An Information Theoretic Approach to EEG-fMRI Integration

Dirk Ostwald

A dissertation submitted to the University of Birmingham in accordance with the requirements of the degree of Doctor of Philosophy in the College of Life Sciences, School of Psychology

June 2010

Abstract

The integration of signals from electro-encephalography (EEG) and functional magnetic resonance imaging (fMRI), acquired simultaneously from the same observer, holds great potential for the elucidation of the neurobiological underpinnings of human brain function. However, the most appropriate way in which to combine the data in order to achieve this goal is not clear. In this thesis, a symmetric and data-driven route to the integration of multimodal functional brain imaging data based on information theory is proposed. As a proof of principle, the framework, which was originally developed in the study of neuronal population codes, is applied in the experimental context of visually evoked responses and the neural underpinnings of visual perceptual decisions. The implications, benefits, and limitations of this theoretical framework for the analysis of simultaneously acquired EEG and fMRI data are discussed.

Acknowlegdements

First and foremost, I would like to thank Dr. Andrew Bagshaw for his supervision and support of this work. Without him this thesis would not have been possible. I further thank Dr. Camillo Porcaro for help with data collection and EEG data pre-processing and Dr. Stephen Mayhew and Nina Salman for help with data collection.

Table of Contents

List of Illustrations

List of Tables

List of Acronyms

ATP	Adenosin-Triphosphate
BCG	Ballistocardiogram
BOLD	Blood-oxygen level dependent
ECG	Electocardiogram
EEG	Electroencephalogram, Electroencephalography
EOG	Electroooculogram
EPI	Echo-Planar Imaging
EPSP	Excitatory Post-Synaptic Potential
ERP	Event Related Potential
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonace Imaging
HRF	Haemodynamic Response Function
IPSP	Inhibitory Post-Synaptic Potential
ITQ	Information Theoretic Quantity
LFP	Local Field Potential
LORETA	Low Resolution Electrotomography
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MR	Magnetic Resonance
MUA	Multi-Unit Activity
ROI	Region Of Interest
sCDP	Standardized Current Density Power
sLORETA	Standardized Low Resolution Electrotomography
SNR	Signal-to-Noise Ratio

TE	Echo Time
TR	Repetition Time
VEP	Visually Evoked Potential

List of Mathematical Symbols

$\mathbb{N}$	Set of natural numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of non-negative real numbers
$\mathbb{R}^n$	Set of n-dimensional real vectors
$\in$	Element of
$\subset$	Subset of
$\sum$	Sum symbol
$\int$	Lebesgue integral
$\sim$	Distributed according to

"The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them."

Sir William Lawrence Bragg (1890 - 1971)

1 Introduction

1.1.General Introduction and outline

The fundamental metaphor for the human brain at the beginning of the twenty-first century is that of an information processing device (Gazzaniga MS, 2004). It is generally believed that the brain represents information about the environment via its perceptual streams, compares and integrates this information with its prior expectations about the state of the world, and acts as such to optimize parameterized representations of its ecological requirements (Friston, 2010). Neurons, the fundamental biological unit of the brain, are thought to represent these different quantities by means of their electrophysiological behaviour (Kandel E.R., 2004).

The general mapping between external stimuli or internal cognitive or motor processes and the neurophysiological state of the brain is usually referred to as the neural code (Dayan and Abbott, 2001). If the brain is regarded as an information processing device, it is a natural intuition to aim to quantify the information the brain can represent about external or internal states. This can serve two purposes: first, it can potentially shed light on which aspects of neuronal activity represent what kind of information, in an attempt to decipher the neural code. Second, it can potentially yield estimates of the maximum amount of information the brain can represent based on its variable biological manifestation.

The neuronal state associated with a given perceptual or cognitive process can be described on many levels. For example, animal preparations allow the intracellular neuronal activity of single cells to be monitored, or the extracellular recording of neuronal population activity. Ethical considerations prohibit the invasive study of intra-cerebral neuronal activity invasively in humans in most cases. However, functional brain imaging techniques allow the

characterization of the neuronal state at the level of neuronal population activity non-invasively in humans (Huettel et al., 2004).

The aim of this thesis is to formalize the information theoretic metaphor for the human brain. To this end an information theoretic framework developed in the study of invasive electrophysiological recordings is applied to the analysis of functional brain imaging recordings in humans. The functional brain imaging technique under consideration here is the combined acquisition of the electroencephalographic (EEG) and functional magnetic resonance imaging (fMRI) data.

The outline of the thesis is as follows: Chapter 1 serves as a general introduction to combined EEG and fMRI. The physical and physiological principles that underlie each imaging modality, the artefacts resulting from simultaneous EEG-fMRI acquisition and their correction, as well as data analytical strategies previously employed to integrate both data types will be reviewed. Finally, Chapter 1 outlines the motivation of applying concepts derived from the information theoretic study of neuronal population codes to combined EEG-fMRI. Chapter 2 then introduces the information theoretic framework that will be employed in subsequent chapters. Specifically, the theoretical concepts of entropy and mutual information and their application to neuroscientific experiments will be discussed. Further, practical considerations about estimating these quantities based on limited experimental data will be reviewed. Chapter 3 presents numerical simulation results based on linear Gaussian models. These are used to demonstrate the power of the framework in the context of common assumptions about functional brain imaging data and to constrain parameters of the analysis of experimental data in subsequent chapters. The first experimental section, Chapter 4, describes an initial application of the framework to EEG-fMRI data acquired under visual stimulation. Further extensions to the experimental application of the framework are presented in Chapters 5 and 6. Specifically, Chapter 5 describes the application of the

information theoretic framework to co-registered EEG and fMRI data in a standardized anatomical space. Additionally, in Chapter 5 the influence of EEG data pre-processing on information quantities is assessed. Chapter 6 provides another extension of the application of the framework to combined EEG-fMRI data, here in terms of a more advanced cognitive-behavioural paradigm. The information theoretic framework is used to characterize the human perceptual decision system based on both separate-session and combined EEG-fMRI recordings. Eventually, the thesis will conclude with a general discussion of the benefits and limitations of applying information theoretic concepts to combined EEG and fMRI data in Chapter 7.

1.2. Physical and physiological basis of the EEG signal

The recording of electrical potentials from electrodes applied to the human scalp surface is known as electroencephalography (EEG). The human electroencephalogram was first recorded by Berger in the 1920's and since then has been widely used in both research and clinical applications (Gloor, 1969; Nunez and Srinivasan R., 2006). Recording and interpreting the EEG comprises an application of electromagnetic concepts to a biological entity, the brain. In Appendix A the fundamentals of the physics of electromagnetism are reviewed to define concepts encountered in the discussion of the physiological basis of the EEG in this section and the discussion of the magnetic resonance signal below.

The origin of the EEG signal can be discussed on at least two levels: first, the underlying cellular neurophysiological mechanisms, which, when synchronized over a large enough number of neurons, lead to the macroscopically observable signal (da Silva, 2010; Logothetis and Wandell, 2004). Second, the mechanisms, which lead to the observation of cortically generated potentials at the scalp surface, and are usually referred to as the EEG

forward problem (Hallez et al., 2007). The following paragraphs will briefly review both levels.

It is a tenet of modern neuroscience, that the basic anatomical and functional unit of nervous systems is the neuron, an electrically excitable cell (Kandel E.R., 2004). The electrophysiological properties of a neuron are determined by its characteristic expression of ion channels embedded in its cell membrane (Hille, 2001). In the absence of external stimulation, ATP-dependent ion transportation mechanisms generate the so-called resting potential across the neuron's cell membrane, which consists of a surplus of negative charge in the intracellular space. Upon stimulation of synaptic compartments, a neuron can exhibit at least two types of electrophysiological activity (da Silva, 2010): on the one hand, fast depolarisations of its membrane based on fast sodium and potassium conductances, which can result in an action potential being transmitted along the neurons axon, and on the other hand, slower membrane potential changes due to synaptic activation, commonly referred to as excitatory or inhibitory post-synaptic potentials (E/IPSPs). EPSPs are usually due to transmembrane inward currents of positively charged ions, such as sodium ions, while IPSPs are due to inward currents of negatively charged ions, such as chloride ions, or outward currents of positively charged ions, such as potassium ions. In bioelectrical terms, an EPSP leads to a current sink at the level of the activated synapse, while and IPSPs leads to a current source. As there is no accumulation of charge in the extracellular medium, inward or outward currents at the level of the synapse are assumed to be compensated by passive transmembrane currents across non-synaptic membrane parts. In sum, the emergence of a current sink or source upon synaptic activation leads to a sink-source current flow configuration in the extracellular medium, which, in combination with the non-zero resistance of the extracellular medium, gives rise to extracellular recordable potential differences (Logothetis and Wandell, 2004) (Figure 1.1. A). Of particular importance for the EEG signal are those potential

differences which are generated along the cell bodies of neurons with so-called ‘open fields’, i.e. the pyramidal layer 4 neurons of the cortex. These cells are arranged in palisades and their apical dendrites expressing the sink-source configurations discussed above are perpendicular to the cortical surface. It is generally believed (da Silva, 2010) that these cells, when activated synchronously in sufficient number, generate coherent electric fields, detectable at the scalp surface (Figure 1.1 B).

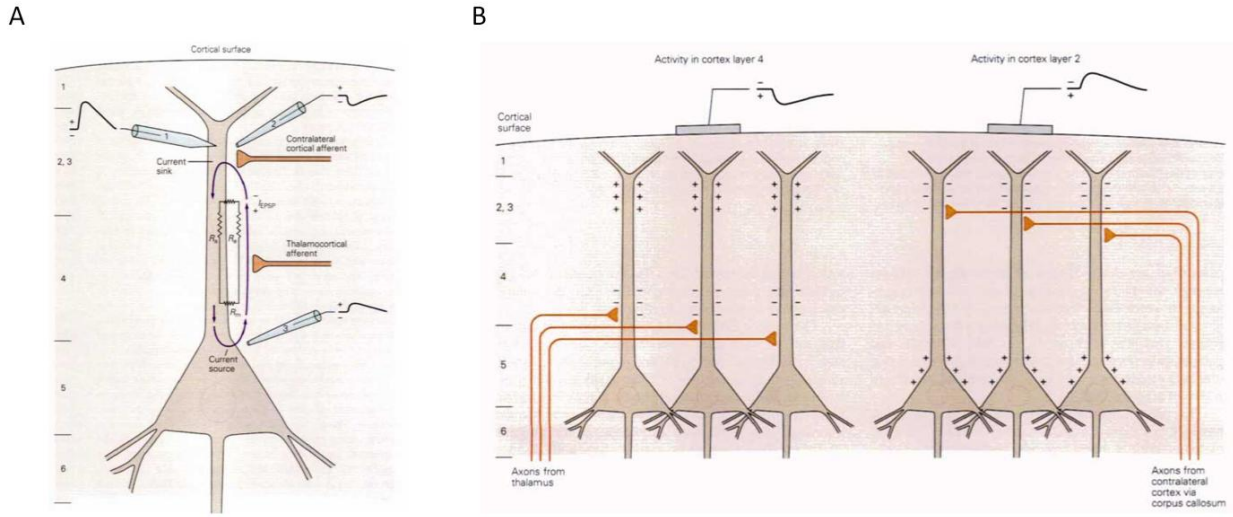


Figure 1.1 Nature of the EEG (reproduced from (Kandel E.R., 2004), Chapter 46: Seizures and Epilepsy, Figures 46-2 and 46-3) A) Emergence of current sinks and sources by excitatory post-synaptic potentials at a cortical pyramidal neuron. At the location of the EPSP (layer 2), current flows into the cell, creating a potential across the cell membrane. Further inferior at the neurons dendritic tree, current flows from the cytoplasm into the extracellular medium by passive transmembrane transport. Due to the relatively larger resistance of the cell membrane compared to the extracellular medium, the potential recorded intracellularly (electrode 1) is larger than that recorded extracellularly (electrodes 2 and 3). B) Hypothetical dependence of the EEG polarity on the location of synaptic activity within the cortical sheet. It is generally believed that the pyramidal neurons with open field configurations generate the electric fields detectable at the scalp surface by EEG electrodes. Here, different kinds of inputs can potentially lead to differential observed EEG activity: while thalamocortical input in layer 4 creates a current sink at the level of the neurons soma and a current source at the apical dendrite, which results in a downward deflection of the scalp EEG, excitatory cortico-cortical input in layer 4 results in the opposite current sink-source configuration leading to an upward deflection of the EEG. It should be noted that this is a highly idealized description. Under normal circumstances each cortical region is subject to many opposing EPSPs and IPSPs as well as thalamo-cortical and cortico-cortical inputs, and hence cortical synaptic events cannot unambiguously be deciphered from EEG recordin

The volume-transduction of electrical fields generated by the dipole sink-source configuration to the scalp surface is referred to as the EEG forward problem (Hallez et al., 2007). In principle, Poisson’s equation, a partial differential equation which governs the

potential at a given point in space due to a current dipole at another part of space, determines the recorded EEG signal for known cellular dipole sources. However, the anatomical complexity of the biological tissues and compartments involved, as well as the difficulty of experimentally establishing their respective conductances, prohibit a straight-forward solution to the forward problem. Further, even if a forward solution based on simplifying assumptions about the respective compartments is obtained, estimating the underlying cortical activity, referred to as the EEG inverse problem, is far from trivial (Grech et al., 2008). Essentially, there exist an infinite number of possible source configurations which can explain a given topographical potential distribution on the scalp. Inverse solutions hence have to be based on a number of simplifying assumptions. One of the many inverse solutions for the EEG problem, low resolution electromagnetic tomography (LORETA) will be discussed in more detail in Chapter 5 and Appendices E and F.

The investigation of the physical and physiological origins of the EEG is an active research area. As conclusion to this section, two recent studies shedding some light on the origins of the EEG from a computational (Murakami and Okada, 2006) and an experimental perspective (Whittingstall and Logothetis, 2009) will be discussed briefly.

In (Murakami and Okada, 2006), the authors employed numerical simulations of realistically three-dimensional single neuron compartment models with a number of physiological relevant membrane conductances. Four principal cell types of the cortex were simulated, stimulated at the soma, and the resulting intracellular current used to compute a current dipole vector Q for each cell type. Of all cell types simulated, pyramidal cells produced the strongest Q values. Based on the quantification that the pyramidal cell current dipole vector Q varied around 0.2 pA, and a majority of these cells showed a current towards the pia mater, the authors predict that a burst discharge by pyramidal cells may be detectable by EEG when 10,000 – 50,000 cells are synchronously active.

In their experimental study, (Whittingstall and Logothetis, 2009) simultaneously recorded scalp surface EEG and intra-cerebral electrophysiological activity from the visual cortex of alert monkeys under visual stimulation. The authors focussed their analysis on the visually evoked potential (VEP), EEG induced frequency modulations, and the multiunit spiking activity of the intracerebrally recorded signal. Using linear correlation measures, the authors provide evidence for a positive co-variation of multiunit activity with negatively oriented aspects of the visually evoked potential (N140), and negative co-variation of MUA with positive aspects of the VEP. With respect to frequency power modulations in the EEG response, the authors find the slow δ- (1-4 Hz) and fast γ-band (40-100 Hz) responses to be associated with modulation of MUA, while no co-variation between the α-band (8-15 Hz) and β-band (15-30 Hz) component and MUA was observed. The peculiar absence of correlations between MUA and the α/β-band is explained by the authors to be in line with the idea that α/β-band activity is mainly driven by neuromodulatory activity and not susceptible to either the stimulus or cortical spiking activity (Belitski et al., 2008, Mazzoni et al., 2008). Further, using a general linear model approach, the authors claim that the strongest increase in MUA is concomitant with an increase in γ-band power coinciding with the negative phase of the δ-band power. Based on their findings, the authors discuss the possibility to predict MUA from the observed EEG-frequency band coupling effect.

In summary, while the scalp recorded EEG provides temporal resolution at the level of neurophysiological events, the identification of the underlying neuronal sources, both in conceptual as well as spatial terms, is an ill-posed problem. Solutions to this problem require additional assumptions about the generation of the EEG signals, which is an active area of research.

1. 3. Physical and physiological basis of the fMRI signal

The possibility to image the metabolic consequences of neuronal activity using magnetic resonance imaging (MRI) is due to two fortunate facts: the development and advances of magnetic resonance imaging and the presence of neurovascular coupling effects in the central nervous system. The aim of this section is to briefly review the physical principles that underlie magnetic resonance imaging and the biological principles that allow the detection of metabolic changes following neuronal activation.

1.3.1 Physical Basis of the fMRI Signal

A full account of the physics of magnetic resonance imaging is beyond the scope of this thesis. Instead, the discussion below will follow an intuitive approach to magnetic resonance imaging based on a standard introductory textbook account (Deichmann et al., 2010). The following discussion will be phrased in terms of proton magnetic resonance, since most current neuroimaging procedures use a proton-based signal. However, it is also possible to obtain signals from other nuclei.

1.3.1.1 The magnetic resonance effect

The magnetic resonance effect, the basis of magnetic resonance imaging, is due to the electromagnetic properties of hydrogen nuclei, i.e. protons. Protons possess magnetic spin, resulting in them having a faint magnetic dipole moment. These dipole moments align themselves with strong external magnetic fields, usually denoted as B_0. One major component of the MR scanner is a large, helium-cooled superconductive electrical coil that generates a strong static magnetic field. MR scanners currently used for fMRI studies in humans have static fields in the range of 1.5 to 7 Tesla. The proton spins present in biological tissues, e.g. in water molecules, introduced into such a strong static field will align either parallel or anti-

parallel with the external static field. A small majority of proton spins assume a parallel alignment configuration, yielding a macroscopic net magnetization vector $M \in \mathbb{R}^3$ which, in a state of equilibrium, is aligned in parallel with the static magnetic field. The presence of a net magnetization vector, M, in the biological sample introduced into the MR scanner can be detected due to the so-called MR effect: the application of an electromagnetic wave at the resonance frequency of the net magnetization vector, the Larmor frequency, yields a spatial tilt of M, away from the external magnetic field. For a single nucleus spin, the Larmor frequency ν is the frequency of electromagnetic radiation which is required to make a spin change from the low-energy, parallel alignment to the high-energy, anti-parallel alignment with the external magnetic field B_0. It is given by

$$\nu = \frac{\gamma}{2\pi} B_0 \tag{1.1}$$

Here, γ denotes the gyromagnetic ratio of the given spin, which is the ratio between the spin's charge and mass, and hence constant for a given nucleus. For protons, γ has been determined to be approximately $2.67 \cdot 10^8 \frac{rad}{T}$. Hence, at $B_0 = 3T$ a frequency of $127.74\,MHz$ is required to change a proton spin from the parallel to the anti-parallel state, i.e. excite it, yielding a tilted net magnetization vector for the entire sample (Huettel et al., 2004).

Upon excitation, the tilted magnetization vector rotates or precesses around the static magnetic field. During precession the spins themselves generate an electromagnetic wave which can be detected by a receiver device. In most current functional brain imaging studies, the electromagnetic transducer is implemented in the main scanner compartment and the receiver devices are implemented in the head coil (e.g. the SENSE (phased-array) head coil employed in the experimental part of this thesis).

1.3.1.2 Spatial encoding of the MR signal

To encode the three-dimensional spatial position of a proton-based MR signal, magnetic field gradients in the x- and y- direction are used for frequency and phase encoding of the signal's position in the plane (slice). A further magnetic field gradient in the z-direction is used for slice selection. Each spatial encoding process will be discussed in turn.

When in supine position in the MR scanner bore, the x-direction commonly corresponds to the left-to-right axis of the subject's brain. For a basic MR sequence, an initial radiofrequency pulse will tilt the net magnetization vector M, upon which the MR signal can be detected. For spatial encoding in the x-direction, a magnetic field gradient is switched on during signal acquisition, leading to a linear variation of the static magnetic field in this direction. The acquired MR signal depends on the proton's Larmor frequency, which in turn depends on the strength of the local magnetic field. Hence, protons exposed to a slightly lower field due to the magnetic gradient will send out a slightly lower signal frequency signal than those exposed to higher field strengths. Based on the MR signal's frequency spectrum the proton's x-coordinates can then be calculated. The x-axis gradient switched on during signal acquisition is referred to as read-gradient.

For spatial encoding in the y-direction, i.e. the anterior-posterior axis of the brain for a subject in supine position, a different principle, namely phase encoding, is used. To enable phase encoding a magnetic field gradient in the y-direction is turned on between application of the radiofrequency pulse and acquisition of the MR signal. While the read-out gradient is turned on, the y-direction gradient is turned off, and the proton's frequency does not vary in the y-direction. However, due to the previous exposure to a linear varying gradient in the y-direction, the proton's phases vary linearly in the y-direction, which can be detected by frequency analysis of the MR signal.

Finally, in order to enable spatial encoding in the z-direction, i.e. the inferior-superior axis of the brain, a third magnetic field gradient is applied prior to and during the transmission of the excitatory radio-frequency pulse. As the strength of the magnetic field determines the proton's Larmor frequency, only the net magnetization vector from those protons with a Larmor-frequency equal to that of the radio-frequency pulse will be excited and contribute to the observed MR signal.

Two aspects of the relaxation process, i.e. the return of the net magnetization vector to equilibrium upon excitation, are of relevance for the creation of images: first, the recovery of the net magnetization in the z-direction (parallel to the static magnetic field) and second, the decay of the net magnetization perpendicular to the z-direction. Both processes are governed by time constants, which are referred to as T1 and T2, respectively. At a static magnetic field of 3 Tesla, the T1 time constants of grey and white matter differ by approximately 300 ms (900 ms for grey matter, 600 ms for white matter (Huettel et al., 2004)). This difference allows image grey levels to be defined corresponding to these two tissue types based on the measured T1 recovery constants. Thus imaging sequences that exploit the T1 contrast are generally used to image brain morphology. The T2 time constants between grey and white matter differ by only approximately 20 ms (100 ms for grey matter, 80 ms for white matter (Huettel et al., 2004)), hence T2 contrast imaging is not usually used to study brain morphology. However, the T2 time constant has other useful properties which are exploited for fMRI as discussed below.

1.3.1.3 Echo planar imaging

MR acquisition sequences used for functional MRI as used in this thesis are usually based upon echo planar imaging (EPI) sequences. These sequences capitalize on the use of

gradient echo sequences and a particular way of data acquisition, both of which will be discussed below.

A gradient echo imaging sequence is best understood as a variation of a gradient echo experiment. As discussed above, in an MR imaging experiment an initial radio-frequency pulse tilts the net magnetization vector M, and the MR signal is acquired in the presence of a read-out gradient in the x-direction. During data acquisition, MR signal decay is observed for the longitudinal (T1) and transverse (T2) net magnetization vector components. If during signal acquisition and after some signal decay, the read-out gradient is reversed, the detected MR signal shows a subsequent re-increase. This effect is known as gradient echo (Figure 1.2. A). The gradient echo can be explained by imagining the proton spins dephasing after the initial radio-frequency pulse, leading to the observed MR signal decay. The gradient inversion after some signal decay, by affecting the respective spin resonance frequency profile in contraposition to the initial read-out gradient, then leads to a rephrasing of spins in the sample and the observed re-build up of the MR signal. Inhomogeneities in the static magnetic field can shorten the observed T2 relaxation time constant further, resulting in what is referred to as the effective transverse relaxation time constant T2*.

A full gradient echo sequence (Figure 1.2 B) consists of the following components: during application of the radio-frequency pulse, a gradient in the z-direction is applied for slice selection, as discussed above. Next, this z-direction gradient is inverted for a slightly shorter time in order to minimize dephasing effects due to the initial slice selection gradient. A little later and overlapping with the negative z-gradient, a y-direction gradient is switched on for phase encoding of the protons in the selected plane. Next, and before the application of the x-direction read-out gradient, a negative x-direction gradient is applied in order to evoke a gradient echo during the signal acquisition time. As the signal acquisition capitalizes on spins being in phase, all effects which lead to local field inhomogeneities and hence spin dephasing,

decrease the apparent transversal decay time T2*, and hence the measured signal. This effect is exploited for BOLD-fMRI as discussed below. Last, this experiment gradient echo sequence has to be repeated with a different y-direction gradient in order to optimize spatial encoding.

The time between the initial radio-frequency pulse and the maximum of the gradient echo signal is referred to as echo time (TE). Importantly, the choice of TE determines the observed T2* contrast: if the TE is chosen too short, spins do not have sufficient time to be affected by local field inhomogeneities and to dephase accordingly, so the T2* contrast is suboptimal. However, if TE is chosen too long, the signal will have decayed before the data acquisition takes place, and again the overall signal to noise ratio will be suboptimal. Standard T2*-weighted sequences employ TEs of approximately 40 ms.

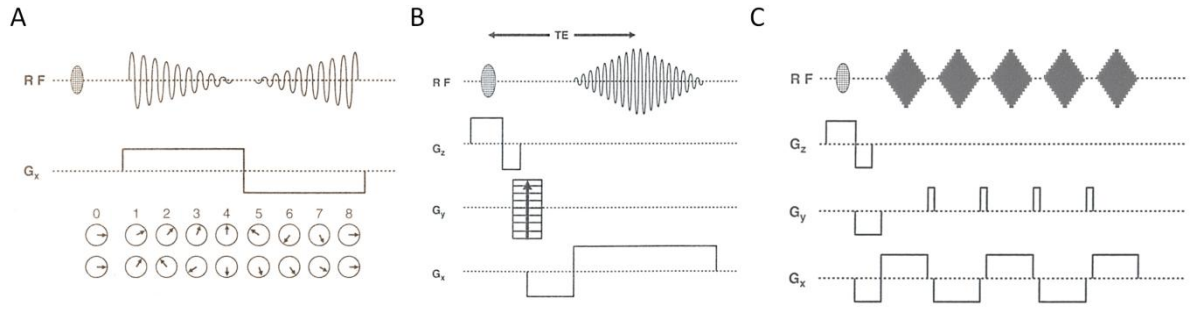


Figure 1.2 Schematic descriptions of the gradient echo effect (A), gradient echo experiment (B) and single-shot EPI sequence (C). Reproduced from (Deichmann et al., 2010), Figures 6, 7 and 9. G_x , G_y and G_z indicate magnetic field gradients in the x, y, and z-direction, respectively. RF indicates the initial excitation pulse and the observed MR signal. A) Gradient echo effect. The inversion of the read-out gradient leads to an increase in MR signal due to re-phasing of the proton spins after signal decay due to de-phasing. B) Gradient echo experiment. The initial negative x-gradient evokes spin-dephasing, while the positive read-out x-Gradient causes the spins to re-phase and the gradient echo to occur. The time difference between the initial RF excitation and the centre of the gradient echo is known as echo time (TE). C) Single-shot EPI sequence. A number of gradient echoes are acquired by repetitions of the read-out gradient inversion. The phase-encoding of y-location is achieved upon only one initial RF pulse by successive turn on of y-direction gradients (blips).

Finally, single-shot echo planar imaging (EPI) (Mansfield, 1977) allows the fast acquisition of gradient echo images without the need for repeated radio-frequency slice excitation pulses for different degrees of y-direction phase encoding. In a typical EPI

sequence, after slice selective excitation, a series of gradient echoes are acquired by successive inversions of the x-direction read gradient. Between successive acquisitions, a short y-gradient pulse, referred to as blip, is applied. As the repeated measurement for phase-encoding in the y-direction is performed with only one excitation pulse, the acquisition time per slice is short. This allows a stack of slices that covers the entire brain to be acquired in approximately 0.5 to 2 seconds. Gradient echo EPI images are strongly T2* weighted and thus strongly affected by local field inhomogeneities, as for example caused by paramagnetic deoxyhmaemoglobin as discussed in the next section.

1.3.2 Physiological basis of the fMRI Signal

The use of gradient echo pulse sequences to observe MR signal variations due to changes in blood oxygenation was first described by Ogawa et al. (Kwong et al., 1992; Ogawa et al., 1990). The effect was ascribed to a magnetic susceptibility change in the blood due to paramagnetic deoxyhaemoglobin causing an intra-voxel dephasing of proton-based MR signals. It is now commonly referred to as the blood-oxygen level dependent (BOLD) signal (Logothetis and Wandell, 2004).

Intuitively, the possibility of observing MR signal changes in response to local neuronal activation using the BOLD signal can be explained as follows (Deichmann et al., 2010): in the neuronal resting state, the local oxygen concentrations are relatively low, hence the deoxyhemoglobin concentration is relatively high. As mentioned above, deoxyhemoglobin is paramagnetic, increasing the static magnetic field, while neuropil is diamagnetic, decreasing the static magnetic field. Due to this difference, magnetic field inhomogeneities arise at the border between blood vessels and brain tissue, which shorten the effective transversal relaxation time T2* and locally reduce the MR signal in T2* weighted

gradient echo images. Local neuronal activation and its concomitant energy demand evokes the transportation of oxygen to the location of neuronal activation, resulting in a local decrease of deoxyhemoglobin and an increase of diamagnetic oxyhemoglobin. As a consequence, the existing field inhomogeneities and the blood/neuropil interface are reduced, and a local increase in the MR signal intensity is observed. It is this increase in MR signal intensity that is exploited in functional MRI studies of the brain.

It is important to note that the neuronal resting state is not an absolute resting state: in fact approximately 60% - 80% of the brain's energy demand is used to establish basic communication among neurons, while the additional energy demand for momentarily evoked activity related to the BOLD contrast may constitute as little 1% of the entire brain energy budget (Raichle et al., 2007, Raichle et al., 2006).

The explanation above provides an intuitive account of the mechanisms underlying functional MRI. However, the details of neurovascular coupling and the basis of the BOLD signal remain elusive. There exists general agreement that neuronal activation triggers a cascade of mechanisms leading to changes in the local cerebral metabolic rate of oxygen consumption ($CMRO_2$), cerebral blood flow (CBF) and cerebral blood volume (CBV). Nevertheless, it is for example less clear, what is meant by 'neuronal activation' in terms of the dynamics of different cortical populations or the precise trigger of the vascular cascade (Stephan et al., 2004). In the following, the standard Balloon model of neurovascular coupling and the current view on the neural correlates of the BOLD contrast will be reviewed briefly (Deichmann et al., 2010).

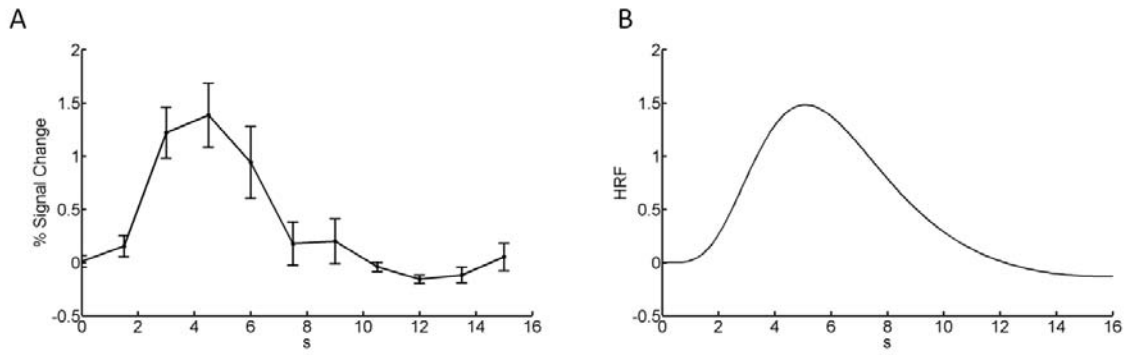


Figure 1.3 Experimental and mathematical haemodynamic response. A) Experimentally measured haemodynamic response as percent signal change with respect to pre-stimulus baseline. Errorbars indicate the SEM across observers (n = 12). The experimental details are described in Chapter V, Figure 5.12 D, high contrast condition. B) Scaled double gamma model of the haemodyamic response as used in GLM approaches to fMRI data analysis and implemented in SPM5 (Friston, 2007). The function is given by $\boldsymbol{h(t)} = \frac{\boldsymbol{l_1 h_1 t^{(h_1-1)} e^{(-l_1 t)}}}{\boldsymbol{\Gamma(h_1)}} - \boldsymbol{c}\frac{\boldsymbol{l_2 h_2 t^{(h_2-1)} e^{(-l_2 t)}}}{\boldsymbol{\Gamma(h_2)}}$, where $\boldsymbol{t}$ is the time in seconds, $\boldsymbol{l_1 = l_2 = 1}$, $\boldsymbol{h_1 = 6}$, $\boldsymbol{h_2 = 16}$, $\boldsymbol{c = 1/6}$ and $\boldsymbol{\Gamma(h)}$ is the gamma function given by $\boldsymbol{\Gamma(h) := \int_0^\infty u^{h-1} e^{-u} du}$ (Lu et al., 2006). Both the experimental and model HRF do not exhibit the initial dip.

The typical haemodynamic response, i.e. the localized MR signal change in T2* weighted EPI volumes in response to external stimulation, has the following time course: a positive signal shift persists for 5 – 10 seconds, peaking usually approximately 4-6 seconds post-stimulus. The entire haemodynamic response might last up to 30 seconds, often with a final signal undershoot, before returning to pre-stimulus baseline. Additionally, and somewhat controversial (Huettel et al., 2004), the signal time course might exhibit an 'initial dip'. This phenomenon is likely of microvascular origin and its observation hence strongly dependent on B0 field strength. Further, it may be obscured by averaging over a large spatial volume and a low sampling rate. Figure 1.3 A. shows an experimental visually evoked BOLD time-course averaged over visual cortex of data used in this thesis, not showing the initial dip phenomenon and a common model for this kind of BOLD response.

In terms of $CMRO_2$, measuring the cerebral oxygen metabolism per brain tissue volume over time, CBF, measuring the delivery of arterial blood to brain tissue over time, and CBV, measuring the blood volume per brain tissue volume, this observed response is

explained as follows (Buxton, 2001; Buxton et al., 2004): upon neural activation, energy demand, and hence $CMRO_2$, increases at the site of neural activity. This in turn leads to a momentary increase in deoxyhemoglobin and hence a T2* signal decrease (this is the origin of the initial dip). Neurovascular coupling mediators such as the glucose metabolism or NO concentration next trigger an increase in both CBF and CBV. Changes in these two quantities have opposing effects on the observed signal: the increase in CBF leads to a transportation of oxygenated blood to the site of neuronal activation and reduces the relative deoxyhaemoglobin concentration. The increase in CBV is concomitant with an increase in the relative deoxyhaemoglobin concentration. However, the fact that the variation in CBF outpaces the effects of both $CMRO_2$ and CBV for some time results in the positive BOLD signal response (Windkessel-effect). After approximately 10 seconds, the return of $CMRO_2$ and CBF to baseline and the delayed relaxation of CBV then explains the observed undershoot. The implementation of an augmented version of this basic Balloon model in terms of differential equations forms the basis of forward modelling approaches to fMRI data analysis, such as dynamic causal modelling (Friston et al., 2003).

BOLD responses can reliably be observed in response to external stimulation, and are under normal conditions triggered by preceding neural activity. However, the question of which aspect of neuronal activity is the primary trigger for BOLD responses remains open. From studies using combined intracortical electrophysiological recordings and metabolic measures such as BOLD and near infra-red spectroscopy the following picture emerges (Logothetis et al., 2001; Mathiesen et al., 1998; Shmuel et al., 2006): intracortically recorded broadband neural activity is usually classified into local field potentials (LFPs), referring to activity in the range of less than 150 Hz, and multi-unit activity (MUA), referring to activity in the range of above 300 Hz. LFPs are thought to reflect synaptic integration processes, such as EPSPs and IPSPs, while MUA is thought to reflect the spiking output of large pyramidal

cells (Logothetis and Wandell, 2004). In general, BOLD activity correlates with both LFPs and MUA, and within a small dynamic range does so in a linear fashion (Shmuel, 2010). Outside this range, saturation effects of both neural and metabolic activity have been observed (Matthiesen et al. 1998; Devor et al., 2003; Jones et al., 2002; Sheth et al. 2004; Hoffmeyer et al., 2007). Studies dissociating the locally highly correlated LFPs and MUA indicate that metabolic/hemodynamic responses are more robustly evoked by LFP activity (regional synaptic input activity), rather than MUA activity (regional spiking output activity). In terms of their spatial co-variation, the point spread function of the haemodynamic response (i.e. the spatial extent of observed BOLD responses to a spatially focussed neural event) has been estimated to lie in the range of 2 – 4 mm (Engel et al., 1997), and is inversely proportional to the B0 field strength.

On the level of the scalp EEG, current experimental evidence points to conditional dependencies (i.e. co-variation at constant stimulus) between EEG frequencies features in the α- (e.g. de Munck et al. 2007), β- (Singh et al., 2002), γ- (e.g. Rosa et al.,2009) range, or an interaction of them (e.g. Kilner et al., 2005, de Munck et al., 2009)) and the BOLD response. However, the relationships between intracranially recorded electrophysiological signals and scalp EEG signals remains elusive, as discussed in section 1.2.

In summary, the details of both neurovascular coupling and the electrophysiological activity that underlies the BOLD signal remain an active research area.

1.4 Simultaneous EEG-fMRI: methodology and artefacts

1.4.1 General methodology

Historically, the concurrent acquisition of EEG and EPI data was motivated by the aim of achieving higher spatial precision in the topographical mapping of epileptic foci (Ives et al., 1993). The simultaneous acquisition of both imaging modalities is required in this context, as the interictal epileptiform discharge, the most common EEG abnormality associated with epilepsy, is a pure EEG phenomenon with no external manifestation (Lemieux and Mulert, 2010).

Recording the EEG inside the MR environment is a challenging task, for both safety and data quality reasons. However, over the last two decades, MR compatible EEG recording systems have become commercially available (Villringer et al., 2010), which has increased the pace at which EEG-fMRI developments are being made. The MR environment requires a number of modifications to clinical EEG recording systems besides the obvious one of being non-ferrous for those components of the recording setup (electrodes, leads, cap, gel, amplifiers, amplifier leads) that enter the MR scanner room (Allen, 2010). Specifically, to minimize electromagnetic induction, the electrode leads are best bunched together and led straight to the amplifier. Further, as movement of electric conductors in the static magnetic field results in induction effects, movement of equipment due to scanner vibration is best minimized by using sandbags to stabilize electrode leads and amplifiers. Due to the large induction effects caused by the time-varying magnetic fields used for MR imaging (see below), the EEG amplifier requires a relatively large dynamic range to avoid signal saturation. Finally, the initial EEG sampling rate is usually an order of magnitude larger than in clinical contexts (in the experimental studies of this thesis 5000 Hz, while for clinical settings often

around 500 Hz (Allen, 2010)), owing to the need of characterizing the imaging artefact to high temporal precision.

With respect to subject safety, a large number of studies have repeatedly shown that if standard MR safety procedures are followed, EEG data can be recorded without causing any harm (Allen, 2010). Over the last decade, it hence has become relatively straight-forward to acquire EEG and fMRI data simultaneously. Thus, current methodological research focuses on questions of data quality and approaches to data integration (Bagshaw and Warbrick, 2007).

1.4.2 EEG artefacts

EEG data recorded in the MRI scanner environment are contaminated by artefacts which result from interactions between the EEG and MRI recording setup and are not observed on EEG data recorded under standard clinical conditions. The two main artefacts observed are referred to as the image acquisition artefact, or MR artefact (Ritter et al., 2010), and the pulse, or ballistocardiogram (BCG) artefact (Debener et al., 2010). In the following, their characteristics, assumed physical origins, and standard correction methods will be reviewed briefly.

1.4.2.1 The image acquisition artefact

As discussed in Appendix A the application of a time-varying magnetic field to an electrical conductor causes the generation of an electric potential in the conductor by means of electromagnetic induction. Further, as discussed in section 1.3.1.3 spatial signal encoding for echo planar imaging capitalizes on the application of radio-frequency pulses and the quick alteration of magnetic field gradients, i.e. time-varying magnetic fields. The changing magnetic fields induce electrical potentials in the electrical circuit formed by EEG electrodes

and leads, the participant and the EEG amplifier. These potentials are recorded by the EEG recording applications and can be more than two orders of magnitude larger than the physiological EEG signal. The largest rate of change of the magnetic field, and hence the largest induced potential, occurs during the application of the slice selection radio-frequency pulse (Ritter et al., 2010). However, the frequency of the radio-frequency pulse is extremely high (64 MHz at 1.5 T (Anami et al., 2003)) and lies outside the range of the conventional MR compatible amplifiers, and hence is largely attenuated. The artefacts induced by the switching of magnetic gradients for spatial MR signal encoding are of lower frequency content and dominate the EEG signal recorded during MR data acquisition. In fact, the EEG frequency spectrum is largely obscured by harmonics of the slice repetition frequency in the range of 10-25 Hz and harmonics of the volume repetition frequency of 0.2 – 2 Hz (Mandelkow et al., 2006). The artefact's temporal profile is largely consistent across different EEG electrodes. Nevertheless, the exact electrode position as well as further induction effects, for example through participant movement, can lead to some degree of variation between EEG channels.

The principle approach to correct the image acquisition artefact exploits its high predictability which is due to the pre-programmed nature of the MR signal acquisition sequence (Allen et al., 2000). During offline EEG data pre-processing, an artefact template is computed as the average EEG signal across a number of MR volume acquisitions of a specific channel, and subtracted from the observed signal. Assuming that the artefact is not correlated with the physiological signal, this subtraction procedure yields the uncontaminated EEG. In order to allow temporally precise sampling of the MR artefact and thus ensure the creation of an appropriate subtraction template, synchronization of the EEG and MR clocks has been shown to be beneficial (Mandelkow et al., 2006). Synchronized acquisition, and ensuring that the actual MR repetition time is a multiple integer of the EEG sampling interval, then results

in the acquisition of a stationary artefact. In order to account for low-frequency variations in the MR artefact shape due to additional effects such as participant's motion, a sliding average subtraction approach can be beneficial. However, as fewer samples are used in this case, the characterization of the artefact might actually deteriorate. Last, residual high-frequency components of the MR scanner artefact can be reduced using low-pass filtering with a cut-off in the 25-50 Hz range. As traditionally there is a prominent interest in the lower EEG frequency band (<50 Hz), especially with respect to event-related potentials, this filtering approach is not necessary detrimental to the study of EEG signals of interest.

1.4.2.2 The Pulse Artefact

The cardiac pulse-related or ballistocardiogram artefact can be observed when recording EEG in the environment in the absence of MR data acquisition. It is of lower amplitude than the MR acquisition artefact, typically in the order of 50μV at 3T and of mesogenous (i.e. exo- and endogenous) origin (Debener et al., 2010): it results from the interaction of the participants active cardiovascular system and the main static B_0 field of the MR scanner. Its main characteristic is its synchrony with the cardiac rhythm of the participant, which is usually captured simultaneously with the EEG. The delay of the BCG artefact of approximately 200 ms with respect to the R-peak of the electrocardiogram (ECG) indicates that it is not merely a volume conducted ECG artefact.

The BCG artefact has a complex topographical and temporal profile (Debener et al., 2010): The peak amplitude of the BCG artefact is variably expressed across different participants, EEG channels and MR scanners. In general EEG channels further removed from the EEG reference electrode show larger BCG artefact amplitudes than those closer to the reference electrode. Further, the size of the artefact is proportional to the strength of the B_0 field (Debener et al., 2007a; Debener et al., 2008). The mean global field power of the EEG

indicates that the BCG artefact starts approximately 150 ms after the Q-wave of the ECG and has two peaks of different topographical profile at 230 and 330 ms, while lasting until approximately 500 ms after the Q-wave. In contrast to the MR artefact the BCG artefact shows a high degree of variability over time, due to its biological nature.

The physical origin of the BCG artefact remains elusive (Debener et al., 2010; Nakamura et al., 2006, Mullinger et al., 2008). While it is clear that it must result from the movement of electrically conductive material in the B_0 field, the type of motion relevant for the BCG artefact is under discussion. Some theories postulate heartbeat-related axial head rotation, while others favour local pulsatile movements of scalp vessels on adjacent electrodes. Most likely, both movement types interact in a complex manner, leading to the observed induction effects.

The principle approach to BCG artefact correction is similar to that of MR artefact correction and was introduced by Allen et al. (Allen et al., 1998): based on successive observation of the BCG artefact of a specific EEG channel, an average artefact template is computed and subtracted from the EEG trace. However, the biological nature of the artefact makes this approach more complex than is the case for the MR artefact: first, the onset of the cardiac cycle as indexed by the R-wave of the ECG has to be identified, and second, the temporal fluctuation of the BCG artefact over time has to be accounted for. A popular method for BCG artefact correction was implemented by Niazy et al. (Niazy et al., 2005). Here, BCG artefact templates are generated based on a channel-wise temporal principal component analysis, which relaxes the requirement of temporal stability for template subtraction.

1.4.3 fMRI Artefacts

Artefacts in the gradient echo EPI images acquired in the presence of an EEG recording system are due to the field inhomogeneities introduced by the magnetically susceptible

materials of the electrode cap, as well as the increased subject discomfort (Carmichael, 2010). The artefacts can be classified into the effects on the static magnetic field, the RF transmission and reception process, as well as general effects on the MR signal SNR. Each will be discussed in turn.

The Ag/AgCl electrodes used in combined EEG-fMRI recordings cause perturbations of the static B_0 field over approximately 10-15 mm at 4T (Mullinger et al., 2008; Stevens et al., 2007), while the electrode gel has a magnetic susceptibility close to that of human tissue. This increased field inhomogeneity with respect to fMRI without an EEG system results in increased EPI image artefacts, including signal drop-out, spatial distortion and potentially ghosting. However, as the inhomogeneity effects are restricted to the locations of the electrodes on the scalp, the EPI image artefacts do not in general affect the imaging of the brain. From a data analytical viewpoint they are hence negligible. The effect the presence of the EEG electrode has on the radio-frequency excitation and reception fields, which potentially could produce regional intensity variation without increased distortion, is weak (Carmichael, 2010). Finally, the combination of both effects (and further nonlinearities as the effect of nyquist ghosting and lower shimming quality (i.e. B0 field inhomogeneities)) potentially lead to a lower SNR of the EPI images. EPI SNR can be assessed for example by comparing the mean signal of a voxel to its variance over time (i.e. multiple volume acquisitions). However, again this effect appears to be negligible for commonly used field strengths and EEG recording systems (Bonmassar et al., 2001a).

In sum, the presence of the electrode cap has some effect on the quality of the EPI images obtained. However, the effects are primarily signal drop-outs at the location of the electrode which are relatively small in diameter (< 2 cm) and hence in general do not affect the imaging of the observer's brain.

1.5. Approaches to EEG-fMRI signal integration

Approaches to the integration of simultaneously acquired EEG-fMRI data can on a first level be classified as data- vs. model-driven approaches. Data-driven approaches are understood as those approaches that quantify aspect of the respective signal modalities and their inter-modality links (Benar et al., 2010; Villringer et al., 2010). These methods are usually of asymmetric nature, i.e. the analysis of one imaging modality is augmented by information derived from the other modality. However, recent developments strive for symmetry, i.e. inter-modality interactions, also for the case of data-driven approaches (Correa et al., 2010; Eichele et al., 2008; Eichele et al., 2009). Model driven approaches are based on computational biophysical models that make quantitative predictions about both EEG and fMRI signals arising from an underlying postulated concept of neural activity (Daunizeau et al., 2010; Valdes-Sosa et al., 2008). Although providing a more principled strategy than data driven methods to EEG-fMRI integration, model driven approaches for EEG-fMRI integration have so far received less attention in the literature and will not be discussed in further detail here.

Two forms of data driven approaches to EEG-fMRI integration dominate the literature to date: EEG-fMRI integration through prediction and EEG-fMRI integration through constraints (Kilner et al., 2005). Integration through prediction usually refers to the use of EEG signal features as additional regressors in a brain-wide general linear model (GLM) analysis of fMRI data. The idea behind this approach is to assess co-variations in the voxel-wise BOLD amplitude with temporally resolved aspects of the EEG, for example the modulation of visually evoked potential components or frequency band power. The integration through prediction approach has successfully been employed in studies in epilepsy research (Carney and Jackson, 2010; Siniatchkin and Dubeau, 2010; Walker et al., 2010), auditory perception (Mayhew et al., 2010), pain perception (Iannetti and Mouraux, 2010;

Mobascher et al., 2009) and cognition (Debener et al., 2006; Eichele et al., 2005; Iannetti and Mouraux, 2010; Karch and Mulert, 2010). Integration through prediction, i.e. the EEG-feature-modulation mass-univariate GLM-fMRI analysis is currently the dominating approach to EEG-fMRI integration. An open question in this line of research is which EEG features are ideally suited for this type of analysis (Bagshaw and Warbrick, 2007; Ostwald et al., 2010; Warbrick et al., 2009).

The integration of EEG and fMRI through constraints is in many ways orthogonal to the integration through prediction approach. While in the integration through prediction approach a standard fMRI data analysis is augmented by EEG data, the reverse is true for integration through constraints. The starting point of this type analysis is a source localization algorithm for EEG whose inverse solution is in some way constrained by the topographical features of the concomitant BOLD signal (Benar et al., 2010). For example, significant localized BOLD activations can be used as spatial initial conditions in dipole modelling approaches. More advanced versions of the integration through constraints were proposed by (Daunizeau et al., 2007) and (Daunizeau et al., 2007; Henson et al., 2010). Here, the authors used a variational Bayesian framework in order to weigh the contribution of EEG and fMRI data to the EEG inverse solution based on their respective variances.

According to the classification above, the information theoretic approach discussed in this thesis can be regarded as a data-driven approach, and is conceptually closer to the integration through prediction GLM framework than the integration through constraints approach. A discussion of the relation of the information theoretic approach with the various approaches to EEG-fMRI integration proposed previously is provided in the general discussion (Chapter 7).

1.6. Motivation of the information theoretic approach to EEG-fMRI integration

The information theoretic framework applied to combined EEG and fMRI signals in this thesis has been pioneered in the analysis of signals from invasive electrophysiological recordings, where it has been applied mainly to spike count data (Panzeri et al., 2008; Schneidman et al., 2003). In this field, information theory has been successfully employed regarding questions of sensory coding (Magri et al., 2009). For example, information theoretic quantities have been used to determine the stimulus selectivity of single neuronal units by revealing which stimulus features modulate neuronal responses most reliably (Adelman et al., 2003; Arabzadeh et al., 2004; Reich et al., 2001). Other studies have shown that spike timing with respect to stimulus onset or network fluctuation plays an important and informative role in neuronal coding, which cannot be extracted from spike counts (Optican and Richmond, 1987; Panzeri et al., 2001). Further, several studies have quantified the impact of cross-neuronal correlations, and their contributions to neuronal population coding across sensory modalities (Panzeri et al., 1999; Schneidman et al., 2003). Last, information theoretic measures have been used in information transfer frameworks to quantify the interactions between neuronal populations (Honey and Sporns, 2008).

At least four previous studies also applied information theoretic concepts to the analysis of fMRI data. In (Fuhrmann et al., 2007), the authors used mutual information metrics to characterize the spatio-temporal profile of event-related BOLD responses during performance of a motor learning task. In (Fuhrmann et al., 2008) a similar approach was applied to BOLD responses acquired under audiovisual stimulation, indicating a temporal hierarchy of information content. Pessoa et al. (Pessoa and Padmala, 2007a) used mutual information to probe the neural correlates of near-threshold fear detection in a decoding framework. Finally, Rolls et al. (Rolls et al., 2010) used an information theoretic framework to predict the subject's affective state from multi-voxel fMRI activity in response to somato-

sensory stimulation. It should be noted, that while this rather small collection of studies has employed information theoretic concepts directly for the description of fMRI signals, a larger number of studies have employed information theoretic concepts in the general context of cognitive neuroscience. For example (Koechlin and Summerfield, 2007) used information theoretic concepts to describe the architecture of executive control in the lateral pre-frontal cortex, which was experimentally supported by standard GLM-fMRI data analyses (Koechlin et al., 2003). Other examples of studies that employed information theoretic concepts in the context of cognitive neuroimaging, but not for the actual analysis of fMRI data, are (Friston et al., 1992, Nevado et al., 2004) and (Chawla et al., 2000)

The application of an information theoretic framework to the study of combined invasive electrophysiological (LFPs) and metabolic measures (BOLD) of neural activity has been proposed in (Panzeri et al., 2008). This reference provides the primary source of inspiration for the proposal to study the application of the same framework to simultaneously acquired EEG and fMRI data in this thesis. The aim of the following section is to discuss some of the intuitive motivations that underpin this attempt.

As will become apparent upon the formal introduction of the information theoretic framework, the quantities involved relate to a number of features of standard experiments in cognitive neuroimaging experiments. Specifically, the framework allows the quantification of information conveyed by each imaging modality about the stimulus in isolation and in their combination. Further, the quantities involved generalize readily to the quantification of inter-modality co-variations, a primary concern in combined EEG-fMRI experiments. In this respect, it has been noted that EEG-fMRI co-variations can occur in an event or stimulus-related and –unrelated fashion (Herrmann and Debener, 2008). A subpart of the information framework provides a formalization and quantification of this intuition. Moreover, with respect to cognitive neuroscience paradigms in which the behavioural outcome of an

internally or externally triggered cognitive process is observed, the framework also enables the probabilistic assessment of response signal-behaviour co-variations.

The information theoretic framework is based on the explicit estimation of the full stimulus-response probability distribution, and hence has potential to contribute to the study of Bayesian approaches to EEG-fMRI data analysis (Daunizeau et al., 2009; Henson et al., 2010). As the approach is data-driven, and not constrained to linear co-variations, the quantification of stimulus-signal and signal-signal co-variations allows the examination of the assumptions of model-based approaches to EEG-fMRI integration.

Finally, the central concept of mutual information relates to the information discriminability achieved from the observation of single neural responses. This ties the framework closely to the current shift from the assessment of response signal trial-averages to the assessment of response signal single-trial variability and its role in sensory coding and inter-modality co-variation (Bagshaw and Warbrick, 2007; Debener et al., 2007b; Makeig et al., 2004). A critical analysis of the premises of information theory to EEG-fMRI integration is part of the general discussion (Chapter 7).

1.7 Summary

In summary, it has been established for almost a century now that electrophysiological activity of the brain can be monitored at high temporal resolution using scalp EEG recordings. The last twenty years have seen the emergence of fMRI, providing high spatial resolution images of the haemodynamic consequences of neural activity. While the relationship of both imaging modalities to intracranially recorded electrophysiological activity remains unclear, it has nevertheless been established that the neurobiological underpinnings of human brain

function can be studied using these non-invasive techniques. The simultaneous acquisition of EEG and fMRI data was initially motivated by the aim of identifying the location of sources of epileptic EEG activity, but has since found additional applications in sensory and cognitive neuroscience. The initial challenge of correcting the artefacts in each modality's signal that are induced by their joint acquisition has mainly been overcome. Nevertheless, obtaining signals that have the quality of the single modality's equivalents remains an active research area.

One current problem in EEG-fMRI is the question of how both modalities can be integrated to obtain new insights into human brain function based on their joint acquisition. The present thesis contributes to this line of research by introducing an information theoretic approach to EEG-fMRI data analysis.

2 Theoretical Foundations

If the human brain is presented with an external or internal stimulus, the neural or metabolic response that can be observed in the EEG or fMRI signal will exhibit variability both for repeated presentation of the same stimulus, as well as for presentations of different stimuli. Modern probability theory is a theoretical framework that can be used as a model or metaphor for experimental situations in which the observed measures display some degree of randomness. Information theory is a branch of probability theory and was pioneered by Claude E. Shannon in the study of communication (Shannon, 1948). Specifically, information theoretic measures are functionals (i.e. functions of functions) of probability mass or density functions. An axiomatic introduction to probability theory based on the notion of probability spaces is provided in Appendix B. This introduction includes the concepts of discrete and continuous random variables and probability mass and density functions, which form the prerequisite for the development of information theory below.

2.1 Introduction to Information Theory

The aim of the following section is to introduce the two central concepts of information theory in this thesis, entropy and mutual information, in a general context. The introduction will capitalize on the notion of discrete random variables, as this treatment is traditionally of more relevance to data analytical problems than a mathematically rigorous treatment based on continuous random variables. Additionally, it is also of greater intuitive appeal. Nevertheless, this section will conclude by introducing the continuous analogues of the most important quantities, providing the basis for the later discussion of the Gaussian method for entropy estimation. The discussion largely follows (Cover and Thomas, 2006).

2.1.1 Entropy

In the following, let X be a discrete random variable with codomain $\mathcal{X}$ and probability mass function

$$p(x) := p_X(x) \coloneqq P(\{X = x\}) \tag{2.1}$$

It should be noted, that using this notation $p(x)$ and $p(y)$ refer to different random variables, in fact being different probability mass functions. The entropy $H(X)$ of a discrete random variable X is defined as the expectation of the random variable $\log_2 \frac{1}{p(x)}$ where X is distributed according to the probability mass function $p(x)$:

$$H(X) = E_P(\log_2 \frac{1}{p(x)}) \tag{2.2}$$

If the logarithm is formed with respect to a base of 2, the entropy is expressed in bits (which is the convention used in this thesis). If the natural logarithm is used, the entropy is expressed in nats. In this thesis, the notation $\log_2$ is used for the logarithm with respect to base 2 and ln is used for the natural logarithm.

Expression (2.2) is equivalent to the two more common expressions of the entropy

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)} = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{2.3}$$

The second equality in (2.3) holds as $\log_2 \frac{1}{x} = -\log_2 x$ for all $x \in \mathbb{R}$. By convention, the expression $0 \log_2 0$ equals zero, which is based on a continuity argument.

Intuitively, entropy is a measure of the variability of a random variable. Its value is independent of the actual values taken by X, and only depends on the distribution of X. This is one difference in comparison to other measures of variability, as for example the variance a of

random variable X, is given by the expectation of the squared deviation of the values of random variable from its expectation

$$V(X) := V_P(X) := E_P((X - E(X))^2) \tag{2.4}$$

The principle motivation for formulating a variance measure based on the expectation of $\log\frac{1}{p(x)}$ lies in the properties that result for the formalization of information below (see section 2.1.5).

The entropy of a binomial variable of the form

$$X = \begin{cases} 0\ with\ probability\ p(0) = \frac{1}{2} \\ 1\ with\ probability\ p(1) = \frac{1}{2} \end{cases} \tag{2.5}$$

is given by

$$H(X) = \frac{1}{2} \cdot \log_2 2 + \frac{1}{2} \cdot \log_2 2 = \frac{1}{2} + \frac{1}{2} = 1\ bit \tag{2.6}$$

The stimulus variables in the experimental chapters of this thesis usually have this distribution.

2.1.2 Joint Entropy and conditional entropy

The joint entropy $H(X, Y)$ of a pair of discrete random variables X and Y with codomains $\mathcal{X}$ and $\mathcal{Y}$, respectively, and a joint distribution denoted as $p(x, y)$ is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 p(x,y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \frac{1}{\log_2 p(x,y)} \tag{2.7}$$

Further, the conditional entropy of a random variable given another random variable is defined as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable. Formally

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \tag{2.8}$$

$$= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(x|y) \log_2 p(x|y)$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 p(x|y)$$

$$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 \frac{1}{p(x|y)}$$

From these definitions it follows immediately that the entropy of a pair of random variables is the sum of the entropy of one random variable and the conditional entropy of the other random variable, conditioned on the first one:

$$H(X,Y) = H(X) + H(Y|X) \tag{2.9}$$

Statement (2.9) is of utmost importance for the application of information theoretic concepts in the analysis of neurophysiological experiments. It is proven as follows:

$$H(X,Y) = -\sum_{x\in\mathcal{X}} \sum_{y\in Y} p(x,y) \log_2 p(x,y) \tag{2.10}$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in Y} p(x,y) \log_2 p(x)p(y|x)$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in Y} p(x,y) \log_2 p(x) - \sum_{x\in\mathcal{X}} \sum_{y\in Y} p(x,y) \log_2 p(y|x)$$

$$= -\sum_{x\in\mathcal{X}} p(x) \log_2 p(x) - \sum_{x\in\mathcal{X}} \sum_{y\in Y} p(x,y) \log_2 p(y|x)$$

$$= H(X) + H(Y|X)$$

If two random variables X and Y are independent, then $H(Y|X) = H(Y)$. According to (2.10) the entropy of the joint distribution is hence the sum of the entropies of the marginal distributions. This is one of the intuitive appeals that motivate the definition of entropy as a measure of variability in the form of (2.2).

2.1.3 Relative entropy and mutual information

The relative entropy, also referred to as the Kullback-Leibler distance, of two probability distributions is a measure of the distance between them. Although not fulfilling all requirements of a mathematical metric, is widely used in applied mathematics as such, for example in variational Bayesian frameworks (Bishop, 2006). For two probability mass functions $p(x)$ and $q(x)$ the relative entropy is defined as

$$D(p||q) = \sum_{x\in\mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)} \tag{2.11}$$

For the definition above, the conventions $0 \log_2 \frac{0}{0} = 0$, $0 \log_2 \frac{0}{q} = 0$ and $p \log_2 \frac{p}{0} = \infty$ apply. In terms of the Kullback-Leibler distance, the mutual information between two random variables is the relative entropy between the joint distribution and the product distribution of the two random variables. Hence, in the case of independent random variables, i.e. the joint distribution being equal to the product distribution, the mutual information is equal to zero. In this view, the mutual information between two variables can be interpreted as a measure of their non-independence. Formally, the mutual information between two random variables X and Y with codomains $\mathcal{X}$ and $\mathcal{Y}$, respectively, is given by

$$I(X;Y) = D\big(p(x,y)||p(x)p(y)\big) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{2.12}$$

2.1.4 Entropy and mutual information

The mutual information between two random variables can be expressed in terms of entropies as well, which will be of major importance in the following section. From definitions (2.3), (2.10) and (2.12) it follows that

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{2.13}$$

$$= \sum_{x\in X}\sum_{y\in y} p(x,y)\log_2\frac{p(x|y)}{p(x)}$$

$$= -\sum_{x\in X}\sum_{y\in y} p(x,y)\log_2 p(x) + \sum_{x\in X}\sum_{y\in y} p(x,y)\log_2 p(x|y)$$

$$= -\sum_{x\in X} p(x)\log_2 p(x) - \left(-\sum_{x\in X}\sum_{y\in y} p(x,y)\log_2 p(x|y)\right)$$

$$= H(X) - H(X|Y)$$

Finally, the entropy of a random variable X conditioned on itself is zero, as

$$H(X|X) = \sum_{x\in X}\sum_{x\in X} p(x)\log_2\frac{1}{p(x|x)} \tag{2.14}$$

$$= \sum_{x\in X}\sum_{x\in X} p(x)\log_2 1$$

$$= \sum_{x\in X}\sum_{x\in X} p(x)\cdot 0$$

$$= 0$$

From this it follows immediately, that the information of a random variable about itself is given by its entropy, as

$$I(X;X) = H(X) - H(X|X) \tag{2.15}$$

This property of mutual information is important when considering the maximum information any neural response can convey about a stimulus or another neural response.

2.1.5 Properties of mutual information

The definitions of entropy and mutual information as provided by (2.2) and (2.12) result in a number of properties of mutual information that appeal to intuitive notions of information. Amongst others, these are that a measure of information should be non-negative and symmetric, be additive for independent random variables and, if passed through an

information processing system, can only decrease but not increase. These properties are briefly reviewed below

Non-negativity of information

For any pair of random variables X and Y, the mutual information $I(X;Y)$ is strictly non-negative, i.e.

$$I(X;Y) \geq 0 \tag{2.16}$$

and equality holds only, if X and Y are independent. This property is a direct consequence of the non-negativity of the Kullback-Leibler distance, i.e.

$$D\big(p(x,y)||p(x)p(y)\big) \geq 0 \tag{2.17}$$

which in turn is a consequence of the non-negativity of probabilities.

Symmetry of information

For any pair of random variables X and Y the mutual information that X conveys about Y is the same as the mutual information that Y conveys about X. While the Kullback-Leibler distance itself is not symmetric, due to the definition of the mutual information the following equalities hold

$$\begin{aligned} I(X;Y) &= \textstyle\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \\ &= \textstyle\sum_{x\in X}\sum_{y\in Y} p(y,x)\log_2\frac{p(y,x)}{p(y)p(x)} \\ &= \textstyle\sum_{y\in Y}\sum_{x\in X} p(y,x)\log_2\frac{p(y,x)}{p(y)p(x)} \\ &= I(Y;X) \end{aligned} \tag{2.18}$$

Additivity of information for independent variables

For a triple of random variables X, Y and Z it can be shown that if Y and Z are independent random variables, then the relation

$$I(X; Y, Z) = I(X; Y) + I(X; Z) \tag{2.19}$$

holds.

Data-processing inequality

It can be shown, that if three random variables X, Y and Z form a Markov chain, i.e. the joint distribution of the three random variables factors according to

$$p(x, y, z) = p(x)p(y|x)p(z|y) \tag{2.20}$$

then the data-processing inequality for mutual information holds, i.e.

$$I(X; Y) \geq I(X; Z) \tag{2.21}$$

Intuitively, in an information processing system, information can hence only decrease by passing through subsequent stages, and not increase.

2.1.6 Differential entropy and mutual information

The discussion of mutual information quantities in the following will largely centre on the case of discrete random variables. However, in a later section, the Gaussian method (Magri et al., 2009) to estimate information theoretic quantities from data will be discussed. This proposed method capitalizes on the analytical information theoretic treatment of a Gaussian probability density function, and hence assumes continuous random variables.

A random variable $X: \Omega \to \Omega'$, where Ω and Ω' are nonempty sets, is said to be continuous, if its cumulative distribution function $F_P(x) = P(\omega \in \Omega | X(\omega) < x), x \in \mathbb{R}$ is continuous. Let $f(x) = F'(x)$ be the probability density function of X. Finally, let $S = \{ x \in \mathbb{R} | f(x) > 0 \}$ be the support set of X. The differential entropy $h(X)$ of a continuous random variable X with probability density function $f(x)$ is defined as

$$h(X) = \int_S f(x) \frac{1}{\log_2 f(x)} dx \tag{2.22}$$

if the respective integral exists. Again, as for the discrete case, the differential entropy is a functional of the probability density function of the random variable. The differential entropy of a collection of random variables $X_1, \dots, X_n$ with joint probability density function $f(x_1, \dots, x_n)$ is defined as

$$h(X_1, \dots, X_n) = \int_{S_{x_1}} \dots \int_{S_{x_n}} f(x_1, \dots, x_n) \frac{1}{\log_2 f(x_1, \dots, x_n)} dx_1 \dots dx_n \tag{2.23}$$

where the integral is taken over the joint support set S_{x_i} of all variables $X_1, \dots, X_n$, if it exists. The continuous analogue to the conditional entropy defined in (2.8) is the conditional differential entropy of two random variables X and Y with a joint density function $f(x, y)$ given by

$$h(X|Y) = \int_{S_x} \int_{S_y} f(x, y) \frac{1}{\log_2 f(x|y)} dx dy \tag{2.24}$$

Finally, the analogue expressions for the relative entropy for two probability density functions and the mutual information for two continuous random variables are given by

$$D(f||g) = \int f \log_2 \frac{f}{g} \tag{2.25}$$

and

$$I(X;Y) = \int_{S_x} \int_{S_y} f(x,y) \frac{f(x,y)}{f(x)f(y)} dxdy \tag{2.26}$$

respectively.

The main relevance of these expressions for the purposes of this thesis is that they allow the entropy of a multivariate Gaussian distribution to be derived analytically as a function of the Gaussian distributions covariance matrix. For a collection of random variables $X_1, \dots, X_n$, a multivariate Gaussian distribution is given by the probability density function $f: \mathbb{R}^n \to \mathbb{R}$, with $\mu \in \mathbb{R}^n$, and $\Sigma \in \mathbb{R}^{n \times n}$ a real, symmetric, positive definite matrix

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \tag{2.27}$$

where $|\Sigma|$ denotes the determinant of Σ and Σ^{-1} the inverse of Σ. Evaluating statement (2.23) for this probability density function yields

$$h(X_1, \dots, X_n) = \frac{1}{2} \log_2 (2\pi e)^n |\Sigma| \tag{2.28}$$

and, hence, the differential entropy of a multivariate Gaussian distribution is a function of the logarithm of the determinant of its covariance matrix Σ.

2.1.7 Summary

In summary, information theory entails the definition of a measure of variability of a random variable, namely entropy, and a measure of the dependence between random variables, namely mutual information. From a mathematical viewpoint, information theoretic measures are functionals of probability mass, or in the case of continuous random variables, probability density functions. The experimental evaluation of information theoretic measures hence is predicated upon the estimation of probability distributions. The definitions of both quantities result in properties that satisfy intuitive notions of uncertainty and information in

real-world situations that exhibit random behaviour. One such situation is the experimental measurement of neurophysiological events using brain imaging data as discussed in detail in the next section.

2.2 Information decomposition in neurophysiological experiments

The application of information theoretic concepts to neurophysiological experiments entails regarding the experimentally observed stimulus and data values as realizations of random variables, i.e. assuming that the mathematical framework of probability theory is an adequate description of a given data acquisition situation. Once this step is made, the theorems that apply for the general concept of random variables in the information theoretic framework hold equivalently for the experimental variables of interest, and assume a specific, neurophysiologically motivated meaning. In the application of information theory to the case of combined EEG-fMRI experiments, the main interest lies on random variables representing the external or internal stimuli, EEG or fMRI signal features, as well as possible behavioural responses, such as reaction times.

Below, an information decomposition frameworks developed in the study of single-unit electrophysiology will be discussed, which is applied to experimental EEG-fMRI data in Chapters 4 and 5 and with some modifications in Chapter 6. Intuitively, this framework has certain similarities to more popular linear variance decomposition schemes (Hays, 1994).

Understanding neural stimulus coding and response signal relationships in a multimodal data acquisition context requires a quantitative description of the respective stimulus-response signal and response signal-response signal interdependencies. In a typical experiment, the observer is presented with a stimulus while an EEG signal and an fMRI signal

are concurrently recorded. For clarity, one differentiates between response signals, response signal modalities and response signal features: By 'signal modality' one understands the type of data acquisition modality (i.e. EEG, fMRI, behaviour), by 'signal' the entirety of the acquired signal modality in its raw form, while by 'signal feature' a given quantified aspect of a signal after data pre-processing is understood. Many different signal features are conceivable, for example the amplitude of evoked haemodynamic response functions in a given brain area for the fMRI signal, or the power of a given frequency band of the EEG extracted from specific electrodes. In this very general context of cognitive neuroimaging experiments, a set of questions follows naturally (Panzeri et al., 2008): Which response signal features convey information about the stimulus or cognitive task? Do different signal features within and across response signal modalities convey redundant or complementary information about the stimulus? Which signal features are most closely related within and across response modalities? Is this relationship stimulus dependent or not?

The information theoretic framework adopted here for multimodal brain imaging experiments has its roots in the study of communication (Shannon, 1948) and was previously applied to the problem of neural coding in an invasive electrophysiological studies (for example in (Belitski et al., 2008; Panzeri et al., 2007; Schneidman et al., 2003)). Presented below are the concepts most relevant to the current application.

2.2.1 Stimulus related information: $\boldsymbol{I(S;R_1), I(S;R_2), I(S;R_1,R_2)}$

A central notion of the framework is the concept of mutual information, i.e. the information that a signal feature conveys about the stimulus or another response signal feature. Consider a neuroimaging experiment with one response signal modality (e.g. EEG). For each stimulus of a given discrete stimulus set $S = \{s_1, \dots, s_{|S|}\}$ (here $|X|$ denotes the

cardinality of a set, i.e. the number $n \in \mathbb{N}$ of elements contained in the finite and countable set X), a response signal feature is determined for each repeat of the stimulus $s_i \in S$. From the observed responses $r_j \in R$, the joint probability distribution $p(s,r)$ can be constructed from the respective entries $p(s = s_i, r = r_j)$, where the response set is denoted as $R = \{r_1, \dots, s_{|r|}\}$. From this joint distribution the marginal distributions $p(s)$ and $p(r)$ can be obtained by summation, and the conditional probability distribution $p(s|r)$ by division. It should be noted that $p(s)$, the stimulus probability distribution, is usually under the control of the experimenter.

As discussed above, Shannon (Shannon, 1948) proposed entropy as a measure of the variability (or uncertainty) of a given random variable, as it satisfies a set of intuitive assumptions (e.g. it is maximised in the case of a uniform distribution indicating highest uncertainty). In the context of neurophysiological experiments, two entropy measures that can be computed from the distributions introduced above are of relevance: the response entropy and the noise entropy (Magri et al., 2009). For the equations given below it is assumed that both the stimulus set S and the response set R are appropriately discretised, i.e. one is dealing with discrete random variables and probability mass functions.

The response entropy, $H(R)$ captures the overall variability of a given response signal feature across all stimuli, and is obtained from the marginal distribution $p(r)$ according to

$$H(R) = \sum_{r \in R} p(r) \log_2 \frac{1}{p(r)} \tag{2.29}$$

On the other hand, the noise entropy, $H(R|S)$, captures the variability of the signal feature that remains after the signal feature variability due to the stimulus is accounted for. The noise entropy is therefore the signal feature variability that exists even in the absence of differential stimulation, and is given by

$$H(R|S) = \sum_{r\in R}\sum_{s\in S} p(r,s)\log_2\frac{1}{p(r|s)} \quad (2.30)$$

The information that the response signal conveys about the stimulus distribution is denoted as the mutual information, $I(S;R)$, between the stimulus and signal feature distribution, and is given as

$$I(S;R) = H(R) - H(R|S) = \sum_{s\in S}\sum_{r\in R} p(s,r)\log_2\left(\frac{p(s,r)}{p(s)p(r)}\right) \quad (2.31)$$

Hence, the mutual information between the response signal R and the stimulus S is the variablity of the response signal minus the variability that remains once the variability not due to the stimulus is accounted for. The variability of the response signal R that remains after this subtraction is hence the variability accounted for by variability in the stimulus S.

As discussed in section 2.1.5, the mutual information as defined above has many properties that correspond well with intuitive connotations of 'information', for example that information is additive with respect to statistically independent variables. Its unit is the 'bit', if a logarithm of base 2 is used. One bit of information means that, on average, observation of the signal feature on a given trial reduces the observer's stimulus uncertainty by a factor of 2 (Cover and Thomas, 2006).

The representation of the mutual information between response signal feature and stimulus as given by equation (2.25) naturally extends to response signal feature distributions of higher dimensionality, i.e. the case of more than one type of response signal feature. In the current application to EEG-fMRI, the main interest is on the case of two response signal features, one from each modality. Substituting the joint distribution $p(s,r)$ by $p(s,r_1,r_2)$ and $p(r)$ by $p(r_1,r_2)$ in equation (2.25), where r_1 and r_2 indicate response signal features from each of the two modalities, yields

$$I(S;R_1,R_2) = \sum_{s\in S}\sum_{r_1\in R_1}\sum_{r_2\in R_2} p(s,r_1,r_2)\log_2\left(\frac{p(s,r_1,r_2)}{p(s)p(r_1,r_2)}\right) \quad (2.32)$$

Equation (2.32) represents the information about the stimulus that is conveyed by the joint distribution of response signal features R_1 (e.g. an EEG signal feature) and R_2 (e.g. an fMRI signal feature).

In the current context of EEG-fMRI experiments, the notion of mutual information allows one to gain insight into the stimulus-response signal relationship for each signal modality individually, i.e. one can compute the $I(S;R_1)$ for an EEG response signal feature and $I(S;R_2)$ for an fMRI response signal feature. However, for the case of two response signal features as in the case of a multimodal, simultaneous neuroimaging experiment one can define additional quantities of interest, which are adopted from (Schneidman et al., 2003) and surround the quantity $I(S;R_1,R_2)$. The basic idea is to quantify the dependencies (or independencies) of the response signals on the stimulus and on each other. Three concepts of dependence are central: information dependence, activity dependence and conditional dependence, each of which will be discussed in turn.

2.2.2 Information dependence and $Synergy$

The concept of 'information dependence' relates to the question of whether particular response features convey redundant or complementary information about the stimulus. Consider for example the case in which the two response signal features are sensitive to orthogonal stimulus features (a case of particular importance in invasive neurophysiological single cell studies). In this case, the information that the combined response signals convey about the stimulus should be the sum of their individual stimulus mutual information

$$I(S;R_1,R_2) = I(S;R_1) + I(S;R_2) \quad (2.33)$$

In addition, two scenarios are possible: the two response signals can be synergistic (i.e. their joint distribution conveys more information about the stimulus than their individual distributions), or redundant (i.e. the stimulus-related information in their joint distribution equals the stimulus-related information in each marginal distribution). These scenarios can be readily captured in the equation for the normalised signal feature synergy:

$$Synergy = \frac{I(S;R_1,R_2)-\left(I(S;R_1)+I(S;R_2)\right)}{I(S;R_1,R_2)} \quad (2.34)$$

$Synergy$ is a quantity that ranges between -1 for fully redundant response signal features, and +1 for fully synergistic response signal features.

2.2.3 Response signal related information: activity and conditional dependence

Activity dependence relates to the question of how much the activity of one response signal feature R_1 depends on the activity of another response signal feature R_2. In the context of EEG-fMRI experiments, this concept is mainly related to the question of which EEG features are most tightly coupled to BOLD signal features (or vice versa). It should be noted that response signal feature dependencies can occur both in the absence as well as the presence of external stimulation (Herrmann and Debener, 2008). Formally, activity dependence is assessed by studying the experimentally obtained joint distribution of stimulus and response signals, $p(s, r_1, r_2)$, and marginalizing appropriately. An obvious measure for the degree of activity dependence between a pair of response signals is the mutual information that one signal feature conveys about the other, given by

$$I(R_1; R_2) = H(R_1) - H(R_1|R_2) = \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} p(r_1, r_2) \log_2 \left(\frac{p(r_1,r_2)}{p(r_1)p(r_2)}\right) \quad (2.35)$$

As an example, R_1 could represent the power at a given frequency band of the EEG signal at a given electrode, while R_2 could relate to the integral of the haemodynamic response function in a corresponding brain area. If the two response signals are independent, by definition the joint response distribution factors to $p(r_1, r_2) = p(r_1)p(\,r_2)$ and the mutual information $I(R_1; R_2)$ is zero.

Finally, conditional dependence refers to the notion of noise correlation, i.e. the question of whether the two response signals are interdependent in the absence of external stimulation. In the multimodal imaging context one would primarily assume that the strength of the neurovascular coupling is independent of the presence or absence of external stimulation, although this might not necessarily be the case. Intuitively, this concept can be understood by considering the conditional distribution of the two response signal features at a given stimulus level s_i, $p(r_1, r_2|s = s_i)$. If the two response signals are dependent on each other due to some factor other than the stimulus (for example, the 'hard-wired' neurovascular coupling), then $p(r_1, r_2|s = s_i)$ will reveal some dependency between the response signal features. If the two response signals are independent or only dependent on each other under stimulation, the conditional distribution will factor to

$$p(r_1, r_2|s = s_i) = p(r_1|s = s_i)p(r_2|s = s_i) \tag{2.36}$$

From these considerations, an obvious measure of response signal conditional dependence is the mutual information between the response signals given the stimulus:

$$I(R_1; R_2|S) = \sum_{r_1 \in R_1} \sum_{r_2 \in R_2} p(r_1, r_2|s) \log_2 \left(\frac{p(r_1, r_2|s)}{p(r_1|s)p(\,r_2|s)} \right) \tag{2.37}$$

and more compactly, its average across stimulus levels, i.e.

$$\langle I(R_1; R_2|S) \rangle_S = \frac{1}{|S|} \sum_{s_i \in S} I(R_1; R_2|s = s_i) \tag{2.38}$$

In sum, the information theoretic framework introduced by (Panzeri et al., 2008; Schneidman et al., 2003) and reviewed here rests on the experimental determination of the joint stimulus-response distribution $p(s, r_1, r_2)$ and the subsequent computation of a set of mutual information quantities which relate response signal features with the stimulus and one another. A summary of the information theoretic quantities is given in the table below.

Information theoretic quantity	**Abbreviation**
Mutual information between stimulus and response signal feature of modality 1 (EEG)	$I(S; R_1)$
Mutual information between stimulus and response signal feature of modality 2 (fMRI)	$I(S; R_2)$
Mutual information between stimulus and response signal features of modalities 1 and 2	$I(S; R_1, R_2)$
Normalized difference between the mutual information between stimulus and response variables of modalities 1 and 2 and the sum of the individual mutual information for response signal feature 1 and response signal feature 2 (Information Dependence)	$Synergy$
Overall mutual information between response signal features of modality 1 (EEG) and modality 2 (fMRI) (Activity Dependence)	$I(R_1; R_2)$
Average stimulus conditional mutual information between response signal features of modality 1 (EEG) and modality 2 (fMRI) (Conditional Dependence)	$\langle I(R_1; R_2 \mid S) \rangle_S$

Table 2.1 Information theoretic quantities. The table presents an overview of the information theoretic quantities related to the information decomposition framework discussed in section 2.2.

2.3 Practical considerations and bias correction

2.3.1 Histogram Estimation

The fundamental problem of quantitative empirical science is that the probability density functions that are assumed to generate experimentally observed data are unknown. Parametric methods to probability density estimation capitalize on the use of parameterized functional forms, such as the Gaussian distribution, and estimate the respective density function parameters from the observed data. Due to its development in the context of neurophysiological experiments, specifically its application to discrete spike-count data, the probability mass functions used to compute the various information quantities introduced in

the previous section have been traditionally estimated non-parametrically (Panzeri and Treves, 1996). The simplest non-parametric approach to probability density estimation is a histogram approach (Bishop, 2006). Intuitively, in order to estimate the probability of a given response feature combination, the response feature space (e.g. $\mathbb{R}^2$ in the case of a bivariate response) is partitioned into discrete, non-overlapping intervals, so-called bins. For example, in the case of $\mathbb{R}^2$, a bin r_i can be written as

$$r_i = \left\{ \binom{x_1}{x_2} \in \mathbb{R}^2 | a_i < x_1 \leq b_i, c_i < x_2 \leq d_i \right\} \tag{2.39}$$

where $a_i, b_i, c_i, d_i \in \mathbb{R}$ denote the bin's boundaries. A probability estimate p_i for each such bin is then obtained by dividing the number of observations n_i that fall into this bin by the total number of observations N and the size of the bin Δ_i, i.e.

$$p_N(r_i) = \frac{n_i}{N\Delta_i} \tag{2.40}$$

Importantly, this standardization of the observed frequency count guarantees, that

$$\sum_{i=1}^{n} p_N(r_i)\Delta_i = 1 \tag{2.41}$$

where n denotes the total number of bins and the subscript N the dependency of the probability mass function on the total number of trials. Hence, p_N is indeed a probability mass function as introduced in (B.25). The estimation quality provided by the histogram estimator defined in this manner is a function of the bin size Δ_i. In general, too small bin sizes Δ_i result in a too variable estimation, while too large bin size Δ_i result in too smooth estimates, which might suppress important characteristics of the true underlying probability distribution. In Chapter 3 a procedure is described that motivates the choice of the bin width Δ_i in the experimental sections of Chapters 4 and 5. Additionally free parameters are the upper and lower limits of the response bins in each dimension. To make efficient use of all sampled data

points, the upper and lower limits of the response grid were set to the maximum and minimum value of each response variable in the data analytical contexts of this thesis.

2.3.2 Panzeri-Treves (PT) bias correction

As discussed in the preceding section, it is important to note, that for the practical application of the information framework the joint stimulus-response signal distribution $p(s, r_1, r_2)$ is not known, but can only be estimated from a finite data set as the observed frequency distribution $p_N(s, r_1, r_2)$. Here N indicates the total number of trials/observations, given by $N = |S| \cdot N_S$, where N_S, indicates the number of repeats per stimulus (for the current purposes, it is assumed that all stimuli of the stimulus set are presented equally often). Accordingly, the subscript N is used for the information theoretic quantities estimated from limited simulated or experimental data, e.g. $I_N(S; R_1)$ refers to the estimated mutual information between stimulus and response signal feature R_1. Dealing with finite (and in the case of EEG-fMRI experiments, relatively small), sample sizes leads to systematic errors (biases) in the values of the computed quantities. The theoretical foundations, as well as methods for the correction of these systematic errors, have been studied extensively for the case of invasive neurophysiological data (Panzeri et al., 2007).

In the current work, for the purposes of assessing the relationship between two variables (i.e. all information theoretic quantities except $I(S; R_1, R_2)$ and $Synergy$), the PT-correction scheme, proposed and implemented by (Panzeri and Treves, 1996) is used. For the computation of bias-corrected $I(S; R_1, R_2)$, the 'shuffling correction' (Montemurro et al., 2007; Panzeri et al., 2007) is employed. The following sections provide a brief overview of these two correction schemes.

Denoting the real mutual information between two random variables by I, it has been shown (Panzeri and Treves, 1996) that the average additional error, or bias, can be expressed as a series expansion in inverse powers of sample size N:

$$\langle I_N \rangle - I = \sum_{m=1}^{\infty} C_m \tag{2.42}$$

Here $\langle I_N \rangle$ indicates the average plug-in mutual information over repeated sampling from the real underlying probability distribution, and C_m represents successive contributions to the asymptotic expansion of the bias. In the following, plug-in mutual information refers to mutual information estimated by applying the respective sum formulae to the estimated probability mass functions without bias correction. Only the first term C_1 of the expansion is, on first approximation, independent of the underlying probability distributions (Panzeri and Treves, 1996). The fact that C_1 is independent of the analytically assumed probability distribution is of major relevance, as it allows to estimate this term in real experiments, where the underlying probability distribution is unknown. C_1 can be estimated from the sampled data as

$$C_1 = \frac{1}{2Nln2}\left(\left(\sum_{s=1}^{S} \bar{R}_s - 1\right) - (\bar{R} - 1)\right) \tag{2.43}$$

Here $\bar{R}_s$ denotes the number of relevant response bins for the stimulus conditional distributions, $p_N(r|s)$, and $\bar{R}$ the number of relevant responses for the marginal response distribution $p_N(r)$. The relevant response bins are defined as the number of bins r_i with non-zero probability. In (Panzeri and Treves, 1996), the authors introduced a Bayesian approach to estimate these two numbers from the data for a single response variable. Following estimation of $\bar{R}_s$ and $\bar{R}$, equation (2.43) can be evaluated and the resulting error term subtracted from the plug-in mutual information estimate, yielding the bias corrected version. This procedure is

implemented in the *bayescount.m* function available from http://stefano.panzeri.googlepages.com/informationbiascorrections)

2.3.3 Shuffling correction

For the case of multidimensional response variables, i.e. the case of response signal features of two modalities, the most powerful bias correction scheme is a shuffling approach (Montemurro et al., 2007; Panzeri et al., 2007). The information contained in the joint response of multiple response variables $R_1, \dots, R_J$ can be computed indirectly as

$$I_{sh}(S;R) = H(R) - H_{ind}(R|S) + H_{sh}(R|S) - H(R|S) \tag{2.44}$$

In equation (2.44), R indicates a multidimensional response variable, which is a two-dimensional vector in the context of most of this thesis. Two new entropy measures, $H_{ind}(R|S)$ and $H_{sh}(R|S)$, are introduced in equation (2.44). In the context of this thesis these refer to the entropies of the observed frequency distributions, $p_{N_{ind}}(r_1, r_2|s)$ and $p_{N_{sh}}(r_1, r_2|s)$, respectively. $p_{N_{ind}}(r_1, r_2|s)$ is the conditional joint response observed frequency distribution that would have been observed, were the two response signal feature distributions independent random variables. In this case, the respective joint distribution would factor according to

$$p_{N_{ind}}(r_1, r_2|s) = p_N(r_1|s)p_N(r_2|s) \tag{2.45}$$

Hence, this non-empirical distribution can be calculated by pair-wise multiplication of the marginal distributions $p_N(r_1|s)$ and $p_N(r_2|s)$. $p_{N_{sh}}(r_1, r_2|s)$, on the other hand, is the conditional joint response observed frequency distribution created by removing response variable correlations in an alternative way, namely shuffling the sampled response variables for a given stimulus individually for each response element, thereby removing trial-by-trial dependencies. The shuffling correction method removes biases in the estimated information

values in the sense that $I_{sh}(S;R)$ has the same value as $I(S;R)$ in the limit of an infinite number of trials, but has a smaller bias for finite number of trials due to the bias cancellation of $H_{ind}(R|S)$ and $H_{sh}(R|S)$ (Panzeri et al., 2007).

Finally, it should be noted that the problem of systematic errors (biases) in summary measures (statistics) of population parameters computed from finite samples is not specific to the framework applied here, but is a rather fundamental problem of quantitative empirical science, and encountered in many contexts, e.g. the computation of standard deviations from mean values (Hays, 1994; Panzeri et al., 2007). One quality criterion for statistical estimator is hence, whether they are biasfree, as for example the well-known BLUE (best linear unbiased estimator) property of the least-square estimator for the parameter vector in the context of the general linear model (Scheffe, 1959; Seber, 1977).

2.3.4 The Gaussian Method

In (Magri et al., 2009), the authors propose to compute information and entropies based on fitting response probabilities to Gaussian functions, i.e. a parametric method for information estimation. Specifically, they state that under the Gaussian hypothesis, the noise and response entropy and the information are given by functions of the observed response variable's variance according to

$$\widehat{H}^{(g)}(R) = \frac{1}{2}\log_2\left((2\pi e)|\sigma^2(R)\,|\right) \tag{2.46}$$

$$\widehat{H}^{(g)}(R|S) = \frac{1}{2}\sum_s p(\mathrm{s})\log_2\left((2\pi e)|\sigma_s^2(R)\,|\right) \tag{2.47}$$

resulting in

$$I^{(g)}(S;R) = \widehat{H}^{(g)}(R) - \widehat{H}^{(g)}(R|S) \tag{2.48}$$

for the stimulus related information in the response variable R. Here, $|\sigma^2(R)|$ and $|\sigma_s^2(R)|$ represent the determinants of the covariance matrices across trials and stimuli and across trials to stimulus s, respectively. Bias-free estimates of these covariance matrices can be obtained using usual variance estimators which do not require the underlying probability distribution to be parameterized and can be computed directly from the data.

The approach of assuming an analytically treatable and potentially experimentally relevant distribution for the response variable has the advantage that the limited sampling bias can be approximated analytically. Specifically, (Goodman, 1963; Misra et al., 2005; Oyman et al., 2003) show that the limited sampling bias for the Gaussian method can be expressed as a function of the number of trials across conditions N_{trials}^{total} and the number of trials for stimulus $s \in S$, N_{trials}^{s} according to

$$Bias_{\widehat{H}^{(g)}(R)} = g_{Bias}\left(N_{trials}^{total}\right) \tag{2.49}$$

and

$$Bias_{\widehat{H}^{(g)}(R|S)} = \sum_{s\in S} p(x) g_{Bias}(N_{trials}^{s}) \tag{2.50}$$

where $g_{Bias}: \mathbb{N} \to \mathbb{R}$ is defined by (Theorem 2.1 in Misra et al., 2005)

$$g_{Bias}(n) = \frac{1}{2\log_2 2}\left(L \ln\left(\frac{2}{n-1}\right) + \sum_{j=1}^{L} \psi\left(\frac{n-j}{2}\right)\right) \tag{2.51}$$

Here, $L \in \mathbb{N}$ denotes the dimensionality of the response array (and hence the assumed Gaussian distribution) and ψ is the digamma function, i.e. $\psi: \mathbb{R} \to \mathbb{R}$ with

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} \tag{2.52}$$

for the gamma function $\Gamma: \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \tag{2.53}$$

Unfortunately, as will be shown based on numerical simulations in Chapter 3, the Gaussian approximation of the response distribution given by (2.46) is not appropriate in most experimental situations of interest, and hence, as it stands, the Gaussian method is not applicable to combined EEG-fMRI data.

3 Numerical Simulations

In the data-driven analytical framework of this thesis, the numerical simulations discussed below serve three purposes. First, the validation of the implementation of the analytical methods, i.e. the programming code. Numerical results of the analysis of real experimental data can be assessed regarding their general validity, for example the order of magnitudes of a given response. However, only if the generation of the analysed data is known explicitly (which for real experimental data is never the case), can the reliability and validity of the data-analytical methods be determined according to some ground truth. Second, as data simulations allow some feature of the data generation to be varied in a controlled manner while observing the response of the data-analytical variables, data simulations are of great educational value. Only by observing the idealized behaviour of the quantities of interest, can the full capability of an analytical framework be appreciated. Last, specifically in the application of information theoretic analyses, the data analytical methods themselves can have parameters which need to be set for a given analysis. Data simulations allow these analytical parameters to be chosen in an informed way, conditioned on the validity of the modelled data with respect to the real experimental data.

In the following, the methods used to generate synthetic data in this thesis will be discussed briefly, namely linear Gaussian models, a specific type of graphical model. Linear Gaussian models can be regarded as forming the basis of much of classical statistics, and as such, form an important conceptual assumption of standard brain imaging data analysis embodied for example in the general linear model. Subsequently, the results of a number of numerical simulations analyzed using the information theoretic framework will be reported. The programming code implementing the simulations is available from http://www.buic.bham.ac.uk/downloads/EEG_FMRI_ITQ/EEG_FMRI_ITQ_Analysis.zip.

3.1 Graphical Models

Probabilistic graphical models represent an important framework in contemporary data analysis and machine learning (Bishop, 2006; Jordan, 1999). In brief, graphical models are graph-theoretic, diagrammatic representations of probability distributions, which enable the intuitive visualization of the structure of a given probabilistic model. Further, in their advanced use, graphical models allow complex manipulations of the underlying mathematical expressions to be expressed in terms of graphical manipulations. Graphical models comprise nodes, or vertices, which are connected by links, or edges. Nodes represent random variables, while links represent probabilistic relationships between the random variables. The specific arrangement of the (possibly directed) links between the nodes of a model represents the factorization of the joint probability distribution of all nodes. The two main classes of probabilistic graphical models are directed graphical models, also known as Bayesian networks, and undirected graphical models, also known as Markov random fields. For the simulations discussed below, a special class of directed graphical models, namely linear Gaussian models, were employed. Figure 3.1 depicts an example of a generic probabilistic graphical model and its corresponding joint probability distribution factorization

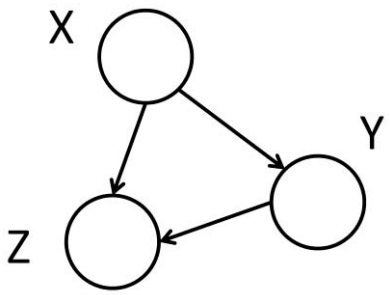


Figure 3.1 Example graphical model (reproduced from (Bishop, 2006)). The directed graphical model depicted here represents the joint probability distribution over three random variables, $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$, denoted as $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. The direction of the links between the nodes expresses a specific factorization of $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ into the product of conditional distributions. For each conditional distribution, the directed links express the conditioning of a 'child' node on the respective 'parent' node. In the example of the Figure, $\boldsymbol{p}(\boldsymbol{x})$ is not conditioned on any of the other two variables, $\boldsymbol{p}(\boldsymbol{y})$ is conditioned on $\boldsymbol{p}(\boldsymbol{x})$ and $\boldsymbol{p}(\boldsymbol{z})$ is conditioned on both $\boldsymbol{p}(\boldsymbol{x})$ and $\boldsymbol{p}(\boldsymbol{y})$. The corresponding factorization of the joint distribution of $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$ represented by the graphical model is hence $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{p}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x})$.

3.2 Linear Gaussian Models

Linear Gaussian models represent the expression of multivariate Gaussian distributions as directed graphical models. Linear Gaussian models form the basis of a large number of data-analytical techniques such as the general linear model, principal component analysis, independent component analysis, factor analysis, and probabilistic linear dynamical systems (Bishop, 2006; Roweis and Ghahramani, 1999). The prominent use of Gaussian models in neuroimaging and other experimental fields is due to their analytical tractability and the continuous nature of the data they are applied to. The strongest motivation for the use of Gaussian models in experimental science stems from the theoretical argument of the central limit theorem, i.e. that the sum of an infinite number of random processes of arbitrary distribution approximate a Gaussian form. Nevertheless, if good reasons exist, why a certain experimental process (e.g. neuroimaging data or behavioural signal) is not normally distributed, the validity of Gaussian models can be questioned. In this section, linear Gaussian models will briefly be reviewed in their general form.

Linear Gaussian models comprise directed graphs with $i = 1, \dots, N$ continuous random variables X_i each having a Gaussian distribution which are represented as nodes. The expectation of each node's distribution is a linear combination of the states of that node's parent nodes. Formally,

$$p(X_i|Pa_i) = N\left(X_i|\sum_{j\in Pa_i} a_{ij}x_j + b_i, \sigma_i^2\right) \tag{3.1}$$

Here Pa_i represents the set of parent nodes of the node representing X_i, and $N(X_i|\mu_i, \sigma_i^2)$ represents a univariate Gaussian distribution with expectation μ_i and variance σ_i^2. a_{ij} and b_i are parameters that govern μ_i, where a_{ij} is a 'coupling strength' parameter between the random variables X_j and X_i.

Based on (3.1), the expectation $E(X_i)$ of each random variable X_i and the covariance $Cov(X_i, X_j)$of two random variables X_i and X_j can be shown to correspond to

$$E(X_i) = \sum_{j \in Pa_i} a_{ij} E(X_j) + b_i \tag{3.2}$$

$$Cov(X_i, X_j) = \sum_{k \in Pa_j} a_{jk} Cov(X_i, X_k) + \delta_{ij} \sigma_j^2 \tag{3.3}$$

where δ_{ij} is the Kronecker delta operator. Starting from the lowest numbered node, expressions (3.2) and (3.3) allow the expectation and covariance of each node and each pair of nodes ($E(X_1), \dots, E(X_N)$ for all $i \in N$ and $Cov(X_i, X_j)$ for all $i, j \in N \times N$) to be determined recursively. These can then in turn be used to populate the expectation vector $\mu \in \mathbb{R}^N$ and covariance matrix $\Sigma \in \mathbb{R}^N$ of a N-dimensional Gaussian distribution, i.e. the probability density function of a random vector $X \in \mathbb{R}^N$ given by

$$p(X) := N(x|\mu, \Sigma) := \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \tag{3.4}$$

where $|\Sigma|$ denotes the determinant and Σ^{-1} the inverse of Σ. Hence, for practical sampling purposes, the model-specific graphical structure can be converted into its corresponding representation as a N-dimensional Gaussian distribution with parameters $\mu \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^N$ and samples obtained from it based on conventional sampling methods.

3.3 Linear Gaussian response signal models

3.3.1 Response signal model outlines

To validate the implementation of the information theoretic framework, constrain the analysis of the experimental data with respect to free data analysis parameters, as well as to study the behavior of the various information theoretic quantities under controlled response signal conditions, a number of analyses on simulated data sets were performed. These data

sets were derived from four simple generative linear Gaussian models, which are referred to as models M0, M1, M2 and M3. Graphical representations of these models are given in Figure 3.2.

Each model consists of a 'stimulus' variable S and two 'response' variables, $R1$ and $R2$, simulating response signal features from both imaging modalities. As in the experimental context, the stimulus variable distribution is under control of the experimenter and hence deterministic and uniform, while the response variables depend on the state of the stimulus variable in different ways, depending on the particular model. An intuitive description of the models is given below.

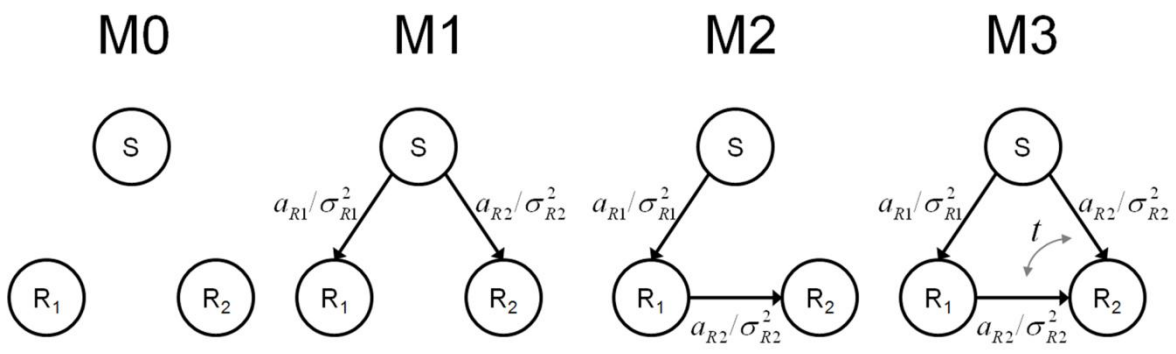


Figure 3.2 Graphical models used in the simulations. In the framework of probabilistic graphical models, each node (circle) represents a random variable, while the links (arrows) express probabilistic relationships between these variables. $\boldsymbol{a_{R1}, a_{R2}, \sigma^2_{R1}, \sigma^2_{R2}}$ and $\boldsymbol{t}$ refer to the model parameters detailed in the main text C. For linear Gaussian models, the distributions of individual variables as well as the distribution over all variables is Gaussian. In the current application only the response variables $\boldsymbol{R1}$ and $\boldsymbol{R2}$ are Gaussian, while the stimulus distribution $\boldsymbol{S}$ is uniform.

Model M0 is a 'reference' or Gaussian null model. For M0 the response variables are independent of the stimulus variable, and of each other. This refers to a case in which the observed EEG and fMRI signal features are not related and non-informative with respect to the stimulus. Model M1 captures a scenario in which both response variables depend on the stimulus, but are independent of one another. This refers to a case in which both EEG and fMRI signal features are modulated by the stimulus, but they do not co-vary with one another. Model M2 captures a scenario, in which response variable $R1$ (e.g. an EEG signal feature) co-

varies with the stimulus, while the other response variable $R2$ (e.g. an fMRI signal feature) is dependent on the stimulus only indirectly via the first variable. Finally, Model M3 represents an intermediate version of models M1 and M2 in which the second response variable $R2$ (e.g. an fMRI signal feature) is sensitive to both changes in the stimulus variable S and the first response variable $R1$, which in terms of simultaneous EEG-fMRI experiments is possibly the most plausible assumption. It should be noted that models M1 and M2 represent special cases of model M3, which was implemented as a weighted linear interpolation between models M1 and M2.

3.3.2 Response signal model details

In the following section, the mathematical details of the models outlined in the previous section will be discussed.

A linear Gaussian model with one 'stimulus' variable S and two 'response' variables, $R1$ and $R2$, can be regarded as a linear mapping $f: S \to R$ from a 'stimulus set' $S = \{s_1, \dots, s_{|S|}\} \subset \mathbb{R}$ to a response set $R = \left\{\begin{pmatrix} r_1^1 \\ r_2^1 \end{pmatrix}, \dots, \begin{pmatrix} r_1^{|R|} \\ r_2^{|R|} \end{pmatrix}\right\} \subset \mathbb{R}^2$, given by

$$f(s) = as + b + \varepsilon \tag{3.5}$$

where $a, b \in \mathbb{R}^2$ and $\varepsilon \sim N(0, \Sigma) \in \mathbb{R}^2$. Here, a is a multiplicative coefficient, b an additive bias, and ε an additional error or noise term which is normally distributed with mean 0 and covariance matrix Σ. Equivalently, and for practical purposes more conveniently, one can obtain samples from the specified models by sampling from a bivariate Gaussian distribution with model specific expectation μ_M and covariance matrix Σ_{M}. Hence, one can write the distribution of the response variables given the stimulus as

$$r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \sim N(r|(\mu_M|s), \Sigma_{\mathrm{M}}) \tag{3.6}$$

where the distribution $N(r|(\mu_M|s), \Sigma_{\mathrm{M}})$ is given by a multivariate Gaussian density function as

$$N(r|(\mu_M|s), \Sigma_{\mathrm{M}}) = \frac{1}{(2\pi)^{N/2}|\Sigma_{\mathrm{M}}|^{1/2}} \exp\left(-\frac{1}{2}\left(r - (\mu_M|s)\right)^T {\Sigma_{\mathrm{M}}}^{-1}\left(r - (\mu_M|s)\right)\right) \quad (3.7)$$

As noted previously, the marginal stimulus distribution $p(s)$ is usually not empirical but rather is under the control of the experimenter. In the simulations (as well as in the real experiment) the marginal stimulus distribution was set to the discrete uniform distribution, i.e.

$$p(s_i) = \frac{1}{|S|} \quad (3.8)$$

In other words, the probability for stimulus s_i to occur on any given trial is the inverse of the cardinality $|S|$of the stimulus set.

In model M0, no dependence of the response variables on the stimulus was postulated. Hence, the dependence of the mean on the stimulus and the interaction between the response variables additive error components were removed by setting

$$\mu_{M0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \Sigma_{\mathrm{M0}} = \begin{pmatrix} \sigma_{R1}^2 & 0 \\ 0 & \sigma_{R2}^2 \end{pmatrix} \quad (3.9)$$

Hence, σ_{R1}^2 and σ_{R2}^2 are the only free parameters of model M0, and both individual response variables $R1$ and $R2$ are normally distributed with mean 0 and respective variances σ_{R1}^2 and σ_{R2}^2.

In model M1, both response variables depend on the stimulus, but not on each other. In order to implement the linearity constraints introduced above, the mean vector and covariance matrix for M1 were set to

$$\mu_{M1} = \begin{pmatrix} a_{R1}s + b_{R1} \\ a_{R2}s + b_{R2} \end{pmatrix} \quad \text{and} \quad \Sigma_{\mathrm{M1}} = \begin{pmatrix} \sigma_{R1}^2 & 0 \\ 0 & \sigma_{R2}^2 \end{pmatrix} \quad (3.10)$$

where $a_{R1}, a_{R2}, b_{R1}, b_{R2}$, σ_{R1}^2 and σ_{R2}^2 are free and independent parameters, and $s \in S$ represents the stimulus value. The joint stimulus response probability distribution of model M1 can then be sampled as

$$p_{M1}(s, r_1, r_2) = N(r|(\mu_{M1}|s), \Sigma_{\mathrm{M1}})) \tag{3.11}$$

For model M2, $R2$ depends on the value of S indirectly by means of $R1$, which depends directly on S. Hence, the mean and covariance matrix for M2 were set to (for a derivation of these results, see (Bishop, 2006), pp. 370)

$$\mu_{M2} = \begin{pmatrix} a_{R1}s + b_{R1} \\ a_{R2}(a_{R1}s + b_{R1}) + b_{R2} \end{pmatrix} \text{ and } \Sigma_{\mathrm{M2}} = \begin{pmatrix} \sigma_{R1}^2 & a_{R2}\sigma_{R1}^2 \\ a_{R2}\sigma_{R1}^2 & {a_{R2}}^2\sigma_{R1}^2 + \sigma_{R2}^2 \end{pmatrix} \tag{3.12}$$

again with free parameters $a_{R1}, a_{R2}, b_{R1}, b_{R2}$, σ_{R1}^2 and σ_{R2}^2. Hence, the joint response probability distribution of model M2 can be sampled from

$$p_{M2}(s, r_1, r_2) = N(r|(\mu_{M2}|s), \Sigma_{\mathrm{M2}})) \tag{3.13}$$

Finally, model M3 was created as an intermediate scenario between models M1 and M2. In model M3, response variable $R2$ is sensitive to both changes in variable S and variable $R1$ depending on a linear weighting constant $t \in [0,1]$, such that for $t = 0$ model M3 is identical with model M1 and for $t = 1$ model M3 is identical to model M2. Hence, the mean vector and covariance matrix of model M3 are for $t \in [0,1]$

$$\mu_{M3} = \begin{pmatrix} a_{R1}s + b_{R1} \\ ta_{R2}(a_{R1}s + b_{R1}) + (1-t)a_{R2}s + b_{R2} \end{pmatrix} \tag{3.14}$$

$$\Sigma_{\mathrm{M3}} = \begin{pmatrix} \sigma_{R1}^2 & ta_{R2}\sigma_{R1}^2 \\ ta_{R2}\sigma_{R1}^2 & t({a_{R2}}^2\sigma_{R1}^2 + \sigma_{R2}^2) + (1-t)\sigma_{R2}^2 \end{pmatrix} \tag{3.15}$$

and the joint response probability distribution of model M3 can be sampled from

$$p_{M3}(s, r_1, r_2) = N(r|(\mu_{M3}|s), \Sigma_{\mathrm{M3}})) \tag{3.16}$$

In the simulation results below, the parameter t was set to $t = 0.5$ to achieve an intermediate scenario between models M1 and M2.

3.4 Numerical Simulation Results

Below, the following simulations are discussed: 1) the analysis of a Gaussian null model (M0), which is of experimental relevance with respect to the implemented bias estimation and correction methods. 2) the analysis of three linear Gaussian models implementing common assumptions about stimulus-EEG-fMRI interdependencies (M1, M2, M3) in order to investigate the differentiating power of the framework, and 3) a validation of the selected number of response bins for the histogram analysis.

In the simulations reported below, the following parameters for each response variable $R1$ and $R2$ were manipulated : a) multiplicative gain factors a_{R1} and a_{R2} which amplify the respective response variable input and b) random additive error terms which are governed by variance terms σ_{R1}^2 and σ_{R2}^2. To sample from a bivariate Gaussian distribution use of the *gsamp.m* function of the netlab toolbox (Nabney, 2002) was made. All simulations were carried out using the parameters of the visual stimulation experiment detailed in Chapter 4, i.e. a stimulus set S of size $|S| = 2$, with $s_1 = 1$ ('low contrast') and $s_2 = 2$ ('high contrast'), and with $N_S = 85$ trials per stimulus condition, unless otherwise stated.

It should be noted, that for two equally distributed stimuli, which each occur with $p(s_i) = 0.5$, the entropy of the stimulus probability distribution is given by

$$H(S) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1\ bit \qquad (3.17)$$

As the response distribution (both in the case of artificial and experimental data) will show larger variability and hence larger entropies, the mutual information between stimulus and responses is therefore bounded from above by 1 bit (Cover and Thomas, 2006).

3.4.1 M0 simulations

In model M0, the response variables are independent of the stimulus and of each other. For an unbiased estimate of the respective information theoretic quantities one hence expects a value of zero, indicating no information content. Figure 3.3 displays the results for 100 model realizations of M0 with parameters $a_{R1} = a_{R2} = \sigma_{R1}^2 = \sigma_{R2}^2 = 1$ and $b_{R1} = b_{R1} = 0$ for the Plug-In (upper panel) and bias corrected (lower panel) versions of the information theoretic quantities (for a definition of these quantities see 2.4.2 and 2.4.3). All quantities are positively biased in the plug-in case, and except for the quantity $I_N(S; R_1, R_2)$ are reduced to approximately zero after bias correction, with the largest variability displayed by $\langle I_N(R_1;\ R_2|S)\rangle_S$, due to its estimation from half of the number samples. For $I_N(S; R_1, R_2)$ and the closely related $Synergy$ the shuffling bias correction results in a bias over-correction of roughly – 0.1 bits. Further simulations revealed that this pattern was independent of the model parameter settings as shown in Figure 3.4.

However, additional simulations (Figure 3.5) revealed that the quality of the bias correction is a function of the experimental setting, i.e. the number of stimuli used and the number of repeats per stimulus. This indicates the bias correction procedures employed are actually proficient, however, mainly so outside the experimental regime (i.e. number of stimuli, number of trials per stimulus) under consideration. This reflects the fact that these procedures have been developed for electrophysiological recordings, where more experimental trials can be acquired in typical experimental data acquisition session.

Nonetheless, for independent Gaussian response variable distributions sampled twice at a sample size of 85, the bias correction for $I_N(S; R_1, R_2)$ was found to be too strong and improvable for $I_N(R_1;\ R_2)$ and $\langle I_N(R_1;\ R_2|S)\rangle_S$.

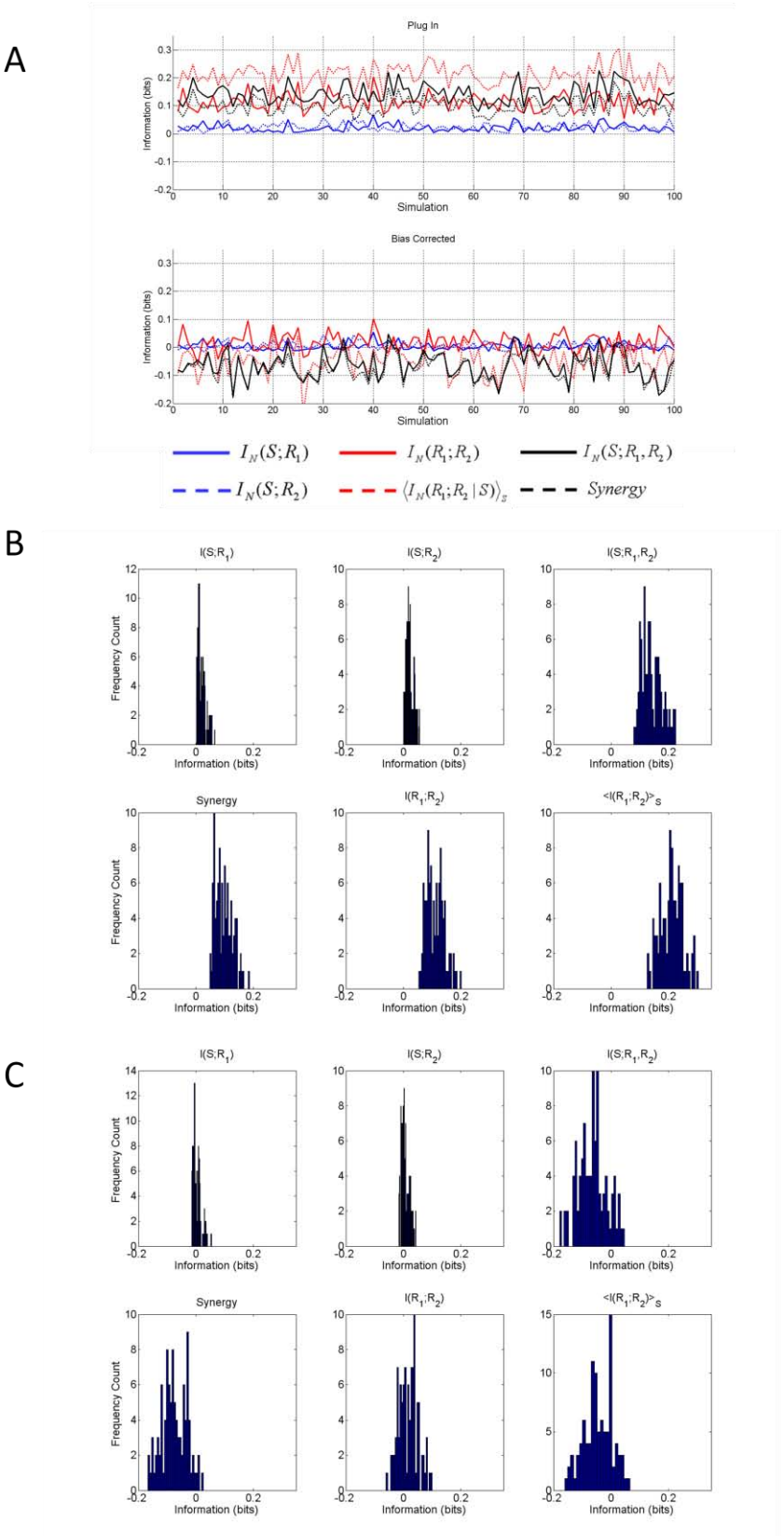


Figure 3.3 M0 simulations. A.Plug-in (upper panel) and bias corrected (lower panel) estimates of information theoretic quantities obtained from model M0 for 100 instantiations with parameter settings $\boldsymbol{a_{R1} = a_{R2} = \sigma^2_{R1} = \sigma^2_{R2} = 1}$ $\boldsymbol{and}$ $\boldsymbol{b_{R1} = b_{R1} = 0}$. B. The same data as in A, upper panel (uncorrected ITQs), displayed as frequency counts. C. The same data as in A, lower panel (corrected ITQs), displayed as frequency counts.

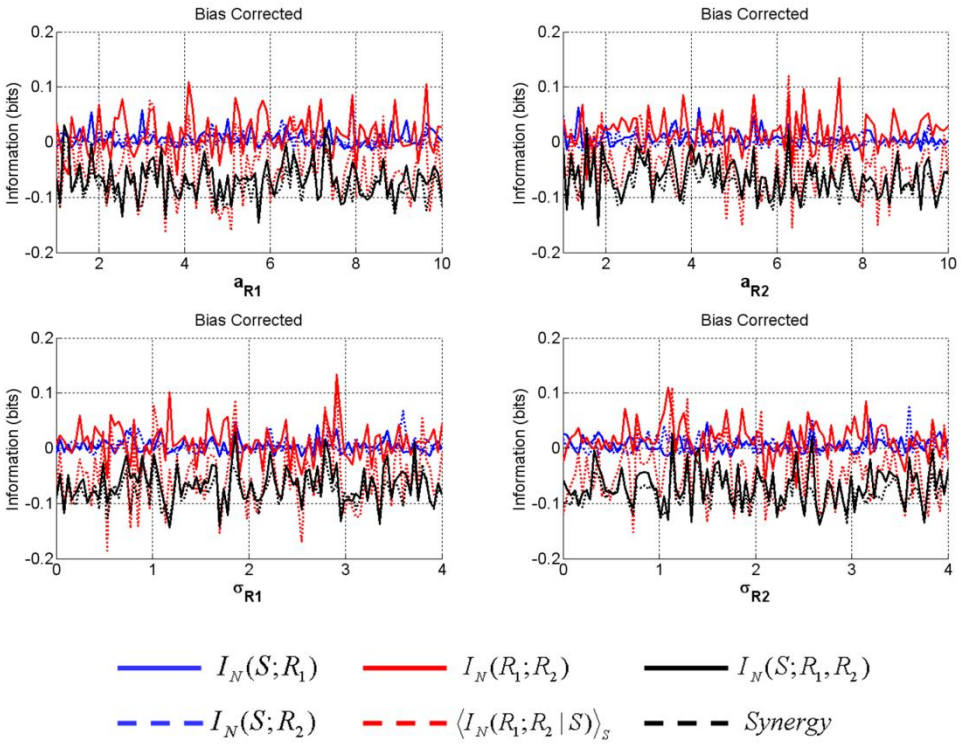


Figure 3.4 Variation of parameters of the Gaussian null model M0. For each of the panels the free parameters a_{R1} and a_{R2} were varied in [1,10] and σ_{R1}^2 and σ_{R2}^2 were varied in]0,4] using 100 models. For each parameter setting the information theoretic quantities that result from a single simulation are shown as a function of the respective parameter (Variation of b_{R1} and b_{R2} is not shown, but is always without effect on the information theoretic quantities as the value of the response variables does not co-vary with the stimulus by means of the additive bias. Hence, varying the additive bias parameter does not affect the difference between the respective stimulus dependent distributions (it just shifts them in parallel) and hence does not affect the information theoretic quantities. This implies the independence of the information theoretic quantities of the absolute value of the response variables). As can be seen, the variation of the parameters does not influence the estimation of the quantities validating the choice of $a_{R1} = a_{R2} = \sigma_{R1}^2 = \sigma_{R2}^2 = 1$ and $b_{R1} = b_{R1} = 0$ for estimation of the $\Delta Bias$ terms.

To take account of this observation in the simulations of the other models and the experimental settings discussed in Chapters 4,5 and 6, the following approach was taken: to amend the observed (negative) biases for the Gaussian Null model, the remaining bias magnitude after bias correction ($\Delta Bias$) for each information theoretic quantity was estimated from 10000 simulations of the model with $a_{R1} = a_{R2} = \sigma_{R1}^2 = \sigma_{R2}^2 = 1$ and $b_{R1} = b_{R1} = 0$. The resulting point estimates are shown in Figure 3.6 together with their estimated 0.999 confidence intervals ($t_\infty = 3.29$). These point estimates (from a total 1000 equivalent

simulations) were added to the bias corrected information theoretic quantities in all following analyses (simulated and experimental data).

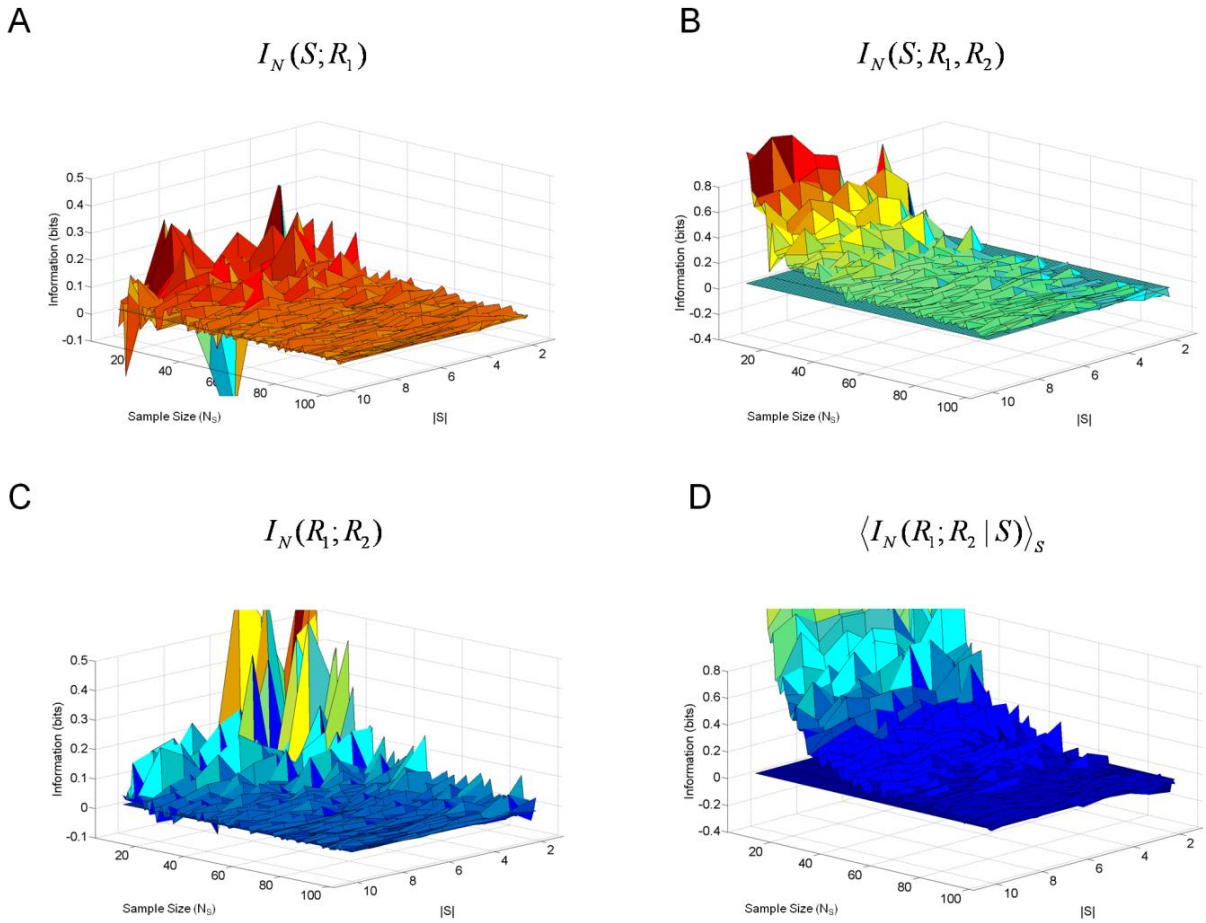


Figure 3.5 Bias and experimental parameters**:** Influence of experimental parameters stimulus set size $|S|$ and sample size per stimulus (N_S, 'Sample size' in the figure) on the bias corrected information theoretic quantities estimated from model M0. Panels A, B and C indicate good bias correction for our experimental regime $|S| = 2, N_S = 85$ while panel D indicates the strong overcorrection observed in the case of shuffling correction of $I_N(S;R_1,R_2)$ for our experimental parameters. However, in different experimental parameter regimes (possibly closer to those commonly employed in invasive neurophysiological studies), the shuffling correction is also efficient and does not overcorrect.

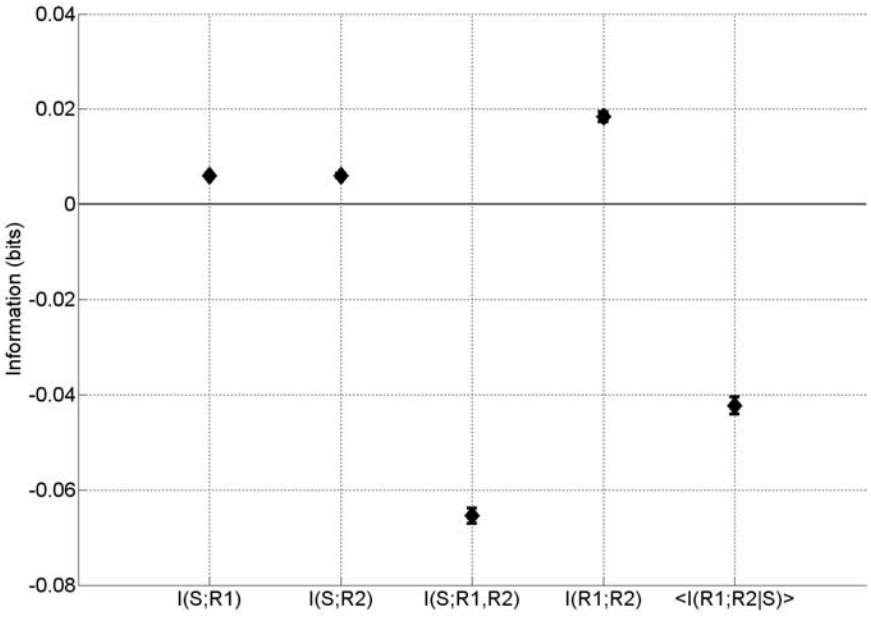


Figure 3.6 Bias Correction: Bias corrected estimates of the information theoretic quantities obtained from 1000 simulations of model M0 with parameter settings as in A). The error bars indicate the 0.999 confidence intervals for these point estimates.

3.4.2 M1 - M3 coupling coefficient simulations

Figure 3.7 shows the results of varying the multiplicative constants a_{R1} and a_{R2} individually and in parallel in the interval [1,10] which effectively covers the low to high SNR regimes in the parameter space chosen for the simulations.

In model M1 (first column of Figure 3.7 A), an increase in a_{R1} leads to an increase in both $I_N(S;R_1)$ and $I_N(S;R_1,R_2)$ to the boundary level of 1 bit, as the corresponding observed frequency distributions for stimulus 1 and stimulus 2 become non-overlapping. $I_N(S;R_2)$ and $I_N(R_1;R_2)$ are virtually zero, and not affected by variation of a_{R1}. As the multiplicative factor at the level of R_1 outweighs the level of variance at the level of R_2, the variables become slightly redundant, as indicated by the negative $Synergy$.

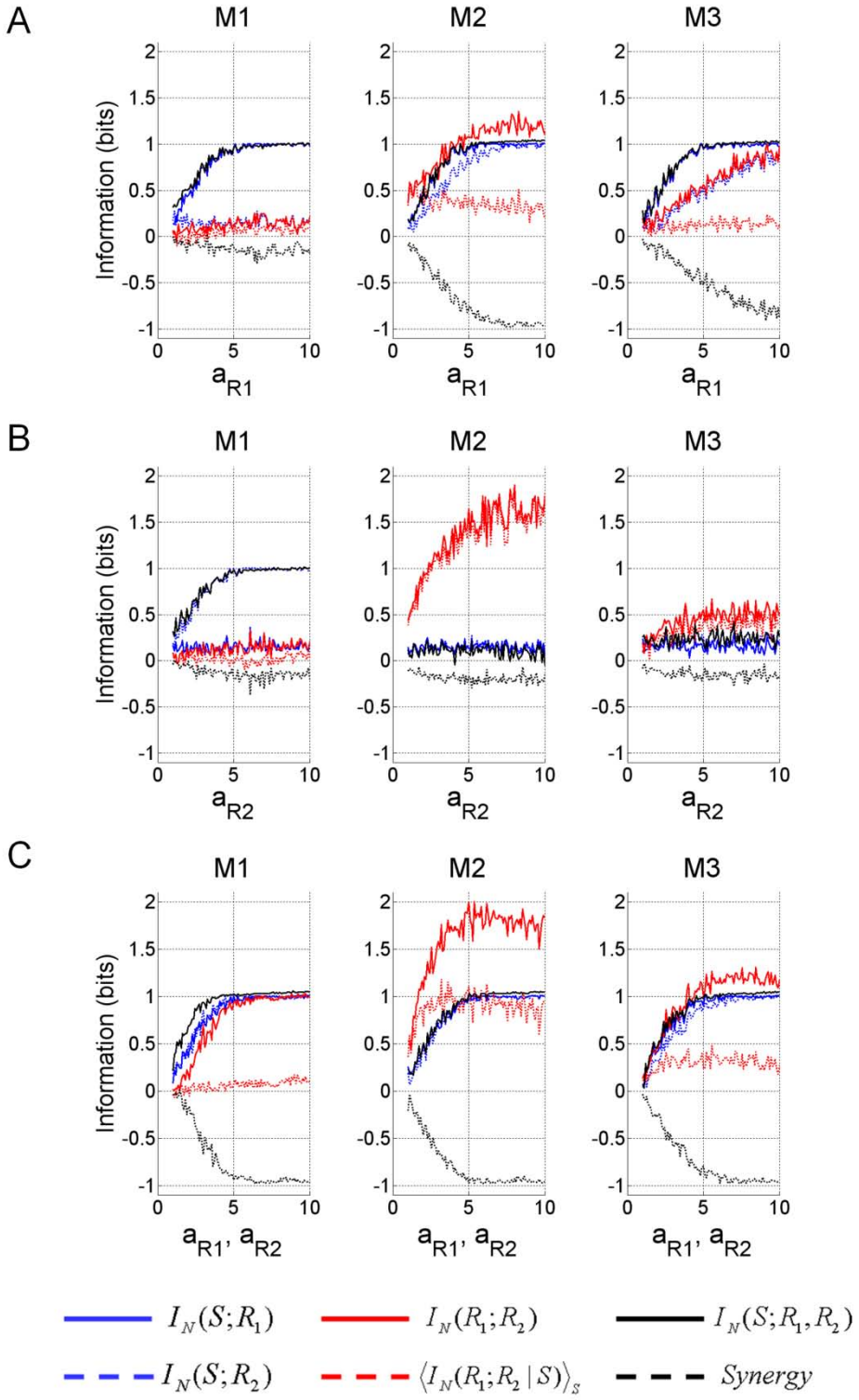


Figure 3.7 Variation of multiplicative gain factors a_{R1} and a_{R2} in models M1-M3. For each panel 100 simulations in the respective parameter space were carried out. A) Variation of a_{R1} only, B) Variation of a_{R2} only, C) parallel variation of a_{R1} and a_{R2}. For each parameter setting the information theoretic quantities that result from a single simulation are shown as a function of the respective parameter

In model M2 (second column of Figure 3.7 A), an increase of parameter a_{R1} leads to an increase of the information about the stimulus contained in both response variables in isolation ($I_N(S;R_1)$ and $I_N(S;R_2)$), and combined ($I_N(S;R_1,R_2)$). Most notably the activity dependence $I_N(R_1;\,R_2)$ reaches a maximum of 1.1 bits, as the response variables become fully redundant.

In model M3 (third column of Figure 3.7 A), variation of parameter a_{R1} corresponds to a modulation of the connection from the stimulus variable S to response variable R_1, while the connections between the stimulus variable and response variable R_2 and between both response variables remain constant. Correspondingly, $I_N(S;R_1)$ and $I_N(S;R_1,R_2)$ increase to the boundary level as in the simulation of M1, while in addition, $I_N(S;R_2)$ and $I_N(R_1;\,R_2)$ increase with larger parameter values of a_{R1} reflecting the introduced link between the two response variables. It is worth noting, that the response variable co-variation is purely stimulus induced, i.e. $\langle I_N(R_1;\,R_2|S)\rangle_S$ remains close to zero.

In model M1 variation of parameter a_{R2} evokes the complementary information theoretic signature with respect to variation of a_{R1} (first column of Figure 3.7 B): here, $I_N(S;R_2)$ increases with $I_N(S;R_1,R_2)$ while the other quantities remain largely unaffected (only the $Synergy$ again decreases slightly).

Varying a_{R2} in model M2 (second column of Figure 3.7 B) reveals a different picture: The information contained in the response variables with respect to the stimulus is unaffected (i.e. $I_N(S;R_1)$, $I_N(S;R_2)$ and $I_N(S;R_1,R_2)$ do not vary with a_{R2}) while activity and conditional dependence rise in parallel, indicating the strong amplification of the random noise component of R_1 by R_2.

Finally, varying a_{R2} in model M3 (third column of Figure 3.7 B) corresponds to increasing both the information input from the stimulus variable S as well as the noise input of R_1 to R_2. Consequently, all quantities show relatively little variation in this complementary scenario. Interestingly, due to the larger entropy of response variable R_1 compared to stimulus variable S, the activity and conditional independence ($I_N(R_1;\ R_2)$, $\langle I_N(R_1;\ R_2|S)\rangle_S$) increase slightly over the stimulus related information in both response variables.

Finally, the variation of a_{R1} and a_{R2} in parallel in the interval [1,10] revealed different information theoretic signatures for all models (Figure 3.7 C). In model M1 (first column) the information in the response variables about the stimulus reaches a maximum of 1 bit for all cases, $I_N(S;R_1)$, $I_N(S;R_2)$and $I_N(S;R_1,R_2)$. Hence the variables become fully redundant with respect to the stimulus, resulting in a negative $Synergy$ of -1. However, the conditional correlation $\langle I_N(R_1;\ R_2|S)\rangle_S$ remains close to zero, as the two variables are not subjected to stimulus independent co-variation. For model M2 (second column) all quantities are dependent on the two multiplicative coefficients and reach extreme values. The two response variables become fully redundant. The information theoretic signature of model M3 (third column) shows an intermediate pattern between the simulations of M1 and M2, as indicated by the behaviour of activity and conditional dependence. Superficially, the observed pattern is reminiscent of the variation of parameter a_{R1} in model M2. However, the activity dependence approaches zero for small parameters values in M3, while it remains above zero for the simulation of model M2. This reflects the fact that a_{R2} is constant in the a_{R1} simulation of model M2, but not in the parallel a_{R1}/a_{R2} simulation of model M3.

3.4.3 Validation of the selected number of response bins

To the best knowledge of the author there exists no investigation that demonstrates how to obtain an optimal trade-off between sensitive information estimation and bias control

for combined EEG and fMRI data. This is an important issue which requires additional theoretical work to provide a fully optimised solution, one of which is introduced in Chapter 7. However, as described above reasonable precautions were chosen not to overestimate the information inherent in the experimental data based on established bias correction methods and the Gaussian null model simulations. The number of response bins was chosen to obtain approximately twice as many trials per stimulus as response bins, which, based on simulations with respect to neurophysiological data (Panzeri et al., 2007; Pola et al., 2003) should yield conservative information estimates. In this section, simulations of models M0, M1, M2 and M3 which further validate this choice will be discussed.

Figures 3.8 A and 3.8 B demonstrate the effects of varying the number of response bins per response variable between discrete values of 2 to 9, resulting in a total number of response bins varying between 4 and 81. As the effect of varying the number of response bins is dependent on the specific structure of the analyzed data, two scenarios were simulated, a low signal-to-noise scenario in which the ratio between parameters a_{R1}, a_{R2} and $\sigma_{R1}^2, \sigma_{R2}^2$. was set to 2 (Figure 3.8 A) and a high signal-to-noise scenario (Figure 3.8 B), with this parameter ratio set to 10. For models M1, M2, and M3 the additional bias correction based on M0 was applied. As indicated by the M0 simulations in both Figures 3.8 A and 3.8 B, increasing the number of response bins leads to diminishing control of the bias mainly for $I_N(S; R_1, R_2)$ and $\langle I_N(R_1;\ R_2|S)\rangle_S$. However, simultaneously the sensitivity of the analysis to actual information in the data increases, as is particularly evident for activity and conditional dependence of model M2 and to a lesser degree model M3. Based on these results, it was reasoned that choosing the number of response bins to equal 36 in total maximizes the sensitivity to information present in the data, while keeping the loss of bias control at bay. Note that this is specific to the experimental scenario of 85 trials per stimulus and two different stimuli.

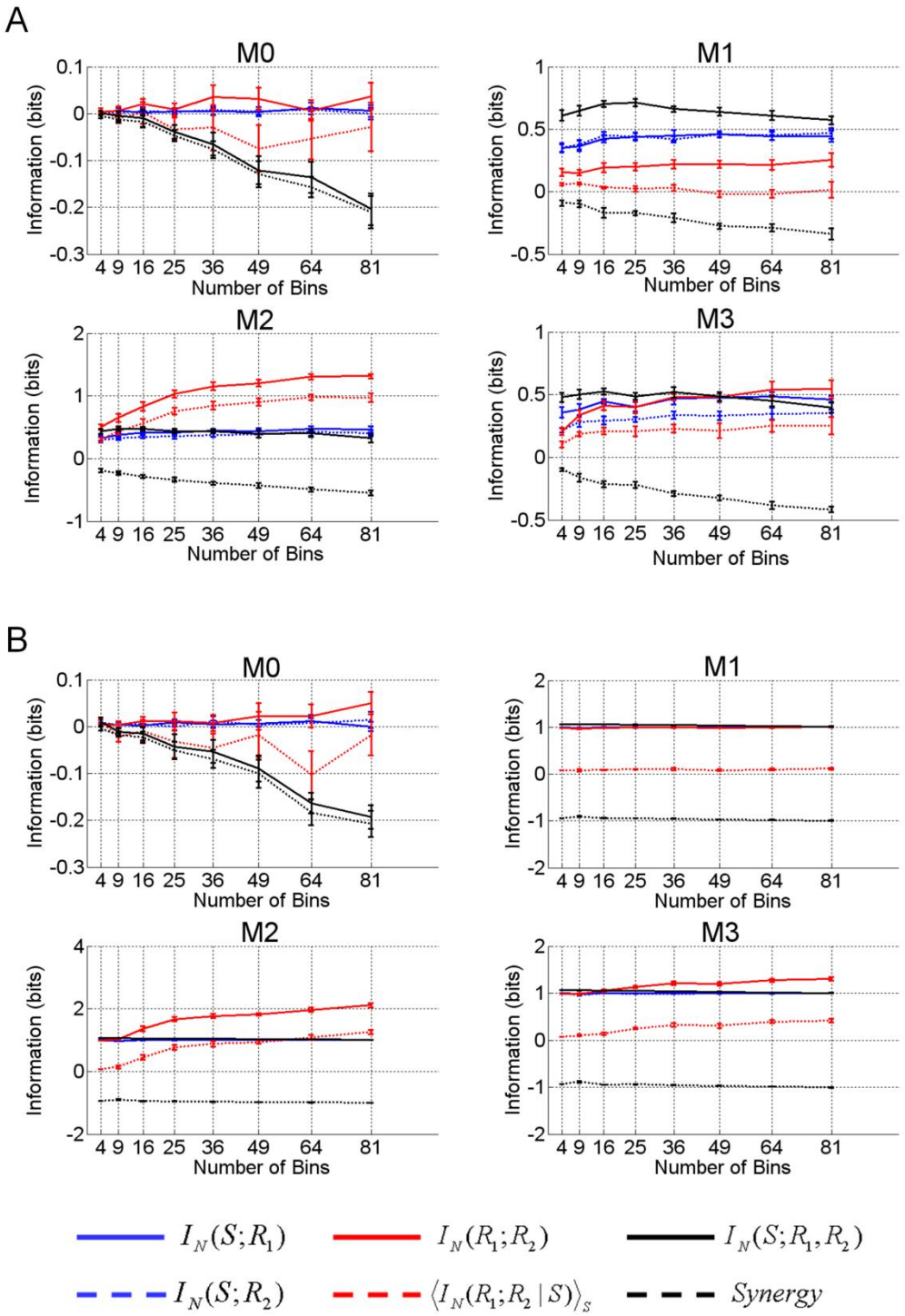


Figure 3.8 Validation of number of response bins: Variation of the number of response bins in the information theoretic analysis of artificial data generated by the set of linear Gaussian models M0 – M3. A) low signal-to-noise scenario $a_{R1} = a_{R2} = 2$, $\sigma_{R1}^2 = \sigma_{R2}^2 = 1$ B) high signal-to-noise scenario ($a_{R1} = a_{R2} = 10$, $\sigma_{R1}^2 = \sigma_{R2}^2 = 1$)

Future studies based on more reasonable assumptions about the underlying probability distributions in the case of combined EEG-fMRI data might provide further insight into the question of response space regularization.

3.5 Validating the Gaussian Method

In (Magri et al., 2009), the authors suggest, that the most appropriate bias correction method for the case of continuous analogue brain signals (in contrast to the discrete nature of spike counts) is the Gaussian method (2.4.4). To investigate this claim and in order to validate the implementation of the Gaussian method in the information breakdown toolbox code provided by (Magri et al., 2009), data were sampled from model M3 with parameter settings $a_{R2} = \sigma_{R1}^2 = \sigma_{R2}^2 = b_{R1} = b_{R1} = 1$, $t = 1$, $p(s_i) = 0.5$, $i = 1{,}2$, 85 trials per stimulus, varying a_{R1} between 1 and 10, and analyzed with the direct and Gaussian method options of the information breakdown toolbox. The results of these simulations are depicted in Figures 3.9 A (direct method) and 3.9 B (Gaussian method).

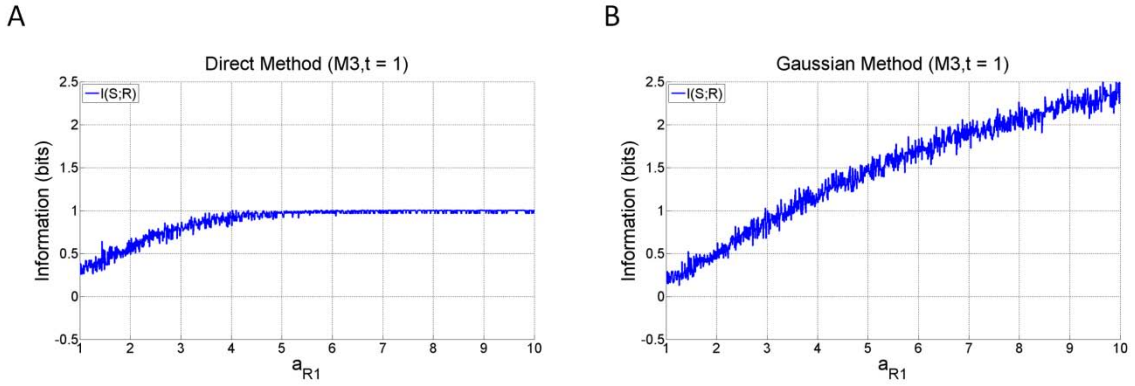


Figure 3.9 Validating the Gaussian Method I: Comparison of stimulus related information estimation for a M3 model simulation for the direct (A) and Gaussian (B) method of the information theoretic toolbox. As evident from the figure, the Gaussian method results in information estimates which are larger than the theoretical upper bound for two stimulus levels, 1 bit. The direct method respects this limit.

The two outcomes are qualitatively similar irrespective of whether a bias correction procedure provided by the toolbox is used or not. The outcome is surprising for two reasons: first, one would expect the direct method and the Gaussian method to give approximately the same results. Second, the Gaussian method results in information estimates above the theoretical limit of 1 bit. As discussed in Chapter 2, the information carried by an n-dimensional response variable about a stimulus variable is bounded from above by the entropy of the stimulus or the response variable, whichever is smaller (Schneidman et al., 2003). Intuitively, in the absence of any noise entropy, i.e. variability in the response at fixed stimulus, $H(R|S)$ is zero, and hence

$$I(S;R) = H(R) - H(R|S) = H(R) - 0 = H(R) \tag{3.17}$$

In this case, the response entropy $H(R)$ is fully induced by the stimulus and hence the variability in the response variable is maximally equal to $H(S)$. Effectively, the result reported above indicates, that the Gaussian method results in information estimates which are theoretically impossible.

To investigate why this is the case, it is instructive to consider the simulated response joint distributions directly, e.g. for a parameter setting of $a_{R1} = 5$ as shown in Figure 3.10. As evident from the figure, the joint distribution $p(r_1, r_2)$, on whose covariance matrix the estimation of the response entropy is based, is non-Gaussian. Specifically, due to the model properties at the given parameter settings it is bimodal, corresponding to a mixture of two Gaussian distributions. However, the computation of the entropy of this distribution according to

$$\widehat{H}^{(g)}(R) = \frac{1}{2}\log_2\left((2\pi e)|\sigma^2(R)\,|\right) \tag{3.17}$$

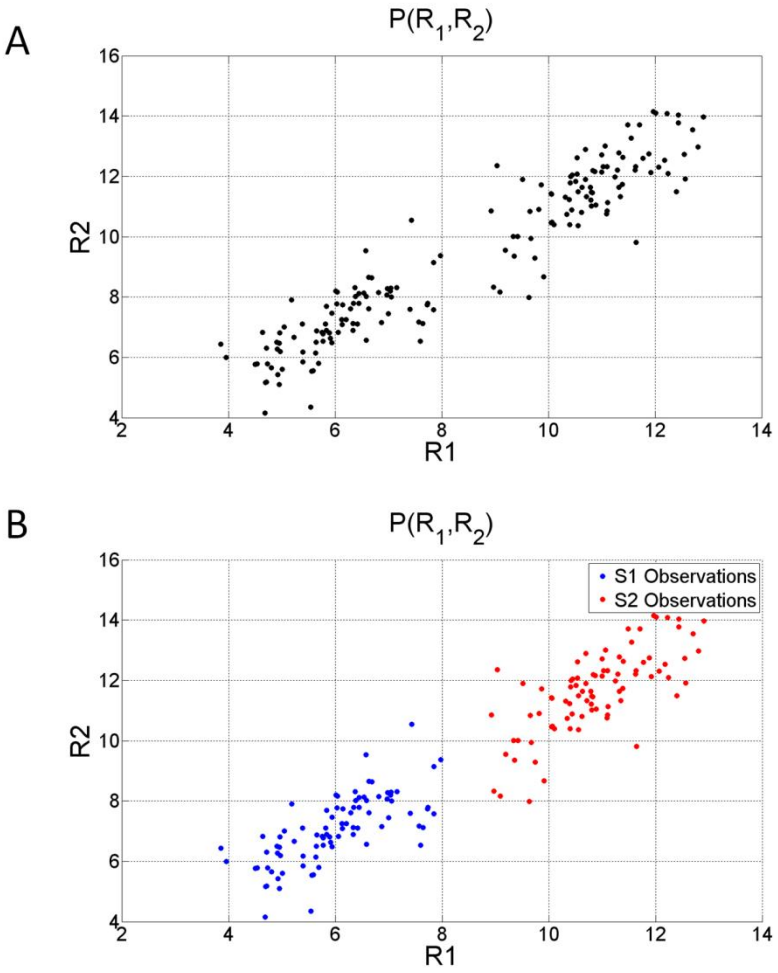


Figure 3.10 Gaussian mixtures: Example of a joint response variable distribution for the M3 simulation discussed above, where $\boldsymbol{a_{R1}} = \mathbf{5}$. Critically, the joint distribution (incomplete data, Bishop, 2006) depicted in A is bimodal, and hence is not well characterized by a Gaussian distribution. The two modes correspond to the means of the conditional response distributions for the two stimuli as evident from B (complete data, Bishop, 2006).

requires the distribution to be Gaussian, and hence uni-modal. It appears, that this assumption can only apply in the case of non-informative signals, i.e. in the case when the stimulus does not change the average response profile.

In (Magri et al., 2009) the authors performed a comparison of the Gaussian and Direct method and concluded that they result in the same information estimates. However, it appears that the computed information estimates reported for the two methods are not identical and in fact shifted to higher values for the Gaussian method (Figure 3.11 A, reproduced from (Magri

et al., 2009)). A similar analysis of the simulated data (variation of parameter a_{R1}) here revealed a linear relationship between the two information estimates can indeed be observed, however only in a specific region of the parameter space of the model under study (Figure 3.11 B)

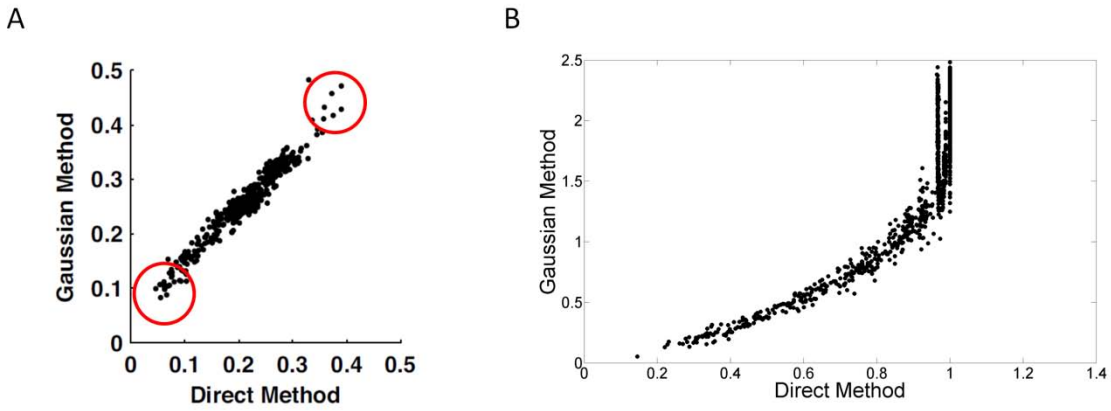


Figure 3.11 Validating the Gaussian Method II: Comparison of the information estimates for Gaussian and direct method from (Magri et al., 2009) (A) and the simulations reported here (B). As evident from A, the information estimates reported by Magri et al. are indeed not identical for direct and Gaussian method, but shifted to larger values by approximately 0.05 bits with respect to the Gaussian method. The linear relationship between both estimates is also observed within a certain parameter space range for the simulations reported here, corresponding to the range of 0 to 0.5 bits. However, the direct method is bounded from above by 1 bit, while the Gaussian method increases further.

To conclude, the Gaussian method, which was intended to facilitate information estimation in the case of continuous response signals, is invalid in the current experimental context and will not be applied for information estimation of real data in this thesis. An alternative approach capitalizing on the use of the analytical properties of the Gaussian distribution and its potential usefulness for non-invasive brain imaging signal is based on Gaussian mixture models (Bishop, 2006) and will be introduced in Chapter 7.

3.6 Discussion

In summary, these simulations indicate the following: First, the PT and shuffling bias correction methods work to reduce the estimated information theoretic quantities from a

limited sample to zero in the case of no posited stimulus-response signal and response signal-response signal dependencies (model M0). However, the shuffling correction of $I_N(S; R_1, R_2)$ (and to a lesser degree the bias correction of $\langle I_N(R_1; R_2|S)\rangle_S$) overcorrects the estimated quantity in the case of a stimulus set of size $|S| = 2$, and $N_S = 85$, which was amended by estimating this overcorrection from repeated simulations, and then adding the estimated overcorrection to the shuffling corrected estimate of $I_N(S; R_1, R_2)$. Although the remaining bias is less severe for the other quantities, the same procedure is applied to the other quantities for consistency. The values estimated from the Gaussian null model are applied to the analysis of the experimental data discussed in Chapter 4, 5 and, with a modification for the number of experimental trials, 6. Second, the assessment of the information theoretic quantities computed from data simulated from models M1, M2 and M3 demonstrate that the data analysis implementation yields results which are compatible with the interdependencies of the response variables posited in the models. In addition, these simulations demonstrate the differentiating power of the theoretical framework as for all parameters variations different behaviours of the information theoretic quantities for the three models were observed. Third, analyzing the models in the experimental scenario under consideration for Chapter 4 and 5 (85 trials per stimulus, two different stimuli) indicates that setting the number of response bins in one dimension to approximately two-thirds of the square-root of the of the number of trials per stimulus yields an optimal trade-off between specificity and sensitivity for the information theoretic analysis.

Clearly, different experimental scenarios (e.g. number of stimulus conditions, number of trials per stimulus conditions) and data acquisition contexts (number of response modalities, dimensionality of the response variable space) require additional simulations to establish the optimality of bin number selection for histogram based information estimation and null model bias correction. Nevertheless, the approach taken here is generic and

generalizes well to other scenarios, conditional on the validity of the linear Gaussian model assumptions implemented.

Finally, the Gaussian method, proposed by (Magri et al., 2009) for the information theoretic analysis of continuous, non-invasive brain imaging signals is not optimal in its current form, as the joint response distribution in any realistic neuroimaging context with a limited number of conditions and experimentally induced effects will not be a (unimodal) Gaussian. A possible remedy for the application of Gaussianity assumptions to the estimation of information by means of Gaussian mixture models will be discussed in Chapter 7.

4 Visual Stimulation Experiment

4.1 Introduction

The visual system lends itself to the application of the information theoretic framework proposed in this thesis, as it is readily accessible during simultaneous EEG-fMRI recordings and well-studied (Becker et al., 2010; Gazzaniga MS, 2004; Kandel E.R., 2004). An extensive literature exists on the neural phenomena following visual stimulation as studied with EEG (Odom et al., 2004), fMRI (Wandell et al., 2007), as well as intra-cortical recordings (Kreiman, 2007). Recent studies using combined EEG and fMRI recordings have mainly followed two approaches (Becker et al., 2010): first, fMRI informed EEG analyses in which dipole solutions to the EEG inverse problem are spatially compared to fMRI activations (Bonmassar et al., 2001b; Di Russo et al., 2002; Di Russo et al., 2007); second, EEG-informed fMRI data analyses, in which spontaneous or task-related EEG activity features are used as predictors in a GLM analysis of fMRI data (e.g. (de Munck et al., 2007a; Goldman et al., 2002; Philiastides and Sajda, 2007)). Despite the number of studies carried out, no study has so far assessed stimulus related information from simultaneous EEG-fMRI recordings. Neither has thus far single-trial visually evoked activity been correlated with the BOLD signal, as has been done in other modalities (Benar et al., 2007; Debener et al., 2005; Eichele et al., 2005). A previous study, which investigated EEG and fMRI single trial variability and their relationship used data recorded in separate sessions (Bagshaw and Warbrick, 2007).

In the current study, a reversing checkerboard stimulus with two different contrast levels was chosen as visual stimulus during simultaneous EEG-fMRI data acquisition. Checkerboards have been established as robust stimuli evoking visually responses in both EEG and fMRI and are used as standard diagnostic tool clinically (Logothetis et al., 2001; Odom et al., 2004). The contrast sensitivity, i.e. the stimulus informativeness of marginal

EEG and fMRI feature distribution, has been established previously and serves as a validation for the methods proposed here (Boynton et al., 1999; Shawkat and Kriss, 2000). The stimulus-informativeness of EEG-fMRI feature joint distributions and its relationship to the informativeness of the respective marginal distributions has so far not been studied.

In the current study, the time domain response signal features subjected to the information theoretic approach are chosen based on a priori spatial constraints and the intuition that the strongest signals are also the most informative ones (Nevado et al., 2004): the maximal visually evoked potential is expected for occipital electrodes, whereas the most visually-responsive voxels will be identified using a GLM approach.

In sum, the aim of the current chapter is to practically apply the information theoretic framework introduced in Chapters 1 – 3 to simultaneously acquired EEG-fMRI data under visual checkerboard stimulation with two contrast levels. The code used in for stimulus presentation and data analysis is available from http://www.buic.bham.ac.uk/downloads/EEG_FMRI_ITQ/EEG_FMRI_ITQ_Analysis.zip.

4.2 Methods

4.2.1 Subjects

Fourteen subjects (6 male, 8 female, mean age 27.2 years, opportunistic sample) were recruited from the University of Birmingham campus and paid for their participation. All observers had normal or corrected to normal vision, no history of neurological disorders and gave written informed consent. The study was approved by the local ethics committee.

4.2.2 Stimuli

Left hemi-field reversing checkerboards were presented at a spatial frequency of 2 cycles per degree of visual angle at two different contrast levels (i.e. the size of the stimulus

set $|S|$ was 2), high ($C_{Michelson}$ = 1) and low ($C_{Michelson}$ = 0.25). These Michelson contrasts refer to the driving luminance ratios, the actually back-projected stimulus contrasts are likely to slightly deviate from these values). Stimuli were presented with a central fixation cross. For the simultaneous EEG-fMRI data acquisition, the stimuli covered the entire MR scanner projection screen left of the midline and subtended approximately 30 degrees of visual angle.

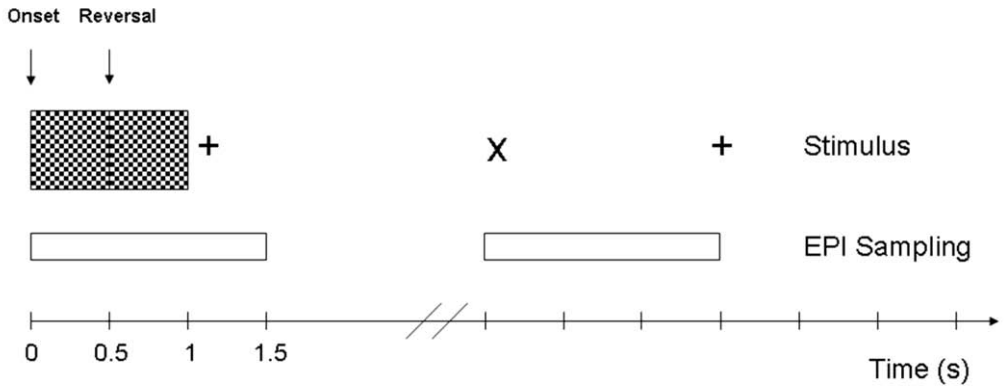


Figure 4.1 Single-trial experimental design of the visual stimulation experiment. A hemi-field checkerboard of either low or high contrast was presented at time 0 and reversed after 500 ms. After another 500 ms the checkerboard was removed. At a subset of the trials, the fixation cross changed to an x at the onset of a randomly sampled EPI volume acquisition, and the observer was asked to indicate this change by a button press.

4.2.3 Design and Procedure

Individual trials (Figure 4.1) of the experiment consisted of a single presentation of the checkerboard stimulus for 1 s with phase reversal after 500 ms followed by a fixation period which was uniformly sampled from 16.5 - 21 s, discretised to 1.5 s (MR repetition time, see below). Individual sessions consisted of 17 trials per condition as well as fixation periods at the beginning (8 s) and at the end (variable), amounting to a total session length of 441 volumes x 1.5 s, i.e. 11 minutes. For each observer, data was acquired during five experimental sessions, yielding $N_S = 85$ trials per condition and observer. The observer was asked to perform a simple task in order to maintain attention: on a random selection of half the trials of a given session, the fixation cross changed from a plus sign to an X during the

fixation period at random time points, discretely (1.5 s) and uniformly sampled from the interval of 4.5 - 16.5 s after stimulus onset. The task was to report the change in fixation by a button press using the index finger of the right hand. Hit rate and number of false alarms were presented to the observer at the end of each session. Stimuli were presented and behavioural data collected using Psychotoolbox3 (Brainard, 1997; Pelli, 1997) for Matlab. Stimulus presentation timing was controlled by the MRI scanner volume trigger.

4.2.4 MRI data acquisition

The experiment was conducted at the Birmingham University Imaging Centre using a 3T Philips Achieva MRI scanner. An initial T1-weighted anatomical (1 mm isotropic voxel) and T2*-weighted functional data were collected with an eight channel phased array SENSE head coil. EPI data (gradient echo-pulse sequence) were acquired from 20 slices (2.5 x 2.5 x 3 mm resolution, TR 1500 ms, TE 35 ms, SENSE factor 2, flip angle 80°, with equidistant temporal slice spacing to facilitate synchronisation of the EEG clock (see below)), providing approximately half brain coverage in the dorsal-ventral direction. Slices were oriented parallel to the AC-PC axis of the observer's brain and positioned to cover the entire occipital cortex.

4.2.5 EEG data acquisition

EEG data were recorded using a 64 channel MR compatible EEG system (BrainAmp MR Plus, Brain Products, Munich, Germany), which incorporates current limiting resistors of 5 kΩ at the amplifier input and in each electrode. The EEG cap consisted of 62 scalp electrodes distributed according to the 10-20 system and two additional electrodes, one of which was attached approximately 2 cm below the left collarbone for recording the ECG, while the other was attached below the left eye (on the lower orbital portion of the orbicularis oculi muscle) for detection of eyeblink artefacts. Data were sampled at 5000 Hz. Impedance

at all recording electrodes was less than 20 kΩ. The EEG data acquisition setup clock was synchronised with the MRI scanner clock using Brain Product's SyncBox, resulting in exactly 7500 data points per EPI-TR interval.

4.2.6 Data pre-processing

Raw EPI data (.par/.rec format) were converted to Niftii format using MRICroN's dcm2nii utility (Rorden and Brett, 2000). SPM5 (Friston, 2007) was used for fMRI data pre-processing. Data pre-processing involved anatomical realignment, slice scan time correction (reference slice 10), re-interpolation to 2 x 2 x 2 mm voxels, anatomical normalization and spatial smoothing (5 mm Gaussian kernel). The experimental data of each individual voxel was modeled using a standard GLM approach (Friston et al., 1994) implementing a high-pass filter to 1/128 Hz. Voxel time-courses were modeled in an event-related fashion using regressors obtained by convolving each stimulus onset unit impulse with a canonical haemodynamic response function and its first temporal derivative. This analysis strategy assumes that the EEG responses to stimulus onset, reversal, and offset are compounded in the related HRF, but was only used to identify visually active voxels with high statistical power. Specific relationships between the individual EEG events and the single-trial haemodynamic response are studied in the information theoretic section. Additional nuisance covariates included the realignment parameters to account for residual motion artefacts. After model estimation, voxels displaying a stimulus (S1 and S2) correlated main effect were identified, and filtered and whitened data was extracted from a sphere of 5 mm radius centred on the global maximum of the resulting statistical parametric map. This was done individually for each observer and the data were averaged over the voxels included in the sphere. A spherical ROI centered on the subjective-specific maximally active voxel was chosen as it allows a straight-forward standardization of data selection across observers and is a relatively common

procedure in neuroimaging (e.g. Noppeney et al., 2010; Stephan et al., 2010). A possible alternative would have been to select a per-subject cluster matched for the number of voxels. However, this procedure would have required specifying subject-specific significance thresholds and topographical voxel inclusion criteria. While potentially optimizing the observed effect sizes, this procedure would have drawbacks with respect to its reproducibility. Last, it is likely that due to the smoothness of the data, both approaches would essentially have resulted in very similar data sets. In the following, this data is referred to as 'advanced pre-processing'. In addition, in order to examine the effect of different pre-processing strategies on information theoretic quantities, we extracted voxel average data from a 5 mm radius sphere centered at the reported MNI coordinates of right primary visual cortex on the MNI coordinates [6 -82 -4] (Hasnain et al., 1998; Wohlschlager et al., 2005). This data is referred to as 'basic pre-processing'.

Raw EEG data were partitioned into data acquisition sessions and exported to .dat format using Brain Vision Analyzer (Brain Products, Munich, Germany). Subsequent pre-processing steps involved gradient artefact removal (Allen et al., 1998; Allen et al., 2000), pulse artefact correction (Optimal Basis Set method (Niazy et al., 2005)) as implemented as a plug-in to EEGLAB (Delorme and Makeig, 2004), low-pass filtering at 25 Hz, and down-sampling to 500 Hz.

Similarly to the advanced and basic pre-processing strategies adopted for the fMRI data, 'advanced pre-processing' EEG data underwent additional artefact removal. Specifically, after concatenation of the runs of an individual observer, a semiautomatic ICA-based procedure (Barbati et al., 2006; Porcaro et al., 2009) was used to identify artefactual non-cerebral components, i.e. eye-movements, residual BCG artefacts, and environmental noise. The computational underpinnings of ICA and details of the procedure employed here are discussed in Appendix C. The final data were obtained by reconstruction from all non-

artefactual components. Single-trial VEPs were then computed from the reconstructed data with respect to a 100 ms pre-stimulus baseline. Data was then extracted from the occipital electrode that showed the highest amplitude P100 in the average VEP across both conditions (selected from electrodes POz, Oz, PO4, PO8 and O2). The 'basic pre-processing' EEG data was extracted from the same electrode omitting the ICA step (i.e. only gradient and BCG artefacts were removed).

4.2.7 Response signal feature extraction

For both modalities the analysis was focussed on a set of single-trial time-domain features (Table 4.1) which were extracted from the EEG electrode or fMRI sphere using both basic and advanced pre-processing strategies as described above. For both modalities, a signal feature value from each trial of each condition was extracted, yielding a matrix of 85 x 2 signal feature values for each modality, which then entered the information theoretic analysis as the sample. For the EEG modality the analysis was focussed on three sets of features, which are grouped into amplitude features, latency features and root mean square (RMS) features. For the amplitude features the analysis was focussed on traditional visual evoked potential (VEP) peaks i.e. the N70, P100 and N140 for both stimulus onset and averaged over the two VEPs generated upon stimulus onset and reversal. In addition, the amplitude difference P100 – N140 for stimulus onset and averaged over stimulus onset and reversal was extracted. For the latency features, the time of the N70, P100, and N140 peaks with respect to stimulus onset from windows of 40 ms centred on the respective peaks was extracted. Lastly, as a measure of the overall power of the single-trial evoked potential, the RMS from a post-stimulus onset window of 50 – 250 ms, from a post-stimulus reversal window of 50 – 250 ms, and across the entire stimulus presentation period of 0 – 1000 ms was computed.

EEG Modality	
Feature	**Definition**
Amplitude Features	
N70 amplitude (O)	min{post-stimulus onset potential, (50 – 90 ms)}
N70 amplitude (OR)	[N70 amplitude + min{post-stimulus reversal potential, (50 – 90 ms)}]/2
P100 amplitude (O)	max{post-stimulus onset potential, (80 – 120 ms)}
P100 amplitude (OR)	[P100 amplitude + max{post-stimulus reversal potential, (80 – 120 ms)}]/2
N140 amplitude (O)	min{post-stimulus onset potential, (120 – 160 ms)}
N140 amplitude (OR)	[N140 amplitude + min{post-stimulus reversal potential, (120 – 160 ms)}]/2
P100 – N140 (O)	P100 amplitude (O) – N140 amplitude (O)
P100 – N140 (OR)	P100 amplitude (OR) – N140 amplitude (OR)
Latency Features	
N70 Latency (O)	Time of N70 amplitude (O) wrt stimulus onset
N70 Latency (OR)	[N70 latency (O) + Time of N70 amplitude (OR) wrt stimulus reversal]/2
P100 Latency (O)	Time of P100 amplitude (O) wrt stimulus onset
P100 Latency (OR)	[P100 latency (O) + Time of P100 amplitude (OR) wrt stimulus reversal]/2
N140 Latency (O)	Time of N140 amplitude (O) wrt stimulus onset
N140 Latency (OR)	[N140 latency (O) + Time of N140 amplitude (OR) wrt stimulus reversal]/2
RMS Features	
VEP RMS (O)	$\sqrt{\frac{1}{100}\sum_{i=1}^{100} x_i^2}$, x_i post-stimulus onset sample (50 – 250 ms)
VEP RMS (OR)	$\sqrt{\frac{1}{100}\sum_{i=1}^{100} x_i^2}$, x_i post-stimulus reversal sample (50 – 250 ms)
VEP RMS (Total)	$\sqrt{\frac{1}{500}\sum_{i=1}^{500} x_i^2}$, x_i post-stimulus onset sample (0 – 1000 ms)
fMRI modality	
Feature	**Definition**
HRF amplitude	max{post-stimulus percent signal change, (0 – 16.5 s)}
HRF latency	Time of 'HRF amplitude' wrt stimulus onset
HRF RMS	$\sqrt{\frac{1}{11}\sum_{i=1}^{11} x_i^2}$, x_i post-stimulus change sample (0 – 16.5 s)

Table 4.1 Overview and definition of single-trial signal features

The feature selection for the EEG modality was restricted to the time-domain for the following reasons: first, the experimental paradigm employed here has traditionally been associated with time-domain, i.e. event-related potential, analyses, owing to the traditional divide in the EEG research community between groups employing time-domain vs. frequency

domain analyses (Kiebel et al., 2008). Second, the goal of the current analyses was to demonstrate the applicability of the information theoretic framework to evoked responses and less an exhaustive feature search for which the experimental paradigm was not necessarily optimized (Rosa et al., 2010). Finally, the relationship between time-locked potentials and time-frequency representations of the same data remains elusive (Porcaro et al., in preparation), and was not intended to be a focus of the current study.

For the fMRI modality the analysis focussed on a set of similar features of the evoked single trial haemodynamic response functions, i.e., the maximum amplitude in BOLD percent signal change with respect to the pre-stimulus baseline (average of the three pre-stimulus onset TR values) and its latency with respect to stimulus onset in seconds was extracted. In addition the post-stimulus RMS as a measure of the overall magnitude of the evoked haemodynamic response was extracted. Table 4.1 lists a complete overview of the extracted features as well as their definition.

In summary, a relatively large, but by no means exhaustive, set of features was chosen to be assessed in the information theoretic analyses. Choosing a relatively large set of features increases the risk of false positive inferences, while at the same time avoids artificial specificity in the reported results. This should be considered when interpreting the information theoretic analyses reported below.

4.2.8 Information theoretic analysis

The same analysis scheme as described for the simulated data in Chapters 2 and 3 was employed here. In brief, after obtaining a data sample, the first step towards an information theoretic analysis is to compute the observed joint frequency distribution, denoted as $p_N(s, r_1, r_2)$, where the subscript N indicates the dependency on the number of samples. The observed joint frequency distribution is obtained by applying a grid of response value bins to

the observed data, counting the samples that fall into each of the response bins and then normalizing by the total number of response bins such that

$$\sum_{s\in S}\sum_{r_1\in R_1}\sum_{r_2\in R_2} p_N(s, r_1, r_2) = 1 \qquad (4.1)$$

where S indicates the (discrete) stimulus set, and R_1 and $R_2$the (discrete) response sets, of each variable, spanning the (discrete) response set $R = R_1 \times R_2$.

Free parameters in the application of the response bin grid are the total number of response bins and the upper and lower limit of the response bins in each dimension. To make efficient use of all sampled data points, the upper and lower limits of the response grid were set to the maximum and minimum value of each response variable. The number of response bins was chosen to be approximately equal to half the number of trials per stimulus N_S (Panzeri et al., 2007; Pola et al., 2003). Hence for each response dimension, the number of bins was set to $0.65 \cdot floor\left(\sqrt{N_S}\right) \approx 6$, resulting in 36 response bins for the bivariate response variable distributions. Specifically, the general recommendation from information theoretic analyses of electrophysiological data is to set the number of response bins of the histogram grid to half of the number of trials per stimulus, which would have been 42.5 bins in total. Assuming discrete response bins and the same number of response bins for each response variable, hence a choice had to be made between either 6 or 7 response bins for each univariate feature. The simulations reported in Chapter 3 indicated that under Gaussian model assumptions a slightly conservative choice of 6 response bins maximizes the sensitivity to information present in the data, while keeping the loss of bias control at bay, which was hence applied to the experimental data. As this choice was consistently applied to all information estimates for all EEG-fMRI feature combinations, it does not affect between-feature combination comparisons.

From the joint observed frequency distribution, the marginal distributions, $p_N(r_1, r_2)$, $p_N(r_1)$, and $p_N(r_2)$, and their corresponding stimulus conditional distributions, $p_N(r_1, r_2|s)$, $p_N(r_1|s)$ and $p_N(r_2|s)$, can be obtained by summation and division, respectively. Based on these observed frequency distributions, the information theoretic quantities introduced above and detailed in Chapter 2 can be computed by applying the respective formulae. These raw values are referred to as plug-in quantities (Panzeri et al., 2007). To correct for systematic biases the *bayescount.m* implementation of the PT-correction scheme (http://stefano.panzeri.googlepages.com/informationbiascorrections) was then used for the case of the univariate information theoretic quantities $I_N(S; R_1)$, $I_N(S; R_2)$, $I_N(R_1; R_2)$, and $\langle I_N(R_1; R_2|S)\rangle_S$. To correct for the systematic bias in the case of the bivariate information theoretic quantity $I_N(S; R_1, R_2)$ a shuffling approach implemented in the main analysis routine (1000 permutations) was employed as detailed in section 2.33.

4.3 Results

4.3.1 Behavioural data

Due to a technical malfunction of the button response box, reliable behavioural data could only be obtained from 9 observers (8 of whom remained in the final sample, see below). For these, the hit rate in the target (fixation cross change) detection task was high (0.95 ± 0.03 (SEM)) and the number of false alarms was low (3.1 ± 1.6 (SEM)) indicating good task performance, fixation and sustained attention.

4.3.2 Single subject and group average EEG and fMRI data

Five reliable data sets (sessions) were acquired for 12 of the 14 observers. For the remaining two observers only four reliable data sets could be acquired due to technical issues.

These two subjects were excluded from further analysis to optimise the within-subject variance homogeneity in the group analyses presented. To evaluate the data quality of the recorded EEG the average visual evoked potentials (VEPs) for low and high contrast stimuli were computed for each subject individually (Figure 4.2), as well as the grand average over subjects (Figure 4.4 A). Reliable P100 peaks could be identified in the reconstructed electrode data of at least one of the occipital electrodes in 10 of the 12 subjects (Figure 4.2). The subject-specific electrode which showed the maximal P100 peak amplitude was then chosen for subsequent single trial feature extraction and the information theoretic analysis.

For the fMRI data, average haemodynamic response functions (HRFs) were computed in terms of percentage signal change with respect to pre-stimulus baseline, extracted from 5 mm radius spheres, as described above. Reliable haemodynamic responses were identified in all subjects and the grand average calculated (Figures 4.3 and 4.4 B).

4.3.3 Information theoretic analyses

Using the theoretical framework introduced above, the aim was to obtain insights into the following questions: 1) Which of the EEG and fMRI time-domain features convey information about the stimulus and how do they compare (analysis of $I_N(S;R)$ for both modalities)? 2) How much information do pair-wise combinations of the time-domain features from both modalities convey about the stimulus (analysis of $I_N(S;R_1,R_2)$), and how does this compare to the information they convey individually (analysis of $Synergy$) 3) How much information do time-domain features of one modality convey about a feature from the other modality, both with respect to differential stimulation (analysis of $I_N(R_1;R_2)$) and in the absence of differential stimulation (analysis of $\langle I_N(R_1;R_2|S)\rangle_S$). 4) Finally, how do

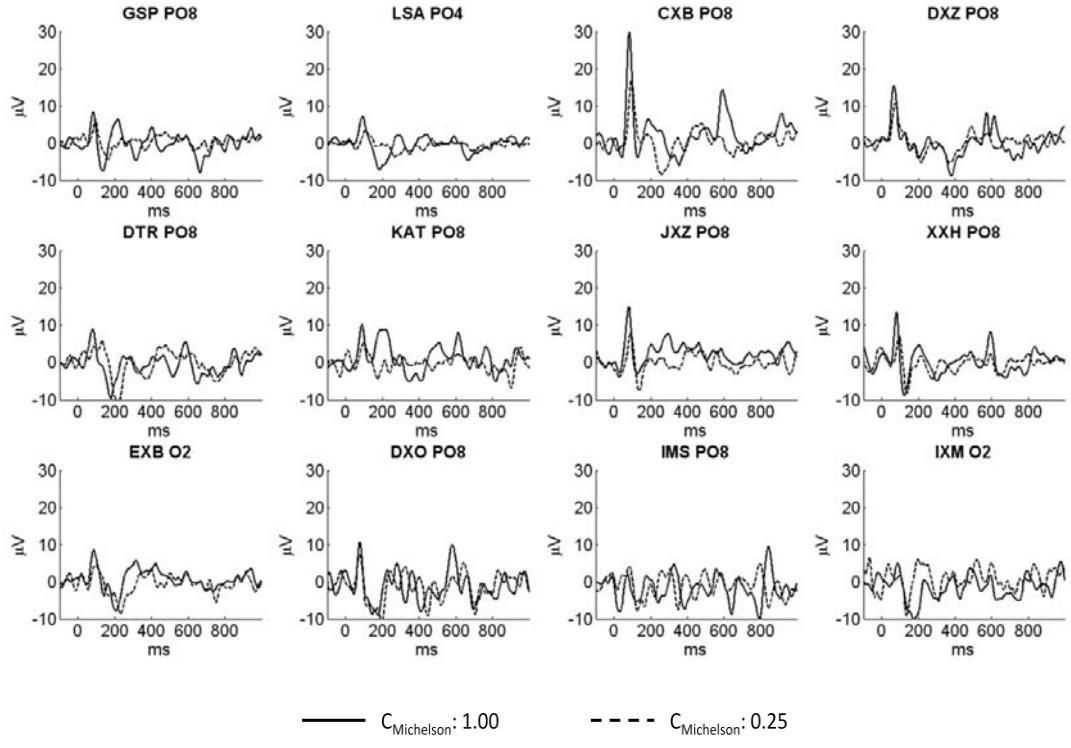


Figure 4.2 Single subject average EEG. Single subject averaged evoked signals for the EEG modality (evoked potentials) extracted from the subject specific maximum P100 amplitude electrode, as indicated in the titles of the subplots consisting of the three letter subject acronym and the electrode label.

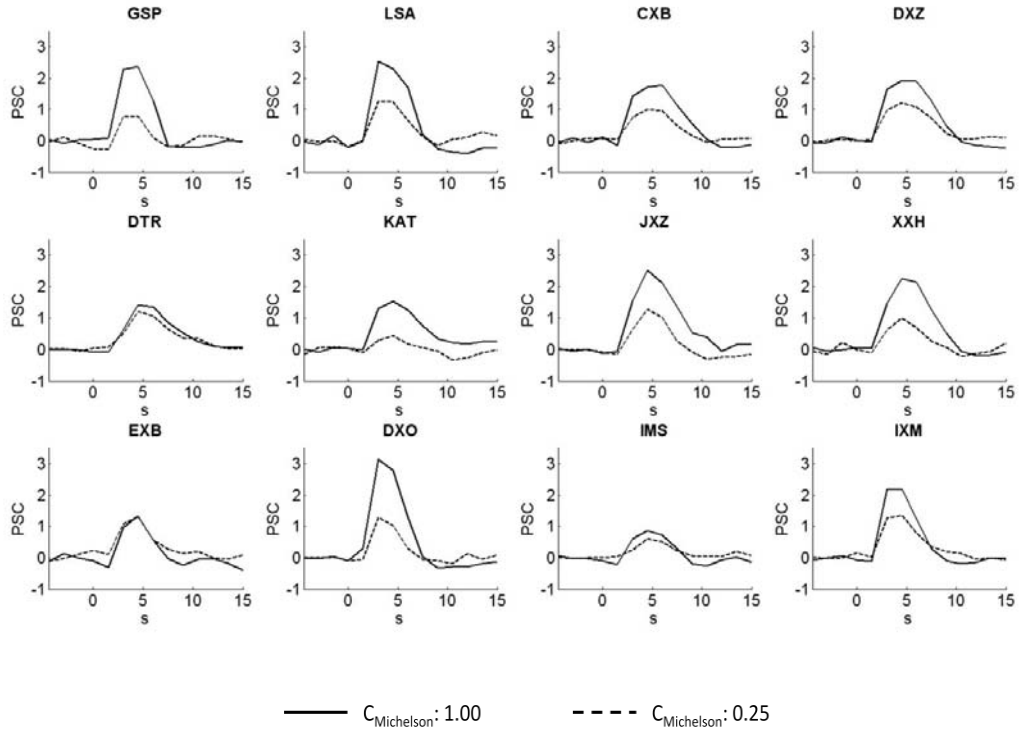


Figure 4.3 Single subject average heamodynamic response functions. The data were extracted from a 5 mm sphere surrounding the most significantly activated voxel. The ordering of the subjects is the same as in Figure 4.2, the titles are the three letter subject's acronyms.

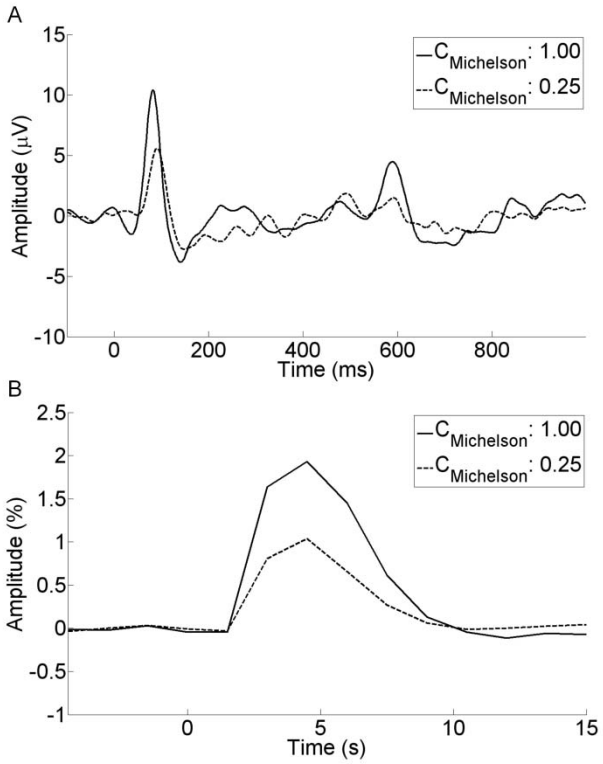


Figure 4.4 Visually evoked responses A) Grand average of visual evoked potentials for all subjects included in the main analyses (n = 12), for the low and high contrast conditions. The stimulus onset was at 0 ms, the stimulus reversal occurred at 500 ms. Data were extracted from the subject specific electrode of maximal P100 amplitude. B) Grand average of haemodynamic response functions of all subjects included in the main analyses (n = 12) for the low and high contrast condition. Data were extracted from a 5 mm sphere centred on the most activated significantly activated voxel across both stimuli.

different data pre-processing strategies affect the estimation of information theoretic quantities? With the exception of question 4) regarding the effect of pre-processing strategies, all information theoretic quantities were calculated using the data which had undergone advanced pre-processing.

Figures 4.5 – 4.9 summarise the results of these analyses averaged over observers (mean ± standard error). For each information theoretic quantity it was statistically evaluated whether it was significantly different from zero (one-sample t-test, two-tailed) and a number of statistical comparisons in the framework of the general linear model (repeated-measures

ANOVA with Huyhn-Feldt correction where appropriate) were performed unless stated otherwise.

A parametric statistical approach was chosen, as the analyses were performed on the between-subject level for which Gaussian assumptions are traditionally employed (Friston, 2007) and common tests for normality (e.g. Kolmogorov-Smirnov) have their own limitations (Hays, 1994). Further validation of the quasi-Gaussian nature of information estimates from Gaussian null-models is provided in Chapter 5, section 5.2.8. A two-tailed approach was chosen to keep the tests relatively conservative (compared to one-tailed comparisons) and to reflect uncertainty about the absolute estimates of information. Due to the exploratory nature of the study and its statistical evaluation as well as the likely dependencies between different information theoretic quantities (co-variation of different EEG and fMRI features prohibiting classical techniques such as Bonferroni correction), corrections for multiple comparison were not performed. In sum, the statistical results should be interpreted as descriptions of the data variance and not as unequivocal rejections of the respective null hypotheses. Future studies might evaluate the relative merits of parametric, non-parametric; bootstrapping and multiple comparison correction procedures for statistical inference on information estimates from combined EEG-fMRI data.

The results relating to each of the research questions are discussed below.

4.3.3.1 $I_N(S;R)$: stimulus-response characteristics

The results for the information that time-domain features convey about the stimulus in isolation are depicted in Figure 4.5. The first three panels depict EEG evoked potential features, grouped according to amplitude, latency and RMS. The final graph depicts the information properties of the fMRI features, namely the HRF amplitude, latency and RMS.

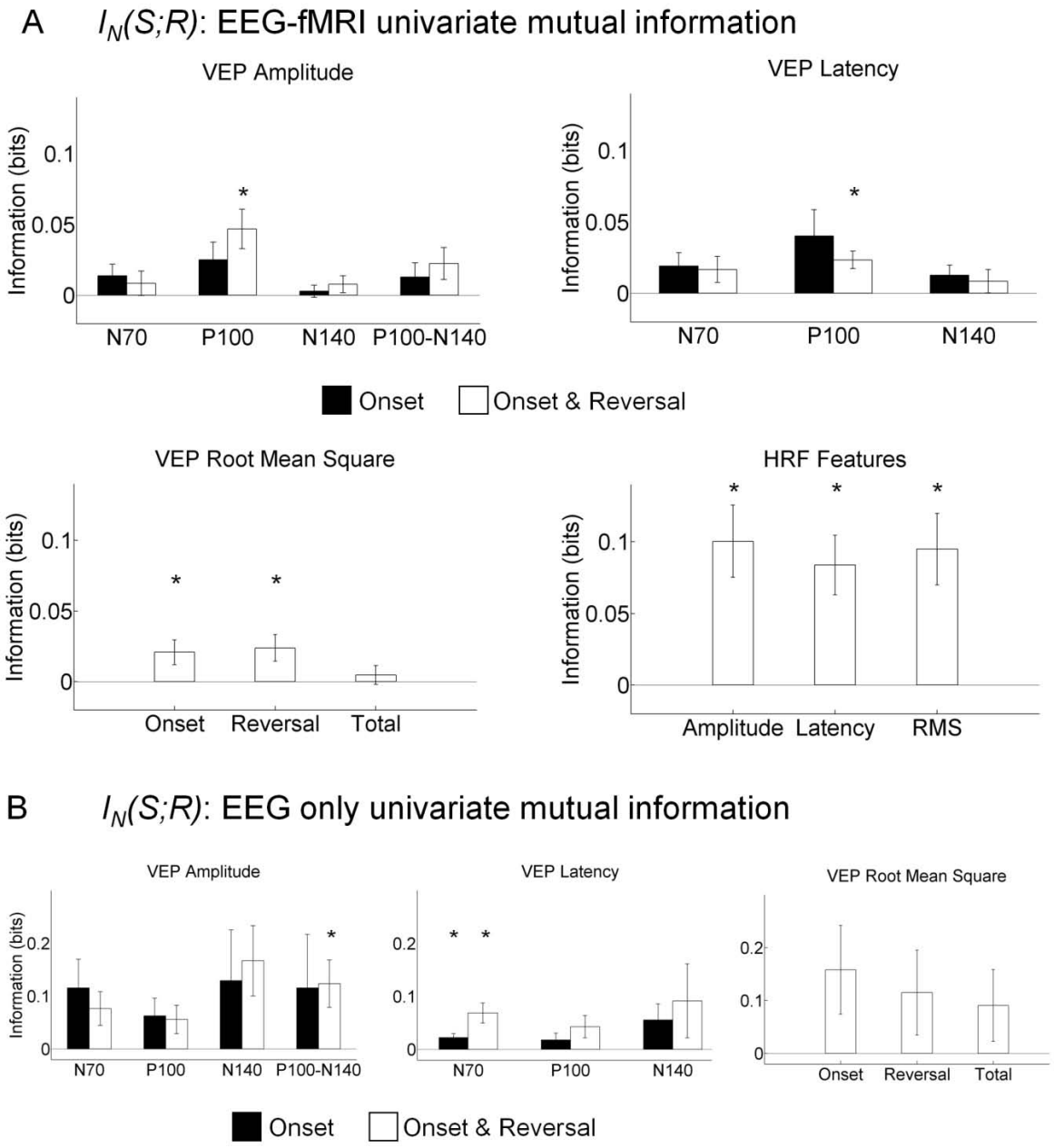


Figure 4.5 EEG-fMRI univariate mutual information. A) Information of EEG and fMRI time-domain features about the stimulus in isolation, averaged over observers ± SEM (n = 12). VEP: Visual evoked potential, HRF: Haemodynamic response function. Onset and Onset & Reversal refer to the EEG features averaged over stimulus onset or stimulus onset and reversal VEPs. Asterisks indicate means significantly larger than zero ($p < 0.05$, two-tailed one sample t-test). B): Information content of EEG time-domain features uncontaminated by the scanner artefact, averaged over observers ± SEM (n = 7). *VEP*: Visual evoked potential. Asterisks indicate means significantly larger than zero ($p < 0.05$, two-tailed one sample t-test). The data underlying this analysis were collected outside the MRI scanner just prior to the scanning session. 90 trials for each of the two conditions (low and high contrast left hemifield checkerboard, 2 Hz presentation rate) were collected with a mean interstimulus-interval of 1.5 seconds sampled uniformly from the interval 1 – 2 seconds.

Within the EEG modality four features conveyed reliable information about the stimulus, namely the onset-reversal P100 amplitude, the onset-reversal P100 latency, the onset and the reversal RMS. Two additional features, the onset P100 amplitude and the combined onset-reversal P100-N140 amplitude showed a trend for reliable information estimates. A one-way ANOVA comparing all EEG feature means revealed significant differences ($F_{(7.7,84.2)} = 2.11$, $p = 0.04$) between the features. Comparing the sub-features in each domain (amplitude, latency, RMS) showed significant differences for the amplitude features ($F_{(3.8, 42.2)} = 3.12$, $p = 0.03$), but not for the latency and RMS features ($F_{(5, 55)} = 1.72$, $p = 0.14$, $F_{(2, 22)} = 2.69$, $p = 0.09$), substantiating the higher information content of the onset-reversal P100 amplitude than the N70 and N140 amplitude features in the amplitude domain.

Within the fMRI modality, the information conveyed by all of the features was reliable and larger than for the EEG features. The differences between the information content of the fMRI modality features were not significant (one-way ANOVA: $F_{(2,22)} = 0.62$, $p = 0.55$).

To evaluate the reliability of the differences between the fMRI and EEG features a two-way ANOVA with factors 'modality' (EEG, fMRI) and 'feature' (amplitude, latency, RMS) was performed, for which the data was collapsed over EEG sub-features. This ANOVA indeed revealed a significant main effect for 'modality' ($F_{(1,11)} = 12.41$, $p < 0.01$), no significant effect for 'feature' ($F_{(2,22)} = 0.42$, $p = 0.66$) and no significant interaction between the two factors ($F_{(2,22)} = 0.67$, $p = 0.52$).

It is worth noting that the smaller stimulus related information values for the EEG time-domain features compared to the fMRI time domain features may be the consequence of the impoverished EEG data quality during simultaneous EEG-fMRI data acquisition and the necessary artefact subtraction methods. In fact, performing the same analysis on data

collected from a subset of the included observers outside of the scanner just before the main experiment, suggested that the information content of EEG time-domain features is at least equal, if not larger than that of fMRI time-domain features (for details, see Figure 4.5 B). However, it has to be noted that due to the variations in experimental paradigm and number of observers the information values for the EEG features between simultaneous EEG-fMRI acquisition and EEG only acquisition are not directly comparable in the current study. It appears likely that the general higher information content across all EEG features is due to the higher differential signal to noise ratio provided by EEG only recordings, of which the P100 is the most robust and recoverable from the simultaneous EEG-fMRI data acquisition. However, as shown in Figures 4.2 and 4.3 there is substantial between-subject variability in the quality of the evoked response inside the scanner, prohibiting reliable comparisons between outside and inside the MR environment EEG in the current study due to the different subject groups. Chapter 6 addresses this question in a more comprehensive manner, and the results indicate good correspondence in the group ERP between both EEG data acquisition schemes when evaluated over the same group of subjects.

4.3.3.2 $I_N(S; R_1, R_2)$ and $Synergy$: feature combinations across modalities

An important question in the integration of simultaneously acquired EEG-fMRI data is whether the joint observation of both variables is in some way more (or less) informative than the observation of each variable in isolation. By estimating $I_N(S; R_1, R_2)$ and $Synergy$ for the pairwise combination of EEG and fMRI time domain features, it was attempted to gain some insight into this question. Figure 4.6 displays the results for $Synergy$ and all possible combinations of the time-domain features of interests. The data are grouped with different EEG feature domains (amplitude, latency, RMS) as rows, and different fMRI feature domains

(amplitude, latency, RMS) as columns. All information values are significantly larger than zero, and appear slightly larger for the combination of EEG amplitude and RMS features with fMRI features than for the combination of EEG latency features.

To assess the reliability of differences between the feature pairings with respect to their information content, two statistical analyses were conducted: first, a two-way ANOVA with factors 'EEG feature' (amplitude, latency, RMS) and 'fMRI feature' (amplitude, latency, RMS), where the data was collapsed over the respective EEG sub-features, and secondly a one-way ANOVA for each of the feature pairing sub-groups. The first of these comparisons yielded no significant main effect for 'EEG feature' ($F_{(2,22)} = 2.53$ $p = 0.10$) and no significant main effect of 'fMRI feature' ($F_{(2,22)} = 1.17$, $p = 0.33$). However, a significant interaction between factors ($F_{(3.5,39.0)} = 3.73$, $p = 0.01$) was observed, indicating a dependency of the information value on the specific EEG-fMRI feature pairing. Within the subgroups, a significant main effect was detected in the pairings of EEG amplitude and fMRI latency ($F_{(7.1, 77.9)} = 2.93$, $p < 0.01$).

Comparing the results of $I_N(S; R_1, R_2)$ to those of the information content of the modality features in isolation ($I_N(S; R)$, Figure 4.5) indicates higher information content in the joint observed frequency distribution. Specifically, the question arises whether the joint distribution of two modality features conveys more or less than the sum of each modality feature's marginal distribution. To this end, the evaluation of the normalised $Synergy$ of the two features is of interest. As the information content of each marginal distribution is always equal to or less than the information content of the respective joint distribution (i.e. $I_N(S; R_1) \leq I_N(S; R_1, R_2)$ and $I_N(S; R_2) \leq I_N(S; R_1, R_2)$) this ranges between –1 and 1.

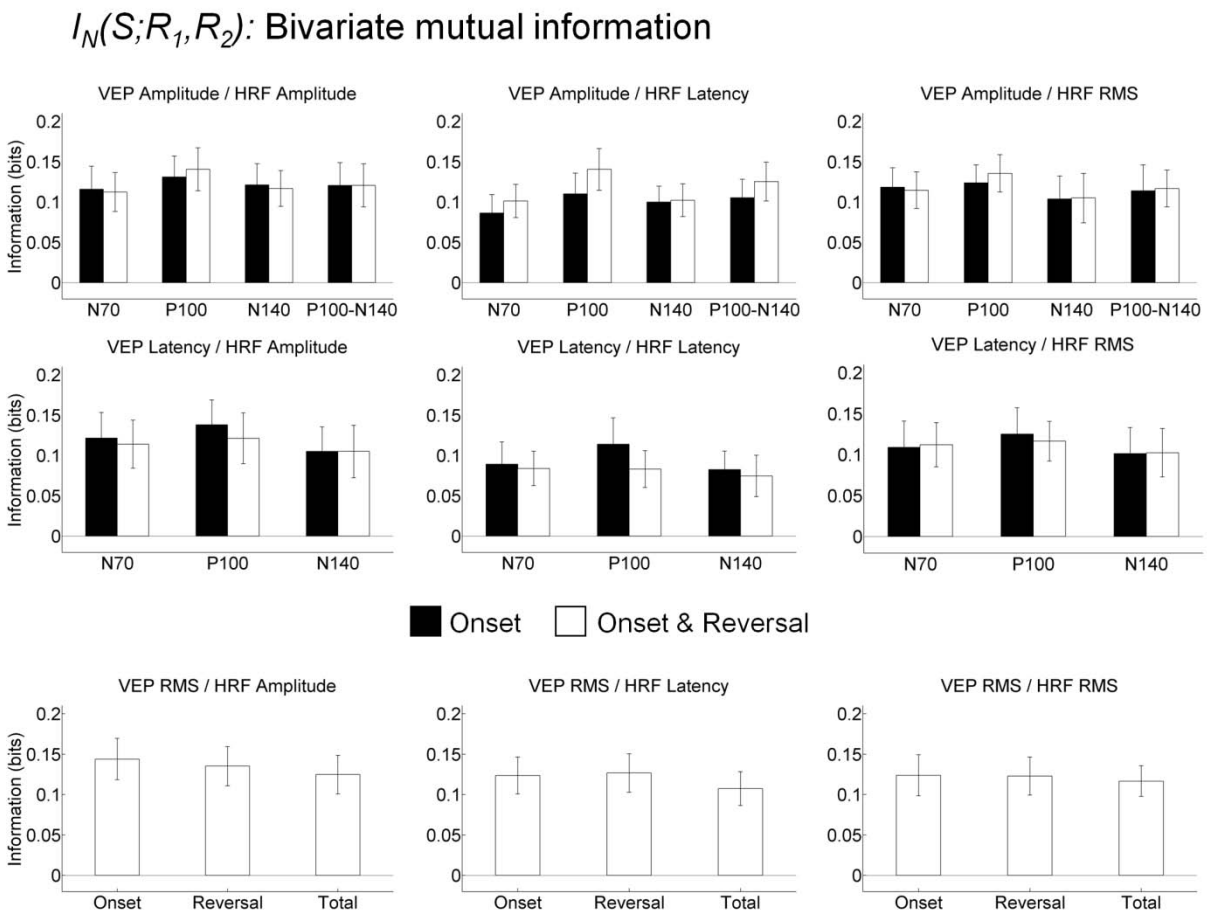


Figure 4.6 Bivariate mutual information**.** Information of combined EEG and fMRI time-domain features about the stimulus, averaged over observers ± SEM (n = 12). VEP: Visual evoked potential, HRF: Haemodynamic response function. Onset and Onset & Reversal refer to the EEG features averaged over stimulus onset or stimulus onset and reversal. All means were significantly different from zero ($p < 0.05$, two-tailed one sample t-test).

In one extreme case the response variable marginal distributions contain no information about the stimulus, and only the joint observation of both variables is informative, in which case the response variables are said to be fully synergistic (i.e. $I_N(S;R_1) = I_N(S;R_2) = 0 \Leftrightarrow Synergy = 1$).

In the other extreme case each marginal distribution contains all information present in the joint distribution, and the variables are said to be fully redundant (i.e. $I_N(S;R_1) =$

$I_N(S;R_2) = I(S;R_1,R_2)0 \Leftrightarrow Synergy = -1$). The normalised $Synergy$ for all feature pairings is shown in Figure 4.7.

The data are again grouped with different EEG features domains (amplitude, latency, RMS) as rows, and different fMRI feature domains (amplitude, latency, RMS) as columns. The normalised $Synergy$ measure is very sensitive to estimation errors (e.g. for information values close to zero, or negative values due to non-optimal bias correction). For this reason, in order to avoid artefactual $Synergy$ values resulting from the inclusion of noisy information values, it was computed only from those observers for whom the respective values of $I_N(S;R_1)$ and $I_N(S;R_2)$ were larger than 0.01 and $I_N(S;R_1,R_2)$ larger than 0.1, reflecting roughly the average values of observed in Figures 4.5 and 4.6. As this resulted in different degrees of freedom and the pairing of different observers across quantities, no explicit statistical comparisons were conducted in this case. Inspecting the data shows that the $Synergy$ varies widely between different feature pairings. With approximately the same number of negative and positive results, the overall pattern of results indicates that for the current experimental paradigm, the response-signal marginal distributions contain the majority of the stimulus-related information and the observed joint distributions do not add much to this. This pattern of results indicates no relevant $Synergy$ between EEG and fMRI features in the current experimental context. The most reliable synergistic effects were observed for the pairings of VEP RMS and HRF latency features, while the most reliable redundancy effects appear for pairings of VEP latency and HRF latency/RMS features.

4.3.3.3 $I_N(R_1;R_2)$ and $\langle I_N(R_1;R_2|S)\rangle_S$: inter-modality comparisons

Next, it was asked how much information time-domain features of each modality convey about each other. Figure 4.8 depicts the mutual information $I_N(R_1;R_2)$ for all possible pairwise combinations of EEG and fMRI features.

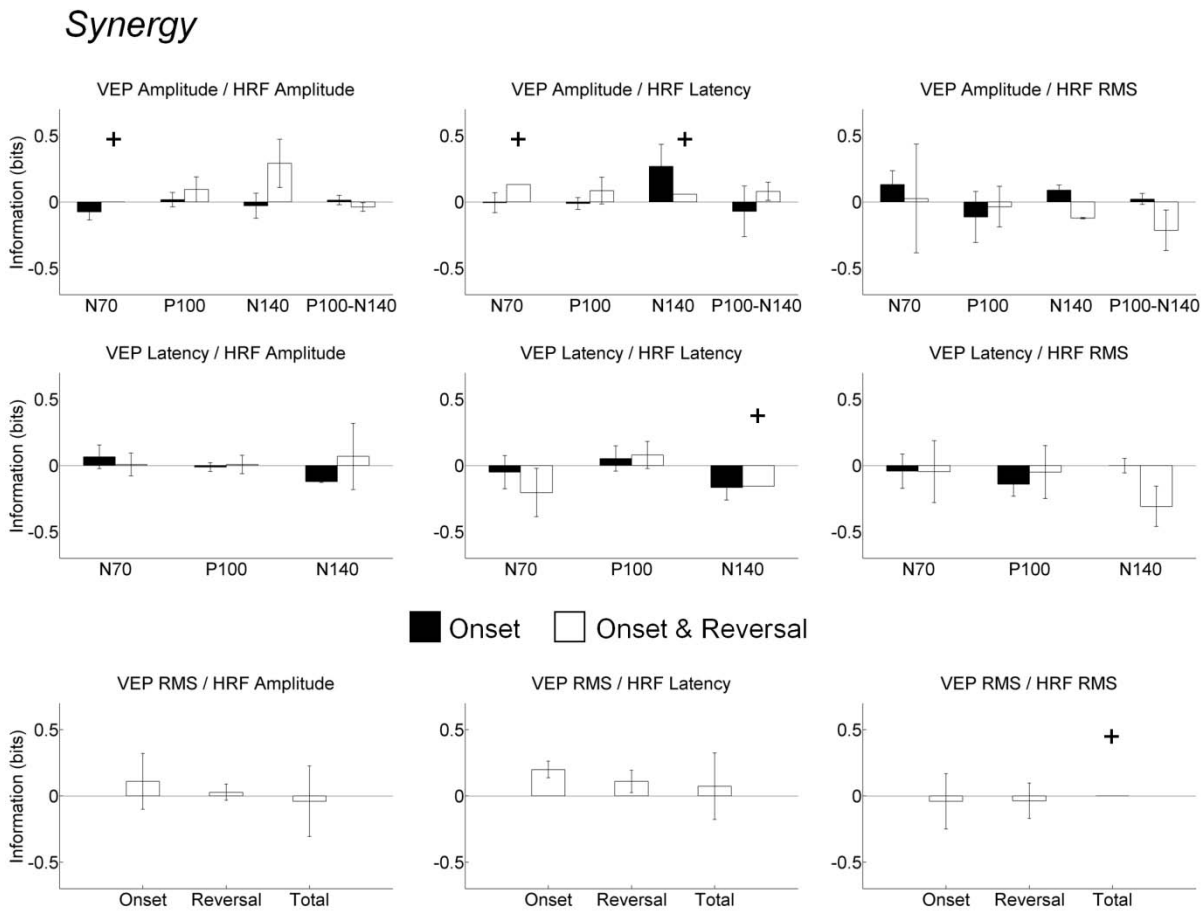


Figure 4.7 Synergy. *Synergy* of combined EEG and fMRI time-domain features about the stimulus, averaged over observers ± SEM. VEP: Visual evoked potential, HRF: Haemodynamic response function. Onset and Onset & Reversal refer to the EEG features averaged over stimulus onset or stimulus onset and reversal. Crosses indicate feature pairings in which only data from one or zero observers survived the data selection criteria discussed in the text.

The first column depicts all pairings of EEG features with HRF amplitude, the second with HRF latency, and the last with HRF RMS. For all feature pairings, the mutual information is small, and only in a few cases significantly different from zero. Specifically, only the combined P100 onset/reversal latency and HRF latency, as well as the VEP onset/ VEP total RMS and HRF RMS showed reliable activity dependence.

The differences between EEG feature domains and fMRI features were evaluated using a two-way repeated measures ANOVA with factors 'EEG feature' (amplitude, latency, RMS) and 'fMRI feature' (amplitude, latency, RMS), collapsing the data over EEG sub-features. This ANOVA yielded no significant main effect of 'EEG feature' ($F_{(2,22)} = 0.30$, p =

0.74), a significant effect of 'fMRI feature' ($F_{(2,22)}$ = 5.20, p = 0.01) and no significant interaction ($F_{(4,44)}$ = 1.17, p = 0.33).

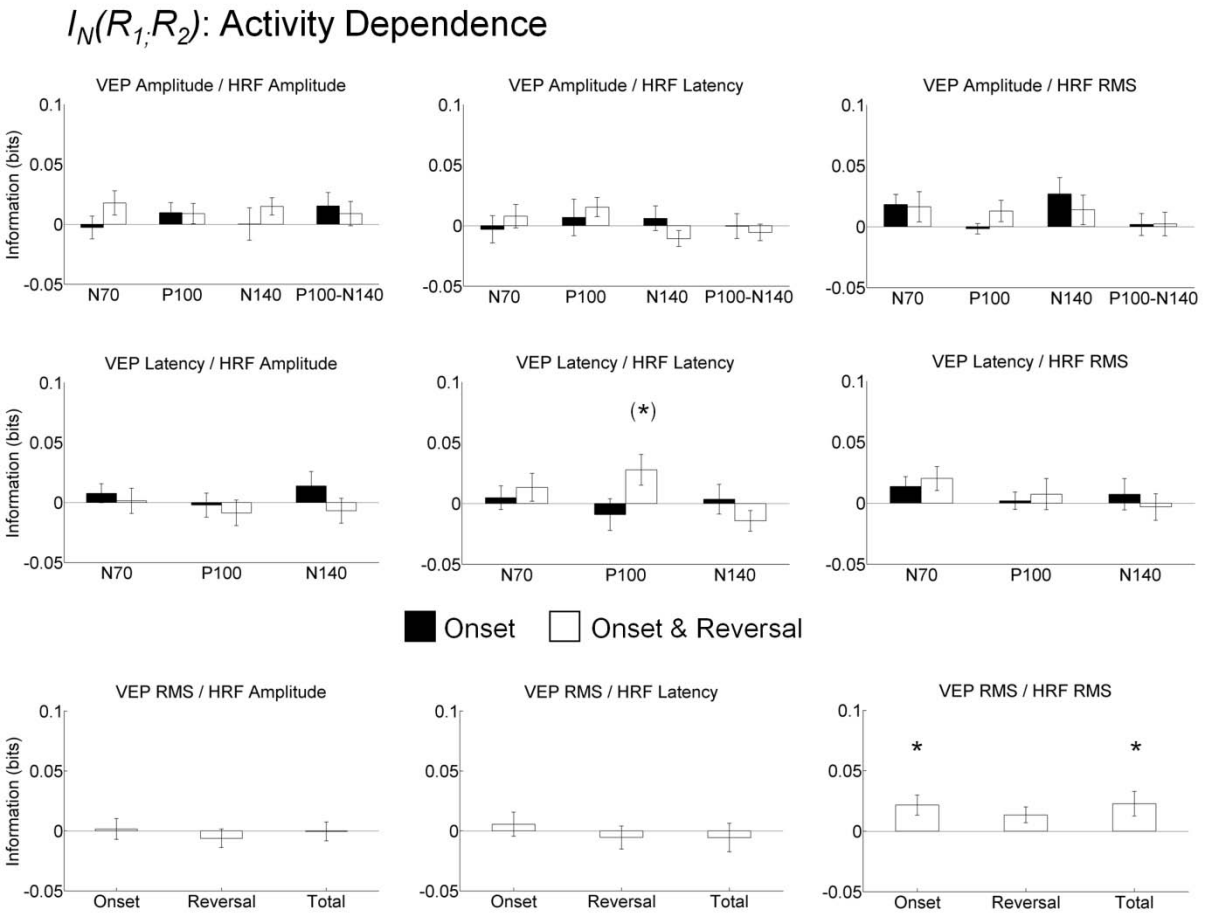


Figure 4.8 Activity dependence**.** Activity dependence of features of EEG and fMRI, averaged over observers ± SEM (n = 12). VEP: visual evoked potential, HRF: Haemodynamic response function. Onset and Onset & Reversal refer to the EEG features averaged over stimulus onset or stimulus onset and reversal. Asterisks indicate means significantly larger than zero (p < 0.05, two-tailed one sample t-test), $^{(*)}$ indicates p = 0.05.

Further, no significant differences between different EEG features in each of the EEG feature domains were detected. Hence, rejecting the null hypothesis that the different EEG and fMRI features are equally informative about each other failed. In sum, the observed mutual information values are low and mostly not significantly different from zero. They do not therefore indicate a clear functional relationship between the EEG and fMRI features, conditional on the specificities of the experimental paradigm.

The mutual information $I_N(R_1;R_2)$ between two time domain features of each modality was determined in the presence of differential stimulation. In principle it is possible that the probabilistic dependence that exists between two features is accounted for by this differential stimulation alone, or, in the other extreme, totally independent of differential stimulation, e.g. due to another source of co-variation (also called noise-correlation). To disentangle these possibilities, the (stimulus-averaged) stimulus conditional mutual information $\langle I_N(R_1;R_2|S)\rangle_S$ can be computed and compared to $I_N(R_1;R_2)$. However, from a practical perspective, it has to be noted that the number of trials available for the estimation of the stimulus conditional observed frequency distributions reduces proportionally to the inverse of the stimulus set cardinality, and hence is likely to be more error prone than the other measures discussed here. In both the current and the next chapter it appears that the conditional dependence is positively biased, regardless of the application of the bias correction methods discussed in Chapter 2. The quality of estimating this quantity clearly indicates strong potential for improvement.

The stimulus conditional mutual information between fMRI and EEG features for all possible comparisons was evaluated (Figure 4.9). Evaluation of differences between the feature pairings in a two-way ANOVA with factors 'EEG feature' (amplitude, latency, RMS, collapsed over subfeatures) and 'fMRI feature' (amplitude, latency, RMS) yielded a significant main effect of 'EEG feature' ($F_{(1.81,19.9)} = 12.79$, $p < 0.01$) indicating the larger variability of $\langle I_N(R_1;R_2|S)\rangle_S$ values compared to $I_N(R_1;R_2)$, where this was not the case. Further, no significant main effect of 'fMRI feature' ($F_{(2,22)} = 0.82$, $p = 0.45$), and no significant interaction ($F_{(4,44)} = 1.26$, $p = 0.30$) were detected. Within each EEG feature domain, no significant differences were detected.

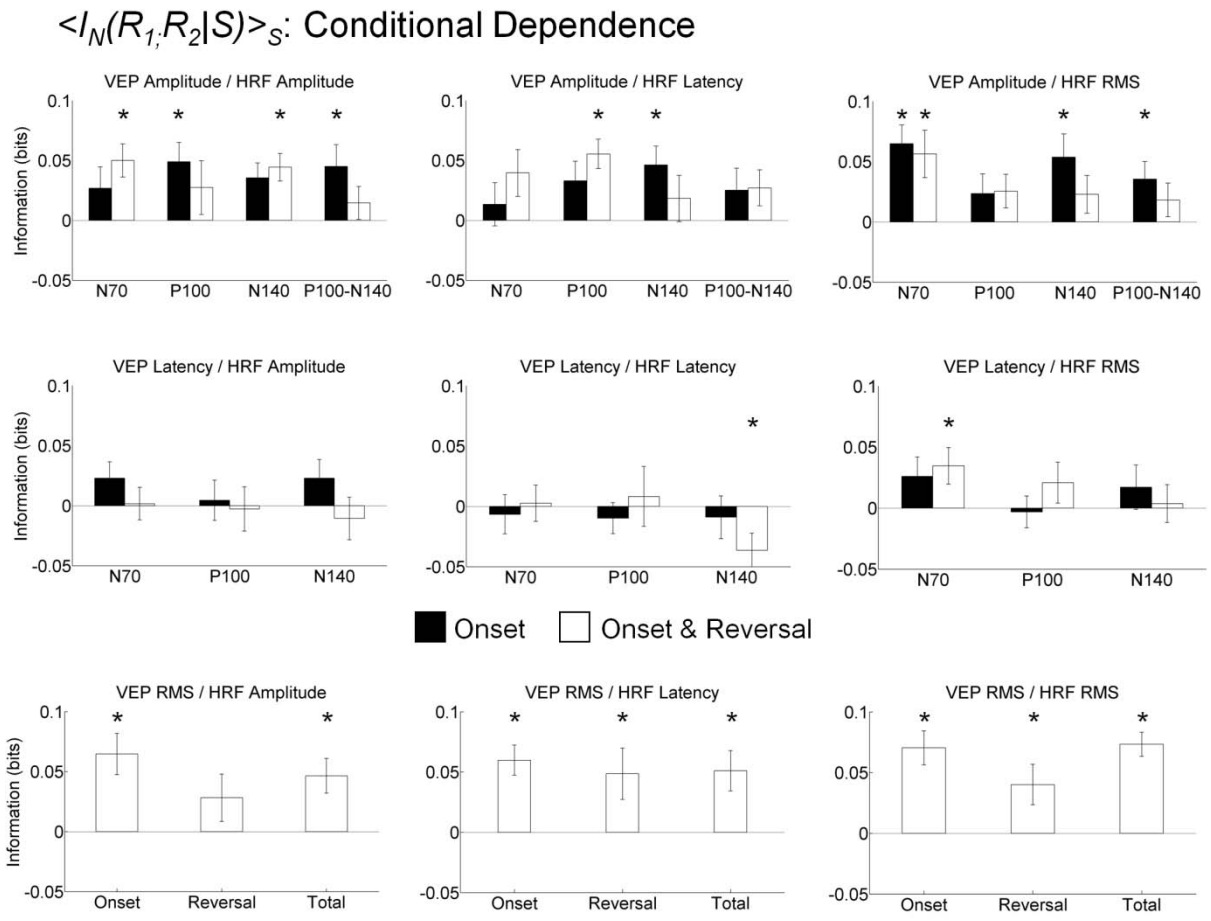


Figure 4.9 Conditional dependence. Stimulus conditional dependence of EEG and fMRI features, averaged over observers ± SEM (n = 12). VEP: visual evoked potential, HRF: Haemodynamic response function. Onset and Onset & Reversal refer to the EEG features averaged over stimulus onset or stimulus onset and reversal. Asterisks indicate means significantly larger than zero ($p < 0.05$, two-tailed one sample t-test).

To compare $I_N(R_1; R_2)$ and $\langle I_N(R_1; R_2|S)\rangle_S$, a three-way repeated measures ANOVA with factors 'information theoretic quantity' ($I_N(R_1; R_2), \langle I_N(R_1; R_2|S)\rangle_S$), 'EEG feature' (amplitude, latency, RMS, collapsed over subfeatures) and 'fMRI feature' (amplitude, latency, RMS) was carried out. This ANOVA revealed a significant main effect of information theoretic quantity ($F_{(1,11)} = 27.5$ $p < 0.01$), indicating some overall differences between $I_N(R_1; R_2)$ and $\langle I_N(R_1; R_2|S)\rangle_S$. For a more detailed assessment of the differences between the unconditional and stimulus conditional mutual information, the analysis was focussed only on those feature pairings for which a significantly non-zero unconditional mutual information was detected. The results of this assessment (paired sample t-tests) are

given in Table 4.2. This analysis shows that the differences between the estimated values of $I_N(R_1;R_2)$ and $\langle I_N(R_1;R_2|S)\rangle_S$ are significant for the RMS features. However, in no case is the activity dependence significantly larger than the conditional dependence indicating, if any, stimulus independent response signal co-variation.

To summarize, the observed mutual information between EEG and fMRI time domain features was quite low and mostly not significantly different from zero, with a few exceptions. The information about the stimulus in the joint distribution of EEG and fMRI features is hence not translated into activity dependence. This is due to the fact that in the current application, the marginal distribution of the fMRI feature is dominating the discriminability. The conditional dependence estimates appear positively biased, indicating the need for improved information estimation procedures for neuroimaging data.

4.3.3.4 Influence of data pre-processing methods

As a final analysis, it was aimed to quantify the effect of the data pre-processing methods on the estimated information theoretic quantities. To this end, the observed information theoretic quantities for the most informative EEG and fMRI time domain features in isolation, the onset-reversal P100 and HRF amplitude for both advanced and basic feature extraction strategies (see Methods), were computed. The results of this analysis are shown in Figure 4.10. Clearly, the advanced pre-processing methods lead to an increase in the estimated information values, which is most evident for those information theoretic quantities that refer to the stimulus. Since both $I_N(S;R_1)$ and $I_N(S;R_1,R_2)$ are fairly low and unreliable for the basic processing strategies, $Synergy$ was not included in this assessment. To assess the reliability of the observed results, a two-way repeated measures ANOVA with factors 'method' (ICA and GLM, raw data and literature V1 coordinates) and 'information theoretic quantity' was carried out.

Feature Combination	Difference	T-Value	DF	p-Value
VEP Amplitude / HRF Amplitude (VEP: Onset N1 – P1)	0.008	0.51	9	0.62
VEP Amplitude / HRF Latency (VEP: Onset and reversal N70)	-0.001	-0.07	9	0.95
VEP Amplitude / HRF RMS (VEP: Onset N70)	-0.014	-0.50	9	0.63
VEP Amplitude / HRF RMS (VEP: Onset N140)	-0.023	-0.98	9	0.36
VEP Latency / HRF RMS (VEP: Onset N140)	0.014	0.65	9	0.53
VEP RMS / HRF Amplitude (VEP: Reversal RMS)	0.009	0.56	9	0.59
VEP RMS / HRF RMS (VEP: Reversal RMS)	-0.007	-0.49	9	0.63
VEP RMS / HRF RMS (VEP: Total RMS)	-0.013	-0.59	9	0.57

Table 4.2 Statistical evaluation (paired sample t-tests, two-tailed) of the differences between activity and conditional dependence for those EEG-fMRI feature pairings showing a significantly non-zero activity dependence

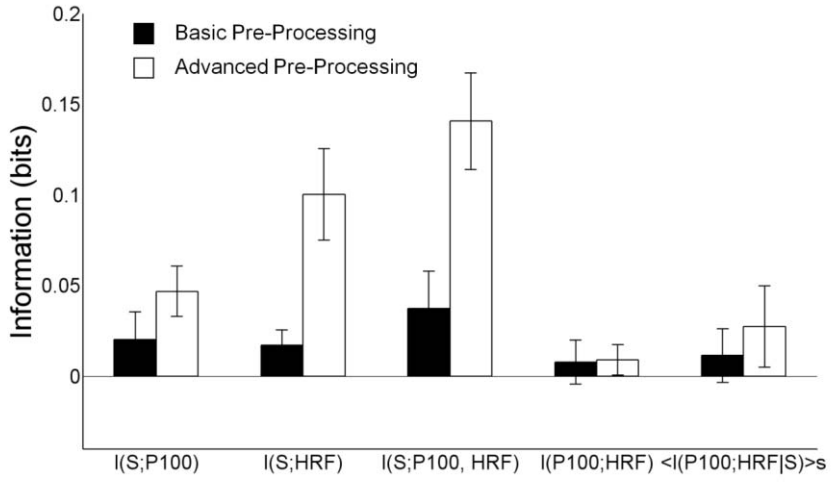


Figure 4.10 Influence of data pre-processing methods. Influence of data pre-processing methods on the estimated information theoretic quantities for the EEG signal feature 'onset-reversal averaged P100 amplitude' (P100) and the fMRI signal feature 'haemodynamic response function amplitude', averaged over observers (n = 12) ± SEM.

This ANOVA revealed a significant main effect of 'method' ($F_{(1,11)} = 17.8$, $p < 0.01$), a significant main effect of 'information theoretic quantity' ($F_{(2.2,\ 24.0)} = 5.70$, $p < 0.01$), as well as a significant interaction between both factors ($F_{(2.19,\ 24.1)} = 4.88$, $p = 0.02$), substantiating the larger information values observed for the advanced pre-processing strategies.

4.4 Discussion

Practically applying the information theoretic framework to a set of EEG-fMRI single trial time-domain features obtained during checkerboard stimulation has provided several interesting results.

First, a subset of the time-domain features chosen was found to be informative with respect to the stimulus in isolation. For the EEG data, the most informative measures were related to the amplitude and latency of the P100 component, which has previously been shown to be reliably modulated by stimulus contrast (Odom et al., 2004; Shawkat and Kriss, 2000). Similarly, the information theoretic framework also identified the effect of stimulus contrast on the haemodynamic response (Bagshaw and Warbrick, 2007; Boynton et al., 1999; Wan et al., 2006). The fact that the methodology proposed here is able to correctly identify the primary features that are affected by stimulus contrast is a useful validation. It is of note that fMRI features were generally more informative than EEG features for simultaneously acquired data. However, as discussed above, this is probably due to the diminished data quality of artefact affected EEG data in the concurrent acquisition of both modalities. The comparison of data acquired inside and outside the scanner suggests that there is some loss of stimulus related information specifically for the lower amplitude peaks of the evoked potential, N70 and N140. These effects are likely to be reduced as the quality of concurrently

recorded EEG data improves. As demonstrated by the comparison of basic and advanced pre-processing strategies, the information theoretic framework provides a quantitative methodology to assess the effect of different artefact removal techniques.

Second, combining EEG and fMRI time-domain features by quantifying the information in their joint distribution was more informative than treating each one in isolation. However, the overall pattern of the $Synergy$ results remains inconclusive, as both synergistic ($Synergy > 0$) and redundant $Synergy < 0$ effects were observed. Further, the absolute $Synergy$ results were relatively small (mainly < 0.4) , implying that most of the information about the stimulus is actually conveyed by the marginal response distributions, rather than their joint distribution. The fact that the exact relation between neuronal activity, surface-electrode EEG phenomena and BOLD signal changes remains debatable complicates the interpretation of the obtained results. For example, if one assumes that EEG and fMRI phenomena represent the same underlying neuronal activity, only redundancy should be observed. The fact that this is not the case can be argued biologically, i.e. that there might be differences in the neuronal activity that the modality features reflect, and only the joint observation from both modalities yields an estimation of the information that is conveyed at the neuronal level. On the other hand, assuming that EEG and fMRI measures reflect the same underlying neuronal information content, the fact that the estimation of $I_N(S; R_1, R_2)$ was amended using a Gaussian null model and observe synergistic effects may indicate that the bias correction procedure chosen is not optimal. Clearly, a better understanding of neuronal information content and its relation to information estimated from non-invasive brain imaging techniques is necessary for conclusive interpretation of synergistic effects.

Third, it was found that the mutual information between time-domain features of the two modalities is relatively low. This finding is in line with previous findings on the

correlation of single-trial time-domain EEG and fMRI data (Bagshaw and Warbrick, 2007) and posits some interesting questions for the 'integration through prediction' approach to EEG-fMRI integration, which capitalizes on the linear co-variation of EEG and fMRI signal features. Possibly, the prediction of the fMRI time-series could be improved if a set of more informative EEG features than the one studied here and used previously to predict the fMRI time-series (Debener et al., 2006; Eichele et al., 2005) could be identified. The current experimental evidence points to conditional dependencies (i.e. co-variation at constant stimulus) between EEG/LFP frequencies features in the α- (e.g. de Munck et al. 2007), β- (Singh et al., 2002), and γ- (e.g Niessing et al., 2005) range, or all of them (e.g. Kilner et al., 2005; de Munck et al., 2009)) and the BOLD response. Potentially, stimulus-induced frequency perturbations of the EEG might exhibit co-variation with evoked BOLD response features. The combination of an appropriate experimental paradigm and the information theoretic approach described here could further address these issues.

Comparing the overall pattern of results obtained from the experimental data to the linear Gaussian model simulations of Chapter 3 indicates the following: first, the experimentally observed mutual information values do not assume their theoretically possible maximum values, reflecting the low signal-to-noise ratio of the data. In terms of the simulations displayed in Figure 3.7, this would correspond to low ratios between values of the 'signal gain factors' a_{R1}/a_{R2} and the 'noise factors' $\sigma_{R1}^2/\sigma_{R2}^2$. Furthermore, the experimentally observed pattern of stimulus related information in both the response marginal distributions and the response joint distribution, no prominent $Synergy$, and relatively low activity and conditional independence is most reminiscent of the low signal-to-noise ratio evaluations of model M1 in the parallel a_{R1}/a_{R2} simulation. Arguing indirectly, this would support a view in which both EEG and fMRI signal features contain stimulus related information, but show little additional single-trial co-variation. It should however be noted

that this interpretation assumes that the EEG and fMRI feature distributions are actually Gaussian and is of course confined to the current experimental paradigm. Also, comparing the information signatures of observed data with the ones created artificially based on very simple Gaussian models is an indirect approach to model evaluation at best. In the context of the current study, the main purpose of the simulations was to constrain the analysis of experimental data, and to demonstrate the potential power of the framework. Future work based on generative forward models for EEG and fMRI and their inversion based on machine learning based approaches appear more promising for a direct assessment of the underlying nature of stimulus-dependent and independent neural activity, and resulting EEG and fMRI signal changes (Valdes-Sosa et al., 2008).

It was shown that the estimated quantities depend on the experimental settings (as shown for the EEG features extracted from data recorded inside and outside MRI scanner) as well as the data pre-processing techniques. As this is quite natural for empirical results, this is not regarded as an drawback of the framework, but rather as a potential application of the methods proposed: the estimated quantities allow a meaningful evaluation of the quality of an experimental setting and hence can be used directly for the optimization of experimental techniques and designs.

In summary, the current chapter demonstrated a reproduction of the contrast sensitivity of spatially selected EEG and fMRI time domain features from an information theoretic perspective. The information about the stimulus in the joint distribution of EEG and fMRI features was found to be approximately equal to the sum of the information in the respective marginal distributions. It was observed that the practical application of the theoretical framework reviewed in Chapter 2 can lead to negative information estimates, which from a theoretical standpoint are not meaningful and complicate the estimation of information dependence. Taking into account the difficulties in estimating conditional

dependencies as discussed above, an important conclusion from the current study is that with the currently available methods, comparisons of information quantities are most meaningful across different features within a given information theoretic quantity, rather than between different information theoretic quantities. This will be a recurrent theme in the two following experimental chapters.

5 Voxel-wise information theoretic EEG-fMRI feature integration

5.1 Introduction

In Chapters 1 - 4 the evaluation of a set of information theoretic quantities (ITQs) for the integration of simultaneously acquired electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data has been proposed. For the experimental evaluation of the information theoretic framework in Chapter 4, the spatial data subsets from which to calculate the ITQs were defined using relatively arbitrary a priori constraints. In the case of EEG, this meant that data were extracted from a single electrode showing the maximal visual evoked potential (VEP), while for fMRI the analysed data came from voxels contained within a sphere surrounding the most responsive voxel of visual cortex. While this approach was a natural starting point for the evaluation of the framework in the application to combined EEG-fMRI data sets, a more principled approach to spatial data selection is desirable. In this Chapter, it is proposed to combine standard fMRI data pre-processing and low-resolution electromagnetic tomography (LORETA, (Frei et al., 2001; Pascual-Marqui, 2002a; Pascual-Marqui, 2002b)) for the evaluation of information theoretic quantities across the entire three-dimensional brain space.

A full three-dimensional evaluation of the set of ITQs requires single-trial estimates of EEG and fMRI signal features at each location (voxel) of the brain (Figure 5.1). For fMRI, these estimates can be obtained using standard data pre-processing including normalization to a standard anatomical space, as the data is inherently three-dimensional. For the EEG, however, a data transform from the electrode recordings at the scalp surface to normalized anatomical three-dimensional brain space is necessary. Several distributed inverse solutions, which can estimate electrical activity features at a pre-defined three-dimensional grid of anatomical locations, exist (see (Michel et al., 2004) for an introduction). LORETA is one such inverse solution, which has been relatively widely used with EEG data in the time-

domain, especially in its modification as sLORETA (Pascual-Marqui, 2002b), but also in the frequency domain (Frei et al., 2001; Pascual-Marqui, 2002a). Using this inverse solution, it hence becomes possible to estimate single-trial electrophysiological features for each location of the three-dimensional brain space. Finally, given that data types are projected into the same and aligned standardized anatomical spaces (e.g. MNI space), the estimation of information theoretic measures can proceed based on the data from corresponding locations.

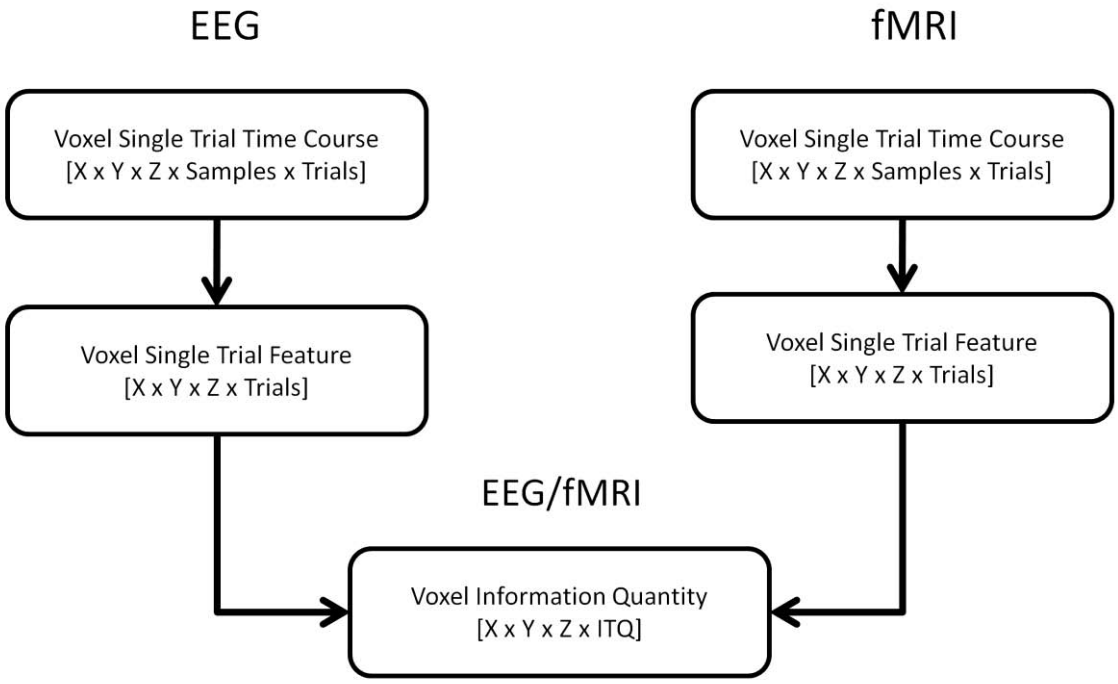


Figure 5.1 Conceptual framework of three-dimensional information theoretic EEG-fMRI feature integration. A necessary prerequisite for the proposed scheme is a voxel single-trial time course for each modality, where voxels from both modalities can be brought into spatial correspondence. In the case of the fMRI modality, the data matches this format upon reconstruction by the MRI scanner console, while for the EEG modality, the sensor readings have to be projected into three-dimensional space based on a solution for the EEG inverse problem. Individually for both modalities, time-domain features are extracted from the time course of each trial and voxel. Upon feature extraction, the features from both modalities at corresponding locations enter the information theoretic analysis. The bracketed expressions represent the underlying numerical arrays, where *X* , *Y* and *Z* indicate the number of voxels in x-,y-, and z-dimension of the data, *Samples* indicates the number of data samples of a single trial of the experiment, and *Trials* represents the number of individual trials across conditions. Last *ITQ* represents the number of information theoretic quantities evaluated.

The aim of this chapter is three-fold: First, to demonstrate the proposed approach to three-dimensional information EEG-fMRI feature integration using data collected in a visual stimulation (checkerboard) experiment as described in Chapter 4. Second, to extend the range of EEG features under consideration in the information theoretic framework to the frequency

domain, reflecting a large body of work that exists on the co-variation of EEG frequency domain features (mainly in the α-frequency range) and the BOLD signal (de Munck et al., 2007b; de Munck et al., 2009; Feige et al., 2005; Goldman et al., 2002; Goncalves et al., 2006; Laufs et al., 2003; Laufs et al., 2006). Third, to undertake a detailed assessment of three EEG pre-processing methods on the estimation of combined EEG-fMRI information, reflecting the question of EEG data quality in simultaneous EEG-fMRI recordings and their influence on inferences about neuro-physiological phenomena (Bagshaw and Warbrick, 2007; Debener et al., 2006; Mantini et al., 2007; Porcaro et al., 2010).

The single-trial data features under consideration in this study are a) for the EEG time-domain, the amplitude of the visually induced standardized current density power in a time-window of 80-120 ms post-stimulus, corresponding to the P100 amplitude in electrode space, b) for the EEG frequency-domain, the relative change of alpha frequency amplitude between a pre-stimulus and post-stimulus time window of ± 1 s, and c) for the fMRI modality, the maximum of the haemodynamic response function in a time-window of 0 – 16.5 s post-stimulus. These features were chosen for the following reasons: first, with respect to the EEG modality, the interplay of visually evoked activity, changes in spontaneous and induced alpha activity, and haemodynamic responses remains largely elusive (Becker et al., 2008; Nierhaus et al., 2009). Second, the amplitude of single-trial haemodynamic responses has been shown to be sensitive to and informative about the stimulus contrast previously (Bagshaw and Warbrick, 2007; Logothetis et al., 2001; Ostwald et al., 2010), and hence is a reasonable choice for the spatial integration framework attempted here.

Finally, for the integration of simultaneously acquired EEG-fMRI data it is vital that the data is of good quality, i.e. that data artefacts caused by the physical interactions of both recording set-ups are attenuated as far as possible. This is especially crucial for the EEG, which is subject to various biological (eye-movements, muscle movements,

ballistocardiogram artefact) and non-biological (gradient artefact) artefacts during simultaneous acquisition with fMRI data. Independent component analysis (ICA) is a blind signal processing technique increasingly used for single-trial EEG-fMRI studies and has been successfully incorporated into the information theoretic analysis of EEG-fMRI data previously (Chapter 4). Functional source separation (FSS), a semi-blind extension of ICA, has recently been demonstrated to reliable improve single-trial EEG data recorded during concurrent EEG-fMRI (Porcaro et al., 2010). In the current study, the influence of these two pre-processing techniques on the voxel-wise estimation of information theoretic quantities is assessed.

All custom written Matlab (The Mathworks, Natick, MA) code used in this study is available from http://www.buic.bham.ac.uk/EEG_FMRI_ITQ/EEG_FMRI_3D_Analysis.zip and the data is available from the author upon request.

5.2 Methods

5.2.1 Experimental paradigm and data acquisition

Both experimental paradigm and data acquisition have been described in Chapter 4, sections 4.2.1 to 4.2.5.

5.2.2 Data pre-processing

The EEG data were preprocessed as described in section 4.2.6, which for the basic pre-processing involved gradient artefact removal (Allen et al., 1998; Allen et al., 2000), pulse artefact correction (Optimal Basis Set method (Niazy et al., 2005)) as implemented as a plug-in to EEGLAB (Delorme and Makeig, 2004), low-pass filtering at 25 Hz, and down-sampling to 500 Hz. In the following, EEG data pre-processed in this manner will be referred

to as 'Basic EEG' data. Further, as described in section 4.2.7, the basic EEG data underwent additional artefact correction using Independent Component Analysis (ICA). The EEG data after this pre-processing step will be referred to as 'ICA EEG' data. Finally, the EEG data underwent another pre-processing procedure, namely functional source separation (Porcaro et al., 2010), the resulting data of which will be referred to as 'FSS EEG'[1]. In brief, for the FSS procedure additional information is included to bias the independent component decomposition algorithm towards solutions that satisfy physiological assumptions. Specifically, the aim of FSS is to enhance the separation of relevant signals by exploiting some a priori knowledge without renouncing the advantages of using only information contained in the original signal (Barbati et al., 2006; Tecchio et al., 2007). The computational underpinnings of FSS are discussed in (Porcaro et al., 2010) and Appendix D. The criterion maximized for the independent component solution in the present study was the VEP amplitude in an 80 – 120 ms post-stimulus time-window. The single-trial VEPs of the basic and retro-projected ICA and FSS EEG data were then subjected to the LORETA algorithms discussed below. Single-trial VEPs were obtained with respect to a 100 ms pre-stimulus baseline for the time-domain features and the entire pre-stimulus period for the frequency domain features, respectively.

As in Chapter 4, raw EPI data (.par/.rec format) were converted to Niftii format using MRICroN's dcm2nii utility (Rorden and Brett, 2000). SPM5 (Friston, 2007) was then used for fMRI data pre-processing. Data pre-processing involved anatomical realignment, slice scan time correction (reference slice 10), anatomical normalization, and re-interpolation to 7 x 7 x 7 mm voxels to allow MNI space alignment with the EEG inverse solutions (see below). Using the SPM5 general linear model approach, additional high-pass filtering (cut-off 1/128

[1] The FSS version of the EEG data were kindly provided by Dr. Camillo Porcaro

Hz) and data whitening (AR(1)-model) was implemented by pre-multiplication with the corresponding filtering matrices.

5.2.3 MNI space alignment

For the computation of ITQs based on both imaging modalities at a specific voxel location, it is essential that both data types relate to the same anatomical location. The MNI space as defined in SPM5 was used as reference space (Friston, 2007). Specifically, the fMRI data were re-interpolated to the resolution of the EEG lead field matrix, i.e. 7 x 7 x 7 mm voxels ranging in the x- dimension between MNI coordinates -63 and 70 mm, in the y-dimension between -105 and 70 mm, and in the z-dimension between -42 and 35 mm.

The projection of scalp EEG data into normalized anatomical space is an intricate problem and several procedures, based on different assumptions about the physiological mechanisms generating the scalp EEG, have been developed (Michel et al., 2004). As the aim of this study was to assign an activity estimate to each voxel of a MNI space, a current density estimation approach was chosen for the inverse solution, as discussed in section 5.2.4. The spatial resolution and correspondence to standardized anatomical spaces of any EEG inverse solution is determined by the configuration of its corresponding forward solution (Chapter 1, section 1.2). EEG forward solutions take the form of lead-field matrices, implementing the linear projection of cortical source activity at each anatomical location (voxel) to the EEG electrodes. A necessary pre-requisite for an EEG inverse solution corresponding to voxel activity estimates in MNI space is hence a forward solution based on MNI space. For the current study a solution to the EEG forward problem, i.e. a lead field matrix, was obtained using Curry 6.0.1 (Compumedics Neuroscan, Charlotte, NC). Specifically, the lead field matrix was based on a standard boundary element model (BEM) created from an MNI space template with a conductivity ratio brain/skull of 1/78 (conductivity parameters 0.33 Sm^{-1} for

scalp, 0.0042 Sm^{-1} for skull, and 0.33 Sm^{-1} for brain tissue). Due to the high computational demand of the inverse algorithms used the resolution of the lead field matrix was set to 7 mm^3 isotropic in PAN (L,P,S) space, which was the highest possible spatial resolution for computing the inverse solutions under Matlab 7.5 (R2007b) on a 64-bit Linux machine.

In the combined EEG-fMRI data analysis discussed below, single-trial EEG features were extracted from the nearest-neighbour of each EPI space voxel based on the MNI voxel coordinates of the lead-field matrix. The alignment between the EPI and EEG MNI spaces derived in the manner discussed above is complicated by the fact that the Curry 6.0.1 software operates on an internal coordinate system (PAN (L,P,S) (Compumedics USA Ltd, 2008)), which is transformed to SPM's MNI space using a rigid-body rotation during export of the lead-field matrix. The re-sampled voxel coordinates as well as the rotation matrix from PAN (L,P,S) space are exported with the lead field matrix. As a consequence, the voxel coordinates of the lead-field matrix, although corresponding to an isotropic 7 mm^3 grid in PAN (L,P,S) space, are slightly displaced with respect to SPM's reference MNI space. The implications of this are discussed below.

In PAN (L,P,S) space the x-axis is oriented right to left and intersecting the left and right peri-auricular points, the y-axis is oriented sagitally anterior-posterior and the z-axis is oriented ventral-dorsally. The orientation of PAN (L,P,S) space amounts to an affine rotation of MNI space (or vice versa). This is conceptually depicted for the two-dimensional projection in Figure 5.2 A. More specifically, the orientation matrix vectors are depicted in Figure 5.2 B, where the PAN (L,P,S) basis vectors are plotted as standard unit vectors in black, and the PAN (L,P,S) to MNI space rotation matrix vectors are plotted in grey. It is easy to see that PAN (L,P,S) and MNI space are inverted with respect to x and y orientation, and tilted with respect to the z-orientation. Figure 5.2 C displays the result of this affine rotation on the MNI space voxel coordinates obtained in a Curry 6.0.1 - generated lead-field matrix.

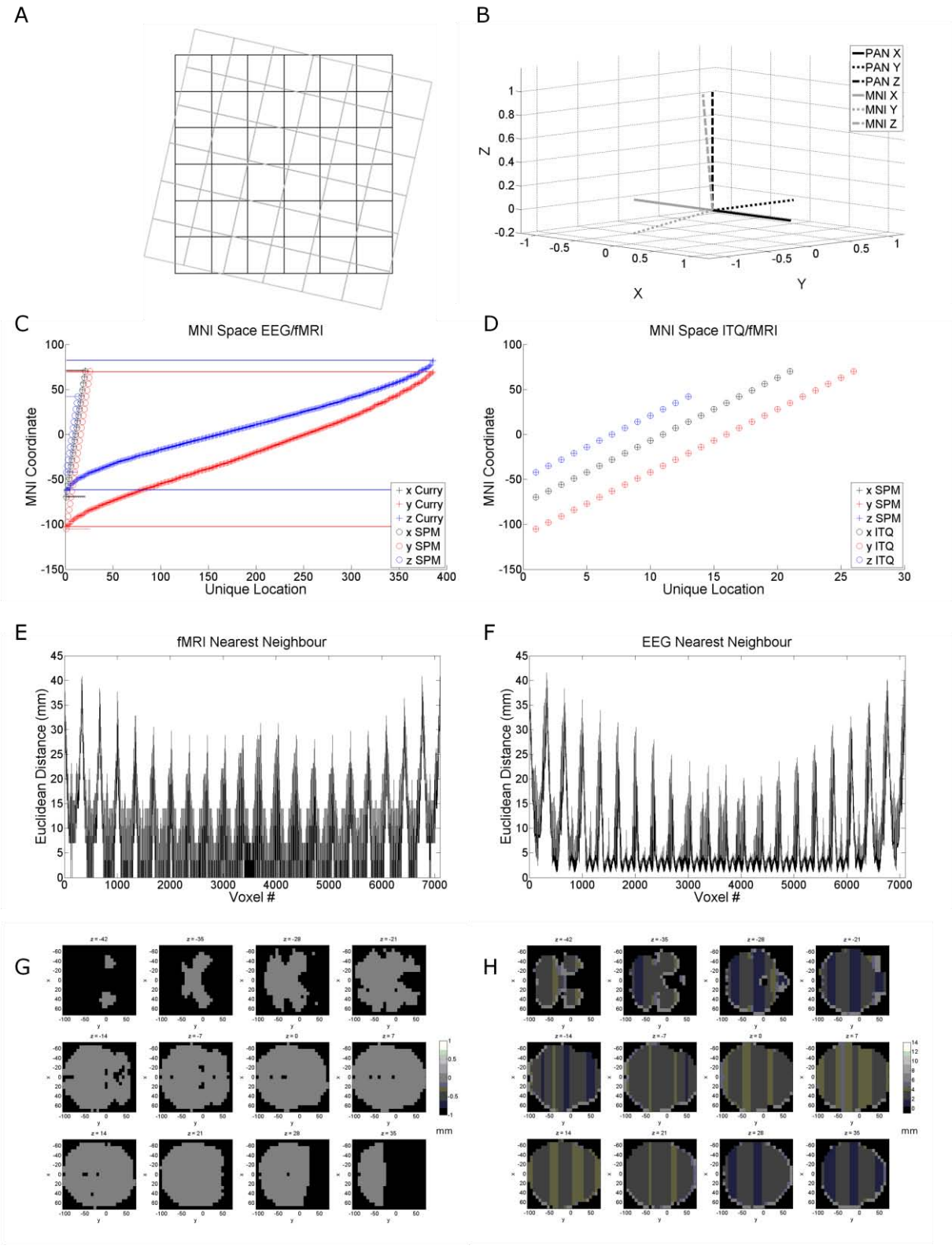


Figure 5.2 EEG-fMRI MNI Space Alignment A) Conceptual depiction of MNI and PAN (L,P,S) misalignment B) Affine translation vectors C) PAN (L,P,S) space resampling performed by Curry 6.0.1 D) Information theoretic MNI space E) EPI data remaining location errors (Euclidean distance in mm) over voxels F) EEG data remaining location errors over voxels (Euclidean distance in mm) G) EPI data remaining location errors in brain space (Euclidean distance in mm) H) EEG data remaining location errors in brain space (Euclidean distance in mm)

These unique coordinates are created, when the software re-samples the PAN (L,P,S) lead field coordinates of a 7 mm isotropic lead field according to MNI space coordinates.

Effectively, the MNI space lead field matrix is not isotropically sampled. However, for the information theoretic analysis the aim was to sample MNI space isotropically in order to be able to image the data with unique 3D locations. It was hence chosen to perform the analysis in 7 mm isotropically sampled MNI space in alignment with SPM5's MNI space as shown in Figure 5.2 D. To unambiguously allocate the data from both the SPM5 and the Curry 6.0.1 MNI space to a location in the information theoretic MNI space, a nearest-neighbour allocation (interpolation) approach was chosen. The nearest neighbour distances of both the spatial EEG and fMRI data for a cubic space of dimensionality 21 x 26 x 13 in x-, y-, and z- direction is shown in Figure 5.2 E and F (data of a representative subject). As the brain space only occupies roughly a central sphere in this cubic space, some of the Euclidean distances are relatively large. For brain voxels in the fMRI modality, the Euclidean distance approaches zero at the locations of information theoretic space sampling, as the SPM5 MNI space forms the basis of the information theoretic MNI space. Due to the non-isotropy of the Curry 6.0 lead-field a remaining error of ~ 1.5 mm remains even in the best cases. However, given the overall low resolution of the analysis this amounts to an error of considerably less than a voxel and is hence negligible. Finally, Figure 5.2 G and H display the remaining Euclidean distance alignment error over voxels in brain space. Panel G displays all voxels in SPM5 MNI space with a remaining Euclidean distance < 3.5 mm and panel H displays in Curry 6.0 MNI space with a remaining Euclidean distance < 7 mm with respect to the information theoretic MNI space. The intersection of those two maps was included in the analysis.

5.2.4 EEG inverse solutions

Two inverse solution algorithms were chosen to obtain estimates of electrical brain activity based on the scalp EEG recordings: first, in the time domain, the standardized current density power for each brain location was estimated using sLORETA (Pascual-Marqui, 2002b). Assuming an instantaneous, distributed, discrete and linear solution to the EEG inverse problem, sLORETA can provide activity estimates for each dipole located at all pre-specified MNI voxel locations over time. The algebraic formulation of sLORETA implemented for the current study is reviewed in Appendix E.

Second, in the frequency domain, voxel-wise frequency power estimates for peri-stimulus time-windows were obtained using a LORETA variant specifically tailored to frequency analyses (Frei et al., 2001; Pascual-Marqui, 2002a). This inverse solution is based on a forward model of the Fourier-transformed impressed source current densities and capitalizes on the evaluation of their cross-spectral matrix. The algebraic formulation of this LORETA-variant implemented for the current study is reviewed in Appendix F. In order to validate the implementation of the LORETA algorithms, a number of analyses were performed on simulated data, as reported below.

5.2.5 EEG inverse solution validation

To validate the sLORETA algorithm, at a single time-point t, a radial and exponentially decaying point source in MNI space, $J(t)$, was simulated, and its profile recovered upon its lead field based projection into electrode space. Specifically, for the j-th voxel, $sCDP_j(t)$ was computed from $V(t)$ according to the specifications in Appendix E for a spatial current density field given by

$$J_j(t) = \begin{cases} (1,1,1)^T & for\ j = i \\ \frac{1}{\|c_i - c_j\|_2^2}(1,1,1)^T & for\ j \neq i \end{cases} \quad (5.1)$$

Here, $J_j(t) \in \mathbb{R}^3$ is the current density vector of the jth voxel, $i \in \mathbb{N}$ is the index of the voxel with maximal signal, $c_k \in \mathbb{R}^3$ denotes the MNI coordinates of the k-th voxel and $\| \ \|_2$ denotes the Euclidean norm. This recovery amounts to the same data projection carried out for each sample time-point of an individual experimental trial.

Figure 5.3 A displays the simulated point source at MNI coordinates [14 -84 0], and panels B, C, D display its recovery using different values for the regularization parameter α, where in B $\alpha = 1$, in C $\alpha = 10$ and in D $\alpha = 0.1$ In the analysis of the experimental data, the regularization parameter was fixed to $\alpha = 1$, as an increase by a factor 10 (panel C) resulted in an overly smooth profile of the recovered point source, while a decrease by a factor 10 (panel D) did not alter the recovery profile much. Equivalent results were obtained for more extreme changes of α (data not shown). Figure 5.4 displays simulated point sources at MNI coordinates [0 0 0] (5.4 A) and [-35 42 0] (5.4 C) and their sLORETA-based reconstructions (5.4 B, and 5.4 D, respectively) for a regularization parameter $\alpha = 1$. For the central source simulation (Figure 5.4 A), the topological recovery of the simulated source is exact, i.e. the voxel coordinates of the maximum of the reconstructed source are identical with the voxel coordinates of the maximum of the simulated source. This is in correspondence with the known zero localization error for sLORETA (Pascual-Marqui, 2002b). For the occipital source the corresponding maxima are displaced by a Euclidean distance of 7 mm, i.e. one voxel, while for the left frontal source the corresponding maxima are displaced by a Euclidean distance of 9.9 mm, i.e. less than 2 voxels. These slight localization errors are caused by the MNI space nonlinearities at the spaces edge and regarded as negligible given the overall low resolution of the current study and the spatial smoothing and spreading property of the inverse solution.

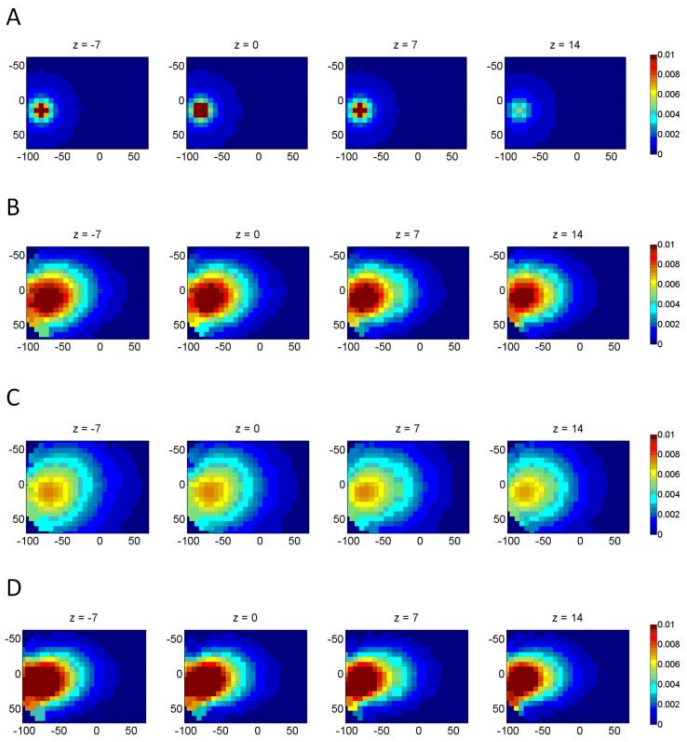


Figure 5.3 Point source simulation and sLORETA based recovery I. Panel A displays the simulated radially decaying point source and panels B, C and D its sLORETA based recovery under variation of the regularization parameter $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha = 1}$, $\boldsymbol{\alpha = 10}$ and $\boldsymbol{\alpha = 0.1}$, respectively.

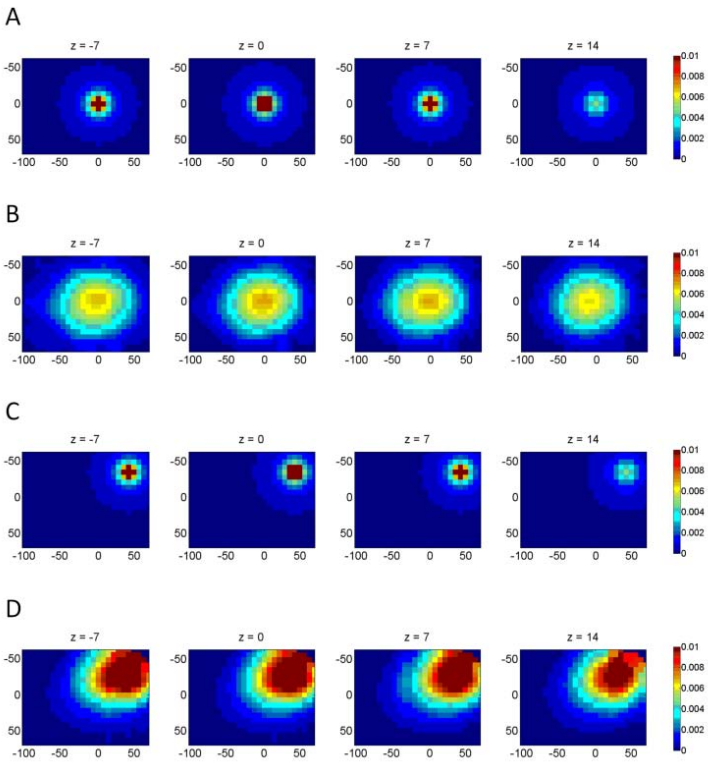


Figure 5.4 Point source simulation and sLORETA based recovery II. Panels A and B display a point source simulation at MNI coordinates [0 0 0] and its sLORETA based recovery, respectively. Panels C and D display a point source simulation at MNI coordinates [-35 42 0] and its sLORETA based recovery, respectively.

Two approaches were chosen to validate the frequency domain LORETA inverse solutions: first, to ensure the projection of the correct frequency profile from electrode to voxel-space, electrode data of known frequency content was simulated and its brain-space frequency spectrum assessed. Second, to ensure appropriate localization, brain sources of known frequency content were simulated and their spatial recovery assessed as for the time-domain solution.

For the first approach, a periodic signal $S(t)$ of the form

$$S(t) = \sum_{i=1}^{3} a_i \sin(2\pi\omega_i t) \tag{5.2}$$

where $a_i, \omega_i \in \mathbb{R}$ was simulated for a time interval of 1 s at a sampling frequency of 125 Hz and allocated to a subset of simulated electrodes. Simulated time-series for 'signal electrodes' are shown in Figures 5.5 A and 5.6 A. The signal time-series were embedded into an electrode space where all other non-signal electrodes were set to independent Gaussian noise (i.e. for each sample point, a value was sampled from a Gaussian distribution with mean 0 and variance 1) as shown in Figures 5.5 B and 5.6 B. Panels C of each figure display the topography of the recovered frequency power for all frequencies of interest in the main experiment, 1-25 Hz. For Figure 5.5, the chosen signal electrodes were located right occipital, O2 and PO8, the three signal frequency components were set to $\omega_1 = 10\,Hz$, $\omega_2 = 16\ Hz$ and $\omega_2 = 20\,Hz$ and the corresponding fourier coefficients to $a_1 = 10$, $a_2 = 9$ and $a_3 = 8$. For Figure 5.6, the chosen signal electrodes were located left temporal, T7, FT7, and C5, the three signal frequency components were set to $\omega_1 = 2\,Hz$, $\omega_2 = 5\,Hz$ and $\omega_2 = 8\,Hz$ and the corresponding Fourier coefficients to $a_1 = 9$, $a_2 = 10$ and $a_3 = 11$. Inspection of the recovered frequency power and topography revealed results consistent with the simulated parameters in both cases. Specifically, the topographical frequency plots exhibit frequency power only for the frequency components ω_i of the simulated signals and the corresponding power profile conform to the quadratic mapping of the respective Fourier coefficients a_i.

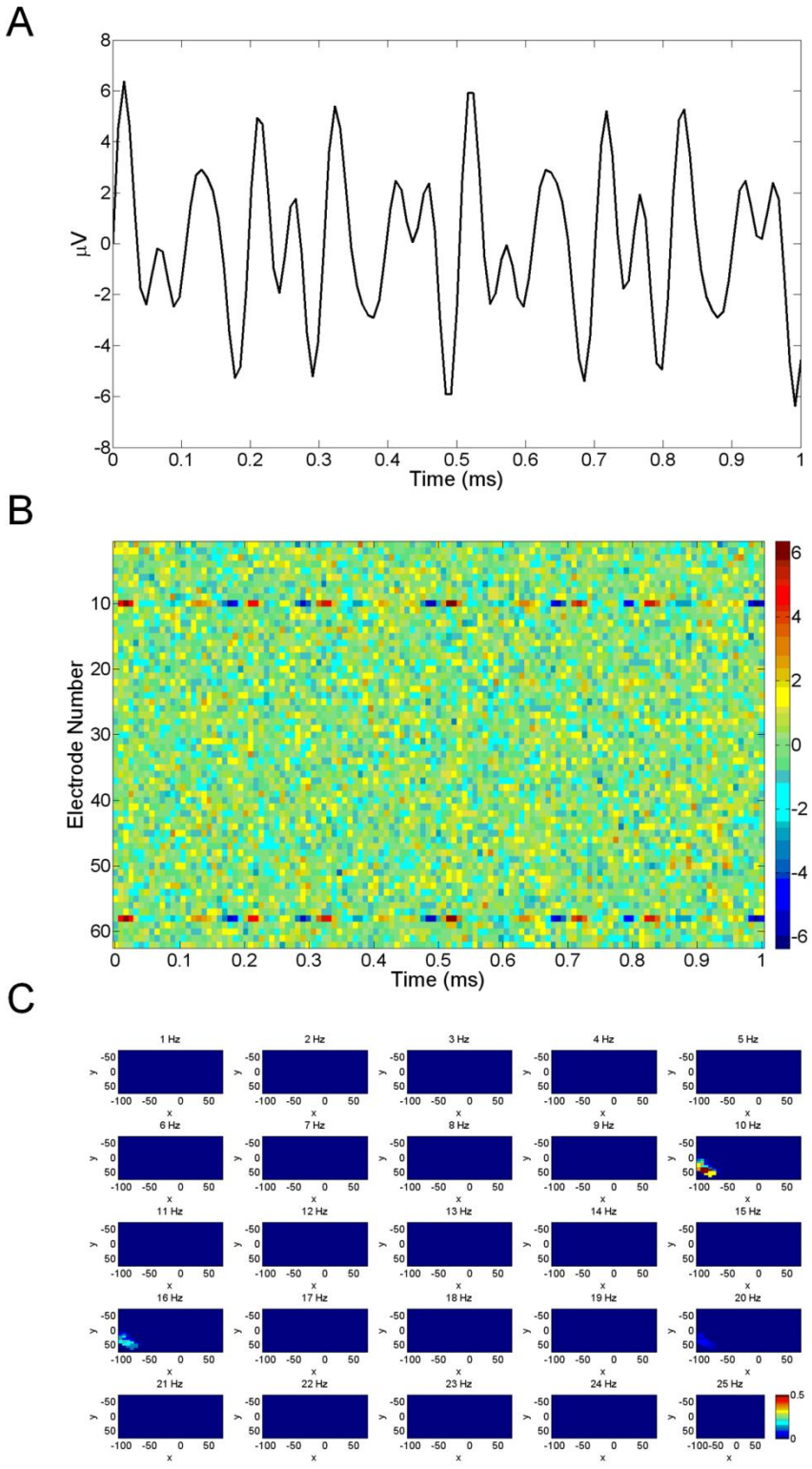


Figure 5.5 Frequency domain LORETA implementation validation I. A periodic signal according to equation (5.2) was simulated at electrodes O2 and PO8, the three signal frequency components were set to $\boldsymbol{\omega_1 = 10\,Hz}$, $\boldsymbol{\omega_2 = 16\,Hz}$ and $\boldsymbol{\omega_2 = 20\,Hz}$ and the corresponding Fourier coefficients to $\boldsymbol{a_1 = 10}$, $\boldsymbol{a_2 = 9}$ and $\boldsymbol{a_3 = 8}$. Panel A displays the simulated signal, panel B its embedding into the entire sensor space, and panel C its frequency domain LORETA based frequency power and topographical profile for z = 0.

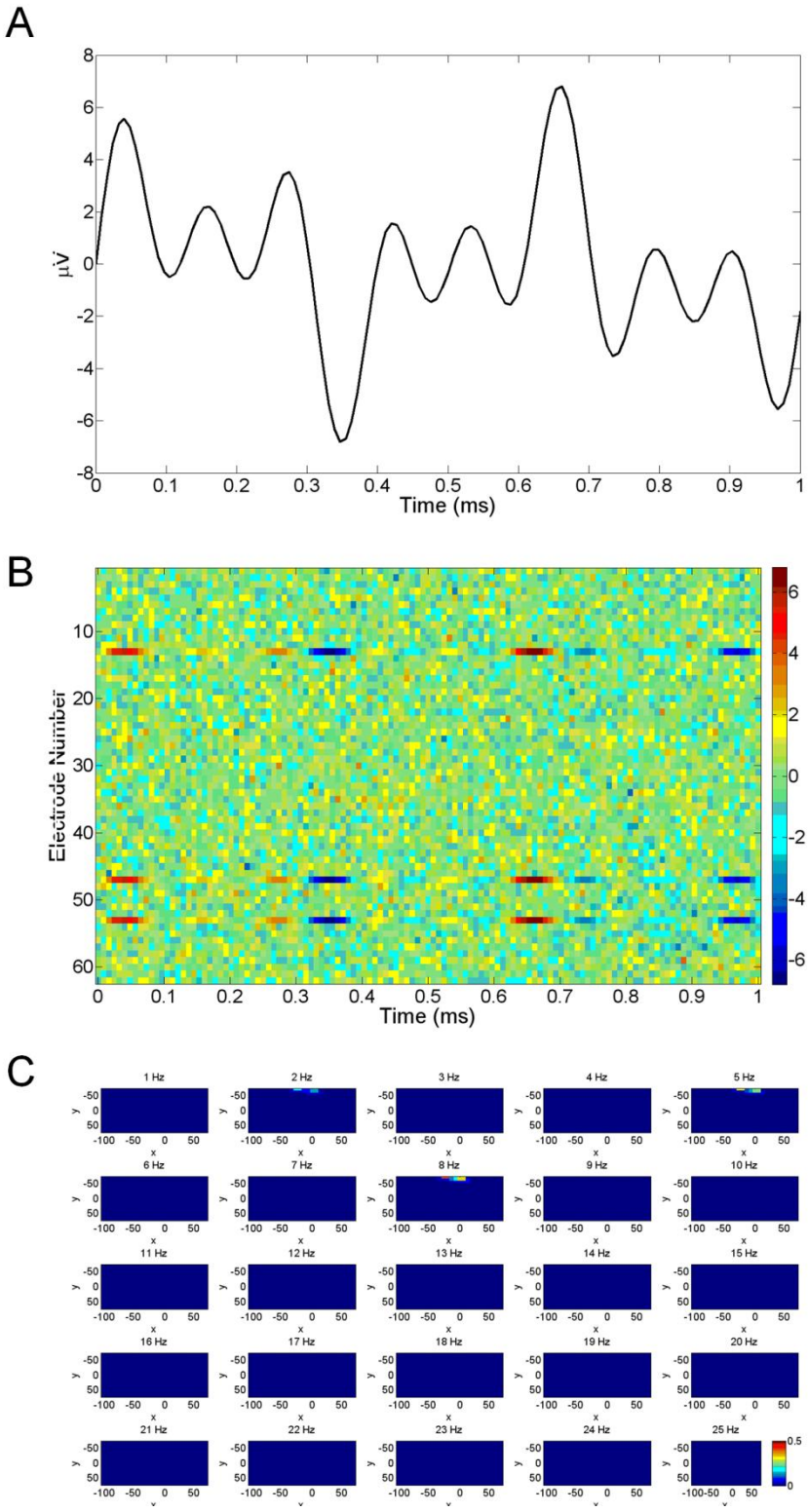


Figure 5.6 Frequency domain LORETA implementation validation II. A periodic signal according to equation (5.2) was simulated at electrodes T7, FT7, and C5, the three signal frequency components were set to $\boldsymbol{\omega_1 = 2\,Hz}$, $\boldsymbol{\omega_2 = 5\,Hz}$ **and** $\boldsymbol{\omega_2 = 8\,Hz\,z}$ and the corresponding fourier coefficients to $\boldsymbol{a_1 = 9}$, $\boldsymbol{a_2 = 10}$ and $\boldsymbol{a_3 = 11}$. Panel A displays the simulated signal, panel B its embedding into the entire sensor space, and panel C its frequency domain LORETA based frequency power and topographical profile for z = -14 .

Further, for left occipital signal electrodes, the brain source projected into a left occipital region, while for right temporal electrodes the brain source is projected into right temporal regions.

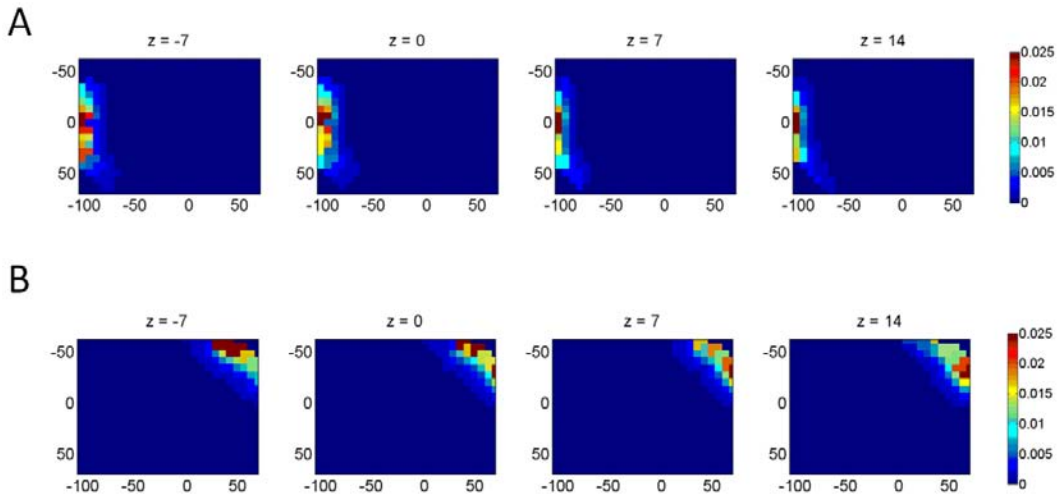


Figure 5.7 Frequency domain LORETA source estimation validation. A) Spatial 10 Hz frequency power profile for the recovery of an exponentially decaying point source located at [14 -84 0] with signal $\boldsymbol{S(t)}$ simulated according to (5.2) for each dipole moment ($\boldsymbol{\omega_1 = 10\,Hz}$, $\boldsymbol{\omega_2 = 16\ \ Hz}$, $\boldsymbol{\omega_2 = 20\,Hz}$, $\boldsymbol{a_1 = 10}$, $\boldsymbol{a_2 = 9}$, $\boldsymbol{a_3 = 8}$) B) Same as in A) with the point source located at [-35 42 0].

For the second approach, the three dipole moments for a radial and exponentially decaying point source of the form (5.1) were set to a signal $S(t)$ of the form (5.2), their frequency domain LORETA solution recovered upon projection of the data into electrode space. Figure 5.7 displays the 10 Hz frequency power topography for two simulated sources located at [14 -84 0] (A) and [-35 42 0] (B), where the signal constants in (5.2) were set to $\omega_1 = 10\,Hz$, $\omega_2 = 16\ \ Hz$, $\omega_2 = 20\,Hz$, $a_1 = 10$, $a_2 = 9$ and $a_3 = 8$. While the frequency power spectrum of the recovered sources was exact, the topography exhibited a slight localisation error (simulated source maximum to recovered source maximum Euclidean distance 26.6 mm, corresponding to less than 4 voxels). Specifically, the recovered sources are displaced to the edge of the simulated brain space. Further simulations revealed that the exact localization error is a function of the simulated signal, i.e. the way that the signal is

projected on the simulated dipole moments. Together with the experimental results reported below, and the general known non-optimality of inverse solution in the frequency domain (Michel et al., 2004), and the spatial smoothing and spreading property the topographical properties of the frequency-domain LORETA inverse solution were hence judged sufficient for the current study.

5.2.6 Feature definition

In the Chapter 4, a large set of time-domain features traditionally associated with the study of VEPs was evaluated. The same approach taken here would lead to a combinatorial explosion of the number of computations performed, as well as the number of information theoretic maps to be evaluated. Instead, the discussion focuses on the following features extracted from the peri-stimulus time-courses of each voxel in each modality: a) for the EEG modality in the time-domain, the maximum of the standardized current density time-course in a time-window of 80 – 120 ms post-stimulus onset, referred to as $sCDP_{P100}$, b) for the EEG modality in the frequency domain, the relative difference of 1 s pre-stimulus and 1 s post-stimulus alpha power (average of frequency power of 8 – 12 Hz) given by

$$\Delta\alpha = \frac{\alpha_{prestimulus} - \alpha_{poststimulus}}{\alpha_{prestimulus}} \tag{5.3}$$

and c) for the fMRI modality, the maximum of the haemodynamic response in a of 0 – 16.5 s post-stimulus time-window, referred to as HRF_{max}, where the haemodynamic response was quantified as BOLD percent signal change with respect to the pre-stimulus baseline (average of the three pre-stimulus onset TR values). The feature definitions are summarized in Table 5.1. It should be noted that given the relatively low-frequency cut-off of the filtered EEG data necessary for MR artefact correction, the $sCDP_{P100}$ does not contain contributions from frequencies larger than 25 Hz.

5.2.7 Information theoretic analysis

The information theoretic analysis (i.e. the computation of the quantities $I_N(S;R_1,R_2)$ $I_N(S;R_1)$ $I_N(S;R_2)$ and $Synergy$, the activity dependence $I_N(R_1;R_2)$ and the conditional dependence $\langle I_N(R_1;R_2|S)\rangle_S$ from experimental data, where S indicates a stimulus variable and R_1 and R_2 indicate two response signal features) has been discussed in detail Chapter 2 (Table 5.2). For the current study, all information theoretic quantities were evaluated for each voxel of the reference anatomical space.

Acronym	Definition
$\boldsymbol{sCDP_{P100}}$	$\max\{post-stimulus\ sCDP\ \ (80-120\ ms)\}$
$\boldsymbol{\Delta\alpha}$	$\frac{\alpha_{prestimulus}-\alpha_{poststimulus}}{\alpha_{prestimulus}}$
$\boldsymbol{HRF_{max}}$	$\max\{post-stimulus\ percent\ signal\ change\ (0-16.5s)\}$

Table 5.1 Feature definitions of Chapter 5. For the information theoretic analyses, each feature was extracted from the data corresponding to each voxel of the respective modality

	Information Theoretic Quantities
I(S;R)	Mutual information between the stimulus distribution and a univariate response feature distribution
$I(S;R_1,R_2)$	Mutual information between the stimulus distribution and the joint distribution of two response features (here EEG and fMRI derived)
Synergy	Normalized difference between $I(S;R_1,R_2)$ and the sum of $I(S;R_1)$ and $I(S;R_2)$
$I(R_1;R_2)$	Mutual information between two response feature distributions (here EEG and fMRI derived), referred to as activity dependence
$I<(R_1;R_2|S)>_S$	Average stimulus conditional mutual information between two response feature distributions (here EEG and fMRI derived), referred to as conditional dependence

Table 5.2 Information theoretic quantities of Chapter 5

With respect to the free parameters in the analysis, namely the upper and lower limits of the discretization grid and the total number of bins, the same choices as discussed in Chapters 2 - 4 were made: the upper and lower limits of the response grid were set to the maximum and minimum value of each response variable and the number of response bins was chosen to be approximately equal to half the number of trials per stimulus resulting in 36 response bins for the bivariate response variable distributions. For bias correction, a combination of PT-correction, shuffling correction ($I_N(S; R_1, R_2)$) and the subtraction of the remaining biases in the analysis of a linear Gaussian null model with the same parameters as the experiment (number of response variables, number of stimuli, number of trials per stimulus) were employed as detailed in Chapters 2,3 and 4.

5.2.8 Statistical evaluation

To statistically evaluate the derived information theoretic maps (effect size maps), voxel-wise null hypothesis testing in combination with false-discovery rate correction for multiple comparison was employed. The t-test strategy used in Chapter 4 for the same problem is not fully appropriate, as the expected ITQ distributions cannot be assumed to be Gaussian, because the information theoretic estimates are not symmetrically distributed about zero. This is due to the fact that, at least in principle, negative information values do not occur. Further, as the computational demands for a permutation based analysis would be very high, ITQ null distributions were derived from a Gaussian null model by sampling 10,000 times from model M0 and evaluating all ITQs. This resulted in ITQ distributions reflecting the case of non-informative Gaussian signals. It should be noted that these distributions are constant under variation of the variance parameters of the M0 model, as they reflect between-simulation, rather than within-simulation, variance (Figure 5.8). Based on these observed frequency distributions,

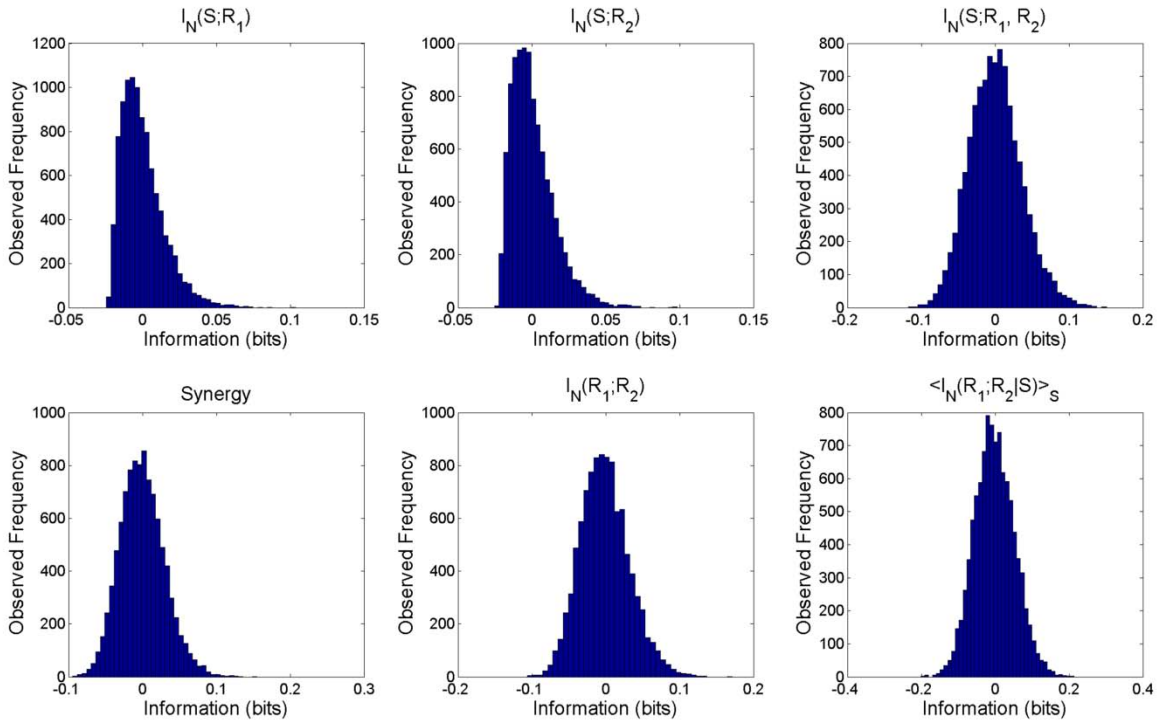


Figure 5.8 M0 derived ITQ probability distributions. The panels display the distribution of estimated ITQ values over 10,000 instantiations of the Gaussian null model M0. As is evident especially from the distribution of $\boldsymbol{I_N(S;R_1)}$ and $\boldsymbol{I_N(S;R_2)}$, these distributions can be non-Gaussian, thereby decreasing the validity of t-tests

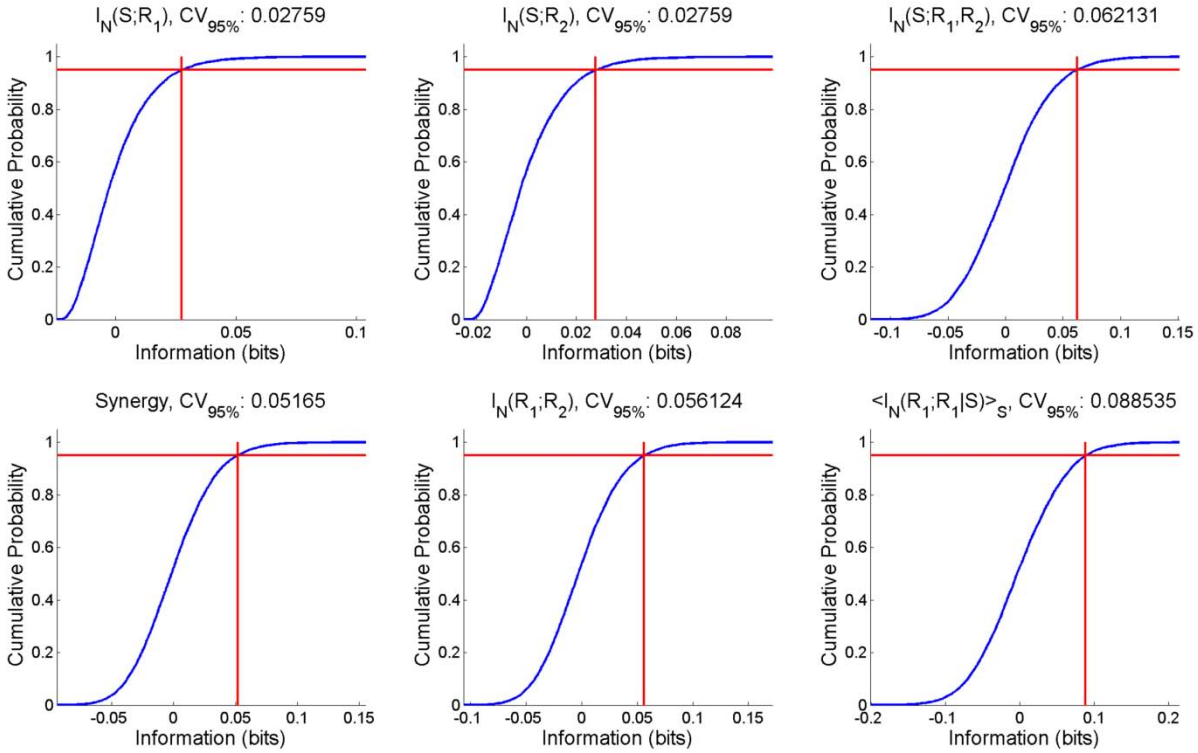


Figure 5.9 M0 derived ITQ cumulative probability distributions. Kaplan Meier estimates for the empirical cumulative distribution functions of the observed frequency distributions of Figure 5.8. The red lines indicate the critical value for a one-tailed alpha rate of 0.05, which is also plotted in each figure's title.

the corresponding Kaplan-Meier estimates of the cumulative probability functions for each ITQ were computed (Figure 5.9), and the critical value for one-tailed hypothesis tests at an alpha error level of 0.05 determined. P-values were then assigned to all MNI space voxels for which data from all observer exits.

To correct for multiple comparisons the resulting statistical maps were thresholded using a false-discovery rate (FDR) procedure (Benjamini et al., 1995; Genovese et al., 2002). Controlling the FDR is a well established procedure for false-positive control in the neuroimaging literature (Chumbley and Friston, 2009; Friston, 2007; Genovese et al., 2002). Briefly, the FDR procedure allows controlling the upper bound of the expectation of the FDR, which is defined as ratio of the number false-positives to the total number of statistical tests. The FDR procedure is reviewed in Appendix G.

5.3 Results

In the following, the results of the whole-brain information theoretic analysis based on the three features of interest introduced above (i.e. $sCDP_{P100}$, $\Delta\alpha$ and HRF_{max}) will be discussed. The derived information theoretic maps are the outcome of a series of data conversions, e.g. for the EEG data from electrode space data to MNI space data and from time-domain data to frequency domain data, for both modalities from raw to feature data. To assess the influence of these various steps on the information theoretic analyses, the results presented here are supplemented by the results displaying the average feature effect sizes and a comparison between the electrode and brain space EEG data. Further, when appropriate, results are presented for all three EEG data pre-processing techniques (basic, ICA, FSS) to demonstrate the influence of EEG data pre-processing on the final information theoretic analysis outcome. An in-depth quantitative comparison of ICA and FSS effects on single-trial

EEG data acquired simultaneously with fMRI in a non-information theoretic framework is also the subject of (Porcaro et al., 2010).

5.3.1 Average feature effects

Evaluating the information of brain imaging data features with respect to the stimulus (or another brain imaging data feature) amounts to evaluating the respective signal's stimulus (or brain imaging data feature) differentiating potential for a single experimental trial. It hence provides a statistic for the stimulus' (or brain imaging data feature) discrimininability in the presence of noise. Nevertheless, in the study of brain imaging signals, more global signal statistics, e.g. the signal's amplitude on average and their sensitivity to various pre-processing methods, are also of relevance. In the following section the average effect sizes for stimulus-induced signal variations for the signal features considered in the information theoretic analysis will be discussed. The EEG data in electrode and MNI space will be considered before the fMRI data.

5.3.1.1 EEG data in electrode space

Figure 5.10 depicts the group average EEG time- and frequency domain data for electrode PO8 for each EEG pre-processing method. The first column depicts the time-domain signal for a peri-stimulus time-window of -1 to 1 seconds. The second column depicts the pre-stimulus induced power spectrum for the time-window -1 to 0 seconds, while the third column depicts the post-stimulus induced power spectrum for the 0 to 1 second time window. In each panel, the black trace indicates high contrast trials, while the grey trace indicates low contrast trials. For each processing method, the visually evoked potential for the checkerboard onset (0 s) and reversal (0.5 s) are clearly visible. Both ICA and FSS pre-processing result in a slight enhancement of the VEP amplitude contrast for the checkerboard onset P100

component. Across both stimulus conditions ICA and FSS pre-processing result in a slight reduction of P100 amplitude (Table 5.3)

The respective power spectra in Figure 5.10 show the band-pass filtering effects (low frequency cut-off 0.25 Hz, high frequency cut-off 25 Hz) of the basic EEG data pre-processing. Further, the overall frequency power of the basic pre-processed EEG data is strongly reduced by the application of ICA, and even further by FSS (Table (5.2)). As discussed in the introduction, the main interest of this study lies in alpha band activity. For most methods and stimulus conditions, the peak in the frequency range 8 – 12 Hz is reduced in the post-stimulus relative to the pre-stimulus time-window.

5.3.1.2 EEG data in MNI space

While the EEG signal depicted in Figure 5.10 represents the data of a single EEG electrode, the features of interest in the information theoretic analysis reported here were extracted from a linear combination of the data from all electrodes, namely their LORETA projections to brain space. For the time-domain, the sLORETA conversion results in the non-negative standardized current density power (sCDP). It should be noted that this is qualitatively different from the electrode potential, which varies in both positive and negative directions. The peak of the sCDP around 100 ms post-stimulus, which results from the conversion of the P100 peak in the respective electrode potentials is hence referred to not as P100, but $sCDP_{P100}$.

Figure 5.11 depicts the topography of the respective features, while Figure 5.12 (A-C) displays the corresponding time and frequency domain representations for a specific voxel, selected as the voxel showing the strongest stimulus-induced effects across EEG pre-processing methods. Specifically, the topography plots depict the average feature amplitude over trials and conditions, normalized by the variability (quantified as the standard deviation)

of the time-course at each voxel. All methods result in the highest $sCDP_{P100}$ feature values being located in right occipital cortex, with a clearer contrast to the remaining voxels apparent for the more advanced pre-processing methods (ICA and FSS, respectively). For the $\Delta\alpha$ feature, the strongest topographically focussed changes in pre- to post-stimulus alpha power are observed for the FSS EEG data.

With respect to the voxel-specific time- and frequency domain data, the following differences in comparison to the electrode data were observed: first, as discussed, the sCDP is strictly non-negative. Second, in contrast to the data of a single electrode, the linear combination of all electrode data after ICA and FSS pre-processing results in a reduction of the average amplitude of the $sCDP_{P100}$ (Table 5.3). Third, the voxel-specific power spectra of the standardized current density power data display more prominent peaks in the alpha range than the single-electrode data. Finally, as for the single electrode data, advanced pre-processing results in a large reduction of the overall frequency and alpha band power.

5.3.1.3 fMRI data in MNI space

The fMRI feature amplitude is depicted in Figure 5.11 D (left column). The fMRI feature was defined as the maximal amplitude in a 16.5 s post-stimulus time-window and hence the topography also includes voxels which show high signal amplitudes not affected by the stimulus, as for example the voxels representing the ventricles. This signal variability (quantified as the standard deviation) over a 32 second peri-stimulus time-window is depicted in Figure 5.11 D (right column) and clearly indicates right occipital voxels as displaying the strongest stimulus induced signal variation. Extraction of the post-stimulus time-courses for a representative occipital voxel (approximated as displaying the strongest stimulus-induced effects at [7 -100 0]) yields the expected hemodynamic response functions peaking at approximately 4 to 5 seconds post-stimulus, as depicted in Figure 5.12 D.

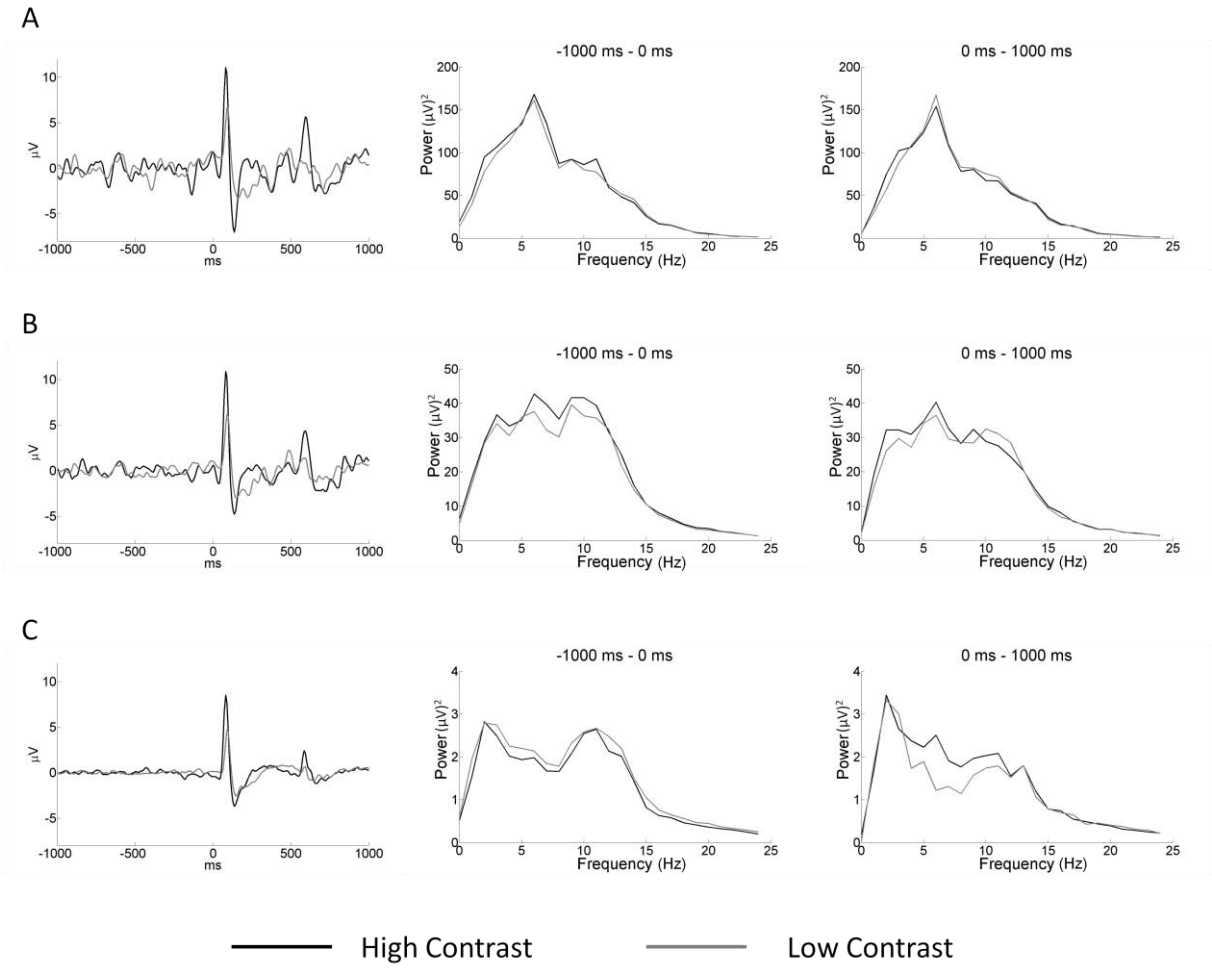


Figure 5.10 Group EEG sensor space data, electrode PO8 for all EEG pre-processing methods (A: basic pre-processing, B: additional ICA pre-processing, C: additional FSS pre-processing). The first column displays the time-domain signals for a peri-stimulus time-window of -1 to 1 seconds, the second column the frequency power spectrum in the pre-stimulus time-window of -1 to 0 seconds, and the third column the frequency spectrum in the post-stimulus time-window of 0 to 1 seconds.

As is evident from Figure 5.11, the strongest stimulus-induced feature effects are topographically slightly displaced for the two imaging modalities. In general the, the strongest feature effects for the EEG data are observed slightly more lateral and anterior in comparison to the fMRI feature maxima (Euclidean distance approximately 2 cm, i.e. three voxels at the current resolution)

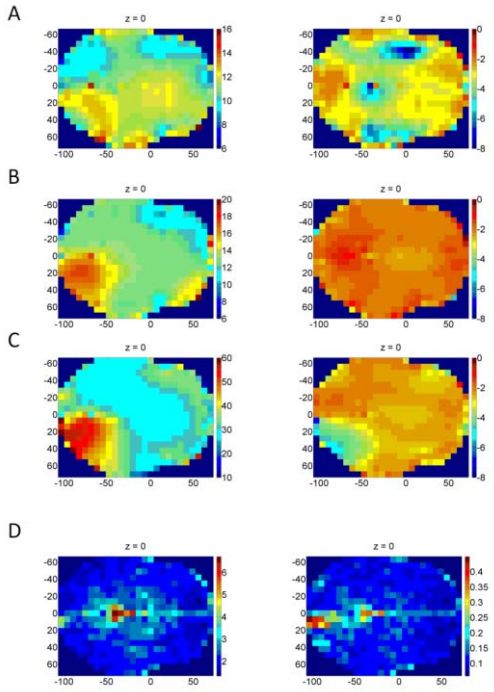


Figure 5.11 Topography of feature amplitudes. Figures A-C depict the respective electrophysiological features, sCDP P100 (left column) and $\Delta\boldsymbol{\alpha}$ (right column), for the different EEG pre-processing methods (A Basic EEG, B ICA EEG, C FSS EEG)). Figure D depicts the amplitude of the HRF feature (left column) as well as the variability of the EPI signal for the same voxels (right column). All feature amplitudes represent the respective group averages across trials and conditions.

5.3.2 Information theoretic quantity maps

In the following, MNI space information theoretic maps derived for all information theoretic quantities and features of interest are presented. For each combination, both group average effect size (information in bits) and group statistical (p-value) maps are presented (n = 12). The scaling of the effect size maps was determined to emphasize topographical features of the resulting maps. The statistical maps are scaled from $p = 0.05$ to $p = 0.00$ (i.e. thresholded at $p < 0.05$ uncorrected) and additionally thresholded at the p-value corresponding to the FDR correction for $q = 0.5$, unless otherwise noted to emphasize topographical features. These thresholded statistical maps are overlaid on a group mean EPI image derived from the first EPI volume of each subject. For all maps, four consecutive slices, depicting approximately the z-directional centre of the acquired EPI stack are displayed. All slices are oriented along the posterior-anterior brain axis from left to right and

the left to right brain axis from top to bottom. All coordinate labels are MNI coordinates. Table 5.4 lists the maximally informative voxel coordinates and extracted information values for all feature and method combinations discussed.

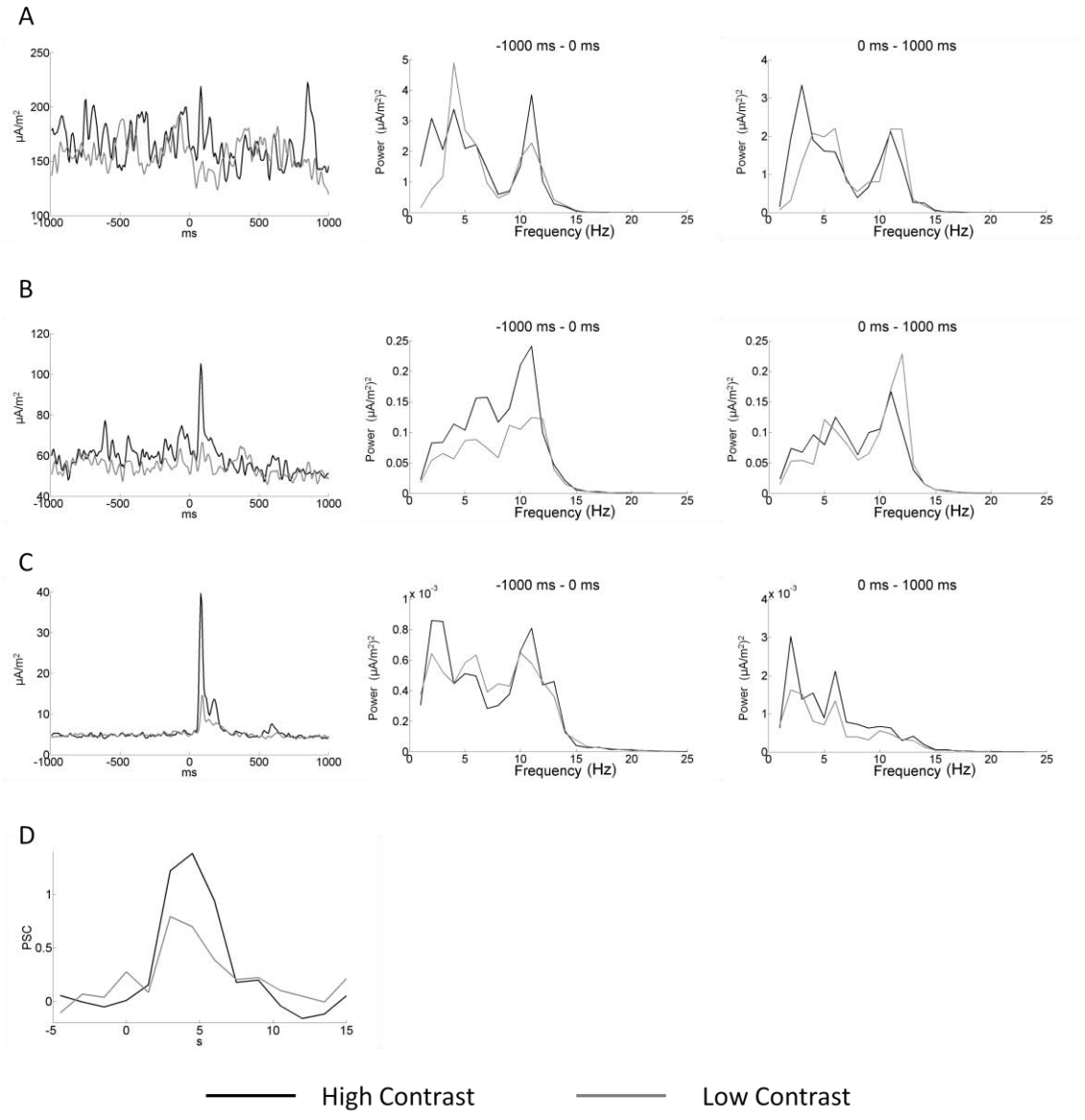


Figure 5.12 EEG and fMRI MNI space data: Peri-stimulus signal time and frequency domain representations extracted from representative right occipital voxels (MNI coordinates EEG data: [25 – 80 0]. MNI coordinates fMRI data: [7 -100 0], coordinates identified as strongly activated voxels for the respective signal topographies). Panels A – C depict the respective electrophysiological time courses (sCDP) for basic EEG (A), ICA EEG (B) and FFS EEG (C). As in Figure 5.10, the first column displays the time-domain signals for a peri-stimulus time-window of -1 to 1 seconds, the second column the frequency power spectrum in the pre-stimulus time-window of -1 to 0 seconds, and the third column the frequency spectrum in the post-stimulus time-window of 0 to 1 seconds. Panel D depicts the peri-stimulus hemodynamic response function.

	Basic	ICA	FSS
Electrode Space			
P100 Amplitude	7.7 (± 1.2)	7.5 (±1.2)	5.6 (±1.4)
Power Spectrum (pre-stimulus)	55.3 (±12.0)	19.8 (±3.2)	1.4 (±0.4)
Alpha Power (pre-stimulus)	94.4 (±24.5)	37.2 (±7.5)	2.2 (±0.5)
Power Spectrum (post-stimulus)	49.0 (±11.2)	17.3 (±2.9)	1.3 (±0.3)
Alpha Power (post-stimulus)	82.1 (±23.9)	39.0 (±6.7)	1.7 (±0.5)
MNI Space			
$sCDP_{P100}$ Amplitude	174.2 (±33.1)	80.9 (±11.3)	23.6 (±8.1)
Power Spectrum (pre-stimulus)	0.9 (± 0.4)	0.05 (±0.01)	$0.2e10^{-4}$ ($\pm 0.8e10^{-5}$)
Alpha Power (pre-stimulus)	1.4 (± 1.0)	0.13 (±0.05)	$0.5e10^{-4}$ ($\pm 0.1e10^{-4}$)
Power Spectrum (post-stimulus)	0.7 (± 0.3)	0.05 (±0.02)	$0.5e10^{-4}$ ($\pm 0.2e10^{-4}$)
Alpha Power (post-stimulus)	1.2 (± 1.0)	0.11 (±0.06)	$0.5e10^{-4}$ ($\pm 0.2e10^{-4}$)

Table 5.3 Quantification of EEG data pre-processing methods effects on time and frequency domain data in electrode and MNI space. All quantifications were performed on the data displayed in Figures 5.10 - 5.12. For the amplitude features (P100 amplitude, $sCDP_{P100}$ amplitude) the entries reflect the average signal amplitude in a time-window of 80 – 100 ms post-stimulus onset across stimuli and subjects ± SEM across subjects For the frequency power features, the entries reflect the average frequency power in the pre- or post-stimulus time-window of -1 to 0 or 0 to 1 s across all frequencies of interest (1-25 Hz) or the alpha band (1-8 Hz) across stimuli and subjects ± SEM across subjects.

5.3.2.1 Stimulus-response characteristics: $I(S;R)$ topography

Figure 5.13 displays the stimulus related information maps for the $sCDP_{P100}$ feature and the three types of EEG data pre-processing (A: Basic, B: ICA, C: FSS). For all cases, a cluster of voxels in the right occipital cortex is informative about the stimulus, most pronounced for z-coordinates of 0 to 14 mm. For the basic EEG data, none of these voxels carries information large enough to be statistically significant, while few voxels survive the statistical correction for the ICA data in the z = 14 plane. The FSS EEG data shows the most pronounced cluster of informative voxels and the highest information estimates of all three pre-processing methods. Thresholding this statistical map at $q = 0.5$ indicated all voxels to be significantly informative. However, using a strongly conservative threshold of $q = 10^{-4}$ again implied only right occipital voxels in the representation of stimulus-related information.

Figure 5.14 depicts the stimulus-related information of the $\Delta\alpha$ frequency power feature for all EEG data pre-processing methods. In comparison to the $sCDP_{P100}$ feature, the

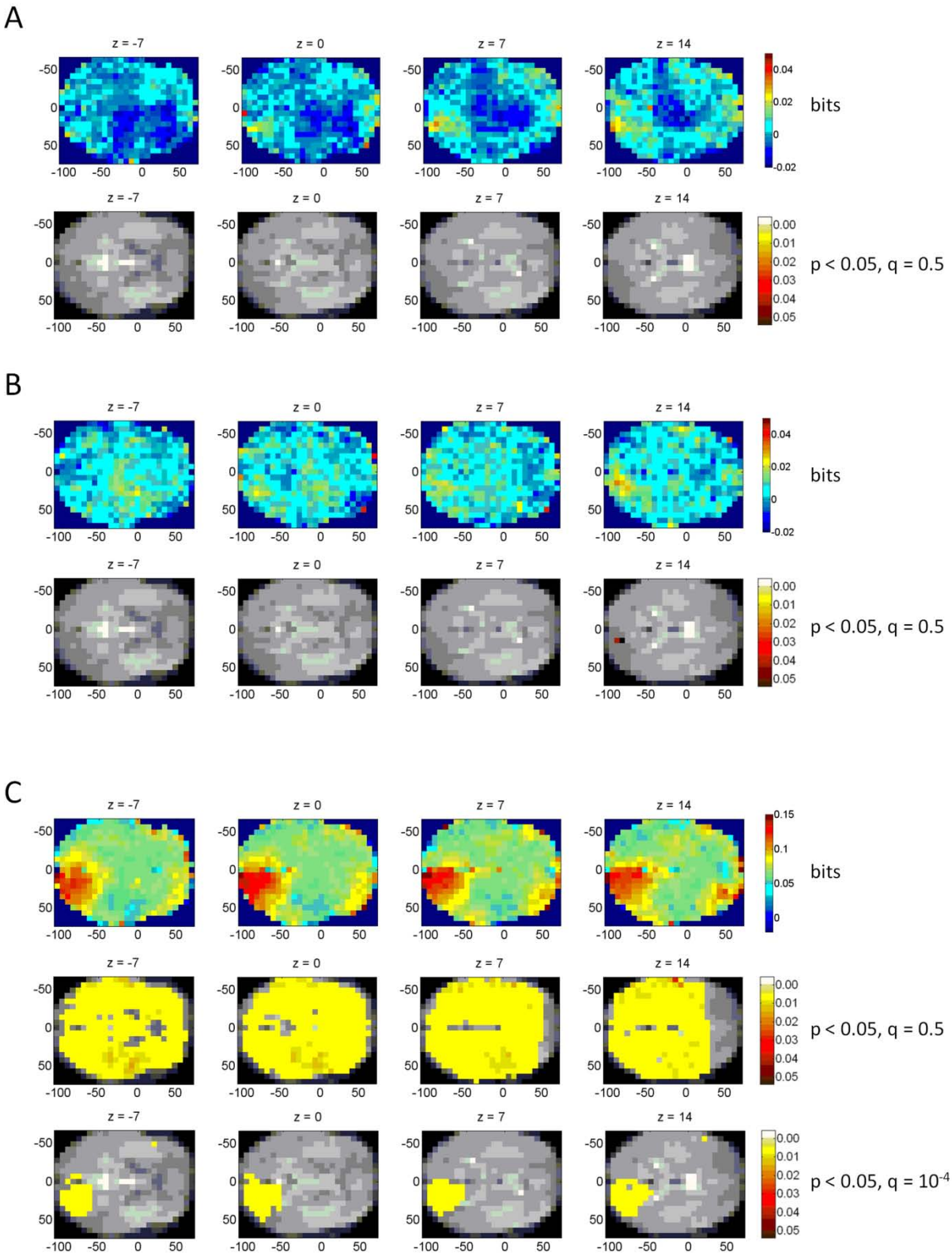


Figure 5.13 $I(S; sCDP_{P100})$: Stimulus-related information of the $sCDP_{P100}$ feature for all pre-processing methods. A) basic EEG pre-processing, B) ICA pre-processing, C) FSS pre-processing. The statistical maps are thresholded at p = 0.05 (uncorrected) and additionally at $q = 0.5$ and additional at $q = 10^{-4}$ in the case of C).

information estimates are lower and less topographically focussed on right occipital cortex. However, as for the time-domain feature, the frequency domain feature largest information estimates are observed for the FFS data pre-processing, with a number of voxels displaying statistically significant results when thresholded at $q = 0.5$.

Finally, Figure 5.15 displays the stimulus-related of the HRF_{max} feature. In comparison to the LORETA derived topographical EEG feature information maps, the informative voxels for the case of fMRI modality are much more focussed and located slightly more ventral in the right occipital cortex (Table 5.3). While for the common threshold of $q = 0.5$ no voxels show significant informativeness, relaxing the threshold to $q = 0.7$ yielded a cluster of significantly informative voxel in right occipital cortex (z = 0).

5.3.2.2 Topography of feature combinations across modalities

Figures 5.16 and 5.17 depict the stimulus related information estimates derived from the joint distributions of $sCDP_{P100}$, HRF_{max} and $\Delta\alpha$, HRF_{max}, respectively (A: Basic EEG, B: ICA EEG, C: FSS EEG). For the standard threshold of $q = 0.5$, all maps display large numbers of significantly informative voxel in a topographically unspecific manner. Using the annotated more conservative q-value thresholds, however, implicates mainly right occipital voxels in the representation of stimulus-related information. In comparison to their univariate equivalents, the estimated information quantities are larger (Table 5.3) and more voxels show significant results at lower FDR correction q-values.

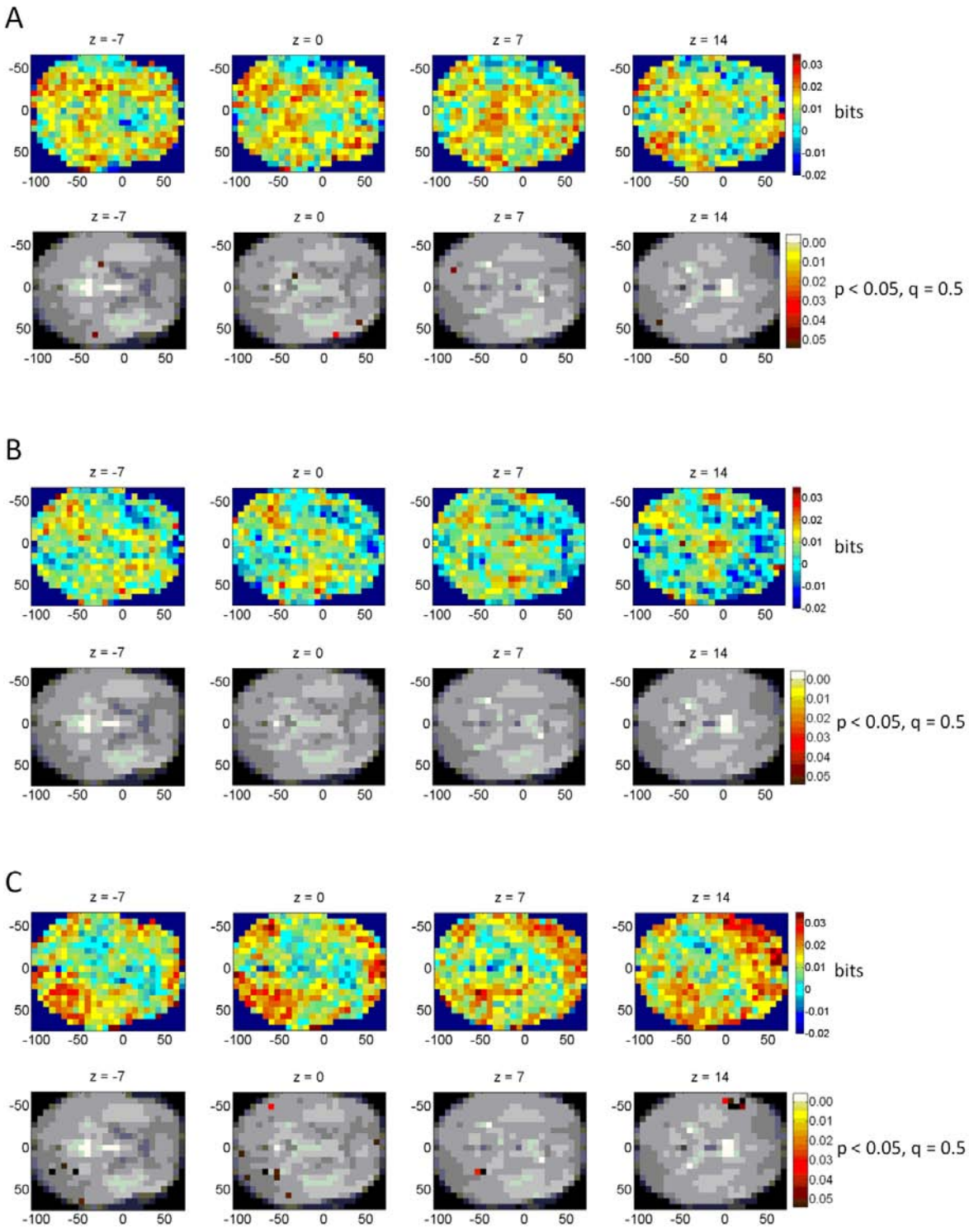


Figure 5.14 $I(S;\Delta\alpha)$ Stimulus-related information of the $\Delta\alpha$ feature for all pre-processing methods. A) basic EEG pre-processing, B) ICA pre-processing, C) FSS pre-processing. The statistical maps are thresholded at p = 0.05 (uncorrected) and additionally at $q = 0.5$ (FDR correction).

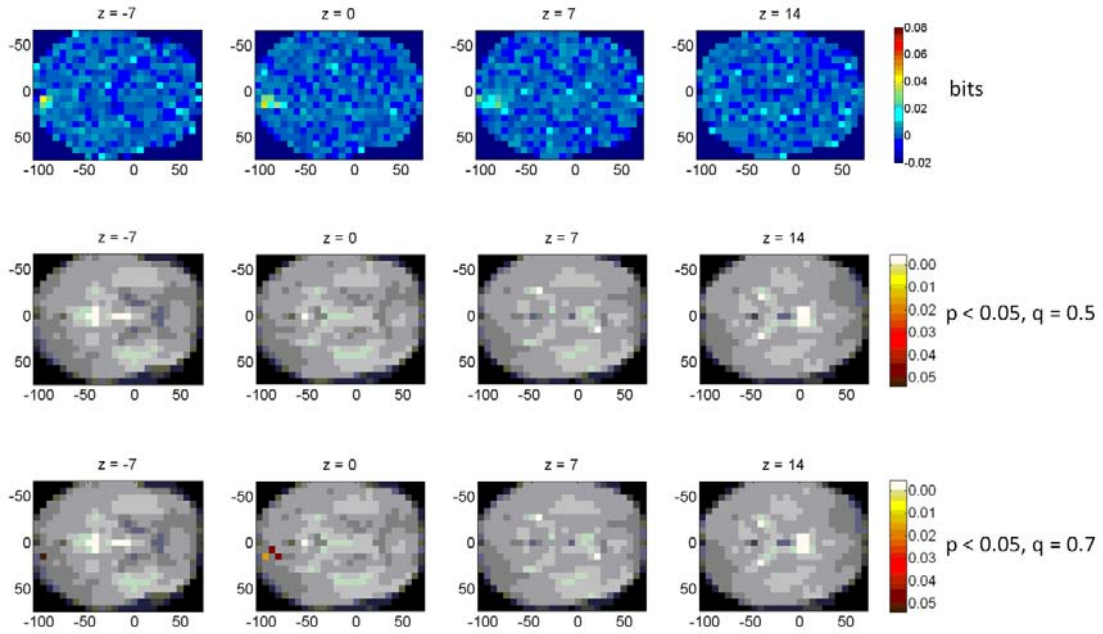


Figure 5.15 $\boldsymbol{I(S, HRF_{max})}$ Stimulus-related information of the $\boldsymbol{HRF_{max}}$ feature. The statistical maps is thresholded at p = 0.05 (uncorrected) and additionally at $\boldsymbol{q = 0.5}$ and $\boldsymbol{q = 0.7}$ (FDR correction).

For example, in the case of the bivariate combination of the basic and ICA pre-processed $sCDP_{P100}$ feature with the HRF_{max} feature, a large number of significantly informative voxels can be observed in right occipital cortex at $q = 0.08$, compared to very few statistically significant informative voxels at $q = 0.5$ in case of the univariate EEG features.

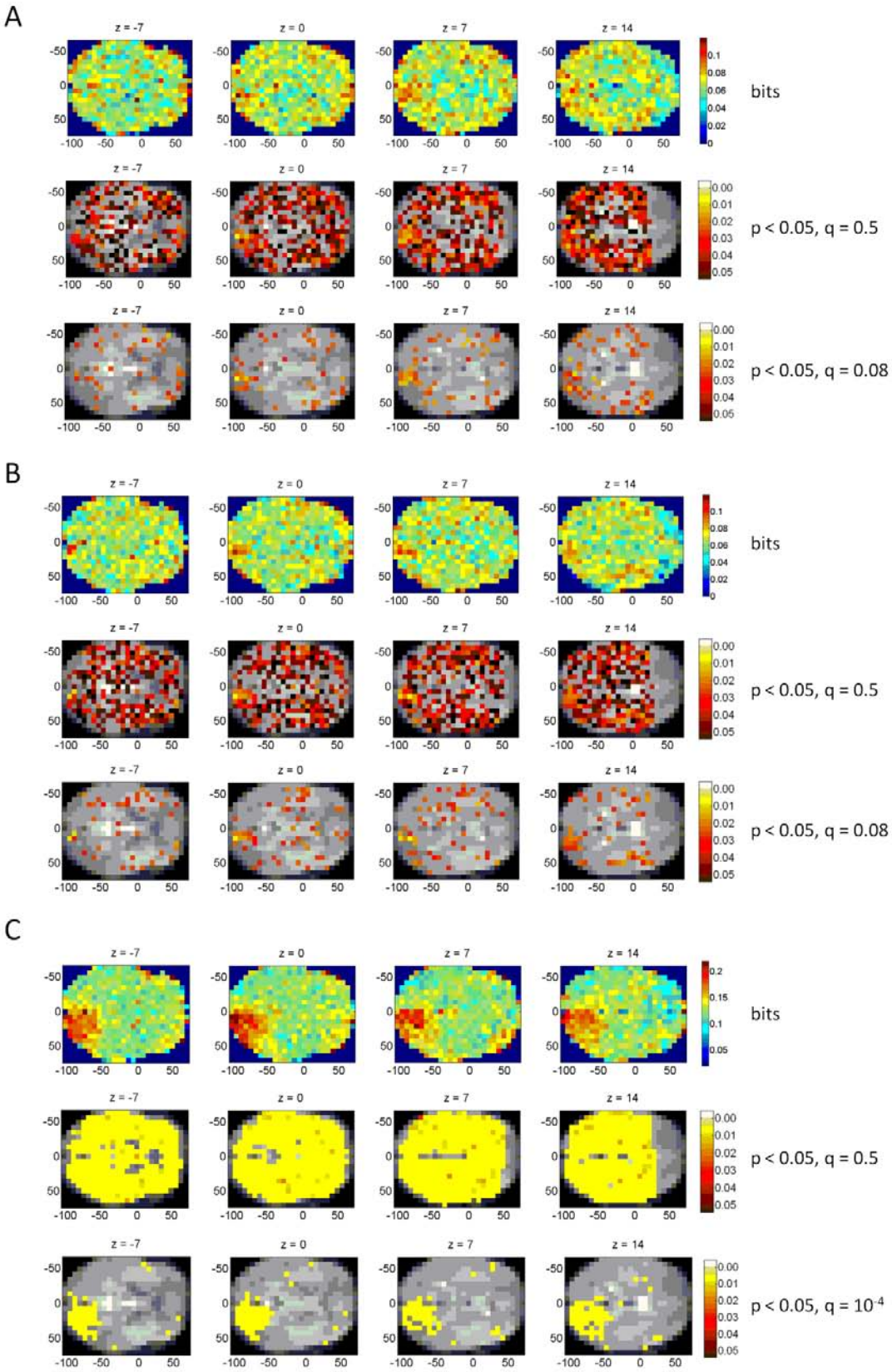


Figure 5.16 $I(S; sCDP_{P100}, HRF_{max})$ Stimulus-related information of the $sCDP_{P100}/HRF_{max}$ feature joint distribution for all EEG pre-processing methods. A) basic EEG pre-processing, B) ICA pre-processing, C) FSS pre-processing. The statistical maps are thresholded at p = 0.05 (uncorrected) / $q = 0.5$ (FDR correction) and additionally at $q = 0.08$, $q = 0.08$ and $q = 10^{-4}$, respectively.

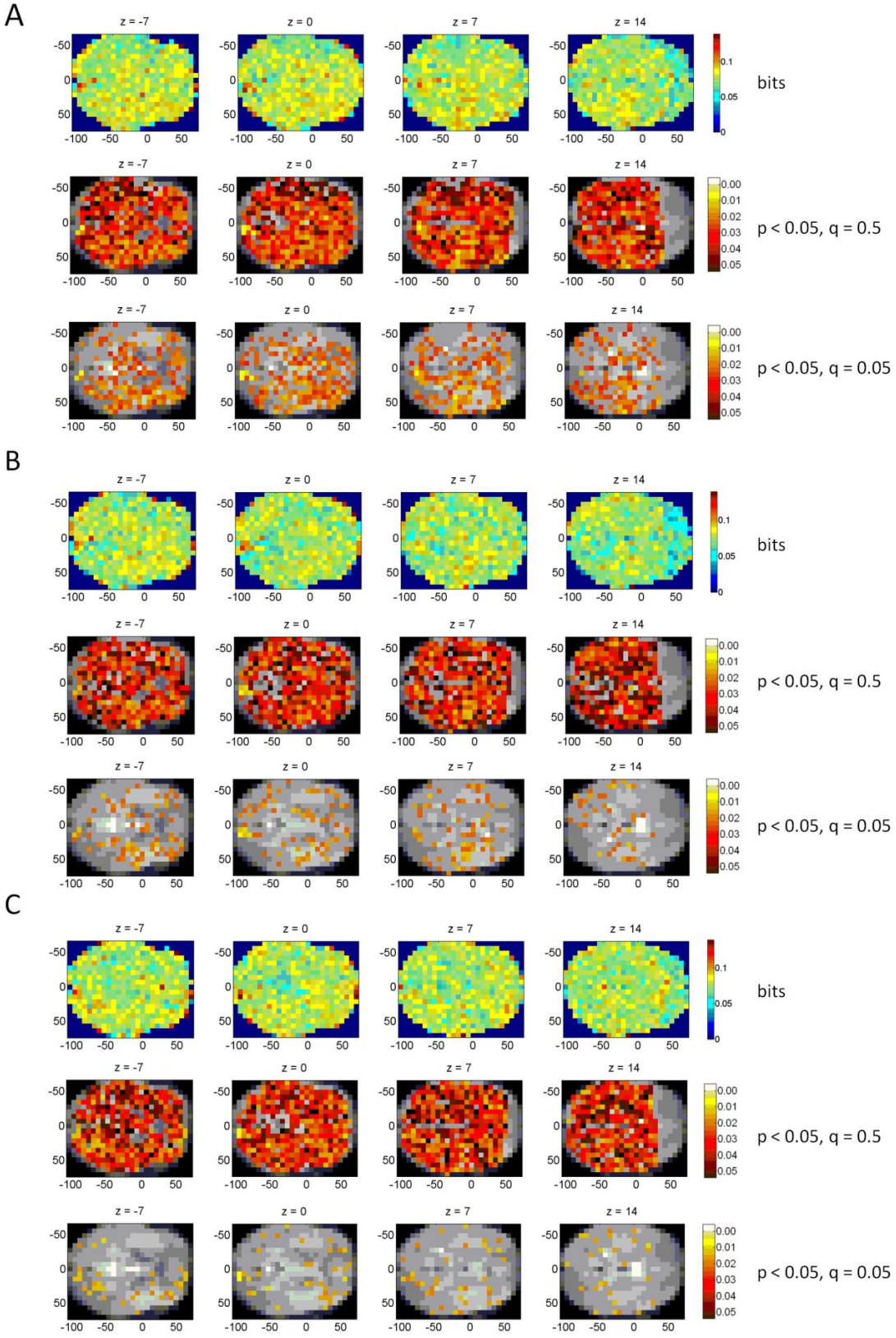


Figure 5.17 $I(S; \Delta\alpha, HRF_{max})$ Stimulus-related information of the $\Delta\alpha$ /HRF_{max}feature joint distribution for all EEG pre-processing methods. A) basic EEG pre-processing, B) ICA pre-processing, C) FSS pre-processing. The statistical maps are thresholded at p = 0.05 (uncorrected) / $q = 0.5$ (FDR correction) and additionally at $q = 0.05$.

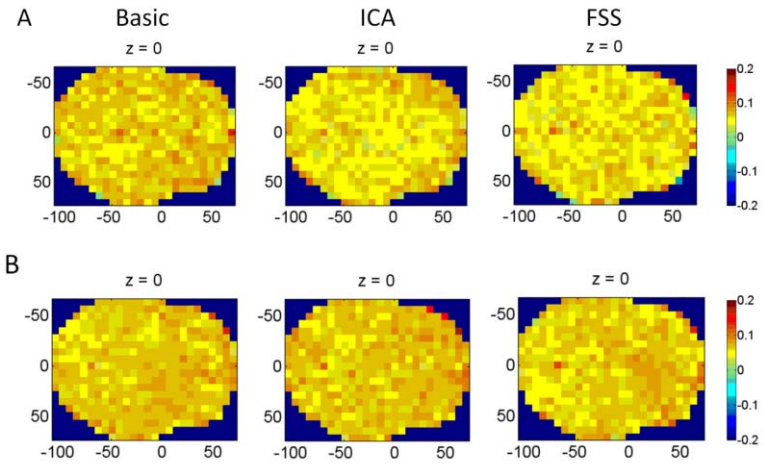


Figure 5.18 Stimulus-related information $Synergy$ between EEG and fMRI derived features. A) $Synergy$ between sCDP P100 and HRF features for all EEG pre-processing methods as indicated in the figure (columns). B) $Synergy$ between $\Delta\alpha$ and HRF for all EEG reprocessing methods as indicated in the figure (columns). As discussed in Chapter 4, the statistical evaluation of this measure is problematic, hence, it is omitted here.

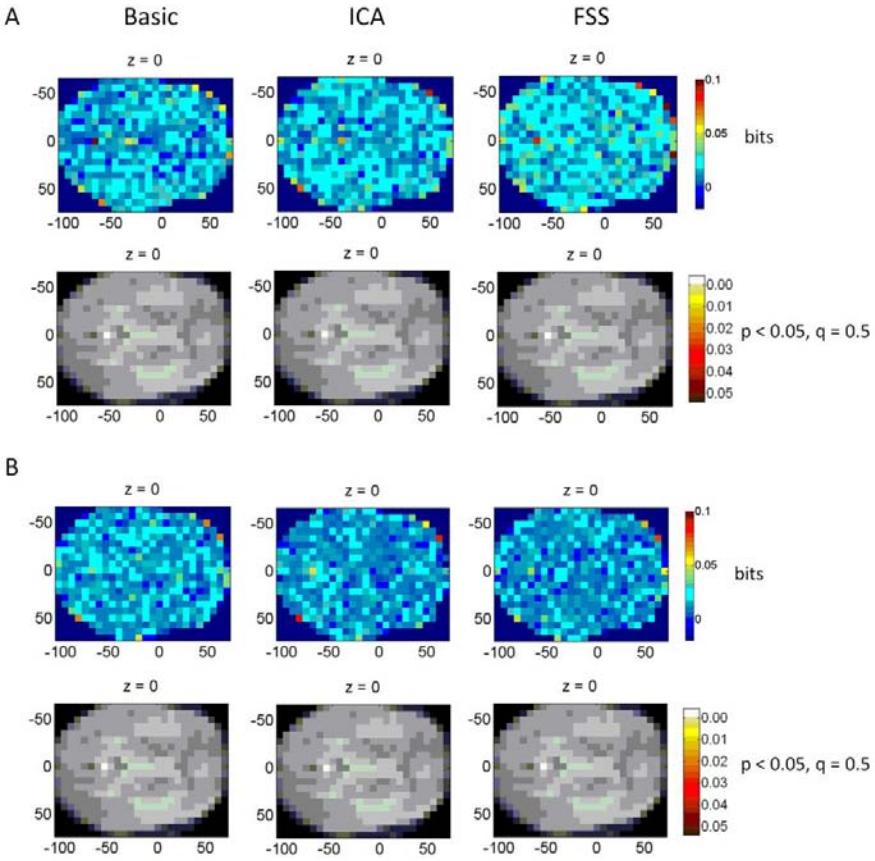


Figure 5.19 Activity dependence between EEG and fMRI derived features. A) $\boldsymbol{I(sCDP_{P100}, HRF_{max})}$ for all EEG pre-processing methods as indicated in the figure (columns). B) $\boldsymbol{I(S; \Delta\alpha, HRF_{max})}$ for all EEG reprocessing methods as indicated in the figure (columns). The statistical maps are thresholded at p = 0.05 (uncorrected) and additionally at $\boldsymbol{q = 0.5}$ (FDR correction).

Comparing the joint distribution information theoretic maps across EEG pre-processing methods indicates that for both time- and frequency-domain bivariate combinations, the information estimates are largest and most topographically focussed for the case of the EEG FSS data (Table 5.3). Overall, compared to the univariate EEG stimulus-related information feature maps, the combination with the fMRI derived feature results in a generally more spatially focussed cluster of most informative voxels, at least in the case of the observed effect sizes (Table 5.3).

Inspection of the $Synergy$ maps for all features and methods (Figure 5.18) revealed a spatially uniform pattern of slightly positive synergistic effects between both modalities. However, it should be noted that the observed $Synergy$ effects, rather than being due to real synergistic effects in the representation of the stimulus, may result from slightly inefficient estimation of information in the case of the univariate feature distributions, as the bias correction method employed was mainly optimized for the case of the bivariate distribution (Chapter 3, (Ostwald et al., 2010)). As a result, the presented maps for the univariate features might slightly underestimate the real univariate information, in terms of its absolute value. However, also in this case the topography (i.e. the spatial differences between the distributions of high and low informative voxels) of both univariate and bivariate information estimates will be preserved, as the $Synergy$ maps are spatially homogenous.

5.3.2.3 Topography of inter-modality comparisons: $\boldsymbol{I(R_1; R_2)}$ and $\boldsymbol{\langle I(R_1; R_2|S)\rangle_S}$

Figure 5.19 depicts the single-trial EEG-fMRI co-variation indexed by its activity dependence $I(R_1; R_2)$. For both features (Figure 5.19 A: $sCDP_{P100}$, B: $\Delta\alpha$) and EEG pre-processing methods (columns), the activity dependence is close to zero with a spatially uniform topography, hence no EEG-fMRI single trial co-variation in right-occipital cortex is

observed. The corresponding statistical maps are, FDR corrected at $q = 0.5$ confirm this null result.

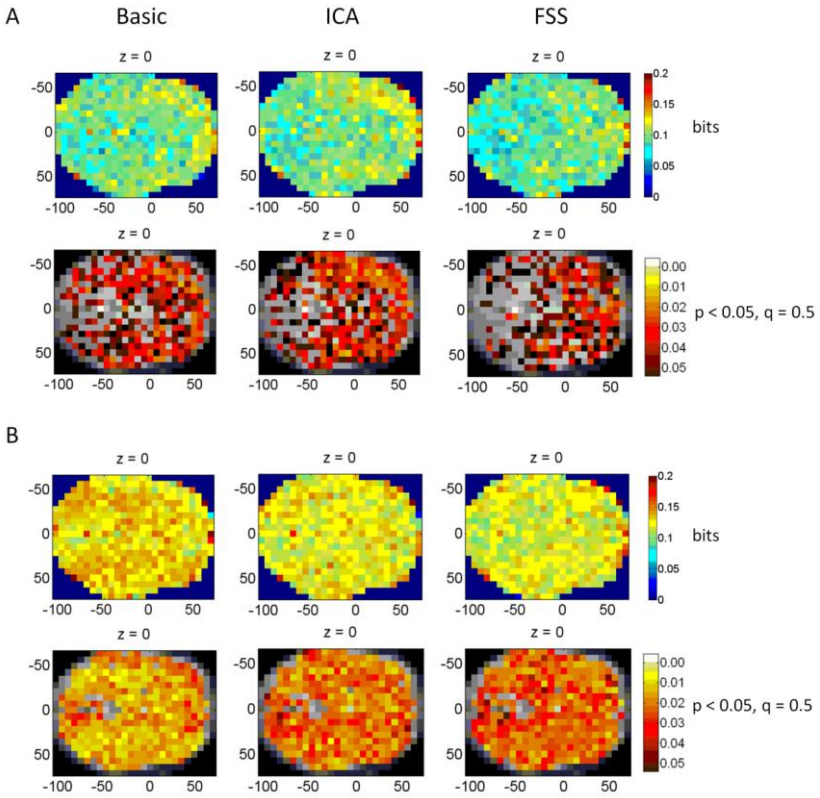


Figure 5.20 Conditional dependence between EEG and fMRI derived features. A) $\langle \boldsymbol{I}(\boldsymbol{sCDP_{P100}}, \boldsymbol{HRF_{max}}|\boldsymbol{S})\rangle_{\boldsymbol{S}}$ for all EEG pre-processing methods as indicated in the figure (columns). B) $\langle \boldsymbol{I}(;\Delta\boldsymbol{\alpha}, \boldsymbol{HRF_{max}}|\boldsymbol{S})\rangle_{\boldsymbol{S}}$ b for all EEG reprocessing methods as indicated in the figure (columns). The statistical maps are thresholded at p = 0.05 (uncorrected) and additionally at $\boldsymbol{q = 0.5}$ (FDR correction).

Figure 5.20 depicts the stimulus conditional single-trial EEG-fMRI co-variation indexed by its conditional dependence $\langle \boldsymbol{I}(\boldsymbol{R_1}; \boldsymbol{R_2}|\boldsymbol{S})\rangle_{\boldsymbol{S}}$. While the conditional dependence as the activity dependence does not indicate strong topographically focussed dependence between the EEG and fMRI features, it has to be noted that the relatively high information estimates for the entire scan volume are most likely due non-optimal bias control for this feature. First, it has to be noted that the estimation of the conditional dependence, due to the smaller number of experimental trials available for its estimation, is more bias-prone than that of the other reported ITQ. Secondly, the observed mismatch between $\boldsymbol{I}(\boldsymbol{R_1}; \boldsymbol{R_2}|\boldsymbol{S})$ and

$\langle I(R_1; R_2|S)\rangle_S$ clearly necessitates the future development of more efficient information estimation procedures for EEG-fMRI activity and conditional dependence.

	MNI X	MNI Y	MNI Z	Information (bits)
$\mathbf{I(S;\ sCDP_{P100})_{RAW}}$	28	-77	35	0.036
$\mathbf{I(S;\ sCDP_{P100})_{ICA}}$	28	-77	28	0.034
$\mathbf{I(S;\ sCDP_{P100})_{FSS}}$	0	-56	35	0.143
$\mathbf{I(S;\ \Delta\alpha)_{RAW}}$	56	14	0	0.035
$\mathbf{I(S;\ \Delta\alpha)_{ICA}}$	56	0	-7	0.028
$\mathbf{I(S;\ \Delta\alpha)_{FSS}}$	-49	-63	0	0.035
I(S;HRF)	14	-98	0	0.045
$\mathbf{I(S;\ sCDP_{P100},\ HRF_{max})_{RAW}}$	14	-98	0	0.116
$\mathbf{I(S;\ sCDP_{P100},\ HRF_{max})_{ICA}}$	14	-98	0	0.113
$\mathbf{I(S;\ sCDP_{P100},\ HRF_{max})_{FSS}}$	14	-98	0	0.230
$\mathbf{I(S;\ \Delta\alpha,\ HRF_{max})_{RAW}}$	14	-98	0	0.132
$\mathbf{I(S;\ \Delta\alpha,\ HRF_{max})_{ICA}}$	14	-98	0	0.118
$\mathbf{I(S;\ \Delta\alpha,\ HRF_{max})_{FSS}}$	14	-98	0	0.138
$\mathbf{I(sCDP_{P100};\ HRF_{max})_{RAW}}$	28	-7	7	0.043
$\mathbf{I(sCDP_{P100};\ HRF_{max})_{ICA}}$	7	14	14	0.045
$\mathbf{I(sCDP_{P100};\ HRF_{max})_{FSS}}$	-28	35	-7	0.051
$\mathbf{I(\Delta\alpha;\ HRF_{max})_{RAW}}$	56	7	14	0.036
$\mathbf{I(\Delta\alpha;\ HRF_{max})_{ICA}}$	-21	-77	35	0.037
$\mathbf{I(\Delta\alpha;\ HRF_{max})_{FSS}}$	-49	7	-21	0.033
$\mathbf{\langle I(sCDP_{P100};HRF_{max}\|S)\rangle_{RAW}}$	42	-42	21	0.144
$\mathbf{\langle I(sCDP_{P100};HRF_{max}\|S)\rangle_{ICA}}$	-42	49	7	0.150
$\mathbf{\langle I(sCDP_{P100};HRF_{max}\|S)\rangle_{FSS}}$	49	-14	-14	0.143
$\mathbf{\langle I(\Delta\alpha\ ;HRF_{max}\|S)\rangle_{RAW}}$	42	-84	-14	0.161
$\mathbf{\langle I(\Delta\alpha\ ;HRF_{max}\|S)\rangle_{ICA}}$	42	-84	-14	0.159
$\mathbf{\langle I(\Delta\alpha\ ;HRF_{max}\|S)\rangle_{FSS}}$	42	-84	-14	0.157

Table 5.4 Maximal informative voxel coordinates and information values for the different feature and method combinations employed, evaluated over the entire scan volume.

5.4 Discussion

In this Chapter, the voxel-wise evaluation of set of information theoretic quantities (ITQs) for the integration of EEG and fMRI data following their projection into a standardized anatomical space has been proposed. The framework was practically applied to a set of three data features, the standardized current density power equivalent of the P100 ($sCDP_{P100}$), the change in alpha power for a 1 s pre- and post-stimulus time window ($\Delta\alpha$) and the maximum of the post-stimulus haemodynamic response function (HRF_{max}) acquired under low and high contrast checkerboard stimulation. The analyses revealed topographically specific stimulus-related information for both the features marginal and joint distributions, for the EEG data predominantly under the use of more advanced pre-processing methods. The inter-modality information analyses remain inconclusive, with no spatial preference on the respective feature combinations activity or conditional dependence.

The approach introduced here has the advantage over the previously proposed use of an information theoretic approach to EEG-fMRI integration (Chapter 4) that it does not require a priori spatial constraints on data selection. Previously, data were selected from a single EEG electrode which the maximum VEP and fMRI voxels contained within a sphere centred on the maximum response in a standard GLM analysis. This contains an implicit, strong assumption regarding the spatial correspondence between EEG and fMRI, namely that there is an underlying relationship between the electrical data recorded over the occipital region and the haemodynamic data recorded within visual cortex. While this assumption is at first sight reasonable, it is not clear that it necessarily holds true when examined at the level of single trial data. As well as issues directly related to neurovascular coupling at the macroscopic level, the signal recorded by a scalp electrode contains contributions from a relatively wide area of cortex (Gloor, 1985) not all of which is likely to have a detectable haemodynamic response (Nunez and Silberstein, 2000). By incorporating EEG source

localisation into the information theoretic framework, the effect of these issues is reduced, potentially allowing a more principled investigation of the relationship between EEG and fMRI. The results presented in this study demonstrate that such an approach is feasible and that the cortical regions known to be involved in the simple task we employed are highlighted as containing stimulus-relevant information.

For the EEG time-domain, the current study analysed the same data as discussed in Chapter 4, with the only difference being pre-processing and feature selection (i.e. based on electrode ICA data in Chapter 4 and using basic, ICA and FSS EEG pre-processing and LORETA in the current study). Perhaps unsurprisingly given the different input data, there are some discrepancies between the studies in terms of magnitude and behaviour of the calculated ITQs. In particular, it was previously found that the fMRI data were more informative than the EEG data regarding the stimulus contrast, whereas the reverse was found in the current study for the FSS pre-processed data. There are two probable, complementary explanations for this. On the one hand, the FSS has previously been shown to greatly improve the quality of data acquired in the MRI scanner (Porcaro et al., 2010). In addition, in the previous study we extracted fMRI data from voxels selected on an individual subject basis as having the highest stimulus response in a GLM analysis. In the current analysis, fMRI data were extracted from coordinates defined by the group activation, which is likely to lead to less optimal data regarding stimulus informativeness. (Essentially, the data selection approach in Chapter 4 followed the functional localizer approach to fMRI data analysis (Saxe et al., 2006), while the data selection approach in the current chapter followed the whole-brain search approach (Friston et al., 2006). The former strategy assumes that topographically optimal between-subject correspondence of data selection is established by selecting the subject-specific data optima, while the latter strategy assumes that topographically optimal between-subject correspondence is established by spatial normalization of the data. The former has the

advantage, that the data entering the group summary statistic are optimized at the single subject level, with the disadvantage that this single-subject optimization is relatively subjective. The latter approach has the advantage that the data entering the group summary statistic is objectively determined (within the constraints of the spatial normalization procedure) and readily reproducible, while it is likely to result in slight misalignments of the subject-specific optimal data. However, for the current chapter, only the latter allowed to achieve the aim of the study, namely a group-level voxel-wise integration of EEG and fMRI data). The issue here is one of inter-subject spatial heterogeneity in the fMRI response, although the low spatial resolution used in the current study should reduce its effect on the calculation of ITQs. The improvement in EEG data quality using FSS will certainly lead to improved calculation of ITQs, as demonstrated previously for the comparison between raw and ICA-processed data (Porcaro et al., 2010).

The results regarding activity dependence indicate that, at the level of the false discovery rate threshold used, there were no brain regions in which the P100 amplitude was informative about the HRF amplitude. This is in accordance with previous work, where the absolute values of activity dependence were low and not significantly different to zero (Whittingstall et al., 2007). This negative result highlights the need for a more detailed investigation of the EEG signal space to identify features which carry information about the haemodynamic response. The spatiotemporal framework outlined here allows this issue to be addressed explicitly, to identify when and where the EEG signal is related to the fMRI signal.

It is noted that the estimation of the conditional dependence across the whole-brain volume appears uniformly positively biased, yielding a uniform pattern of statistical significance at the current false-discovery threshold. This result is most likely due to non-optimal information estimation and bias correction. As discussed in Chapter 7, the application

of Gaussian mixture models for the estimation of information from continuous, analogue brain signals might enable optimized estimation of this quantity

Two recent studies have proposed approaches to combined EEG-fMRI data with some similarities to the scheme proposed here (Esposito et al., 2009; Martuzzi et al., 2009). However, while both previous studies propose the projection of EEG electrode data into an anatomical space in alignment with the fMRI data, they differ with respect to how the integration of the two modalities is achieved. Specifically, in (Martuzzi et al., 2009) the authors use a different approach to estimate intracranial source activations, in their case estimated intracranial local field potentials (eLFPs) based on the ELECTRA algorithm (Grave de Peralta et al., 2000; Grave de Peralta et al., 2004). In order to integrate data from both modalities, the authors then compute a similarity index between the statistical activation maps obtained from EEG and fMRI data analysis for (occipital) regions of interest. While this is a sensible approach to compare the outcome of the application of statistical procedures to data from both modalities, the information theoretic framework allows a more direct spatial evaluation of the information provided by the joint observation of the two modalities with respect to the stimulus in the quantities $I_N(S; R_1, R_2)$ and $Synergy$. In (Esposito et al., 2009) the authors use another projection of EEG sensor data into anatomical space, namely a cortically constrained distributed inverse model based on the minimum L2-norm approach (Dale et al., 2000). To integrate EEG and fMRI data, the authors then extract single-trial features from the estimated source time-courses and enter these into a mass-univariate GLM analysis. Again, the procedures proposed in the current study are more direct, in that they assume a voxel-by-voxel correspondence of the reconstructed EEG and fMRI signal. However, the general approach of integrating EEG-fMRI data by working in a standard, common space appears to be feasible and have some advantages over more restricted methods.

There are a number of ways in which the procedures described here can be improved in future research. First, as discussed in above a bias correction approach tailored to the peculiarities of EEG and fMRI data is desirable. Second, computational constraints limited the analysis to a resolution of 7 mm isotropic, but this is a purely practical issue. Higher resolution analyses will be possible with improved hardware and potentially distributed processing. However, it should be noted that given the relatively low resolution of EEG source localisation, it is debatable whether a considerably higher spatial resolution would be beneficial. Last, there a considerable number of inverse solutions in addition to LORETA, as well as additional improvements to be gained by the use of individual geometry in the forward model (Fuchs et al., 1998; Fuchs et al., 2001). Improvements in the spatial accuracy of the EEG source localisation might help to optimise the proposed information theoretic analysis scheme, while inverse solutions more amenable to the localisation of spectral data, such as beamforming (Hillebrand and Barnes, 2005), would allow an additional level of inter-modality comparison.

It should be noted that the currently proposed analysis scheme for simultaneous EEG-fMRI data embodies the assumption that neural activity changes (as reflected in the EEG) and BOLD activity changes spatially co-localize in a bijective manner. Accounting for partial volume effects, this co-variation is assumed to be expressed within approximately a 10 mm diameter sphere. While this assumption appears reasonable (Logothetis et al., 2001) a potential extension of the current work would be to estimate information theoretic quantities for all possible voxel combinations across brain-space (i.e. to include off-diagonal elements in a matrix spanned by the vectorized EEG and fMRI MNI spaces) in the analysis. While increasing the computational demand, such an analysis should converge in finite time. A potential data-analytical challenge would be to effectively correct for false-positives in this square matrix, also accounting for local co-variation and the only quasi-Gaussian nature of the

information theoretic quantity distributions, which prevent a straight-forward application of Gaussian random field theory models (Friston, 2007). Finally, a further challenge would be the interpretation of these voxel-by-voxel information theoretic matrices with respect to measures of functional and effective connectivity.

In conclusion, the feasibility of integrating EEG-fMRI data within an information theoretic framework across the entire three-dimensional brain space and at the group level was demonstrated. Regions demonstrating significant stimulus-related information were anatomically well localised with both data acquisition modalities. No regions were identified in which the EEG features were activity dependent regarding the amplitude of the HRF, corroborating the previous results at the electrode level (Chapter 4) The formalism described here provides a methodology for identifying spatiotemporal or spatiospectral relationships between EEG and fMRI in terms of information content which utilises all of the available EEG data, rather than that selected a priori from a single electrode. This approach may help to uncover a more detailed picture of the regional co-variation between electrical and haemodynamic measures of brain function.

6 Information theoretic characterization of the human perceptual decision system

6.1 Introduction

The metaphor for the human brain at the beginning of the 21st century is that of a dynamic, information processing device (Gazzaniga MS, 2004). By means of its neural activity, the brain is thought to represent information about the external state of the world, internal expectancies about incoming sensory information, as well as the formation and execution of decisional processes (Friston, 2010). A first step in understanding how neural activity can represent this variety of processes is to quantify the spatio-temporal dynamics of information representation in the brain, i.e. to answer questions about the *where* and *when* of information encoding about *which* aspects of the external and internal states of the world. It is generally believed that knowledge about the information-carrying features of neuronal activity will lead to a better understanding of the dynamical principles that underlie brain function.

Perceptual decisions are arguably one of the core cognitive functions (Gold and Shadlen, 2007) and can be defined as the selection of one among a set of possible interpretations of a sensory event (Heekeren et al., 2004; Heekeren et al., 2006; Heekeren et al., 2008; Philiastides and Sajda, 2007). Previous research indicates that perceptual decisions are based on the accumulation of sensory evidence over time (Ditterich, 2006; Ratcliff and McKoon, 2008). Electrophysiological research in primates has lead to the concept of neural integrators thought to implement such accumulative processes (Ratcliff et al., 2009; Shadlen and Newsome, 2001), while functional brain imaging studies have identified areas potentially involved in the processing of perceptual decisions (Heekeren et al., 2004; Heekeren et al., 2006; Mulert et al., 2008; Philiastides and Sajda, 2007). However, these areas, here referred to as the human perceptual decision system (Figure 6.1 A), have as yet not been characterized

in a principled manner with respect to the information they represent about external and internal decision variables, and their roles in the decision process remain largely elusive (Heekeren et al., 2008).

A theoretical framework that has been applied successfully in the invasive, electrophysiological study of information representation in animal models is that of information theory (Magri et al., 2009; Panzeri et al., 2008; Schneidman et al., 2003). Information theory, and its core quantity of mutual information, allows probabilistic inferences about which neuronal activity features discriminate between external, internal or behavioural variables in their single-trial responses. As such, it has high ecological validity, because the brain has to successfully decide on appropriate responses to incoming sensory information based on neural activity responses to single events, rather than averages (Quian and Panzeri, 2009).

Using a recent adaption of this framework for the analysis of combined EEG and fMRI data (Ostwald et al., 2010), the aim of the current study is to study the representation of information for perceptual decisions in the human brain at both high temporal and spatial resolution. Observers performed a visual perceptual decision task (face vs. car categorization), while in a full-factorial design stimulus informativeness and the observer's spatial attention were manipulated allowing the representation of external and internal state variables to be discerned. Further, measuring combined EEG-fMRI and psychophysical data allows the characterisation of the relative roles of topographical and temporal signal features in the representation of information for perceptual decisions. Specifically, upon identification of spatial (fMRI) and temporal (EEG) signal features of interest, the informativeness of the joint and marginal probability distributions of these features with respect to the experimental variables, whether these are related to the stimuli or the subject's behaviour, is assessed. While the relationship between electroencephalographic and haemodynamic signatures and

their correspondence to invasively recorded electrophysiological phenomena remains elusive, the question of EEG-fMRI co-variation is not a focus of this chapter.

Recent work in the brain imaging field has underlined the importance of single-trial as opposed to average data analyses (Debener et al., 2007b; Makeig et al., 2004). For example, recent approaches to EEG-fMRI integration by prediction capitalize on single-trial linear co-variations of the EEG and fMRI data features (Benar et al., 2010; Debener et al., 2006). However, simultaneous EEG-fMRI recording has detrimental effects on EEG data quality. While recent approaches allow the MR scanner induced EEG artefacts largely to be corrected, EEG data quality remains lower than for EEG recordings outside the MR environment. Given the subtle nature of the expected EEG effects, in addition to combined EEG-fMRI data acquisition, EEG data were also acquired for the same paradigm outside the MR environment. This allowed a principled investigation of the advantage of performing combined EEG-fMRI recordings, as well as determining the effect of poorer quality EEG on the quantities calculated.

In sum, the aim of the current study is to bring together the advances in combined EEG-fMRI recordings during performance of an ecologically meaningful cognitive task with the perspective provided by an information theoretic approach, to study the single-trial informativeness of the signals' joint distributions about a range of experimental variables of interest. The results of the analyses reported here, like any data-driven methodology, do not offer a mechanistic description of the decision process, however, they yield a data-driven perspective for future formulations of whole-brain spatio-temporal models of the neurobiological underpinnings of perceptual decisions.

All custom written Matlab (The Mathworks, Natick, MA) code used in this study is available from http://www.buic.bham.ac.uk/EEG_FMRI_ITQ/EEG_FMRI_PD_Analysis.zip and the data is available from the author upon request.

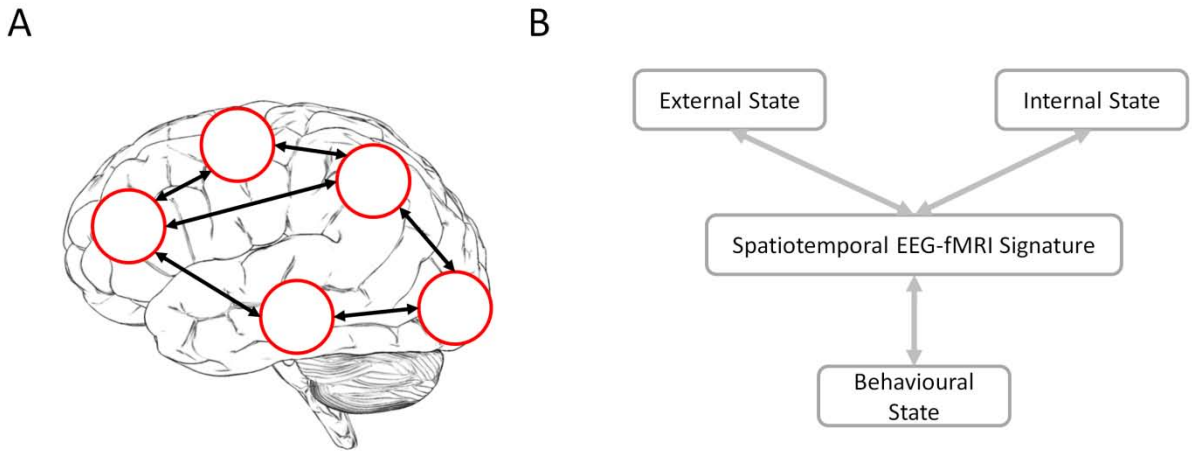


Figure 6.1 Conceptual framework of Chapter 6 A. Human perceptual decision system. Brain imaging studies have implicated a number of brain regions in the processing of visual perceptual decisions. These areas include primary sensory (striate cortex), intermediate (lateral-occipital, temporal, parietal) and higher (frontal, pre-frontal) association cortices. While known to be involved in the perceptual decision process, their respective functional roles in the dynamical representation of information remains elusive B. Variables of interest. A system view of the human brain implicates neuronal processes in the representation of information about external states (stimuli), internal states (attention, memory), as well as behavioural states (outcome of a perceptual decision). In the current study, these variables have been operationalized as stimulus category and informativeness (external state), spatial attention/prioritization (internal state) and the observer's decision and response time on a given experimental trial (behavioural state). The method of choice for studying the neuronal representation of information about these variables non-invasively in the human brain with both high temporal and spatial resolution is combined EEG-fMRI.

6.2 Methods

6.2.1 Subjects

Seventeen subjects (8 female, mean age 25.9 years, range 20 – 33 years, 2 left-handed) were recruited from the University of Birmingham campus and paid for their participation. All observers had normal or corrected to normal vision, no history of neurological disorders and gave written informed consent. The study was approved by the local ethics committee.

Complete data sets (2 experimental runs of EEG alone, 5 experimental runs of simultaneous EEG-fMRI) were acquired from 13 of the 17 subjects (for two of the four

subjects who did not have complete data, incomplete EEG data were recorded outside the MR scanner, for one of the four incomplete psychophysical data were recorded inside the MR scanner, while for other subject the EEG data recorded inside the MR scanner were strongly contaminated by movement artefacts of the reference electrode due to contact with the head-coil). All information theoretic analyses (outside and inside MR scanner EEG data, fMRI data, and EEG-fMRI data) reported are based on the 13 complete data sets. To identify regions of interest for the information theoretic analyses with maximum detection power, a total of 16 fMRI data sets were included in the GLM-RFX analysis. One subject's fMRI data was excluded from the GLM analysis because no psychophysical data could be recorded.

6.2.2 Experimental design and paradigm

In a 2 x 2 factorial within-subject design, observers performed a perceptual decision task, similar to that described in (Philiastides et al., 2006; Philiastides and Sajda, 2006; Philiastides and Sajda, 2007). On each trial of the experiment, a visual stimulus either depicting a face or a car was presented in one visual hemifield (left/right eccentricity of stimulus centre 11 degrees of visual angle, stimulus extension 9 degrees of visual angle) for 200 ms and the observer was asked to indicate via a button press, whether a face or a car stimulus was presented. For the button presses, observers used their right index and middle finger for the two categories, and the mapping from stimulus category to response button was counterbalanced across observers. The informativeness of the visual stimulus was manipulated by altering the phase coherence of its spatial frequency spectrum resulting in low and high informative trials (see below and Figure 6.2A). On half of the trials, a cueing arrow shown continuously for 1 s prior to the stimulus indicated in which hemifield the stimulus

would be presented. The observers were asked to allocate their spatial attention to the respective hemifield, while maintaining steady fixation (spatial prioritization condition). On the other half of the trials, the two-headed cueing arrow was un-informative and the stimulus was presented randomly in either hemifield (no spatial prioritization condition). Face and car stimuli were equally distributed over the four experimental conditions. Observers were asked to respond as quickly and accurately as possible, with particular emphasis on speeded responses and the maintenance of fixation on a central fixation cross throughout the experiment. Analyses of eye-movement data (Appendix H) obtained during the combined EEG-fMRI data acquisition sessions indicate that good fixation was achieved by the observers. The order of stimulus presentation was randomized

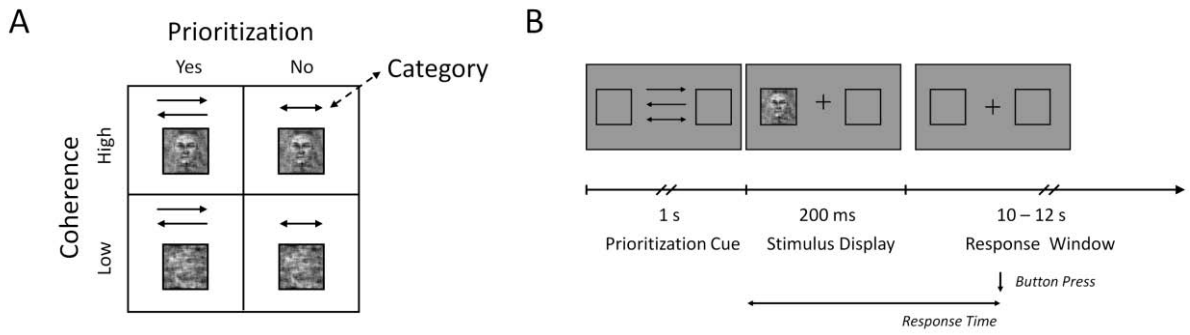


Figure 6.2 Experimental Design and Paradigm. A. 2 x 2 factorial experimental design with factors informativeness (high, low) and spatial prioritization (yes, no). On each trial of the experiment, the observer is presented a face or car stimulus, which has been manipulated according to visual informativeness and the observer is prompted to either spatially prioritize the stimulus display or not. The stimulus category (face or car), which the observer is asked to discriminate, is manipulated orthogonally to the other factors. B. Single experimental trial outline. Prior to the presentation of the stimulus, either a one-headed arrow indicated the hemifield of the subsequent stimulus presentation, or a two-headed arrow was uninformative in this respect. The cueing arrow was shown continuously for 1 s pre-stimulus, the stimulus itself for 200 ms. The observer was asked to respond as fast and as accurate as possible with no restrictions on the response window. The inter-trial interval was 0-300 ms for the EEG only and 10-12 s for the combined EEG-fMRI recordings.

For the EEG only recordings, data from 72 trials for each of the four conditions (half of them face stimuli) were recorded with an inter-trial stimulus randomized between 0 – 300 ms. The data acquisition was split into two experimental runs of approximately 10 min each. For the combined EEG-fMRI recordings data from 90 trials for each of the four conditions

(half of them face stimuli) were recorded with an inter-trial stimulus discretely randomized between 10 s and 12 s (5 or 6 TRs). This long inter-trial stimulus was chosen to obtain reliable recordings of single-trial haemodynamic responses. The 90 trials were split into five experimental runs, each lasting approximately 14 minutes. Prior to the EEG recordings the observers also completed two practice runs to familiarise themselves with the task.

6.2.3 Stimuli

The stimulus set consists of 18 pictures of cars and 18 pictures of faces, similar to the stimulus set used in (Philiastides et al., 2006; Philiastides and Sajda, 2006; Philiastides and Sajda, 2007). The car images were obtained from http://liinc.bme.columbia.edu/mainTemplate.htm?liinc_downloads.htm while the face images were obtained from the Max Planck face database (Troje and Bulthoff, 1996). The image categories were matched for the number of frontal, and left and right lateral views. The full original stimulus set is shown in Figure 6.3.

Figure 6.3 Original stimulus set. The upper three rows display face stimuli, the lower three rows car stimuli.

All images were converted to bitmap format (.bmp) and the corresponding 256 x 256 matrices saved with 8 bit depth. The two stimulus sets were matched for their mean driving luminance and contrast as assessed by a one-way ANOVA with factor 'image category' and levels 'face' and 'car' (mean driving luminance: $F_{(1,34)} = 0.08$, $p = 0.78$, contrast: $F_{(1,34)} = 0.22$, $p = 0.64$)

To manipulate the informativeness of the images, the spatial phase spectra were linearly weighted with a phase spectrum of a Gaussian noise image using the weighted mean phase technique as described in (Dakin et al., 2002). With the original phase of an image given by ϕ_{image}, the final phase ϕ_{final} was computed as follows:

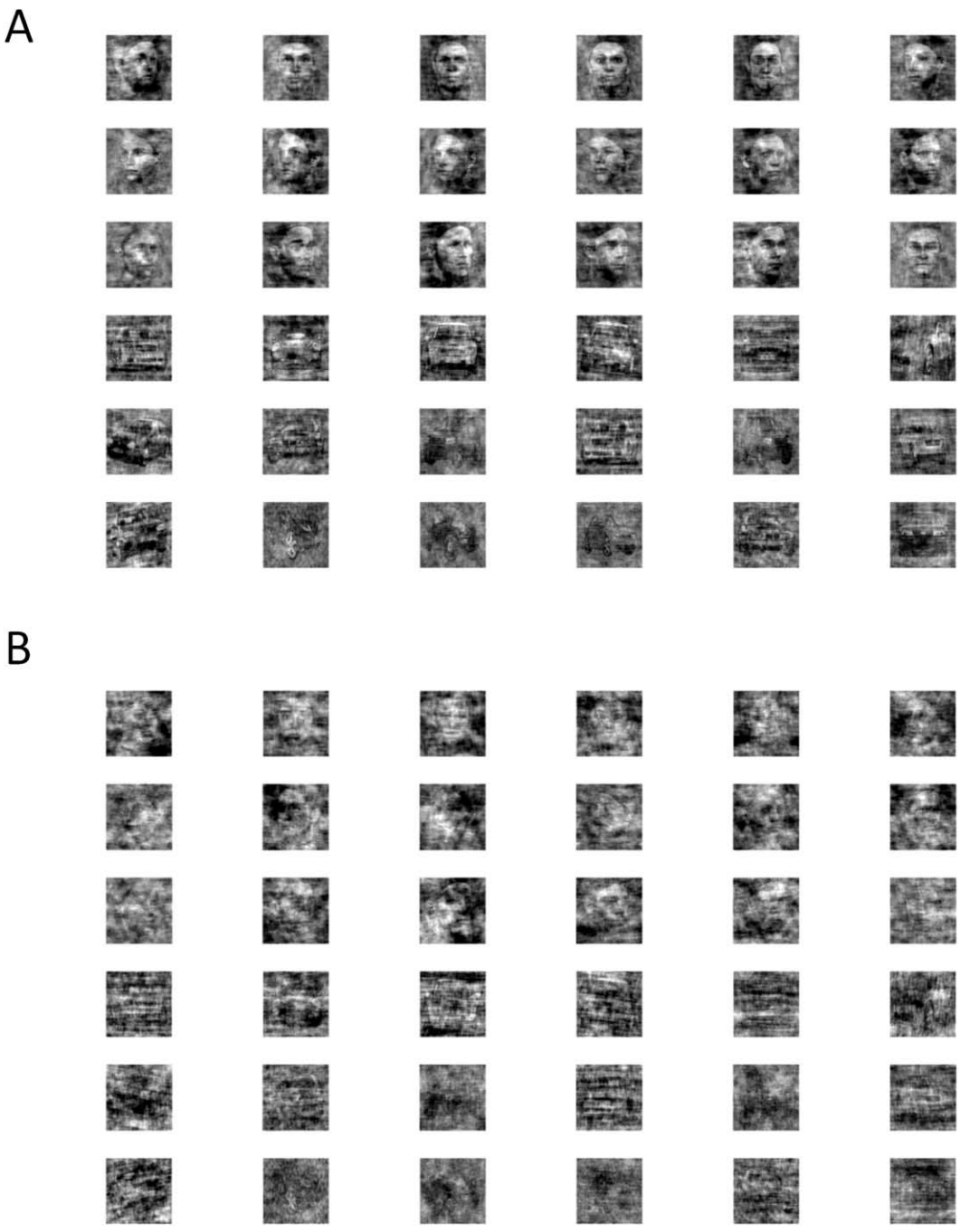
$$\phi_{final} = \begin{cases} \arctan(S/C) & C > 0 \\ \arctan(S/C) + \pi & C < 0,\ S > 0 \\ \arctan(S/C) - \pi & C < 0,\ S < 0 \end{cases} \tag{6.1}$$

where

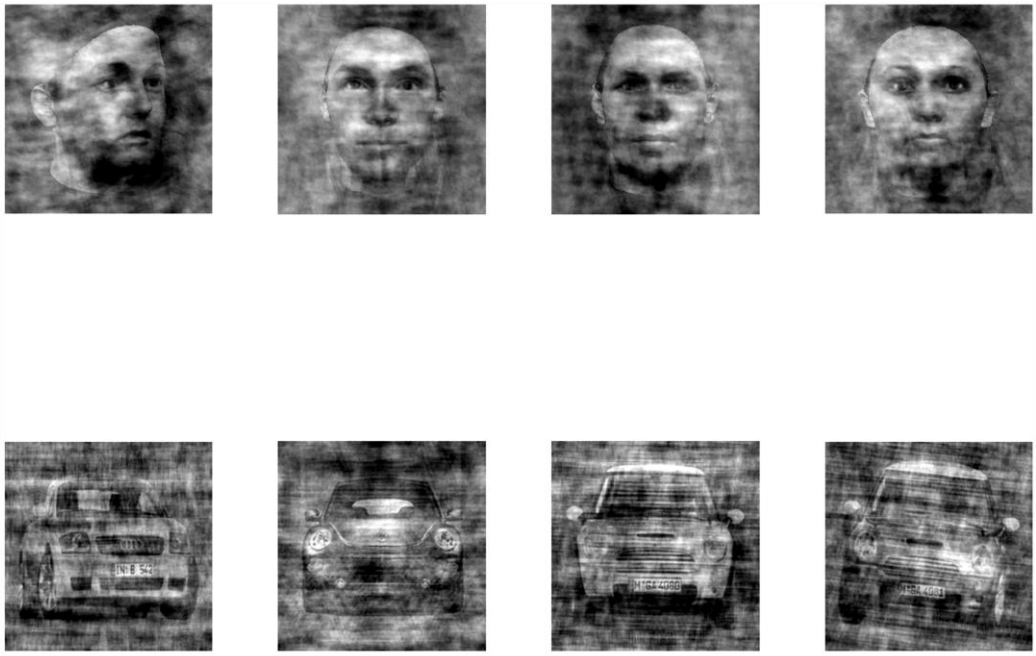
$$S = w \sin(\phi_{image}) + (1 - w) \sin \phi_{noise} \tag{6.2}$$

$$C = w \cos(\phi_{image}) + (1 - w) \cos \phi_{noise} \tag{6.3}$$

and ϕ_{noise} is the phase of uniform random noise and $w \in [0,1]$ is the signal-to-noise weighting coefficient. Based on a psychophysical pilot study (Appendix I), stimuli of weighting coefficients $w_1 = 0.90$ (high informativeness) and $w_2 = 0.50$ (low informativeness) were chosen for the experiment in order to elicit reliable differences in the response times for either stimulus class, while still allowing accurate performance of the task. The full phase coherence manipulated stimulus sets for high and low informativeness are depicted in Figure 6.4.


A
B

C



D

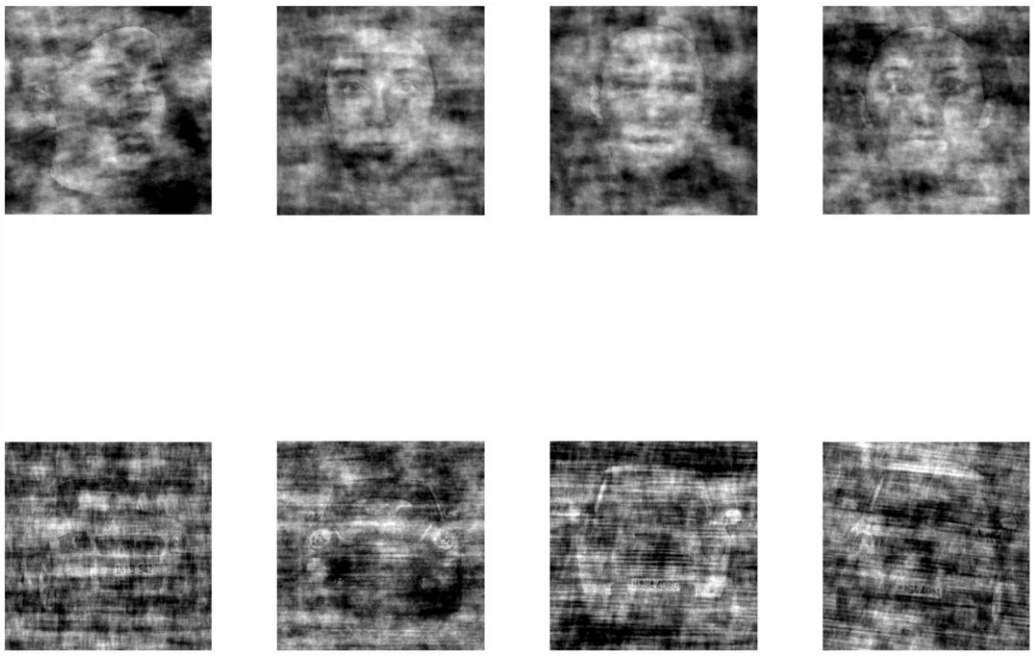


Figure 6.4 A. High informativeness stimulus set, $\boldsymbol{w_1 = 0.90}$ The upper three rows display face stimuli, the lower three rows car stimuli B. Low informative stimulus set, $\boldsymbol{w_2 = 0.50}$. The upper three rows display face stimuli, the lower three rows car stimuli. C High informativeness stimulus set selection at larger scale D. Low informative stimulus set selection at larger scale,, $\boldsymbol{w_2 = 0.50}$.

6.2.4 Data acquisition

EEG data were recorded using a 64 channel MR compatible EEG system (BrainAmp MR Plus, Brain Products, Munich, Germany), which incorporates current limiting resistors of 5 kΩ at the amplifier input and in each electrode. The EEG cap consisted of 62 scalp electrodes distributed according to the 10-20 system (Jasper, 1958) and two additional electrodes, one of which was attached approximately 2 cm below the left collarbone for recording the ECG, while the other was attached below the left eye (on the lower orbital portion of the orbicularis oculi muscle) for detection of eyeblink artefacts. Data were sampled at 5000 Hz. Impedance at all recording electrodes was less than 20 kΩ. For simultaneous EEG-fMRI recordings, the EEG data acquisition setup clock was synchronised with the MRI scanner clock using Brain Product's SyncBox, resulting in exactly 10,000 data points per EPI-TR interval (see details of the fMRI sequence below). The EEG set-up was identical for the outside and inside MR scanner recordings. In the following, the EEG data set recorded outside the MR scanner will be referred to as EEG, while the EEG data set recorded simultaneously with the fMRI data will be referred to as EEG-fMRI data set.

The simultaneous EEG-fMRI experiment was conducted at the Birmingham University Imaging Centre using a 3T Philips Achieva MRI scanner. An initial T1-weighted anatomical (1 mm isotropic voxels) and T2*-weighted functional data were collected with an eight channel phased array SENSE head coil. EPI data (gradient echo-pulse sequence) were acquired from 32 slices (3 x 3 x 4 mm resolution, TR 2000 ms, TE 35 ms, SENSE factor 2, flip angle 80°, with equidistant temporal slice spacing to facilitate synchronisation of the EEG clock (Mandelkow et al., 2006). Slices were oriented parallel to the AC-PC axis of the observer's brain and positioned to cover the entire brain space.

Eye-movements were monitored for four observers while performing the perceptual decision task for the combined EEG-fMRI data acquisition in the MR scanner. Eye

movements were recorded using the ASL 6000 Eye-tracker (Applied Science Laboratories, Bedford, MA).

6.2.5 Data pre-processing

EEG data acquired outside the MRI scanner were partitioned into data acquisition sessions, band-pass filtered from 0.5 to 25 Hz, down-sampled to 500 Hz and exported to .dat format using Brain Vision Analyzer 2.0 (Brain Products, Munich, Germany). EEG data acquired simultaneously with EPI data were partitioned into data acquisition sessions and the MR gradient and ballistocardiogram artefact removed using Brain Vision Analyzer 2.0, band-pass filtered from 0.5 to 25 Hz, and down-sampled to 500 Hz. After exporting to .dat format, the data were imported to Matlab (The Mathworks, Natick, MA) using an EEGLab (Delorme and Makeig, 2004) routine. Subsequently, to identify artefactual non-cerebral EEG components (i.e. eye-movements, muscular movement, environmental noise, and, in the case of the EEG data acquired simultaneously with the EPI data, residual MR and BCG artefacts) a semiautomatic ICA-based procedure was employed (Barbati et al., 2006; Porcaro et al., 2009) The computational underpinnings and details of this procedure are discussed in Appendix E. Finally, all trials with maximum or minimum amplitudes outside a physiological range of $-100\ \mu V$ to $100\ \mu V$ were discarded prior to further analyses.

Raw EPI data (.par/.rec format) were converted to Niftii format using MRICroN's dcm2nii utility (Rorden and Brett, 2000). SPM5 (Friston, 2007) was used for fMRI data pre-processing. Data pre-processing involved anatomical realignment, slice scan time correction (reference slice 16), re-interpolation to 2 x 2 x 2 mm voxels, anatomical normalization and spatial smoothing (5 mm Gaussian kernel).

6.2.6 EEG data analysis

EEG data were analyzed in electrode space using custom written Matlab code. Specifically, event-related potentials were computed using a 100 ms pre-stimulus baseline and 500 ms post-stimulus window. As stimuli were presented lateralized, data were analyzed from the contra-lateral electrodes and collapsed over hemispheres for subsequent analyses. Grand averages of even-related potentials were computed across all trials of a given condition and subjects for electrodes O1, O2, PO3, PO4, PO7 and PO8.

Upon the identification of time-windows of interest based on the grand average EEG data (Figure 6.6) and previous studies (Benar et al., 2010; Karch and Mulert, 2010; Luck and Hillyard, 2000; Philiastides et al., 2006; Philiastides and Sajda, 2006; Rossion and Jacques, 2008), single-trial amplitude estimates were extracted from the EEG time-course for five discrete, non-overlapping time-windows of interest covering the entire -100 to 500 ms peri-stimulus presentation. For the EEG only data, these time-windows consisted of the intervals -100 to 58 ms, 60 to 120 ms, 122 to 154 ms, 156 to 370 ms, and 372 to 500 ms. As the equivalent neuronal and behavioural responses were slightly delayed for the combined EEG-fMRI data acquisition, the corresponding time-windows for the EEG-fMRI data were determined as -100 to 58 ms, 60 to 140 ms, 142 to 188 ms, 190 to 400 ms and 402 to 500 ms. For each time-window, except the third, the maximum amplitude on each single trial was extracted. For the third time-window, which encompasses a negative potential deflection, the minimum amplitude was extracted.

The electrodes for which these time-domain features were extracted comprised a set of eight parieto-occipital electrodes (O1/2, PO3/4, PO7/8, P1/2, P3/4, P5/6, P7/8 and TP7/8), whose selection was based on the topography of the grand average event-related potential (Figure 6.6. panels C and D). Upon feature extraction and information estimation, the

information theoretic results for the EEG marginal features and EEG-fMRI joint features were averaged across these electrodes to yield the final information estimate.

6.2.7 fMRI data analysis

To identify fMRI regions of interest (ROIs), the experimental data of each individual voxel was modelled using the standard univariate GLM approach in SPM5 (Friston, 2007). A total of 16 experimental regressors were used, corresponding to the 8 stimulus conditions (2 coherence x 2 prioritization x 2 stimulus category levels) and 2 presentation locations (left and right visual hemifield). Voxel time-courses were modelled in an event-related fashion using regressors obtained by convolving each stimulus onset unit impulse with a canonical haemodynamic response function and its first temporal derivative. Additional nuisance covariates included the realignment parameters to account for residual motion artefacts and session specific means. A mixed-effects analysis was then implemented using a summary statistics approach to allow inferences at the population level (Frison and Pocock, 1992; Mumford and Nichols, 2009): upon estimation of the model parameters for each subject, a subject-specific contrast image for each effect of interest was computed. Contrast vectors for the following effects of interest were used: all stimuli > fixation, left hemifield stimuli > right hemifield stimuli, right hemifield stimuli > left hemifield stimuli, high coherence stimuli > low coherence stimuli, low coherence stimuli > high coherence stimuli (Heekeren et al., 2004; Heekeren et al., 2006) and face stimuli > car stimuli. The contrast images were then subjected to a one-sample t-test on the second level.

For the information theoretic analysis filtered and whitened data were extracted from a sphere of 2 mm radius centred on the subject specific peak for the relevant contrast. The subject specific peak for each ROI was uniquely identified by visual inspection as the

coordinates of the peak of the significantly activated cluster that was closest to the group random effects analysis coordinates. The average deviation across ROIs and observers from the group coordinates was 15 (± 1 SEM) mm. Upon time-course extraction, single-trial event-related haemodynamic responses were computed as percent signal change with respect to a baseline comprising 2 pre-stimulus data points. The single-trial fMRI amplitude feature was then determined as the maximum over the 10 s post-stimulus period. The HRF amplitude was chosen as the only fMRI data feature in the current chapter as a) it was shown to be marginally more informative compared to the other fMRI data features in chapter 4, b) the fMRI data features of chapter 4 do not vary substantially, and c) to simplify the analysis and prohibit exponential growth in the number of EEG-fMRI data feature combinations.

6.2.8 Information theoretic EEG-fMRI feature integration

The estimation of information theoretic quantities from EEG-fMRI data features has been discussed in detail in chapter 2, section 2.4 and chapter 4, section 4.2.8. Here, only those aspects that are specific to the current experimental paradigm will be discussed.

To elucidate the spatiotemporal information representation signature for perceptual decisions, mutual information quantities were formulated relating to external, internal and behavioural state variables. Here, the external and internal variables are equivalent to experimental manipulations, i.e. stimulus variables. Specifically, the interest lies on a) the informativeness of the stimulus (spatial coherence) b) the stimulus category (face or car) and c) the observer's attentional state, parameterized by spatial prioritization. Let S denote the respective stimulus variable, then the following quantities were computed with respect to the different response variable features:

$$I_N(S;R) = \sum_{s\in S}\sum_{r\in R} p_N(s,r)\log_2\left(\frac{p_N(s,r)}{p_N(s)p_N(r)}\right) \quad (6.4)$$

where the variable R indicates either an EEG or fMRI data feature, and

$$I_N(S;R_1,R_2) = \sum_{s\in S}\sum_{r_1\in R_1}\sum_{r_2\in R_2} p_N(s,r_1,r_2)\log\left(\frac{p_N(s,r_1,r_2)}{p_N(s)p_N(r_1,r_2)}\right) \quad (6.5)$$

where the variable R_1 indicates an EEG and the variable R_2 an fMRI data feature.

Intuitively, the quantity $I(S;R)$ in (6.4) represents the relative distance of the observed stimulus-response joint distribution $p_N(s,r)$ from its factorized counterpart $p_N(s)p_N(r)$ which embeds the assumption of stimulus-response variable independence for a univariate response feature. Similarly, $I_N(S;R_1,R_2)$ in (6.5) represents the same distance for a bivariate response feature, here comprising an EEG and an fMRI feature.

Analogously, with respect to the observer's behaviour, interest lies on 1) the subject's response time, and 2) the subject's categorical decision. Let B denote the respective behavioural variable, then

$$I_N(B;R) = \sum_{b\in B}\sum_{r\in R} p_N(b,r)\log_2\left(\frac{p_N(b,r)}{p_N(b)p_N(r)}\right) \quad (6.6)$$

where the variable R indicates either an EEG or fMRI data feature

$$I_N(B;R_1,R_2) = \sum_{b\in B}\sum_{r_1\in R_1}\sum_{r_2\in R_2} p_N(b,r_1,r_2)\log\left(\frac{p_N(b,r_1,r_2)}{p_N(b)p(r_1,r_2)}\right) \quad (6.7)$$

where the variable R_1 indicates an EEG and the variable R_2 an fMRI data feature. The expressions with respect to the stimulus or the subject's behaviour are obviously analogous. However, it has to be noted that the marginal stimulus distributions $p(s)$ are determined by the experimenter and are uniform, while the marginal behaviour distributions $p(b)$ are experimentally observed, resulting in larger experimental uncertainty for the latter.

To estimate the information of a given EEG or fMRI feature or a feature combination about each of the variables of interest, the trials associated with this variable were sorted according to the respective variable categories and collapsed over all other stimulus categories. For example, to estimate $I_N(S; R_1, R_2)$ with respect to the stimulus informativeness, trials were grouped into low stimulus s_1 and high stimulus spatial coherence s_1, the joint observed probability distributions $p_N(s, r_1, r_2)$ estimated and the informativeness of the signal features with respect to s_1 and s_2 assessed. An analogous procedure was carried out for the information about the stimulus category (face vs. car) and about the observer's attentional state (spatial prioritization vs. no spatial prioritization).

For the behavioural state variables, a similar procedure was carried out in a slightly modified manner: first, regarding the information about response times, all trials across all conditions were considered, and the joint probability distribution $p_N(b, r_1, r_2)$ estimated, where b represents the continuous random variable response time. The small number of trials on which observers did not respond within 1 s of stimulus onset (average of 10.8% per subject) were excluded from the analysis to render the estimation of the joint probability distribution less prone to outliers (the histogram grid is adjusted to include the maximum and minimum values of each response variable (Chapter 2, section 2.3), hence single outliers can have strong effects on the overall response space partitioning, which is to be avoided).

Finally, with respect to the observer's decision, only low coherence trials were considered in order to decouple the distribution of the observer's decision as much as possible from the physical stimulus category, i.e. $p_N(b, r_1, r_2)$ was estimated where b represents the distribution of the discrete random variable decision (face vs. car) on low spatial coherence trials. It should be noted that the current experimental paradigm was not optimized to study the informativeness of features with respect to the observer's perceptual state as the high

accuracy of performance indicates that for most of the trials, the physical stimulus category and the observer's perception matched. Future studies using near-threshold paradigms (Pessoa and Padmala, 2007b) might elucidate the informativeness of joint EEG-fMRI signal features about the observer's perceptual state in more detail. For the current study, it follows that the estimation of information about the observer's decision is more error prone compared to the other variables, as it proceeds based on half of the number of trials.

The information theoretic analysis for expressions (6.4) and (6.5) was performed as described in the previous Chapters. Briefly, based on the single-trial signal feature values, the respective probability distributions were estimated non-parametrically using a two-dimensional histogram approach with the number of bins set to $floor(\sqrt[2]{N_C})$, where N_C denotes the number of trials per condition (Panzeri et al., 2007; Pola et al., 2003). Entropy and mutual information values were then computed using the respective equations and bias corrected for limited sample sizes using a combination of PT-, shuffling (1000 permutations)- and Gaussian null model (1000 simulations)-correction (Chapters 2,3 and 4). For expressions (6.6) and (6.7) the estimation of the the marginal behavioural variable distributions is required. This entails a three-dimensional histogram analysis with the number of bins set to $floor(\sqrt[3]{N})$ where N denotes the total number of trials evaluated. Given the non-Gaussianity of response time distributions (Ratcliff and McKoon, 2008) and binary nature of the perceptual decision, the pure Gaussian null model bias correction employed for stimulus-response signal relationships is no longer appropriate in this case. Hence, the respective null models for bias correction were based on sampling from three independent random variables (1000 simulations), two of which were Gaussian, representing the response signals, while the third, representing the behavioural variable, was either a Gamma distribution (response time null model) or a Bernoulli distribution (decisional variable null model).

It should be noted that the selected of the number of response bins in the current application is relatively high and non-conservative, i.e. it maximizes sensitivity to informative aspects in the data while it decreases specificity (i.e. it increases the risk of false-positives). Bias control procedures were in turn employed to keep an over-estimation of information at bay as described above (i.e. to decrease the risk of false positives). However, the uncertainty about the absolute value of information in the current analysis is reflected in the fact, that in the following only between-feature information comparisons are evaluated (i.e. the analysis focuses on relative information estimates) and no tests are performed for the difference of the information values from zero (i.e. the analysis does not focus on absolute information estimates) .

Statistical comparisons of the estimated information quantities were carried out using one or two-way repeated measures ANOVA models with Greenhouse-Geisser correction when appropriate, i.e. a significant result of Mauchy's test for sphericity followed by pairwise comparisons based on the estimated marginal means (least-significant differences) in SPSS (SPSS Inc, Chicago, IL)

6.3 Results

In the following, traditional psychophysical, event-related potential and fMRI-GLM analyses presented are prior to the information theoretic analyses. These serve the following purposes: 1) to make the reported IT results more comparable to similar studies of perceptual decisions, 2) to determine whether the experimental manipulations resulted in behavioural modulations, 3) to guide the identification of data features of interest, i.e. time-windows of interest for the EEG data and regions of interest for the fMRI data based on group results, and

4) to allow data quality assessment and inspection prior to single-trial feature distribution estimation.

Subsequently, the information represented in the data features of interest about the external, internal and behavioural variables of interest is presented successively for the EEG domain, the fMRI domain and finally the combined EEG-fMRI domain.

6.3.1 Psychophysical results

Figure 6.5 depicts the psychophysical results for both EEG only and simultaneous EEG-fMRI experiments. In both cases, faster median response times were observed for high informative compared to low informative and spatially prioritized compared to not spatially prioritized stimuli (Figure 6.5 A). Equivalently, response accuracies increased with stimulus informativeness and spatial prioritization (Figure 6.5 B). The observed behavioural pattern was identical between EEG only and simultaneous EEG-fMRI experiment. However, the MRI scanner environment lead to an overall increase in response times and decrease in performance accuracy (mean response time across conditions: 441 (± 22 SEM) ms EEG vs. 740 (± 38 SEM) EEG-fMRI, accuracy across conditions: 90 (± 2 SEM) % correct EEG vs. 84 (± 2 SEM) % EEG-fMRI). Possible endogenous sources for this baseline shift to longer response times and lower accuracies include the noisy scanner environment, the uncomfortable scanning position, and fatigue, as the simultaneous EEG-fMRI data acquisition always followed the EEG data collection outside the MR environment. Possible exogenous sources include the lower quality of the visual projection as well as potential signal delays due to differences in the response button set-up and fibre optic conduction. Impairment in behavioural performance in psychophysical tasks for inside the MR scanner compared to outside the MR scanner are have been reported repeatedly (Noppeney, U.,personal

communication, Eichele T., personal communication), and their origins could be addressed in future studies.

To quantitatively assess the reliability of the experimental manipulation on the behavioural responses, a two-way repeated measures ANOVA with factors stimulus informativeness and spatial prioritization was carried out. For response times on correct response trials, this ANOVA revealed significant main effects of stimulus informativeness (EEG: $F_{(1,12)} = 39.6$, $p < 0.001$, EEG-fMRI: $F_{(1,12)} = 17.4$, $p = 0.001$) and spatial prioritization (EEG: $F_{(1,12)} = 12.1$, $p = 0.005$, EEG-fMRI: $F_{(1,12)} = 30.6$, $p < 0.001$). However, no significant interaction was observed (EEG: $F_{(1,12)} = 1.1$, $p = 0.31$, EEG-fMRI: $F_{(1,12)} = 0.1$, $p = 0.81$). For accuracy, significant main effects of stimulus informativeness (EEG: $F_{(1,12)} = 54.9$, $p < 0.001$, EEG-fMRI: $F_{(1,12)} = 146.7$, $p = 0.001$) and spatial prioritization (EEG-fMRI: $F_{(1,12)} = 7.8$, $p = 0.01$, not for the EEG data set recorded outside the MR scanner: $F_{(1,12)} = 3.0$, $p = 0.10$) were observed, but no significant interaction (EEG: $F_{(1,12)} = 0.1$, $p = 0.70$, EEG-fMRI: $F_{(1,12)} = 2,9$, $p = 0.11$).

Together these results indicate that the experimental manipulation reliably evoked differential behavioural responses, while both experimental factors appear to act on independent substrates as no significant interaction was observed. In general, the psychophysical behaviour inside the MR scanner is more error prone and longer response times are observed. This is likely due to the diminished quality of the stimulus projection (rear-projection instead of LCD display), the scanner noise and the more uncomfortable recording setup.

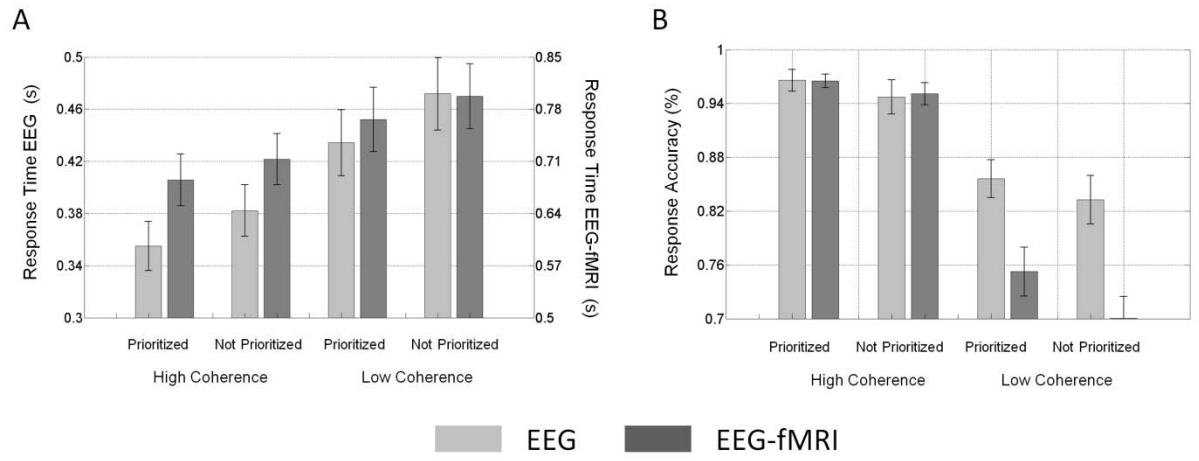


Figure 6.5 Psychophysical Results A. Response Times. Bars depict the average median response times across observers, error bars the SEM. Response accuracy. Bars depict the average median response times across observers, error bars the SEM. The light grey bars (EEG) represent the EEG data set recorded outside the MR scanner, the dark grey bars (EEG-fMRI) represent the EEG data set acquired simultaneously with the fMRI data.

6.3.2 Time course analysis of EEG data

To assess the data quality, to select electrode regions relevant for the current study and to identify time-windows of interests, a traditional event-related potential (ERP) analysis was performed. Figure 6.6 depicts the grand average EEG time-courses for a set a of parieto-occipital electrodes for both the EEG only (Panels A and C) and combined EEG-fMRI (Panels B and D) study. Given the hemifield presentation of the stimulus, the data were extracted from electrodes O2, PO4 and PO8 for left hemifield trials, from electrodes O1, PO3, and PO7 for right hemifield trials, and collapsed according to the experimental conditions. In line with similar previous studies (Philiastides et al., 2006; Philiastides and Sajda, 2006) no substantial potential deflections were observed after 500 ms post-stimulus, hence the focus of the analyses is on the -100 to 500 ms peri-stimulus time window.

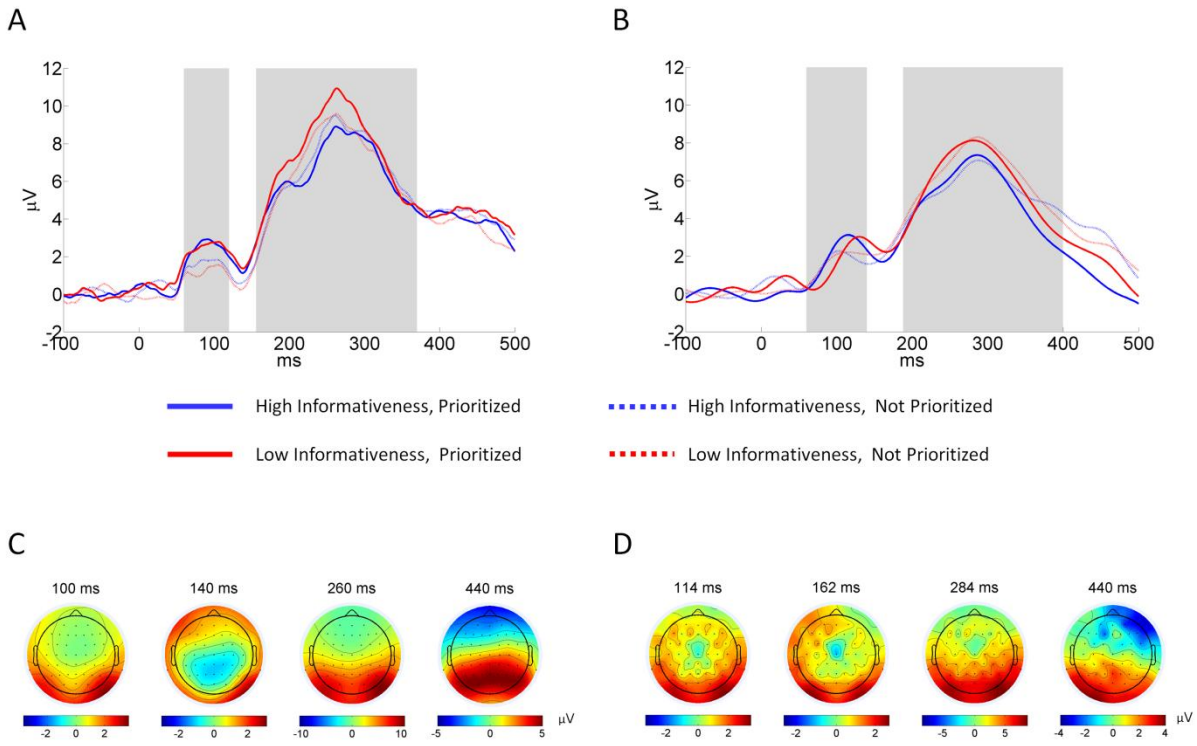


Figure 6.6 ERP analysis. A and B.EEG grand average time-courses for contralateral trials from electrodes O2, PO4, PO8 (left hemifield trials) and O1, PO3, PO7 (right hemifield trials) for the EEG only (A) and combined EEG-fMRI (B) data acquisition. C and D. The pattern of shaded and non-shaded areas reflects the five non-overlapping time-windows used for the subsequent information theoretic analyses. Topography plots of the entire EEG electrode set potentials for time-points 100, 140, 260 and 440 ms post-stimulus. The main positive deflections at all time-points are observed for a set of parieto-occipital electrodes (O2, O1, PO8, PO7, P8, P7, PO4, PO3).

For both data sets and all conditions, an early (approximately 100 ms post-stimulus, referred to as P100) and a late (approximately 270 ms post-stimulus onset, referred to as P300) positive potential deflection can clearly be identified. These positive deflections are separated by an intermediate negative deflection (approximately 140 ms post-stimulus, referred to as N140), which is more prominent in the EEG only data. With respect to the experimental conditions, the most obvious effects are that spatial prioritization of the stimulus led to an increase in the P100 amplitude, while increasing the task difficulty by lowering the stimulus informativeness increased the P300 amplitude. These effects are observed for both the EEG only data acquisition set (Figure 6.6 A), as well as with slightly diminished prominence for the simultaneous EEG-fMRI data set (Figure 6.6 B). The stimulus condition

effect on the N140 deflection is less apparent from these grand averages: while in the EEG only data, a more negative peak is observed for the not prioritized conditions, for the simultaneous EEG-fMRI data more negative amplitudes are observed for the high coherence stimulus condition instead. Differences in the information theoretic analysis for this time bin between EEG and EEG-fMRI data hence might be due to the lower data quality of the EEG data acquired inside the MR scanner. On the level of these time-courses, no tests for significant differences between time-courses were performed, and are unlikely to reach statistical significance on the group level. However, it should be noted that the information theoretic statistics described below are formally equivalent to likelihood ratio test statistics (Friston, 2006) and hence effectively amount (under a specific set of assumptions) to differential signal to variability ratios.

Topography plots of the entire electrode set of the grand average across conditions for time-points corresponding to the most prominent potential deflections (Figure 6.6, lower panels) show that the strongest positive deflections for all time-points are observed for posterior, parieto-occipital electrodes. For the simultaneous EEG-fMRI data set, the topography plots display the spatial potential distribution. However, the data appear slightly noisier than the EEG only topography plots, reflecting residual MR and BCG artefacts.

As detailed in the methods section, based on this spatio-temporal pattern of potential deflections, the peri-stimulus EEG time-course of an electrode set of interest was divided into five non-overlapping time-windows reflecting 1) the pre- and early post-stimulus baseline, 2) the positive deflection around 100 ms (P100), 3) the negative deflection around 140 ms (N140), 4) the positive deflection around 270 ms (P300) and 5) the remaining time. The respective time-windows are indicated by the pattern of shaded and not-shaded areas underlying the time-courses. The topography of responses across all of these time windows motivated the selection of a parieto-occipital electrode set as electrophysiological region of

interest for the subsequent information theoretic analyses, including the following, bilateral electrodes: O1/2, PO7/8, P7/8, PO3/4, P5/6, TP7/8, P1/2 and P3/4.

6.3.3 General linear model analysis of fMRI data

To identify regions of interest for the subsequent information theoretic analyses, a group level GLM analysis of the fMRI data set was performed. The aim of this analysis was to explicitly identify areas previously implicated in perceptual decision processes (Heekeren et al., 2004; Heekeren et al., 2006; Mulert et al., 2008).

Figures 6.7 and Table 6.1 depict the results of the second-level GLM analysis of the fMRI data. Of all possible main effect contrasts for the current factorial design, the main effect of prioritization is omitted here, as no significant activation was detected for this contrast at the group level. This potentially due to the fact that the task demand was high for both the prioritized and non-prioritized conditions. For the contrasts of left vs. right and right vs. left hemifield stimulus presentation, lateralized activity was detected in the occipital cortex, while higher level cortices did not display lateralized activity. This provides some validation for the use of single hemispheric signal features for occipital (striate cortex, extra-striate cortex, lateral occipital sulcus) regions of interest in the information theoretic analyses reported below.

The set of regions identified as significantly activated ($p < 0.001$ (uncorrected)) for the high coherence vs. low coherence and low coherence vs. high coherence including superior frontal sulcus, pre-central sulcus, anterior cingulate gyrus, insula and frontal-eye fields, all of which have previously been implicated in the processing of visual perceptual decisions (Heekeren et al., 2004). While some of these regions do not reach full family-wise corrected statistical significance in the current study (Table 6.1), their implication in the perceptual decision process reported previously motivated their selection. As the motivation of the GLM

analysis was to determine ROIs for the IT analysis, a relatively liberal threshold was used to avoid missing informative voxels.

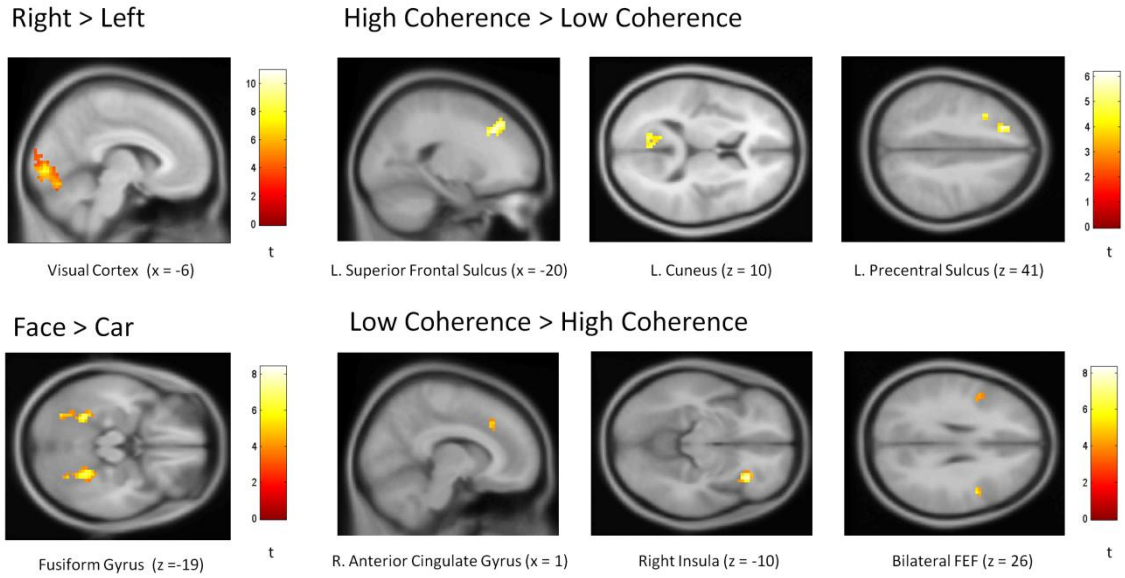


Figure 6.7 Illustration of the group-level fMRI GLM results. The respective statistical parametric maps (thresholded at $p < 0.001$ (uncorrected), extent threshold 15 - 20 voxels) are overlaid on the group MNI template.

Further, given the known role of the intra-parietal sulcus in cognitive tasks (Corbetta and Shulman, 2002; Shulman et al., 2002), the most active voxels in this region for the low vs. high coherence contrast were also identified and selected, although they were not significantly activated even at $p < 0.001$ (uncorrected)). Finally, as observers performed a face vs. car categorization task, face responsive cortex of the fusiform gyrus (fusiform face area (FFA), (Haxby et al., 2002)) was identified using the face vs. car stimulus contrast.

6.3.4 Feature extraction

The information theoretic analyses reported below capitalise on the evaluation of the probability distributions of signal features.

Region	x	y	z	z-score peak	p-value
Right > Left					
L. Striate Cortex	-15	-81	-15	5.66	<0.001
L. Extra-Striate Cortex	-27	-72	-15	5.45	<0.001
L. Lateral occipital sulcus	-46	-80	10		
Left > Right					
R. Striate Cortex	12	-84	-3	5.32	<0.001
R. Extra-Striate Cortex	21	-75	-12	5.73	<0.001
R. Lateral Occipital Sulcus	51	-75	9	4.34	<0.001
High Coherence > Low Coherence					
L. Superior Frontal Sulcus	-21	27	39	4.22	0.001
L. Cuneus	-15	-69	12	4.29	<0.001
L. Pre-Central Sulcus	-33	12	45	3.79	0.083
Low Coherence > High Coherence					
R. Anterior Cingulate Gyrus	12	21	39	3.90	0.102
R. Insula	33	24	-9	5.01	<0.001
R. Frontal Eye Field	48	3	27	4.02	0.083
L. Frontal Eye Field	-45	6	24	3.40	0.156
R. Intra-Parietal Sulcus	30	-60	42	2.32	1.000
L. Intra-Parietal Sulcus	-24	-69	30	3.21	0.961
Face > Car					
L. Fusiform Gyrus	-27	-51	-24	5.04	< 0.001
R. Fusiform Gyrus	27	-45	-21	4.74	< 0.001

Table 6.1 Group-level fMRI-GLM results. The MNI coordinates, z-scores of the cluster peak voxel and family-wise corrected p-values (cluster level) are displayed.

These distributions are estimated non-parametrically from the extracted single-trial feature data. Figure 6.8 displays an example for a single subject and shows the single-trial time-courses for the electrode and brain regions from which the single-trial estimates were obtained. Inspection of the plots indicates that on most individual trials, a reliable ERP/HRF could be observed. As the last column of averages indicates, the profiles of potential deflections across conditions vary over electrodes, but are qualitatively similar.

Likewise, Figures 6.9 displays the extracted feature distributions across the experimental conditions. As can be seen, the distributions for the respective features overlap.

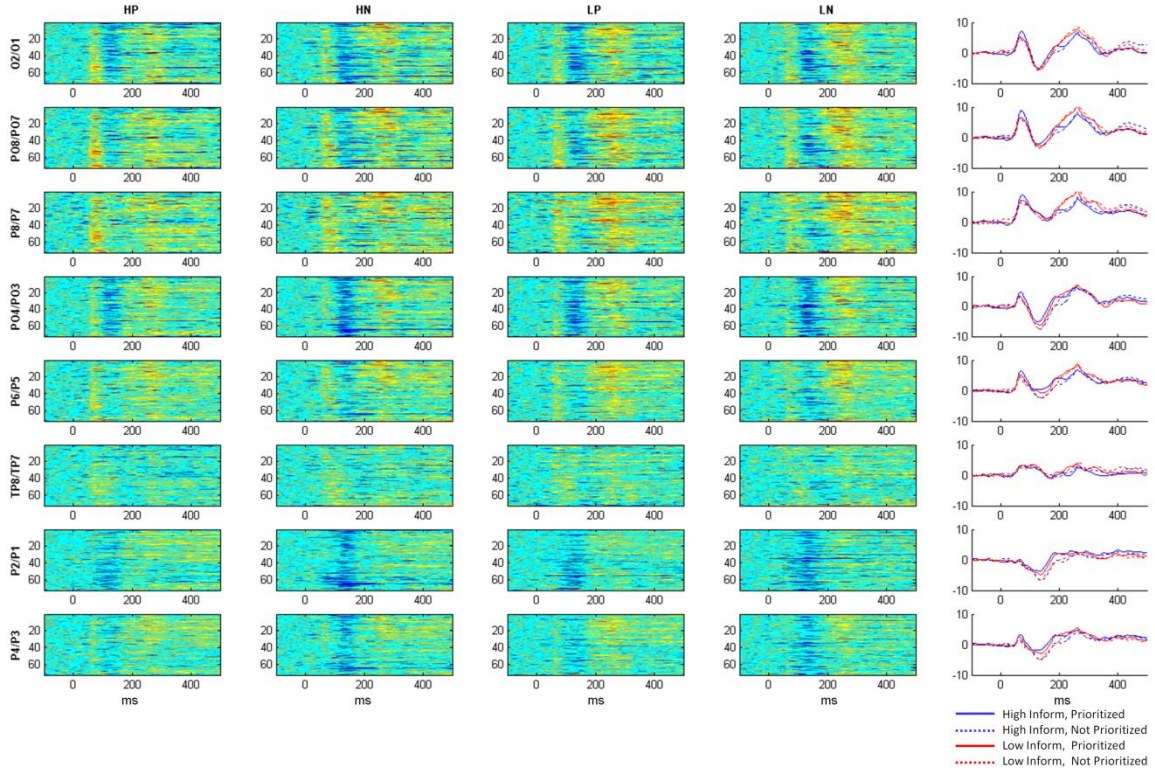


Figure 6.8 Single subject single-trial EEG plots. The rightmost column displays the experimental condition averages with the same colour coding as in Figure 6.6. (HP: High Informativeness, Prioritized, HN: High Informativeness, Not Prioritized, LP: Low Informativeness, Prioritized, LN: Low Informativeness, Not Prioritized).

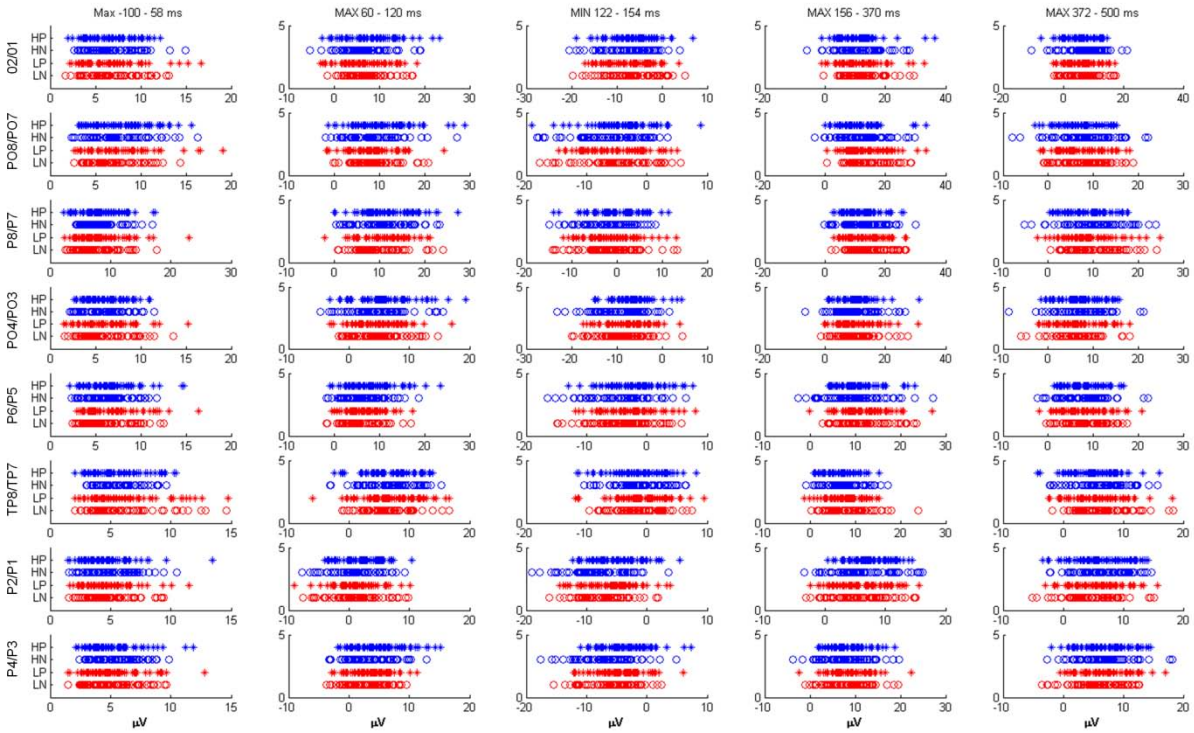


Figure 6.9 Single subject EEG feature distributions. The y-axis annotation represent the four experimental conditions (HP: High Informativeness, Prioritized, HN: High Informativeness, Not Prioritized, LP: Low Informativeness, Prioritized, LN: Low Informativeness, Not Prioritized).

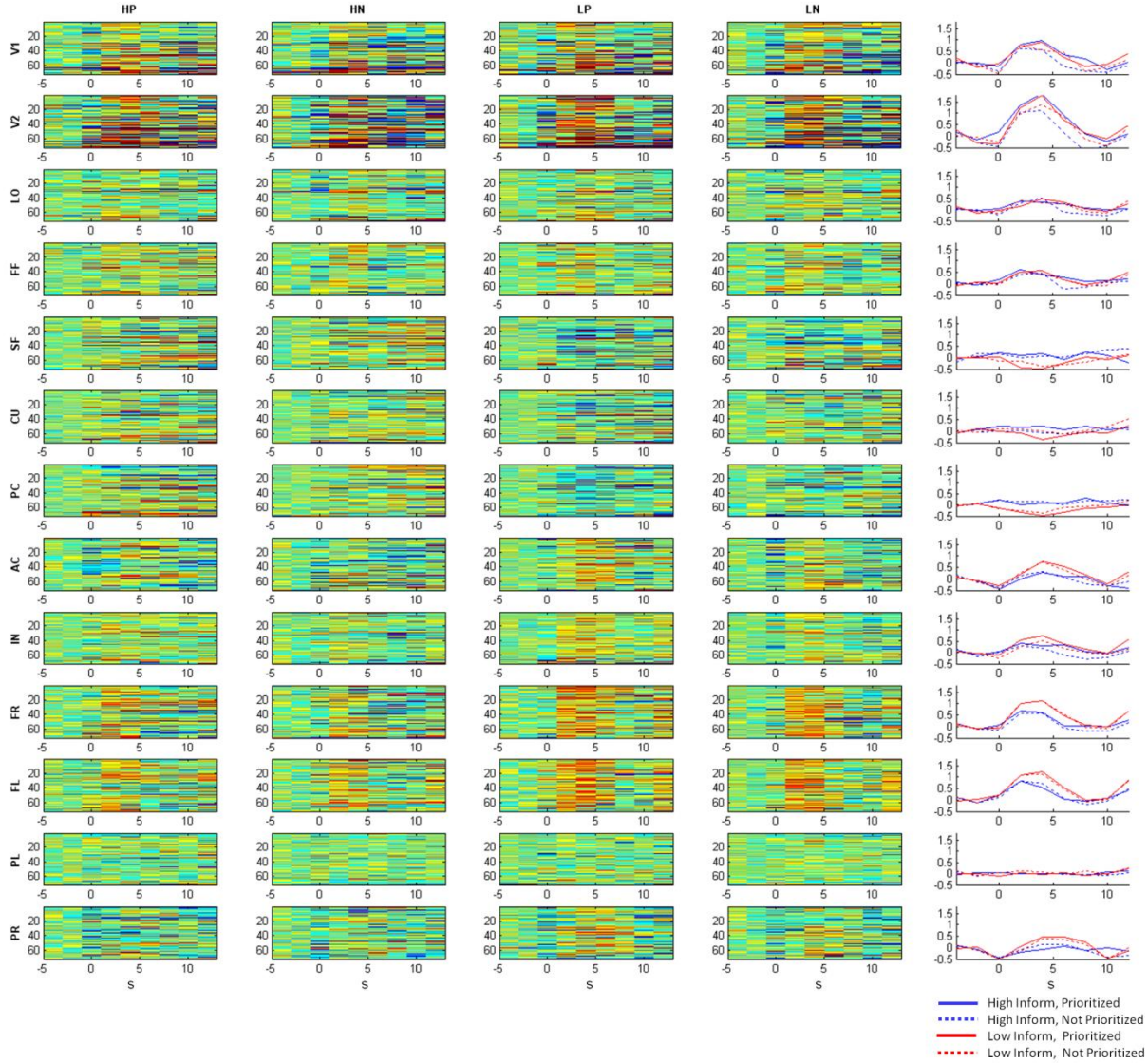


Figure 6.10 Single subject single-trial fMRI plots. The rightmost column displays the experimental condition averages with the same colour coding as in Figures 6.8 and 6.9. (HP: High Informativeness, Prioritized, HN: High Informativeness, Not Prioritized, LP: Low Informativeness, Prioritized, LN: Low Informativeness, Not Prioritized. The ROIs are abbreviated as V1: Striate Cortex, V2: Extrastriate Cortex, LO: Lateral Occipital Complex, CU: Cuneus, FF: Fusiform Gyrus, PL: L. Intra-Parietal Sulcus, PR: R.Intra-Parietal Sulcus, PC:Post-Central Gyrus, AC: R.Anterior Cingulate, IN:R. Insula, FL:L. Frontal Eye Field, FR: R. Frontal Eye Field, SF: L. Superior Frontal Gyrus)

Likewise, Figures 6.10 and 6.11 display the extracted feature distributions across the experimental conditions for the fMRI modality in an analogous manner to Figures 6.8 and 6.9.

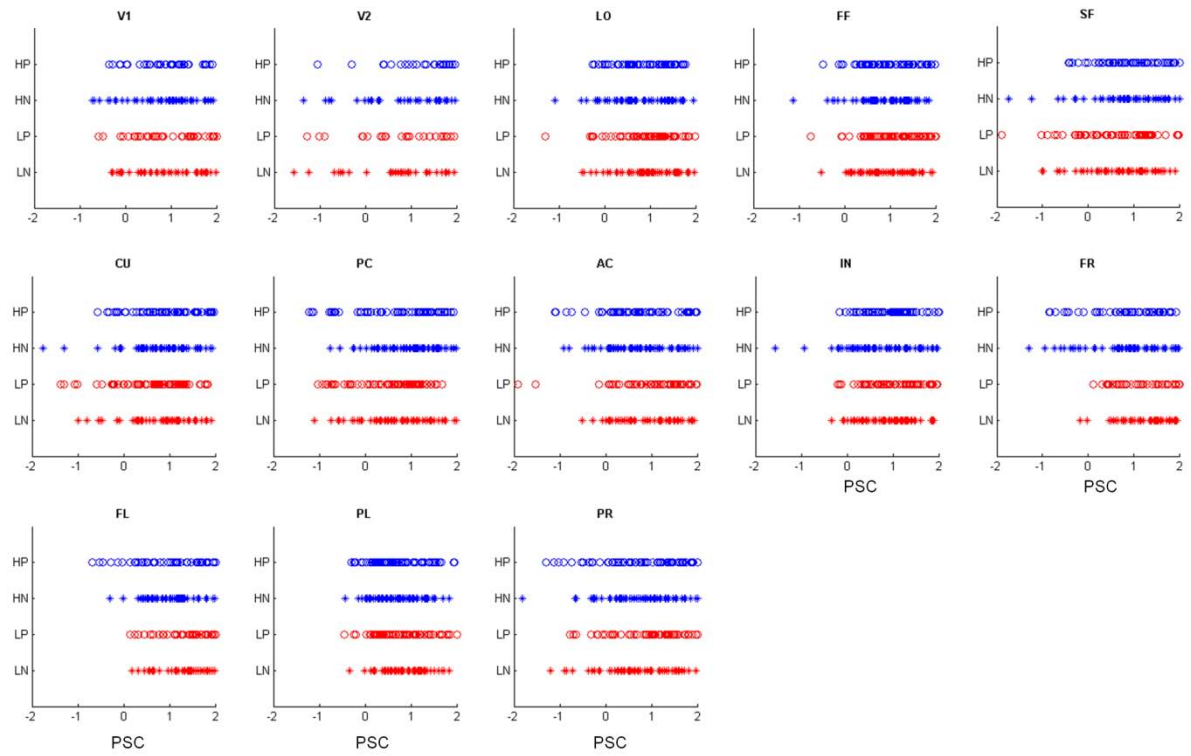


Figure 6.11 Single subject FMRI feature distributions. The y-axis annotation represent the four experimental conditions (HP: High Informativeness, Prioritized, HN: High Informativeness, Not Prioritized, LP: Low Informativeness, Prioritized, LN: Low Informativeness, Not Prioritized. The ROIs are abbreviated as V1: Striate Cortex, V2: Extrastriate Cortex, LO: Lateral Occipital Complex, CU: Cuneus, FF: Fusiform Gyrus, PL: L. Intra-Parietal Sulcus, PR: R.Intra-Parietal Sulcus, PC:Post-Central Gyrus, AC: R.Anterior Cingulate, IN:R. Insula, FL:L. Frontal Eye Field, FR: R. Frontal Eye Field, SF: L. Superior Frontal Gyrus, PSC: percent signal change)

6.3.5 EEG-fMRI integrated information theoretic analysis

The aim of the single-trial information theoretic analyses reported here is to gain insight into the following questions: 1) which EEG and fMRI features convey information about the external stimulus variables, stimulus informativeness and stimulus category, 2) which EEG and fMRI features convey information about the internal stimulus variable, namely spatial prioritization, and finally 3) which EEG and fMRI features convey information about the behavioural response variables response time and category of the decision.

The information conveyed by the brain imaging data features can be evaluated separately, i.e. the information conveyed by the EEG (both EEG and EEG-fMRI data sets) and fMRI marginal distributions, and based on their joint distribution obtained from the

simultaneous EEG-fMRI data acquisition. In the following, the information conveyed by the marginal distributions will be discussed before the information conveyed by the joint distribution is considered. The EEG features are assumed to convey information on a time-scale relevant for the decision on a single-trial, i.e. in the first 500 ms after stimulus onset. Hence the information estimated for the respective EEG time-windows is referred to as temporal information representation (Figure 6.12). While there is a spatial component to the EEG data, as parieto-occipital electrodes were selected, this is relatively unspecific, when compared to the fMRI modality. The fMRI features are assumed to convey information on the relevant spatial scale. As the maximum of the haemodynamic response is approximately 3 to 4 seconds after the decision process has terminated, it is assumed that no temporally information relevant for the decision on a single-trial is represented in the haemodynamic response. Hence the information estimated for the respective fMRI regions of interest is referred to as spatial information representation (6.13). Finally, the information represented in the joint distributions of EEG and fMRI features is referred to as spatiotemporal information (Figure 6.14).

It should be noted that the focus of the information theoretic comparison lies on relative comparisons between spatio-temporal features within a given information theoretic quantity with respect to a given experimental variable. Absolute comparisons between different information quantities about different experimental are precluded due to the uncertainty about the absolute information estimates given the varying bias correction schemes. Further, a relatively liberal approach with respect to Type I errors is pursued in the case of statistical comparisons, as the focus of the analysis is exploration of the spatiotemporal information space rather than on establishing spatiotemporal specificities.

6.3.5.1 Temporal information representation

Figure 6.12 displays the information about the external, internal and behavioural state variables for the EEG features of interest. As EEG data were acquired both without and with simultaneous fMRI data, the results of both data sets are presented. The columns of the figure represent the different stimulus and behavioural variables of interest (external, internal and behavioural state). For each stimulus/behavioural variable of interest, the average information estimate across subjects ± SEM is depicted for each of the five time windows of interest. Overall, it can be observed that the information estimates for both data sets, while being lower for the EEG data acquired simultaneously with fMRI data, show similar patterns. The lower information estimates for the EEG data acquired inside the MR scanner reflect the lower single-trial SNR for the artefact-corrected EEG data set.

For both data sets, the earliest time-window around stimulus onset yields the lowest information estimates across features for most stimulus/behavioural variables, in agreement with the fact that information about the stimulus can only be extracted upon stimulus-presentation. With respect to the external state, the information about the stimulus informativeness increases from 140 ms onwards to reach its maximum in the final time-window. This effect is present in both data sets, although it is slightly diminished and delayed for the data acquired in the MR environment. For the EEG data acquired outside the MR scanner, a trend for a significant effect of the time-window on the information estimate was observed ($F_{(4,48)} = 2.24$, $p = 0.07$)). The information estimate in the 450 ms time-window is marginally significantly different from the information estimate in the 0 ms time-window ($p = 0.06$) and significantly different from the information estimate in the 100 ms time-window ($p = 0.04$). Similarly, for the EEG data acquired simultaneously with the fMRI data, a marginally significant main effect of time-window is observed ($F_{(4,48)} = 2.42$, $p = 0.06$), while the comparisons between the fifth and the first, second and third time windows approach

statistical significance, or are statistical significant ($p = 0.08$, $p = 0.03$, and $p = 0.02$, respectively).

Hence, information about the stimulus informativeness, or equivalently, the difficulty of the decision, is found to be represented rather late in the EEG response. This is in concordance with the observed grand average effect of stimulus informativeness and previous studies on the task-difficulty and reaction time sensitivity of the P300 deflection (Benar et al., 2007; Mulert et al., 2008). It should be noted, that for the current experimental design and analysis strategy, task difficulty, reaction time and P300 amplitude co-vary, and the activity dependent information theoretic analysis discussed above do not dissociate these three concepts. A possible means for their dissociation would be to assess the stimulus-conditional mutual information between P300 amplitude and response time, or the response-time conditional mutual information between task difficulty and P300 amplitude.

The other external stimulus attribute that was manipulated in the experimental paradigm is the stimulus category. Category-selective responses for faces compared to other stimuli have been described previously (Philiastides and Sajda, 2006; Rossion and Jacques, 2008). In the current study, for both EEG data sets, the largest estimate for represented information about the stimulus-category is observed in the 260 ms time-window, i.e. in the interval of 150 to 370 ms post-stimulus onset. This is in concordance with the maximal discriminative time-windows for a similar stimulus set identified by (Philiastides and Sajda, 2006). This finding is substantiated by a significant effect of time-window on the information estimates for the EEG acquired outside the MR scanner ($F_{(4,48)} = 6.01$, $p = 0.001$) and statistically significant differences for the fourth time window in comparison to all others ($p < 0.05$), except the fifth ($p = 0.08$). For the EEG data acquired inside the MR scanner, the main effect of time-window was not significant ($F_{(4,48)} = 1.53$, $p = 0.20$). However, the information estimate for the fourth time-window shows marginal statistically significant differences from

that the first and second windows ($p = 0.05$ and $p = 0.06$, respectively). The most information about the category of the stimulus was thus observed in a time-window 150 – 370 ms post-stimulus and declining thereafter.

In comparison to the information about the stimulus, the information estimates about the internal state, i.e. the spatial prioritization of the stimulus, are expressed earlier, with effects from 100 ms post-stimulus onwards. This is in line with the well-known attentional modulation effect on the P100 (Luck and Hillyard, 2000). For the current data sets, the largest information about the attentional state of the observer is observed in the early time-windows of 100 and 140 ms and then, after a decrease in the fourth time-window, again in the last time-window. For the data recorded outside of the scanner, no statistically significant main effect of time-window was observed, consistent with the observation that the information estimates are similar for the time-windows from 100 ms on. The pairwise comparison between the fifth and the first time-window was marginally significant ($p = 0.05$), while the others were not. Again, for the EEG data acquired inside the MR scanner no significant main effect of time-window was observed, while the pairwise comparison between the fifth and the first time-window was the most reliable ($p = 0.16$).

These statistical comparisons indicate the subtle nature of prioritization effects in the current study. Most previous studies on attentional modulation contrast trials during which the stimulus was attended, with those trials in which the contra-lateral visual hemifield was attended (Mangun and Hillyard, 1991), in fact creating a contrast between attended and actively inhibited stimuli. For the current study the experiment was designed to enable the comparison between spatially prioritized stimuli and non-prioritized, but not actively suppressed stimuli, which was considered of larger ecological validity. While the findings of early attentional modulation of visually evoked potentials are similar to those reported

previously, they are of a less prominent nature, reflecting the differences in experimental design.

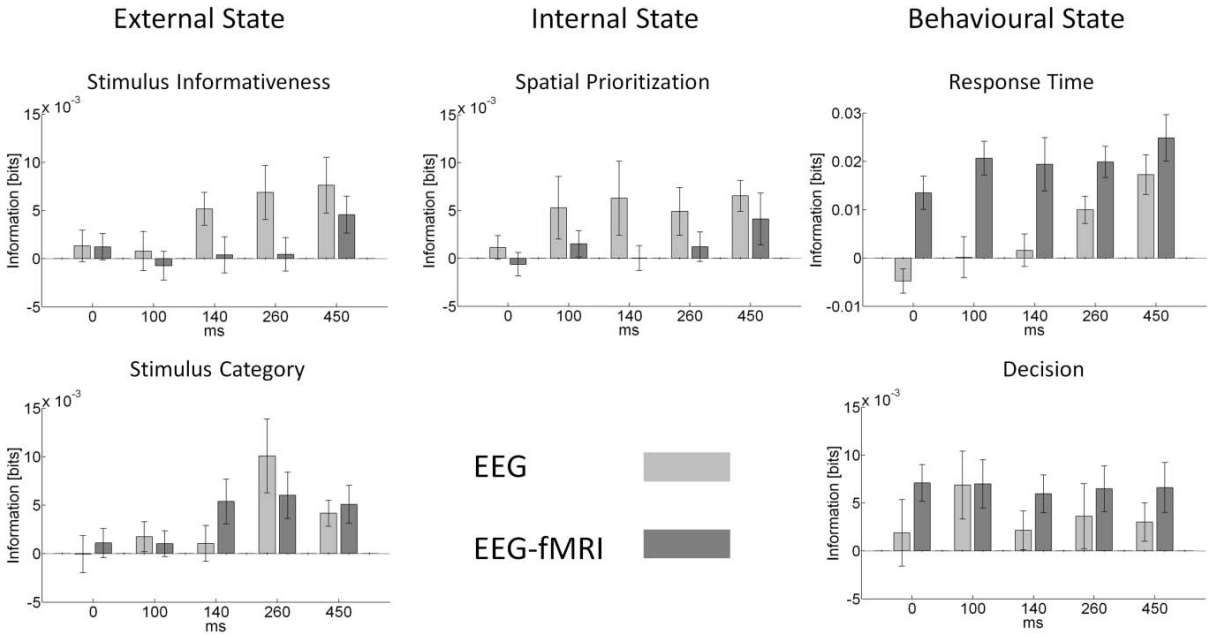


Figure 6.12 Temporal information representation. The data is ordered columnwise according to the variables of interest, external, internal and behavioural state. For each variable of interest, data from the five time-windows identified based on the grand average is displayed. The light bars represent information estimates from the EEG data set acquired outside the MR environment, while the dark gray bars represent information estimates from the simultaneous EEG-fMRI data recordings. All bars reflect group averages (n = 13) and the error bars indicate the SEM.

Finally, with respect to the behavioural state, for the response time, the information estimates increase with time. This effect was significant for the EEG only dataset ($F_{(4,48)}$ = 5.9, p = 0.001), but not for the EEG data acquired simultaneously with the fMRI data ($F_{(4,48)}$ = 1.6, p = 0.18). This pattern is reminiscent of that observed for the information about stimulus informativeness, which is to be expected given the longer response times for low informative trials. Unfortunately, these two processes cannot be dissociated in the current paradigm. For the observer's decision, all EEG time-windows appear equally informative. Consistent with this observation one-way ANOVAs for both behavioural variables and EEG data sets indicated no statistically significant effects for both data sets (EEG: $F_{(4,48)}$ = 0.40, p = 0.80), EEG-fMRI: $F_{(4,48)}$ = 0.04, p = 0.99). As the estimated information values for later time windows do not appear particularly different from those at the earliest time-point, it may be

that the number of trials was too low, or the electrode set chosen not appropriate, to detect a decisional effect on the basis of the marginal EEG data distributions.

In summary, the following picture of temporal representation of information for visual perceptual decisions emerges: in concordance with previous studies, information about the state of the stimulus is represented in the EEG response later than that about the subject's attentional state. With respect to behaviour, later time windows represent more information about the response time, in agreement with their involvement in the representation of uncertainty about the stimulus. The same patterns of results observed for both data sets, although they are more reliable for the EEG only data set. However, it is apparent that the observed single-trial information differences appear larger than those observed on the signals grand averages. This motivates the future evaluation of novel methodologies for the improvement of EEG quality in combined EEG-fMRI recordings on the single-trial level (Porcaro et al., 2010).

6.3.5.2 Spatial information representation

Figure 6.13 displays the information about the external, internal and behavioural state variables for each fMRI region of interest. The columns of the figure represent the different stimulus and behavioural variables of interest (external, internal and behavioural state). For each stimulus/behavioural variable of interest, the average information estimate across subjects ± SEM is depicted for each of the regions of interest identified based on the group fMRI-GLM analysis. Overall, the information of relevance for the perceptual decision task employed in this study appears to be spatially distributed across the cortex. Both anterior (higher) and posterior (lower) cortical areas are implicated in the representation of stimulus

related information, while information about the internal state is slightly more strongly represented in posterior brain areas.

In more detail, the areas primarily implicated in the representation of information about the stimulus informativeness or task difficulty comprise a network of frontal (insula, left frontal eye field, pre-central), parietal (left intraparietal sulcus) and occipital-temporal (extrastriate and fusiform) cortex. The most information about the stimulus informativeness is represented in insular cortex, with an estimate approaching a significant difference with respect to the cuneus and lateral-occipital cortex ($p = 0.07$, $p = 0.04$, respectively). Insular cortex has been implicated in perceptual decision making previously (e.g. Thielscher et al., 2007). Moreover, due to the involvement of insular cortex in a wide variety of cognitive processes (e.g. interoception, self-recognition, emotional awareness, time perception, attention, cognitive control and performance monitoring), it has been proposed that insular cortex plays a pivotal role in the neurobiological representation of awareness (Craig, 2009). In the current study insular cortex appears to be involved in the representation of task difficulty at the single trial level, which could potentially be reconciled with this view, in the sense that insular cortex activity differentiates different states of the stimulus-dependent awareness induction.

Nevertheless, the distributed nature of the represented information is substantiated by the absence of an overall main effect of region of interest on the information about stimulus informativeness ($F_{(12,\ 144)} = 0.78$, $p = 0.67$). It should be noted that most areas were selected according to the contrast of high vs. low and low vs. high stimulus coherence, i.e. on the basis of being informative about stimulus coherence in the sense of a GLM contrast.

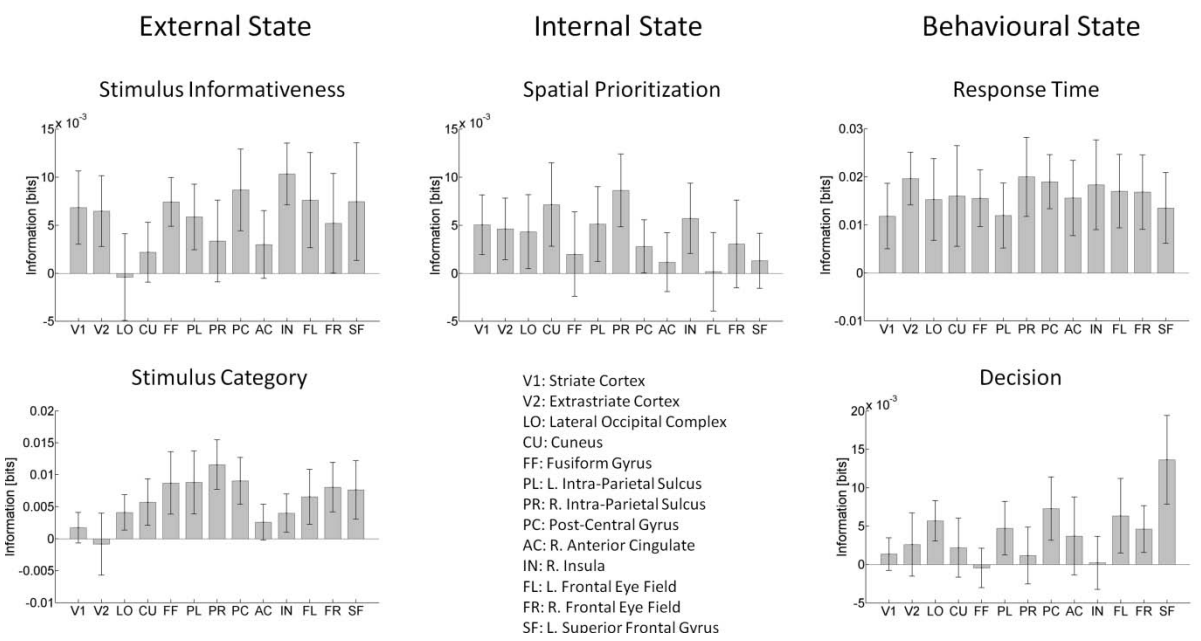


Figure 6.13 Spatial information representation. The data is ordered columnwise according to the variables of interest, external, internal and behavioural state. For each variable of interest, data from all regions of interest identified based on the GLM group analysis is displayed. All bars reflect group averages (n = 13) and the error bars indicate the SEM.

Categorical information about the stimulus appears to be mainly represented in a network of frontal (superior frontal gyrus, frontal eye fields, pre-central) and parietal (intra-parietal sulcus) regions, with some contribution from the fusiform gyrus. This finding is in concordance with the known roles of the IPS and superior frontal gyrus or dorso-lateral prefrontal cortex (DLPFC) in perceptual decisions (Heekeren et al., 2004; Heekeren et al., 2006; Shadlen and Newsome, 2001). While again, the information appears to be distributed across regions of interest ($F_{(12,\ 144)} = 1.01$, $p = 0.43$), the pairwise comparison of the information estimate for the right intra-parietal shows significantly higher information estimates in comparison to extrastriate visual ($p = 0.03$), and anterior cingulate cortex ($p = 0.04$).

A pair of occipito-parietal regions encompassing the cuneus and right parietal sulcus is most informative about the observer's state, i.e. the spatial prioritization of the stimulus. For

the right intra-parietal sulcus, this comparison reaches marginal significance with respect to the right intra-parietal ($p = 0.04$), right frontal eye field ($p = 0.04$) and superior frontal gyrus ($p = 0.06$). This result is in line with previous studies using fMRI to study spatial attention (Corbetta et al., 2000; Corbetta et al., 2002; Corbetta and Shulman, 2002; Hopfinger et al., 2000; Kastner et al., 1999). The involvement of a dorsal frontoparietal network of regions implicated in spatial attention in these studies is substantiated by the absence of a significant main effect of region of interest on information estimate ($F_{(5.4, 64.7)} = 0.74$, $p = 0.60$).

For the behaviour related information, no significant main effects of region of interest, or pairwise comparisons between regions were observed for response time ($F_{(5.0, 60.5)} = 0.24$, $p = 0.94$). The overall representation of the response time appears rather spatially unspecific. This is consistent with GLM-based fMRI studies suggesting that BOLD activity in a large-network of brain areas co-varies with the observer's response time (Noppeney et al., 2010). With respect to the subject's decisional variable, the largest information values are observed for the superior frontal and the pre-central gyrus.. This implicates a shift of stimulus categorical information towards more frontal regions in comparison to the representation of physical stimulus category discussed above. The pairwise comparison of the superior frontal gyrus with the insular cortex, right intra-parietal, fusiform gyrus and cuneus reaches statistical significance ($p < 0.05$) while no overall main effect of region of interest is observed ($F_{(4.5, 53.7)} = 1.0$, $p = 0.38$). While it is tempting to speculate that this indicates a more high level cognitive, rather than low level perceptual determination of the decision on low coherence trials, it should be noted that physical and perceptual stimulus attributes were not completely dissociated in the current study.

In summary, the following picture of spatial representation of information for visual perceptual decisions emerges: with respect to the stimulus, information about differences in stimulus informativeness appears to be represented distributed throughout the cortical regions

studied, while information about the (physical) stimulus category shows a maximum for parietal cortices. Occipito-parietal areas are implicated in the representation of information about the observer's attentional state, while no clear pattern emerges with respect to the speed of the response. On low coherence trials, the superior frontal gyrus is most informative about the observer's decision.

It should be noted that, as with any between-ROI comparison, differences between regions can potentially be confounded by differences in the local haemodynamic response, as well as possible EPI data quality inhomogeneities across the field of view. The information theoretic measures employed here essentially asses the differential (between conditions) SNR of each region, and hence are at least somewhat independent of the ROI-specific time-course SNR. However, it cannot be excluded that the observed differential effects between ROIs arise from an interaction between differences in neural activity and the differences of non-neural origin mentioned above.

6.3.5.3 Spatiotemporal information representation

Figure 6.14 displays the spatiotemporal information surfaces related to the external, internal and behavioural state variables for the combined EEG and FMRI feature variables of interest. The observed data points for the respective EEG time-window × fMRI region-of-interest pairings are depicted as the vertices of the surfaces. Each of the information data points (group average, n = 13) has been estimated from the respective joint distribution $p_N(v, r_1, r_2)$, where v represents the state variable of interest, r_1 the EEG amplitude in the respective time-window and r_2 the fMRI signal amplitude for the respective region of interest. It is hence determined by the signal features' joint distribution, i.e., both dependencies

between the signal features and the variable of interest and the dependencies between the signal features (both stimulus conditional and non-conditional) themselves.

Overall, the results implicate a complex, spatiotemporal pattern of information representation for perceptual decisions in the human brain. For all variables and regions the information estimates are lowest at stimulus onset. Thereafter, information can be observed to flow and accumulate in a distributed manner across time and brain space, the details of which will be discussed in the following.

The information represented about the stimulus informativeness across all regions of interest differs significantly over time ($F_{(4,48)} = 8.8$, $p < 0.001$), but not over space ($F_{(12,144.4)} = 0.5$, $p = 0.90$). For most regions, the information about the stimulus informativeness over time shows a rebound-pattern: following high information estimates early, a decrease is observed for the 140 ms time-window, followed by a later increase. A significant interaction is not observed ($F_{(8.4,\ 101.7)} = 0.6$, $p = 0.74$). With respect to the regions involved at the final time-point, both high (insula, pre-central) and low (extrastriate visual cortex) show the largest information estimates.

With respect to the perceptual decision task, the most important question concerns the representation of information about the stimulus category on a given experimental trial. The most prominent finding from the spatiotemporal information surface with respect to this variable is the parallel increase of information in both high-level (superior frontal gyrus, frontal eye fields) and low-level (striate, extrastriate cortex) areas over time. This is surprising at least with respect to the low-level areas, as the marginal distributions discussed above did not indicate this. This effect might hence be strongly driven by the joint analysis of occipital electrodes and fMRI regions of interest. Again, the main effect of time-window was significant $F_{(2.0,\ 24.2)} = 4.2$, $p = 0.02$), the main effect for region of interest and the interaction

are not ($F_{(12,\ 144)}$ = 0.7, p = 0.68, $F_{(8.3,\ 99.7)}$ = 0.9, p = 0.46, respectively). The largest information estimate of the frontal areas for the left frontal eye field is observed at the latest time point considered, i.e., immediately before the initiation of the observer's motor response. Consistent with previous studies implicating the superior frontal gyrus or DLPFC in the representation of a decision variable (Heekeren et al., 2004; Heekeren et al., 2006; Philiastides and Sajda, 2007), this region shows a build-up of information over time. Finally, at the latest time-point, the fusiform gyrus is also informative about the stimulus category, however, it does not show the incremental build-up of information seen in the frontal areas. It is tempting to speculate that the observed behaviour of information representation for this low-level area could be explained by recurrent feedback from higher areas (Friston, 2010).

With respect to the internal state, the information surface indicates that early in the decision process mostly low and mid-level cortical areas (extra-striate, the cuneus and the anterior cingulate gyrus) are involved, while later in the decision process both low and high level areas. Again, the main effect of time-window is significant ($F_{(4,48)}$ = 7.8, p < 0.001), but not the main effect of regions of interest and interaction ($F_{(5.2,68.9)}$ = 0.8, p = 0.50, $F_{(7.6,90.9)}$ = 0.7, p = 0.65, respectively).

Three areas, the extrastriate visual cortex, the cuneus and the left frontal eye field are implicated in the representation of information about the response time throughout the decision process, indicating a sustained process involved for response speed in these areas. Overall, the main effect of time-window is significant ($F_{(4,48)}$ = 7.9, p = < 0.001), the main effect of region of interest and the interaction are not ($F_{(3.1,38.3)}$ = 0.43, p = 0.74, $F_{(7.5,90.5)}$ = 0.74, p = 0.64, respectively).

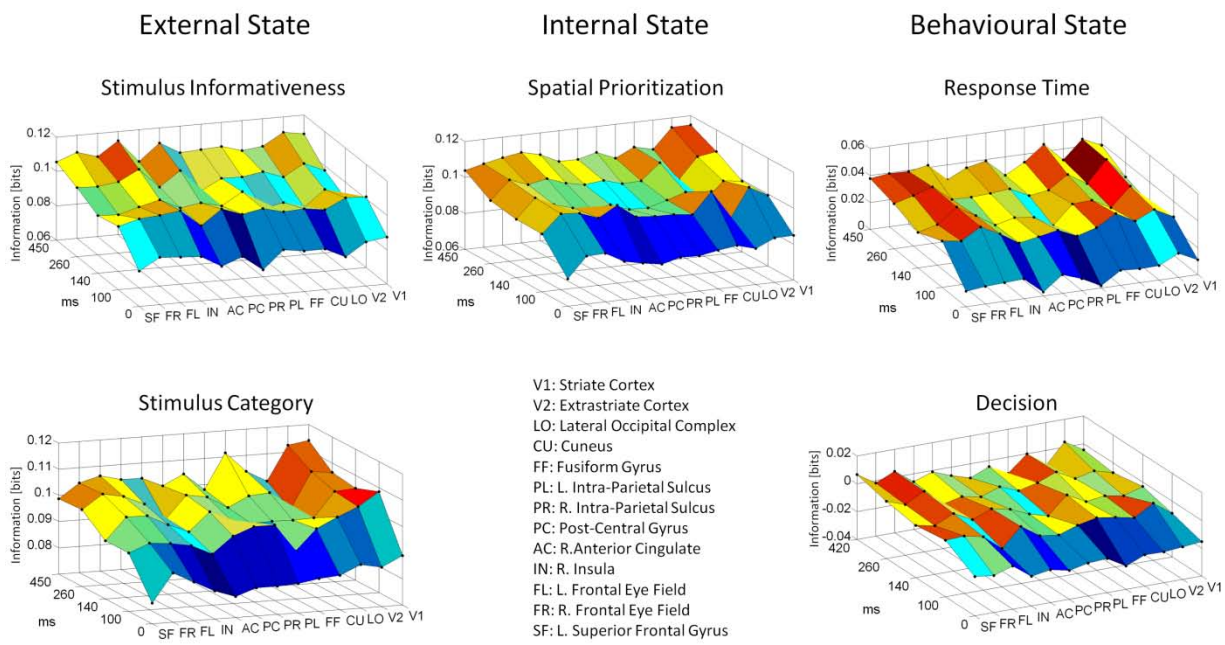


Figure 6.14 Spatiotemporal information representation. The data is ordered columnwise according to the variables of interest, external, internal and behavioural state. For each variable of interest, data from all joint distributions of all features of interest combinations, i.e. EEG time-windows ×fMRI regions of interest is displayed. The vertices (black dots) of the surface reflect the group averages.

Finally, comparing the representation of the decisional state to the representation of the physical stimulus category shows some differences: first, for the physical stimulus category, both high and low level areas show larger information estimates both early and late during the decision process, while for the observer's decisional variable, this effect is stronger for the high level areas (superior frontal gyrus and left frontal eye field). Interestingly, the most positive deflection for some areas can be observed for mid-temporal time-windows, as for example in the anterior cingulate and the fusiform gyrus. However, it has to be noted that, given the supra-threshold nature of the experimental design, a clear dissociation between the subjects' perceptual state and the physical stimulus property at low spatial coherence cannot be obtained in the framework of the current study. Statistical evaluation revealed the usual pattern of significant main effect of time-window ($F_{(4,48)} = 4.1$, $p = 0.005$) and non-significant

effects of region of interest ($F_{(4.8,\ 58.3)}$ = 0.5, p = 0.75) and interaction ($F_{(7.7,\ 92.7)}$ = 1.3, p = 0.24).

In summary, the following picture of spatiotemporal representation of information for visual perceptual decisions emerges: with respect to the external state variables, both low and high level cortical areas are involved in the representation of information with a temporal rebound pattern mainly observed for the informativeness of the stimulus. For the stimulus category, both high and low level areas increase the information content over time, the specific areas being complementary to those implicated in the representation of stimulus informativeness. Regarding the representation of information about the observer's internal state, additional mid-level cortical areas appear of relevance. A set of three brain regions is informative about the observer's response time throughout and the decision process. Finally, with respect to the categorical decision, the data indicate a stronger involvement of high level cortical areas over time compared to the representation of the physical stimulus category, which implicated both higher/anterior and lower/posterior areas.

6.4 Discussion

In summary, the current study supports the view of the brain representing information about external, internal and behavioural states in a highly distributed, parallel, and dynamical manner. In concordance with this view, no single brain region or single time-point in the first 500 ms of the perceptual decision process was identified to be of sole relevance. In general, most information is represented in both low (visual cortex) and high (frontal cortex) level regions towards the time of the execution of the decision, with the possible exception of information about the internal state. Finally, some dissociation between the representation of

the physical stimulus category and the observer's perceptual interpretation could be identified with a shift of information representation to higher cortical areas in the latter.

What does the current study add with respect to previous studies on visual perceptual decision making? First, in employing an information theoretic framework, the emphasis of the current study is on the information that is represented in the neuronal response on the single-trial level, not averaged over multiple observations. It is hence describing the perceptual decision process at the ecologically most meaningful level, as the brain has to make optimal decisions upon single representation of the perceptual evidence. Second, using simultaneous EEG-fMRI recordings, the current study uses state-of-the art brain imaging methodology to assess the joint EEG-fMRI signal feature probability distributions. Most previous studies (with the exception of (Mulert et al., 2008)) employed non-simultaneous or one modality data acquisition schemes and hence are susceptible to between session effects, such as changes in the observer's vigilance, attention or learning effects. While not a focus of the current communication, future evaluations of the same data set will also allow insight to be gained into the between-modality dependencies that contribute to information encoding. Third, the current study explicitly manipulated the observer's internal state by adding a spatial prioritization/attention component to the perceptual decision process. In (Philiastides and Sajda, 2007), the authors proposed a spatio-temporal diagram of the processes involved in perceptual decision making based on an EEG-informed analysis of an fMRI data. The current study proposes the following additions to this scheme: a) the regions implicated in early temporal visual perception are modulated by the observer's internal state and represent both top-down and bottom up factors of perceptual decisions and b) the implication of higher cortical areas in the representation about the observer's decision emphasizes the idea of recurrent feedback loops in the entire network.

Some notes of caution on the interpretation of the results of this study are necessary. First and foremost, the problem of entropy estimation, or in other words information bias correction, for EEG-fMRI experiments remains unresolved. However, reasonable precautions have been taken not to overestimate information based on the PT- and shuffling correction schemes as well as probabilistic null models. The main focus of this study is also not so much on the absolute, but on the relative information content between different signal features and feature combinations (see also Chapter 4, section 4.4). Nevertheless, future applications of the information theoretic framework to EEG-fMRI data sets should strive to optimize entropy estimation for the specifics of continuous, analogue-type signals (Magri et al., 2009). Second, any analysis of univariate features is sensitive to the feature selection process. Here, a route informed by the signals group averages was taken. For some of the comparisons, namely those in which the selection criteria were not orthogonal to the comparison of interest, this entails the danger of circular analysis, which was noted in the discussion of the results (Kriegeskorte et al., 2009). Further, the focus of this study is on the signal features amplitude in the time-domain, and many other features (e.g., EEG frequency components or HRF basis function parameters) are conceivable. Finally, especially with respect to the observer's perceptual state, near-threshold paradigms might be more powerful to elucidate spatiotemporal dissociations in the representation of physical and perceptual information. While in the current study this was partly achieved by focussing on the low informative stimulus trials, future studies using even less physically informative stimuli might be more suited to this question. Finally, only the first 500 ms of the perceptual decision process were assessed, and often the highest information estimates were obtained for the final time-window. The focus on the first 500 ms is partly justified by the fact that the observer's response has been made by this time-point and by the behaviour of the grand-average ERP, which returns to approximately baseline at this time. However, working memory and error

monitoring processes following the decision and response presumably require information representation about the perceptual decision process. Future studies might elucidate the brain's spatiotemporal information representation profile with respect to these.

Has the combined acquisition of EEG and fMRI data and analysis in an information theoretic framework expanded our understanding of the neurobiological correlates of the perceptual decision process? In this regard, it has to be noted that the construction of joint EEG-fMRI feature distribution and hence the evaluation of their combined information representation is only possible from combined recordings. Also, the information theoretic framework applied extends beyond commonly applied EEG-fMRI GLM analysis (e.g. (Karch and Mulert, 2010; Mulert et al., 2008)), in that it asks, how well certain stimulus or behavioural attributes can be dissociated from a single trial of the observed response, and not only which areas show temporal co-variation between EEG and fMRI features. The experiment and analyses reported here hence represent the most principled approach to the question of information representation about the perceptual decision process to date. The results are necessarily complex, but reinforce and extend previous ideas about the spatiotemporal neurobiological correlates of perceptual decisions. A question that has not been a focus of this chapter regards the role of the context-dependency of EEG-fMRI noise correlations, their link to neurovascular coupling processes as well as the underlying neurophysiology. However, the data acquisition and analysis framework reported here open the door to future explorations of these important questions.

In summary, the current study reinforces the notion of networks of brain areas dynamically being involved in the process of visual perceptual decisions. As the methodology employed does not allow the direction of information flow to be inferred, this stresses the need for comprehensive spatiotemporal models of functional/effective connectivity in the

analysis of combined EEG and fMRI data. The data discussed here might serve as a guide for the spatiotemporal complexity these models will need to achieve.

7 General Discussion

The general discussion is arranged as follows: after summarizing and discussing the main contributions of this thesis, the approach to EEG-fMRI integration proposed here is compared to other approaches in the literature. The thesis closes with some proposals on the future development of the application of information theoretic concepts to combined EEG and fMRI data.

7.1 Contributions

Comprising both theoretical and experimental aspects, this thesis has contributed to the EEG-fMRI literature on two levels: the methodological and the neuroscientific, each of which will be discussed in turn.

7.1.1 Methodological contributions

The main contribution of this thesis is the conceptual and practical application of an information theoretic approach to the analysis of simultaneously acquired EEG-fMRI data. While proposed for the case of invasive electrophysiological recordings and functional brain imaging signals in (Panzeri et al., 2008), to the author's knowledge this thesis comprises the first conceptual and practical application of the information theoretic framework to combined EEG-fMRI data. On the theoretical level, this application allows the precise definition and differentiation of the colloquial terms 'information' and 'informative signal' which are prevalent in the EEG-fMRI literature (Bagshaw and Warbrick, 2007; Mulert and Lemieux, 2010). Most notably, the framework allows the differentiation between signal features which are informative about the external stimulus they represent, both in isolation or in combination with the complementary imaging modality, and those which are informative about other

signal features, both in the presence or absence of external stimulation. Accordingly, this thesis emphasizes that signals from the EEG and fMRI modalities can be integrated with respect to the stimulus-related information they provide, $I_N(S; R_1, R_2)$, or the information they provide with respect to each other, $I_N(R_1; R_2)$ and $\langle I_N(R_1; R_2|S)\rangle_S$, as dicussed in Chapter 2. The approach is particularly tailored to simultaneous EEG-fMRI recordings, as the joint stimulus-response signal distribution for non-independent response signal features can only be estimated based on simultaneous recordings. It is important to note that the same information theoretic quantities can be evaluated from separate EEG and fMRI data acquisition sessions only under the assumption that the respective EEG and fMRI features are independent, i.e. that their joint distribution can be approximated by the product of the marginal distributions.

Upon conceptually establishing the information theoretic framework in Chapter 2, Chapter 3 investigated the sensitivity of the set of information theoretic quantities to a set of simulated scenarios that were intended to have some resemblance to experimental EEG-fMRI settings. Specifically, these simulations allowed the strong theoretical appeal of the framework to differentiate different assumptions about EEG-fMRI stimulus-response signal relationships to be demonstrated. Furthermore, the simulations revealed some issues (bias correction, probability distribution estimation) that need to be considered in the practical application of the framework. The data simulation approach that was employed identify these issues was the use of linear Gaussian model simulations. Linear Gaussian models are very prominent in the study of brain imaging signals, especially fMRI, as they form the basis of the mainstay of fMRI data analysis, the general linear model. Given the discrete nature of classical invasive electrophysiological data (spike-counts), these models have so far not achieved great attention in the information theoretic literature. One contribution of this thesis is to bring together these two lines of thought.

Chapter 4 demonstrated the application of the information theoretic framework to a basic visual stimulation experiment. It was noted that in the practical application, different response features can dominate the information content in the joint response distribution due to their different SNRs. This is an important insight with respect to future simulation studies aimed at optimizing the quality of information estimation. On the methodological level, Chapter 4 also demonstrated the use of the information theoretic framework for the question of EEG data pre-processing. The question of EEG data quality from combined EEG-fMRI recordings is still a prominent line of research, and while some progress has been made, the EEG data recorded inside the MR environment after pre-processing is still not necessarily of equal quality to EEG data recorded in isolation. Especially with regard to the question of single-trial data quality, the framework applied here allows a principled investigation of the influence of different data pre-processing strategies on the informativness of the recovered data.

Chapter 5 extended the application of the framework to the spatial domain, by instead of using single electrode and fMRI ROI data as in Chapter 4, introduced a voxel-wise information estimation scheme. With respect to more conventional questions of EEG-fMRI data integration, the analysis of EEG and fMRI signal features upon projection into a common anatomical space using a current density approach for EEG data has not been explored in any great detail so far. This is surprising, as the voxel-wise dependencies of EEG and fMRI features form a natural counterpart to the invasive study of local electrophysiological and haemodynamic dependencies. While conceptually appealing and sensible from the viewpoint of using all EEG and fMRI data rather than a highly constrained pre-selected data subset, this approach is not without its limitations. Specifically, given the ill-posed nature of the EEG inverse problem, the estimated electrophysiological single-trial voxel time-courses and frequency power estimates are conditional on the validity of the forward and inverse solutions

employed. Improvements in the application of information theoretic voxel-wise EEG-fMRI integration are hence to be expected from improved source reconstruction methods in the EEG domain.

Finally, with respect to the study of one of the core cognitive functions of the brain, Chapter 6 proposed the conceptualization of experimental modulations into external, internal and behavioural state variables, in order to achieve a principled approach to study the neurobiological information encoding during perceptual decisions. Chapter 6 introduced the inclusion of behavioural variables into the information theoretic framework, thereby increasing the dimensionality of the joint distribution that required estimation from two (EEG and fMRI) to three (EEG, fMRI and behaviour). While conceptually straightforward, practically this has the consequence that the augmented space is less densely sampled than its lower-dimensional counterpart, given approximately the same amount of experimental observations. This fact is known in the machine learning literature as the curse of dimensionality (Bishop, 2006). Future applications of the information theoretic framework to higher dimensional response signal spaces (as for example in multi-voxel pattern fMRI analysis (Norman et al., 2006)) will need to address this issue, and can potentially benefit from dimensionality-reduction using parameterized probabilistic models.

Throughout this thesis, the problem of accurately estimating the information theoretic quantities (entropy and mutual information) has been addressed. Using both simulated and experimental data sets, it has become obvious that the estimation or bias correction schemes developed in the application of information theoretic concepts to invasive electrophysiological data are sub-optimal for the case of non-invasive brain imaging data. Moreover, their use required additional, relatively unprincipled ad-hoc solutions, such as the application of probabilistic null models in Chapters 4, 5, and 6. Also, a recently proposed method for the estimation of information from analogue brain signals (Magri et al., 2009) has

been shown to be not fully appropriate in Chapter 3. It is one important contribution of this thesis to have identified these practical difficulties and to hint at the methodological improvements that need to be achieved for a more principled application of information theory to combined EEG-fMRI data. Some proposals in this respect are discussed in the final part of this general discussion.

7.1.2 Neuroscientific contributions

With respect to the checkerboard stimulation experiment, the neuroscientific contribution of this thesis can be summarized as follows: while both EEG and fMRI time domain features are informative about stimulus contrast on the single-trial level as shown previously (Boynton et al., 1999; Shawkat and Kriss, 2000), no feature combination could be identified which exhibited strong single-trial co-variation between the modalities (Chapters 4 and 5). This is surprising, as the visually evoked responses from both modalities are thought to reflect at least partially overlapping neuronal activity (Logothetis et al., 2001; Wandell et al., 2007) and both signals show the same stimulus modulation (e.g. with contrast). However, due to the dominance of either one feature in the information contained in the joint distribution (i.e. in Chapter 4 the fMRI features, in Chapter 5 the FSS pre-processed EEG data), the observed informativeness of the joint distribution with respect to the stimulus is not mirrored by consistent activity dependence.

What can explain the observed single-trial mismatch? One argument is that the set of signal features analyzed within this thesis is not exhaustive, and many more features (for example, time-domain integral features (Porcaro et al., 2010), features from the EEG frequency domain, or single trial model-based (basis function parameters) features) could be considered. Another argument concerns the still elusive relationship between neural activity,

EEG potentials recorded at the scalp and haemodynamic responses. It is very well conceivable that the electrophysiological and haemodynamic features considered here in fact do not reflect the same underlying neuronal activity. This would even provide a stronger motivation for the acquisition of combined EEG and fMRI data to study the neuronal underpinnings of cognitive processes, as both imaging modalities could be used to describe complementary facets of the neural response. There is some evidence from combined invasive electrophysiological and imaging studies, that the BOLD signal mainly reflects local field potentials, while the EEG evoked potential is closer related to multi-unit activity (Logothetis et al., 2001; Logothetis and Wandell, 2004; Whittingstall et al., 2007). Based on these results, combined EEG-fMRI would hence be able to reflect the two most relevant electrophysiological signals in a complementary manner, and it could naturally be expected that dissociations at the level of MUA and LFP are reflected as dissociations between the EEG and fMRI signal.

With respect to the neurobiological underpinnings of visual perceptual decisions, the results of Chapter 6 reinforce the notion that the brain represents information in a dynamical and distributed fashion. Using the combination of simultaneous EEG-fMRI data acquisition and the information theoretic framework, the results of this chapter provide evidence for dissociations in both the temporal and spatial domain with respect to the experimental variables of interest. For example, in the temporal domain, it was found that internal states are represented prior to external states, while in the spatial domain, the observer's decision is more dependent on higher cortical areas than the representation of physical evidence. Extending beyond commonly applied EEG-fMRI GLM approaches to identify cortical regions involved in perceptual decisions, the theoretical and experimental framework allowed a characterization about where and when certain stimulus or behavioural attributes can be dissociated from a single trial of the neurobiological response. Overall, however, the results

strongly support a dynamic network view of the brain (Friston, 2010). They might hence serve as basis for future model-driven analyses of the same data set in an effective connectivity framework with the aim of a more comprehensive view of the neurobiological dynamics underlying perceptual decision making (David et al., 2006; Friston et al., 2003; Kiebel et al., 2006). A first step in this direction will be the analysis of the EEG and fMRI data sets in a dynamic causal modelling framework (Friston, 2007), based on corresponding anatomical nodes, i.e. the regions of interest considered in Chapter 6. Not only might this elucidate the role of pathway modulations in perceptual decisions, but also shed some light on the relationships of effective connectivity estimates obtained from electrophysiological and haemodynamic data.

7.2 Comparison to other approaches for EEG-fMRI integration

The question arises what the current treatment of multi-modal brain imaging data adds with respect to previous integrative approaches. A full formal treatment of this question would require an analytical comparison of the different measures of dependencies for random variables employed by these approaches (i.e. mutual information in comparison to linear correlation and regression) conditioned on different assumptions about the actual relationships, which is beyond the scope of this discussion. Nevertheless, some similarities and dissimilarities with other approaches should be pointed out.

In Chapter 1, Section 1.5 previous signal integration approaches were categorized as follows (Kilner et al., 2005): 1) Integration through prediction (using a property of the EEG signal to predict changes in the BOLD response in the statistical framework of the general linear model), 2) integration through constraints (e.g. constraining EEG source localization based on fMRI results), 3) and integration through common forward models (a

comprehensive model or set of models which describes the mechanism generating the two signals and the links between them).

With respect to approach 1) and response signal–signal relationships, the framework proposed in this thesis is more tailored to explicitly address the question of which signal features are mutually dependent, initially free of additional assumptions. The success of approach 1) is based on the assumption that extracted EEG features linearly co-vary with a pre-defined haemodynamic response function under additive Gaussian noise. This assumption is validated by obtaining meaningful statistical parametric maps (Debener et al., 2006; Eichele et al., 2005; Mayhew et al., 2010). The framework proposed here explicitly states quantitative (effect sizes) degrees of dependency, while in the current implementation it is based on some prior feature extraction method. In this vein, the information theoretic framework also allows the explicit differentiation of event-related and non event-related signal co-variation (Debener et al., 2006; Herrmann and Debener, 2008) by means of the stimulus unconditional and stimulus conditional mutual information measures, which is usually not addressed in approach 1). Lastly, also with respect to stimulus–response signal relationships, the framework proposed here provides a unified approach using the same mutual information quantities as for the response signal–signal relationships.

With respect to approach 2), again the success of this method is based on meaningful co-localization results (Henson et al., 2010; Whittingstall et al., 2007), while the evaluation of the framework presented here is more explicit. A potential application of the measures proposed here within approach 2) might be the evaluation of obtained co-localization results with respect to the various mutual information measures, or indeed the use of mutual information measures explicitly as constraint of the source localisation. The information

theoretic analyses in a common anatomical framework reported in Chapter 5 can be regarded as a first step into this direction.

Finally approach 3), the integration of both modalities through common forward models fitted to empirical data in a full Bayesian treatment, appears to be the most principled approach to real data fusion, although probably the most distant in terms of practical implementation (Daunizeau et al., 2010). The framework presented here can add to this line of research by explicitly taking into account the response signal probability distributions. Studying the signal feature distributions of real data sets as well as their relationships can potentially be useful to the more realistic specification of prior distributions in a Bayesian generative model context.

7.3 Future directions

This thesis employed an information theoretic framework developed in the analysis of invasive electrophysiological spike-count data for the analysis of continuous brain imaging signals. On the conceptual level, the information partitioning schemes adapted from the analysis of electrophysiologial data evaluate how much information an observer of the neuronal response can obtain about the stimulus. However, the focus of an information theoretic approach to EEG-fMRI is slightly different, since other aspects are also relevant. For example, it is important to know whether EEG and fMRI signals complement each other in the spatiotemporal representation of information. These questions need to be investigated from a theoretical standpoint, by formulating an information partitioning scheme for multi-modal brain imaging signals.

Another central aim for the future application of information theoretic concepts to combined EEG-fMRI must be the theoretical and practical development of analytical methods

that are tailored to the particularities of non-invasive brain imaging data. Specifically, this encompasses the continuous Gaussian nature of the signals and the experimental restriction to relatively few trials and stimulus classes.

On the practical level of estimating information theoretic quantities from EEG-fMRI data it has been proposed to facilitate entropy estimation using the analytical differential entropy of Gaussian distributions, whose covariance matrices can more readily be estimated from experimental data than full probability distributions (Magri et al., 2009). However, this is based on the assumption that the stimulus-response signal distributions are Gaussian, which is not necessarily the case for small number of stimuli. An appropriate conceptual extension of Gaussian distributions in this case are Gaussian mixture models (McLachlan and Peel, 2000). Entropy estimation based on Gaussian mixture models has so far not been explored for brain imaging signals, but given the analogue nature of the data, this holds great potential.

Specifically, a Gaussian mixture approach for information estimation from combined (i.e. at least bivariate) EEG-fMRI response features would proceed as follows: in a first step, a single multivariate Gaussian density is fitted to each stimulus conditional distribution using maximum likelihood. In the second step, the entropies of these conditional distributions and the entropy of the mixture of Gaussians across stimuli are evaluated based on the differential entropy solutions of Gaussian random variables. In the third step, the analytical entropy bias estimates for a Gaussian mixture with $K \in \mathbb{N}$ components and $N_S \in \mathbb{N}$ experimental trials is determined and subtracted from the estimated differential entropies. Finally, the information theoretic quantities are evaluated as before. The crucial theoretical advancement that is required for this scheme is the analytical determination of the entropy bias of a Gaussian mixture model. This is complicated by the fact that no closed form for the entropy of a Gaussian mixture exists, due to the summation over components in the logarithmic term (Huber et al., 2010). However, by combining the theoretical results regarding entropy

estimation in (Misra et al., 2005) and the analytical approximation of Gaussian mixture models in (Huber et al., 2010) this is achievable and might also prove efficient in higher dimensional spaces (e.g. with applications to fMRI multi-voxel pattern analysis).

From a neuroscientific viewpoint, future work on existing or novel data sets should specifically address the question of EEG-fMRI co-variation and its contribution to information encoding. For example, it would be desirable to be able to show that EEG-fMRI feature co-variation is context-dependent. First, this would demonstrate the full power of the information theoretic framework in the application to EEG-fMRI. Second, and more importantly, this would pave the way to demonstrate synergistic effects of combined EEG-fMRI recordings. These in turn would argue strongly for the sensitivity of EEG and fMRI to complementary aspects of the neuronal response. Given that so far, at least for the visual system, no single EEG feature time-domain feature (i.e. aspect of the VEP) could be identified that strongly co-varies with the BOLD signal of visual cortex in the absence of differential stimulation, it is very well conceivable that EEG and fMRI do not simply reflect the same underlying neurophysiology in an evoked response context (Becker et al., 2010). Being able to show that EEG and fMRI measure indeed reflect different aspect of the neurobiological response would hence provide strong motivation for their combined (simultaneous or non-simultaneous) use in the study of the human brain.

Finally, as noted in the introduction of Chapter 5, some experimental evidence exists on the co-variation of EEG α-frequency power (e.g. de Munck et al., 2007, Wu et al., 2010) and the BOLD signal, as well as co-variations between different EEG frequency power bands and the BOLD signal (de Munck et al., 2009). An interesting research question for future studies on EEG-fMRI neurovascular coupling hence concerns the relationship between visually evoked potentials and ongoing or induced α-frequency power on the one hand (Becker et al., 2008), and the relationship of both to the BOLD response on the other. The

information theoretic approach described here could be applied in this context, by explicitly dissociating activity and conditional dependencies between the various features.

7.4 Conclusion

The central theme of this thesis is to view combined EEG and fMRI data from an information theoretic perspective, i.e. to identify relationships of experimental and signal variables of interest and quantify their non-independence using mutual information. The work hence is based on a model-free, low-complexity application of probabilistic concepts to the data analysis of non-invasive brain imaging signals. A number of technical difficulties that will need to be overcome in the quantification of the signals' informativeness have been identified, which are very close to the very nature of scientific inference, namely the estimation of unknown probability distributions.

Based on this thesis, future studies will improve the methodology for estimated information of neuroimaging signals, address the formal equivalences between neuroimaging analyses of the scientific mainstream and information theoretic approaches, and establish the estimation of information as a principled data analytical method in cognitive functional neuroimaging.

Appendices

Appendix A: Fundamentals of electromagnetism

The aim of this Appendix is to provide an introduction to the fundamentals of electromagnetism, including the notions of *electric field, electric potential, electric current, electric current density, magnetic field* and *magnetic flux*. The discussion is held intuitively, rather than providing a formally rigorous account. This Appendix largely follows a standard textbook account of electromagnetism (Kraus J.D. and Fleisch D.A., 1999).

Electric field

The electric field of an electric charge Q_1, the basic electrical quantity measured in Coulombs $[C]$, is defined as the as the force that Q_1 exerts on a unit charge. A charge Q_2 $[C]$ at a given distance $r[m]$ in a medium with dielectric constant $\varepsilon[F/m]$ experiences a vector force given by Coulomb's law as

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon^2} r \tag{A.1}$$

where $r \in \mathbb{R}^3$ represent the distance vector between Q_1 and Q_2. As the electric field is defined per unit charge, scalar division of (A.1) by the charge Q_2 yields the electrical field measured in Newtons per Coulomb, or Volts per metre.

$$E = \frac{F}{Q_2} = \frac{Q_1}{4\pi\varepsilon^2} r \tag{A.2}$$

The electric field is a vector field, i.e. a mapping $E: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, as it allocates a force vector to each point in space surrounding the charge Q_1.

Electric potential

The EEG measures time-varying electric potentials, measured in Volts, between recording electrodes and a reference electrode applied to the scalp surface. The electric potential between two points $a \in \mathbb{R}^3$ and $b \in \mathbb{R}^3$ in space is defined by the potential difference, i.e. the work per unit charge that is done when moving a hypothetical test charge against an electric field E. Informally, the potential difference between points a and b can be written

$$\Delta V = E\Delta x \tag{A.3}$$

where $\Delta x \in \mathbb{R}^3$ denotes the difference vector between a and b. Formally, the potential between two points is given as the line integral

$$V = \int_a^b E \cdot dL \tag{A.4}$$

between the two points. Here, $E \cdot dL$ denotes the dot product between the electric field vector and the direction vector dL between the points a and b. The absolute potential at a given point in space is given by hypothetically moving a test charge from infinity to that point c in space, i.e. by the line integral

$$V = \int_\infty^c E \cdot dL \tag{A.5}$$

The electric potential is a scalar field, i.e. a mapping V: $\mathbb{R}^3 \rightarrow \mathbb{R}$, as it allocates to each point in space a scalar quantity. This quantity is the electric potential at that point due to the electric field of a charge of a specific value. According to the linear superposition principle, the electric potentials due to more than one charge are given by the algebraic sum of the individual potentials at that point. The definition of the electric potential as a scalar field

$V: \mathbb{R}^3 \rightarrow \mathbb{R}$ also allows the electric field to be defined as the gradient of the potential field, i.e. the function of spatial partial derivatives of the potential

$$E = grad\, V = \begin{pmatrix} \frac{\partial V}{\partial x} \\ \frac{\partial V}{\partial y} \\ \frac{\partial V}{\partial z} \end{pmatrix} \tag{A.6}$$

where $\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}$ and $\frac{\partial V}{\partial z}$ denote the partial derivatives of the potential V in x, y and z direction, respectively.

Electric current and electric current density

Electric current constitutes the movement of electric charge in a conductive medium. In biological systems such as the human body, current is carried by charged atomic particles, such as sodium, potassium, calcium or chloride ions. Electric current is defined as the number of particles of volume density $\rho\,[Cm^{-3}]$ passing the uniform cross section $A\,[m^2]$ of a conductive medium at a given reference point per second. Electric current is measured in Coulombs per second $[Cs^{-1}]$ or Ampere $[A]$ given by

$$I = v_d \rho A \tag{A.7}$$

Here, v_d is the charge drift velocity given by the product of the accelerating force of the local electric field E and a medium-dependent mobility constant, and is measured in metres per second $[ms^{-1}]$.

The current density, an important quantity in the context of the EEG inverse problem, is derived from equation (A.7) by division by the area

$$J = \frac{I}{A} = v_d \rho \tag{A.8}$$

Thus, current density, measured in Amperes per square metre $[Am^{-2}]$, denotes the flow of electric current per unit area. Alternatively, the charge drift velocity can be viewed as being determined by the potential difference between the points that enclose the conductive medium. In short, given the potential V over the length of a conductor with resistance $R[\Omega]$, the current through an ideal conductor (i.e. a conductor whose resistance is independent of its current) is given by Ohm's law according to

$$I = \frac{V}{R} \tag{A.9}$$

Magnetic fields

A conductor carrying a current I is surrounded by a magnetic field H, measured in Amperes per metre $[Am^{-1}]$. As for the electric field E, the magnetic field H is a vector field $H: \mathbb{R}^3 \to \mathbb{R}^3$, allocating a force vector acting on magnetic objects to each point in space. According to the Biot-Savart law, the differential magnetic field dH at a point $p \in \mathbb{R}^3$ evoked from a short section dL of a current-carrying conductor with current I is given by

$$dH = \frac{IdLsin\theta}{4\pi r^2} \tag{A.10}$$

where r indicates the distance between the conductor and the respective point $p \in \mathbb{R}^3$ and θ the angle of the conductor section dL with respect to the point $p \in \mathbb{R}^3$.

Magnetic flux

The magnetic flux through a surface area is obtained by integrating the normal component of the magnetic field multiplied by a medium-dependent permeability component μ over the area (ds) and is measured in Webers $[W]$

$$\psi_m = \iint \mu H \cdot ds \tag{A.11}$$

which for a uniform field assumes the simple form

$$\psi_m = \mu HA \tag{A.12}$$

where $A[m^2]$ indicates the area over which the magnetic field H extends. Normalizing the magnetic flux per unit area yields the magnetic flux density B, measured in Webers per square metre $[Wm^{-2}]$, or Teslas $[T]$

$$B = \frac{\psi_m}{A} \tag{A.13}$$

The magnetic flux density is a well-known characteristic of the MR scanner static field and is also commonly referred to as magnetic field. An important artefact of combined EEG-fMRI recordings is the so-called image acquisition or gradient artefact, which is discussed in detail in section (1.4.2.1). The image acquisition artefact is the manifestation of electromagnetic induction in the EEG recording wires caused by fast changing magnetic fields. In brief, a changing magnetic flux ψ_m through a closed loop produces a voltage V at the loop terminals given by

$$V = -\frac{d\psi_m}{dt} \tag{A.14}$$

where the voltage is the integral of the induced electric field E around the loop. The relationship between loop area (integration over ds), change in magnetic flux density over time $-\frac{\partial B}{\partial t}$ and induced electric potential is stated by Faraday's law as

$$V = \oint E \cdot dL = -\iint \frac{\partial B}{\partial t} \cdot ds \tag{A.15}$$

where $\oint E \cdot dL$ denotes the line integral of the electric field E around the loop, measured in Volts, and, equivalently, $\iint \frac{\partial B}{\partial t} \cdot ds$ is the surface integral of the rate of change of magnetic

flux $\frac{\partial B}{\partial t}$ over the loop area A. If the loop is closed, the induced voltage causes a current to flow.

Appendix B: Introduction to probability theory

The aim of this Appendix is to present the most fundamental and relevant concepts from probability theory for the development of information theory and lay the foundation for the practical application of these concepts to the experimental situation under consideration, namely the analysis of combined EEG-fMRI data.

Modern probability theory is grounded like all modern mathematics in the two concepts of sets (collections of well-dissociable entities) and mappings (rules that allocate members of one set to members of another set). Specifically, since the work of Borel and Kolmogorov (Borel, 1963; Kolmogorov, 1974), probability theory is often regarded as a branch of measure theory, where measure theory is concerned with the question of how to quantify the sizes of sets. In particular, a frequentist view of probability theory is concerned with modelling random experiments using measure theory. Random experiments are understood as processes whose outcomes are uncertain, but which can (at least in principle) be repeated infinitely often. The acquisition of EEG and/or fMRI data in response to the repeated presentation of a stimulus can be regarded as a random experiment, motivating the application of both probability and information theory to its study. The fundamental mathematical models of a random experiment are the *probability space* and the *random variable*. Probability spaces dissociate the notions of outcome and events in random experiments, and formulate a model of probability based on a *measure*, i.e. a mapping that allocates real numbers to sets. *Random variables* are mappings that allow probability measures to be transferred between sets. The definition of a *probability distribution* follows directly from those of probability spaces and random variables. Eventually, the primary goal of the current section is to systematically develop and prepare the introduction of *probability density* and *probability mass functions*, which form the basis of the development of the central information theoretic concepts of *mutual information* and *differential mutual information* in section 2.2.

The following discussion follows standard texts on the measure theoretic development of probability theory, here especially (Moeschlin O.et al., 2003; Skorokohd, 2005)

B.1 Probability spaces and random variables

Mathematically, a random experiment is modelled by a probability space, which is defined as follows:

Probability space (B.1)

A triple $(\Omega, \mathcal{A}, P)$ is called probability space, if the following conditions hold

i) Ω is a nonempty set

ii) $\mathcal{A}$ is a σ-Algebra on Ω

iii) $P: \mathcal{A} \to \mathbb{R}$ is mapping with the following properties

1. $P(A) \geq 0 \ (A \in \mathcal{A})$
2. $P(\Omega) = 1$
3. For every sequence $(A_n | \, n \in \mathbb{N})$ of pair-wise exclusive sets from $\mathcal{A}$, the following condition (σ-additivity) holds

$$P(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$$

P is called a probability measure on the σ-Algebra $\mathcal{A}$.

Here, a sequence of sets is a special case of a family of sets: If I is a nonempty set of indices, $\mathcal{K}$ is a system of sets (defined as a set of sets) and $f: I \to \mathcal{K}$ is a mapping, which allocates to each $i \in I$ an element $f(i) := A_i \in \mathcal{K}$, then $(A_i | \, i \in I)$ is called a family of sets from $\mathcal{K}$. In the special case that the nonempty set of indices is $\mathbb{N}$, the family of sets $(A_i | \, i \in \mathbb{N}) := (A_i)$ is called a sequence of sets. A sequence of sets $(A_i | \, i \in \mathbb{N})$ is called pair-wise exclusive if $A_i \cap A_j = \emptyset \ (i, j \in \mathbb{N}, i \neq j)$. Finally, $\sum_{n \in \mathbb{N}} A_n$ denotes the infinite, countable, disjunctive union of the pair-wise exclusive sets A_n, $n \in \mathbb{N}$ (also called the sum of the sets A_n, $n \in \mathbb{N}$)

defined as $\sum_{n\in\mathbb{N}} A_n \coloneqq \bigcup_{n\in\mathbb{N}} A_n$. The definition of a probability space draws on the concept of a σ-Algebra, which is defined below.

σ-Algebra (B.2)

A system $\mathcal{A}$ of subsets of a set Ω is called a σ-Algebra on Ω, if the following conditions hold

i) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

ii) For every sequence (A_n) of sets in $\mathcal{A}$ the union $\bigcup_{n\in\mathbb{N}} A_n$ is an element of $\mathcal{A}$

Here A^c denotes the complement of A with respect to $\mathcal{A}$, defined as

$$A^c \coloneqq \mathcal{A} - A \coloneqq \{x \in \mathcal{A} | x \notin A\} \tag{B.3}$$

If $\mathcal{A}$ is a σ-Algebra on Ω, then $(\Omega, \mathcal{A})$ is called a measurable space. The elements of $\mathcal{A}$ are called measurable sets or events.

The most important σ-Algebra for the purposes of this thesis is the Borel σ-Algebra $\mathcal{B}^n$, which is the standard σ-Algebra on $\mathbb{R}^n$. The Borel σ-Algebra is generated by the system $\mathcal{I}^n$ of all finite semi-closed intervals on $\mathbb{R}^n$, defined as the system of all sets of the kind $[a, b[= \{x \in \mathbb{R}^n | a \leq x < b\} \subset \mathbb{R}^n, (a, b \in \mathbb{R}^n, a \leq b)$. A σ-Algebra is said to be generated by a set $\mathcal{K}$ as follows: Let $\mathcal{K}$ be system of sets on a set Ω. Then there exists a σ-Algebra on Ω which contains $\mathcal{K}$, namely the power set $\mathcal{P}(\Omega)$, the set of all sets in Ω, which, as can be shown, is always a σ-Algebra on Ω. If one defines $\sigma_\Omega(\mathcal{K})$ as the intersection of all σ-algebras on Ω which contain $\mathcal{K}$, then this intersection is a σ-algebra and is the smallest σ-algebra on Ω which contains $\mathcal{K}$. In other words, if $\mathcal{A}$ is an arbitrary σ-Algebra on Ω which contains $\mathcal{K}$, it follows that $\mathcal{K} \subset \sigma_\Omega(\mathcal{K}) \subset \mathcal{A}$. An intuitively helpful but formally false notion of $\mathcal{B}^n$ is to

consider it the equivalent of the power set of $\mathbb{R}^n$, i.e. the set of all sets in $\mathbb{R}^n$. The standard set and σ-algebra considered for the purposes of the thesis is $\Omega = \mathbb{R}^n$ and $\mathcal{A} = \mathcal{B}^n$.

As a model of a random experiment, a probability space can be interpreted as follows: the set Ω represents the outcomes of the experiment, while the σ-Algebra $\mathcal{A}$ represents events, i.e. collections of outcomes. For example, when recording an EEG signal, the potentially observable voltage values can be regarded as the outcomes, while the observed EEG signal falling into a prespecified range of values would be regarded as an event.

The definition of a random variable, which is often helpful in the description of random experiments, as it can relate an observed value to an underlying probability space, is based on the concept of measurable mappings. To prepare the definition of a random variable, the concepts of a measure, measurable mapping and image measure are introduced next.

Measure (B.4)

Let $\mathcal{K}$ be a system of sets over a set Ω. A mapping $\mu: \mathcal{K} \to \overline{\mathbb{R}}$ is called a measure on $\mathcal{K}$, if

i) $\mu(\emptyset) = 0$

ii) $\mu(A) \geq 0 \ (A \in \mathcal{K})$

iii) For every sequence $(A_n | \, n \in \mathbb{N})$ of pair-wise exclusive sets in $\mathcal{K}$, with $\sum_{n=1}^{\infty} A_n \in \mathcal{K}$ the following condition (σ-additivity) holds

$$\mu(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

The triple $(\Omega, \mathcal{A}, \mu)$ is called a measure space, while the double $(\Omega, \mathcal{A})$ is called a measurable space.

As can be seen, unlike for a probability measure, the definition of a measure does not require the condition $P(\Omega) = 1$, but requires $\mu(\emptyset) = 0$.

Measurable mapping and image measure (B.5)

Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces. A mapping $T: \Omega \to \Omega'$ is called $\mathcal{A} - \mathcal{A}'$- measurable, if one of the following equivalent conditions holds

i) $T^{-1}(A') = \{\omega \in \Omega | T(\omega) \in A'\} \in \mathcal{A} \ \ (A' \in \mathcal{A}')$

ii) $T^{-1}(\mathcal{A}') := \{T^{-1}(A') | A' \in \mathcal{A}'\} \subset \mathcal{A}$

Here, $T^{-1}(A')$ or $T^{-1}(\mathcal{A}')$ is also called the preimage of the respective sets under the mapping $T: \Omega \to \Omega'$.

Using the concept of a measurable mapping $T: \Omega \to \Omega'$ it is possible to define an image measure on $\mathcal{A}'$ based on a given measure on $\mathcal{A}$ as follows: Let $T: \Omega \to \Omega'$ be a measurable mapping and μ a measure on $\mathcal{A}$. Then

$$\mu'(A') := \mu(T^{-1}(A')) \ (A' \in \mathcal{A}') \tag{B.6}$$

defines a measure μ' on $\mathcal{A}'$. This measure is called the image measure of μ under T and is also denoted as $T(\mu)$ or μ_T. Before reintroducing the concept of an image measure below for the special case of a probability measure, the notion of a random variable is defined next.

Random variable (B.7)

Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces. An $\mathcal{A} - \mathcal{A}'$- measurable mapping $X: (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ is called an $\mathcal{A} - \mathcal{A}' -$ random variable. If $(\Omega', \mathcal{A}') = (\mathbb{R}^n, \mathcal{B}^n)$, the random variable is called real random variable.

In the description of random variables, the following notation is often used for $X^{-1}(A') = \{\omega \ \in \Omega | X(\omega) \in A'\}$:

$$\{X \ \in A'\} := X^{-1}(A') = \{\omega \ \in \Omega | X(\omega) \in A'\} \tag{B.8}$$

Here, $X \in A'$, where X is a mapping and A' is a set of $\mathcal{A}'$ obviously does not have a direct sense. Nevertheless, if P is a probability measure on $\mathcal{A}$, then

$$P\{X \in A'\} \coloneqq P(\{X \in A'\}) = P(\{\omega \in \Omega | X(\omega) \in A'\}) \tag{B.9}$$

is interpreted as the probability, that the image $X(\bar{\omega})$ of a randomly chosen $\bar{\omega} \in \Omega$ lies in A'. Of course, if $\bar{\omega}$ is chosen at random, $X(\bar{\omega})$ is also chosen at random. X itself however, despite being named a random variable, is neither a variable, nor dependent on chance. It is a mapping given by its respective definition of allocating an output to a given input. Having introduced the notion of a random variable and an image measure, the definition of a probability distribution now follows naturally:

Probability Distribution (B.10)

Let $X: (\Omega, \mathcal{A}, P) \rightarrow: (\Omega', \mathcal{A}')$ be a random variable. The image measure P_X of P is called the probability distribution of X with respect to P. The following identity holds

$$P_X(A') = P\left(X^{-1}(A')\right) = P(\{X \in A'\}) \quad (A' \in \mathcal{A}')$$

In order to link the measure theoretic concept of a probability distribution with real functions such as the Gaussian kernel, the concept of probability density functions is helpful.

Probability Density function (B.11)

Let $(\Omega, \mathcal{A})$ be a measurable space and μ and ν be measures on $\mathcal{A}$. A function $f \in \overline{\mathcal{M}}^+$ with

$$\nu(A) = \int_A f d\mu \quad (A \in \mathcal{A})$$

is called a μ-density of ν. In the special case that ν is a probability measure, f is called its probability density function with respect to μ.

In the definition above, $\overline{\mathcal{M}}^+$ denotes the set of all non-negative measurable functions $f: (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ and the integral in the definition of the density function is defined as follows: let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $f: \Omega \to \mathbb{R}_+$ a non-negative $\mathcal{A}$- measurable function. If $(f_n | n \in \mathbb{N})$ is a sequence of step functions which converges to f, then the augmented real number

$$\int f d\mu \coloneqq \lim_{n\to\infty} \int f_n d\,\mu \tag{B.12}$$

which is independent of the specific choice of the sequence $(f_n | n \in \mathbb{N})$ is called the μ-integral of f on Ω. Last, the integral of a step function, $f_n: \Omega \to \mathbb{R}_+$, $f = \sum_{i=1}^{m} \alpha_i 1_{A_i}$ which takes on the values α_i on the finite number of sets A_i is given by

$$\int f_n d\mu \coloneqq \sum_{i=1}^{m} \alpha_i \mu(A_i) \tag{B.13}$$

B.2 Product probability spaces

So far, only random experiments with a single outcome (i.e. for example an EEG potential or an MRI value) have been considered. However, one of the mainstays of information theory is to be able to make probabilistic statements about the relation between different random variables, whose values are observed simultaneously and whose joint probability distributions can be constructed. In the following, the concepts of a joint probability distribution and its relation to product measures will be introduced.

Joint distribution and independent random variables (B.14)

Let $(\Omega, \mathcal{A})$ and $(\Omega_\mathrm{i}, \mathcal{A}_i)$ $(i \in \mathbb{N}_n)$ denote measurable spaces and P a probability measure on $\mathcal{A}$. If $X_i: \Omega \to \Omega_\mathrm{i}$ are $\mathcal{A} - \mathcal{A}_i$ random variables (i.e. measurable mappings), then the mapping

$$X \coloneqq (X_1, \dots, X_n): \Omega \to \times_{\mathrm{i}=1}^{\mathrm{n}} \Omega_\mathrm{i}$$

is a $\mathcal{A} - \bigotimes_{i=1}^{n} \mathcal{A}_i$ random variable. The distribution P_X on $\bigotimes_{i=1}^{n} \mathcal{A}_i$ of X defined by P is called the joint distribution of the X_i with respect to P.

The independence of a family $(X_i | i \in \mathbb{N}_n)$ of random variables $X_i: (\Omega, \mathcal{A}, P) \to (\Omega_\mathrm{i}, \mathcal{A}_i)$ is a property of their joint distribution. More specifically, the random variables $X_1, \dots, X_n$ are said to be independent, if their joint distribution P_X is equal to the product of their individual distributions P_{X_i}. Formally, a family $(X_i | i \in \mathbb{N}_n)$ of random variables $X_i: (\Omega, \mathcal{A}, P) \to (\Omega_\mathrm{i}, \mathcal{A}_i)$ is called independent, if

$$P(\{\omega \in \Omega | X_i(\omega) \in A_i (i \in \mathbb{N}_n)\}) = \prod_{i=1}^{n} P(\{\omega \in \Omega | X_i(\omega) \in A_i\}) \quad (A_i \in \mathcal{A}_i \ (i \in \mathbb{N}_n)) \text{ (B.15)}$$

Using the short notation introduced above, this definition can also be written

$$P\{X_i \in A_i (i \in \mathbb{N}_n)\} = \prod_{i=1}^{n} P\{X_i \in A_i\} \quad \text{(B.16)}$$

If the random variables $X_1, \dots, X_n$ are independent, their joint distribution is equal to the product measure of their individual distributions. This can be seen as follows:

For $A_i \in \mathcal{A}_i$

$$P(\{\omega \in \Omega | \, \mathrm{X_i}(\omega) \in \mathrm{A_i}\}) = P\left(X_i^{-1}(\mathrm{A_i})\right) = P_{\mathrm{X_i}}(\mathrm{A_i}) \quad \text{(B.17)}$$

and

$$P(\{\omega \in \Omega | \, \mathrm{X_i}(\omega) \in \times_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i}\}) = P\left(X^{-1}(\times_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i})\right) = P_{\mathrm{X}}(\times_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i}) \quad \text{(B.18)}$$

If the family of random variables $(X_i | i \in \mathbb{N}_n)$ is independent the condition

$$P_{\mathrm{X}}(\times_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{A_i}) = \prod_{i=1}^{n} P_{\mathrm{X_i}}(\mathrm{A_i}) \quad (A_i \in \mathcal{A}_i) \quad \text{(B.19)}$$

holds.

Product measure

Let $(\Omega_i, \mathcal{A}_i, P_i)$ $(i \in \mathbb{N}_n)$ be probability measure spaces. Then the probability measure on $\bigotimes_{i=1}^{n} \mathcal{A}_i$ defined by

$$P(\times_{i=1}^{n} A_i) = P(A_1 \times \ldots \times A_n) = P_1(A_1) \cdot \ldots \cdot P_n(A_n) = \prod_{i=1}^{n} P_i(A_i) \quad \text{(B.20)}$$

for all $A_i \in \mathcal{A}_i$ is called the product measure of on $\bigotimes_{i=1}^{n} \mathcal{A}_i$ and denoted by $\bigotimes_{i=1}^{n} P_i$.

A set of random variables $(X_i | i \in \mathbb{N}_n)$ is hence independent if

$$P_{\mathrm{X}} = \bigotimes_{i=1}^{n} P_{X_i} \quad \text{(B.21)}$$

The notion of independent families of random variables and product measures is intimately linked with the concept of conditional probability: an event A in independent of an event B if the conditional probability $P(A|B)$ equals the unconditional probability$P(A)$. The conditional probability distribution of a random variable X conditioned on an event B is defined below and used extensively in the information theoretic framework introduced in section 2.1.

Conditional probability distributions

Let X be a continuous random variable with probability distribution P_X, where P is given by a probability density function $f: \Omega \rightarrow \mathbb{R}_+$. Let $B \in \mathcal{A}$ be an event of the σ-algebra $\mathcal{A}$ with $P(B) \neq 0$. Then the conditional density function of P_X is defined by

$$f(x|E) = \begin{cases} \frac{f(x)}{P(B)} \; if \; x \in B \\ 0 \quad if \; x \notin B \end{cases} \quad \text{(B.22)}$$

It follows, that for any event $A \in \mathcal{A}$

$$P(A|B) = \int_A f(x|B) d\mu \quad \text{(B.23)}$$

B.3 Discrete random variables

Information theoretic concepts are usually formulated in the context of discrete random variables, a special case of the general random variable concept discussed so far. Below, the concepts of discrete random variables, probability mass functions and discrete conditional and marginal probability distributions are briefly introduced.

Discrete random variable (B.24)

Discrete random variables are random variables whose codomain consists of either finite or at least countable elements (A set S is called countable, if there exists a bijective function $f: \mathbb{N} \to S$).

The analogue for probability density functions for the case of discrete random variables are probability mass functions.

Probability mass function (B.25)

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X \in \bar{\mathcal{M}}^+$, $X: \Omega \to \Omega'$ a discrete random variable with range $X(\Omega) = \{\omega_i' | i \in I\}$ with $I \subset \mathbb{N}$ and $\omega_i' \neq \omega_j'$ $(i, j \in I, i \neq j)$. Then the sets $\Omega_\mathrm{i} \subset \Omega$, $\Omega_\mathrm{i} = \{\omega | \mathrm{X}(\omega) = \omega_i', i \in I\}$ are disjoint sets and according to the definition of the probability measure P

$$P(\cup_i \Omega_\mathrm{i}) = \sum_i P(\Omega_\mathrm{i}) = \sum_i P(\{\omega | X(\omega) = \omega_i'\}) := \sum_i P(\{X = \omega_i'\}) := \sum_i P\{X = \omega_i'\} = 1$$

The probability mass function of the discrete random variable $X: \Omega \to \Omega'$ is then defined as the mapping $p_X: \Omega' \to [0,1]$ according to

$$p_X(\omega_i') = P\{X = \omega_i'\} = P(\{\omega | X(\omega) = \omega_i'\}) \tag{B.26}$$

In analogy to the general case, the discrete joint distribution of discrete random variables $X \coloneqq (X_1, \dots, X_n): \Omega \to \times_{i=1}^{n} \Omega_i$ is given by

$$p_X\left(\omega_i^{'} = \left(\omega_{1_i}^{'}, \dots, \omega_{n_i}^{'}\right)\right) = P\left\{\left(X_1 = \omega_{1_i}^{'}, \dots, X_n = \omega_{n_i}^{'}\right)\right\} \tag{B.27}$$

$$= P\left(\omega = \left\{(\omega_1, \dots, \omega_{1n}) | X(\omega) = \omega_i^{'}\right\}\right)$$

For two discrete random variables X_1 and X_2 the following conditions holds

$$P\left\{X_1 = \omega_{1_i}^{'}\right\} = \sum_{\omega_{2_i}^{'} \in \Omega_2} P\left\{X_1 = \omega_{1_i}^{'}, X_2 = \omega_{2_i}^{'}\right\} \tag{B.28}$$

Here, $P\left\{X_1 = \omega_{1_i}^{'}\right\}$ is called the marginal probability. In general, marginal probability distributions of discrete joint distributions can be obtained by summing over the other discrete random variables.

Discrete conditional probability

Let $(\Omega = \{\omega_1, \dots, \omega_n\}, \mathcal{A}, P)$ be a discrete probability space with distribution function $m(\omega_j)$ assigned to P (in the case of discrete outcome spaces, the distribution function of a measure P on Ω is defined by $(A) = \sum_{\omega_j \in A} m(\omega_j)$). For any $B \in \mathcal{A}$ with $P(B) \neq 0$, a distribution function of the form

$$m\left(\omega_j \middle| B\right) = \frac{m(\omega_j)}{P(B)} \tag{B.29}$$

can be defined, which is known as the conditional distribution P given the event B. For any event $A \in \mathcal{A}$, the conditional probability of A given B is then defined as

$$P(A|B) = \sum_{\omega_j \in A \cap B} m\left(\omega_j \middle| B\right) = \sum_{\omega_j \in A \cap B} \frac{m(\omega_j)}{P(B)} = \frac{P(A \cap B)}{P(B)} \tag{B.30}$$

In summary, modern probability theory rests on the interpretation of probabilities as measures of sets. A lot of the more advanced probabilistic terms and integral definitions are motivated by the question of how such a framework can be developed in the presence of the continuous underlying field of the real numbers. $\mathbb{R}$ is continuous in the sense that it has the property that between any two elements of $\mathbb{R}$ there lie an infinity of other elements of $\mathbb{R}$.

Random variables, which assume a relatively unspecific meaning in the context of experimental research, are defined as mappings between probability spaces. The important concept to note here is that the values a random variable can assume are deterministically given by the form of the random variable. The randomness in experimental contexts is given by the random selection of an input argument $\omega \in \Omega$ according to the size of the subset A of Ω that ω lies in, which is measured by the probability measure P.

The measure defined for the ur-image of the values a random variable can assume is the random variables probability distribution. Finally, it is important to realize that ordinary real functions of the form $f: \mathbb{R} \to \mathbb{R}$ can assume meaning as probability density functions by defining probability measures by means of an integral operation.

Appendix C: Independent component analysis

Independent component analysis (ICA) (Hyvaerinen et al., 2001) is a generative, latent variable model that describes how the observed data are generated by a process of mixing the underlying unknown sources; the sources (ICs) are assumed to be statistically independent and non-Gaussian. Since the observed mixed signals will tend to have more Gaussian amplitude distributions, ICA strives to find a separation matrix that minimizes the Gaussianity of the results, thus optimally separating the signals. For this purpose, it is assumed that the set of observed EEG signals is generated by the mixing model:

$$X = AS \tag{C.1}$$

where $X \in \mathbb{R}^{N_C \times N_S}$ (N_E: number of channels, N_S: number of samples) denotes the EEG data set, $A \in \mathbb{R}^{N_C \times N_C}$ is an unknown full-rank mixing matrix and $S \in \mathbb{R}^{N_C \times N_S}$ is the matrix containing the unknown time-series of statistically independent non-Gaussian sources. The model is approached by processing channel signals by an ICA demixing system described in the form:

$$\hat{S} = WX \tag{C.2}$$

where $\hat{S} \in \mathbb{R}^{N_C \times N_S}$ is an set of estimated independent components and $W \in \mathbb{R}^{N_C \times N_C}$ a separation matrix, i.e., the estimate of the inverse of the unknown mixing matrix A, up to permutation and scaling (denoted by the hat):

$$W = \hat{A}^{-1} \tag{C.3}$$

To estimate W, the FastICA algorithm (Hyvaerinen, 1999) was employed.

In the case of a large number of channels (64 in this case) a direct extraction of all independent components would have been extremely time consuming and component

selection extremely challenging. Hence, dimensionality of the data set was reduced by selecting a subset of independent components that explained at least 95% of the data variance (Salustri C. et al., 2005). After the identification of a set of $N_{ICA} \in \mathbb{N}_{N_C}$ independent components of interest, the data at the scalp electrodes were obtained by retro-projecting the selected independent components into channel space by

$$X_{ICA} = A_{ICA}\hat{S}_{ICA} \tag{C.4}$$

where $A_{ICA} \in \mathbb{R}^{N_C \times N_{ICA}}$ is the estimated mixing vector for the $N_{ICA} \in \mathbb{N}_{N_C}$ selected sources in $\hat{S}_{ICA} \in \mathbb{R}^{N_{ICA} \times N_S}$ and $X_{ICA} \in \mathbb{R}^{N_C \times N_S}$ is the resulting $\hat{S}_{ICA}$ retro-projection on the channels space.

In summary, the ICA approach to EEG data cleaning is based on the decomposition of the electrode-time-series matrix into a weighted sum of a set of independent components. Some of these independent components are then excluded in the re-composition of the EEG electrode-time-series due to them being regarded as artefactual.

Appendix D: Functional source separation

The FSS procedure starts from a basic Independent Component Analysis (ICA) model (Hyvaerinen et al., 2001) in which the set of EEG signals $X \in \mathbb{R}^{N_C \times N_S}$ (N_C: number of channels, N_S: number of samples) are assumed to be obtained as a linear combination of statistically independent non-Gaussian sources $S \in \mathbb{R}^{N_C \times N_S}$ by means of an unknown mixing matrix $A \in \mathbb{R}^{N_C \times N_C}$

$$X = AS \tag{D.1}$$

The sources in the rows of S are estimated (up to arbitrary scaling and permutation) by independent components $\hat{S} \in \mathbb{R}^{N_C \times N_S}$ as

$$\hat{S} = WX \tag{D.2}$$

where the unmixing matrix $W \in \mathbb{R}^{N_C \times N_C}$ is estimated along with the components.

In the FSS procedure, additional information is included to bias the decomposition algorithm towards solutions that satisfy physiological assumptions. The aim of FSS is to enhance the separation of relevant signals by exploiting some a priori knowledge without renouncing the advantages of using only information contained in the original signal. A modified (with respect to standard ICA) contrast function is defined as

$$F = J + \lambda R_{FS} \tag{D.3}$$

where J is the statistical index normally used in ICA, while R_{FS} accounts for the prior information used to extract a single source. According to the weighting parameter λ it is possible to adjust the relative weight of these two aspects. In this study, λ was chosen equal to 1000 in all cases, as detailed in (Porcaro et al., 2008). The functional constraint R_{FS} was defined as

$$R_{FS} = \sum_{t_k - \Delta_1 t_k}^{t_k + \Delta_2 t_k} |EA(FS_{P100}, t)| - \sum_{-100}^{0} |EA(FS_{P100}, t)| \tag{D.4}$$

with the evoked activity, EA , computed by averaging signal epochs of the source $FS_{P100} \in \mathbb{R}^{1\times N_S}$ triggered on the visual stimulus ($t = 0$); t_k is the time point with the maximum electric potential around 100 ms after the stimulus onset on the maximal original EEG channel; $\Delta_1 t_k$ ($\Delta_2 t_k$) is the time point corresponding to a signal amplitude of 50% of the maximal value before (after) t_k. The baseline was computed in the time interval from -100 to 0 ms. The precise value of each latency t_k was chosen for each subject, corresponding to the maximum electric potential in the time interval of interest (80 – 120 ms post-stimulus).

The source was then retro-projected to obtain its electric potential distribution at the scalp electrodes according to

$$EEG_{FS} = A_{P100} FS_{P100} \tag{D.5}$$

where $A_{P100} \in \mathbb{R}^{N_C\times 1}$ is the estimated mixing vector (the first column of $A \in \mathbb{R}^{N_C\times N_C}$) for the functional source. $EEG_{FS} \in \mathbb{R}^{N_C\times N_S}$ are the retro-projections on the electrodes of the estimated $FS_{P100} \in \mathbb{R}^{1\times N_S}$.

Appendix E: sLORETA

This Appendix reviews the algebraic formulation of sLORETA following (Pascual-Marqui, 2002b). Assuming linear superposition of potentials, the electrode potentials at a given time point $t \in \mathbb{R}$ at the scalp surface are given by

$$V(t) = LJ(t) + c1 \tag{E.1}$$

Denoting the number of electrode as N_E and the number of voxels as N_V, equation (E.1) describes the electrode potentials $V(t) \in \mathbb{R}^{N_E \times 1}$ as the matrix product of the lead field matrix $L \in \mathbb{R}^{N_E \times (3N_V)}$ and the current density vector $J(t) \in \mathbb{R}^{3N_V \times 1}$ plus a constant offset given by $c1$, where $c \in \mathbb{R}$ and $1 \in \mathbb{R}^{N_E}$ is a vector of ones. The experimentally obtained data is represented by $V(t)$, while the data projected into three-dimensional brain space is an estimate for $J(t)$.

The lead field matrix $L \in \mathbb{R}^{N_E \times (3N_V)}$ is a time-invariant solution to the EEG forward problem and has the following structure

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1N_V} \\ l_{21} & l_{22} & \dots & l_{2N_V} \\ \dots & \ddots & \dots & \vdots \\ l_{N_E1} & l_{N_E1} & \dots & l_{N_EN_V} \end{pmatrix} \tag{E.2}$$

with $l_{ij} \in \mathbb{R}^{1 \times 3}$ for $i = 1, \dots, N_E$ and $j = 1, \dots, N_V$. The entries l_{ij} in the lead field matrix represent the projection from the dipole moments at the j-th voxel to the i-th electrode. Specifically,

$$l_{ij} = \begin{pmatrix} l_{ij}^x & l_{ij}^y & l_{ij}^z \end{pmatrix} \tag{E.3}$$

where l_{ij}^x is the scalp electric potential at the i-th electrode, due to a unit strength x-oriented dipole at the j-th voxel, and correspondingly for l_{ij}^y and l_{ij}^z.

sLORETA itself is based on the explicit minimization of the error function

$$F\big(J(t)\big) = \|V(t) - LJ(t) - c1\|^2 + \alpha\|J(t)\|^2 \quad \text{(E.4)}$$

where $\alpha \geq 0\, \epsilon\, \mathbb{R}$ denotes a regularization parameter and $\|\ \ \|$ denotes an l_p-Norm. The explicit solution to this minimization problem is given by the matrix product

$$\widehat{J(t)} = (L^T H[HLL^T H + \alpha H]^{\perp})V(t) \quad \text{(E.5)}$$

where $\perp$ denotes the Moore-Penrose pseudo-inverse, and H= $I - \frac{11^T}{1^T 1} \in \mathbb{R}^{N_E \times N_E}$ denotes a centring matrix.

Further, the variance of the estimated current density is given by

$$S_J = L^T(LL^T + \alpha H)^{\perp} L \quad \text{(E.6)}$$

sLORETA than corresponds to the following estimate of standardized current density power

$$sCDP_j(t) = \widehat{J(t)}_l^T \left(\left[S_J\right]_{jj}\right)^{-1} \widehat{J(t)}_l \quad \text{(E.7)}$$

where $\widehat{J(t)}_l\, \epsilon\, \mathbb{R}^{3\times 1}$ is the current density estimate at the j-th voxel and $\left[S_J\right]_{jj} \epsilon\, \mathbb{R}^{3\times 3}$ is the j-th diagonal block of matrix S_J.

In essence, given the data of electrical potentials at all electrodes $V(t)$ at time $t\, \epsilon\, \mathbb{R}$, the standardized current density estimate at all voxels is given by matrix products of $V(t)$ and the lead-field generalized matrix inverse.

Appendix F: Frequency-domain LORETA

This Appendix reviews the algebraic formulation of the frequency domain LORETA variant following(Pascual-Marqui, 2002a).

Let $F \in \mathbb{C}^{N_T \times N_T}$ denote the discrete Fourier transform operator for a time-domain signal of N_T time points and let $\hat{J} \in \mathbb{R}^{3N_V \times N_T}$ denote the matrix of all estimated primary impressed current densities at all voxels $1, \dots, N_V$ and time points $1, \dots, N_T$. Further, let $V \in \mathbb{R}^{N_E \times N_T}$ denote the matrix of all sensor potentials at all time points of an EEG epoch of interest. The Fourier transform of the current densities is then given by

$$\hat{J}\, F = TVF \tag{F.1}$$

where $T = (L^T H[HLL^T H + \alpha H]^{\perp}) \in \mathbb{R}^{3N_V \times N_T}$. This formulation of the pseudo-inverse assumes that the discrete Laplacian matrix B and the lead field normalizing matrix W in (Pascual-Marqui, 2002a) are set to the respective identity matrices. Equivalently, let $\hat{J}_\omega$ and V_ω denote the frequency domain representation of the impressed current densities $\hat{J}$ and sensor potentials V, respectively, i.e. the columns of $\hat{J}_\omega$ and V_ω correspond to all discrete frequencies. Then (5.13) can be equivalently written as

$$\hat{J}_\omega = TV_\omega \tag{F.2}$$

where $\hat{J}_\omega \in \mathbb{C}^{3N_V \times N_T}$ and $V_\omega \in \mathbb{C}^{N_E \times N_T}$. For a single discrete frequency ω_i expression (F.2) is equivalent to

$$\hat{J}_{\omega_i} = TV_{\omega_i} \tag{F.3}$$

with $\hat{J}_{\omega_i} \in \mathbb{C}^{3N_V \times 1}$ and and $V_{\omega_i} \in \mathbb{C}^{N_E \times 1}$. As for any discrete Fourier transform, although there are N_T discrete frequencies, half of them are conjugates of the other half, and hence redundant. Let $V_{\omega_i} \in \mathbb{C}^{N_E \times 1}$ denote the discrete Fourier transform of a given EEG epoch with N_T samples and frequency ω_i. From equation (5.15) the cross-spectral matrix of the primary impressed current densities $S_{\hat{J}_{\omega_i}} \in \mathbb{C}^{3N_V \times 3N_V}$ can then be derived as follows

$$S_{\hat{J}_{\omega_i}} = \frac{1}{\varphi}\sum_{i=1}^{N} \hat{J}_{\omega_i}\hat{J}_{\omega_i}^{*} \tag{F.4}$$

where $*$ denotes the transpose and complex conjugate operator. From (5.15) it then follows that

$$S_{\hat{J}_{\omega_i}} = \frac{1}{\varphi}\hat{J}_{\omega_i}\hat{J}_{\omega_i}^{*} = \frac{1}{\varphi}TV_{\omega_i}\left(TV_{\omega_i}\right)^{*} = \frac{1}{\varphi}TV_{\omega_i}V_{\omega_i}^{*}T^{*} = \frac{1}{\varphi}TV_{\omega_i}V_{\omega_i}^{*}T^{T} \tag{F.5}$$

as T is a real matrix and $\varphi = 2\pi N_T \in \mathbb{R}$. The spectra at each voxel, i.e. the spectra of the electrical neuronal activity time series estimated via LORETA at the j-th voxel corresponds to the diagonal elements of the matrix $diag\left(S_{\hat{J}_{\omega_i}}\right)$, where the $diag$ operator takes a Hermitian matrix and returns a real diagonal matrix with elements corresponding to the diagonal elements of the original Hermitian matrix (which are real in the first place).

Appendix G: False-discovery rate correction

This appendix reviews the procedure for false-discovery rate correction following (Genovese et al., 2002). False-discovery rate correction applies to the case in which a large number of statistical tests are carried out, in the present case one for informativness at each voxel. Let V denote the total number of voxels being tested in this analysis and D_i and D_u the number of voxels declared informative and uninformative, respectively. The false-discovery rate is defined as

$$FDR = \frac{V_{ui}}{D_i} \tag{G.1}$$

where V_{ui} denotes the number of voxels which are declared informative but are actually uninformative. The false-discovery rate is controlled by specifying a rate $q \in [0,1]$ and ensuring that on average, i.e. over replications of the experiment, the false-discovery rate is not bigger than q, in short

$$E(FDR) \leq q \tag{G.2}$$

where $E(\cdot)$ denotes the expectation.

The procedure to control the false-discovery rate then is as follows

1. Select a FDR bound $q \in [0,1]$ for the current set of multiple tests
2. Order the corresponding P values of the tests in monotone non-descending order

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(V)} \tag{G.3}$$

3. Let r be the largest i for which

$$P_{(i)} \leq \frac{i}{V} q \tag{G.4}$$

4. Declare the voxels $v_{(i)}, \ldots, v_{(r)}$ informative.

Appendix H: Perceptual decision paradigm eye-movement data

Eye-movement data were recorded from 8 observer's partaking in the combined EEG-fMRI data acquisition using the long-range ASL 6000 Eye-tracker (Applied Science Laboratories, Bedford, MA) at a sampling frequency of 60 Hz. Eye-tracking data was exported using the Eyenal software (Applied Science Laboratories, Bedford, MA) and imported into Matlab (The Mathworks, Natick, MA). For each subject, samples for which both the pupil circumference and the corneal reflex were not detected were excluded from further analysis. These samples correspond to blinks and recording setup noise. Two observers were excluded from further analysis as the number of invalid samples was too substantial. For the remaining subjects, the session time-series was partitioned into experimental trials comprising the onset of the attention cue (arrow) at 0 seconds, the onset of the stimulus at 1 second and the remaining post-stimulus 2 second period. Mean eye-movement traces around fixation (corrected to 0 degree of visual angle) are shown in Figure H.1 for the stimulus conditions and H.2 for left- and right-hemifield trials, respectively. Data are displayed for both the horizontal and the vertical eye-position (upper panels). Additionally, Figures H.1 and H.2 display the SEM across trials averaged over observers for both horizontal and vertical eye position (lower panels).

For none of the eye-position time-series systematic variability upon the onset of the prioritization cue (at 0 s) or stimulus (at 1 s) could be detected, indicating steady fixation throughout the experimental trial. It should be noted that the centre of the peripherally presented stimulus was at 11 degrees of visual angle. Towards the end of the time-series investigated, the variability of the vertical eye position increases slightly, potentially indicating eye-blinks. Based on these data it is unlikely that observer's did not maintain steady fixation, and condition effects could be explained by eye-movements.

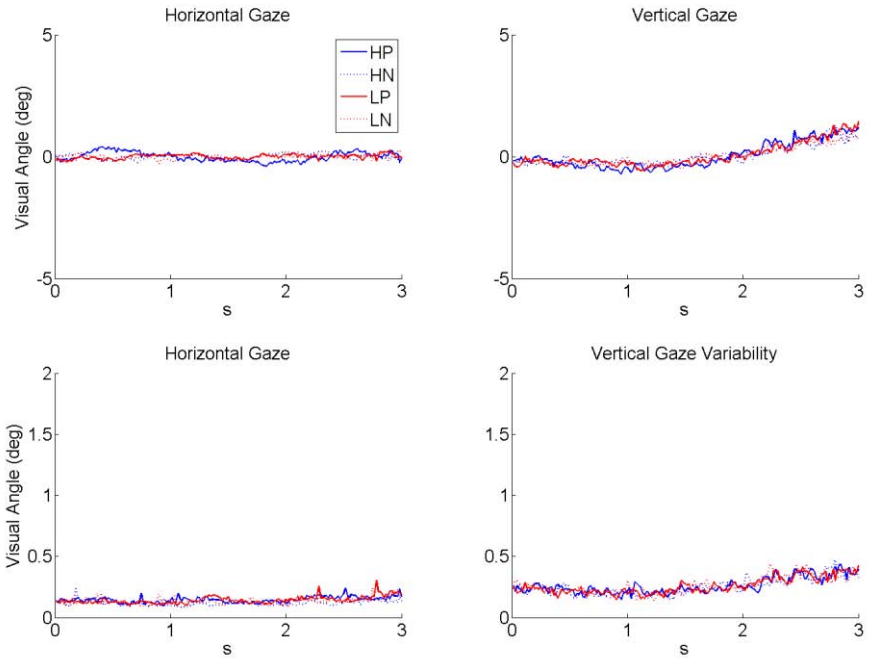


Figure H.1 Eye-movements according to stimulus condition. The panels display the eye-tracking time-series of a 3 s peri-stimulus time-window, corresponding to the onset of the prioritization cue at 0 s and the stimulus onset at 1 s. The two upper panel display the mean (n = 6) eye-position in degrees of visual angle for the horizontal and vertical eye-position. The lower two panels display the mean (n = 6) SEM of the eye-position across trials (HP: High informativeness, prioritized, HN: High Informativeness, not prioritized, LP: Low informativeness, prioritized, LN: Low informativeness, not prioritized)

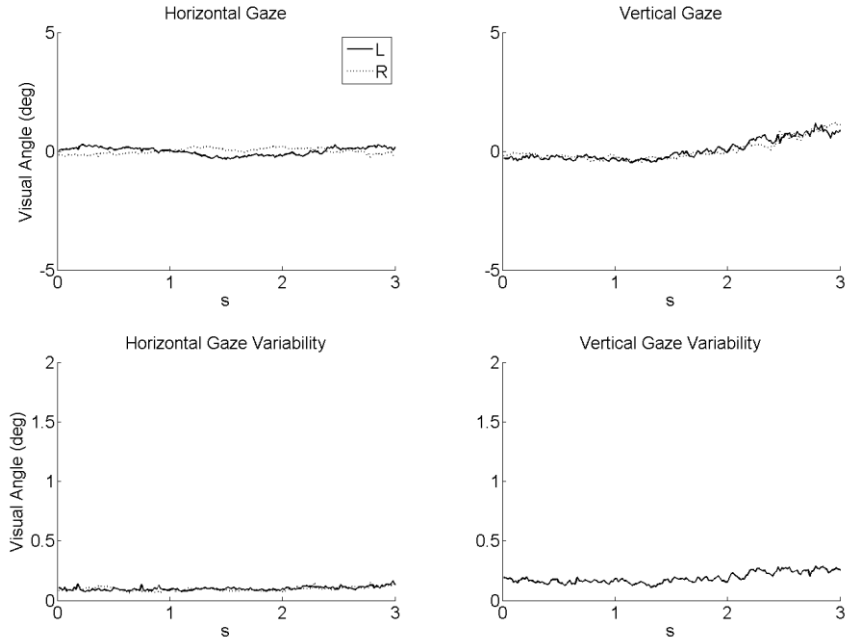


Figure H.2 Eye-movements according to stimulus presentation side. The panels display the eye-tracking time-series of a 3 s peri-stimulus time-window, corresponding to the onset of the prioritization cue at 0 s and the stimulus onset at 1 s. The two upper panel display the mean (n = 6) eye-position in degrees of visual angle for the horizontal and vertical eye-position. The lower two panels display the mean (n = 6) SEM of the eye-position across trials (L: left hemifield trials, R: right hemifield trials)

Appendix I: Perceptual decision psychophysical pilot study

To establish that the given stimulus and behavioural manipulations of the perceptual decision task discussed in sections 6.2.2 and 6.2.3 was successful in evoking a differential behavioural response pattern (response times and accuracy effects), a psychophysical pilot study according to the specification in 6.2.2 for the EEG only recordings was conducted with 9 participants (mean age 27.3 years, range 22 – 37 years). Three of the participants also participated in the main EEG-fMRI experiment approximately four months later. The results of the pilot psychophysical study are shown in Figure I.1. As for the main experiment, an increase in stimulus informativeness and spatial prioritization of the stimulus' location led to faster response times and higher response accuracy. Specifically, a two-way repeated measures ANOVA for the median response times including all trials revealed a significant main effect of stimulus coherence ($F_{(1,8)}$ = 20.6, p = 0.002), a significant main effect of prioritization ($F_{(1,8)}$ = 8.3, p = 0.02) and no significant interaction ($F_{(1,8)}$ = 1.9, p = 0.21). Similarly, for median response times on correct response trials only, a significant main effect of stimulus coherence ($F_{(1,8)}$ = 22.2, p = 0.002), a significant main effect of prioritization ($F_{(1,8)}$ = 7.8, p = 0.02) and no significant interaction ($F_{(1,8)}$ = 1.8, p = 0.21) were detected. Finally, for response accuracy, a two-way repeated measures ANOVA revealed a significant main effect of stimulus coherence ($F_{(1,8)}$ = 22.6, p = 0.001), a significant main effect of prioritization ($F_{(1,8)}$ = 3.2, p = 0.11) and no significant interaction ($F_{(1,8)}$ = 2.0, p = 0.19).

The paradigm was hence judged adequate for the subsequent EEG-fMRI data acquisition.

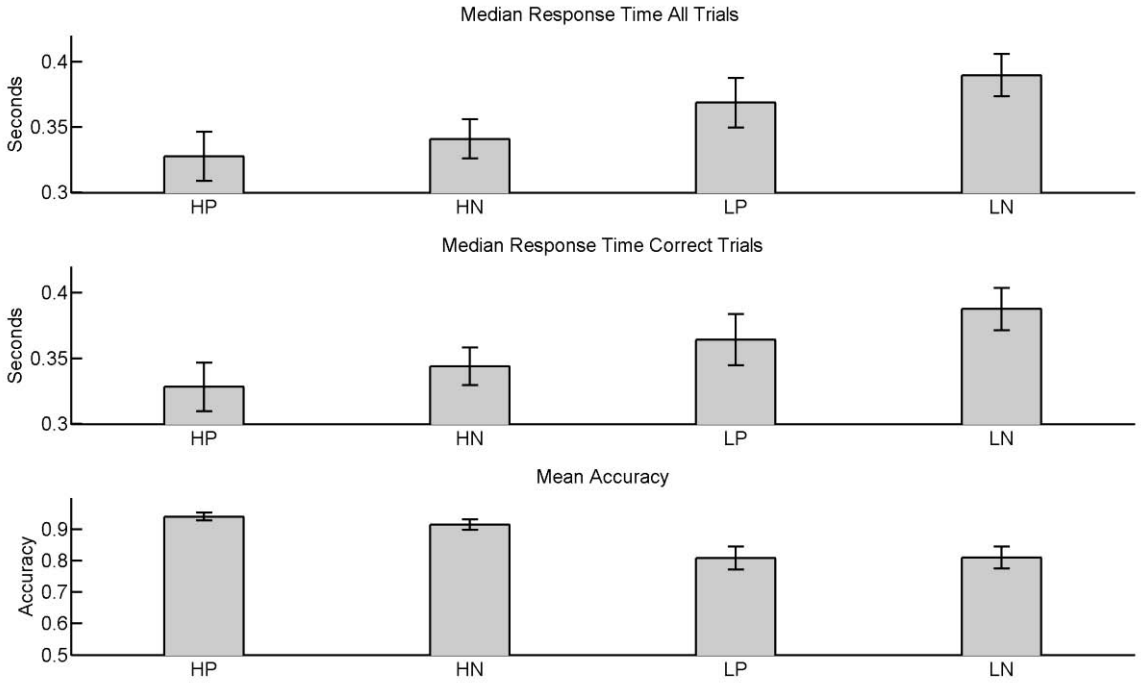


Figure I.1 Psychophysical pilot study results. Uppermost panel: Average median response time for all experimental trials across observers ± SEM. Middle panel: Average median response time for correct response trial across observers ± SEM. Lowermost panel: Average response accuracy across observers ± SEM (HP: High Informativeness, Prioritized, HN: High Informativeness, Not Prioritized, LP: Low Informativeness, Prioritized, LN: Low Informativeness, Not Prioritized

References

Adelman, T.L., Bialek, W., Olberg, R.M., 2003. The information content of receptive fields. Neuron 40, 823-833.

Allen, P.J., 2010. EEG Instrumentation and Safety. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 115-134.

Allen, P.J., Josephs, O., Turner, R., 2000. A method for removing imaging artifact from continuous EEG recorded during functional MRI. Neuroimage 12, 230-239.

Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L., 1998. Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction. Neuroimage 8, 229-239.

Anami, K., Mori, T., Tanaka, F., Kawagoe, Y., Okamoto, J., Yarita, M., Ohnishi, T., Yumoto, M., Matsuda, H., Saitoh, O., 2003. Stepping stone sampling for retrieving artifact-free electroencephalogram during functional magnetic resonance imaging. Neuroimage 19, 281-295.

Arabzadeh, E., Panzeri, S., Diamond, M.E., 2004. Whisker vibration information carried by rat barrel cortex neurons. J.Neurosci. 24, 6011-6020.

Bagshaw, A.P., Warbrick, T., 2007. Single trial variability of EEG and fMRI responses to visual stimuli. Neuroimage 38, 280-292.

Barbati, G., Sigismondi, R., Zappasodi, F., Porcaro, C., Graziadio, S., Valente, G., Balsi, M., Rossini, P.M., Tecchio, F., 2006. Functional source separation from magnetoencephalographic signals. Hum.Brain Mapp. 27, 925-934.

Becker, R., Ritter, P., Villringer, A., 2010. Visual System. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 401-418.

Becker, R., Ritter, P., Villringer, A., 2008. Influence of ongoing alpha rhythm on the visual evoked potential. Neuroimage. 39, 707-716.

Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M.A., Logothetis, N.K., Panzeri, S., 2008. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. J.Neurosci. 28, 5696-5709.

Benar, C.G., Bagshaw, A.P., Lemieux, L., 2010. Experimental Design and Data Analysis Strategies. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 221-258.

Benar, C.G., Schon, D., Grimault, S., Nazarian, B., Burle, B., Roth, M., Badier, J.M., Marquis, P., Liegeois-Chauvel, C., Anton, J.L., 2007. Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI. Hum.Brain Mapp. 28, 602-613.

Benjamini, Y., Hochbeg, Y., 1995. Controlling the False-Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J.R. Statist. Soc. B. 57(1) 289-300

Bishop, M.B., 2006. Pattern Recognition and Machine Learning. Springer.

Bonmassar, G., Hadjikhani, N., Ives, J.R., Hinton, D., Belliveau, J.W., 2001a. Influence of EEG electrodes on the BOLD fMRI signal. Hum.Brain Mapp. 14, 108-115.

Bonmassar, G., Schwartz, D.P., Liu, A.K., Kwong, K.K., Dale, A.M., Belliveau, J.W., 2001b. Spatiotemporal brain imaging of visual-evoked activity using interleaved EEG and fMRI recordings. Neuroimage 13, 1035-1043.

Borel, E., 1963. Probability and Certainty. Walker, New York.

Boynton, G.M., Demb, J.B., Glover, G.H., Heeger, D.J., 1999. Neuronal basis of contrast discrimination. Vision Res. 39, 257-269.

Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vision 10, 433-436.

Buxton, R.B., 2001. The elusive initial dip. Neuroimage 13, 953-958.

Buxton, R.B., Uludag, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. Neuroimage 23 Suppl 1, S220-S233.

Carmichael, D., 2010. Image Quality Issues. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 173-200.

Carney, P., Jackson, G., 2010. EEG-fMRI in Idiopathic Generalised Epilepsy (Adults). In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 333-348.

Chawla, D., Lumer, E.D., Friston, K.J., 2000. Relating macroscopic measures of brain activity to fast, dynamic neuronal interactions. Neural Computation. 12, 2805-2821

Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage. 44, 62-70.

Compumedics USA Ltd, 2008. CURRY User Guide: Multi-modal neuroimaging for CURRY 6.

Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat.Neurosci. 3, 292-297.

Corbetta, M., Kincade, J.M., Shulman, G.L., 2002. Neural systems for visual orienting and their relationships to spatial working memory. J.Cogn Neurosci. 14, 508-523.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat.Rev.Neurosci. 3, 201-215.

Correa, N.M., Eichele, T., Adali, T., Li, Y.O., Calhoun, V.D., 2010. Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. Neuroimage 50, 1438-1445.

Cover, T.M., Thomas, J.M., 2006. Elements of Information Theory, 2 ed. John Wiley and Sons.

Craig, A.D., 2009. How do you feel - now? The anterior insula and human awareness. Nature Reviews Neuroscience 10, 59-70

da Silva, F.L., 2010. EEG: Origin and Measurement. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 19-38.

Dakin, S.C., Hess, R.F., Ledgeway, T., Achtman, R.L., 2002. What causes non-monotonic tuning of fMRI response to noisy images? Curr.Biol. 12, R476-R477.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55-67.

Daunizeau, J., Friston, K.J., Kiebel, S.J., 2009. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. Physica D. 238, 2089-2118.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pelegrini-Issac, M., Lina, J.M., Benali, H., 2007. Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. Neuroimage 36, 69-87.

Daunizeau, J., Laufs, H., Friston, K.J., 2010. EEG-fMRI Information Fusion: Biophysics and Data Analysis. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 511-526.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage. 30, 1255-1272.

Dayan, P., Abbott, F., 2001. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems, 1 ed. The MIT Pres, Cambridge, MA.

de Munck, J.C., Goncalves, S.I., Huijboom, L., Kuijer, J.P., Pouwels, P.J., Heethaar, R.M., Lopes da Silva, F.H., 2007a. The hemodynamic response of the alpha rhythm: an EEG/fMRI study. Neuroimage 35, 1142-1151.

de Munck, J.C., Goncalves, S.I., Huijboom, L., Kuijer, J.P., Pouwels, P.J., Heethaar, R.M., Lopes da Silva, F.H., 2007b. The hemodynamic response of the alpha rhythm: an EEG/fMRI study. Neuroimage 35, 1142-1151.

de Munck, J.C., Goncalves, S.I., Mammoliti, R., Heethaar, R.M., Lopes da Silva, F.H., 2009. Interactions between different EEG frequency bands and their effect on alpha-fMRI correlations. Neuroimage 47, 69-76.

Debener, S., Kranczioch, C., Gutberlet, I., 2010. EEG Quality: Origin and Reduction of the EEG Cardiac-Related Artefact. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 135-152.

Debener, S., Mullinger, K.J., Niazy, R.K., Bowtell, R.W., 2008. Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength. Int.J Psychophysiol. 67, 189-199.

Debener, S., Strobel, A., Sorger, B., Peters, J., Kranczioch, C., Engel, A.K., Goebel, R., 2007a. Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: removal of the ballistocardiogram artefact. Neuroimage 34, 587-597.

Debener, S., Ullsperger, M., Siegel, M., Engel, A.K., 2006. Single-trial EEG-fMRI reveals the dynamics of cognitive function. Trends in Cognitive Sciences 10, 558-563.

Debener, S., Ullsperger, M., Siegel, M., Engel, A.K., 2007b. Towards single-trial analysis in cognitive brain research. Trends Cogn Sci. 11, 502-503.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D.Y., Engel, A.K., 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. Journal of Neuroscience 25, 11730-11737.

Deichmann, R., Noeth, U., Weiskopf, N., 2010. The Basics of Functional Magnetic Resonance Imaging. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 39-62.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134, 9-21.

Devor, A., Dunn, A.K., Andermann, M.L, Ulbert, I. Boas, D.A., Dale, A.M., 2003. Coupling of total hemoglobin concentration, oxygenation, and neural activity in rat somatosensory cortex. Neuron. 39, 353-359

Di Russo, F., Martinez, A., Sereno, M.I., Pitzalis, S., Hillyard, S.A., 2002. Cortical sources of the early components of the visual evoked potential. Hum.Brain Mapp. 15, 95-111.

Di Russo, F., Pitzalis, S., Aprile, T., Spitoni, G., Patria, F., Stella, A., Spinelli, D., Hillyard, S.A., 2007. Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. Hum.Brain Mapp. 28, 323-334.

Ditterich, J., 2006. Computational approaches to visual decision making. Novartis.Found.Symp. 270, 114-126.

Eichele, T., Calhoun, V.D., Debener, S., 2009. Mining EEG-fMRI using independent component analysis. Int.J Psychophysiol. 73, 53-61.

Eichele, T., Calhoun, V.D., Moosmann, M., Specht, K., Jongsma, M.L., Quiroga, R.Q., Nordby, H., Hugdahl, K., 2008. Unmixing concurrent EEG-fMRI with parallel independent component analysis. Int.J Psychophysiol. 67, 222-234.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M.L.A., Quiroga, R.Q., Nordby, H., Hugdahl, K., 2005. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. Proceedings of the National Academy of Sciences of the United States of America 102, 17798-17803.

Engel, S.A., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb.Cortex 7, 181-192.

Esposito, F., Mulert, C., Goebel, R., 2009. Combined distributed source and single-trial EEG-fMRI modeling: Application to effortful decision making processes. Neuroimage.

Feige, B., Scheffler, K., Esposito, F., Di, S.F., Hennig, J., Seifritz, E., 2005. Cortical and subcortical correlates of electroencephalographic alpha rhythm modulation. J Neurophysiol. 93, 2864-2872.

Frei, E., Gamma, A., Pascual-Marqui, R., Lehmann, D., Hell, D., Vollenweider, F.X., 2001. Localization of MDMA-induced brain activity in healthy volunteers using low resolution brain electromagnetic tomography (LORETA). Hum.Brain Mapp. 14, 152-165.

Frison, L., Pocock, S.J., 1992. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat.Med. 11, 1685-1704.

Friston, K.J., 2007. Statistical Parametric Mapping. Academic Press, London.

Friston, K.J., Frith, C.D., Passingham, R.E., Dolan, R.J., Liddle, P.F., Frackowiak, R.S.J., 1992. Entropy and Cortical Activity: Information Theory and PET Findings. Cereb. Cortex. 2, 259-267

Friston, K.J., Jezzard, P., Turner, R., 1994. Analysis of Functional MRI Time-Series. 1 ed, pp. 153-171.

Friston, K.J., 2010. The free-energy principle:a unified brain theory? Nat.Rev.Neurosci. 11, 127-138.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273-1302.

Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., Henson, R.M, 2006. A critique of functional localizers. Neuroimage. 30 (2006) 1077-1087

K.J. Friston, K.J, Rotshtein, P., Geng, J.J., Sterzer,P., Henson, R.M, 2006. A critique of functional localizers

Fuchs, M., Wagner, M., Kastner, J., 2001. Boundary element method volume conductor models for EEG source reconstruction. Clin.Neurophysiol. 112, 1400-1407.

Fuchs, M., Wagner, M., Wischmann, H.A., Kohler, T., Theissen, A., Drenckhahn, R., Buchner, H., 1998. Improving source reconstructions by combining bioelectric and biomagnetic data. Electroencephalogr.Clin.Neurophysiol. 107, 93-111.

Fuhrmann, A.G., Hein, G., Tsai, N., Naumer, M.J., Knight, R.T., 2008. Temporal characteristics of audiovisual information processing. J Neurosci. 28, 5344-5349.

Fuhrmann, A.G., Sun, F.T., Handwerker, D., D'Esposito, M., Knight, R.T., 2007. Spatio-temporal information analysis of event-related BOLD responses. Neuroimage 34, 1545-1561.

Gazzaniga MS, 2004. The Cognitive Neurosciences III. MIT Press.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 15, 870-878.

Gloor, P., 1969. The work of Hans Berger. Electroencephalography and clinical neurophysiology 27, 649.

Gloor, P., 1985. Neuronal generators and the problem of localization in electroencephalography: application of volume conductor theoery to electroencephalography. J.Clin.Neurophysiol. 2, 327-354.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu.Rev.Neurosci. 30, 535-574.

Goldman, R.I., Stern, J.M., Engel, J., Jr., Cohen, M.S., 2002. Simultaneous EEG and fMRI of the alpha rhythm. Neuroreport 13, 2487-2492.

Goncalves, S.I., de Munck, J.C., Pouwels, P.J., Schoonhoven, R., Kuijer, J.P., Maurits, N.M., Hoogduin, J.M., Van Someren, E.J., Heethaar, R.M., Lopes da Silva, F.H., 2006. Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: inter-subject variability. Neuroimage 30, 203-213.

Goodman, N.R., 1963. The distribution of the determinant of a complex Wishart distributed matrix. Annals of Mathematical Statistics 34, 178-180.

Grave de Peralta, M.R., Gonzalez Andino, S.L., Morand, S., Michel, C.M., Landis, T., 2000. Imaging the electrical activity of the brain: ELECTRA. Hum.Brain Mapp. 9, 1-12.

Grave de Peralta, M.R., Murray, M.M., Michel, C.M., Martuzzi, R., Gonzalez Andino, S.L., 2004. Electrical neuroimaging based on biophysical constraints. Neuroimage. 21, 527-539.

Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., Vanrumste, B., 2008. Review on solving the inverse problem in EEG source analysis. J.Neuroeng.Rehabil. 5, 25.

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De, C.W., Vergult, A., D'Asseler, Y., Camilleri, K.P., Fabri, S.G., Van, H.S., Lemahieu, I., 2007. Review on solving the forward problem in EEG source analysis. J.Neuroeng.Rehabil. 4, 46.

Hasnain, M.K., Fox, P.T., Woldorff, M.G., 1998. Intersubject variability of functional areas in the human visual cortex. Human Brain Mapping 6, 301-315.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2002. Human neural systems for face recognition and social communication. Biol.Psychiatry 51, 59-67.

Hays, W.L., 1994. Statistics. Wadsworth Pub Co.

Heekeren, H.R., Marrett, S., Bandettini, P.A., Ungerleider, L.G., 2004. A general mechanism for perceptual decision-making in the human brain. Nature 431, 859-862.

Heekeren, H.R., Marrett, S., Ruff, D.A., Bandettini, P.A., Ungerleider, L.G., 2006. Involvement of human left dorsolateral prefrontal cortex in perceptual decision

making is independent of response modality. Proc.Natl.Acad.Sci.U.S.A 103, 10023-10028.

Heekeren, H.R., Marrett, S., Ungerleider, L.G., 2008. The neural systems that mediate human perceptual decision making. Nat.Rev.Neurosci. 9, 467-479.

Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Hum.Brain Mapp.

Herrmann, C.S., Debener, S., 2008. Simultaneous recording of EEG and BOLD responses: A historical perspective. International Journal of Psychophysiology 67, 161-168.

Hille, B., 2001. Ion Channels of Excitable Membranes, 3 ed. Sinauer Associates Inc., New York.

Hillebrand, A., Barnes, G.R., 2005. Beamformer analysis of MEG data. Int.Rev.Neurobiol. 68, 149-171.

Hoffmeyer, H.W., Enager, P., Thomsen, K.J., Lauritzen M.J., 2007. Nonlinear neurovascular coupling in rat sensory cortex by activation of transcallosal fibers. J Cereb Blood Flow Metab. 27:575–587.

Honey, C.J., Sporns, O., 2008. Dynamical consequences of lesions in cortical networks. Hum.Brain Mapp. 29, 802-809.

Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top-down attentional control. Nat.Neurosci. 3, 284-291.

Huber, M., Bailey, T., Durrant-Whyte, H., Hanebeck, U., 2010. On Entropy Approximation for Gaussian Mixture Random Vectors.

Huettel, S.A., Song, A.W., McCarthy, G., 2004. Functional Magnetic Resonance Imaging, 1 ed. Sinauer Associates, Inc., Sunderland.

Hyvaerinen, A., 1999. Fast and robust fixed point algorithms for independent component analysis. IEEE Trans.Neural.Netw. 10, 626-634.

Hyvaerinen, A., Karhunen, J., Oja, E., 2001. Independent component analysis. John Wiley and Sons, New York.

Iannetti, G.D., Mouraux, A., 2010. Combining EEG and fMRI in Pain Research. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 365-377.

Ives, J.R., Warach, S., Schmitt, F., Edelman, R.R., Schomer, D.L., 1993. Monitoring the patient's EEG during echo planar MRI. Electroencephalogr.Clin.Neurophysiol. 87, 417-420.

Jasper, H., 1958. The ten-twenty system of the Internationl Federation. Electroencephalogr. Clin.Neurophysiol. 10, 371-375.

Jordan, M.I., 1999. Learning in Graphical Models, Jordan, M.I. ed. MIT Press.

Jones, M., Berwick, J., Mayhew, J., 2002. Changes in blood flow, oxygenation, and Volume following extended stimulation of rodent barrel cortex. Neuroimage. 15, 474-487

Kandel E.R., S.J.H.J.T.M., 2004. Principles of Neural Science, 4 ed. McGraw-Hill, New York.

Karch, S., Mulert, C., 2010. Cognition. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 419-450.

Kastner, S., Pinsk, M.A., De, W.P., Desimone, R., Ungerleider, L.G., 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22, 751-761.

Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. Neuroimage. 30, 1273-1284.

Kilner, J.M., Mattout, J., Henson, R., Friston, K.J., 2005. Hemodynamic correlates of EEG: A heuristic. Neuroimage 28, 280-286.

Koechlin, E., Summerfield C., 2007. An information thoeretical approach to prefrontal exectutive funtion. Trends in Cog. Sci. 11(6), 229-235

Koechlin, E., Ody, C., Kounheir, F., 2003. The architecture of cognitive control in the human prefrontal cortex. Science. 302, 1181-1184

Kolmogorov, A.N., 1974. Foundations of the Theory of Probability. English transl. of 1933 German ed., 2 ed. Chelsea Publishing CO., New York.

Kraus J.D., Fleisch D.A., 1999. Electromagnetics with Applications, 5 ed. McGraw-Hill Book Co - Singapore, Singapore.

Kreiman, G., 2007. Single unit approaches to human vision and memory. Curr.Opin.Neurobiol. 17, 471-475.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat.Neurosci. 12, 535-540.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., ., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc.Natl.Acad.Sci.U.S.A 89, 5675-5679.

Laufs, H., Holt, J.L., Elfont, R., Krams, M., Paul, J.S., Krakow, K., Kleinschmidt, A., 2006. Where the BOLD signal goes when alpha EEG leaves. Neuroimage 31, 1408-1418.

Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., Krakow, K., 2003. EEG-correlated fMRI of human alpha activity. Neuroimage 19, 1463-1476.

Lemieux, L., Mulert, C., 2010. Preface. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, p. vii-ix.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150-157.

Logothetis, N.K., Wandell, B.A., 2004. Interpreting the BOLD signal. Annu.Rev.Physiol 66, 735-769.

Lu, Y., Bagshaw, A.P., Grova, C., Kobayashi, E., Dubeau, F., Gotman, J., 2006. Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. Neuroimage 32, 238-247.

Luck, S., Hillyard, S., 2000. The Operation of Selective Attention at Multiple Stages of Processing: Evidence from Human and Monkey Electrophysiology. In: Gazzaniga MS (Ed.), The New Cognitive Neurosciences. The MIT Press, Cambridge, pp. 687-700.

Magri, C., Whittingstall, K., Singh, V., Logothetis, N.K., Panzeri, S., 2009. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. BMC.Neurosci. 10, 81.

Makeig, S., Debener, S., Onton, J., Delorme, A., 2004. Mining event-related brain dynamics. Trends Cogn Sci. 8, 204-210.

Mandelkow, H., Halder, P., Boesiger, P., Brandeis, D., 2006. Synchronization facilitates removal of MRI artefacts from concurrent EEG recordings and increases usable bandwidth. Neuroimage 32, 1120-1126.

Mangun, G.R., Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J.Exp.Psychol.Hum.Percept.Perform. 17, 1057-1074.

Mansfield, P., 1977. Multiplanar image formation using NMR spin echoes. J Phys C Solid State Phys 10, 55-58.

Mantini, D., Perrucci, M.G., Cugini, S., Ferretti, A., Romani, G.L., Del, G.C., 2007. Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. Neuroimage 34, 598-607.

Martuzzi, R., Murray, M.M., Meuli, R.A., Thiran, J.P., Maeder, P.P., Michel, C.M., Grave de Peralta, M.R., Gonzalez Andino, S.L., 2009. Methods for determining frequency- and region-dependent relationships between estimated LFPs and BOLD responses in humans. J.Neurophysiol. 101, 491-502.

Mathiesen, C., Caesar, K., Akgoren, N., Lauritzen, M., 1998. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. J Physiol 512 (Pt 2), 555-566.

Mayhew, S.D., Dirckx, S.G., Niazy, R.K., Iannetti, G.D., Wise, R.G., 2010. EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. Neuroimage 49, 849-864.

Mazzoni, A., Panzeri, S., Logothetis, N.K., Brunel, N., 2008 Encoding of Naturalistic Stimuli by Local Field Potential Spectra in Networks of Excitatory and Inhibitory Neurons. PLoS Comput Biol 4(12) e1000239

McLachlan, G., Peel, D., 2000. Finite Mixture Models, 1 ed. John Wiley & Sons, Inc., New York.

Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de, P.R., 2004. EEG source imaging. Clin.Neurophysiol. 115, 2195-2222.

Misra, N., Singh, H., Demchuk, E., 2005. Estimation of the entropy of a multivariate normal distribution. Journal of Multivariate Analysis 92, 324-342.

Mobascher, A., Brinkmeyer, J., Warbrick, T., Musso, F., Wittsack, H.J., Saleh, A., Schnitzler, A., Winterer, G., 2009. Laser-evoked potential P2 single-trial amplitudes covary with the fMRI BOLD response in the medial pain system and interconnected subcortical structures. Neuroimage 45, 917-926.

Moeschlin O.et al., 2003. Wahrscheinlichkeitstheorie I (Mass- und Wahrscheinlichkeitstheorie). FernUni Hagen, Hagen.

Montemurro, M.A., Senatore, R., Panzeri, S., 2007. Tight data-robust bounds to mutual information combining shuffling and model selection techniques. Neural Comput. 19, 2913-2957.

Mulert, C., Lemieux, L., 2010. EEG-fMRI, 1 ed. Springer, New York.

Mulert, C., Seifert, C., Leicht, G., Kirsch, V., Ertl, M., Karch, S., Moosmann, M., Lutz, J., Moller, H.J., Hegerl, U., Pogarell, O., Jager, L., 2008. Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. Neuroimage 42, 158-168.

Mullinger, K., Debener, S., Coxon, R., Bowtell, R., 2008. Effects of simultaneous EEG recording on MRI data quality at 1.5, 3 and 7 tesla. Int.J Psychophysiol. 67, 178-188.

Mumford, J.A., Nichols, T., 2009. Simple group fMRI modeling and inference. Neuroimage. 47, 1469-1475.

Murakami, S., Okada, Y., 2006. Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J.Physiol 575, 925-936.

Nabney, I.T., 2002. Netlab: Algorithms for Pattern Recognition. Springer, London.

Nakamura, W., Anami, K., Mori, T., Saitoh, O., Cichocki, A., Amari, S., 2006. Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis. IEEE Trans.Biomed.Eng 53, 1294-1308.

Nevado, A., Young, M.P., Panzeri, S., 2004. Functional imaging and neural information coding. Neuroimage 21, 1083-1095.

Niazy, R.K., Beckmann, C.F., Iannetti, G.D., Brady, J.M., Smith, S.M., 2005. Removal of FMRI environment artifacts from EEG data using optimal basis sets. Neuroimage 28, 720-737.

Nierhaus, T., Schon, T., Becker, R., Ritter, P., Villringer, A., 2009. Background and evoked activity and their interaction in the human brain. Magn Reson.Imaging 27, 1140-1150.

Niessing, J., Ebisch, B., Schmidt, K.E., Niessing, M., Singer, W., Galuske, R.A.W., 2005. Hemodynamic Signals Correlate Tightly with Synchronized Gamma Oscillations. Science. 309 948-951

Noppeney, U., Ostwald, D., Werner, S., 2010. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. J.Neurosci. 30, 7434-7446.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10, 424-430.

Nunez, P.L., Silberstein, R.B., 2000. On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? Brain Topogr. 13, 79-96.

Nunez, P.L., Srinivasan R., 2006. Electric Fields of the Brain: The Neurophysics of EEG, 2 ed. Oxford University Press, New York.

Odom, J.V., Bach, M., Barber, C., Brigell, M., Marmor, M.F., Tormene, A.P., Holder, G.E., Vaegan, 2004. Visual evoked potentials standard (2004). Doc.Ophthalmol. 108, 115-123.

Ogawa, S., Lee, T.M., Nayak, A.S., Glynn, P., 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson.Med. 14, 68-78.

Optican, L.M., Richmond, B.J., 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J.Neurophysiol. 57, 162-178.

Ostwald, D., Porcaro, C., Bagshaw, A.P., 2010. An information theoretic approach to EEG-fMRI integration of visually evoked responses. Neuroimage. 49, 498-516.

Oyman, O., Naber, R.U., Bolcskei, H., Paulraj, A.J., 2003. Characterizing the statistical properties of mutual information in MIMO channels. IEEE Transactions on Signal Processing 51, 2784-2795.

Panzeri, S., Magri, C., Logothetis, N.K., 2008. On the use of information theory for the analysis of the relationship between neural and imaging signals. Magn Reson.Imaging 26, 1015-1025.

Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., Diamond, M.E., 2001. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29, 769-777.

Panzeri, S., Schultz, S.R., Treves, A., Rolls, E.T., 1999. Correlations and the encoding of information in the nervous system. Proc.Biol.Sci. 266, 1001-1012.

Panzeri, S., Senatore, R., Montemurro, M.A., Petersen, R.S., 2007. Correcting for the sampling bias problem in spike train information measures. J.Neurophysiol. 98, 1064-1072.

Panzeri, S., Treves, A., 1996. Analytical estimates of limited sampling biases in different information measures. Network-Computation in Neural Systems 7, 87-107.

Pascual-Marqui, R.D., 2002a. Notes on some physics, mathematics, and statistics for time and frequency domain LORETA.

Pascual-Marqui, R.D., 2002b. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find.Exp.Clin.Pharmacol. 24 Suppl D, 5-12.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision 10, 437-442.

Pessoa, L., Padmala, S., 2007a. Decoding near-threshold perception of fear from distributed single-trial brain activation. Cereb.Cortex 17, 691-701.

Pessoa, L., Padmala, S., 2007b. Decoding near-threshold perception of fear from distributed single-trial brain activation. Cereb.Cortex 17, 691-701.

Philiastides, M.G., Ratcliff, R., Sajda, P., 2006. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J.Neurosci. 26, 8965-8975.

Philiastides, M.G., Sajda, P., 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb.Cortex 16, 509-518.

Philiastides, M.G., Sajda, P., 2007. EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. J.Neurosci. 27, 13082-13091.

Pola, G., Thiele, A., Hoffmann, K.P., Panzeri, S., 2003. An exact method to quantify the information transmitted by different mechanisms of correlational coding. Network. 14, 35-60.

Porcaro, C., Barbati, G., Zappasodi, F., Rossini, P.M., Tecchio, F., 2008. Hand sensory-motor cortical network assessed by functional source separation. Hum.Brain Mapp. 29, 70-81.

Porcaro, C., Coppola, G., Di, L.G., Zappasodi, F., Siracusano, A., Pierelli, F., Rossini, P.M., Tecchio, F., Seri, S., 2009. Hand somatosensory subcortical and cortical sources assessed by functional source separation: an EEG study. Hum.Brain Mapp. 30, 660-674.

Porcaro, C., Ostwald, D., Bagshaw, A.P., 2010. Functional source separation improves the quality of single trial visual evoked potentials recorded during concurrent EEG-fMRI. Neuroimage 50, 112-123.

Quian, Q.R., Panzeri, S., 2009. Extracting information from neuronal populations: information theory and decoding approaches. Nat.Rev.Neurosci. 10, 173-185.

Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20, 873-922.

Ratcliff, R., Philiastides, M.G., Sajda, P., 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc.Natl.Acad.Sci.U.S.A.

Reich, D.S., Mechler, F., Victor, J.D., 2001. Formal and attribute-specific information in primary visual cortex. J.Neurophysiol. 85, 305-318.

Ritter, P., Becker, R., Freyer, F., Villringer, A., 2010. EEG Quality: The Image Acquisition Artefact. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 153-172.

Rolls, E.T., Grabenhorst, F., Parris, B.A., 2010. Neural systems underlying decisions about affective odors. J Cogn Neurosci. 22, 1069-1082.

Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. 12 ed, pp. 191-200.

Rosa, M., Kilner, J., Blankenburg, F., Josephs, O., Penny, W., 2010. Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. Neuroimage. 49, 1496-1509

Rossion, B., Jacques, C., 2008. Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. Neuroimage. 39, 1959-1979.

Roweis, S., Ghahramani, Z., 1999. A unifying review of linear gaussian models. Neural Comput. 11, 305-345.

Salustri C., Barbati G., Porcaro C., 2005. Fetal Magnetocardiographic Signals Extracted by 'Signal Subspace' Blind Source Separation. IEEE Trans.Biomed.Eng 52, 1140-1142.

Saxe, R., Brett M., Kanwisher N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage. 30(4):1088-96

Scheffe, H., 1959. The Analysis of Variance, 1 ed. John Wiley & Sons, Inc., New York.

Schneidman, E., Bialek, W., Berry, M.J., 2003. Synergy, redundancy, and independence in population codes. Journal of Neuroscience 23, 11539-11553.

Seber, G.E.F., 1977. Linear Regression Analysis, 1 ed. John Wiley & Sons, New York.

Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J.Neurophysiol. 86, 1916-1936.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379-423.

Shawkat, F.S., Kriss, A., 2000. A study of the effects of contrast change on pattern VEPS, and the transition between onset, reversal and offset modes of stimulation. Doc.Ophthalmol. 101, 73-89.

Sheth, S.A., Nemoto, M., Guiou, M., Walker, M., Pouratian, N., Toga, A.W., 2004. Linear and nonlinear relationships between neuronal activity, oxygen metabolism, and hemodynamic responses. Neuron, 42, 347-355

Shmuel, A., 2010. Locally Measured Neuronal Correlates of Functional MRI Signals. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 63-82.

Shmuel, A., Augath, M., Oeltermann, A., Logothetis, N.K., 2006. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nat.Neurosci. 9, 569-577.

Shulman, G.L., d'Avossa, G., Tansy, A.P., Corbetta, M., 2002. Two attentional processes in the parietal lobe. Cereb.Cortex 12, 1124-1131.

Siniatchkin, M., Dubeau, F., 2010. EEG-fMRI in Children with Epilepsy. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 349-362.

Singh, K.D., Barnes, G.R., Hillebrand,A., Forde, E.M.E., Williams, A.L., 2002. Task-Related Changes in Cortical Synchronization Are Spatially Coincident with the Hemodynamic response. Neuroimage. 16, 103-114

Skorokohd, A.V., 2005. Basic Principles and Applications of Probability Theory. Springer, Berlin.

Stephan, K.E., Harrison, L.M., Penny, W.D., Friston, K.J., 2004. Biophysical models of fMRI responses. Curr.Opin.Neurobiol. 14, 629-635.

Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E.M, Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modelling. Neuroimage. 49, 3099-3109

Stevens, T.K., Ives, J.R., Klassen, L.M., Bartha, R., 2007. MR compatibility of EEG scalp electrodes at 4 tesla. J Magn Reson.Imaging 25, 872-877.

Tecchio, F., Porcaro, C., Barbati, G., Zappasodi, F., 2007. Functional source separation and hand cortical representation for a brain-computer interface feature extraction. J.Physiol 580, 703-721.

Thielscher A, Pessoa L., 2007. Neural correlates of percpetual choice an decisio making during fear-disgust discrimination. J Neurosc. 14, 27(11):2908-17.

Troje, N.F., Bulthoff, H.H., 1996. Face recognition under varying poses: the role of texture and shape. Vision Res. 36, 1761-1771.

Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Sotero, R.C., Iturria-Medina, Y., eman-Gomez, Y., Bosch-Bayard, J., Carbonell, F., Ozaki, T., 2008. Model driven EEG/fMRI fusion of brain oscillations. Hum.Brain Mapp.

Villringer, A., Mulert, C., Lemieux, L., 2010. Principles of Multimodal Functional Imaging and Data Integration. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 3-18.

Walker, M., Chaudhary, U.J., Lemieux, L., 2010. EEG-fMRI in Adults with Focal Epilepsy. In: Mulert, C., Lemieux, L. (Eds.), EEG-fMRI, 1 ed. Springer, Berlin, pp. 309-332.

Wan, X., Riera, J., Iwata, K., Takahashi, M., Wakabayashi, T., Kawashima, R., 2006. The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism. Neuroimage 32, 616-625.

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56, 366-383.

Warbrick, T., Mobascher, A., Brinkmeyer, J., Musso, F., Richter, N., Stoecker, T., Fink, G.R., Shah, N.J., Winterer, G., 2009. Single-trial P3 amplitude and latency informed event-related fMRI models yield different BOLD response patterns to a target detection task. Neuroimage 47, 1532-1544.

Whittingstall, K., Logothetis, N.K., 2009. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. Neuron 64, 281-289.

Whittingstall, K., Stroink, G., Schmidt, M., 2007. Evaluating the spatial relationship of event-related potential and functional MRI sources in the primary visual cortex. Hum.Brain Mapp. 28, 134-142.

Wohlschlager, A.M., Specht, K., Lie, C., Mohlberg, H., Wohlschlager, A., Bente, K., Pietrzyk, U., Stocker, T., Zilles, K., Amunts, K., Fink, G.R., 2005. Linking retinotopic fMRI mapping and anatomical probability maps of human occipital areas V1 and V2. Neuroimage 26, 73-82.