature v and the labels to be decoded. Thus a loss in prediction accuracy between original and permuted data is as measure for the importance of that feature. Mathematically the Permutation Importance for feature p can be notated as following:

$$I_P(v) = \frac{1}{n_{\text{trees}}} \sum_{\text{all trees}} \left(\text{acc}(OOB_b) - \text{acc}(\widetilde{OOB}_b^v) \right) \quad (3.13)$$

Where OOB_b are the OOB-samples of tree b and \widetilde{OOB}_b^v are the OOB-samples of tree b with the values of feature v permuted. The function $\text{acc}(X)$ calculates the prediction accuracy of the data X . The $\text{acc}(X)$ -function can be modified to calculate a class specific – or an over-all accuracy, this calculates the permutation importance for a specific class or an unspecific over-all importance as the Gini importance.

Different studies were made to assess the differences between the different importance measures: e.g. in [3] standard importance measures (the Gini importance and permutation importance on the OOB-samples) are compared. They report similar performance but advise to use the Gini importance for small sample size because the OOB-sample is unreliable small.

3.4.3 Special Issues for RF Application on fMRI-Data

The following experimental results using a RF-classifier for MVPA of fMRI-data can be found in literature:

Langs et al. [77] demonstrated that an RF classifier can be used to detect distributed multivariate activation patterns. The RF classifier is particularly useful since the independent random sub-sampling of features for building each node allows highly correlated but predictive features to be included into the classifier. This ensures to detect all informative features rather than a subset which is sufficient for an accurate decoding. Compared to a “traditional” method such as Support Vector Machine (SVM) in conjunction with fMRI it is shown that the Gini Importance of a RF classifier delivers a more stable activation pattern. Whereas the classification performance is similar. Thus it is advised to use RF if voxel identification is the primary aim.

Genuer et al. [44] perform a similar experiment and use an RF-classifier for decoding fMRI-data. They also compare RF-classifier to SVM as reference method. As in [77] Genuer et al. [44] observe a similar result: The SVM classification rate is only slightly below the RF classification rate but the sparsity of feature selection in the RF reveals a better interpretable pattern of jointly informative regions.

With the information given above it can be summarized why the RF-algorithm is useful in fMRI-data analysis: Bagging is based on the combination of weak learners, each of them is assumed to be unbiased but unstable in giving the right prediction. Taking back in mind Equation (3.7) which explains the idea behind bagging, two conclusions can be made. To achieve an overall stable result first n_{tree} has to be chosen, enough to eliminate the term $(1 - \rho) \frac{\sigma^2}{n_{\text{trees}}}$. More intuitively the number of trees smoothens the hard cuts by a single tree decision boundary [106]. The number of trees increases the computational expense for training the classifier, but the training can easily be parallelized. To reduce the term $\rho\sigma^2$ the individual trees have to be decorrelated. This is done by the random subsampling of the features for each split, this allows “weaker” predictor variables, otherwise outplayed by their highly predictive competitors, to enter the ensemble [106]. In conjunction with variable importance measures this exactly corresponds to the finding of Langs et al. [77] where it is reported that RF stable detects all involved voxels rather than just the most predictive ones.

3.5 Summary and Limitations of the State-of-the-Art fMRI-Analysis Methods

In standard fMRI-analysis whole-brain analysis is either done with the GLM or in a searchlight MVPA. The first method belongs to the activation based methods, which yield to identify extended regions of homogenous activation. The second method is an information based approach, which in contrast to the GLM, searches for local informative coding patterns.

By capturing the informational content of multiple voxels in the MVPA, higher sensibility compared to the GLM is reported [52]. In the classical whole brain MVPA the informative patterns are extracted from local searchlight-volumes. This standard procedure is a simple ap-

proach to overcome the curse of dimensionality, but beforehand prohibits to make use of the informative patterns, which spatially exceed the local volume.

This limitation to local patterns motivates to develop a method which is capable of detecting distributed patterns. As shown in [77] the RF-classifier can serve as the basis classifier for such a MVPA. The available RF-variable importance measures, namely the the Gini importance and the permutation importance, will be used and extended in the proposed method.

The Empathic Brain

The title of this chapter is taken from the book “The Empathic Brain” [62,64] by Christian Keyers. In his book he gives an insight into his research on the neuronal basis of empathy. Based on this book and further literature research this chapter gives an overview on the underlying neurological processes of empathic processes in the brain for a better understanding of the question of interest from a psychological point of view. This information will be used to verify whether the results of the proposed method are in agreement with findings of recent studies. For a better understanding of the cited studies and a better understanding of the results in this thesis the last section in this chapter gives an overview on the most important areas of the brain that are involved in empathy.

4.1 Definition and Factors of Empathy

The human ability to share affective states of others, is a crucial component in emotional experience and social interaction, which allows us to understand other’s feelings, motivations and actions [7]. Singer and Lamm [102] define *empathy* as the occurrence “*when an observer perceives or imagines someone else’s (i.e., the target’s) affect and this triggers a response such that the observer partially feels what the target is feeling*”.

According to Zaki and Ochsner [115] behavioral research has examined empathy in three major factors:

1. **Experience sharing:** shared self-other representations, emotional contagion, affective empathy
2. **Mentalizing:** theory of mind, cognitive empathy, perspective taking
3. **Prosocial concern:** empathic motivation and concern, sympathy

The authors claim that neuroscience research mainly focuses on the first two points.

4.2 Study of Principal Neuronal Mechanisms for Empathy

fMRI-studies show that brain networks active in observing affective states are also involved in firsthand experience, this means that shared networks are involved in a neuronal mechanism of empathy [7]. In particular the Anterior Insula (AI), dorsal- Anterior Cingulate Cortex (dACC) and anterior- Mid. Cingulate Cortex (aMCC) play a central role in different empathic states [31, 73]. Keyers [62] argues that the mirror neuron system is involved in empathy. Mirror neurons were originally discovered in monkey and premotor cortex, these neurons discharge when a monkey does a particular action but also when the monkey observes other individuals doing a similar action [25, 97]. This provides a neuronal mechanism for a shared representation in doing and understanding actions.

4.2.1 Empathy of Pain

Pain is robust in inducing empathy, thus empathy for pain is studied frequently [7]. Lamm et al. [73] gives a meta-analysis on 32 studies. Here two different types of paradigm design are investigated:

- **Picture-based:** The participants watch visual displays depicting limbs of target persons in painful situations. In this type of paradigm sensory-motor processes and motor mimicry might have a strong contribution to the neural process, this process may happen without full self-other awareness. [73] An example for a picture-based fMRI-study can be found in [74].
- **Cue-based:** Abstract visual symbols (cues) are displayed to the participant in the scanner to indicate whether the target person receives painful or non-painful stimuli. The target persons are seated next to the scanner. No explicit depictions of painful situations or any expression of pain is encountered by the participants. [73] An example for a cue-based fMRI-study can be found in [54].

Differences in the involved activations were observed between these two paradigm types [73]: During the other-related condition only in the picture based paradigm primary somatosensory cortex (S1) and secondary somatosensory cortex (S2) activation could be found. The picture-based paradigm also showed higher activation compared to cue-based design in the mirror-neuron network. Cue-based design showed higher activation in networks known to be involved in “Theory of Mind” and “mentalizing”¹. Bilateral AI, Medial Cingulate Cortex (MCC) and bilateral inferior parietal cortex (IPC) are frequently found in empathic pain experience in both paradigms [73].

Corradi et al. [18] used the method of searchlight MVPA, a method especially useful in analyzing the local fine grained multivariate activation pattern, to answer the question if these patterns identified as coding for empathy consist of a unique distributed population of bimodal neurons. Alternatively the neurons can be intermingled but independent populations of neurons in these areas. In this study the authors use local volumes of 125 voxels each with a volume of 27 mm³ per voxel. A picture based paradigm is used. In contrast to other studies they

¹ “ability to infer and represent beliefs and desires” [7]

also control for the negative valence in painful pictures by including pictures with negative but painless content. They conclude that Middle Insula (MI) and MCC show similar patterns for felt and seen pain, this is also true for the AI but this pattern seems to be more general shared across negative–aversive content. For detailed comments on this study see [111].

4.2.2 Empathy in non Pain Experiments

The data investigated in this thesis contain valences of “pleasant”, “neutral” and “unpleasant” tactile stimuli, and non painful conditions. Thus studies where empathy for non–painful scenario is investigated are of special interest.

Also in various other empathic experiments other than pain the central role of AI and Anterior Cingulate Cortex (ACC)/MCC has been shown in various studies, see [7] for review non–pain studies. Studies of special interest for the investigated data are given in the following:

Positive and negative empathic emotions were e.g. studied by Jabbi et al. [56], subjects in their study observed disgusting, neutral and pleasant facial expressions related to taste. They revealed that the AI and adjacent frontal operculum is not only involved in negative emotion but also involved in positive emotions. In a very similar experiment only an overlap between disgust in self and other was observed in AI and ACC, but no overlap between self and other is reported for pleasure [113].

Sense of touch in social cognition is also studied: posterior Insular Cortex (pIC) is reported to be positively modulated when being self touched, but negatively modulated (suppressed compared to baseline) when observing others being touched [43]. Another interesting finding is that shared neuronal activity can be found in S2 for *any* type of touch, regardless if the touch is animate or inanimate² [27].

4.3 Empathy Related Studies of Interest and MVPA

Except in the Corradi et al.–study [18] information–based MVPA approaches in empathy research are not used. Pattern classification is named in [7] for being of further interest in empathy research. This is done in this thesis. To further point out the usefulness of this tool some closely related studies are given here:

With a searchlight MVPA it was found out that the medial PreFrontal Cortex (mPFC) and the left Superior Temporal Sulcus (STS) contain informative patterns coding for specific emotions (e.g. anger, fear, joy) regardless of the used sensory cues (faces, bodies, voices) [91]. This study did not explicitly test if these patterns are also consistent across self and empathic experience.

Decoding multivoxel patterns restricted to certain ROIs, the emotional expression of faces was successfully decoded [48]: Fusiform Face Area (FFA) as well as Early Visual Cortex (EVC) independently carry informative patterns for the emotional expression of observed faces.

²e.g. they show a tree branch touching a chair

4.4 Map of the Empathic Brain

Various locations in the brain were mentioned as being involved in empathy in this chapter. MVPA gives rise to various other regions being involved in processing emotions. To ease the orientation some important areas are marked in Figure 4.1, Table 4.1 gives a short description.

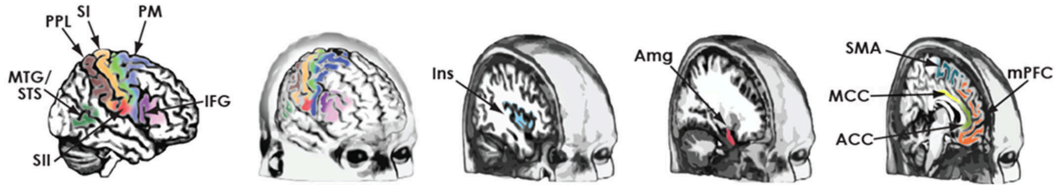


Figure 4.1: Map of the empathic brain: Graphics taken from [64] and [63] with permission from the author. MCC is not marked in the original source. See table 4.1 for explanations.

| | Full Name | Function |
|------|------------------------------|---|
| ACC | Anterior Cingulate Cortex | links emotion and action [62] |
| Amg | Amygdala | pain, fear [21, 72], positive [99] emot. processing |
| IFG | Inferior Frontal Gyrus | programming complex actions and language [62] |
| Ins | Insula | senses inner state of body, controls visceral response; emotions [62] |
| MCC | Middle Cingulate Cortex | reported as important part for empathy of pain [18, 31, 73] and also non-pain empathy [7] |
| mPFC | medial PreFrontal Cortex | cognitive processing of the state of self and others [62]; codes for emotion [91] |
| MTG | Middle Temporal Gyrus | perception socially relevant entities e.g. faces [62] |
| PM | PreMotor Cortex | plans action [62] |
| PLL | Posterior Parietal Lobe | integrates senses' info. and programs action [62] |
| S1 | primary somatosensory cortex | somatosensory signals (projection of body) [16] |
| S2 | secondary som.sens. cortex | sensation of touch [27, 62] |
| SMA | Supplementary Motor Area | plans and controls action [62] |
| STS | Superior Temporal Sulcus | codes for emotion [91] |

Table 4.1: Acronyms in Figure 4.1. Note: „Function“ may not give a full functional description

4.5 Summary

Empathy is the ability to share affective states of others and is crucial for social interaction. Neuronal mechanisms of empathy are of interest for the psychologist in basic research. One factor frequently investigated and a central point of this thesis, is to identify the neuronal basis which is shared between self perception and the observation of others (empathy condition). The

way of presenting the empathy condition effects the neuronal response. Thus results from this thesis should be compared with other picture-based fMRI-studies. A comprehensive summary on the areas which are known to be involved in empathy is given in Table 4.1.

Methods for Detecting Shared Distributed Patterns of Neural Activity

In a short summary of the motivation in Chapter 1, the method has to perform the following analysis-steps:

- Identification of the multivariate neuronal activation patterns in fMRI data across the entire cortex, that distinguish between three different emotions (namely pleasant, neutral and unpleasant) induced by visuo-tactile stimuli during fMRI acquisition in self perception, as well as in empathic perception.
- Identification of those networks that exhibit similar patterns in self- and empathic- perception.
- Evaluation of the information encoded in individual voxels, and in long-range cliques of voxels.

This chapter first outlines the contribution of this method in addition to a standard RF-training. The description of the method is broken down in the following steps:

1. **Study design** describes the fMRI-study, in which subjects a) perceive visuo-tactile stimuli which are known to induce different emotions and b) observe others being touched with the objects inducing these emotions.
2. **Data are prepared** for training a classifier.
3. **Training** a suitable multivariate classifier to decode the emotional state from the fMRI-data in the self- as well as in the empathy-session.
4. Using standard **variable importance** measures, voxels showing important activation for decoding the emotional state are identified.
5. **Shared patterns** across both sessions are identified
6. The **importance of cliques of voxels** is calculated.

5.1 Contribution of the Proposed Method

Compared to other MVPA of fMRI-data the proposed method of this thesis provides two major contributions.

5.1.1 Contribution 1: Identification of Distributed Shared Patterns

One factor on the basis of which empathy is investigated, are shared neuronal representations between self and other [115]. For examination of such shared representations, experiments as the following are done: 1.) self-session: a participant gets a certain stimulus to experience e.g. pain, pleasure, disgust, ... and 2.) other-session: one observes or gets a cue for another person experiencing such stimuli. In the latter case the participant will show empathy with the other. Neuronal representations present in both conditions are of interest in terms of “shared representations”.

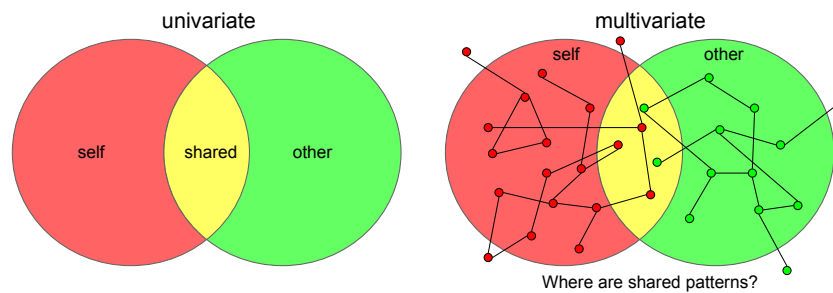


Figure 5.1: Shared activation: In the univariate case (left) a population of voxels (red) is identified in the self-session to be significantly activated, another population (green) is significantly activated in the other-session. The significance was calculated univariately, thus the intersection of voxels active in both conditions (yellow) show shared activation in both sessions. In the multivariate analysis (right) the concept has to be refined. In the graph some voxels were picked out, the connections between them indicate that the voxels belong to one multivariate pattern. Again voxels important in both conditions can be found, but they are part of different patterns. Thus in this hypothetical example the four features in the yellow intersection are not shared between the patterns coding for self- or the other-session.

The identification of **shared patterns** between self-perception and empathy brings up an important question: What are shared patterns?

The question is easy to answer for the univariate case, see Figure 5.1 (left half): each voxel’s activation is analyzed in a univariate way. A voxel with a significant activation contrast between two conditions in either the self- or the other-session is informative for its own. Thus a voxel active in both sessions can be marked as shared. Simple the intersection of both populations of univariately activated voxels can be defined as *shared* activated. In the multivariate analysis the importance measures, as described in Section 3.4.2, give a measure for each single voxel. In contrast to the univariate case these voxels were not analyzed independently from all other

voxels, this implies that they might not be informative on their own, but only in a complex conjunction of other voxels. If one would simply take the intersection of important voxels in both sessions, it would be neglected that each voxel is part of a pattern. Intuitively it is not useful to take the condition “important in both conditions” as the criteria for *shared*. Shared voxels should be marked as “shared” if they are part of the same multivariate pattern. Figure 5.1 (right half) gives a simplified illustration of the issue.

The main idea of finding the shared patterns is to test if a classifier, trained to decode the emotion from fMRI-measurements in self-perception, is also able to decode the emotion from fMRI-measurements when the subject feels empathy.

5.1.2 Contribution 2: Quantification of Informativeness of Combined Features

The proposed method substantially differs searchlight approach in the fact, that it reveals distributed patterns instead of local informative patterns. This immediately poses the questions if those long-ranged conjunctions of features, which cannot be captured by the searchlight approach, carry useful information. Only if such connections can be found the method is a useful extension to the searchlight approach.

For addressing this question the Gini Importance was modified to measure the decrease in impurity induced by a pair of features. Further this importance of pairs of features can be pooled across extended regions of anatomical areas in the brain. With the knowledge on the functional association of these areas further insight in the neuronal processes of empathy can be gained.

5.2 Study Design: A Paradigm for Self – Perception and Empathy

A new developed paradigm by Silani et al. [101] was used to assess touch associated positive and negative feelings in self and empathic perception. A visuo-tactile stimulus followed by a rating-block was used for this experiment. The data used for demonstrating the proposed methods of this thesis are from an experiment using the the same visuo-tactile stimuli with a subsequent rating-block after each stimulus block. Data are generously provided by Univ.-Prof. Dr. Claus Lamm¹.

5.2.1 Experimental Setup and Stimuli

30 participants were scanned in two sessions: 1.) self-session, 2.) other-session (empathy):

1.) For the **self-session** the stimuli consisted of pictures of objects or animals and of simultaneous tactile stimulation by stroking the participant’s left palm with a material with a surface texture similar to that of the displayed object. Different materials and objects were tested in a pilot experiment. The final set of 10 pleasant, 10 neutral and 10 unpleasant stimuli is listed in Table 5.1. Pictures were matched for size, luminance and semantic category for all three types of stimuli. Tactile stroking was performed on the left hand palm. Simultaneously the matching

¹Department of Basic Psychological Research and Research Methods – University of Vienna

visual stimulus was presented using a back projection system. The visuo–tactile stimulus block lasted for 4 s. The stimulus block was followed by a 5 s rating block where the participants had to rate the emotional response with a track ball as input device. The next stimulus started after a 2.5 s Inter Stimulus Interval (ISI).

2.) For the **other–session** the procedure was exactly the same except that no tactile stimuli were provided. Participants had to rate the emotion of others being touched with the objects displayed on the screen.

| Picture | Material | Type of Stimulus |
|----------------|-----------------|-------------------------|
| Dog | Wool 1 | Pleasant |
| Cotton plant | Cotton 1 | Pleasant |
| Cat | Fur 1 | Pleasant |
| Brush | Brush 1 | Pleasant |
| Sheep | Wool 3 | Pleasant |
| Rabbit | Fur 2 | Pleasant |
| Swan | Feather 1 | Pleasant |
| Cotton ball | Cotton 2 | Pleasant |
| Rose | Silk | Pleasant |
| Chick | Feather 2 | Pleasant |
| Pen | Pen | Neutral |
| Branch | Branch 1 | Neutral |
| Stork | Wood stick | Neutral |
| Walnut | Walnut | Neutral |
| Turtle | Shell | Neutral |
| Box | Cardboard | Neutral |
| Elk | Branch 2 | Neutral |
| Dog 2 | Brush 2 | Neutral |
| Peanut | Peanut | Neutral |
| Wild pig | Brush 3 | Neutral |
| Mushroom 1 | Slime 1 | Unpleasant |
| Tongue | Silicon | Unpleasant |
| Stinkbug | Stinkbug | Unpleasant |
| Mushroom 2 | Slime 2 | Unpleasant |
| Catfish | Slime 3 | Unpleasant |
| Maggots | Slime 4 | Unpleasant |
| Liver | Slime 5 | Unpleasant |
| Slug | Slime 6 | Unpleasant |
| Spider | Spider | Unpleasant |
| Oyster | Slime 7 | Unpleasant |

Table 5.1: List of materials and pictures used to induce different types of stimuli

5.2.2 Technical Details on the fMRI Data – Acquisition and Preprocessing

Functional images were acquired with $T_R = 2.06s$ at resolution of $3mm \times 3mm \times 3.3mm$ and with $n_m = 180$ volumes per session. Data were provided movement-, slicetime- corrected and normalized. From each subject one T1 weighted structural image was acquired.

5.3 Data Preparation for Classification

The aim of the data preparation explained in this section is twofold: First all subjects' BOLD-measurements have to be aligned to a common atlas, second the stimulus paradigm has to be converted to labels which can be used to train the classifier.

5.3.1 Cortical Alignment of the fMRI-Data

FREESURFER is a freely available suit of tools for the analysis in neuroimaging. It automatically creates cortical surface models and an anatomical parcellation of macroscopically visible structures. These surface models can be used to align data across subjects. [33,93]

For a surface based group registration the subject's cortical surface is registered on a surface atlas. By registration of the functional images on the anatomical image all data can be registered on the same surface atlas where group analysis can be performed. Fischl et al. [32] report that a surface based intersubject alignment is more accurate than a volume based group analysis. Desai et al. [22] also show a higher precision for surface-based alignment compared to volumetric approaches in the case of localizing auditory cortex activation. The reduced geometric variability of activated regions can be explained by dimensionality reduction and a better registration between subjects [110]. Thus this surface based approach is a very good method if the neuroscientific question is limited to the cortical surface; deep brain regions can not be analyzed with this approach [93]. It is also shown that surface based approach is usable in conjunction with MVPA [88].

Figure 5.2 provides a graphical illustration of the following preprocessing steps: **a)** From the scanning-institution for each of the 30 subjects ($n_{\text{subj}} = 30$) one anatomical T1-weighted image and two functional volumetric time series, one for the self-session and the second for the other-session (empathy-session) are provided. The functional data are slice-time and motion corrected. Each function time series contains $n_m = 180$ volumes. **b)** The anatomical data are processed using FreeSurfer's `recon-all` command, it comprises 31 processing steps, containing various preprocessing steps, segmentation, registration and parcellation steps. All steps are listed in [38]. Furthermore the functional volumes are registered on the structural volume, which allows to map all functional data on the same common atlas. FS4 with 2562 surface-nodes per hemisphere is used². **c)** As output for each subject two matrices are obtained. Z_{self} contains the BOLD measurement of the self-session and is of the size $V \times n_m = 5124 \times 180$. Z_{other} is

²FS4 with 2562 surface-nodes per hemisphere was used for all final evaluations. In some points of the thesis also experiments with FREESURFER `fsaverage5-atlas` (FS5) with 10242 nodes per hemisphere was used.

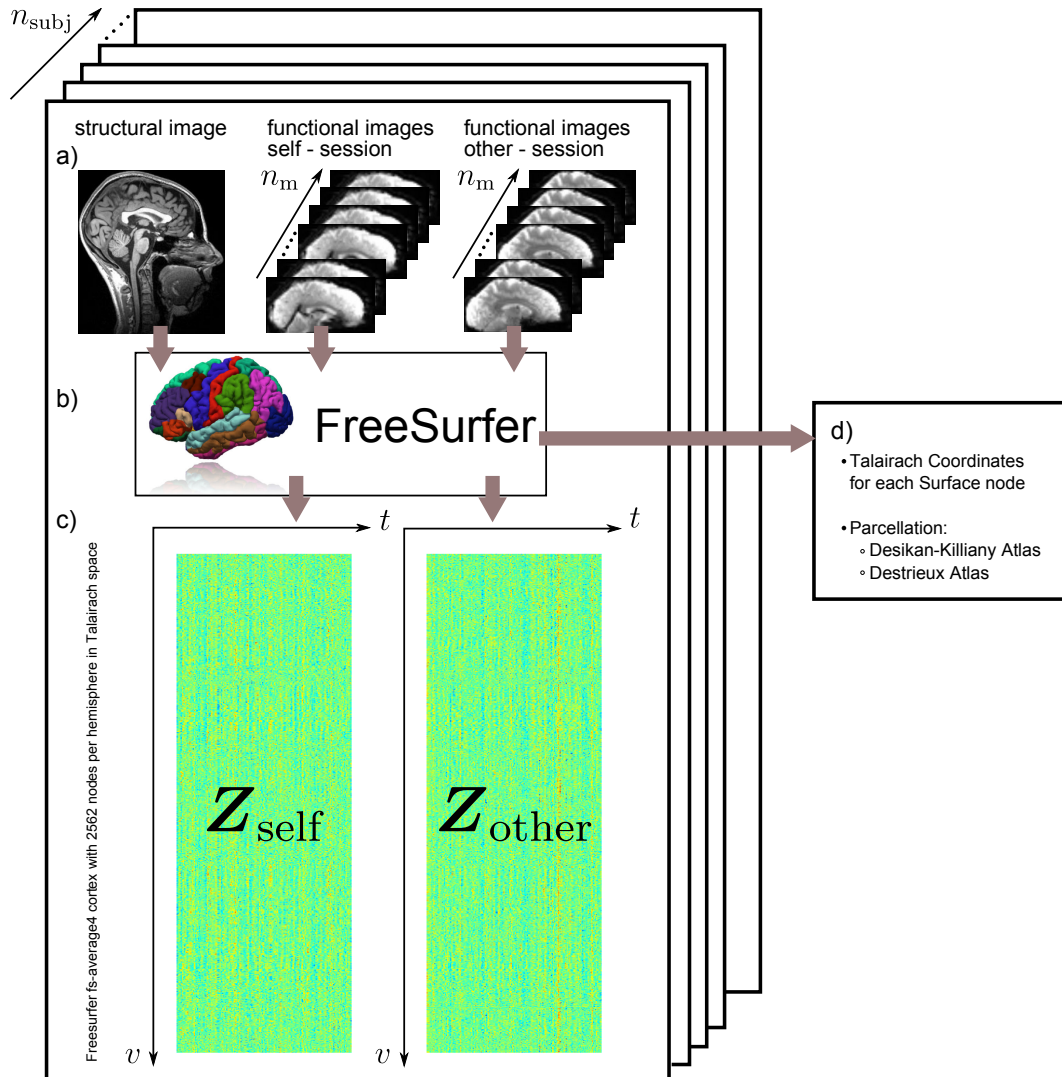


Figure 5.2: Preprocessing with FREESURFER: a) Anatomical image and the two functional images are fed into FREESURFER. Note: each depicted 2D-image represents a whole volumetric image. After cortex surface segmentation the subject’s cortical surface is mapped onto FS4. Functional images are aligned with the structural image. c) For each of the n_{subj} subjects two matrices, Z_{self} for the self-session and Z_{other} for the other-session, of BOLD data are calculated, where each of the 5124 features corresponds to one of the 5124 surface nodes on the group atlas. Each matrix has the size $V \times n_m$, where $V = 5124$ features and n_m the number of measurements in time t . d.) FREESURFER also outputs Talairach Coordinates for each node and anatomical parcellation-labels according to two different atlases. FREESURFER-logo taken from [34]

the same size and contains the BOLD–data of the other–session. Motion parameter obtained at motion–correction step as well as linear trends are regressed out using a linear model. In total all further analysis are done on 5124 surface nodes V with 180 measurements per session n_m in two sessions on 30 subjects n_{subj} . This results in $5124 \cdot 180 \cdot 2 \cdot 30 = 55,339,200$ BOLD measurements. **d)** FreeSurfer also outputs Talairach [108]-Coordinates for each surface node and anatomical parcellation labels [35, 36, 75, 76]. These atlases are available: 1.) Desikan-Killiany Atlas [23] providing 32 automatic labels in each hemisphere 2.) Destrieux Atlas [24] providing 74 automatic labels per hemisphere.

5.3.2 Preparing Labels for the Classifier

For classifier–training a set of observations \mathbf{X} and a response vector \mathbf{y} containing a label l for each observation have to be provided. For this purpose the BOLD–measurements \mathbf{Z} are directly taken as the training data \mathbf{X} ; the response vector \mathbf{y} is derived from the paradigm via the HRF–function $HRF(t)$. For all analysis the HRF–function is taken from the FMRISTAT–toolbox [114]. It is modeled as a difference of two gamma density functions, parameters are chosen according to [46].

During each scanning session the onset– and duration–times for each block of stimulus or rating are recorded. From this the stimulus function $s_l(t)$ can be derived where l ranges from 1 to 3 for neutral, pleasant and unpleasant tactile and visual stimuli and from 4 to 6 for the corresponding blocks of rating each type of stimulus. We assume that each stimulus function $s_l(t)$ causes a measurable hemodynamic BOLD–response $h_l(t)$, which can be obtained by

$$h_l(t) = s_l(t) \otimes HRF(t) \quad (5.1)$$

The stimuli–functions (dashed lines) and the corresponding hemodynamic responses (full lines) are shown in Figure 5.3. A constant threshold–function th is introduced. th was chosen in such a way that the the resulting durations between rating– and stimuli–label is approximately equal the ISI–length of the paradigm. The label–function $l(t)$ is then given by:

$$l(t) = \begin{cases} 0 & \text{if } \max(h_i(t)) < th \\ 1 & \text{if } \max(h_i(t)) = h_1(t) \\ 2 & \text{if } \max(h_i(t)) = h_2(t) \\ 3 & \text{if } \max(h_i(t)) = h_3(t) \\ 4 & \text{if } \max(h_i(t)) = h_4(t) \\ 5 & \text{if } \max(h_i(t)) = h_5(t) \\ 6 & \text{if } \max(h_i(t)) = h_6(t) \end{cases} \quad (5.2)$$

where $\max(h_i(t))$ is the value of the function $h_1(t) \cdots h_6(t)$ with the highest amplitude. The function $l(t)$ is defined continuously and given in the lower trace of Figure 5.3. For the labels at each acquired volume the function $l(t)$ is sampled with a sampling interval of T_R . The vector containing the sampled values from $l(t)$ is denoted \mathbf{l} and has the length n_m .

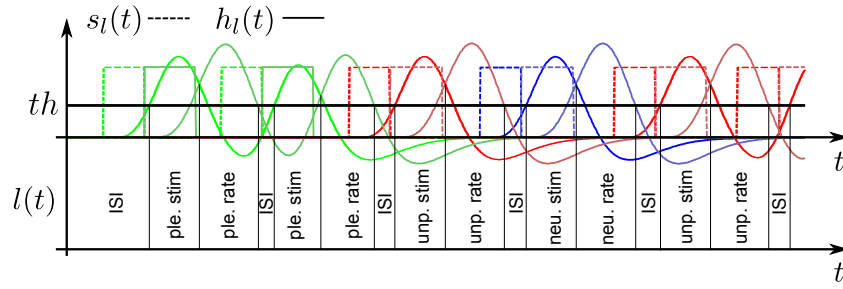


Figure 5.3: Labeling each point of time with the experimental condition: Hemodynamic responses $h_l(t)$ are derived from the stimulus-functions $s_l(t)$ by convolution with the hemodynamic response functions. Points in time where each response is below the threshold th are labeled with $l = 0$ *ISI inter stimulus interval*, otherwise the labels are $l = 1$ for *neutral stimulus*, $l = 2$ for *pleasant stimulus*, $l = 3$ for *unpleasant stimulus*, $l = 4$ for *rating of neutral stim.*, $l = 5$ for *rating of pleasant stim.* and $l = 6$ for *rating of unpleasant stim.*. The shown sample depicts the first 12 blocks of the self-session in subject #1.

Summary Training Data for Classification

With the calculation given above a label l can be assigned to each measurement. Thus the surface aligned BOLD-measurements Z can be used as training data X and the calculated label vector l is used as the target vector y . This is done for both, the self-session (X_{self} and y_{self}) and the other-session (X_{other} and y_{other}) for each subject. The individual data matrices and response vectors are stitched together to a group data matrices $X_{\text{self group}}$ and $X_{\text{other group}}$ and the group response vectors $y_{\text{self group}}$ and $y_{\text{other group}}$ as shown in Figure 5.4.

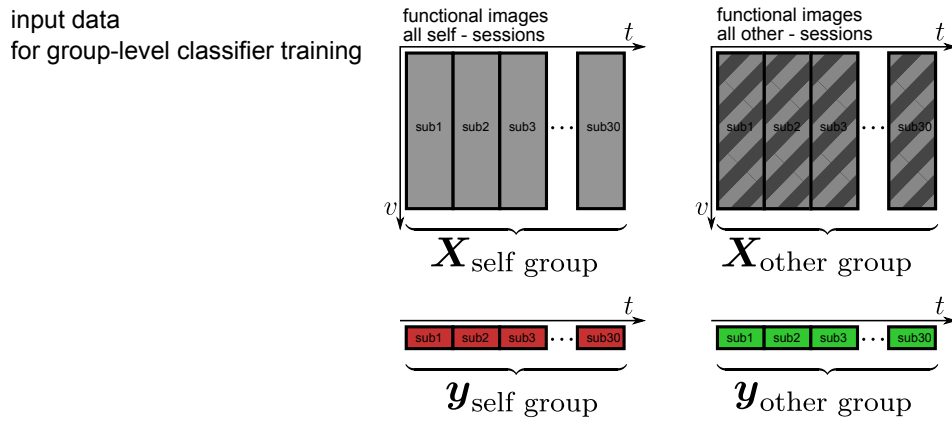


Figure 5.4: Structure of the group-matrix: 30 subjects' $V \times n_{\text{obs}}$ matrices are stitched to one $V \times 30n_{\text{obs}}$ group-matrix. This is done for both sessions.

Additional Notes for the Classical GLM Analysis

For comparison a GLM-analysis is performed. As predictor-functions (see Section 3.1) the functions $h_1(t)$ to $h_6(t)$ are used as predictors for the BOLD response of the stimuli and ratings.

5.4 Classifier Training for Decoding the Emotional State from fMRI-Data

For detecting the distributed multivariate patterns, which code for the emotional state, an RF-classifier is trained to decode the emotional state (labels response vector \mathbf{y}) from the measurements \mathbf{X} . This procedure is done for the data of both scanning sessions. From the two training two classifier models are generated:

- **Self-Model:** Classifier training to decode $\mathbf{y}_{\text{self group}}$ from $\mathbf{X}_{\text{self group}}$
- **Other-Model:** Classifier training to decode $\mathbf{y}_{\text{other group}}$ from $\mathbf{X}_{\text{other group}}$

For this work a MATLAB implementation of the RF was used, which is available at [57]. Classifier was trained with standard parameters and $m = \sqrt{V}$, as advised for classification. For n_{tree} a number in the order of V is useful [77]. This was confirmed in experiments with the data of this thesis: classification results were stable with $n_{\text{tree}} > 0.1V$, no further increase in stability of the variable importance measured was observed with $n_{\text{tree}} > \approx 0.5V$. Finally $n_{\text{tree}} = 5,000$ was chosen for FS4 with $V = 5,124$ and $n_{\text{tree}} = 20,000$ for FS5 where $V = 20,484$.

5.5 Measuring the Variable Importance of Voxels Within a Stimulus Category

Right after training the Gini importance measure is available. The Gini importance per se is not class specific, since it is the accumulation of decreases in impurity from the RF-training and in the training only the over all class impurity is evaluated.

A slightly **modified Gini importance** measure compared to the “standard” *Gini importance*, which is assessed by the classifier during training, is defined in the following. It is adapted to be assessable after training. This is particularly useful to measure the variable importance for a dataset completely independent from the data where the classifier was trained on. Another modification is made to get class specific importances for each label l . To satisfy the first requirement simply the whole dataset, and not just the in-bag samples for each tree, are evaluated by each tree. The second requirement is fulfilled by modifying the labels for the data: To assess the Gini-importance for a specific label l all data-points with a label $\neq l$ are marked with the same label. This results in a two class labeling with the label of interest vs. all other labels. Thus the decrease in impurity just calculates the purity of the data-points labeled with the specified label l vs. labels $\neq l$, but not a decrease in impurity between other classes than l . The calculation is exactly the same as given by Equation (3.10) and (3.12). Comparing the modified Gini importance after training, with the Gini importance from training, a different behavior can be expected. During training ΔI is evaluated only on the subsampled data points. In the assessment

after training all observations are fed into the evaluation algorithm.

The RF-implementation used also outputs the permutation importance, which is evaluated during training on the OOB-samples. When evaluating the permutation importance after training it will suffer from a slight bias, because now it also includes in-bag samples.

5.6 Measuring the Variable Importance of Voxels Shared Across Stimulus Category

To fulfill the requirement of finding voxels of shared patterns across stimulus category, we make use of the classifier-models. Since the classifier model can decode the correct label it “knows” the multivariate pattern typically coding for a specific label. Results show that the classifier trained e.g. on the self-session is also able to decode the emotion from the empathy-session (for an overview on various decoding performances see the overview in Section 7.1). The experiment of using the classifier trained on one session to decode the complimentary session will be termed as *model transfer* in the following. This implies that the *self-model* finds the patterns of the self-session in the data from the empathy-session. Exactly these patterns which allow the classifier to decode the data in the complimentary session can be considered as “shared” across both conditions. By assessing any variable importance measures (the modified Gini importance or the permutation importance) on the self-model by feeding in the empathy-data, we now see how important the features are for decoding the correct “empathy” label when using the self-model.

In this procedure assessing the variable importances after training is uncritical since all data from the other-session are independent from the training data and therefore OOB.

5.7 Measuring the Variable Importance of Cliques of Voxels

As given in Table 3.1 multivariate whole brain analysis is usually done with the searchlight approach. To see if wide-spread long-range patterns (outside a local searchlight volume) are important and contributive to a correct classification, the connection importance between pairs has to be measured. For this a new method was developed which is explained in the following:

Keeping in mind the algorithm of growing a random forest, the selected split-variable and value is assessed on the data reaching this node. The fraction of data reaching a node depends on the parent node’s split. This means each split-variable is conditional on the previous split, see also for a discussion on the condition of the parent node [106]. For the proposed method of the importance for pairs of variables this dependency between parent and child node as well as the concept of the Gini importance will be used. For a clear understanding the method will be rolled out in two steps.

5.7.1 Step 1: Joint Gini Importance

To get the gain in label purity in the data induced by the connection of two features, the Gini importance $I_G(v)$ is not simply summed over all nodes with the same split feature v but broken up to all nodes with a split feature v and the same parent split feature u . We define the joint Gini importance $I_{JG}(v, u)$ as:

$$I_{JG}(v, u) = \sum_{\text{all trees}} \left(\sum_{\substack{w : v^*(w) = v \wedge \\ w_{\text{parent}} : v^*(w_{\text{parent}}) = u}} \Delta I(w) \right) \quad (5.3)$$

where w_{parent} is the parent node number of node w . From Equation (3.12) and (5.3) it follows immediately that

$$I_G(v) = \sum_{\text{all } u} I_{JG}(v, u) \quad (5.4)$$

This means the sum of the joint Gini importance $I_{JG}(v, u)$ over all possible features for u results in the Gini importance $I_G(v)$. This implies if the feature v is predictive for its own it effects all joint importances where v is involved. In other words: The joint importance $I_{JG}(v, u)$ is biased for pairs involving a feature v which is predictive for its own. Since we want just the importance of a connection, without a bias of uniquely important features the value has to be corrected in second step:

5.7.2 Step 2: Correction for Unconditioned Importance

To address the fact that a variable could be predictive on its own an additional measure is introduced: $\Delta i(w, \mathbf{X}_{w_{\text{parent}}})$ measures the decrease in impurity by the split of the observed node w , if it would split the whole parent data $\mathbf{X}_{w_{\text{parent}}}$ and not just the data \mathbf{X}_w conventionally assigned to node w . We denote this as the „unconditional“ decrease in impurity where the node’s decrease in impurity is not conditional on the parent’s split in node w_{parent} . The unconditional decrease impurity can be assessed with the definition of the decrease in impurity (Equation (3.9)) and is calculated by:

$$\Delta i(w, \mathbf{X}_{w_{\text{parent}}}) = \left(i(\mathbf{X}_{w_{\text{parent}}}) - \frac{n_{w_1}}{n_{w_{\text{parent}}}} \cdot i(\mathbf{X}_{w_1}) - \frac{n_{w_2}}{n_{w_{\text{parent}}}} \cdot i(\mathbf{X}_{w_2}) \right) \quad (5.5)$$

Note: $\Delta i(w, \mathbf{X}_{w_{\text{parent}}}) \neq \Delta i(w_{\text{parent}}, \mathbf{X}_{w_{\text{parent}}})$ – for the calculation the parent–data $\mathbf{X}_{w_{\text{parent}}}$ are split with the split–criteria of the node w and not on the parent–node w_{parent} . Analog to the Gini importance a weighting has to be made to account for the amount of data. The weighted decrease in „unconditional“ Gini–importance is weighted equally as the conditional weighted decrease in impurity $\Delta I(w, \mathbf{X}_w)$, which is the number of data at node w . Thus $\Delta I(w, \mathbf{X}_{w_{\text{parent}}})$ can be calculated as:

$$\Delta I(w, \mathbf{X}_{w_{\text{parent}}}) = \Delta i(w, \mathbf{X}_{w_{\text{parent}}}) \cdot n_w \quad (5.6)$$

Analogous to Equation (3.11) but accounting for the fact that the conditional and unconditional weighted decrease in impurity has to be calculated with the same weighting factor n_w Equation (5.6) in conjunction with Equation (5.5) we get:

$$\Delta I(w, \mathbf{X}_{w_{\text{parent}}}) = \left(i(\mathbf{X}_{w_{\text{parent}}}) - \frac{n_{w_{1\text{parent}}}}{n_{w_{\text{parent}}}} \cdot i(\mathbf{X}_{w_{1\text{parent}}}) - \frac{n_{w_{2\text{parent}}}}{n_{w_{\text{parent}}}} \cdot i(\mathbf{X}_{w_{2\text{parent}}}) \right) n_w \quad (5.7)$$

See Figure 5.5 for definition of the variable names.

If the (conventional) conditional weighted decrease in impurity $\Delta I(w, X_w)$ is greater than the decrease in the unconditional case $\Delta I(w, X_{w_{\text{parent}}})$ (meaning if $\Delta I(w, X_w) > \Delta I(w, X_{w_{\text{parent}}})$) the parent split on node w_{parent} was useful for the split in node w since the decrease in impurity in the data X_w is higher than splitting the parent–data $X_{w_{\text{parent}}}$ with the same split criteria. The parent split was useful for its child node w for further purifying the data. For $\Delta I(w, X_w) \leq \Delta I(w, X_{w_{\text{parent}}})$ the parent split was not useful; the split in node w would lead to the same or higher decrease in impurity if it would split the data without the upstream parent split. In the latter case the combination of the split feature and the parent’s split feature is not informative for gaining pure classes. Thus finally the reduction in impurity for the combination of feature v and u can be calculated with:

$$I_{CJG}(v, u) = \sum_{\text{all trees}} \left(\sum_{\substack{w : v^*(w) = v \wedge \\ w_{\text{parent}} : v^*(w_{\text{parent}}) = u}} \Delta I(w, \mathbf{X}_w) - \Delta I(w, \mathbf{X}_{w_{\text{parent}}}) \right) \quad (5.8)$$

This measure will be called *Conditional Joint Gini* importance. Child-Parent and Parent-Child connections are equally important, thus the purity gain by feature v and u is given by the sum $I_{CJG}(v, u) + I_{CJG}(u, v)$.

Special Issues in the Conditional Joint Gini Evaluation

The usage of V features leads to $\frac{V^2-V}{2}$ possible connections. In the case used of $V = 5,124$ this results in 26,252,814 possible connections. One tree in the forest contains approximately 280 splits for the used data–set. This means that at least nearly 100,000 trees would be necessary to capture each connection once. Since the Gini split criteria captures important features more likely, it is also more likely that informative pairs are selected more frequently than by chance, since the Gini split criteria preferably chooses informative voxels. But with a moderate number of trees, which is mandatory for computational reasons, the value of important connections will still be based on a few captured pairs. Due to the random subsampling of features in the RF–training these values can be expected to be noisy.

Another issue is that 26,252,814 connections are uninterpretable. This can be managed by calculating connections between bigger areas in the brain: e.g. Venkataraman et al. [112] used the surface parcellation of FREESURFER’s segmentation [35]. Summation of $I_{CJG}(v, u)$

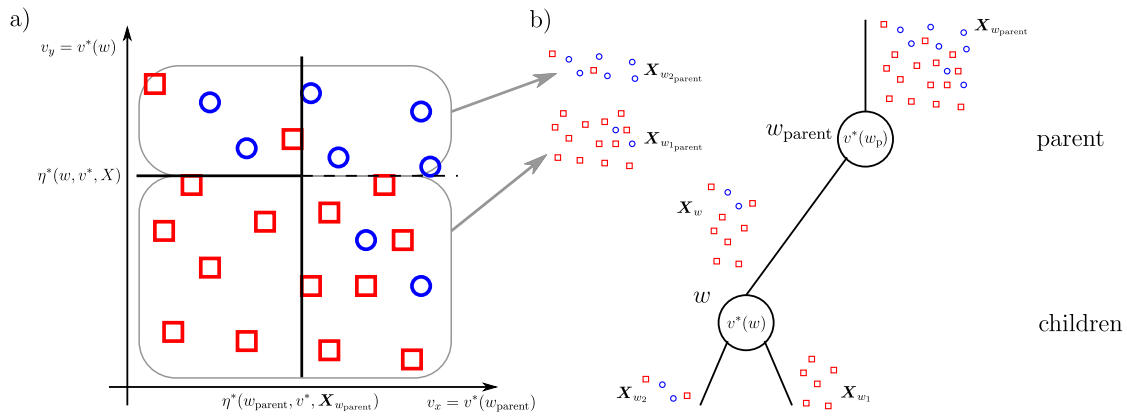


Figure 5.5: Nomenclature of the data-sets: a) Full parent data $\mathbf{X}_{w_{\text{parent}}}$ in two dimensions of the feature-space; b) In each branch the corresponding parts of the data and their names are given. $\mathbf{X}_{w_{1\text{parent}}}$ and $\mathbf{X}_{w_{2\text{parent}}}$ are no split data-set of the tree, but arise from hypothetical unconditioned split of the full parent data perpendicular to the feature $v^*(w)$ at a value of $\eta^*(w, v^*, \mathbf{X}_w)$

across all features v in on parcel to all features u in a other parcel delivers stable results, but the anatomical parcellation is problematic since the parcel size differs strongly, accounting for a severe variation in contributing node to node connections. Thus the brain was parcellated in arbitrary cells of equal size. To avoid cutting apart functional similar nodes a parcellation procedure as given in Algorithm 5.1 was used. $n_p = 75$ cells per hemisphere are used. This is the same amount of parcels used in the Destrieux-atlas [24], thus the mean parcels size is the same as in this atlas. 20,000 trees are necessary for a stable result.

input : matrix \mathbf{Z} (size $V \times n_m$) containing BOLD measurements
input : surface mesh with $\frac{V}{2}$ surface nodes per hemisphere
output: a surface parcellation

- 1 Randomly spread n_p seeds per hemisphere by maximizing;
- 2 **repeat**
- 3 | measure the seed to seed distance in the mash with Dijkstra's algorithm [26];
- 4 | relocate seeds to maximize distances;
- 5 **until** all seed to seed distances converge to maximal distance;
- 6 Assign each seed to a parcel;
- 7 **repeat**
- 8 | **foreach** parcel **do** find all unoccupied neighbor nodes;
- 9 | select node with highest correlating BOLD response between node and parcel-mean;
- 10 | assign selected node to the parcel ;
- 11 **until** each node is assigned to a parcel;

Algorithm 5.1: parcellation-algorithm: Algorithm grows n_p of equal size with preference of BOLD-correlation within a parcel

5.8 Summary

The main contribution and novelty compared to state of the art methods is twofold: First the method is able to detect distributed multivariate patterns shared across stimuli category, second the informational content cliques of features can be measured. For finding the shared patterns in between self-perception and observing others the experimental setup has to provide two independent data sets. The classifier is trained on both, the self-data and the other-data. Variable importance measures are used to identify important features of the multivariate patterns. The main idea for detecting shared patterns is to use the model trained on the first session to decode the data from the second session. The importance of cliques of voxels is assessed by a newly introduced measure.

Experiments and Results

This chapter is organized as follows: First the results from a classical **GLM analysis** are given for comparison to the new method used in this thesis. The **decoding performance** is investigated. The results of comparing different **variable importance** measures are given. Further the permutation importance was used for cortical **visualization** of informative activated regions for each fMRI-scanning session and for the shared distributed patterns. Compared to the GLM, the MVPA is able to capture informative patterns across multiple voxels. Thus the information gained from combining features is separately evaluated, to give a closer insight into the connected importance. The **joint importances** are visualized in a schemaball-plot.

For each experiment the structure sticks to the following organization:

1. The **aim** of the experiment is explained.
2. The **data** used for the experiment are given.
3. The **experiment's procedure** is outlined.
4. Outline of the **evaluation**: used measures, visualizations and colormaps
5. The experiment's **results** are given and explained.

General Explanation of Cortical Visualizations: In the following sections cortical visualizations of the hemispheres are organized as follows: Left Hemisphere (LH) and Right Hemisphere (RH) are plotted separately. Views are orientated as given by

| | | | |
|------------------|------------------|------------------|------------------|
| LH medial view | LH left view | RH right view | RH medial view |
| LH superior view | LH inferior view | RH inferior view | RH superior view |

Table 6.1: Ordering of the views for LH and RH in all plots

For visualization the triangular mesh of 2,562 surface nodes was smoothed in shape before rendering with a mesh of 10,242 nodes. Values, depicted as colors, were interpolated between the original 2,562 surface nodes, no smoothing was applied.

6.1 Classical GLM Analysis

For comparison to the presented multivariate analysis, a univariate GLM-analysis is performed.

The GLM-analysis is performed on the BOLD-measurements from all 30 subjects. The GLM is calculated separately for the data from the self-session Z_{self} and for the data from the other-session Z_{other} . Typically fMRI group-modeling is done in a two step approach [83]. In the MVPA all subjects' BOLD-measurements are stitched together as shown in Figure 5.4. This corresponds to a fixed-effect model. For statistical agreement with the MVPA-model, also a fixed-effect model is used for the GLM group analysis. Thus the measurement matrices $X_{\text{self group}}$ and $X_{\text{other group}}$ are used. This model is valid for the subject population [84].

The GLM is modeled and calculated exactly as given in Section 3.1. Since the classifier was trained on identifying the different types of stimuli, the corresponding GLM-modeling is the activation contrast of one stimulus compared to the other stimuli. The following contrast vectors are used:

- **neutral vs. pleasant and unpleasant:** $c^T = (2 \quad -1 \quad -1 \quad 0 \quad 0 \quad 0)$
- **pleasant vs. neutral and unpleasant:** $c^T = (-1 \quad 2 \quad -1 \quad 0 \quad 0 \quad 0)$
- **unpleasant vs. neutral and pleasant:** $c^T = (-1 \quad -1 \quad 2 \quad 0 \quad 0 \quad 0)$

Note: The first three estimators are the stimuli paradigm functions, the last three estimators correspond to the ratings.

For visualization the color map is defined as following: Voxels with insignificant activation contrast are not colored. False Discovery Rate (FDR)-correction was done by the procedure of Benjamini et al. [5, 6]. $p = 0.01$ is used to be in agreement with results from [73, 113]. Significant voxels are colored according to the value of $c^T\beta$, same scaling factor is used for all plots. Scaling factor is chosen to fit values in the range -1 (blue) to +1 (red). For visualization of shared activation voxels, significantly activated in both, the GLM on the data $X_{\text{self group}}$ and on the data $X_{\text{other group}}$, are colored in the colormap chosen.

Results: Figure 6.1 – 6.3 depict the activation contrast for all three stimulus types in the self-session (self perceived stimulus $X_{\text{self group}}$). Figure 6.4 – 6.6 depict the same contrast for the other-session (participants observe others perceiving the stimuli from the self-session). In both sessions the neutral stimulus has a predominant negative activation contrast (blue), this means the colored areas are significantly weaker activated for neutral stimuli than for the two other stimulus types.

For the unpleasant stimulus the majority of voxels with a significant activation contrast show a higher activation than in the case of the two other stimulus types. The pleasant stimuli show the smallest amount of voxels with a significant activation contrast. Across all stimulus categories the amount of significantly activated voxels is lower in the other-session.

Figure 6.7 – 6.9 show the voxels with significant activation contrast in both sessions. These voxels' activation is shared between self-perception and observing others (empathy).

These results will serve as a comparison analysis for the proposed method.

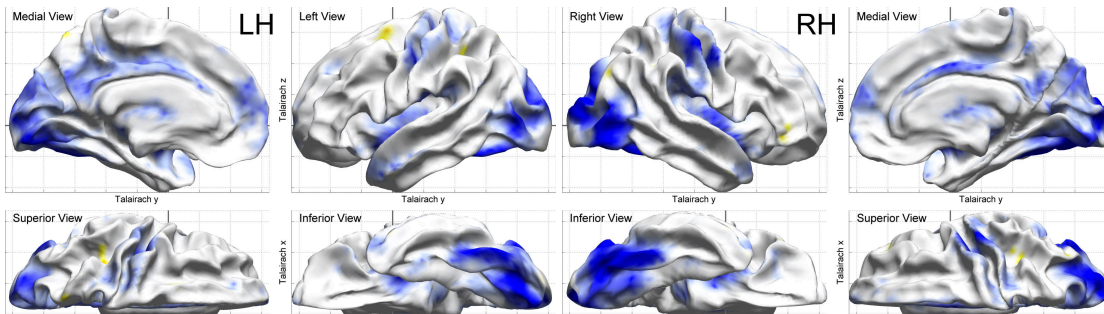


Figure 6.1: GLM analysis: Voxels with a significant activation contrast for neutral vs. pleasant and unpleasant stimuli in the self-session

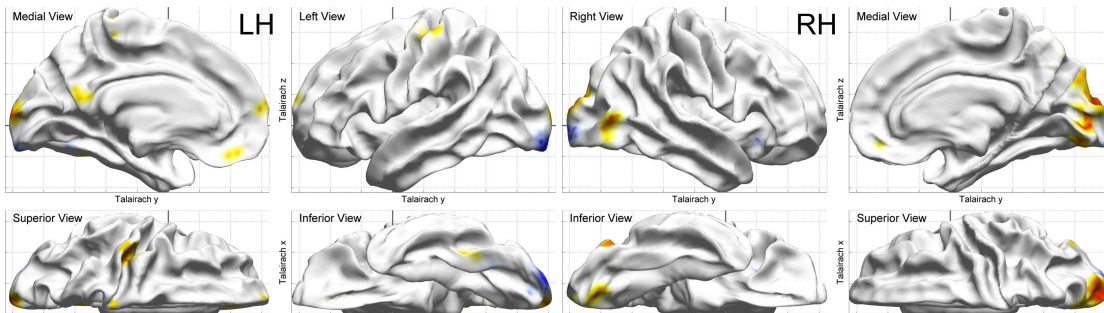


Figure 6.2: GLM analysis: Voxels with a significant activation contrast for pleasant vs. neutral and unpleasant stimuli in the self-session

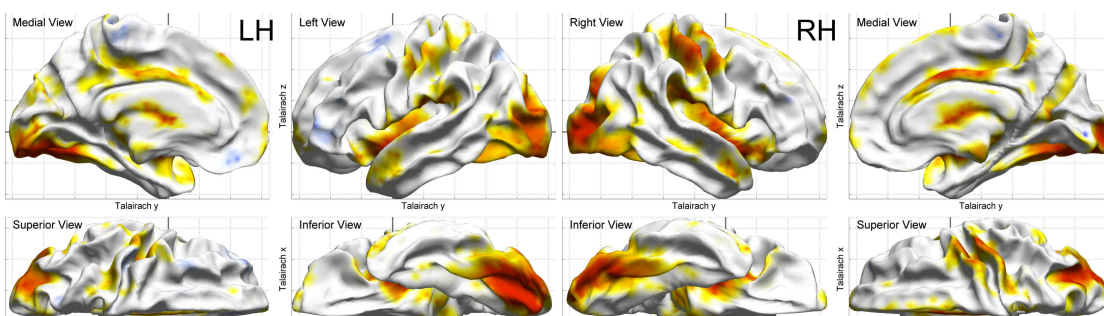


Figure 6.3: GLM analysis: Voxels with a significant activation contrast for unpleasant vs. neutral and pleasant stimuli in the self-session

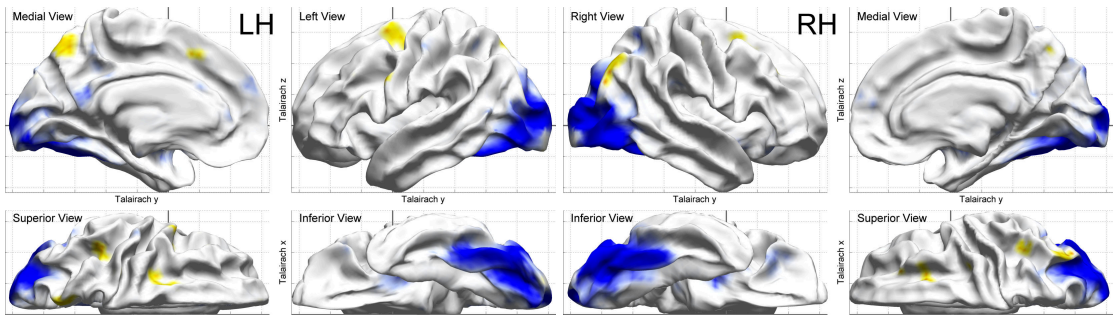


Figure 6.4: GLM analysis: Voxels with a significant activation contrast for neutral vs. pleasant and unpleasant stimuli in the other-session

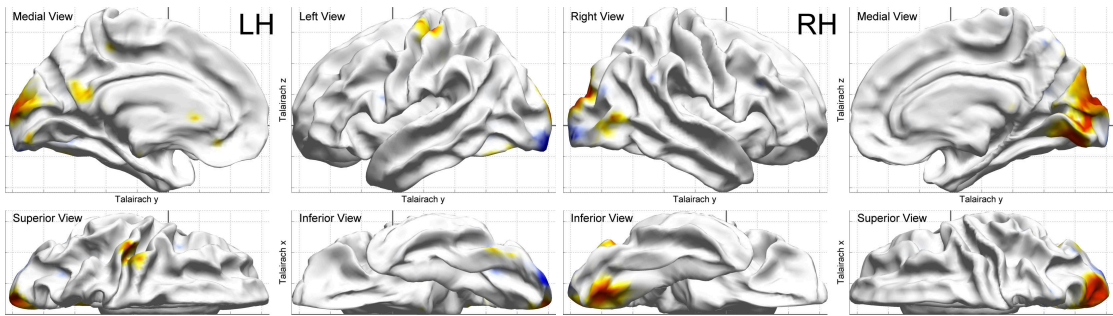


Figure 6.5: GLM analysis: Voxels with a significant activation contrast for pleasant vs. neutral and unpleasant stimuli in the other-session

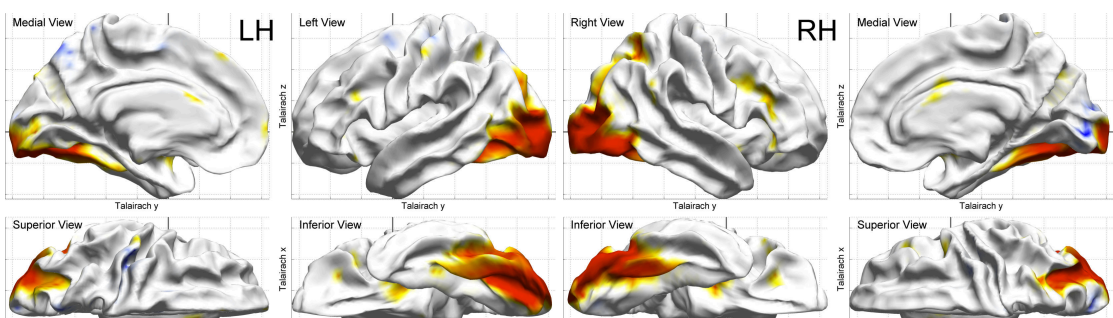


Figure 6.6: GLM analysis: Voxels with a significant activation contrast for unpleasant vs. neutral and pleasant stimuli in the other-session

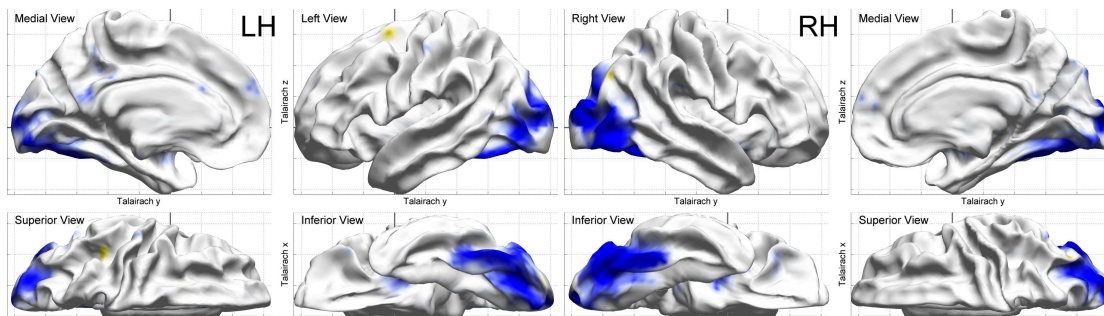


Figure 6.7: GLM analysis: Voxels with a significant activation contrast for neutral vs. pleasant and unpleasant stimuli in both, the self-session and the other-session (shared activation)

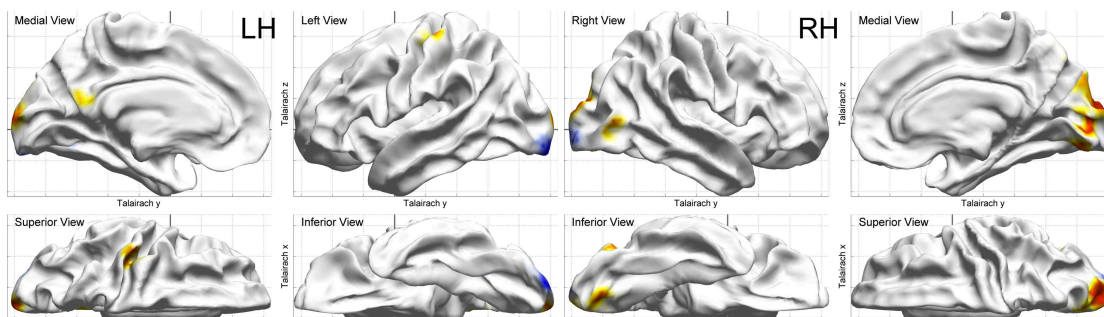


Figure 6.8: GLM analysis: Voxels with a significant activation contrast for pleasant vs. neutral and unpleasant stimuli in both, the self-session and the other-session (shared activation)

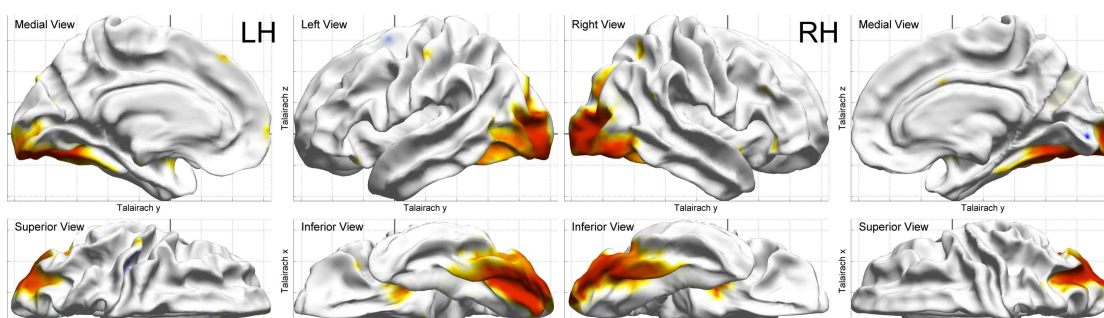


Figure 6.9: GLM analysis: Voxels with a significant activation contrast for unpleasant vs. neutral and pleasant stimuli in both, the self-session and the other-session (shared activation)

6.2 Classification Performance

It is crucial to validate the classifier’s decoding performance. A high classification performance implies that the classifier is able to capture the multivariate coding pattern of the data, which is a prerequisite of the method. The classification performance has to be checked for the in-session training and decoding and for the *model transfer* (training classifier on the data from one scanning session and using it to decode the data from the complementary scanning session). Finally it has to be answered which surface mesh resolution (FS4 with 2,562 nodes/hemisphere or FS5 with 10,242 nodes/hemisphere) is useful for the group-level classification.

The classifier is trained on $\mathbf{X}_{\text{self group}}$ and $\mathbf{X}_{\text{other group}}$ to decode the target vectors $\mathbf{y}_{\text{self group}}$ and $\mathbf{y}_{\text{other group}}$. Both models are used to decode the data from both sessions. For the experiments where the training session is the same as the session to decode, a 30-fold cross validation leaving out each subject once is used.

The classifier can be trained on the self-session or on the other-session. This model can be used to decode the emotions in either the self- or the other-session. This results in four possible combinations, which are all performed on FS4 resolution level. Because of the high computational effort for the FS5 resolution only two combinations are repeated in high-resolution.

For evaluation the following measures were analyzed in each experiment: For a simple comparison the over all **correct classification rate** (relative amount of correct classified stimuli) was evaluated. The correct classification rate is insufficient for a full description of the classification accuracy. Two evaluations are given for each classifier:

1. **Single subject class-sensitivities:** The classifier sensitivity for each class *neu-stim*, *ple-stim* and *unp-stim* (*neutral*, *pleasant*, *unpleasant*) was evaluated. These values tell how sensitive the classifier is for a certain class. The standard error (std.err.) is calculated for each class to estimate the stability of the sensitivity of each class across the group.
2. **Confusion Matrix:** The confusion matrix is evaluated for each experiment. The shown matrix depicts the accumulated values across all subjects. Colormap for relative values is consistent across all experiments.

Results: A compact overview on the correct classification rates is given in Table 6.2. No significant difference between FS4 and FS5 can be found, thus FS4 is used to reduce computational expense of all evaluations. The complete results are given in Appendix B.2. The performance of the finally used FS4 resolution is shown in Figure 6.10 – 6.13. Remarkably the decoding performance for decoding the emotions from the other-session is similar regardless whether the classifier is trained on the self-session (see Figure 6.14 b)) or on the other-session (see Figure 6.14 c)). This is a first indicator for a high amount of informative multivariate coding patterns which are shared across self-perception and observing others (empathy).

| Model training on | | | Self | Self | Other | Other |
|-------------------|--------------------|-------|------|-------|-------|-------|
| Decoding of | | | Self | Other | Other | Self |
| Res. | n_{trees} | Model | | | | |
| FS4 | 5000 | Group | 72% | 69% | 68% | 66% |
| FS5 | 20000 | Group | 71% | 71% | – | – |

Table 6.2: Overview on the group-model correct classification rates: Resolution (Res.). The colored experiment is the further used resolution.

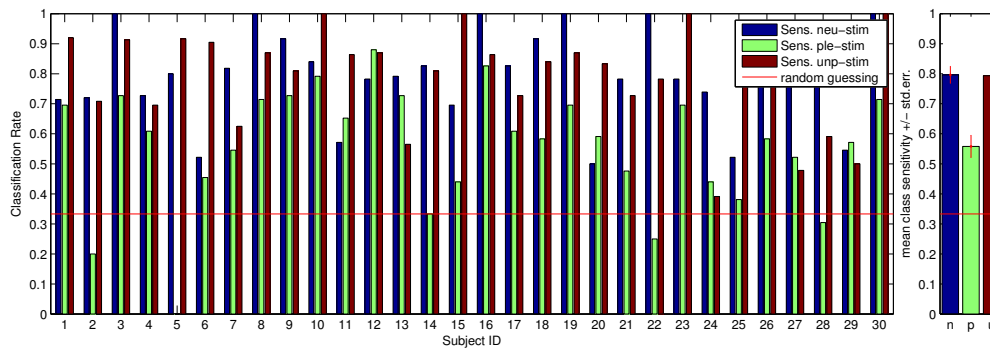


Figure 6.10: Self-model decoding self-session. Figure shows the class-sensitivities on the left out subject and mean class-sensitivity \pm std.err.

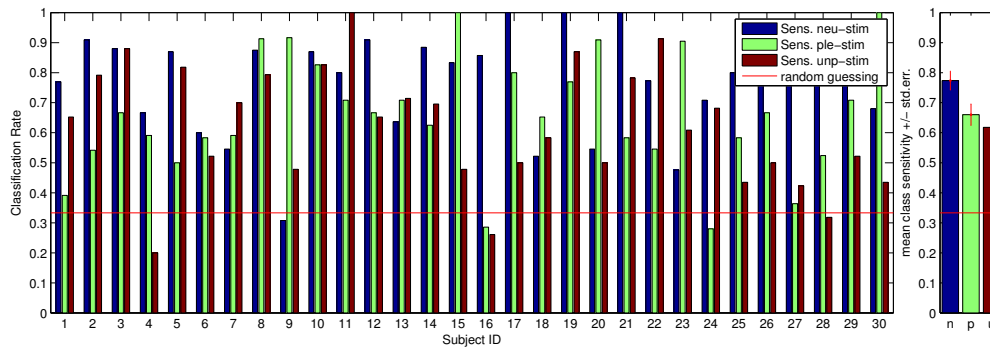


Figure 6.11: Self-model decoding other-session. Figure shows the class-sensitivities on the left out subject and mean class-sensitivity \pm std.err.

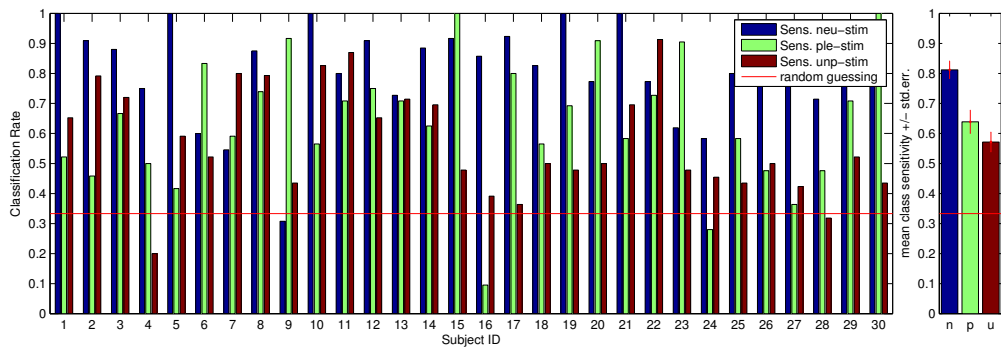


Figure 6.12: Other-model decoding other-session. Figure shows the class-sensitivities on the left out subject and mean class-sensitivity \pm std.err.

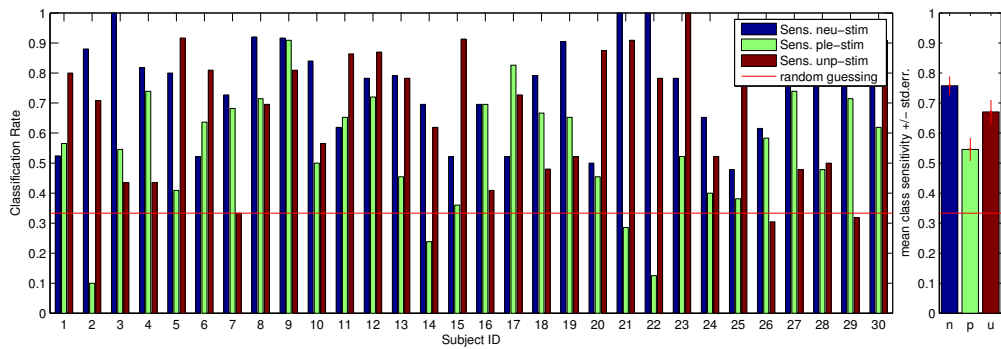


Figure 6.13: Other-model decoding self-session. Figure shows the class-sensitivities on the left out subject and mean class-sensitivity \pm std.err.

6.3 Comparison of Variable Importance Measures

Variable importance measures are used to detect the informative voxels. The behavior of these measures is crucial for interpreting the results from the MVPA.

6.3.1 Importance of Voxels Within a Stimulus Category

The RF-algorithm used [57] outputs two variable measures after training:

- **Gini Importance:** from Gini-split criteria in tree-learning, not class-specific
- **Permutation Importance:** during training on OOB-samples, class-specific

After training, the following variable importance measure can be assessed:

- **Modified Gini Importance:** on all samples, class-specific
- **Permutation Importance:** calculation as during training, but on all test samples

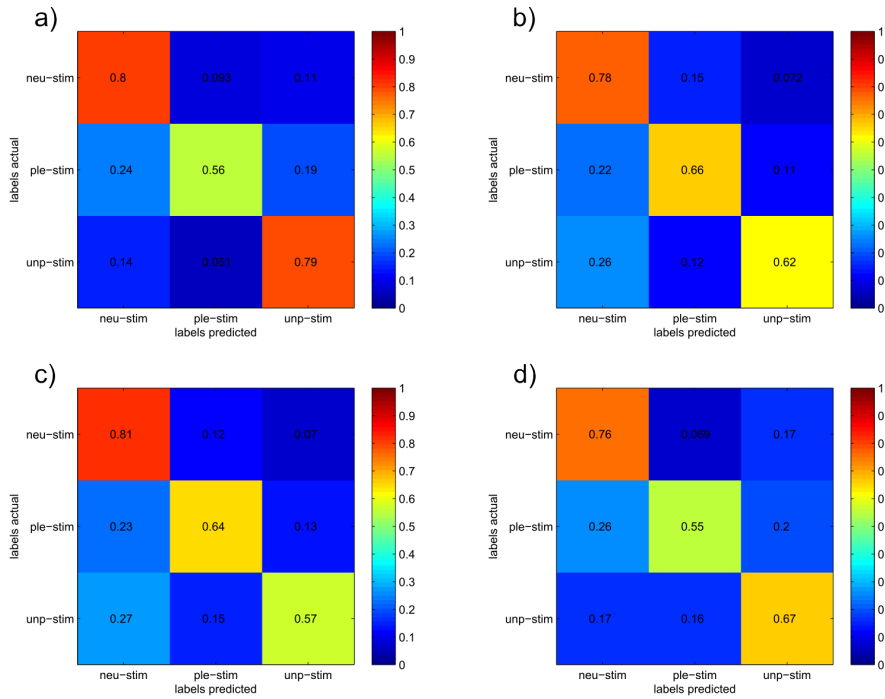


Figure 6.14: Confusion matrices: a) self-model decoding self-session; b) self-model decoding other-session, c) other-model decoding other-session and d) other-model decoding self-session

In the following, experiments are conducted to investigate their behavior. Further the labels are randomly permuted to break the dependency between data and labels. In this case the training data should not be predictive for the permuted response vector. These experiments show the behavior of the variable importance measures for the „non predictive case“.

The full data $X_{\text{self group}}$ and $y_{\text{self group}}$ are used to train the RF-classifier. For the „non predictive case“ a permuted response vector $\widetilde{y}_{\text{self group}}$ is used for variable importance assessment.

All class-specific importance measures are evaluated: During training the **permutation importance** on the OOB-samples is assessed. After training the **modified Gini importance** and the **permutation importance** are assessed on all training samples $X_{\text{self group}}$ and the original labels $y_{\text{self group}}$ and on all training samples $X_{\text{self group}}$ and the permuted vector $\widetilde{y}_{\text{self group}}$.

As explained in Section 5.5, the only **class-specific** and **unbiased** (unbiased because assessed on OOB-samples) importance measure is the permutation importance. Thus it is used as reference method. For evaluation scatterplots are drawn for each stimulus type, where each dot represents one feature. The abscissa coordinate value is the importance of the reference method (permutation importance on OOB-samples); the ordinate value is the importance of the measure under comparison. Four comparison experiments are conducted:

1. The permutation importance on the OOB–samples is compared to the permutation importance when evaluated on all samples. See Figure 6.17.
2. By randomly permutating the labels, the dependency between labels and data is broken. Thus the variables should not be important (predictive) for the labels. Performing the same experiment as in 1.), but with permuted labels, the importance for the „non predictive case“ can be evaluated. See Figure 6.16.
3. The modified class specific Gini Importance is compared to the class specific OOB–permutation importance. See Figure 6.17.
4. Again the behavior of the Gini Importance is evaluated for the non predictive case. See Figure 6.18.

Results: For some features, especially for the case of pleasant stimuli, including the in–bag samples in the permutation importance evaluation (see Figure 6.15) results in higher variable importances ranks than in the OOB–evaluation. This can also be seen in the Gini importance evaluation in Figure 6.17, this bias was expected from theoretical considerations (see Section 5.5). In the non–predictive case both variable measures show a dependency of the variable importance in the „non predictive case“ to the variable importance on the original data (see Figure 6.16 and 6.18).

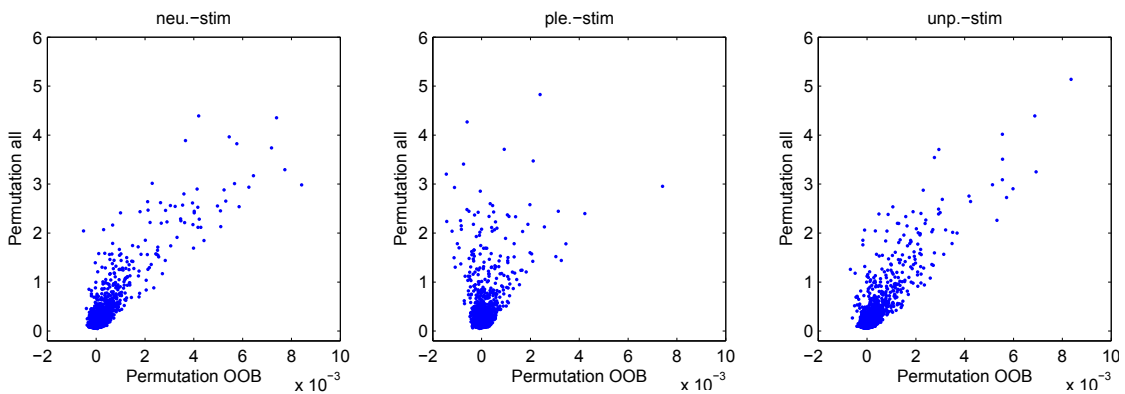


Figure 6.15: Permutation importance on OOB–samples compared to permutation importance on all samples

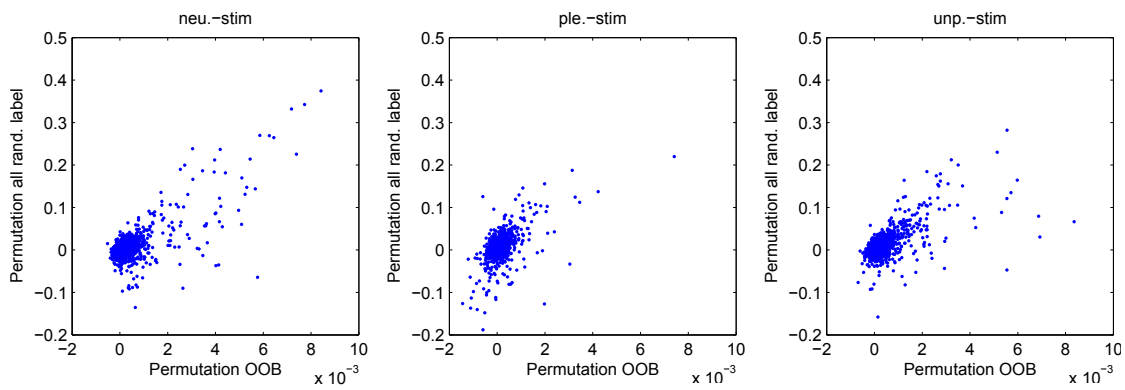


Figure 6.16: Permutation importance on OOB-samples compared to permutation importance on all samples with permuted labels

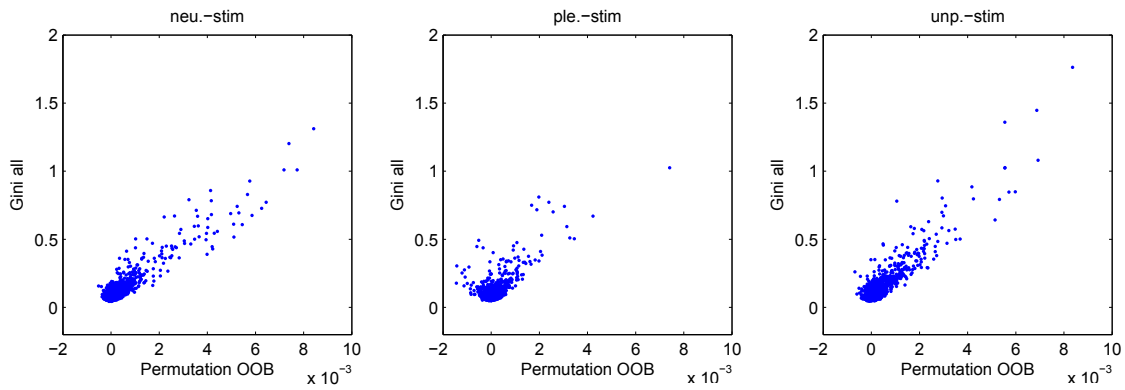


Figure 6.17: Permutation importance on OOB-samples compared to modified Gini importance on all samples

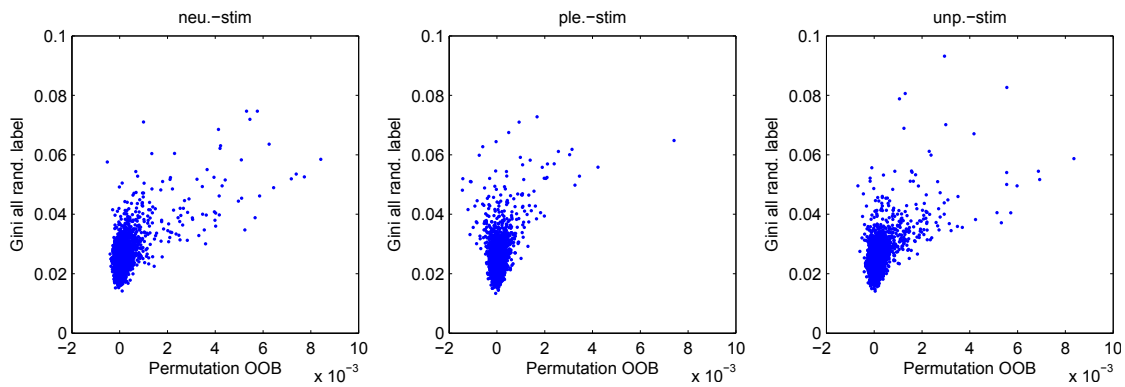


Figure 6.18: Permutation importance on OOB-samples compared to modified Gini importance on all samples with permuted labels

6.3.2 Importance of Voxels Shared Across Stimulus Category

For the „model transfer“ (training the classifier on one session and decoding the data from complementary session) the importance of the variables can be assessed in two procedures when decoding the data from the complementary session.

- **Modified Gini Importance:** on all samples, class-specific (see Section 5.5)
- **Permutation Importance:** calculation as during training, but on all test samples

In the following, experiments are conducted to investigate the relation between these two variable measures. Further the labels are randomly permuted to break the dependency between data and labels. In this case the data from the complementary session, the session to be decoded, should not be predictive for the permuted response vector. These experiments show the behavior of the variable importance measures for the „non predictive case“.

For training the full data from the self-session $\mathbf{X}_{\text{self group}}$ and the response vector $\mathbf{y}_{\text{self group}}$ is used. The trained classifier is used to decode the data from the other-session $\mathbf{X}_{\text{other group}}$. For the variable importance assessment the original target vector $\mathbf{y}_{\text{other group}}$ and permuted vector $\widetilde{\mathbf{y}}_{\text{other group}}$ are used.

After training the model on the self-session data, the importance measures are evaluated on the data from the other-session. Data are visualized in a scatter plot, with each point representing one feature, dot coordinates are the feature's importances. Four experiments are conducted in this experiment:

1. The permutation importance is compared with the class specific modified Gini importance. See Figure 6.19 upper row. For demonstration, the permutation importance is compared with the Gini importance for incongruent classes. See Figure 6.19 lower row.
2. The permutation importance is compared with the permutation importance in the „non predictive case“, where the labels are randomly permuted before the variable importance assessment. See Figure 6.20.
3. The Gini importance is compared with the Gini importance in the „non predictive case“. See Figure 6.21.

Results: The comparison of modified Gini importance, which is introduced in this thesis, with the permutation importance show highly correlated feature importance values (see Figure 6.19 upper row). For demonstration, the importances are plotted across inconsistent stimulus types, which reveals that some features are predictive for one stimulus type, but not for the other (see Figure 6.19 lower row). Comparing the variable importance measure for the „unpredictive case“ with the original case again (as in Section 6.3.1) a dependency between both importance measures can be found: For the permutation importance this can especially be seen for neutral stimuli (leftmost plot in Figure 6.20); for the modified Gini importance the dependency is visible in all categories (see Figure 6.21).

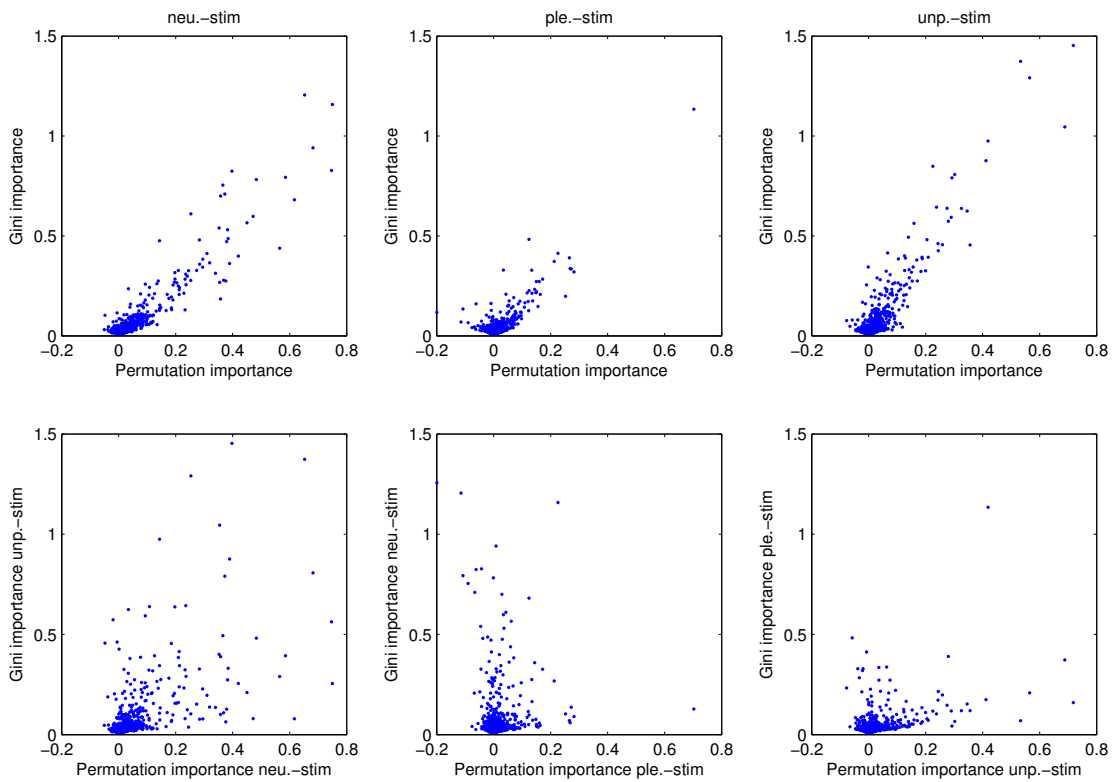


Figure 6.19: Upper row: Permutation importance compared with the modified class-specific Gini importance. Lower Row: For demonstration, permutation importance and the modified class-specific Gini importance are compared across inconsistent classes.

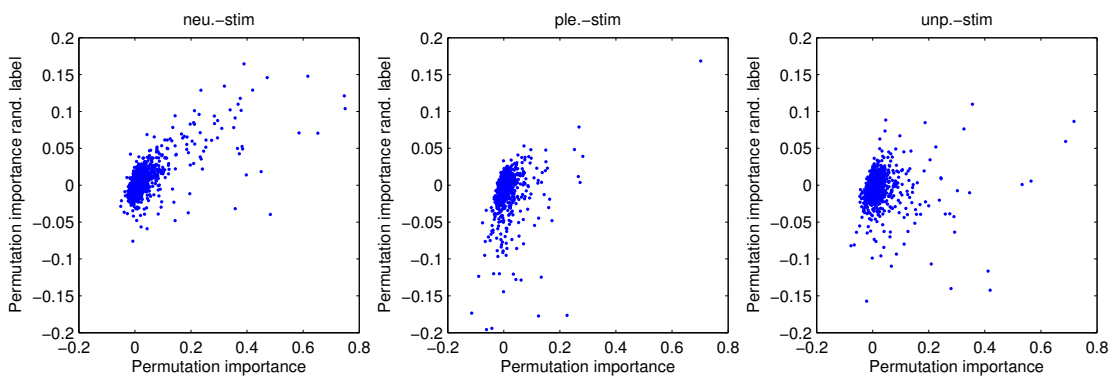


Figure 6.20: Permutation importance is compared with the permutation importance for randomly permuted features.

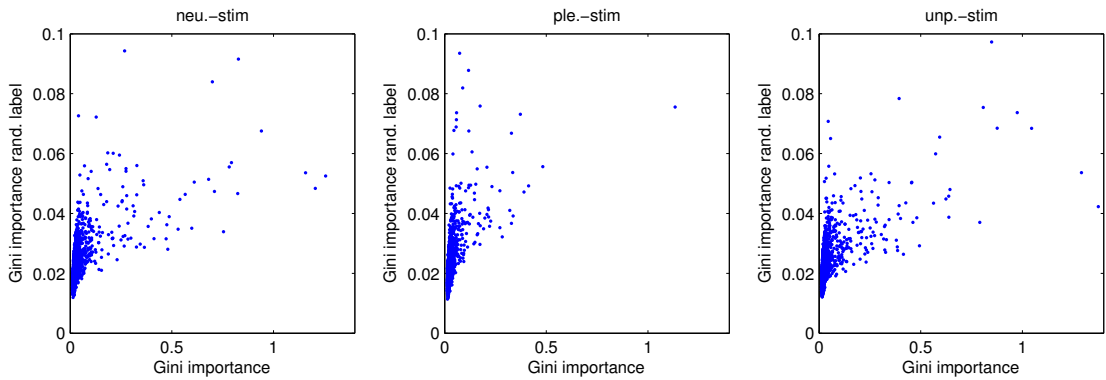


Figure 6.21: Gini importance is compared with the Gini importance for randomly permuted features.

6.4 Visualization of Detected Informative Voxels

In the following, the location of predictive features is visualized on the cortical surface. Since from the training model the only reliable class-specific importance measure is the OOB permutation importance (see Section 7.2 for a detailed discussion), the permutation importance was used for visualization.

Outline of the Plots: Colormap is set as following: three colormaps are used (see Table 6.3), the colormap is chosen according to the correlation with the stimulus, to see if the areas are activated, inhibited or in an undefinable (but predictive) state during the corresponding stimulus. The color-saturation is scaled with the importances. A saturated color means high importance, a light color depicts weaker importances. To distinguish between important and unimportant importance values a threshold value was chosen. This value was set to the 95-percentile from the distribution of the importances of the variable importance when assessed with permuted labels in the „unpredictive case“¹. Technical properties in rendering and visualization as well as structure of the views is exactly as described in the beginning of this chapter.




| Colormap | Correlation |
|---|---|
|  | $0.1 \cdot \max(c) < c < \max(c)$ |
|  | $0.1 \cdot \min(c) < c < 0.1 \cdot \max(c)$ |
|  | $\min(c) < c < 0.1 \cdot \min(c)$ |

Table 6.3: Assignment of the colormaps: The appropriate colormap is chosen according to the value of c according to the list above. c is the correlation between the voxel’s BOLD response z and the estimated hemodynamic response h . h is either h_1 , h_2 or h_3 according to the stimulus type of investigation.

¹See section 7.2 for critical discussion on that strategy

6.4.1 Voxels Informative Within a Stimulus Category

The purpose of this procedure is to visualize the location of voxels being informative in decoding the emotional state.

The detection for informative voxels is done in both sessions. $\mathbf{X}_{\text{self group}}$ and $\mathbf{y}_{\text{self group}}$ is used for training the classifier for the self-session and $\mathbf{X}_{\text{other group}}$ and $\mathbf{y}_{\text{other group}}$ is used for training the classifier for the other-session.

During training the OOB-permutation importance was assessed to detect the informative voxels in each session.

Importance values were visualized on renderings of the cortical surface. Visualization was performed as given in the beginning of this section.

Results: In general, the voxels marked as being informative are highly consistent to the activated voxels in the GLM-analysis (see Section 6.1). This is true for the self-session (Figure 6.22 – 6.24) and for the informative voxels in the other-session (Figure 6.25 – 6.27). Differences will be discussed in conjunction with other evaluations (see Section 7.3 for the discussion). Note: The difference in the coloring comes from the fact that the GLM outputs the activation contrast (weaker or stronger activated than in the other stimulus types), whereas the colormap of the MVPA-visualizations depicts the correlation with the corresponding stimuli.

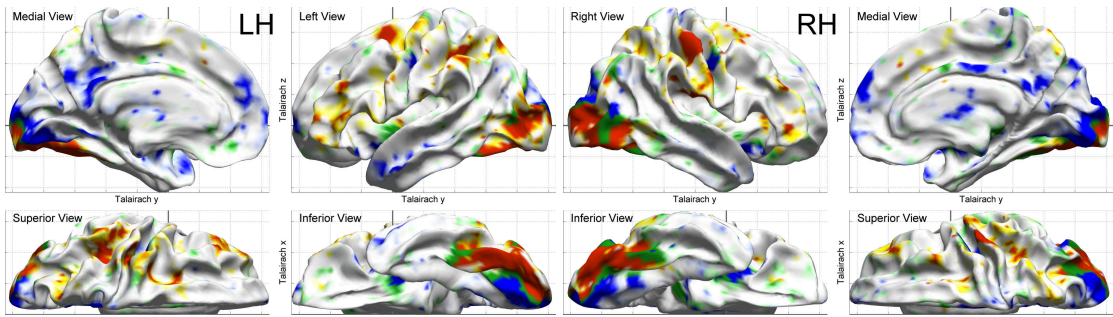


Figure 6.22: MVPA: Important features for identification of neutral stimuli in the self-session, RF is trained on the group self-data

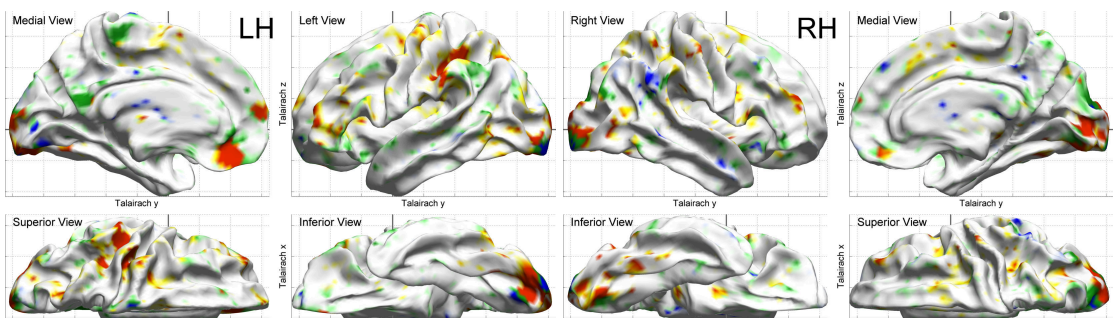


Figure 6.23: MVPA: Important features for identification of pleasant stimuli in the self-session, RF is trained on the group self-data

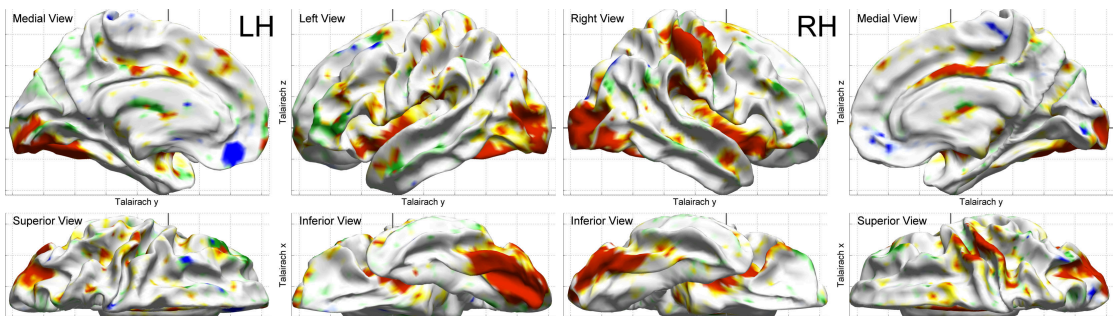


Figure 6.24: MVPA: Important features for identification of unpleasant stimuli in the self-session, RF is trained on the group self-data

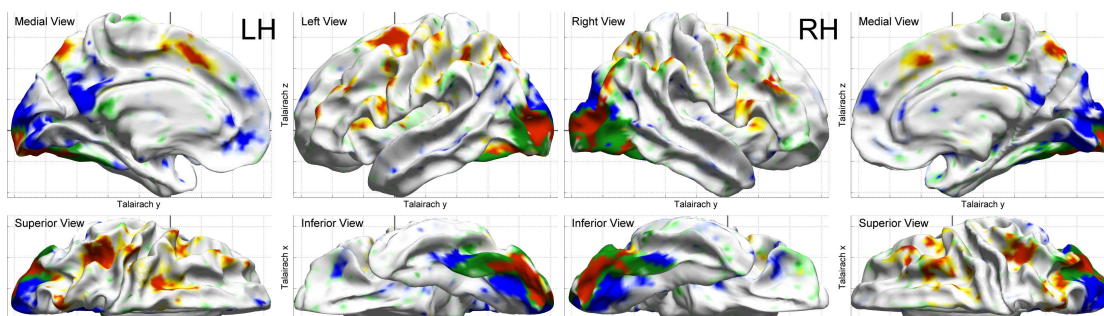


Figure 6.25: MVPA: Important features for identification of neutral stimuli in the other-session, RF is trained on the group other-data

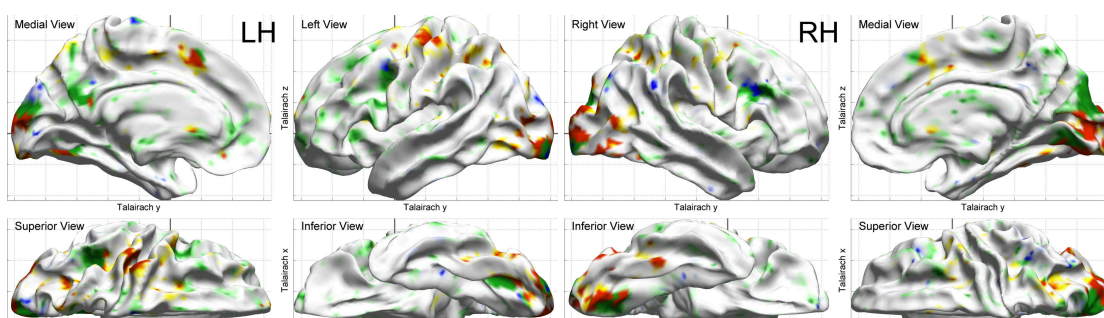


Figure 6.26: MVPA: Important features for identification of pleasant stimuli in the other-session, RF is trained on the group other-data

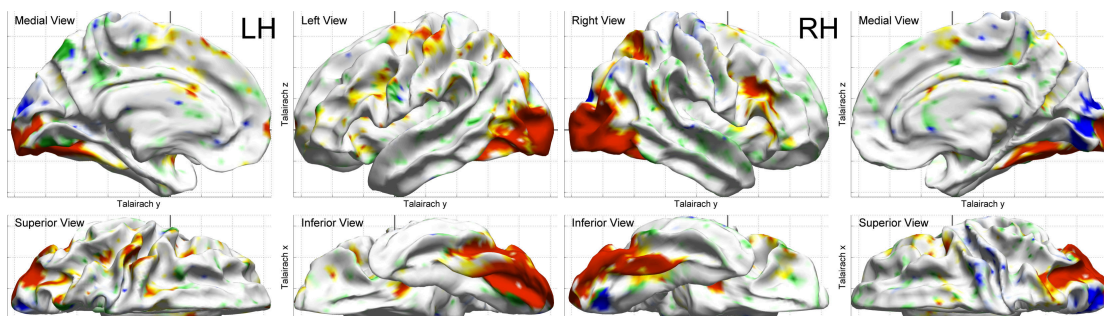


Figure 6.27: MVPA: Important features for identification of unpleasant stimuli in the other-session, RF is trained on the group other-data

6.4.2 Voxels Informative Shared Across Stimulus Category

The aim of this experiment is to visualize the location of voxels part of distributed multivariate coding patterns which are shared across two self-perception (self-session) and observing others (other-session). These are the results from the first main contribution of this work as explained in Section 5.1.1.

The classifier training is done separately on both sessions. For the self-session $\mathbf{X}_{\text{self group}}$ and $\mathbf{y}_{\text{self group}}$ is used for training the classifier. $\mathbf{X}_{\text{other group}}$ and $\mathbf{y}_{\text{other group}}$ is used for training the second model.

The model trained on the $\mathbf{X}_{\text{self group}}$ data to decode $\mathbf{y}_{\text{self group}}$ is named „self-model“. The training on the $\mathbf{X}_{\text{other group}}$ data to decode $\mathbf{y}_{\text{other group}}$ is named „other-model“. After training the „self-model“ was used to decode $\mathbf{X}_{\text{other group}}$ and the „other-model“ was used to decode $\mathbf{X}_{\text{self group}}$. At both decoding procedures the permutation importance was assessed to detect the informative voxels.

Importance values are visualized on renderings of the cortical surface. Visualization is performed as given in the beginning of this section. With that the „self-model“ decoding the data from the other-session reveals which of the coding patterns for the self-session are also informative in decoding the data from the other-session. These patterns are shared across both sessions. The same considerations can be made for the „other-model“ decoding the data from the self-session.

Results: The method reveals that a high amount of the coding patterns from the „self-model“ are also informative for decoding the emotional state from the other-session (see Figure 6.28 – 6.30), this is also true for the „model transfer“ in the other direction (see Figure 6.31 – 6.33). The highest difference between both experiments can be seen in the missing detected activation on the S1 for neutral- (Figure 6.31) and unpleasant stimuli (Figure 6.33) in the „other-model“ decoding self-session experiment compared to the experiment „self-model“ decoding other-session. This issue is discussed in detail in Section 7.3.3.

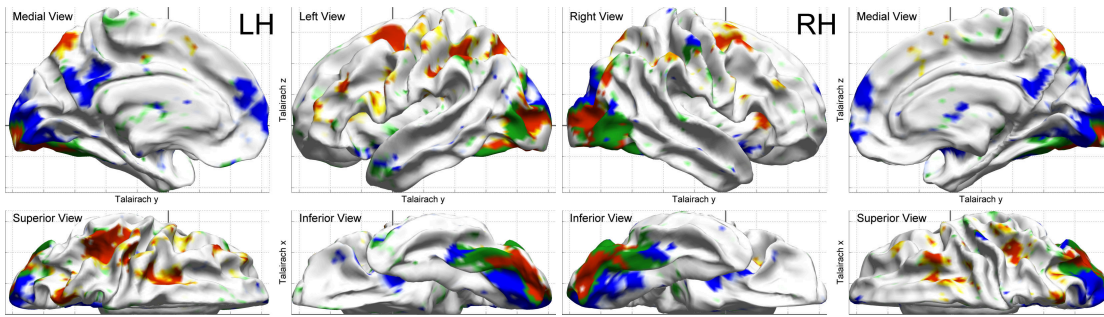


Figure 6.28: MVPA: Important features for identification of neutral stimuli in the other-session, RF is trained on the group self-data (shared patterns for neutral stimuli)

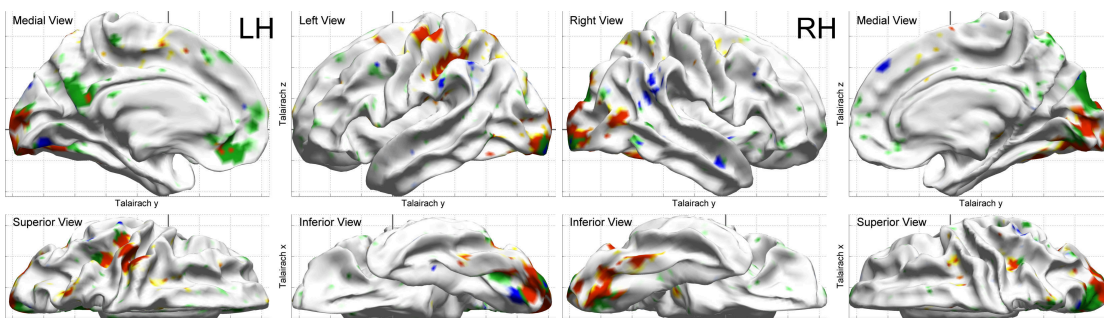


Figure 6.29: MVPA: Important features for identification of pleasant stimuli in the other-session, RF is trained on the group self-data (shared patterns for pleasant stimuli)

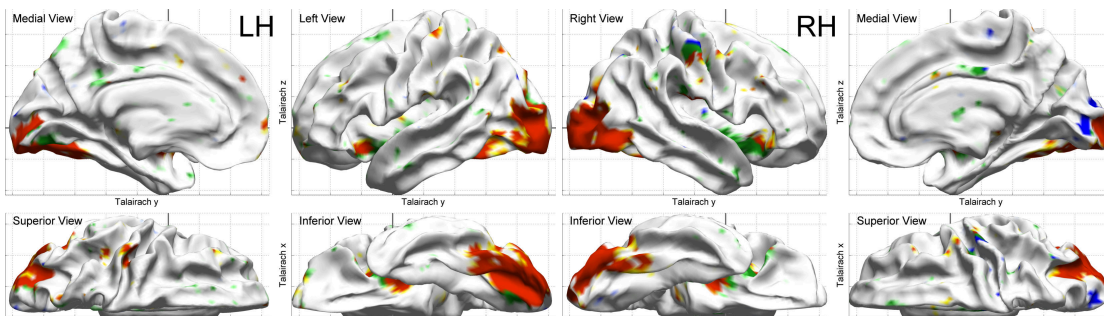


Figure 6.30: MVPA: Important features for identification of unpleasant stimuli in the other-session, RF is trained on the group self-data (shared patterns for unpleasant stimuli)

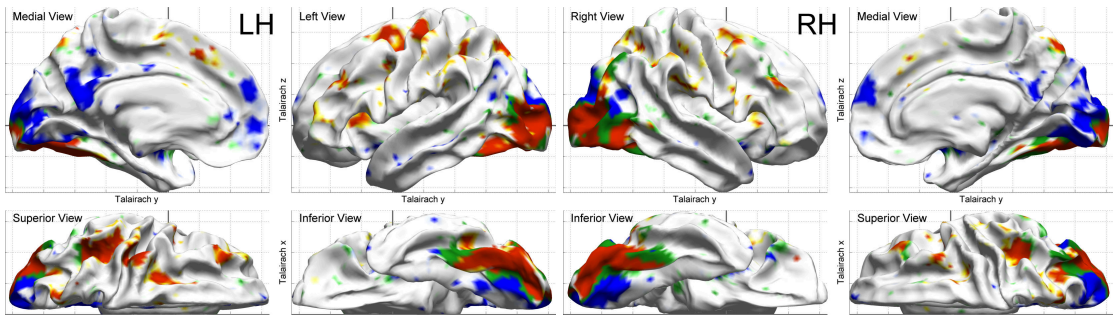


Figure 6.31: MVPA: Important features for identification of neutral stimuli in the self-session, RF is trained on the group other-data (shared patterns for neutral stimuli)

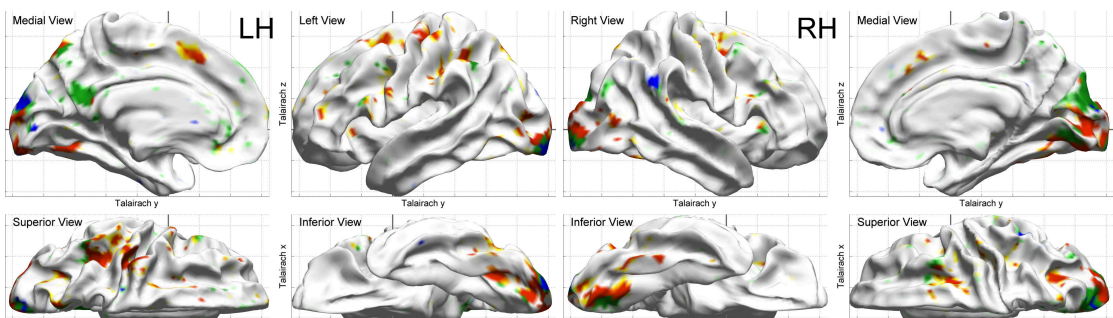


Figure 6.32: MVPA: Important features for identification of pleasant stimuli in the self-session, RF is trained on the group other-data (shared patterns for pleasant stimuli)

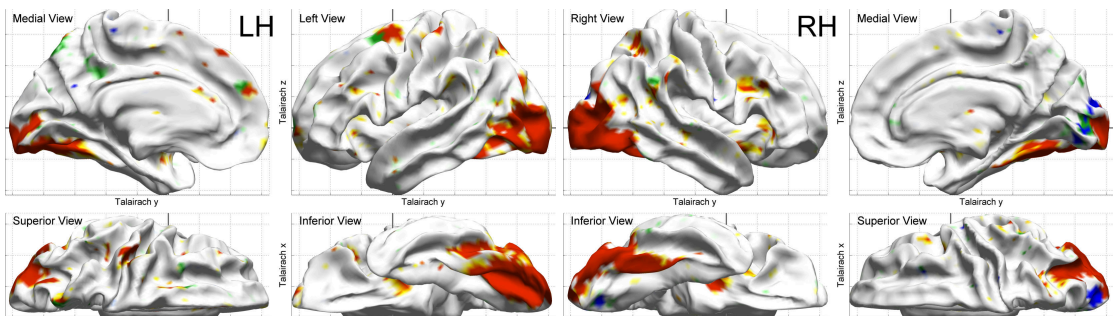


Figure 6.33: MVPA: Important features for identification of unpleasant stimuli in the self-session, RF is trained on the group other-data (shared patterns for unpleasant stimuli)

6.5 Identifying Informative Cliques of Voxles

When comparing the single variable importance from the previous section with the GLM–results, a high consistency can be seen (see discussion in Section 7.3). This can be expected, since the multivariate–approach should also be sensitive to the univariate features. For better revealing the differences, the results in this section focus only on the multivariate properties. For that, the newly introduced Conditional Joint Gini importance is used. This is the second main contribution of this work as explained in Section 5.1.2.

Importance of Long Range Connections: In contrast to the searchlight algorithm the proposed method here is also capable of capturing long–ranged connections. Thus the newly introduced importance measure first is used to answer the question if long–ranged connections are important.

Importances of Joint Parcels: Further the spatial organization of the informative long–ranged connections are of interest. As outlined in Section 5.7.2, the full connection matrix comprises more than 26 million possible connections. Thus the joint importances were pooled across all connections between two defined brain–parcels. This reduction of the amount of connections make a visualization possible.

Training was done on the self–session data ($\mathbf{X}_{\text{self group}}, \mathbf{y}_{\text{self group}}$). For the „*model transfer*“ this trained classifier was used to decode the other–session data $\mathbf{X}_{\text{other group}}$. The model transfer was only done in one direction:

After training the „self model“ the Conditional Joint Gini importance was evaluated for the self–session data and for the „model transfer“ on the data from the „other–session“.

To check if the long–ranged connections are of interest for classification, all positive Conditional Joint Gini Importances is plotted in a scatter plot according to their euclidean distance. Due to the high variation of dot density in the scatter plot the method for visualization was adapted from [28].

For visualization of the spatial organization, the voxel–pair’s joint importance was pooled across the parcels as explained in Section 5.7.2. The informativeness by connecting the data from two voxels being part of two different parcels is depicted in a schemaball, where bright yellow lines indicate a high information content by compaining voxels from the two abjected parcels. The schemaball is given for all three stimulus types and for the self–model decoding the self–session and the self–model decoding the other–session (shared patterns).

Results: The distribution of the conditional joint Gini importance from a pair of features is similar for all stimulus types. Thus only one exemplary plot for decoding neutral stimuli is given in Figure 6.34.

One exemplary **schemaball**² which will further be used in the discussion is given in the Figure 6.35; all available schema balls are given in the appendix in Figure C.1–C.6. The parcels are ordered according to their median y –coordinate: from the bottom of the diagram with the most

²Source code for visualization was taken from Oleg Komarov [66] and adapted for fitting the problem.

occipital parcels to the top with the most frontal parcels. Parcels from the left hemisphere are ordered in the left half, and the right hemisphere is ordered in the right half of the circle. Parcels are annotated with the name of the anatomical parcel according to the Destrieux atlas [24], with which they show the highest overlap. Parcels which have the highest connectedness to other parcels are marked with colored dots, where dark-red is the highest connected parcel and dark-blue the last parcel of the selection. In the brain surface plot below the schemaball the high connected parcels are marked in the corresponding color.

Additional information for correct interpretation of the diagrams: These diagrams deliver a **complementary** information to the importance visualization in Section 6.4.1 and 6.4.2, where the importance visualization depicts the importance, regardless if the voxels are informative on their own or in conjunction with other voxels. Due to the conception of the conditional joint Gini importance the plots in this sections just depict the important features which are impotent in conjunction with other voxles. Voxels informative on their own will not be visible in this plot due to the normalization done by Equation (5.8).

Example reading the diagram: For a better understanding of the plots in the next six pages an example on interpretation is given here on Figure 6.35. Exemplary the connection between *L. G orbital* (market red on the top) and *R. S temporal sup* (marked yellow on the bottom) is taken. For finding the parcel *L. G orbital* one has to search in the four leftmost plots on the bottom of Figure 6.35, since it is located on the left hemisphere. Since the parcel is on the top, the corresponding red parcel is located in frontal direction. It is best visible on the left side of the inferior view of the LH. *R. S temporal sup* has to be found on the right hemisphere. The bright yellow line between the two parcels tells us now that many features in *L. G orbital* are informative in conjunction with features in *R. S temporal sup*.

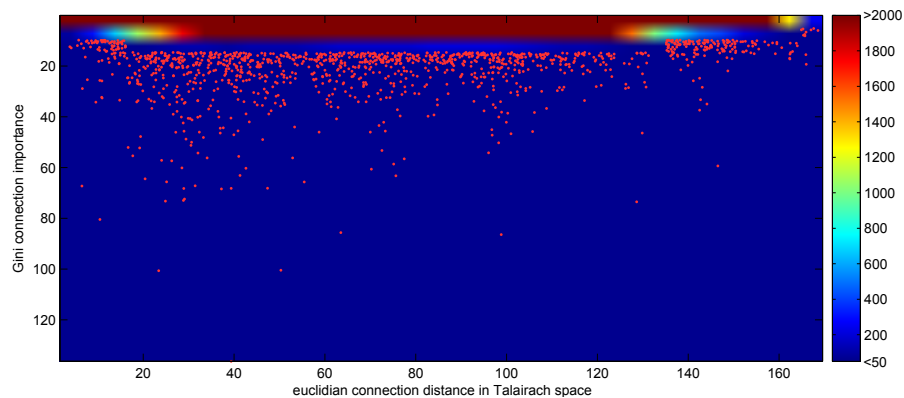


Figure 6.34: Scatter plot of conditional joint Gini importance vs. the euclidean distance between the two connected variables. If the scatter plot density exceeds a density of 50 points per $5\text{mm} \times 5$, the dot density is color coded according to the color scale. Example taken from the Joint importance for decoding neutral stimuli in the self-session data.

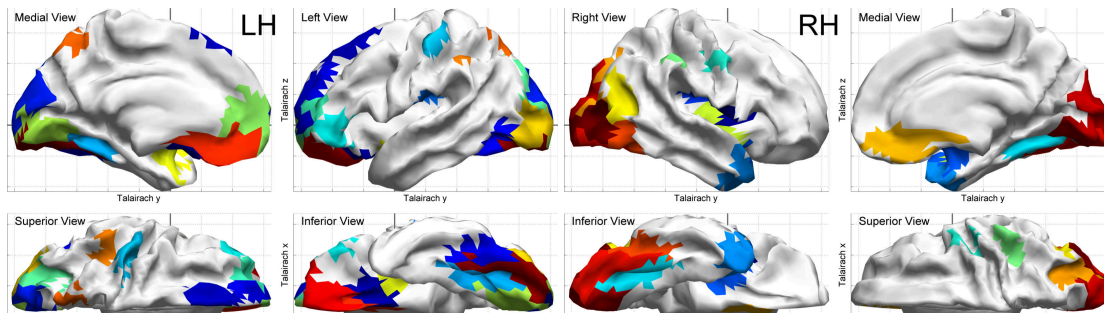
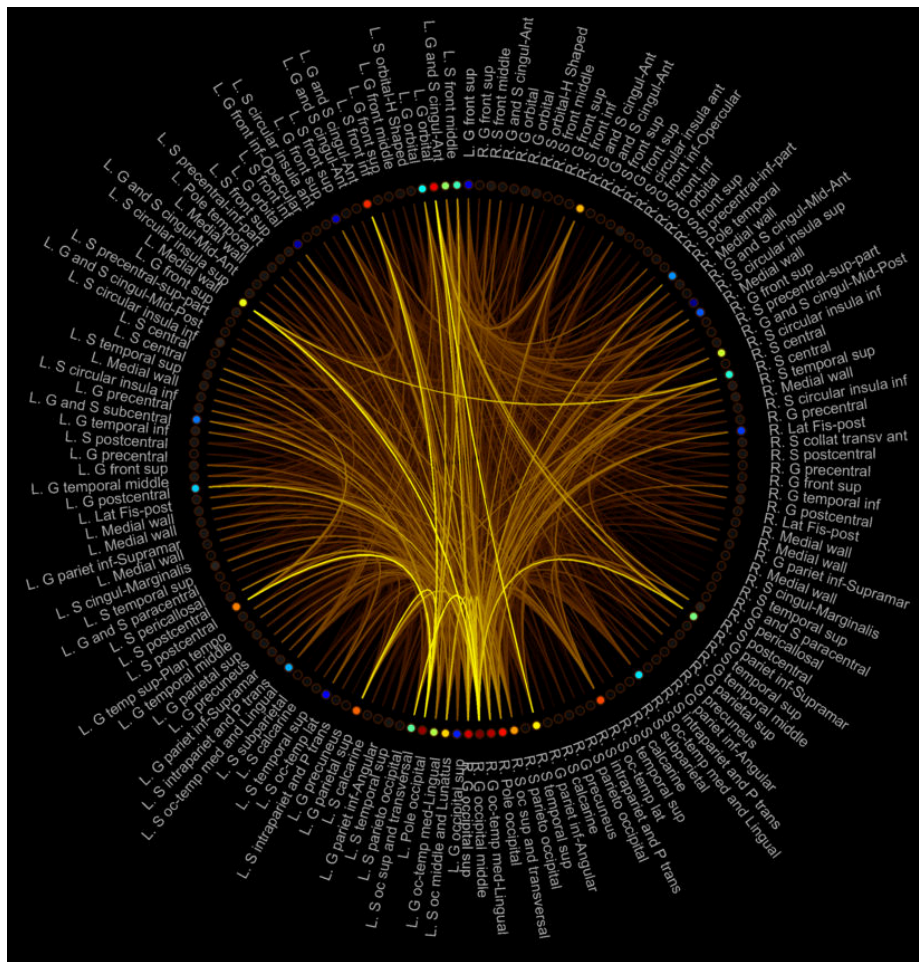


Figure 6.35: MVPA: Importance of joints of parcels for decoding unpleasant stimuli in the self-session, RF is trained on self-data; plots below depict location of highly connected parcels

6.6 Summary of Main Results

The classification performance evaluation reveals that the classifiers are able to successfully decode the data from the session they are trained on, as well as the data from the complimentary session. The modified Gini importance and the permutation importance in general show a high consistency. The comparison of the importance values with the „unpredictive case“ (target vector is randomly permuted) allows to distinguish between important and unimportant features, where a slight dependence between the importance in the unpredictable case and the importance in the original (predictive) case is prohibitive for a sharp distinction between important and unimportant features. The MVPA reveals a higher amount of shared neural representation than the GLM. The importance of cliques of voxels show that informative multivariate patterns are not restricted to local areas.

Discussion

All experiments' results are discussed in this chapter in detail: At first the validity and behavior of the decoding of the emotional states is discussed. Then the experiments of Section 6.3.1 and 6.3.2 are discussed to study the behavior of the investigated variable importance measures. In the following the results from visualizing the variable importance of the proposed method is compared with the results from classical GLM-analysis, to point out the benefits of the proposed method. The key novelty, the detection of distributed shared coding patterns across a stimulus category, is further investigated by usage of the newly introduced measure for the importance of pairs of voxels. Here the discussion focuses on how to interpret these data as the joint importance should not be confounded with the terminus „connectivity“ which is a term used differently in neuroscience.

The sections are comprised of questions which are answered in a discussion and summarized in a short conclusion. General considerations and remarks are given at the beginning of the sections.

7.1 Validity of the Classifier Models

Beside the decoding methods described in Chapter 5, an alternative way of preparing the data and labels as well as a classifier training on a single subject level is tested. The alternative methods described in Appendix A perform worse than the finally chosen approach, all results are summarized in Table B.1. Since single-subject analysis are common in MVPA [85, 103]. The validity of the single-subject models is discussed here:

Validity of single-subject models

The single-subject classification performs substantially worse than the group analysis. The fMRI-recordings are too short (≈ 6 min 20 s), providing too few samples (10 stimuli of each class). This also explains why the model-transfer (self-training, empathy-decoding) performs

better than the in-session decoding: Due to the permutation schema, see Figure A.2, the training data are reduced to $\frac{3}{4}$ in the 4-fold crossvalidation. Since self- and empathy-session are independent measurements, $\frac{4}{4}$ of the self-session can be used for training the decoder for the empathy-session.

Validity of group models

The group model delivers substantially better results than the single-subject models and the performance is well above chance. For comparison no similar experiments are known in literature. But the results are good compared to the expectations given for a group-level decoding in a recent MVPA reviews. **Conclusion:** The RF-classifier is a useful approach for brain-pattern decoding. The group-level decoding may profit from the bagging strategy, where features are selected from a random subsample of all features. As outlined in Section 3.4.3 this permits features with correlated but weaker predictability also to be incorporated in the decision trees. In FS4 features are spaced approximately in 5.5mm distance. This is the order of alignment error in a surface based group alignment [67]. Thus a single feature in the group-model cannot be expected to be on a consistent surface-node in all subjects. From that point it is useful to work with an ensemble of correlated features rather than with the single most predictive feature, which may not be consistently aligned across all subjects.

What do the confusion matrices tell us?

Here the four procedures of combining the model-training and data to be decoded are discussed. **(S→S)** Model training on the self-data and decoding of the self-data: This procedure has the highest correct classification rate. The class-specific correct classification rates reveal that the classifier has the highest problem in decoding the pleasant stimuli. The class sensitivity is equally high for neutral and unpleasant stimuli.

(S→O) Model training on the self-data and decoding of the other-data: When comparing the decoding performance of S→S with S→O, there is a noticeable preference for neutral-stimulus. A possible explanation is that the decoding of neutral stimuli may be done by finding an absence of emotional content (see discussion further below in Section 7.3.1). If the emotions are weaker in the empathic perception, the stimuli may more likely be labeled as „neutral“.

(O→O) Model training on the other-data and decoding of the other-data: In this case the same behavior as for S→O can be observed: the neutral-stimuli are the preferred classifier output. Remarkably the training on the self-session delivers equally high decoding performance for the other-data, as training on the other-data (performance S→O = performance O→O). This is a first indication that there is a high amount of shared neuronal representation between self perception and empathic perception.

(O→S) Model training on the other-data and decoding of the self-data: In contrast to O→O, this scenario shows the same decoding behavior as S→S: Exactly as in S→S, pleasant stimuli are worse to be detected.

Conclusion: The decoding behavior depends on the data and is independent from the model. Pleasant stimuli are badly detected compared to the other stimulus types in the self-data (S→S as well as O→S poor detection of pleasant stimuli). In the other-data the decoder is biased to neutral stimuli, regardless on which data the decoder is trained (S→O as well as O→O pref-

erence for predicting neutral stimuli). Yet there is no explanation for the first phenomena, the latter observation is likely to arise from a decrease in „emotionality“ in the other-session, meaning that the emotions are stronger if participants feel the tactile stimuli themselves. The model transfer $S \rightarrow O$ works better than the $O \rightarrow S$ transfer, which suggests that the neuronal pattern in the other-data is more fuzzy and thus it is harder to train the classifier on this session.

What resolution to choose?

Comparing the results from FS4 and FS5 no significant change in classification performance is visible. This is to be expected for the cortical alignment errors given in literature [67]. Due to the lacking amount of training samples a more detailed fine-grained pattern cannot be found in the single-subject models. The RF-classifier is able to handle the dimensionality of the FS5-atlas (>20k features) without deterioration of the classification performance. Due to group-alignment errors the resolution of the FS4-atlas (>5k features) is sufficient to depict the pattern on a group-level scale. To evaluate if the proposed method is capable of capturing a more fine-grained pattern, further experiments namely high resolution fMRI-scans as well as longer scanning sessions are necessary for this purpose. **Conclusion:** From the experiments carried out FS4 is a useful resolution for a group level analysis, which also corresponds to the accuracy of a cortical group-alignment.

What experimental setup is needed for group decoding?

The group model trained on 29 subjects is able to decode emotional states on never seen subjects. A cortical alignment and sampling on 2,562 surface nodes per hemisphere is sufficient. The computational effort for training a group model on 30 subjects in FS4 with 3,000 trees needs approximately 16 minutes on a 64bit system with 76GB RAM parallelized on 8 cores of Intel Xenon X5650 processors with 2.67 GHz clock frequency. This is handy compared to training 20,000 trees on FS5 which takes about 3 hours. Compared to other studies recordings of 6 to 7 min per participant and session are very convenient and sufficient for training a classifier in a group model. **Conclusion:** In contrast to the poor performance of the single-subject models, the experiments showed that the short scanning trails on 30 subjects are sufficient for a decoding on a group level.

7.2 Discussion on the Variable Importance Measures

How do in-bag samples influence the permutation importance compared to OOB-assessment?

In Figure 6.15 the permutation importance of the OOB-assessment during training is compared with the permutation importance assessed on all samples. For the two classes with high predictability the two measures are consistent. For the *unpleasant stimuli*, which have a class specific prediction accuracy about 15% lower than for the two other classes, permutation importance on all samples frequently shows high importance value for variables which are unimportant in the OOB-assessment. This phenomenon is not unexpected: during training some variables could have been chosen for the best split variable for the in-bag samples just because of a random noise, this (random) split is not predictive for the OOB evaluation, thus the selected split variable is not important. **Conclusion:** Despite of the fact that all-sample and OOB-sample eval-

uation deliver consistent values, for the variable importance in neutral and unpleasant stimuli the OOB-concept is the better choice because it cannot be biased by noise, which occasionally can serve as a good split condition. If no independent test-set is available for variable importance assessment, one should stick to the OOB-importance estimation evaluated during training.

How does the permutation importance behave compared to Gini importance?

This comparison is done in literature frequently, for that see the depiction on the state-of-the-art in RF-variable-importance measures in Section 3.4.2. Summarizing Gini importance and permutation importance are known to deliver well comparable variable rankings, where the permutation importance is known to spread the variable measure more uniformly [49]. It has to be noted that in literature the Gini importance from RF training is used, which per se is not class specific. In this work a modified class-specific Gini importance was introduced in Section 5.5. To evaluate the validity of this concept a comparison with the permutation importance was made: Modified Gini- and permutation importance is plotted for each feature across consistent classes in Figure 6.19 upper trace and across inconsistent classes in Figure 6.19 lower trace. The plots in Figure 6.19 lower trace are more scattered than the plots in Figure 6.19 upper trace, where the dots would ideally align along a line if there is a linear dependency between both importance measures. **Conclusion:** In consistency with literature Gini-importance and permutation importance deliver similar results where the newly introduced modified class-specific Gini importance is shown to be a usable modification. In this thesis furthermore the permutation importance was used since during training only this measure is evaluated on OOB-samples.

Is there a common threshold-value for separating important and unimportant features?

For this purpose the labels were randomly permuted before testing. This breaks the dependency between the data and the labels, this is further referred to as „unpredictive case“. After the permutation of the labels the variable importance in assessing the random labels is calculated. As a first expectation one could think that all variables are equally (un)important. For all experiments with the Gini importance (see Figure 6.18 and 6.21) the importance in the „unpredictive case“ depends on the importance for predicting the original labels. This issue is easy to explain: Each decrease in impurity ($\Delta i(w, \mathbf{X}_w)$) when splitting a dataset is non-negative. It is impossible that the parent dataset has a higher purity than fractions of it. For completely random data the impurity in parent and child will be the same and thus $\Delta i(w, \mathbf{X}_w) \approx 0$ any random fluctuations from the perfect randomness will cause positive values for $\Delta i(w, \mathbf{X}_w)$. In tree-building, predictive variables are more likely to be selected and they tend to accumulate a higher amount of positive values, making them more important even in the „unpredictive case“. The same effect is also visible for the permutation importance (see Figure 6.16 and 6.20). Since the permutation of the a features can lead to a decrease in prediction performance, but also to an increased performance when the permuted feature occasionally performs better, the summation over all trees partly cancels out the effect. Indeed the effect is not very pronounced. I.e. in the importance for unpleasant stimuli (Figure 6.20 rightmost picture) the effect is not visible at all. **Conclusion:** Assessing a significance for the importance of a feature is a weakness of the method. Regarding the computational effort it would be viable to compute the distribution for the „unpredictive case“ as it was done for the shown experiments. The results give evidence

that a feature's importance in the „unpredictive case“ is dependent on the feature itself¹. This makes it impossible to assess the significance of a single feature by comparing it with the random distribution of all features in the „unpredictive case“. A workaround would be to find the random distribution for each single feature by permutating the labels several times. This method is proposed in [1]. Due to the high dimensionality $V = 5,124$ and the complex RF-model, with $n_{\text{tree}} = 5,000$ trees the variable importance assessment is extensive to compute and thus up to now this approach is not practicable.

7.3 RF Variable Importance Compared to GLM

Remarks on the comparison with the GLM

As discussed in detail in Section 3.3 the GLM is an activation-based and MVPA is an information-based approach which is per se different and not directly comparable. Furthermore the procedure of using the 95-percentile from the „non predictive“ permutation importance assessment should not be interpreted as 95% significance level. As explained above the random-value for the importance value in the “non predictive case“ depends on the feature itself, thus the common used value will be too high or too low for some features. Comparing the GLM-analysis for self-perception (Figure 6.1 – 6.3) with the importance values in the S→S MVPA (Figure 6.22 – 6.24), shows that there are activations marked in the GLM but not in the MVPA, activations marked in the MVPA but not in the GLM and activations marked in both analysis. This reflects the fact known from literature that both methods partly tell complementary information [59].

Why compare the multivariate RF-analysis with the GLM?

For comparing this newly proposed method it has to be brought in line with the established methods. A comprehensive overview was given in Table 3.1. The proposed method performs a whole-brain analysis, which on the first hand makes it comparable to univariate GLM-analysis and multivariate searchlight analysis. As discussed above this method is incomparable with MVPA-searchlight, as this method is designed to find fine grained local informative patterns. This is prohibited in the experiments in this thesis since all analysis were done on a group-level on an anatomical alignment, where only an alignment precision of 2 to 7mm can be exacted due to inter subject variability [67]. **Conclusion:** The proposed method is a whole brain analysis but delivers complementary information to the searchlight-approach. Thus the comparison is done with GLM, where the remarks given above should be kept in mind.

7.3.1 Self Perception

Comparing the results from the GLM (Figure 6.1 – 6.3) and MVPA (Figure 6.22 – 6.24) a very consistent pattern can be found, whereas in the MVPA analysis a quite high amount of small distributed activated spots can be found. Beside that also the GLM shows some activated patterns not found in MVPA. The speckled activation in the MVPA is difficult to interpret and may be due to the suboptimal set of threshold-value for masking unimportant nodes. An activated area with a much higher contrast in MVPA than in GLM is discussed in the following:

¹the dependency can be on various properties of the feature: e.g. selection frequency, importance

How to interpret informative voxels found in the MVPA, which show no activation in the GLM-analysis? Example: Coding for emotional content

When comparing the medial view of the left hemisphere in the MVPA Figure 6.22 – 6.24 it is noticeable that an area in the frontal cortex (see Figure 7.1 for marking of the exact area) seems to be involved in the emotional content. This corresponds to findings in literature [91, 100, 116]. The area is not informative for neutral stimuli activated for pleasant stimuli and deactivated for unpleasant stimuli. These areas are also visible in the GLM-analysis, but with a much weaker contrast. For easier comparison see Figure 7.1, where the area is marked with a red circle in both MVPA and GLM analysis. Interestingly the parcel (L. G and S cingul-Ant) containing the marked area, shows informative joint importances with parcels in the visual cortex (see Figure 6.35 for unpleasant stimuli). **Conclusion:** The marked area is not informative on its own (weak contrast in the GLM) but informative in conjunction with the patterns in the visual cortex (high contrast in the MVPA).

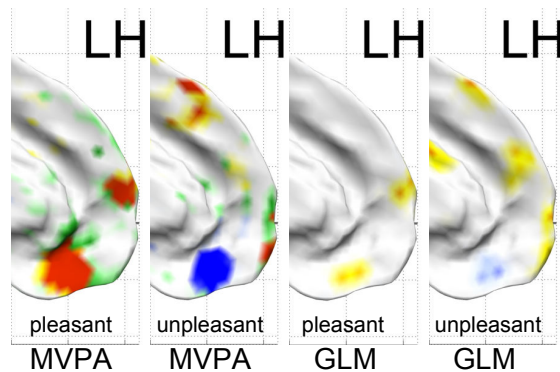


Figure 7.1: Comparison of activation coding for emotional content in self-perception. Images are cuttings from Figure 6.23, 6.24, 6.2 and 6.3

7.3.2 Empathic Perception

For that a comparison is done between areas with a significant activation contrast between the different stimulus-types in the other-session with GLM-analysis (Figure 6.4 and 6.6) and the importance in the O→O decoding (Figure 6.25 to 6.27). Again a very consistent activation between both methods can be found. Again speckled activations are visible in MVPA which are not present in the GLM. Mainly the activation is visible on the visual cortex in the GLM but also all other activations in the GLM-analysis can consistently be found in the MVPA.

7.3.3 Shared Networks

The MVPA finds areas carrying informative activation for distinction between the stimuli. To reveal shared coding patterns the method as discussed in Section 5.1 was used. In contrast to the GLM the MVPA delivers two sets of shared patterns. First the S→O experiment reveals

patterns from the self-model which are also informative in decoding the other-data. Second the O→S experiment shows which of the patterns in the other-model are predictive for decoding self-perceived emotions.

Due to the low amount of significantly activated areas in the other-session in the GLM-analysis, the cut set between activation in self-perception and empathic-perception is quite small. Shared significant activation contrast between the stimuli can mainly be seen in the visual cortex, on a small subarea on the left S1, on the amygdala for unpleasant stimuli and for neutral and pleasant stimuli in the area of dorsal and ventral part of the posterior cingulate cortex (see Figure 4.1 for definition of the areas and [24] for anatomical brain parcel names and location).

What information is gained by the S→O experiment?

The decoding performance of S→O is better than the decoding of the O→S experiment. Thus it can be considered as the more informative decoding procedure.

First of all shared activations from the GLM analysis can also be found in the shared multivariate representation.

In Chapter 4 shared representations between self and other are listed, most of them are from studies where GLM was used. These shared representations can also be found by the proposed method:

- Shared circulate cortex activity can be seen in unpleasant stimuli (Figure 6.30)
- Insula activity is visible in neutral (Figure 6.28) and unpleasant stimuli
- Activity in mPFC is found for pleasant (Figure 6.29) and neutral stimuli.
- Activation in the right S1 is informative for decoding neutral and unpleasant stimuli. Since tactile stimuli were applied on the left hand palm, a contralateral activation corresponds to the finding that also sensory related activity is shared in picture-based paradigms [73].
- Strong activation in the PreMotor Cortex (PM) is found for pleasant stimuli
- The amygdala is positively activated for unpleasant stimuli and deactivated for neutral stimuli
- Posterior Parietal Lobe (PLL) can be found in neutral and pleasant stimuli.

Additionally speckled activity is found in all three conditions. Further the activation in the dorsal and ventral part of the posterior cingulate cortex spreads out to the precuneus in all stimulus types. **Conclusion:** In the used data the proposed method reveals the same shared representations as the GLM on the same dataset, further activations can be found which are in agreement with literature. Further activations will be of interest in further detailed investigations to get new insight in neuronal empathy processing which is not revealable in a univariate analysis. In future experiments special emphasis has to be given on an improvement in significance testing and false discovery rate reductions to remove false detected activation.

Why is there a difference between S→O and O→S shared activation?

Intuitively the shared multivariate pattern across both sessions should be the same regardless in which direction the model is transferred. The fact that the S→O experiment reveals patterns informative for decoding the type of stimulus, which are not detected in the O→O experiment reveals that the feature selection in the RF-training is not perfect. E.g. the S1 activation is

not selected as important feature in the O→O experiment. For practical application one should take the model with the higher in-session decoding performance, in this case the self-model, for decoding the complimentary session and finding shared representations. **Conclusion:** The experiments show that a suboptimal decoding performance at the model transfer indicates that not all shared representation will be found.

7.4 Importance of Joint Voxels and the Importance of Long Range Connections

The higher sensitivity for MVPA compared to univariate analysis as it is reported in literature [52] arises from the combination of features (see e.g. the procedure outlined in Figure 3.1 which gives an descriptive example on how the combination of features can increase sensitivity compared to a univariate analysis). To uncover the information gained by connections of features the conditional joint Gini importance is newly introduced in this thesis. Per conception of this measure only the conjunction of two features is investigated. Of course coding patterns could contain more complex patterns across a high amount of voxels, but these patterns are captured by the conjunction of multiple pairs. To avoid confusion with the terminus *connectivity* commonly used in neurosciences, general remarks are given before discussion.

Remarks on the interpretation of the joint Gini importance:

For interpretation the joint Gini importance and the informative connection of parcels plotted in Figure C.1– C.6 it is mentioned that these connections do not follow the common usage of „connections“ in neuroscientific research. The terms *functional connectivity* and *effective connectivity* are used in neuroscience [42]. Friston et al. defines these terms: functional connectivity is the „temporal correlations between spatially remote neurophysiological events“ [40] and effective connectivity is „the influence one neuronal system exerts over another“ [39]. In contrast to functional connectivity features which are connected with a high conditional joint Gini importance will not be correlated since the two channels have to carry complementary information to be jointly more informative than one or the other feature in an univariate fashion. Thus the connections should not be interpreted as functionally connected in terms of „doing similar things“ but as being jointly informative „being together part of a activity pattern“. This goes along the controversy discussed in Section 3.3. The joint Gini importance goes in line with a „information based analysis“. As outlined in [50] this does not necessarily lead to interpretable results. Despite of that fact some findings can be made.

Are whole brain analysis capturing long-ranged connections of interest?

The main difference of the proposed method to established whole brain analysis (searchlight approach) is, that the newly introduced method in this thesis captures connections across any length. Figure 6.34 demonstrates that node to node connections are informative on any connection length. Also the schemaballs show that many informatively connected parcels are in distance locations. **Conclusion:** The data show that the whole-brain analysis without restrictions to certain ROIs is useful to capture the full complexity of a coding pattern in the human brain.

What is the contribution provided by the schemaballs?

One example was already given in the coding for emotional content: For voxels, which are uninformative alone, the schemaball shows the area in which voxels with complementary information are located. E.g. voxels in the frontal cortex were only detected with low contrast in the univariate analysis but in conjunction with voxels in the visual cortex, these voxels are informative for the type of stimulus. The same is true for the right S1 activation found with MVPA but not with GLM. The corresponding anatomical labels of S1 are *gyrus postcentral* and the adjacent sulci *sulcus central* and *sulcus postcentral*. For parcels overlapping these anatomical areas, higher connection to other parcels can be found. Furthermore and common in all schemaball plots the visual cortex (parcels in the bottom middle of the diagrams) show a high interconnectedness, meaning that the different parts in the visual cortex provide complementary information on the visual stimulus. **Conclusion:** The schemaballs provide useful information on the location of voxels which provide complementary information for refining the classification. As for other information-based analysis the neurological interpretation is sometimes difficult.

7.5 Summary

It could be shown that the data used are not suitable for single-subject models due to the short recording length in the fMRI. But this thesis demonstrates a successful way for dealing with the data by performing all analysis on a group-level on a FS4-resolution level. The decoding models from the self-session performs slightly better than the decoding model from the empathy session. For the psychologically interesting shared patterns the proposed method shows a increased sensitivity compared to the classical GLM-analysis. Areas with weak contrast in the GLM but high contrast in the MVPA consistently show high joint importances with other parcels.

Conclusion and Future Work

8.1 Major Findings of this Thesis

It could be demonstrated on real data that the RF-classifier is extremely useful in performing whole brain multivariate pattern analysis. The shown procedure works on a group level with 30 subjects and is able to decode the emotional state of visuo-tactile stimuli on subjects not included in the training set. The classification model is robust enough to also encode the correct emotional state in a different scanning session where the participants perceived the emotion empathically.

8.1.1 Benefit of RF in fMRI Data Analysis

The RF-training procedure incorporates correlated features in the model, and does not restrict the feature selection to the most predictive features. Standard RF-variable importance measures are shown to be useful in identifying informative voxels. The results are consistent with a GLM-analysis.

8.1.2 Insight into Empathic Mechanisms

By training the classifier model on the data of the self-perception scanning session and using this classifier to decode the type of stimulus in the empathy scanning session, this thesis demonstrates a method of using the shared multivariate patterns across both conditions to successfully decode stimuli in the empathy scanning session. By subsequent variable importance assessment for the empathy-data in the self-decoding-model the shared representations between self-perception and empathy could be identified. The results are in line with previous findings from studies investigating the neuronal substrate of empathy in the human brain. The comparison between the newly proposed MVPA and the GLM shows a much higher sensitivity for the MVPA. The exact comparison in detail is problematic. The MVPA probes for informative patterns, whereas

the GLM probes for activated voxels. These two analysis methods rely on different models and different statistics and are thus not directly comparable.

8.1.3 Benefits Compared to State-of-the-Art Whole-Brain MVPA

A newly introduced importance measure reveals the conjunctions between features, which make features together more informative in decoding compared to only the one or the other feature alone. Visualizing these connections, the spatial location of jointly informative voxels can be shown. This reveals that connections of any length can be informative. Thus the widely used MVPA-searchlight approach loses lots of informative conjunctions by the restriction of the patterns to local volumes. Compared to the searchlight-method the proposed method does not necessarily require a high resolution which makes it 1.) usable for group-level approaches and 2.) suitable for short single subject scanning sessions.

8.2 Further Extensions to the Proposed Method

The identification of shared multivariate whole-brain neuronal representations will be of further interest for neuro psychological research. This is planned with supervision and psychological expertise from the University of Vienna, where this thesis should support as the methodological description of the new method. One idea for further improvement of the method arose during the end of the work:

A More Elaborate Significance Test for MVPA As described in the discussion, the separation of unimportant and important voxels should be done with threshold-values calculated separately for each single voxel. This could be done by numerous random permutation tests as proposed in [1]. This is computationally problematic for the huge datasets because many permutations are needed to find the correct null-distribution for each feature. Therefore a more effective implementation of variable importance assessment would be needed, and more importantly a modification of the used RF-implementation is necessary, since some evaluations have to be assessed during training on the OOB-samples, otherwise the results are biased by the in-bag samples. This would lead to a higher statistical power of the analysis and may reduce the uninformative parts of speckle-activation without losing informative activations.

Alternative Data Preparation for Classification and Model Trainings

In this appendix an alternative method for preparing the data and labels for classifier training is presented. Further it is tried to train the classifiers on a single-subject level. The complete results to the methods described here can be found in Appendix B.1 for the single subject models and Appendix B.2 for the group models.

A.1 Alternative Data Preparation Strategy

Alternatively to the strategy given in section 5.3.2 one can build one training example per experimental block instead of one training example for each point in time.

A.1.1 Preparing Labels for Classification with a GLM Block Design

This design is used for pattern estimation in MVPA fMRI-studies e.g. [18]. In this approach a GLM is modeled in a classical style but a separate estimator is used for each block in the recording session. If the experiment consists of n_B individual blocks, a number of n_B β -values can be obtained for each voxel in the volume. These β -values can be assigned to a label l (again ranging from 1 to 6) according to the experimental condition of the block. The processing for this design is outlined in the following in more detail:

Building the Predictor-Functions

The stimulus-function $s_l(t)$ is now split up in separate functions $s_{li}(t)$ for each block, where l corresponds to the experimental condition and i is a consecutive index to identify each separate block of the same stimulus type. In all sessions i ranges from 1 to 10 for each type of stimulus. This results in $n_B = 6 \cdot 10 = 60 = n_{\text{stim}}$ stimulus block-functions for each scanning session. As an example for such a set, 60 stimulus functions $s_{li}(t)$ are plotted in Figure A.1. By convolving

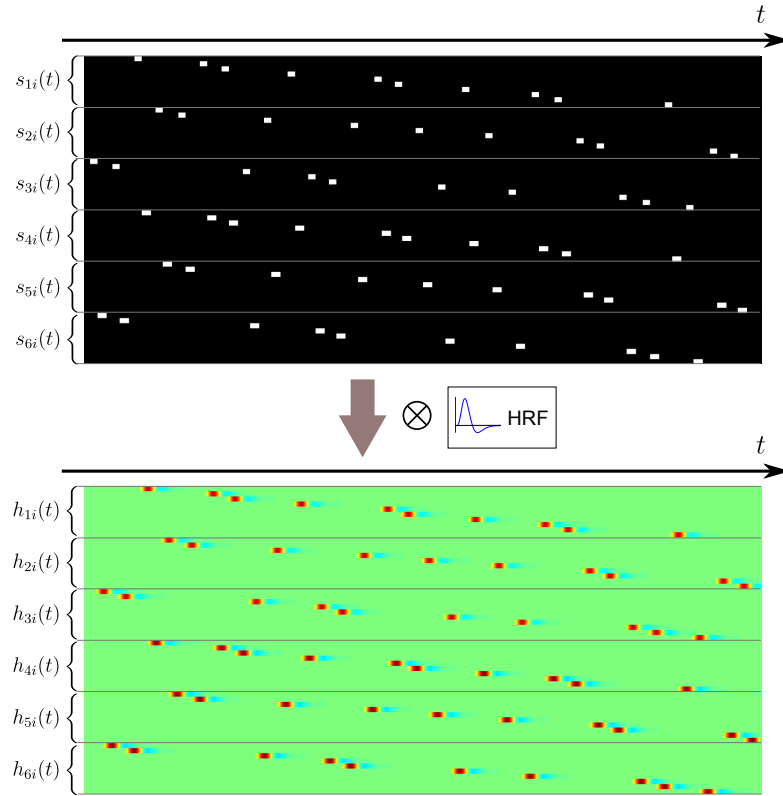


Figure A.1: Building a GLM predictor function for each block: 60 stimulus functions $s_{li}(t)$. l ranges from 1 to 3 for the three different types of stimuli and 4 to 6 for the respective rating blocks; i ranges from 1 to 10 for every l . Each function $s_{li}(t)$ is convolved with the hemodynamic response function to estimate the 60 predictor-functions $h_{li}(t)$ for the GLM model. Colorcode: black = no stimulus/rating, white = stimulus/rating; green = 0, hot colors positive, cold colors negative values

each function $s_{li}(t)$ with the hemodynamic response function, n_B different estimator functions $h_{li}(t)$ can be obtained. This process is illustrated in Figure A.1. All functions $h_{li}(t)$ are sampled with a sampling interval of T_R , resulting in 60 vectors each with 180 ($=n_{\text{obs}}$) entries. Combining these vectors delivers a $n_{\text{stim}} \times n_{\text{obs}}$ matrix, which is the transposed predictor matrix M^T for the GLM. Such a predictor matrix with one predictor per block is also given in [70].

Training Data for the GLM Strategy

When solving the GLM according Equation (3.3) for each voxel v with the predictor matrix given above, one obtains a matrix \hat{B} of the size $V \times n_B$. Herein each row contains the estimated beta values $\hat{\beta}$ for each of the n_B stimulus blocks. Further we define \mathbf{l}_b as the block label vector, which contains the label l for each of the n_B stimulus blocks.

We use the matrix of β -values $\hat{\beta}$ as the training data \mathbf{X} , the block label vector \mathbf{l}_b is used as

response vector y . Again this is done for both, the self-session and the other-session.

A.1.2 Decoding Each Point in Time vs. Decoding Experimental Blocks

There are two possible targets for decoding: In strategy 1 the classifier is trained to decode the experimental condition of each point in time as shown in section 5.3.2, whereas in strategy 2 the classifier is trained to decode the condition of a whole experimental block.

It can be expected that the classification results of the block design are better, since this preprocessing incorporates the information of the block length, whereas the decoding of each point in time permits inconsistent decodings within one block. For a fair comparison the information of the block-length was included in the “strategy 1” decoding by summing up the tree’s votes across all observations within an experimental block. This leads to a majority-vote for each block. The resulting three decoding methods are listed in Table A.1.

Advantages and Disadvantages for Continuous Labeling

Advantages:

- + High temporal resolution since a label is assigned to each single volume.
- + BOLD signal can be directly used as feature value in the classifier training. Thus the testing can be made without the knowledge of the block length.
- + High amount of training points since each acquired volume is one observation.

Disadvantages:

- Interstimulus interval has to be used to prevent temporally overlapping hemodynamic responses.
- Responses of stimuli- and rating-blocks temporally overlap. Thus at some points in time the labels are inaccurate since they belong to mixed experimental conditions.
- Selection of the threshold-value th changes the amount of training-points. If it is set too low the classifier is trained on observations where there is barely a response from the stimulus. When taking a high threshold-value useful observations for training get lost.

Advantages and Disadvantages for GLM Block Design

Advantages:

- + This method can handle overlapping hemodynamic responses correctly since the β -values as features correctly depict the correspondence of each voxel to each separate stimulus block. [85]
- + Often used in other studies → better comparability if the same preprocessing is used.

Disadvantages:

- No direct decoding is possible. The onsets and durations of each stimulus block have to be known to calculate the β -values for each block. Only with the previous GLM-model the β -values can be fed into the classifier to decode a labeling for each block.

| Name | | Training | Math Symbols | Decoding |
|----------|----|---|----------------------------|--|
| Method 1 | M1 | BOLD-response for each point in time | $X = Z$ $y = l$ | majority of vote for each point in time |
| Method 2 | M2 | BOLD-response for each point in time | $X = Z$ $y = l$ | majority of vote across all observations per block |
| Method 3 | M3 | β -values for each experimental block | $X = \hat{B}$ $y = l_b$ | majority of vote for each block |

Table A.1: List of decoding methods

A.2 Training Single Subject Models

For this purpose the RF-classifier is trained on each subject's individual data matrix X_{self} and X_{other} with the target vectors y_{self} and y_{other} . Except the cross validation schema all evaluations are exactly as in the group-model. The single subject cross-validation schema self-model decoding the self-data is a 4-fold cross validation as shown in Figure A.2 (red box). For the *model transfer* experiment the model is trained on the full self-session and tested on the empathy-session. This procedure is done separately for each of the 30 subjects.

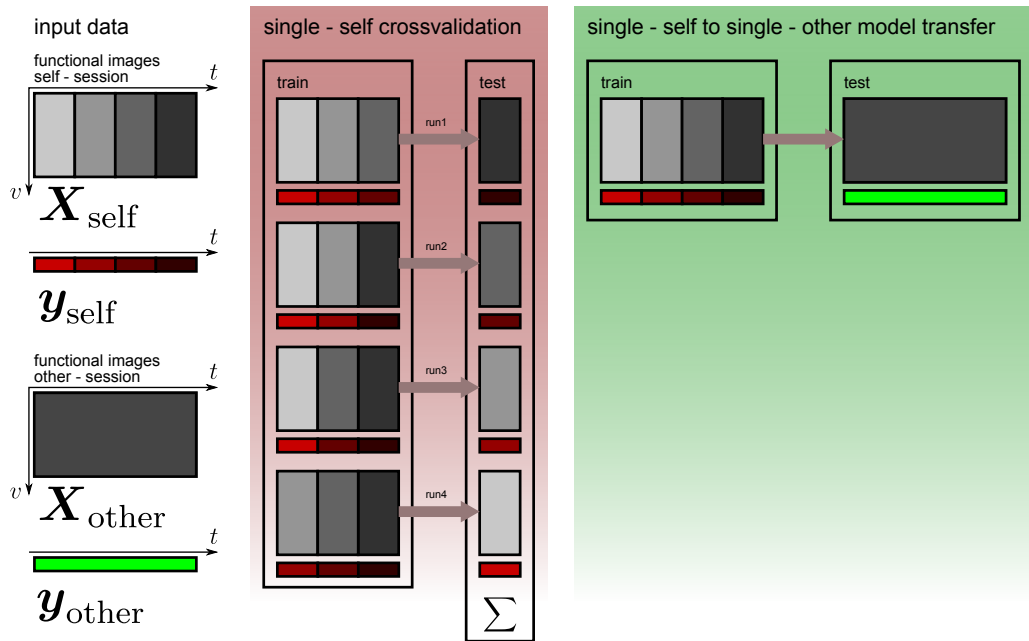


Figure A.2: Training and validation schema for single subject analysis: The self-session is split into 4 consecutive blocks. For the self-session cross validation the classifier is trained on $\frac{3}{4}$ of the data and tested on the left out $\frac{1}{4}$; The validation result is summed up over 4 runs for each block left out for testing once. The model transfer test is done by training on the full self-session and decoding the data from the other-session

Decoding Performance

This appendix gives the full results of all conducted classification performance experiments. An overview on all correct classification rates is given in Table B.1.

| Model training on | | | | Self | Self | Other | Other | |
|-------------------|--------------------|--------|----|------|-------|-------|-------|---|
| Decoding of | | | | Self | Other | Other | Self | |
| Res. | n_{trees} | Model | M. | | | | | |
| FS4 | 5000 | Group | M1 | 57% | 57% | 57% | 54% | ◇ |
| FS4 | 5000 | Group | M2 | 72% | 69% | 68% | 66% | ◇ |
| FS4 | 5000 | Group | M3 | 61% | 57% | 63% | 58% | ◇ |
| FS5 | 20000 | Group | M1 | 57% | 57% | – | – | ◇ |
| FS5 | 20000 | Group | M2 | 71% | 71% | – | – | ◇ |
| FS5 | 20000 | Group | M3 | 62% | 59% | – | – | ◇ |
| FS4 | 5000 | Single | M2 | 44% | 56% | – | – | □ |
| FS5 | 20000 | Single | M2 | 44% | 58% | – | – | □ |

Table B.1: Overview on the correct classification rates: Resolution (Res.) FS4 or FS5; Method (M.) see Table A.1 for definition; ◇ accuracy of group model; □ average accuracy over all single–subject models

B.1 Single Subject Classification Accuracy

It can be seen from Table B.1 that method M2 performs best. Since the single subject classification performance is very low, only the results of M2 were evaluated for the single subject analysis.

B.1.1 Self-Model decoding Self-Session

M2 – 5000 Trees – Self-Model decoding Self-Session

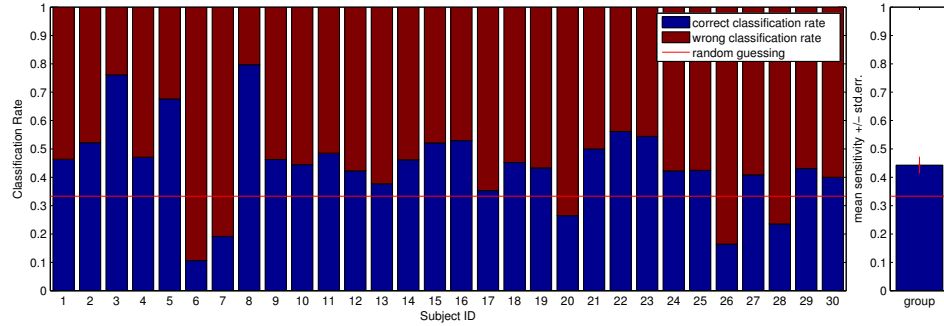


Figure B.1: correct classification rates for each subject and mean \pm std.err.

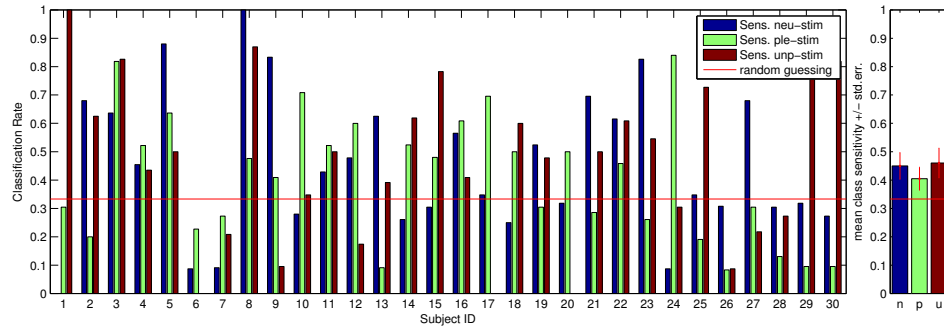


Figure B.2: class-sensitivities for each subject and mean class-sensitivity \pm std.err.

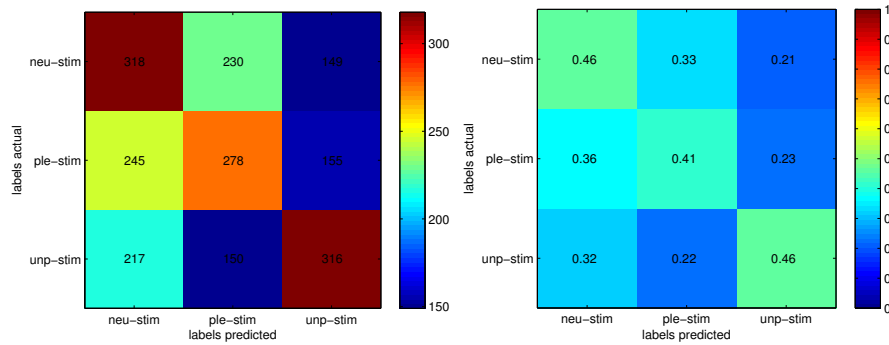


Figure B.3: confusion matrix: absolute counts for each class (left) and relative values (right) accumulated over all subjects

M2 – 2000 Trees – Self-Model decoding Self-Session

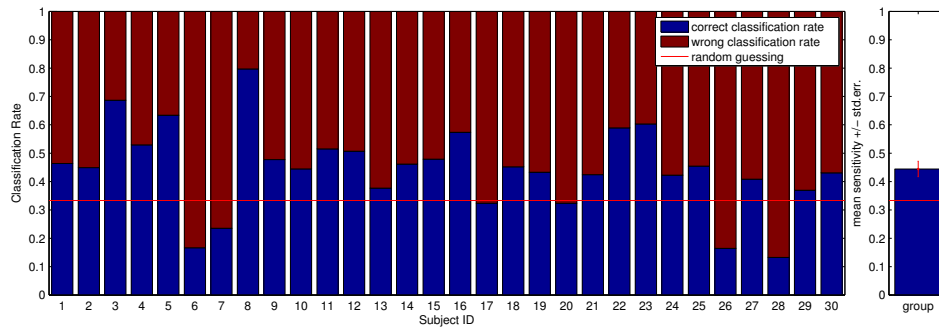


Figure B.4: correct classification rates for each subject and mean \pm std.err.

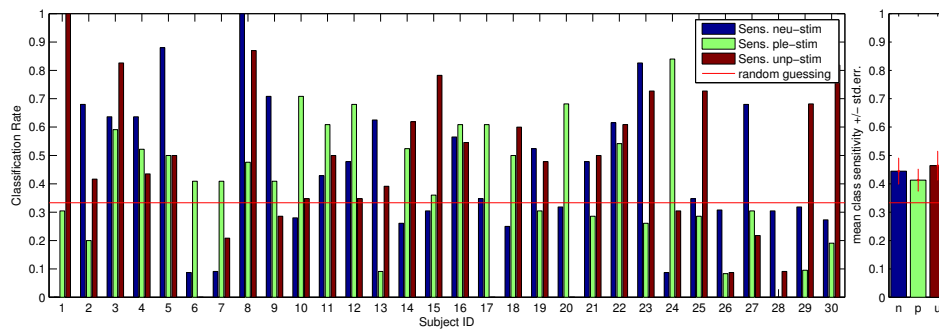


Figure B.5: class-sensitivities for each subject and mean class-sensitivity \pm std.err.

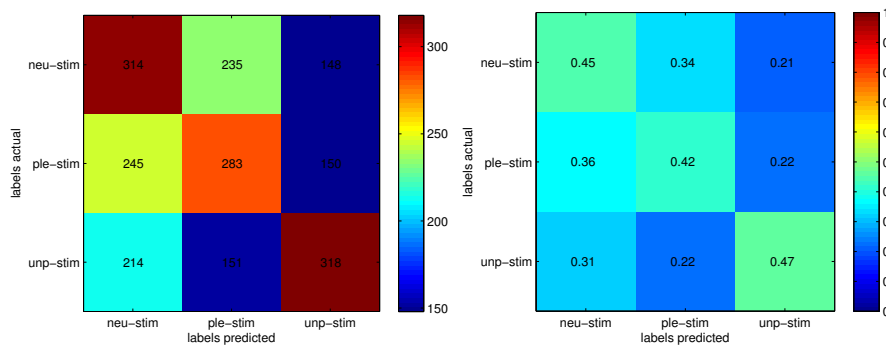


Figure B.6: confusion matrix: absolute counts for each class (left) and relative values (right) accumulated over all subjects

B.1.2 Self-Model decoding Other-Session

M2 – 5000 Trees – Self-Model decoding Other-Session

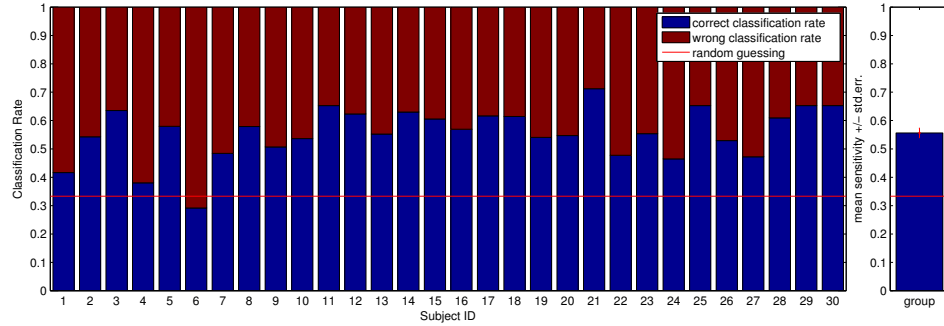


Figure B.7: correct classification rates for each subject and mean \pm std.err.

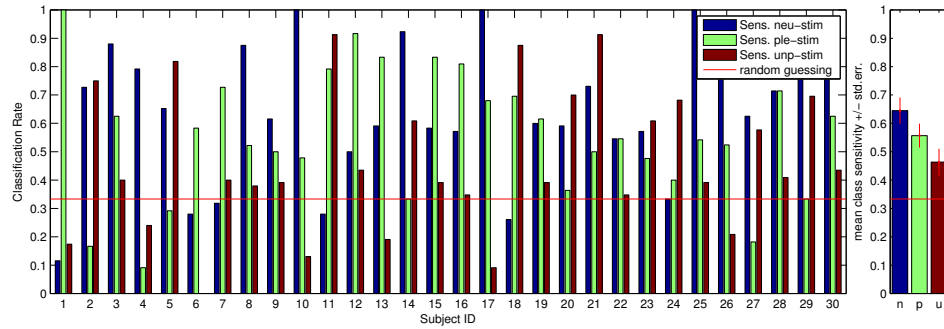


Figure B.8: class-sensitivities for each subject and mean class-sensitivity \pm std.err.

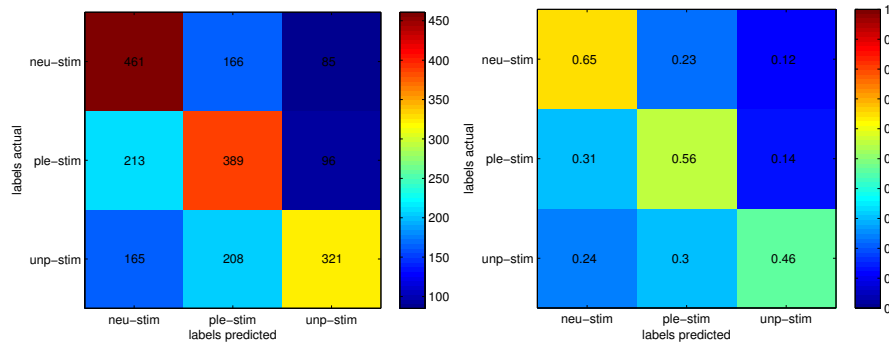


Figure B.9: confusion matrix: absolute counts for each class (left) and relative values (right) accumulated over all subjects

M2 – 2000 Trees – Self–Model decoding Other–Session

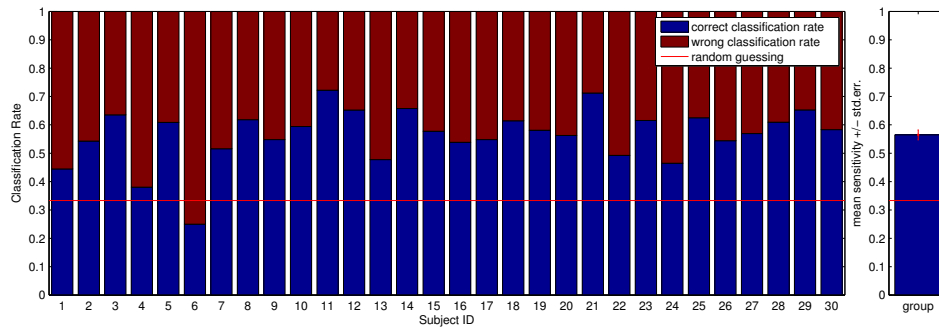


Figure B.10: correct classification rates for each subject and mean \pm std.err.

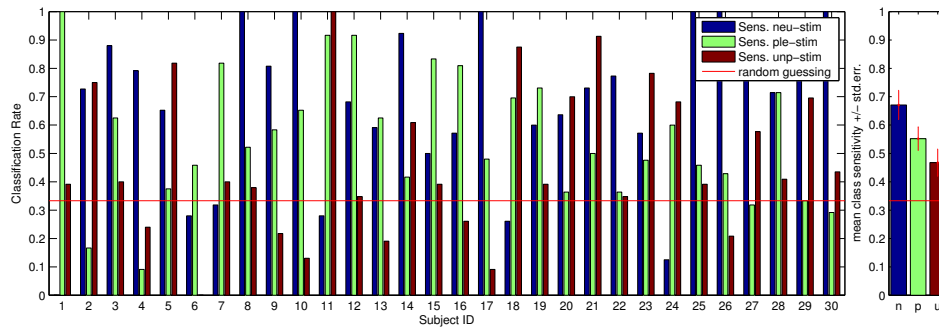


Figure B.11: class-sensitivities for each subject and mean class-sensitivity \pm std.err.

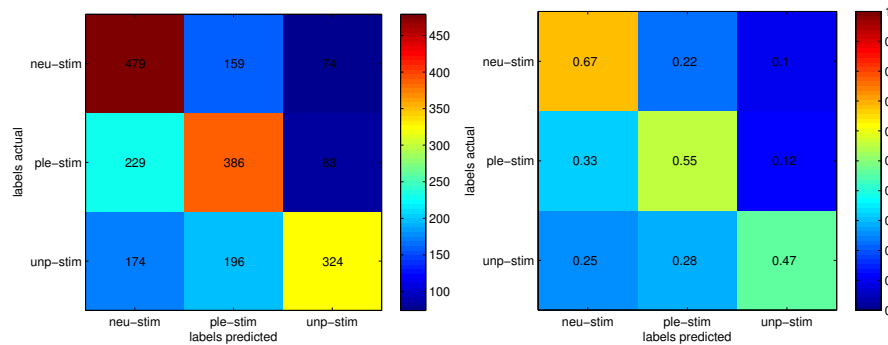


Figure B.12: confusion matrix: absolute counts for each class (left) and relative values (right) accumulated over all subjects

B.2 Group Classification Accuracy

In the following the complete results for the decoding performance assessment are given: For FS4 each decoding method M1, M2 and M3 (see table A.1) every possible combination (Self-Model decoding Self-Session, Self-Model decoding Other-Session, Other-Model decoding Other-Session and Other-Model decoding Self-Session) was evaluated. For FS5 only Self-Model decoding Self-Session and Self-Model decoding Other-Session was evaluated.

B.2.1 Self-Model decoding Self-Session

M1 – 5000 Trees – Self-Model decoding Self-Session

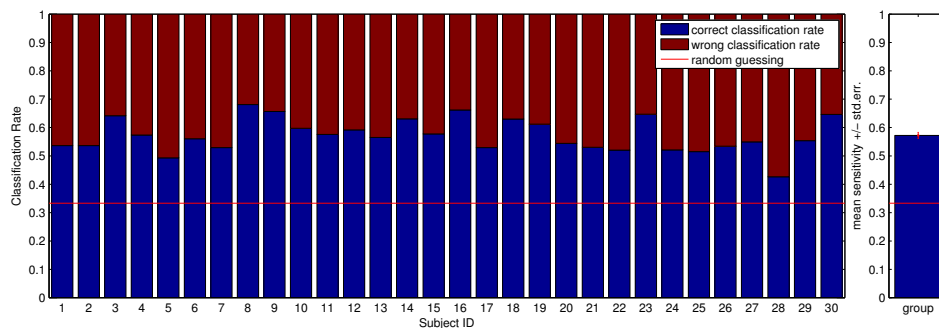


Figure B.13: M1: correct classification rates on left out subject and group mean \pm std.err.

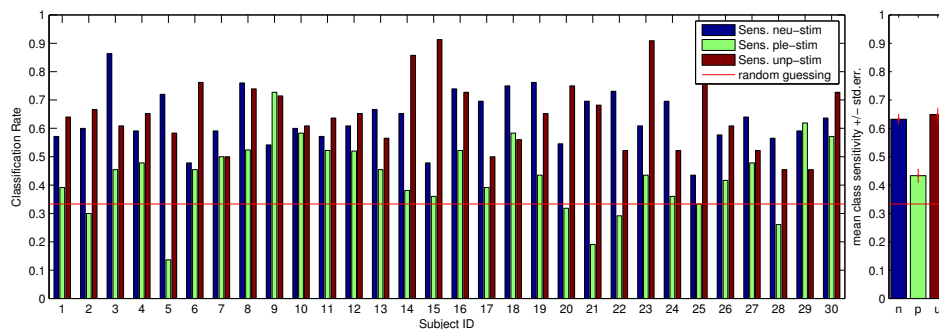


Figure B.14: M1: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

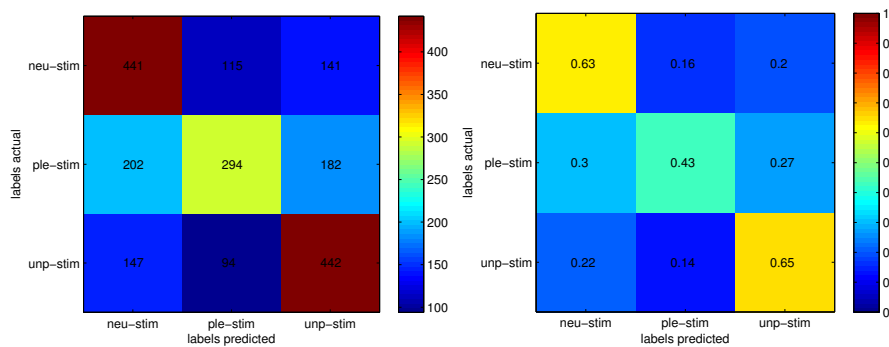


Figure B.15: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 5000 Trees – Self-Model decoding Self-Session

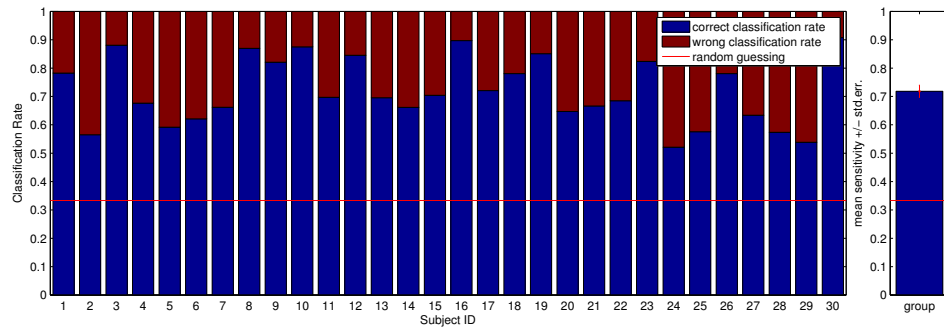


Figure B.16: M2: correct classification rates on left out subject and group mean \pm std.err.

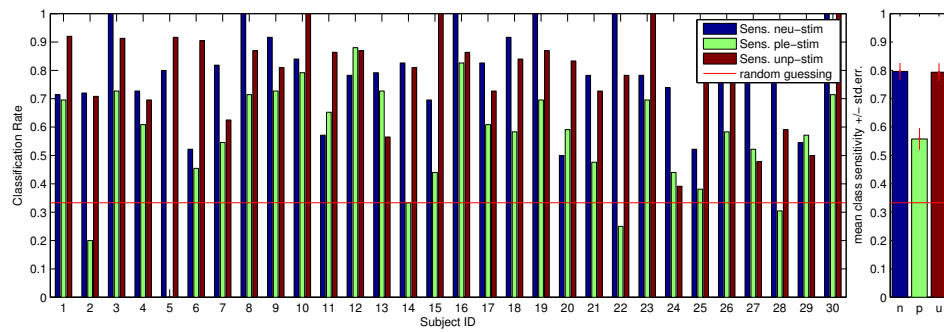


Figure B.17: M2: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

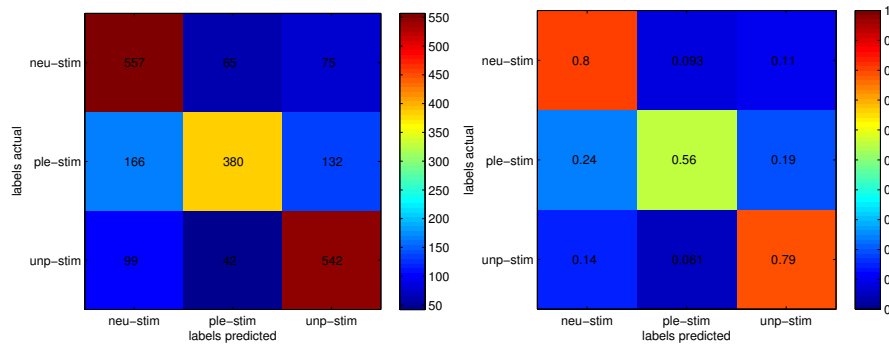


Figure B.18: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 5000 Trees – Self-Model decoding Self-Session

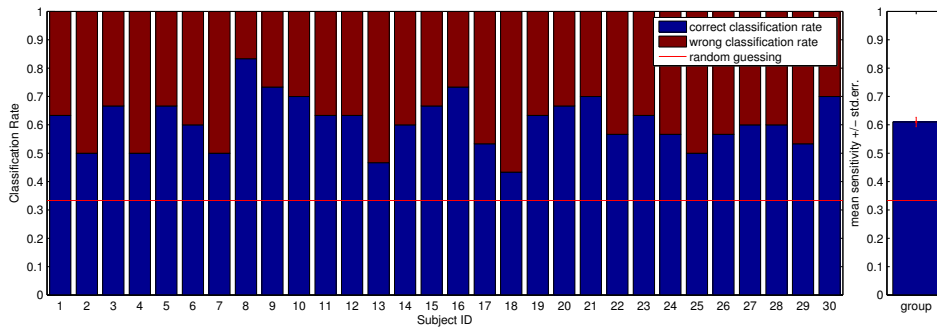


Figure B.19: M3: correct classification rates on left out subject and group mean \pm std.err.

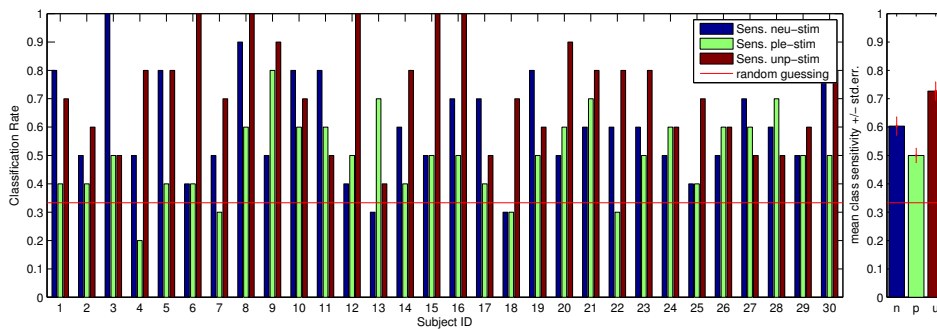


Figure B.20: M3: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

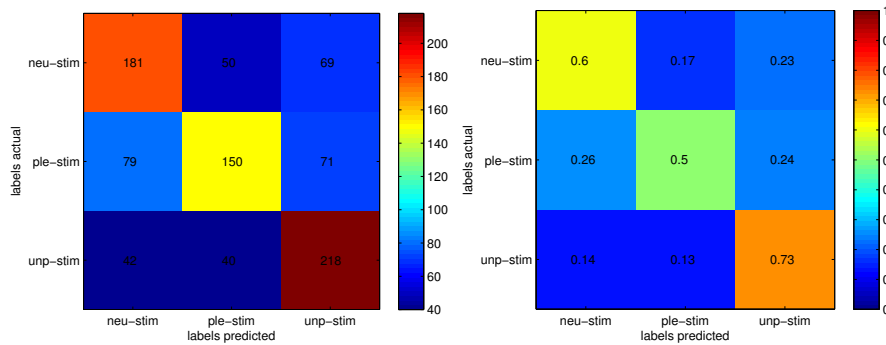


Figure B.21: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

M1 – 20000 Trees – Self-Model decoding Self-Session

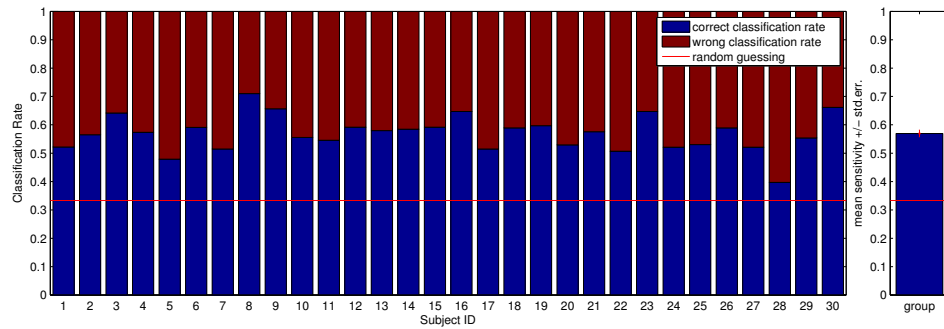


Figure B.22: M1: correct classification rates on left out subject and group mean \pm std.err.

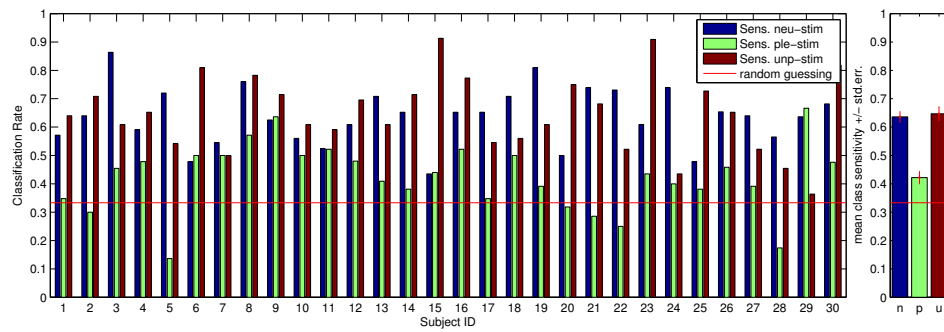


Figure B.23: M1: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

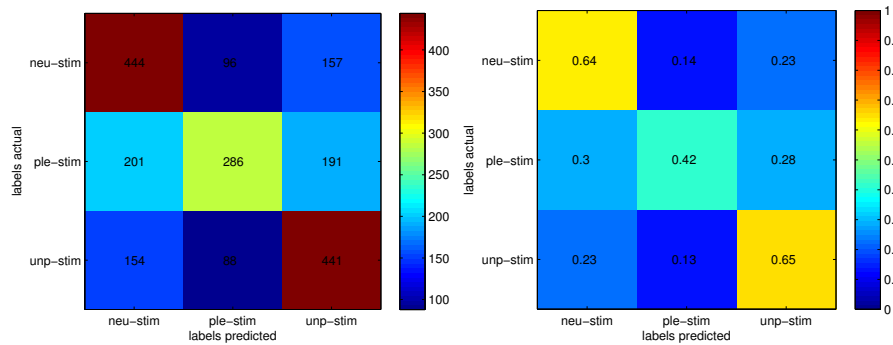


Figure B.24: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 2000 Trees – Self-Model decoding Self-Session

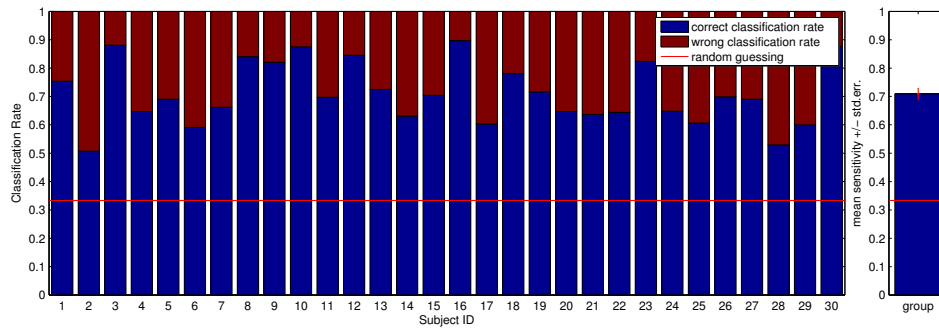


Figure B.25: M2: correct classification rates on left out subject and group mean \pm std.err.

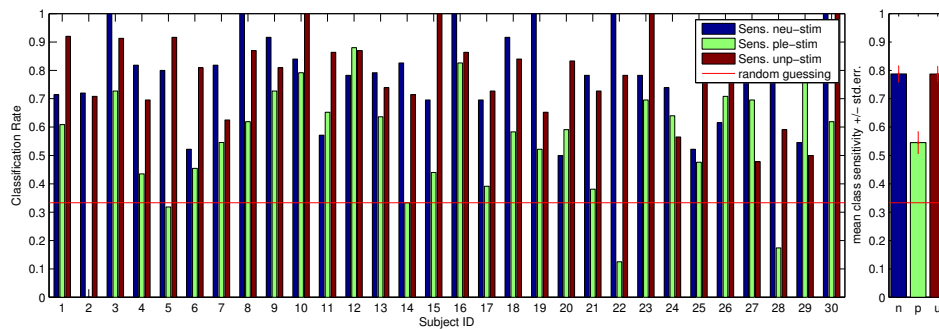


Figure B.26: M2: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

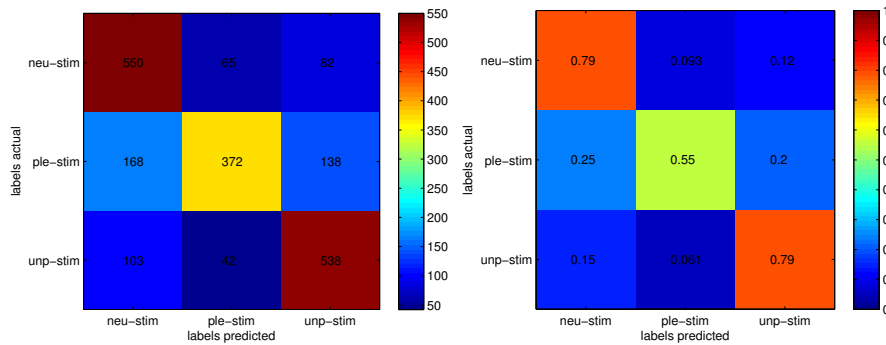


Figure B.27: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 20000 Trees – Self-Model decoding Self-Session

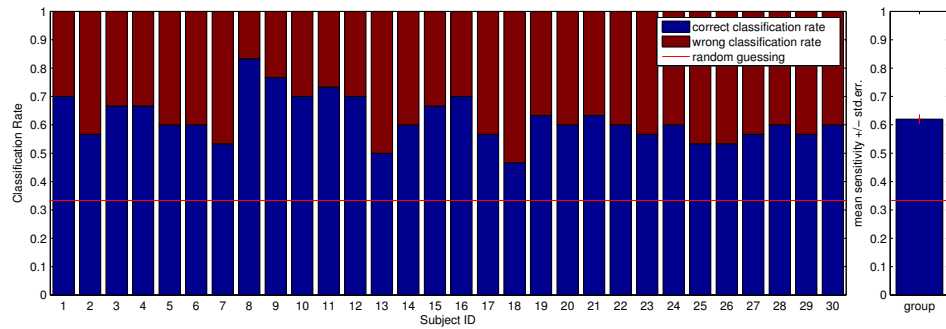


Figure B.28: M3: correct classification rates on left out subject and group mean \pm std.err.

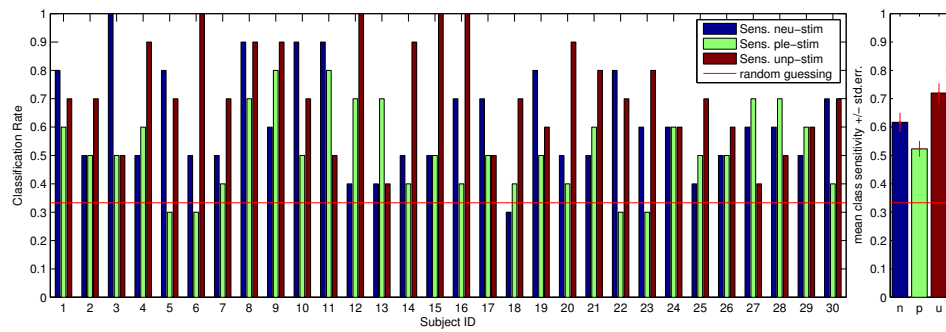


Figure B.29: M3: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

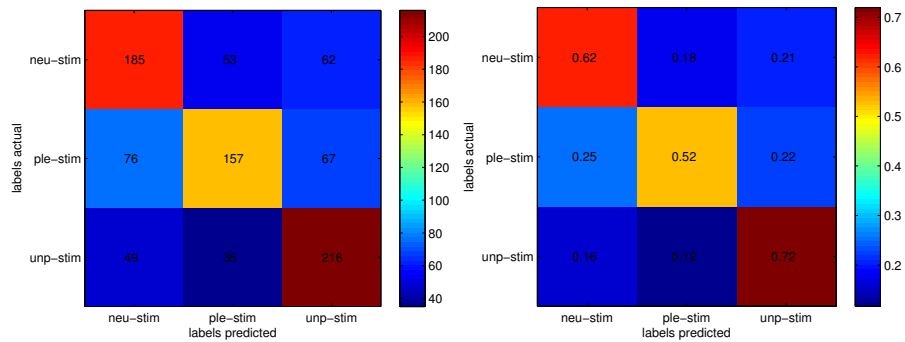


Figure B.30: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

B.2.2 Self-Model decoding Other-Session

M1 – 5000 Trees – Self-Model decoding Other-Session

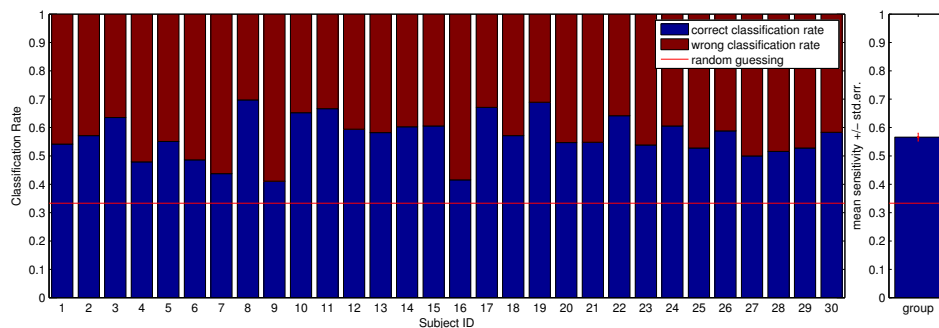


Figure B.31: M1: correct classification rates on each subject and group mean \pm std.err.

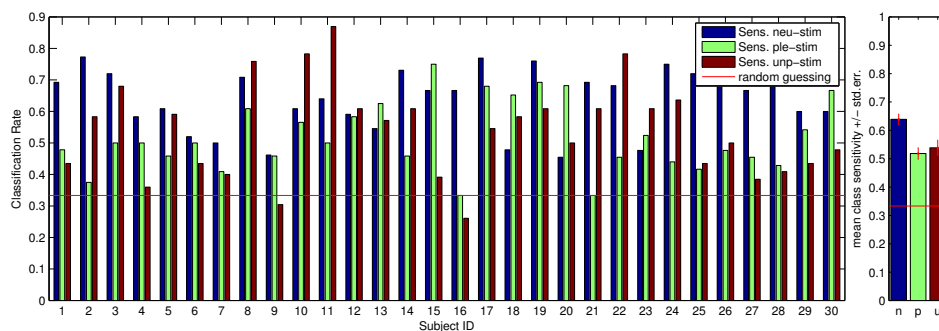


Figure B.32: M1: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

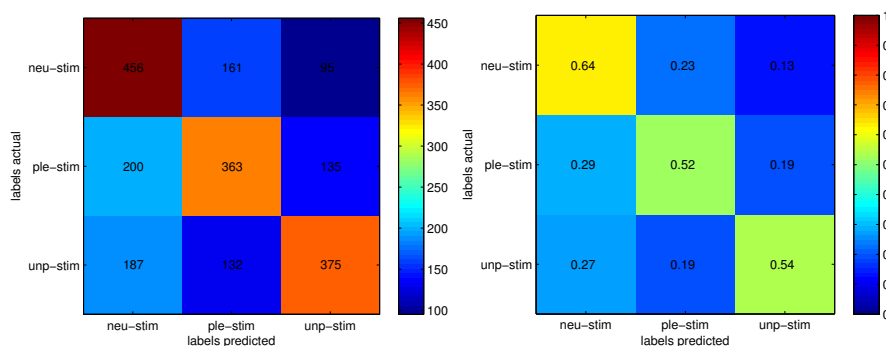


Figure B.33: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 5000 Trees – Self-Model decoding Other-Session

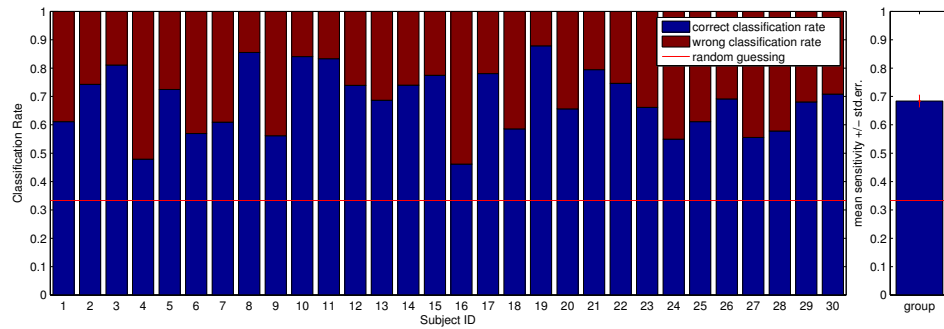


Figure B.34: M2: correct classification rates on each subject and group mean \pm std.err.

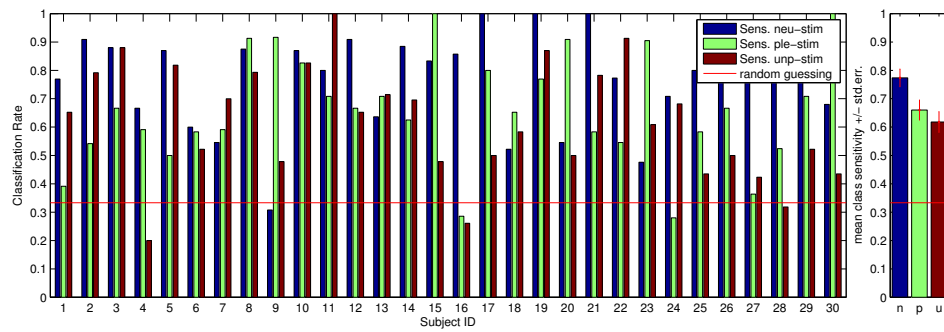


Figure B.35: M2: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

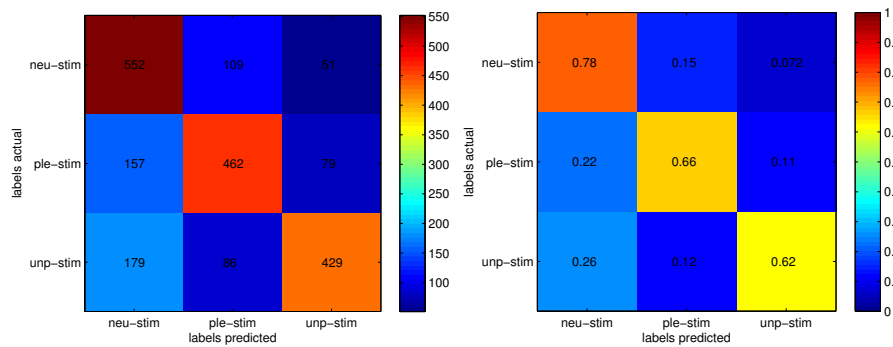


Figure B.36: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 5000 Trees – Self-Model decoding Other-Session

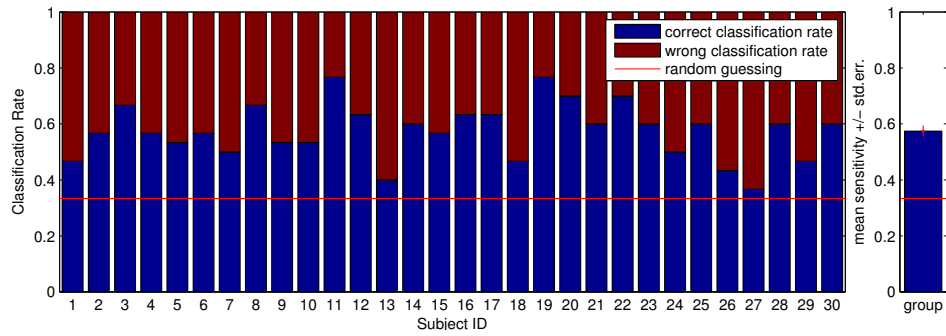


Figure B.37: M3: correct classification rates on each subject and group mean \pm std.err.

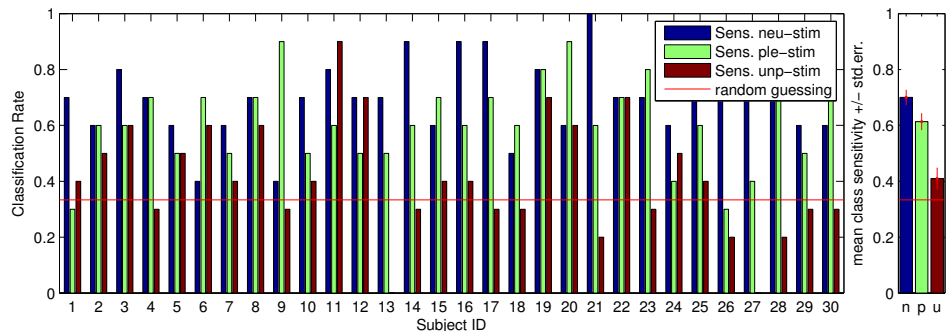


Figure B.38: M3: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

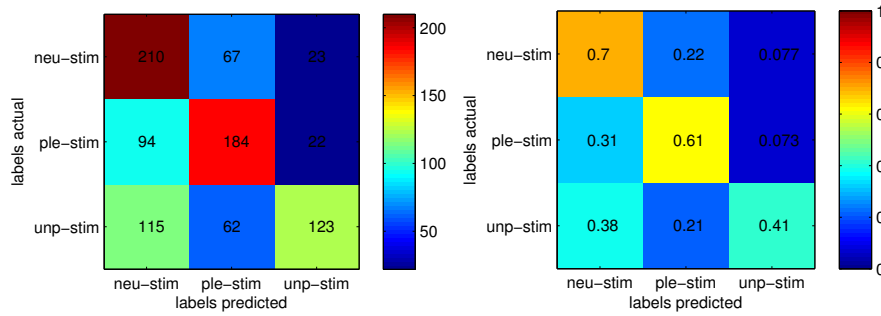


Figure B.39: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

M1 – 2000 Trees – Self-Model decoding Other-Session

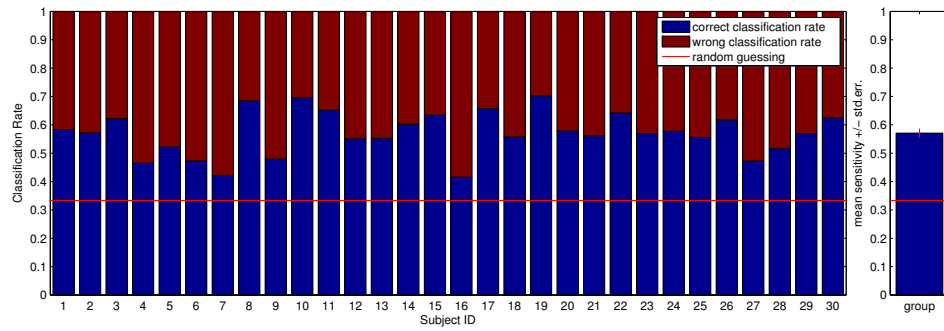


Figure B.40: M1: correct classification rates on each subject and group mean \pm std.err.

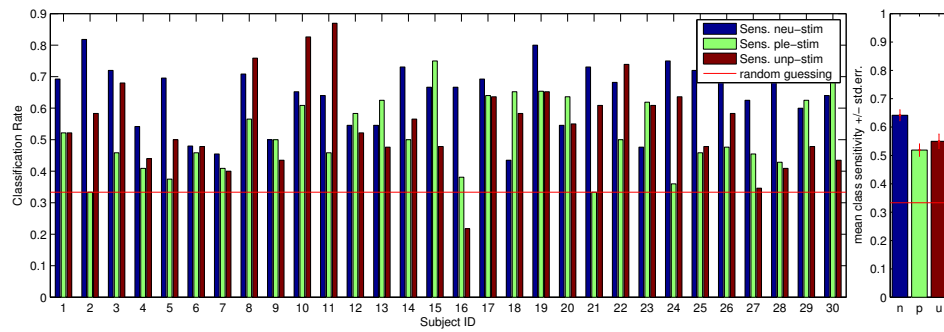


Figure B.41: M1: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

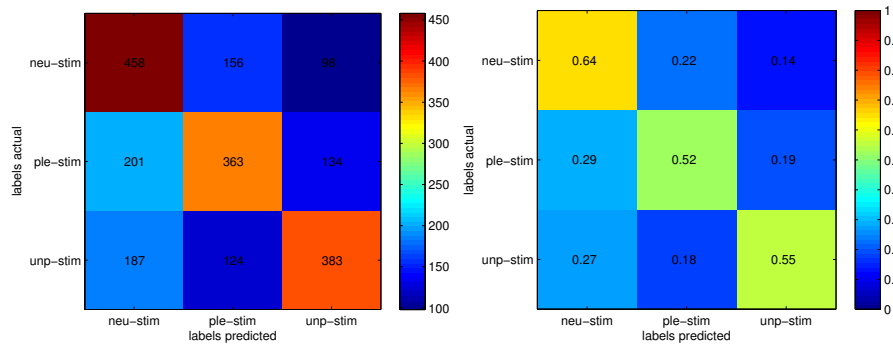


Figure B.42: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 20000 Trees – Self–Model decoding Other–Session

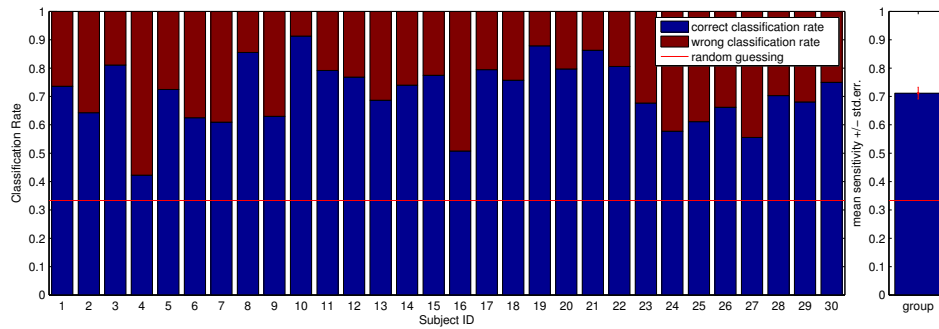


Figure B.43: M2: correct classification rates on each subject and group mean \pm std.err.

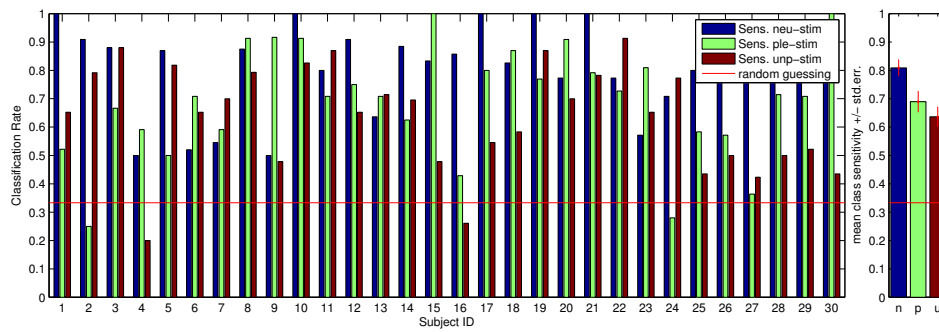


Figure B.44: M2: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

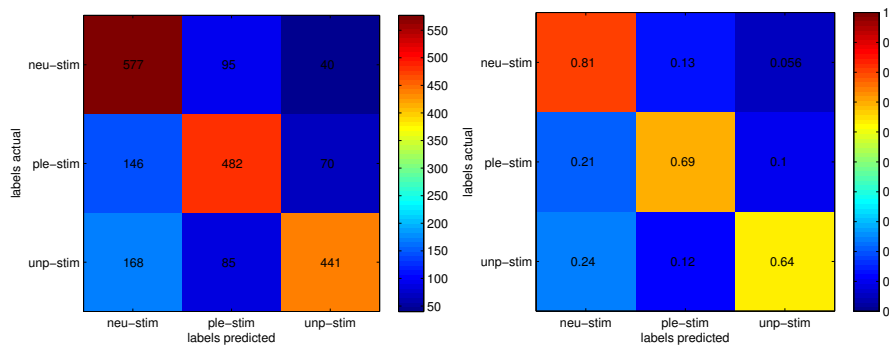


Figure B.45: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 2000 Trees – Self-Model decoding Other-Session

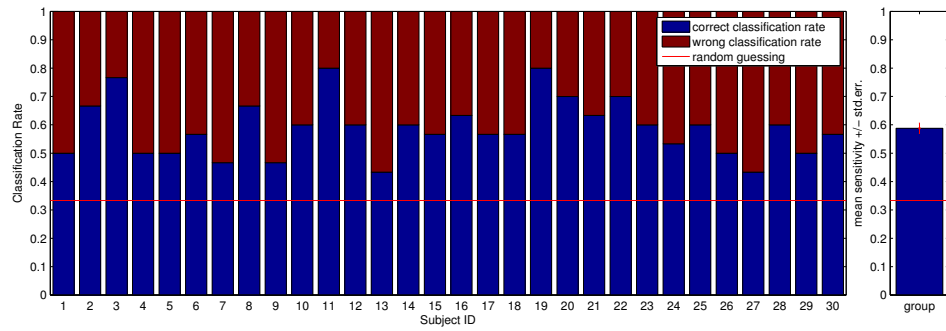


Figure B.46: M3: correct classification rates on each subject and group mean \pm std.err.

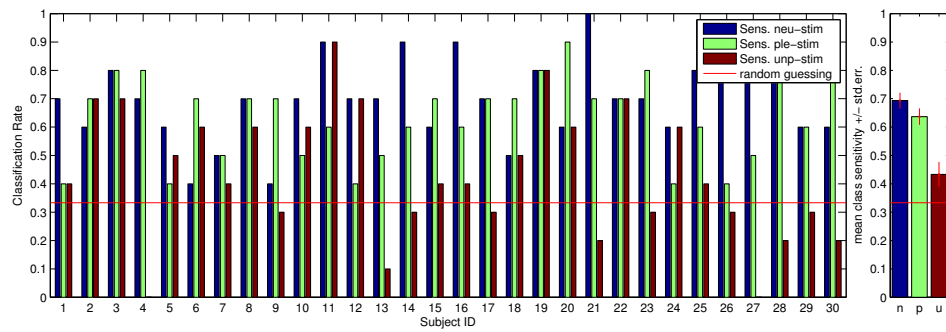


Figure B.47: M3: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

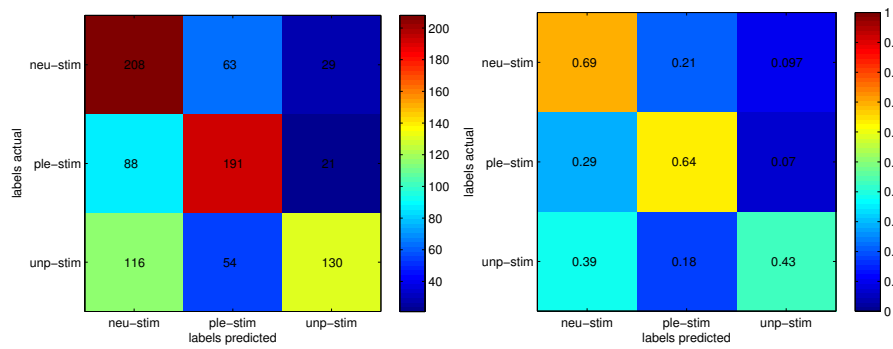


Figure B.48: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

B.2.3 Empathy-Model decoding Other-Session

M1 – 5000 Trees – Empathy-Model decoding Other-Session

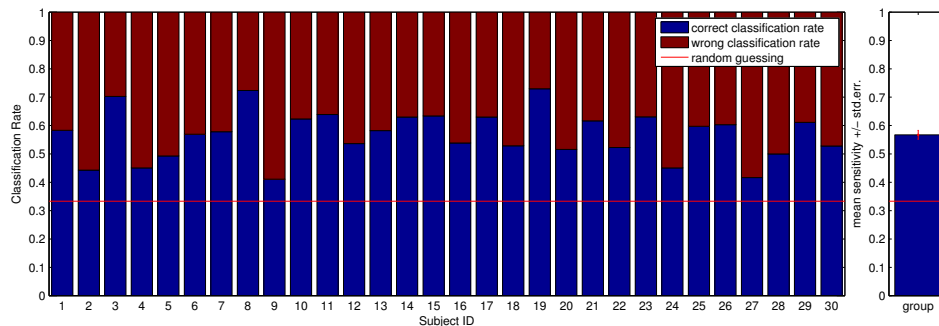


Figure B.49: M1: correct classification rates on left out subject and group mean \pm std.err.

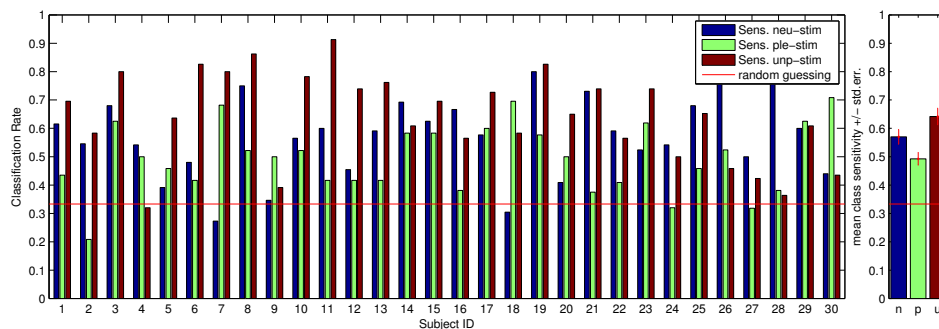


Figure B.50: M1: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

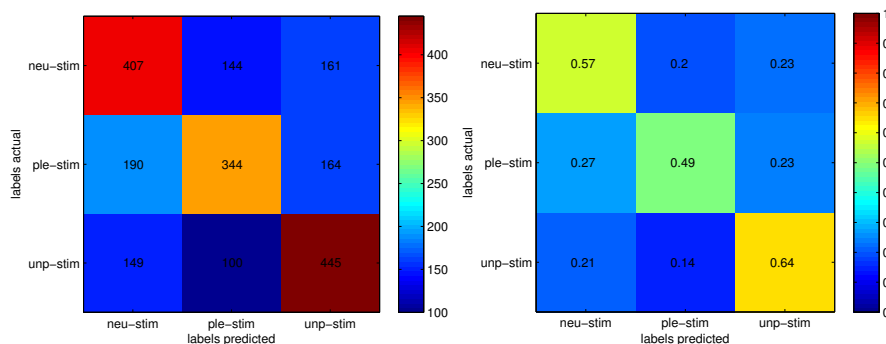


Figure B.51: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 5000 Trees – Other–Model decoding Other–Session

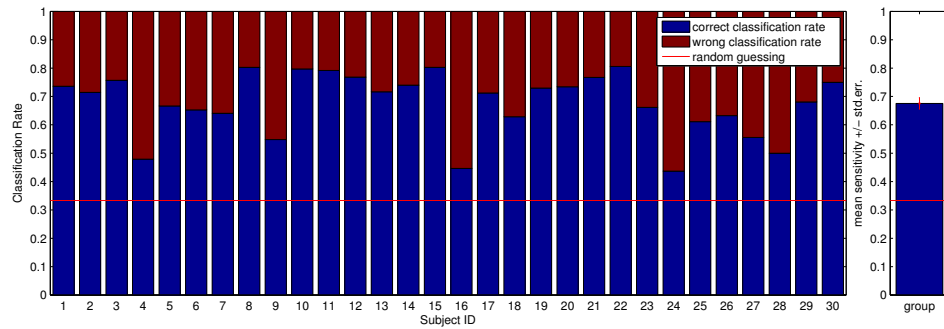


Figure B.52: M2: correct classification rates on left out subject and group mean \pm std.err.

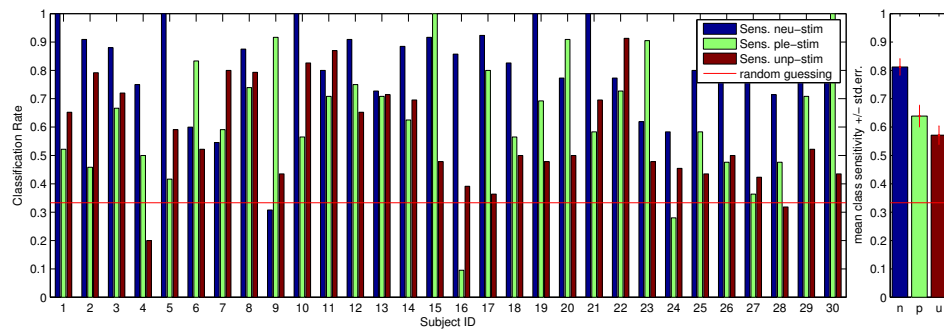


Figure B.53: M2: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

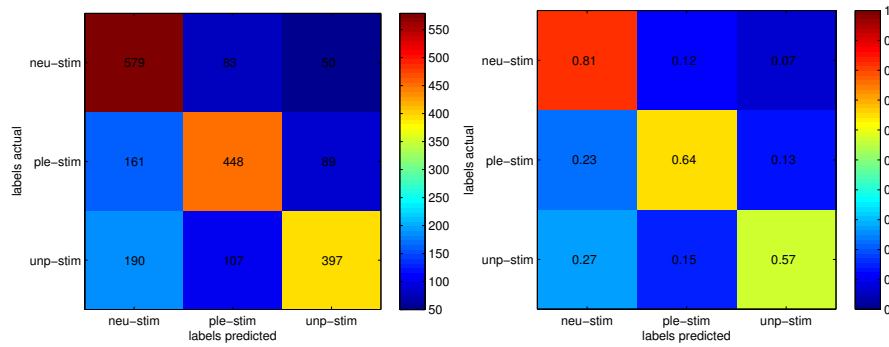


Figure B.54: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 5000 Trees – Other–Model decoding Other–Session

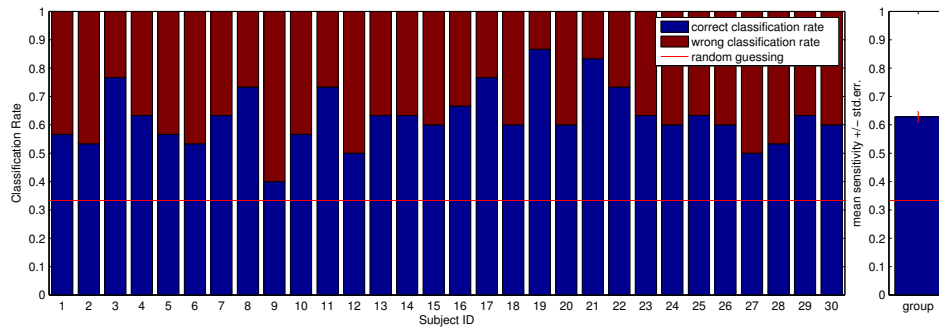


Figure B.55: M3: correct classification rates on left out subject and group mean \pm std.err.

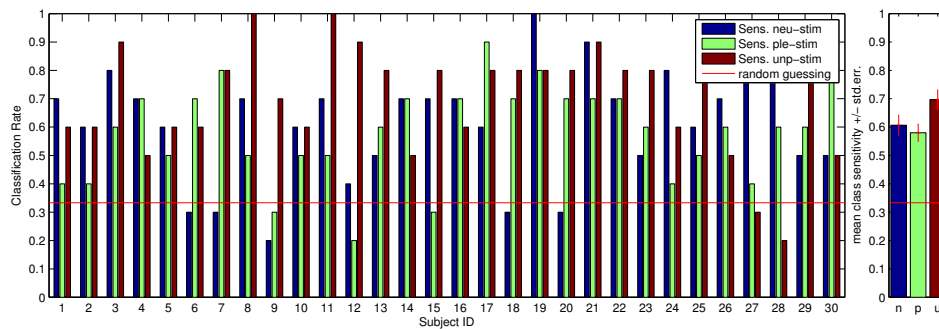


Figure B.56: M3: class-sensitivities on left out subject and mean class-sensitivity \pm std.err.

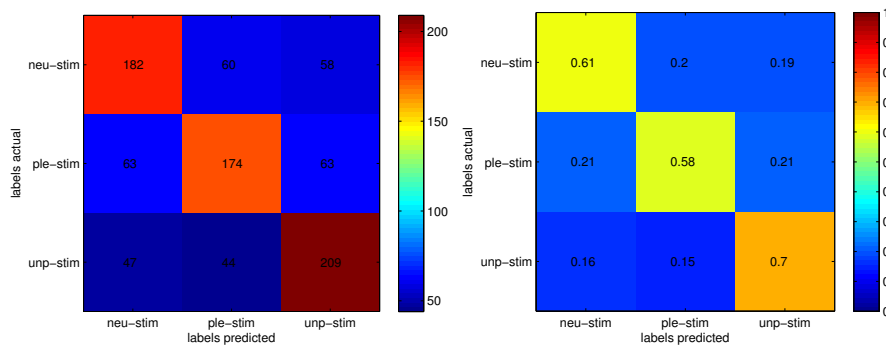


Figure B.57: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

B.2.4 Other-Model decoding Self-Session

M1 – 5000 Trees – Other-Model decoding Self-Session

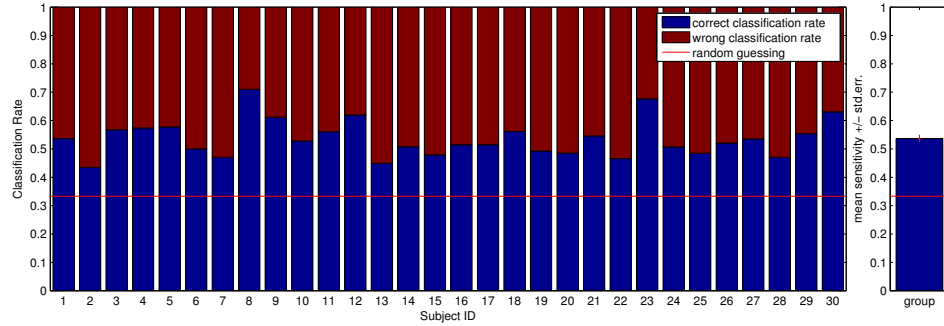


Figure B.58: M1: correct classification rates on each subject and group mean \pm std.err.

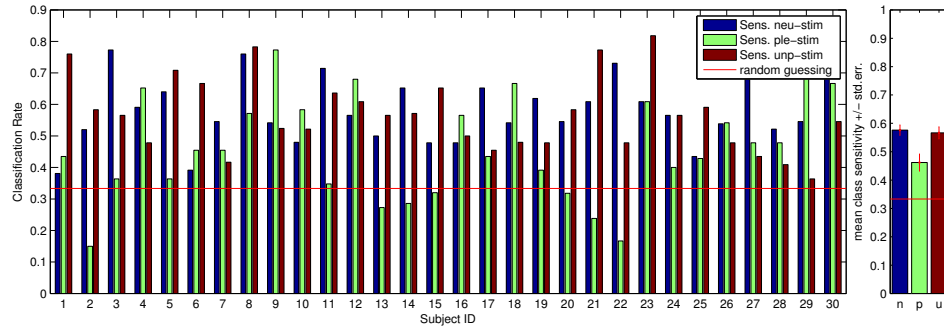


Figure B.59: M1: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

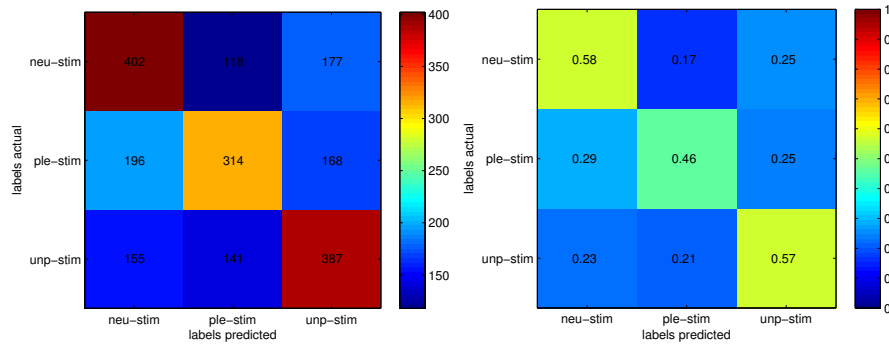


Figure B.60: M1: confusion matrix: absolute counts for each class (left) and relative values (right)

M2 – 5000 Trees – Other–Model decoding Self–Session

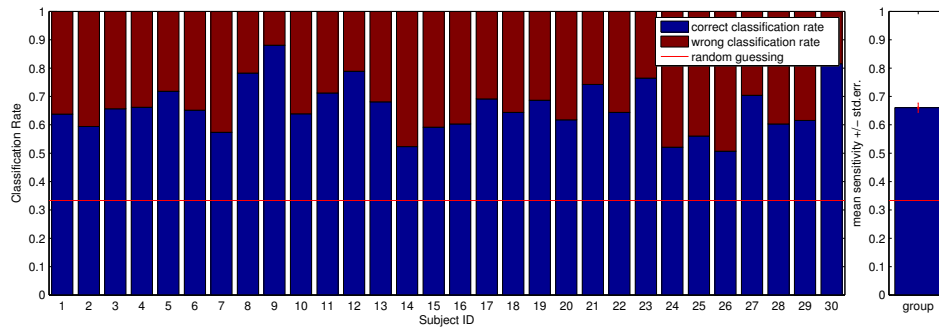


Figure B.61: M2: correct classification rates on each subject and group mean \pm std.err.

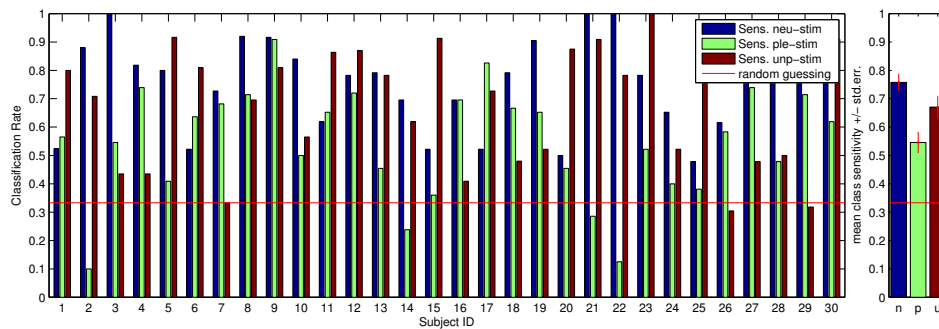


Figure B.62: M2: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

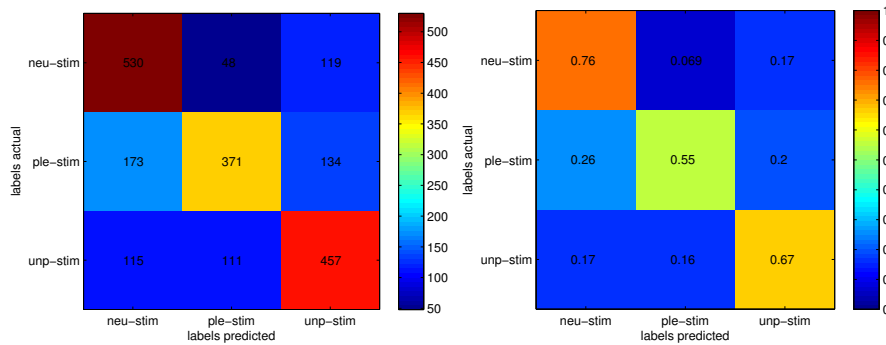


Figure B.63: M2: confusion matrix: absolute counts for each class (left) and relative values (right)

M3 – 5000 Trees – Other–Model decoding Self–Session

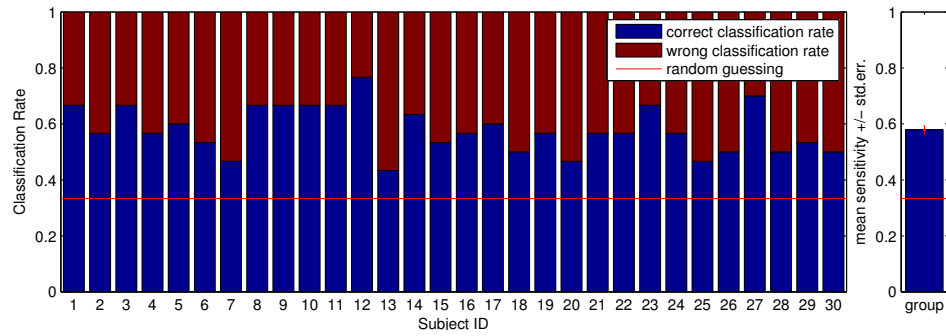


Figure B.64: M3: correct classification rates on each subject and group mean \pm std.err.

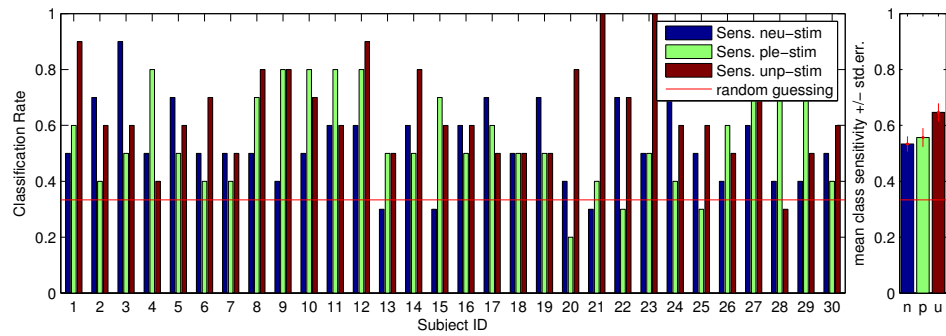


Figure B.65: M3: class-sensitivities on each subject and mean class-sensitivity \pm std.err.

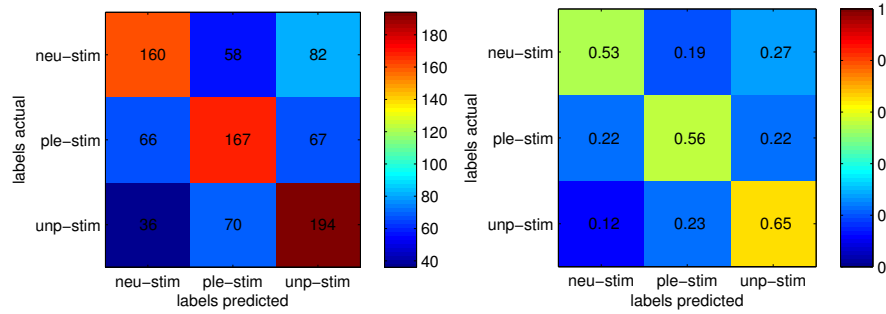


Figure B.66: M3: confusion matrix: absolute counts for each class (left) and relative values (right)

Schemaballs

This appendix contains the full set of all schemaballs derived from the self-model. Figure C.1, C.3 and C.5 depict the schemaballs for joint importances within a stimulus category; Figure C.2, C.4 and C.6 depict the schemaballs for joint importances across stimulus categories.

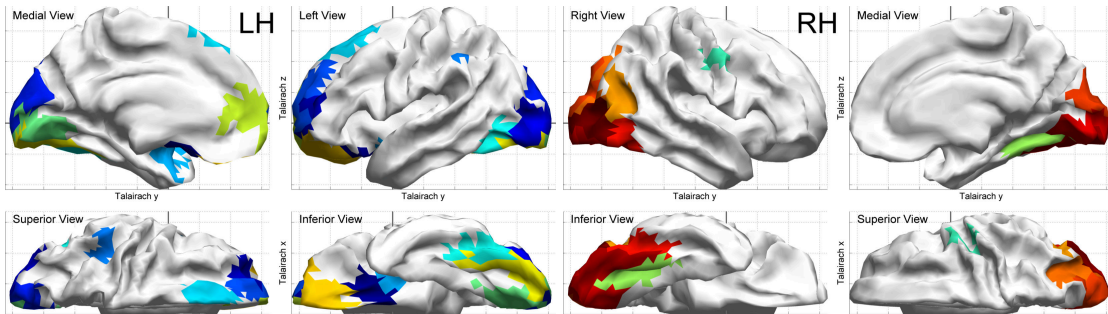
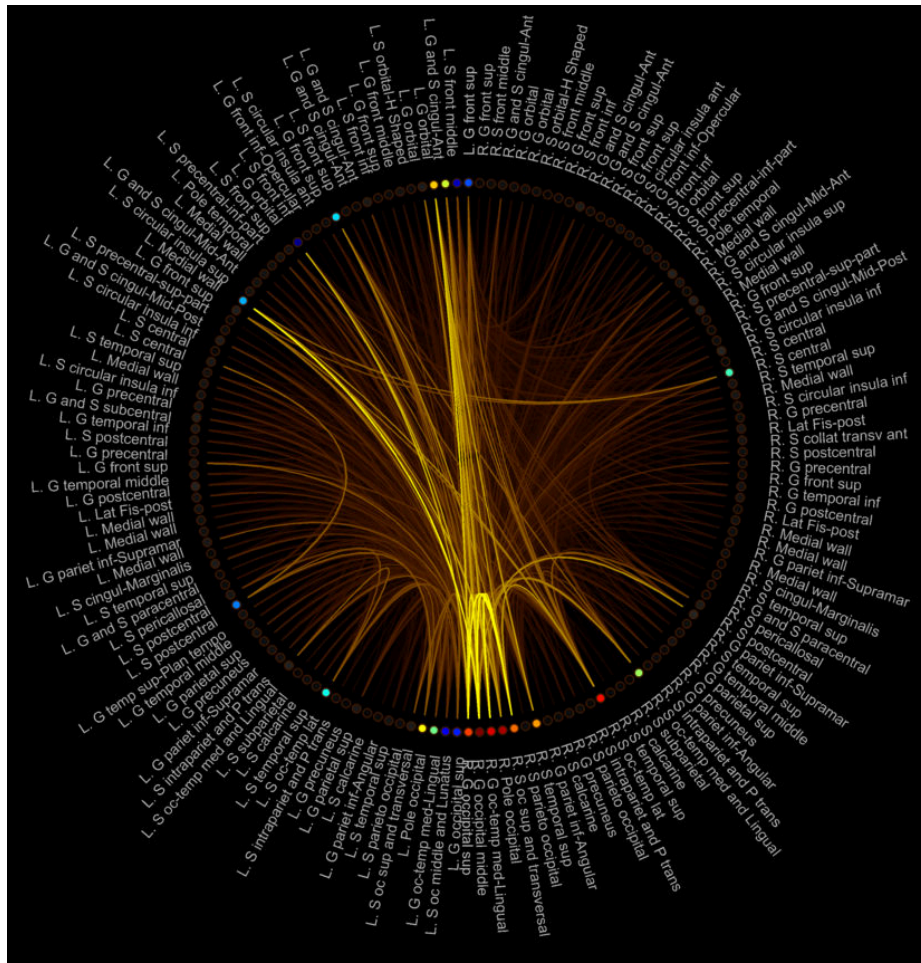


Figure C.1: MVPA: Importance of joints of parcels for decoding neutral stimuli in the self-session, RF is trained on self-data; plots below depict location of highly connected parcels

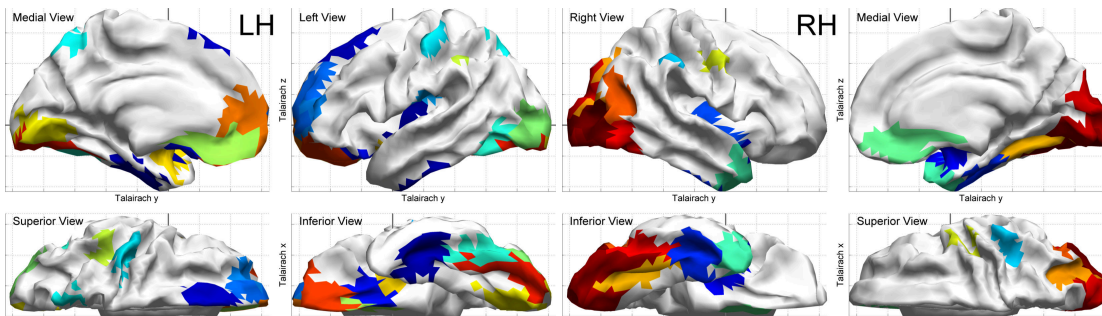
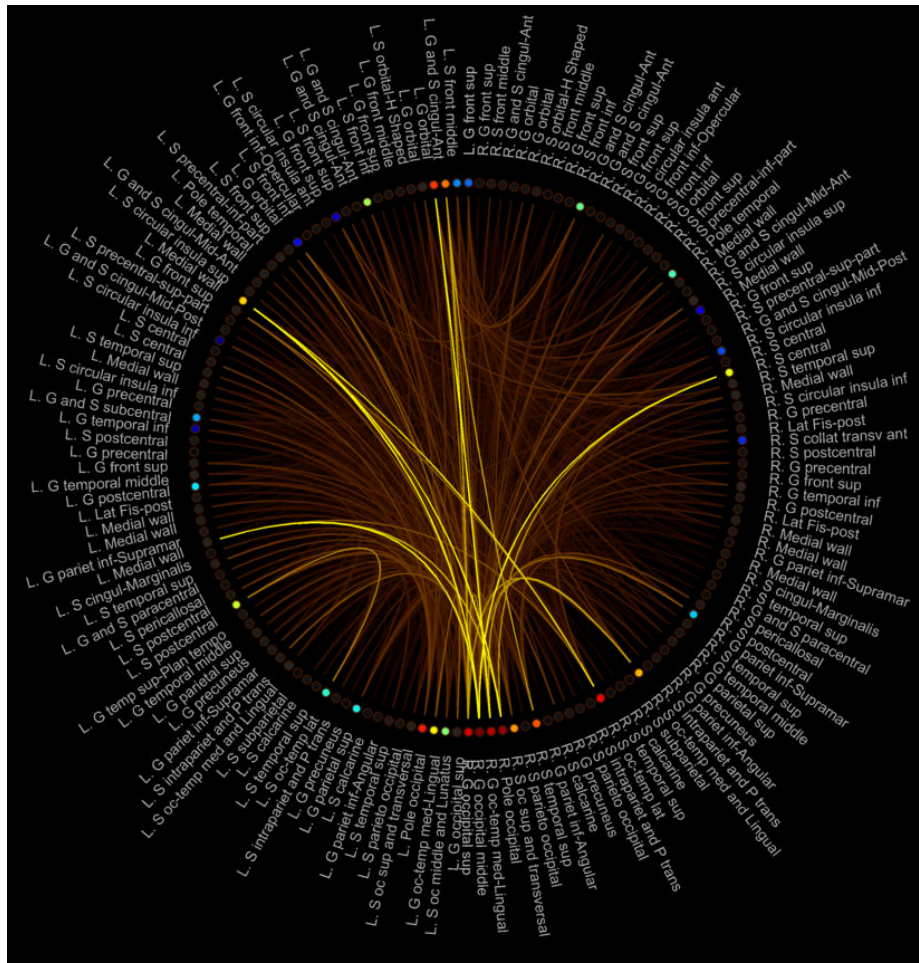


Figure C.2: MVPA: Importance of joints of parcels for decoding neutral stimuli in the other-session, RF is trained on self-data (shared patterns for neutral stimuli); plots below depict location of highly connected parcels

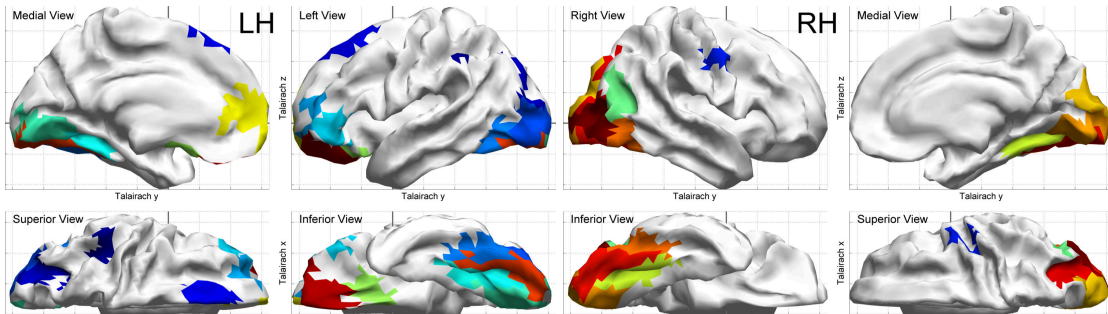
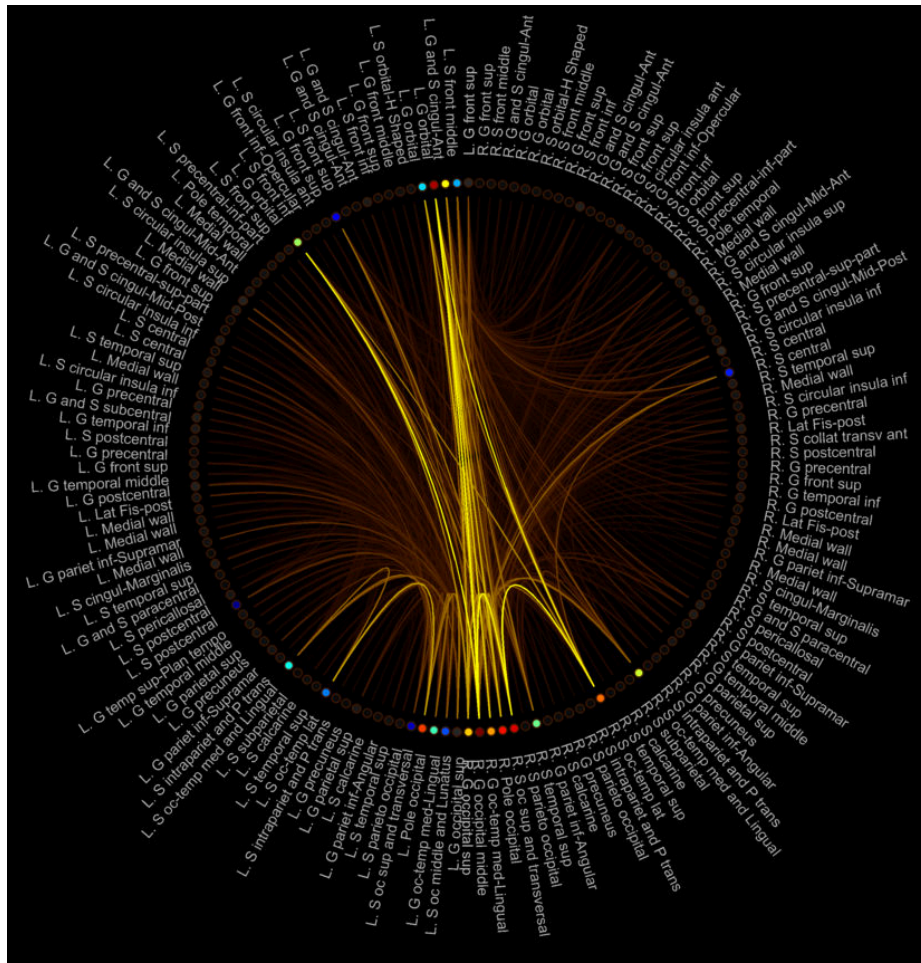


Figure C.3: MVPA: Importance of joints of parcels for decoding pleasant stimuli in the self-session, RF is trained on self-data; plots below depict location of highly connected parcels

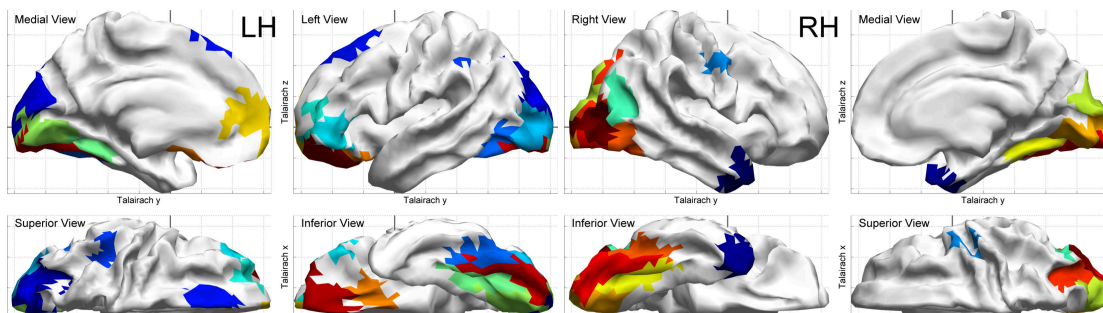
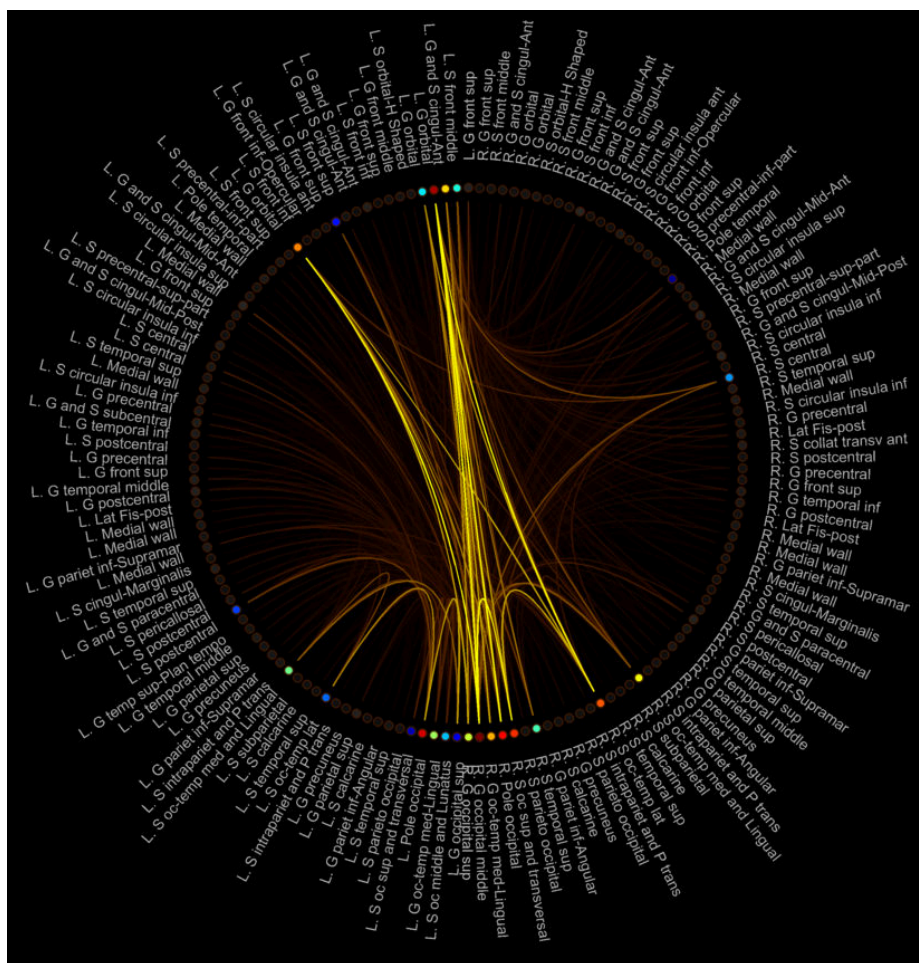


Figure C.4: MVPA: Importance of joints of parcels for decoding pleasant stimuli in the other-session, RF is trained on self-data (shared patterns for pleasant stimuli); plots below depict location of highly connected parcels

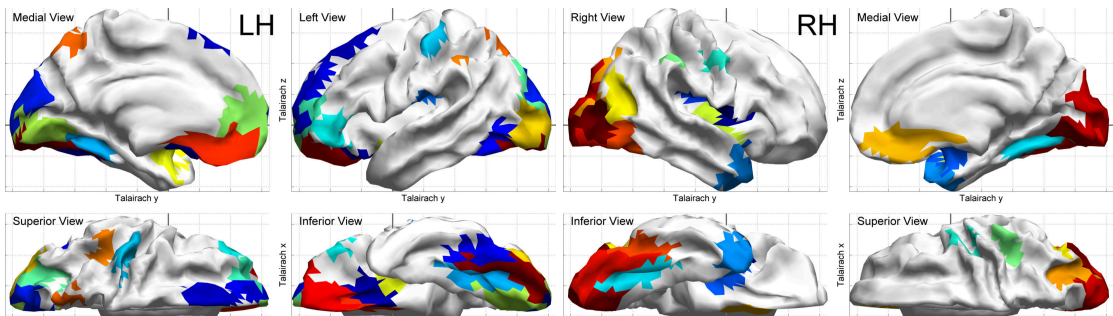
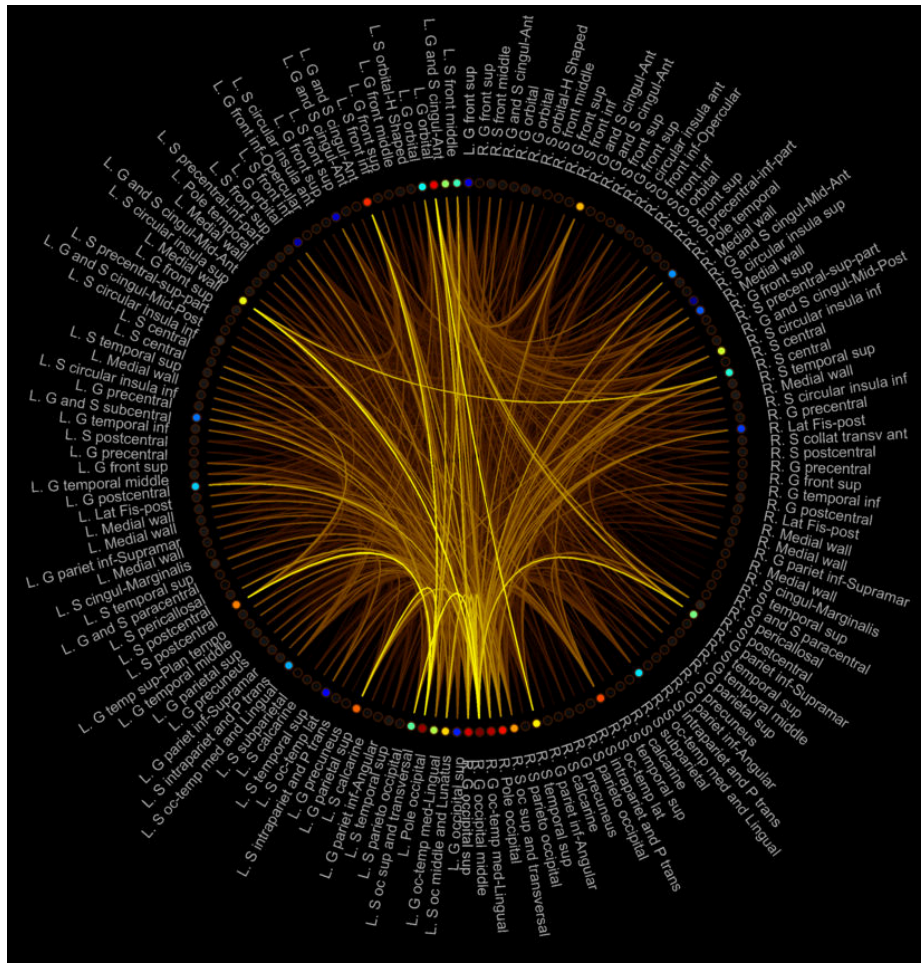


Figure C.5: MVPA: Importance of joints of parcels for decoding unpleasant stimuli in the self-session, RF is trained on self-data; plots below depict location of highly connected parcels

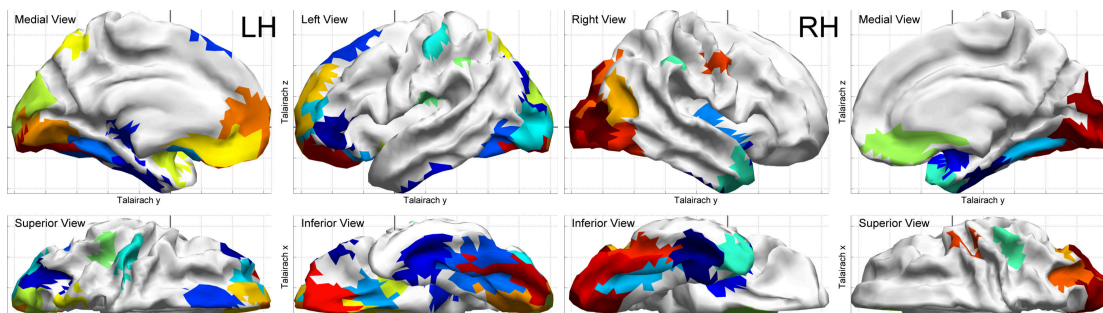
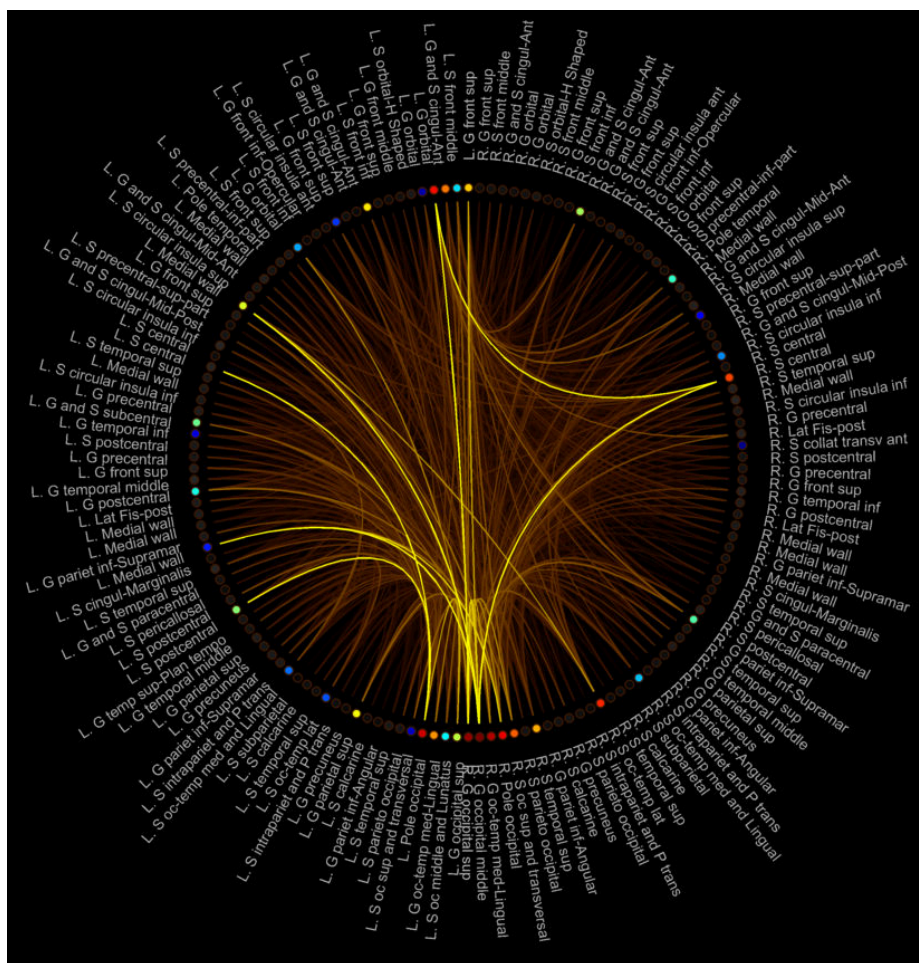


Figure C.6: MVPA: Importance of joints of parcels for decoding unpleasant stimuli in the other-session, RF is trained on self-data (shared patterns for unpleasant stimuli); plots below depict location of highly connected parcels

List of Acronyms

- ACC** Anterior Cingulate Cortex
- AI** Anterior Insula
- aMCC** anterior– Mid. Cingulate Cortex
- BOLD** Blood Oxygenation Level Dependent
- dACC** dorsal– Anterior Cingulate Cortex
- EVC** Early Visual Cortex
- FDR** False Discovery Rate
- FFA** Fusiform Face Area
- fMRI** functional Magnetic Resonance Imaging
- FS4** FREESURFER *fsaverage4*–atlas
- FS5** FREESURFER *fsaverage5*–atlas
- FWHM** Full Width Half Maximum
- GLM** General Linear Model
- HRF** Hemodynamic Response Function
- i.d.** identically distributed
- i.i.d.** independent identically distributed
- IPC** inferior parietal cortex

ISI Inter Stimulus Interval
LH Left Hemisphere
MCC Medial Cingulate Cortex
MI Middle Insula
mPFC medial PreFrontal Cortex
MRI Magnetic Resonance Imaging
MVPA MultiVariate Pattern Analysis
OOB Out Of Bag
PCA Principal Component Analysis
pIC posterior Insular Cortex
PLL Posterior Parietal Lobe
PM PreMotor Cortex
RF Random Forest
RH Right Hemisphere
ROI Region of Interest
S1 primary somatosensory cortex
S2 secondary somatosensory cortex
std.err. standard error
STS Superior Temporal Sulcus
SVM Support Vector Machine

List of Math Symbols

$\Delta I(w, \mathbf{X}_w)$ weighted reduction in Gini impurity in node w for the data \mathbf{X}_w

$\Delta i(w, \mathbf{X}_w)$ decrease in impurity at node w when splitting the data \mathbf{X}_w into two children \mathbf{X}_{w_1} and \mathbf{X}_{w_2}

$\eta^*(w, v^*, \mathbf{X}_w)$ split-value of node w

$i(\mathbf{X}_w)$ Gini impurity for the data \mathbf{X}_w

$I_G(v)$ Gini importance of feature v

$I_{CJG}(v, u)$ conditional joint Gini importance of feature v with parent split feature u

$I_{JG}(v, u)$ joint Gini importance of feature v with parent split feature u

n_w total number of examples at node w = number of columns in \mathbf{X}_w

$v^*(w)$ split-feature of node w

$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix}$ vector of weights in the GLM

$\hat{\mathbf{B}}$ matrix of estimated β -values for each feature and predictor, size is $V \times n_B$

$\hat{\mathbf{z}}$ estimated BOLD-response (estimate of \mathbf{z})

$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_q \end{pmatrix}$ contrast-vector for hypothesis test in the GLM GLM

- l vector of time discrete samples of the label function $l(t)$, length is n_m
- l_b block label vector, containing the label l for each of the n_B stimulus blocks
- $M = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q)$ is a $n_{\text{obs}} \times p$ model predictor matrix, or design matrix each column represents one model predictor
- X data matrix, size is $V \times n_{\text{obs}}$
- X^* subsample of observations or training a tree in the RF with size $V \times n_{\text{obs}}^*$
- $X_{\text{other group}}$ data matrix accumulating all observations across all subjects' data matrixes X_{other}
- X_{other} data matrix X for the other-session
- $X_{\text{self group}}$ data matrix accumulating all observations across all subjects' data matrixes X_{self}
- X_{self} data matrix X for the self-session
- X_w data assigned to node w
- X_{w_1}, X_{w_2} children datasets of the parent dataset X_w ; $X_w = X_{w_1} \cup X_{w_2}$; $X_{w_1} \cap X_{w_2} = \{\}$
- y response vector containing a label l for each observation in X , vector length is n_{obs}
- $y_{\text{other group}}$ target vector accumulating all subjects' target vectors y_{other}
- y_{other} response vector y for the other-session
- $y_{\text{self group}}$ target vector accumulating all subjects' target vectors y_{self}
- y_{self} response vector y for the self-session
- z vector of n_{obs} time discrete BOLD measurements
- Z_{other} matrix of surface aligned BOLD-measurements of the other-session, size is $V \times n_m$
- Z_{self} matrix of surface aligned BOLD-measurements of the self-session, size is $V \times n_m$
- $\hat{\beta}$ estimated value of β
- $\widetilde{y_{\text{other group}}}$ permuted version of the vector $y_{\text{other group}}$
- $\widetilde{y_{\text{self group}}}$ permuted version of the vector $y_{\text{self group}}$
- $\widetilde{OOB_b^v}$ the OOB-samples of tree b with the values of feature v permuted
- b a specific tree in the forest
- B_0 static magnetic field in MRI
- $e(t)$ error-function in the GLM

| | |
|--------------------|---|
| $h_l(t)$ | hemodynamic response of the stimulus $s_l(t)$ |
| $HRF(t)$ | HRF–function |
| L | number of different labels |
| l | label number ranging from 1 to L $l = 1$ for neutral stimuli $l = 2$ for pleasant stimuli $l = 3$ for unpleasant stimuli |
| $l(t)$ | function which assigned a label l to each point in time t |
| m | number of randomly selected features for each node in the RF–algorithm |
| $m(t)$ | predictor–function in GLM–model; \mathbf{m} is a vector of sampled points in time |
| M_0 | equilibrium magnetization in z -direction (direction of B_0) |
| M_z | magnetization in z -direction |
| M_{xy} | magnetization in plane perpendicular z -direction |
| n_B | number of experimental blocks |
| n_m | number of acquired volumes during one fMRI–scanning session. |
| n_{obs} | number of observations |
| n_{obs}^* | number of subsampled observations for training a tree in the RF |
| n_p | number of parcels per hemisphere |
| n_{trees} | number of trees in the RF |
| OOB_b | the OOB–samples of tree b |
| q | number of predictors in the GLM |
| $s_l(t)$ | stimulus function of stimulus type l |
| T_1 | spin–lattice relaxation time |
| T_2 | spin–spin relaxation time |
| T_2^* | experimentally observed decay in M_{xy} |
| th | threshold value for labeling |
| V | dimensionality of the feature space |
| v, u | specific features |

w a specific node in a decision tree
 w_{parent} parent node of node w
 $z(t)$ BOLD-measurement in one voxel

Bibliography

- [1] Andre Altmann, Laura Tolosi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, May 2010.
- [2] Edson Amaro, Jr and Gareth J. Barker. Study design in fMRI: basic principles. *Brain Cogn*, 60(3):220–232, Apr 2006.
- [3] Kellie J Archer and Ryan V Kimes. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4):2249–2260, 2008.
- [4] F. Gregory Ashby. *Statistical Analysis of fMRI Data*. MIT press, 2011. ISBN: 9780262015042.
- [5] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [6] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165–1188, 2001.
- [7] Boris C. Bernhardt and Tania Singer. The neural basis of empathy. *Annu Rev Neurosci*, 35:1–23, 2012.
- [8] F. Bloch. Nuclear Induction. *Phys. Rev.*, 70:460–474, Oct 1946.
- [9] P. A. Bottomley, T. H. Foster, R. E. Argersinger, and L. M. Pfeifer. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age. *Med Phys*, 11(4):425–448, 1984.
- [10] Leo Breiman. Random Forests. In *Machine Learning*, pages 5–32, 2001.
- [11] Leo Breiman and Adele Cutler. Random Forest Version 5.1. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm, June 2004. last access 05.12.2013.
- [12] Joseph D. Bronzino and Donald R. Peterson. *The Biomedical Engineering Handbook*. Crc Pr I Llc, 2014.

- [13] Richard B. Buxton. *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, 2009. ISBN-13: 978-0521899956.
- [14] Richard B. Buxton. The physics of functional magnetic resonance imaging (fMRI). *Rep Prog Phys*, 76(9):096601, Sep 2013.
- [15] Yi Chen, Praneeth Namburi, Lloyd T. Elliott, Jakob Heinzle, Chun Siong Soon, Michael W L. Chee, and John-Dylan Haynes. Cortical surface-based searchlight decoding. *Neuroimage*, 56(2):582–592, May 2011.
- [16] David Clark. *The brain and behavior an introduction to behavioral neuroanatomy*. Cambridge University Press, New York, 2010.
- [17] Bryan R. Conroy, Benjamin D. Singer, J Swaroop Guntupalli, Peter J. Ramadge, and James V. Haxby. Inter-subject alignment of human cortical anatomy using functional connectivity. *Neuroimage*, 81:400–411, Nov 2013.
- [18] Corrado Corradi-Dell’Acqua, Christoph Hofstetter, and Patrik Vuilleumier. Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J Neurosci*, 31(49):17996–18006, Dec 2011.
- [19] C. Davatzikos, K. Ruparel, Y. Fan, D. G. Shen, M. Acharyya, J. W. Loughead, R. C. Gur, and D. D. Langleben. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28(3):663–668, Nov 2005.
- [20] Jean Decety and John T. Cacioppo, editors. *The Oxford Handbook of Social Neuroscience (Oxford Library of Psychology)*. Oxford University Press, USA, 2011.
- [21] Jean Decety, Chenyi Chen, Carla Harenski, and Kent A. Kiehl. An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Front Hum Neurosci*, 7:489, 2013.
- [22] Rutvik Desai, Einat Liebenthal, Edward T. Possing, Eric Waldron, and Jeffrey R. Binder. Volumetric vs. surface-based alignment for localization of auditory cortex activation. *Neuroimage*, 26(4):1019–1029, Jul 2005.
- [23] Rahul S. Desikan, Florent Segonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, Jul 2006.
- [24] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, Oct 2010.
- [25] G. di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti. Understanding motor events: a neurophysiological study. *Exp Brain Res*, 91(1):176–180, 1992.

- [26] Edsger. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [27] Sjoerd J H. Ebisch, Mauro G. Perrucci, Antonio Ferretti, Cosimo Del Gratta, Gian Luca Romani, and Vittorio Gallese. The sense of touch: embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. *J Cogn Neurosci*, 20(9):1611–1623, Sep 2008.
- [28] Paul H C. Eilers and Jelle J. Goeman. Enhancing scatterplots with smoothed densities. *Bioinformatics*, 20(5):623–628, Mar 2004.
- [29] Joset A. Etzel, Valeria Gazzola, and Christian Keysers. An introduction to anatomical ROI-based fMRI classification analysis. *Brain Res*, 1282:114–125, Jul 2009.
- [30] Joset A. Etzel, Jeffrey M. Zacks, and Todd S. Braver. Searchlight analysis: promise, pitfalls, and potential. *Neuroimage*, 78:261–269, Sep 2013.
- [31] Yan Fan, Niall W. Duncan, Moritz de Greck, and Georg Northoff. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neurosci Biobehav Rev*, 35(3):903–911, Jan 2011.
- [32] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, 8(4):272–284, 1999.
- [33] Bruce Fischl. FreeSurfer. *Neuroimage*, 62(2):774–781, Aug 2012.
- [34] Bruce Fischl. FreeSurfer Software Suite. <http://surfer.nmr.mgh.harvard.edu/>, 2013. last access: 11.11.2013.
- [35] Bruce Fischl, David H. Salat, Andre J W. van der Kouwe, Nikos Makris, Florent Segonne, Brian T. Quinn, and Anders M. Dale. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23 Suppl 1:S69–S84, 2004.
- [36] Bruce Fischl, Andre van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Segonne, David H. Salat, Evelina Busa, Larry J. Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, and Anders M. Dale. Automatically parcelating the human cerebral cortex. *Cereb Cortex*, 14(1):11–22, Jan 2004.
- [37] R. A. FISHER. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, 7(2):179–188, 1936.
- [38] FreeSurferWiki. FreeSurfer recon-all. <http://ftp.nmr.mgh.harvard.edu/fswiki/recon-all>, 2 2013. last access: 10.11.2013.
- [39] K. J. Friston, C. D. Frith, and R. S. J. Frackowiak. Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping*, 1(1):69–79, 1993.

- [40] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. Frackowiak. Functional connectivity: the principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab*, 13(1):5–14, Jan 1993.
- [41] K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1(2):153–171, 1994.
- [42] Karl J. Friston. Functional and effective connectivity: a review. *Brain Connect*, 1(1):13–36, 2011.
- [43] Vittorio Gallese and Sjoerd Ebisch. Embodied Simulation and Touch: The Sense of Touch in Social Cognition. *Phenomenology and Mind*, 4:269–291, 2013.
- [44] R. Genuer, V. Michel, E. Eger, and B. Thirion. Random Forests based feature selection for decoding fMRI data. *Proceedings of the 19th COMPSTAT*, 1:1071–1078, 2010.
- [45] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. Variable selection using random forests. *Pattern Recogn. Lett.*, 31(14):2225–2236, October 2010.
- [46] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9(4):416–429, Apr 1999.
- [47] Daniel A. Handwerker, John M. Ollinger, and Mark D’Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651, Apr 2004.
- [48] Bronson Harry, Mark A. Williams, Chris Davis, and Jeesun Kim. Emotional expressions evoke a differential response in the fusiform face area. *Front Hum Neurosci*, 7:692, 2013.
- [49] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, second edition edition, 2008.
- [50] Stefan Haufe, Frank Meinecke, Kai Gorgen, Sven Dahne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, (0):–, 2013. In Press, Uncorrected Proof.
- [51] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, Sep 2001.
- [52] James V. Haxby. Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*, 62(2):852–855, Aug 2012.
- [53] John-Dylan Haynes and Geraint Rees. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7):523–534, Jul 2006.
- [54] Grit Hein, Giorgia Silani, Kerstin Preuschoff, C Daniel Batson, and Tania Singer. Neural responses to ingroup and outgroup members’ suffering predict individual differences in costly helping. *Neuron*, 68(1):149–160, Oct 2010.

- [55] Alan J Izenman. *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Springer, 2008.
- [56] Mbemba Jabbi, Marte Swart, and Christian Keysers. Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage*, 34(4):1744–1753, Feb 2007.
- [57] Abhishek Jaiantilal. randomforest-matlab. <https://code.google.com/p/randomforest-matlab/>, 2013. last access 05.12.2013.
- [58] Peter Jezzard, Paul M. Matthews, and Stephen M. Smith, editors. *Functional MRI: An Introduction to Methods*. Oxford University Press, 2002. ISBN: 978-0198527732.
- [59] Koji Jimura and Russell A. Poldrack. Analyses of regional-average activation and multi-voxel pattern information tell complementary stories. *Neuropsychologia*, 50(4):544–552, Mar 2012.
- [60] Yukiyasu Kamitani and Yasuhito Sawahata. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *Neuroimage*, 49(3):1949–1952, Feb 2010.
- [61] Yukiyasu Kamitani and Frank Tong. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5):679–685, May 2005.
- [62] Christian Keysers. *The Empathic Brain*. CreateSpace Independent Publishing Platform, 2011. ISBN 9781463769062.
- [63] Christian Keysers. The Empathic Brain – Why We Care About the People of Syria. *Psychology Today*, August:online, 2011. <http://www.psychologytoday.com/blog/the-empathic-brain/201108/why-we-care-about-the-people-syria> last access 16.12.2013.
- [64] Christian Keysers. *Unser empathisches Gehirn*. Verlagsgruppe Random House GmbH, translation by hainer kober edition, 2013. ISBN 9783641121891.
- [65] Andreas Kleinschmidt. Different analysis solutions for different spatial resolutions? Moving towards a mesoscopic mapping of functional architecture in the human brain. *Neuroimage*, 38(4):663–665, Dec 2007.
- [66] Oleg Komarov. Schemaball. <http://www.mathworks.com/matlabcentral/fileexchange/42279-schemaball>, Jun 2013. last access: 23.12.2013.
- [67] Nikolaus Kriegeskorte and Peter Bandettini. Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4):649–662, Dec 2007.
- [68] Nikolaus Kriegeskorte and Peter Bandettini. Combining the tools: activation- and information-based fMRI analysis. *Neuroimage*, 38(4):666–668, Dec 2007.
- [69] Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proc Natl Acad Sci U S A*, 103(10):3863–3868, Mar 2006.

- [70] Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci*, 2:4, 2008.
- [71] Ph.d. Kirkland Kyle. *Biological Sciences: Notable Research and Discoveries (Frontiers of Science)*. Facts on File, 1 edition, 1 2010.
- [72] Claus Lamm, C Daniel Batson, and Jean Decety. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J Cogn Neurosci*, 19(1):42–58, Jan 2007.
- [73] Claus Lamm, Jean Decety, and Tania Singer. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, 54(3):2492–2502, Feb 2011.
- [74] Claus Lamm, Andrew N. Meltzoff, and Jean Decety. How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *J Cogn Neurosci*, 22(2):362–376, Feb 2010.
- [75] J. L. Lancaster, L. H. Rainey, J. L. Summerlin, C. S. Freitas, P. T. Fox, A. C. Evans, A. W. Toga, and J. C. Mazziotta. Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp*, 5(4):238–242, 1997.
- [76] J. L. Lancaster, M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten, and P. T. Fox. Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp*, 10(3):120–131, Jul 2000.
- [77] Georg Langs, Bjoern H. Menze, Danial Lashkari, and Polina Golland. Detecting stable distributed patterns of brain activation using Gini contrast. *Neuroimage*, 56(2):497–507, May 2011.
- [78] G. Leuba and L. J. Garey. Comparison of neuronal and glial numerical density in primary and secondary visual cortex of man. *Exp Brain Res*, 77(1):31–38, 1989.
- [79] Abdelhak Mahmoudi, Sylvain Takerkart, Fakhita Regragui, Driss Boussaoud, and Andrea Brovelli. Multivoxel pattern analysis for FMRI data: a review. *Comput Math Methods Med*, 2012:961257, 2012.
- [80] Masaya Misaki, Youn Kim, Peter A. Bandettini, and Nikolaus Kriegeskorte. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, 53(1):103–118, Oct 2010.
- [81] TomM. Mitchell, Rebecca Hutchinson, RaduS. Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just, and Sharlene Newman. Learning to Decode Cognitive States from Brain Images. *Machine Learning*, 57(1-2):145–175, 2004.

- [82] Janaina Mourao-Miranda, Arun L W. Bokde, Christine Born, Harald Hampel, and Martin Stetter. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage*, 28(4):980–995, Dec 2005.
- [83] Jeanette A. Mumford and Thomas Nichols. Simple group fMRI modeling and inference. *Neuroimage*, 47(4):1469–1475, Oct 2009.
- [84] Jeanette A. Mumford and Russell A. Poldrack. Modeling group fMRI data. *Soc Cogn Affect Neurosci*, 2(3):251–257, Sep 2007.
- [85] Marieke Mur, Peter A. Bandettini, and Nikolaus Kriegeskorte. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc Cogn Affect Neurosci*, 4(1):101–109, Mar 2009.
- [86] Douglas C. Noll. A Primer on MRI and Functional MRI. Technical Report V 2.1, Departments of Biomedical Engineering and Radiology – University of Michigan, 2001. <http://web.eecs.umich.edu/~dnoll/primer2.pdf>.
- [87] Kenneth A. Norman, Sean M. Polyn, Greg J. Detre, and James V. Haxby. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*, 10(9):424–430, Sep 2006.
- [88] Nikolaas N. Oosterhof, Tobias Wiestler, Paul E. Downing, and Joern Diedrichsen. A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, 56(2):593–600, May 2011.
- [89] Hans P. Op de Beeck. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage*, 49(3):1943–1948, Feb 2010.
- [90] Alice J. O’Toole, Fang Jiang, Herve Abdi, Nils Penard, Joseph P. Dunlop, and Marc A. Parent. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J Cogn Neurosci*, 19(11):1735–1752, Nov 2007.
- [91] Marius V. Peelen, Anthony P. Atkinson, and Patrik Vuilleumier. Supramodal representations of perceived emotions in the human brain. *J Neurosci*, 30(30):10127–10134, Jul 2010.
- [92] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1 Suppl):S199–S209, Mar 2009.
- [93] Russell A. Poldrack, Jeanette A. Mumford, and Thomas E. Nichols. *Handbook of Functional MRI Data Analysis*. Cambridge University Press, 1 edition edition, 2011. ISBN 0521517664.
- [94] Jean-Baptiste Poline and Matthew Brett. The general linear model and fMRI: does love last forever? *Neuroimage*, 62(2):871–880, Aug 2012.

- [95] Sean M. Polyn, Vaidehi S. Natu, Jonathan D. Cohen, and Kenneth A. Norman. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756):1963–1966, Dec 2005.
- [96] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.*, 69:37–38, Jan 1946.
- [97] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annu Rev Neurosci*, 27:169–192, 2004.
- [98] Daniel B. Rowe and Steven W. Morgan. Computing FMRI Activations: Coefficients and t-Statistics by Detrending and Multiple Regression. Technical report, Division of Biostatistics – Medical College of Wisconsin, 2002.
- [99] Karine Sergerie, Caroline Chochol, and Jorge L. Armony. The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*, 32(4):811–830, 2008.
- [100] Amitai Shenhav, Lisa Feldman Barrett, and Moshe Bar. Affective value and associative processing share a cortical substrate. *Cogn Affect Behav Neurosci*, 13(1):46–59, Mar 2013.
- [101] Giorgia Silani, Claus Lamm, Christian C. Ruff, and Tania Singer. Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *J Neurosci*, 33(39):15466–15476, Sep 2013.
- [102] Tania Singer and Claus Lamm. The social neuroscience of empathy. *Ann N Y Acad Sci*, 1156:81–96, Mar 2009.
- [103] Kerri Smith. Brain decoding: Reading minds. *Nature*, 502(7472):428–430, Oct 2013.
- [104] S. M. Smith. Overview of fMRI analysis. *Br J Radiol*, 77 Spec No 2:S167–S175, 2004.
- [105] Carolin Strobl, Anne-Laure Boulesteix, and Thomas Augustin. Unbiased split selection for classification trees based on the Gini index. *Computational Statistics & Data Analysis*, 52(1):483–501, 2007.
- [106] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307, 2008.
- [107] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8:25, 2007.
- [108] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging (Thieme Classics)*. Thieme, 1988.

- [109] Jody Tanabe, David Miller, Jason Tregellas, Robert Freedman, and Francois G. Meyer. Comparison of detrending methods for optimal fMRI preprocessing. *Neuroimage*, 15(4):902–907, Apr 2002.
- [110] Alan Tucholka, Virgile Fritsch, Jean-Baptiste Poline, and Bertrand Thirion. An empirical comparison of surface-based and volume-based group studies in neuroimaging. *Neuroimage*, 63(3):1443–1453, Nov 2012.
- [111] Elia Valentini and Katharina Koch. Fine-grained analysis of shared neural circuits between perceived and observed pain: implications for the study of empathy for pain. *J Neurophysiol*, 108(7):1805–1807, Oct 2012.
- [112] Archana Venkataraman, Marek Kubicki, Carl-FredrikF. Westin, and Polina Golland. Robust Feature Selection in Resting-State fMRI Connectivity Based on Population Studies. *Conf Comput Vis Pattern Recognit Workshops*, 1:63–70, 06 2010.
- [113] Bruno Wicker, Christian Keysers, Jane Plailly, Jean Pierre Royet, Vittorio Gallese, and Giacomo Rizzolatti. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40(3):655–664, Oct 2003.
- [114] Keith J. Worsley. FMRISTAT - A general statistical analysis for fMRI data. <http://www.math.mcgill.ca/keith/fmristat/>, June 2006. Toolbox for MATLAB – last access 23.12.2013.
- [115] Jamil Zaki and Kevin Ochsner. The neuroscience of empathy: progress, pitfalls and promise. *Nat Neurosci*, 15(5):675–680, May 2012.
- [116] David H. Zald and Scott Rauch, editors. *The Orbitofrontal Cortex*. Oxford University Press, USA, 2006.